Improving the use and assessment of state-space models of population dynamics: with application to Oregon Coastal coho salmon

Sam McKechnie

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics, The University of Auckland, 2013

Abstract

Use of state-space modeling for stock assessments of salmon population dynamics has been limited compared to other fish and wildlife species, and this thesis aims to facilitate their application by investigating advantages of this approach, and resolving some methodological limitations. Other extensions such as the inclusion of abundance data from multiple stage-classes that have seldom been considered previously are also developed. A case study is explored where state-space, stage-structured models ('life-history models') are fitted to coho salmon data, and advantages over traditional models are investigated.

Unfortunately, selecting among a set of candidate models can be difficult for state-space models owing to the technical nature of some model selection tools, and choices about which part/s of the model should be focused on. It is shown that currently used measures of data-level deviance could be improved by use of partially-marginalized deviance measures that allow differences in process equations between candidate models to be detected.

Whether the selected best model/s are adequate for inference is seldom assessed, or if it is, methodology is usually vague, and the power of detecting inadequacies has not been reliably assessed to date. It is shown that full assessment of model adequacy requires a partially marginalized extension of the more widely utilized posterior predictive checks (which are shown to have low power). Furthermore, the relative properties of test variables constructed based on different combinations of data and/or parameters/latent states are investigated.

Finally, a significant limitation of state-space models in practice is the difficulty in ensuring all parameters in the model, and most notably the variance components parameters, are identifiable.

Previous studies have attempted to constrain variance parameters by, for example, specifying the observation error variance as a constant after estimating sampling variation of the monitoring program that produced the abundance data. The latter technique was utilized herein, but before this was possible, methods for estimating the variance of spawner population size had to be developed. Novel estimators were developed to overcome this long-standing weakness of salmon monitoring programs, and a suite of estimators were compared in an extensive simulation study to determine the most robust methods to be applied when fitting state-space models to salmon data.

Acknowledgments

Firstly, thank you to my supervisor Russell for overseeing my research. Your guidance and technical skills were certainly much appreciated and this thesis would have been considerably poorer without you supervision. Thank you also to the statistics department - Nancy, Alex, Aileene and Karen in the office, my many officemates in the various offices I have had over the years and other students and staff in the department. The Oregon Department of Fish and Wildlife and Chris Jordan especially, kindly provided the empirical data that was used in this thesis.

Thank you especially to my wider family for their support including my sisters, grandparents and aunt, though especially my parents who have always been extremely supportive. Finally my partner Sarah and daughter Lila who have copped the brunt of the financial burden of student life and the long hours necessary to complete the thesis, thank you for always being there.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General introduction

Models used for the assessment of salmon stocks have remained relatively unchanged for several decades. Primarily, simple stock-recruitment models are fitted using linear regression, despite numerous problems with this approach (Ludwig and Walters, 1981; Walters, 1985; Hilborn, 2009). Recently, several authors have recommended state-space versions of these stock assessment models (Lessard et al., 2008; Su and Peterman, 2011), a development that has been occurring for some time in the wider field of fisheries science (Maunder, 2003), and also terrestrial ecology (King et al., 2009). If modeling of salmon stocks was to move away from restrictive traditional stock-recruitment approaches, other model variations (in addition to state-space structures) become more feasible. Perhaps the most important is the addition of age- and stage-structure (so called 'life-history models'), which has been heavily promoted recently (Hilborn, 2009; Knudsen and Michael, 2009), but they have seldom been applied in practice. This can be partly attributed to the rarity of datasets containing the abundance of fish in multiple parts of the life-cycle.

A dataset with potential for supporting life-history models, has been collected by the Oregon Department of Fish and Wildlife whereby coho salmon abundance is monitored at both the juvenile and spawner stages (and also at the smolt stage in some populations) for multiple stocks

in the Oregon Coastal Natural (OCN) group of populations. This opportunity to explore more complex models opens up the potential for a number of questions, for example; Can state-space, life-history models be reliably fitted to extremely noisy salmon abundance data? What extra ecological inferences can be made once different parts of the life-cycle can be modeled explicitly? Will predictive ability of models improve over traditional techniques?

Before addressing such questions these models must be fitted and compared within a reliable statistical framework. Prior to the development of spatially extensive, state-space, life-history models to the OCN data, a preliminary case-study was implemented for a long-term high quality dataset collected for a small population in Lobster Creek in the Alsea catchment. In the early development of suitable models a number of outstanding issues became apparent, that make reliable fitting of these types of models challenging. The goals of this thesis were realigned towards investigating and attempting to overcome several methodological issues that were limiting effective modeling of salmon population dynamics.

State-space models in general can be far more difficult to fit than more traditional models (Schnute, 1994). In addition to general statistical considerations such as how to choose among candidate models, and how to determine if a best model/s fits the data adequately, it often has to be determined whether the data can support a the desired model at all. Together, these issues are the basis of this thesis.

Perhaps the most important findings are that common methods for selecting among models, and assessing the adequacy of selected model/s, are deficient. Initially it was found that model selection is compromised when differences between models occur in certain parts of the state-space model structure (Chapter 4) and that a better way to approach selection would be to compare models based on the fit of those parts of the model structure that differ between them. Partial marginalization was used to develop an alternative selection criteria that improves discriminatory ability. The usual posterior predictive checking approach to assessing model adequacy was also found to have low power (Chapter 5). An alternative approach was tested and its derivation was subsequently found to be very similar in principle to the partially marginalized model selection

method.

These tools were then applied to models developed for the Lobster Creek dataset (Chapters 2 and 3), which can be considered a case study for developing similar models for the full, spatially extensive dataset for the wider OCN. However, before any of the desired Lobster Creek candidate models could be reliably fitted, it was necessary to constrain the variance components parameters, as has been established in other situations (Schnute, 1994; King et al., 2008). Unfortunately, the usual methods of obtaining prior information on the observation error variances were elusive for Lobster Creek because of the absence of techniques to estimate sampling variation for the spawner stage. Chapter 6 attempts to rectify this long standing problem, and facilitated effective fitting of the models in the case study presented in Chapters 2 and 3. The developments presented here will provide a more robust base from which modeling of OCN and other stocks can be extended.

In order to address these specific model developments, background information is provided about coho salmon biology, population structure and the modeling of their population dynamics, both past and present.

1.2 Coho salmon biology

Coho salmon (*Onchorhynchus kisutch*) are one of seven Pacific salmon species, many of which are extremely important ecologically, culturally and commercially across their distributional range. They play a critical role in nutrient cycling (Bilby et al., 2001; Naiman et al., 2002), are a dominant prey species of several threatened species (Sigler et al., 2009; Ward et al., 2009), support recreational, commercial and indigenous fisheries (Johnson, 1984; Hilborn, 2006; Martell et al., 2008) and are iconic species for conservation. All species of Pacific salmon share some general life-cycle features including andronomy, where spawning occurs in freshwater with immature fish migrating to the ocean to grow before returning to freshwater to spawn, and semelparity, where all individuals die after spawning. Each species also tends to exhibit high productivity and low survival rates (Bradford, 1995).

Coho are distributed over large areas of Russia, parts of Asia and the west coast of North America, from California north. Ideal spawning habitat is relatively small, low-gradient streams with suitable spawning gravels. The different Pacific salmon species exhibit varying plasticity of life-cycle between individuals, with coho following a relatively strict cycle across most of their distributional range (Groot and Margolis, 1991). Adult coho return to freshwater, typically their natal stream, to spawn in autumn to early winter ($\sim$ November–February). Eggs and alevins reside in the spawning gravels overwinter, and fry emerge ($\sim$ March–May) from the gravel and reside in-stream for around a year before migrating to the ocean as smolts ($\sim$ March–July).

Understanding of smolt movements after ocean entry is relatively poor, however despite coastal areas are their preferred habitat, it is known that some individuals may migrate hundreds, or even thousands of kilometers from their natal stream (Weitkamp and Neely, 2002). Evidence suggests that there is some geographical structuring of the migration with some groups of populations favoring distinct coastal areas over others (Weitkamp and Neely, 2002). Once maturity is reached after about 1.5 years in the ocean, adults return to freshwater to spawn, after which they die. In a variation to this cycle a small proportion of males mature early after 0.5 years in the ocean and return to spawn at 2 years of age. These precocious males, or 'jacks', are considerably smaller than 3-yr-old males and attempt to 'sneak' matings without the usual fights for dominance undertaken by the older males (Koseki and Fleming, 2006). Coho exhibit strong homing tendencies with most spawners returning to natal streams, although a small proportion stray to other parts of the catchment, or between catchments. This straying is an important mechanism in mitigating localized extinctions (Waples et al., 2009).

The past two centuries have been characterized by widespread declines of many stocks of all Pacific salmon species with numerous groups of Pacific salmon populations on the west coast of the USA listed under the Endangered Species Act (Good et al., 2005). About 27 percent of coho salmon populations in the Pacific Northwest of the USA are estimated to have become extinct over the last several centuries (Gustafson et al., 2007). This level is higher than observed for several other species and perhaps reflects their long juvenile freshwater residence, where they

are exposed to habitat modification, and the fact that their strict life-cycle can result in annual spawner abundance being dominated by only one year-class. The latter makes coho populations vulnerable to catastrophic events.

1.3 Structuring of coho salmon populations for conservation: Oregon Coastal Natural Evolutionary Significant Unit

The widespread range contractions and declines in coho abundance have prompted considerable management intervention including prioritizing conservation effort amongst groups of coho populations (Pinsky et al., 2009). In response to the introduction of the U.S. Endangered Species Act, populations of each Pacific salmon species were assigned to discrete evolutionarily significant units (ESU). These are defined as groups of populations that are reproductively isolated from other groups, and whose conservation is important for the evolutionary potential of the species. The OCN group of populations is an ESU that has long been the subject of intensive scientific research (Johnson, 1984; Nickelson et al., 1986) and the data used in this thesis come from Lobster Creek which is a tributary of the Alsea River, one of the populations within the OCN ESU (Fig 1.1). This unit has been further classified into gene conservation groups (GCGs), subsets of populations sharing similar genetics (Kostow, 1995). Originally three GCGs were identified (Kostow, 1995), the Mid-North Coast, Umpqua and South Coast GCGs, but subsequent genetic analyses resulted in one addition, the Lakes GCG (Ford et al., 2004). For monitoring purposes the Mid-North Coast GCG is separated into two units, North Coast and Mid-Coast, resulting in a total of five stratum within the ESU. The focal unit for many conservation and monitoring objectives is the population, defined as a set of individuals in a geographical region that are largely reproductively isolated from other such groups (Wainwright et al., 2008). For coho salmon these units typically relate to river catchments. Fifty six populations were identified within the OCN by the Oregon Coast Coho Salmon working group based on biological, geographical and ecological landscape characteristics. Populations are classified as independent (n=21) or dependent

(n=35) based on whether there would be a high or low probability of population persistence over a 100 yr period in the absence of demographic contributions from other populations, respectively.

1.4 Diagnosing threats to coho salmon persistence

While declines in abundance of some coho populations have been profound, diagnosing the exact mechanisms involved has been difficult. This is partly because their complicated life-cycle and exposure to several habitats has resulted in declines being influenced by multiple factors (Bradford and Irvine, 2000). Although identifying factors affecting population fluctuations, or declines, is often possible, actually quantifying the relative roles of more than one agent is generally more difficult, given that experimental manipulation of individual threats is usually prohibitive. Consequently, most inferences are based on observational data that are susceptible to the usual issues involved in correlative techniques (Holland, 1986). Despite these shortcomings it is generally accepted that there are four factors to which most declines in coho populations, and Pacific salmon species in general, can be attributed; modification of freshwater habitat, addition of hatchery-reared fish, harvesting by commercial and recreational fisherman and fluctuations in climatic conditions.

- **Habitat modification.** Extensive modification of river catchments where coho spawn has occurred across their range, especially in the Pacific Northwest (Bilby and Mollot, 2008). Waterways have been channeled, dredged, diverted and have suffered from water abstraction. Riparian zones have been cleared, and extensive parts of some catchments have been utilized for forestry and agriculture, or have suffered from urbanization. Consequently, coho populations in these catchments have been adversely affected. Eggs and alevins are sensitive to heightened sediment loads in spawning streams resulting from modification (Scrivener and Brownlee, 1989). Fry and smolt prefer to occupy pool habitat, beaver ponds and off-channel sloughs, especially for overwintering, all of which have been reduced by waterway management (Nickelson et al., 1992; Nickelson and Lawson, 1998; Bradford and Irvine, 2000), and most of the freshwater stages are presumably susceptible to resulting

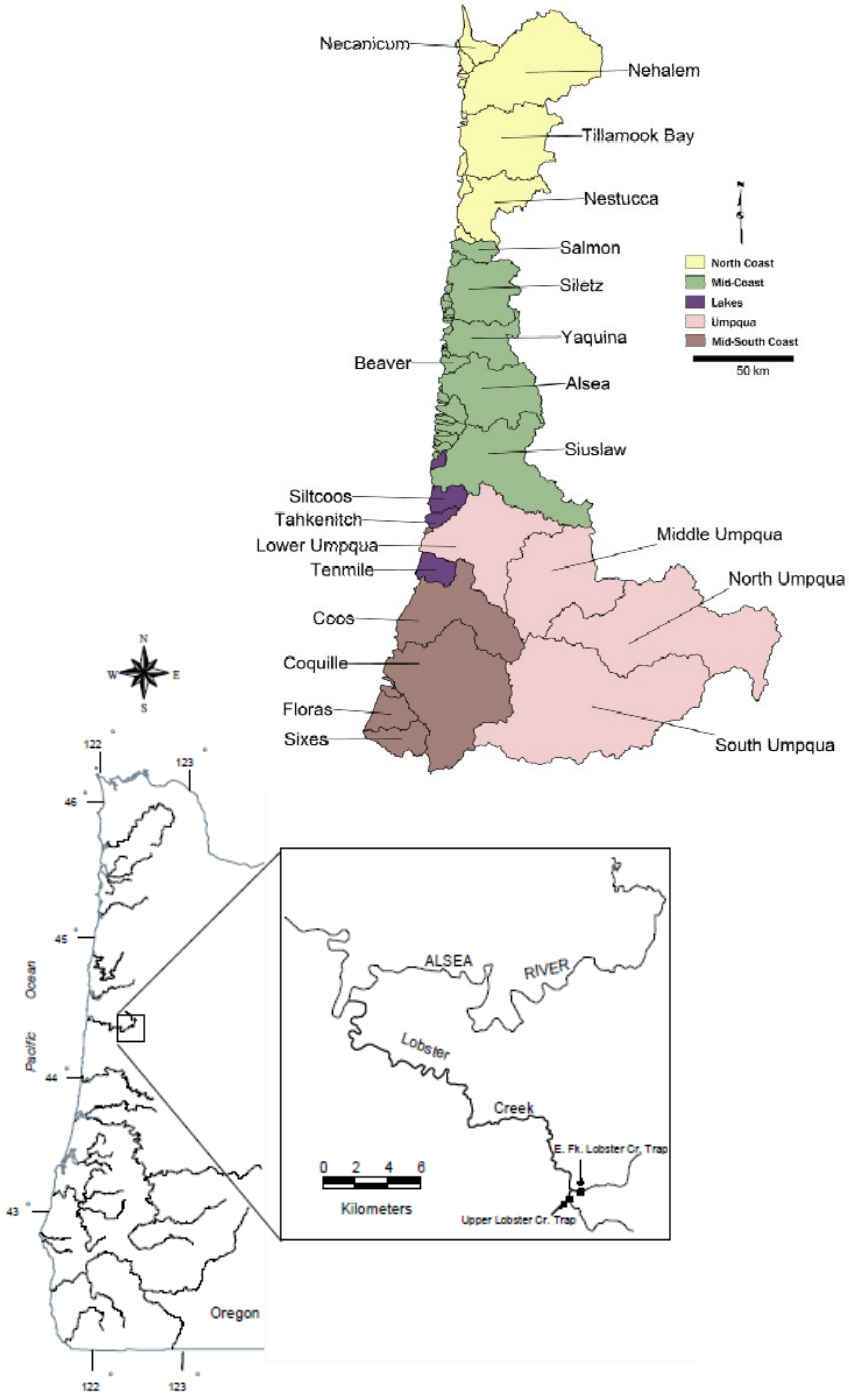

Figure 1.1: Map of the location of Lobster Creek within the Alsea River catchment, OCN ESU.

changes in water temperatures and chemical inputs.

- **Impact of hatchery-reared fish.** Impacts of hatchery-reared fish on the dynamics of wild populations are well documented for a range of salmonid species (Chilcote, 2003; Hindar et al., 2006; Ford and Myers, 2008). Two related threats exist, the presence of salmon farms, and direct releases of hatchery-raised fry and smolts. No salmon farms exist in close proximity to the OCN range although there is a long history of releases of hatchery-reared fish. Releases aim to either enhance fisheries for commercial purposes or to supplement threatened wild stocks. In either case, some hatchery-reared fish subsequently mate with wild fish as a result of straying from hatchery release points in the former. While impacts of hatchery fish on absolute population abundance can be positive through increased adults returning to spawn in streams (Oosterhout et al., 2005), observed impacts on wild fish have generally been negative (Nickelson, 2003; Wainwright et al., 2008; Buhle et al., 2009). Mechanisms involved are often difficult to identify but can include competition for limited resources between released and wild juveniles and smolts, transmission of disease and parasites from hatcheries to the wild, and the reduced fitness often observed in hatchery-reared fish (Fleming and Gross, 1993) which can be propagated to the wild population via genetic introgression. There is growing evidence that hatchery-reared fish have had a significant adverse impact on the abundance of coho salmon in the OCN ESU. This is indicated by a negative relationship between productivity of populations and both the number of smolt released into a population (Nickelson, 2003), and the proportion of spawners in the population that are from hatchery origins (Wainwright et al., 2008), and negative density dependent effects of hatchery-reared spawners at the population level (Buhle et al., 2009).

- **Impact of fisheries.** Coho populations of the Pacific Northwest have been subjected to a long history of exploitation by humans. In Oregon this was initially subsistence harvesting by indigenous peoples but developed into large-scale commercial and recreational fisheries in the 1800's (Johnson, 1984). Commercial activity in the early years focused on in-stream net

fisheries, predominantly on the Columbia River, with ocean-based troll fisheries developing in the early 1900's (Johnson, 1984). An important recreational fishery of Oregon coho also developed, especially after the mid-1900's, and became a significant proportion of total numbers harvested (Johnson, 1984). Recreational fishing is concentrated on offshore and tidewater areas.

It is difficult to estimate the impact that fisheries have had on coho populations due to uncertainties in estimating total catch, the origin of coho harvested and levels of density dependence occurring in the populations. However, there is some evidence that impacts have at times been severe. For instance, exploitation rates of OCN have been estimated to be exceptionally high at times during recent decades (PFMC 2000), at levels suggested to be unsustainable using population models (Nickelson and Lawson, 1998). While commercial and recreational fisheries are now tightly regulated, with impacts of fishing on OCN abundance now limited by restricting allowable exploitation rates below prescribed levels (PFMC 2000), the efficacy of this management strategy relies on robust preseason forecasts of run size and accurate estimates of escapement after fishing occurs.

- **Fluctuating climatic conditions.** Salmon are sensitive to prevailing oceanic conditions during the marine-rearing stage of their life-cycle and numerous studies have documented strong correlations between marine survival and various physical and biological variables in the marine environment, for all Pacific salmon species (Adkison et al., 1996; Pyper et al., 1999; Mueter et al., 2002). Fluctuations in these variables are typically driven by large-scale climatic events (Ware and Thomson 2000). Coho salmon appear to have increased marine survival when sea surface temperatures are low, mixed layer depths are shallow and sea levels are low, among other conditions (Coronado and Hilborn, 1998; Beamish et al., 2000; Hobday and Boehlert, 2001; Logerwell et al., 2003; Quinn et al., 2005). Mechanisms underlying these correlations are difficult to identify but are presumably related to changes in food abundance and/or quality, predator abundance or physiological

costs. Recent fluctuations in climate regimes have resulted in extreme variation (over an order of magnitude) in coho marine survival (Coronado and Hilborn, 1998), making these processes perhaps the most important in determining levels of spawner returns. In contrast, relationships between climate and the success of freshwater stages of the coho life-cycle are poorly studied, although there is some evidence that part of the significant variation in freshwater survival between years may also be influenced by climatic influences on physical variables such as flow levels and stream temperature (Lawson et al., 2004).

1.5 Previous models of coho salmon dynamics

Optimal management of salmon populations requires scientific advice in the form of accurate and precise estimates of spawning stock size, robust preseason (and in some cases within-season) forecasts of spawning stock size, sound knowledge of determinants of population dynamics, and the likely responses of populations to management interventions. Most salmon management therefore relies heavily on constructing and fitting population models to data. Historically these models have been relatively simple, and in some cases are inadequate in providing advice required by management agencies. For example, despite a long history of salmon monitoring, and consequently rich time-series data for some populations, it has proven difficult to separate and quantify the relative roles of different mechanisms involved in declines (Knudsen and Michael, 2009). Consequently, prescribing appropriate management actions has been difficult. Despite inherent problems in formulating management advice, decisions need to be made, and making the most robust inferences, given the available data, becomes a priority.

Numerous studies have fitted stock-recruitment models to OCN coho salmon spawner-abundance data (Chilcote, 1999; Bradford et al., 2000; Wainwright et al., 2008; Buhle et al., 2009), and other coho stocks (e.g. Chen and Holtby, 2002). Typically these modeling exercises have attempted to estimate reference points and set harvest limits (Bradford et al., 2000), assess population status and viability (Wainwright et al., 2008) or assess relationships between population dynamics and external variables such as oceanic conditions (Buhle et al., 2009). The Ricker stock-recruitment

model is the most widely used and is easily fitted to transformed data using simple linear regression techniques. While there have been some variations to the basic Ricker model, for example extending the Ricker model to allow density dependent parameters to be modeled hierarchically across populations (Chen and Holtby, 2002; Barrowman et al., 2003), and various extra terms are added to the model to account for environmental affects on survival (Chilcote, 1999; Wainwright et al., 2008; Buhle et al., 2009), the general approach has changed little over recent decades. Unfortunately there are several shortcomings of the simple Ricker model (and other spawner-recruit functions) which can have serious consequences for inferences and management advice. Two major concerns are that widely used statistical techniques for fitting models are inadequate (Walters, 1985) and current models lack the age-structure to make reliable inferences about dynamics of coho salmon throughout their life-cycle (Hilborn, 2009). Recently developed models provide the potential to overcome these traditional problems and form part of the work in this thesis.

1.6 Improving spawner-recruitment models using a state-space formulation

While linear regression is a simple technique for fitting to data, its use in modeling spawner-recruit relationships is beset with problems. For example, the independent variable (spawner abundance) is assumed to be measured without error, when in reality abundance of spawning salmon is typically estimated with considerable uncertainty. Unless this is accounted for in some manner, parameter estimates will tend to be biased (King et al., 2009; Su and Peterman, 2011), and in severe circumstances spawner-recruit relationships will be obscured (Ludwig and Walters, 1981). Problems also arise from the time-series nature of spawner-recruit data that is not explicitly accounted for using linear regression (Walters, 1985).

While these problems can be tackled with various corrective methods, such as errors-in-variables regression, and Walters' time-series bias correction (Walters, 2009), there is evidence that corrections can be ineffective (Peterman et al., 2003). Furthermore, most traditional meth-

ods acknowledge only observation error in their structure. They therefore ignore the presence of process error, the deviation of population trajectory away from deterministic dynamics (Carlin et al., 1992), which is undoubtedly an important factor in the extreme variation present in observed salmon abundance data (Peterman et al., 2003).

State-space models are a form of latent variable time-series model that are widely used in modeling animal population dynamics. These models have advantages over traditional models by providing a more realistic representation of the processes that generated the data. They consist of three elements, initial conditions, the process model and the observation model. Stochasticity enters the model in both the process model (process error) and observation model (observation error). In animal population dynamics the process model specifies the current unobserved state/s (number of individuals in various age-classes for example) based on previous state/s, and the process error allows for variation in the demographic rates between years. The observation equation represents the sampling process, thus relating observed data to the true unobserved states. So essentially state-space models attempt to partition variance based on the fact that process error is propagated through time whereas observation error is not. This can result in improved inferences for parameters that are known to be biased in traditional models (King et al., 2009; Su and Peterman, 2011) and is also beneficial for making predictions, as only the appropriate components of variation are propagated into the future (Clark, 2003).

Early examples of state-space models in the fisheries literature include Pella (1993) and Schnute (1994), who relied on the Kalman Filter for model fitting. More recently a range of state-space fisheries models have been fitted within a Bayesian framework (e.g. Meyer and Millar, 1999; Millar and Meyer, 2000a; Rivot et al., 2004) which provides considerably more flexibility in model specification than the Kalman Filter. These types of models are now a routine tool in marine fisheries stock assessment throughout the world (Hilborn, 2003), although their use in stock assessments of salmon is far less frequent (but see Rivot and Parent, 2001; Rivot et al., 2004; Lessard et al., 2008 and others for examples of state-space salmon models).

Part of the relatively slow shift towards state-space modeling for salmon populations has been

the ease with which traditional models can be fitted. Salmon life-cycles are generally far more strict than other fish species in that often most individuals progress through different stages of the life-cycle at the same rate. Simple stock-recruitment models that are clearly inappropriate for other species may still be reasonable for salmon. Furthermore, these simple models can be fitted using linear regression, which is simple to carry out, is familiar to most fisheries scientists and is available in all common statistical packages. State-space models on the other hand typically requires writing bespoke code which can be intimidating to some, even though the advantages of such an approach are clear. However, once modeling moves away from traditional techniques not only will these potential advantages be realized, but other variations in structure become feasible. For instance, age- or stage-structured models can be explored, especially if abundance data from multiple parts of the life-cycle is available.

1.7 Life-history models of salmon population dynamics

The quality and length of time-series data varies considerably between stocks of Pacific salmon species. A unifying theme however is a strong reliance on spawner-abundance data and relatively simple stock-recruitment models by research and management agencies, which effectively ignore detail within the life-cycle (Hilborn, 2009). While this in-part reflects the fact that most salmon data is restricted to spawner abundance, it severely restricts the ability of analysts to partition different sources of mortality, or estimate the timing of any density dependence. There have been recent attempts to construct full life-history models incorporating functional relationships between the various salmon life-history stages and also with external variables such as habitat and climate (Scheuerell et al., 2006). In general these have relied on estimating individual parameter values using separate data sources, or in some cases expert opinion, and combining them into a full simulation model. This is performed in a rather ad hoc manner rather than embedding inference and parameter estimation in a statistical framework. While this is a valuable exercise in understanding the study system, and is sometimes the only option when faced with a paucity of data, significant advantages can be gained from formal model fitting. Previous attempts at

fitting life-history models for salmon populations appear to be restricted to a small number of studies including Rivot et al. (2004) and Lessard et al. (2008). However data from multiple life-history stages certainly exists for a number of of other populations and species (Bradford, 1995; Barrowman et al., 2003).

Recently significant efforts to increase the frequency of adopting life-history models in salmon research has been encouraged (Hilborn, 2009; Knudsen and Michael, 2009), but until these models become more widely used there remain uncertainties in their potential utility and the level of model complexity that can be supported, given the inherent variability of salmon data. In the case of OCN coho salmon, formulation of integrated models will be governed by the diverse range of data already available for analyses. The potential for improving inferences is vast however, including separation of mortality into stage-specific components and linking these vital rates to covariates, improving precision of forecasts of escapement and determining optimal monitoring programs to achieve certain objectives, as just some examples. Now that the data is available, the toolbox of techniques has become relatively well developed and the motivation for investigating these modeling options has been established, the potential of fitting state-space life-history models to OCN coho salmon data is ready to be assessed.

1.8 Structure of state-space models

1.8.1 Model structure

At its most basic, a state-space model of animal abundance describes the structure of two linked processes. These are the unobserved dynamics of the system, which have a time-series component (the state equation, or process model), and the sampling of how these unobserved states are linked to observed data via the sampling processes (the observation equation, or observation model). An example of a Bayesian state-space model of this system with additive errors (which will be

heavily used in this thesis) can be defined by three equations

$$n_t = f(n_{t-1}, \boldsymbol{\psi}) + \epsilon_t \ , \quad \text{for} \quad t = 1, 2, ..., T \tag{1.1}$$

$$y_t = g(n_t, \boldsymbol{\xi}) + \nu_t \ , \qquad \text{for} \quad t = 1, 2, ..., T \tag{1.2}$$

$$n_0 \sim h \tag{1.3}$$

where n_t is the unobserved (latent) abundance of the animal, y_t is the observed abundance, both in year t, and the $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ are vectors of model parameters. The $\boldsymbol{n}$ have at times been referred to as parameters, states or auxiliary variables (King, 2012), though throughout this thesis they will be denoted latent states. The functions f and g represent the deterministic components of the process and observation models, and ϵ_t and ν_t are the process and observation errors, respectively. Often (and for demonstration purposes herein) ϵ_t and ν_t are assumed to be normally distributed with mean zero and variances τ^2 and σ^2, which are denoted the process and observation error variances, respectively. In (1.1) and (1.2), errors are assumed to be additive though in practice could take a multitude of forms. The initial state, n_0,i- must receive a prior distribution, h, under a Bayesian framework.

State-space implementation of a traditional spawner-recruit model for salmon could be represented by (1.1)–(1.3) with the function $f(n_{t-1}, \boldsymbol{\psi})$ taken to be a Ricker function (for example), where n_{t-1} would become the number of spawners in the brood year that produced the number of recruits n_t. For example, in coho salmon n_t would be a (Ricker) function of the number of spawners present three seasons before i.e. $n_t \sim f(n_{t-3}, \boldsymbol{\psi}) + \epsilon_t$.

Similarly, (1.1)–(1.3) can easily be extended to allow the stage-structure necessary for a sequential life-history model in which case the equations would be

$$n_{i,t} = f_i(n_{i-,t-1}, \boldsymbol{\psi}_i) + \epsilon_{i,t} \ , \quad \text{for} \quad i = 1, ..., I \ , \ t = 1, 2, ..., T \tag{1.4}$$

$$y_{i,t} = g_i(n_{i,t}, \boldsymbol{\xi}_i) + \nu_{i,t} \ , \qquad \text{for} \quad i = 1, ..., I \ , \ t = 1, 2, ..., T \tag{1.5}$$

$$n_{i,0} = h_i \tag{1.6}$$

where subscript $i = 1, ..., I$ now indexes stage, i.e. $y_{i,t}$ is the observed abundance of fish in stage-

class i in year t. Note that $i-$ denotes the preceding stage-class instead of the more natural $i-1$, as $n_{1,t}$ is a function of $n_{3,t-1}$, not $n_{0,t-1}$ as would be suggested by the latter. Importantly, this structure allows the modeling of stage-specific processes $f_i(n_{i,t-1}, \boldsymbol{\psi}_i)$, which in the cases explored for coho salmon in Chapter 2 represent the spawning by adults and subsequent survival of eggs to the juvenile stage f_1, the survival of juveniles to the smolt stage f_2, and finally the survival of smolts to the spawner stage when they return to breed f_3. The structure in (1.4)–(1.6) is only possible due to semelparity of salmon and a strict life-cycle which allows cohorts to move through stages in totality. Violation of these assumptions requires more complicated models such as those needed for some sockeye (Lessard et al., 2008) and chinook salmon (Scheuerell et al., 2006) populations, where it is necessary to account for variation in maturation between individuals for example.

1.8.2 Posterior distributions

If the state-space model is fitted in a Bayesian framework, Bayes theorem needs to be implemented, which for a very simple model with a vector of parameters $\boldsymbol{\theta}$ the posterior distribution of these parameters is given by

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})} \\ &\propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta}) \end{aligned} \tag{1.7}$$

where $p(\boldsymbol{\theta})$ is the prior distribution of the parameters, $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function and $p(\boldsymbol{y})$ is the normalization constant. Bayes theorem can be applied to the state-space model with structure defined in (1.1)–(1.3) for example, so that the posterior distribution is given by

$$\begin{aligned} p(\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2|\boldsymbol{y}) &\propto p(\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2)p(\boldsymbol{y}|\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) \\ &= p(n_0, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2)\prod_{t=1}^{T} p(n_t|n_{t-1}, \boldsymbol{\psi}, \tau^2)p(y_t|n_t, \boldsymbol{\xi}, \sigma^2) \end{aligned}$$

where the prior distribution of the parameters $p(n_0, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2)$ includes a prior on the initial latent state in year 0, and $p(y_t|n_t, \boldsymbol{\xi}, \sigma^2)$ is the likelihood for the observation process. Note that

the $p(n_t|n_{t-1}, \boldsymbol{\psi}, \tau^2)$ can be regarded as the prior specification of n_t, and owing to the conditional independence of the current state from other states (and parameters $\boldsymbol{\xi}$ and τ^2), given the previous states, this is given by

$$p(n_t|n_0, ..., n_{t-1}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) = p(n_t|n_{t-1}, \boldsymbol{\psi}, \tau^2) \ . \tag{1.8}$$

Two general inferences that are often made from the joint posterior distribution involve calculation of marginal distributions for parameters and/or latent states. Parameters such as survival of individuals from one year to the next, or maximum population growth rate in density-dependent models are often of interest for management decisions, while the latent states themselves can also be the focus of a study, for instance for robust estimation of population trends of an endangered species. Obtaining the posterior distribution of the parameters or latent states requires marginalization, for example the joint posterior of the parameters can be derived by marginalizing over the latent states

$$p(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2|\boldsymbol{y}) = \int p(\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2|\boldsymbol{y}) \mathrm{d}\boldsymbol{n} \tag{1.9}$$

and to obtain the marginal posterior distribution of individual parameters within $\boldsymbol{\theta} = (\boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2)$, further marginalization is required.

1.9 Approximating posterior distributions using MCMC

Carrying out the high dimensional integrations necessary for determining the normalization constant in (1.7) and when marginalizing over individual parameters or states (e.g. 1.9), is challenging for complex state-space models. For a very restrictive class of models that includes some linear state-space models with normally distributed observation and process errors obtaining posterior distributions may be possible using analytical techniques (Calder et al., 2003). However, this is certainly not the case for most of the models considered in this thesis. Therefore, more general numerical estimation techniques that are especially useful in situations where non-linear and non-normal error structures are encountered are needed. These techniques are described

herein, as these are the techniques used throughout the thesis to approximate the posterior distributions of parameters and latent states.

Numerical techniques for fitting Bayesian state-space models include methods such as sequential particle filters that have seen some use (Newman et al., 2006, 2008) and the more widely used Markov chain Monte Carlo (MCMC) algorithms Gilks et al. (1996). The popularity of the latter can be attributed to their generality for a wide range of conditions, though also because they are the basis for user-friendly software packages such as the various forms of BUGS (Bayesian inference Using Gibbs Sampling; Lunn et al., 2009), and ADMB (Automatic Differentiation Model Builder; Fournier et al., 2012), among others. These packages circumvent the cumbersome derivation of the full conditional distributions required when bespoke code is written in lower-level languages. The BUGS language is used throughout this thesis, and although it can be very inefficient compared to lower-level code, the ease and speed with which models can be expanded and modified often compensate for this drawback.

As suggested by its name, BUGS relies heavily on the Gibbs sampler (Geman and Geman, 1984) to sample from the approximate posterior distribution of the parameters and latent states. The general outline of this approach is provided here. The Gibbs sampler relies on being able to specify the conditional distribution of each individual parameter or state, given all other parameters, states and data. By way of example, the conditional distributions of latent states in the model (1.1)–(1.3), given the conditional independence highlighted in (1.8), are

$$
\begin{aligned}
&p(n_0|n_1, \boldsymbol{\psi}, \tau^2) \propto p(n_0)p(n_1|n_0, \boldsymbol{\psi}, \tau^2) \\
&p(n_t|y_t, n_{t-1}, n_{t+1}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) \\
&\qquad \propto p(y_t|n_t, \boldsymbol{\xi}, \sigma^2)p(n_t|n_{t-1}, \boldsymbol{\psi}, \tau^2)p(n_{t+1}|n_t, \boldsymbol{\psi}, \tau^2) \quad \text{for } t = 1, ..., T-1 \\
&p(n_T|y_T, n_{T-1}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) \propto p(y_T|n_T, \boldsymbol{\xi}, \sigma^2)p(n_T|n_{T-1}, \boldsymbol{\psi}, \tau^2) \ .
\end{aligned}
$$

Similarly, conditional distributions can also be constructed for parameters σ^2 and τ^2 and each component of parameter vectors $\boldsymbol{\xi}$, and $\boldsymbol{\psi}$.

If $\boldsymbol{\theta}$ is a vector of all u unknown quantities (parameters and states), $\boldsymbol{\theta} = (\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2)$,

the joint distribution of the components is $p(\theta_1, ..., \theta_u)$ and the conditional distribution of the ith component is denoted $p(\theta_i|\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_u, \boldsymbol{y})$, then the Gibbs sampler is given by the following steps. Specify initial values for each component, $(\theta_1^0, ..., \theta_u^0)$ then generate new values for each i component of the model in turn, for each $j = 1, ..., v$ iteration of the sampler as

$$\begin{array}{llll}
\theta_1^1 \sim p(\theta_1|\theta_2^0, ..., \theta_u^0, \boldsymbol{y}) & \ldots & \theta_i^1 \sim p(\theta_i|\theta_1^1, ..., \theta_{i-1}^1, \theta_{i+1}^0, ..., \theta_u^0, \boldsymbol{y}) & \ldots \\
\vdots & & & \\
\theta_1^j \sim p(\theta_1|\theta_2^{j-1}, ..., \theta_u^{j-1}, \boldsymbol{y}) & \ldots & \theta_i^j \sim p(\theta_i|\theta_1^j, ..., \theta_{i-1}^j, \theta_{i+1}^{j-1}, ..., \theta_u^{j-1}, \boldsymbol{y}) & \ldots \\
\vdots & & & \\
\theta_1^v \sim p(\theta_1|\theta_2^{v-1}, ..., \theta_u^{v-1}, \boldsymbol{y}) & \ldots & \theta_i^v \sim p(\theta_i|\theta_1^v, ..., \theta_{i-1}^v, \theta_{i+1}^{v-1}, ..., \theta_u^{v-1}, \boldsymbol{y}) & \ldots
\end{array}$$

thus, new values are generated from their conditional distribution where the values for all other components are those generated either in the preceding or current iteration. Specific techniques for sampling from the conditional distributions $p(\theta_i|\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_u, \boldsymbol{y})$ are dependent on their form. This can vary from direct sampling in conjugate examples with closed-form distributions, to adaptive rejection sampling, slice sampling, or application of the Metropolis-Hastings algorithm when the full conditional is unavailable in closed form (Lunn et al., 2009). The resulting samples are a Markov chain that approximate the joint distribution $p(\theta_1, ..., \theta_u)$ (Gelman et al., 2003). Thus, once convergence has occurred, and samples early in the chain that may have been influenced by the starting values have been discarded, the remaining samples can be used for posterior inferences, and calculating summary statistics such as means and credible intervals, correlations among components, and derived functions of parameters.

1.10 Development of state-space salmon models and the structure of this thesis

State-space models have become widely used in both fisheries science (Millar and Meyer, 2000b; Maunder, 2003; Peterman et al., 2003) and terrestrial ecology (de Valpine and Hastings, 2002; Calder et al., 2003; Newman et al., 2006) although their uptake has been more limited when

fitting models of salmon population dynamics (though see Rivot and Parent, 2001; Peterman et al., 2003; Rivot et al., 2004; Su and Peterman, 2011). Despite their widespread use, and the general acknowledgment that they have the potential to improve inferences over observation- or process-error models (Calder et al., 2003; King et al., 2009; Su and Peterman, 2011), there are a number of areas where research could be focused to improve their use in modeling animal population dynamics in general, and more specifically the case of fitting state-space models life-history models to datasets of salmon abundance. This thesis aims to address several of these outstanding problems.

1.10.1 Chapters 2 and 3: The potential of Bayesian life-history models of salmon dynamics

Chapters 2 and 3 are a case study for fitting state-space life-history models to coho salmon abundance data in a pair of streams in coastal Oregon. It is the motivating example for the methodological improvements developed in the other chapters. The main aims of this chapter were to investigate whether these models could be effectively fitted to the temporally noisy datasets typical of salmon populations in practice. The value of adding additional abundance data (at the juvenile and smolt stages) was determined with respect to the accuracy and precision of the predictive ability of models, and the breadth of questions that could be answered using dynamical models. Methods of model comparison evaluated in Chapter 4, model adequacy evaluated in Chapter 5 and estimators of spawner abundance (and its variance) assessed in Chapter 6 were utilized in an effort to make inferences in this chapter as robust as possible.

1.10.2 Chapter 4: Multilevel fit of Bayesian state-space population dynamics models

Much research on state-space animal population dynamics models has focused on the significant challenge of how to reliably fit models to data. This focus appears to be at the expense of other aspects of statistical inference such as selecting among models and assessing goodness of fit. In many studies it appears that once a model is successfully fitted inferences are often made without consideration about whether it is a 'good' model, or better than other models that could

be considered. When active selection among models does occur, the properties of the selection tool used is not always well known. As state-space models become more popular and are fitted by a growing number of analysts with differing degrees of statistical nous there has been a tendency in recent years towards the use of easily carried-out selection methods such as the Deviance Information Criteria (DIC) that are a default output of BUGS programs. Chapter 4 investigates whether this technique is suitable in these situations, focusing especially on models that differ in the structure of the process model such as (1.1) in the simple example presented above.

1.10.3 Chapter 5: Assessing the adequacy of state-space models of population dynamics

In the same manner that model selection has received less research focus than aspects of model fitting, so too has the process of assessing whether a selected model/s are adequate for the specified objectives of the research. The most frequently used techniques of evaluating the adequacy of Bayesian models are posterior predictive checks (Rubin, 1984), although their implementation for state-space population dynamics models has previously been vague (or more commonly no assessment is carried out), and the suitability of the many choices available for carrying out the checks for these models remain untested. Chapter 5 attempts to address some of these issues by formulating checks of different parts of state-space models. This includes investigating not only posterior predictive distributions, but also partially marginalized predictive distributions that are part-way between prior and posterior predictive distributions, and which appear to have merit for testing inadequacies in components of the process model of state-space models in particular. Furthermore, different test variables are investigated based on different combinations of data and/or states/parameters in a bid to achieve the main aim of the chapter which is to assess whether predictive checks can reliably inform the analyst about the adequacy or otherwise of state-space models.

1.10.4 Chapter 6: Comparison of estimators of salmon escapement for periodic count data

The focus of Chapter 6 is more narrow than in Chapters 4 and 5. The strength of state-space models is their ability to partition variation into its rightful components, however this comes at a cost in that identifiability of all parameters can be difficult to achieve for many models and datasets (Schnute, 1994; Bolker, 2008; Snover, 2008). This is especially the case for variance components parameters (e.g. σ^2 and τ^2 for (1.1)–(1.3)), and often restrictions are placed on these parameters in the form of informative priors (Meyer and Millar, 1999), specifying the ratio of observation to process error variance (Schnute, 1994), or specifying the observation error variance as a constant after estimating sampling variation of the monitoring program that produced the abundance data (King et al., 2008, 2009). Early versions of the models developed for coho salmon in Chapter 2 indicated difficulty in estimating the variance component parameters, and so the latter approach of estimating sampling variability was attempted. However, the methods of estimating the variance of spawner abundance for periodic count data were largely undeveloped at the time of model-fitting. This chapter investigated a range of estimators (including several developed in collaborative work during the thesis; e.g. Millar et al., 2012) using simulation to decide the most appropriate to use for providing the relevant inputs required for successful model fitting in Chapter 2.

The thesis concludes with a general discussion (chapter 7) of findings in the individual chapters and the links between them. It also puts the work in the context of the wider field of state-space modeling of animal population dynamics. As is often the case with a body of work such as this, a number of questions have been raised by the research presented herein and so some fertile avenues for future research are outlined.

Chapter 2

The potential of Bayesian life-history models of coho salmon dynamics in Lobster Creek, Oregon: Methodological development

2.1 Introduction

Modeling of fish population dynamics is often conducted with sparse data, which can only support the simplest of models. However, there are many situations where the assumptions of simple models will be violated (Walters and Ludwig, 1981; Walters, 1985), sometimes with severe consequences for management decisions (Walters and Ludwig, 1981; Su and Peterman, 2011). Sparsity is particularly a problem for semelparous (reproduces once before dying), andronomous (migrate into the ocean for part of the life-cycle before returning to freshwater to spawn) species as data are generally restricted to a single, annual count of recruiting, breeding-age individuals as they return to spawn. The spawner stage of the life-cycle is where estimation of population size is most feasible and where individuals are most easily observed by the public, making this stage the focus of conservation actions. Furthermore, recreational and commercial fisheries target these mature individuals, and so this is the stage where harvest impacts are assessed (Quinn and Deriso, 1999). Large parts of the life-cycle are therefore effectively unobserved, which limits the breadth of questions that can be answered from these data. For instance it is impossible to partition variation in survival between freshwater and marine phases, which reduces the power

of environmental impact assessments in spawning streams, as just one example.

There are a number of monitoring programs that have estimated abundance of salmon over a range of life stages (e.g. Bradford et al., 1997; Barrowman et al., 2003; Lessard et al., 2008). In some cases they have been included in stage-structured population models (Lessard et al., 2008), but often analyses have been ad-hoc or focused on stages in isolation (Bradford et al., 2000). Even if data from additional stages are available there has been a tendency to ignore it and continue to fit traditional-type models.

There are several factors that may have led to the continued reliance on traditional models. Presumably, much can be attributed to the ease with which traditional models can be fitted. For example, the Ricker model is undoubtedly the most popular model fitted to stock-recruitment data for Pacific salmon, and can be routinely fitted using simple linear regression techniques after log-transformation of the data (Quinn and Deriso, 1999). Being part of all common statistical packages, these routines are accessible to biologists with even minimal statistical training. In contrast, more sophisticated techniques generally require coding specific to the individual analyses. Furthermore, traditional stock-recruitment models assume vital rates are common to all individuals in a brood year which, although very limiting for species with plastic life-cycles (Lessard et al., 2008), is less of an issue for species such as coho and pink salmon that exhibit strict life-cycles. There is also a lack of evidence that more complex stage-structured models lead to better estimates of abundance, and improved management decisions, and whether they can be reliably fitted to available data. Thus, until more examples of life-history models occur in the literature, it is difficult to assess how limiting traditional approaches are.

Recently emphasis has been placed on exploring the efficacy of 'life-history models' of andronomous fish populations (Hilborn, 2009; Knudsen and Michael, 2009), which partition individuals into discrete age- or stage-classes, with separate vital rates (survival, movement etc.) linking the classes. Moving away from the restrictive assumptions of traditional stock-recruitment analysis also allows the possibility of improving other aspects of model structure. For example, multiple sources of stochasticity, including process and observation error, can be accounted for if

models are fitted within a state-space modeling framework (Carlin et al., 1992; Peterman et al., 2003; Su and Peterman, 2011). Partitioning error into its rightful components will not only allow the underlying dynamics of the system to be estimated more appropriately, but predictions of future dynamics and decision analyses will also be more robust, given that only appropriate sources of error will be propagated forward through time (Clark, 2003). Furthermore, the state-space framework simultaneously overcomes several obstacles encountered in traditional stock-recruitment models, including the errors-in-variables (Walters and Ludwig, 1981), and time-series bias problems (Walters, 1985).

Often a Bayesian approach will be adopted to make these models tractable, which simultaneously permits the incorporation of prior information about certain model parameters. This is closely linked to the ongoing development of integrated models in the wider field of fisheries science (Maunder, 2003), which combine disparate datasets into coherent models, potentially enhancing their inferential power. The framework for fitting state-space models is extremely flexible, allowing numerous extensions to basic population models, such as incorporating ecological hypotheses directly into the model in a structural manner. An example would be directly modeling survival between stages as a function of habitat covariates. This is especially pertinent given the large volume of research dedicated to identifying the ecological determinants of variation in survival of salmon during their freshwater residency (Solazzi et al., 2000; Lawson et al., 2004; Johnson et al., 2005; Petrosky and Schaller, 2010).

One such example is Solazzi et al. (2000), who investigated the impact of changes in the amount of rearing habitat available to coho salmon, on abundance and survival in Lobster Creek. They tested their hypotheses on the raw observed abundance data, and survival rates derived directly from them. This approach suffers similar problems to those explained above, not least that observed abundance is assumed to be measured without error. That study allows more sophisticated techniques to be directly compared to their simple approach, with the strengths and weaknesses of each then able to be assessed. Solazzi et al. (2000) utilized a dataset of coho salmon abundance from Lobster Creek, Oregon, which is long-term, of high quality, and consists

of counts at three stages of the life-cycle. It therefore lends itself well to a case study of developing alternatives to simple traditional model-building techniques.

The goals of this chapter are manifold. Firstly, structured models that partition the coho salmon population into important life-stages for coho salmon are constructed in Section 2.2.2.1. These are then fitted to the Lobster Creek dataset within a Bayesian state-space modeling framework to partition stochasticity into its appropriate components. Stage-specific vital rates, including the timing and magnitude of compensatory parameters, are investigated, and inferences are compared to previous studies that fitted traditional models to only spawner abundance data. Extra variables are added to basic models in Section 3.2.5 to assess whether habitat effects can be detected during the freshwater components of the life-cycle, especially in the presence of substantial environmental stochasticity. The value of including juvenile and/or smolt data in model-fitting is then assessed by constructing a suite of models in Section 2.2.4 that differ in the number of stages modeled, and comparing the predictive performance of each (Section 3.2.6). This case study will provide a reference for fitting similar models to datasets that are currently being collected which have a substantially wider spatial extent.

2.2 Methods

2.2.1 Datasets

2.2.1.1 Study site

Lobster Creek is a tributary of the Alsea River on the mid-coast of Oregon (Fig 1.1; Section 1.3). Annual rainfall in the Alsea Basin is typically 150–250 cm, and summer water temperature is between 11–17^{o}C. In its upper reaches Lobster Creek divides into the East Fork (EF) and Upper Mainstem (MF). Downstream smolt traps are located in the lower reaches of both forks and the study reaches extend from the smolt traps to the upper extent of coho salmon occupation (approximate reach lengths; EF=3.5 km, MF=4.7 km).

Datasets analyzed herein are taken from ODFW survey reports (Johnson et al., 2005; Lorion, 2011), which also provide details of motivation for sampling and methodology. The details

relevant to this study are provided below.

2.2.1.2 Habitat manipulation experiment

In the summer of 1991 the rearing habitat in MF was manipulated in an attempt to increase survival of the freshwater stages of the coho life-cycle (Solazzi et al., 2000). Logs were placed in the stream channel to create dam pools and additional woody debris were placed in pools to increase habitat complexity. Off-channel alcove pools were also excavated. In February 1996 flooding led to extensive scouring of MF and most of the added rearing habitat was washed out of the study reach. Similar flood damage occurred in EF in the winter of 1999, and in response creation of rearing habitat was undertaken in the summer of 1999 using debris jams although this modification was minor compared to that undertaken in MF and aimed to repair damage rather than enhance it above previous levels.

2.2.1.3 Habitat surveys

Surveys of the extent of rearing habitat were carried out in Lobster Creek over part of the study period, using the methodology of Hankin and Reeves (1988). Summer habitat was assessed during late August–early September over 1988-2002 and in 2006, while winter surveys were assessed during December–January in 1990-91, 1991-92, 1995-96, and 2004-05, in both streams. An additional winter survey was conducted in MF in 1993-94. Rearing habitat was estimated as the sum of the areas of alcove, dam and beaver pools (Solazzi et al., 2000). Summer habitat surveys were used as an index of rearing habitat as this was when monitoring was most frequent (Fig 2.1). The dynamics of winter rearing habitat appeared to be similar to summer habitat, although data was less frequent then.

Rearing habitat would ideally be quantified by a continuous covariate that could be linked to salmon vital rates within the models, however the rearing habitat index contains numerous missing values, especially at the beginning and end of the time-series, which are difficult to accommodate in state-space models. Furthermore, the combined area of disparate rearing habitats is a crude measurement of habitat resources available to affecting coho salmon, and there will

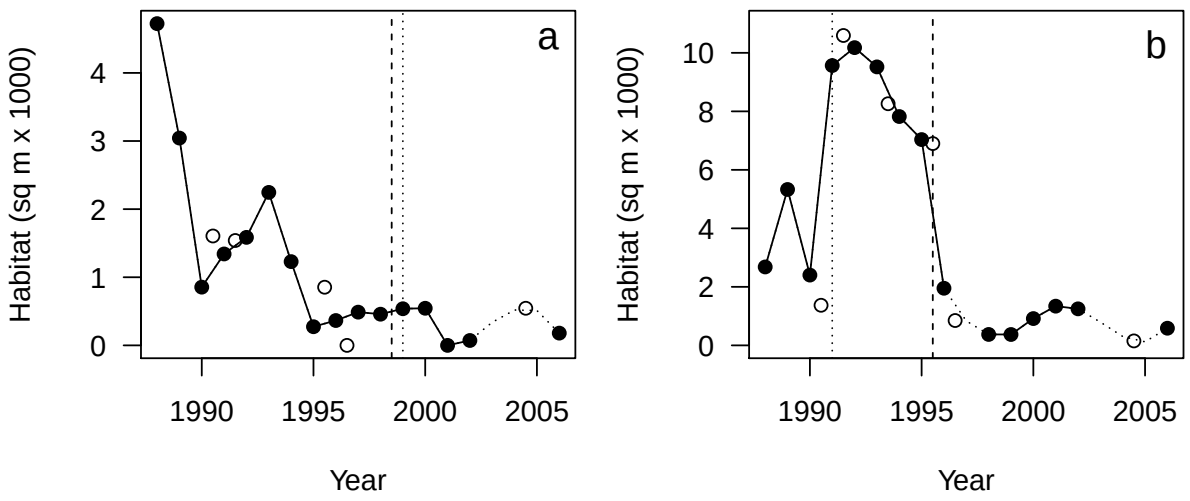

Figure 2.1: Area of rearing habitat in EF (a) and MF (b), Lobster Creek, from 1988–2006. Rearing habitat is the sum of dam, beaver and alcove pools. Solid and open dots represent summer and winter habitat surveys respectively, and the dotted lines represent the interpolation of data for missing years achieved using cubic splines fitted to all data. Vertical dotted and dashed lines indicate the timing of habitat augmentation and flooding events, respectively.

almost certainly be a significant amount of sampling error in its estimation. An alternative approach was utilized whereby the index of rearing habitat (Fig 2.1) was simplified to a binary variable indicating the presence (one), or absence (zero), of additional rearing habitat. Additional habitat was considered to be present immediately after the habitat addition experiment in 1991 in MF until it was destroyed in 1996, thus the index was zero for all years in EF and all years other than 1991–1995 in MF. This index variable was denoted $I_{j,t}$ for stream $j = 1, 2$, in year $t = 1, 2, ..., T$.

2.2.1.4 Juvenile abundance

The size of summer rearing populations of juveniles was estimated during August and September using a combination of techniques (Lorion, 2011). From 1988–2002 the number of juveniles was counted by snorkeling in every third pool. In a subset of these pools electro fishing was undertaken to estimate the proportion of fish missed during snorkeling, and this 'correction factor' was used to correct the latter. In non-pool habitat (riffles, glides, rapids) electro fishing was conducted using the removal method (Seber and Le Cren, 1967) with two or more passes. An overall population estimate for each reach is provided in Lorion (2011), and was made by multiplying

the mean pool count by the number of pools and by multiplying mean densities in each non-pool habitat by their respective areas, then adding these two metrics together.

From 2003–2011 only pool counts were conducted and total population estimates were made by scaling the pool estimates by the ratio of pool counts to total population size in the years 1988–2002 (Lorion, 2011). Estimated abundance of juveniles and its variance in stream j in year t were denoted $y_{1,j,t}$ and $\varsigma^2_{1,j,t}$, respectively, where the 1 in the first subscript denotes stage-class 1 (juveniles).

2.2.1.5 Smolt abundance

From 1988–2011 the number of smolt migrating downstream during spring was estimated using modified incline plane traps which were placed at the lower limit of the study reaches in both streams. Traps were operated from the first week in March until smolt ceased to be captured, usually around early June (Lorion, 2011). Traps were checked daily over this period. Trap efficiency was estimated by marking a sample of individuals, releasing them upstream of the trap, and then recording the proportion caught in the trap over the rest of the trapping season. Raw counts were corrected to account for trap efficiency and the corrected counts were denoted $y_{2,j,t}$, for stream j in year t, where the 2 in the first subscript denotes stage-class 2 (smolt). Variance estimates given in Lorion (2011) were obtained using the bootstrap routine presented by (Thedinga et al., 1994), and were denoted $\varsigma^2_{2,j,t}$.

2.2.1.6 Spawner abundance

The number of spawners returning to both streams was estimated by conducting periodic visual counts (the interval between counts was generally $\sim$ 7–10 d) and analyzing the data using trapezoidal area-under-the-curve (AUC) methodology (English et al., 1992; Irvine et al., 1992). Counts of spawners were made by an observer/s walking the length of habitable stream in the study reach. Firstly, 'fish-days', $F_{j,t}$, in stream j, in year t, was estimated using the trapezoidal AUC estimator (Irvine et al., 1992; $\widehat{F}_T$ in Section 6.2.2.2). To simplify notation, the fish-days

for a single stream in a single year is given by

$$\begin{aligned}\widehat{F} &= 0.5\sum_{r=1}^{R-1}(d_{r+1}-d_r)(c_r+c_{r+1}) \\ &= 0.5\left[(d_2-d_1)c_1+(d_R-d_{R-1})c_R+\sum_{r=2}^{R-1}(d_{r+1}-d_{r-1})c_r\right]\end{aligned}$$

where d_r and c_r are the day of spawning season and count of spawners for sampling occasion $r = 1, ..., R$. This calculation represents the sum of trapeziums under a curve connecting individual counts in a periodic count dataset (see Fig 6.1). When non-zero counts were encountered on the first or last sampling occasion (7 out of 50 samples) zeros were added l (stream-life; the number of days between a fish arriving and dying) days prior to, or after, the non-zero count (as described in Section 6.4.6) which is equivalent to the standard practice (Bue et al., 1998; Parsons and Skalski, 2009, Chapter 6).

An estimate of population size of spawners in stream j, in year t, denoted $y_{3,j,t}$ (equivalent to $\widehat{E}_T$ in Section 6.2.2.2) where the 3 in the first subscript denotes stage-class 3 (spawners), can then be derived by correcting for stream-life l and observer efficiency v (the probability that a spawner is observed during sampling) as

$$\widehat{y}_{3,j,t} = \widehat{F}_{j,t} \times z$$

where $z = 1/(l \times v)$. Estimates of l ($\mu = 11.670, \sigma = 0.847$; Zhou, 2000) and v ($\mu = 0.755, \sigma = 0.035$; Solazzi, 1984; Zhou, 2000) were assumed, and have been used previously for correcting estimates of fish-days from Lobster Creek (Johnson et al., 2005; Lorion, 2011). Variance of the fish-days estimator, $\widehat{\text{Var}}(\widehat{F}_{j,t})$ was estimated using the TAUC variance estimator ($\widehat{\text{Var}}(\widehat{F}_T)$; Section 6.2.2.2). Variance of escapement, was estimated by taking into account the uncertainty in the estimates of l and v, using sequential application of the delta method (Millar and Jordan In Press) by first calculating

$$\widehat{\text{Var}}(k) = \frac{l^2\widehat{\text{Var}}(v) + v^2\widehat{\text{Var}}(l)}{z^2}$$

and then the second application gives

$$\tau^2_{3,j,t} = \widehat{F}^2_{j,t}\widehat{\text{Var}}(z) + z^2\widehat{\text{Var}}(\widehat{F}_{j,t})$$

where $\varsigma^2_{3,j,t}$ denotes the variance estimate for stream j, in year t, for stage-class 3 (spawners).

2.2.2 Construction of models

2.2.2.1 Coho salmon model structure

State space models of coho salmon population dynamics were developed from the basic components introduced in Section 1.8. Compared to many age- and stage-structured models used to represent other animal species (Caswell, 2001), coho salmon models are relatively simple owing to their strict 3-year life-cycle, which allows all individuals in a cohort to transition through the modeled life-stages in totality. Furthermore, their mortality soon after spawning prevents the need to recycle individuals back into the population following the spawner stage, and that stage is the only one contributing to the next generation through reproduction. As a consequence the entire population can be modeled by the transitions between the three stages of the life-cycle for which estimates of abundance are available for Lobster Creek. For simplicity models are presented on the log-scale, and thus a general representation of the process model is

$$\log n_{i,j,t} = \log \tilde{n}_{i,j,t-1} + \epsilon_{i,j,t} \quad , \quad \text{where} \quad \log \tilde{n}_{i,j,t} = f_i(\log n_{i-,j,t-1}, \boldsymbol{\psi}_i) \tag{2.1}$$

where f_i are stage-specific functions (with the stage-class-specific vectors of parameters, $\boldsymbol{\psi}_i$) linking abundance in each stage-class to abundance in the preceding stage-class (denoted $i-$ as specified in Section 1.8), in the preceding year, and $\log n_{i,j,t}$ and $\epsilon_{i,j,t}$ are the abundances of fish (latent states) and process errors, respectively, both for stage-class i, in stream j in year t. Thus $\log n_{1,j,t}$, $\log n_{2,j,t}$ and $\log n_{3,j,t}$ denote the log-abundance of juveniles, smolt and spawners, respectively. Note that the time between sampling the three stages was not exactly one year, for instance the time periods between spawner-fry and fry-smolt are less than a year, and between smolt-spawner well over a year. Thus t indexes the year with which the counts are associated, rather than an exact census date as assumed in other types of models (e.g. Caswell, 2001). The deterministic components of the process model, $\log \tilde{n}_{i,j,t}$ in (2.1), are important quantities for model comparisons carried out in Section 2.2.6.

The f_i are functions that represent the deterministic processes affecting the transition of fish

between the stages from year $t-1$ to year t. For example, the production of eggs by spawners that survive to become juveniles (f_1) or the survival from juvenile to smolt stages (f_2). These functions can take a variety of forms depending on the assumed dynamics of the system. Often the process errors at each stage, $\epsilon_{1,j,t}$, $\epsilon_{2,j,t}$ and $\epsilon_{3,j,t}$ would be considered independent over streams and over years within streams. In these cases the errors could be modeled in each stream separately as normally distributed with mean zero and stream-specific process error variance parameters (see model M^*_{BLL} in appendix Section 8.1 for an example of this structure). It will be shown in Section 3.2.2 that this assumption conflicts with the data. Consequently, process errors are assumed to have a temporal component common to both streams, and thus were modeled as normally distributed on the log-scale such that

$$\epsilon_{i,j,t} = \omega_{i,t} + \varpi_{i,j,t} \ , \quad \omega_{i,t} \sim N(0, \tau_i^2) \ , \quad \varpi_{i,j,t} \sim N(0, \delta_i^2) \tag{2.2}$$

where $\omega_{i,t}$ are the temporal, and $\varpi_{i,j,t}$ are the spatial components of process error, each with a mean of zero and stage-class-specific variances τ_i^2, and δ_i^2, respectively.

The observed abundances of fish, $y_{i,j,t}$, were linked to the latent states via observation equations which represent the sampling processes. These are absolute estimates of abundance rather than the relative indices often used in fisheries, thus by again modeling on the log-scale the general observation equation is given by

$$\log y_{i,j,t} = \log n_{i,j,t} + \nu_{i,j,t} \tag{2.3}$$

where $\nu_{i,j,t}$ are the observation errors, which are assumed to be normally distributed on the log-scale, i.e. $\nu_{i,j,t} \sim \mathrm{N}(0, \sigma^2_{i,j,t})$. Variance parameters $\sigma^2_{i,j,t}$, are the stage-class-, stream- and year-specific sampling variances for $\log y_{i,j,t}$ from the monitoring programs at the juvenile (Section 2.2.1.4), smolt (Section 2.2.1.5) and spawner stages (Section 2.2.1.6). These were calculated from $\varsigma^2_{i,j,t}$ such that the back-transformed distribution would maintain the same CV as estimated in those sections. Note that these enter the model as constants rather than parameters to be estimated, owing to the difficulties that can sometimes be encountered with partitioning observation and process errors in state space models (Schnute, 1994; King et al., 2008, 2009; see

Chapter 6 for further details). To complete the Bayesian state-space model formulation, prior distributions are specified for all parameters including the initial latent states, details of which are given in Section 2.2.5.

2.2.2.2 Initial process model variations

There are numerous options for parameterizing the process model, including both linear and non-linear dynamics. The most simple model considered, assumes abundance to be simply a linear function of the abundance of the preceding stage, in the preceding year. The three process equations of this model are

$$\log n_{1,j,t} = \log \Phi^J + \log n_{3,j,t-1} + \epsilon_{1,j,t} \tag{2.4}$$

$$\log n_{2,j,t} = \log \Phi^P + \log n_{1,j,t-1} + \epsilon_{2,j,t} \tag{2.5}$$

$$\log n_{3,j,t} = \log \Phi^S + \log n_{2,j,t-1} + \epsilon_{3,j,t} \tag{2.6}$$

where Φ^J is the rate of production of juveniles by spawners in the preceding year, and Φ^P and Φ^S are survival rates to the smolt and spawner stages, from the preceding stage in the preceding year, respectively. Note that Φ^J could be a wide range of non-negative numbers (note that this 'survival' parameter can, and will generally be, larger than 1 because multiple juveniles are produced by each individual spawner) and so is modeled on the log scale, while Φ^P and Φ^S are constrained between zero and one. Consequently, the latter are modeled on the logit-scale with associated logit-transformed parameters φ^P and φ^S as follows

$$\Phi^P = \text{logit}^{-1}\left(\varphi^P\right)$$

$$\Phi^S = \text{logit}^{-1}\left(\varphi^S + \lambda O_t\right)$$

where λ is the extra survival parameter, and O_t is a covariate calculated as the log of survival of hatchery fish from their release as smolt until their return as spawners. This approach has previously been adopted by Chilcote (1999) and Wainwright et al. (2008) to help explain variation in marine survival, and the covariate is known as the Oregon Production Index (OPI). Because

survival from smolt to spawners is extremely low this relationship results in Φ^S being proportional to $\exp(O_t)$.

Nonlinear dynamics were investigated by utilizing two functions often implemented in fisheries modeling, the Beverton-Holt (Beverton and Holt, 1957) and Ricker (Ricker, 1954) functions (an example that exhibits the differences between them is given in Fig. 3.7). Using the juvenile stage for demonstration purposes a process equation with Beverton-Holt dynamics at this stage is given by

$$\log n_{1,j,t} = \log \alpha^J + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_j^J}\right) + \epsilon_{1,j,t} \tag{2.7}$$

where α^J is the number of juveniles produced per spawner at low spawner abundance, K_j^J is the number of spawners at which the number of juveniles produced is half the asymptotic level (the carrying capacity of juveniles) and this carrying capacity is at $\alpha^J \times K_j^J$. The production parameter, α^J, was assumed to be common to both streams as this parameter is more likely to be related to physiological limits of spawners etc., while the compensation parameter, K_j^J, was assumed to be stream-specific, as the factors limiting juvenile abundance such as the extent of rearing habitat are likely to vary between streams.

Similarly, a model with Ricker dynamics for the juvenile process equation is given by

$$\log n_{1,j,t} = \log \alpha^J + \log n_{3,j,t-1} - \beta_j^J n_{3,j,t-1} + \epsilon_{1,j,t}$$

where α^J is the number of juveniles produced per spawner at low spawner abundance and β_j^J determines the rate (again stream-specific) at which compensation occurs as spawner abundance increases. The number of spawners at which the number of juveniles produced is maximized can thus be given by $1/\beta_j^J$. In contrast to the Beverton-Holt function, overcompensation occurs in the Ricker function, with the number of juveniles decreasing to an asymptote at zero at high spawner abundances. The initial set of models (Table 2.1) was constructed using models with different combinations of linear, Ricker and Beverton-Holt process equations, and these will be presented in Section 2.2.2.3.

2.2.2.3 Model notation and expansion

Throughout, models (M) are denoted by the form of their process equations at the three stages, with subscripts indicating linear (L), Beverton-Holt (B) or Ricker (R) functions. For example, M_{BLL} has a Beverton-Holt function for the juvenile process equation and linear functions for the smolt and spawner equations, which correspond to (2.7), (2.5) and (2.6), respectively. Similarly, M_{LLR} combines the linear process equations (2.4), (2.5) and a Ricker function at the spawner stage (Table 2.1). The addition of habitat covariates to process models is presented in Section 2.2.3, which adds another level (denoted h) to the subscript. Thus, a model denoted M_{B_hLL} would have a habitat covariate added to (2.7) but would otherwise be identical to M_{BLL}, while $M_{B_hL_hL}$ would also have a habitat covariate added to (2.5). Lastly, models in Section 2.2.4 involve excluding data at selected stages (and hence exclude the process equations at those stages also; Table 2.1). Excluded stages are denoted N, thus by way of example, model M_{BNL} is a model fitted to juvenile and spawner data only (smolt data excluded).

Model refinement was undertaken in an iterative manner starting from the most simple model M_{LLL}. The adequacy of the model was assessed using a suite of predictive checks (Chapter 5), and when inadequacy was identified the model structure was modified accordingly. Model fit, as measured by the posterior mean of partial deviance ($\overline{D(\theta)}_m$, and its stage-class-specific components; Section 4.2.5), examination of residuals and direct evaluation of parameter estimates were also utilized when deciding whether a model should be expanded. The partial deviance and partially marginalized predictive checks are novel tools for model building and were developed (in Chapters 4 and 5) to overcome the inadequacies of the standard data level deviance and more widely applied posterior predictive checks. Expansion ended when a model appeared to adequately describe the features of the data considered important (Chapter 5). This process was compared to the more rigid approach of selecting among models using the marginalized deviance information criterion (DIC_m; Section 4.2.5). This involved fitting a full set of candidate models differing only in the structure of their process equations. For any given model in this set,

density dependence was assumed to occur in at most one stage (to prevent overfitting), and this assumption was tested using predictive checks. This resulted in seven models in the set (initial models; Table 2.1).

Finally, models initially fitted to the data did not allow for the correlation in process errors between stream assumed in (2.2). These models with uncorrelated process errors were quickly identified as inadequate for modeling the data and are not presented in detail here. An example of this type of model is given in appendix Section 8.1 and is denoted M^*_{BLL} where the '*' indicates the different process error structure to M_{BLL} presented above.

Table 2.1: Comparison of the deterministic components ($\log \tilde{n}_{i,j,t}$) of the juvenile, smolt and spawner process equations for the models considered for formal model comparison. The three sets of models refer to three stages of model fitting.

Model	Juvenile	Smolt	Spawner
Initial			
M_{LLL}	$\log \Phi^J + \log n_{3,j,t-1}$	$\log \Phi^P + \log n_{1,j,t-1}$	$\log \Phi^S + \log n_{2,j,t-1}$
M_{RLL}	$\log \alpha^J + \log n_{3,j,t-1} - \beta_j^J n_{3,j,t-1}$	$\log \Phi^P + \log n_{1,j,t-1}$	$\log \Phi^S + \log n_{2,j,t-1}$
M_{BLL}	$\log \alpha^P + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_j^P}\right)$	$\log \Phi^P + \log n_{1,j,t-1}$	$\log \Phi^S + \log n_{2,j,t-1}$
M_{LRL}	$\log \Phi^J + \log n_{3,j,t-1}$	$\log \alpha^P + \log n_{1,j,t-1} - \beta_i^P n_{1,j,t-1}$	$\log \Phi^S + \log n_{2,j,t-1}$
M_{LBL}	$\log \Phi^J + \log n_{3,j,t-1}$	$\log \alpha^P + \log n_{1,j,t-1} - \log\left(1 + \frac{n_{1,j,t-1}}{K_j^P}\right)$	$\log \Phi^S + \log n_{2,j,t-1}$
M_{LLR}	$\log \Phi^J + \log n_{3,j,t-1}$	$\log \Phi^P + \log n_{1,j,t-1}$	$\log \alpha^S + \log n_{2,j,t-1} - \beta_j^S n_{2,j,t-1} + \lambda O_t$
M_{LLB}	$\log \Phi^J + \log n_{3,j,t-1}$	$\log \Phi^P + \log n_{1,j,t-1}$	$\log \alpha^S + \log n_{2,j,t-1} - \log\left(1 + \frac{n_{2,j,t-1}}{K_i^S}\right) + \lambda O_t$
Habitat			
M_{B_hLL}	$\log \alpha^J + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_j^J(1+\gamma I_{j,t})}\right)$	$\log \Phi^P + \log n_{1,j,t-1}$	$\log \Phi^S + \log n_{2,j,t-1}$
M_{BL_hL}	$\log \alpha^J + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_j^J}\right)$	$\log \acute{\Phi}^P + \log n_{1,j,t-1}$	$\log \Phi^S + \log n_{2,j,t-1}$
$M_{B_hL_hL}$	$\log \alpha^J + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_j^J(1+\gamma I_{j,t})}\right)$	$\log \acute{\Phi}^P + \log n_{1,j,t-1}$	$\log \Phi^S + \log n_{2,j,t-1}$
Exclusion			
M_{NBL}	excluded	$\log \alpha^P + \log n_{3,j,t-2} - \log\left(1 + \frac{n_{3,j,t-2}}{K_j^P}\right)$	$\log \Phi^S + \log n_{2,j,t-1}$
M_{BNL}	$\log \alpha^J + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_j^J}\right)$	excluded	$\log \Phi^S + \log n_{1,j,t-2}$
M_{NNB}	excluded	excluded	$\log \alpha^S + \log n_{3,j,t-3} - \log\left(1 + \frac{n_{3,j,t-3}}{K_i^S}\right) + \lambda O_t$

Notes: Survival rates $\Phi^P, \acute{\Phi}^P$ differed in their parameterization. For most models smolt survival was modeled on the logit scale, $\Phi^P = \text{logit}^{-1}(\varphi^P)$, although for M_{BL_hL} and $M_{B_hL_hL}$ survival was modeled as a function of the habitat covariate, e.g. $\acute{\Phi}^P = \text{logit}^{-1}(\varphi^P + \eta I_{j,t})$. Similarly $\Phi^S = \text{logit}^{-1}(\varphi^S + \lambda O_t)$ for all models.

2.2.3 Developing models with habitat effects

Once an adequate model had been constructed in Section 2.2.2.3, a further extension of the models involved investigating the possibility that habitat modification significantly influences coho salmon dynamics. *A priori*, habitat manipulation was hypothesized to increase egg–juvenile and juvenile–smolt survival, by increasing the availability and quality of rearing habitat. Habitat covariates were added to the juvenile and smolt process equations in-turn, and then simultaneously, and improvements in fit and the examination of posterior densities of relevant parameters were used to determine whether the resulting models improved inferential performance.

The initial expansion of models in Section 2.2.2.3 (results given in Section 3.2.2) led to the acceptance of M_{BLL} as the most appropriate for the data. This model suggested compensation occurring at the juvenile stage, thus it seems logical that an increase in rearing habitat will lead to a higher carrying capacity rather than simply increasing survival rates. A linear relationship was therefore assumed between the carrying capacity ($\alpha^J \times K_j^J$) and the index of rearing habitat ($I_{j,t}$; Section 2.2.1.3) which is achieved by reparameterizing the Beverton-Holt juvenile process equation (2.7). The smolt process equation was also modified to investigate the possibility that rearing habitat benefits this part of the life-cycle. No compensation was identified at this stage during the model expansion in Section 2.2.2.3 and so it was assumed that habitat would affect survival, rather than compensatory dynamics. The two modified process equations are thus given by

$$\log n_{1,j,t} = \log \alpha^J + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_j^J\,(1+\gamma I_{j,t})}\right) + \epsilon_{1,j,t} \tag{2.8}$$

$$\log n_{2,j,t} = \log \acute{\Phi}^P + \log n_{1,j,t-1} + \epsilon_{2,j,t}\ , \quad \text{where} \quad \acute{\Phi}^P = \text{logit}^{-1}(\varphi^P + \eta I_{j,t}) \tag{2.9}$$

where γ and η are parameters determining the strength of the relationships between the habitat covariate and the carrying capacity of the Beverton-Holt function (2.8), and survival of fish from the juvenile to the smolt stage (2.9), respectively. Three models are possible depending on whether habitat was assumed to affect the juvenile and/or the smolt stages; M_{B_hLL} consisting of (2.8), (2.5) and (2.6), M_{BL_hL} consisting of (2.7), (2.9) and (2.6), and $M_{B_hL_hL}$ consisting of

(2.8), (2.9) and (2.6) (habitat models; Table 2.1).

2.2.4 Developing models with stages excluded

A major question about the performance of life-history models is how much improvement they make to inferences over models fitted to only spawner abundance data. To assess the value of juvenile and smolt abundance data with respect to the performance of the model in out-of-sample prediction of spawner abundance, these data sources were excluded from modeling and the process equations were modified to account for their absence. Due to only minor improvements in fit of the habitat models (at least at the spawner stage where prediction is focused) presented in Section 2.2.3 over M_{BLL} from Section 2.2.2.3, the latter was used as the base model for this investigation. The performance of models in the absence of spawner data was not considered as this is generally the focal stage of the life-cycle and is invariably the first stage to be incorporated into monitoring programs. This left three possible scenarios; juvenile data excluded (M_{NBL}), smolt data excluded (M_{BNL}) and both juvenile and smolt data excluded (M_{NNB}), with the full deterministic components of the process equations for these models displayed in Table 2.1.

To represent the absence of juvenile data, a model (M_{NBL}) was considered where the juvenile process equation was removed completely and the smolt stage was modeled directly as a function of the spawner stage two years previously i.e.

$$\log n_{2,j,t} = \log \alpha^P + \log n_{3,j,t-2} - \log\left(1 + \frac{n_{3,j,t-2}}{K_j^P}\right) + \epsilon_{2,j,t} \tag{2.10}$$

and spawners remain a function of the smolt one year previous ((2.6); Table 2.1). Similarly, a model representing the absence of smolt data (M_{BNL}) was constructed by modifying the process model for spawners to be a function of juvenile abundance two years earlier

$$\log n_{3,j,t} = \log \Phi^S + \log n_{1,j,t-2} + \epsilon_{3,j,t} \ , \quad \text{where} \quad \Phi^S = \text{logit}^{-1}(\varphi^S + \lambda O_t) \ .$$

Finally, the absence of both juvenile and smolt data is represented in a further model (M_{NNB}) by again modifying the spawner process model such that spawners in year t are a function of

spawners three years prior, i.e.

$$\log n_{3,j,t} = \log \alpha^S + \log n_{3,j,t-3} - \log \left(1 + \frac{n_{3,j,t-3}}{K_i^S}\right) + \lambda O_t + \epsilon_{3,j,t} \ . \tag{2.11}$$

Note that the absence of juvenile data in (2.10) and (2.11) result in the density dependent equations in M_{BLL} being excluded also, hence the compensatory dynamics are shifted to the next stage-class in both these models. Partially marginalized predictive checks of models where no compensation was included (e.g. models denoted M_{NLL} and M_{NNL}) showed that the assumptions of linearity in these models conflicted with the data hence they were not presented herein or considered further.

Measurement of the value of data from the juvenile and smolt stages is difficult using the exclusion method as model comparison using likelihood-based measures such as $\overline{D(\theta)}$ or DIC becomes impossible due to different datasets being used for different models. Leave-one-out cross validation was therefore used to assess the predictive performance of models with and without different combinations of data from stages. This was carried out as follows; observed spawner abundance from one year, t (for both streams simultaneously), was excluded from the dataset for fitting the model being considered, spawner counts in this year were therefore considered missing values and the model formulation was modified accordingly, by removing the observation model component for this year. The model was then fitted, model predictions of spawner abundance (the latent states, $n_{3,j,t}$) in this year were stored and compared to the observed counts ($y_{3,j,t}$). The excluded observations were then added back to the dataset and the next years spawner counts were excluded, and so on, until data from each year had been excluded once. Due to the presence of missing values for juvenile and smolt abundance in the first two years of the time-series, spawner counts were only excluded from year five (1990) onwards, resulting in a sample size of 21 years in which data was excluded.

2.2.5 Prior distributions for model parameters

Prior distributions for all parameters in the models were specified so that they could be considered "vague" or "weakly informative" depending on the definitions given by various authors (Punt

and Hilborn, 1997; Gelman et al., 2003). They were developed with the aim of minimizing their influence on the resulting posterior distributions. Independent prior distributions were assumed for all parameters. Normally distributed priors were assumed to be suitable for most parameters, sometimes on the log-scale. Parameters $\log n_{i,j,u_i}$, $\log \alpha_i$, $\log K_j^J$, $\log K_j^P$, $\log K_j^S$, $\log \phi^J$, φ^P, φ^S, γ, η and λ were given given normal priors with mean zero and variances set at a large enough value for each parameter so as to be approximately uniform over the parameter space deemed to be biologically plausible. Note that $\log n_{i,j,u_i}$ represent the initial states for each stage-class, where $u = (u_1, u_2, u_3) = (3, 4, 4)$ because the process equations were modeled over different time periods depending on the presence of observations in the previous stage-class. For juveniles this period was $t = 3, 4, ..., 26$, for smolt it was $t = 4, 5, ..., 26$ and for spawners it was $t = 4, 5, ..., 25$.

The approach of Gelman (2006) was adopted to formulate priors for the components of process error variance, by specifying half-Cauchy distributions on the standard deviations, τ_i and δ_i, with scale parameters of one, which is slightly larger than the expected standard deviations. Finally, the priors for the Ricker compensation parameters, β_j^J, β_j^P and β_j^S were set to be uniformly distributed such that $\sim \mathrm{U}(0, 10)$.

2.2.6 Calculation of DIC and model computation

The model expansion approach in Section 2.2.2.3 was further validated by comparing the relative performance of candidate models on the basis of DIC as presented in Chapter 4. DIC can be calculated using the standard formula

$$\mathrm{DIC} = \overline{D(\boldsymbol{\theta})} + p_D \tag{2.12}$$

where p_D is the effective number of parameters calculated as $p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$, $\overline{D(\boldsymbol{\theta})}$ is the posterior mean deviance, and $D(\bar{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior means of the parameters. As seen in Section 4.2.3, DIC and each of its components can be calculated at both the data (DIC_d) and partially marginalized (DIC_m) level, and for each individual stage (indexed by superscript i) separately, in addition to all stages combined (superscript C). Thus DIC_m^1 and $\overline{D(\boldsymbol{\theta})}_m^1$ denote DIC and posterior mean deviance for the juvenile stage at the marginal level,

while DIC_m^C and $\overline{D(\boldsymbol{\theta})}_m^C$ are the same measures for all stages combined at the data level, for example. Derivation of DIC_d and DIC_m are shown in detail in Section 4.2.5 and so only the relevant equations for the full model are shown here.

Deviance at the data level assesses the fit of the latent states to the observed data (e.g. it is calculated using (4.19) in Chapter 4), was calculated for all stages combined as

$$D_d^C = 2\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{t=u_i}^{U_i} \log\left(\sqrt{2\pi\sigma_{i,j,t}^2}\right) + \frac{(\log y_{i,j,t} - \log n_{i,j,t})^2}{2\sigma_{i,j,t}^2}$$

where $u = (u_1, u_2, u_3) = (3, 3, 1)$ indexes the first year of observations for each stage, as estimates of abundance for juveniles and smolt only began in 1988 ($t = 3$). Similarly, $U = (U_1, U_2, U_3) = (26, 26, 25)$ indexes the last year of observations as no estimate of spawner abundance in 2011 ($t = 26$) were yet available at the time of analysis. DIC at the data level, for the full model, DIC_d^C was calculated by inserting D_d^C into (2.12).

The marginalized partial deviance is based on the distribution of $y_{i,j,t}$ conditioned on abundance in year $t-1$ (e.g. it is calculated using (4.21) - see Section 4.2.5 for the derivation), and was calculated

$$D_m^C = -2\sum_{i=1}^{3}\sum_{t=u_i}^{U_i} \log\left(\frac{1}{2\pi\kappa_{i,1,t}\kappa_{i,2,t}\sqrt{1-\rho_{i,t}^2}}\exp\left(-\frac{1}{2(1-\rho_{i,t}^2)}\left[\frac{(\log y_{i,1,t} - \log\tilde{n}_{i,1,t})^2}{\kappa_{i,1,t}^2} + \frac{(\log y_{i,2,t} - \log\tilde{n}_{i,2,t})^2}{\kappa_{i,2,t}^2} - \frac{2\rho_{i,t}(\log y_{i,1,t} - \log\tilde{n}_{i,1,t})(\log y_{i,2,t} - \log\tilde{n}_{i,2,t})}{\kappa_{i,1,t}\kappa_{i,2,t}}\right]\right)\right) \tag{2.13}$$

where $\boldsymbol{u} = (u_1, u_2, u_3) = (3, 4, 4)$ and $\boldsymbol{U} = (U_1, U_2, U_3) = (26, 26, 25)$, and

$$\kappa_{i,1,t}^2 = \sigma_{i,1,t}^2 + \tau_i^2 + \delta_i^2$$

$$\kappa_{i,2,t}^2 = \sigma_{i,2,t}^2 + \tau_i^2 + \delta_i^2$$

$$c_{i,t} = \rho \times \kappa_{i,1,t} \times \kappa_{i,2,t} = \tau_i^2$$

$$\rho_{i,t} = \frac{\tau^2}{\kappa_{i,1,t} \times \kappa_{i,2,t}} \ .$$

By inserting D_m^C into (2.12) the partial DIC for the full model, DIC_m^C, can be calculated.

Models were fitted using the software OpenBUGS version 3.0.3 (Lunn et al., 2009) which was implemented via the statistical software R version 2.14.0 (R Development Core Team, 2010) using the contributed packages `BRugs` and `R2WinBUGS` (Sturtz et al., 2005). Two chains of the MCMC sampler were run for 90 000 iterations, with a burnin of 30 000, and the sample was thinned to 1 in 6, resulting in 20 000 samples being stored to disk. Convergence of the sampler was assessed during trial runs using a suite of diagnostic plots and statistics (Geweke, 1992; Brooks and Gelman, 1998) to confirm that stationarity was achieved before the burnin ended, for all models. The total number of iterations and thinning rate were established to ensure that the effective number of samples stored was greater than 1 000 for all parameters. Unless stated otherwise parameter estimates are presented as posterior means and the limits of 95% central credible intervals.

Chapter 3

The potential of Bayesian life-history models of coho salmon dynamics in Lobster Creek, Oregon: Presentation of results

3.1 Linkage to other chapters

The results presented in this chapter build on the methodology presented in Chapter 2, which were in-turn influenced by each of the other chapter which will come later in the thesis. More explicitly, the need to investigate model fit using partially marginalized deviance rather than relying on the more commonly estimated data level deviance is the topic of Chapter 4, and its comparative performance is investigated in more detail there, using a variety of empirical and simulated data. The predictive checking framework that will be utilized extensively in sections 3.2.1–3.2.4 is a result of Chapter 5 where the difficulties in detecting inadequacies of state-space models using well known posterior predictive checks is highlighted. An alternative partially marginalized predictive distribution is presented there and is shown to have more desirable properties for the models being investigated in this chapter. Lastly, models with uninformative prior distributions on the observation error variances are shown to be difficult to fit reliably in Section 8.2. The topic of Chapter 6 is the determine the best methods for estimating observation error variance for the spawner stage when periodic count datasets are available. The results of that chapter are used in calculating $\sigma^2_{i,j,t}$ in Section 2.2.2.1, and hence are a key component of

model fitting results presented in Section 3.2 below.

3.2 Results

3.2.1 Inadequacy of simple models

Strong relationships between raw counts in subsequent stages, in subsequent years, were observed, although there appears to be significant variation in some, particularly between the abundance of spawners and the abundance of smolts in the preceding year (Fig 3.1 e,f). There appears to be some nonlinearity in the relationship between juveniles and the number of individuals spawning in the preceding year (Fig 3.1 a,b).

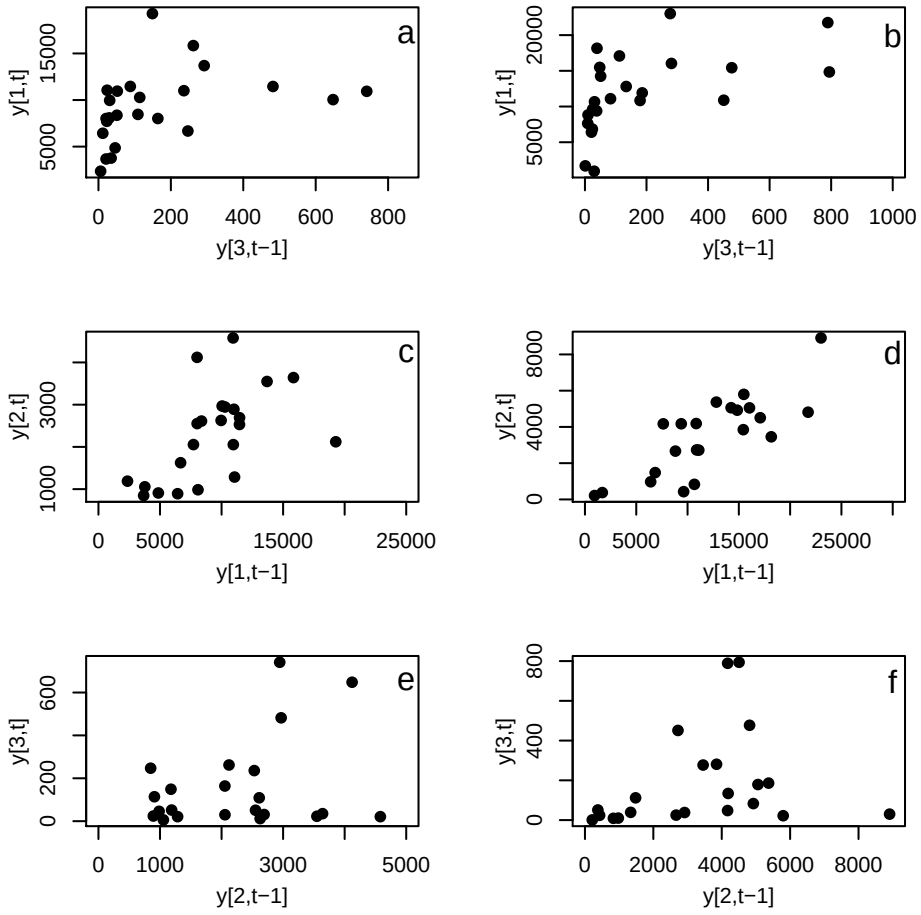

Figure 3.1: Plots displaying the relationships between counts at different stage-classes i, in year t, as a function of counts of the same cohort in the preceding stage-class $(i-1)$ in the preceding year $(t-1)$. Labels $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$ denote juveniles, smolt and spawners, respectively. Panels a, c, e show counts in EF and the remainder are for MF.

Simple models fitted to the dataset were clearly inadequate. The initial assumption that process errors were independent between steams appeared to be violated at the juvenile and spawner, and to a lesser extent the smolt stage. Using model M^*_{BLL} (which assumes no correlation between the process errors of the two streams; see appendix Section 8.1 for specification of this model) as an example, the correlation coefficient between the process errors simulated from the partially marginalized predictive distribution were considerably lower than for the model fitted to the observed data (Fig 3.2; this test variable is tested by $P^{(\Theta)}_{\epsilon,i,cor}$ in (5.2.8) in Section 5.2.8). When the variance components in (2.2) were added to the process error structure, the predictive distributions became more consistent with the data (Fig 3.3).

It was also necessary to include information on the magnitude of the observation errors. For instance, although a model constructed with uninformative priors on the observation error variances ($\sigma^2_{i,j,t}$) passed the tests of convergence of the MCMC sampler, the resulting posterior distributions for the variance components exhibited some worrying characteristics. This included significant amounts of the posterior density being clustered at zero for several of the variance parameters (appendix Fig 8.1). Using the estimates of $\sigma^2_{i,j,t}$ derived from monitoring (Sections 2.2.1.4–2.2.1.6) and assuming that the values were known, resulted in coherent posterior inferences for the variance component parameters, and all models presented herein therefore assume this structure.

3.2.2 Model expansion

A model with linear process equations at each stage (M_{LLL}) led to datasets simulated from the partially marginalized predictive distribution that displayed discrepancies that were at odds with the discrepancies for the observed data. Most notably, the observed process errors were highly negatively correlated with population size while the replicated process errors were not, suggesting non-linearity in the system that was not accounted for by the model (Fig 3.4; this test variable is tested by $P^{(\Theta)}_{\epsilon,i,den}$ in (5.2.8). The most obvious source of non-linearity appeared to be at the juvenile stage (Fig 3.1), and this also appeared to be where the most substantial discrepancy

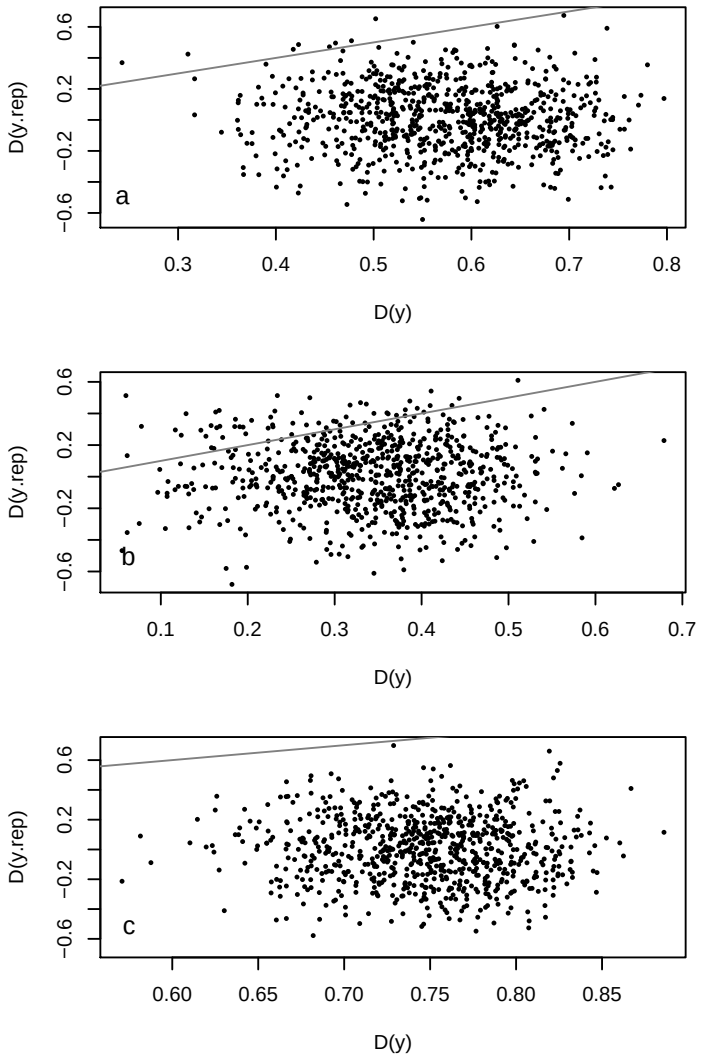

Figure 3.2: Predictive checks testing for correlation in process errors between streams (the graphical equivalent to $\tilde{P}^{(\Theta)}_{\epsilon,i,cor}$ in Section 5.2.8). The observed discrepancy, $D(\boldsymbol{y})$, is compared to $D(\boldsymbol{y.rep})$ which is the discrepancy measured for latent states simulated from the partially marginalized predictive distribution of model M^{*}_{BLL} (appendix Section 8.1), at the juvenile (a), smolt (b) and spawner (c) stages. Lines represent equality between the measures.

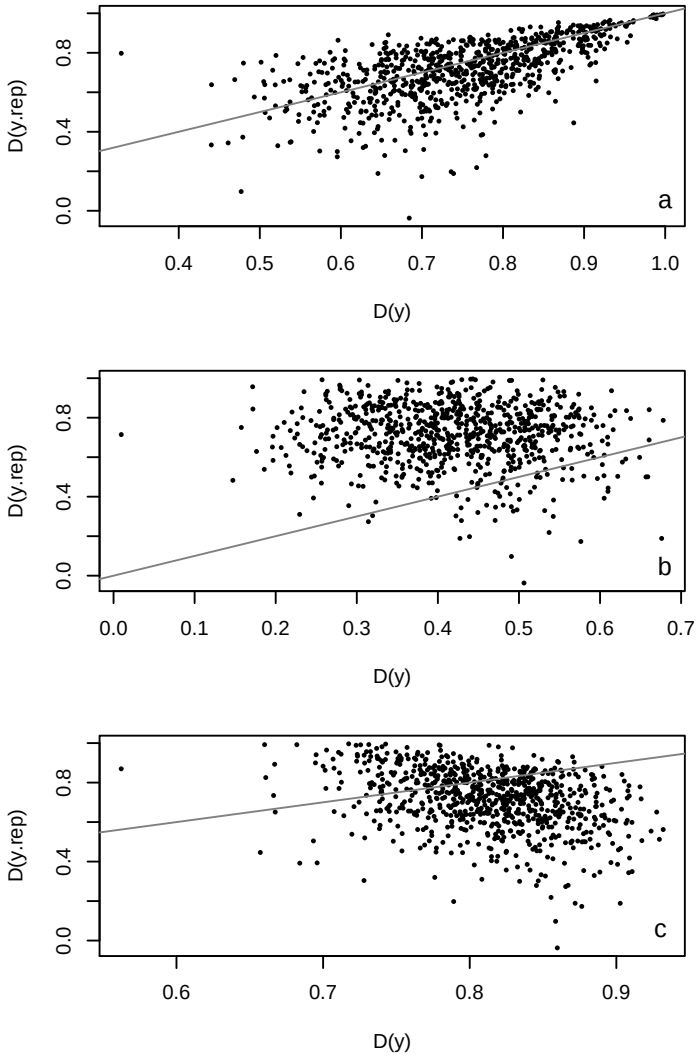

Figure 3.3: Predictive checks testing for correlation in process errors between streams (the graphical equivalent to $\tilde{P}^{(\Theta)}_{\epsilon,i,cor}$ in Section 5.2.8). The observed discrepancy, $D(\boldsymbol{y})$, is compared to $D(\boldsymbol{y.rep})$ which is the discrepancy measured for latent states simulated from the partially marginalized predictive distribution of model M_{BLL}, at the juvenile (a), smolt (b) and spawner (c) stages. Lines represent equalit between the measures.

between model and data occurred (Fig 3.4 a,b). If a Beverton-Holt function was assumed at the juvenile stage (M_{BLL}) partially marginalized predictive checks suggested consistency with the observed data (Fig 3.5). However, if a Ricker function was assumed (M_{RLL}), the inconsistencies between the observed and replicated process errors remained at the juvenile stage (Fig 3.6). This is largely the result of the Ricker function overestimating production of juveniles at moderate abundance of spawners, and predicting overcompensation at high spawner abundance, which is not supported by the data (Fig 3.7). In contrast the Beverton-Holt function appears to fit the relationship well and full posterior mean partial deviance of M_{BLL} was $\overline{D(\boldsymbol{\theta})}^{C}_{m}$=205.8, considerably lower than for the fully linear model, M_{LLL} ($\overline{D(\boldsymbol{\theta})}^{C}_{m}$=275.8) and the model with a Ricker function at the juvenile stage, M_{RLL} ($\overline{D(\boldsymbol{\theta})}^{C}_{m}$=230.5).

Once non-linearity at the juvenile stage was explicitly modeled, there was little evidence for substantial discrepancy between model predictions and data with respect to further non-linearity at either the smolt or spawner stages (Fig 3.5). Furthermore, investigations of other discrepancies between data and the model using posterior and partially marginalized predictive checks and analysis of residuals (Chapter 5) failed to identify any obvious inadequacies in the model.

Changing the form of the process equation at one stage had little effect on the fit of the model at other stages. For example, most of the improvement in fit of the models with Beverton-Holt-type compensation at the juvenile stage, not surprisingly occurred at that stage, and all models in the initial set displayed relatively similar fit to the data at the subsequent smolt stage ($\overline{D(\boldsymbol{\theta})}^{2}_{m}$=48.9–52.2; Table 3.1).

When the marine survival covariate $\boldsymbol{O}$ and associated parameter λ were removed from M_{BLL} the fit was considerably poorer with full posterior mean partial deviance, $\overline{D(\boldsymbol{\theta})}^{C}_{m}$, increasing from 205.8 to 214.4. Most notable periods where the covariate improved fit at the spawner stage were 1994–1998, the last few years of the time-series, and most obviously in 1992 in EF (Fig 3.8).

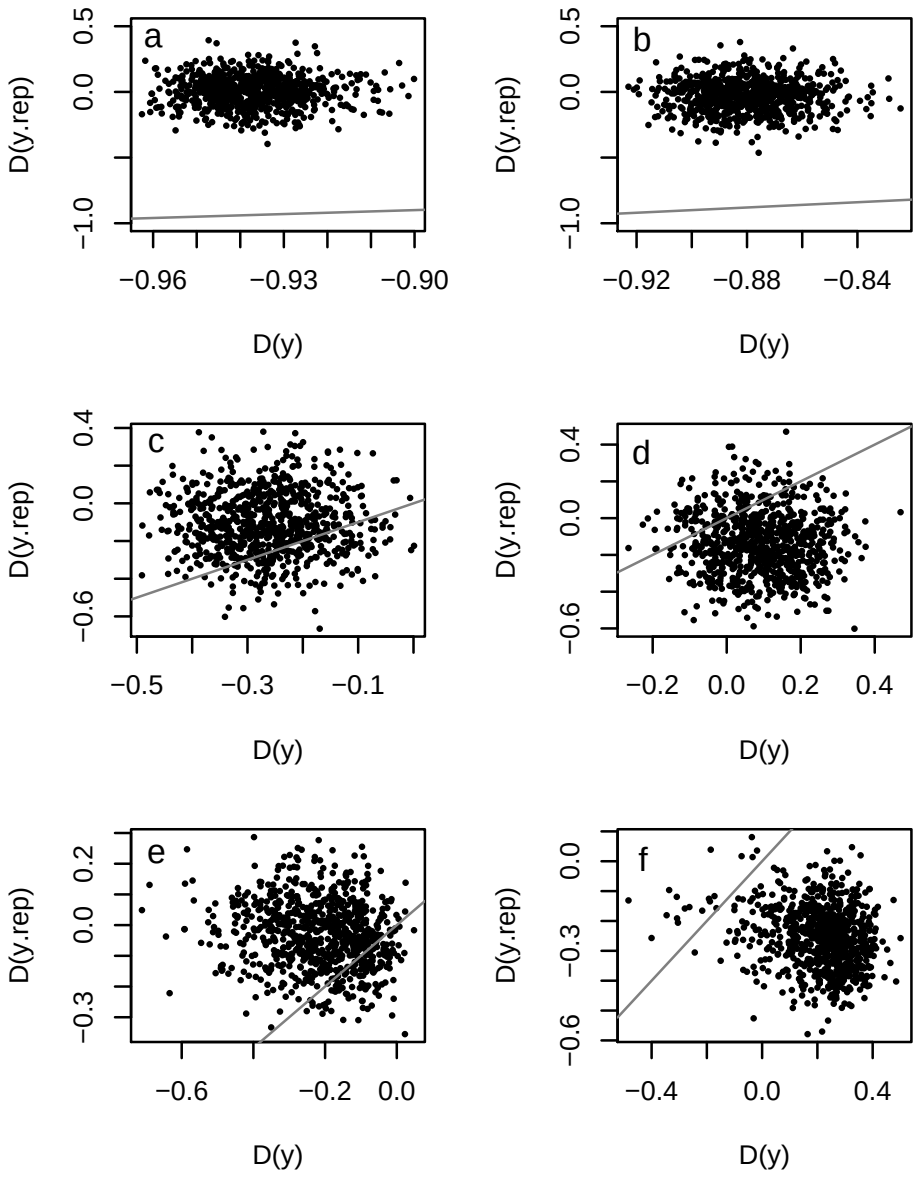

Figure 3.4: Predictive checks testing for correlation between process errors $\epsilon_{i,j,t}$ and the deterministic estimates of population size $\log \tilde{n}_{i,j,t}$ (the graphical equivalent to $\tilde{P}^{(\Theta)}_{\epsilon,den}$ in Section 5.2.8). The observed discrepancy, $D(\boldsymbol{y})$, is compared to $D(\boldsymbol{y.rep})$ which is the discrepancy measured for latent states simulated from the partially marginalized predictive distribution of model M_{LLL}. Lines represent equality between the measures.

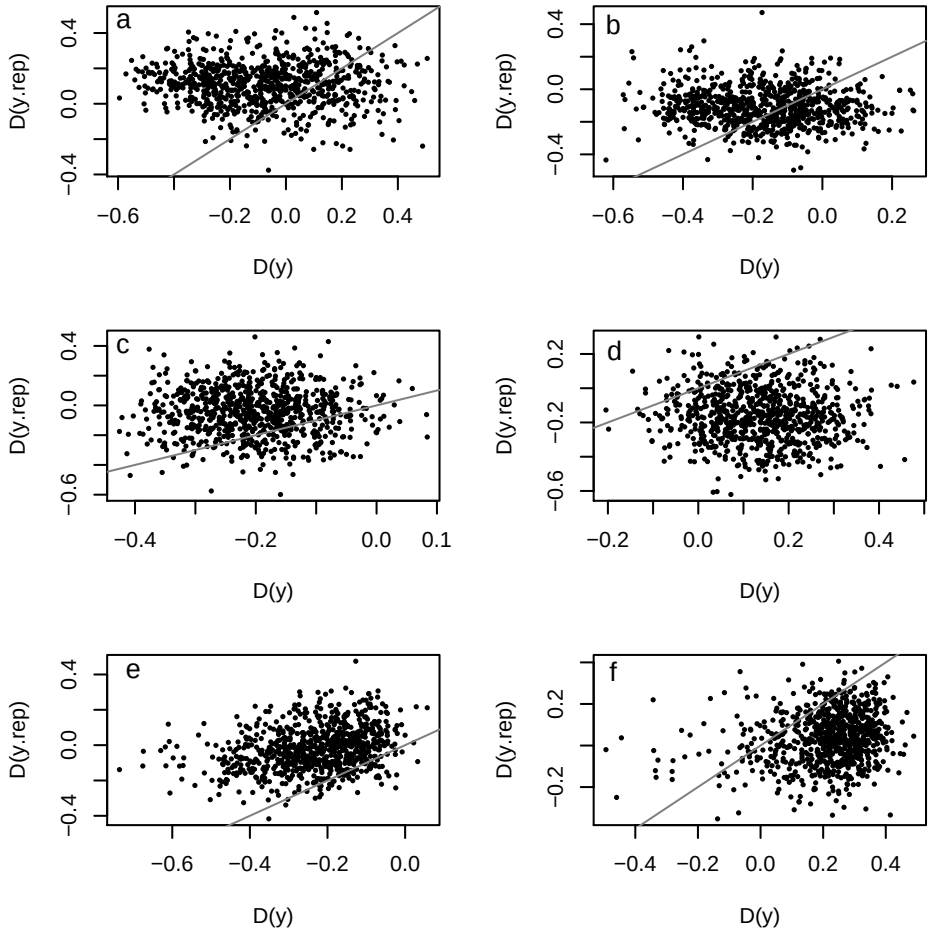

Figure 3.5: Predictive checks testing for correlation between process errors $\epsilon_{i,j,t}$ and the deterministic estimates of population size $\log \tilde{n}_{i,j,t}$ (the graphical equivalent to $\tilde{P}^{(\Theta)}_{\epsilon,den}$ in Section 5.2.8). The observed discrepancy, $D(\boldsymbol{y})$, is compared to $D(\boldsymbol{y.rep})$ which is the discrepancy measured for latent states simulated from the partially marginalized predictive distribution of model M_{BLL}. Lines represent equality between the measures.

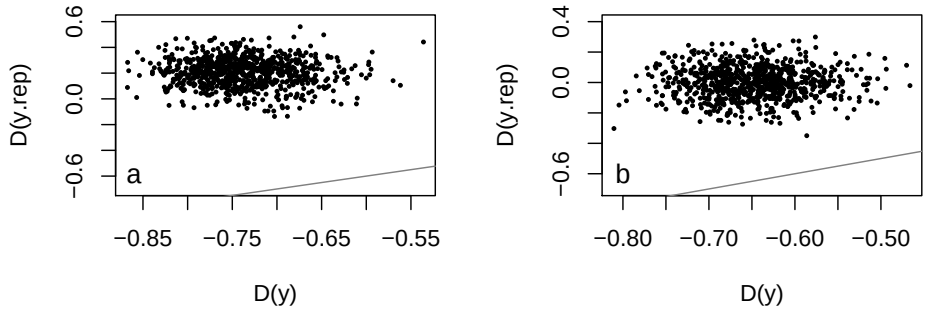

Figure 3.6: Predictive checks testing for correlation between process errors $\epsilon_{1,j,t}$ and the deterministic estimates of population size $\log \tilde{n}_{1,j,t}$ for the juvenile stage. The observed discrepancy, $D(\boldsymbol{y})$, is compared to $D(\boldsymbol{y.rep})$ which is the discrepancy measured for latent states simulated from the partially marginalized predictive distribution of model M_{RLL}. Lines represent equality between the measures.

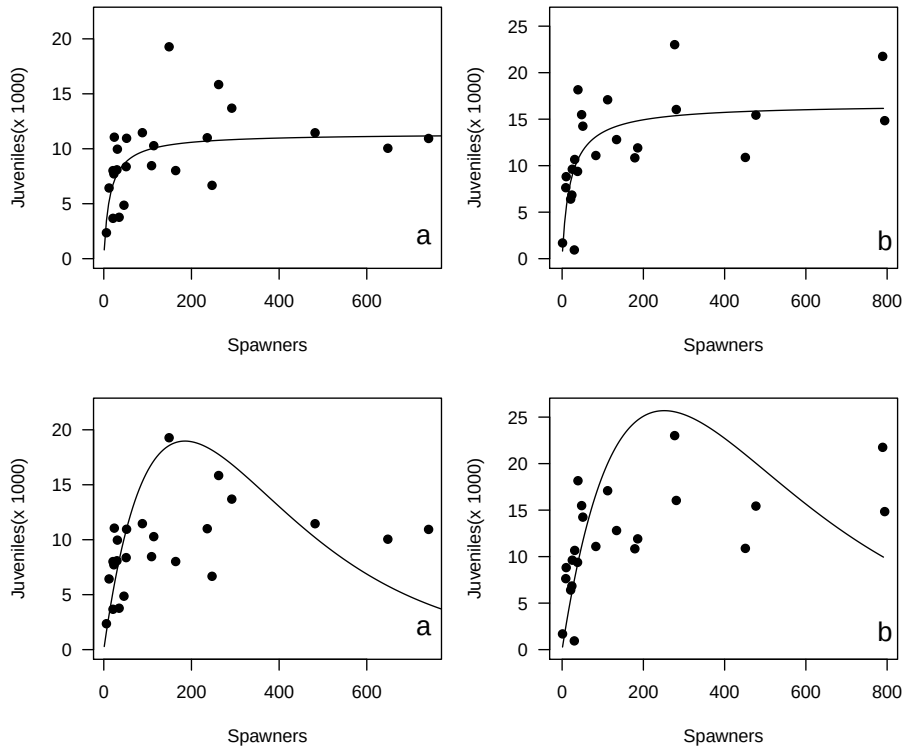

Figure 3.7: Comparison of models assuming Beverton-Holt (M_{BLL}; a,b) and Ricker (M_{RLL}; c,d) functions at the juvenile stage. Dots are the observed values, $y_{i,j,t}$, and lines are the posterior mean estimates for the respective models for the relationship between spawners and juveniles, in EF (a and c), and MF (b and d).

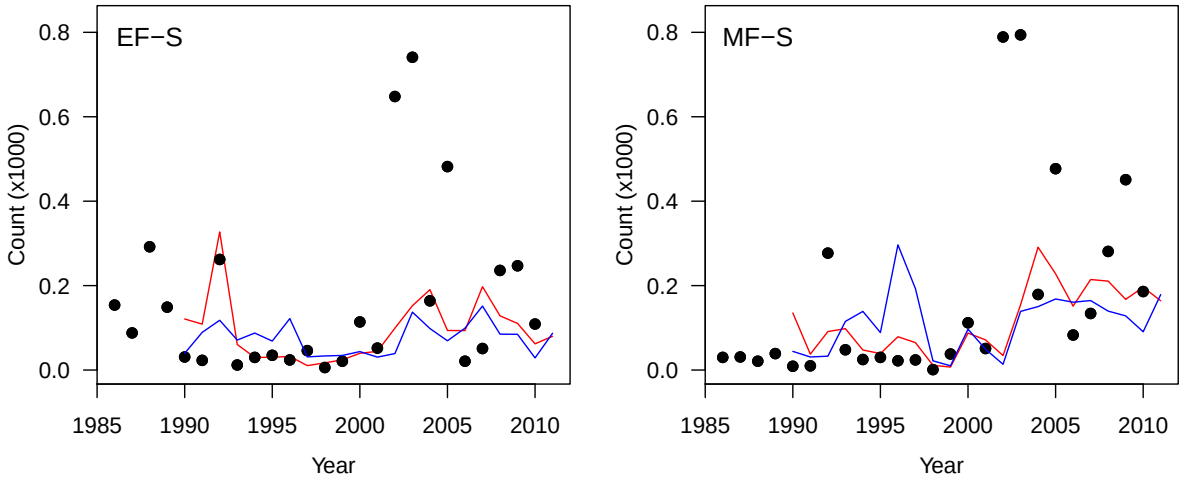

Figure 3.8: Estimated values of the latent states representing abundance of salmon in the spawner stage in East Fork (EF) and Mainstem (MF), Lobster Creek for the best fitting state-space model (M_{BLL}; red lines) and the same model with the ocean survival covariate, $\boldsymbol{O}$ excluded (blue lines) from 1986–2011. Dots are the observed counts and lines are posterior means.

3.2.3 Model comparison using DIC

Formal comparison of models using DIC_m largely supported the results of the expansion method of model building, with M_{BLL} identified as the most parsimonious among the candidate models without habitat covariates (Table 3.1). The compensation occurring between the spawner and juvenile stages in this model was essential for fitting the data well, as evidenced by all high ranking models ($\Delta\text{DIC}_m^C < 59.0$) including compensation at this stage of the life-cycle (Table 3.1). Furthermore, the Beverton-Holt function was again suggested to be more suitable at the juvenile stage than the Ricker function, which fitted the data poorly by comparison ($\Delta\text{DIC}_m^C = 30.4$).

Models with nonlinear functions in the process equations for the smolt and spawner stages ($M_{LBL}, M_{LRL}, M_{LLB}, M_{LLR}$) fitted the data poorly in comparison to M_{BLL} ($\Delta\text{DIC}_m^C > 59.0$) suggesting that any compensation at the smolt or spawner stages was far less important than compensation at the juvenile stage.

Table 3.1: Comparison of Deviance Information Criterion (DIC) across a suite of models of coho salmon dynamics that differ in the structure of the process equations. The posterior mean deviance of the model is denoted $\overline{D(\boldsymbol{\theta})}$ and pD is a measure of the effective number of parameters in the model. The partial set compares DIC at the partially marginalized level, DIC_m, while the data set compares DIC at the data level, DIC_d. Superscripts 1, 2, 3 and C indicate juveniles, smolt, spawners and the complete model (all stages combined).

		Juveniles			Smolts			Spawners			Total			
Level	Model	$\overline{D(\boldsymbol{\theta})}^1$	pD^1	DIC^1	$\overline{D(\boldsymbol{\theta})}^2$	pD^2	DIC^2	$\overline{D(\boldsymbol{\theta})}^3$	pD^3	DIC^3	$\overline{D(\boldsymbol{\theta})}^C$	pD^C	DIC^C	$\Delta\mathrm{DIC}^C$
Partial	M_{BL_hL}	37.9	6.8	51.5	42.3	7.6	57.5	115.9	4.2	124.2	196.1	18.6	233.3	0.0
	$M_{B_hL_hL}$	39.3	7.5	54.2	42.2	7.7	57.6	115.6	4.2	124.1	197.1	19.4	235.9	2.6
	M_{BLL}	38.6	6.5	51.7	51.4	6.0	63.5	115.7	4.2	124.2	205.8	16.8	239.3	6.0
	M_{B_hLL}	39.3	7.6	54.5	50.5	6.1	62.8	115.4	4.2	123.9	205.3	17.9	241.1	7.9
	M_{RLL}	65.8	8.8	83.4	50.0	6.6	63.1	114.7	4.2	123.1	230.5	19.6	269.6	36.4
	M_{LLR}	106.4	6.2	118.7	49.3	6.5	62.4	105.1	6.0	117.2	260.8	18.8	298.3	65.0
	M_{LLB}	109.2	6.1	121.5	49.4	6.6	62.6	107.3	6.3	119.8	265.9	19.0	303.9	70.7
	M_{LLL}	113.1	6.1	125.3	48.9	6.8	62.6	113.8	4.4	122.7	275.8	17.4	310.5	77.3
	M_{LRL}	112.9	6.2	125.2	51.3	6.5	64.2	114.0	4.4	122.7	278.2	17.0	312.2	78.9
	M_{LBL}	112.8	6.2	125.2	52.2	6.6	65.4	114.1	4.3	122.8	279.1	17.1	313.4	80.1
Data	M_{RLL}	-105.3	41.4	-22.6	-194.4	46.0	-102.4	-54.6	41.9	29.3	-354.3	129.3	-95.7	0.0
	M_{LLL}	-106.7	42.4	-21.9	-194.4	45.9	-102.7	-55.8	42.6	29.3	-357.0	130.8	-95.3	0.4
	M_{LRL}	-107.5	43.3	-21.0	-193.9	45.8	-102.4	-56.0	42.6	29.3	-357.4	131.7	-94.1	1.6
	M_{LLB}	-106.7	42.2	-22.3	-195.7	46.2	-103.2	-50.5	41.1	31.6	-352.9	129.5	-93.9	1.8
	$M_{B_hL_hL}$	-101.0	39.5	-22.0	-193.5	46.0	-101.5	-59.8	45.0	30.2	-354.3	130.5	-93.3	2.4
	M_{B_hLL}	-101.3	39.9	-21.4	-194.4	46.1	-102.2	-60.0	45.3	30.6	-355.7	131.3	-93.0	2.6
	M_{LBL}	-107.6	43.6	-20.3	-194.0	46.0	-102.0	-56.0	42.7	29.3	-357.5	132.3	-93.0	2.7
	M_{BLL}	-100.6	39.4	-21.7	-194.5	46.1	-102.2	-59.3	45.4	31.4	-354.4	130.9	-92.5	3.2
	M_{BL_hL}	-99.9	39.0	-22.0	-193.3	45.9	-101.5	-58.9	45.2	31.5	-352.2	130.1	-92.0	3.6
	M_{LLR}	-106.4	42.2	-22.0	-195.7	46.1	-103.4	-46.4	40.0	33.6	-348.5	128.3	-91.8	3.9

Table 3.2: Estimates of the standard deviations of the temporal (τ) and spatial components of process errors (δ) for model M_{BLL}. Displayed are the posterior means and standard deviations for the three stage-classes. Also shown are the observation error standard deviations (ς) that were fixed constants in the model, with the mean and standard deviation shown being calculated over all years over both streams.

	τ		δ		ς	
Stage-class	mean	sd	mean	sd	mean	sd
Juveniles	0.36	0.08	0.23	0.08	0.12	0.04
Smolt	0.26	0.11	0.36	0.07	0.07	0.11
Spawners	1.05	0.21	0.55	0.12	0.22	0.10

3.2.4 Fit and diagnostics for the best model

The latent states, $n_{i,j,t}$, of the best model among the initial set (M_{BLL}) appeared to fit the data for each stage very well (Fig 3.9). This is a consequence of the low observation error variances assumed at the three stages which resulted from the intense monitoring programs. These directly led to high precision in the estimates of the underlying abundances at each stage, though especially for smolts. Substantial variation in abundance occurred over the modeling period for each stage (Fig 3.9). There was a pattern observed for spawners and smolts, and to a lesser extent juveniles, of high abundance at the beginning and end of the time series, with very low abundance observed during ∼1995–2000. Mean abundance of spawners increased approximately 133- and 623-fold between the lowest observed abundance in 1998 and the highest in 2003, in the EF and MF respectively.

Due to the the high precision in the estimates of abundances, most unexplained variation in the data was attributed to process error. For both the juvenile and smolt stages, process errors showed similar patterns, with temporal process error standard deviation (τ), and the spatial process error standard deviation (δ) being of similar magnitude for these stage-classes, with roughly half the variation attributed to temporal and the other spatial components (Table 3.2). In contrast, both components of process error were considerably higher for the spawner stage, with a larger proportion being attributed to temporal, rather than spatial variation for this stage-class (Table 3.2).

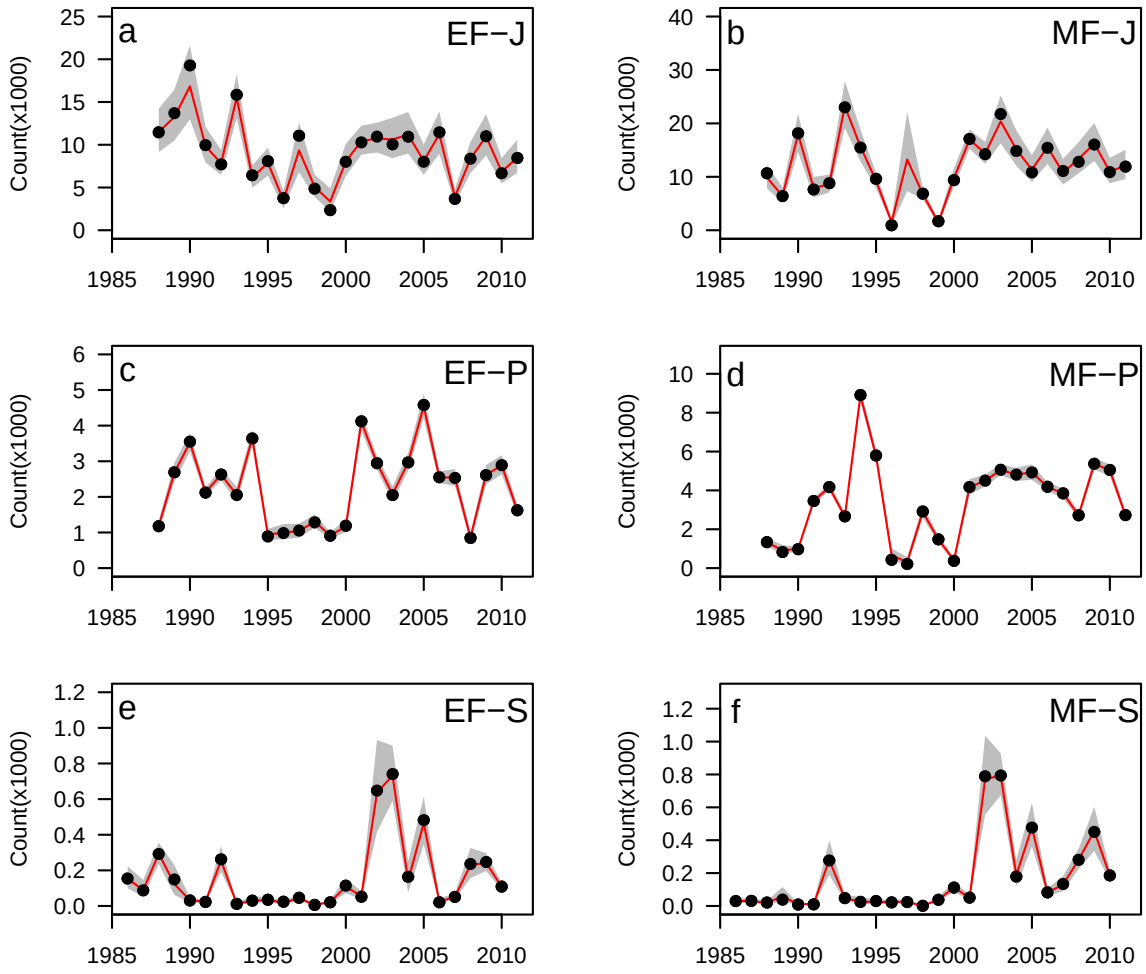

Figure 3.9: Estimated values of the latent states ($n_{i,j,t}$) representing abundance of coho salmon from the best fitting state-space model (M_{BLL}) from 1986–2011. Counts are displayed in the thousands, and represent the juvenile (a,b), smolt (c,d) and spawner stages (e,f), for the EF (a,c,e) and MF (b,d,f) of Lobster Creek. Dots are the observed counts, red lines are posterior means and the gray regions represent 95% credible intervals of the latent states.

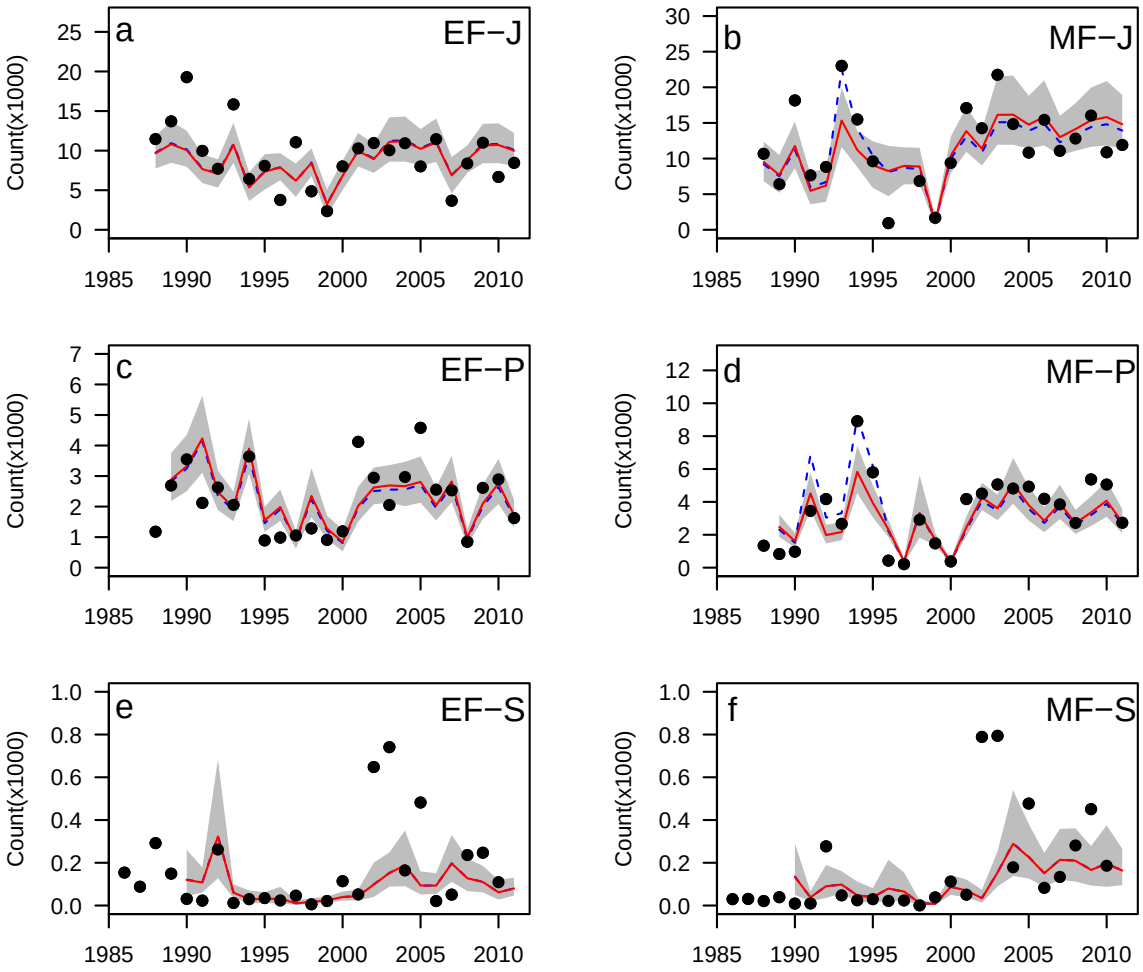

Figure 3.10: Posterior mean estimates of coho salmon abundance based on the deterministic components of the process equations of two models for Lobster Creek from 1986–2011. Abundance is displayed in the thousands, and represents the juvenile (a,b), smolt (c,d) and spawner stages (e,f), for the EF (a,c,e) and MF (b,d,f). Dots are the observed counts, and the red solid lines and gray regions are the posterior means and 95% credible intervals, respectively, for M_{BLL}. The blue dashed lines are the posterior means for an identical model but with habitat covariates at both the juvenile and smolt stages ($M_{B_hL_hL}$). Blue dashed lines are almost identical to the red lines for the spawner stage so are obscured.

Comparing the estimated deterministic components of the process model ($\log \tilde{n}_{i,j,t}$; (2.1) in Section 2.2.2.1) of M_{BLL} to the observed data indicates the ability of the model to explain variation in abundance one year ahead (Fig 3.10). Fit appears adequate and a large part of the variation in abundance for each stage-class appears to be explained by abundance of the preceding stage, in the preceding year. The most notable deviations occur at the spawner stage in both streams from 2002–2011, when observations of spawner numbers are substantially higher than the expectations from the deterministic process model (Fig 3.10). Although the observed spawner numbers were high over that period, observed smolt numbers were close to average, and while the values for survival of hatchery smolt were high ($\boldsymbol{O}$), they were similar to the period from 1986–1991 when the observed survival from smolts to spawners was much lower (Fig 3.10).

These patterns were largely reflected in general predictive checks (appendix Section 8.4). For the most part, partially marginalized predictions of replicate data, $\tilde{y}^{rep}_{i,j,t}$, from M_{BLL} appear to be largely comparable with the observed data for the juvenile and smolt stages, with only a few individual years showing consistent discrepancies. However, systematic inconsistencies between observed and replicated data were detected for the spawner stage (appendix Section 8.4). The model appears unable to achieve the high observed abundances in the early 2000's in both streams and consistently produces replicated data higher than the observed abundance during parts of the low-abundance phase in the mid-1990's in MF.

3.2.5 Fit of habitat models

The habitat models with a covariate at the smolt stage (M_{BL_hL}, $M_{B_hL_hL}$) explained more of the variation in the data ($\overline{D(\boldsymbol{\theta})}^C_m$=196.1–197.1) than M_{BLL} ($\overline{D(\boldsymbol{\theta})}^C_m$=205.8) and resulted in lower DIC values (ΔDIC^C_m >3.4–6.0; Table 4.4). The parameter estimates of γ and η also suggested that there was evidence for relationships between demography and habitat modification, however there was considerable uncertainty in posterior distributions of both parameters. The estimates of η by $M_{B_hL_hL}$ predicted that survival of smolt in years with enhanced rearing habitat to be 1.65 (1.12–2.35) times that in normal years, with 99.3% of the posterior density greater than

1 (Fig 3.11). There was more uncertainty in the estimation of γ, with the asymptote of the Beverton-Holt function (the carrying capacity) for juveniles estimated to be 1.68 (0.88–3.36) times higher during years with enhanced habitat, and 93.3% of the posterior density was greater than 1. The model with only a covariate at the juvenile stage displayed similar fit to M_{BLL} as measured by $\overline{D(\boldsymbol{\theta})}_m^C$ (205.3 vs 205.8) and DIC_m (241.1 vs 239.3).

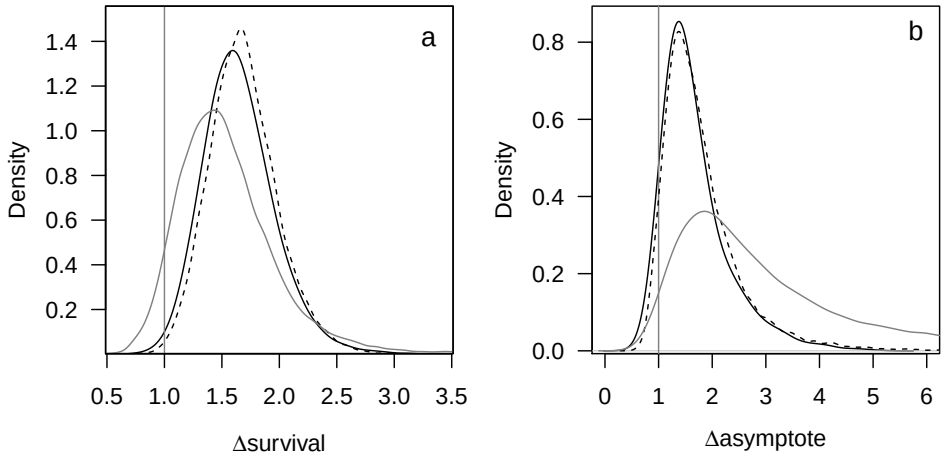

Figure 3.11: The relative increase in (a) survival of smolts, and (b) asymptote of the Beverton-Holt function describing compensation at the juvenile stage, when rearing habitat was enhanced in Main Fork in comparison to normal conditions. Lines displayed are the estimated posterior densities for model $M_{B_hL_hL}$ (solid lines), the same model with independent process errors between streams (M^*_{BLL}; gray lines) and for models with only one habitat parameter (dotted lines; a, M_{BL_hL}, b, M_{B_hLL}).

The improvements in fit of $M_{B_hL_hL}$ over M_{BLL} are displayed in Figure 3.10. The addition of habitat parameters at both the juvenile and smolt stages allowed the deterministic component of the process model to attain levels close to the observed counts in MF in the years following the habitat addition. In contrast, M_{BLL} underestimated abundance in these years by a substantial amount. This is most obvious in 1993 and 1994 for juveniles, and 1994 and 1995 for smolts.

It is interesting to note that failure to account for correlation in process errors between streams affected estimates of habitat effects. Estimates of parameters γ and η, and thus the change in survival and change in carrying capacity, in M^*_{BLL} showed greater uncertainty and/or a greater proportion of their density below one than M_{BLL} (Fig 3.11; gray lines).

Table 3.3: Results of leave-one-out cross validation tests of the performance of models where juvenile and/or smolt data were excluded from the dataset, assessed by comparing predicted spawner abundance in 'missing data' years with observed abundance. Measurements are bias, standard deviation, root mean squared error and mean half width of 95% central credible intervals of posterior estimates of the latent states in the spawner stage-class.

	EF				MF				Total			
Model	Bias	SD	RMSE	HW	Bias	SD	RMSE	HW	Bias	SD	RMSE	HW
M_{BLL}	46	228	232	505	71	185	198	651	59	206	214	578
M_{NBL}	67	251	259	565	80	203	218	680	73	225	237	623
M_{BNL}	63	247	255	565	68	184	196	660	65	215	225	613
M_{NNB}	126	276	304	803	95	277	293	814	111	274	295	808

3.2.6 Importance of data from each stage

Cross validation indicated that all models tested had relatively poor performance with respect to predicting spawner abundance in the years where data was excluded (Table 3.3). Mean bias was greater than 59 spawners for all models and very high variation was present in estimates of spawner abundance in the excluded years (mean 95% credible interval half-widths 578). All summary statistics were relatively similar for M_{NBL} and M_{BNL}, although they were all worse than observed for the model for the full dataset, M_{BLL}. Performance of M_{NNB} at predicting the latent abundances was considerably worse than the other three models, with bias, RMSE and credible interval half-widths about 1.9, 1.4 and 1.4 times larger than the equivalent measures for M_{BLL}

3.3 Discussion

3.3.1 General description of dynamics of spawner abundance

Fluctuations in abundance of Pacific salmon are highly correlated across many populations, and there are even correlations among populations of different species (Adkison et al., 1996; Mueter et al., 2002). The dynamics of spawner abundance in Lobster Creek broadly follow those described for coho salmon elsewhere in the Pacific North-West (Beamish et al., 2000; Bradford and Irvine, 2000), and within the OCN ESU (Wainwright et al., 2008; Buhle et al., 2009), to which Lobster Creek belongs. In general, the early–mid-1990's were a period of poor

marine survival of coho salmon (Beamish et al., 2000, 2008; Koslow et al., 2002), with declines in abundance occurring to the point where many populations reached historically low abundances (Wainwright et al., 2008; Buhle et al., 2009). In contrast, the 2000's have returned to more favorable conditions and, while displaying marked variability, escapements have rebounded to more healthy levels again. Escapement of spawners in Lobster Creek followed this pattern very closely, with the exception being that abundance early in the time-series in MF (1986–1991) was very low and similar to the levels observed during the generally poor period exhibited in EF and other populations (1993–1998).

The mechanisms involved in these dynamics have been well discussed previously (e.g. Bradford and Irvine, 2000) and for this reason will not be presented in detail herein. Long-term declines since at least the early 1900's have been attributed to the synergistic effects of overharvesting by fisheries (Johnson, 1984), unfavorable climatic conditions (Beamish et al., 2000), modification of spawning habitat - both in-stream and via changes in surrounding land use (Hartman et al., 1996; Bilby and Mollot, 2008), and a negative influence of hatchery-released fish (Chilcote, 2003; Nickelson, 2003; Buhle et al., 2009). However, most of the annual variation in abundance of spawners during the time-series available for Lobster Creek can presumably be attributed to prevailing climatic conditions, especially during the ocean stage of the life-cycle (Beamish et al., 2000). Survival rates of smolts to spawners in Lobster Creek varied by more than an order of magnitude, displaying just how much spawner abundance can vary even when smolt abundance is relatively constant. In general, ocean conditions that include features such as cold sea-surface temperatures, weak downwellings and later spring transition dates, among others, appear to be favorable for the survival of coho salmon once they enter the marine environment (Hobday and Boehlert, 2001; Botsford and Lawence, 2002; Logerwell et al., 2003). These conditions appear to be controlled by a combination of large-scale climatic and ocean current systems, and local weather and sea patterns (Mueter et al., 2002). There is also evidence that relatively abrupt changes, or regime shifts, occur in these processes, which has led to the distinct periods of favorable (2000's) and unfavorable (1990's) conditions for coho and several other pacific salmon

species (Beamish et al., 2000; Irvine and Fukuwaka, 2011).

3.3.2 Changes in juvenile and smolt abundance

The dynamics of juvenile and smolt abundance in Lobster Creek generally followed that of spawners, although changes between the favorable and unfavorable periods were less extreme, especially for juveniles. In fact in some years during the early–mid-1990's juvenile abundances reached moderate levels, exhibiting the potential for high productivity even when escapement of spawners is very low. This demonstrates that the freshwater and oceanic stages are 'decoupled' from each other to some degree and reiterates that to gain a full understanding of abundance throughout the life-cycle and the role of coho in both habitats, then monitoring in both is essential. Inferring the status of freshwater stages from spawner counts will clearly be unsound.

The reasonable fit of the deterministic components of both the juvenile and smolt process models indicates that juvenile and smolt abundance can largely be explained by a Beverton-Holt-type non-linear relationship and a linear relationship with abundance of the previous stage in the previous year, respectively. Without compensation in the former relationship, the deterministic components of the process model would predict juvenile abundance in years following high spawner abundance well above the observed counts, for instance during many of the years 2001–2011. The remaining unexplained process error in the juvenile and smolt stages could have been introduced from a variety of sources, for example environmental variation related to flow rates, water temperatures, depredation etc. Including covariates to account for these processes has the potential to improve the inferential and predictive properties of the model, but requires formulation of appropriate hypotheses and measures of the covariates themselves.

3.3.3 Inflated process errors at the spawner stage

Despite the adequate performance of the juvenile and smolt process models, both temporal and spatial components of the process errors were moderate at both stages. By comparison though, process errors at the spawner stage were considerably larger, especially the temporal component, indicating that the deterministic components of the spawner process model showed

limited ability in explaining variation in the observed counts. Furthermore, this inflated process error was evident not only in the estimates of the variance components but also the predictive performance of the models in the cross validation, which displayed relatively high uncertainty, and the variability in partially marginalized predictive distributions for the latent states of spawners. The latter resulted in statistical checks of model adequacy that suggested that the best models showed weakness for this stage, and there were discrepancies between the data and replicated data simulated from the model (Appendix Section 8.4). While this is not ideal, instead of 'rejecting' the model it merely suggests that profitable future work would include investigation of the (presumably) environmental determinants of this unexplained error.

The inadequacy of the spawner process model appears to be due to the OPI covariate being unable to predict the very high abundances observed during several periods of the time-series. This was most notable during some of the years between 2002–2011, when the deterministic components substantially underestimated the observed abundance. The OPI was high for the early and late periods but dropped to a low level in the early–mid-1990's. While spawner abundance followed a similar pattern, it was substantially lower during the early period than towards the end of the time-series, especially in MF. This presumably prevented smolt to spawner survival being directly proportional to OPI, resulting in the bias. OPI therefore appears to be an inadequate covariate for marine survival of the wild populations in Lobster Creek. The deficiencies of OPI are perhaps not surprising given the difficulties in its estimation, the large-scale nature of it's estimation, and the possibility of considerable differences in the survival abilities of hatchery and wild stocks, especially over a range of environmental conditions (Einum and Fleming, 2001; Buhle et al., 2009).

3.3.4 Using models for prediction

One of the consequences of high process error is that predictive performance will be very imprecise (as evidenced by the cross validation exercise), which makes setting management targets for future runs highly uncertain. This would make these models appear to be unsuitable for

this purpose when compared to currently implemented methods, however it may be that they are just more realistic in their treatment of uncertainties occurring in the system. The latest preseason prediction of OCN coho abundance made by the Pacific Fishery Management Council (Pacific Fishery Management Council, 2012) is based on multi-modal inference over a set of generalized additive models that differ in the explanatory variables used to model past abundances. The latter largely involve a range of physical oceanic variables with only one model containing the number of spawners three years earlier as a predictor variable, and no consideration of the underlying population dynamics. While this approach (and earlier versions e.g. Pacific Fishery Management Council, 2009) have been moderately successful in the past (Pacific Fishery Management Council, 2012) due to the profound effects of ocean conditions on escapement levels, formal decision analysis for management decisions such as setting harvest limits requires robust uncertainty estimates (Punt and Hilborn, 1997). The latter are presumably significantly underestimated in currently implemented predictive models due to the absence of process uncertainty.

If state-space life-cycle models similar to Lobster Creek were to be used for prediction and setting management targets for OCN coho salmon, then improving the deterministic components of the spawner process model to reduce predictive uncertainty at this stage, should be the highest priority, as their use with the current levels of uncertainty are questionable. This may be achieved by expanding the range of covariates considered when explaining marine survival, such as specific physical measurements of ocean conditions or perhaps broad-scale climatic indices, for example. Due to the computational demands of state-space models, vast numbers of candidate models are prohibitive, thus considerable thought would be required to pare the set down to a manageable number. Prior studies of relationships between marine survival for coho salmon and physical processes (e.g. Hobday and Boehlert, 2001; Botsford and Lawence, 2002; Logerwell et al., 2003) could be useful in this regard.

3.3.5 Locating timing of compensation with life-history models

Rarely, if ever, have the vital rates for the transitions between the three most distinct stages of the coho life-cycle been investigated using structural models, although Bradford et al. (2000) investigate three stages in an *ad hoc* manner. For coho, as with other Pacific salmon, the most prevalent use of modeling has been spawner-only stock-recruitment analyses, often aiming to identify the strength of compensation and setting management reference points (e.g. Adkison et al., 1996; Barrowman and Myers, 2000). Compensation is almost always assumed, and generally compensation parameters are found to be strong, although it should be noted that alternative, density-independent models are seldom considered.

Several studies of coho salmon which utilized observation error-only models have detected compensation between the spawner and smolt stages (when juvenile abundance estimates are unavailable; Bradford et al., 2000; Barrowman et al., 2003). Others have investigated seasonal habitat selection during the juvenile stage in the OCN region, and have suggested that compensation was likely to be occurring during this part of the life-cycle (Nickelson et al., 1992; Nickelson and Lawson, 1998). This was attributed to the limitation in the availability of suitable rearing habitat and the territorial behavior of juveniles at that time. Models of coho salmon in Lobster Creek generally support these inferences and suggest that the majority of compensation is that stream occurred by the time juveniles were surveyed in their first summer after emergence.

3.3.6 Detecting responses to perturbations

Habitat covariates were successfully incorporated into the state-space model, and detected increases in the capacity of juveniles and survival of smolts, despite the presence of substantial process error. Results largely agree with the inferences of Solazzi et al. (2000) who suggested that survival from the juvenile to smolt stages, and to a slightly lesser degree, the abundance of juveniles, benefited from additional rearing habitat. That the current study suggested a potential increase in the carrying capacity of juveniles with additional rearing habitat (although there was significant uncertainty around this difference), indicates the advantages of using a structural

modeling approach where various derived parameters, in this case carrying capacity, can be easily accommodated into testing. In contrast, Solazzi et al. (2000) only assessed changes in juvenile abundance which do not account for the size of the spawner population in the preceding year, which could add significant variation to the data being compared.

There are several other advantages of incorporating covariates or indicator variables into structural models over alternative, more *ad hoc* approaches. The former partitions variation between different sources, lessening the potential for random variation to be attributed to the 'effect' being tested. For example, survival (to be related to the covariate) is measured between estimates of latent variables in adjoining years rather than between raw, observed counts. Furthermore, habitat effects on multiple vital rates and abundance parameters can be investigated simultaneously due to the flexibility of the Bayesian state-space formulation, rather than applying multiple univariate tests with different response variables (Solazzi et al., 2000). Lastly it is difficult to envisage how traditional models/analyses relying on linear regression could detect relationships between compensation parameters (at the juvenile stage in this study for example) and covariates without undertaking an *ad hoc* two-step process.

3.3.7 Predictive advantages of life-history models

When juvenile and/or smolt data were excluded from the dataset used for modeling, large differences in predictive performance were observed using a cross validation approach. Traditional stock-recruitment models are most closely related to the spawner-only situation of model M_{NNB}, although they generally assume only observation error in the response variable, and most often assume a Ricker instead of a Beverton-Holt model. Notably this was the worst performing of the 'exclusion' models tested, exhibiting the highest values for all summary statistics in the cross validation exercise, suggesting that including juvenile and smolt data in models has substantial benefits for prediction. Interestingly, the consequences of excluding only juvenile data (M_{NBL}), and only smolt data (M_{BNL}), were similar and it is difficult to address which would be more suitable to include in monitoring programs if funding necessitated making a choice between them.

It could be argued that measuring smolt abundance allows survival over the entire life-cycle to be partitioned into freshwater and oceanic components most effectively, although the estimation of density dependence is extremely important for determining where in the life-cycle management can be directed to release the population from limitation, for example. In the case of the latter, measuring abundance close to the timing of density dependence seems sensible and so monitoring of juvenile abundance may be desirable.

These inferences relate solely to the predictive performance of models with respect to spawner abundance. For example, there is a strong possibility that the predictive performance of M_{NBL} would be considerably worse than M_{BLL} if the cross validation was performed at the smolt stage instead. In addition, there are other benefits to including data from all stages such as the ability to reliably partition survival between different stages and identifying the true location of compensation. Which stages are most important to monitor will depend on the objectives of the study. The modeling can provide some guidance however, as the results indicate that juvenile and smolt data can both improve prediction of spawner abundance, and their availability allowed the investigation of a range of hypotheses such as the effects of various factors during different parts of the life-cycle that would otherwise be impossible.

3.3.8 Conclusions

The state-space life-history model for the Lobster Creek dataset demonstrated the flexibility of the approach, especially when fitted within a Bayesian framework. In general these models can be modified to test a wide variety of hypotheses, with the main constraint being whether all parameters in the model are identifiable. For the purposes of the Lobster Creek model this was achievable, and despite relatively high levels of uncertainty, the effects of habitat modification on coho salmon dynamics appear detectable. This dataset is unusual however in that monitoring was undertaken to high standards over a long time frame (Lorion, 2011). Datasets for the wider OCN coho populations exhibit some challenging features, including much more complicated monitoring programs, which would require careful construction of state-space models. However,

the potential gains that can be made in inferential and predictive ability as displayed by the Lobster Creek model suggest that these efforts may be worthwhile.

Chapter 4

Multilevel fit of Bayesian state-space population dynamics models

4.1 Introduction

A major challenge when studying animal population dynamics using time-series of abundance data is the presence of multiple sources of uncertainty. In the past, analyses failed to partition uncertainty into its appropriate components, usually in response to meeting the assumptions of relatively crude statistical models (Polacheck, Hilborn, and Punt, 1993). Increasing sophistication of modeling techniques over the last two decades has lead to the present situation, where complex models of population dynamics incorporating several sources of stochasticity, are routinely fitted to data. Key to these developments have been the adoption of state-space models (Carlin, Polson, and Stoffer, 1992), which link sub-model/s describing the hidden population processes in the system (process models), with those describing how these hidden states have been observed (observation models), and integrated models (which generally include state-space structures) which allow the joint analysis of multiple sources of demographic and abundance data (Maunder, 2003).

These models have undoubtedly improved many aspects of inferences that are made from this type of data, including, superior estimation of demographic parameters (Besbeas et al., 2002, 2003; Thomas et al., 2005), in some cases parameters that would not otherwise be identifiable (Abadi et al., 2012), and more robust prediction (Clark, 2003). The costs are, firstly, that high dimensionality of the models make them difficult to fit, and generally precludes the use of tra-

ditional statistical software. For this reason, the flexibility of the Bayesian framework when fitting models using Markov chain Monte Carlo (MCMC) techniques has led to it being widely adopted by a growing number of analysts in statistical ecology, and especially for state-space models of population dynamics (Gimenez et al., 2009). Secondly, the presence of hierarchical structures makes selecting among a set of candidate models difficult, at least using an information theoretical-type (Akaike, 1973) approach. Using information theory, model selection typically involves trading off model fit and complexity, however for hierarchical models the latter is difficult to define, as what constitutes a 'likelihood' is not unique (Spiegelhalter et al., 2002). This is not to say that information theory offers the optimal solution to model choice, especially in light of alternative Bayesian techniques, such as Bayes factors (Kass and Raftery, 1995), reversible jump MCMC (RJMCMC) (Green, 1995), minimum message length (Wallace, 2005) and posterior predictive loss (Gelfand and Ghosh, 1998), which often implicitly account for model complexity. However, there is little consensus, even among statisticians, as to which is the best approach for certain situations. Many of the alternatives will not generally be considered by ecologists owing to the technical nature of their calculation, or practical problems in their implementation. Furthermore, there has been a recent strong emphasis placed on information theoretical approaches by both frequentist (Burnham and Anderson, 2002) and Bayesian (McCarthy, 2007; Gimenez et al., 2009) ecologists.

One variant of the information theoretic approach involves comparing models on the basis of the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). DIC is calculated by adding the mean posterior deviance from MCMC samples, $\overline{D(\boldsymbol{\theta})}$ (a measure of model adequacy), and a penalty function, the estimate of the effective number of parameters in the model, p_D (model complexity). The statistic p_D is itself calculated as $\overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$ where $D(\bar{\theta})$ is the deviance calculated at the posterior point estimate of the parameters (usually the mean) and is sometimes termed the 'plug-in' estimate.

DIC is well suited to the analysis of a wide range of models because the complexity of the model (as estimated by p_D) can be estimated even in the presence of hierarchical structures,

it is trivial to calculate from MCMC samples, and is an output from WinBUGS, the most utilized software for carrying out ecological analyses in a Bayesian framework. Unsurprisingly, it has quickly become popular in a number of applied disciplines (e.g. Berg, Meyer, and Yu, 2004; Wilberg and Bence, 2008; Shriner and Yi, 2009), including modeling animal population dynamics (e.g. Schaub et al., 2007; Abadi et al., 2012; Rhodes et al., 2011). In a recent review of integrated models in terrestrial ecology, three of thirteen studies undertaking integrated modeling in a Bayesian framework utilized DIC for selecting among models (Schaub and Abadi, 2011), with many additional ('non-integrated') state-space applications also adopting this approach (e.g. Holdo et al., 2009; Davies and Jonsen, 2011; Moore and Barlow, 2011; Zhou et al., 2011). Presumably, this number will increase markedly in years to come as state-space and integrated modeling continues to move away from investigations of statistical properties towards practical ecological applications (Schaub and Abadi, 2011).

Unfortunately, DIC has several contentious properties. Among others, questions have been raised about its theoretical justification, the choice of the 'plug-in' estimator and the fact it is only a point estimate, the sensitivity to reparameterization that can result from the use of $\bar{\boldsymbol{\theta}}$ as a plug-in, and critically, the choice that has to be made as to which parameters in the model are to be focused on (Dawid, 2002; Meng and Vaida, 2006; Smith, 2002; Spiegelhalter et al., 2002; Celeux et al., 2006). The latter is particularly relevant for models with hierarchical structures. For instance, different values of DIC can be calculated depending on what is considered the likelihood and what is considered prior information for the model, and consequently, what constitutes $\boldsymbol{\theta}$ in the above description. Indeed, Bolker (2008) predicted 'that ecologists will rapidly begin to misuse deviance information criterion (DIC), since it is computed automatically by WinBUGS and is the easiest method of Bayesian hierarchical model selection.' Herein it is established that this is already occurring for state-space models recently fitted to animal abundance data, where the calculation of DIC has been misplaced within the model.

The problem with naive use of DIC occurs when the differences in models being compared occur in the process model. The standard estimate of DIC is based on the fit of the latent states

to the observations of abundance, however substantial process errors can compensate for an inadequate process model. This may result in poor performance in selecting among models, and erroneous metrics of the distance of the best model from others in the set. This was recognized by Moore and Barlow (2011) who compared a suite of state-space models fitted to transect survey data for whales. They noted large process errors present in the system that resulted in similar fit between models being compared, and undertook an *ad hoc* approach to model selection by refitting the models with only observation error included, and then comparing them on the basis of DIC. Similarly, Schaub et al. (2012) also raise the issue of DIC being unsuitable for certain classes of hierarchical integrated models and again select among models based on simplified versions, before making inferences from the equivalent full model. This is clearly an undesirable situation, and so a more robust approach is presented here to deal with these problems.

This unreliability of DIC when selecting among models is similar to the problems encountered by Millar (2009) for overdispersed count data, and the random effects one-way analysis of variance example in Spiegelhalter et al. (2002). In both those cases the focus of the DIC was shifted from the data level to the next level up in the hierarchy by integrating out the data level random effects to obtain a marginalized likelihood. Here an analogous approach is suggested. While the potential of refocusing information criteria has been suggested elsewhere (Spiegelhalter et al., 2002; Meng and Vaida, 2006; Plummer, 2006; Millar, 2009, also see Vaida and Blanchard, 2005 for a classical treatment of the same issue), this approach does not appear to have been previously utilized in the current context. This may be because the notion of model focus is not as meaningful in state-space models due to their temporal component. Indeed, it is shown that the marginalized likelihood does not correspond to a true likelihood and hence can not be interpreted as the likelihood arising from a re-focused model. Nonetheless, the predictive criterion resulting from marginalization would appear ideally suitable for many situations where state-space models are applied, and it is suggested herein that it could potentially be interpreted as a partial likelihood in the sense that it utilizes only the relevant portion of the likelihood. Schaub et al. (2012) correctly recognize that the default calculation of DIC in WinBUGS is inappropriate

for state-space models with extra hierarchical structure beyond one level of stochasticity, in each of the process and observation models, however it is shown herein that marginalization will also often be required for application of standard state-space models. This suggests that for many of the occasions DIC has been previously used in these situations, it has been done so incorrectly (e.g. Schaub et al., 2007; Holdo et al., 2009; Rhodes et al., 2011; Abadi et al., 2012).

Several simple simulated examples are initially presented in this chapter to clarify situations when the calculation of DIC must be shifted. The process of calculating DIC at the higher level is demonstrated, and marginalization is carried out to illustrate the consequences of misuse of the standard form of DIC. The results of the simple examples are then extended to consider the dataset of abundance of a coho salmon population in Lobster Creek, Oregon. Datasets are then simulated to represent realistic dynamics in this system, and a candidate set of models based on those considered in practice are fitted to each dataset to compare the discriminatory behavior of DIC calculated at multiple levels. The results of these simulations will provide a basis for which the utility of DIC in selecting among state-space population dynamics models can be assessed, for not only the Lobster Creek situation but also other applications of these models.

4.2 Methods

4.2.1 Calculation of DIC

A detailed outline of the derivation of DIC and its properties is given by (Spiegelhalter et al., 2002), however the details relevant to this study are presented below. The deviance of a model with parameters $\boldsymbol{\theta}$ is defined as

$$D(\boldsymbol{\theta}) = -2 \log p(\boldsymbol{y}|\boldsymbol{\theta}) + 2 \log l(\boldsymbol{y})$$

where $\boldsymbol{y}$ denotes the observed data and $l(\boldsymbol{y})$ is a standardizing function. As $l(\boldsymbol{y})$ is a function of the data alone, and is common to all models, it is ignored for the purposes of model selection

and the discussion below. DIC is calculated as

$$\begin{aligned} \text{DIC} &= 2\overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) \\ &= \overline{D(\boldsymbol{\theta})} + p_D \end{aligned} \tag{4.1}$$

where p_D is the effective number of parameters calculated as $p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$, with $\overline{D(\boldsymbol{\theta})}$ being the posterior mean deviance, and $D(\bar{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior estimates (often the means) of the parameters $\boldsymbol{\theta}$. Using this formulation p_D can be interpreted as a measure of the difficulty of fitting the model, or the reduction in uncertainty attributable to estimation.

4.2.2 Structure of state-space models

At its most basic, a state-space model of animal abundance describes the dynamics of two linked processes. These are the unobserved dynamics of the system, which have a time-series component (the state equation, or process model), and the dynamics of how these unobserved states are linked to observed data via the sampling processes (the observation equation, or observation model). An example of such a model (previously presented in Section 1.8) is given by three equations

$$n_t = f(n_{t-1}, \boldsymbol{\psi}) + \epsilon_t \ , \quad \text{for} \quad t = 1, 2, ..., T \tag{4.2}$$

$$y_t = g(n_t, \boldsymbol{\xi}) + \nu_t \ , \quad \text{for} \quad t = 1, 2, ..., T \tag{4.3}$$

$$n_0 \sim h$$

where n_t is the latent state and y_t is the observation, both in year t, and $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ are vectors of model parameters. In general the states and observations could be multidimensional, but herein they will be univariate and correspond to the true abundance of an animal and its observed abundance, respectively. Functions f and g represent the deterministic components of the process and observation models, and ϵ_t and ν_t are the process and observation errors, respectively. Often ϵ_t and ν_t are assumed to be normally distributed with mean zero and variances τ^2 and σ^2, which are denoted the process and observation error variances, respectively. In (4.2) and (4.3) errors are assumed to be additive but, in practice, could take a multitude of forms. The prior

distribution, h, is placed on the initial state n_0 under a Bayesian framework. To keep future notation concise the deterministic component of the process model will at times be denoted $\tilde{n}_t$, i.e. $\tilde{n}_t = f(n_{t-1}, \boldsymbol{\psi})$.

4.2.3 Focus of DIC for state-space models

4.2.3.1 Standard DIC for state-space models

The joint density function (of data and parameters) for a Bayesian state-space model with normally distributed process and observation errors is easily obtained as the product of terms from sequentially conditioning on the elements of the sequence $(n_0, n_1, y_1, n_2, y_2, ..., n_t, y_t)$, whereby each element is conditioned on all preceding elements. Letting $n_{[-t]}$ denote all elements in the sequence preceding n_t (and similarly for $y_{[-t]}$), it is the case that

$$p(n_t|n_{[-t]}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) = p(n_t|n_{t-1}, \boldsymbol{\psi}, \tau^2)$$
$$p(y_t|y_{[-t]}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) = p(y_t|n_t, \boldsymbol{\xi}, \sigma^2) \ .$$

The joint density function is therefore

$$p(\boldsymbol{y}, \boldsymbol{n}, n_0, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) = p(\boldsymbol{y}|\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2)p(\boldsymbol{n}|n_0, \boldsymbol{\psi}, \tau^2)p(n_0, \boldsymbol{\psi}, \boldsymbol{\xi}, \tau^2, \sigma^2) \ , \tag{4.4}$$

where $\boldsymbol{n} = (n_1, ..., n_t)$, and

$$p(\boldsymbol{y}|\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2) = \prod_{t=1}^{T} p(y_t|n_t, \boldsymbol{\xi}, \sigma^2) \ , \tag{4.5}$$

and

$$p(\boldsymbol{n}|n_0, \boldsymbol{\psi}, \tau^2) = \prod_{t=1}^{T} p(n_t|n_{t-1}, \boldsymbol{\psi}, \tau^2) \ .$$

DIC of state-space models is currently implemented using the data likelihood from (4.5), and is denoted DIC_d herein.

4.2.3.2 Partial DIC for state-space models

It is much less natural, but nonetheless still valid, to obtain the joint density function in (4.4) by sequentially conditioning on the elements of the sequence $(n_0, y_1, n_1, y_2, n_2, ..., y_t, n_t)$ in that

order. Then, it is the case that

$$p(n_t|n_{[-t]}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) = p(n_t|n_{t-1}, y_t, \boldsymbol{\psi}, \tau^2)$$

$$p(y_t|y_{[-t]}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) = p(y_t|n_{t-1}, \boldsymbol{\xi}, \sigma^2) \ .$$

Hence

$$p(\boldsymbol{y}, \boldsymbol{n}, n_0, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\psi}, \tau^2) = p^*(\boldsymbol{y}|\boldsymbol{n}, \boldsymbol{\xi}, \sigma^2)p^*(\boldsymbol{n}|n_0, \boldsymbol{y}, \boldsymbol{\psi}, \tau^2)p(n_0, \boldsymbol{\psi}, \boldsymbol{\xi}, \tau^2, \sigma^2) \ ,$$

where

$$p^*(\boldsymbol{y}|\boldsymbol{n}, n_0, \boldsymbol{\xi}, \sigma^2) = \prod_{t=1}^{T} p(y_t|n_{t-1}, \boldsymbol{\xi}, \sigma^2) \ , \tag{4.6}$$

and

$$p^*(\boldsymbol{n}|n_0, \boldsymbol{y}, \boldsymbol{\psi}, \tau^2) = \prod_{t=1}^{T} p(n_t|n_{t-1}, y_t, \boldsymbol{\psi}, \tau^2) \ . \tag{4.7}$$

It is proposed herein that the density in (4.6) be used to implement DIC. This term has the intuitive appearance of a partially marginalized likelihood and the associated DIC will be denoted DIC_m. However, it is important to note that it has no formal interpretation as a likelihood because $\boldsymbol{y}$ are also contained in (4.7). It is for this reason that (4.6) is termed a partial likelihood herein. Similarly, it would not be appropriate to say that the focus of the deviance has been changed.

It is interesting to take a closer look at (4.7), which is ignored by using partial deviance. It could be argued that this term represents the degree to which the process equation has been altered by having observed y_t. This is precisely the concern that was raised in Section 4.1, namely that the standard DIC can be compromised when process errors compensate for poor deterministic process equations. Therefore, the level at which calculation of DIC is most suitable is dependent on context, although it will be shown that DIC_m is more suitable for situations where models in a candidate set differ in the structure of their process model, which is frequently the case in the applications of these models (e.g. Schaub et al., 2007; Holdo et al., 2009; Abadi et al., 2012; Moore and Barlow, 2011; Rhodes et al., 2011).

4.2.4 Testing criterion with simulation

Herein, the general approach to testing the performance of the two forms of DIC is to simulate replicate datasets from a known model (which will be referred to as the true model), fit a set of candidate models that includes the true model and one or more alternative models, to the data, and observe the ability of DIC_d and DIC_m in identifying the true, data-generating model.

4.2.4.1 Poisson-gamma example

This example represents the situation where an animal population occurs at low abundance. In these cases sampling programs typically result in count data that are incorporated within a state-space model by assuming a discrete observation model based on Poisson or negative binomial sampling distributions for example (e.g. Abadi et al., 2012). Process models can take many forms and typically involve both deterministic and stochastic elements with numerous choices to be made about the structure of each, depending on the dynamics of the population and the aims of the study. In light of this, an example is presented here to represent an unstructured population at low abundance, with a count-based sampling process.

The observation and process models are assumed to follow Poisson and gamma distributions, respectively, which makes for simple analytical calculation of the partially marginalized DIC while also providing a similar structure (albeit, much simplified) to models that may be considered in practice. These equations are

$$y_t|n_t \sim \text{Pois}(n_t) \ , \quad t = 1, 2, ..., T \ ,$$

and

$$n_t|\alpha, n_{t-1}, \boldsymbol{\beta} \sim \text{Gamma}(\alpha, \delta_t(\alpha, n_{t-1}, \boldsymbol{\beta})) \ , \quad t = 1, 2, ..., T \tag{4.8}$$

where y_t and n_t are again the observed and latent abundances, and α and $\delta_t(\alpha, n_{t-1}, \boldsymbol{\beta})$ are the shape and rate parameters of the gamma density, respectively. The latter are calculated $\delta_t(\alpha, n_{t-1}, \boldsymbol{\beta}) = \frac{\alpha}{\tilde{n}_t}$, where $\tilde{n}_t$ is the deterministic component of the process model, and $\boldsymbol{\beta}$ are parameters in this deterministic component. Three competing models were considered for the

dynamics of $\tilde{n}_t$, the Ricker (M^R; (4.9)), Gompertz (M^G; (4.10)) and an exponential growth model (M^E; (4.11))

$$\tilde{n}_t = \beta_0 n_{t-1} \exp(-\beta_1 n_{t-1}) \tag{4.9}$$

$$\tilde{n}_t = \beta_0 n_{t-1} \exp(-\beta_1 \log n_{t-1}) \tag{4.10}$$

$$\tilde{n}_t = \beta_0 n_{t-1} \tag{4.11}$$

where β_0 is the maximum population growth rate and β_1 is a compensation parameter that quantifies the reduction in the rate of population growth due to density dependent processes (limited food etc.).

To test the performance of DIC at different levels of the hierarchy the following algorithm was implemented; 1) simulate a vector of latent states ($\boldsymbol{n}^*$) from the Ricker process model, where the '*' denotes that they are simulated values (4.8) and (4.9), 2) simulate a new vector of data ($\boldsymbol{y}^*$) from $\sim$ Poisson($\boldsymbol{n}^*$), 3) fit the three competing models to the simulated dataset, 4) calculate DIC at both the data and partially marginalized levels of the hierarchy, and store, 5) repeat the process N times and calculate summary statistics. Parameter values used for the simulations were set at $\beta_1 = 0.005$, $n_1 = 10$, and $T = 30$, for all scenarios, while the values of α and β_0 were varied. Four scenarios were considered, corresponding to all combinations of values of $\alpha = \{50, 80\}$ and $\beta_0 = \exp\{0.3, 0.6\}$ to allow different population dynamics, and levels of observation and process error in the simulated data. For each scenario $N = 250$ simulations were carried out.

Data were generated in the statistical software package R 2.14.0 (R Development Core Team 2010) and the competing models were fitted in OpenBUGS 3.0.3 (Lunn et al., 2009) which were linked using the packages `R2WinBUGS` and `BRugs`. Prior distributions were specified as $n_0 \sim \text{LN}(2.5, 0.25)$, $\log \beta_0 \sim \text{N}(0, 10\,000)$, and $\beta_1 \sim \text{Uniform}(-5, 5)$ for all scenarios. The prior for α was specified as $\alpha \sim \text{LN}(3.867, 0.09)$ and $\alpha \sim \text{LN}(4.337, 0.09)$ when scenarios involved true values of α of 50 and 80, which are moderately informative, and allow convergence of the sampler for all model parameters. For each simulated dataset each model was fitted by running

two MCMC chains for 40 000 iterations, with samples stored for 1 in every 2 samples after a burnin of 20 000, resulting in a total of 20 000 samples available for inferences. Convergence typically occurred much faster than 20 000 iterations but this burnin was specified conservatively to ensure that convergence occurred before storage of samples for all simulated datasets. The total number of iterations and thinning rate were established to ensure that the effective number of samples stored was greater than 1 000 for all parameters.

The identical processes were repeated for another set of scenarios though with datasets generated from M^E, rather than than M^R. Parameter values were kept the same except that β_0 took on the values $\beta_0 = \{0.05, 0.09\}$ and β_1 is not used in this generation model. The simulation process was not repeated with data generated from M^G because that model is very similar to M^R and results were expected to be consistent.

The data level density is the default choice of DIC in WinBUGS and for the Poisson-gamma state space model this is given by

$$p(y_t|n_t) \sim \text{Poisson}(n_t)$$

and, consequently the deviance at this level is calculated

$$D_d = -2\sum_{t=1}^{T} \log p(y_t|n_t) = -2\sum_{t=1}^{T} y_t \log n_t - n_t - \log y_t! \; . \tag{4.12}$$

The deviance D_d can then be used in (4.1) to calculate DIC_d.

The evaluation of DIC_m for the Poisson-gamma model requires one-dimensional integration to be able to calculate (4.6) which can be achieved analytically and results in a partial deviance based on the negative-binomial distribution as follows

$$\begin{aligned} D_m &= -2\sum_{t=1}^{T} \log p(y_t|n_{t-1}, \alpha, \boldsymbol{\beta}) \\ &= -2\sum_{t=1}^{T} \log \int_0^{\infty} p(y_t|n_t) p(n_t|n_{t-1}, \alpha, \boldsymbol{\beta}) \mathrm{d}n_t \\ &= -2\sum_{t=1}^{T} \log \left[\frac{(y_t + \alpha - 1)!}{y_t!(\alpha - 1)!} \left(\frac{\alpha}{\tilde{n}_t + \alpha} \right)^{\alpha} \left(\frac{\tilde{n}_t}{\tilde{n}_t + \alpha} \right)^{y_t} \right] \end{aligned} \tag{4.13}$$

where $\tilde{n}_t$ is calculated from (4.9)–(4.11). The partial DIC is then given by using D_m in the calculation of (4.1).

It must be noted that there are different parameterizations that can be used when calculating $D(\bar{\boldsymbol{\theta}})$ from (4.12) and (4.13). In WinBUGS the default method is to use the posterior means of the stochastic parents of $\boldsymbol{n}$ and $\tilde{\boldsymbol{n}}$, if they exist. In the case of the latter this would involve the posterior means of $(n_{t-1}, \alpha, \boldsymbol{\beta})$. A more natural approach, and one that appears to be more stable, is to calculate the posterior means of $\tilde{\boldsymbol{n}}$ directly and use them in (4.13) to calculate D_m, and that is the technique utilized for all models considered herein.

4.2.4.2 Log-normal example

State-space models of very abundant species often employ log-normal observation and process models (e.g. Meyer and Millar, 1999). Using similar notation to the Poisson-gamma example, and undertaking the calculations on the log-scale, the observation and process models are given by

$$\log y_t | \log n_t, \sigma^2 \sim \mathrm{N}(\log n_t, \sigma^2) \ , \quad t = 1, 2, ..., T \ , \tag{4.14}$$

$$\log n_t | \log n_{t-1}, r, K, \tau^2 \sim \mathrm{N}(\log \tilde{n}_t, \tau^2) \ , \quad t = 1, 2, ..., T \tag{4.15}$$

where σ^2 and τ^2 are the observation and process error variances, respectively, and the deterministic component of the process model, $\log \tilde{n}_t$, was considered to be derived from either of two competing functions. The parameterization of the discrete logistic model (M^L) commonly encountered in terrestrial ecology is given by

$$\log \tilde{n}_t = \log \left[n_{t-1} + r n_{t-1} \left(1 - \frac{n_{t-1}}{K} \right) \right]$$

where K and r are parameters representing the carrying capacity and intrinsic maximum population growth rate, respectively. An alternative model, denoted the logistic-stepped model (M^S), accommodates a change in carrying capacity part way through the time-series to represent, for instance, a sudden increase in a food source that was previously limiting population growth.

This model is formulated

$$\log \tilde{n}_t = \log \left[n_{t-1} + r n_{t-1} \left(1 - \frac{n_{t-1}}{K\,(1+\gamma I_t)} \right) \right] \tag{4.16}$$

where I_t is an indicator variable taking the value of 1 or 0 in year t when the factor releasing the population from limitation is present or absent, respectively, and γ is a parameter determining the strength of the relationship between the carrying capacity and the indicator variable.

The same simulation algorithm as the Poisson-gamma example was utilized. Datasets were simulated from M^S ((4.14), (4.15) and (4.16)), and both competing models were then fitted to each dataset. Parameter values used for the simulations were $K = 50\,000$, $r = 0.4$, $n_1 = 10\,000$, $\gamma = 0.5$, $I_t = 0$ for $t = 1, 2, ..., 25$ and 1 for $t = 26, 27, ..., 40$, $T = 40$ and $\sigma = 0.1$ for all scenarios, while τ was varied over the set $\tau = \{0.02, 0.07, 0.1, 0.15, 0.2\}$, resulting in five scenarios that had a range of process to observation error ratios. Prior distributions for the parameters in the models were; $K \sim \text{Gamma}(50, 0.001)$, $r \sim \text{Uniform}(0, 1)$, $\log n_0 \sim \text{N}(9, 4)$, $\gamma \sim \text{N}(0, 4)$, $\sigma \sim \text{LN}(-2.548, 0.49)$ and $\tau \sim \text{Uniform}(0, 10)$. In these types of models identifiability of parameters such as the variance components is often compromised unless prior information is available. Several of these prior distributions are moderately informative to ensure robust estimates of posterior distributions. This would correspond to situations in practice where prior information is received from past studies of the same, or congeneric, species etc. For each scenario, 250 simulations were performed.

Another set of scenarios was investigated where datasets were simulated from M_L and both competing models were again fitted and compared. Parameter values for five scenarios were identical to above except that parameter γ was absent from the simulation model.

The data level density for this example is

$$p(\log y_t | \log n_t, \sigma^2) \sim \text{N}(\log n_t, \sigma^2)$$

and the corresponding deviance is therefore

$$\begin{aligned} D_d &= -2\sum_{t=1}^{T} \log p(\log y_t | \log n_t, \sigma^2) \\ &= T\log(2\pi\sigma^2) + 2\sum_{t=1}^{T} \frac{(\log y_t - \log n_t)^2}{2\sigma^2} \end{aligned}$$

Using D_d in (4.1) allows the calculation of DIC_d.

The marginalization required to calculate the partial deviance is explicit in this case since the observation and process equations are both linear and normally distributed. Specifically, $y_t|n_{t-1}$ is normal with mean $\log \tilde{n}_t$ and variance $\tau^2 + \sigma^2$. The partial deviance is therefore

$$\begin{aligned} D_m &= -2\sum_{t=1}^{T} \log p(\log y_t | \log \tilde{n}_t, \sigma^2, \tau^2) \\ &= -2\sum_{t=1}^{T} \log \int_{-\infty}^{\infty} p(\log y_t | \log n_t, \sigma^2) p(\log n_t | \log \tilde{n}_t, \tau^2) \mathrm{d}\, n_t \\ &= T\log\left(2\pi\left(\sigma^2 + \tau^2\right)\right) + 2\sum_{t=1}^{T} \frac{(\log n_t - \log \tilde{n}_t)^2}{2(\sigma^2 + \tau^2)} \ . \end{aligned} \tag{4.17}$$

Using D_m from (4.17) in (4.1) allows the calculation of DIC_m.

4.2.5 Lobster Creek coho salmon example

4.2.5.1 Evaluation of DIC on an empirical dataset

A set of candidate models which differed only in the form of the deterministic components of their process equations were fitted to the Lobster Creek dataset, with the details of their structure presented in Sections 2.2.2–2.2.3 and Table 2.1. The necessary details are repeated here. The observation model for this example was

$$\log y_{i,j,t} | \log n_{i,j,t}, \sigma^2_{i,j,t} \sim N(\log n_{i,j,t}, \sigma^2_{i,j,t}) \ , \qquad i = 1,2,3,\ j = 1,2$$

where $\log y_{i,j,t}$ are the observed log-abundances, $\log n_{i,j,t}$ are the latent states, and $\sigma^2_{i,j,t}$ are observation error variances, each for stage-class i, in stream j, in year t. It should be noted that observations of abundance for the juvenile and smolt stages ($y_{1,j,t}$ and $y_{2,j,t}$) were only available for $t = 3, 4, ..., 26$, while observations of abundance for spawners ($y_{3,j,t}$) were only available for $t = 1, 2, ..., 25$ (Sections 2.2.1.4–2.2.1.6). The variances $\sigma^2_{i,j,t}$ were derived from the sampling

variances estimated as part of the monitoring programs (see Sections 2.2.1.4–2.2.1.6) and these were included in the model as constants rather than parameters to be estimated (e.g. King et al., 2008).

The general process model can be represented by

$$\log n_{i,j,t} = f_i(\log n_{i-,j,t-1}, \boldsymbol{\psi}_i) + \epsilon_{i,j,t} \ , \quad i = 1, 2, 3, \ j = 1, 2$$

where it must again be noted that $i-$ denotes the preceding stage-class instead of the more natural $i-1$, as $n_{1,t}$ is a function of $n_{3,t-1}$, not $n_{0,t-1}$ as would be suggested by the latter. The f_i are stage-specific functions representing the deterministic processes (with the stage-class-specific vectors of parameters $\boldsymbol{\psi}_i$) linking abundance in each stage-class to abundance in the preceding stage-class in the preceding year, and $\epsilon_{i,j,t}$ are the process errors. The process equations were modeled over different time periods depending on the presence of observations in the previous stage-class. For juveniles this period was $t = 3, 4, ..., 26$, for smolt it was $t = 4, 5, ..., 26$ and for spawners it was $t = 4, 5, ..., 25$. The process errors were in turn modeled

$$\epsilon_{i,j,t} = \omega_{i,t} + \varpi_{i,j,t} \ , \quad \omega_{i,t} \sim N(0, \tau_i^2) \ , \quad \varpi_{i,j,t} \sim N(0, \delta_i^2) \tag{4.18}$$

where $\omega_{i,t}$ are the temporal, and the $\varpi_{i,j,t}$ are the spatial components of process error, each with a mean of zero and stage-class-specific variances τ_i^2, and δ_i^2, respectively. The deterministic components of the process model, denoted $\log \tilde{n}_{i,j,t}$, were calculated $\log \tilde{n}_{i,j,t} = f_i(\log n_{i-1,j,t-1}, \boldsymbol{\psi}_i)$. The details of the structure of $f_i(\log n_{i-1,j,t-1}, \boldsymbol{\psi}_i)$ for each individual model considered are given in Table 2.1.

Although the Lobster Creek example has additional stage-classes, and more complicated process models, it is fundamentally similar to the log-normal example (Section 4.2.4.2) and this is reflected in the similarities when calculating DIC_d and DIC_m. The data level density for these models was

$$p(\log y_{i,j,t} | \log n_{i,j,t}, \sigma_{i,j,t}^2) \sim \mathrm{N}\left(\log n_{i,j,t}, \sigma_{i,j,t}^2\right)$$

and, consequently the deviance at this level is calculated

$$\begin{aligned} D_d &= -2\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{t=u_i}^{U_i}\log p\left(\log y_{i,j,t}|\log n_{i,j,t},\sigma^2_{i,j,t}\right) \\ &= 2\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{t=u_i}^{U_i}\log\left(\sqrt{2\pi\sigma^2_{i,j,t}}\right)+\frac{(\log y_{i,j,t}-\log n_{i,j,t})^2}{2\sigma^2_{i,j,t}} \end{aligned} \tag{4.19}$$

where $\boldsymbol{u}=(u_1,u_2,u_3)=(3,3,1)$ as estimates of abundance for juveniles and smolt only began in 1988 ($t=3$), and $\boldsymbol{U}=(U_1,U_2,U_3)=(26,26,25)$ as no estimates of spawner abundance for 2011 ($t=26$) were yet available at the time of analysis. Again, DIC_d is calculated using D_d in (4.1).

Calculating the partial deviance again requires marginalization, which is very similar to the log-normal example, and is thus available in closed form. The partial deviance is consequently calculated by

$$\begin{aligned} D_m &= -2\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{t=u_i}^{U_i}\log p(\log y_{i,j,t}|\log n_{i-1,j,t-1},\boldsymbol{\psi}_i,\sigma^2_{i,j,t},\tau^2_i,\delta^2_i) \\ &= -2\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{t=u_i}^{U_i}\log\int_{-\infty}^{\infty}p(\log y_{i,j,t}|\log n_{i,j,t},\sigma^2_{i,j,t})p(\log n_{i,j,t}|\log n_{i-1,j,t-1},\boldsymbol{\psi}_i,\tau^2_i,\delta^2_i)\,\mathrm{d}\,n_{i,j,t}\ . \end{aligned} \tag{4.20}$$

To evaluate (4.20) it is useful to note that the process errors in (4.18) can be reparameterized in the form of a bivariate normal distribution for the two streams, and so after partial marginalization a bivariate normal distribution will also result, and is given by

$$p(\log y_{i,j,t}|\log n_{i-1,j,t-1},\boldsymbol{\psi}_i,\sigma^2_{i,j,t},\tau^2_i,\delta^2_i)\sim \mathrm{N}(\boldsymbol{\mu}_{i,t},\boldsymbol{\Sigma}_{i,t})$$

where the mean vector, $\boldsymbol{\mu}_{i,t}$, and the covariance matrix, $\boldsymbol{\Sigma}_{i,t}$, for stage i and year t, are given by

$$\boldsymbol{\mu}_{i,t}=\begin{pmatrix}\log\tilde{n}_{i,1,t}\\ \\ \log\tilde{n}_{i,2,t}\end{pmatrix},\quad\text{and}\quad\boldsymbol{\Sigma}_{i,t}=\begin{pmatrix}\kappa^2_{i,1,t} & c_{i,t}\\ \\ c_{i,t} & \kappa^2_{i,2,t}\end{pmatrix}$$

and

$$\kappa_{i,1,t}^2 = \sigma_{i,1,t}^2 + \tau_i^2 + \delta_i^2$$

$$\kappa_{i,2,t}^2 = \sigma_{i,2,t}^2 + \tau_i^2 + \delta_i^2$$

$$c_{i,t} = \rho \times \kappa_{i,1,t} \times \kappa_{i,2,t} = \tau_i^2$$

$$\rho_{i,t} = \frac{\tau^2}{\kappa_{i,1,t} \times \kappa_{i,2,t}} \ .$$

Consequently, the partially marginalized deviance is given by

$$D_m = -2 \sum_{i=1}^{3} \sum_{t=u_i}^{U_i} \log \left(\frac{1}{2\pi\kappa_{i,1,t}\kappa_{i,2,t}\sqrt{1-\rho_{i,t}^2}} \exp\left(-\frac{1}{2(1-\rho_{i,t}^2)} \left[\frac{(\log y_{i,1,t} - \log \tilde{n}_{i,1,t})^2}{\kappa_{i,1,t}^2} + \right. \right. \right.$$
$$\left. \left. \left. \frac{(\log y_{i,2,t} - \log \tilde{n}_{i,2,t})^2}{\kappa_{i,2,t}^2} - \frac{2\rho_{i,t}(\log y_{i,1,t} - \log \tilde{n}_{i,1,t})(\log y_{i,2,t} - \log \tilde{n}_{i,2,t})}{\kappa_{i,1,t}\kappa_{i,2,t}} \right] \right) \right) \tag{4.21}$$

where $\boldsymbol{u} = (3, 4, 4)$ and $\boldsymbol{U} = (26, 26, 25)$. Inserting D_m into (4.1) results in the calculation of DIC_m.

Note that DIC_m is calculated over a slightly more restricted range when using (4.21), as in practice the process model is typically initiated in the year following the first observation in the time-series (e.g. Schaub et al., 2007; Abadi et al., 2012), and this was the case in Chapter 2. For completeness DIC_d was calculated over the restricted range used in (4.21) for the calculation of DIC_m, in addition to the default calculation in (4.19). However, the relative comparison of DIC_d and DIC_m was almost identical for both metrics, and consequently only the latter is presented here.

4.2.5.2 Lobster Creek simulations

The simulation process used model M_{BLL} as the generation model and followed the same algorithm as adopted in the previous examples. Data were generated from M_{BLL} by simulating 250 datasets directly from the approximate posterior predictive distributions stored in the model fitting in Chapter 2. The set of candidate models, including the data-generating model, were fitted to each dataset, and models were compared using both DIC_d and DIC_m.

Only three alternative models were considered owing to the considerable computational demands of the repeated application of MCMC on simulated datasets. The set was constructed to include examples of models that differed by small and large amounts from the true model, to compare the discriminatory behavior of DIC in these situations. These included models where instead of a Beverton-Holt function there was a Ricker (M_{RLL}), or linear (M_{LLL}) function at the juvenile stage, and a model which was identical to M_{BLL} except that survival from the juvenile to smolt stages was modeled as a function of rearing habitat (M_{BL_hL}; see Section 2.2.2.3 for more details of these models, including the priors used).

4.3 Results

4.3.1 Poisson-gamma example

The proportion of simulations where M^R was correctly selected over the alternatives varied significantly between scenarios. In general DIC struggled to distinguish between M^R and M^G at both the data and partially marginalized levels when $\log \beta_0 = 0.3$ (Table 4.1), although DIC_m always performed better than DIC_d, and when $\log \beta_0 = 0.6$ it nearly always correctly selected M^R. DIC_m correctly identified model M^R as superior to M^E in all simulations, across all scenarios (Table 4.1). Similar behavior was exhibited by DIC_d when $\log \beta_0 = 0.6$, however M^E was incorrectly selected in a moderate proportion of simulations when $\log \beta_0 = 0.3$ (P=0.06–0.11).

At the data level, p_D increased when $\log \beta_0$ was raised from 0.3 to 0.6, and decreased slightly when α was raised from 50 to 80 (Table 4.1). At the partially marginalized level, p_D decreased slightly when $\log \beta_0$ was raised from 0.3 to 0.6, but remained similar when α was 50 or 80. The difference in p_D between the exponential growth and the other models was never consistently close to the expected value of one less for the former at either the data or partially marginalized level.

Mixed results were exhibited when datasets were generated from M^E. In general, DIC at both levels struggled to correctly select M^E over the alternative models. However, DIC_m still always performed better at correctly selecting the true model over the others, with P ranging

Table 4.1: Performance of DIC_d and DIC_m in correctly detecting the true (data generating model) as more parsimonious than two alternative models, across four scenarios where data were generated from models with different values of the parameters $\log \beta_0$ and/or α. The top and bottom panels show results for scenarios where data was simulated from M^R and M^E respectively.

				Data Level				Partially Marginalized			
Generating Mod	$\log \beta_0$	α	Fitted Model	$\overline{D(\theta)}$	p_D	DIC_d	P	$\overline{D(\theta)}$	p_D	DIC_m	P
M^R	0.3	50	M^R	196.6	14.4	211.0		216.4	6.2	222.6	
			M^G	197.1	14.1	211.2	0.60	217.2	6.7	223.8	0.87
			M^E	198.2	14.0	212.2	0.89	221.1	7.4	228.6	1.00
	0.3	80	M^R	197.6	12.2	209.7		212.1	6.0	218.2	
			M^G	198.1	11.9	210.1	0.68	213.0	6.6	219.5	0.90
			M^E	199.5	12.2	211.7	0.94	217.9	7.5	225.4	1.00
	0.6	50	M^R	220.2	19.5	239.6		253.0	4.6	257.6	
			M^G	221.8	18.9	240.7	0.87	256.9	5.5	262.3	0.97
			M^E	223.8	19.8	243.6	1.00	270.6	6.0	276.6	1.00
	0.6	80	M^R	220.6	17.1	237.7		245.4	4.6	250.0	
			M^G	222.9	16.5	239.4	0.90	249.8	5.7	255.5	0.99
			M^E	225.5	18.2	243.7	1.00	266.7	6.4	273.1	1.00
M^E	0.09	50	M^E	190.5	12.4	202.9		209.2	7.2	216.4	
			M^R	190.3	12.7	203.0	0.58	209.3	7.5	216.9	0.76
			M^G	190.4	12.6	203.1	0.56	209.5	7.5	217.0	0.76
	0.09	80	M^E	188.9	10.3	199.2		202.1	7.1	209.2	
			M^R	188.8	10.5	199.3	0.66	202.3	7.5	209.8	0.78
			M^G	188.9	10.5	199.4	0.68	202.3	7.4	209.7	0.72

Notes: Parameter estimates are means over the 250 simulated datasets for each scenario. P is calculated as the proportion of simulations in which the Ricker model was correctly detected as better than the model in that row. All standard errors for $\overline{D(\theta)}$, p_D and DIC were less than 0.5, 0.1 and 0.5 respectively.

between 0.72–0.78, in contrast to 0.56–0.68 for DIC_d. Estimates of p_D were more similar across all models at the data level than at the partially marginalized level, where p_D for M^R and M^G were ~ 0.5 higher than for M^E.

4.3.2 Log-normal example

Performance of DIC_d was worse than DIC_m across all scenarios when data were generated from M^L, though particularly when the simulated ratio of process to observation error was lowest (Table 4.2). The data generating M^S was actually selected against in nearly all simulations (100% of simulations when $\tau = 0.02$ and 0.07) for these scenarios . In these situations the deterministic components of the process model explain the data poorly, resulting in process errors being inflated, relative to the true model. As the ratio of process to observation error variance determines the shrinkage of the $\boldsymbol{n}$ towards $\boldsymbol{y}$, the $\boldsymbol{n}$ will sometimes actually fit the data better for the alternative, rather than the true model. An example of this phenomenon is given in Figure 4.1.

As the ratio of process to observation error increased a number of trends occurred; the performance of DIC_d improved, the performance of DIC_m declined, the data level p_D for M^S increased, while for M^L it declined slightly, until p_D was almost equivalent when $\tau = 0.2$, p_D at the partially marginalized level declined slightly for both models, $\overline{D(\theta)}$ of M^S model decreased slightly at the data level but increased moderately for M^L, $\overline{D(\theta)}$ at the partially marginalized level increased for both models, though at a slightly greater rate for M^S (Table 4.2). The difference in p_D between models at the partially marginalized level remained relatively close to the expected difference of one for all scenarios, in contrast to the difference at the data level where it varied considerably.

Statistics were more similar between models when data were generated from M^L, however DIC_m still performed better ($P = 0.77 - 0.88$) than DIC_d ($P = 0.44 - 0.49$), across all scenarios (Table 4.3). The estimated p_D and DIC at the data level were very similar between M^S and M^L. At the partially marginalized level the difference in p_D was close to the expected value of

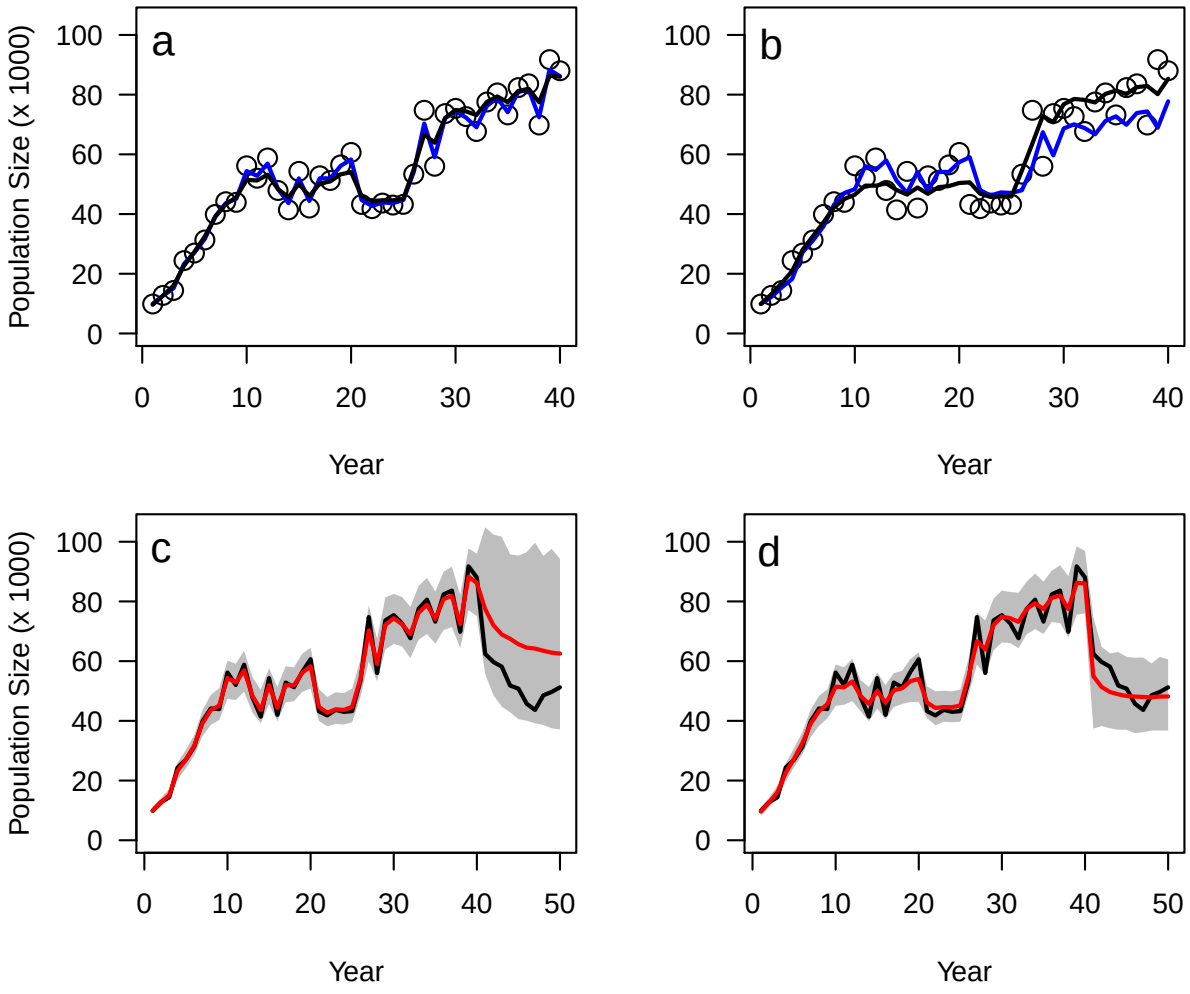

Figure 4.1: Example of a dataset simulated from the log-normal M^S under a scenario with $(\tau = 0.07, \sigma = 0.1)$. Open dots are the simulated observations ($\boldsymbol{y}^*$). Displayed are the posterior estimates of the latent states ($\boldsymbol{n}$; a) and the deterministic components of the process model ($\tilde{\boldsymbol{n}}$; b), for the two competing models, M^L (blue line) and M^S (black line). Panels c and d display the true simulated abundance (black lines) and the posterior means of the estimated latent states (red lines) and the limits of their 95% credible intervals (shaded areas), for M^L (c) and M^S (d) models, respectively. Models were fitted to the simulated observations for years 1–40 only and from years 41–50 represent model predictions of $\boldsymbol{n}$ beyond the fitting period. The indicator variable I_t was zero for the years 1–25 and 41-50, and one for years 26-40.

Table 4.2: Performance of DIC_d and DIC_m in correctly detecting M^S (data generating model) as more parsimonious than M^L, across five scenarios where data was generated from models with different values for the observation and process error standard deviations, σ and τ respectively.

			Data Level				Partially Marginalized			
σ	τ	Fitted Model	$\overline{D(\theta)}$	p_D	DIC_d	P	$\overline{D(\theta)}$	p_D	DIC_m	P
0.10	0.02	M^S	-85.3	11.3	-74.0		-66.7	7.8	-58.9	
		M^L	-116.2	21.9	-94.3	0	-47.6	7.2	-40.4	1.00
0.10	0.07	M^S	-88.1	14.2	-73.9		-53.0	8.2	-44.8	
		M^L	-111.7	21.6	-90.1	0	-46.2	7.0	-32.2	1.00
0.10	0.10	M^S	-93.9	16.6	-77.3		-41.0	8.1	-33.0	
		M^L	-108.3	21.2	-87.0	0.15	-31.9	6.8	-25.1	0.97
0.10	0.15	M^S	-96.4	18.4	-78.0		-22.8	7.4	-15.4	
		M^L	-103.4	20.9	-82.4	0.32	-23.0	6.3	-10.4	0.90
0.10	0.20	M^S	-95.7	19.1	-76.6		-6.2	6.9	0.7	
		M^L	-98.8	20.5	-78.3	0.46	-1.6	5.8	4.2	0.81

Notes: Parameter estimates are means over the 250 simulated datasets for each scenario. P is calculated as the proportion of simulations in which M^S was correctly detected as better than M^L. All standard errors for $\overline{D(\theta)}$, p_D and DIC were less than 1.4, 0.3 and 1.2 respectively.

Table 4.3: Performance of DIC_d and DIC_m in correctly detecting M^L (data generating model) as more parsimonious than M^S across five scenarios where data were generated from models with different values for the observation and process error standard deviations, σ and τ respectively.

			Data Level				Partially Marginalized			
σ	τ	Fitted Model	$\overline{D(\theta)}$	p_D	DIC_d	P	$\overline{D(\theta)}$	p_D	DIC_m	P
0.10	0.02	M^S	-85.6	11.4	-74.2	0.44	-66.5	7.7	-58.8	0.88
		M^L	-84.4	10.6	-73.8		-67.0	6.6	-60.4	
0.10	0.07	M^S	-93.0	14.9	-78.1	0.44	-52.7	8.1	-44.6	0.81
		M^L	-92.2	14.6	-77.6		-52.9	7.1	-45.7	
0.10	0.10	M^S	-92.1	16.3	-75.8	0.49	-41.0	7.9	-33.2	0.77
		M^L	-92.7	16.2	-76.4		-40.8	6.9	-33.9	
0.10	0.15	M^S	-93.5	17.8	-75.7	0.48	-22.2	7.5	-14.7	0.83
		M^L	-93.4	17.7	-75.7		-22.3	6.5	-15.8	
0.10	0.20	M^S	-94.4	18.6	-75.7	0.48	-6.7	6.9	0.2	0.78
		M^L	-94.3	18.8	-75.5		-6.6	5.9	-0.7	

Notes: Parameter estimates are means over the 250 simulated datasets for each scenario. P is calculated as the proportion of simulations in which M^L was correctly detected as better than M^S. All standard errors for $\overline{D(\theta)}$, p_D and DIC were less than 1.5, 0.3 and 1.2 respectively.

Table 4.4: Comparison of DIC_d and DIC_m of a suite of models of coho salmon dynamics that differ in the structure of the process equations that were fitted to the Lobster Creek dataset. Details of the structure of models are given in Section 2.2.2.2.

	Data Level				Partially Marginalized			
Fitted Model	$\overline{D(\theta)}$	pD	DIC_d	ΔDIC_d	$\overline{D(\theta)}$	pD	DIC_m	ΔDIC_m
M_{BL_hL}	-352.2	130.1	-92.0	0.0	196.1	18.6	233.3	0.0
$M_{B_hL_hL}$	-354.3	130.5	-93.3	-1.2	197.1	19.4	235.9	2.6
M_{BLL}	-354.4	130.9	-92.5	-0.5	205.8	16.8	239.3	6.0
M_{B_hLL}	-355.2	131.4	-92.4	-0.3	205.3	17.9	241.1	7.9
M_{RLL}	-354.3	129.3	-95.7	-3.6	230.5	19.6	269.6	36.4
M_{LLR}	-348.5	128.3	-91.8	0.3	260.8	18.8	298.3	65.0
M_{LLB}	-352.9	129.5	-93.9	-1.9	265.9	19.0	303.9	70.7
M_{LLL}	-357.0	130.8	-95.3	-3.2	275.8	17.4	310.5	77.3
M_{LRL}	-357.4	131.7	-94.1	-2.0	278.2	17.0	312.2	78.9
M_{LBL}	-357.5	132.3	-93.0	-0.9	279.1	17.1	313.4	80.1

one, and the DIC was slightly higher for M^S, as would be expected with the addition of an extra parameter that explains little of the variation in the data.

4.3.3 Lobster Creek coho salmon example

All models were estimated to have similar fit to the empirical dataset when compared using DIC_d (Table 4.4). The maximum ΔDIC_d was 3.6 even though process models among the different models represent very different dynamics. At the partially marginalized level, DIC_m suggested a group of models that all exhibited Beverton-Holt-type density dependence at the juvenile stage, had similar fit to the data. All other models had either Ricker-type density dependence or linear dynamics at the juvenile stage and showed very poor fit relative to the best-fitting models. The differences in p_D among models appear to be more intuitive at the partially marginalized, than the data level.

Both DIC_d and DIC_m correctly selected the true model, M_{BLL} over the Ricker model, M_{RLL}, and linear model, M_{LLL}, for nearly every simulated dataset (Table 4.5). DIC_m performed better at selecting M_{BLL} over M_{BL_hL} than DIC_d but the proportion of simulations where the true model was considered worse than M_{BL_hL} was more frequent than comparisons with the other models (0.43 and 0.18 for DIC_d and DIC_m, respectively).

Table 4.5: Performance of DIC_d and DIC_m in correctly detecting the true model, M_{BLL} as more parsimonious than alternative models for datasets simulated to represent the population dynamics of coho salmon in Lobster Creek.

	Data Level				Partially Marginalized			
Fitted Model	$\overline{D(\theta)}$	p_D	DIC_d	P	$\overline{D(\theta)}$	p_D	DIC_m	P
M_{BLL}	-232.4	127.3	-105.4		226.5	16.9	243.4	
M_{RLL}	-225.9	129.4	-96.5	0.98	270.6	19.1	289.7	1.00
M_{LLL}	-230.6	132.0	-98.5	0.99	319.8	15.9	335.7	1.00
M_{BL_hL}	-232.4	127.4	-105.0	0.57	226.8	17.6	244.4	0.82

Notes: Parameter estimates are means over the 250 replicate simulations. P is calculated as the proportion of simulations in which M_{BLL} was correctly detected as better than the model in that row. All standard errors for $\overline{D(\theta)}$, p_D and DIC were less than 0.3, 0.3 and 0.3 at the data level, and 2.6, 0.4 and 2.5 at the partially marginalized level, respectively.

Differences in $\overline{D(\theta)}$ and DIC were more pronounced at the partially marginalized than the data level. Even when models could be discriminated between, estimates of ΔDIC between the true and alternative models were often small at the data level (Fig 4.2). Consequently, if models are considered to be comparable if they are within $\Delta\text{DIC}= 2$ of each other (as was tentatively suggested by Spiegelhalter et al. (2002) though see Plummer (2008) for discussion of comparing models based on ΔDIC), then the proportion of times M_{BLL} would be considered 'significantly' better than M_{RLL} and M_{LLL} becomes more different at the two levels of DIC. For example the proportion of simulations where M_{BLL} would be selected over M_{RLL} and M_{LLL} decreased to 0.94 and 0.90, respectively for DIC_d but remained at 1 for both comparisons when DIC_m was used. The differences in p_D between models were also closer to the values expected *a priori* at the partially marginalized level than at the data level. For example it would be expected that M_{RLL} would have the same, M_{BL_hL} would have one more, and M_{LLL} would have one less parameter than M_{BLL}.

4.4 Discussion

4.4.1 DIC for state-space models

DIC is a predictive criterion, aiming to identify models with superior out-of-sample predictive performance. State-space models are interesting in this respect in that they often overfit the

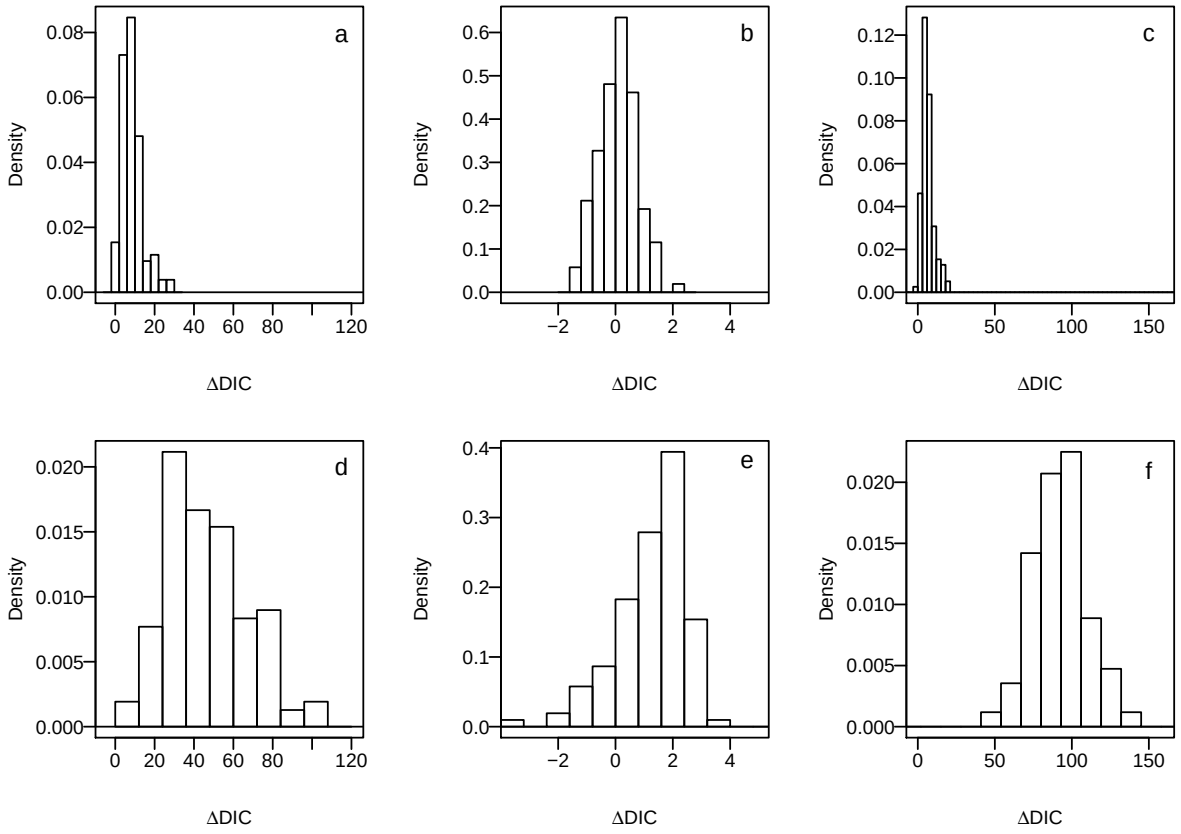

Figure 4.2: Histograms of the distribution of $\Delta\mathrm{DIC}$ between models fitted to the simulated Lobster Creek coho salmon datasets. The $\Delta\mathrm{DIC}$ measures the difference in DIC between the data generating model (M_{BLL}) and the Ricker model (M_{RLL}; a, d), smolt survival model (M_{BL_hL}; b, e) and linear model (M_{LLL}; c, f), measured at the data level (a–c) and the partially marginalized level (d–f). Histograms represent 250 simulated datasets.

data, because the stochastic elements of their structure can allow apparently good fit to the data, but result in poor predictive performance. There is generally a trade off between the observation and process errors and this will determine the relative fits of the latent states, $\boldsymbol{n}$, and the deterministic components of the process model, $\tilde{\boldsymbol{n}}$.

For the log-normal example, the posterior mean of the latent states at time t is the precision-weighted mean of the prior log population size, of which $\log \tilde{n}_t$ is the mean, and the observed population size, $\log y_t$, at time t. It follows that if process error variances are high relative to observation error variances then the $\boldsymbol{n}$ will be estimated to be close to the observations, for all competing models, even if the deterministic components $\tilde{\boldsymbol{n}}$ are a poor description of the system. Consequently, all models will fit well and have similar DIC values at the data level. This is observed in the scenarios with higher τ^2/σ^2 ratios in Table 4.2 and for the empirical dataset from Lobster Creek.

In practice, DIC calculated at the data level will therefore struggle to discriminate between models when the τ^2/σ^2 ratio is high (as observed by Moore and Barlow, 2011), which will generally occur when populations fluctuate a lot (e.g. fisheries) and/or when the drivers of variation are not well known, for instance when unidentified (and unmodeled) environmental conditions drive annual variation in abundance. This was observed for the Lobster Creek dataset where most models in the candidate set fitted similarly well, despite the substantial differences in the dynamics that they represent. The overfitting of models results in poor predictive performance of 'well fitting' models, as evidenced by the systemic problems of the simple logistic model (M^L) in projecting the future dynamics of the population in the log-normal example (Fig 4.1). If DIC_d is utilized in situations where the simulations suggest that DIC_m is more suitable, the risk of spurious selection of models with inadequate deterministic components is high. In these cases predictions of future states will not only be biased, but inflated process errors will result in much less precise estimates than could be justified from the data.

When differences between models occur in the process equations, the more intuitive results at the partially marginalized level also extend to the estimates of effective numbers of parameters

in the models. As just one example, in the scenarios where models had an extra parameter that explained little of the variation in the data, estimates of p_D at the partially marginalized level were typically around one higher for the model with an extra parameter, as was expected. At the data level results were often counterintuitive. Examples of this include the Lobster Creek simulation example when comparing M_{BLL} with a model with an extra parameter that had very little explanatory power (M_{BL_hL}), and also the simple log-normal example when data were simulated from M^L, and parameter γ in M^S provided no extra explanatory ability.

When the deterministic components of the process model are poor at explaining variation in the data, inflated process errors will result. This leads to shrinkage of $\boldsymbol{n}$ towards the observations and higher estimated number of effective parameters at the data level, relative to adequate models. This was noted for M^L in the log-normal example when the simulated τ^2/σ^2 ratio was low, and for many simulations the (inadequate) M^L was estimated to fit the data better than the true model at the data level (Table 4.2, Fig 4.1). In practice these models would hopefully be identified as inadequate using diagnostic tools (such as posterior and partially marginalized predictive checks; Chapter 5), as otherwise the inadequacies of these 'worse' models would allow them to be actively selected over better models, rather than just being selected by chance alone. The latter can occur when there is low power to differentiate between models, for instance when there are large process error variances in the system and models have difficulty in distinguishing signal from noise, such as when $\tau = (0.15, 0.2)$ in Table 4.2.

4.4.2 Are the examples representative?

It could perhaps be argued that the Poisson-gamma and log-normal examples are somewhat contrived, in that these issues would be readily diagnosed by assessing model adequacy in conjunction with undertaking model selection. For example, model inadequacies should be detected by predictive checks after constructing appropriate discrepancy measures (Rubin, 1984; Gelman et al., 2005; Chapter 5). However, there is a tendency to assess model adequacy after completion of model comparison (Burnham and Anderson, 2002), and often on only the best model. This

could result in inadequate models remaining in the candidate set during comparison, which has obvious risks owing to the sensitivity of the components of DIC_d to these inadequacies. In these situations it must be remembered that p_D, and consequently DIC, are only valid if they fit the data adequately, and so models such as the exponential growth model in the Poisson-gamma example may produce p_D estimates at odds with intuition. Lastly, the examples presented are simplified versions of candidate sets commonly compared in practice (e.g. Colchero et al., 2009), for which similar results are to be expected. This suggests they do have merit in supporting the results observed for the Lobster Creek example which was derived from an empirical dataset.

4.4.3 Where to focus in practice?

Simulations suggest that when differences between models occur in the process equations, calculating DIC at the partially marginalized level leads to more robust discrimination between models. Introductions to DIC often state that the appropriate choice of level is determined by the parameters that are of interest (Spiegelhalter et al., 2002; Gelman et al., 2003), or by how predictions are to made e.g. are predictions for new observations from existing, versus entirely new groups, in the simple random effects model example (Gelman et al., 2003). For state-space models, it is natural for the latent states to be of interest, as these are generally the focus of management strategies, and so it is easy to see why DIC has generally been adopted at the data level (in addition to it being the default calculation in BUGS). The results herein suggest it is more profitable for choice of focus to be made based on where the differences between models in the candidate set occur, rather than these previous arguments that do not appear to have been successfully heeded by ecologists.

Following this argument, while differences in the process model will require marginalization of DIC, there are perhaps situations where the DIC_d will be more appropriate, such as when two or more models have the same process model but different observation models. This situation appears to be less common in practice (though see Davies and Jonsen, 2011), as interest in population biology is often directed at the underlying dynamics of the population, rather than

the sampling process by which the data was gathered.

It should also be noted that the most severe consequences of using DIC_d (instead of DIC_m) for selecting among models with different process equations will occur when the process model can compensate for any deficiencies by increasing process error. This clearly occurs when normally distributed process errors are assumed, however it is expected that less dispersed process models such as the binomial and Poisson processes used in some situations (e.g. Abadi et al., 2012) will have a restricted ability in allowing latent states to fit the data well when the deterministic components fit the data poorly. For these scenarios DIC_d will potentially perform better than indicated by the examples presented herein. However, DIC_m is still expected to perform at least as well as DIC_d in these situations given that differences between models still occur in the deterministic part of the process model, and regardless, it is questionable how often a process model without overdispersion of some kind will be suitable for typical animal populations.

4.4.4 Alternatives to DIC for selecting among models

There are many alternatives to DIC for selecting between hierarchical Bayesian models, which differ not only in their philosophies, but also their difficulty of calculation and computational requirements. In general, it appears that ecologists apply DIC on practical grounds, due to the simplicity of its calculation at the data level rather than any philosophical advantages over alternative techniques. Marginalization has significant consequences for this choice. For some combinations of process and observation models analytical marginalization will be possible, and DIC can still be calculated directly from the MCMC output. In fact, it can be fully implemented within BUGS using the `cut` function, if the analyst so desires. However, there are advantages to calculating any form of DIC (including DIC_d) external to BUGS as it allows more flexible calculation of $D(\bar{\theta})$. For instance, deviance can be assessed at the posterior means of the latent states (n_t) directly, rather than as a function of the posterior means of the stochastic parents of n_t, as it is implemented in BUGS. There are situations where correlated parameters lead to (avoidable) negative p_Ds and unreliable DIC_d in BUGS, and a similar problem appears to

have been encountered by Davies and Jonsen (2011) who report negative p_Ds for state-space biomass dynamics models of fish. If marginalization is not tractable, numerical integration will be essential, and the computational advantages of DIC over alternative methods will quickly diminish. In these situations it seems unlikely that ecologists will continue to select DIC for purely practical rather than philosophical reasons.

The most frequently cited alternatives include various information criteria (AIC, BIC etc.; Burnham and Anderson, 2002), posterior predictive loss (Gelfand and Ghosh, 1998), and Bayes factors (Kass and Raftery, 1995). Information criteria and the posterior predictive loss approach also rely on explicitly calculating deviance and thus suffer from the same issues of focus as DIC. Furthermore, the challenge of measuring model dimension for state-space models will continue to be a limitation of the information criterion other than DIC. The *ad hoc* application of AIC by specifying a certain number of parameters (e.g. Lessard et al., 2008), which often assumes that each latent state can be counted as one parameter, is clearly a risky endeavor.

Bayes factors differ from DIC philosophically, in that they attempt to identify a 'true' model Kass and Raftery (1995), rather than a model with optimal out-of-sample prediction. In practice they can be difficult to compute and suffer from instability when very diffuse priors are used for parameters in some models in the set. This is frequently the case in practice for ecological models and is certainly so for the set of models considered for coho salmon in Lobster Creek, which makes the use of Bayes factors problematic for this dataset. Furthermore, model development is often an iterative process where models are expanded or contracted as different aspects of the model, or problem, are explored. This is not suited to the simultaneous analysis required by Bayes factors where the full candidate set must be known *a priori*. An advantage of Bayes factors is they readily allow implementation of multimodal inference (Burnham and Anderson, 2002; King et al., 2009). However, they are unlikely to be used frequently in practice as numerical routines to calculate them have to be coded explicitly rather than relying on prescribed routines in higher-level languages such as BUGS. Note though, that for a (very) restricted set of conditions WinBUGS can implement RJMCMC which can be used to calculate Bayes factors using its

JUMP function (King et al., 2009), but unless/until a general form becomes available in BUGS or other easily implemented software, this technique is unlikely to see much use by practitioners.

Given the challenges of many of the alternative model-choice methods, and the results shown here, DIC could still have a role to play in selecting among state-space models, however the practice of indiscriminately applying DIC at the data level clearly needs to change. Analysts will have to consider the purpose of their modeling, and the ensuing range of practical and philosophical issues when selecting which methods will achieve the aims of their study. If this still leads them to utilizing DIC for selecting among state-space population models, it should at least be calculated at the appropriate level for the models in their candidate set.

4.4.5 Applying DIC to data from Lobster Creek

In the case of the Lobster Creek data- and model-set (e.g. Section 2.2), DIC_m is more appropriate, and while there are clearly some limitations to DIC methods in general, there appeared to be utility in applying DIC_m in this situation. This was especially the case for models with large differences in performance such as the linear vs non-linear models. It appears to have much lower power for selecting between models that are very similar in structure, for instance one extra parameter that has little ability in explaining extra variation. This is perhaps not surprising as process errors were very high for all life-stages (Section 3.2.4) which can make it very difficult to choose between models, as was also observed in the Poisson-gamma example to a degree. Again, DIC_m cannot be considered in isolation for Lobster Creek, but it appears to provide a strong base from which it can be combined with other techniques such as predictive checks (Chapter 5) and the model expansion approach advocated by Gelman et al. (2003). Even simple techniques such as directly observing the posterior distributions of parameters such as γ and η in $M_{B_h L_h L}$ (Section 3.2.5) may complement more formal methods in selecting among models. When models become complicated, and especially in the presence of multiple sources of stochasticity that can reach high levels, choosing between models is likely to be difficult even with formal measures of performance available. In fact, choice will often come down to the judgment of the analyst and

the objectives of the study, and an approach that considers multiple aspects of model fit and comparison would appear to be a the most suitable choice in these situations.

In most cases, great benefit would be gained from spending more time investigating the adequacy of models in the set using diagnostic techniques, rather than overvaluing the importance of single measures of model parsimony. Within such a framework, DIC would contribute to inferences, but not at the expense of a suite of contributing considerations. This more holistic approach was hence adopted in Chapter 2 where DIC_m is used as just one diagnostic when selecting among the potential models in the candidate set.

Chapter 5

Assessing the adequacy of state-space models of population dynamics

5.1 Introduction

The flexibility of state-space models and their advantages over more traditional models have seen their widespread application in modeling animal population dynamics, both in fisheries science (Millar and Meyer, 2000b; Maunder, 2003; Peterman et al., 2003) and terrestrial ecology (de Valpine and Hastings, 2002; Calder et al., 2003; Newman et al., 2006). In the former these models were quickly utilized for applied work such as stock assessments (McAllister et al., 1994; Fargo and Richards, 1998; McAllister and Kirkwood, 1998), while development in terrestrial ecology appears to have been more circumspect. For example, early work by several groups (Besbeas et al., 2002, 2003; de Valpine and Hastings, 2002) mainly focused on statistical aspects of fitting models, their advantages over traditional models, and limitations to their application. Only recently have these models begun to be applied to terrestrial systems in large numbers.

In both fisheries science and terrestrial ecology, perhaps the most investigated area of statistical inference for state-space models has been the algorithms for fitting models (McAllister et al., 1994; Meyer and Millar, 1999; Millar and Meyer, 2000b; Newman et al., 2008). Selection among candidate models has received some work (King and Brooks, 2002b; King et al., 2009) and was the topic of Chapter 4. However, once the best model/s have been identified there is no guarantee that they are 'good', or 'adequate' models, that is, that they fit the data well and are suitable for the objectives of the model-fitting exercise. Despite this, the vast majority of appli-

cations of Bayesian state-space models to animal abundance data fail to suitably address model adequacy. While there are likely some studies where formal model checks have been carried out but not adequately presented in publications, it appears more common for checks to simply not be made at all. In fact, in the recent review of integrated models by Schaub and Abadi (2011) it was suggested that goodness-of-fit tests for these types of models have not been developed. Furthermore, where techniques such as posterior predictive checks (Rubin, 1984) are actually used, the necessary details of their implementation are often vague, and not repeatable, with no discussion of the considerable range of choices that have to be made when assessing the fit of models in a state-space setting.

The contention of Schaub and Abadi (2011) that goodness-of-fit tests do not exist for state-space models ignores the flexible, largely simulation-based approaches that have become popular in a wide range of Bayesian models (Rubin, 1984; Gelman et al., 2003; Sinharay and Stern, 2003). Perhaps the most frequently utilized of these methods of assessing the adequacy of Bayesian models, are posterior predictive checks (Rubin, 1984; Gelman et al., 2003). These checks are based on the concept of calculating a test statistic on the observed data, then simulating replicated datasets from the model and recalculating the test statistic for each simulated dataset. If the model is adequate with regards to the aspect of the model being tested by the statistic, then the observed value will be similar to the values calculated from simulation. This is often summarized by a p-value measured as the proportion of times the statistic for the simulated data is as extreme, or more extreme than the observed statistic (Rubin, 1984). If a large proportion of the simulated values are larger or smaller than the observed value then the model is clearly not consistent with the data, and is perhaps inadequate for the objectives of the study if they are sensitive to the aspect of the model being tested. More generally, the posterior predictive approach includes graphical comparison of observed data and data simulated under the model (Gelman, 2003, 2004) without formal quantification of the consistency between them, although this can be performed after examining the plot/s if desired.

The posterior predictive check approach has been advocated by King (2012) in the context

of state-space models, and has been relatively widely adopted in terrestrial applications (e.g. Rhodes et al., 2011; Cubaynes et al., 2012), but less so in fisheries applications (though see Zhou et al., 2009). The appeal of posterior predictive checks can largely be attributed to their ease of computation when models are fitted using numerical techniques, and the flexibility with which different aspects of a models adequacy can be assessed. Alternative techniques, such as cross validation and its variants have heavy computational demands that make them less attractive to analysts. For these reasons it seems likely that posterior predictive checks will continue to be the method of choice for assessing the adequacy of state-space animal population dynamics models, and provided they are shown to display favorable properties and improve inferences from modeling studies, then their uptake will ideally increase.

The question of whether posterior predictive checks have utility is certainly a valid one. These checks have been shown to be conservative in certain situations although the contention that this can be attributed to double use of the data (Bayarri and Berger, 2000) has been contested by proponents of the method (e.g. Stern, 2000). This is related to the long-standing criticism of posterior predictive checks that they do not always share the properties of traditional null hypothesis tests which seek to test whether the model is true, and thus it is desirable to have a specified Type I error rate. This requires the p-value to be asymptotically uniform(0,1) under the null model, but this is often not the case for posterior predictive checks (Bayarri and Berger, 2000; Robins et al., 2000). The goal of achieving uniformity has in some cases prompted the calibration of p-values (Steinbakk and Storvik, 2009) which can increase power in 'rejecting' models. Alternatively, others regard calibration and the desire of uniformity to be unnecessary, and in fact a misinterpretation of the objectives of modeling. Gelman et al. (2003) and Gelman (2007) suggest interpreting p-values directly as probabilities of future data being as, or more extreme, than the the observed data. This latter philosophy is appealing as state-space models are used for a certain purpose, be it prediction of a future population abundance, estimation of current abundance for setting harvest limits etc., and so concern should be placed on whether the model allows its stated function to be carried out effectively, rather than whether it is the

'true' model.

In addition to the choice of whether p-values should be uniformly distributed under the null, there are several issues that occur in state-space, and more generally, hierarchical models, that influence predictive checks. Firstly, decisions must be made regarding which structures in the model are held constant at the values observed or estimated by the model, and which have new values generated from appropriate sampling or process distributions. In the context of state-space models for example, should replicate data be generated from the sampling distribution $p(y_t|n_t, \sigma^2)$ (see Sections 1.8 and 4.2.2) with σ^2 and latent abundances n_t, in year t, kept at the posterior estimates of the model (this is the standard posterior predictive distribution). Or alternatively, should replicate data be generated after first drawing new values from the distribution $p(n_t|n_{t-1}, \boldsymbol{\psi}, \tau^2)$ where n_{t-1}, $\boldsymbol{\psi}$ and τ^2 are kept at the posterior estimates of the model (this is an 'intermediate' or 'partial' predictive distribution; Bayarri and Castellanos, 2007; Sinharay and Stern, 2003; Steinbakk and Storvik, 2009; herein denoted the partially marginalized distribution). Similarly, there is the choice about how to construct test statistics or discrepancies. Initial development of posterior predictive checks involved test statistics calculated from only the observed or replicated data (Rubin, 1984) and these appear to still be the most widely applied test variables in practice. Gelman et al. (1996) extended the work of Rubin (1984) to allow discrepancies to be a function of both data and model parameters, while more recently at least two methodological studies have used discrepancies based solely on parameters (Sinharay and Stern, 2003; Steinbakk and Storvik, 2009). Gelman et al. (1996) suggest that the choice of how to replicate data will depend on the part/s of the model of interest, and it seems intuitive that this will also govern the choice of which level of discrepancy (data, data/parameter, parameter) will be most suitable, in much the same way that the focus of DIC has to be actively selected (Chapter 4).

If use of posterior predictive checks is to be encouraged in practice, then a better understanding of their properties in these situations is urgently required. The goal of this chapter is to improve knowledge of their behavior when applied to state-space animal population dynamics

models. The key questions that need to be answered include; what are the most suitable predictive distributions (posterior or partially marginalized) for testing adequacy of different parts of the model? Are test variables best constructed from data, parameters or functions of both? What types of test functions are best able to detect common inadequacies of state-space models? Once the best performing predictive checks have been identified, how much power do they have for detecting inadequacies?

In general this thesis concerns state-space models with normally (usually on the log-scale) distributed observation and process models. These models have been shown to be relatively 'robust' even in situations where data is suspected to have been generated from process that differ significantly from normal distributions. For example, Brooks et al. (2004) demonstrated their flexibility in approximating count-based processes except when population size was very small, and Sinharay and Stern (2003) produced similar results for more simple hierarchical models. Consequently, most focus of this chapter is placed on detecting inadequacies in the deterministic components of the process model, and consequently, by the structure of process errors.

The general approach closely follows that used in Chapter 4, and so simulation is called upon to test whether the deficiencies of inadequate models fitted to data simulated from known processes can be detected using predictive checks. While constructing appropriate test variables will often be context-dependent, simulation certainly has the potential to identify some essential properties that can be generalized across situations. Before simulation can be addressed, prior and posterior predictive checks and the variations that are possible for state-space and hierarchical models are first presented in Sections 5.2.1 and 5.2.3. Test variables (test statistics and discrepancy functions, e.g. Gelman et al., 1996) and Bayesian p-values are defined in Section 5.2.2 and a number of simulation examples are presented in Sections 5.2.6–5.2.7. Replicate datasets for each example are generated from both posterior predictive and partially marginalized predictive distributions, and a range of test statistics and discrepancies are developed to test different aspects of the model structure. Finally, data are simulated to represent the dynamics of coho salmon in Lobster Creek as a means to test the predictive methods and test variables

identified as promising in the earlier examples in a more realistic setting.

5.2 Methods

5.2.1 Prior and posterior predictive distributions

Prediction of future data is easily achieved using a Bayesian model, although there may be choice over what part of the model is conditioned upon when making these predictions. To illustrate this point and to present the predictive distributions available, notation similar to that used by Gelman et al. (1996) and Gelman (2003) is adopted throughout this chapter. The prior predictive distribution (Box, 1980) of replicate data $\boldsymbol{y^{rep}}$ is given by

$$p(\boldsymbol{y^{rep}}) = \int p(\boldsymbol{y^{rep}}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{5.1}$$

where $p(\boldsymbol{\theta})$ is the prior distribution for $\boldsymbol{\theta}$. The prior predictive distribution is seldom used for model checking in practice as prior distributions are often specified so as to be uninformative, and are sometimes improper which can can make calculation of (5.1) impossible. Furthermore, the usual scenario in practice is that relatively weak prior values have minimal influence on the posterior distribution and thus the sensitivity of (5.1) to the prior distribution in these situations is at odds with the focus of the modeling, that is, the prior model is generally not intended to be a good specification of $p(\boldsymbol{y})$, and hence model checking is not a relevant issue.

An alternative predictive distribution that has found widespread use in model checking is the posterior predictive distribution of $\boldsymbol{y^{rep}}$, which is given by

$$p(\boldsymbol{y^{rep}}|\boldsymbol{y}) = \int p(\boldsymbol{y^{rep}}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})\mathrm{d}\boldsymbol{\theta} \tag{5.2}$$

where $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior distribution of the parameters.

Rarely can (5.1) and (5.2) be calculated analytically, and certainly not for non-linear state-space models like those considered in Chapter 2. Posterior predictive derivation typically proceeds by first fitting models using numerical techniques such as MCMC, which results in a sample of $\boldsymbol{\theta}_i$, $i = 1, ..., n$, from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, and then generating the replicated data

$\boldsymbol{y^{rep}}$ from these samples using the distribution $p(\boldsymbol{y^{rep}}|\boldsymbol{\theta})$. The prior predictive approach would use samples of $\boldsymbol{\theta}$ from the prior, rather than posterior distributions.

5.2.2 Test variables and Bayesian p-values

Test summaries can refer to a wide range of diagnostics, from uni- and multivariate test statistics to graphical representations which may be constructed from data, parameters or functions of both. Throughout these will collectively be referred to as test variables. The original derivation of the posterior predictive check involved calculation of test statistics for the observed data $T(\boldsymbol{y})$, and replicate data $T(\boldsymbol{y^{rep}})$ generated from the posterior predictive distribution (Rubin, 1984), which is given by (5.2). $T(\boldsymbol{y^{rep}})$ is a reference distribution for the test statistic and the associated Bayesian p-value is then given by

$$p = P[T(\boldsymbol{y^{rep}}) \geq T(\boldsymbol{y})|\boldsymbol{y}] \ . \tag{5.3}$$

This framework was extended by Gelman et al. (1996) to allow more general discrepancy statistics, $D(\boldsymbol{y}, \boldsymbol{\theta})$, to be calculated from functions of both data and parameters, with the associated p-value calculated

$$p = P[D(\boldsymbol{y^{rep}}, \boldsymbol{\theta}) \geq D(\boldsymbol{y}, \boldsymbol{\theta})|\boldsymbol{y}] \ . \tag{5.4}$$

In practice the p-values (5.3) and (5.4) are calculated numerically using the following algorithm (Sinharay and Stern, 2003). The model is fitted using numerical methods giving the approximate posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ in the form of a sample of $i = 1, ..., n$ vectors of parameters ($\boldsymbol{\theta}_i$). A replicate $\boldsymbol{y^{rep}}_i$ for each of the $i = 1, ..., n$ samples is then generated from the sampling distribution $p(\boldsymbol{y^{rep}}|\boldsymbol{\theta})$ giving n pairs ($\boldsymbol{y^{rep}}_i$,$\boldsymbol{\theta}_i$). To calculate (5.3) the single observed test statistic $T(\boldsymbol{y})$ is calculated for the observed data $\boldsymbol{y}$ and this is compared to the reference distribution created by calculating the n test statistics, one for each of the replicates of the data $\boldsymbol{y^{rep}}_i$, and thus $P[T(\boldsymbol{y^{rep}}) \geq T(\boldsymbol{y})|\boldsymbol{y}]$ is easily computed.

In contrast, the discrepancy in (5.4) is by nature a function of both data and parameters, and so different values of D can result for each $\boldsymbol{\theta}_i$. Thus n discrepancies form the observed

distribution, one for each of the samples of $\boldsymbol{\theta}$, and are denoted $D(\boldsymbol{y}, \boldsymbol{\theta}_i)$. This is compared to a reference distribution of the same size where the discrepancy is calculated for each pair ($\boldsymbol{y^{rep}}_i$,$\boldsymbol{\theta}_i$), denoted $D(\boldsymbol{y^{rep}}_i, \boldsymbol{\theta}_i)$, and the p-value can again be easily computed as

$$p = \frac{1}{n}\sum_{i=1}^{n} I[D(\boldsymbol{y^{rep}}_i, \boldsymbol{\theta}_i) \geq D(\boldsymbol{y}, \boldsymbol{\theta}_i)|\boldsymbol{y}]$$

where I is an indicator function.

5.2.3 Extension to hierarchical and state-space models

Distinguishing between prior and posterior predictive distributions becomes difficult in hierarchical models where definitions of priors and likelihoods become less clear. This is a similar situation to the problem of determining what part of the model to focus on when assessing model fit (Chapter 4). For demonstration purposes consider a simple hierarchical model with posterior distribution given by

$$p(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})$$

where $\boldsymbol{\theta}$ are data level parameters and $\boldsymbol{\psi}$ are hyperparameters. The prior and posterior predictive distributions of $\boldsymbol{y^{rep}}$, as suggested by (5.1) and (5.2), are given for this model by

$$p(\boldsymbol{y^{rep}}) = \int p(\boldsymbol{y^{rep}}|\boldsymbol{\theta})p(\boldsymbol{\theta}, \boldsymbol{\psi})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} \tag{5.5}$$

$$p(\boldsymbol{y^{rep}}|\boldsymbol{y}) = \int p(\boldsymbol{y^{rep}}|\boldsymbol{\theta})p(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{y})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} \tag{5.6}$$

where $p(\boldsymbol{\theta}, \boldsymbol{\psi})$ in (5.5) and $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{y})$ in (5.6) are the joint prior distribution and joint posterior distribution of all parameters, respectively, and $p(\boldsymbol{y^{rep}}|\boldsymbol{\theta}, \boldsymbol{\psi})$ has been reduced to $p(\boldsymbol{y^{rep}}|\boldsymbol{\theta})$ due to the conditional independence of $\boldsymbol{y^{rep}}$ and $\boldsymbol{\psi}$, given $\boldsymbol{\theta}$.

Now the predictive distributions (5.5) and (5.6) can be further extended to the basic state-space model described in Section 1.8 and Section 4.2.2. Using the notation given in those sections the prior and posterior predictive distributions for replicate abundance at time t can be given by

$$p(y_t^{rep}) = \int p(y_t^{rep}|n_t, \sigma^2)p(n_t, \sigma^2)\mathrm{d}n_t\mathrm{d}\sigma^2 \tag{5.7}$$

$$p(y_t^{rep}|\boldsymbol{y}) = \int p(y_t^{rep}|n_t, \sigma^2)p(n_t, \sigma^2|\boldsymbol{y})\mathrm{d}n_t\mathrm{d}\sigma^2 \tag{5.8}$$

where $p(n_t, \sigma^2)$ and $p(n_t, \sigma^2|\boldsymbol{y})$ are the joint prior and posterior distributions of the latent states and the observation error variance. In most cases some of the prior distributions such as $p(\sigma^2)$ will be uninformative and so the prior predictive distribution becomes difficult to compute, and again is not generally of interest. Note that for clarity the parameters $\boldsymbol{\xi}$ shown in Sections 1.8 and 4.2.2 are excluded from (5.8)–(5.7) as the observation equations in this chapter are a function of only $\boldsymbol{n}$ and σ^2.

Due to the multiple levels of state-space, and more generally hierarchical models, there is also a partially marginalized predictive distribution that utilizes the posterior values for the hyperparameters but the prior values for the data level 'parameters' (such as the latent states). This distribution can be considered to be part-way between (5.7) and (5.8). The partially marginalized predictive distribution has been investigated by several authors who have given it different names and sometimes undertaken the calculations in slightly different manners (Bayarri and Castellanos, 2007; Sinharay and Stern, 2003; Gelman, 2007; Steinbakk and Storvik, 2009). As suggested by Gelman et al. (1996), all these predictive distributions are variants of posterior predictive distributions, the only difference being at which level/s either new parameters are generated from the model or parameter values are kept at the values resulting from calculation of the joint posterior distribution. Furthermore, the full Bayesian interpretation of the partially marginalized predictive distribution will be considered in this chapter rather than the 'empirical Bayes' method of Bayarri and Castellanos (2007).

For the simple state-space example the partially marginalized predictive distribution is given by

$$p(\tilde{y}_t^{rep}|\boldsymbol{y}) = \int p(\tilde{y}_t^{rep}|\tilde{n}_t, \sigma^2, \tau^2) p(\tilde{n}_t, \sigma^2, \tau^2|\boldsymbol{y}) \mathrm{d}\tilde{n}_t \mathrm{d}\sigma^2 \mathrm{d}\tau^2 \tag{5.9}$$

where $\tilde{y}_t^{rep}$ is the replicate data (it is given this notation to distinguish it from y_t^{rep}, the data generated from the posterior predictive approach) generated conditionally on $\tilde{n}_t$, σ^2 and τ^2, and $\tilde{n}_t$ is the deterministic component of the process model (see sections 2.2.2.1 and 4.2.4.1 for examples; note also that $\tilde{n}_t$ is a function of parameters $\boldsymbol{\psi}$, hence they do not appear explicitly in (5.9)).

The method of calculating a posterior predictive test variable using (5.8) closely follows that described in Section 5.2.2. Replicate data in year t is generated from the sampling distribution $p(y_t^{rep}|n_t, \sigma^2)$ where the values of n_t and σ^2 are samples from the posterior distribution estimated using MCMC. Similarly, if the distribution $p(y_t^{rep}|\tilde{n}_t, \sigma^2, \tau^2)$ is available in closed form then replicate data for the partially marginalized predictive distribution in (5.9) can be generated directly using samples of $\tilde{n}_t$, σ^2 and τ^2 from their posterior distributions. If not, a two-stage process can be implemented by firstly generating new n_t from the conditional prior distribution $p(n_t|\tilde{n}_t, \tau^2)$, and then replicate data can be generated from the sampling distribution $p(y_t^{rep}|n_t, \sigma^2)$ using the newly generated n_t, where $\tilde{n}_t$, τ^2 and σ^2 are again samples from the posterior distribution.

5.2.4 Test variables for predictive checks of state-space models

The possibility of generating reference datasets from different predictive distributions significantly increases the choices that can be made when constructing test variables. Especially when it is considered that these variables can be constructed from data, parameters or both, and the parameters can be from different levels of the model. It has to be decided at which level/s adequacy of the model needs to be examined, in the same way that focus of DIC had to be defined when comparing model-fit in Chapter 4. If the adequacy of the process model or the full model (both observation and process models) is of interest it seems that predictive checks will have to include at least some examples based on the partially marginalized predictive distribution. In general it would appear that two levels of distributions with scale parameters (e.g. the log-normal/log-normal state-space model; (4.14)–(4.15) in Section 4.2.4.2) would result in posterior predictive checks based on (5.8) lacking power in detecting model inadequacy due to process and/or observation errors being inflated in poor (inadequate) models, as occurs when poor models fit data well at the data-level (Chapter 4). This has been observed in other studies of posterior predictive checks for simple hierarchical models (Sinharay and Stern, 2003; Bayarri and Castellanos, 2007).

For clarity all p-values investigated will be given notation that differentiate between whether

Table 5.1: Notation of p-values used for assessing the adequacy of state-space models.

Notation	Test Variable	Description
$P^{(y)}$	$T(\boldsymbol{y})$	Replicate data $\boldsymbol{y^{rep}}$ generated from (5.8); posterior values for $\boldsymbol{n}$ used
$P^{(y,\Theta)}$	$D(\boldsymbol{y}, \boldsymbol{\Theta})$	Replicate data $\boldsymbol{y^{rep}}$ generated from (5.8); posterior values for $\boldsymbol{n}$ used
$P^{(\Theta)}$	$D(\boldsymbol{\Theta})$	Replicate data $\boldsymbol{y^{rep}}$ generated from (5.8); posterior values for $\boldsymbol{n}$ used
$\tilde{P}^{(\tilde{y})}$	$T(\tilde{\boldsymbol{y}})$	Replicate data $\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}$ generated from (5.9); posterior values of $\tilde{\boldsymbol{n}}$ used
$\tilde{P}^{(\tilde{y},\Theta)}$	$D(\tilde{\boldsymbol{y}}, \boldsymbol{\Theta})$	Replicate data $\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}$ generated from (5.9); posterior values of $\tilde{\boldsymbol{n}}$ used
$\tilde{P}^{(\Theta)}$	$D(\boldsymbol{\Theta})$	Replicate data $\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}$ generated from (5.9); posterior values of $\tilde{\boldsymbol{n}}$ used

Notes: $T(\boldsymbol{y})$ and $T(\tilde{\boldsymbol{y}})$ are test statistics calculated from data, only, $D(\boldsymbol{\Theta})$ are discrepancies that are functions of parameters, only, and $D(\boldsymbol{y}, \boldsymbol{\Theta})$ and $D(\tilde{\boldsymbol{y}}, \boldsymbol{\Theta})$ are discrepancies that are functions of both data and parameters. Note that for simplicity Θ includes latent states as well as 'traditional' parameters.

replicate data has been generated from the posterior (P) or the partially marginalized predictive distribution ($\tilde{P}$), and whether the test variable is a function of the data ($\boldsymbol{y}$ or $\tilde{\boldsymbol{y}}$), parameters ($\boldsymbol{\Theta}$) or both (($\boldsymbol{y}, \boldsymbol{\Theta}$) or ($\tilde{\boldsymbol{y}}, \boldsymbol{\Theta}$)), as described in Table 5.1. Note that latent states are classified as parameters for the purposes of discussion of construction of test variables throughout this chapter.

5.2.5 General approach to testing

The general approach to testing the effectiveness of posterior predictive checks closely followed the simulation-based methods utilized for evaluating DIC in Section 4.2.4. Datasets were generated, and the generating model ('true' model) and one or more alternative models were then fitted to the data. The alternative models differed from the true model by having different structures such that they might be considered 'inadequate' for modeling the data ('inadequate' models). Posterior and partially marginalized predictive checks were then developed and were applied for all models to evaluate whether they could identify the inadequacy of the alternative model while remaining consistent with the data for the true model.

5.2.6 Simulations from a log-normal logistic model

5.2.6.1 A simple logistic model

The data generation model (denoted M_{ln}^{L}) for this example consisted of a logistic model with log-normally distributed observation and process errors (the same model was previously used in Section 4.2.4.2). Consequently, the observation and process models are given by

$$\log y_t = \log n_t + \nu_t \tag{5.10}$$

$$\log n_t = \log \tilde{n}_t + \epsilon_t \tag{5.11}$$

where ν_t and ϵ_t are the observation and process errors, respectively, both in year $t = 1, ..., T$, and are assumed to be normally distributed such that $\nu_t \sim \mathrm{N}(0, \sigma^2)$ and $\epsilon_t \sim \mathrm{N}(0, \tau^2)$, where σ^2 and τ^2 are the observation and process error variances, respectively. The deterministic component of the process model, $\log \tilde{n}_t$, was given by the logistic model as

$$\log \tilde{n}_t = \log \left[n_{t-1} + r n_{t-1} \left(1 - \frac{n_{t-1}}{K} \right) \right] \tag{5.12}$$

where r and K are parameters representing intrinsic maximum population growth rate and the carrying capacity, respectively.

An exponential growth model (M_{ln}^{E}) was considered as an alternative model to test whether the failure to account for the non-linear structure in the generation model could be detected using posterior, and partially marginalized predictive checks. The deterministic component of the process equation for this model was given by

$$\log \tilde{n}_t = \log \left(r n_{t-1} \right)$$

where r is the population growth rate parameter.

Several scenarios were considered for this example which were achieved by varying parameters n_1, σ and τ in the simulation model. All other parameters were kept constant at; $T = 30$, $r = 0.4$, and $K = 50\,000$. The seven scenarios considered the sets of parameter values (in order of their application in scenarios); $n_1 = \{10\,000, 10\,000, 10\,000, 10\,000, 10\,000, 25\,000, 40\,000\}$,

$\sigma = \{0.05, 0.1, 0.2, 0.1, 0.1, 0.1, 0.1\}$ and $\tau = \{0.1, 0.1, 0.1, 0.05, 0.2, 0.1, 0.1\}$. This resulted in the population increasing to carrying capacity (50 000) from different starting population sizes over the time series (Fig 5.1), and also allowed for a variety of observation to process error variance ratios.

Prior distributions for the parameters in the models were; $K \sim$ Gamma$(50, 0.001)$, $r \sim$ Uniform$(0, 1)$, $\log n_0 \sim$ N$(9, 4)$ and $\tau \sim$ Uniform$(0, 10)$. The prior for σ was set at $\sigma \sim$ LN$(-3.243, 0.490)$, LN$(-2.548, 0.490)$, and LN$(-1.857, 0.490)$ when the value of σ that was used to simulate data was 0.05, 0.1 and 0.2, respectively. For each simulated dataset each model was fitted by running two MCMC chains for 20 000 iterations after a burnin of 20 000, with samples stored for 1 in every 2 samples, resulting in a total of 20 000 samples available for inferences. Convergence typically occurred much faster than 20 000 iterations but this burnin was specified conservatively to ensure that that convergence occurred before storage of samples for all simulated datasets. A total of 250 simulations were carried out for each scenario.

5.2.6.2 Construction of test variables

A number of checks were investigated for both posterior and partially marginalized predictive distributions, including test variables that were functions of data, parameters, and both data and parameters. Several of the test variables were formulated based on hypothesized dynamics of the specific models. For instance it is expected that when a linear model is fitted to data from a non-linear system, population growth rate at low and high abundance will be under- and overestimated, respectively. This inadequacy could potentially be identified by test variables based on correlations between data or parameters and population size.

By way of example, the correlation coefficient between the estimated process errors ($\boldsymbol{\epsilon}$) and deterministic estimates of population size ($\tilde{\boldsymbol{n}}$) are expected to be high, as the $\boldsymbol{\epsilon}$ will be predominantly positive at low population abundance to compensate for the deterministic components of M_{ln}^E underestimating abundance, and conversely will be predominantly negative at high population abundance. The correlation coefficients calculated for data replicated under the model are

expected to be lower as M_{ln}^{E} assumes $\boldsymbol{\epsilon}$ are unrelated to population size, and hence an extreme p-value is likely (this is the basis for $\tilde{P}_{\epsilon,den}^{(\Theta)}$ (5.14), and the similar $P_{\nu,den}^{(y,\Theta)}$ (5.13) and $\tilde{P}_{\tilde{y},den}^{(\tilde{y})}$ (5.15), which utilize different data/parameters as test variables).

Other test variables were more general and could perhaps be suitable for upcoming examples, or other situations where different inadequacies in models are encountered. The full set of p-values is given below

$$P_{\nu,den}^{(y,\Theta)} = P\left[\text{corr}\left(\boldsymbol{\nu}^{\boldsymbol{rep}}, \log \boldsymbol{n}\right) \geq \text{corr}\left(\boldsymbol{\nu}, \log \boldsymbol{n}\right) | \boldsymbol{y}\right] \tag{5.13}$$

$$\tilde{P}_{\epsilon,den}^{(\Theta)} = P\left[\text{corr}\left(\boldsymbol{\epsilon}^{\boldsymbol{rep}}, \log \tilde{\boldsymbol{n}}\right) \geq \text{corr}\left(\boldsymbol{\epsilon}, \log \tilde{\boldsymbol{n}}\right) | \boldsymbol{y}\right] \tag{5.14}$$

$$\tilde{P}_{\tilde{y},den}^{(\tilde{y})} = P\left[\text{corr}\left(\tilde{\boldsymbol{y}}^{\boldsymbol{rep}} - \boldsymbol{y}, \log \tilde{\boldsymbol{n}}\right) \geq 0 | \boldsymbol{y}\right] \tag{5.15}$$

$$P_{\nu,aut}^{(y,\Theta)} = P\left[\rho\left(\boldsymbol{\nu}^{\boldsymbol{rep}}\right) \geq \rho\left(\boldsymbol{\nu}\right) | \boldsymbol{y}\right] \tag{5.16}$$

$$\tilde{P}_{\epsilon,aut}^{(\Theta)} = P\left[\rho\left(\boldsymbol{\epsilon}^{\boldsymbol{rep}}\right) \geq \rho\left(\boldsymbol{\epsilon}\right) | \boldsymbol{y}\right] \tag{5.17}$$

$$P_{n,dev}^{(y,\Theta)} = P\left[-2\, p(\boldsymbol{y}^{\boldsymbol{rep}} | \boldsymbol{n}) \geq -2p\left(\boldsymbol{y} | \boldsymbol{n}\right) | \boldsymbol{y}\right] \tag{5.18}$$

$$\tilde{P}_{\tilde{n},dev}^{(\tilde{y},\Theta)} = P\left[-2\, p(\tilde{\boldsymbol{y}}^{\boldsymbol{rep}} | \tilde{\boldsymbol{n}}) \geq -2p\left(\boldsymbol{y} | \tilde{\boldsymbol{n}}\right) | \boldsymbol{y}\right] \tag{5.19}$$

$$P_{var}^{(y)} = P\left[\text{Var}\left(\boldsymbol{y}^{\boldsymbol{rep}}\right) \geq \text{Var}\left(\boldsymbol{y}\right) | \boldsymbol{y}\right] \tag{5.20}$$

$$\tilde{P}_{var}^{(\tilde{y})} = P\left[\text{Var}\left(\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}\right) \geq \text{Var}\left(\boldsymbol{y}\right) | \boldsymbol{y}\right] \tag{5.21}$$

where corr and ρ indicate the Pearson correlation coefficient and autocorrelation functions respectively, $\boldsymbol{\nu}^{\boldsymbol{rep}} = \log \boldsymbol{y}^{\boldsymbol{rep}} - \log \boldsymbol{n}$, $\boldsymbol{\nu} = \log \boldsymbol{y} - \log \boldsymbol{n}$, $\boldsymbol{\epsilon}^{\boldsymbol{rep}} = \log \boldsymbol{n}^{\boldsymbol{rep}} - \log \tilde{\boldsymbol{n}}$, $\boldsymbol{\epsilon} = \log \boldsymbol{n} - \log \tilde{\boldsymbol{n}}$ and $-2p(\boldsymbol{y}|\boldsymbol{n})$ and $-2p(\boldsymbol{y}|\tilde{\boldsymbol{n}})$ are the data level and marginalized deviances (calculated using the same formulas as in Section 4.2.4.2) of the parameters in question.

P-values $P_{\nu,den}^{(y,\Theta)}$ and $\tilde{P}_{\epsilon,den}^{(\Theta)}$ test for correlation between population size and the observation and process errors, respectively, that are not consistent with the assumptions of the model. Similarly, $\tilde{P}_{\tilde{y},den}^{(\tilde{y})}$ attempts to identify this same deficiency, though using test variables based only of replicate data from the partially marginalized predictive distribution rather than as a function of parameters as in (5.13) and (5.14). P-values $P_{\nu,aut}^{(y,\Theta)}$ and $\tilde{P}_{\epsilon,aut}^{(\Theta)}$ test for autocorrelation in the observation and process errors, respectively. P-values $P_{n,dev}^{(y,\Theta)}$ and $\tilde{P}_{\tilde{n},dev}^{(\tilde{y},\Theta)}$ are general tests

for inadequacy and are very similar to the likelihood-based discrepancy proposed by King and Brooks (2002a) where the deviance is calculated at the the data level and the marginalized level, for the former and latter respectively. P-values $P_{var}^{(y)}$ and $\tilde{P}_{var}^{(\tilde{y})}$ test for differences in the overall variance of data between the observed and replicated datasets when replications come from the posterior predictive (5.8) and partially marginalized predictive distributions (5.9), respectively.

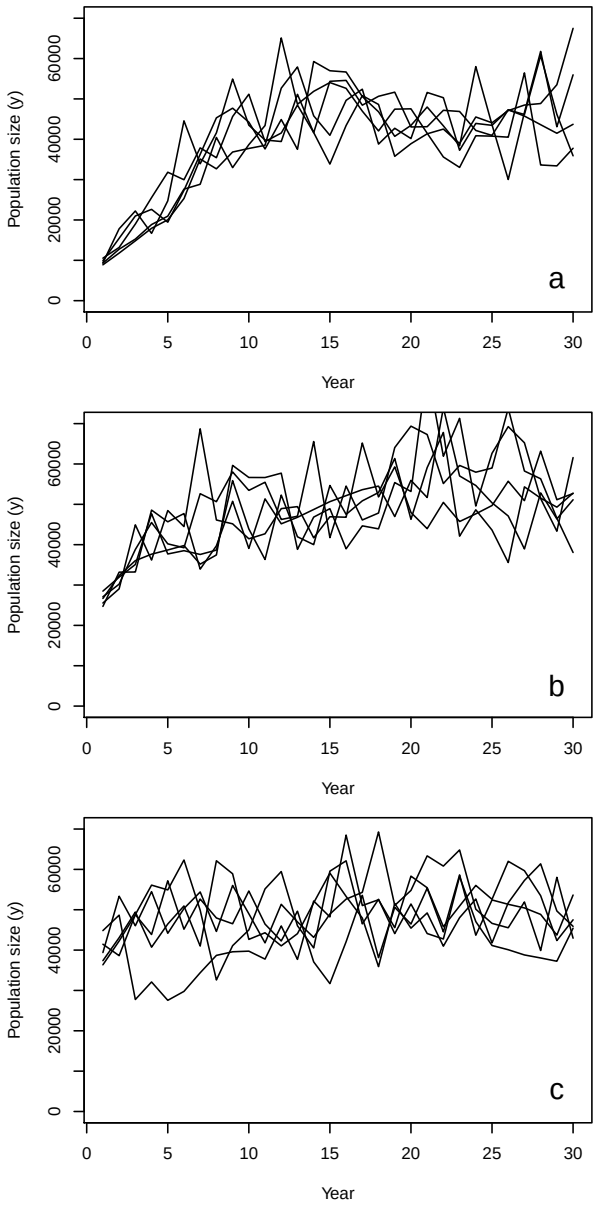

Figure 5.1: Examples of datasets ($\boldsymbol{y}$) simulated from M_{ln}^{L} in Section 5.2.6.1, with initial population sizes of 10 000 (a), 25 000 (b) and 40 000 (c).

It should be noted that the p-values resulting from (5.13)–(5.21) are considered extreme if

they are close to either zero or one, in contrast to traditional p-values that are only considered extreme when they approach zero. For example, (5.14) is expected to be close to one in situations where M_{ln}^{E} is fitted to data generated from M_{ln}^{L} as correlation between $\boldsymbol{\epsilon}^{\boldsymbol{rep}}$ and $\mathbf{log}\,\tilde{\boldsymbol{n}}$ is expected to be higher than between $\boldsymbol{\epsilon}$ and $\mathbf{log}\,\tilde{\boldsymbol{n}}$. Thus, as the p-value approaches one it can be interpreted simply that the test variable for data replicated under the model is almost always higher than for the observed data. A p-value close to zero obviously has the opposite interpretation in that the test variable for data replicated under the model is almost always lower than for the observed data. An example of the latter is expected to be (5.21) when M_{ln}^{E} is fitted to data generated from M_{ln}^{L}, as the process errors are expected to be inflated for this inadequate model in a similar manner to that observed in Chapter 4.

5.2.6.3 Simulations from the logistic-stepped model with a covariate

The next example used the logistic-stepped model first encountered in Section 4.2.4.2 as the true (data-generating) model. This model is very similar to the logistic model in Section 5.2.6.1, with observation and process errors identical to (5.10) and (5.11), respectively. The deterministic components of the process model were however modified to accommodate a change in carrying capacity part way through the time-series to represent, for instance, a sudden increase in a food source that was previously limiting population growth. The deterministic component of the process model is given by

$$\log \tilde{n}_t = \log \left[n_{t-1} + r n_{t-1} \left(1 - \frac{n_{t-1}}{K\left(1 + \gamma I_t\right)} \right) \right]$$

where I_t is an indicator variable taking the value of either 1 or 0 in year $t = 1, ..., T$ when the factor releasing the population from limitation is present or absent, respectively, and γ is a parameter determining the strength of the relationship between the carrying capacity and the indicator variable. The purpose of this example was to test the ability of predictive checks in identifying the presence of time-varying carrying capacity. Hence, in addition to the true model being fitted to the simulated data the more simple, 'inadequate' model, M_{ln}^{L} (Section 5.2.6.1), was also fitted.

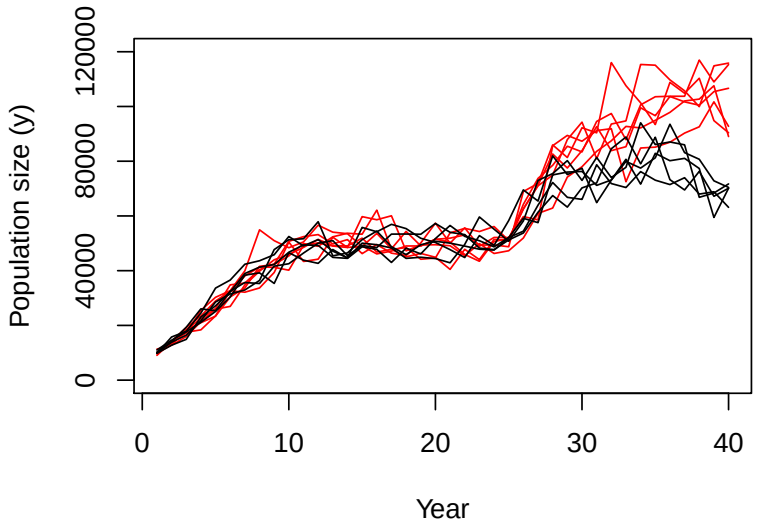

Figure 5.2: Examples of datasets ($\boldsymbol{y}$) simulated from M_{ln}^{S} in Section 5.2.6.3, with initial population sizes of $n_0 = 10\,000$, σ=0.05 and τ=0.05, showing the difference between simulations with γ=0.5 (black lines) and γ=1 (red lines). Each line is an individual simulated dataset.

Most parameters were kept constant at the following values for all scenarios; $T = 40$, $n_1 = 10\,000$, $r = 0.4$, $K = 50\,000$, and $I_t = 0$ for $t = 1, 2, ..., 25$, and 1 for $t = 26, 27, ..., 40$. Parameters σ, τ and γ were varied to produce eight scenarios where the sets of parameter values (in order of their application in scenarios) were; $\sigma = \{0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.05, 0.2\}$, $\tau = \{0.05, 0.05, 0.1, 0.1, 0.2, 0.2, 0.1, 0.1\}$ and $\gamma = \{0.5, 1, 0.5, 1, 0.5, 1, 1, 1\}$.

All priors for these models were identical to those in Section 5.2.6.1 and the additional parameter γ had a prior specified as $\gamma \sim \mathrm{N}(0, 4)$. Model fitting details were the same as in Section 5.2.6.1 and a total of 250 simulations were carried out for each scenario.

An extra test variable based on the behavior of the process errors was added to those considered in Section 5.2.6.1. It was hypothesized that M_{ln}^{L} would overestimate $\log \tilde{n}_t$ when the population level was around the lower carrying capacity at K and subsequently underestimate $\log \tilde{n}_t$ when the higher carrying capacity at $K \times (1 + \gamma)$ was reached later in the time-series when $I = 1$. Consequently, test variables based on parameters ($\tilde{P}_{\epsilon,ratio}^{(\Theta)}$) and data ($\tilde{P}_{\tilde{y},ratio}^{\tilde{y}}$) were constructed using the partially marginalized predictive distribution to investigate whether the ratio

of errors between the two periods differs between the observed and predicted data/parameters. The p-values of these variables are given by

$$\tilde{P}^{(\Theta)}_{\epsilon,ratio} = P\left[\left(\frac{1}{11}\sum_{t=30}^{40}\epsilon_t^{rep} - \frac{1}{11}\sum_{t=10}^{20}\epsilon_t^{rep}\right) \geq \left(\frac{1}{11}\sum_{t=30}^{40}\epsilon_t - \frac{1}{11}\sum_{t=10}^{20}\epsilon_t\right)|\boldsymbol{y}\right]$$

$$\tilde{P}^{\tilde{y}}_{\tilde{y},ratio} = P\left[\left(\frac{1}{11}\sum_{t=30}^{40}\tilde{y}_t^{rep} - y_t\right) - \left(\frac{1}{11}\sum_{t=10}^{20}\tilde{y}_t^{rep} - y_t\right) \geq 0|\boldsymbol{y}\right]$$

where the times $t = 10, ..., 20$ and $t = 30, ..., 40$ represent the periods where the population is expected to be near the low and high carrying capacities, respectively.

5.2.7 Simulations from a log-normal exponential growth model

This example involved simulating datasets from an exponential growth model with log-normally distributed observation and process errors. The true model and a model with a Poisson (instead of log-normal) process model were then fitted to each dataset to test the ability of predictive checks to identify situations where the process model is underdispersed. The observation and process equations of the simulation model (denoted M^E_{ln}) were again given by (5.10) and (5.11) from Section 5.2.6.1 and the deterministic component of the process model, $\log \tilde{n}_t$, was given by the simple exponential growth model, (5.12). The alternative ('inadequate') model, denoted M^E_{po}, shared the log-normal observation model (5.10) and the deterministic component of the process model (5.12) however it differed from M^E_{ln} in that the process equation was modeled as a Poisson process rather than having log-normal process errors, i.e.

$$n_t|r, n_{t-1} \sim \text{Pois}(rn_{t-1}) \ .$$

Most parameters were kept constant for all scenarios at; $T = 30$, $n_1 = 10\,000$, $r = 1.035$. The error standard deviation parameters were however varied, with scenarios having observation error standard deviations of $\sigma = \{0.1, 0.1, 0.1, 0.05, 0.2\}$ and process error standard deviations of $\tau = \{0.05, 0.1, 0.2, 0.1, 0.1\}$ for the five different scenarios, thus giving varying observation to process error ratios.

Priors for the models were similar to those in Sections 5.2.6.1 and 5.2.6.3 with $\log n_0 \sim \text{N}(9, 4)$ and $\tau \sim \text{Uniform}(0, 10)$, while parameter r can potentially take higher values than in previous ex-

amples and hence it was given the prior $r \sim \text{Uniform}(0, 10)$. When the value of σ used to simulate data was 0.05, 0.1 and 0.2, the prior for this parameter was $\sigma \sim \text{LN}(-3.243, 0.490), \text{LN}(-2.548, 0.490)$, and $\text{LN}(-1.857, 0.490)$, respectively. The prior for the observation error standard deviation for (M_{po}^{E}) was allowed to be uninformative as identifiability is not such an issue in this model, hence $\sigma \sim \text{U}(0, 10)$. Again the model fitting details were the same as in Section 5.2.6.1 and a total of 250 simulations were carried out for each scenario.

One additional test variable was considered along with those described in Section 5.2.6.1, and is given by

$$\tilde{P}_{\tilde{y},aut}^{(\tilde{y})} = P[\rho(\tilde{\boldsymbol{y}}^{\boldsymbol{rep}} - \boldsymbol{y}) \geq 0 | \boldsymbol{y}]$$

where ρ is again an autocorrelation function. This p-value tests for autocorrelation in the differences between data simulated from the partially marginalized predictive distribution and the observed data, and is an attempt to investigate the same deficiencies as $\tilde{P}_{\nu,aut}^{(\tilde{y},\Theta)}$ and $\tilde{P}_{\epsilon,aut}^{(\Theta)}$ but using only data instead of functions of parameters.

5.2.8 Simulations from models of Lobster Creek coho salmon

Most interest in the adequacy of models of coho salmon population dynamics in Lobster Creek (Chapter 2) is concerned with the structure of the process model, especially the deterministic components, $\log \tilde{n}_{i,j,t}$. Simulations were therefore carried out to test the utility of a variety of predictive checks for assessing adequacy of these models. Simulations followed the general methods presented in Sections 5.2.6–5.2.7 whereby datasets were simulated from a known model, and that model, and 'inadequate' alternative models, were fitted to the data before assessing their adequacy with a range of discrepancy measures.

The data generation model was essentially identical to that used to simulate data in Section 4.2.5.2. It was based on model M_{BLL} (Section 2.2.2.3) and generates datasets for all three stages of the coho life-cycle, in both streams, for 26 years. Three models were then fitted to each simulated dataset; the true model (M_{BLL}), a model with linear functions at each stage (M_{LLL}; Section 2.2.2.3) and a model (M_{BLL}^{*}; appendix Section 8.1) where process errors were assumed

to have stream-specific process error variances, rather than the temporal process error structure ($\omega_{i,t}$; see (2.2) in Section 2.2.2.1) which were assumed in the former two models. By fitting these two alternative models the predictive checks could be used to assess the adequacy of fitting a linear model to a non-linear system, and the adequacy of assuming independent process errors in a system where temporal correlations exist between streams.

The performance of the various predictive checks on the simple examples in Sections 5.2.6–5.2.7 were used to choose appropriate test variables for the Lobster Creek simulation. Correlations between process errors and population density were examined in the same manner as (5.14) in Section 5.2.6.1. P-values were calculated for each stage-class separately and using the juvenile stage as an example they are given by

$$\tilde{P}^{(\Theta)}_{\epsilon,1,den} = P\left[\text{corr}\left(\epsilon^{rep}_{1,j,t}, \log\tilde{n}_{1,j,t}\right) \geq \text{corr}\left(\epsilon_{1,j,t}, \log\tilde{n}_{1,j,t}\right)|\boldsymbol{y}\right]$$

where the 1 denotes the juvenile stage-class and the other subscripts denote the ith stream and tth year.

Direct correlations in process errors between streams were compared for both the observed and the partially marginalized predictive process errors, to test the adequacy of assuming independent process errors between streams ($\epsilon_{i,1,t}$ is EF and $\epsilon_{i,2,t}$ is MF). A new test variable was constructed to achieve this, and again using the juvenile stage as an example, the p-values for this test are given by

$$\tilde{P}^{(\Theta)}_{\epsilon,1,cor} = P\left[\text{corr}\left(\epsilon^{rep}_{1,1,t}, \epsilon^{rep}_{1,2,t}\right) \geq \text{corr}\left(\epsilon_{1,1,t}, \epsilon_{1,2,t}\right)|\boldsymbol{y}\right]$$

where corr denotes the Pearson correlation coefficient.

5.3 Results

5.3.1 Log-normal logistic simulations

All p-values based on posterior predictive distributions, and those based on test variables involving observation errors (ν), autocorrelation functions and deviance calculations for the partially marginalized predictive distributions, suggested that both the true and inadequate models were

consistent with the data (Table 5.2). Only $\tilde{P}^{\Theta}_{\epsilon,den}$, $\tilde{P}^{(\tilde{y})}_{\tilde{y},den}$ and $\tilde{P}^{(\tilde{y})}_{var}$ identified inconsistencies between the inadequate model (M^E_{ln}) and the data for this example. Tail area probabilities were most extreme (P$\sim$ 0 or 1) for these tests for scenarios where $n_1 = 10\,000$ and simulated error variances (σ^2 and/or τ^2) were low. P-values became less extreme as n_1, σ or τ increased. $\tilde{P}^{\Theta}_{\epsilon,den}$ remained very extreme (P>0.96) for all scenarios, however $\tilde{P}^{(\tilde{y})}_{\tilde{y},den}$ and $\tilde{P}^{(\tilde{y})}_{var}$ were slightly less extreme than the former for all scenarios, and p-values suggested much less conflict between the model and data (P=0.87 and 0.14, respectively) when $n_1 = 40\,000$. For these three test variables, and in fact all those investigated for this example, p-values for the true model (M^L_{ln}) were close to 0.5, correctly indicating an adequate model. In all cases there was very little between-simulation variation in p-values (all standard errors <0.02), hence only the mean values are displayed in Table 5.2.

An example of the behavior of replicated data and parameters for both inadequate and adequate models is displayed in Figures 5.3 and 5.4, respectively. As predicted, the posterior estimates of the deterministic abundances ($\tilde{\boldsymbol{n}}$; black lines) from M^E_{ln} tended to be lower than the latent states ($\boldsymbol{n}$; red lines) when population size was low at the beginning of the time-series, and higher at the end when the population was fluctuating around the carrying capacity at high abundance (Fig 5.3). Thus, when new replicate latent states ($\boldsymbol{n^{rep}}$; blue lines) are predicted from this model they are randomly spread around the deterministic abundances, and the differences between the 'observed' (estimated) and simulated parameters can be detected with well constructed partially marginalized predictive checks (though not posterior predictive checks, as the $\boldsymbol{n}$ remain at the same values for both the observed and replicated data/parameters in those cases). It can be seen that the variance of the replicate latent states appears higher than the 'observed' latent states (tested for by $\tilde{P}^{(\tilde{y})}_{var}$), and process errors ($\boldsymbol{\epsilon} = \boldsymbol{n} - \tilde{\boldsymbol{n}}$) will be evenly distributed around $\tilde{\boldsymbol{n}}$ for the replicate states, but will decrease from largely positive to largely negative values for the 'observed' states (tested for by $\tilde{P}^{(\Theta)}_{\epsilon,den}$).

Figure 5.4 shows that the dynamics of both the 'observed' and replicated latent states are relatively similar for the true model M^L_{ln}. Note also that process variation is considerably reduced

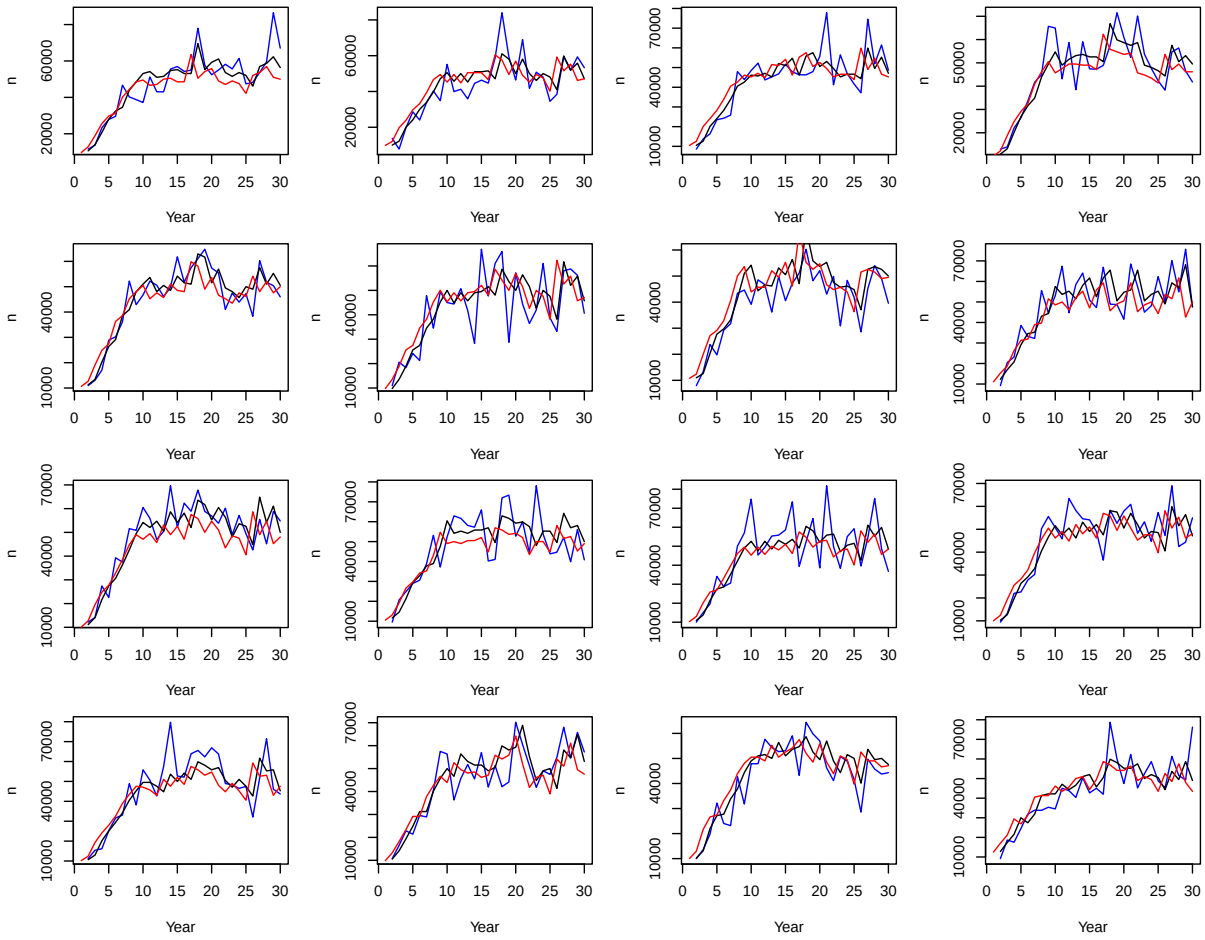

Figure 5.3: Plots of estimated latent states ($\boldsymbol{n}$) for M_{ln}^{E} for a single dataset simulated from M_{ln}^{L} with $n_1 = 10\,000$, σ=0.1 and τ=0.1. Lines indicate posterior estimates of latent states ($\boldsymbol{n}$; red lines), posterior estimates of the deterministic states ($\tilde{\boldsymbol{n}}$; black lines) and replicated latent states from the partially marginalized predictive distribution ($\boldsymbol{n}^{\boldsymbol{rep}}$; blue lines). Each panel represents one of the $i = 1, ..., n$ samples randomly selected from the posterior distribution approximated using MCMC.

compared to M_{ln}^{E}, which is unsurprising given it is also the data generating model.

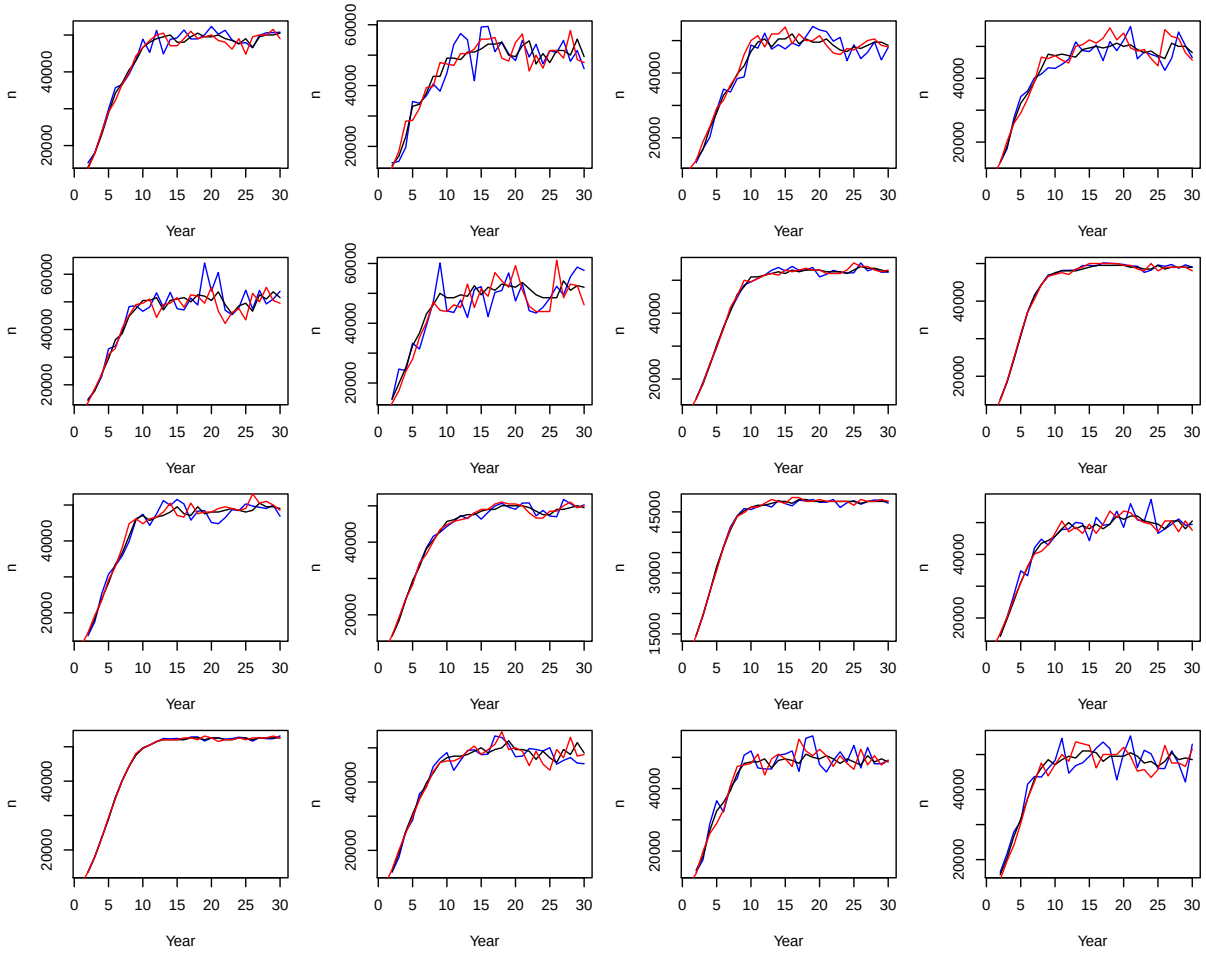

Figure 5.4: Plots of estimated latent states ($\boldsymbol{n}$) for M_{ln}^{L} for a single dataset simulated from M_{ln}^{L} with $n_1 = 10\,000$, σ=0.1 and τ=0.1. Lines indicate posterior estimates of latent states ($\boldsymbol{n}$; red lines), posterior estimates of the deterministic states ($\tilde{\boldsymbol{n}}$; black lines) and replicated latent states from the partially marginalized predictive distribution ($\boldsymbol{n}^{\boldsymbol{rep}}$; blue lines). Each panel represents one of the $i = 1, ..., n$ samples randomly selected from the posterior distribution approximated using MCMC.

5.3.2 Log-normal logistic-stepped simulations

Most of the test variables, including those that performed well in the previous example, suggested that both M_{ln}^{S} and M_{ln}^{L} were adequate models for the simulated data (Table 5.3). Only the example-specific p-values $\tilde{P}_{\epsilon,ratio}^{(\Theta)}$ and $\tilde{P}_{\tilde{y},ratio}^{(\tilde{y})}$ which were both based on partially marginalized predictive distributions suggested that M_{ln}^{L} was in anyway inadequate. The p-values for these test variables were less extreme when the covariate value was low (γ=0.5) and when either observation or process errors (σ or τ) were high. When those conditions were met $\tilde{P}_{\epsilon,ratio}^{(\Theta)}$ and

Table 5.2: Posterior predictive p-values for the true model (M_{ln}^L) and an inadequate model (M_{ln}^E) fitted to datasets simulated for the log-normal logistic example. Scenarios were investigated using different values for the initial population size (n_1), observation error variance (σ^2) and process error variance (τ^2). P-values less than 0.1 or greater than 0.9 are in bold.

n_1	σ	τ	Model	$\tilde{P}_{\nu,den}^{(\tilde{y},\Theta)}$	$\tilde{P}_{\epsilon,den}^{(\Theta)}$	$\tilde{P}_{\tilde{y},den}^{(\tilde{y})}$	$\tilde{P}_{\nu,aut}^{(\tilde{y},\Theta)}$	$\tilde{P}_{\epsilon,aut}^{(\Theta)}$	$P_{n,dev}^{(y,\Theta)}$	$\tilde{P}_{\tilde{n},dev}^{(\tilde{y},\Theta)}$	$P_{var}^{(y)}$	$\tilde{P}_{var}^{(\tilde{y})}$
10 000	0.1	0.1	M_{ln}^L	0.50	0.52	0.51	0.50	0.54	0.50	0.46	0.50	0.47
			M_{ln}^E	0.58	**1.00**	**0.99**	0.51	0.43	0.47	0.43	0.41	**0**
25 000	0.1	0.1	M_{ln}^L	0.51	0.57	0.57	0.49	0.53	0.50	0.46	0.48	0.41
			M_{ln}^E	0.61	**0.99**	**0.96**	0.52	0.59	0.49	0.45	0.39	**0.03**
40 000	0.1	0.1	M_{ln}^L	0.51	0.63	0.61	0.48	0.51	0.50	0.46	0.48	0.39
			M_{ln}^E	0.57	**0.96**	0.87	0.50	0.60	0.51	0.48	0.44	0.14
10 000	0.1	0.05	M_{ln}^L	0.51	0.50	0.50	0.55	0.58	0.50	0.47	0.49	0.48
			M_{ln}^E	0.59	**1.00**	**1.00**	0.51	0.41	0.47	0.42	0.40	**0**
10 000	0.1	0.2	M_{ln}^L	0.51	0.57	0.58	0.48	0.50	0.50	0.46	0.49	0.42
			M_{ln}^E	0.56	**0.99**	**0.98**	0.50	0.52	0.49	0.45	0.44	**0.01**
10 000	0.05	0.1	M_{ln}^L	0.51	0.54	0.54	0.49	0.49	0.51	0.46	0.49	0.45
			M_{ln}^E	0.57	**1.00**	**1.00**	0.49	0.25	0.49	0.44	0.43	**0**
10 000	0.2	0.1	M_{ln}^L	0.50	0.58	0.55	0.44	0.42	0.47	0.43	0.49	0.41
			M_{ln}^E	0.57	**0.99**	**0.97**	0.47	0.43	0.46	0.42	0.42	**0.02**

$\tilde{P}_{\tilde{y},ratio}^{(\tilde{y})}$ were as low as 0.79 and 0.78, respectively, but when γ=1 and σ and τ were less than 0.2 then p-values were generally above 0.9 (Table 5.3).

Table 5.3: Posterior predictive p-values for the true model (M_{ln}^S) and an inadequate model (M_{ln}^L) fitted to datasets simulated for the log-normal logistic-stepped example. Scenarios were investigated by varying the values for change in carrying capacity (γ), observation error variance (σ^2) and process error variance (τ^2). P-values less than 0.1 or greater than 0.9 are in bold.

γ	σ	τ	Model	$\tilde{P}_{\nu,den}^{(\tilde{y},\Theta)}$	$\tilde{P}_{\epsilon,den}^{(\Theta)}$	$\tilde{P}_{\nu,aut}^{(\tilde{y},\Theta)}$	$\tilde{P}_{\epsilon,aut}^{(\Theta)}$	$P_{n,dev}^{(y,\Theta)}$	$\tilde{P}_{\tilde{n},dev}^{(\tilde{y},\Theta)}$	$P_{var}^{(y)}$	$\tilde{P}_{var}^{(\tilde{y})}$	$\tilde{P}_{\epsilon,ratio}^{(\Theta)}$	$\tilde{P}_{\tilde{y},ratio}^{(\tilde{y})}$
0.5	0.1	0.05	M_{ln}^S	0.51	0.57	0.54	0.57	0.50	0.48	0.48	0.46	0.46	0.46
			M_{ln}^L	0.53	0.70	0.53	0.52	0.48	0.45	0.46	0.30	**0.93**	**0.91**
1	0.1	0.05	M_{ln}^S	0.51	0.53	0.53	0.56	0.50	0.47	0.48	0.46	0.47	0.47
			M_{ln}^L	0.53	0.69	0.53	0.47	0.48	0.44	0.46	0.30	**0.94**	**0.93**
0.5	0.1	0.1	M_{ln}^S	0.52	0.55	0.50	0.53	0.50	0.47	0.48	0.44	0.48	0.48
			M_{ln}^L	0.53	0.67	0.51	0.51	0.49	0.46	0.46	0.32	0.89	0.89
1	0.1	0.1	M_{ln}^S	0.50	0.56	0.50	0.54	0.50	0.47	0.49	0.44	0.46	0.46
			M_{ln}^L	0.52	0.67	0.51	0.48	0.49	0.45	0.48	0.33	**0.92**	**0.91**
0.5	0.1	0.2	M_{ln}^S	0.52	0.64	0.48	0.48	0.50	0.47	0.48	0.35	0.41	0.41
			M_{ln}^L	0.51	0.63	0.49	0.50	0.50	0.47	0.49	0.37	0.79	0.78
1	0.1	0.2	M_{ln}^S	0.51	0.65	0.47	0.48	0.50	0.47	0.49	0.35	0.42	0.42
			M_{ln}^L	0.51	0.63	0.49	0.48	0.49	0.46	0.49	0.36	0.87	0.86
1	0.05	0.1	M_{ln}^S	0.51	0.57	0.49	0.49	0.50	0.46	0.49	0.42	0.44	0.44
			M_{ln}^L	0.52	0.70	0.49	0.39	0.50	0.45	0.47	0.29	**0.91**	**0.91**
1	0.2	0.1	M_{ln}^S	0.52	0.56	0.54	0.55	0.50	0.47	0.48	0.43	0.46	0.46
			M_{ln}^L	0.53	0.56	0.52	0.64	0.50	0.47	0.48	0.38	0.85	0.85

Table 5.4: Posterior predictive p-values for the true model (M_{ln}^E) and an inadequate model (M_{po}^E) fitted to datasets simulated for the Poisson process error example. P-values less than 0.1 or greater than 0.9 are in bold.

σ	τ	Model	$\tilde{P}_{\nu,den}^{(\tilde{y},\Theta)}$	$\tilde{P}_{\epsilon,den}^{(\Theta)}$	$\tilde{P}_{\nu,aut}^{(\tilde{y},\Theta)}$	$\tilde{P}_{\epsilon,aut}^{(\Theta)}$	$\tilde{P}_{\tilde{y},aut}^{(\tilde{y})}$	$P_{var}^{(y)}$	$\tilde{P}_{var}^{(\tilde{y})}$
0.1	0.05	M_{ln}^E	0.51	0.65	0.54	0.59	0.65	0.49	0.39
		M_{po}^E	0.48	0.51	0.22	0.48	0.35	0.52	0.51
0.1	0.1	M_{ln}^E	0.52	0.79	0.50	0.57	0.63	0.48	0.27
		M_{po}^E	0.50	0.57	**0.08**	0.48	0.18	0.49	0.49
0.1	0.2	M_{ln}^E	0.50	0.85	0.49	0.54	0.62	0.49	0.18
		M_{po}^E	0.48	0.54	**0.01**	0.49	**0.10**	0.55	0.55
0.05	0.1	M_{ln}^E	0.50	0.77	0.49	0.54	0.40	0.49	0.26
		M_{po}^E	0.48	0.54	**0.02**	0.48	0.12	0.48	0.48
0.2	0.1	M_{ln}^E	0.53	0.78	0.55	0.59	0.35	0.46	0.30
		M_{po}^E	0.51	0.54	0.19	0.49	0.35	0.45	0.45

5.3.3 Poisson process model simulations

Only $\tilde{P}_{\nu,aut}^{(\tilde{y},\Theta)}$ and $\tilde{P}_{\tilde{y},aut}^{(\tilde{y})}$ detected any deficiencies in M_{po}^E when fitted to data simulated from M_{ln}^E (Table 5.4). Both these test variables were based on partially marginalized predictive distributions and aimed to detect autocorrelation in errors of the fitted model that are not assumed in the structure of that model. However, even p-values for these test variables were only extreme for scenarios with higher process error variances (τ >0.05 for $\tilde{P}_{\nu,aut}^{(\tilde{y},\Theta)}$ and > 0.1 for $\tilde{P}_{\tilde{y},aut}^{(\tilde{y})}$). The p-value $\tilde{P}_{\tilde{y},aut}^{(\tilde{y})}$, which is based on the replicated data alone, had substantially less power in detecting autocorrelated errors than $\tilde{P}_{\nu,aut}^{(\tilde{y},\Theta)}$, which is based on both data and parameters. Both these p-values were more and less extreme when observation error variances for the scenario were lower and higher, respectively.

An example of the inadequacy of M_{po}^E is shown in Figure 5.5. For this scenario (σ=0.1 and τ=0.1) it is clear that the autocorrelation of observation errors for data replicated under the model ($\boldsymbol{\nu^{rep}} = \tilde{\boldsymbol{y}}^{\boldsymbol{rep}} - \boldsymbol{n^{rep}}$) using a partially marginalized predictive distribution is substantially less than the autocorrelation of the observation errors when the model is fitted to the observed data ($\boldsymbol{\nu} = \boldsymbol{y} - \boldsymbol{n}$).

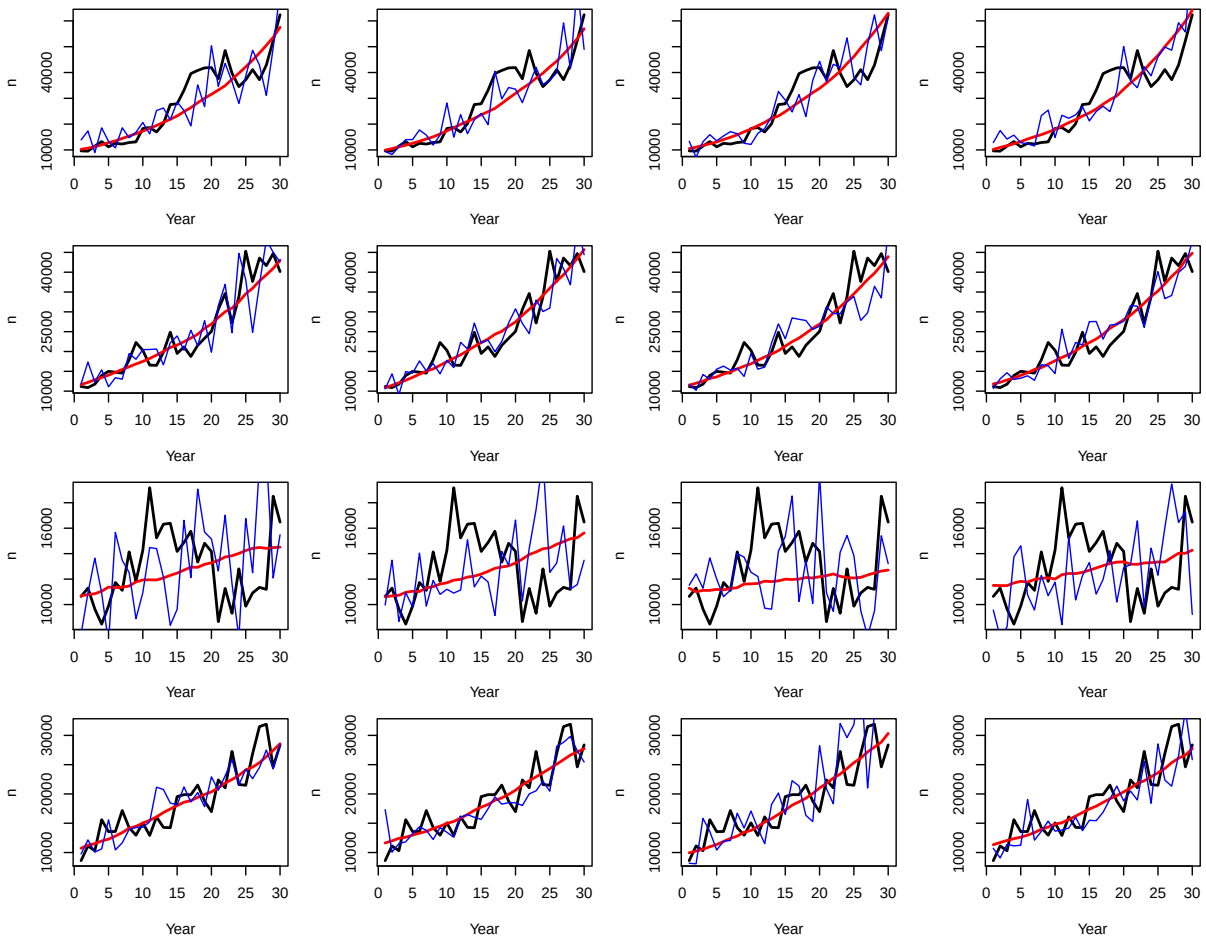

Figure 5.5: Plots of observed data (abundance - $\boldsymbol{y}$; black lines) simulated from M_{ln}^{E} and posterior estimates of the latent states ($\boldsymbol{n}$; red lines) for M_{po}^{E} fitted to the data. Blue lines are replicate data ($\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}$) generated from the partially marginalized predictive distribution under M_{po}^{E}. Each row represents one simulated dataset, and each panel within the row represents one of the $i = 1, ..., n$ samples randomly selected from the posterior distribution approximated using MCMC for that dataset. Simulation parameters used to generate the datasets shown were; $n_1 = 10\,000$, σ=0.1 and τ=0.1.

Table 5.5: Posterior predictive p-values for three models fitted to datasets simulated for the Lobster Creek example, the true model (M_{BLL}), a model without a non-linear function for the juvenile stage (M_{LLL}), and a model assuming independent process errors between streams (M^*_{BLL}). P-values less than 0.1 or greater than 0.9 are in bold.

Model	$\tilde{P}^{(\Theta)}_{\epsilon,1,den}$	$\tilde{P}^{(\Theta)}_{\epsilon,2,den}$	$\tilde{P}^{(\Theta)}_{\epsilon,3,den}$	$\tilde{P}^{(\Theta)}_{\epsilon,1,cor}$	$\tilde{P}^{(\Theta)}_{\epsilon,2,cor}$	$\tilde{P}^{(\Theta)}_{\epsilon,3,cor}$
M_{BLL}	0.47	0.52	0.43	0.53	0.51	0.54
M_{LLL}	**0**	0.43	0.41	0.54	0.51	0.54
M^*_{BLL}	0.47	0.52	0.44	**0.95**	0.83	**1.00**

5.3.4 Lobster Creek simulations

P-values were largely as expected given the known deficiencies of the fitted models, with M_{LLL} identified as inadequate in accounting for non-linearity at the juvenile stage ($\tilde{P}^{(\Theta)}_{\epsilon,1,den}$=0), while the linear functions at the smolt and spawner stages were correctly identified as adequate (Table 5.5). Model M^*_{BLL} was identified as inadequate in modeling correlation of process errors between streams for the juvenile and spawner stages ($\tilde{P}^{(\Theta)}_{\epsilon,1,cor}$=0.95, $\tilde{P}^{(\Theta)}_{\epsilon,3,cor}$=1; Table 5.5), although the predictive checks suggested that this inadequacy was not present at the smolt stage ($\tilde{P}^{(\Theta)}_{\epsilon,2,cor}$=0.83), despite the data at this stage also being simulated with correlated process errors.

5.4 Discussion

5.4.1 Posterior and partially marginalized predictive checks

A number of aspects of predictive checks were investigated in this chapter including whether they should be carried out based on partially marginalized predictive distributions, where new values for the latent states $\boldsymbol{n}$ are drawn from the conditional prior, or posterior predictive distributions where the states are drawn from their posterior distributions. Test statistics for both types of predictive check were also calculated based on data, parameters/states and functions of both to assess whether this affected the power of the approach in detecting inadequate models. The examples used reflect the general conditions encountered when modeling animal population dynamics, and more specifically the case of salmon stock assessment. Consequently, emphasis was mainly placed on detecting deficiencies in the deterministic and stochastic components of

the process model as this is typically where most choices in the structure of the model have to be made. In these situations a number of generalizations can be made from this simulation study.

Partially marginalized predictive checks performed much better than posterior predictive checks when deficiencies in the model were located in the deterministic components of the process model. In some cases posterior predictive checks were not even considered as their utility is clearly limited. For instance, a posterior predictive equivalent of $\tilde{P}^{\Theta}_{\epsilon,den}$ could be calculated for the examples in Sections 5.2.6 and 5.2.6.3, however it would be insensitive to the inadequacies in the deterministic components of the process model being investigated. This is because the process errors will have already compensated for the inadequacies and generating replicate data from the posterior estimates of $\boldsymbol{n}$ (via (5.8)) will nearly always be consistent with the observed data. A possible exception to this is the Poisson process model example (Section 5.2.7) where an appropriate posterior predictive check might detect the inadequacy of M^{E}_{po} as in this case it was located in the stochastic, rather than deterministic components of the process model. However, there appears to be little reason to adopt this approach when inadequacy is easily detected by the partially marginalized predictive checks.

Because partially marginalized predictive checks involve simulating new values for the latent states ($\boldsymbol{n}$) it seems logical to formulate test variables based on the parameters/states at this level of the model, as this is where interest often lies. Test variables based only on underlying parameters of simple hierarchical models have also been considered by Sinharay and Stern (2003) and Steinbakk and Storvik (2009), and these types of test variables ($\tilde{P}^{\Theta}$) performed well for the state-space examples considered herein. For example, when linear models were hypothesized to be inadequate in Sections 5.2.6.1 and 5.2.8, it was relatively easy to construct test variables based on the expected behavior of the process errors ($\boldsymbol{\epsilon}$) alone, while it is often more difficult to construct traditional test variables that are a function of data alone. This has been noted in the past by others (Steinbakk and Storvik, 2009), and is perhaps one reason behind the frequent reliance on omnibus test variables such as the chi-squared (Gelman et al., 1996), Freeman-Tukey (Brooks et al., 2000) and likelihood statistics (King and Brooks, 2002a).

5.4.2 Comparing test variables at different levels of the model

As was shown in the examples considered herein, it is often relatively simple to reformulate test variables from being functions of parameters, to functions of data alone. This can be achieved by investigating the differences between the replicated and observed data and applying the test for the hypothesized deficiency on this difference, rather than directly on the parameters of interest. In this chapter many of the $\tilde{P}^{\Theta}$ were constructed based on the process errors, $\boldsymbol{\epsilon}$. To calculate similar p-values as a function of data alone (e.g. $\tilde{P}^{\tilde{y}}$) the difference $\tilde{\boldsymbol{y}}^{\boldsymbol{rep}} - \boldsymbol{y}$ was first calculated. The equivalent test function was then applied (e.g. correlation with population size), and the value of the test was compared to zero, rather than a reference distribution as occurs with the other test variables. This approach seemed to work relatively well because the general inadequacies of the deterministic process models are propagated through the generation of $\boldsymbol{n}^{\boldsymbol{rep}}$ from $p(\boldsymbol{n}|\tilde{\boldsymbol{n}}, \tau^2)$ and then further onto $\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}$ from $p(\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}|\boldsymbol{n}, \sigma^2)$, and results in differences between $\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}$ and $\boldsymbol{y}$ in these models. It is difficult to see how this would ever be the desirable option however, as stochasticity is induced for both of these distributions which has the potential to ameliorate conflict between data and models if process and/or observation error variances are large.

Using Section 5.2.6.1 as an example of the differences between test variables as functions of data or parameters, tests $\tilde{P}^{(\Theta)}_{\epsilon,den}$ and $\tilde{P}^{(\tilde{y})}_{\tilde{y},den}$ aim to identify the same discrepancy between the fitted model and replications under the model, namely that non-linearity present in the data cannot be reliably modeled by a linear model. This will be evidenced by unusual (given the assumptions of the model) patterns in the process errors, and by extension differences between $\boldsymbol{y}$ and $\tilde{\boldsymbol{y}}^{\boldsymbol{rep}}$. The most significant difference between $\tilde{P}^{(\Theta)}_{\epsilon,den}$ and $\tilde{P}^{(\tilde{y})}_{\tilde{y},den}$ is that the latter is simulated with two distributions with scale parameters (τ and σ) rather than just the stochasticity of the process model, and so it is expected that by chance some replicated test variables will be more likely to be similar to the observed test variables than in the case of $\tilde{P}^{(\Theta)}_{\epsilon,den}$. This is reflected in the results, for instance $\tilde{P}^{(\tilde{y})}_{\tilde{y},den}$ has less power than $\tilde{P}^{(\Theta)}_{\epsilon,den}$ in Section 5.3.1, $\tilde{P}^{(\tilde{y})}_{\tilde{y},ratio}$ has slightly less power than

$\tilde{P}^{(\tilde{y},\Theta)}_{\nu,aut}$ in Section 5.3.2, and $\tilde{P}^{(\tilde{y})}_{\tilde{y},aut}$ had less power than $\tilde{P}^{(\tilde{y})}_{\tilde{y},aut}$ in Section 5.3.3. The difference in power between $\tilde{P}^{\Theta}$ and $\tilde{P}^{\tilde{y}}$ will be partly determined by the form of the test variable. For instance, a variable that is a mean over a number of data/parameter/state values (e.g. multiple years) will be more robust than a statistic based on only one data point/parameter/state.

5.4.3 The use of omnibus tests

Omnibus tests are often utilized as a means of conveniently applying posterior predictive checks to new modeling situations however they were observed to have poor power in the examples examined here. However, it must be noted that little effort was placed in determining the most powerful omnibus test that could be applied for the examples considered, given that formulating test variables for specific aspects of the models was the main aim of this study. It seems likely that they will have limited potential unless the inadequacies of the model are severe, especially if the inadequacy occurs somewhere in the process model. Well thought-out discrepancies that are specific to the situation being considered certainly appear to be the best approach and also directly allow the assessment of whether the objectives of the modeling can be met (if the discrepancies are constructed with these objective in mind). However, it is certainly more difficult to develop specific test statistics, and it appears that omnibus test variables are generally attractive because they circumvent most of the hardest aspects of measuring the adequacy of a model. That is, determining exactly what the purpose of the model is, what aspects of the model consequently need to be adequate, how to construct appropriate tests for these aspects, and how much difference between observed data and data replicated under the model is required before the model is modified or an alternative considered. The effort and careful thought needed in addressing these factors will however certainly enhance the analysts understanding of the model and its adequacy for the dataset at-hand.

5.4.4 Interpreting p-values and graphical summaries

Emphasis throughout this chapter has been placed on p-values rather than graphical comparison of observed data/parameters and those generated under the model. This perhaps reflects the

situation in practice where the later are rarely utilized, or at least presented in published work, but this should not be taken to mean that p-values are superior. In fact, they are entirely compatible with each other. The main reason for the heavy use of p-values herein is the ease with which their calculation can be automated and summarized using simulation. However, well constructed graphical summaries have the ability to assess multiple aspects of the model, some of which may not have been considered *a priori*, and it is generally very easy to confirm a visual discrepancy by also calculating p-values. In fact, several of the test variables considered herein were only constructed after first examining appropriate graphical summaries.

Once a p-value has been calculated, or a graphical comparison of observed and replicated data/parameters constructed and interpreted, the consequences for inferences or further model development may not be immediately obvious. Firstly, the lack of extreme p-values does not guarantee an adequate model, it merely suggests that it is sufficient for replicating the aspect of the data examined by the test. For this reason, robust model checking will involve constructing test variables (or graphs) for all of the aspects of the model most important for achieving the objectives of the study. Similarly, the most cited weakness of posterior predictive checks is their conservatism for hierarchical models (Bayarri and Castellanos, 2007; Steinbakk and Storvik, 2009). The two levels of flexible distributions (distributions with scale parameters) in most of the models explored allow the latent states to nearly always fit the data well, and so data, and test variables from them, replicated under the model with latent states fixed at the 'observed' values (posterior predictive distribution), will nearly always be close to those observed. This is the same situation as observed for DIC_d in Chapter 4. In contrast to posterior predictive checks, tests based on the partially marginalized distributions were relatively powerful when they are carefully constructed, which goes some way to appeasing the doubts of Bayarri and Castellanos (2007) and Steinbakk and Storvik (2009).

It should also be noted that even if conservatism of predictive tests is encountered it is not necessarily a problem (Sinharay and Stern, 2003; Gelman, 2007). It merely suggests the model is able to predict data with characteristics similar to observed dataset even when the model is

known to be 'incorrect'. It seems to be the case that the deficiencies commonly encountered in state-space applications often occur at a level not tested by posterior predictive checks, which test the full model with respect to producing data for new observations from the same groups (in a simple hierarchical model context). Sinharay and Stern (2003) suggest that this indicates 'robust' models, and this appears to be true for those models investigated herein in that normal distributed errors for both process and observation models produce very flexible models. Thus the deficiencies that are often of interest in the process model will have to be investigated using partially marginalized predictive checks.

The decisions that have to be made when extreme p-values (or graphical summaries) are observed are similar to those addressed above for tests that indicate consistency between the model and data. An extreme p-value does not automatically indicate an unusable model, that is, one that has to be modified or necessitates an alternative to be considered. We know before fitting it to data that the model is wrong and so the question becomes, how wrong does it have to be before it becomes unusable? In some cases an extreme p-value may indicate substantial conflict between model and data but will still be usable for the purposes of the study. Consider the test for non-linearity in Section 5.2.6.1. If the purpose of the modeling is to predict future abundance then extreme p-values identifying inadequacy of the linear (exponential growth) model should be heeded, as predictions will overestimate and underestimate the true abundance when population size is low and high respectively. This will consequently impact on robust management of harvest regulations, conservation planning etc. In contrast, graphical summaries in Section 8.4 that display information similar to $\tilde{P}_{var}^{\tilde{y}}$ suggest that the process model is inadequate for describing variation in abundance of coho spawners in Lobster Creek, presumably due the inability to model the affects of climate variation adequately. If the purpose of that model was to separate survival of fish into freshwater and marine components then the model will still have utility. Prediction of future abundance may well be compromised to a degree though. It should be noted though that the inflated process errors for that model will make prediction intervals wide but as stated in Chapter 2, if prediction is deemed important for that system would benefit greatly from

development of this aspect of the model. One possible method of helping determine whether the magnitude of conflict is practically important is to put more emphasis on confidence intervals of test statistics for the replications under the model, rather than calculating only p-values.

5.4.5 Predictive checks and model selection

It is possible to question the use of predictive checks by arguing that many aspects of inadequacy being identified could perhaps be evident in more traditional model assessment, such as examination of residuals etc. In addition, if a certain inadequacy is being tested then generally an alternative model is being considered (at least in the analysts mind) and so it could be asked why should it not be rephrased as a model selection problem using DIC, Bayes factors etc.? This point could of course be made for model checking of traditional models as well, and it has previously been suggested that checking of a single model should only be undertaken in the early, exploratory periods of an analysis before formal methods are used (e.g. Piccinato, 2000). Even if this were true, it can be difficult to compare state-space models when they differ only in their deterministic components (Chapter 4), let alone when they differ at multiple levels such as error structures etc. which are also often the focus of predictive checks (e.g. $\tilde{P}^{(\Theta)}_{\epsilon,cor}$ in Section 5.2.8). In fact, many analysts adopt an approach that expands models based entirely on predictive checks (Gelman et al., 2003; Gelman, 2004; Kery and Schaub, 2011) rather than using formal model selection techniques at all during statistical analysis (this was the basis of Section 3.2.2 in Chapter 2).

5.4.6 Conclusions

In the context of state-space models of animal population dynamics, predictive checks appear to have substantial potential and ideally their use will become more widespread. More explicit description of which predictive distributions, and what type of test variables have been utilized is needed in published work before checks of model adequacy will become standard in the field, as currently there appears to be little emphasis on establishing whether fitted models (or best model/s among a candidate set) can meet the objectives of those studies. Based on this chapter

the path towards identifying adequate model/s would have several steps. The objectives of the study must be well defined, the part/s of the model that are to be focused on must be decided (this will partly be determined by which units of replication are of interest), appropriate test variables must be constructed and assessed using p-values or graphical summaries, and finally it must be evaluated whether any conflict between data and models is significant enough to warrant further model development.

Chapter 6

Comparison of estimators of salmon escapement for periodic count data

6.1 Introduction

Escapement of salmon is usually defined as the number of adult fish that avoid mortality and return to spawn in a certain reach, or waterway (Hilborn and Walters, 1992; Quinn and Deriso, 1999). This measure is frequently the focus of management actions, being an estimate of population size at the life-history stage most-valued by fisherman, conservation managers and the general public. It is also the basis of most assessments of stock status, and allows quantification of the link between subsequent generations of individuals (Beverton and Holt, 1957; Quinn and Deriso, 1999). In contrast to species with more complicated life-cycles, data for a salmon stock is usually restricted to this single measure, at this single stage-class.

Unfortunately, escapement can be difficult to estimate. A fundamental issue is that individuals return to the study reach and spawn over an extended period of time. Because they are semelparous, and because mortality of individuals can occur quickly after spawning, this means that not all individuals will be present at the site simultaneously. This is not necessarily a problem, as under ideal conditions a full census (or very close to it) can be achieved for certain waterways by using weirs or dams to enumerate the number of fish passing daily (e.g. Irvine et al., 1992; Bue et al., 1998; Szerlong and Rundio, 2008). Typically though, this approach will be logistically prohibitive, except for well funded studies with low spatial extent. Alternative techniques are usually adopted when spatial scale of an assessment is increased.

Very commonly, alternatives to censuses involve estimating abundance at a sample of days throughout the season, often using visual survey methods (e.g Irvine et al., 1992). Resulting data are termed periodic counts, and a number of techniques are available to estimate escapement from them. There are other methods available, such as mark-recapture studies, however their use is numerically insignificant in practice compared to periodic count methods (Knudsen, 2000; Parsons and Skalski, 2009).

Perhaps the most popular escapement estimators for periodic count data are based on trapezoidal 'area-under-the-curve' (TAUC) techniques (Ames, 1984; English et al., 1992), where the total number of 'fish-days' is estimated by calculating the area under a curve connecting the counts over the sampling period (Fig 6.1), and escapement is derived by accounting for observer efficiency (the proportion of fish in the steam that are detected) and the stream-life (the time between entry into the study reach and death) of spawning fish. An alternative to this estimator fits a parametric model to the counts with the model parameterized such that escapement is a parameter estimated during maximization of the likelihood (Hilborn et al., 1999). This technique assumes that the arrival of salmon over the spawning season follows a Gaussian curve and has been denoted the likelihood AUC, parametric AUC and HARR (Hilborn arrival model) by Holt and Cox (2008), Parsons and Skalski (2009) and Millar et al. (2012), respectively, despite it not strictly being an AUC technique. Herein this estimator will be denoted HARR. This estimator has the advantage that variance in escapement can be derived from standard maximum likelihood (ML) techniques. Recently, Millar et al. (2012) introduced a similar method (denoted GAUC herein) that fits a Gaussian curve directly to the observed periodic counts within a generalized linear model (GLM) framework. The area under the fitted model is then calculated, and escapement is derived by accounting for observer efficiency and stream-life in the same manner as when using the TAUC method. The estimates produced by GAUC are almost identical to the estimator of Hilborn et al. (1999) except that they are more stable numerically and can be carried out easily in many statistical software packages (Millar et al., 2012).

Another class of methods is the average spawner (AS) estimators, that treat the periodic

counts as a random sample from the full spawning run and multiply the mean count by spawning season duration to estimate fish-days. Escapement is again derived from fish-days using the same techniques as for AUC methods. These estimators have received favorable reviews recently (Holt and Cox, 2008; Parsons and Skalski, 2009) but have seldom been applied in practice.

Finally, the peak count (PC) method can perhaps be considered a special case of an estimator for periodic counts, in that the estimate of escapement is simply the maximum observed count over the periodic count dataset (Parsons and Skalski, 2009). It is not an absolute estimate of escapement however, and thus it is subject to the associated risks that come with basing management decisions on indices of abundance (Anderson, 2001). Note that there are methods for calibrating these relative estimates against absolute escapement (reviewed by Parsons and Skalski, 2009), but they are rarely used in practice. Despite the weaknesses of peak counts, they are, or are among, the most widely used estimators for some salmon species (Knudsen, 2000; Parsons and Skalski, 2009).

Once periodic count data has been obtained there is clearly some choice about which estimator to apply, but little is known about their relative properties. Ideally, choice of estimator would be based on it exhibiting several essential features including low bias, the availability of valid variance estimates which result in confidence intervals with good coverage properties, weak assumptions, and robustness to violation of these assumptions. Ease of implementation is also important, as analyses are often conducted by salmon biologists with little statistical background.

Of the desirable properties of the estimators, the lack of widely used variance estimators is perhaps the most significant weakness in those currently being applied. The inadequacies of relying on only point estimates are widely acknowledged in statistics, but have been largely ignored in studies monitoring salmon escapement. Associated estimates of uncertainty are essential for valid inferences from studies using periodic count data. For example comparing escapement between sites or times, conducting impact assessments on salmon abundance and basing population viability assessments on escapement estimates will be weak when based solely on point estimates.

An immediate example of the benefits of estimating uncertainty, and the motivation for this

chapter, comes when fitting dynamic models to salmon abundance data (e.g. Rivot and Parent, 2001; Peterman et al., 2003; Su and Peterman, 2011, and Chapter 2 herein). In practice it can be difficult to partition the variation in the data between observation and process error when state-space models are fitted to time-series of counts (Schnute, 1994). Often prior information on one or both of these variances, or their ratio, is necessary to achieve identifiability of all model parameters (Schnute, 1994; King et al., 2008, 2009). It is difficult to formulate reasonable prior distributions for observation error variance (for example, parameters $\sigma^2_{3,j,t}$ in Section 2.2.2.1) for state-space models of salmon dynamics when estimators of variance in escapement are either unavailable, or whose reliability has not been assessed. Development of robust variance estimators will potentially be the key to facilitating the successful fitting of state-space models to some datasets of salmon abundance.

There are several reasons why the variance of escapement estimates has seldom been calculated in studies using periodic count data. Most importantly, variance estimators have simply not been available for TAUC and peak count estimators, and these are the most frequently applied techniques in practice (Knudsen, 2000; Parsons and Skalski, 2009). However, there are also difficulties in identifying exactly what the estimated variance represents.

Uncertainty in escapement estimates is introduced from a number of sources including uncertainty in 1) observer efficiency, 2) the detection of fish, 3) the selection of study reaches within the population of interest, 4) the selection of days that are sampled, and 5) in stream-life estimates used to convert fish-days into escapement (Parsons and Skalski, 2009). It has not always been made clear which of these sources of uncertainty have been accounted for in the various variance estimators applied in the past.

A further issue when defining what variance of estimators of escapement represents is that there are major differences in the assumptions of the estimators that can broadly be categorized into design- and model-based approaches. The parametric versions of AUC (Holt and Cox, 2008; Parsons and Skalski, 2009; Millar et al., 2012) are model-based estimators which treat the set of counts as a single realization of underlying stochastic processes and attempt to estimate fish-

days as a function of a model with one or more parameters. The robustness of the estimates is conditional on the model being an adequate representation of the underlying processes. In contrast, design-based approaches (TAUC and AS methods) treat the abundances as fixed, with stochasticity occurring in the selection of the days to be sampled. Design-based estimators have the advantage of generally being unbiased for parameters such as population means and totals, independent of the nature of the underlying population, however their precision will depend on the suitability of the design for the population structure (chapter 8 of Cochran, 1977). For example, high variance in escapement is expected when the temporal structure of fish arrivals in streams is effectively ignored by treating periodic counts as a random sample over the spawning season. This distinction is seldom made with respect to estimating salmon escapement, and has led to confusion regarding the properties of some estimators (Parsons and Skalski, 2009). There are certainly potential advantages in both approaches. For instance design-based estimators may hold promise is situations where spawning runs have multiple pulses of fish entering the stream, which are hard to model adequately. However, sampling of escapement is rarely designed with the assumptions of estimators, and their variance, in mind. Indeed, most frequently a systematic sample is obtained for logistical convenience which has consequences for the validity of most design-based estimators (chapter 8 of Wolter, 2007). The practical significance of design choices such as this has yet to be determined for period count estimators.

Clearly, there are a number of challenges facing practitioners when analyzing periodic count data and guidance for choosing appropriate estimators and their implementation should be a high priority if salmon monitoring and inferences made from resulting estimates are to be improved. These challenges are reflected in the fact that choice among estimators currently appears to be based on convenience, or at times institutional standard practice, rather than on statistical grounds. This seems likely to continue until robust comparisons of the properties of all suitable estimators are implemented and the results disseminated. Undertaking this comparison and focusing on the estimators of escapement, and the associated estimates of variance in escapement, will be the focus of this chapter, and will simultaneously guide the choice of estimator for

producing the escapement estimates (and their variance) used to fit the state-space models in Chapter 2.

Previous studies have compared estimators across different classes of data, for example AUC versus mark-recapture methods (Korman et al., 2002; Parken et al., 2003), while others have focused on several estimators within a single class, such as Millar et al. (2012), who assessed properties of several AUC estimators. Holt and Cox (2008) simulated the dynamics of salmon runs to test the ability of AUC, AS and PC estimators in detecting trends in salmon abundance. They identified AS estimators as most suited to their objectives, although these were slightly different to the usual application of escapement estimates, in that indices of abundance can sometimes be adequate for trend detection. Furthermore, their study contained methodological errors that unfairly influenced the comparison of estimators (as pointed out by Millar et al., 2012). However, their results do highlight the possibility that some widely-used estimators have poor inferential properties compared to alternatives. The most general assessment of escapement estimators was carried out by Parsons and Skalski (2009, 2010). They conducted an extensive review that focused on the statistical properties of a suite of estimators of salmon escapement (across several data-types such as periodic counts, passage counts, mark-recapture data etc.), including statistical bias, validity of assumptions and the availability of associated variance estimates. Among the estimators for periodic count data, they identified the AS method as having the most desirable statistical properties, followed by TAUC. The latter was suggested to be statistically biased and did not account for uncertainty from the selection of sampling days.

While these previous comparisons of estimators, especially those of Parsons and Skalski (2009, 2010), are a valuable addition to the literature, they need to be augmented by assessing the practical (in addition to statistical) properties of a comprehensive set of estimators that appear to have the most potential for periodic count data. For instance, a complete assessment of estimators must investigate their bias and precision, and coverage rates of confidence intervals under conditions encountered in practice. This is a difficult assignment as it requires estimators to be applied to datasets where the true escapement is known, and to make the results general across

species and different types of waterways a wide variety of datasets would be needed. Unfortunately, few datasets amenable to this approach are available (though see Irvine et al., 1992; Bue et al., 1998), and their generality is questionable. One alternative approach to approximating estimator performance under field conditions is to carry out simulations that represent the dynamics of salmon spawning runs and the sampling process through which periodic counts are obtained. This approach was previously adopted by Holt and Cox (2008) and Millar et al. (2012), and is utilized in this chapter to assess the practical properties of the candidate set of escapement estimators for periodic count data with the aim of complementing the theoretical work of Parsons and Skalski (2009).

The structure of this chapter is as follows. The important features of periodic count datasets such as the definitions of spawning season and fish-days are presented in Section 6.2.1. The definition of variance in the estimate of escapement for the various methods and the conditions assumed in its estimation are then addressed. The suite of estimators under consideration are introduced in Sections 6.2.2–6.2.4, their derivations are presented, and the availability and form of associated variance estimators are especially noted. Three classes of estimator are investigated, AUC (Sections 6.2.2.1, 6.2.2.2), AS (Sections 6.2.3.1–6.2.3.3) and PC (section 6.2.4) methods.

Several variants are compared within classes. For instance, several variance estimates for AS methods for standard systematic count data are compared, in an attempt to reduce the bias known to occur in these situations (Wolter, 2007). Similarly, AS methods are most suited to random sampling designs (Parsons and Skalski, 2009) and so the possibility of collecting periodic count data using simple, and stratified random sampling is explored. Technically, periodic count data refers to sets of counts collected by systematic, or roughly systematic sampling, though for the purposes of this chapter the definition is extended to also include random samples which share most of the same fundamental properties. Furthermore, comparisons included in this chapter explore recent advances in AUC estimators that have been made since the review by Parsons and Skalski (2009), including development of the GAUC estimator and proposed variance estimates for TAUC techniques Millar et al. (2012).

Once estimators have been defined, the simulation-based approach of assessing the estimators is introduced in Section 6.2.6, by presenting the structure of the data-generating model and the range of scenarios investigated in the simulations. The results of the simulations are presented in Sections 6.3.1–6.3.3 where the performance of estimators are compared with regard to bias, precision and the coverage rates of their confidence intervals. These comparisons include assessment of differences in performance of estimators for systematic and random sampling-based designs (Section 6.3.2). Finally, the relative merits of all estimators are discussed in Section 6.4. In situations where periodic counts have already been collected via a systematic sample, this comparison will provide guidance for salmon biologists in selecting the most appropriate estimator for the specific conditions encountered in their studies. Where data is yet to be collected, the comparison among estimators, and between alternative sampling designs, will lend itself to practitioners making informed decisions about designing data collection and analysis to meet the objectives of their study.

6.2 Methods

6.2.1 Definitions for the estimation of escapement

6.2.1.1 Definition of periodic counts

The length of the spawning season (the period between the 1st fish arriving in the study reach and the last fish dieing) must be well defined for several of the estimators of escapement. Exceptions include the GAUC (and HARR - which is not calculated herein) estimator which estimates season-length implicitly along with abundance, and PC methods where season length is not utilized. Season-length (in days) for the remaining estimators is denoted N, while the days of the season are denoted d_j for $j = 1, ..., N$. In most applications the season length is defined (sometimes implicitly) as the time between midday of the first day, d_1, and midday of last day, d_N, i.e. $N = d_N - d_1$ (Irvine et al., 1992). Periodic counts are typically collected via a systematic design where counts are made at approximately regular intervals (typically around a week) after the initial count. Counts are denoted c_i where $i = 1, ..., n$ indexes the sampling occasion over

the total of n counts. Sampling stops at c_n when surveyors deem that the spawning season has been completed, usually after one or more zero counts have been observed.

6.2.1.2 Definition and calculation of fish-days and escapement

Most of the estimators examined herein require an estimate of the number of fish present in the stream on each day of the spawning season, based on the periodic count sample $\boldsymbol{c} = c_1, ..., c_n$. In general there will be issues in detecting fish such that $\boldsymbol{c}$ will be negatively biased due to observers consistently missing a fraction of the individuals present on the sampling day. If the proportion detected on sampling occasion i is denoted v_i (termed observer efficiency), then a set of corrected counts would be calculated $\tilde{\boldsymbol{c}} = c_1/v_1, ..., c_n/v_n$. In general however, only estimates of observer efficiency ($\widehat{v}_i$) will be available. The different estimators then generally use these counts to estimate the number of fish present on every day of the season, with the sum over all days denoted fish-days $\widehat{F}$. This is not an estimate of true escapement as individuals are present on multiple days. The estimate of fish-days is divided by an estimate of stream-life (the length of time between a fish entering the study area and it dying; $\widehat{l}$) to estimate escapement, E, e.g.

$$\widehat{E} = \frac{\widehat{F}}{\widehat{l}} \ . \tag{6.1}$$

Estimates of observer efficiency will in most cases only be available for the whole season (and often these will have been estimated in different streams and/or years) rather than for each $i = 1, ..., n$ survey. This simplifies calculations and in these cases an equivalent estimate to (6.1) results from calculating fish-days from the uncorrected counts, $\boldsymbol{c}$, and then correcting fish-days for observer efficiency and stream-life simultaneously, e.g.

$$\widehat{E} = \frac{\widehat{F}}{\widehat{l} \times \widehat{v}} \tag{6.2}$$

where $\widehat{v}$ is an estimate of observer efficiency for all counts. The choice between (6.1) and (6.2) is made independent of which estimator is being used, among those considered herein. Thus, only (6.2) is utilized in this chapter as observer efficiency is considered constant and known throughout (see Section 6.2.1.3) for the comparison of estimators.

6.2.1.3 Variance of fish-days and escapement

Each of the estimators considered herein begins by estimating the variance of $\widehat{F}$, denoted $\mathrm{Var}(\widehat{F})$. The exceptions to this are PC estimators which have no valid variance estimator, and the HARR estimator which includes escapement as a parameter in the model and estimates its variance using standard ML techniques. Because the GAUC variant has been found to be superior to the very similar HARR (Millar et al., 2012), the latter is not considered herein and thus the slightly different technique for estimating variance for that estimator is not presented.

The model-based AUC estimators (GAUC) treat the sampling-days as fixed and the counts are considered one realization from stochastic processes of the system, thus variance estimates will consequently represent the variation in estimated fish-days that would occur over these potential realizations. In contrast, the design-based AS estimators treat the counts as fixed and uncertainty in estimates of fish-days represents variation induced from the selection of days to be surveyed (the TAUC estimator presented in Section 6.2.2.2 is a hybrid of the approaches in that the estimate of escapement is design-based while the variance in this estimate is model-based). The variance represents the variation in escapement over a set of hypothetical replicate samples that could have been collected via identical sampling processes. In practice, the variance estimated by both approaches, $\widehat{\mathrm{Var}}(\widehat{F})$, will represent variation in the detectability of individuals and process variation in the numbers of fish entering and dying throughout the spawning season.

It should be noted that Parsons and Skalski (2009) apply finite population correction factors when using AS estimators but these are not implemented herein. Sampling of fish abundance occurs at some time during the day and is used as an approximation for the number of fish present on that given discrete day, even though the arrival and death of fish are continuous processes. Likewise, when the inferences from the sample are extended to the full season to estimate fish-days, the choice of day as a unit is relatively arbitrary. Separating the continuous time of the season into different discrete units, say half-days etc., may well be just as justifiable as full days, and clearly this would affect the influence of a finite population correction factor.

In general, interest is ultimately directed at estimating escapement ($\widehat{E}$; (6.2)), and ideally also a robust estimate of its variance, $\mathrm{Var}(\widehat{E})$. The latter must account for uncertainty in v and l in addition to uncertainty in the estimate of fish-days, $\mathrm{Var}(\widehat{F})$. These additional sources of uncertainty cannot be accounted for with periodic count data alone. Previous studies have only been able to include these components of variation by incorporating prior information on variation in v and/or l (Hilborn et al., 1999), or utilizing additional data such as from tagging studies (Szerlong and Rundio, 2008) or replicate counts at each sampling occasion (Parken et al., 2003). Without these data, or prior values, variance estimates are restricted to measuring uncertainty in fish-days and this is the basis for the comparison of estimators in this chapter. While reliable inferences about escapement rely on these extra sources of uncertainty being included, the methods for accommodating them will generally be the same across estimators, for example variance in fish-days will first be estimated and then the additional sources of uncertainty will be incorporated using techniques such as analytical approaches based on the delta method (Millar et al., 2012), bootstrapping (Parken et al., 2003) or Bayesian methods. This chapter therefore focuses on the estimation of uncertainty in fish-days only, as this is the limitation of periodic count data, and is also where most differences between estimators will exist. Consequently, inferences from this study are expected to be general to situations in practice where uncertainty in v and l is addressed, and the absence of their uncertainty herein reflects the desire for clarity in the most important differences between estimators.

6.2.2 AUC estimators

6.2.2.1 AUC using a Gaussian spawner model

The GAUC method was proposed by Millar et al. (2012) as a very similar, but more easily (and reliably) implemented alternative to the HARR estimator of Hilborn et al. (1999). If the counts are again denoted c, then this estimator assumes that the expected count of spawners on day d, follows the Gaussian curve

$$\mathrm{E}(c_d) = a \exp\left(-\frac{(d-m)^2}{2\tau^2} \right) \tag{6.3}$$

where a is the maximum height of the curve, m is the day of peak spawners, and τ is the standard deviation.

Fish-days is the area under the curve of spawner abundance, and because the normal density function integrates to unity the GAUC estimator of fish-days, F_G, is simply

$$F_G = a\int_{-\infty}^{\infty} \exp\left(-\frac{(d-m)^2}{2\tau^2}\right)\mathrm{d}d = a\sqrt{2\pi}\tau \ .$$

Calculation of $\widehat{F}_G$ is facilitated by re-expressing (6.3) as a log-linear model, by taking logs on both sides. Estimation occurs from a sample of counts which are now denoted c_i, where $i = 1, ..., n$ indexes the sampling occasion and d_i is the day-of-season on which the count is made. The log-linear model with a quadratic term is thus given by

$$\begin{aligned} \log\left(\mathrm{E}\left(c_i\right)\right) &= \log a - \frac{(d_i-m)^2}{2\tau^2} \\ &\equiv \beta_0 + \beta_1 d_i + \beta_2 d_i^2 \ , \end{aligned} \tag{6.4}$$

where the regression coefficients correspond to

$$\begin{aligned} \beta_0 &= \log a - \frac{m^2}{2\tau^2} \\ \beta_1 &= \frac{m}{\tau^2} \\ \beta_2 &= -\frac{1}{2\tau^2} \ . \end{aligned}$$

The above equations can easily be inverted to obtain a and τ as functions of β_0, β_1 and β_2, and allow $\widehat{F}_G$ to be expressed as a function of the estimated regression coefficients,

$$\widehat{F}_G = \sqrt{\frac{\pi}{-\widehat{\beta}_2}}\exp\left(\widehat{\beta}_0 - \frac{\widehat{\beta}_1^2}{4\widehat{\beta}_2}\right) \ . \tag{6.5}$$

Equation (6.4) is a log-linear model with regression coefficients $(\beta_0, \beta_1, \beta_2)$, which is easily implemented in a generalized linear modeling framework by assuming the counts are Poisson-distributed. The quasi-likelihood, overdispersion correction for GLMs (Millar, 2011) is utilized to allow the variance of the counts to be proportional to, rather than equal to (under the standard Poisson GLM), the expected count, e.g.

$$\mathrm{Var}(c_i) = \phi E(c_i)$$

where ϕ is an overdispersion parameter, itself calculated

$$\hat{\phi} = \frac{1}{n-3}\sum_{i=1}^{n}\frac{(c_i - \hat{c}_i)^2}{\hat{c}_i} \ .$$

Alternative approaches such as mixed models could also be considered although for each incremental increase in model complexity, the probability of methods being taken up by practitioners decreases, and so they are not considered herein. The quasi-likelihood GLM will produce ML estimates $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$, from which $\widehat{F}_G$ is obtained from (6.5). The approximate variance of this estimator, denoted $\widehat{\text{Var}}(\widehat{F}_G)$ is obtained from the covariance matrix of $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$, via application of the delta method (Millar et al., 2012).

The final step in estimating escapement involves using (6.2) to divide fish-days by the product of stream-life (l) and observer efficiency (v) (which are assumed known herein), i.e.

$$\widehat{E}_G = \frac{\widehat{F}_G}{l \times v} \ . \tag{6.6}$$

By assuming that l and v are known, the estimate of variance of the GAUC escapement estimator is given by

$$\widehat{\text{Var}}(\widehat{E}_G) = \frac{\widehat{\text{Var}}(\widehat{F}_G)}{l^2 v^2} \ . \tag{6.7}$$

Note that in practice uncertainty in the estimation of l and v can be readily be incorporated at this stage using a variety of methods suggested in Section 6.2.1.3, and by Parsons and Skalski (2009) and Millar et al. (2012).

6.2.2.2 Trapezoidal AUC with model-based variance estimator

Trapezoidal area-under-the-curve (Irvine et al., 1992) is the most frequently applied of the AUC methods, and is calculated

$$\widehat{F}_T = 0.5\sum_{i=1}^{n-1}(d_{i+1} - d_i)(c_i + c_{i+1}) \tag{6.8}$$

$$= 0.5\left[(d_2 - d_1)c_1 + (d_n - d_{n-1})c_n + \sum_{i=2}^{n-1}(d_{i+1} - d_{i-1})c_i\right] \tag{6.9}$$

where the c_i are again the counts on the $i = 1, ..., n$ sampling occasions and d_i are the days of these counts. If c_1 and c_n are fixed zeros then this reduces to

$$\widehat{F}_T = 0.5 \sum_{i=2}^{n-1} (d_{i+1} - d_{i-1}) c_i \ , \tag{6.10}$$

and so if the c_i are independent

$$\mathrm{Var}(\widehat{F}_T) = 0.25 \sum_{i=2}^{n-1} (d_{i+1} - d_{i-1})^2 \mathrm{Var}(c_i)$$

which is the variance formula in Irvine et al. (1992). An obvious estimator of $\mathrm{Var}(\widehat{F}_T)$ is

$$\widehat{\mathrm{Var}}(\widehat{F}_T) = 0.25 \sum_{i=2}^{n-1} (d_{i+1} - d_{i-1})^2 \widehat{\mathrm{Var}}(c_i) \ . \tag{6.11}$$

However, until the recent proposal of Millar and Jordan (In Press), a valid estimator of $\widehat{\mathrm{Var}}(c_i)$ has been elusive. They showed that by fitting the GLM model using (6.4) in the GAUC approach then $\widehat{\mathrm{Var}}(c_i)$ could be obtained as

$$\widehat{\mathrm{Var}}(c_i) = \widehat{\phi} \widehat{c}_i$$

and, (6.11) can be simplified to

$$\widehat{\mathrm{Var}}(\widehat{F}_T) = 0.25 \, \widehat{\phi} \sum_{i=2}^{n-1} (d_{i+1} - d_{i-1})^2 \, \widehat{c}_i \ .$$

6.2.3 Average spawner estimators

6.2.3.1 AS estimators for systematic samples

The average spawner (AS) method has rarely been used in practice, and details of its implementation are scarce. For instance, it is not immediately obvious how to handle zeros or how to define season-length. Counts are again denoted c_i where $i = 1, ..., n$ indexes the sampling occasion. To be consistent with AUC methods, it seems logical to assume that $N = d_n - d_1$, for a set of n periodic counts, and d_1 and d_n are the day-of-season of tail counts c_1 and c_n, which equal zero. For demonstration purposes it is assumed that the sampling interval $d_{i+1} - d_i$ is constant throughout the season, but this is not essential.

A natural AS estimator of fish days for systematic samples (F_{sys}) is

$$\widehat{F}_{sys} = \bar{c}N = \frac{N}{n}\sum_{i=1}^{n} c_i = \frac{N}{\sum_{i=1}^{n} p_i}\sum_{i=1}^{n} p_i c_i \tag{6.12}$$

where $p_i = d_{i+1} - d_{i-1}$, as sampling intervals are assumed to be constant $p_i = p_{i+1} = ... = p_n \equiv p$. Note that this is the estimator proposed by Parsons and Skalski (2009) as alternative to (6.9), in an attempt to overcome the difficulties of estimating a variance for the latter. Because the systematic sample is assumed to have zeros at either tail, the contributions of the first and last counts, c_1 and c_n, to the sum in (6.12) are $p \times c_1$ and $p \times c_n$ respectively, from which it is immediate that $p/2$ days before the 1st count and $p/2$ days after the last count are included in the calculation (Fig 6.1). This amounts to the spawning season being extended beyond the usual definition of its limits ($d_n - d_1$), or equivalently, all counts are being considered (erroneously) to have equal weighting. The tail counts should have half the weighting of the others otherwise the estimator will be negatively biased.

The AS estimator will be unbiased if weights for non-tail counts are twice those for the 1st and last counts, i.e. $w_i = b$ for $i = 2, 3, ..., n-1$ and $w_i = b/2$ for $i = 1$ and n. The estimator of fish-days is now given by

$$\widehat{F}_{sys} = \bar{c}_w N \ , \quad \text{where} \quad \bar{c}_w = \frac{\sum_{i=1}^{n} w_i c_i}{\sum_{i=1}^{n} w_i} \tag{6.13}$$

where $\bar{c}_w$ is the weighted mean of the counts.

The corrected AS estimator in (6.13) is now equivalent to the TAUC estimator (F_T) for a constant sampling interval, because if we let $w_i = 2p$ for $i = 2, 3, ..., n-1$, and $w_i = p$ for $i = 1$, then

$$\begin{aligned}
\widehat{F}_{sys} &= \sum_{i=1}^{n} w_i c_i \\
&= p(c_1 + c_n) + \sum_{i=2}^{n-1} 2pc_i \\
&= 0.5\left[(d_2 - d_1)c_1 + (d_n - d_{n-1})c_n + \sum_{i=2}^{n-1}(d_{i+1} - d_{i-1})c_i\right]
\end{aligned}$$

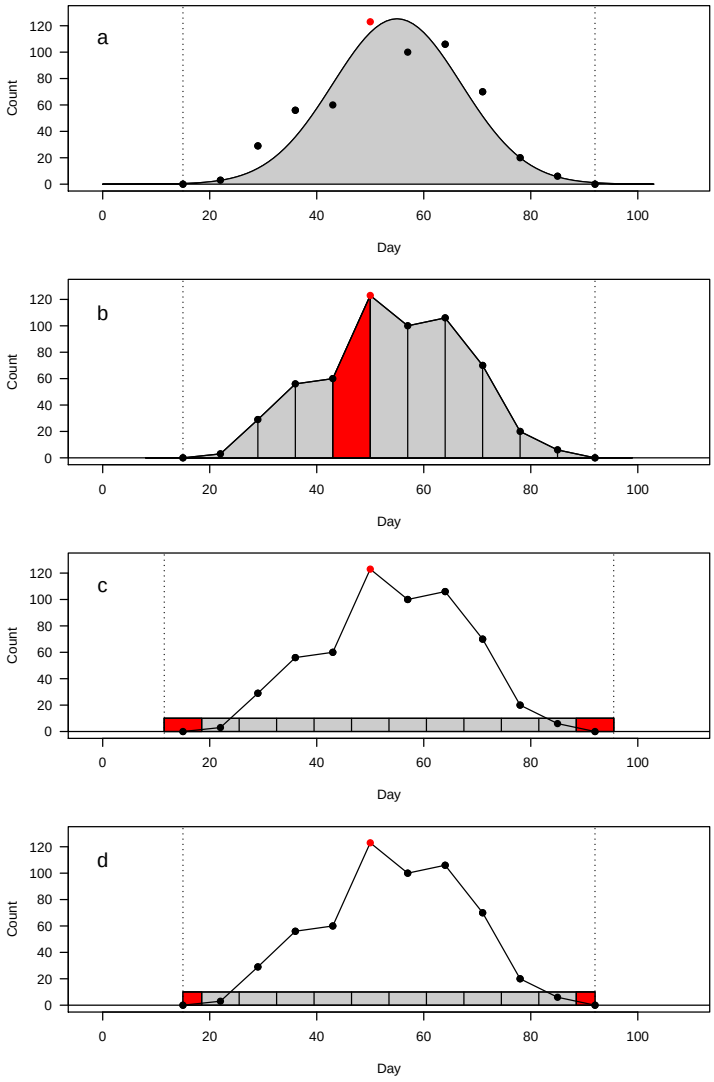

Figure 6.1: A hypothetical example of a periodic count dataset and visual display of estimators of fish-days. Black dots are the counts, the red dot is the peak count estimate and the dotted vertical lines are the implicit limits of the spawning season. The line in (a) is the fitted curve using the GAUC method and the gray area under the curve is the estimate of fish-days. The TAUC method in (b) sums the area of individual trapeziums (an example is given in red) under the curve connecting the counts to give an estimate of fish-days. The AS method in (c) gives equal weights for each count which implicitly extends the spawning season beyond the 1st and last count. The weights are indicated by the width of the gray blocks and the red blocks are the inappropriate weights of the tail counts. The AS method with the tail-counts correctly weighted at half the weights of the internal counts is given by (d).

which is identical to (6.9). If sampling intervals vary, the results hold, but the weights must be adjusted accordingly if the estimator is to equivalent to F_T.

Estimating $\text{Var}(\widehat{F}_{sys})$ is difficult given the absence of unbiased estimators available for systematic samples (Wolter, 2007). However, several methods of overcoming this deficiency have been proposed in other fields (Wolter, 2007; Fewster, 2011), and also within the salmon escapement literature (Skalski et al., 1993; Reynolds et al., 2007; Skalski et al., 2009), though it should be noted that these latter applications differ significantly from the periodic count datasets addressed herein.

The most obvious estimator assumes that the systematic sample can be treated as a simple (weighted) random sample. It is assumed that the variance of each count is equal, $\text{Var}(c_1) = \text{Var}(c_2) =, ..., = \text{Var}(c_n)$, and the weights are normalized to sum to one, and so the associated variance of fish-days is given by

$$\widehat{\text{Var}}_1(\widehat{F}_{sys}) = N^2\widehat{\text{Var}}\left(\sum_{i=1}^{n} w_i c_i\right) = N^2\sum_{i=1}^{n} w_i^2\widehat{\text{Var}}(c_i) = N^2\frac{\sum_{i=1}^{n}(c_i-\bar{c})^2}{n-1}\sum_{i=1}^{n} w_i^2 \ .$$

This estimator is known to be poor in cases when there is significant correlation structure in the sample (Wolter, 2007; Millar and Olsen, 1995), as can be expected to occur in salmon periodic count data.

A series of alternative variance estimators are listed by Wolter (2007), in an effort to account for various potential forms of correlation that may be present in a sample. Two that appear to be well suited to periodic count data are Wolter (2007)'s v_2 and v_4, which can be calculated for the weighted case as

$$\widehat{\text{Var}}_2(\widehat{F}_{sys}) = N^2\sum_{i=1}^{n} w_i^2\sum_{i=2}^{n}\frac{(c_i-c_{i-1})^2}{2(n-1)} \tag{6.14}$$

$$\widehat{\text{Var}}_3(\widehat{F}_{sys}) = N^2\sum_{i=1}^{n} w_i^2\sum_{i=3}^{n}\frac{(c_i-2c_{i-1}+c_{i-2})^2}{6\,(n-2)} \ . \tag{6.15}$$

These estimators invoke a post-stratification of the systematic design the data were collected under. Both estimators assume overlapping 'stratum' to allow an increase in differences compared to non-overlapping designs (such as Wolter's v_3). While several other variance estimators may

appear to have potential for periodic count data, e.g. estimators based on higher order contrasts than (6.14)–(6.15), or non-overlapping designs, they are not considered in this study owing to concerns that they lack the degrees of freedom necessary to provide valid estimates for periodic count data, that can at times be sparse.

6.2.3.2 AS estimators for random samples

If a simple random sampling scheme is used to gather the periodic count data, rather than a systematic design, then the estimator of fish-days $\widehat{F}_{rs}$ is simple to calculate, and does not suffer from the biasness of (6.12). Using the same notation as the systematic design (Section 6.2.3.1), fish-days $\widehat{F}_{rs}$ is given by

$$\widehat{F}_{rs} = \bar{c}N \ , \quad \text{where} \quad \bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i$$

and the estimate of variance is given by

$$\widehat{\text{Var}}(\widehat{F}_{rs}) = N^2 \frac{s^2}{n}$$

where s^2 is simply the sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (c_i - \bar{c})^2 \ .$$

6.2.3.3 AS estimators for stratified random samples

If periodic count data are collected based on a stratified random sampling design, and the stratum are denoted k for $1, ..., K$, n_k and N_k are number of days sampled and the total number of days both in stratum k, respectively, $c_{i,k}$ are the counts for sample $i = 1, ..., n_k$ in stratum k and s_k^2 is the sample variance for the n_k samples in stratum k. An estimate of total fish-days $\widehat{F}_{srs}$ is then calculated

$$\widehat{F}_{srs} = \sum_{k=1}^{K} N_k \bar{c}_k \ , \quad \text{where} \quad \bar{c}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} c_{i,k}$$

where $\bar{c}_k$ is the stratum mean count, and variance is given by

$$\widehat{\text{Var}}(\widehat{F}_{srs}) = \sum_{k=1}^{K} N_k^2 \frac{s_k^2}{n_k}$$

where s_k^2 is the sample variance for the n_k samples in stratum k.

$$s_k^2 = \frac{1}{n_k - 1} \sum_{k=1}^{n_k} (c_{i,k} - \bar{c}_k)^2 \ . \tag{6.16}$$

6.2.4 Peak count method

The peak count escapement estimate is simply

$$E_p = \max\{c_1, ..., c_n\}$$

where $c_1, ..., c_n$ are the periodic counts for surveys $i = 1, ..., n$. Note that although the estimator is denoted E_p this is an index of escapement only, not an absolute estimate. No estimate of variance of escapement is available for this estimator.

6.2.5 Estimating escapement and construction of confidence intervals

For each of the AUC and AS estimators in Sections 6.2.2.1–6.2.3.3, the appropriate estimates of fish-days ($\widehat{F}$) were substituted into (6.6) instead of F_G, to estimate escapement. The estimates of fish-days variance, $\widehat{\text{Var}}(\widehat{F})$, were substituted into (6.7) instead of $\widehat{\text{Var}}(\widehat{F}_G)$, to allow estimation of variance of the escapement estimate. Estimates of escapement and its variance for each technique were denoted; TAUC - E_T and $\widehat{\text{Var}}(\widehat{E}_T)$ (Section 6.2.2.2), AS for systematic samples - $\widehat{E}_{sys}$ and $\widehat{\text{Var}}(\widehat{E}_{sys})$ (Section 6.2.3.1), AS for simple random samples - $\widehat{E}_{rs}$ and $\widehat{\text{Var}}(\widehat{E}_{rs})$ (Section 6.2.3.2), AS for stratified random samples - $\widehat{E}_{srs}$ and $\widehat{\text{Var}}(\widehat{E}_{srs})$ (Section 6.2.3.3).

Confidence intervals (95%) were constructed as $\widehat{E} \pm t_{df} \sqrt{\widehat{\text{Var}}(E)}$, where t_{df} is the appropriate student-t quantile calculated using degrees of freedom specific to the estimators. For $\widehat{\text{Var}}_1(\widehat{E}_{sys})$, $\widehat{\text{Var}}_2(\widehat{E}_{sys})$, $\widehat{\text{Var}}_3(\widehat{E}_{sys})$ and $\widehat{\text{Var}}(\widehat{E}_{rs})$ degrees of freedom were given by $df = n - 1$, while for $\widehat{\text{Var}}(\widehat{E}_T)$ and $\widehat{\text{Var}}(\widehat{E}_G)$ they were given by $df = n - 3$.

The Satterthwaite approximation was utilized for $(\widehat{E}_{srs}, \widehat{\text{Var}}(\widehat{E}_{srs}))$ as variances for each strata are unlikely to be equal for periodic count data. These approximations were calculated

$$df = \left(\sum_{k=1}^{K} a_k s_k^2 \right)^2 \Big/ \sum_{k=1}^{K} \frac{\left(a_k s_k^2\right)^2}{(n_k - 1)} \quad \text{where} \quad a_k = \frac{N_k \left(N_k - n_k\right)}{n_k}$$

where the s_k^2 are the variances of counts from stratum k, which are defined in (6.16).

6.2.6 Evaluation using simulation

6.2.6.1 Baseline simulation

Periodic count data were generated under a model similar to that given by Millar et al. (2012). Normally distributed arrival times were assumed such that the cumulative number of fish entering (S_d^A) and dying (S_d^D), in the hypothetical study reach on day d were

$$S_d^A = E \int_0^d \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(u-m)^2}{2\tau^2}\right) \mathrm{d}u \tag{6.17}$$

$$S_d^D = E \int_0^{d-l} \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(u-m)^2}{2\tau^2}\right) \mathrm{d}u \tag{6.18}$$

respectively, where E is escapement, l is a constant stream-life and m and τ are the mean and standard deviation of the arrival time distribution (in days). The number of fish present on day d is calculated $S_d^P = S_d^A - S_d^D$ and $\boldsymbol{S^P} = (S_1^P, ..., S_{N^*}^P)$ is a vector of abundances over the N^* days in the simulated spawning season. N^* was set at 365 days to ensure the season covered the entire range of days where fish predicted by (6.17)–(6.18) were non-negligible. A fixed stream-life of 10 days was chosen for all scenarios as a roughly average salmon stream-life, given meta-analyses by Bue et al. (1998), Lady and Skalski (1998) and Holt and Cox (2008). It should be noted that escapement estimates are usually corrected for observer efficiency, v, however this parameter was set at unity for all scenarios as it influences all estimators equally, and thus can be ignored without unduly affecting comparisons.

The distributions of m and τ were based upon the posterior distributions produced in the Bayesian meta-analysis of pink salmon escapement of Su et al. (2001). Parameter values for individual simulations were generated according to their hierarchical model such that

$$\begin{aligned} m &\sim N(\mu_m, \gamma_m^2) \ , \\ \tau &\sim \mathrm{LN}(2.6, 0.01) \ , \end{aligned} \tag{6.19}$$

where LN denotes a log-normal distribution, and

$$\begin{aligned} \mu_m &\sim N(208, 2.25) \ , \\ \gamma_m^2 &\sim \chi_{22}^2 \ , \end{aligned}$$

where χ^2_{22} denotes a chi-square distribution with 22 degrees of freedom. Note that the expected day of peak arrival is day-of-year 208, which corresponds to 27 July in a non-leap year. This was the peak arrival for pink salmon from Su et al. (2001) but for simulations the date of the year is essentially arbitrary and will not affect inferences for coho or other salmon species.

Escapement levels were set at $E = 100$, 1 000 and 10 000 for the main set of simulation scenarios. These cover situations ranging from scarce to very productive populations (or large study areas). Comparative performance of estimators is not expected to change substantially for $E > 10\,000$, e.g. in situations representing super-abundant populations such as pink or sockeye salmon in various river systems (e.g. Hilborn, 2006; Holt and Peterman, 2007), and this was confirmed with preliminary simulations.

'Observable' stream counts were simulated on all days-of-year (1–365), and were denoted $\boldsymbol{c^*} = (c^*_1, ..., c^*_{365})$. These are the vector of counts from which the periodic count sample was selected using the sampling designs outlined in Section 6.2.6.3. This is a much longer duration of sampling than would occur in practice, but it ensures that the sampling period covers the full period where salmon are present, for all simulations. These simulated counts were generated by adding a sub-model representing the process of sampling true abundance, S^p_d, on each day d, by way of visual survey methods. The sampling process was set such that the variance of the generated counts was proportional to their mean, by simulating c^*_d from a negative binomial distribution with mean S^p_d and an appropriately chosen size parameter calculated $\Omega_d = S^p_d/(q-1)$. This is in contrast to the more usual form where the variance is a quadratic function of the mean (e.g. Millar, 2011). Several different values of the proportionality constant were considered, though in the main set of scenarios these were $q = 1, 2$ and 5 for escapements of 100 or under, and $q = 1, 10$ and 100 for escapements over 100. Examples of the relationships between variation in the simulated data and the expected count are shown in Figure 6.2 for each value of q. These values were chosen to allow simulated datasets to be similar to the wide range of datasets observed in practice.

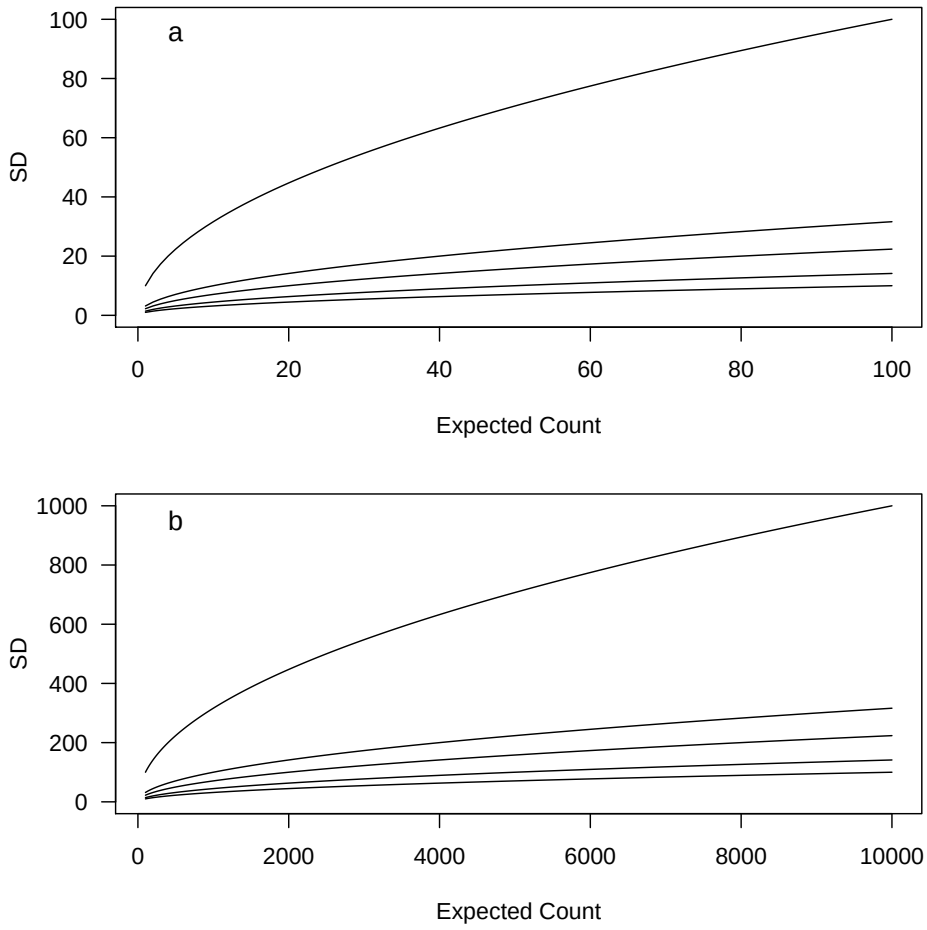

Figure 6.2: Relationship between standard deviation (SD) of the negative binomial sampling process model and the expected count. Panel (a) shows the relationships for low expected counts (1–100) while (b) shows the same relationships at high expected abundance (>100). Lines from lowest to highest in both panels represent distributions with q=1,2,5,10 and 100.

6.2.6.2 Multimodal arrival distribution

Multimodality in peak count datasets are often observed for a variety of salmon species (e.g. Holt and Cox, 2008; Szerlong and Rundio, 2008), and are generally attributed to clustered entry of fish to the study reach related to changes in environmental conditions, such as flow rate. To account for this phenomenon, the arrival distribution (6.17) was modified to be a mixture of two or more normal distributions. Components of the mixture are indexed by $h = 1, ..., H$, and the proportions of the total escapement occurring in each component were determined randomly, by simulating $H - 1$ iid Uniform(0,1) random variables and ordering them from smallest to largest,

$$0 \equiv U_{[0]} < U_{[1]} < U_{[2]} < ... < U_{[H-1]} < U_{[H]} \equiv 1 \ ,$$

where $U_{[j]}, j = 1, ..., H-1$ denotes the jth smallest. The hth arrival pulse contributed escapement E_h, given by

$$E_h = (U_{[h]} - U_{[h-1]})E \ , \quad h = 1, ..., H \ .$$

The largest arrival pulse had an arrival time distribution as specified in the unimodal case (6.19). The lesser components were distributed such that the date of peak arrival, m_h, had a standard deviation of a little over two weeks,

$$m_h \sim N(208, 15^2) \ . \tag{6.20}$$

The standard deviation of arrival times was set proportional to E_h such that the largest pulse would have the same standard deviation as the baseline case (6.19), and the smaller pulses would tend to be of shorter duration,

$$\tau_h \sim \frac{E_h}{E_{\max}} \mathrm{LN}(2.6, 0.1^2) \ ,$$

where $E_{\max}$ is the escapement from the largest arrival pulse. Across the scenarios explored h varied between one and four.

6.2.6.3 Sampling schemes

The next step in the simulations was to select a sample of counts from the vector of potentially observable counts ($\boldsymbol{c}^*$) to which to apply the candidate estimators. For each replication of the

simulation model, three datasets were selected and stored; a systematic (SYS), a simple random (RS) and a stratified random sample (SRS). In general the term periodic count refers to a systematic, or at least roughly systematic sample of counts. However, as pointed out by Parsons and Skalski (2009), there may be advantages to randomly selecting a sample of counts in either a simple or stratified manner. The SYS sample was simply selected as every pth count from day t=1 onwards. To ensure that the RS sample size was roughly equivalent to the SYS sample, a sample of size $n = N^*/p$ (rounded appropriately) counts were randomly selected from $\boldsymbol{c}^*$. To achieve an equivalent-sized SRS sample, $\boldsymbol{c}^*$ was split into stratum of size $2N^*/p$ days and two samples were randomly selected within each stratum.

The systematic sample is the main focus herein as it is the most common design in practice. These datasets were analyzed with estimators E_T, E_G and E_{sys} (with the three associated variants of the variance estimator for the latter; $\widehat{\text{Var}}_1(\widehat{E})$, $\widehat{\text{Var}}_2(\widehat{E})$ and $\widehat{\text{Var}}_3(\widehat{E})$). The estimator E_{rs} was applied to the simple random samples, and E_{srs} was applied to the stratified random samples. Thus in the results section the estimates $\widehat{E}_{rs}$ and $\widehat{E}_{srs}$ are calculated from different data to the other estimators, though each dataset is selected from the same vector of potentially observable counts, $\boldsymbol{c}^*$.

Due to the unrealistically long season-length assumed for most scenarios, many zero counts usually occur before and after the true spawning season in the simulated datasets. Excess zeros were trimmed to leave only one zero count immediately before the first non-zero count, and one after the last non-zero count, for all three sampling schemes (SYS, RS and SRS). For simulated datasets where an even number of counts were selected in the SRS sample, each stratum had two counts, i.e. $n_k = 2$. For simulated datasets with an odd number of counts selected, the first stratum accommodated the extra count and the size of this stratum in days was extended to be 1.5 times the size of the other strata (e.g. $N_1 = 1.5N_k$, and $n_1 = 3$).

6.2.6.4 Parameter Scenarios

Given the number of different values to be investigated for each of the four parameters that were allowed to vary between scenarios, a factorial design was utilized to allow for all combinations of parameter values, resulting in 81 scenarios. These cover variations in $E = \{100, 1\,000, 10\,000\}$, $q = \{1, 10, 100\}$, $H = \{1, 2, 3\}$ and $p = \{3, 7, 11\}$ and form the basis of the evaluations of all estimators (for SYS, RS and SRS samples).

Further scenarios were investigated for a subset of estimators. For instance the implications of non-constant p was only investigated on estimators fitted to systematic sample data. Non-constant p was achieved by allowing time between surveys to randomly vary between a minimum of three and maximum of 11 days, with a mean of seven days. This was accomplished by setting $d_i - d_{i-1} = 3 + \epsilon_i$, where $\epsilon_i \sim \text{Binomial}(8, 0.5)$.

A further subset of scenarios was simulated for only systematic samples and the associated estimators to investigate estimator performance at very low abundance of spawners. This aimed to evaluate whether there are practical difficulties of fitting ML-based estimators, such as E_G (and consequently, E_T due to it relying on E_G to calculate $\widehat{\text{Var}}(\widehat{E}_T)$), and whether AS estimators for systematic samples may be more stable in these situations. Escapement was set at $E = \{25, 50\}$ to achieve this, with the other parameters varied being; $q = \{1, 100\}$, $H = \{1, 3\}$ and $p = \{3, 7, 11\}$. These scenarios were not investigated for the simple and stratified random samples as E_{rs} and E_{srs} do not have convergence issues for any datasets.

6.2.6.5 Measurement of estimator performance

Estimators were compared across $n^* = 10\,000$ simulated datasets for each scenario. The basis of comparisons were statistics of estimator performance that included, 1) mean bias measured across simulations as $\bar{B}(\widehat{E}) = 1/n^* \sum_{i=1}^{n^*} \widehat{E}_i - E$, where $\widehat{E}_i$ is the estimate of escapement for the particular estimator in question, for simulation i, and E is the true escapement from which the dataset was simulated, 2) root mean squared error (RMSE), which is the expected value of the squared difference between $\widehat{E}$ and E, and can be calculated RMSE $= \sqrt{\bar{B}(\widehat{E})^2 + \text{Var}(\widehat{E})}$,

where $\mathrm{Var}(\widehat{E})$ is the variance of the escapement estimate across the simulations, 3) coverage of confidence intervals of each estimator, where the intervals are estimated as in Section 6.2.5 and coverage is defined as the proportion of simulations where the interval contains the true escapement E, 4) mean half-width of the confidence intervals.

For many scenarios E_G would produce a small number of unstable results, generally overestimating E, and thus inflating $\bar{B}(\widehat{E})$ and $\mathrm{Var}(\widehat{E})$. For comparisons of estimators that included E_G, alternative measures of estimator performance were used; median bias, median confidence interval half-width and median absolute deviation which is calculated $MAD = \text{median}\left(|\widehat{E}_i - E|\right)$.

6.3 Results

6.3.1 Comparison of AS method variance estimators for systematic samples

The AS variance estimator most suitable for systematic samples was selected from $\widehat{\mathrm{Var}}_1(\widehat{E}_{sys})$, $\widehat{\mathrm{Var}}_2(\widehat{E}_{sys})$ and $\widehat{\mathrm{Var}}_3(\widehat{E}_{sys})$ before comparing $\widehat{E}_G$, $\widehat{E}_T$ and $\widehat{E}_{sys}$. With the exception of a restricted set of scenarios (E=1 000 and high q), $\widehat{\mathrm{Var}}_1(\widehat{E}_{sys})$ performed poorly on the simulated data (Table 6.1). Median confidence interval half-widths for this estimator were almost always the highest among the three. This resulted in large coefficients of variation for this estimator, generally in the vicinity of ~ 0.25, and for a large proportion of scenarios coverage rates were either 100% or very close to it, far higher than the nominal rate.

Across the range of scenarios considered, $\widehat{\mathrm{Var}}_3(\widehat{E}_{sys})$ consistently performed better than the alternatives, displaying the most narrow median confidence interval half-widths and coverage rates closest to the nominal rate (Table 6.1). The later was mostly due to coverage of $\widehat{\mathrm{Var}}_3(\widehat{E}_{sys})$ being lower than the excessively conservative confidence intervals of the other two estimators, although when confidence intervals were too narrow (as occurred in some scenarios with low escapement (E=100) and/or very high overdispersion), coverage of $\widehat{\mathrm{Var}}_3(\widehat{E}_{sys})$ was often furthest from the nominal rate. Although the differences between estimators were generally slight in these scenarios. In general however, confidence intervals were still too conservative, with coverage rates still above nominal when overdispersion was low to medium and escapement was 1 000 or

Table 6.1: Comparison of three estimators ($\widehat{\theta}$) of variance of AS estimates of escapement for systematic samples of periodic counts. Datasets were simulated in scenarios where the number of modes of the arrival distribution of spawners (H) varied between one and three, and overdispersion of the sampling process q varied between one and 100. Performance measures are the median standard error, median 95% confidence interval half-width and percentage coverage of confidence intervals across simulations for different scenarios.

			E=100		E=1 000		E=10 000	
H	q	$\widehat{\theta}$	HW	Cvg	HW	Cvg	HW	Cvg
1	low	$\widehat{\mathrm{Var}}_1(\widehat{E})$	57.5	100.0	596.8	100.0	6163.7	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	27.4	99.5	213.0	100.0	2041.7	100.0
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	19.1	94.8	94.2	99.7	750.4	100.0
1	med	$\widehat{\mathrm{Var}}_1(\widehat{E})$	58.4	100.0	593.6	100.0	6065.7	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	32.0	98.0	270.0	99.6	2116.0	100.0
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	25.5	93.8	187.9	95.0	932.1	99.7
1	high	$\widehat{\mathrm{Var}}_1(\widehat{E})$	62.2	98.7	695.6	95.7	6025.9	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	42.9	94.7	545.7	91.5	2692.4	99.6
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	38.6	91.2	520.3	89.1	1878.3	94.6
2	low	$\widehat{\mathrm{Var}}_1(\widehat{E})$	52.7	100.0	550.7	100.0	5722.2	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	27.8	99.2	221.1	99.8	2146.3	99.8
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	21.1	96.3	126.6	99.2	1172.1	99.5
2	med	$\widehat{\mathrm{Var}}_1(\widehat{E})$	53.8	99.8	546.5	100.0	5619.7	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	32.6	98.0	273.7	99.3	2216.6	99.8
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	27.2	94.5	207.2	96.1	1274.6	99.2
2	high	$\widehat{\mathrm{Var}}_1(\widehat{E})$	58.4	98.3	666.2	95.4	5546.0	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	43.0	95.1	550.9	92.1	2733.9	99.3
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	39.5	92.3	527.2	90.6	2073.5	96.5
3	low	$\widehat{\mathrm{Var}}_1(\widehat{E})$	49.4	100.0	512.6	100.0	5350.8	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	27.3	99.1	219.9	99.8	2124.5	99.9
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	21.5	96.7	136.7	99.4	1256.5	99.5
3	med	$\widehat{\mathrm{Var}}_1(\widehat{E})$	50.7	99.7	512.6	100.0	5242.3	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	32.0	97.7	271.5	99.1	2184.2	99.8
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	27.3	94.7	214.5	96.6	1356.5	99.4
3	high	$\widehat{\mathrm{Var}}_1(\widehat{E})$	55.7	98.1	640.7	95.0	5242.2	100.0
		$\widehat{\mathrm{Var}}_2(\widehat{E})$	42.7	94.7	540.8	92.0	2725.8	99.1
		$\widehat{\mathrm{Var}}_3(\widehat{E})$	39.6	92.4	520.9	90.3	2141.2	96.4

Notes: Parameter estimates are calculated over 10 000 simulations for each scenario, and coverage is the percentage of simulations where the true escapement was contained within the confidence interval of the estimator. The sampling interval was set at seven days for all scenarios presented. Low, medium and high overdispersion was set at 1, 2 and 5 for E=100, and 1, 10 and 100 for $E > 100$.

10 000. Estimator $\widehat{\mathrm{Var}}_3(\widehat{E}_{sys})$ was consequently selected as the variance estimator for $\widehat{E}_{sys}$ when comparing its performance with AUC and random sample AS estimators (Table 6.2).

6.3.2 Comparison of estimators of escapement and its variance

Bias was low for all estimators across nearly all scenarios investigated (Table 6.2). Consequently, MAD was largely determined by the variability of each estimator across simulations. The relative MAD of the estimators did not change substantially over the range of escapement levels simulated.

Escapement level did however strongly influence the coverage rates of all estimators, though especially the estimators for systematic samples ($\widehat{E}_T$, $\widehat{E}_{sys}$, $\widehat{E}_G$). In general coverage of most estimators was slightly lower then the nominal rate when E=100 (except when q=1). The exception was E_{srs} which had coverage rates consistently higher than 95%. Coverage of all estimators increased substantially at higher escapements, to the point where confidence intervals were often far too conservative. When $E = 10\,000$ coverage rates for all estimators, for almost all scenarios, were higher than the nominal rate, often by a significant margin. The exception was $\widehat{E}_{rs}$ which had confidence intervals very close to the nominal rate across all scenarios at that escapement level.

Overdispersion had perhaps the strongest influence on estimator performance with higher overdispersion leading to increased bias and MAD, and lower coverage rates, for all estimators. This was particularly evident when escapement was low. In contrast, increasing the number of modes in the arrival distribution of spawners from one through three did not substantially affect bias or the MAD of any of the estimators. However, coverage rates generally increased with additional modes for $\widehat{E}_T$ and $\widehat{E}_G$, but remained relatively stable for the AS estimators. These patterns relating to modality were relatively consistent across the different levels of escapement.

Overall it is difficult to identify a best-performing estimator from this set of simulation scenarios, with each displaying different characteristics that resulted in their comparative performance being scenario-specific. In general, estimator performance was relatively similar within-scenarios

Table 6.2: Performance of escapement estimators ($\widehat{\theta}$) with different numbers of modes in the arrival distribution H, and levels of overdispersion q. Performance measures are median bias, MAD of the estimator and coverage of confidence intervals.

			E=100			E=1 000			E=10 000		
H	q	$\widehat{\theta}$	Bias	MAD	Cvg	Bias	MAD	Cvg	Bias	MAD	Cvg
1	low	$\widehat{E}_T$	0.1	5.7	94.8	-0.4	17.8	94.9	0.9	56.2	94.6
		$\widehat{E}_{sys}$			94.8			99.7			100.0
		$\widehat{E}_G$	0.1	5.7	94.8	-0.3	17.8	94.8	1.0	56.3	94.7
		$\widehat{E}_{rs}$	1.5	17.7	95.4	4.5	175.8	95.6	-45.2	1859.6	95.4
		$\widehat{E}_{srs}$	0.1	7.6	98.2	-0.4	54.2	98.7	-19.4	501.9	98.7
1	med	$\widehat{E}_T$	-0.6	7.8	93.7	-0.4	56.3	92.7	4.4	178.7	93.0
		$\widehat{E}_{sys}$			93.8			95.0			99.7
		$\widehat{E}_G$	-0.1	8.0	93.9	-0.2	55.9	92.7	5.2	178.8	93.2
		$\widehat{E}_{rs}$	1.5	17.8	94.9	19.1	180.8	95.6	12.6	1796.9	95.3
		$\widehat{E}_{srs}$	-0.6	9.7	98.0	-3.2	75.3	98.1	-2.6	536.0	98.8
1	high	$\widehat{E}_T$	-1.3	12.7	91.9	-15.8	177.5	87.5	-31.6	576.1	90.9
		$\widehat{E}_{sys}$			91.2			89.1			94.6
		$\widehat{E}_G$	-0.4	12.8	92.2	-9.1	178.5	88.1	-29.5	575.5	90.9
		$\widehat{E}_{rs}$	0.7	20.0	93.7	-4.8	233.6	92.2	148.0	1790.9	95.4
		$\widehat{E}_{srs}$	-1.3	13.9	97.1	-23.5	181.6	95.9	-27.8	752.3	98.5
2	low	$\widehat{E}_T$	-0.3	5.7	97.9	-0.4	18.6	99.3	-1.8	66.0	99.9
		$\widehat{E}_{sys}$			96.3			99.2			99.5
		$\widehat{E}_G$	0.7	5.8	98.0	3.1	19.3	99.3	12.8	71.6	99.9
		$\widehat{E}_{rs}$	1.1	15.9	95.1	7.9	161.4	95.7	-110.6	1636.3	95.1
		$\widehat{E}_{srs}$	-0.6	7.6	98.3	-1.8	52.2	98.5	-20.1	509.8	98.3
2	med	$\widehat{E}_T$	-0.6	7.8	96.7	-1.8	56.4	97.0	-4.7	182.5	99.3
		$\widehat{E}_{sys}$			94.5			96.1			99.2
		$\widehat{E}_G$	0.8	8.0	97.0	4.5	58.1	97.1	20.4	187.4	99.2
		$\widehat{E}_{rs}$	0.8	16.3	95.1	13.9	165.2	95.4	-68.1	1640.1	95.7
		$\widehat{E}_{srs}$	-0.6	9.7	98.0	-4.6	77.3	98.4	-27.8	548.3	98.4
2	high	$\widehat{E}_T$	-0.6	12.5	94.0	-20.0	174.7	90.0	-19.4	557.7	96.6
		$\widehat{E}_{sys}$			92.3			90.6			96.5
		$\widehat{E}_G$	1.3	12.9	94.8	4.7	179.9	91.1	28.4	567.1	96.7
		$\widehat{E}_{rs}$	0.6	19.2	93.7	-10.0	231.0	91.6	45.8	1703.2	95.5
		$\widehat{E}_{srs}$	-1.3	13.9	96.9	-27.7	188.6	95.7	-41.1	760.4	98.3
3	low	$\widehat{E}_T$	-0.1	5.7	98.8	-0.4	18.6	99.9	-1.9	71.9	100.0
		$\widehat{E}_{sys}$			96.7			99.4			99.5
		$\widehat{E}_G$	1.3	6.0	98.9	4.6	19.9	99.9	24.3	80.8	100.0
		$\widehat{E}_{rs}$	0.8	15.3	95.3	-1.6	151.2	95.5	-144.1	1493.3	95.5
		$\widehat{E}_{srs}$	0.1	7.6	98.4	-1.8	52.2	98.3	12.1	503.1	98.3
3	med	$\widehat{E}_T$	-0.6	7.8	97.0	-3.5	57.2	98.2	-1.9	187.5	99.8
		$\widehat{E}_{sys}$			94.7			96.6			99.4
		$\widehat{E}_G$	1.4	8.3	97.3	6.7	59.1	98.3	38.2	197.5	99.8
		$\widehat{E}_{rs}$	0.8	15.7	94.8	6.4	154.7	95.1	-56.0	1552.4	95.7
		$\widehat{E}_{srs}$	-0.6	9.7	97.9	-4.6	76.7	98.3	-5.4	530.4	98.2
3	high	$\widehat{E}_T$	-1.3	12.7	94.7	-28.4	180.9	90.5	-13.1	569.3	97.8
		$\widehat{E}_{sys}$			92.4			90.3			96.4
		$\widehat{E}_G$	2.0	13.3	95.7	1.8	185.7	91.9	70.6	580.3	97.9
		$\widehat{E}_{rs}$	-0.1	18.0	93.4	-8.6	223.5	91.4	64.6	1572.6	95.8
		$\widehat{E}_{srs}$	-1.3	13.9	96.8	-25.6	183.1	95.6	-12.8	760.2	98.2

Notes: 10 000 simulations were undertaken for each scenario, and coverage is the percentage of simulations where the true escapement was contained within the confidence interval of the estimator. Coverage for E_{sys} was determined using confidence intervals constructed from $\widehat{\text{Var}_3}(\widehat{E})$. The sampling interval was set at seven days for all scenarios presented. Low, medium and high overdispersion was set at 1, 2 and 5 for E=100, and 1, 10 and 100 for $E > 100$.

and they displayed the same general trends across different levels of E, H and q as explained above. Perhaps the most notable general differences occur in MAD between the estimators for different sample-types. MAD was consistently highest for $\widehat{E}_{rs}$ across most scenarios, followed by $\widehat{E}_{srs}$, with the three remaining estimators for the systematic sample exhibiting the lowest values.

6.3.3 Influence of sampling interval

Comparison of estimators for different lengths of sampling interval are only displayed for a subset of the scenarios simulated (q=low and H=2; Table 6.3), as a representative sample of the full set of scenarios simulated. The patterns attributed to the sampling interval were also common to scenarios with other overdispersion levels and number of modes in arrival distribution, and their exclusion simply provides clarity in the tabular comparison of the estimators.

The estimators displayed different responses to changes in the length of sampling intervals. Coverage rates for estimators $\widehat{E}_T$, $\widehat{E}_G$ and $\widehat{E}_{rs}$ decreased when p was increased from 3 through 11. The magnitude of the decrease was dependent on the escapement scenario, decreasing by 1.5–2.8 percentage points between p of 3 and 11, for an escapement of 100, and 0.6–3.9 percentage points for an escapement of 10 000 (Table 6.3). Coverage for $\widehat{E}_{sys}$ increased when p increased from 3 to 11, by 2.7, 4.4 and 4.2 percentage points when escapement was 100, 1 000 and 10 000, respectively. Coverage for $\widehat{E}_{srs}$ increased when p increased from 3 to 11, by 2.6, 2.3 percentage points when escapement was 100 and 1 000, respectively, and decreased by 0.1 percentage point when escapement was 10 000. Whether these changes shifted coverage rates for the estimators closer or further away from the nominal rate was scenario-dependent, as coverage rates for the base-case of p=7 could be higher or lower than nominal, depending on the estimator and escapement-level.

Bias and MAD of most estimators increased by small to moderate levels when p increased from 3 to 11. This was most obvious for $\widehat{E}_{rs}$ in absolute terms, although bias and MAD were high for this estimator even when $E = 100$ and so in relative terms the change with sampling interval was similar to the other estimators.

Table 6.3: Performance of five estimators ($\widehat{\theta}$) of escapement for simulated datasets with three levels of sampling intervals, p=3, 7 and 11 days. Performance measures are median bias, median absolute deviation of the estimator and coverage of confidence intervals constructed from estimator variance. Overdispersion was set at 2 when E=100 and at 10 when $E > 100$, and the number of modes in the arrival distribution was set at two for all scenarios presented.

		E=100			E=1 000			E=10 000		
p	$\widehat{\theta}$	Bias	MAD	Cvg	Bias	MAD	Cvg	Bias	MAD	Cvg
3	$\widehat{E}_T$	-0.1	5.2	97.1	-1.0	37.1	98.0	2.3	116.9	99.5
	$\widehat{E}_{sys}$			93.6			94.3			95.8
	$\widehat{E}_G$	1.6	5.6	97.4	5.9	38.7	98.1	27.3	124.4	99.5
	$\widehat{E}_{rs}$	-0.3	9.6	96.7	-0.8	94.0	97.3	7.3	946.7	97.7
	$\widehat{E}_{srs}$	-0.1	5.5	95.6	-1.3	39.7	95.8	-10.5	184.3	98.0
7	$\widehat{E}_T$	-0.6	7.8	96.7	-1.8	56.4	97.0	-4.7	182.5	99.3
	$\widehat{E}_{sys}$			94.5			96.1			99.2
	$\widehat{E}_G$	0.8	8.0	97.0	4.5	58.1	97.1	20.4	187.4	99.2
	$\widehat{E}_{rs}$	0.8	16.3	95.1	13.9	165.2	95.4	-68.1	1640.1	95.7
	$\widehat{E}_{srs}$	-0.6	9.7	98.0	-4.6	77.3	98.4	-27.8	548.3	98.4
11	$\widehat{E}_T$	-1.0	10.0	95.6	-3.4	71.4	96.2	-3.2	220.1	98.9
	$\widehat{E}_{sys}$			96.3			98.7			100.0
	$\widehat{E}_G$	0.1	10.0	95.9	0.1	71.4	96.3	12.2	223.4	98.9
	$\widehat{E}_{rs}$	0.9	21.8	93.9	-4.3	215.2	93.6	-325.5	2107.2	93.8
	$\widehat{E}_{srs}$	-1.0	14.2	98.2	-4.5	120.0	98.1	-21.9	937.3	97.9

Notes: Parameter estimates are calculated over 10 000 simulations for each scenario, and coverage is the percentage of simulations where the true escapement was contained within the confidence interval of the estimator. Coverage for $\widehat{E}_{sys}$ was determined using confidence intervals constructed using $\widehat{\mathrm{Var}}_3(\widehat{E})$.

Table 6.4: Performance of two estimators ($\widehat{\theta}$), $\widehat{E}_T$ and $\widehat{E}_{sys}$, in estimating the true escapement of simulated, systematic, periodic count datasets with variable sampling intervals. Performance measures are mean bias, estimator standard deviation and mean root mean squared error. The sampling interval was randomly varied around a mean of seven days, while the number of modes of the arrival distribution was set at one or three, and overdispersion of the counts varied between one and 100.

			E=100		E=1 000		E=10 000	
H	q	$\widehat{\theta}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	low	$\widehat{E}_T$	0.0	8.5	0.2	27.0	-0.6	89.7
		$\widehat{E}_{sys}$	0.2	9.8	1.5	62.6	6.2	596.1
1	med	$\widehat{E}_T$	0.2	12.0	0.1	84.5	2.7	267.1
		$\widehat{E}_{sys}$	0.3	12.9	1.1	99.0	12.8	636.6
1	high	$\widehat{E}_T$	0.2	18.8	-1.9	266.0	1.1	852.0
		$\widehat{E}_{sys}$	0.4	19.3	-0.4	268.2	16.3	1008.3
3	low	$\widehat{E}_T$	0.0	8.6	-0.5	28.9	-2.7	139.5
		$\widehat{E}_{sys}$	0.1	9.6	0.3	56.8	-2.9	541.2
3	med	$\widehat{E}_T$	0.0	12.0	-0.3	85.8	-2.4	286.0
		$\widehat{E}_{sys}$	0.1	12.8	0.8	96.7	-2.3	590.0
3	high	$\widehat{E}_T$	-0.2	19.0	-5.9	268.3	-1.2	850.9
		$\widehat{E}_{sys}$	-0.1	19.4	-4.8	270.2	4.9	977.2

Notes: Parameter estimates are calculated over 10 000 simulations for each scenario. Low, medium and high overdispersion was set at 1, 2 and 5 for E=100, and 1, 10 and 100 for $E > 100$.

When variability in the sampling interval was simulated, $\widehat{E}_T$ showed very similar behavior to equivalent scenarios with a constant sampling interval. However, RMSE for $\widehat{E}_{sys}$ was inflated compared to scenarios with constant sampling intervals, resulting in this estimator having a higher RMSE than $\widehat{E}_T$ for most scenarios. This was particularly apparent in scenarios with higher escapements ($E > 100$) and low and medium overdispersion.

6.3.4 Evaluation of stability of estimators

Invalid escapement estimates can result from either non convergence of the ML optimizer (for E_G and $\mathrm{Var}(\widehat{E}_T)$), or in simulations where only zero counts occur in the periodic count sample, by chance. The latter occurrence results in non-convergence of E_G (and $\widehat{\mathrm{Var}}(\widehat{E}_T)$) and produces estimates of escapement of zero for the design-based estimators (E_T, E_{sys}, E_{rs} and E_{srs}) but this was very rarely observed (a maximum of 11% of simulated datasets contained only zeros for one scenario, Table 6.5, for all other scenarios this value was less than 2% and was usually zero) and hence values are not given in Table 6.5. Non-convergence was only problematic at lower escapement levels, with estimates being valid for almost all simulations, for all scenarios with $E \geq 50$ (Table 6.5). For scenarios with escapement of less than 50, the occurrence of invalid estimates was heavily scenario-specific. Estimates were invariably valid when overdispersion was low, but the proportion that were invalid quickly increased for any scenario with high overdispersion. This was compounded in situations where the number of modes increased, the frequency of sampling decreased and escapement level decreased. As few as 0.72 of estimates were valid in the worst-case scenario (E=10, q=5, H=3, p=11). The mean number of samples in the simulated periodic count datasets ($\bar{n}$) showed very similar patterns, with $\bar{n}$ lower for scenarios with lower proportions of estimates valid, and vis versa (Table 6.5).

6.3.5 Evaluation of peak count bias

The peak count estimator was negatively biased for all scenarios investigated, usually severely (Fig 6.3). It appeared to be most sensitive to the interaction between escapement level and overdispersion level for a given scenario. The estimator was least biased when overdispersion

Table 6.5: Comparison of proportion, P, of simulations where valid estimates of escapement were achieved for $\widehat{E}_G$, and the mean sample size of periodic counts, $\bar{n}$, achieved, for different scenarios with low escapement levels ($E = 10$–$1\,000$).

			E=10		E=25		E=50		E=100		E=1 000	
p	H	q	P	$\bar{n}$	P	$\bar{n}$	P	$\bar{n}$	P	$\bar{n}$	P	$\bar{n}$
3	1	low	1.00	22.1	1.00	25.4	1.00	27.7	1.00	29.8	1.00	35.7
		high	0.89	18.3	1.00	22.2	1.00	24.7	1.00	26.9	1.00	27.4
	3	low	0.97	25.3	1.00	28.2	1.00	30.2	1.00	31.9	1.00	37.1
		high	0.77	21.6	0.95	25.4	0.99	27.5	1.00	29.5	1.00	29.9
7	1	low	1.00	9.6	1.00	11.2	1.00	12.3	1.00	13.3	1.00	16.0
		high	0.92	7.6	1.00	9.6	1.00	10.9	1.00	12.0	1.00	12.2
	3	low	0.98	11.0	1.00	12.5	1.00	13.5	1.00	14.3	1.00	16.5
		high	0.83	8.8	0.96	11.0	0.99	12.2	1.00	13.2	1.00	13.4
11	1	low	0.99	6.5	1.00	7.7	1.00	8.4	1.00	9.1	1.00	10.9
		high	0.74^{Ψ}	5.1	0.99	6.6	1.00	7.4	1.00	8.2	1.00	8.3
	3	low	0.98	7.4	1.00	8.6	1.00	9.2	1.00	9.7	1.00	11.2
		high	0.72	5.7	0.98	7.4	0.99	8.3	1.00	9.0	1.00	9.1

Notes: Parameter estimates are over $10\,000$ simulations for each scenario. The proportion of simulations that are not viable for $\widehat{E}_G$ are those where only zero counts were included in the simulated dataset or the ML optimizer failed to converge for that simulated dataset. Low, medium and high overdispersion was set at 1, 2 and 5 for $E \leq 100$, and 1, 10 and 100 for E=1 000.

Ψ For this scenario 11% of simulated periodic count datasets contained only zero-counts. For all other scenarios this value was below 2%, and was usually zero.

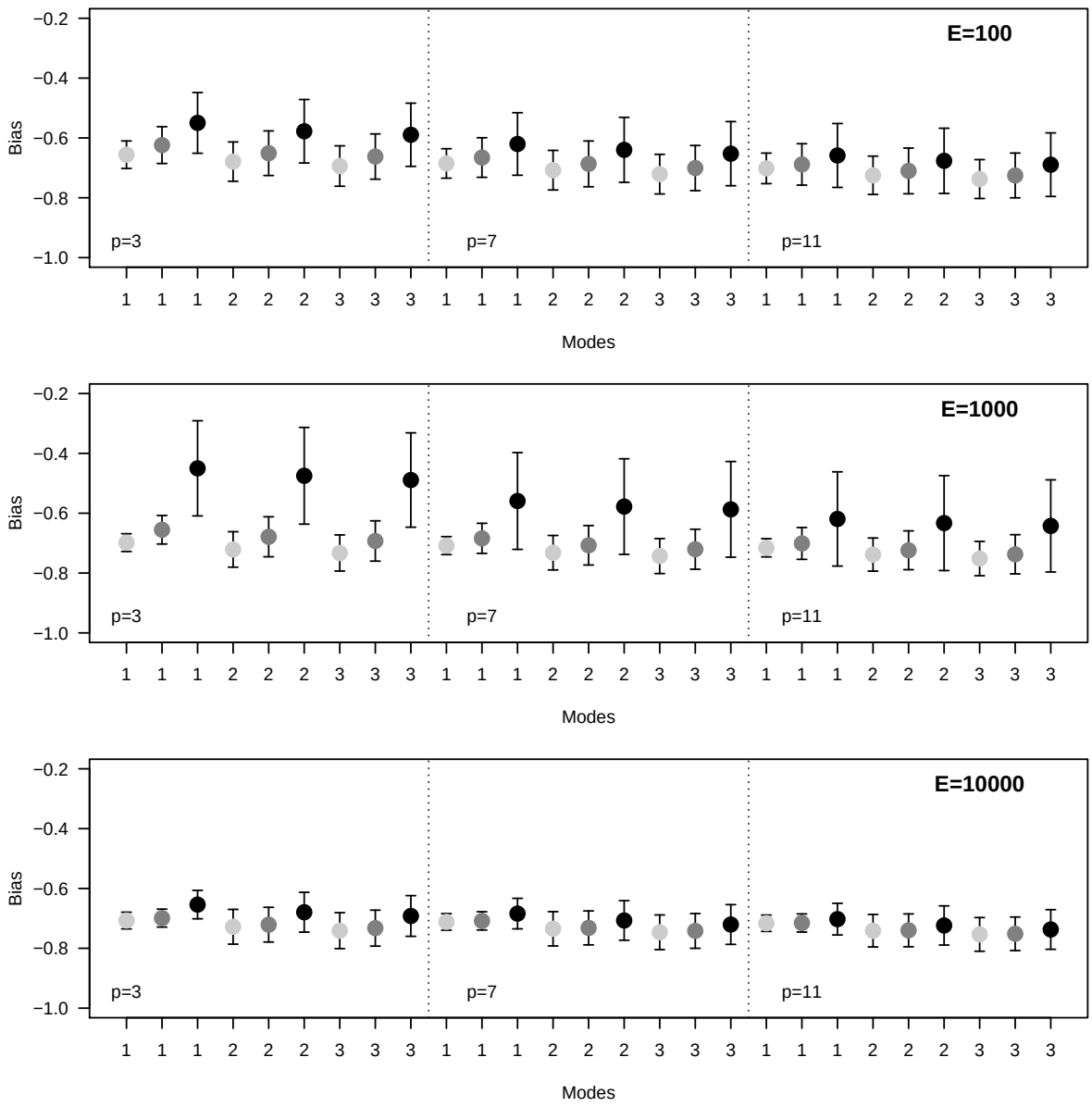

Figure 6.3: Mean relative bias of the peak count estimator of escapement for simulated periodic count data. Relative bias on the y-axis is calculated $(\widehat{E}_P - E)/E$ where E and $\widehat{E}_p$ are the true and estimated escapements over the $10\,000$ simulations for that scenario. The x-axis indicates the number of modes for the scenario. The three graphs display the results for scenarios with escapements E ranging from 100 (top graph) to $10\,000$ (bottom graph). The three panels indicate the sampling interval p specified for the scenarios, ranging from three to 11 days from left to right. Light gray, dark gray and black dots indicate scenarios with low, medium and high overdispersion, q, which were 1, 2, and 5 for E=100, and 1, 10 and 100 for $E \geq 1\,000$. Error bars represent $\pm$ one standard deviation.

was high relative to the escapement level, which is most evident when E=1 000 and q=100 (Fig 6.3). When the number of modes of arrival for a scenario increased from one to three E_p displayed only very slightly higher bias, and similarly, estimates of escapement became slightly more negatively biased when the sampling interval increased from three to 11 days. Relative bias of the peak count estimator stabilized at around -70% for all scenarios when E was 10 000. For some scenarios, most notably those with high overdispersion, there was moderate variation in the estimates over the simulations within scenarios.

6.4 Discussion

6.4.1 Comparative performance of estimators for systematic samples

The results of this simulation-based study show that it is difficult to make broad generalizations about the comparative performance of estimators for systematic periodic counts. The three estimators ($\widehat{E}_T$, $\widehat{E}_G$, $\widehat{E}_{sys}$) displayed relatively similar performance across all scenarios. Certainly, the recently cited advantages of AS estimators over the widely implemented trapezoidal AUC estimator (Holt and Cox, 2008; Parsons and Skalski, 2009) were not realized in the simulations. In fact, the unweighted AS estimator in (6.12) in Section 6.2.3.1 was found to be negatively biased when zero-tail counts were present, while in situations where no zero counts are encountered (e.g. the example given in Parsons and Skalski, 2009) it will be positively biased and it is also difficult to see how season length could be determined in these cases.

The corrected AS estimator and its associated estimators of variance were found to be viable alternatives to $\widehat{E}_T$, however there appears to be no clear advantage in suggesting practitioners switch to them from their current estimators, providing they start estimating Var($\widehat{E}_T$). Parsons and Skalski (2009) cited bias, and lack of an associated variance estimator, as weaknesses of $\widehat{E}_T$, which have both been allayed to some degree in this chapter. $\widehat{E}_T$ was shown to be effectively unbiased in practice for a wide range of scenarios, and is actually equivalent to the corrected (weighted) version of the AS estimator for systematic samples ($\widehat{E}_{sys}$) when the sampling interval was constant. The only differences in performance statistics between these two estimators, at

least for the scenarios with constant sampling intervals, occurred due to the (generally rare) incidence of $\widehat{\text{Var}}(\widehat{E}_T)$ not being calculable for some simulated datasets.

In the case of variable sampling intervals within a spawning season, $\widehat{E}_T$ performed adequately, while $\widehat{E}_{sys}$ displayed inflated RMSE compared to equivalent scenarios with a constant sampling interval. This result is logical, as $\widehat{E}_T$ explicitly accounts for variable weights of counts in its calculation, while the standard estimator for $\widehat{E}_{sys}$ based on fish-days estimator (6.13) assumes equal weightings for the non-tail-counts. Clearly, if non-constant sampling intervals are encountered in practice it must be accepted that $\widehat{E}_{sys}$ will be worse-performing than $\widehat{E}_T$. If appropriate weightings are added to (6.13) to remedy this, then it will again be equivalent to $\widehat{E}_T$.

The simulations also highlighted that the performance of $\widehat{E}_G$ was also very similar to $\widehat{E}_T$, and it appears that neither estimator can generally be considered superior. In the context of the discussion in this chapter most focus between $\widehat{E}_G$ and $\widehat{E}_T$ estimators will be placed on the latter, given the similar behavior between them, and the fact that $\widehat{E}_T$ has been, and will presumably continue to be, the most widely used estimator for periodic count data.

6.4.2 Comparative performance of variance estimators for systematic samples

A previous weakness of $\widehat{E}_T$ was that it did not produce a variance estimate, although this has recently been addressed by the estimator of Millar and Jordan (In Press). However, comparing its performance with variance estimators for AS methods is challenging owing to the difficulties in calculating the latter for systematic samples. Wolter (1984, 2007) and Fewster (2011) discuss general approaches to confronting the analysis of systematic samples. They suggest that most options can be separated into three approaches based on whether estimators; ignore the systematic design and utilize random sample-based estimators, assume various post-sampling stratification schemes, or attempt to describe the underlying structure of the data. The estimators compared for systematic periodic count data follow this demarcation well. The most simple variance estimator, $\widehat{\text{Var}}_1(\widehat{E}_{sys})$, falls into the first category, $\widehat{\text{Var}}_2(\widehat{E}_{sys})$ and $\widehat{\text{Var}}_3(\widehat{E}_{sys})$ the second, and the AUC-based techniques are an attempt at modeling the underlying structure encountered over

the spawning season.

The different approaches of the variance estimators of the different methods led to greater differences in performance than in the escapement estimators themselves. If structure in the sample is known to be present *a priori*, and this is generally the case for the majority of salmon periodic counts, then ignoring it will reduce the inferential potential of the data. The poor performance of $\widehat{\text{Var}}_1(\widehat{E}_{sys})$ is an example of when failing to admit to the underlying structure leads to overly conservative variance estimates. The other two approaches address the structure, though by opposing means. The estimators $\widehat{\text{Var}}_2(\widehat{E}_{sys})$ and $\widehat{\text{Var}}_3(\widehat{E}_{sys})$ rely on post stratification to approximate the structure post-hoc, by using higher-order differences to account for correlation between adjacent counts. In contrast, $\widehat{\text{Var}}(\widehat{E}_G)$ and $\widehat{\text{Var}}(\widehat{E}_T)$ rely on positing a model for the structure (in this case a model based on unimodal, normally distributed arrivals and deaths), and estimate uncertainty in the model formally using likelihood-based techniques. Providing the model is reasonably accurate and general, this latter approach seems logical, although it is clearly sensitive to violation of the model. This was evident when the performance of $\widehat{\text{Var}}(\widehat{E}_G)$ and $\widehat{\text{Var}}(\widehat{E}_T)$ changed substantially with variation in the number of modes of the simulation model (and hence dynamics differed from that assumed by the model), while the other design-based variance estimators remained relatively stable.

Although there were clear differences between the best AS estimator $\widehat{\text{Var}}_3(\widehat{E}_{sys})$ and the AUC estimators, they showed similar general patterns in performance over the scenarios. Whether $\widehat{\text{Var}}_3(\widehat{E}_{sys})$ was more, or less conservative than $\widehat{\text{Var}}(\widehat{E}_T)$ and $\widehat{\text{Var}}(\widehat{E}_G)$, was scenario-dependent which resulted in the coverage of no estimator being consistently closer to the nominal rate. Consequently, it appears that there are no general advantages of implementing $\widehat{\text{Var}}_3(\widehat{E}_{sys})$ over $\widehat{\text{Var}}(\widehat{E}_T)$ if variation in the dynamics of spawning runs in practice reflect the variation explored in the simulations. It should also be noted that it is difficult to justify the variance estimators for E_{sys} as a statistically superior alternative to $\widehat{\text{Var}}(\widehat{E}_T)$ given that they are themselves an approximation to an appropriate variance for systematic samples.

6.4.3 Performance of AS estimators for random samples

Simulation also provided the opportunity to explore the potential of alternative sampling designs to which AS methods are more suited, including simple random, and stratified random designs (Parsons and Skalski, 2009). However, if the three designs are compared solely on the basis of performance in the simulations, then $\widehat{E}_{sys}$ would be selected over $\widehat{E}_{srs}$, and especially over $\widehat{E}_{rs}$, due to the inflated MAD of the latter two. As bias of all three estimators was relatively similar, most of the differences in MAD originate from the higher precision of $\widehat{E}_{sys}$ across simulations.

The decrease in precision of estimators from systematic to stratified random, and further to simple random sampling, is well recognized (Cochran, 1977). This is due to the potential of some samples in random designs being highly clustered and/or covering a restricted area of the sample space, by chance. Similarly, for a given sampling intensity, stratified random designs will always have samples with some adjacent counts being closer together than for systematic designs, as the latter has the maximum possible spread in timing of counts for that sampling intensity. However, as noted in Section 6.4.1 it is far more difficult to formulate a justifiable variance estimate for the systematic sample estimator and this is a significant weakness when comparing it to randomized designs.

Perhaps a more promising design-based sampling method for the salmon spawning season would be a properly implemented systematic design, for example the multiple-start systematic sample (Wu, 1986; Wolter, 2007). This would combine both the high precision of $\widehat{E}_{sys}$ and the availability of a valid, theoretically justified, variance estimator. Again though, the probability of this approach being utilized seems at least as, though probably even more unlikely, than for random and stratified random designs, especially within the constraints of many monitoring programs, which limit the sample sizes of surveys.

Interestingly, the performance of both random sample variance estimators reflected the performance of the escapement estimators, in that they did not display marked improvement over the systematic variance estimators (either AS or AUC). The exception to this generalization

is perhaps the performance of $\widehat{\text{Var}}(\widehat{E}_{rs})$ which often, though not always, produced confidence intervals with coverage closer to the nominal rate than other estimators.

While the assumptions of the random-sample-based estimators are more easily met, the practical success of the systematic estimators in the simulations and the preference shown by practitioners for these estimators, suggest that these will continue to be the choice for salmon biologists. Certainly, sampling designs for salmon monitoring in the past appear to be driven by practical rather than statistical considerations. The absence of variance in estimates of escapement in the vast majority of salmon research and monitoring are testament to this. In the absence of clear practical advantages of random sample estimators over systematic sample alternatives, their use will likely remain minimal. This is especially the case given the disadvantages of random sampling perceived by practitioners. For example, systematic sampling makes planning work schedules for upcoming seasons easier, and it will seem inefficient to many practitioners if, by chance, random samples in a given year fall very close together, or on subsequent days, in a random design.

6.4.4 Performance of the peak count estimator

The last estimator to be considered is the peak count estimator. This is not a completely satisfactory estimator, owing to its inability to produce an absolute escapement estimate (unless they are calibrated to true abundance somehow), and an absence of a viable variance estimator. However, it is one of the more numerically dominant estimators in practice (Knudsen, 2000; Parsons and Skalski, 2009), and the simulations investigated provided a platform to test its efficiency. Such relative indices of abundance such as peak counts are less desirable for informing management (Anderson, 2001), especially for applications that require absolute abundance such as calculating the proportion of the population removed by fisheries. Often relative indices are deemed adequate for trend detection, where assumptions are more easily met, and this was the focus of Holt and Cox (2008)'s study. This assumes proportionality between peak counts and true abundance, and while this assumption should clearly be assessed with empirical and simulated

data, this has seldom occurred. While Holt and Cox (2008) provide some valuable advice on the use of peak counts for trend detection, little comparison of peak counts and true abundance and this relationship over different scenarios, was presented in their study.

The simulations showed that the assumption of proportionality was reasonable at high escapement where little difference in relative bias occurred between scenarios, and there was little variation between simulations within scenarios. However at lower escapements relative bias was affected by different levels of overdispersion and for some scenarios there was significant variation in bias between simulations within some scenarios. If this variation occurred between sites or years in practice, then peak counts will be unreliable as a relative abundance index. At high escapements, the peak count estimator stabilized at about 30% of the true escapement level across all scenarios. This is largely related to the dynamics of the spawning runs assumed in the simulations. For example, a much steeper arrival distribution and a longer stream-life would result in peak counts being closer to the true level. However, the simulations show that for dynamics similar to those commonly observed in practice, peak counts may easily account for a relatively small proportion of the fish contributing to the run, with severe consequences for applications such as estimating fishing mortality.

In reality, peak counts will be an even more unstable estimator of abundance than indicated here as detection probability will often be less than one, and this was not accounted for in the simulations. For this reason the changes in bias between scenarios must be considered relative rather than absolute. Furthermore, variation in detectability will contribute to variation in the proportionality between sites or years and will further limit the value of the peak count method, even for the analysis of trends. This has been noted in several studies (e.g. Shardlow et al., 1987; Bue et al., 1998), and corrections for detection are less frequently applied for this estimator than, for example, $\widehat{E}_T$ (Parsons and Skalski, 2009).

6.4.5 Limitations of trapezoidal AUC and its variance

The estimator $\widehat{E}_T$ (in conjunction with $\widehat{\text{Var}}(\widehat{E}_T)$) appears to be a reasonable practical option for estimating escapement. There are several limitations to its implementation however. One disadvantage of $\widehat{\text{Var}}(\widehat{E}_T)$ compared to AS variance estimators is the possibility that it will be incapable of estimating variance for some datasets. It is derived from the GLM model (6.4) fitted using maximum likelihood that forms the basis of $\widehat{E}_G$, and thus exhibits the same susceptibility to non-convergence exhibited in Section 6.3.4. However, the estimator of escapement itself, $\widehat{E}_T$, is the standard trapezoidal AUC method it is very stable, in contrast to $\widehat{E}_G$. In fact, it shows the same behavior as $\widehat{E}_{sys}$, in that it is theoretically capable of estimating escapement for any dataset. Clearly though, when $\widehat{E}_T$ is used in conjunction with its ML-based variance estimator, $\widehat{\text{Var}}(\widehat{E}_T)$, it will require care in ensuring data is adequate for estimation.

From the simulations explored, non-convergence of $\widehat{\text{Var}}(\widehat{E}_T)$ would appear limited to situations involving very low escapements in conjunction with high overdispersion, which will not generally be the case except for some of the more endangered salmon stocks. In those cases the most pragmatic approach is perhaps a two-step analysis where $\widehat{E}_T$ and $\widehat{\text{Var}}(\widehat{E}_T)$ are calculated first, and in the event of instability an alternative AS variance estimator could be utilized, perhaps $\widehat{\text{Var}}_3(E_{sys})$. Such a composite estimator was examined in the simulations but was not presented as its performance was obviously part-way between the two individual estimators, with the exact position depending on the frequency of non-valid estimates encountered for $\widehat{\text{Var}}(\widehat{E}_T)$, in each scenario.

An intermediate step between switching from $\widehat{\text{Var}}(\widehat{E}_T)$ to $\widehat{\text{Var}}(E_{sys})$ for challenging datasets, would be to still utilize $\widehat{E}_T$, and calculate variance using a more simple GLM formulation. For example, an intercept-only model with the two extra coefficients excluded. This results in a much more stable estimator given the extra degrees of freedom, and is generally capable of fitting the same datasets as $\widehat{E}_{sys}$. However, due to the fact that the link function of a GLM ensures that the expected value is determined on the untransformed scale (giving the sample mean, $\bar{c}$), this

estimator (and consequently its variance) is equivalent to the naive, equal weighting version of $\widehat{E}_{sys}$ given by (6.12), and its subsequent use would be a poor choice for systematic samples.

Another potential weakness of $\widehat{E}_T$ is that its variance relies on the parametric model at the core of $\widehat{E}_G$ being a reasonable representation of the true dynamics of salmon arrivals and deaths. This is highlighted by the larger differences in performance statistics for AUC compared to AS variance estimators when the number of modes in the arrival distribution changed. This is an example of the true dynamics differing from the assumed model for $\widehat{\text{Var}}(\widehat{E}_T)$, while the AS estimator makes fewer assumptions about the dynamics and so is less affected when modality changes. In most situations where arrival distributions are approximately unimodal then the model presented should be adequate, though the potential of more flexible models (splines etc.) is a topic for further research.

6.4.6 Presence of non-zero tail-counts

A problem often encountered in periodic count sampling is a failure for monitoring to extend over the entire spawning season (Bue et al., 1998; Hilborn et al., 1999). In these situations non-zero counts are observed for the first and/or last count/s (denoted non-zero tail-counts), and this is evidence that the monitoring program is deficient. No methods can reliably compensate for this inadequacy, but various assumptions can be utilized. For example, a correction factor has been used in the past for TAUC estimators (Hilborn et al., 1999) and details of this method are presented in Section 6.2.2.2. This approach can also be generalized to other estimators.

The purpose of this chapter was to compare the estimators for the data for which they were developed not how they perform when poor monitoring programs are implemented, and thus comparisons were restricted to datasets with zero tail-counts. The application of estimators when non-zero tail-counts are encountered has been discussed elsewhere (Hilborn et al., 1999; Szerlong and Rundio, 2008; Millar et al., 2012), and in these cases E_G would appear to be a suitable choice, however it must be stressed that monitoring that results in non-zero tail-counts must be avoided if at all possible as all estimates will become worse in these situations.

6.4.7 Practical implementation of escapement estimators

The set of estimators examined herein did not include carcass and redd counts, mark-recapture techniques and the original HARR alternative to E_G (Holt and Cox, 2008; Parsons and Skalski, 2009; Millar et al., 2012). Carcass and redd counts were not favored by Parsons and Skalski (2009) due to inherent bias and/or violation of their assumptions, and for most species (with the exception of chinook salmon) they are rarely implemented in practice (Knudsen, 2000; Parsons and Skalski, 2009). While mark-recapture holds promise it is not ideally suited to large-scale monitoring programs and is again less common than periodic count methods. Lastly, the HARR estimator has effectively been made redundant by the very similar (and easier to implement) GAUC estimator of Millar et al. (2012). For these reasons the AUC and AS estimators that were investigated appear to be more suitable for most sets of periodic count data currently being collected, and thus some general issues of their practical implementation need to be addressed.

If TAUC methods and systematic sampling continue to be favored by practitioners, it is difficult to address exactly how these estimates should be used for inference. Quite possibly this will be case-specific. For example, the variance estimate $\widehat{\mathrm{Var}}(\widehat{E}_T)$ only addresses part of the uncertainty in escapement. Total uncertainty can be separated into the components affecting single sites, which are uncertainty in; detection of fish, choice of days to be surveyed, observer efficiency, and stream-life. While other sources are introduced when escapement for a waterway involves multiple sites (either in time or space), this can usually be handled in a standard manner, and so uncertainty at single sites is the focus herein. Robust inferences clearly require variance estimates for escapement that reflect this total variability, otherwise proceeding with simplified estimators will produce overly optimistic confidence intervals. Several previous studies have attempted to account for multiple sources of uncertainty in escapement as possible, by various means.

Parsons and Skalski (2009) assume independence among different sources of variance and suggest the use of the standard total variance formula, with variances from each component as-

sumed to be available, perhaps from different datasets. Hilborn et al. (1999) and Su et al. (2001) used prior information on steam-life and/or observation efficiency parameters and implement models using penalized likelihood and Bayesian methods, respectively. These approaches allow uncertainty in these parameters to be propagated through to the final estimates of escapement. Other studies have included estimates of detection during individual counts (Szerlong and Rundio, 2008) or uncertainty estimated from repeated sampling (Parken et al., 2003). However, each of these methods of combining sources of uncertainty requires auxiliary information in the form of data, or prior distributions. Unfortunately the probability of collecting additional data at all study sites to allow simultaneous estimation of different sources of variance seems unlikely given that only periodic count data are collected for the majority of monitoring programs. Furthermore, the fact that very few salmon biologists estimate any sort of variance in escapement suggests that it will be even more difficult to encourage attempts to account for multiple sources.

This study did not attempt to overcome the issue of accounting for all uncertainty sources. The specific target was to compare the various options available for estimating variance in fish-days relating to the observation of fish on sampling days (not including the uncertainty in estimating observer efficiency v). The optimal approach of extending this variance estimate to an appropriate total variance estimate is likely to be case-specific. In many cases, the data, or prior knowledge required, will simply not be available. Sometimes *ad hoc* approaches such as using previous estimates of residence time and observer efficiency and their respective variances, from the same, or other study sites for that species, can be applied. This is not without its risks however, as the applicability to the new dataset will remain unknown. This is especially the case given variation in these parameters observed within and between studies (e.g. Bue et al., 1998; Holt and Cox, 2008). The emphasis will be on the analyst to assess the best method of combining information, and in some cases acknowledging when inadequacies in the data are insurmountable. In all cases the question will need to be asked, do we have enough information available to meet the objectives of the research or monitoring?

6.4.8 Escapement estimates for state-space modeling

This comparison of estimators of escapement and its variance has consequences not only for monitoring programs, but also for fitting state-space population models to time-series of salmon escapement estimates. The advantage of these models over more traditional population, or stock-assessment models is that they have the ability to partition uncertainty in the time-series into components relating to the sampling of abundance (observation error) and uncertainty related to the inherent variability in the dynamics of the animal population (process error). This ability results in a suite of advantages for inference and prediction (Clark, 2003). Unfortunately, for many datasets the information content in the data is too low to reliably partition uncertainty into these components. It is simply the case that not all parameters in these relatively complicated models are identifiable. Often the only way to fit adequate models is to constrain the variance parameters, by specifying their ratio (Schnute, 1994), specifying informative priors for one or both error variances when fitting models under a Bayesian framework (Meyer and Millar, 1999), or by setting one of them as a known constant/s (King et al., 2008, 2009). Often this last option is exercised as the point estimates of abundance (the observations) used for modeling sometimes have associated variances already estimated, and these can be set as the observation error variances in each year (King et al., 2008, 2009).

In fact, the motivation for exploring the estimators and their variance here originated from a lack of an estimate that could be used as prior information for the observation error variance parameter in state-space salmon models. Preliminary modeling of OCN coho data suggested that identifiability was difficult to achieve when models were fitted to time-series that contained only point estimates of escapement. The development of viable estimators of variance in $\widehat{E}_T$, and their performance on simulated data, means that there is now the option of investigating the potential of applying state-space models to the assessment of OCN salmon stocks, by using $\widehat{\mathrm{Var}}(E_T)$ as a base from which observation error variance in the state-space models can be constrained, as was explored in Chapter 2.

Chapter 7

General Discussion

7.1 Goodness of fit of state-space salmon models

Much of the foundational work on state-space models of animal population dynamics focused on implementation of model fitting (McAllister et al., 1994; Millar and Meyer, 2000b) and determining whether all parameters could be identified for certain classes of models and datasets (Schnute, 1994; Magnusson and Hilborn, 2007). Now however, these models are widely applied for empirical problems in ecology and their proliferation in use has raised questions about their robustness in certain situations (Snover, 2008). It appears that in the rush to apply what are to many analysts, attractive models, some of the foundations of robust model-fitting have suffered. This is particularly true of assessment of model adequacy (Chapter 5), which is seldom addressed explicitly in published work, and to a lesser degree the selection among models (Chapter 4). With the advent of on-line supplementary materials for journal articles, space restrictions are no longer an excuse to prevent analysts including assessments of model adequacy in papers, particularly in graphical form. This would allow fellow researchers to judge the robustness of model-fitting and would lead to explicit tests of adequacy to become more prevalent. In time, widespread use would lead to not only more applied cases, but also methodological interest Together, these may allow generalizations to be made across studies (e.g. certain test variables) that will make the choices of analysts easier in the future. Publication of unreliable models (e.g. Chaloupka and Balazs, 2007; Snover, 2008) will likely be reduced significantly when model adequacy becomes more widely understood and tested.

7.2 Selecting among state-space salmon models

Selecting among even simple statistical models has long been contentious and analysts often fall into different camps based on various philosophical and/or practical considerations (Kass and Raftery, 1995; Burnham and Anderson, 2002). The issues are compounded when selecting among state-space models due to the difficulties in defining model complexity and what constitutes model fit, as was seen in Chapter 4. Unfortunately, there are clear weaknesses in all commonly considered model selection tools.

It is expected that the problems with DIC that are highlighted in Chapter 4 will lead to less reliance, or at least more circumspect use of this method in the future. Similar problems will occur with *ad hoc* application of other easily calculated information criteria such as AIC and BIC, in addition to difficulties in specifying a model complexity penalty for those methods. Consequently, the model expansion method advocated by Gelman et al. (2003); Gelman (2004) and Kery and Schaub (2011) is appealing in that it avoids formal model selection *per se*, is relatively easy to carry out and is a natural way to build models. For instance, it is often not until a model is constructed and fitted that its inadequacies become clear, or alternative forms are considered. That is, modeling animal population dynamics is often an iterative process which fits this modeling approach well. However, it is not clear how considerations such as model uncertainty (typically handled by way of multimodal inference) could be carried out in this framework or whether this would even be desired.

Perhaps the most logical alternatives to those that attempt to trade off fit and complexity are Bayes factors and RJMCMC. The computational challenges of these techniques are more formidable however. The main reason that use of state-space models in fisheries and terrestrial ecology has taken off is that they have become relatively easy to fit for fisheries scientists and biologists using higher-level languages. Ease of computation also appears to be the reason for the widespread use of DIC in ecology. Thus, until specific routines become easier to implement in software packages it seems unlikely that either RJMCMC or Bayes factors will be widely

implemented in practice as analysts have to balance the robustness of their analyses with the time constraints in carrying them out. Continual software development in these areas, perhaps even software specific to dynamical modeling of animal population dynamics (the early stages of development of such a tool has been indicated by Newman et al., 2008), would be very beneficial for progress in the field, especially if it could be combined in some way with semi-automated generation of various predictive distributions and related diagnostics for assessing adequacy, as suggested by Gelman (2004). There appears to be ample potential for a comparison of model selection techniques for state-space models, especially those with structures similar to models commonly fitted in studies of animal population dynamics, and such studies would provide extremely valuable guidance for practitioners in this field.

7.3 The value of state-space life-history models

The inadequacies of traditional salmon stock-recruitment models fitted using linear regression have long been known (Walters and Ludwig, 1981; Walters, 1985), but emphasis on moving towards alternatives has become more prevalent only recently (Peterman et al., 2003; Lessard et al., 2008; Hilborn, 2009; Su and Peterman, 2011). Traditional models will likely continue to be the method of choice by a lot of fisheries biologists in spite of their limitations, unless more case studies such as Chapter 2 continue to be presented. Unfortunately most datasets in practice are not of as high quality as that from Lobster Creek, and most are restricted to monitoring of escapement. Where possible it appears that fitting state-space versions of stock-recruitment models will be desirable given the potential reduction in bias of parameter estimates (King et al., 2009; Su and Peterman, 2011) and their more parsimonious treatment of uncertainty when predicting future population size, amongst other factors. However, if state-space models cannot be fitted due to problems with identifiability the best pathway to robust inference is difficult to prescribe, though perhaps the advice given by Bolker (2008) is pertinent. He suggests considering several different methods of fitting observation-error-only or process-error-only models and comparing the results, while techniques such as simulating data similar to the observed dataset

and assessing the consequences of violation of assumptions of more simple models are alternative options.

Ensuring identifiability is particularly challenging in cases where flexible distributions with scale parameters are used in both the process and observation models. This is partly the reason for the issues that were encountered in Chapters 4 and 5, as the flexibility of the models allow inadequate process models to be compensated for and it is therefore difficult to select among models and detect inadequacies based on the fits of the latent states alone. In situations where distributions with fixed dispersion (such as Poisson or Binomial) would be used these problems are expected to be less severe, and this was observed to a degree in the Poisson process model example in Sections 5.2.7 and 5.3.3. In these cases the widely used data level DIC and posterior predictive checks may well detect the best model amongst the set, and detect model inadequacy, but they will remain inferior to the partially marginalized DIC and predictive checks when the process equation is the part of the model of interest. Furthermore, animal population dynamics can be highly variable, are often poorly understood and estimating population abundance is often beset with difficulties, which makes the suitability of distributions with fixed dispersion highly questionable in practice. However, if data alone cannot support more challenging flexible models, restrictions (priors etc.) may have to be placed on problematic parameters.

The methods developed in Chapter 6 are expected to help analysts to estimate sampling variance at the spawner stage when periodic count data are available, and this is expected to facilitate the fitting of some state-space models that would otherwise be unstable. One aspect of the method of using the estimated variances as the observation error variances in the model in Chapter 2 that warrants further study, is whether there are alternative methods of constraining these parameters. In effect, the method in Chapter 2 (fixed observation error variances) places an extremely informative prior at exactly the values estimated on the empirical data. This prevents the posterior distribution of the variance parameters to change even if they conflict with the data. Two methods that may have utility in relaxing this constraint are 1) allowing each variance estimate in year t to come from a common higher-level distribution with appropriate

hyperparameters and/or 2) placing priors (rather than fixed values) on the variance parameters in each year t, σ_t^2, in the form of scaled chi-squared distributions with hyperparameters given by the empirically estimated variance, $\widehat{\sigma_t^2}$, and the degrees of freedom of the periodic count dataset in that year, i.e.

$$\sigma_t^2 \sim \widehat{\sigma_t^2}\frac{\chi_{df}^2}{df} \ .$$

Both of these techniques would aim to provide relatively strong prior information to ensure stable posteriors, but at the same time allow the posterior distribution to be modified if they conflict substantially with the data.

In addition to introducing state-space structures to traditional models, the extra life-cycle complexity added in Chapter 2 allowed a number of advantages over traditional approaches. It is increasingly important that agencies assess whether the objectives of their monitoring can be achieved by only sampling at the spawner stage. This is obviously a poor approach when focus is placed on factors influencing dynamics during freshwater residency, owing to the vast amounts of uncertainty introduced by variation in marine survival that reduce the inferential power of using spawner abundance as a response variable. It is difficult to prescribe where extra monitoring effort should be focused for future studies however, as the aims and optimal monitoring programs will be case-specific. The most obvious extra stage to monitor if resources were limited is the smolt stage as then survival can be partitioned into its two most general components, freshwater and marine survival.

7.4 Expanding life-history models to spatially extensive datasets

The limited spatial extent and lack of replication of the study reaches in Lobster Creek, and the unknown degree of its generality to other coho salmon populations mean that inferences cannot be extended beyond the study area. The consistency of monitoring and the high quality of the the dataset (consistent monitoring effort, absence of missing values etc.) did make it a valuable dataset for trialling new analytical techniques however. The obvious next step in developing models of OCN coho salmon is to construct life-history models based on the GRTS

dataset (Stevens, 2002) which is collected over the full set of populations in the ESU. Previous modeling efforts for these populations have been limited to traditional techniques. Recently, Wainwright et al. (2008) utilized a variety of modeling approaches to assess future viability of coho salmon populations in the OCN ESU. These included fitting Ricker stock-recruitment models to spawner abundance data (assuming no observation error) using the traditional linear regression method. Similarly, Buhle et al. (2009) fit similar models to the OCN stock-recruitment dataset, though they modify model-structure to account for the effects of hatchery-bred fish on population dynamics. Again, observation error was unaccounted for.

Upscaling state-space models to data from the full set of OCN populations is theoretically desirable. Parameter estimates will generally be less biased than observation-error-only or process-error-only models (King et al., 2009; Su and Peterman, 2011), providing the model parameters are identifiable, and the model is suitable for the dynamics of the system being modeled. Not only would they allow improved inferences about the critical factors determining population dynamics, such as oceanic conditions, freshwater habitat etc., but prediction of future abundance will be more robust given the ease with which prediction can be carried out in the Bayesian framework, and the fact that only relevant sources of stochasticity will be projected forward in time (Punt and Hilborn, 1997; Clark, 2003). This latter advantage is particularly relevant for coho salmon owing to the importance placed on prediction of abundance prior to the spawning run, which includes ensuring the sustainability of commercial and recreational (Pacific Fishery Management Council, 2012).

Care will be necessary when expanding life-history models to the GRTS dataset. There are a number of factors that may pose more of a challenge to successful model fitting than were encountered for the Lobster Creek data, such as the prevalence of zero counts, the low counts in many reaches, missing data and the shorter time-series of counts available. The most obvious way to model the data would be at the population level where a life-history model is fitted to a single time-series of abundance for each population for the relevant stages of the life-cycle. The abundance estimates would be derived from either the design-based methods underpinning the

GRTS monitoring scheme (Stevens, 2002) or model-based methods such as hierarchical models for count data. The alternative would be to fit population models at the individual reach-level as this is where the demographic processes are occurring, and modeling at this level has the potential to account for some of the reach-level variation that is typical of many salmon datasets (e.g. Pess et al., 2010; Steel et al., 2012). Whether such models could be fitted reliably in practice remains to be seen.

Chapter 8

Appendices

8.1 A model with stream-specific process error variances

A model, denoted M^*_{BLL}, that was very similar to M_{BLL} is used to demonstrate an alternative structure of the process errors. The deterministic components were identical to M_{BLL} and are given by

$$\log n_{1,j,t} = \log \alpha^P + \log n_{3,j,t-1} - \log\left(1 + \frac{n_{3,j,t-1}}{K_i^P}\right) + \epsilon_{1,j,t}$$

$$\log n_{2,j,t} = \log \Phi^P + \log n_{1,j,t-1} + \epsilon_{2,j,t} \quad , \quad \text{and} \quad \log \Phi^P = \text{logit}^{-1}(\varphi^P)$$

$$\log n_{3,j,t} = \log \Phi^S + \log n_{2,j,t-1} + \epsilon_{3,j,t} \quad , \quad \text{and} \quad \log \Phi^S = \text{logit}^{-1}(\varphi^S + cO_{i,t})$$

where all parameters have the same interpretations as in Section 2.2.2.2 except the process errors which are modeled; $\epsilon_{1,j,t} \sim \text{N}(0, \sigma^2_{1,j})$, $\epsilon_{2,j,t} \sim \text{N}(0, \sigma^2_{2,j})$ and $\epsilon_{3,j,t} \sim \text{N}(0, \sigma^2_{3,j})$, where $\sigma^2_{i,j}$ are process error variances for each i stage and j stream.

8.2 A model with uninformative priors on all variance components

All models presented in Chapter 2 specify the values for the age-class-, stream- and year-specific observation error variances, $\sigma_{i,j,t}$. An alternative approach is to ignore these estimates and specify 'uninformative' priors on observation error variances. For such models the general observation equation is

$$\log y_{i,j,t} = \log n_{i,j,t} + \nu_{i,j,t} \tag{8.1}$$

where the observation errors $\nu_{i,j,t}$ are normally distributed on the log-scale such that $\nu_{i,j,t} \sim \mathrm{N}(0, \sigma_i^2)$, where σ_i^2 are age-class-specific observation error variances. The priors for these variance parameters were uniform distributions $\sigma_i^2 \sim \mathrm{U}(0, 10)$ (very similar results were also obtained using Cauchy priors such as presented in Section 2.2.5). An example of posterior distributions of variance parameters from a model using this approach, that shares an identical process model the best fitting model in Chapter 2 (M_{BLL}), is shown in Figure 8.1.

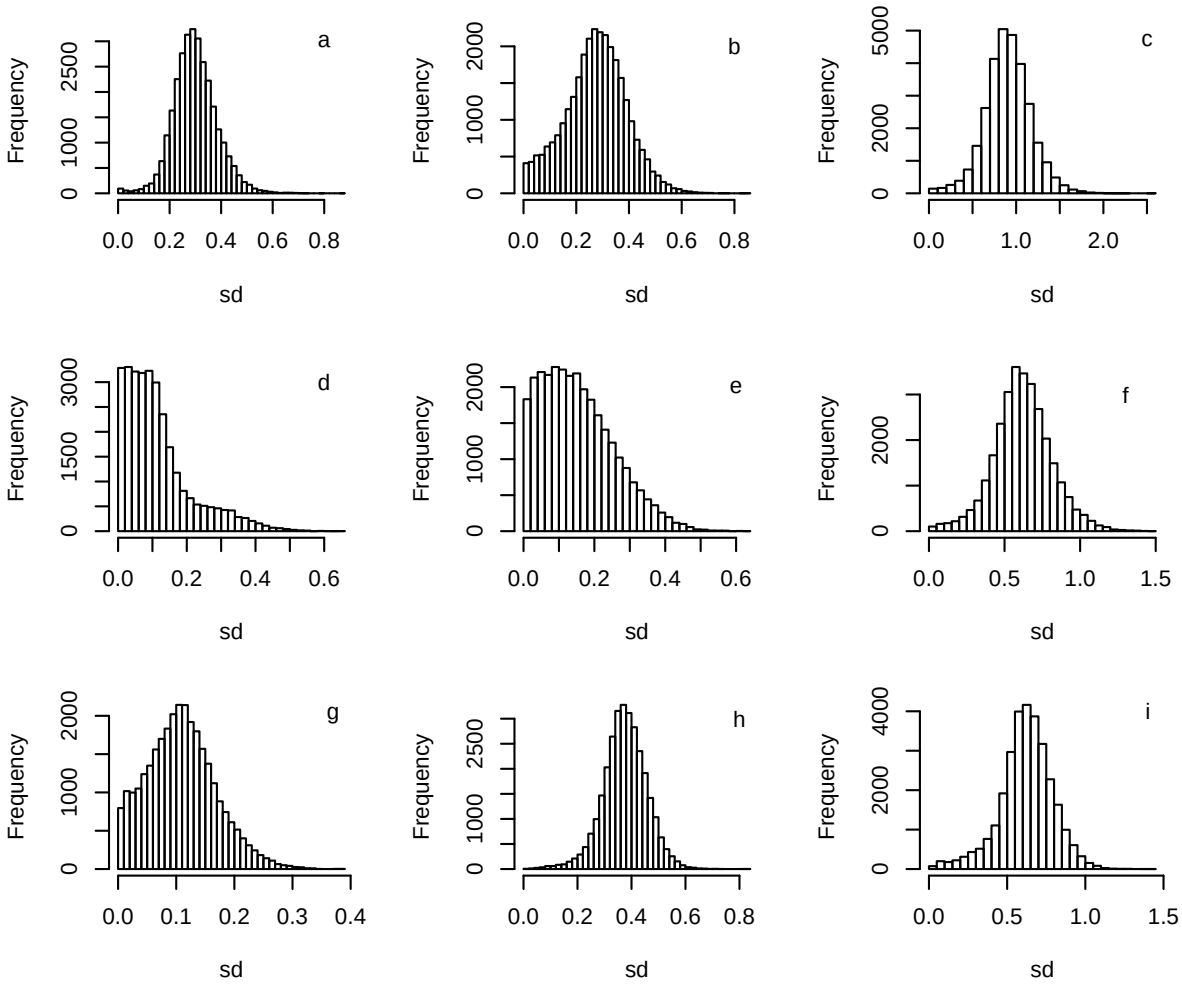

Figure 8.1: Posterior distributions of variance components parameters for a model with uninformative priors on all variance parameters. Parameters are standard deviations for the temporal components of process errors (τ_i; a–c), the spatial components of process errors (δ_i; d–f), and the observation errors (σ_i; g–i). Standard deviation parameters are specific to juveniles (a,d,g), smolt (b,e,h) and spawners (c,f,i).

8.3 Correction of the TAUC estimator for non-zero tail-counts

If non-zero tail counts are encountered in periodic count datasets then (6.8)–(6.9) in Section 6.2.2.2 will be negatively biased. A correction factor has previously been used (Hilborn et al., 1999) to minimize this bias by predicting the number of fish-days before the first ($\widehat{F}_T^{-1}$) or after the last count ($\widehat{F}_T^{+n}$), based on the stream-life of fish. Assuming both the first (c_1) and last (c_n) counts were non-zero, these, along with the corrected estimate of total fish-days, are given by

$$\widehat{F}_T^{-1} = \frac{c_1 l}{2} \tag{8.2}$$

$$\widehat{F}_T^{+n} = \frac{c_n l}{2} \tag{8.3}$$

$$\widehat{F}_T = 0.5 \left[\widehat{F}_T^{-1} + \widehat{F}_T^{+n} + \sum_{i=1}^{n-1} (d_{i+1} - d_i)(c_i + c_{i+1}) \right] \tag{8.4}$$

where l is again stream-life. Of course only one or other of (8.2) or (8.3) would be applied if only one non-zero count was encountered at the start or end of the periodic counts. The estimate of fish-days given by (8.4) is equivalent to extending the periodic count dataset by one count at each end by imputing a zero count l days before the first count (denoted c_0) and l days after the last count (denoted c_{n+1}). This then allows the usual calculation (6.8) of fish-days, though for the extended sample, e.g.

$$\begin{aligned} \widehat{F}_T &= 0.5 \left[(c_0 + c_{n+1})\hat{l} + \sum_{i=1}^{n-1} (d_{i+1} - d_i)(c_i + c_{i+1}) \right] \\ &= 0.5 \sum_{i=1}^{n} (d_{i+1} - d_i) c_i \end{aligned}$$

where $d_1 - d_0$ and $d_{n+1} - d_n = \hat{l}$. By redefining the correction factor in this way it can also be used when AS estimators (Section 6.2.3) are fitted to data with non-zero tail-counts, and makes them directly comparable to the TAUC estimator in these situations.

8.4 Graphical partially marginalized predictive checks for M_{BLL}

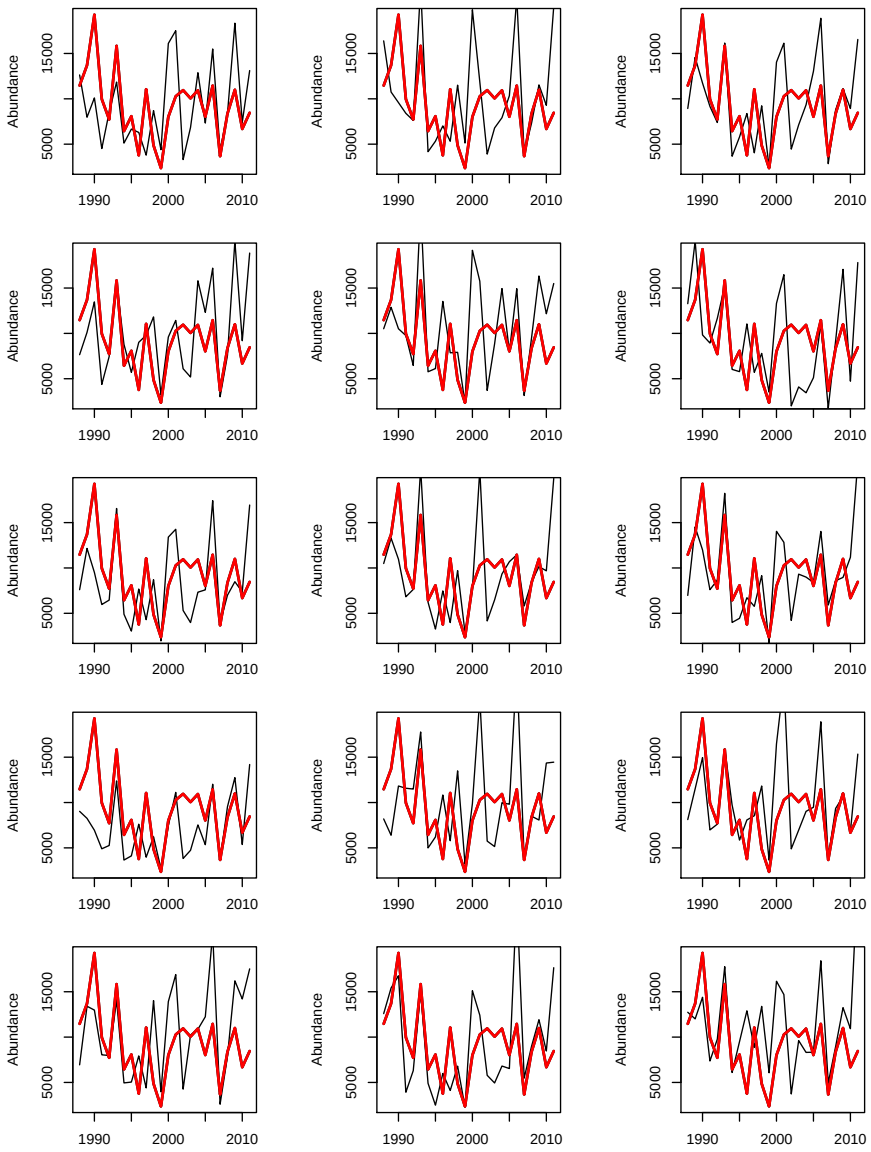

Figure 8.2: Examples of partially marginalized predictive dynamics of juvenile abundance in EF, Lobster Creek. The red line is the observed data ($y_{1,1,t}$) and the black lines are individual predictions of replicated data ($\tilde{y}^{rep}_{1,1,t}$) with each figure displaying one replicates randomly selected from the full set of simulated datasets.

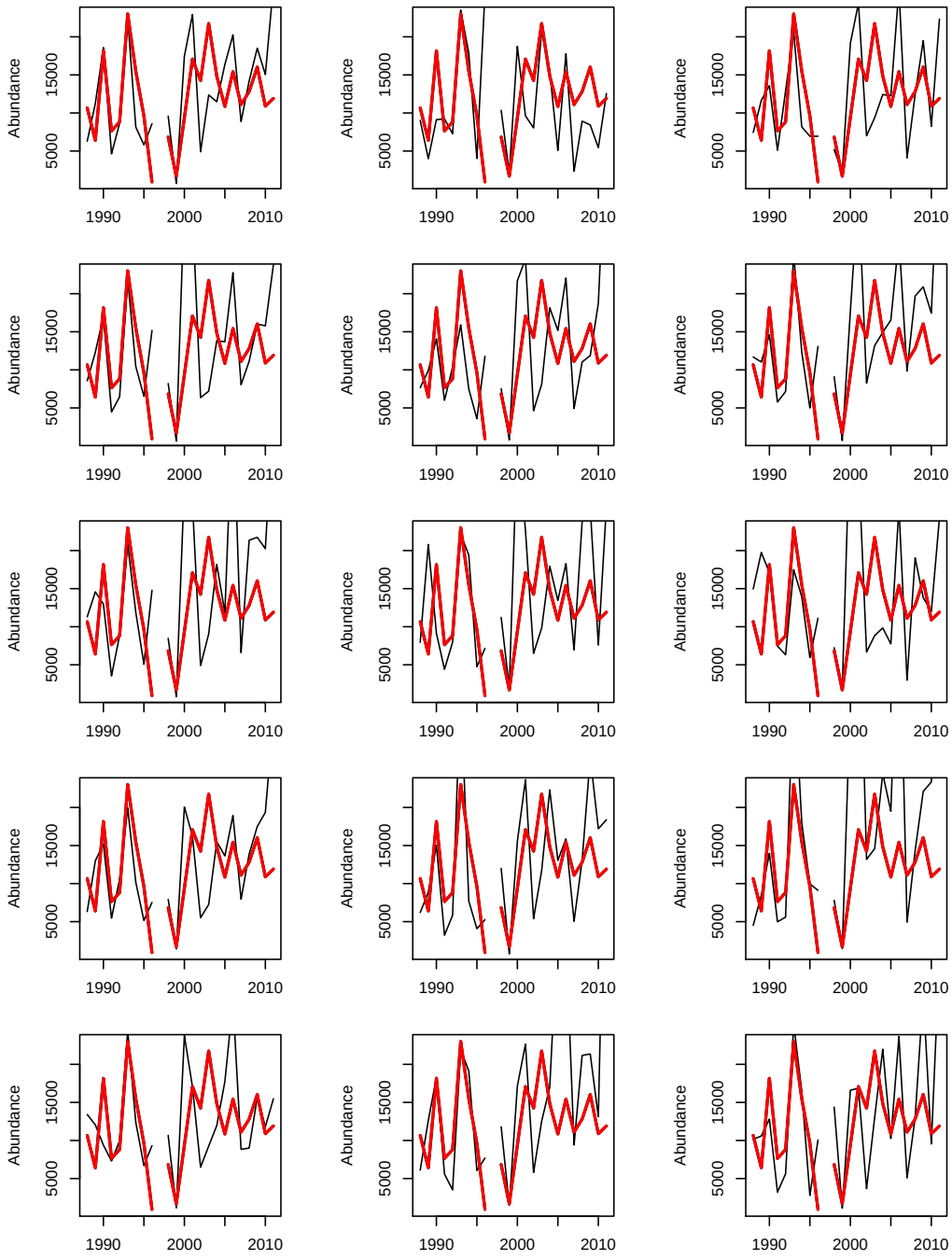

Figure 8.3: Examples of partially marginalized predictive dynamics of juvenile abundance in MF, Lobster Creek. The red line is the observed data ($y_{1,2,t}$) and the black lines are individual predictions of replicated data ($\tilde{y}^{rep}_{1,2,t}$) with each figure displaying one replicates randomly selected from the full set of simulated datasets.

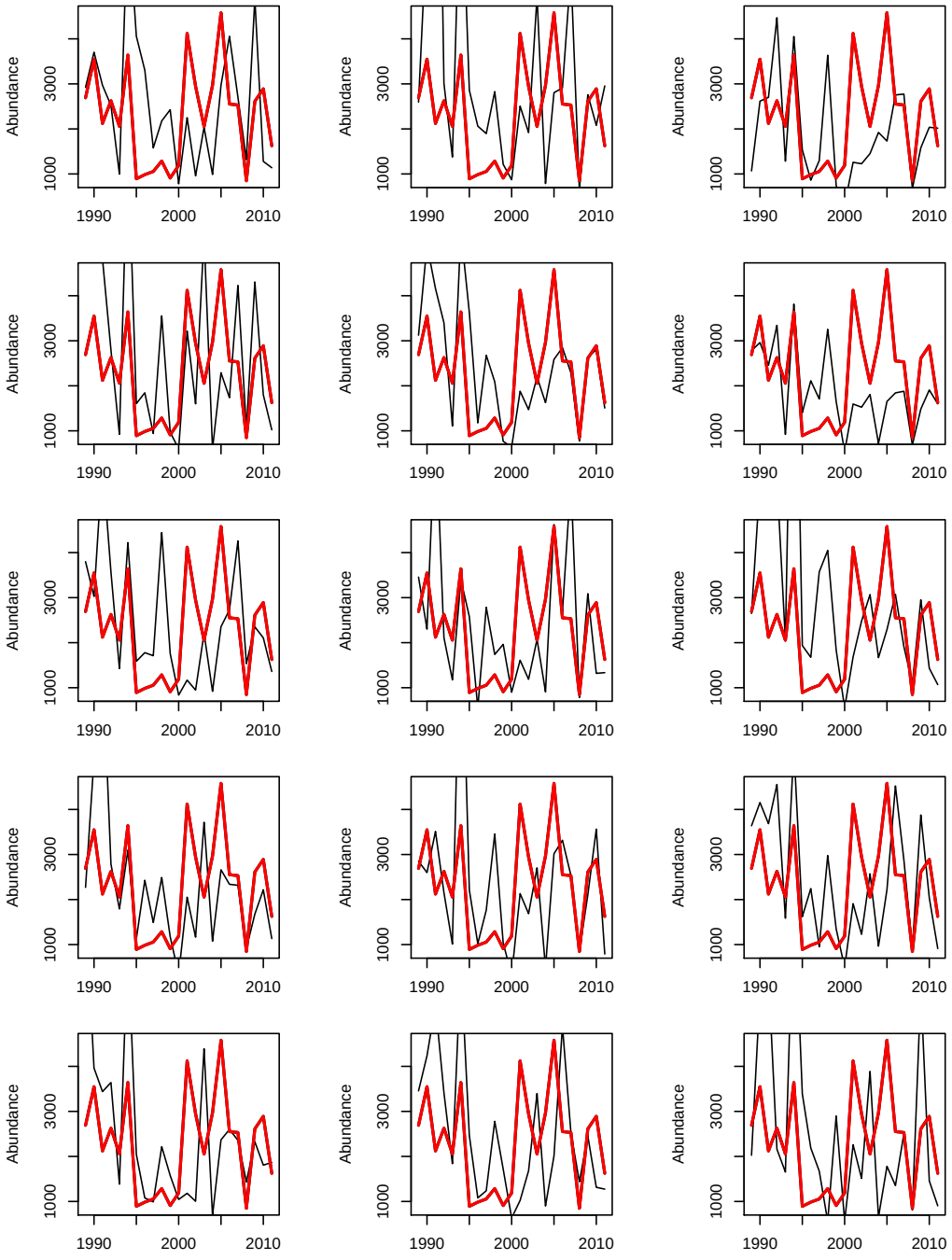

Figure 8.4: Examples of partially marginalized predictive dynamics of smolt abundance in EF, Lobster Creek. The red line is the observed data ($y_{2,1,t}$) and the black lines are individual predictions of replicated data ($\tilde{y}^{rep}_{2,1,t}$) with each figure displaying one replicates randomly selected from the full set of simulated datasets.

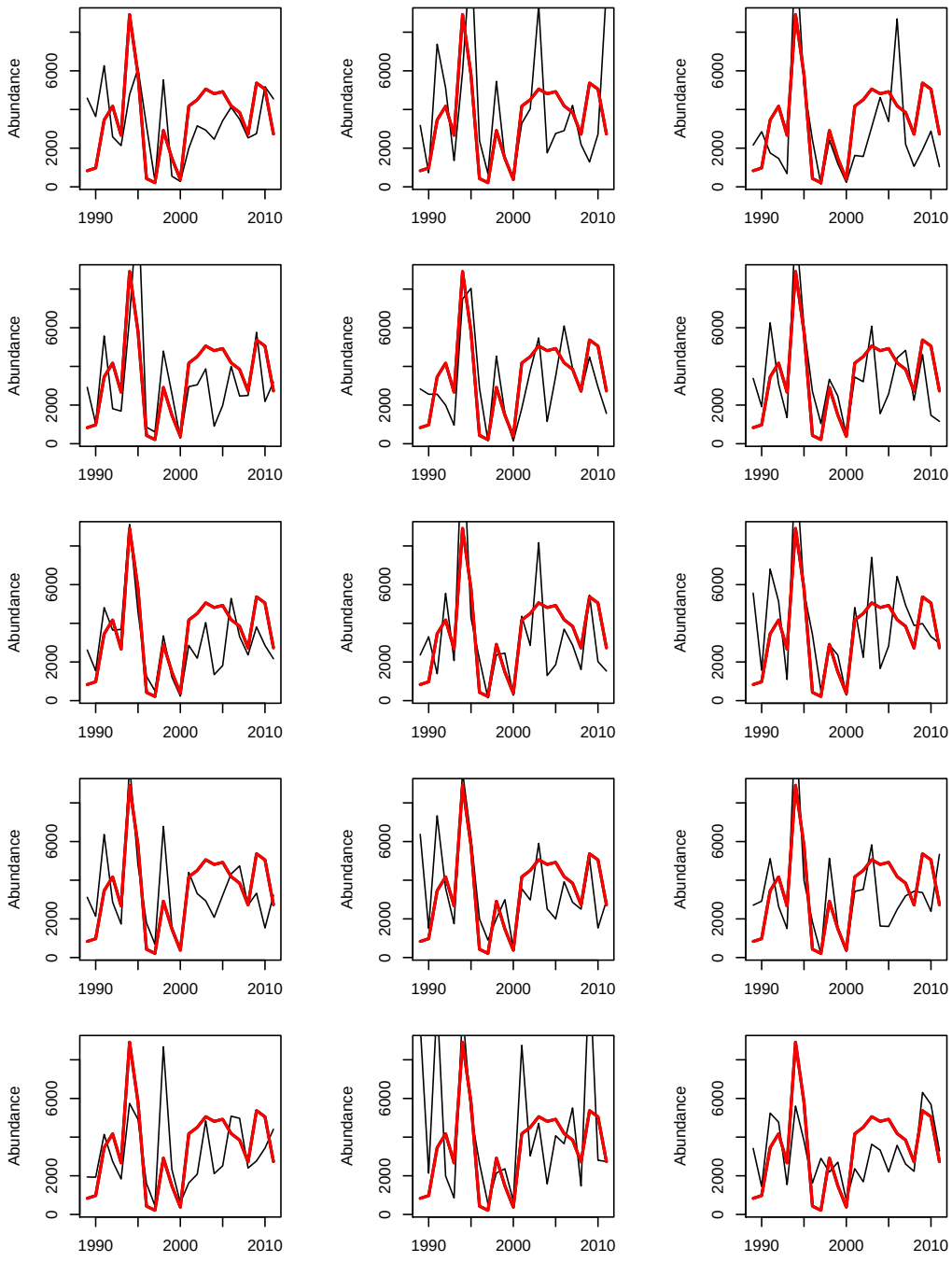

Figure 8.5: Examples of partially marginalized predictive dynamics of smolt abundance in MF, Lobster Creek. The red line is the observed data ($y_{2,2,t}$) and the black lines are individual predictions of replicated data ($\tilde{y}^{rep}_{2,2,t}$) with each figure displaying one replicates randomly selected from the full set of simulated datasets.

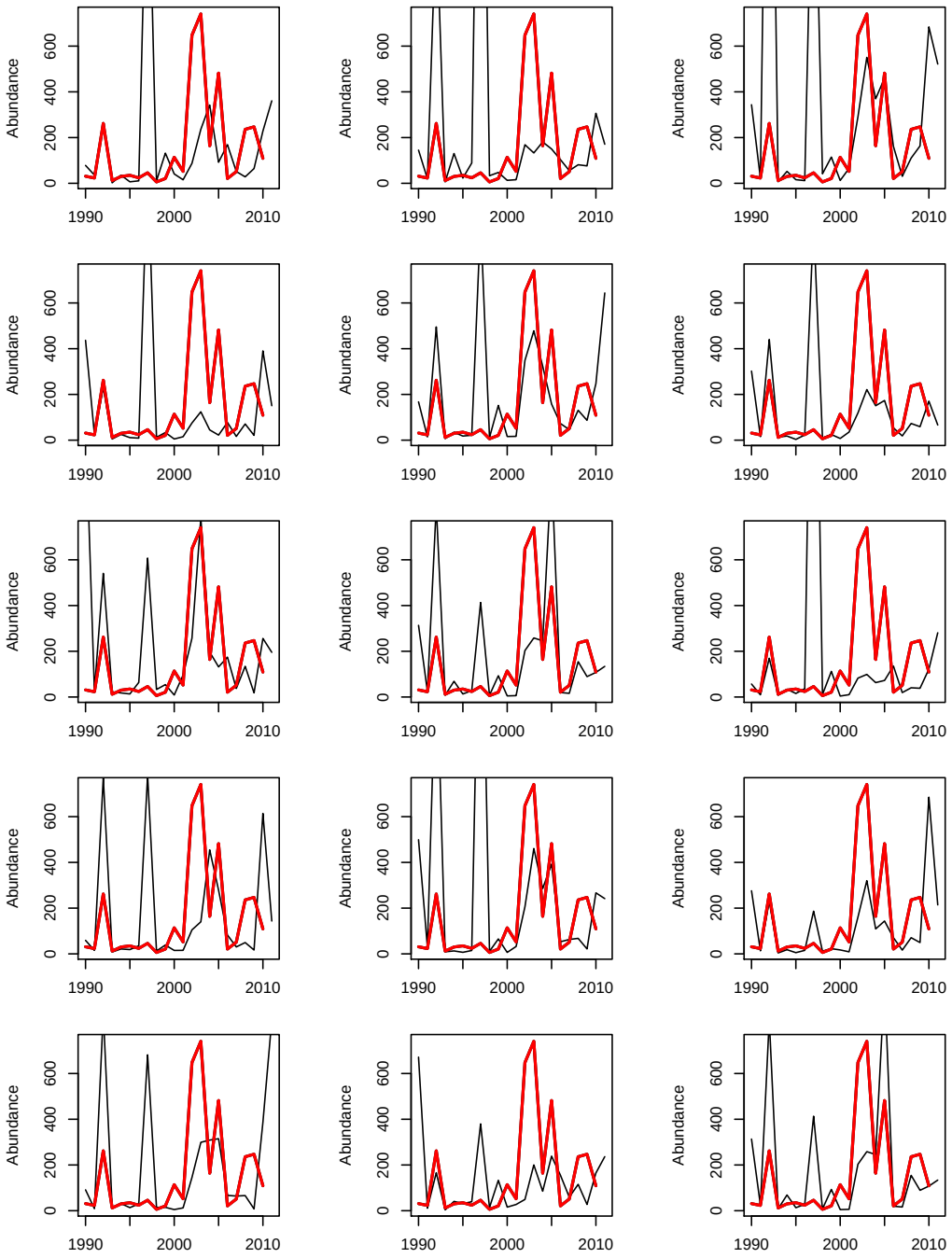

Figure 8.6: Examples of partially marginalized predictive dynamics of spawner abundance in EF, Lobster Creek. The red line is the observed data ($y_{3,1,t}$) and the black lines are individual predictions of replicated data ($\tilde{y}^{rep}_{3,1,t}$) with each figure displaying one replicates randomly selected from the full set of simulated datasets.

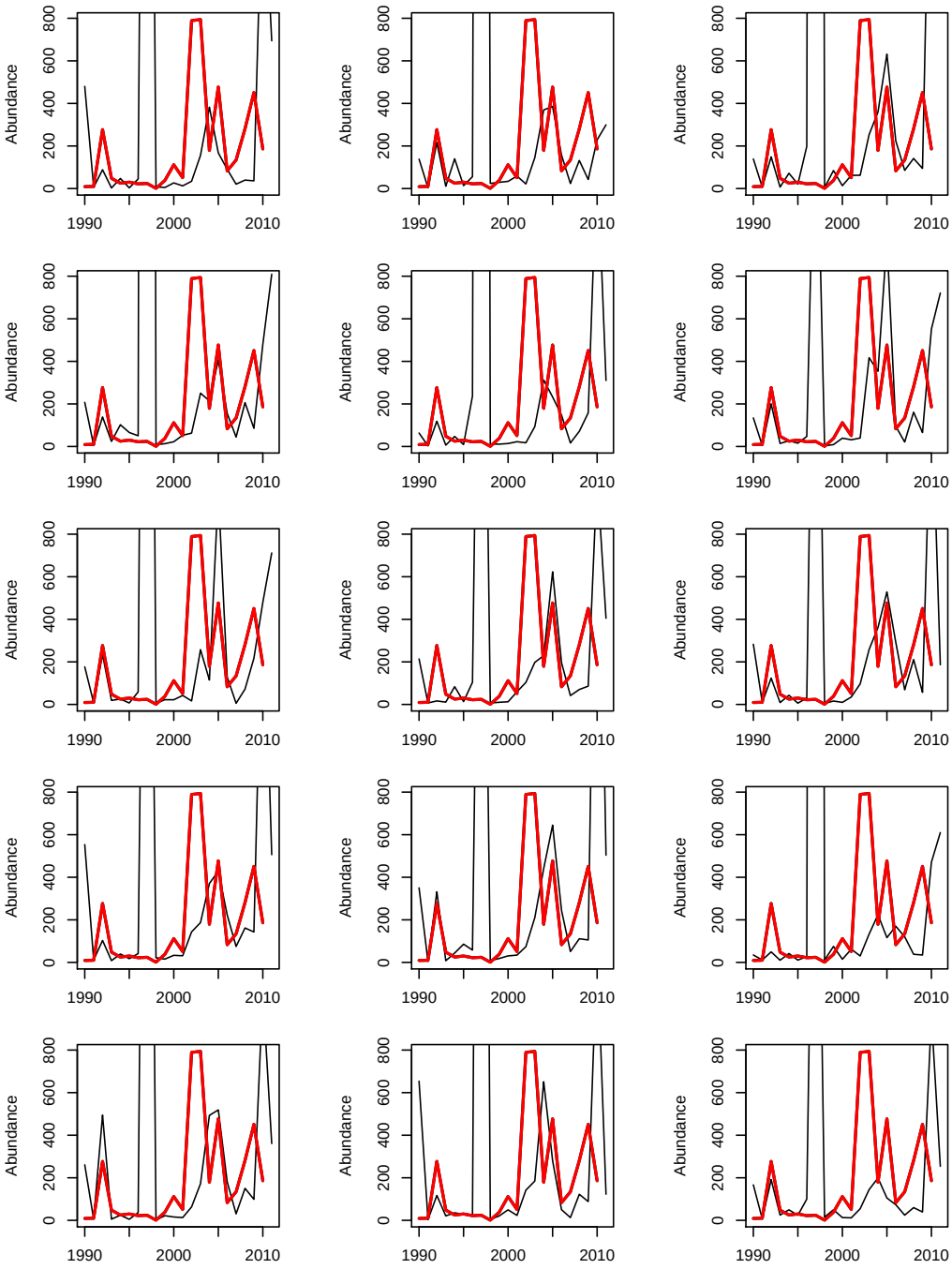

Figure 8.7: Examples of partially marginalized predictive dynamics of spawner abundance in MF, Lobster Creek. The red line is the observed data ($y_{3,2,t}$) and the black lines are individual predictions of replicated data ($\tilde{y}^{rep}_{3,2,t}$) with each figure displaying one replicates randomly selected from the full set of simulated datasets.

Bibliography

Abadi, F., Gimenez, O., Ullrich, B., Arlettaz, R., and Schaub, M. (2012). Estimation of immigration rate using integrated population modeling. *Journal of Applied Ecology* **47,** 393–400.

Adkison, M. D., Peterman, R. M., .Lapointe, M. F., Gillis, D., and Korman, J. (1996). Alternative models of climatic effects on Sockeye salmon, *Oncorhynchus nerka*, productivity Bristol Bay, Alaska, and the Fraser River, British Columbia. *Fisheries Oceanography* **5,** 137–152.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium on Information Theory*, pages 267–281.

Ames, J. (1984). Puget Sound chum salmon escapement estimates using spawner curve methodology. In *Proceedings of the workshop on stream indexing for salmon escapement estimation. Canadian Technical Report of Fisheries and Aquatic Science, No. 1326*, pages 133–148.

Anderson, D. (2001). The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin* **29,** 1294–1297.

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

Barrowman, N. J., Myers, R. A., Hilborn, R., Kehler, D. G., and Field, C. A. (2003). The variability among populations of coho salmon in the maximum reproductive rate and depensation. *Ecological Applications* **13,** 784–793.

Bayarri, M. J. and Berger, J. O. (2000). P values for composite null models. *Journal of the American Statistical Association* **95,** 1127–1142.

Bayarri, M. J. and Castellanos, M. E. (2007). Bayesian checking of the second levels of hierarchical models. *Statistical Science* **22,** 322–343.

Beamish, R. J., Noakes, D. J., McFarlane, G. A., Pinnix, W., Sweeting, R., and King, J. (2000). Trends in coho marine survival in relation to the regime concept. *Fisheries Oceanography* **9,** 114–119.

Beamish, R. J., Sweeting, R. M., Lange, K. L., and Neville, C. M. (2008). Changes in the population ecology of hatchery and wild coho salmon in the Strait of Georgia. *Transactions of the American Fisheries Society* **137,** 503–520.

Berg, A., Meyer, R., and Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics* **22,** 107–120.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58,** 540–547.

Besbeas, P., Lebreton, J. D., and Morgan, B. J. T. (2003). The efficient integration of abundance and demographic data. *Journal of the Royal Statistical Society Series C-Applied Statistics* **52,** 95–102.

Beverton, R. J. H. and Holt, S. J. (1957). On the dynamics of exploited fish populations. Fishery Investigations Series II Volume XIX, Ministry of Agriculture Fisheries and Food.

Bilby, R. E., Fransen, B. R., Walter, J. K., Cederholm, C. J., and Scarlett, W. J. (2001). Preliminary evaluation of the use of nitrogen stable isotope ratios to establish escapement levels for Pacific salmon. *Fisheries* **26,** 6–14.

Bilby, R. E. and Mollot, L. A. (2008). Effect of changing land use patterns on the distribution of coho salmon (*Oncorhynchus kisutch*) in the Puget Sound region. *Canadian Journal of Fisheries and Aquatic Sciences* **65,** 2138–2148.

Bolker, B. M. (2008). *Ecological models and data in R.* Princeton University Press, Princeton, New Jersey.

Botsford, L. W. and Lawence, C. A. (2002). Patterns of co-variability among California Current chinook salmon, coho salmon, Dungeness crab, and physical oceanographic conditions. *Progress in Oceanography* **53,** 283–305.

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society Series A-Statistics in Society* **143,** 383–430.

Bradford, M. J. (1995). Comparative review of Pacific salmon survival rates. *Canadian Journal of Fisheries and Aquatic Sciences* **52,** 1327–1338.

Bradford, M. J. and Irvine, J. R. (2000). Land use, fishing, climate change, and the decline of Thompson River, British Columbia, coho salmon. *Canadian Journal of Fisheries and Aquatic Sciences* **57,** 13–16.

Bradford, M. J., Myers, R. A., and Irvine, J. R. (2000). Reference points for coho salmon (*Oncorhynchus kisutch*) harvest rates and escapement goals based on freshwater production. *Canadian Journal of Fisheries and Aquatic Sciences* **57,** 677–686.

Bradford, M. J., Taylor, G. C., and Allan, J. A. (1997). Empirical review of coho salmon smolt abundance and the prediction of smolt production at the regional level. *Transactions of the American Fisheries Society* **126,** 49–64.

Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2000). Bayesian animal survival estimation. *Statistical Science* **15,** 357–376.

Brooks, S. P. and Gelman, A. (1998). Alternative methods for monitoring convergence of iterative simulation. *Journal of Computational and Graphical Statistics* **7,** 434–455.

Brooks, S. P., King, R., and Morgan, B. J. T. (2004). A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation* **27,** 515–529.

Bue, B. G., Fried, S. M., Sharr, S., Sharp, D. G., Wilcock, J. A., and Geiger, H. J. (1998). Estimating salmon escapement using area-under-the-curve, aerial observer efficiency, and stream-life estimates: the Prince William Sound pink salmon example. *North Pacific Anadromous Fish Commission Bulletin* **1,** 240–250.

Buhle, E. R., Holsman, K. K., Scheuerell, M. D., and Albaugh, A. (2009). Using an unplanned experiment to evaluate the effects of hatcheries and environmental variation on threatened populations of wild salmon. *Biological Conservation* **142,** 2449–2455.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, 2nd edition.

Calder, C., Lavine, M., Muller, P., and Clark, J. S. (2003). Incorporating multiple sources of stochasticity into dynamic population models. *Ecology* **84,** 1395–1402.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association* **87,** 493–500.

Caswell, H. (2001). *Matrix population models: construction, analysis, and interpretation*. Sinauer, Sunderland, 2nd edition.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). Deviance Information Criteria for missing data models. *Bayesian Analysis* **1,** 651–673.

Chaloupka, M. and Balazs, G. (2007). Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green turtle stock. *Ecological Modelling* **205,** 93–109.

Chen, D. G. and Holtby, L. B. (2002). A regional meta-model for stock-recruitment analysis using an empirical Bayesian approach. *Canadian Journal of Fisheries and Aquatic Sciences* **59,** 1503–1514.

Chilcote, M. W. (1999). *Conservation status of Lower Columbia River coho salmon*, volume 99-3. Oregon Department of Fish and Wildlife, Corvallis, Oregon, USA.

Chilcote, M. W. (2003). Relationship between natural productivity and the frequency of wild fish in mixed spawning populations of wild and hatchery steelhead (*Oncorhynchus mykiss*). *Canadian Journal of Fisheries and Aquatic Sciences* **60,** 1057–1067.

Clark, J. S. (2003). Uncertainty in ecological inference and forecasting. *Ecology* **84,** 1349–1350.

Cochran, W. G. (1977). *Sampling techniques*. Wiley, 3rd edition.

Colchero, F., Medellin, R. A., Clark, J. S., Lee, R., and Katul, G. G. (2009). Predicting population survival under future climate change: density dependence, drought and extraction in an insular bighorn sheep. *Journal of Animal Ecology* **78,** 666–673.

Coronado, C. and Hilborn, R. (1998). Spatial and temporal factors affecting survival in coho salmon (*Oncorhyncus kisutch*) in the Pacific Northwest. *Canadian Journal of Fisheries and Aquatic Sciences* **55,** 2067–2077.

Cubaynes, S., Doutrelant, C., Gregoire, A., Perret, P., and Faivre, B. (2012). Testing hypotheses in evolutionary ecology with imperfect detection: capture-recapture structural equation modeling. *Ecology* **93,** 248–255.

Davies, T. D. and Jonsen, I. D. (2011). Identifying nonproportionality of fishery-independent survey data to estimate population trends and assess recovery potential for cusk (*Brosme brosme*). *Canadian Journal of Fisheries and Aquatic Sciences* **68,** 413–425.

Dawid, A. P. (2002). Discussion of Speigelhalter et al. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64,** 624.

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

Einum, S. and Fleming, I. A. (2001). Implications of stocking: ecological interactions between wild and released salmonids. *Nordic Journal of Freshwater Research* **75,** 56–70.

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

Fargo, J. and Richards, L. J. (1998). A modern approach to catch-age analysis for Hecate Strait rock sole (*Pleuronectes bilineatus*). *Journal of Sea Research* **39,** 57–67.

Fewster, R. M. (2011). Variance estimation for systematic designs in spatial surveys. *Biometrics* **67,** 1518–1531.

Fleming, I. A. and Gross, M. R. (1993). Breeding success of hatchery and wild coho salmon (*Oncorhyncus kisutch*) in competition. *Ecological Applications* **3,** 230–245.

Ford, J. S. and Myers, R. A. (2008). A global assessment of salmon aquaculture impacts on wild salmonids. *PLoS Biology* **6,** e33.

Ford, M. J., Teel, D., Doornik, D. M. V., Kuligowski, D., and Lawson, P. W. (2004). Genetic population structure of central oregon coast coho salmon (*Oncorhynchus kisutch*). *Conservation Genetics* **5,** 797–812.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimisation Methods and Software* **27,** 233–249.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85,** 1–11.

Gelman, A. (2003). A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review* **71,** 369–382.

Gelman, A. (2004). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics* **13,** 755–779.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1,** 515–533.

Gelman, A. (2007). Comment: Bayesian checking of the second levels of hierarchical models. *Statistical Science* **22,** 349–352.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. CRC Press, London, UK.

Gelman, A., Mechelen, I. V., Verbeke, G., Heitjan, D. F., and Meulders, M. (2005). Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics* **61,** 74–85.

Gelman, A., Meng, X. L., and Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* **6,** 733–807.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6,** 721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernado, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 4*, Oxford, UK. Clarendon Press.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. Chapman and Hall, New York.

Gimenez, O., Bonner, S., King, R., Parker, R. A., Brooks, S. P., Jamieson, L. E., Grosbois, V., Morgan, B. J. T., and Thomas, L. (2009). WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo methods. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling demographic processes in marked populations*, volume 3 of *Environmental and Ecological Statistics*, pages 883–915, Berlin, Germany. Springer.

Good, T. P., Waples, R. S., and Adams, P. (2005). Updated status of federally listed ESUs of west coast salmon and steelhead. NOAA Technical Memorandum NMFS-NWFSC-66, U.S. Department of Commerce.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82,** 711–732.

Groot, C. and Margolis, L. (1991). *Pacific salmon life histories.* UBC Press, Vancouver, B.C.

Gustafson, R. G., Waples, R. S., Myers, J. M., Weitkamp, L. A., Bryant, G. J., Johnson, O. W., and Hard, J. J. (2007). Pacific salmon extinctions: Quantifying lost and remaining diversity. *Conservation Biology* **21,** 1009–1020.

Hankin, D. G. and Reeves, G. H. (1988). Estimating total fish abundance and total habitat area in small streams based on visual estimation methods. *Canadian Journal of Fisheries and Aquatic Sciences* **45,** 834–844.

Hartman, G. F., Scrivener, J. C., and Miles, M. J. (1996). Impacts of logging in Carnation Creek, a high-energy coastal stream in British Columbia, and their implication for restoring fish habitat. *Canadian Journal of Fisheries and Aquatic Sciences* **53,** 237–251.

Hilborn, R. (2003). The state of the art in stock assessment: where we are and where we are going. *Scientia Marina* **67,** 15–20.

Hilborn, R. (2006). Fisheries success and failure: The case of the Bristol Bay salmon fishery. *Bulletin of Marine Science* **78,** 487–498.

Hilborn, R. (2009). Life history models for salmon management: the challenges. In Knudsen, E. E., , and Michael, J. H., editors, *Pacific salmon environmental and life history models: advancing science for sustainable salmon in the future*, volume Symposium 71, Bethesda, Maryland. American Fisheries Society.

Hilborn, R., Bue, B. G., and Sharr, S. (1999). Estimating spawning escapements from periodic counts: a comparison of methods. *Canadian Journal of Fisheries and Aquatic Sciences* **56,** 888–896.

Hilborn, R. and Walters, C. J. (1992). *Quantitative fisheries stock assessment: Choice, dynamics and uncertainty.* Chapman and Hall, New York.

Hindar, K., Fleming, I. A., McGinnity, P., and Diserud, O. (2006). Genetic and ecological effects of salmon farming on wild salmon: modelling from experimental results. *ICES Journal of Marine Science* **63,** 1234–1247.

Hobday, A. J. and Boehlert, G. W. (2001). The role of coastal ocean variation in spatial and temporal patterns in survival and size of coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences* **58,** 2021–2036.

Holdo, R. M., Sinclair, A. R. E., Dobson, A. P., Metzger, K. L., Bolker, B. M., Ritchie, M. E., and Holt, R. D. (2009). A disease-mediated trophic cascade in the Serengeti and its implications for ecosystem c. *PLOS Biology* **7,** e1000210.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81,** 945–960.

Holt, C. A. and Peterman, R. M. (2007). Missing the target: uncertainties in achieving management goals in fisheries on Fraser River, British Columbia, sockeye salmon (*Onchorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences* **63,** 2722–2733.

Holt, K. R. and Cox, S. P. (2008). Evaluation of visual survey methods for monitoring Pacific salmon (*Onchorhynchus spp.*) escapement in relation to conservation guidelines. *Canadian Journal of Fisheries and Aquatic Sciences* **65,** 212–226.

Irvine, J. R., Bocking, R. C., English, K. K., and Labelle, M. (1992). Estimating coho salmon (*Oncorhynchus kisutch*) spawning escapements by conducting visual surveys in areas selected

using stratified random and stratified index sampling designs. *Canadian Journal of Fisheries and Aquatic Sciences* **49,** 1972–1981.

Irvine, J. R. and Fukuwaka, M. (2011). Pacific salmon abundance trends and climate change. *ICES Journal of Marine Science* **68,** 1122–1130.

Johnson, K. (1984). A history of coho fisheries and management in Oregon through 1982. Information Report 84-12, Oregon Department of Fish and Wildlife.

Johnson, S. L., Rodgers, J. D., Solazzi, M. F., and Nickelson, T. E. (2005). Effects of an increase in large wood on abundance and survival of juvenile salmonids (*Oncorhynchus spp.*) in an oregon coastal stream. *Canadian Journal of Fisheries and Aquatic Sciences* **62,** 412–424.

Kass, R. E. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association* **90,** 773–795.

Kery, M. and Schaub, M. (2011). *Bayesian population analysis using WinBUGS.* Academic Press, Waltham, USA.

King, R. (2012). A review of Bayesian state-space modelling of capture-recapture-recovery data. *Interface Focus* **2,** 190–204.

King, R. and Brooks, S. P. (2002a). Bayesian model discrimination for multiple strata capture-recapture data. *Biometrika* **89,** 785–806.

King, R. and Brooks, S. P. (2002b). Model selection for integrated recovery/recapture data. *Biometrics* **58,** 841–851.

King, R., Brooks, S. P., Mazzetta, C., Freeman, S. N., and Morgan, B. J. T. (2008). Identifying and diagnosing population declines: a Bayesian assessment of lapwings in the UK. *Journal of the Royal Statistical Society Series C-Applied Statistics* **57,** 609–632.

King, R., Morgan, B. J. T., Gimenez, O., and Brooks, S. P. (2009). *Bayesian analysis for population ecology.* Chapman and Hall, Boca Raton, FL, US.

Knudsen, E. E. (2000). Managing pacific salmon escapements: the gaps between theory and reality. In Knudsen, E. E., Steward, C. R., MacDonald, D. D., Williams, J. E., and Reiser, D. W., editors, *Sustainable fisheries management: pacific salmon*, pages 237–272, Boca Raton, Florida. Lewis Publishers.

Knudsen, E. E. and Michael, J. H. (2009). Introduction: the past, present and future of Pacific salmon life history models. In Knudsen, E. E., , and Michael, J. H., editors, *Pacific salmon environmental and life history models: advancing science for sustainable salmon in the future*, volume Symposium 71, Bethesda, Maryland. American Fisheries Society.

Korman, J., Ahrens, R. N. M., Higgins, P. S., and Walters, C. J. (2002). Effects of observer efficiency, arrival timing, and survey life on estimates of escapement for steelhead trout (*Oncoryhnchus mykiss*) derived from repeat mark-recapture experiments. *Canadian Journal of Fisheries and Aquatic Sciences* **59,** 1116–1131.

Koseki, Y. and Fleming, I. A. (2006). Spatio-temporal dynamics of alternative male phenotypes in coho salmon populations in response to ocean environment. *Journal of Animal Ecology* **75,** 445–455.

Koslow, J. A., Hobday, A. J., and Boehlert, G. W. (2002). Climate variability and marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon production area. *Fisheries Oceanography* **11,** 65–77.

Kostow, K. (1995). The natural production program. Biennial report on the status of wild fish conservation polices, Oregon Department of Fish and Wildlife.

Lady, J. M. and Skalski, J. R. (1998). Estimators of stream residence time of pacific salmon (*Oncorhynchus spp.*) based on release–recapture data. *Canadian Journal of Fisheries and Aquatic Sciences* **55,** 2580–2587.

Lawson, P. W., Logerwell, E. A., Mantua, N. J., Francis, R. C., and Agostini, V. N. (2004). Environmental factors influencing freshwater survival and smolt production in Pacific Northwest

coho salmon (*Onchorynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences* **61,** 360–373.

Lessard, R. B., Hilborn, R., and Chasco, B. E. (2008). Escapement goal analysis and stock reconstruction of sockeye salmon populations (*Oncorhynchus nerka*) using life-history models. *Canadian Journal of Fisheries and Aquatic Sciences* **65,** 2269–2278.

Logerwell, E. A., Mantua, N., Lawson, P. W., Francis, R. C., and Agostini, V. N. (2003). Tracking environmental processes in the coastal zone for understanding and predicting Oregon coho (*Oncorhynchus kisutch*) marine survival. *Fisheries Oceanography* **12,** 554–568.

Lorion, C. (2011). *Summary of habitat and fish monitoring data from East Fork and Upper Mainstem Lobster Creeks: 1988–2011.* Oregon Department of Fish and Wildlife, Corvallis, Oregon, USA.

Ludwig, D. and Walters, C. J. (1981). Measurement errors and uncertainty in parameter estimates for stock and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences* **38,** 711–720.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine* **28,** 3049–3067.

Magnusson, A. and Hilborn, R. (2007). What makes fisheries data informative? *Fish and Fisheries* **8,** 337–358.

Martell, S. J., Walters, C. J., and Hilborn, R. (2008). Retrospective analysis of harvest management performance for Bristol Bay and Fraser River sockeye salmon (*Oncorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences* **65,** 409–424.

Maunder, M. N. (2003). Paradigm shifts in fisheries stock assessment: from integrated analysis to Bayesian analysis and back again. *Natural Resource Modeling* **16,** 465–475.

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

McAllister, M. K., Pikitch, E. K., Punt, A. E., and Hilborn, R. (1994). A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* **51,** 2673–2687.

McCarthy, M. A. (2007). *Bayesian methods for ecology.* Cambridge University Press, Cambridge, UK.

Meng, X. L. and Vaida, F. (2006). Comment on article by Celeux et al. *Bayesian Analysis* **1,** 687–698.

Meyer, R. and Millar, R. B. (1999). BUGS in bayesian stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences* **56,** 1078–1086.

Millar, R. B. (2009). Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics* **65,** 962–969.

Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB.* Wiley, London.

Millar, R. B., McKechnie, S., and Jordan, C. E. (2012). Simple estimators of salmonid escapement and its variance using a new area-under-the-curve method. *Canadian Journal of Fisheries and Aquatic Sciences* **69,** 1002–1015.

Millar, R. B. and Meyer, R. (2000a). Bayesian state-space modeling of age-structured data: fitting the model is just the beginning. *Canadian Journal of Fisheries and Aquatic Sciences* **57,** 43–50.

Millar, R. B. and Meyer, R. (2000b). Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *Applied Statistics* **49,** 327–342.

Millar, R. B. and Olsen, D. (1995). Abundance of large toheroa (*Paphies ventricosa* Gray) at Oreti Beach, 1971–90, estimated from two-dimensional systematic samples. *New Zealand Journal of Marine and Freshwater Research* **29,** 93–99.

Moore, J. E. and Barlow, J. (2011). Bayesian state-space model of fin whale abundance trends from a 1991-2008 time series of line-transect surveys in the california current. *Journal of Applied Ecology* **48,** 1195–1205.

Mueter, F. J., Peterman, R. M., and Pyper, B. J. (2002). Opposite effects of ocean temperature on survival rates of 120 stocks of Pacific salmon (*oncorhynchus* spp.) in northern and southen areas. *Canadian Journal of Fisheries and Aquatic Sciences* **59,** 456–463.

Mueter, F. J., Ware, D. M., and Peterman, R. M. (2002). Spatial correlation patterns in coastal environmental variables and survival rates of salmon in the north-east Pacific Ocean. *Fisheries Oceanography* **11,** 205–218.

Naiman, R. J., Bilby, R. E., Schindler, D. E., and Helfield, J. M. (2002). Pacific salmon, nutrients, and the dynamics of freshwater and riparian ecosystems. *Ecosystems* **5,** 399–417.

Newman, K. B., Buckland, S. T., Lindley, S. T., Thomas, L., and Fernandez, C. (2006). Hidden process models for animal population dynamics. *Ecological Applications* **16,** 74–86.

Newman, K. B., Fernandez, C., Thomas, L., and Buckland, S. T. (2008). Monte Carlo inference for state-space models of wild animal populations. *Biometrics* **65,** 572–583.

Nickelson, T. (2003). The influence of hatchery coho salmon (*Oncorhynchus kisutch*) on the productivity of wild coho salmon populations in Oregon coastal basins. *Canadian Journal of Fisheries and Aquatic Sciences* **60,** 1050–1056.

Nickelson, T. E. and Lawson, P. W. (1998). Population viability of coho salmon, *Oncorhynchus kisutch*, in Oregon coastal basins: application of a habitat-based life cycle model. *Canadian Journal of Fisheries and Aquatic Sciences* **55,** 2383–2392.

Nickelson, T. E., Rogers, J. D., Johnson, S. L., and Solazzi, M. F. (1992). Seasonal changes in habitat use by juvenile coho salmon (*Oncorhynchus kisutch*) in Oregon coastal streams. *Canadian Journal of Fisheries and Aquatic Sciences* **49,** 783–789.

Nickelson, T. E., Solazzi, M. F., and Johnson, S. L. (1986). Use of hatchery coho salmon (*Oncorhynchus kisutch*) to rebuild wild populations in Oregon coastal streams. *Canadian Journal of Fisheries and Aquatic Sciences* **43,** 2443–2449.

Oosterhout, G. R., Huntington, C. W., Nickelson, T. E., and Lawson, P. W. (2005). Potential benefits of a conservation hatchery program for supplementing Oregon coast coho salmon (*Oncorhynchus kisutch*) populations: a stochastic model investigation. *Canadian Journal of Fisheries and Aquatic Sciences* **62,** 1920–1935.

Pacific Fishery Management Council (2009). Preseason report I: Stock abundance analysis for 2009 ocean salmon fishery regulations. Preseason report, Pacific Fishery Management Council.

Pacific Fishery Management Council (2012). Preseason report I: Stock abundance analysis and environmental assessment part 1 for 2012 ocean salmon fishery regulations. Preseason report, Pacific Fishery Management Council.

Parken, C. K., Bailey, R. E., and Irvine, J. R. (2003). Incorporating uncertainty into area-under-the-curve and peak count salmon escapement estimation. *North American Journal of Fisheries Management* **23,** 78–90.

Parsons, A. and Skalski, J. R. (2009). A statistical critique of estimating salmon escapement in the Pacific Northwest. In *The design and analysis of salmonid tagging studies in the Columbia Basin*, volume XXIV, Portland, Oregon. Bonneville Power Administration.

Pella, J. J. (1993). Utility of structural time series models and the kalman filter for predicting consequences of fishery actions. Report number 93-02, 571–593, Alaska Sea Grant College Program.

Pess, G. R., Montgomery, D. R., Steel, E. A., Bilby, R. E., Feist, B. E., and Greenberg, H. M. (2010). Landscape characteristics, land use, and coho salmon (*Oncorhynchus kisutch*) abundance, Snohomish River, Wash., U.S.A. *Canadian Journal of Fisheries and Aquatic Sciences* **59,** 613–623.

Peterman, R. M., Pyper, B. J., and Grout, J. A. (2003). Comparison of parameter estimation methods for detecting climate-induced changes in productivity of pacific salmon (*Oncorhynchus spp.*). *Canadian Journal of Fisheries and Aquatic Sciences* **60,** 809–824.

Peterman, R. M., Pyper, B. J., and MacGregor, B. W. (2003). Use of the Kalman filter to reconstruct historical trends in productivity of Bristol Bay sockeye salmon (*Oncorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences* **60,** 809–824.

Petrosky, C. E. and Schaller, H. A. (2010). Influence of river conditions during seaward migration and ocean conditions on survival rates of Snake River Chinook salmon and steelhead. *Ecology of Freshwater Fish* **19,** 520–536.

Piccinato, L. (2000). Asymptotic distribution of P values in composite null models: comment. *Journal of the American Statistical Association* **95,** 1166–1167.

Pinsky, M. L., Springmeyer, D. B., Goslin, M. N., and Augerot, X. (2009). Range-wise selection of catchments for Pacific salmon conservation. *Conservation Biology* **23,** 680–691.

Plummer, M. (2006). Comment on article by Celeux et al. *Bayesian Analysis* **1,** 681–686.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics* **1,** 523–539.

Polacheck, T., Hilborn, R., and Punt, A. E. (1993). Fitting surplus production models - comparing methods and measuring uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences* **50,** 2597–2607.

Punt, A. E. and Hilborn, R. (1997). Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries* **7,** 35–63.

Pyper, B. J., Peterman, R. M., Lapointe, M. F., and Walters, C. J. (1999). Patterns of covariation in length and age at maturity of British Columbia and Alaska sockeye salmon (*Oncorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences* **56,** 1046–1057.

Quinn, T. J. and Deriso, R. B. (1999). *Quantitative fish dynamics.* Oxford University Press, New York.

Quinn, T. J., Dickerson, B. R., and Vollestad, L. A. (2005). Marine survival and distribution patterns of two Puget Sound hatchery populations of coho (*Oncorhynchus kisutch*) and chinook (*Oncorhynchus tshawytscha*) salmon. *Fisheries Research* **76,** 209–220.

R Development Core Team (2010). *R: A language and environment for statistical computing.* R Project for Statistical Computing, Vienna, Austria.

Reynolds, J. H., Woody, C. A., Gove, N. E., and Fair, L. F. (2007). Efficiently estimating salmon escapement uncertainty using systematically sampled data. In *Proceedings of the American Fisheries Society Symposium*, volume 54, pages 121–129.

Rhodes, J. R., Ng, C. F., de Villiers, D. L., Preece, H. J., McAlphine, C. A., and Possingham, H. P. (2011). Using integrated population modelling to quantify the implications of multiple threatening processes for a rapidly declining population. *Biological Conservation* **144,** 1081–1088.

Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Research Board of Canada* **11,** 559–623.

Rivot, E. and Parent, E. (2001). How robust are bayesian posterior inferences based on a ricker model with regards to measurement errors and prior assumptions about parameters. *Canadian Journal of Fisheries and Aquatic Sciences* **58,** 2284–2297.

Rivot, E., Prevost, E., Bagliniere, J. L., and Parent, E. (2004). A bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data. *Ecological Modelling* **179,** 463–485.

Robins, J. M., Van Der Vaart, A., and Ventura, V. (2000). The asymptotic distribution of p values in composite null models. *Journal of the American Statistical Association* **95,** 1143–1156.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* **12,** 1151–1172.

Schaub, M. and Abadi, F. (2011). Integrated population models: a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology* **152,** 227–237.

Schaub, M., Gimenez, O., Sierro, A., and Arlettaz, R. (2007). Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology* **21,** 945–955.

Schaub, M., Reichlin, T. S., Abadi., F., Kery, M., Jenni, L., and Arlettaz, R. (2012). The demographic drivers of local population dynamics in two rare migratory birds. *Oecologia* **168,** 97–108.

Scheuerell, M. D., Hilborn, R., Ruckelshaus, M. H., Bartz, K. K., Lagueux, K. M., Haas, A. D., and Rawson, K. (2006). The Shiraz model: a tool for incorporating anthropogenic effects and fish-habitat relationships in conservation planning. *Canadian Journal of Fisheries and Aquatic Sciences* **63,** 1596–1607.

Schnute, J. T. (1994). A general framework for developing sequential fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences* **51,** 1676–1688.

Scrivener, J. C. and Brownlee, M. J. (1989). Effects of forest harvesting on spawning gravel and inclubation survival of chum (*Onchorhynchus keta*) and coho salmon (*O. kisutch*) in Carnation Creek, British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* **46,** 681–696.

Seber, G. A. F. and Le Cren, E. D. (1967). Estimating population parameters from catches large relative to the population. *Journal of Animal Ecology* **36,** 631–643.

Shardlow, T., Hilborn, R., and Lightly, D. (1987). Components analysis of insteam escapement methods for Pacific salmon (*Oncorhynchus spp.*). *Canadian Journal of Fisheries and Aquatic Sciences* **44,** 1031–1037.

Shriner, D. and Yi, J. N. (2009). Deviance information criterion (DIC) in Bayesian multiple QTL mapping. *Computational Statistics and Data Analysis* **53,** 1850–1860.

Sigler, M. F., Tollit, D. J., Vollenweider, J. J., Thedinga, J. F., Csepp, D. J., Womble, J. N., Wong, M. A., Rehberg, M. J., and Trites, A. W. (2009). Stellar sea lion foraging response to seasonal changes in prey available. *Marine Ecology Progress Series* **388,** 243–261.

Sinharay, S. and Stern, H. S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference* **111,** 209–221.

Skalski, J. R., Buchanan, R. A., Townsend, R. L., Steig, T. W., and Hemstrom, S. (2009). A multiple-release model to estimate route-specific and dam passage survival at a hydroelectric project. *North American Journal of Fish Management* **29,** 670–679.

Skalski, J. R., Hoffmann, A., Ransom, B. H., and Steig, T. W. (1993). Fixed-location hydroacoustic monitoring designs for estimating fish passage using stratified random and systematic sampling. *Canadian Journal of Fisheries and Aquatic Sciences* **50,** 1208–1221.

Smith, J. (2002). Discussion of Speigelhalter et al. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64,** 619–620.

Snover, M. L. (2008). Comments on 'Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock'. *Ecological Modelling* **212,** 545–549.

Solazzi, M. F. (1984). Relationships between visual counts of coho, chinook and chum salmon from spawning fish surveys and the actual number of fish present. Information Report 84-7, Oregon Department of Fish and Wildlife.

Solazzi, M. F., Nickelson, T. E., Johnson, S. L., and Rodgers, J. D. (2000). Effects of increasing winter rearing habitat on abundance of salmonids in two coastal Oregon streams. *Canadian Journal of Fisheries and Aquatic Sciences* **57,** 906–914.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64,** 583–616.

Steel, E. A., Jensen, D. W., Burnett, K. M., Christiansen, K., Firman, J. C., Feist, B. E., Anlauf, K. J., and Larsen, D. P. (2012). Landscape characteristics and coho salmon (*Oncorhynchus kisutch*) distributions: explaining abundance versus occupancy. *Canadian Journal of Fisheries and Aquatic Sciences* **69,** 457–468.

Steinbakk, G. H. and Storvik, G. (2009). Posterior predictive p-values in Bayesian hierarchical models. *Scandinavian Journal of Statistics* **2,** 320–336.

Stern, H. S. (2000). Asymptotic distribution of P values in composite null models: comment. *Journal of the American Statistical Association* **95,** 1157–1159.

Stevens, D. L. (2002). Sampling design and statistical analysis methods for the integrated biological and physical monitoring of oregon streams. Opsw-odfw-2002-07, Oregon Department of Fish and Wildlife.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software* **12,** 1–16.

Su, Z. and Peterman, R. M. (2011). Performance of a Bayesian state-space model of semelparous

species for stock-recruitment data subject to measurement error. *Ecological Modelling* **224,** 76–89.

Su, Z. M., Adkison, M. D., and Alen, B. W. V. (2001). A hierarchical Bayesian model for estimating historical salmon escapement and escapement timing. *Canadian Journal of Fisheries and Aquatic Sciences* **58,** 1648–1662.

Szerlong, R. G. and Rundio, D. E. (2008). A statistical modeling method for estimating mortality and abundance of spawning salmon from a time series of counts. *Canadian Journal of Fisheries and Aquatic Sciences* **65,** 17–26.

Thedinga, J. F., Murpy, M. L., Johnson, S. W., Lorenz, J. M., and Koski, K. V. (1994). Determination of salmonids smolt yield with rotary-screw traps in the Situk River, Alaska, to predict effects of glacial flooding. *North American Journal of Fish Management* **14,** 837–851.

Thomas, L., Buckland, S. T., Newman, K. B., and Harwood, J. (2005). A unified framework for modelling wildlife population dynamics. *Australian and New Zealand Journal of Statistics* **47,** 19–34.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed effects models. *Biometrika* **92,** 351–370.

Wainwright, T. C., Chilcote, M. W., Lawson, P. W., Nickelson, T. E., Huntington, C. W., Mills, J. S., Moore, K. M. S., Reeves, G. H., Stout, H. A., and Weitkamp, L. A. (2008). Biological recovery criteria for the Oregon Coast coho salmon evolutionarily significant unit. NOAA Technical Memorandum NMFS-NWFSC-91, U.S. Department of Commerce.

Wallace, C. S. (2005). *Statistical and inductive inference by minimum message length.* Springer, New York.

Walters, C. J. (1985). Bias in the estimation of functional relationships from time series data. *Canadian Journal of Fisheries and Aquatic Sciences* **42,** 185–187.

Walters, C. J. (2009). A partial bias correction factor for stock-recruitment parameters in the presence of autocorrelated environmental effects. *Canadian Journal of Fisheries and Aquatic Sciences* **47,** 516–519.

Walters, C. J. and Ludwig, D. (1981). Effects of measurement errors on the assessment of stock-recruitment relationships. *Canadian Journal of Fisheries and Aquatic Sciences* **38,** 704–710.

Waples, R. S., Beechie, T., and Pess, G. R. (2009). Evolutionary history, habitat disturbance regimes, and anthropogenic changes: What do these mean for resilience of Pacific salmon populations? *Ecology and Society* **14,** 3.

Ward, E. J., Holmes, E. E., and Balcomb, K. C. (2009). Quantifying the effects of prey abundance on killer whale reproduction. *Journal of Applied Ecology* **46,** 632–640.

Weitkamp, L. and Neely, K. (2002). Coho salmon (*Onchorhynchus kisutch*) ocean migration patterns: insight from marine coded-wire tag recoveries. *Canadian Journal of Fisheries and Aquatic Sciences* **59,** 1100–1115.

Wilberg, M. J. and Bence, J. R. (2008). Performance of deviance information criterion model selection in statistical catch-at-age analysis. *Fisheries Research* **93,** 212–221.

Wolter, K. M. (1984). An investigation of some estimators of variance for systematic sampling. *Journal of the American Statistical Association* **79,** 781–790.

Wolter, K. M. (2007). *Introduction to variance estimation.* Springer, New York.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* **14,** 1261–1295.

Zhou, S. (2000). Stock assessment and optimal escapement of coho salmon in three Oregon coastal lakes. Information report 2000-07, Oregon Department of Fish and Wildlife.

Zhou, S., Punt, A. E., Deng, R., and Bishop, J. (2011). Estimating multifleet catchability coefficients and natural mortality from fishery catch and effort data: comparison of Bayesian

state-space and observation error models. *Canadian Journal of Fisheries and Aquatic Sciences* **68,** 1171–1181.

Zhou, S., Punt, A. E., Deng, R., Dichmont, C. M., Ye, Y., and Bishop, J. (2009). Modified hierarchical bayesian biomass dynamics models for assessment of short-lived invertebrates: a comparison for tropical tiger prawns. *Marine and Freshwater Research* **60,** 1298–1308.