Bayesian models of age and growth in sharks

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Information Sciences

by
James Bristow

Supervisors:
Dr. Adam Smith
Dr. Brit Finucci

School of Mathematical and Computational Sciences
Massey University
Auckland
New Zealand

2023

Bayesian models of age and growth in sharks

James Bristow
School of Mathematical and Computational Sciences, Massey University

2023

Abstract

Marine ecosystems are under increasing pressure from factors such as overfishing, that have lead to reported declines in chondrichthyan populations across the globe. Approximately 80% of New Zealand chondrichthyian species have no specific management or monitoring, and a lack of species-specific information has led to uncertainty concerning population trends over time.

Biological parameters such as growth rates are typically incorporated into fisheries growth models, the most common of which is the von Bertalanffy growth model (VBGM). The common VBGM is monophasic, and is therefore a continuous and monotone curve over all stages of growth. The VBGM has received some criticism concerning the assumption of monophasic growth. One critique is that a single curve fails to account for changes in energy reallocation and growth that occur during the transition between the juvenile and mature life stages of the shark. Conversely, a biphasic growth model may better model the patterns of both juvenile and mature sharks. The biphasic VBGM (BVBGM) has offered promising results when applied to chondrichthyans in previous growth studies.

Estimating the ages of chondrichthyans is typically performed by trained readers who count the growth bands deposited within the vertebral centra. However, the reading of vertebrae is subjective, error-prone, and time consuming. While vertebral band pairs are the most commonly used structure for age estimation, there are various sources of bias and uncertainty inherent to the reading process. Deep learning models such as convolutional neural networks (CNNs) have demonstrated promise for the automated interpretation of bony fish otolith growth zones, though the application of CNNs for the automated reading of chondrichthyan vertebrae has yet to be explored.

The principal goal of this thesis is to advance methodologies for estimating the growth parameters of sharks via the application of Bayesian models, thereby offering more robust management and conservation of chondrichthyes against various factors such as overfishing.

We first explored the potential of biphasic growth models on five species of New Zealand chondrichthyes: *Centrophorus squamosus*; *Isurus oxyrinchus*; *Lamna nasus*; *Mustelus lenticulatus*; and *Prionace glauca*. We compared two monophasic and two biphasic growth models using the Pareto-smoothed importance sampling approximation of leave-one-out cross-validation (PSIS-LOO) metric. Biphasic models appeared

to provide superior fit for both males and females in the majority of cases, and we were able to improve upon prior examples from the literature where parameter estimates were noted to be biased or poor. Our overall results demonstrated that the popular monophasic VBGM should not be chosen *a priori* as the only candidate model to describe the growth of chondrichthyans. Instead, we should consider multiple alternative growth models, as informed by statistical evidence and domain expertise.

We next explored the feasibility of automating the age estimation of chondrichthyans by training CNNs on an image dataset of *Isurus oxyrinchus* vertebrae. We evaluated three Bayesian deep learning methods for uncertainty quantification: DeepEnsembling, mixture of Laplace approximations (MoLA), and multi-stochastic weight average Gaussian (MultiSWAG) in terms of predictive power and model calibration. We found that MultiSWAG offered marginally superior predictive performance and model calibration relative to DeepEnsembling and MoLA. Moreover, predictions produced by MultiSWAG typically closely matched the estimates provided by human readers, though the perfomance of these deep learning models tended to degrade for older age classes.

We argue that our results demonstrate promise for emulating trained readers, leading to potential efficiency gains and cost savings. However, we note that there is a lack of evidence that our models are directly counting the bands of the vertebrae, and that further refinement of our CNNs may be required.

Our findings have demonstrated the ability of Bayesian methods to perform principled uncertainty quantification and parameter estimation within the context of age and growth modelling. Our Bayesian growth models were able to quantify the epistemic uncertainty of our parameter estimates, offering more robust estimation of growth parameters and superior model fit.

Additionally, we were able to incorporate prior information into our growth models, as informed by the available literature and domain expertise. Likewise, the application of Bayesian deep learning facilitated the quantification of epistemic and aleotoric uncertainty for our age estimates, while also offering superior predictive performance and well-calibrated prediction intervals. We showed that Bayesian CNNs could be used to efficiently automate the interpretation of vertebral growth bands for the purposes of age estimation.

This study has contributed to the research of New Zealand sharks, marine conservation, and fisheries management by improving methods used to measure and interpret growth parameters. We hope it contributes to the management and persistence of these species for future generations.

Keywords: Chondrichthyes, Sharks, Bayesian modelling, Somatic growth, Age studies, Growth studies

Table of contents

List of Figures

List of Tables

Acknowledgements

With special thanks to my patient and supportive supervisors Adam Smith and Brit Finucci, and to my colleagues at Massey University. This research involved many generous conversations and work sessions with friends, family, and collaborators who have supported me throughout this process. In particular, I would like to thank both Alex Burton and Hayden Rabel for their advice and input.

This work was supported by a Fisheries New Zealand Scholarship in Quantitative Fisheries Science. I would like to thank both William Gibson and Pamela Mace for their assistance during the completion of my thesis. Without their generosity, this research would not have been completed.

With thanks to Andy McKenzie, Richard O'Driscoll, Samik Datta, Bradley Moore, Richard Saunders, Matt Dunn, Kameron Christopher, and Teresa A'mar at the National Institute of Water and Atmospheric Research (NIWA). Your expertise and advice was invaluable during the development of this work.

I would also like to thank all participants of the Ministry for Primary Industries (MPI) fisheries observers programme for collecting the samples needed for the success of this research.

The COVID-19 pandemic brought about many unexpected challenges during this journey. In spite of numerous stumbling blocks and doubts, this journey has ultimately been a fulfilling one. I would like to humbly thank all those who helped to bring this thesis to life.

1 General introduction

1.1 Introduction

Marine ecosystems are under increasing pressure from several factors that have led to reported declines of many fish populations around the globe. Such factors include habitat destruction, pollution, climate change, and overfishing (Dulvy et al., 2008, 2021; Jorgensen et al., 2022). The ramifications of fishing activity on chondrichthyans (sharks, rays, and chimaeras) is a focal point of international research (Bradley & Gaines, 2014; Jorgensen et al., 2022; Pacoureau et al., 2021; Stevens et al., 2000). Approximately one third of chondrichthyan fishes are threatened with extinction (Dulvy et al., 2021), primarily as a consequence of overfishing (Dulvy et al., 2008). The life-history traits of chondrichthyans makes many species especially vulnerable to fishing pressure. These traits include slow growth, late age of sexual maturity, long reproduction cycles, and low fecundity (Jorgensen et al., 2022; Pacoureau et al., 2021; Stevens et al., 2000).

Fisheries contribute significantly to the economies of high-latitude nations such as Denmark, Iceland, Norway, and New Zealand (Allison et al., 2009). Eleven chondrichthyan species are managed under the New Zealand Quota Management System (QMS), and the concentration of fisheries activity on these species is particularly intense (Ford et al., 2018). However, approximately 80% of New Zealand chondrichthyans have no species-specific management or monitoring; these are principally deepwater chondrichthyan species, and some are caught in considerable numbers as by-catch in commercial fisheries (Finucci et al., 2019). Finucci et al. (2019) reported that this lack of species-specific information has led to uncertainty in abundance trends of New Zealand deepwater chondrichthyans over time. It is therefore imperative that populations of at-risk species are closely monitored to ensure

their continued sustainability (Dulvy et al., 2021; Jabado et al., 2018).

1.1.1 Fisheries growth models

Biological parameters such as growth rates (Pardo et al., 2013; Trippel & Harvey, 1991) and length-at-maturity estimates (Lappalainen et al., 2016; Trippel & Harvey, 1991) are commonly incorporated into fisheries growth models. Within fisheries science, the most commonly applied growth model is the von Bertalanffy growth model (VBGM), which provides informative parameters including growth rates and maximum lengths (Cailliet et al., 2006a; Contreras-Reyes et al., 2021; Flinn & Midway, 2021b; McFarlane et al., 2010; Minte-Vera et al., 2016; Moe, 2015; Ricker, 1975).

The VBGM is monophasic, and therefore represents growth as a continuous and monotone curve over all life stages (Contreras-Reyes et al., 2021). Though the VBGM is undoubtedly popular, the fundamental assumption that a monophasic curve can model the entire lifetime growth of sharks has received some criticism. One such critique is that a single curve will fail to account for changes in energy allocation and growth that occur at the onset of maturity, resulting in models that fit poorly to the smallest or largest of individuals (Moe, 2015; Quince et al., 2008). Hence, a biphasic growth model that represents growth as two distinct mechanistic phases may better characterise growth patterns for both the juvenile and mature life stages of sharks (Araya & Cubillos, 2006; Moe, 2015; Quince et al., 2008).

Soriano et al.'s (1992) biphasic VBGM (BVBGM) is one example of a biphasic growth model which has demonstrated promise when applied to chondrichthyes (Acuña et al., 2001; Araya & Cubillos, 2006; Aversa, Dans, Garcia, et al., 2011; Olmeda-de la Fuente et al., 2022). A Bayesian approach to biphasic growth modelling is advantageous, as prior information concerning potential contributors to phasic shifts, including changes in diet,

habitat, or reallocation of resources for reproductive investment, can easily be incorporated into the model (Contreras-Reyes et al., 2021).

1.1.2 Age estimation via vertebral reading

A common approach to the age estimation of chondrichthyans is the counting of vertebral growth bands by trained readers (Cailliet et al., 2006c; Carrier et al., 2012; Francis et al., 2007a; Harry, 2018). Despite its popularity, vertebrae reading is subjective (Francis, 2015a; Francis & Ó Maolagáin, 2016), error-prone (Francis & Ó Maolagáin, 2016) and time consuming (Moore et al., 2019, 2021).

The band pairs within the vertebral centra have become the most commonly used structure for age estimation studies (Carrier et al., 2018; Harry et al., 2022; Natanson et al., 2018). However, there are various sources of bias and uncertainty within the reading of these band pairs, and more recent age studies provide some doubt as to whether band counts provide an accurate measure of age, particularly for older sharks where systematic age underestimation has been identified (Francis et al., 2007b; Harry, 2018; Jorgensen et al., 2022; Natanson et al., 2018; Raoult et al., 2018). For instance, it has been argued that the vertebral centra grows proportionately with somatic growth, and is only indirectly correlated with age (Natanson et al., 2018).

Bayesian models offer a principled approach to the quantification of uncertainty (McElreath, 2020). Likewise, deep learning models such as convolutional neural networks (CNNs) have demonstrated promising results for the automated counting of bony fish otolith growth zones. However, the application of Bayesian CNNs for automating the reading of chondrichthyan vertebrae has yet to be explored.

1.1.3 Aims

The overarching goal of this thesis is to advance methodologies for estimating the growth parameters of chondrichthyans via the application of Bayesian models. This thesis makes two specific advancements with respect to Bayesian growth modelling and automated age estimation. Chapter 2 compares monophasic and biphasic curves, and evaluates their goodness of fit for the growth modelling of five species of chondrichthyans. Chapter 3 explores the application of CNNs for the automated reading of vertebral bands, and assesses three deep learning methods in terms of predictive accuracy and model calibration against an image dataset of vertebrae. Finally, Chapter 4 contains an overall summary and discussion concerning the findings of this research.

2 Comparing monophasic and biphasic growth curves for New Zealand chondrichthyans

Abstract

The von Bertalanffy growth model (VBGM) is a common growth function for the modelling of chondrichthyans. However, several theoretical works have suggested that growth may better be described using a biphasic growth model, where growth is comprised of two mechanistic phases induced by changes in resource reallocation for reproduction, diet, or habitat. Here, we compare monophasic and biphasic VBGMs, by fitting them to five chondrichthyan species: *Centrophorus squamosus*; *Isurus oxyrinchus*; *Lamna nasus*; *Mustelus lenticulatus*; and *Prionace glauca*. Data for these five species were collected as incidental catch in New Zealand's surface longline (SLL) fisheries. To account for spatiotemporal variation and incorporate prior knowledge into our growth models, we adopt a Bayesian multi-level modelling approach. The fit of our Bayesian models is assessed using the Pareto-smoothed importance sampling approximation of leave-one-out cross-validation (PSIS-LOO) metric. Overall, our biphasic growth models appear to provide a better goodness of fit to the available data, offering statistical evidence that the growth of chondrichthyans may better be modelled using biphasic growth curves. We conclude that the monophasic VBGM should not be selected *a priori* as the only model to describe the growth of chondrichthyans. Instead, careful consideration of alternative models is recommended, as the VBGM may not always fit best.

2.1 Introduction

The assessment of fish stocks depends on knowledge of biological growth parameters, such as growth rates (Pardo et al., 2013; Trippel & Harvey, 1991) and length-at-maturity estimates (Lappalainen et al., 2016; Trippel & Harvey, 1991). More accurate parameter estimates generally yield more accurate stock assessments, and help inform conservation and fisheries management decisions (Flinn & Midway, 2021b; Pardo et al., 2013; K. Wilson et al., 2018). Moreover, these biological parameters are useful for calculating the maximum intrinsic population growth rates (r_{max}) of chondrichthyans, which is useful in understanding the biological sensitivity of species (S. Chen & Watanabe, 1989; Pardo et al., 2016). Hence, appropriate models of length-at-age relationships that yield robust estimates of growth parameters are key to effective management and long-term sustainability of fisheries.

2.1.1 The von Bertalanffy growth function

By far, the equation most commonly used within fisheries science to model growth patterns and length-at-age relationships is the von Bertalanffy growth model (VBGM), which provides informative parameter estimates such as growth rates and maximum lengths (Cailliet et al., 2006a; Contreras-Reyes et al., 2021; Flinn & Midway, 2021b; McFarlane et al., 2010; Minte-Vera et al., 2016; Moe, 2015; Ricker, 1975).

Von Bertalanffy hypothesised that net growth of an individual is roughly a one-third power function of size, assuming that changes in body weight resulted from differences between anabolic and catabolic processes (Minte-Vera et al., 2016; Pardo et al., 2013). Anabolic processes scale with $\beta = 2/3$ of body mass, while catabolic proccesses scale linearly with body mass (Minte-Vera et al., 2016). These metabolic processes are often modelled using a monophasic growth curve, and more specifically, as a continuous and monotone curve over

all life stages (Contreras-Reyes et al., 2021). These monophasic growth curves are typically straightforward to fit, and are usually applied to cross-sectional or longitudinal size-at-age data from single populations (Contreras-Reyes et al., 2021; K. Wilson et al., 2018).

Despite the popularity of the VBGM, the assumption that a monophasic curve can represent the lifetime growth of organism has received some criticism. Several critics argue that the VBGM should only be fitted to adults of the species (Araya & Cubillos, 2006; Day & Taylor, 1997; Lester et al., 2004; Ricker, 1975; Soriano et al., 1992). The foundation of this argument is that a single curve will fail to capture changes in energy allocation (and growth, as a consequence) that occur upon the onset of maturity, and may often result in a poor fit for the smallest and largest of sharks (Moe, 2015; Quince et al., 2008). Furthermore, fitting a monophasic VBGM across all life phases can also bias parameter estimates (Ricker, 1975), therefore a biphasic growth model may better describe growth during the juvenile and mature life stages (Araya & Cubillos, 2006; Day & Taylor, 1997; Lester et al., 2004; Moe, 2015; Quince et al., 2008).

2.1.2 Biphasic growth

A number of authors (Day & Taylor, 1997; Lester et al., 2004; Quince et al., 2008) have suggested applying a correction due the inability of the VBGM to account for the costs of reproduction. This correction may come in the form of a biphasic growth curve, where growth patterns prior to the age-at-maturity have a different functional form than those growth patterns following the post-maturation phase. Hence, a biphasic growth curve can be applied to fish that experience two distinct growth periods distinguished by some transition point (Figure 1), typically characterised by different growth rates and patterns.

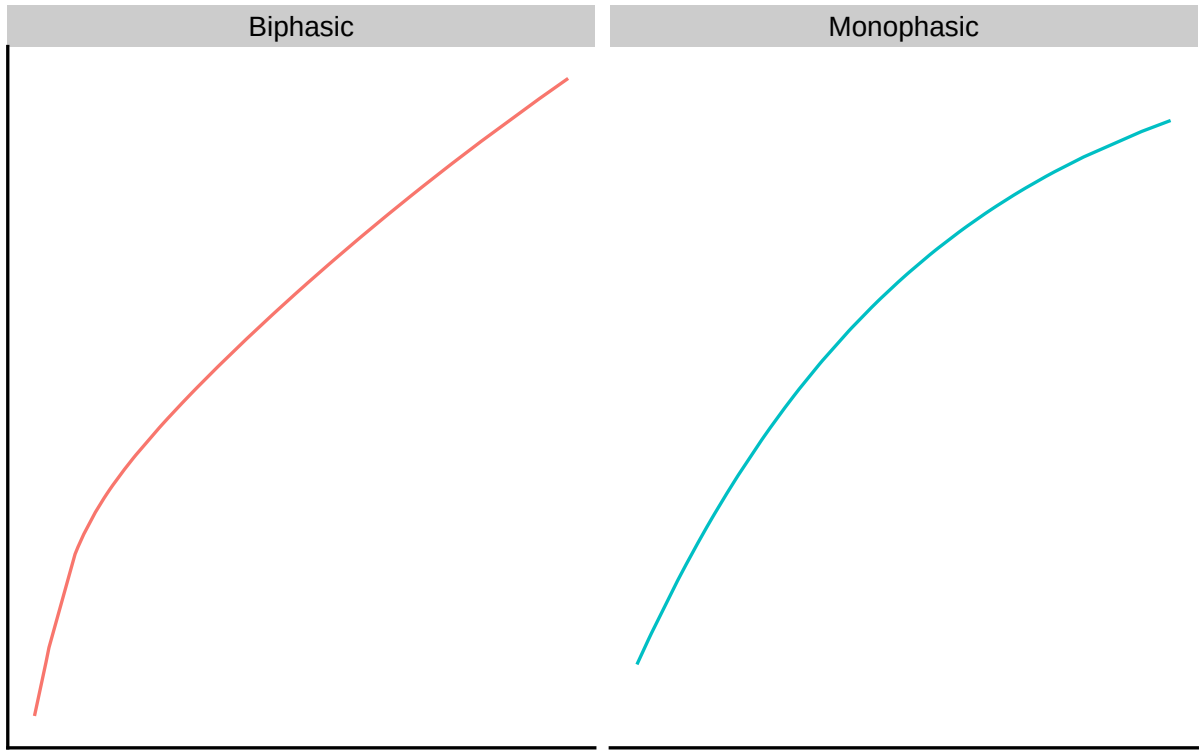


Figure 1: An example of a monophasic and biphasic growth curve.

There are more than twenty identified biphasic growth models which may be categorised by both their mathematical and mechanistic properties (Contreras-Reyes et al., 2021; K. Wilson et al., 2018). Biphasic models may be either continuous or discontinuous at the phasic transition point (Contreras-Reyes et al., 2021; Minte-Vera et al., 2016; K. Wilson et al., 2018). Several biological processes have been theorised to account for this phasic transition, including the reallocation of resources from growth into reproduction (Craig et al., 1998; Minte-Vera et al., 2016; Ohnishi et al., 2012); changes in growth with age (Alós et al., 2010; Condrey et al., 1988; Hoese et al., 1991); changes in diet (Soriano et al., 1992); and habitat shift (Laslett et al., 2002; Porch et al., 2002; Tracey & Lyle, 2005).

In comparison to teleostei, reproductive investment is relatively high for most chondrichthyans, so any transition following the age-at-maturity is more likely to be identified

from growth and age data alone (Contreras-Reyes et al., 2021; Frisk et al., 2001). Moreover, female chondrichthyans devote a significant amount of energy to reproduction, utilising a variety of energy investment tactics to support their offspring during their growth (from egg laying to the formation of a pseudoplacenta) (Cotton et al., 2015; Finotto et al., 2023). As chondrichthyans generally do not provide maternal care after birth, a female's sole chance to contribute resources directly to her offspring takes place during the formation of eggs or throughout pregnancy, with juvenile sharks depending significantly on the maternal resources they acquire in utero (Finotto et al., 2023).

Soriano et al.'s (1992) biphasic extension to the VBGM, known as the biphasic VBGM (BVBGM), has been applied to a number of chondrichthyes in several growth studies. Acuna et al. (2001) made use of the BVBGM to model the growth of three pelagic sharks: *Lamna nasus*, *Isurus oxyrinchus*, and *Prionace glauca*. A multi-model approach composed of 13 candidate models was adopted to describe the growth of the coastal shark, *Sphyrna tiburo*, from which the BVBGM was found to have the best fit (Olmeda-de la Fuente et al., 2022). Similarly, the BVBGM was found to better fit length-at-age data for the deepwater skate, *Dipturus chilensis*, than that of the common VBGM (Aversa, Dans, Garcia, et al., 2011). In an extensive evaluation, Araya and Cubillos (2006) successfully applied the BVBGM to 16 species of sharks and rays, finding statistical evidence in favour of biphasic growth for the majority of these species, particularly when separate models were fitted to males and females.

2.1.3 Changes in growth rates

Several hypotheses have been put forward to explain the decrease in growth rates that occur near the age at first sexual maturity, and the consequent changes for the parame-

ter estimates of the VBGM. Length-at-age selective fishing mortality (Harry et al., 2022; Walker et al., 1998; Walker, 1998); selective sampling bias (Contreras-Reyes et al., 2021; Harry et al., 2022; Walker et al., 1998; Walker, 1998); misestimation of age (Harry, 2018; Walker et al., 1998); and resource reallocation for reproduction (Harry, 2018; Harry et al., 2013; Natanson, 1993) have all been identified as potential factors. The theory put forth by Soriano et al. (1992) was that transitions in growth were related to changes in diet, and their proposed BVBGMs were intended to explicitly model the different growth patterns of juveniles and adults in response to these changes.

This biphasic phenomenon has been observed in a number of species. For instance, Natanson et al. (2002) identified a decrease in growth rates for male and female *Lamna nasus* within the Northeast Atlantic Ocean after reaching sexual maturity, though Araya and Cubillos (2006) noted that only the monophasic VBGM was fitted when a BVBGM may have been more appropriate. Skomal and Calliet (2003) similarly identified a biphasic growth pattern for *Prionace glauca* that coincided with the age-at-maturity for females, but argued that this result may have occurred due to a smaller sample size of female specimens. Finally, Natanson and Cailliet (1990) identified a change in growth for *Squatina californica* that occurred at the length-at-maturity for the species, though they did not fit a VBGM either (Araya & Cubillos, 2006).

In contrast with the VBGM, the BVBGM is able to account for any decreases in growth rates that occur at different life stages (for example, the age-at-maturity). Moreover, while several examples are present within the literature where the BVBGM was fitted to pelagic populations located in the Atlantic Ocean (Acuña et al., 2001; Araya & Cubillos, 2006; Contreras-Reyes et al., 2021), the BVBGM has more rarely been applied to populations located in the Pacific (Aversa, Dans, García, et al., 2011; Olmeda-de la Fuente et al., 2022).

2.1.4 Aims

The aim of this study is to explore the potential of biphasic growth models for New Zealand chondrichthyes by comparing the performance of both biphasic and monophasic models. We first fitted two monophasic and two biphasic von Bertallanffy growth models to males and females of five species of Chondrichthyes (*Centrophorus squamosus*, *Isurus oxyrinchus*, *Lamna nasus*, *Mustelus lenticulatus*, and *Prionace glauca*), and subsequently compared model accuracy using modern Bayesian multilevel modelling approaches. We finally discuss the implications of our results for the fisheries management of these species.

2.2 Theory and review of literature

2.2.1 Bayesian modelling

A robust approach to modelling growth is to numerically approximate parameter distributions using a combination of Bayesian inference and Markov-Chain Monte Carlo (MCMC) algorithms (Smart & Grammer, 2021; K. Wilson et al., 2018). Moreover, Bayesian estimation permits us to incorporate prior information from domain experts into our models (Gelman & Hill, 2006; McElreath, 2020). Bayesian estimation for biphasic growth modelling is particularly promising, as prior information regarding potential contributors to phasic shifts, such as changes in diet, habitat, or reallocation of resources for reproductive investment, can be incorporated into these models (Contreras-Reyes et al., 2021).

2.2.1.1 Bayesian model assessment

Historically, in the field of fisheries science, a single growth model was often applied to data on body size at various ages. Nevertheless, numerous model types have been suggested and assessed for the purpose of estimating growth, and they are most commonly asymptotic,

sigmoidal, or biphasic in shape (Flinn & Midway, 2021b). However, employing an incorrect growth model has been demonstrated to be problematic (Flinn & Midway, 2021a; Minte-Vera et al., 2016; Pardo et al., 2013). For instance, a sub-optimal model may produce sub-optimal parameter estimates that will carry through to any subsequent modelling, such as for stock assessment modelling (Flinn & Midway, 2021a).

The uncertainty surrounding how well different growth models represent the growth patterns of various species has prompted the development of innovative approaches, models, and procedures for model selection. Fisheries researchers increasingly apply a variety of growth models to size-at-age data and use information-theoretic (IT) methods to select the best model (Flinn & Midway, 2021a; Katsanevakis, 2006; Katsanevakis & Maravelias, 2008; Smart & Grammer, 2021).

Bayesian estimation is an alternative to IT approaches for selecting among multiple growth models (Flinn & Midway, 2021a; Katsanevakis, 2006; Katsanevakis & Maravelias, 2008; Smart & Grammer, 2021). Suppose we have dataset y composed of independent data $y_1, ..., y_n$ conditioned upon our parameters θ; hence $p(y|\theta) = \prod_{i=1}^{n} p(y_i|\theta)$. Consider also the prior distribution $p(\theta)$, and posterior distribution $p(\tilde{y}_i|y) = \int p(\tilde{y}_i|\theta)p(\theta|y)d\theta$. We then define the expected log pointwise predictive density (ELPD) as a measure of predictive accuracy given n observations (McElreath, 2020; Vehtari et al., 2017):

$$ELPD = \sum_{i=1}^{n} \int p_t(\tilde{y}_i) \log p(\tilde{y}_i|y) d\tilde{y}_i \tag{1}$$

where $p_t(\tilde{y}_i)$ is the probability distribution for the true data-generating process of $\tilde{y}_i$. $p_t(\tilde{y}_i)$ is unknown and approximated using leave-one-out (LOO) cross-validation. While exact cross-validation requires us to refit the model to multiple training sets, LOO can instead be

approximated using importance sampling to produce importance weights (Gelfand, 1996). These importance weights may be regularised using Pareto smoothed importance sampling (PSIS) (Doll & Jacquemin, 2019; Vehtari et al., 2015, 2017).

The assessment and comparison of models can be performed using the Pareto-smoothed importance sampling approximation of leave-one-out cross-validation (PSIS-LOO) estimate for the ELPD (Vehtari et al., 2015, 2017):

$$\hat{ELPD}_{psis-loo} = \sum_{i=1}^{N} log(\frac{\Sigma_{s=1}^{s} w_i^s p(y_i|\theta^s)}{\Sigma_{s=1}^{s} w_i^s}) \tag{2}$$

where w_i^s for each of i is a vector of importance weights for $s = 1, ..., S$ posterior simulations; and θ^s are the posterior simulations for the parameters. The PSIS-LOO are used to estimate pointwise out-of-sample prediction accuracy by evaluating the ELPD via posterior simulations of parameter values (Doll & Jacquemin, 2019; Gelman & Hill, 2006; McElreath, 2020; Vehtari et al., 2015, 2017).

2.2.1.2 Bayesian multilevel models

Bayesian multilevel methods have seen practical use for modelling the life history traits of chondrichthyes (Dell'Apa et al., 2018; Jiao et al., 2009, 2011; MacNeil et al., 2020; Takahashi et al., 2017), while recent advances in Bayesian multilevel modelling offer ways of incorporating prior information into fish growth models (Zhou et al., 2020; Zhu et al., 2016).

For multilevel models, data from all units within a group are used to inform the individual parameter estimates of each unit (Gelman & Hill, 2006; Vincenzi et al., 2020) through a process known as partial pooling (Gelman & Hill, 2006; McElreath, 2020). Partial pooling

enables data-poor groups to "borrow strength" from those groups which are most similar, which is advantageous when sample sizes are small or data are sparse (Gelman & Hill, 2006; McElreath, 2020).

2.2.1.3 Bayesian imputation

The Bayesian imputation method was utilised to account for missing age estimates. Within the Bayesian inference framework, missing data are represented as parameter estimates of the posterior (Gelman et al., 2013). Assuming that we are utilising a latent variable approach to impute missing data, we can use the following expression:

$$x_i \sim \mathcal{N}(\mu_i^\star, \sigma_i^{\star 2}) \tag{3}$$

where x_i are the unobserved values, which we treat as Gaussian distributed latent variables with some mean $\mu_i^\star$ and variance $\sigma_i^{\star 2}$. With this approach, imputed values are then drawn from the posterior.

2.2.2 Chondrichthyan species

2.2.2.1 Life history traits

2.2.2.1.1 *Centrophorus squamosus*

Centrophorus squamosus is a medium-sized deepwater species found in the Atlantic, Pacific, and Indian Oceans (Finucci et al., 2020). It is caught as both target and bycatch within deepwater fisheries using various fishing gear, and as a consequence, overfishing and habitat degradation have caused population declines in several regions (Ebert et al., 2013; Finucci et al., 2020; Parker & Francis, 2012a). The species is listed as Globally Endangered by the

International Union for Conservation of Nature (IUCN), as it is estimated to have undergone population declines in the Northeast Atlantic and Indo-Pacific, with some population increase in the Southwest Pacific (Ebert et al., 2013; Finucci et al., 2020).

Centrophorus squamosus can grow up to 166 cm total length (TL); the length-at-maturity for males is roughly 100 - 110 cm, while for females it is 110 - 125 cm TL (Ebert et al., 2013; Finucci et al., 2020; Weigmann, 2016). Reproduction is aplacental viviparous with relatively small litter sizes of 5 - 8 pups and a size-at-birth of 30 - 40 cm TL (Ebert et al., 2013; Finucci et al., 2020). Note also that there is regional variation in growth parameters - within Atlantic populations, the female age-at-maturity is estimated to be 35 years and the maximum age is 70 years (Clarke et al., 2002). Conversely, in Pacific populations, the female age-at-maturity is estimated to be 20.8 years while the maximum age is 42 (Parker & Francis, 2012a).

2.2.2.1.2 *Isurus oxyrinchus*

Isurus oxyrinchus is a large, pelagic species found in the Atlantic, Indian, and Pacific Oceans (Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Jabado, et al., 2019), particularly within tropical and warm-temperate waters at depths to 888 metres (Abascal et al., 2011; Ebert et al., 2013; Weigmann, 2016). Length-at-maturity is estimated to be between 166 - 204 cm TL for males and 265 - 312 cm TL for females (Bishop et al., 2006; Pratt Jr & Casey, 1983; Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Jabado, et al., 2019; Varghese et al., 2017). The length at birth is estimated to be 60 - 70 cm TL, and litter sizes are between 4 - 25 pups (Compagno, 2001; Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Jabado, et al., 2019; Weigmann, 2016). The IUCN has listed the shortfin mako shark as an Endangered species, and a number of countries have implemented regulations

to manage the species (Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Jabado, et al., 2019).

2.2.2.1.3 *Lamna nasus*

Lamna nasus is a large (up to 357 cm TL) pelagic shark, occurring at depths of 1,809 m within the North Atlantic and Southern Hemisphere at temperate and cold-temperate waters (Ebert et al., 2013; Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019a; Weigmann, 2016). The species is not abundant within the North Pacific, where their congener *Lamna ditropis* is instead present (Francis et al., 2008). Notably, there are separate populations within the North Atlantic and Southern Hemisphere (Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019a; Weigmann, 2016). Due to overfishing, steep population declines have occurred within the North Atlantic, with fewer declines for populations within the Southern Hemisphere (Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019a). This has resulted in the species being classified as Vulnerable by the IUCN Red List (Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019a).

Lamna nasus has been noted to have low fecundity, with small litter sizes of 1 - 5 pups and a late age-at-maturity (Ebert et al., 2013; Francis et al., 2008; Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019a; Weigmann, 2016). It may reach a maximum size of 357 cm TL, although maximum size is most probably larger for North Atlantic populations, as *Lamna nasus* greater than 230 cm TL are rare within the Southwest Pacific (Francis et al., 2008; Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019a).

Females mature at roughly 170 - 195 cm TL in the Southern Hemisphere, while males

mature at around 140 - 165 cm TL (Francis et al., 2008; Francis & Duffy, 2005). Age estimates have been validated using radiocarbon assays (Francis et al., 2007a), and the age underestimation of older specimens (using vertebral band count methods) was noted to be a significant concern (Francis et al., 2007a; Harry, 2018). The female age at maturity is 16.5 years, with a maximum age of 60 years (Francis et al., 2008; Francis, 2015b).

2.2.2.1.4 *Mustelus lenticulatus*

Mustelus lenticulatus is a small shark endemic to New Zealand, Stewart Island, the Chatham Islands, and within the Snares Islands (Francis & Francis, 1992; Francis & Maolagáin, 2000). Declining catch rates and stock sizes indicated that the species was being overfished (Francis, 1988). In 1986, the Quota Management System (QMS) was introduced and commercial landing saw a reduction by more than one half (Francis & Francis, 1992).

The species uses estuaries and inshore localities as nurseries for newborns where water clarity is typically low (Francis, 2013), which offers some protection against visual predators (Francis et al., 2012). *Mustelus lenticulatus* nurseries can be found in the southernmost regions of the North Island, within both Wellington Harbour and Porirua Harbour (Francis et al., 2012). Predators are rare in Porirua Harbour, and the area is rich in food resources, providing a safe location for juveniles (Francis et al., 2012; Francis, 2013). A spring-summer inshore migration of *Mustelus lenticulatus* towards these nurseries has been reported, with a substantial skew towards males, leading to male-biased populations and evidence of multiple paternity (Boomer et al., 2013). The species is ovoviviparous, with a gestation period of roughly 11 months (Francis & Mace, 1980). Fecundity is known to increase with the body length of female sharks, which mature at a TL of 106 cm (Massey & Francis, 1989).

Conversely, males mature at around 89 cm TL (Massey & Francis, 1989). The maximum observed lengths of females is 137 cm, while for males it is 115 cm (Francis & Mace, 1980).

2.2.2.1.5 *Prionace glauca*

Prionace glauca is a large, pelagic shark found up to 1000 m deep throughout all oceans in tropical and temperate waters (Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019b ; Ebert et al., 2013; Last & Stevens, 2009; Weigmann, 2016). It has the fastest known growth rate among all pelagic sharks, with high fecundity and pups that mature young (Ebert et al., 2013; Last & Stevens, 2009; Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019b; Weigmann, 2016). Population declines have occurred within the Atlantic and Indian Oceans, and while populations have increased in the Pacific, the species has still been classified as Near Threatened by the IUCN (Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019b).

The species occurs near inshore regions where the continental shelf is most narrow (Last & Stevens, 2009), and expresses migratory movement patterns consistent with reproduction and prey distributions (Nakano & Stevens, 2008; Rigby, Barreto, Carlson, Fernando, Fordham, Francis, Herman, et al., 2019b). Male sharks mature at 183 - 218 cm TL, while females mature at 183 - 221 cm TL (Carrera-Fernández et al., 2010; Nakano, 1994; Pratt Jr, 1979). The median length at maturity for male, New Zealand *Prionace glauca* was estimated to be 190 - 195 cm (Francis & Duffy, 2005). Clasper lengths (the means to determine maturity in males) for this species continuously increase across the majority of the length range, only reaching an asymptote for larger sharks, so it is challenging to determine the maturity of the shark from clasper lengths alone (Francis & Duffy, 2005; Pratt Jr, 1979). According to Francis and Maolagáin (2016), the 190 − 195 cm size at maturity interval

from Francis and Duffy (2005) remains the best available estimate to date.

2.3 Methods

2.3.1 Monophasic growth models

The original 3-parameter VBGM can be expressed as:

$$L(t) = L_\infty - (L_\infty - L_0)e^{-kt} \tag{4}$$

where $L(t)$ is the length as a function of time; L_∞ is the asymptotic length-at-age (cm); k is the Brody growth coefficient (per year) determining the rate at which the fish will reach the asymptotic length-at-age; and L_0 is the size at birth (Von Bertalanffy, 1938, 1960).

The 3-parameter VBGM that is most commonly used within fisheries science can be expressed with the familiar equation (Beverton, 1957):

$$L(t) = L_\infty(1 - e^{-k(t-t_0)}) \tag{5}$$

where t_0 is the theoretical time at length zero, and the other parameters are as previously defined (Cailliet et al., 2006a).

2.3.2 Biphasic growth models

Soriano et al. (1992) proposed two continuous, biphasic extensions to the common VBGM to account for potential dietary shifts that may occur when juveniles mature into adults. For each of the BVBGMs, a new factor $A(t)$ is proposed:

$$A(t) = 1 - \frac{h}{1 + (t - t_h)^2} \tag{6}$$

where t_h is the age in which the phasic shift occurs; and h is the magnitude of the maximum difference in length-at-age between the common VBGM and the BVBGM (Araya & Cubillos, 2006).

Here, we make use of two parameterisations of Soriano et al.'s (1992) BVBGMs. The first of which is:

$$L(t) = L_\infty A(t)(1 - e^{-k(t-t_0)}) \tag{7}$$

where $A(t)$ modifies L_∞ as age increases.

The second parameterisation of Soriano et al.'s (1992) BVBGM is:

$$L(t) = L_\infty (1 - e^{-kA(t)(t-t_0)}) \tag{8}$$

where $A(t)$ modifies the value of k as age increases.

2.3.3 Species under analysis

Published length-at-age data were obtained from Fisheries New Zealand's fish aging database for five species of shark, belonging to three orders (Table 1). Growth parameters for each species were identified from previous age and growth studies conducted by the National Institute of Water and Atmospheric Research (NIWA), and these parameters were used to inform the priors of our models.

Table 1: Species and data sources

Species	Order	Source
Centrophorus squamosus	Squaliformes	Parker & Francis (2012b)
Isurus oxyrinchus	Lamniformes	Francis (2016)
Lamna nasus	Lamniformes	Francis (2015a)
Mustelus lenticulatus	Carcharhiniformes	Francis & Maolagáin (2000)
Prionace glauca	Carcharhiniformes	Francis & Ó Maolagáin (2016)

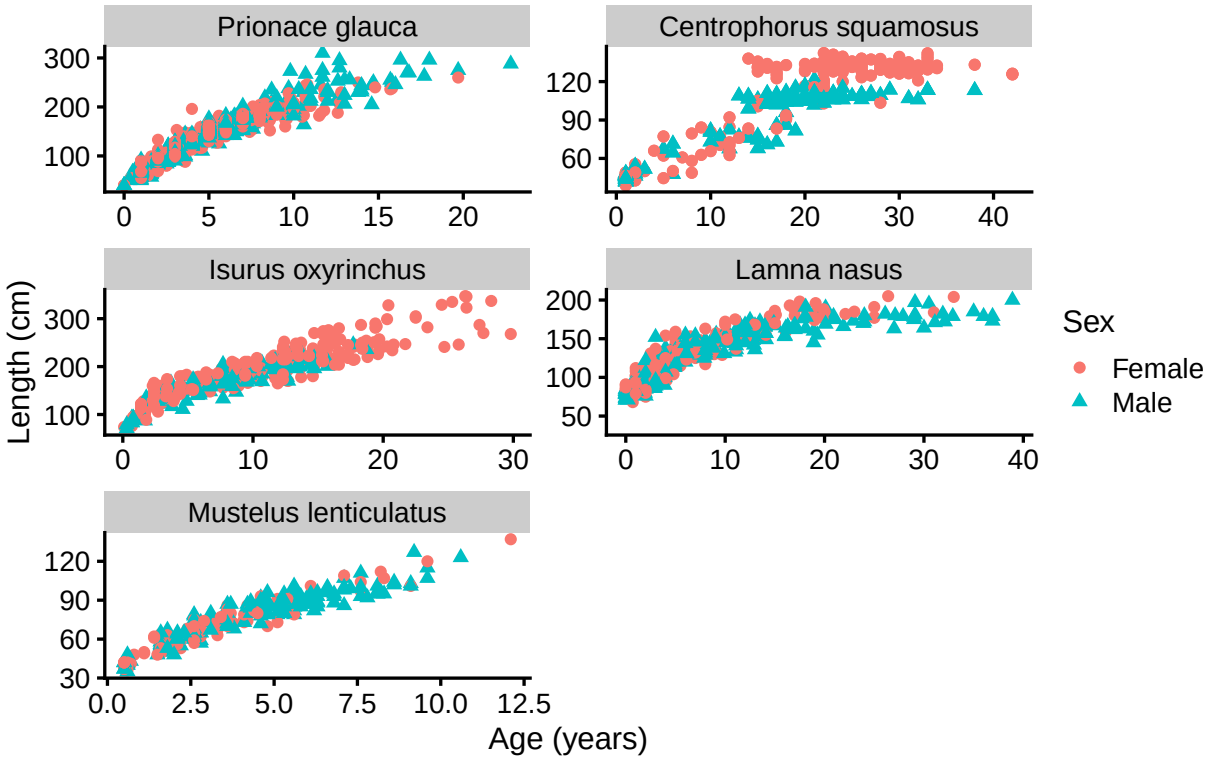


Figure 2: Length-at-age data by species and sex.

We fitted separate growth models by sex for each species to account for differences in the age-of-maturity and growth patterns of both male and female chondrichthyans (Figure 2). A similar approach of fitting separate growth models by sex has been adopted in several

other studies on biphasic growth (Acuña et al., 2001; Araya & Cubillos, 2006; Contreras-Reyes et al., 2021; Mejía-Falla et al., 2014). Note also that the pooling of both sexes can obscure evidence of biphasic growth (Araya & Cubillos, 2006).

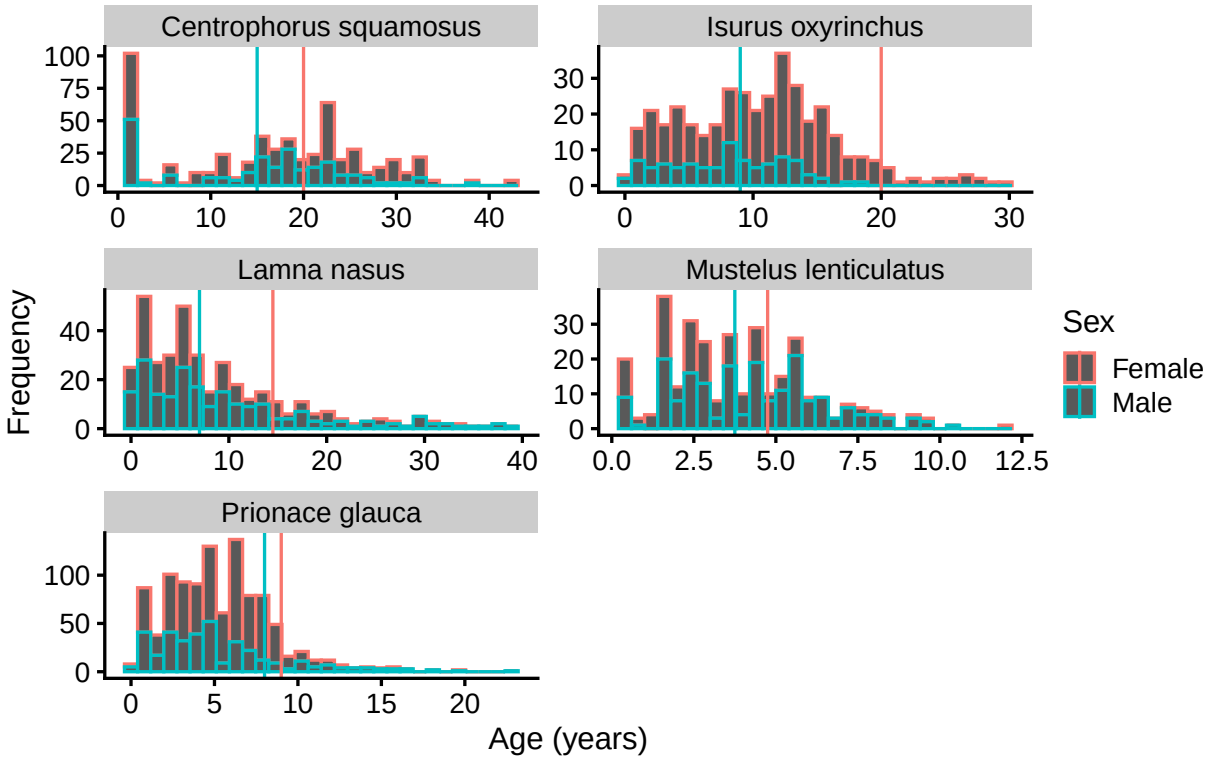


Figure 3: Frequency distribution of ages by species and sex. Vertical lines represent age-at-maturity.

Pelagic sharks, including *Isurus oxyrinchus*, *Lamna nasus*, and *Prionace glauca*, are frequently caught as incidental catch in New Zealand's surface longline (SLL) fisheries and, to a smaller degree, in midwater trawl fisheries (Francis, 2015a, 2016; Francis & Ó Maolagáin, 2016). Samples for these species were dominated by juveniles (Figure 3), and selectivity bias was noted to be an issue in cases where mature adults were discarded without being measured (Francis, 2015a, 2016; Francis & Ó Maolagáin, 2016). Samples of *Mustelus lenticulatus* were collected principally from research trawl surveys in the west coast South

Island (WCSI) and the east coast South Island (ECSI). Samples from the WCSI contained very few rig that were older than 8 years of age, where differences in growth would become more apparent (Francis & Maolagáin, 2000). Aging data for *Centrophorus squamosus* were collected from multiple surveys within Chatham Rise and the Sub-Antarctic plateau; and the age-frequency distribution was composed of an abundance of 1 year old fish, with an absence of juveniles until age 10 (Parker & Francis, 2012b).

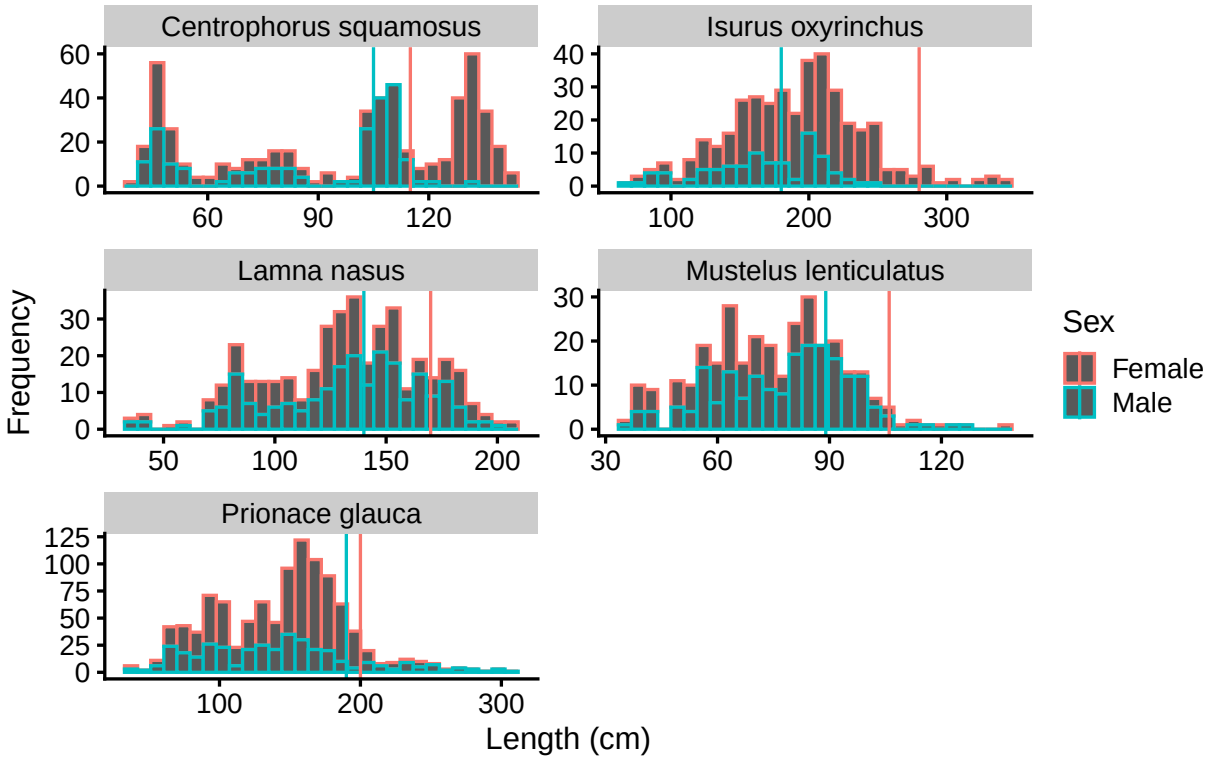


Figure 4: Frequency distribution of lengths by species and sex. Vertical lines represent length-at-maturity.

The length distribution of *Centrophorus squamosus* was trimodal in nature, with a large number of juveniles near a mode of 30 cm, another mode of 110 cm composed of males and females, and a third mode of female sharks near 130 cm (Figure 4). The lengths of male and female *Isurus oxyrinchus* were centred around a mode 200 cm, and females were

noticeably larger than males. The majority of male and female *Lamna nasus* were below 150 cm, with a mode of 140 cm for males and females, meaning that most sampled sharks were immature. The distribution of male *Mustelus lenticulatus* was left-skewed, with a mode length of 80 cm, while females were similarly left-skewed with an additional mode at 65 cm. Finally, the majority of *Prionace glauca* were below 200 cm in length; females had a mode length of 170 cm, while males were slightly smaller with a mode of 160 cm.

It was noted that the accuracy of length measurements in samples of *Isurus oxyrinchus*, *Lamna nasus*, and *Prionace glauca* were affected by the inability of observers to measure every captured shark (Francis, 2015a, 2016; Francis & Ó Maolagáin, 2016). There was also evidence suggesting that the size distribution of unmeasured, discarded sharks could differ from that of measured, discarded sharks, as larger sharks are less likely to be hauled on board (Francis, 2015a, 2016; Francis & Ó Maolagáin, 2016).

2.3.3.1 Growth parameters and datasets

2.3.3.1.1 *Centrophorus squamosus*

Table 2: von Bertalanffy growth parameter estimates of *Centrophorus squamosus* for total length (Parker & Francis, 2012a)

	L_∞	k	t_0	L_0	A_M
$Male$	136.46	0.05	−6.68	$30-45$	15.44
$Female$	192.18	0.03	−6.67	$30-45$	20.82

VBGM parameter estimates of *Centrophorus squamosus* were produced from trawl surveys within Chatham Rise and several regions within the Sub-Antarctic plateau (Auckland,

Snares Shelf, Campbell Plateau, Puysegur), as was the age at maturity (A_M) (Table 2) (Cox & Francis, 1997; Parker & Francis, 2012a). Growth curve fits were noted to be poor due to several factors (Parker & Francis, 2012a):

1. A small sample size.
2. Few juvenile samples between 2 and 15 years old.
3. A disunity between samples below and above 15 years of age.
4. Most medium fish were sampled from Chatham Rise, while both younger and older specimens were sampled from the Sub-Antarctic.
5. The proportion of female *Centrophorus squamosus* was typically greater than males (70% on the Chatham Rise and 60% in the Sub-Antarctic), and it was suggested that this female bias may be due to a lack of juvenile samples, and also due to adult females reaching a greater size than males (Parker & Francis, 2012b).
6. The differences in fishing effort histories between Chatham Rise and the Sub-Antarctic.

Referring to the original study by Parker and Francis (2012a), these factors resulted in a steeper VBGM curve, with a lack of upper asymptote for either of the sexes, leading to underestimated ages for mature specimens and overestimated ages for juveniles. Despite this, the authors noted that the resulting growth curves seemed sensible, and that there was some variation in the growth patterns of the two sexes (Parker & Francis, 2012a).

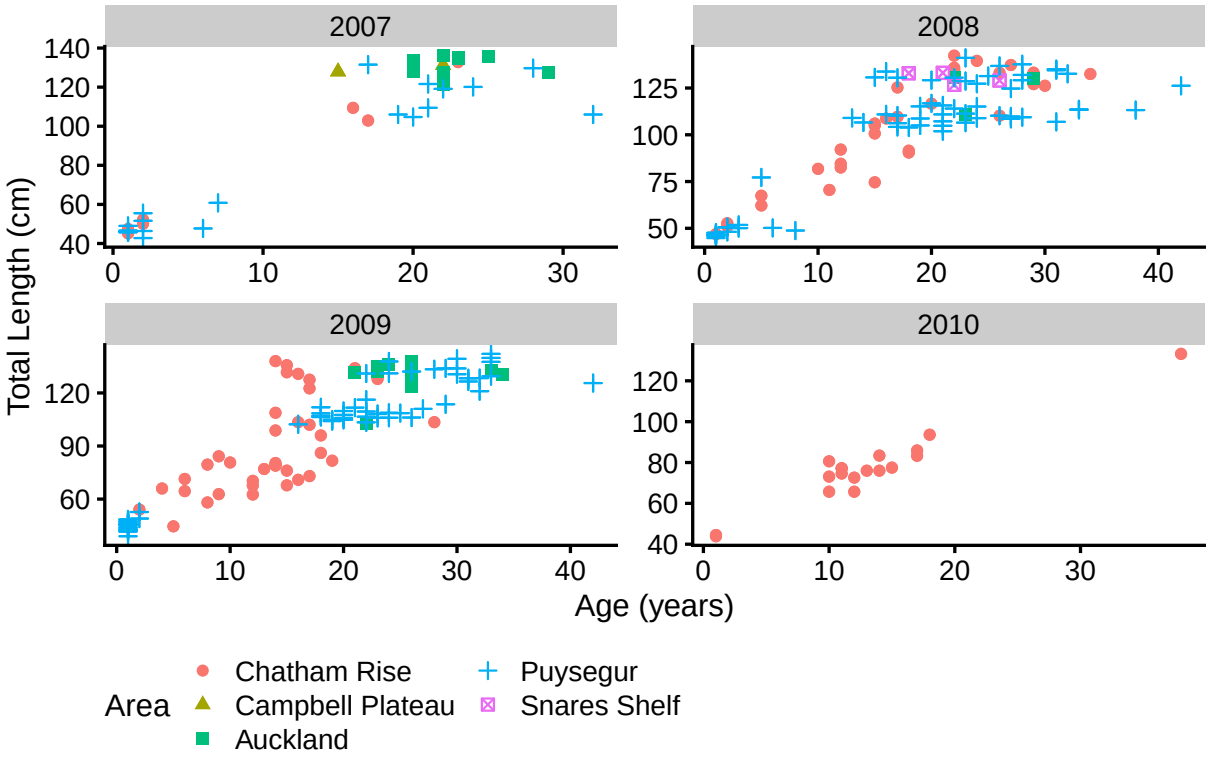


Figure 5: Variation of *Centrophorus squamosus* across locations over time.

Spatiotemporal variation in the distribution of ages and TLs was present within the surveyed *Centrophorus squamosus* dataset (Figure 5). For the 2007 observations, we can see two clusters at the extremes of length and age. In 2008 and 2009, the data were more evenly distributed across different age groups, and more closely resembled the shape of a logistic curve. Only Chatham Rise was observed in 2010.

For this work, a dataset of 309 females and 231 males was provided, spanning over 4 years and 5 locations. There were no missing data, and TL in cm was treated as the standard measurement.

2.3.3.1.2 *Isurus oxyrinchus*

Table 3: von Bertalanffy growth parameter estimates of *Isurus oxyrinchus* for fork length (Bishop et al., 2006)

	L_∞	k	t_0	L_0	A_M	A_J
Male	302.30	0.052	−9.0	61	$7-9$	2
Female	820.10	0.013	−11.30	61	$19-21$	2

The two most informative studies regarding the growth parameters of *Isurus oxyrinchus* within the Pacific are those of Bishop et al. (2006) and Francis (2016). In the case of Bishop et al. (2006), parameter estimates for monophasic VBGMs were noted to be poor for both sexes, particularly when the growth curves were forced to fit through the estimated size at birth of 61 cm fork length (FL) (Table 3) (Bishop et al., 2006). For both studies (Bishop et al., 2006; Francis, 2016), the fitted growth curves of females demonstrated a near-linear growth pattern across age groups and failed to asymptote. Referring to Table 3, the estimate for L_∞ was 820.1 cm FL - this estimate was deemed to be implausible, as it was substantially greater than length measurements within the dataset or estimates from other studies (Cerna & Licandeo, 2009; Natanson et al., 2006; M. Ribot-Carballal et al., 2005; Semba et al., 2009). The maximum observed female age of 347 cm FL from Bishop et al. (2006) was instead identified as a potential prior for the Bayesian growth models.

Francis (2016) noted that males and females had significantly different growth curves, though this difference would only become apparent after roughly 16 years. This is slightly below the estimated A_M of $19-21$ from Bishop et al. (2006). Moreover, males and females grow rapidly during their first 2 years before subsequently slowing, nearly doubling in size from their length at birth (Francis, 2016). We denote this transition from the juvenile life

stage to the maturing stage as A_J (refer to Table 3).

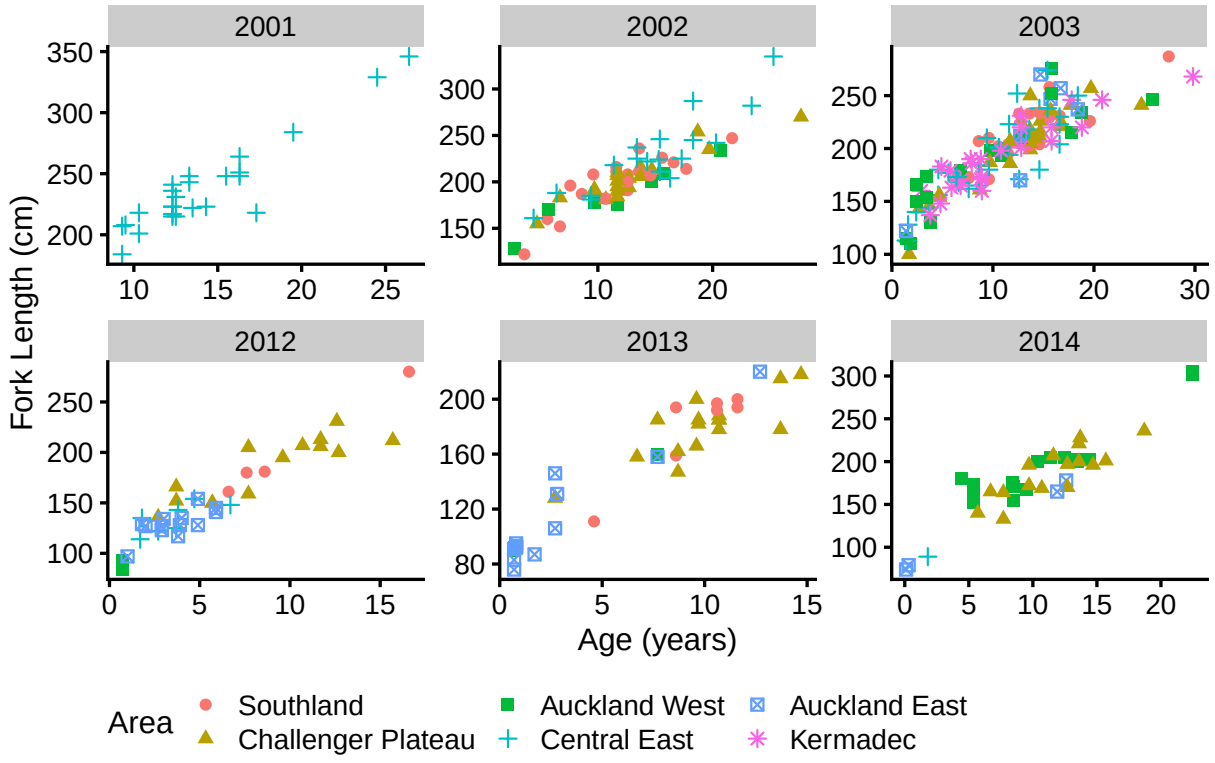


Figure 6: A filtered subsample of *Isurus oxyrinchus* across locations over time (n > 10, for each year).

The data for this study were collected from SLL vessels targeting tunas and swordfish, and the distribution of ages was dominated by juveniles (Figure 6). Samples were taken from waters across the northeast of New Zealand (the Kermadec Islands), the west of New Zealand (Challenger Plateau), the North Island (the east and west coasts of Auckland; and the Central East fisheries region), and the southernmost South Island (Southland). From 1994 - 2001, samples of *Isurus oxyrinchus* were sampled only from Central East; by contrast, there was substantially greater diversity in sampling locations between 2002 - 2014. In particular, observations caught during 2002 and 2003 spanned a wide area and included a greater range of age groups. By contrast, we can see a relative sparsity in

specimens between 2012 - 2014, with a large cluster of juveniles caught during 2012 and 2013. A grouping of maturing sharks (5 - 15 years old) were taken in during 2014.

Growth models were fitted to a dataset of 295 females and 93 males. Samples for females were taken over 15 years and 7 locations, and 7 years and 6 locations for males. There were no missing values. FL in cm was treated as the standard measurement.

2.3.3.1.3 *Lamna nasus*

Table 4: von Bertalanffy growth parameter estimates of *Lamna nasus* for fork length (Francis, 2015b)

	L_∞	k	t_0	L_0	A_M
$Male$	185.77	0.13	−4.22	79.70	$6.3 - 8.2$
$Female$	210.86	0.09	−6.10	86	$13.0 - 16.3$

Francis (2015b) found that male and female *Lamna nasus* had substantially different growth patterns Table 4, though the growth rates of the two sexes were nearly identical up to 10 years of age, which was consistent with the results of previous studies (Francis et al., 2007a; Francis & Duffy, 2005). Estimates for L_0 were substantially greater than the previous reported length of 58–67 cm FL, and this difference may be due to size-selectivity from longline hooks or an absence of new-born juveniles within the sampled locations (Francis, 2015b).

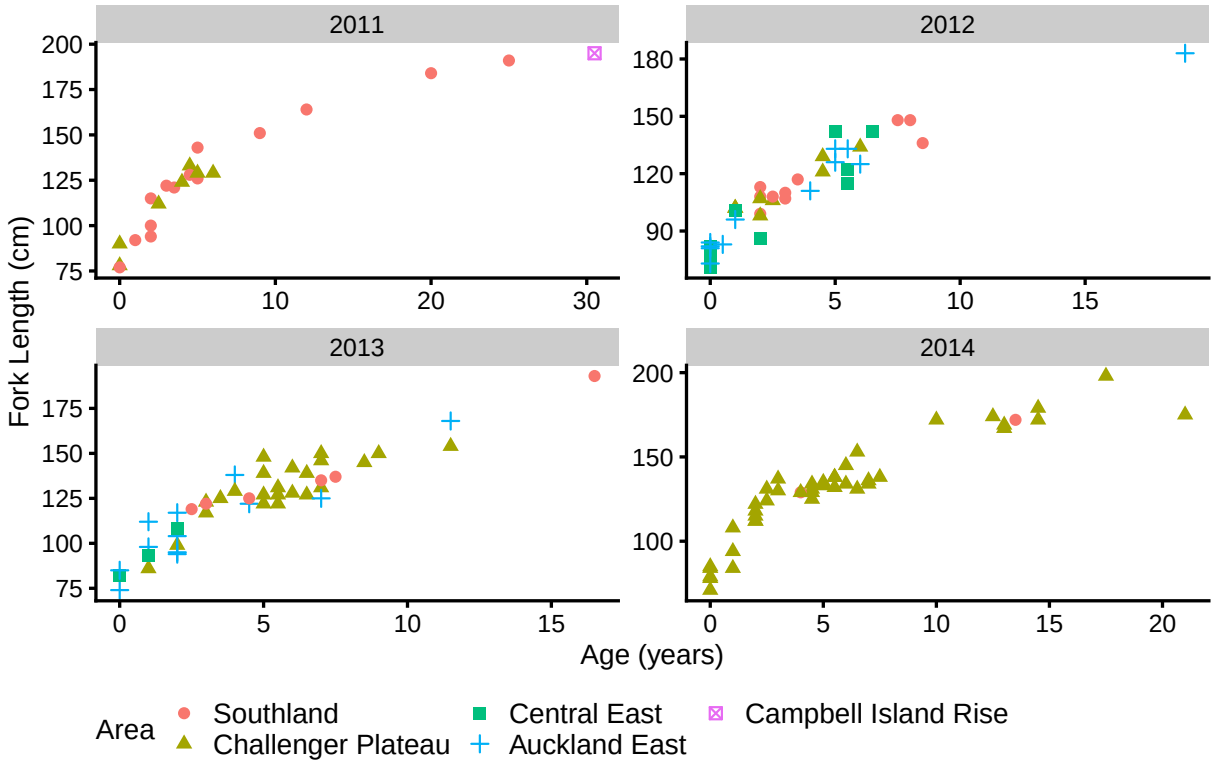


Figure 7: A filtered subsample of *Lamna nasus* across locations over time (n > 10, for each year).

Samples of *Lamna nasus* were collected between 1997 - 2014 from SLL vessels from tuna fisheries (Figure 7). Specimens from between 1993 - 2003 did not contain location data, and the majority of missing age estimates were taken during this time. A large number of juveniles (2 years and younger) and maturing sharks were sampled from Southland and the Central Plateau during 2011, with few older sharks and the very oldest taken from the southern waters of New Zealand (Campbell Island Rise). In 2012, the majority of caught *Lamna nasus* were under 10 years, with the majority of juveniles being from Central East and a single mature shark from Auckland East. In 2013, most juveniles were sampled from Auckland East, with the majority of maturing sharks taken from the Central Plateau. In 2014, the vast majority of *Lamna nasus* were sampled from Central Plateau, with a single

shark from Campbell Island Rise.

A dataset of 177 females and 223 males was made available. Data spanned 12 years and 6 locations. Roughly 6% of the age measurements were missing, and 37% of the sampled areas were not recorded. All length data were measured using FL in cm.

2.3.3.1.4 *Mustelus lenticulatus*

Table 5: von Bertalanffy growth parameter estimates of *Mustelus lenticulatus* for total length (Francis & Francis, 1992)

	L_∞	k	t_0	L_0	A_M
$Male$	118.70	0.16	−2.02	34.2	3.7
$Female$	90.7	0.42	−0.77	25.3	4.7

Francis and Francis (1992) stated that their growth parameter estimates (Table 5) were likely biased towards smaller and younger *Mustelus lenticulatus*, as larger sharks are able to avoid trawl nets. This may mean that L_∞ is underestimated while k is overestimated. Likewise, the authors noted that female sharks were undersampled, which may explain why the L_∞ estimate is smaller for females than males (Francis & Francis, 1992). For most species of *Mustelus*, females typically live longer and grow larger than males (Francis & Francis, 1992; Francis & Mace, 1980), though these differences in growth patterns generally only become pronounced after eight years of age (Francis & Francis, 1992).

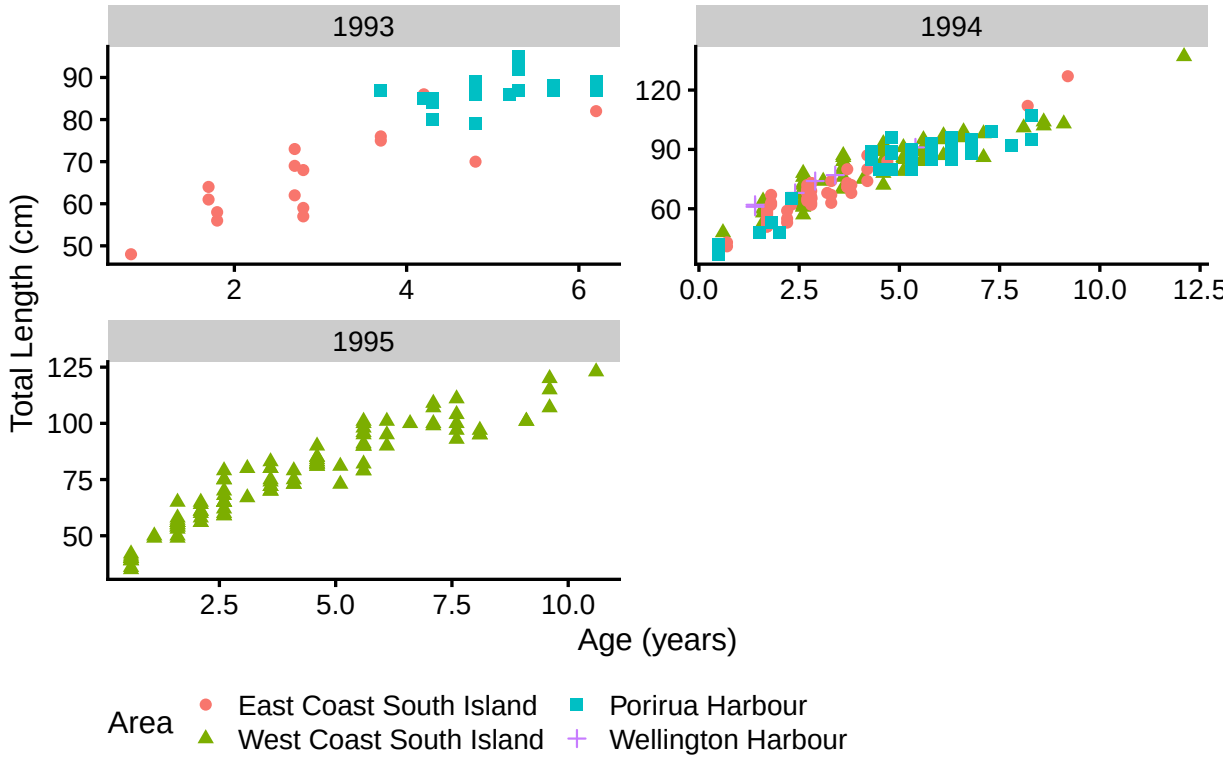


Figure 8: Variation in *Mustelus lenticulatus* across locations over time.

Data for *Mustelus lenticulatus* were gathered between 1993 - 1995 (Figure 8). In 1993, the majority of younger juveniles were taken from the East Coast South Island, while the majority of older juveniles were sampled from Porirua Harbour. Data from East Coast South Island made up the majority of under 5 year olds in 1994, while 5-10 year olds were sampled from Porirua Harbour and West Coast South Island, with a single 12 year old *Mustelus lenticulatus*. In 1995, data from the West Coast South Island spanned across a variety of age groups.

Growth curves were fitted to a *Mustelus lenticulatus* dataset of 116 females and 194 males across 3 years and 4 areas. TL in cm was treated as the standard measurement of length. There were no missing values present.

2.3.3.1.5 *Prionace glauca*

Table 6: von Bertalanffy growth parameter estimates of *Prionace glauca* for fork length (Manning & Francis, 2005)

	L_∞	k	t_0	L_0	A_M
$Male$	342.90	0.0878	−1.2570	54	8
$Female$	267.49	0.1259	−1.0467	54	$7-9$

Growth parameter estimates for *Prionace glauca* were produced from data collected across New Zealand's Exclusive Economic Zone (EEZ) Table 6 (2005). Females appear to have a faster growth rate and lower mean asymptotic length than males (Manning & Francis, 2005), contradicting overseas studies concerning Atlantic populations of *Prionace glauca* (Skomal & Natanson, 2003). Manning and Francis (2005) suggest that samples for larger females were missing from the data potentially due to selectivity bias from SLL observers, likely resulting in biased growth estimates for females. However, the authors noted that growth rates and longevity were broadly comparable with overseas studies (Manning & Francis, 2005). Referring to A_M estimates, Manning and Francis (2005) stated that *Prionace glauca* in the New Zealand EEZ seemed to have a later maturity than sharks from the Atlantic and Indian Oceans, though samples taken from the EEZ were most probably part of a larger South Pacific population.

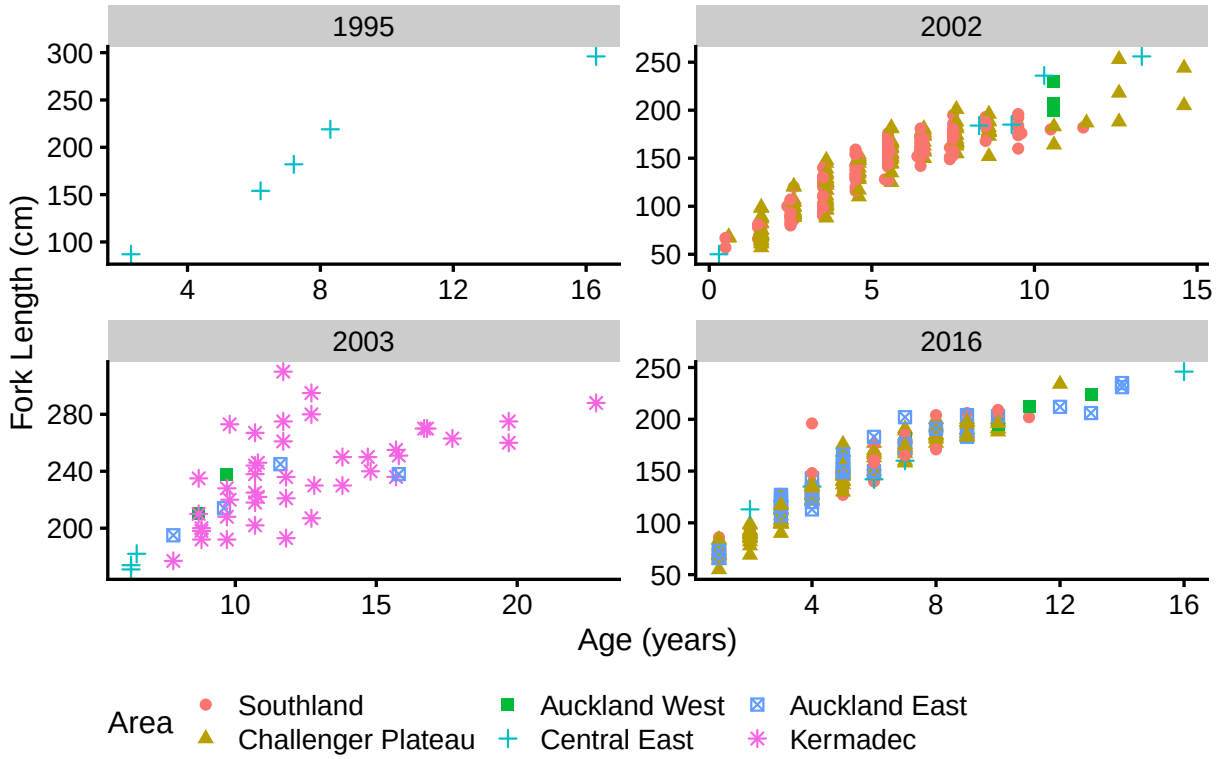


Figure 9: A filtered subsample of *Prionace glauca* across locations over time (n > 4, for each year).

Prionace glauca were collected between 1994-2016 (Figure 9). Between 1994-1999, observations were taken only from the Central East. In 2002, the vast majority of sharks were caught from the Central Plateau and Southland, the majority of which were under 10 years of age. In 2003, specimens were caught from the Kermadec Islands, with a greater proportion of older and larger *Prionace glauca* and some relatively larger, maturing sharks (roughly 10-13 years of age). In 2016, samples were observed across a greater range of locations and age groups, though the majority were still under 15 years old.

For the present study, a relatively large dataset of 690 females and 362 males was made available. All body length measurements were standardised to FL in cm. The data spanned across 8 years and 7 locations, with 5 missing FL measurements, and 37% of observations

lacking a recorded area.

2.3.4 Bayesian Model parameterisation

All Bayesian models were fitted using *Stan*, a probabilistic programming language for statistical modelling and Bayesian inference (Carpenter et al., 2017). All analyses were performed in the R statistical computing language (R Core Team, 2022). The *brms* package (Bürkner, 2017) was used to specify the models. Here, we have adopted the Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS) algorithms to perform Bayesian inference (Hoffman et al., 2014).

For each model, we sampled 6000 draws from the posterior using 4 parallelised Markov chains, with 1500 iterations during the warm-up phase. The implementation of PSIS-LOO from the *loo* package was used for Bayesian model assessment (Vehtari et al., 2017). *brms* was utilised to impute missing data using a latent variable approach (Bürkner, 2017; Carpenter et al., 2017).

2.3.4.1 Common parameterisation

For each of our four candidate models, we assumed that length followed a Gaussian distribution, with relatively diffuse student's t priors for our scale parameter σ. Here, we use the common VBGM as an example:

$$\begin{aligned} L(t) &\sim \mathcal{N}(\mu, \sigma) \\ \mu &= L_{\infty}(1 - e^{-k(t-t_0)}) \\ \sigma &\sim \mathcal{T}_{+}(0, 50, \nu = 3) \end{aligned} \tag{9}$$

The other three candidate models had a comparable parameterisation. The linear components of our multilevel models (the growth parameters of interest) were also assumed to be Gaussian, and we specifically used a non-centred parameterisation for each linear component. Our resulting growth models can be divided into two categories: (1) Multilevel models for data without missing cases; and (2) Imputation models for data with missing cases.

2.3.4.2 Multilevel models

For *Centrophorus squamosus*, *Isurus oxyrinchus*, and *Mustelus lenticulatus*, we assigned varying intercepts by both year and area for our six growth parameters $\theta_G = \{L_\infty, k, t_0, L_0, h, t_h\}$. Using L_∞ as an example:

$$\begin{aligned} L[i]_\infty &= \bar{\alpha} + z[i]\sigma_\alpha + x[i]\sigma_\gamma \\ z_j &\sim \mathcal{N}(0,1) \quad \text{for } j = 1, ..., J \text{ years} \\ \sigma_\alpha &\sim \mathcal{T}_+(0, 50, \nu = 3) \\ x_k &\sim \mathcal{N}(0,1) \quad \text{for } k = 1, ..., K \text{ areas} \\ \sigma_\gamma &\sim \mathcal{T}_+(0, 50, \nu = 3) \end{aligned} \tag{10}$$

where z is the vector of standardised intercepts for each year, x is the vector of standardised intercepts for each area, σ_α is the standard deviation that determines the degree of pooling across years, and σ_γ is the standard deviation that determines the degree of pooling across areas. Additionally, we specified population-level priors for our intercept term $\bar{\alpha}$ (see Section 2.3.4.4). Varying intercepts were also provided for the other five growth parameters k, t_0, L_0, h, t_h using the same formulation as L_∞ above.

2.3.4.3 Imputation models

The data for *Lamna nasus* and *Prionace glauca* contained missing values for both age and area. Missing age measurements were treated as parameters of the model, and were imputed using the approach describe in Section 2.2.1.3. The form of the common VBGM with missing age measurements is:

$$\begin{aligned} L[i](t) &\sim \mathcal{N}(\mu_i, \sigma) \\ \mu_i &= L_\infty(1 - e^{-k(t[i]_{true} - t_0)}) \\ \sigma &\sim \mathcal{T}_+(0, 50, \nu = 3) \\ t[i]_{true} &\sim \mathcal{N}(\mu[i]_t, \sigma[i]_t) \\ \mu[i]_t &\sim \mathcal{T}(0, 5, \nu = 3) \\ \sigma[i]_t &\sim \mathcal{T}_+(0, 5, \nu = 3) \end{aligned} \tag{11}$$

where $t[i]_{true}$ is the true age for each observation indexed by i, which is modelled as a Gaussian distributed latent variable with mean $\mu[i]_t$ and standard deviation $\sigma[i]_t$. As with every other parameter of the model, the distribution of each latent variable is jointly learnt using Bayesian inference.

Unlike the models described in Section 2.3.4.2, we were not able to incorporate varying intercepts by area, as we could not impute any missing areas during the model fitting process. We would need to treat any missing areas as discrete parameters of the model, and the gradient-based HMC algorithm is not able to sample from discontinuous densities (Betancourt, 2017).

2.3.4.4 Prior specification

We specified population-level priors for the intercept terms of our growth parameters (see Table 7), each of which was informed by the studies discussed in Section 2.2.2.1.

Table 7: Population-level priors by sex and species.

(a) Males

Species	L_∞	k	t_0	L_0	t_h	h
Centrophorus squamosus	$\mathcal{N}(136.46, 15)$	$\mathcal{N}_+(0.05, 0.1)$	$\mathcal{N}(-6.68, 1)$	$\mathcal{N}(37.5, 3)$	$\mathcal{N}(15.44, 1.5)$	$\mathcal{N}_+(0.15, 0.05)$
Isurus oxyrinchus	$\mathcal{N}(302.30, 20)$	$\mathcal{N}_+(0.052, 0.1)$	$\mathcal{N}(-9.0, 1.5)$	$\mathcal{N}(61, 5)$	$\mathcal{N}(2, 1)$	$\mathcal{N}_+(0.2, 0.1)$
Lamna nasus	$\mathcal{N}(185.77, 20)$	$\mathcal{N}_+(0.13, 0.1)$	$\mathcal{N}(-4.22, 1)$	$\mathcal{N}(79.70, 10)$	$\mathcal{N}(7.25, 1.5)$	$\mathcal{N}_+(0.2, 0.1)$
Mustelus lenticulatus	$\mathcal{N}(118.70, 15)$	$\mathcal{N}_+(0.16, 0.1)$	$\mathcal{N}(-2.02, 0.5)$	$\mathcal{N}(34.2, 5)$	$\mathcal{N}(3.7, 1)$	$\mathcal{N}_+(0.15, 0.05)$
Prionace glauca	$\mathcal{N}(342.90, 25)$	$\mathcal{N}_+(0.0878, 0.2)$	$\mathcal{N}(-1.2570, 1.5)$	$\mathcal{N}(54, 10)$	$\mathcal{N}(8, 1)$	$\mathcal{N}_+(0.2, 0.1)$

(b) Females

Species	L_∞	k	t_0	L_0	t_h	h
Centrophorus squamosus	$\mathcal{N}(192.18, 25)$	$\mathcal{N}_+(0.03, 0.05)$	$\mathcal{N}(-6.67, 1.5)$	$\mathcal{N}(37.5, 3)$	$\mathcal{N}(20.82, 1.5)$	$\mathcal{N}_+(0.15, 0.05)$
Isurus oxyrinchus	$\mathcal{N}(347, 20)$	$\mathcal{N}_+(0.013, 0.1)$	$\mathcal{N}(-11.30, 1.5)$	$\mathcal{N}(61, 5)$	$\mathcal{N}(2, 1)$	$\mathcal{N}_+(0.2, 0.1)$
Lamna nasus	$\mathcal{N}(210.86, 20)$	$\mathcal{N}_+(0.09, 0.1)$	$\mathcal{N}(-6.10, 1)$	$\mathcal{N}(86, 10)$	$\mathcal{N}(14.65, 1.25)$	$\mathcal{N}_+(0.2, 0.1)$
Mustelus lenticulatus	$\mathcal{N}(90.7, 15)$	$\mathcal{N}_+(0.42, 0.1)$	$\mathcal{N}(-0.77, 0.5)$	$\mathcal{N}(25.3, 5)$	$\mathcal{N}(4.7, 1)$	$\mathcal{N}_+(0.15, 0.05)$
Prionace glauca	$\mathcal{N}(267.49, 25)$	$\mathcal{N}_+(0.1259, 0.2)$	$\mathcal{N}(-1.0467, 1)$	$\mathcal{N}(54, 10)$	$\mathcal{N}(8, 1)$	$\mathcal{N}_+(0.2, 0.1)$

Notably, we used truncated Gaussian distributions for the priors of k and h, constraining them both to the domain of positive reals. For k, this guarantees that negative growth over time is not possible. For h, this ensures that the rate of growth decreases following the transition point at t_h. By contrast, a negative value of h would imply that growth increases after point t_h. For *Isurus oxyrinchus*, we specified t_h as the transition from juvenile to maturing age (2 years old). For all the other candidate species, t_h was treated as the age-at-maturity. Prior predictive checks were performed for all of our models.

2.4 Results

Table 8 summarises the performance of the candidate growth models by species and sex, as ranked by the PSIS-LOO estimate of the ELPD. Note that a higher PSIS-LOO score implies a better fitting model. Refer to Appendix I: Growth Model Metrics for a more detailed display of the PSIS-LOO metrics.

Table 8: PSIS-LOO scores for each model. The highest scores have been bolded and italicised, indicating the best fitting model for each species and sex.

(a) Males

Species	Original VBGM	Common VBGM	Biphasic L_∞	Biphasic k
Centrophorus squamosus	-804.32	-798.2	***-748.04***	-777.8
Isurus oxyrinchus	-383.68	-383.69	-383.33	***-382.9***
Lamna nasus	-778.76	-778.73	***-769.83***	-773.25
Mustelus lenticulatus	-611.3	-610.7	***-608.56***	-609.02
Prionace glauca	-1420.91	-1421.21	-1418.04	***-1414.69***

(b) Females

Species	Original VBGM	Common VBGM	Biphasic L_∞	Biphasic k
Centrophorus squamosus	-1234.88	-1445.52	-1163.16	***-1152.23***
Isurus oxyrinchus	-1266.28	-1265.58	-1273.2	***-1263.5***
Lamna nasus	-639.96	-638.38	-631.54	***-627.52***
Mustelus lenticulatus	-359.61	***-358.5***	-359.77	-358.74
Prionace glauca	-2585.02	-2585.35	***-2576.46***	-2581.5

We can see that biphasic growth models have a better fit for the majority of cases. In four cases, the biphasic L_∞ model had the greatest fit, while in five other cases, the biphasic k model was best. Only in one case, did a monophasic growth model have the best fit.

2.4.1 *Centrophorus squamosus*

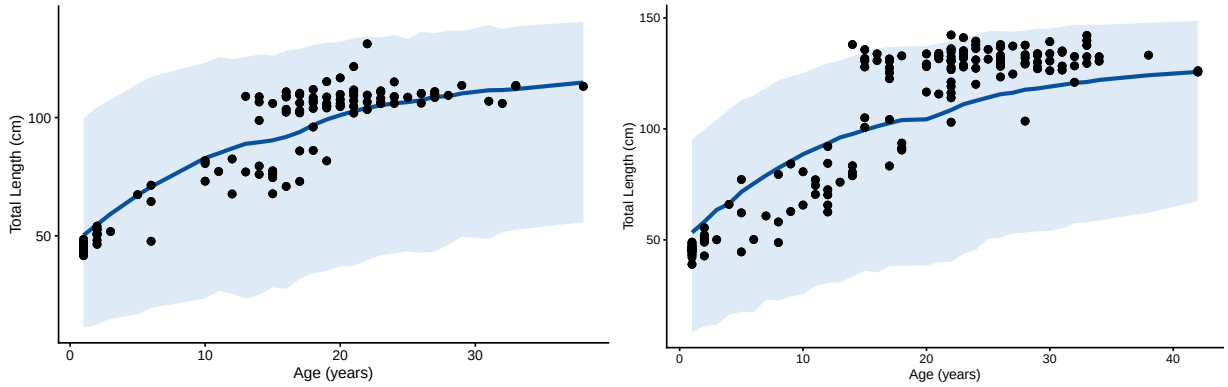
Table 9: Growth parameter estimates of *Centrophorus squamosus* by sex (95% credible intervals).

(a) Biphasic L_∞ estimates for males. (b) Biphasic k estimates for females.

Parameter	Estimate	Error	2.5% Q	97.5% Q	Estimate	Error	2.5% Q	97.5% Q
L_∞	119.64	10.1	101.7	141.94	134.41	2.92	128.99	140.38
k	0.15	0.05	0.04	0.26	0.07	0.03	0.02	0.14
t_0	-6.25	1.03	-8.25	-4.17	-6.71	1.22	-9.13	-4.12
t_h	15.4	1.48	12.49	18.32	20.58	1.46	18.03	23.72
h	0.21	0.09	0.04	0.4	0.18	0.1	0.04	0.39
σ	4.21	1.01	2.15	6.07	3.18	2.63	0.51	7.34

In comparison to the parameter estimates of Parker and Francis (2012a), the biphasic growth models each produced a smaller estimate of L_∞ and a greater estimate of k (Table 9). Note also that the data for female *Centrophorus squamosus* were able to overwhelm the prior for L_∞, leading to an estimate that was relatively smaller than 192.18 cm. The posterior of t_0 and t_h were not substantially different than the specified prior of each.

Our resulting growth curves offered some similarities and differences from those of Parker and Francis (2012a) (Figure 10). Namely, the growth curve for female *Centrophorus*



(a) Males

(b) Females

Figure 10: Growth curves for *Centrophorus squamosus* (95% posterior predictive intervals)

squamosus was able to reach an upper asymptote, unlike the resulting growth curve of the aforementioned study (Parker & Francis, 2012a). Moreover, we can clearly a prominent change in the shape of the curve for male *Centrophorus squamosus* at around 15 years of age. However, the credible intervals of the posterior predictive distributions for each model were relatively wide, and we can see that the growth models do not form the shape of a logistic curve. This is likely due to some of the limitations noted by Parker and Francis (2012a), as discussed in Section 2.3.3.1.1. Overall, the fitted curves appear to overestimate length for younger sharks, and underestimate the upper asymptote.

2.4.2 *Isurus oxyrinchus*

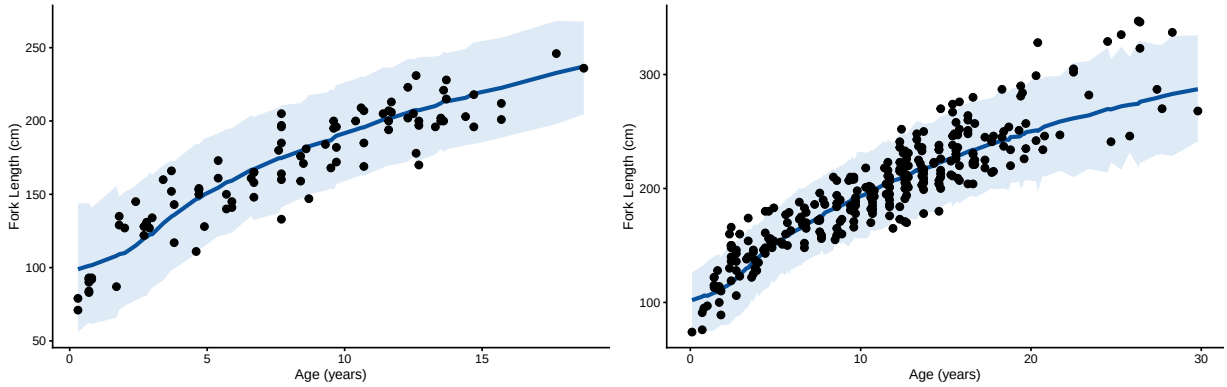
Table 10: Growth parameter estimates of *Isurus oxyrinchus* by sex (95% credible intervals).

(a) Biphasic k estimates for males. (b) Biphasic k estimates for females.

Parameter	Estimate	Error	2.5% Q	97.5% Q	Estimate	Error	2.5% Q	97.5% Q
L_∞	303.24	17.47	270.12	338.98	357.68	16.93	325.74	391.35
k	0.06	0.01	0.04	0.08	0.04	0.01	0.03	0.05
t_0	-7.00	1.11	-9.36	-4.99	-8.32	0.98	-10.47	-6.60
t_h	2.03	0.92	0.34	3.95	2.01	0.94	0.94	0.33
h	0.20	0.09	0.03	0.39	0.20	0.10	0.03	0.40
σ	3.32	2.02	0.67	7.29	1.45	0.98	0.48	3.33

The parameter estimates of male *Isurus oxyrinchus* were largely comparable to those produced by Bishop et al. (2006), though the biphasic k model offered a slightly larger value for t_0 (Table 10). By contrast, the estimates for female *Isurus oxyrinchus* were relatively different. Bishop et al. (2006) provided an estimate of 820.10 cm for L_∞, while the biphasic k model resulted in a substantially smaller asymptotic length of 357.68 cm. Our estimate of k (0.04) was notably larger than 0.013, and our value of t_0 (-8.32 years) was greater than the previous estimate of -11.30. The estimates for h and t_h were quite close to the specified prior values. Moreover, this was the only species where both sexes shared the same best performing model: The biphasic k model.

As noted by Bishop et al. (2006), *Isurus oxyrinchus* grow rapidly during their first two years. The t_h and h parameters are able to alter the rate of growth at roughly two years of age, leading to a slight change in the trajectory of the curve (Figure 11). Due to the shape



(a) Males

(b) Females

Figure 11: Growth curves for *Isurus oxyrinchus* (95% posterior predictive intervals)

of the data, the growth model fitted by Bishop et al. (2006) on female *Isurus oxyrinchus* produced an almost linear, straight line past two years of age, resulting in a large estimate of L_∞. By contrast, we were able to generate a much flatter curve that could plausibly eventually reach some upper asymptote, leading to a smaller estimate of L_∞.

2.4.3 *Lamna nasus*

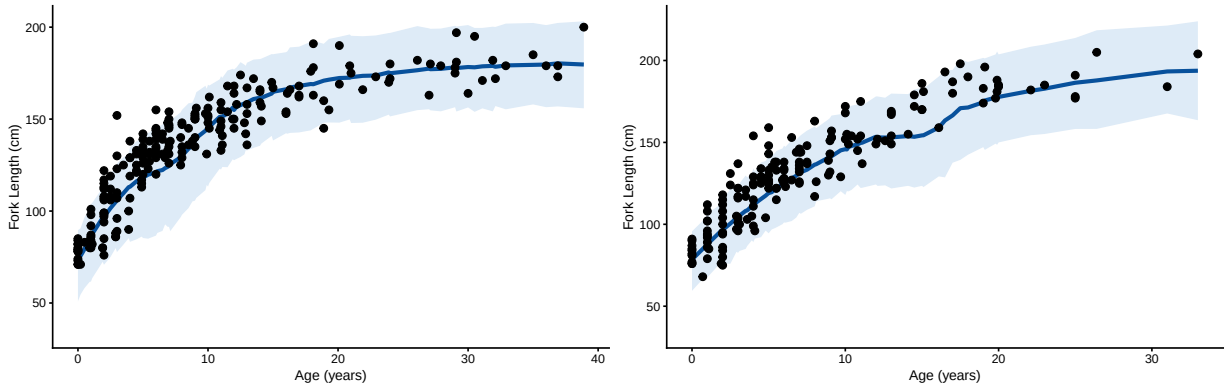
Table 11: Growth parameter estimates of *Lamna nasus* by sex (95% credible intervals).

(a) Biphasic L_∞ estimates for males.

(b) Biphasic k estimates for females.

Parameter	Estimate	Error	2.5% Q	97.5% Q	Estimate	Error	2.5% Q	97.5% Q
L_∞	181.03	3.96	173.82	189.45	205.63	6.82	192.05	219.57
k	0.13	0.02	0.09	0.15	0.08	0.01	0.05	0.09
t_0	-4.07	0.33	-4.76	-3.47	-6.27	0.54	-7.20	-5.17
t_h	7.28	1.45	4.45	10.13	14.72	0.99	12.74	16.37
h	0.18	0.07	0.05	0.34	0.22	0.08	0.04	0.37
σ	1.74	1.06	0.82	2.41	2.42	1.35	1.44	3.84

The parameter estimates for male *Lamna nasus* were largely comparable to those of Francis (2015b), though the biphasic L_∞ model produced a smaller value for L_∞: 181.03 cm compared to 185.77 cm (Table 11). Similarly, the biphasic k model resulted in smaller estimates of L_∞ (205.63 cm to 210.86 cm) and k (0.08 to 0.09) for female *Lamna nasus*. However, the parameter estimates of t_0 and t_h were relatively close to the specified prior values of each.



(a) Males

(b) Females

Figure 12: Growth curves for *Lamna nasus* (95% posterior predictive intervals)

The growth curves for both male and female *Lamna nasus* fitted well to the data, and were able to produce an upper asymptote (Figure 12). The estimate of h was particularly larger for females than males (0.22 compared to 0.18), and this is reflected in the plot above - females saw a greater decrease in growth at 15 years of age than males did at 7 years.

2.4.4 *Mustelus lenticulatus*

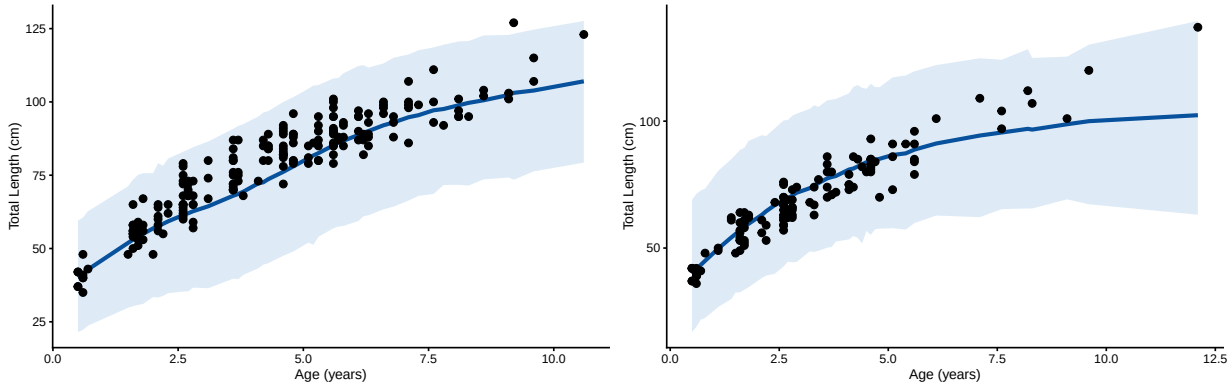
Table 12: Growth parameter estimates of *Mustelus lenticulatus* by sex (95% credible intervals).

(a) Biphasic L_∞ estimates for males. (b) Common VBGM estimates for females.

Parameter	Estimate	Error	2.5% Q	97.5% Q	Estimate	Error	2.5% Q	97.5% Q
L_∞	118.85	8.86	101.41	137.46	105.7	11.18	82.26	126.88
k	0.19	0.04	0.1	0.28	0.3	0.1	0.16	0.53
t_0	-1.87	0.39	-2.7	-1.16	-1.2	0.45	-1.99	-0.23
t_h	3.65	0.99	1.66	5.55				
h	0.21	0.09	0.03	0.39				
σ	2.11	1.13	1.21	3.19	4.38	1.2	1.97	6.66

The biphasic L_∞ model fitted to male *Mustelus lenticulatus* resulted in comparable growth parameter estimates to those from the study by Francis and Francis (1992) (Table 12). However, the biphasic L_∞ model produced larger parameter estimates for k and t_0. Notably, the common VBGM was the best fitting model for female *Mustelus lenticulatus*; this was the only case in this study where a monophasic model resulted in a better fit than a biphasic model. The L_∞ estimate of the common VBGM was larger than that of Francis and Francis (1992), while the k estimate was smaller. The authors had previously stated that their own L_∞ estimate for females was likely too small, while their k estimate was likely too large due to undersampling of females and oversampling of juveniles (Francis & Francis, 1992).

We can see that the shape of the growth curve for females closely resembles that of a logistic curve, though the same cannot be said for males (Figure 13). The credible intervals for both



(a) Males

(b) Females

Figure 13: Growth curves for *Mustelus lenticulatus* (95% posterior predictive intervals)

males and females were relatively wide, reflecting greater uncertainty within these models. The mean of the posterior predictive distribution of the biphasic L_∞ model on the left would suggest that the growth trajectory of male *Mustelus lenticulatus* is being underestimated; the solid blue curve only appears to pass through smaller sharks within each age group. Likewise, there appears to be some underestimation of older females. Therefore, the upper asymptote is underestimated for both sexes. For males, there is a very slight change in the shape of the mean growth curve at 3.65 years, which can be attributed to the influence of the h and t_h parameters.

2.4.5 *Prionace glauca*

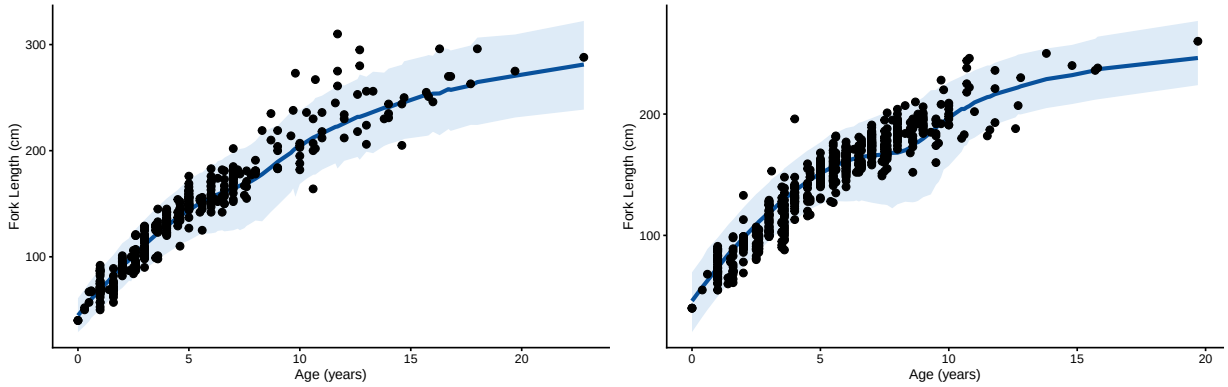
Table 13: Growth parameter estimates of *Prionace glauca* by sex (95% credible intervals).

(a) Biphasic k estimates for males. (b) Biphasic L_∞ estimates for females.

Parameter	Estimate	Error	2.5% Q	97.5% Q	Estimate	Error	2.5% Q	97.5% Q
L_∞	311.04	15.59	283.69	345.42	260.28	9.87	244.46	283.55
k	0.1	0.01	0.08	0.12	0.14	0.01	0.12	0.17
t_0	-1.62	0.26	-2.25	-1.16	-1.34	0.26	-1.84	-0.75
t_h	7.89	1	5.99	9.88	8.18	0.96	6.16	9.88
h	0.19	0.09	0.03	0.38	0.19	0.09	0.03	0.39
σ	0.58	0.17	0.32	1.25	2.21	0.44	1.35	3.09

The two biphasic growth models generated smaller estimates of L_∞, larger estimates of k, and smaller estimates of t_0 than those of Manning and Francis (2005) (Table 13). However, the values for t_h and h did not differ greatly from the specified prior distributions. Females had a lower L_∞ estimate and a greater k estimate than males, suggesting a faster growth rate and lower mean asymptotic length, which was consistent with the findings of Manning and Francis (2005).

The Bayesian models fitted well to the *Prionace glauca* data (Figure 14). We can see that both curves were able to produce an upper asymptote, and that there was a change in the shape of the curve at roughly 8 years for both males and females. Despite the comparable h estimates, the change in growth was more prominent for females than males. For males, the credible intervals appear to noticeably widen as age increases, reflecting both the lack of older, male sharks within the dataset, as well as the heteroscedastic, fan-shaped pattern



(a) Males

(b) Females

Figure 14: Growth curves for *Prionace glauca* (95% posterior predictive intervals)

within the data. Though there were also fewer older females within the dataset, the fan-shaped pattern does not appear, and consequently, the credible intervals do not appear to widen substantially as age increases.

2.5 Discussion

Several authors (Day & Taylor, 1997; Lester et al., 2004; Quince et al., 2008; Ricker, 1979) note that growth can be divided into a sequence of stages within the life history of a fish, with some changes in development occurring at each stage. Examples of these life stages include the transition from the juvenile stage to the maturing stage, the onset of sexual maturity, changes in habitat, or changes in behaviour (Araya & Cubillos, 2006). These authors (Day & Taylor, 1997; Lester et al., 2004; Quince et al., 2008; Ricker, 1979) further suggested that there were no strict reasons for choosing the VBGM to model the relationship between growth and maturity, as there are discontinuities in growth between the pre-maturity and post-maturity phase. Indeed, it has been argued that growth should be modelled via two separate equations - one for pre-maturity and one for post-maturity, after which surplus energy is reallocated for reproductive processes. The trajectory of the

growth curve should display some fundamental change at the transition point between life stages, such that we can distinguish between growth patterns before and after this transition (Day & Taylor, 1997).

We evaluated two separate suggestions provided by the aforesaid authors (Day & Taylor, 1997; Lester et al., 2004; Quince et al., 2008). The first of which was a more general suggestion regarding the use of biphasic growth models to account for transitions between different life stages, which we applied to *Isurus oxyrinchus* (as its pre-maturation and post-maturation growth patterns were largely linear and straight-line beyond the juvenile life stage). The second suggestion pertained to the use of BVBGMs to account for changes in growth during the pre-maturation and post-maturation stages, which we applied to *Centrophorus squamosus*, *Lamna nasus*, *Mustelus lenticulatus*, and *Prionace glauca*.

Our overall findings were largely consistent with those of previous studies on the biphasic growth of sharks (Acuña et al., 2001; Araya & Cubillos, 2006; Aversa, Dans, Garcia, et al., 2011; Contreras-Reyes et al., 2018; Olmeda-de la Fuente et al., 2022). There was strong statistical evidence in favour of the BVBGMs for our candidate species, with only a single case where a monophasic growth model proved to have a better fit to the data. As with Araya and Cubillos (2006) and Contreras-Reyes et al. (2018), our estimates of L_∞ were generally smaller than those of previous studies, while our estimates of k were larger, providing evidence of faster growth rates and lower average asymptotic body lengths. There were some examples from the literature where parameter estimates were stated to be biased or poor (Bishop et al., 2006; Francis & Francis, 1992). We were able to improve upon these examples with smaller values for overestimated parameters and larger values for underestimated parameters.

In contrast with Bishop et al. (2006), Francis (2016), and Parker and Francis (2012a), our

growth models consistently produced an upper asymptote. However, this was the case for both our monophasic and biphasic models (comparison not shown), suggesting that this difference was not driven by the choice of model. Though we cannot make any conclusive statements as to why our results differed, a key distinction is that we applied a Bayesian approach via MCMC for the purposes of parameter estimation, while the aforementioned authors adopted an optimisation approach via non-linear least squares. In concordance with Parker and Francis (2012a), our model fits for *Centrophorus squamosus* were relatively poor owing to the lack of data for particular age classes and limited prior knowledge of the species. In other cases, our models typically fitted relatively well to the data, resulting in curves that closely resembled the shape of logistic curves.

Keeping with the suggestions of Araya and Cubillos (2006), we fitted separate models to each species by sex. The authors (Araya & Cubillos, 2006) noted that females and males generally reach their first maturity at different ages, where females typically attain maturity at older ages with larger body sizes (Cortés, 2000; Ebert et al., 2013; Weigmann, 2016). This pattern would likely be obscured if we pooled data for the two sexes together (Araya & Cubillos, 2006). In the majority of cases, we detected differences in growth patterns between the two sexes at the transition point specified by t_h. However, h and t_h were found to be relatively insensitive to the provided growth data for both sexes and all species, suggesting that the data may not be particularly informative for these two parameters.

In the case of biphasic growth models, the dominant explanation for changes in growth have been energy reallocation for reproductive investment, changes in habitat, and diet (Contreras-Reyes et al., 2018; K. Wilson et al., 2018). For Soriano et al.'s (1992) BVBGMs, the phasic transition point was derived empirically using changes in diet for *Lates niloticus*. This justification may not be appropriate for chondrichthyans, with a more likely

explanation being energy reallocation for reproductive investment (Araya & Cubillos, 2006; Contreras-Reyes et al., 2018; Minte-Vera et al., 2016). For *Centrophorus squamosus*, *Lamna nasus*, *Mustelus lenticulatus*, and *Prionace glauca*, our theory was that reproductive investment acted as the principal driver for this transition. For *Isurus oxyrinchus*, it is known that juveniles grows rapidly to maturity for both Pacific and Atlantic populations (Bishop et al., 2006; Natanson et al., 2006), though we do not assign a particular reason (such as reproductive investment) for the decrease in growth that occurs after two years of age. We did identify some spatiotemporal variability in the distribution of sampled *Isurus oxyrinchus*, with a large proportion of juveniles located in Auckland East (refer to Figure 6), so habitat shift (and possible consequent changes in diet) could be a contributory factor for this transition. Refer to Appendix II: Varying intercept estimates for growth models for our estimates across different years and areas.

2.6 Recommendations

The results of our study provide evidence that the monophasic VBGM should not be chosen *a priori* as the only model to describe the growth patterns of chondrichthyans. Careful consideration of alternative models is recommended, as the VBGM may not always fit best (Cailliet et al., 2006a). There is substantial diversity in the growth patterns of the five candidate species selected for this study, though there are some commonalities including decreases in growth rates beyond some transition point such as the age-at-maturity. Therefore, a multi-model approach including biphasic growth models may offer more robust growth parameter estimates for New Zealand chondrichthyes.

The parameter estimates of the VBGM are often incorporated into stock assessments (Minte-Vera et al., 2016; Pardo et al., 2013; Trippel & Harvey, 1991), and the inclusion of

growth models that fail to account for any decreases in growth may propagate bias or error into said stock assessments (Araya & Cubillos, 2006; Minte-Vera et al., 2016; K. Wilson et al., 2018). More precise estimates of growth can be utilised within highly-structured stock assessments, thereby improving the interpretation of length-composition data (Minte-Vera et al., 2016). Likewise, these robust estimates can be used for the assessment of data-poor stocks, such as those of deepwater species, via "life-history invariants" methods (Prince et al., 2015). Minte-Vera et al. (2016) further propose that the improved performance of biphasic models that account for additional reproductive costs of older chondrichthyans may be understated when used within stock assessment models that are fit to length-composition data, as there is frequently a lack of validation data for these older sharks.

Our work offers growth parameter estimates that are potentially more robust than those of previous research efforts, as we have incorporated changes in growth rates over time within our models. With respect to future research, more recent approaches for assessing data-poor stocks have included life-history ratios and life-history invariants to maximise the information gained from length-frequency data (Hordyk et al., 2015; Prince et al., 2015). The life-history invariants are derived from the parameters of the VBGM, and more robust parameter estimates from biphasic models may consequently lead to more robust life-history invariants. Likewise, various sources of sampling bias were also identified by the studies associated with each candidate species, potentially leading to frequency distributions of length and age containing asymmetry and heavy tails (Lopez Quintero et al., 2017). Such sampling bias is usually the result of size-selectivity during the sampling process (Montenegro & Branco, 2016), favouring smaller and faster growing individuals within each age group (Contreras-Reyes et al., 2018; Taylor et al., 2005). While we have principally chosen Gaussian-distributed models, more heavy-tailed distributions that are

capable of accommodating for skew, such as the t-distribution, may be more appropriate (Contreras-Reyes et al., 2018; Lopez Quintero et al., 2017; Montenegro & Branco, 2016).

Our research offers statistical evidence in favour of biphasic growth of chondrichthyans. We argue that the monophasic VBGM may not always fit best, and found that our two BVBGMs typically provided a better fit to the data. Despite this, the PSIS-LOO metric used here is simply a measure of predictive accuracy on unseen data (Vehtari et al., 2015, 2017). Wang et al. (1995) argue that the selection of a growth model is subjective, and proposed that a pragmatic choice based upon prior studies and experience, as opposed to goodness-of-fit to the data, should be used. Conversely, Haddon (2010) suggests that the model which best explains the underlying growth process should be used. Indeed, the overall "best" model might vary by context, and the statistical results and biological interpretation of a model may often conflict with one another (Araya & Cubillos, 2006; G. Cailliet et al., 2006; G. Cailliet & Goldman, 2004). Despite this, we are optimistic that our work will foster greater discussion regarding the application of Bayesian biphasic models for growth modelling.

3 An investigation of automated vertebrae reading for *Isurus oxyrinchus* using Bayesian deep learning

Abstract

Vertebral band counting is a common approach to the aging of chondrichthyans. However, this method is both time-consuming and prone to human-error. There may be opportunities for a novel application of deep learning algorithms to automate the age estimation of chondrichthyans. Convolutional neural networks (CNNs) have previously been applied for the age interpretation of bony fish otoliths, though the automated interpretation of vertebral band patterns has yet to be investigated. Here, we compared three deep learning methods for the automation of age estimation to an image dataset of *Isurus oxyrinchus* vertebral bands: DeepEnsembling, mixture of Laplace approximations (MoLA), and multi-stochastic weight average Gaussian (MultiSWAG). The mean squared error (MSE) and Gaussian negative log-likelihood (NLL) metrics were used to assess predictive performance, while the coverage rates of our prediction intervals were used to assess model calibration. Overall, MultiSWAG appeared to offer marginally better predictive power and well-calibrated prediction intervals, relative to DeepEnsembling and MoLA. Moreover, MultiSWAG predictions generally closely matched the age estimates provided by human readers, though predictive accuracy and certainty tended to be poorer for older age classes. We conclude that the application of MultiSWAG for the automated vertebrae reading of *Isurus oxyrinchus* demonstrated promising initial results. However, we found no conclusive evidence that the deep learning models were explicitly counting the bands of the shark vertebrae when the integrated gradients method was used to interpret our predictions. Further refinement is required to ensure that our CNNs are treating vertebral bands as the most discriminative

feature for age estimation purposes. Greater investigation is required to determine if our results are generalisable, and if we can automate the process of vertebral band counting for a variety of species and datasets.

3.1 Introduction

The counting of vertebral growth bands is a common method for the age estimation of chondrichthyans (Cailliet et al., 2006c; Carrier et al., 2012; Francis et al., 2007a; Harry, 2018), such as *Isurus oxyrinchus* (Bishop et al., 2006; Francis, 2016; Kinney et al., 2016; Natanson, n.d.; M. C. Ribot-Carballal et al., 2005), yet vertebrae aging is subjective (Francis, 2015a; Francis & Ó Maolagáin, 2016), and can be both error-prone (Francis & Ó Maolagáin, 2016) and time consuming (Moore et al., 2019, 2021). Thus, new means of interpreting vertebral growth bands are required to improve accuracy in counts and reduce staff time allocated to this process. There may be opportunities to use machine learning algorithms to automate the process of counting vertebral growth bands (Moore et al., 2019, 2021). As far as we are aware, this automated approach has yet to be applied to chondrichthyans.

3.1.1 Sources of uncertainty in age and growth studies

Otoliths are formations found in the inner ear chambers of all teleost (bony) fish, which aid in perceiving sound and maintaining balance and spatial awareness - these otoliths contain growth marks which act as indicators of age (Campana, 1992). By contrast, chondrichthyans lack the well-conserved otoliths that teleostei have (Carrier et al., 2018; Harry et al., 2022). It was Ridewood and Williams (1921) who first proposed using the calcification patterns of vertebrae for the age determination of chondrichthyans, principally owing to the lack of alternative aging methods at that time. However, the authors (Ridewood & MacBride, 1921) cautioned that vertebrae were of limited use for aging or taxonomic classification, and noted that the calcified masses of vertebrae contained considerable variation and were more closely associated with the physiological purposes of the vertebral column than of age.

Despite this, opaque and translucent band pairs within the vertebral centra have become the most commonly used structure for age estimation studies (Carrier et al., 2018; Harry et al., 2022; Natanson et al., 2018). In earlier aging studies, it was often assumed that the deposition of growth bands was annualised (Carrier et al., 2018), though validated data regarding the periodicity of vertebral band pair formation is available for only a limited number of species (Ardizzone et al., 2006; G. Cailliet & Goldman, 2004; Carrier et al., 2018; Harry et al., 2022; Kneebone et al., 2008). For some species, there may be limited-to-no correspondence between band pairs and time (Carrier et al., 2018; Natanson et al., 2018). Likewise, this correspondence may be inconsistent over the lifetime of other species, resulting in the underestimation of age by readers (Francis et al., 2007b; Harry, 2018; Harry et al., 2013; Natanson & Deacy, 2019; Passerotti et al., 2014).

Band pair deposition has been linked to the somatic growth of chondrichthyans (Carrier et al., 2018; Harry et al., 2022; Natanson et al., 2008; Natanson & Cailliet, 1990), and in many cases the band pair counts differ along the vertebral column (Chidlow et al., 2007; Harry et al., 2022; Huveneers et al., 2013; Natanson et al., 2008). For example, the deposition of band pairs has been connected with ontogenetic changes in at least one population of *Isurus oxyrinchus* (Natanson et al., 2018; Wells et al., 2013). For a number of species, somatic growth has been linked to increases in band counts (Chidlow et al., 2007; Harry et al., 2022; Natanson et al., 2008; Natanson et al., 2018; Natanson & Cailliet, 1990), which suggests that centra grow in proportion to somatic growth. This means that the deposition of vertebral bands is only an indirect measure of age, and is more directly associated with the bodily growth of the shark (Natanson et al., 2018). Natanson et al. (2018) further suggest that vertebral bands are associated with structural properties of the centrum and the body shape (e.g. physical support and flexibility) of chondrichthyans.

Hence, the interpretation of vertebrae is a subjective procedure with inherent uncertainty involved in the reading process (Bowlby & Gibson, 2020; Francis, 2015a, 2016, 2016; Moore et al., 2021). More recent age and growth studies have resulted in some doubt as to whether growth zone counts accurately represent age, particularly for older specimens where age underestimation has been a cause of concern (Francis et al., 2007b; Harry, 2018; Jorgensen et al., 2022; Natanson et al., 2018; Raoult et al., 2018).

3.1.2 Automated otolith reading

The literature on the application of computer vision algorithms for the automated age estimation of vertebrae is sparse. By contrast, there is substantial research interest in applying these methods for the automated reading of otoliths (Carbini et al., 2008).

Automated otolith reading using machine learning has been explored for many years (Carbini et al., 2008; Fablet, 2006; Fablet & Le Josse, 2005; S. G. Robertson & Morison, 1999; S. Robertson & Morison, 2001; Takashima et al., 2000; Troadec, 1991), although predictions have typically been less precise then estimates made by experienced readers (Fablet & Le Josse, 2005; Fisher & Hunter, 2018; S. G. Robertson & Morison, 1999). Several machine learning approaches have been applied to this task, including support vector machines (SVMs) (Bermejo et al., 2007; Fablet & Le Josse, 2005), Bayesian parametric models (Fablet, 2006), and artificial neural networks (ANNs) (S. G. Robertson & Morison, 1999; S. Robertson & Morison, 2001).

Robertson et al. (1999) trialed ANNs to replicate the estimates of experienced readers for *Acanthopagrus butcheri*, *Macruronus novaezelandiae*, and *Pagrus auratus*. Their results compared favorably with experienced readers for *A. butcheri* and *P. auratus*, though by contrast, there were high error rates for *M. novaezelandiae* (Moore et al., 2021; S. G.

Robertson & Morison, 1999). Robertson et al. (2001) further noted that the predictive performance of their ANNs were improved upon by incorporating geometric features such as body length into their models. Fablet and Le Josse (2005) applied statistical learning methods to age *Pleuronectes platessa*, finding accurate classification rates that closely matched the level of agreement between different readers.

More recently, deep learning models such as convolutional neural networks (CNNs) have produced state-of-the-art results for computer vision tasks (Gu et al., 2018; Li et al., 2021; O'Shea & Nash, 2015), and their application for the automation of otlith reading has demonstrated great success (Fisher & Hunter, 2018; Moore et al., 2019, 2021; Myers et al., 2020; Ordoñez et al., 2022; Politikos et al., 2021, 2022). Conversely, the automated interpretation of chondrichthyan vertebrae is under-explored within the literature. Hence, there is an opportunity to evaluate methods that may facilitate this automation.

Moen et al. (2018) treated age estimation as a regression task, and trained the Inception V3 CNN architecture on otoliths sampled from *Reinhardtius hippoglossoides*. Martinsen et al. (2022) utilised heat maps to identify the most discriminative features of otolith images for *R. hippoglossoides*. Domain adaptation methods, such as Adversarial Generative Adaptation, allowed CNNs that were trained on otolith images from Moen et al.'s study (2018) to generalise well to data sourced from other otolith image laboratories (Ordoñez et al., 2022). More recently, the Contrastive Language-Image Pretraining (CLIP) architecture demonstrated promising results for few-shot learning, allowing a deep learning model to be trained on a small dataset of annotated images for *Pleuronectes platessa*, *Gadus morhua*, *Melanogrammus aeglefinus*, and *R. hippoglossoides* (Sigurardóttir et al., 2023). Finally, recent work within New Zealand training Inception V3 on otolith images of *P. auratus* and *M. novaezelandiae* has shown encouraging results, erring by 1.2 - 1.3 years on average

(Moore et al., 2019, 2021).

3.1.3 Aims

The aim of this chapter is to explore the use of CNNs for automating the process of age estimation via the interpretation of chondrichthyan vertebrae, and for comparing the accuracy of our age estimates from CNNs to those of human readers. Moreover, owing to the uncertainty inherent within the vertebrae reading process, a principled approach to well-calibrated uncertainty quantification is required to foster confidence in the predictions made by these CNNs. We perform a comparison between three probabilistic deep learning methods for the quantification of uncertainty: DeepEnsembling (Lakshminarayanan et al., 2017), mixture of Laplace approximations (MoLA) (Eschenhagen et al., 2021), and multi-stochastic weight average Gaussian (MultiSWAG) (A. G. Wilson & Izmailov, 2020). Finally, we incorporate interpretable machine learning via integrated gradients to determine the most discriminative features identified by our models.

Our model species is *Isurus oxyrinchus*, which is frequently caught in New Zealand as bycatch by surface longline (SLL) vessels targeting tuna (Bishop et al., 2006; Francis et al., 2001; Francis, 2016, 2016; Griggs & Baird, 2013). Band counts were performed by trained vertebrae readers on images using the aging technique for *Lamna nasus* (Natanson et al., 2002), which has also produced reliable age estimates for *Isurus oxyrinchus* (Bishop et al., 2006). A vertebral band pair is composed of one light and one dark band, and the assumption was made that these band pairs represented one year of growth (Bishop et al., 2006).

We contribute to the current literature by applying and evaluating state-of-the-art methods for the automated estimation, uncertainty quantification, and interpretation of vertebrae

reading to *Isurus oxyrinchus*. In what follows, we provide an overview of the theory behind CNNs, integrated gradients, and probabilistic deep learning, with a focus on scalable Bayesian deep learning methods for uncertainty quantification.

3.2 Theory and review of literature

3.2.1 Convolutional neural networks

In 1989, LeCun et al. (1989) published the pioneering paper on the LeNet-5 architecture, the predecessor to the modern CNN (Gu et al., 2018). While the LeNet-5 model was able to effectively classify handwritten digits due to its ability to automatically extract informative features from the raw pixels of image data (Russakovsky et al., 2015), its performance was limited for more complex task due to computational bottlenecks and a lack of training data (Gu et al., 2018). In 2012, the AlexNet architecture achieved state-of-the-art performance in the ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al., 2015), revolutionising the field of computer vision. While the architecture of AlexNet was similar to LeNet, its greater depth allowed it to more effectively recognise patterns within the image data (Alzubaidi et al., 2021; Gu et al., 2018).

In the present day, CNNs are a state-of-the-art deep learning architecture (Li et al., 2021), and are primarily used for image recognition and processing tasks (Alzubaidi et al., 2021; Gu et al., 2018; Guo et al., 2016; LeCun et al., 1989; Li et al., 2021; O'Shea & Nash, 2015). They are composed of computational units known as neurons, which are grouped together into layers (Figure 15) (O'Shea & Nash, 2015; Simonyan & Zisserman, 2014). Examples include convolutional, pooling, and fully connected layers, which are then combined together to form a hierarchical representation of the input data (Simonyan & Zisserman, 2014). The neuron weights (parameters) are estimated during the training process, which

consists of two steps: Forward propagation, where the CNN makes a prediction based upon the supplied input data; and back propagation, where the CNN "learns from its mistakes" by evaluating the gradient of some loss function, which is then used to update the neuron weights (O'Shea & Nash, 2015).

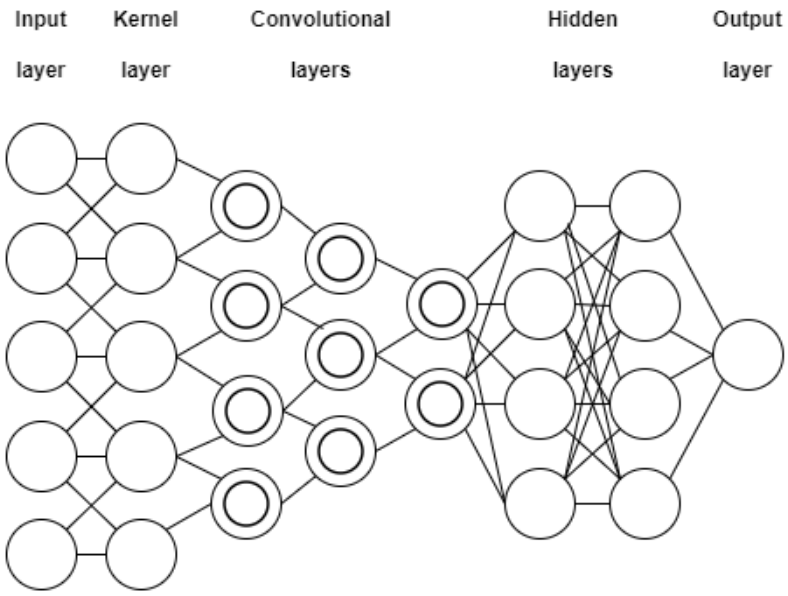


Figure 15: An example of a deep convolutional neural network architecture.

The locality property refers to the ability of a CNN to identify local patterns in the data via the use of convolutional layers, which apply kernels to local regions of image data, extracting increasingly finely grained spatial information and local features (LeCun et al., 1989; O'Shea & Nash, 2015). This enables CNNs to identify patterns and features, such as textures and edges, regardless of their location within the input image, and CNNs are therefore translation invariant (Krizhevsky et al., 2017; O'Shea & Nash, 2015).

3.2.2 Probabilistic deep learning

3.2.2.1 Uncertainty quantification

The uncertainty of a model can be separated into both reducible (epistemic) and irreducible (aleatoric) components (Der Kiureghian & Ditlevsen, 2009; Hora, 1996; Hüllermeier & Waegeman, 2021).

Broadly speaking, aleatoric uncertainty refers to the concept of randomness, specifically the inherent variability in an experiment's outcome from random factors (Der Kiureghian & Ditlevsen, 2009; Hora, 1996; Hüllermeier & Waegeman, 2021). In such an experiment, the data generating process contains a stochastic element that cannot be reduced by acquiring more information (Hüllermeier & Waegeman, 2021). As a consequence of this irreducible uncertainty, even the best of models cannot provide a fully confident answer with no uncertainty.

By contrast, epistemic uncertainty is related to a lack of understanding or knowledge about the most suitable model; this type of uncertainty represents the decision-maker's ignorance, focusing on their epistemic state rather than any inherent randomness in the underlying phenomenon (Der Kiureghian & Ditlevsen, 2009; Hora, 1996; Hüllermeier & Waegeman, 2021).

A variety of methods have been proposed for performing uncertainty quantification for deep learning. These include scalable ensembling methods such as DeepEnsembling (Lakshminarayanan et al., 2017), alongside Bayesian approaches using Laplace approximation (Daxberger et al., 2021; Mackay, 1992; Ritter et al., 2018a); variational inference (Gal & Ghahramani, 2015a, 2015b; Graves, 2011; Osawa et al., 2019); approximate Bayesian inference using Monte-Carlo dropout (Folgoc et al., 2021; Gal & Ghahramani, 2016; Pop & Fulop, 2018); stochastic gradient Markov chain Monte Carlo (MCMC) (T. Chen et al., 2014; Gürbüzbalaban et al., 2021; Havasi et al., 2018; Nemeth & Fearnhead, 2021; Zou et al., 2019); and full-batch MCMC (Cobb & Jalaian, 2021; Foreman et al., 2021; Hoffman,

2017; Neal, 2012).

Here, we do not intend to provide an exhaustive list of uncertainty quantification methods available for deep learning. Instead, we discuss some of the more common approaches to performing Bayesian deep learning, and contrast them with the popular non-Bayesian DeepEnsembling method (Lakshminarayanan et al., 2017).

3.2.2.2 DeepEnsembling

Deep learning models have achieved state-of-the-art performance within many computer vision tasks, and yet despite this, they perform poorly at quantifying predictive uncertainty (Hüllermeier & Waegeman, 2021; Lakshminarayanan et al., 2017; Rahaman et al., 2021). Due to their overparameterised nature, deep learning models can easily overfit to training data (Ganaie et al., 2022; Rahaman et al., 2021; Srivastava et al., 2014), and tend to be overly confident in their predictions (Hüllermeier & Waegeman, 2021; Lakshminarayanan et al., 2017; Rahaman et al., 2021).

DeepEnsembling describes a modelling approach whereby multiple deep learning models, or ensemble members, are fitted to training data, and their results are then combined to achieve greater performance and more robust uncertainty estimates than that of a single model (Fort et al., 2019; Ganaie et al., 2022; Lakshminarayanan et al., 2017; Lee et al., 2015; A. G. Wilson & Izmailov, 2020). This approach takes advantage of the diverse representations of different ensemble members to learn different properties of the data (Fort et al., 2019; Ganaie et al., 2022; Lakshminarayanan et al., 2017; A. G. Wilson & Izmailov, 2020).

Lakshminarayanan et al. (2017) demonstrate that DeepEnsembling provides well-calibrated uncertainty estimates for regression and classification tasks, and that ensembles offer more

accurate and less overconfident estimates, particularly for out-of-distribution (OOD) or shifted data (Abe et al., 2022; Ovadia et al., 2019). By calibration, we mean that a predicted probability $\hat{p}$ corresponds well with its true probability p (Nixon et al., 2019). Using 95% prediction intervals as an example, we would expect the proportion of future observations that lie within said intervals to be approximately 95%, but with a poorly calibrated model, we may have too overly narrow or wide prediction intervals for this to be the case.

Let us consider a univariate regression problem $x \in \mathbb{R}^{\mathbb{D}}$ denotes the $\mathbb{D}$ dimensional features and $y \in \mathbb{R}^1$ denotes the targets. Let $\mathcal{D} = \{x, y\}$ be our set of features and targets. Assume we have a parametric model m of the form $p(y|x, \theta)$, where θ denotes the weights and biases of a deep learning model. Given that our parametric model has Gaussian-distributed outputs $y \sim \mathcal{N}(\mu, \sigma)$, we find the optimal weights of our model using maximum likelihood estimation:

$$P\left(y_i|x_i\right) = \left(2\pi\sigma^2\left(x_i\right)\right)^{-\frac{1}{2}} \cdot exp\left(-\frac{\left(y_i - \mu\left(x_i\right)\right)^2}{2\sigma^2\left(x_i\right)}\right) \tag{12}$$

or equivalently by minimising the negative log loss (NLL) of a heteroscedastic Gaussian $-\ln\left(\mathcal{L}\right)$:

$$-\ln\left(\mathcal{L}\right) \propto \sum_{i=1}^{N} \frac{1}{2\sigma_m^2\left(x_i\right)}(y_i - \mu_m\left(x_i\right))^2 + \frac{1}{2}\ln\left(\sigma_m^2\left(x_i\right)\right) + \frac{1}{2}\ln\left(2\pi\right) \tag{13}$$

Our parametric model therefore quantifies the aleatoric uncertainty $\sigma_m^2(x_i)$ of observation i. For the DeepEnsembling approach (Lakshminarayanan et al., 2017), $\mathcal{M}$ different models are initialised using $\mathcal{M}$ different random weight sets. The ensemble prediction is a mixture of Gaussians $\hat{y}_{ens} = \frac{1}{M}\sum_{m=1}^{M}\hat{y}$, with an ensemble mean:

$$\mu_{ens} = \frac{1}{M}\sum_{m=1}^{M}\mu_m \tag{14}$$

and ensemble variance:

$$\sigma^2_{ens} = \frac{1}{M}\sum_{m=1}^{M}\sigma^2_m + \frac{1}{M}\sum_{m=1}^{M}(\mu_m - \mu_{ens})^2 \tag{15}$$

where $\sigma^2_{ale} = \frac{1}{M}\sum_{m=1}^{M}\sigma^2_m$ is the aleotoric uncertainty, and $\sigma^2_{epi} = \frac{1}{M}\sum_{m=1}^{M}(\mu_m - \mu_{ens})^2$ is the epistemic variance (Lakshminarayanan et al., 2017).

3.2.2.3 A loss landscape perspective

Consider our priors $p(\theta)$ and our predicted values y^*, our loss function $\mathcal{L}$, and our optimisation objective $\mathcal{L}(\theta, y, y^*)$. Then $p(\theta|\mathcal{D})$ is the resulting posterior.

An exact evaluation of the posterior is typically infeasible for deep learning models due to both the high-dimensionality of the parameter space Ω, and non-convexity of the loss-landscape (Figure 16) (Fort et al., 2019; Izmailov et al., 2020; Lakshminarayanan et al., 2017; Maddox et al., 2019; Pearce et al., 2018; A. G. Wilson & Izmailov, 2020).

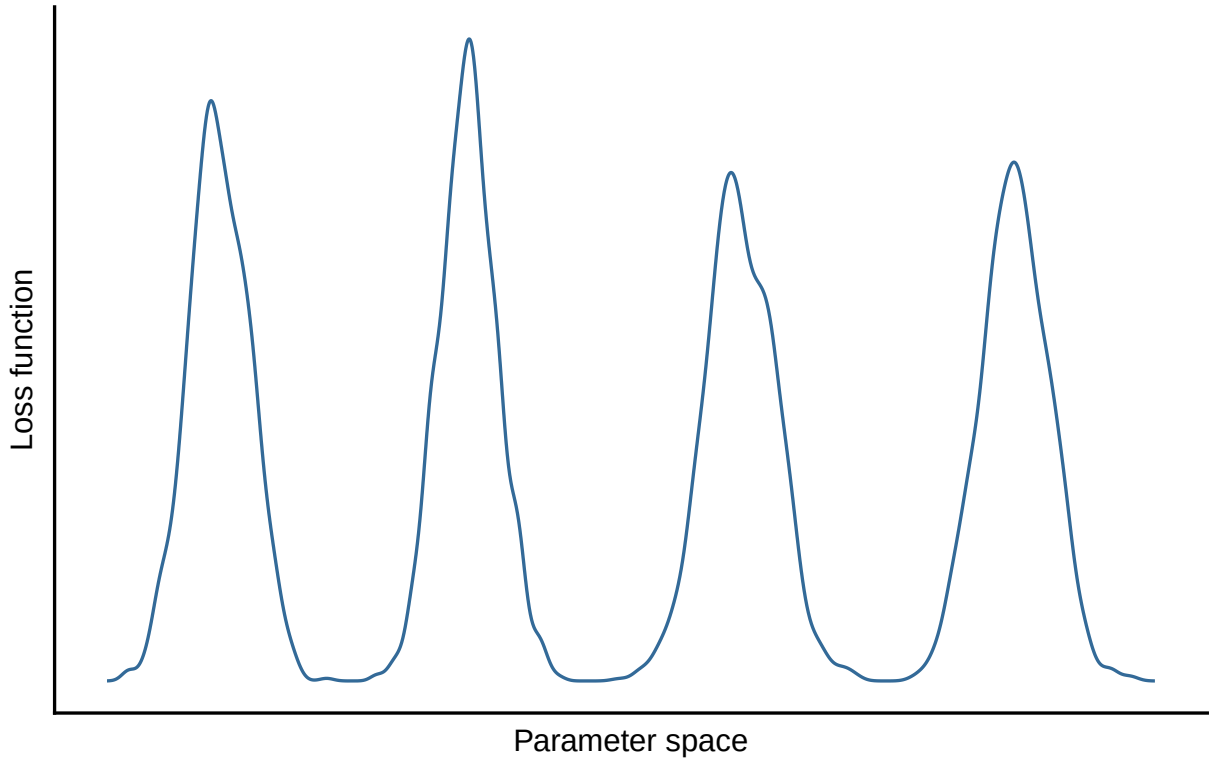


Figure 16: A multimodal loss landscape.

By contrast, it is generally easier to perform a maximum-a-posteriori (MAP) estimate $\hat{\theta}_{\text{MAP}}$, which corresponds to a mode of the posterior (Fort et al., 2019; Pearce et al., 2018). In the case of deep learning models, the MAP estimate can easily be found using standard gradient descent algorithms (Pearce et al., 2018; Ritter et al., 2018a). The MAP solution can be expressed as:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} p(\mathcal{D}|\theta)p(\theta) \tag{16}$$

However, the MAP solution provides only a single point-estimate (Fort et al., 2019; Maddox et al., 2019; Pearce et al., 2018; H. Wang & Yeung, 2020; A. G. Wilson, 2020), and whether such a method is suitable for Bayesian inference depends upon how sharply peaked the posterior is (A. G. Wilson, 2020). Deep learning models are typically underspecified, and

so their likelihoods are relatively diffuse meaning that the posterior will not be sharply peaked (A. G. Wilson, 2020; A. G. Wilson & Izmailov, 2020). Hence, the MAP estimate is a poor approximation of the posterior (Pearce et al., 2018; A. G. Wilson, 2020), and only considers a single setting of the parameters when performing predictions (Maddox et al., 2019).

For DeepEnsembling (Lakshminarayanan et al., 2017), $\mathcal{M}$ repeated optimisation procedures could potentially lead to $\mathcal{M}$ different solutions, assuming that the loss is non-convex (Fort et al., 2019; Lakshminarayanan et al., 2017; Pearce et al., 2018). Wilson (2020) argues that this provides a strong case for Bayesian deep learning, as this ensembling process can be viewed as repeated MAP training using random parameter initialisations, leading to the identification of different local optima (A. G. Wilson, 2020). The author further states that DeepEnsembling can be viewed as approximate Bayesian model averaging (BMA), and that the inclusion of multiple point masses with high likelihood in an ensemble leads to a more accurate approximation of the posterior predictive distribution (A. G. Wilson, 2020)

3.2.3 Bayesian deep learning

Bayesian methods quantify epistemic uncertainty with probability distributions over model parameters, and subsequently marginalise them to form a posterior predictive distribution (Maddox et al., 2019; McElreath, 2020). This contrasts with the deterministic DeepEnsembling method, where each weight is a single-point estimate. Here, we specify a prior distribution $p(\theta)$, encoding our prior distributional beliefs with respect to our model parameters. Our beliefs change in light of new information through the process of Bayesian updating, resulting in a posterior distribution. The likelihood of our model is $p(\mathcal{D}|\theta)$. To evaluate the posterior $p(\theta|\mathcal{D})$, we take the product of our prior and likelihood, over the probability

of observing our available data $p(\mathcal{D})$:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} \tag{17}$$

This is Bayes' theorem, a foundational formula for Bayesian inference.

In the 1990s, Bayesian methods were the state-of-the-art-approach to training deep learning models (Mackay, 1992; Neal, 2012), though they have become largely intractable for more modern deep learning architectures due to their large number of parameters.

Here, we discuss several common approaches to performing Bayesian estimation within deep learning.

3.2.3.1 Markov chain monte carlo

MCMC is considered to be the gold-standard for performing Bayesian inference (Nemeth & Fearnhead, 2021), and (under mild conditions) theoretically guarantees the asymptotically exact recovery of the posterior as the number of posterior draws increases (Neiswanger et al., 2013; Roberts & Rosenthal, 2004). MCMC algorithms approximate the posterior using samples produced from one ore more ergodic Markov chains, whose invariant distribution is the posterior itself (Nemeth & Fearnhead, 2021).

MCMC was also once considered the gold-standard for Bayesian deep learning (Maddox et al., 2019), owing to the pioneering work on the Hamiltonian Monte Carlo (HMC) algorithm by Neal (Betancourt, 2017; Maddox et al., 2019; Neal, 2012). HMC specifies a Hamiltonian function with respect to the target distribution that we wish to sample, where the potential energy and kinetic energy terms are parameterised by momentum auxiliary variables (Betancourt, 2017; T. Chen et al., 2014).

Given that our parameter space is continuous, HMC facilitates efficient Markov state transitions towards new and unexplored neighbourhoods within the state space (Betancourt, 2017). This is achieved by updating the momentum variables (Betancourt, 2017; T. Chen et al., 2014), thereby simulating a Hamiltonian dynamical system through the use of *Hamilton's equations*. However, HMC is a gradient-based method (Betancourt, 2017; Hoffman et al., 2014), and requires full gradients for its sampling procedure (Maddox et al., 2019); as such, it is computationally intractable for modern, overparameterised deep learning models (Maddox et al., 2019). These computational challenges have led to greater interest in new, scalable MCMC algorithms, which typically achieve greater efficiency by either parallelising the MCMC scheme or subsampling from the data (Nemeth & Fearnhead, 2021). Here, we focus on the latter approach.

3.2.3.1.1 Stochastic gradient methods

Chen et al. (2014) extended the HMC algorithm with the development of stochastic gradient HMC (SGHMC), a subsampling algorithm that promotes greater scalability via the inclusion of stochastic gradients. The authors incorporate a friction component to the momentum update, assume that the injected noise is Gaussian, and demonstrate that the second-order Langevin dynamics ensure that the target distribution remains an invariant one (T. Chen et al., 2014).

Suppose that we are interested in sampling from a target density $\pi(\theta)$, and that the unnormalized density can be expressed as:

$$\pi(\theta) \propto \exp\{-U(\theta)\} \tag{18}$$

where $U(\theta)$ is a continuous and differentiable potential function. For the MCMC algorithm,

we sample from $\pi(\theta)$ by simulating a stochastic process with π as the target distribution (Hastings, 1970; Nemeth & Fearnhead, 2021). Then the Langevin diffusion is expressed by the stochastic equation:

$$d\theta(t) = -\frac{1}{2}\nabla U(\theta(t))dt + dB_t \tag{19}$$

where $\nabla U(\theta(t))$ is the drift term, and B_t denotes Brownian motion at time t which has π as its invariant distribution (Grenander & Miller, 1994; Nemeth & Fearnhead, 2021). This defines the dynamics of a Markov process over infinitely small time-intervals (Grenander & Miller, 1994; Nemeth & Fearnhead, 2021), where the Langevin diffusion has approximate dynamics for some time interval $h > 0$ given by:

$$\theta(t+h) \approx \theta(t) - \frac{h}{2}\nabla U(\theta(t)) + \sqrt{h}Z, \quad k = 0, \ldots, K, \tag{20}$$

where Z is a vector of independent standard Gaussian variables. Using Equation 20, we may sample from the Langevin diffusion. Given a time period of $T = Kh$ for some integer K, we set θ_0 to be the initial state of the process, and simulate from Equation 20 to obtain draws from the process at $h, 2h, ..., Kh$ (Nemeth & Fearnhead, 2021).

Langevin diffusion has π as its invariant distribution, so this stochastic process acts as a theoretical basis for an MCMC algorithm, known as stochastic gradient Langevin dynamics (SGLD) (Grenander & Miller, 1994; Nemeth & Fearnhead, 2021). Both SGHMC and SGLD theoretically sample from the posterior in the limit of infinitely small step sizes, but in practice, the use of finite learning rates introduces approximation errors (Maddox et al., 2019; Mandt et al., 2017). Moreover, these stochastic sampling methods contain many hyper-parameters and are difficult to tune (Maddox et al., 2019). Finally, subsampling in-

troduces bias into the trajectory of the Hamiltonian sampler, fundamentally compromising the HMC algorithm (Betancourt, 2015).

3.2.3.2 Variational inference

Hinton et al. (1993) first suggested the use of variational Bayesian methods to perform approximate Bayesian inference within deep learning. However, variational methods had not seen widespread adoption for many years after this proposal, possibly due to the difficulty in evaluating an analytical solution to the posterior, or even to efficiently sample from it (Graves, 2011). The objective behind variational inference is to estimate the conditional probability distribution of some latent variables z, given a set of observed variables x (Blei et al., 2017). Here, we specify a family of candidate densities, and aim to identify the density closest to the optimal conditional by minimising the Kullback-Leibler (KL) divergence (Blei et al., 2017):

$$q^*(z) = \arg\min_{q(z)\in Q} KL(q(z)\|p(z|x)) \tag{21}$$

where $q^*(z)$ is the optimal candidate, $q(z)$ is our candidate density, Q is the set of all candidates, and $p(z|x)$ is our target conditional. In practice, it is not possible to minimise Equation 21 as it is contains $p(z|x)$ (the unknown conditional that we are attempting to approximate) (Seoh, 2020), but we can instead minimise the KL divergence by maximising an evidence lower bound (ELBO) (Blei et al., 2017; Seoh, 2020):

$$\text{ELBO}(q) = \mathbb{E}_{q(z)}[\log p(x,z)] - \mathbb{E}_{q(z)}[\log q(z)] \tag{22}$$

where $p(x,z)$ is the joint distribution of the observed data x and latent variables z, and all

other parameters are as defined.

Graves (2011) proposed a Gaussian variational approximation of the posterior, with Gaussian priors over the model weights. This was then followed by the development of the reparameterization trick, a method to sample from the posterior of deep latent variable models, such as variational autoencoders (VAEs) (Kingma et al., 2015). The reparameterization trick rewrites a random variable as a deterministic function of some other random variable, which has an easy-to-compute gradient, thereby reducing the difficulty in performing optimisation (Kingma et al., 2015). Variational Bayesian methods have shown some promise in moderately sized models (Maddox et al., 2019), but appear to scale poorly for larger models (Abdar et al., 2021; He et al., 2016; Maddox et al., 2019).

3.2.3.3 Monte Carlo dropout

Dropout is a regularisation method, where model weights and their connections are randomly selected and disabled during the training process (Srivastava et al., 2014), which reduces overfitting by preventing weights from co-adapting to the data (Srivastava et al., 2014). Gal and Ghahramani (2016) interpreted Monte Carlo (MC) dropout as an approximate variational method for Bayesian inference via the use of a spike and slab variational distribution (Abdar et al., 2021; Maddox et al., 2019). Here, dropout is enabled during testing time, and model weights are repeatedly randomly subsampled, resulting in a distribution of predictions (Gal & Ghahramani, 2016).

However, it has been noted that the model priors are specified using only the dropout rate, meaning that it is challenging to explicitly control these priors in a principled Bayesian fashion (Gal, 2016). Moreover, MC dropout has been criticised for failing to accurately quantify either epistemic or aleotoric uncertainty (Alarab et al., 2021), and may lead to

overly-confident predictions (Gal, 2016).

3.2.3.4 Laplace approximation

The Laplace approximation dates back to the 18th century (Laplace, 1774), and was first proposed by MacKay (1992) as a method for training Bayesian deep learning models. Suppose we have a Gaussian posterior $N(\theta^*, I(\hat{\theta}_{\text{MAP}})^{-1})$ where $I(\hat{\theta}_{\text{MAP}})^{-1}$ is the inverse of the Fisher information matrix. In the case of the Laplace approximation, the posterior is approximated by fitting a Gaussian centred at the MAP estimate, with a covariance matrix corresponding to its local curvature (Daxberger et al., 2021; Gal, 2016; Ritter et al., 2018a).

Conventional second-order optimisation techniques employ the Hessian matrix or a positive semi-definite approximation to produce parameter updates (Ritter et al., 2018a). They appear in the form:

$$\Delta = C^{-1}g \tag{23}$$

where C is the chosen curvature matrix, and g is the gradient of the loss function. However, the curvature matrix is intractable for high-dimensional deep learning models, meaning that the order of C may be in the terabytes (Ritter et al., 2018a). For a single observation, the diagonal blocks of the curvature matrices are Kronecker factored (Botev et al., 2017; Martens & Grosse, 2015; Ritter et al., 2018a):

$$H_\lambda = \frac{\partial^2 E}{\partial \text{vec}(\theta_\lambda) \partial \text{vec}(\theta_\lambda)} = Q_\lambda \otimes H_\lambda \tag{24}$$

where H_λ is the Hessian for the weights in layer λ, $Q_\lambda = \alpha_{\lambda-1}\alpha_{\lambda-1}^T$ is the covariance of

the activations $\alpha_{\lambda-1}\alpha_{\lambda-1}^T$, and $H_\lambda = \frac{\partial^2 E}{\partial h_\lambda \partial h_\lambda}$ is the pre-activiated Hessian (Ritter et al., 2018a).

While MacKay (1992) had originally adopted the use of the diagonal approximation to the inverse of the Hessian due to computational limitations, Ritter et al. (2018a) suggested employing either a diagonal or block Kronecker factored (KFAC) approximation to the Hessian matrix, and KFAC approximation was successfully used for online learning to prevent catastrophic forgetting (Ritter et al., 2018b).

3.2.3.4.1 Mixture of Laplace approximations

MoLA extends the Laplace approximation to perform post-hoc uncertainty quantification (Eschenhagen et al., 2021). Here, a Gaussian approximation of the posterior is fitted for each ensemble member m within some ensemble $\mathcal{M}$, forming a mixture of Gaussians of equally weighted components. A mixture of Gaussians acts as a universal approximator of any smooth density with some error (Carreira-Perpinan, 2000). Hence, we use this mixture to efficiently approximate a high-dimensional and complex multi-modal posterior. For the original implementation of MoLA by Eschenhagen et al. (2021), last-layer Laplace approximation is performed, meaning that a Gaussian approximation is constructed only for the weights of the last layer of the deep learning model.

3.2.3.5 SWAG

Here, we discuss the stochastic weight average Gaussian (SWAG) algorithm, and its two predecessors: Stochastic gradient descent (SGD) and stochastic weight average (SWA). The SGD algorithm (Amari, 1993) has the following rule:

$$\Delta\theta_t = -\eta_t \left(\frac{1}{B} \sum_{i=1}^{B} \nabla_\theta \log p(y_i|f_\theta(x_i)) - \frac{\nabla_\theta \log p(\theta)}{N} \right), \tag{25}$$

where η is the learning rate, the ith input and target are x_i and y_i respectively, the size of the training set is N, the batch size is B, and the model f has the weight vector θ^2. The loss of the negative log likelihood $-\sum_i \log p(y_i|f_\theta(x_i))$ is combined with regulariser $logp(\theta)$. This results in a maximum likelihood (ML) estimate, which fails to account for the epistemic uncertainty of the model (Maddox et al., 2019).

Maddox et al. (2019) interpret the SWA algorithm as a method for estimating the mean of the invariant distribution of SGD iterates. Here, we run SGD with a constant learning rate, starting from a pre-trained solution, and then average the weights of each model that it traverses (Izmailov et al., 2018; Maddox et al., 2019). The SWA solution can be expressed as follows:

$$\theta_{\text{SWA}} = \frac{1}{T} \sum_{i=1}^{T} \theta_i \tag{26}$$

where for epoch i, the models weights will be θ_i, and the total number of epochs are T. Given a high learning rate schedule, the SGD algorithm will explore the set of set of possible solutions, rather than converging to a single point mass (Izmailov et al., 2018; Maddox et al., 2019). Izmailov et al. (2018) argue that SWA provides greater generalisability and is more robust to distribution shift than a conventional SGD approach.

3.2.3.5.1 SWAG-diagonal

For SWAG, the SWA algorithm is extended to estimate the covariance of an invariant distribution, resulting in a Gaussian approximation of the posterior (Maddox et al., 2019).

To fit a simple diagonal format for the covariance matrix, a running average is kept for the second uncentred moment of each weight (Maddox et al., 2019). To calculate the covariance, the following identity is used:

$$\bar{\theta}^2 = \frac{1}{T}\sum_{i=1}^{T}\theta_i^2, \quad \Sigma_{\text{diag}} = \text{diag}(\bar{\theta}^2 - \theta_{\text{SWA}}^2) \tag{27}$$

where the squares of θ_{SWA}^2 and θ_i^2 are applied elementwise (Maddox et al., 2019). This results in the approximate Gaussian posterior $N(\theta_{\text{SWA}}^2, \Sigma_{diag})$.

3.2.3.5.2 SWAG: Low rank plus diagonal covariance structure

The diagonal covariance approximation is standard in Bayesian deep learning (Blundell et al., 2015; Kirkpatrick et al., 2017; Maddox et al., 2019), yet for the purposes of Bayesian inference, it can be overly restrictive (Maddox et al., 2019). For SWAG, this covariance approximation is extended via a flexible low-rank plus diagonal posterior approximation. The sample covariance matrix of the SGD iterates can be expressed as the sum of outer products of rank T (Maddox et al., 2019):

$$\Sigma = \frac{1}{T-1}\sum_{i=1}^{T}(\theta_i - \theta_{\text{SWA}})(\theta_i - \theta_{\text{SWA}})^T \tag{28}$$

The value of θ_{SWA} is inaccessible during training, so the sample covariance is approximated using $\Sigma \approx \frac{1}{T-1}\sum_{i=1}^{T}(\theta_i - \bar{\theta}_i)(\theta_i - \bar{\theta}_i)^T = \frac{1}{T-1}DD^T$, where D is the deviation matrix comprised of columns $D_i = (\theta_i - \bar{\theta}_i)$, and $\bar{\theta}$ is the running estimate of the parameter means obtained from the first i samples (Maddox et al., 2019).

Only the last K epochs of training are used for the D_i vectors, such that the rank of the covariance matrix is limited (Maddox et al., 2019); K is the rank of the resulting

approximation and is a hyperparameter of SWAG. $\hat{D}$ is the matrix with columns equal to D_i, for $i = T-K+1, ..., T$.. This results in the low rank approximation $\Sigma_{low-rank} = \frac{1}{K-1}.\hat{D}\hat{D}^T$, which is combined with Σ_{diag}. This results in the Gaussian approximation of the posterior:

$$N\left(\theta_{\text{SWA}}, \frac{1}{2} \cdot (\Sigma_{\text{diag}} + \Sigma_{\text{low-rank}})\right) \tag{29}$$

From this point onwards, all references to SWAG refer to the low-rank plus diagonal approximation. We sample from SWAG using the following identity:

$$\tilde{\theta} = \theta_{\text{SWA}} + \frac{1}{\sqrt{2}} \cdot \Sigma_{\text{diag}}^{\frac{1}{2}} z_1 + \frac{1}{\sqrt{2(K-1)}} \hat{D} z_2, \quad \text{where} \quad z_1 \sim N(0, I_d), \quad z_2 \sim N(0, I_K) \tag{30}$$

where d is the number of parameters.

Maddox et al. (2019) argue that this method approximates fully Bayesian inference, and not simply MAP optimisation for some point mass. A Gaussian approximation of the posterior is produced from SGD iterates, $p(\theta|\mathcal{D}) \approx N(\theta; \mu, \Sigma)$ which is then sampled to perform Bayesian model averaging (BMA), and Gaussian priors are specified using different regularisers during the optimisation procedure (Maddox et al., 2019).

3.2.3.5.3 MultiSWAG

It has been noted that scalable methods for performing Bayesian inference within deep learning (variational inference, the Laplace approximation, and SWAG) are only able to identify a single mode within the posterior, and may therefore perform worse than DeepEnsembling for multimodal BMA, owing to the non-convexity of the posterior within high-dimensional settings (A. G. Wilson, 2020; A. G. Wilson & Izmailov, 2020). Wilson and Izmailov (2020)

combined the DeepEnsembling approach with SWAG to develop the MultiSWAG method. As with MoLA, the posterior is approximated using a mixture of Gaussians, where each equally-weighted Gaussian component is centred about a different mode of the posterior (Figure 17) (Eschenhagen et al., 2021; A. G. Wilson & Izmailov, 2020).

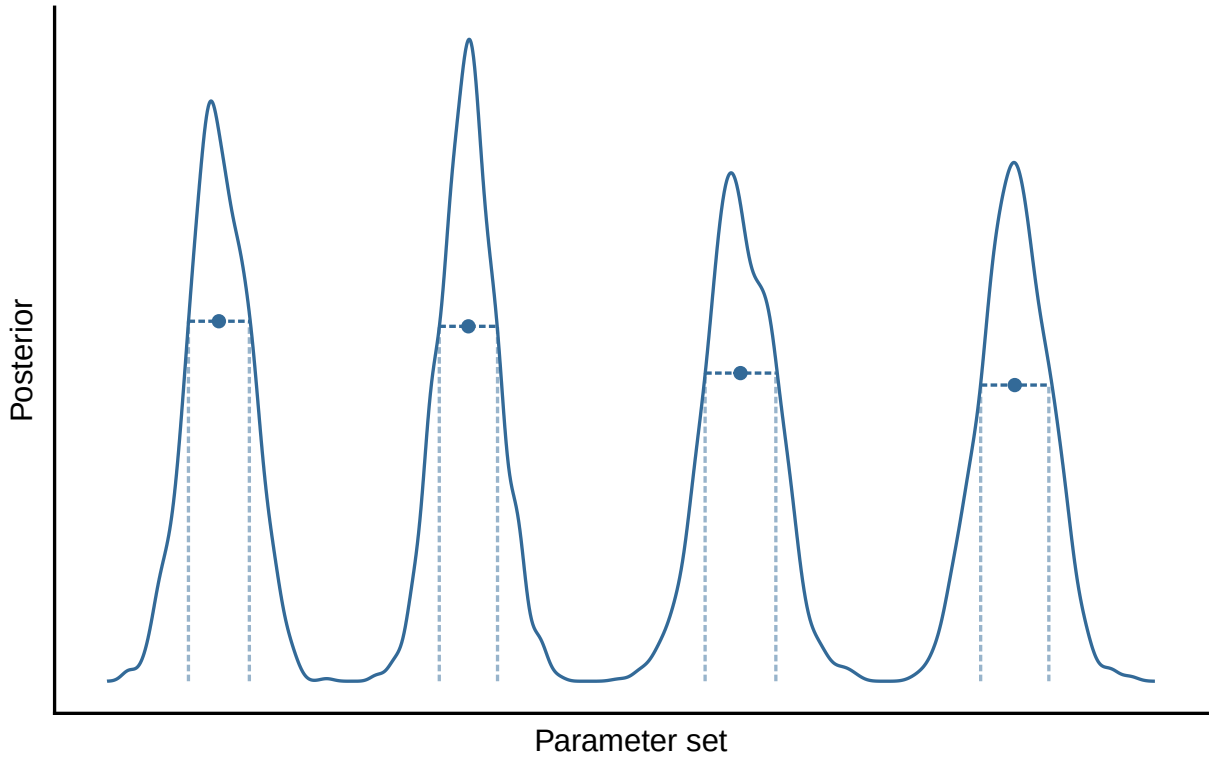


Figure 17: A multimodal posterior approximated using a mixture of Gaussians.

In the case of variational and Laplacian approximation methods, multiple samples are drawn about a single mode of the posterior, whilst in the MultiSWAG and MoLA approaches, samples are drawn across multiple modes, providing a more accurate approximation of the posterior (Eschenhagen et al., 2021; A. G. Wilson & Izmailov, 2020).

The MultiSWAG method has been utilised for several tasks, including climate prediction (Haddad, 2020); large scale cosmological structure modelling (Slav, 2021); predicting the dissolution of compact planetary systems (Cranmer et al., 2021); identifying desirable

molecular compounds within a chemical library (Hwang et al., 2020); and for identifying artificial satellites from spectroscopic images (Gazak et al., 2022).

3.2.4 Integrated gradients

The integrated gradients algorithm is a method for attributing the predictions of a deep learning model to its input features, thereby improving the interpretability of the model (Qi et al., 2019; Sundararajan et al., 2017). Integrated gradients calculates the contribution of each feature via the integration of the model's gradients with respect to the input along a straight-line path from a baseline value to the actual input (Sundararajan et al., 2017). The integrated gradients of input x can be expressed as follows:

$$IG_i(x) \approx (x_i - x_i') \times \frac{1}{m} \sum_{k=1}^{m} \frac{\partial F(x' + \frac{k}{m} \times (x - x'))}{\partial x_i} \tag{31}$$

where m is the number of steps in the Riemman approximation of the integral (Sundararajan et al., 2017). According to Sundararajan et al. (2017), advantages of integrated gradients include:

- Sensitivity: Assuming that a model changes its prediction by altering one feature while holding all others constant, then the attribute of that feature is non-zero.
- Implementation invariance: If there are two deep learning models which are functionally equivalent, meaning that the two models will produce the same outputs for the same input, then the attributions of these models will also be the same.
- Completeness: The sum of attributions will be equal to the difference in the model's output at the input and baseline.

3.3 Methods

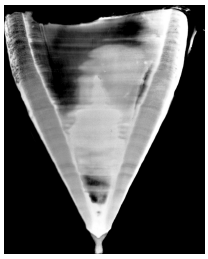
3.3.1 Vertebrae images of *Isurus oxyrinchus*

The National Institute of Water and Atmospheric Research (NIWA) conducted several age and growth studies (Bishop et al., 2006; Francis, 2016; Francis & Ó Maolagáin, 2016) using samples of *Isurus oxyrinchus* from two primary sources: Recreational gamefishing competitions and Ministry for Primary Industries (MPI) observers aboard tuna SLLs. Samples were collected from gamefishing competitions between 1994 and 2003 from New Zealand's North Island, while observers collected samples between 2002 and 2003 primarily from the South Island (Bishop et al., 2006).

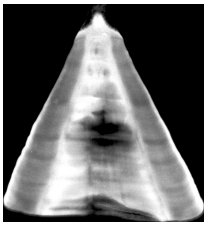
Images of the vertebral sections were processed and analysed during the aforementioned studies, and a dataset of 146 *Isurus oxyrinchus* vertebrae images belonging to 131 sharks was provided by NIWA for the present study. The majority of these images were composed of half bow-tie section pairs (Figure 18). Section pairs were subsequently separated, resulting in a complete dataset of 266 half bow-tie section images. In the case of one image, the vertebral bands were hardly visible, so this image was excluded from the training set. Images were scaled to a resolution of 224 x 224 x 3.

The right-skewed dataset was heavily dominated by juveniles and maturing sharks, with relatively few adults (Figure 19); Francis (2016) had highlighted the potential occurrence of sampling bias for data procured from SLL observers, and noted that selection bias, in particular, may be present. With such an unbalanced distribution of ages, we would therefore expect our deep learning models to demonstrate greater uncertainty when supplied with image data composed of older sharks.

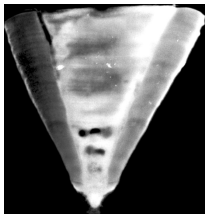
(a) Sample 301: Section A.



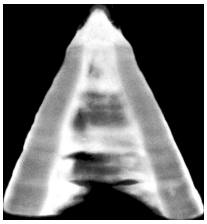
(b) Sample 301: Section B.



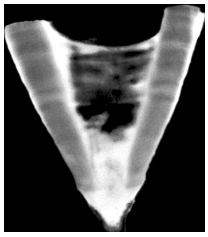
(c) Sample 305: Section A.



(d) Sample 305: Section B.

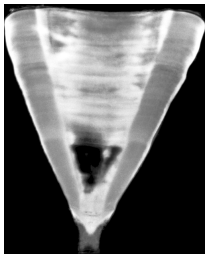


(e) Sample 333: Section A.



(f) Sample 333: Section B.

(g) Sample 363: Section A.



(h) Sample 363: Section B.

Figure 18: Half bow-tie section pairs for *Isurus oxyrinchus* vertebrae.

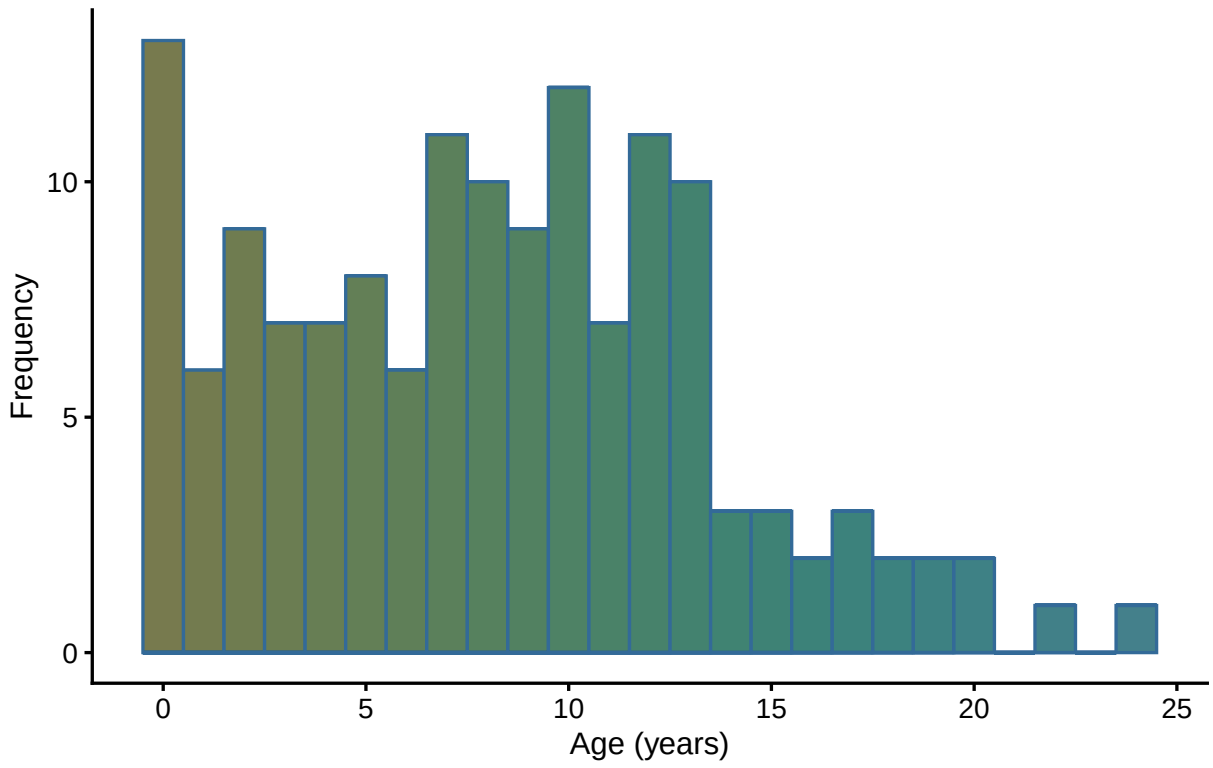


Figure 19: Frequency distribution of ages for the vertebrae dataset.

For model evaluation, we divided our dataset using a stratified 4-fold cross-validation procedure, accounting for each shark's specimen ID to ensure that it was not possible for our CNNs to be trained on one section pair and evaluated against another. Stratification was by age class, though this stratification was not possible for sharks older than 13 as there were too few per-class images available, so we retained those images in the training set. A batch size of 16 was specified for the training set.

3.3.1.1 Offline augmentation

It is generally agreed upon that large datasets are required for deep learning models, and that overfitting is likely to occur in the small-data regime (Buslaev et al., 2020; Shorten & Khoshgoftaar, 2019). Augmentation methods combat overfitting by introducing greater

variation into the data by artificially expanding the training set with synthetic observations, under the assumption that additional information can be gleaned from these augmented samples (Buslaev et al., 2020; Shorten & Khoshgoftaar, 2019). Geometric and colour transformations are examples of common augmentation techniques. Each augmented image was mapped to their original, non-augmented image using the specimen ID value, to ensure that no data leakage occurred during the cross-validation process - meaning that specimens within the validation dataset were not also found in the training set.

The Albumentations library (Buslaev et al., 2020) was used to perform various offline augmentations on our image dataset (Figure 20). From preliminary, empirical experiments, we found that the use of offline augmentations dramatically sped up the rate of convergence for our CNNs. For each of our 266 images, we generated 25 augmented images, leading to an augmented dataset of 6650 images. Beyond simple flips, rotations, and tone changes, the following augmentation methods were used: Elastic deformation (Simard et al., 2003); FancyPCA (Krizhevsky et al., 2017); CoarseDropout (DeVries & Taylor, 2017); Contrast limited adaptive histogram equalization (Zuiderveld, 1994); and GaussianBlur (Gedraite & Hadad, 2011).

3.3.2 Modelling approach

For each fold of our cross-validation procedure, we trained one DeepEnsembling ensemble, one MoLA ensemble, and one MultiSWAG ensemble. We specified ensemble sizes of $\mathcal{M} = 5$ for each. Each ensemble member used the ResNet-50 network architecture (He et al., 2016) as the backbone network (Figure 21). ResNet-50's classification head was replaced by a fully-connected layer with two outputs, one output being the mean μ of a Gaussian distribution, with the other being its standard deviation σ. The model weights were randomly

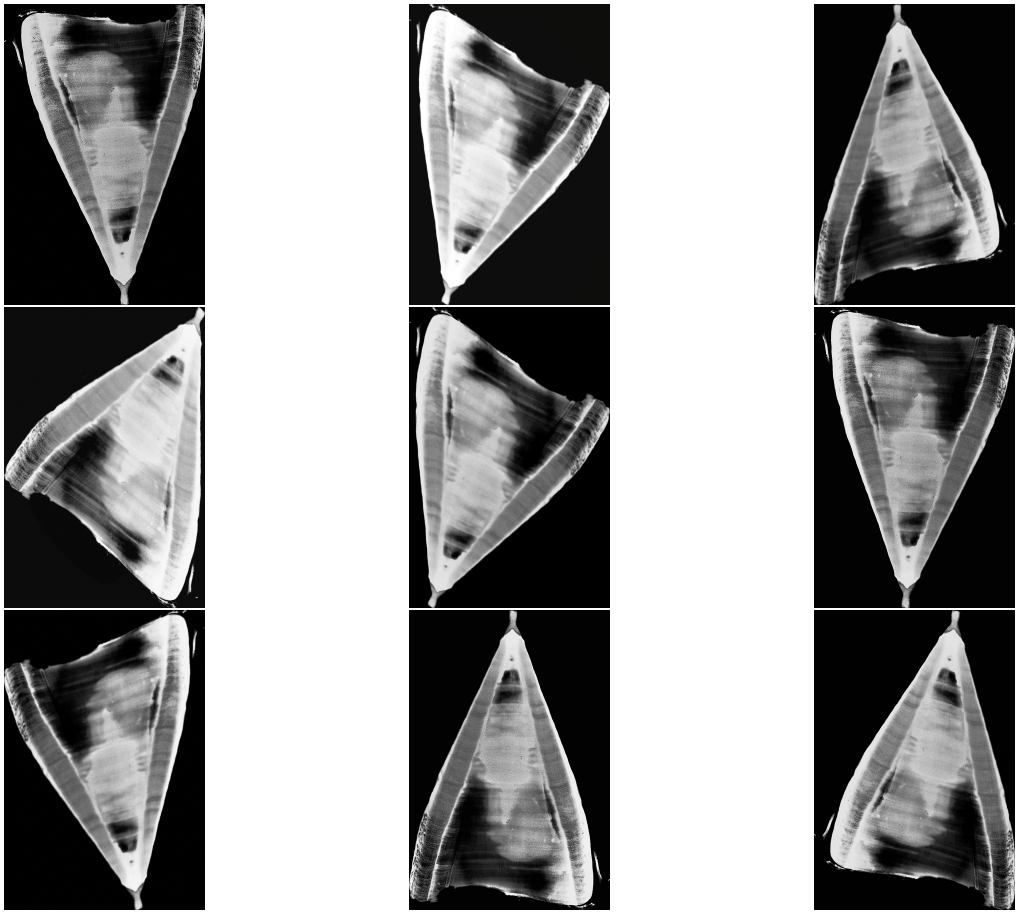


Figure 20: Augmented images for vertebral sample 301: Section A.

initialised, and models were optimised using the Gaussian NLL loss function. As we did not know the true ages of the *Isurus oxyrinchus* samples, the objective of the models was to most closely match the age estimates of the vertebrae readers.

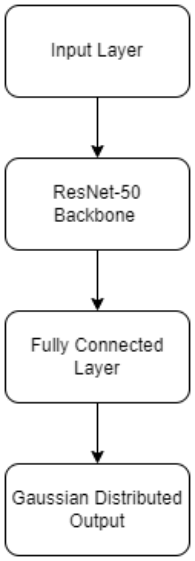


Figure 21: Ensemble model architecture.

For DeepEnsembling, we used the stochastic gradient descent optimiser (Robbins & Monro, 1951), with an initial learning rate of 1e-3, a weight decay of 1e-4, and a cosine annealing learning rate scheduler (Loshchilov & Hutter, 2016). Models were trained for 1000 epochs. We calculated the ensemble mean and variance using Lakshminarayanan et al.'s (2017) method, as described in Section 3.2.2.2.

When training the MoLA ensemble, we adopted the same approach as DeepEnsembling. We then performed post-hoc Bayesian inference using the KFAC approximation. Unlike Eschenhagen et al. (2021), we performed full-layer Laplace approximation using the *laplace* library (Daxberger et al., 2021). We drew 20 samples per ensemble member, for a total of 100 draws.

Finally, in the case of MultiSWAG, we used a comparable procedure as Wilson et al. (2020):

1. We trained 5 ensemble members in the same fashion as the DeepEnsembling model, until each member had converged to a mode within the parameter space.
2. For each ensemble member, we then continued training with an increased step size, causing the optimiser to take a random walk around the mode (Cranmer et al., 2021; Maddox et al., 2019).
3. Parameter values were sampled during this random walk, and each parameter's variance was approximated using a low rank plus diagonal covariance structure.
4. We trained each model for an additional 1000 epochs, with one sample per epoch.

We then took 20 samples from the Gaussian approximation of the posterior for each MultiSWAG ensemble member, leading to a total of 100 draws. For both MoLA and MultiSWAG, we were able to construct an approximate posterior predictive distribution. We could then perform averaging to produce point estimates when any direct comparisons with DeepEnsembling were required.

3.3.2.1 Model calibration

We had previously noted that deep learning models are often poorly calibrated (refer to Section 3.2.2.2 and Section 3.2.2.3), and are typically overly-confident in their predictions. To compare the DeepEnsembling, MoLA, and MultiSWAG methods, we calculated the coverage rates of their 90% and 95% prediction intervals on our testing data. The ensemble with coverage rates that were closest to 90% and 95% respectively were assumed to be better calibrated. We repeated this process 3 times, once per fold of the cross-validation procedure.

3.3.2.2 Integrated gradients for model interpretation

The *Captum* library offers several algorithms for interpreting the outputs made by deep learning models with respect to their inputs (Kokhlikyan et al., 2020). Here, we have used *Captum* to interpret the predictions of our CNNs via integrated gradients, with particular interest in identifying the most informative features of our vertebrae images. *Captum* generates images with a gradient of lightly and darkly coloured pixels - darker coloured pixels imply that they contribute more greatly to the predictions of the model.

3.4 Results

3.4.1 Model performance

Table 14: Model performance for three ensemble methods on testing data, with mean scores and standard deviations for the negative log-likelihood (NLL) and mean-squared error (MSE) across the 4-fold cross-validation procedure. The best scores have been bolded and italicised, indicating the best performing model

Model	NLL ± SD	MSE ± SD
DeepEnsembling	-1141.972 (58.094)	2.871 (0.236)
MoLA	-1187.357 (59.710)	2.747 (0.237)
MultiSWAG	***-1237.539 (59.690)***	***2.733 (0.236)***

The MultiSWAG method produced models with the greatest predictive performance, as assessed using the NLL and mean squared error (MSE) metrics, followed by MoLA and DeepEnsembling (Table 14). The MSEs for each of our three models were between 2.7-2.8 years. Moreover, the application of either MoLA or MultiSWAG lead to a comparable improvement in model performance, relative to the use of DeepEnsembling alone. However, the differences between models are not significant when accounting for the standard

deviation across folds.

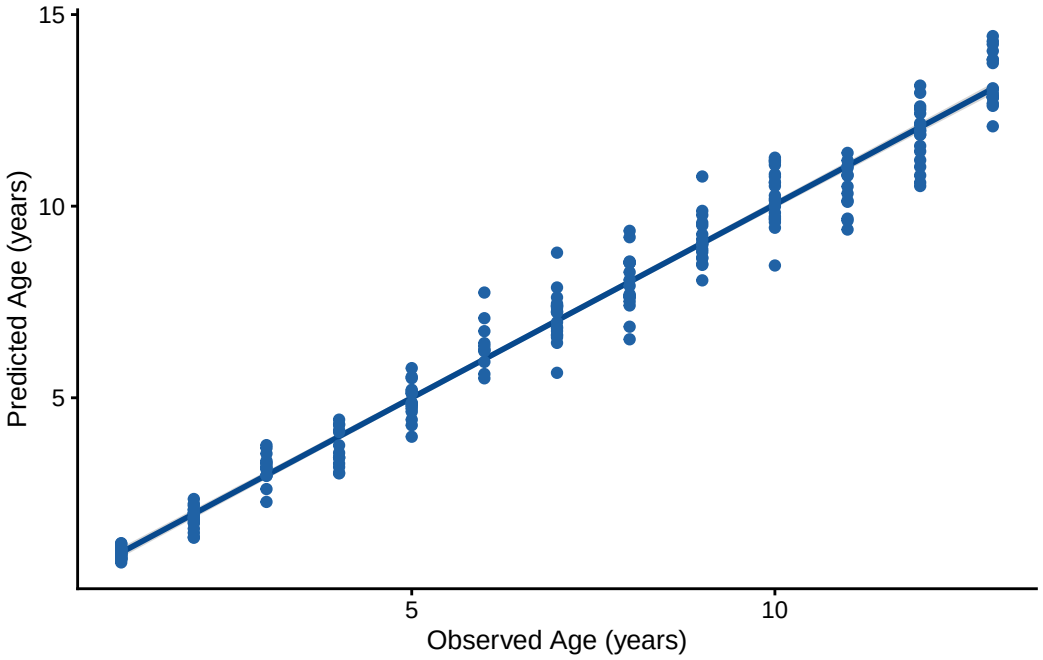


Figure 22: Predicted ages of MultiSWAG against observed reader estimates of age, using the testing data and predictions for each of our four folds.

The predictions outputted by the MultiSWAG ensemble closely match the estimated ages produced by the vertebrae readers, though it appears that the ages for 11 year olds were overestimated (Figure 22). Likewise, the reader estimates for 6 years olds appeared to be underestimated, so there was some evidence of bias in the estimates for particular age classes. Overall, the error in these predictions increases as age increases. This is potentially due to the smaller sample sizes available for older age classes.

3.4.2 Model calibration

Table 15: Empirical coverage rates of three ensemble models with 90% and 95% prediction intervals, with mean coverage rates and standard deviations across the 4-fold cross-validation procedure. The values closest to the expected coverage probabilities have been bolded and italicised, indicating the most well-calibrated model

Model	90% ± SD	95% ± SD
DeepEnsembling	82.013 (2.099)	87.994 (1.971)
MoLA	86.923 (2.394)	92.701 (1.902)
MultiSWAG	***87.001 (1.980)***	***92.898 (1.916)***

MultiSWAG was the most well-calibrated method, while DeepEnsembling was the least (Table 15). However, the prediction intervals were frequently overly narrow for all three methods, leading to the under-coverage of unseen testing data when compared with the expected coverage probabilities. Nevertheless, both MoLA and MultiSWAG lead to some improvement in model calibration, with coverage rates that were closest to the expected coverage probabilities.

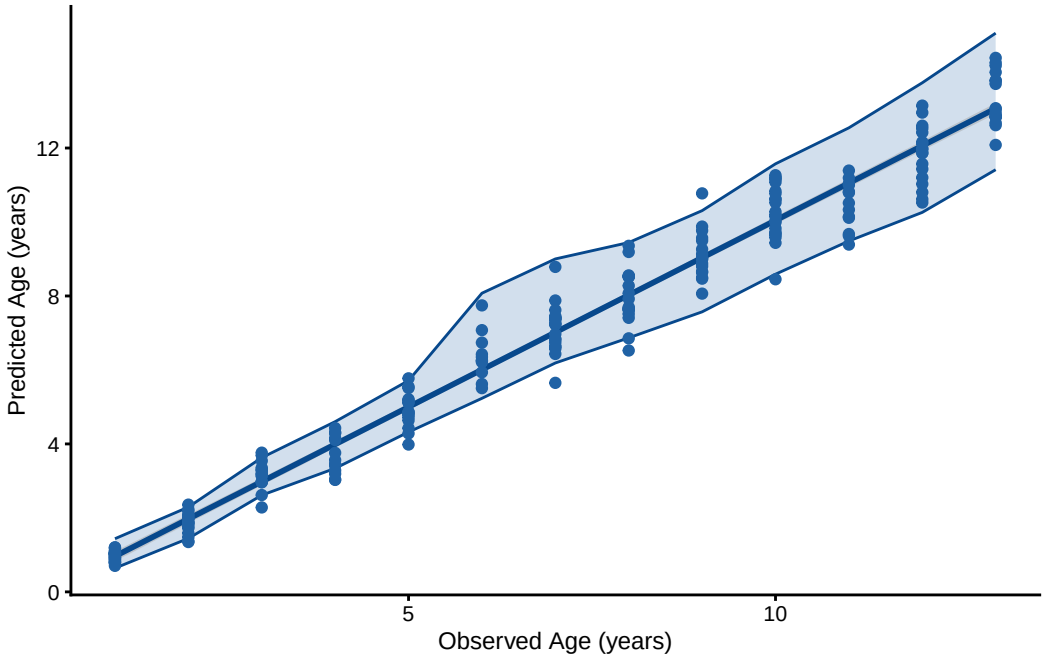


Figure 23: 90% prediction intervals for the MultiSWAG ensemble, with the combined testing data and prediction intervals from the 4 folds.

The prediction intervals tended to be wider for older age estimates and narrower for younger ones (Figure 23). There were a greater proportion of observations that fell below the lower-bound of the prediction intervals (74%), as opposed to the upper bound. Hence, the MultiSWAG model seems more likely to overestimate reader ages than underestimate them.

3.4.3 Model interpretation

Darker colours appear mostly along the bow tie sections of each vertebrae, suggesting that they contribute heavily to the predictions made by the CNNs (Figure 24). Despite this, it is not clear that the vertebral bands themselves are the principal contributors to the predicted ages produced by these models. Assuming that the models were directly counting the vertebral bands within each image, we would expect to see a banding pattern

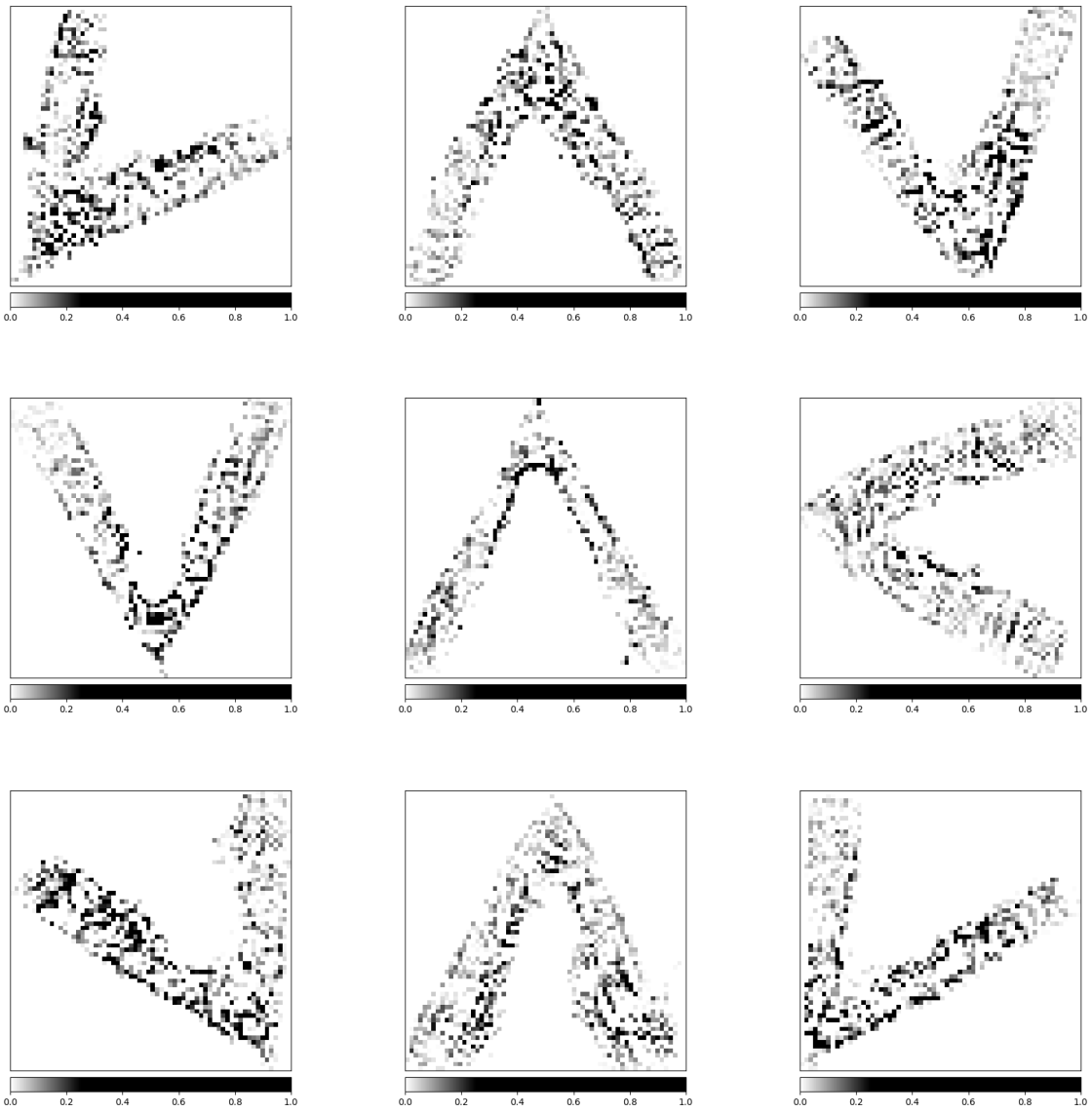


Figure 24: Feature importances for nine vertebrae images. Darker pixel colours represent greater contributions to the predictions made by the model.

highlighted within each section. Hence, we cannot definitively claim that these CNNs are directly counting the bands contained within each vertebrae image.

3.5 Discussion

This study is the first of its kind to demonstrate the ability of CNNs to predict ages from images of chondrichthyan vertebrae. Using a dataset of *Isurus oxyrinchus* images (Bishop et al., 2006), we trained several ResNet-50 networks and used Bayesian estimation, ensembling, and four-fold cross-validation to minimise over-fitting and quantify the uncertainty in our predictions. Our models were able to reproduce human-read estimates of age with a mean square error of 2.7 to 2.8 years.

We compared three deep learning methods: DeepEnsembling; MoLA; and MultiSWAG, and found that MultiSWAG offered marginally greater predictive performance and well-calibrated coverage rates relative to its two alternatives. Hence, MultiSWAG may be useful for providing effective age estimation with uncertainty quantification for future attempts at automated vertebrae reading. However, the Bayesian deep learning methods used here were chosen due to their relative speed and scalability; other fitting methods, such as MCMC, stochastic MCMC, variational Bayes, or Monte Carlo dropout, might prove to out-perform MultiSWAG in future. Indeed, the "gold-standard" performance of MCMC is of particular note due to its asymptotically exact inference (Neiswanger et al., 2013; Nemeth & Fearnhead, 2021; Roberts & Rosenthal, 2004).

The integrated gradients method showed that the bow tie section of vertebrae images were the principal feature identified by the models. This is consistent with the behaviour of human readers, who primarily use counts of bands from the bow tie section to estimate the age of sharks (Carrier et al., 2018; Harry et al., 2022; Natanson et al., 2018). Despite this,

we could not conclusively state that our models were actually counting the vertebral bands depicted within our images. For instance, it it known that CNNs can be biased towards texture information (Hermann et al., 2020), so the texture of the bow tie sections may have been influencing the model predictions. Further work to refine our models towards identifying shape features, such as the shape of vertebral band zones, may be required to facilitate the automated counting of vertebral bands.

The small sample sizes of chondrichthyan datasets remains a concern for age and growth studies (Carrier et al., 2018; Harry et al., 2022), in spite of our use of offline augmentation. Due to overparameterisation, deep learning methods typically require large datasets to avoid overfitting (Buslaev et al., 2020; Shorten & Khoshgoftaar, 2019). Further investigation with a larger dataset composed of more age classes may lead to superior model performance, and further effort will be required to produce such a large database of vertebrae images. However, the self-supervised CLIP architecture has demonstrated state-of-the-art results for few-shot learning (Sigurardóttir et al., 2023), and may provide a prudent avenue for training models on smaller datasets in cases where it is infeasible to secure more data.

Domain adaptation methods may be useful for training models across different datasets (Ordoñez et al., 2022), such that we can apply automated vertebral reading to *Isurus oxyrinchus* samples stored at different vertebrae image laboratories. Domain adaptation may also allow our models to generalise to different species of chondrichthyans, particularly those where data are more scarce, as is the case with deepwater species (Finucci et al., 2019). However, the periodicity of band deposition varies by species, as does the correspondence between band deposition and time; for instance, band deposition occurs annually in the case of *Isurus oxyrinchus*, which may not be the case for other species. Further work must be taken to ensure that this inter-species variation in periodicity and correspondence is

accounted for in our models.

Alternative aging methods that are not reliant on growth band counts have been enabled by new technology, such as near-infrared spectroscopy, eye lens dating, and genomics techniques (Carrier et al., 2018; Harry et al., 2022). Rigby et al. (2014) investigated the use of near-infrared spectroscopy for the aging of deep water chondrichthyans, while Nehmens et al. (2021) investigated the potential for the genetic analysis of telomeres to be used for the age determination of the deep water *Etmopterus granulosus*. Likewise, eye lens radiocarbon analysis has been applied for the age estimation of *Somniosus microcephalus* (Nielsen et al., 2016) and *Centroscyllium fabricii* (Hedeholm et al., 2021). Indeed, the need for validated aging data has been extensively discussed within the literature (Cailliet et al., 2001; Cailliet et al., 2006b; G. Cailliet & Goldman, 2004; Carrier et al., 2018; Goldman et al., 2012; Harry et al., 2022).

Multi-task and multi-modal techniques may be key for integrating different types of data (images, spectroscopic measurements, radiocarbon-validated eye lens data, and genomics datasets) and different features (length, weight, sex, and species) into a single model. Robertson et al. (2001) had previously found that the integration of additional geometric features such as body length lead to greater predictive performance when automating the reading of otoliths. Indeed, there could be value in exploring whether the same would apply to vertebrae. Moreover, multi-task and multi-modal learning would enable deep learning models to not only perform age estimation, but also predict different attributes of interest, such as length, weight, or species.

A key motivation for this study was to determine if approaches and models that had been applied successfully to automating otolith interpretation, such as CNNs, could find similar success with vertebrae. In a relevant study on automated otolith interpretation, Moen et

al. (2018) highlight cost savings as a primary incentive for the exploration of automated aging algorithms, further noting that the principal cost-driver for otolith reading is not the reading process itself, but the preparation procedure performed prior to the actual reading. Under that argument, the authors state that any cost savings of an automated approach would be marginal. Similarly, Fisher and Hunter (2018) conclude that digital image analysis provided little improvement in cost-effectiveness over a manual approach to otolith reading, though CNNs were not included in their study. However, Moen et al. (2018) instead argued that skilled otolith readers require many years of training, which should be incorporated into any cost calculations. Under that argument, a trained CNN could "emulate" the behaviour of a human otolith reader, and fisheries sampling programmes could be scaled up without needing to train any additional readers (Moen et al., 2018).

Comparable arguments could be made for vertebrae readers. Even if the principal cost-driver of vertebrae reading is the preparation procedure, the ability to emulate experienced vertebrae readers may lead to efficiency gains, greater scalability, and decreased costs. An additional compounding factor is the greater difficulty in estimating the ages of chondrichthyans using vertebral band counts, relative to interpreting bony fish otolith zones (Carrier et al., 2018; Harry et al., 2022). Hence, the ability to mimic the behaviour of an experienced vertebrae reader is particularly invaluable. As the technology behind automated vertebrae reading matures, a comparable study to that of Fisher and Hunter (2018) will no doubt be warranted.

Our work offers new insights into the feasibility of automated vertebrae reading of chondrichthyans. We believe that our initial findings for *Isurus oxyrinchus* demonstrate promise. However, further refinement and investigation is required to determine if we can automate the process of vertebral band counting for a variety of species and datasets; several poten-

tial avenues of investigation have been suggested within this work. We are optimistic that the positive results found for automated otolith reading can equally be matched for shark vertebrae.

4 General discussion

4.1 Discussion

One third of chondrichthyan species are threatened by extinction due to overfishing (Dulvy et al., 2021). Life-history traits such as slow growth, late age of sexual maturity, long reproduction cycles, and low fecundity (Jorgensen et al., 2022; Pacoureau et al., 2021; Stevens et al., 2000) make chondrichthyans particularly vulnerable to overfishing. Eleven species are managed under the New Zealand Quota Management System (QMS) (Ford et al., 2018), though roughly 80% of New Zealand chondrichthyans have no species-specific management or monitoring, resulting in uncertainty with respect to their population trends over time (Finucci et al., 2019).

This thesis aimed to advance Bayesian methods for estimating the growth parameters of sharks, thereby leading to more robust management and conservation of chondrichthyes against factors such as overfishing. To accomplish this, we fitted and evaluated monophasic and biphasic growth models, and explored the application of deep learning for automated age estimation via vertebrae reading.

4.1.1 Fisheries growth models

Chapter 2 explored the potential of biphasic growth models on five species of New Zealand chondrichthyes, by comparing the performance of both biphasic and monophasic models using the Pareto-smoothed importance sampling approximation of leave-one-out cross-validation (PSIS-LOO) metric (Doll & Jacquemin, 2019; Vehtari et al., 2015, 2017). We found that biphasic growth models provided superior statistical fit for both males and females in the majority of cases, with only a single case where a monophasic model gave

greater performance. We were able to improve upon growth estimates from previous studies where parameter estimates were noted to be biased or poor (Bishop et al., 2006; Francis & Francis, 1992) by offering smaller values for previously overestimated parameters and larger values for previously underestimated parameters. Our overall results were consistent with previous studies on the biphasic growth of sharks (Acuña et al., 2001; Araya & Cubillos, 2006; Aversa, Dans, Garcia, et al., 2011; Contreras-Reyes et al., 2018; Olmeda-de la Fuente et al., 2022), and offered evidence that the popular monophasic von Bertalanffy growth model (VBGM) should not be chosen *a priori* as the only candidate model to describe the growth of chondrichthyans. Rather, we should carefully consider alternative growth models, as informed by both statistical evidence (Haddon, 2010) and the biological knowledge of domain experts (Y.-G. Wang et al., 1995).

4.1.2 Age estimation via vertebral reading

In Chapter 3, our objective was to explore the feasibility of automated age estimation for chondrichthyans by using deep learning to "emulate" the behaviour of trained vertebrae readers. We stated that automated otolith interpretation methods had a notably richer history within the literature (Carbini et al., 2008), and compared and contrasted the challenges of vertebrae reading with otolith reading. We applied convolutional neural networks (CNNs) (O'Shea & Nash, 2015) to learn the most discriminative features present within an image dataset of *Isurus oxyrinchus* vertebrae. We further evaluated three different deep learning methods in terms of predictive accuracy and model calibration, by comparing their mean squared error (MSE), Gaussian negative log-likelihood (NLL), and coverage error scores. We found that the multi-stochastic weight average Gaussian (MultiSWAG) method (A. G. Wilson & Izmailov, 2020) offered marginally greater predictive power and well-calibrated prediction intervals, when compared to DeepEnsembling (Lakshminarayanan et al., 2017)

and mixture of Laplace approximations (MoLA) (Eschenhagen et al., 2021). To the best of our knowledge, this study is the first of its kind to demonstrate the ability of CNNs to predict ages from images of chondrichthyan vertebrae. We argue that our initial results offered promise for emulating the behaviour of vertebrae readers. However, we believe that further refinement is required for future studies. By using the integrated gradients method, we could not conclusively state that our CNNs were directly counting the bands of our shark vertebrae, and noted that additional calibration may be needed to achieve this desirable behaviour. Moreover, the small sample sizes of chondrichthyan datasets remains a concern for age and growth studies (Carrier et al., 2018; Harry et al., 2022), as deep learning models generally require large datasets to avoid overfitting (Buslaev et al., 2020; Shorten & Khoshgoftaar, 2019). Future work with a larger dataset composed of a broader range of age classes may lead to greater model performance, and further effort will be needed to collate a larger database of vertebrae images.

4.1.3 Bayesian modelling

Our findings demonstrated the capability of Bayesian models to offer a principled approach to uncertainty quantification and parameter estimation within the context of age and growth modelling. This is particularly beneficial in situations where available data is limited, as is the case with deepwater species (Finucci et al., 2019). In the case of Chapter 2, we were able to incorporate prior knowledge concerning the growth parameters of interest from both domain experts and the previous literature. This allowed us to more easily quantify the epistemic uncertainty of our parameter estimates through interpretable probability distributions, leading to a more robust estimation of growth parameters and superior model fit in several cases. For Chapter 3, the application of Bayesian deep learning enabled us to quantify both the epistemic and aleotoric uncertainty of our age estimates, while offer-

ing marginally superior predictive performance and calibration against the non-Bayesian DeepEnsembling method. Overall, we provided promising initial results that Bayesian deep learning could be used to efficiently automate the counting of vertebral bands for age estimation, while quantifying the uncertainty inherent to the reading process.

This work contributes to our knowledge and understanding of chondrichthyan life history, while improving methods used to measure and interpret these parameters. Knowledge gained in this thesis can be directly applied and incorporated to chondrichthyan conservation and management applications, including stock assessments. We hope this work contributes to greater management and persistence of these species for future generations.

References

Abascal, F. J., Quintans, M., Ramos-Cartelle, A., & Mejuto, J. (2011). Movements and environmental preferences of the shortfin mako, isurus oxyrinchus, in the southeastern pacific ocean. *Marine Biology*, *158*, 1175–1184.

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, *76*, 243–297.

Abe, T., Buchanan, E. K., Pleiss, G., Zemel, R., & Cunningham, J. P. (2022). Deep ensembles work, but are they necessary? *Advances in Neural Information Processing Systems*, *35*, 33646–33660.

Acuña, E., Cid, L., Pérez, E., Kong, I., Araya, M., Lamilla, J., Peñailillo, J., Bennett, X., González, M., Villarroel, J., et al. (2001). Estudio biológico de tiburones (marrajo dentudo, azulejo y tiburón sardinero) en la zona norte y central de chile. *Informes Técnicos FIP, FIP/IT*, *2000-23*, 1–112.

Alarab, I., Prakoonwit, S., & Nacer, M. I. (2021). Illustrative discussion of mc-dropout in general dataset: Uncertainty estimation in bitcoin. *Neural Processing Letters*, *53*, 1001–1011.

Allison, E. H., Perry, A. L., Badjeck, M.-C., Neil Adger, W., Brown, K., Conway, D., Halls, A. S., Pilling, G. M., Reynolds, J. D., Andrew, N. L., et al. (2009). Vulnerability of national economies to the impacts of climate change on fisheries. *Fish and Fisheries*, *10*(2), 173–196.

Alós, J., Palmer, M., Balle, S., Grau, A. M., & Morales-Nin, B. (2010). Individual growth pattern and variability in serranus scriba: A bayesian analysis. *ICES Journal of Marine*

Science, *67*(3), 502–512.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*, 1–74.

Amari, S. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, *5*(4-5), 185–196.

Araya, M., & Cubillos, L. A. (2006). Evidence of two-phase growth in elasmobranchs. *Environ Biol Fish*, *77*(3-4), 293–300. https://doi.org/10.1007/s10641-006-9110-8

Ardizzone, D., Cailliet, G. M., Natanson, L. J., Andrews, A. H., Kerr, L. A., & Brown, T. A. (2006). Application of bomb radiocarbon chronologies to shortfin mako (Isurus oxyrinchus) age validation. *Environ Biol Fish*, *77*(3-4), 355–366. https://doi.org/10.1007/s10641-006-9106-4

Aversa, M. I., Dans, S. L., Garcia, N. A., & Crespo, E. A. (2011). Growth models fitted to dipturus chilensis length-at-age data support a two-phase growth. *Revista Chilena de Historia Natural*, *84*(1), 33–49.

Aversa, M. I., Dans, S. L., García, N. A., & Crespo, E. A. (2011). Growth models fitted to Dipturus chilensis length-at-age-data support a two phase growth. *Rev. Chil. Hist. Nat.*, *84*(1), 33–49. https://doi.org/10.4067/S0716-078X2011000100003

Bermejo, S., Monegal, B., & Cabestany, J. (2007). Fish age categorization from otolith images using multi-class support vector machines. *Fisheries Research*, *84*(2), 247–253.

Betancourt, M. (2015). The fundamental incompatibility of scalable hamiltonian monte carlo and naive data subsampling. *International Conference on Machine Learning*, 533–540.

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv Preprint arXiv:1701.02434*.

Beverton, R. (1957). On the dynamics of exploited fish populations. *Fish. Invest. Lond. Ser.*, *2*, 19.

Bishop, S., Francis, M., Duffy, C., & Montgomery, J. (2006). Age, growth, maturity, longevity and natural mortality of the shortfin mako shark (isurus oxyrinchus) in new zealand waters. *Marine and Freshwater Research*, *57*(2), 143–154.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877.

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural network. *International Conference on Machine Learning*, 1613–1622.

Boomer, J. J., Harcourt, R. G., Francis, M., Walker, T. I., Braccini, J. M., & Stow, A. J. (2013). Frequency of multiple paternity in gummy shark, mustelus antarcticus, and rig, mustelus lenticulatus, and the implications of mate encounter rate, postcopulatory influences, and reproductive mode. *Journal of Heredity*, *104*(3), 371–379.

Botev, A., Ritter, H., & Barber, D. (2017). Practical gauss-newton optimisation for deep learning. *International Conference on Machine Learning*, 557–565.

Bowlby, H. D., & Gibson, A. J. F. (2020). Implications of life history uncertainty when evaluating status in the northwest atlantic population of white shark (carcharodon carcharias). *Ecology and Evolution*, *10*(11), 4990–5000.

Bradley, D., & Gaines, S. D. (2014). Counting the cost of overfishing on sharks and rays. *Elife*, *3*, e02199.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, *11*(2), 125.

Cailliet, G. M., Andrews, A. H., Burton, E. J., Watters, D. L., Kline, D. E., & Ferry-Graham, L. A. (2001). Age determination and validation studies of marine fishes: Do deep-dwellers live longer? *Experimental Gerontology*, *36*(4-6), 739–764. https://doi.org/10.1016/S0531-5565(00)00239-4

Cailliet, G. M., Smith, W. D., Mollet, H. F., & Goldman, K. J. (2006a). Age and growth studies of chondrichthyan fishes: The need for consistency in terminology, verification, validation, and growth function fitting. *Environ Biol Fish*, *77*(3-4), 211–228. https://doi.org/10.1007/s10641-006-9105-5

Cailliet, G. M., Smith, W. D., Mollet, H. F., & Goldman, K. J. (2006b). Age and growth studies of chondrichthyan fishes: The need for consistency in terminology, verification, validation, and growth function fitting. *Environ Biol Fish*, *77*(3-4), 211–228. https://doi.org/10.1007/s10641-006-9105-5

Cailliet, G. M., Smith, W. D., Mollet, H. F., & Goldman, K. J. (2006c). Age and growth studies of chondrichthyan fishes: The need for consistency in terminology, verification, validation, and growth function fitting. *Environmental Biology of Fishes*, *77*, 211–228.

Cailliet, G., & Goldman, K. (2004). Age determination and validation in chondrichthyan fishes. In "biology of sharks and their relatives."(eds JC carrier, JA musick and MR heithaus.) Pp. 399–447. *CRC Press: Boca Raton, FL*, *10*, 9780203491317.

Cailliet, G., Smith, W., Mollet, H., & Goldman, K. (2006). Chondrichthyan growth studies: An updated review, stressing terminology, sample size sufficiency, validation, and curve fitting. *Special Volume from Symposium of the American Elasmobranch Society*, 000–

000.

Campana, S. E. (1992). Measurement and interpretation of the microstructure of fish otoliths. *Canadian Special Publication of Fisheries and Aquatic Sciences*, *117*(1), 59–71.

Carbini, S., Chessel, A., Benzinou, A., Fablet, R., Mahé, K., & De Pontual, H. (2008). A review of image-based tools for automatic fish ageing from otolith features. *Approche Systémique Des pêches, 5-7 Septembre 2008, Boulogne-Sur-Mer*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1).

Carreira-Perpinan, M. A. (2000). Mode-finding for mixtures of gaussian distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1318–1323.

Carrera-Fernández, M., Galván-Magaña, F., & Ceballos-Vázquez, B. P. (2010). Reproductive biology of the blue shark prionace glauca (chondrichthyes: Carcharhinidae) off baja california sur, méxico. *Aqua*, *16*(3), 101–110.

Carrier, J. C., Heithaus, M. R., & Simpfendorfer, C. A. (2018). *Shark research: Emerging technologies and applications for the field and laboratory*.

Carrier, J. C., Musick, J. A., & Heithaus, M. R. (2012). *Biology of sharks and their relatives*. CRC press.

Cerna, F., & Licandeo, R. (2009). Age and growth of the shortfin mako (isurus oxyrinchus) in the south-eastern pacific off chile. *Marine and Freshwater Research*, *60*(5), 394–403.

Chen, S., & Watanabe, S. (1989). Age dependence of natural mortality coefficient in fish population dynamics. *Nippon Suisan Gakkaishi*, *55*(2), 205–208.

Chen, T., Fox, E., & Guestrin, C. (2014). Stochastic gradient hamiltonian monte carlo.

International Conference on Machine Learning, 1683–1691.

Chidlow, J. A., Simpfendorfer, C. A., & Russ, G. R. (2007). Variable growth band deposition leads to age and growth uncertainty in the western wobbegong shark, orectolobus hutchinsi. *Marine and Freshwater Research*, *58*(9), 856–865.

Clarke, M., Connolly, P., & Bracken, J. (2002). Age estimation of the exploited deepwater shark centrophorus squamosus from the continental slopes of the rockall trough and porcupine bank. *Journal of Fish Biology*, *60*(3), 501–514.

Cobb, A. D., & Jalaian, B. (2021). Scaling hamiltonian monte carlo inference for bayesian neural networks with symmetric splitting. *Uncertainty in Artificial Intelligence*, 675–685.

Compagno, L. J. (2001). *Sharks of the world: An annotated and illustrated catalogue of shark species known to date* (Vol. 2). Food & Agriculture Org.

Condrey, R., Beckman, D., & Wilson, C. (1988). Management implications of a new growth model for red drum. Appendix d. *US Dept. Commerce Cooperative Agreement NA87-WC-H-06122. Marine Fisheries Initiative (MARFIN) Program. Louisiana Department of Wildlife and Fisheries, Seafood Division, Finfish Section, Baton Rouge, Louisiana.*

Contreras-Reyes, J. E., Quintero, F. O. L., & Wiff, R. (2018). Bayesian modeling of individual growth variability using back-calculation: Application to pink cusk-eel (genypterus blacodes) off chile. *Ecological Modelling*, *385*, 145–153.

Contreras-Reyes, J. E., Wiff, R., Soto, J., Donovan, C. R., & Araya, M. (2021). Biphasic growth modelling in elasmobranchs based on asymmetric and heavy-tailed errors. *Environ Biol Fish*, *104*(5), 615–628. https://doi.org/10.1007/s10641-021-01100-z

Cortés, E. (2000). Life history patterns and correlations in sharks. *Reviews in Fisheries Science*, *8*(4), 299–344.

Cotton, C. F., Grubbs, R. D., Dyb, J. E., Fossen, I., & Musick, J. A. (2015). Reproduction and embryonic development in two species of squaliform sharks, centrophorus granulosus and etmopterus princeps: Evidence of matrotrophy? *Deep Sea Research Part II: Topical Studies in Oceanography*, *115*, 41–54.

Cox, G. J., & Francis, M. (1997). *Sharks and rays of new zealand*. Canterbury University Press.

Craig, P., Choat, J., Axe, L., & Saucerman, S. (1998). Population biology and harvest of the coral reef surgeonfish acanthurus lineatus in american samoa. *Oceanographic Literature Review*, *2*(45), 407.

Cranmer, M., Tamayo, D., Rein, H., Battaglia, P., Hadden, S., Armitage, P. J., Ho, S., & Spergel, D. N. (2021). A bayesian neural network predicts the dissolution of compact planetary systems. *Proceedings of the National Academy of Sciences*, *118*(40), e2026053118.

Daxberger, E., Kristiadi, A., Immer, A., Eschenhagen, R., Bauer, M., & Hennig, P. (2021). Laplace redux-effortless bayesian deep learning. *Advances in Neural Information Processing Systems*, *34*, 20089–20103.

Day, T., & Taylor, P. D. (1997). Von bertalanffy's growth equation should not be used to model age and size at maturity. *The American Naturalist*, *149*(2), 381–393.

Dell'Apa, A., Pennino, M. G., Bangley, C. W., & Bonzek, C. (2018). A hierarchical bayesian modeling approach for the habitat distribution of smooth dogfish by sex and season in inshore coastal waters of the US northwest atlantic. *Marine and Coastal Fisheries*, *10*(6), 590–605.

Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, *31*(2), 105–112.

DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv Preprint arXiv:1708.04552*.

Doll, J. C., & Jacquemin, S. J. (2019). Bayesian model selection in fisheries management and ecology. *Journal of Fish and Wildlife Management*, *10*(2), 691–707.

Dulvy, N. K., Baum, J. K., Clarke, S., Compagno, L. J., Cortés, E., Domingo, A., Fordham, S., Fowler, S., Francis, M., Gibson, C., et al. (2008). You can swim but you can't hide: The global status and conservation of oceanic pelagic sharks and rays. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *18*(5), 459–482.

Dulvy, N. K., Pacoureau, N., Rigby, C. L., Pollom, R. A., Jabado, R. W., Ebert, D. A., Finucci, B., Pollock, C. M., Cheok, J., Derrick, D. H., et al. (2021). Overfishing drives over one-third of all sharks and rays toward a global extinction crisis. *Current Biology*, *31*(21), 4773–4787.

Ebert, D. A., Fowler, S. L., & Compagno, L. J. (2013). *Sharks of the world: A fully illustrated guide*. Wild Nature Press.

Eschenhagen, R., Daxberger, E., Hennig, P., & Kristiadi, A. (2021). Mixtures of laplace approximations for improved post-hoc uncertainty in deep learning. *arXiv Preprint arXiv:2111.03577*.

Fablet, R. (2006). Semi-local extraction of ring structures in images of biological hard tissues: Application to the bayesian interpretation of fish otoliths for age and growth estimation. *Canadian Journal of Fisheries and Aquatic Sciences*, *63*(6), 1414–1428.

Fablet, R., & Le Josse, N. (2005). Automated fish age estimation from otolith images using statistical learning. *Fisheries Research*, *72*(2-3), 279–290.

Finotto, L., Walker, T. I., & Reina, R. D. (2023). Influence of female reproductive state and of fishing-capture stress on the oxygen uptake rate of a viviparous elasmobranch.

Journal of Experimental Zoology Part A: Ecological and Integrative Physiology.

Finucci, B., Bineesh, K., Cheok, J., Cotton, C., Dharmadi, K., DW, N., FC, P., N, R., CL, T., S, W., & TI. (2020). N., Rigby, CL, tanaka, s. & walker, TI 2020. *Centrophorus Squamosus. The IUCN Red List of Threatened Species*.

Finucci, B., Duffy, C. A., Francis, M., Gibson, C., & Kyne, P. M. (2019). The extinction risk of new zealand chondrichthyans. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*(5), 783–797.

Fisher, M., & Hunter, E. (2018). Digital imaging techniques in otolith data capture, analysis and interpretation. *Marine Ecology Progress Series*, *598*, 213–231.

Flinn, S. A., & Midway, S. R. (2021a). Trends in Growth Modeling in Fisheries Science. *Fishes*, *6*(1), 1. https://doi.org/10.3390/fishes6010001

Flinn, S. A., & Midway, S. R. (2021b). Trends in growth modeling in fisheries science. *Fishes*, *6*(1), 1.

Folgoc, L. L., Baltatzis, V., Desai, S., Devaraj, A., Ellis, S., Manzanera, O. E. M., Nair, A., Qiu, H., Schnabel, J., & Glocker, B. (2021). Is MC dropout bayesian? *arXiv Preprint arXiv:2110.04286*.

Ford, R., Francis, M., Holland, L., Clark, M., Duffy, C., Dunn, M., Jones, E., & Wells, R. (2018). Qualitative (level 1) risk assessment of the impact of commercial fishing on new zealand chondrichthyans: An update for 2017. *New Zealand Aquatic Environment and Biodiversity Report*, *201*, 103.

Foreman, S., Jin, X.-Y., & Osborn, J. C. (2021). Deep learning hamiltonian monte carlo. *arXiv Preprint arXiv:2105.03418*.

Fort, S., Hu, H., & Lakshminarayanan, B. (2019). Deep ensembles: A loss landscape perspective. *arXiv Preprint arXiv:1912.02757*.

Francis, M. (1988). The new zealand rig fishery: Catch statistics and composition, 1974-85. *New Zealand Fisheries Technical Report (New Zealand)*.

Francis, M. (2013). Temporal and spatial patterns of habitat use by juveniles of a small coastal shark (mustelus lenticulatus) in an estuarine nursery. *PloS One*, *8*(2), e57021.

Francis, M. (2015b). Size, maturity and age composition of porbeagle sharks observed in new zealand tuna longline fisheries. *New Zealand Fisheries Assessment Report*, *16*, 30.

Francis, M. (2015a). *Size, maturity and age composition of porbeagle sharks observed in New Zealand tuna longline fisheries*. Ministry for Primary Industries.

Francis, M. (2016). *Size, maturity and age composition of mako sharks observed in New Zealand tuna longline fisheries*. Ministry for Primary Industries, Manatū Ahu Matua.

Francis, M., Campana, S., & Jones, C. (2007a). Age under-estimation in New Zealand porbeagle sharks (Lamna nasus): Is there an upper limit to ages that can be determined from shark vertebrae? *Mar. Freshwater Res.*, *58*(1), 10. https://doi.org/10.1071/MF06069

Francis, M., Campana, S., & Jones, C. (2007b). Age under-estimation in New Zealand porbeagle sharks (Lamna nasus): Is there an upper limit to ages that can be determined from shark vertebrae? *Mar. Freshwater Res.*, *58*(1), 10. https://doi.org/10.1071/MF06069

Francis, M., & Duffy, C. (2005). *Length at maturity in three pelagic sharks (lamna nasus, isurus oxyrinchus, and prionace glauca) from new zealand*.

Francis, M., & Francis, R. (1992). Growth rate estimates for new zealand rig (mustelus lenticulatus). *Marine and Freshwater Research*, *43*(5), 1157–1176.

Francis, M., Griggs, L. H., & Baird, S. J. (2001). Pelagic shark bycatch in the new zealand tuna longline fishery. *Marine and Freshwater Research*, *52*(2), 165–178.

Francis, M., Lyon, W., Jones, E., Notman, P., Parkinson, D., & Getzlaff, C. (2012). Rig nursery grounds in new zealand: A review and survey. *New Zealand Aquatic Environment and Biodiversity Report No*, *95*, 50.

Francis, M., & Mace, J. T. (1980). Reproductive biology of mustelus lenticulatus from kaikoura and nelson. *New Zealand Journal of Marine and Freshwater Research*, *14*(3), 303–311.

Francis, M., & Maolagáin, C. (2016). *Size, maturity and length composition of blue sharks observed in new zealand tuna longline fisheries*. Ministry for Primary Industries.

Francis, M., & Maolagáin, C. Ó. (2000). Age, growth and maturity of a new zealand endemic shark (mustelus lenticulatus) estimated from vertebral bands. *Marine and Freshwater Research*, *51*(1), 35–42.

Francis, M., Natanson, L., & Campana, S. (2008). The biology and ecology of the porbeagle shark, lamna nasus. *Sharks of the Open Ocean: Biology, Fisheries and Conservation*, 105–113.

Francis, M., & Ó Maolagáin, C. (2016). *Size, maturity and length composition of blue sharks observed in New Zealand tuna longline fisheries*. Ministry for Primary Industries.

Frisk, M. G., Miller, T. J., & Fogarty, M. J. (2001). Estimation and analysis of biological parameters in elasmobranch fishes: A comparative life history study. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(5), 969–981.

Gal, Y. (2016). *Uncertainty in deep learning*.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *International Conference on Machine Learning*, 1050–1059.

Gal, Y., & Ghahramani, Z. (2015a). On modern deep learning and variational inference.

Advances in Approximate Bayesian Inference Workshop, NIPS, *2*.

Gal, Y., & Ghahramani, Z. (2015b). Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv Preprint arXiv:1506.02158*.

Ganaie, M. A., Hu, M., Malik, A., Tanveer, M., & Suganthan, P. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151.

Gazak, J. Z., McQuaid, I., Swindle, R., Phelps, M., & Fletcher, J. (2022). SpectraNet: Learned recognition of artificial satellites from high contrast spectroscopic imagery. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 4012–4020.

Gedraite, E. S., & Hadad, M. (2011). Investigation on the effect of a gaussian blur in image filtering and segmentation. *Proceedings ELMAR-2011*, 393–396.

Gelfand, A. (1996). *Model determination using sampling-based methods. Markov chain monte carlo in practice. WR gilks, s. Richardson and DJ spiegelhalter*. Chapman & Hall/CRC Interdisciplinary Statistics.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Goldman, K. J., Cailliet, G. M., Andrews, A. H., & Natanson, L. J. (2012). *Assessing the Age and Growth of Chondrichthyan Fishes*.

Graves, A. (2011). Practical variational inference for neural networks. *Advances in Neural Information Processing Systems*, *24*.

Grenander, U., & Miller, M. I. (1994). Representations of knowledge in complex systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(4), 549–581.

Griggs, L., & Baird, S. (2013). Fish bycatch in new zealand tuna longline fisheries 2006–07 to 2009–10. *New Zealand Fisheries Assessment Report*, *13*, 73.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48.

Gürbüzbalaban, M., Gao, X., Hu, Y., & Zhu, L. (2021). Decentralized stochastic gradient langevin dynamics and hamiltonian monte carlo. *The Journal of Machine Learning Research*, *22*(1), 10804–10872.

Haddad, Y. Y. (2020). *Probabilistic deep learning on spheres for weather/climate applications*.

Haddon, M. (2010). *Modelling and quantitative methods in fisheries*. CRC press.

Harry, A. V. (2018). Evidence for systemic age underestimation in shark and ray ageing studies. *Fish Fish*, *19*(2), 185–200. https://doi.org/10.1111/faf.12243

Harry, A. V., Smart, J. J., & Pardo, S. A. (2022). Understanding the age and growth of chondrichthyan fishes. *Biology of Sharks and Their Relatives*, 177–202.

Harry, A. V., Tobin, A. J., & Simpfendorfer, C. A. (2013). Age, growth and reproductive biology of the spot-tail shark, carcharhinus sorrah, and the australian blacktip shark, c. Tilstoni, from the great barrier reef world heritage area, north-eastern australia. *Marine and Freshwater Research*, *64*(4), 277–293.

Hastings, W. K. (1970). *Monte carlo sampling methods using markov chains and their applications*.

Havasi, M., Hernández-Lobato, J. M., & Murillo-Fuentes, J. J. (2018). Inference in deep

gaussian processes using stochastic gradient hamiltonian monte carlo. *Advances in Neural Information Processing Systems*, *31*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Hedeholm, R., Qvist, T., Frausing, M., Olsen, J., Nielsen, J., & Grønkjær, P. (2021). Age of black dogfish (centroscyllium fabricii) estimated from fin spines growth bands and eye lens bomb radiocarbon dating. *Polar Biology*, *44*, 751–759.

Hermann, K., Chen, T., & Kornblith, S. (2020). The origins and prevalence of texture bias in convolutional neural networks. *Advances in Neural Information Processing Systems*, *33*, 19000–19015.

Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, 5–13.

Hoese, H., Beckman, D., Blanchet, R., Drullinger, D., & Nieland, D. (1991). A biological and fisheries profile of louisiana red drum sciaenops ocellatus. *Fishery Management Plan Series*, *4 Part 1*.

Hoffman, M. D. (2017). Learning deep latent gaussian models with markov chain monte carlo. *International Conference on Machine Learning*, 1510–1519.

Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, *15*(1), 1593–1623.

Hora, S. C. (1996). Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. *Reliability Engineering & System Safety*, *54*(2-3), 217–223.

Hordyk, A., Ono, K., Valencia, S., Loneragan, N., & Prince, J. (2015). A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. *ICES Journal of Marine Science*, *72*(1), 217–231.

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*, 457–506.

Huveneers, C., Stead, J., Bennett, M. B., Lee, K. A., & Harcourt, R. G. (2013). Age and growth determination of three sympatric wobbegong sharks: How reliable is growth band periodicity in orectolobidae? *Fisheries Research*, *147*, 413–425.

Hwang, D., Lee, G., Jo, H., Yoon, S., & Ryu, S. (2020). A benchmark study on reliable molecular supervised learning via bayesian learning. *arXiv Preprint arXiv:2006.07021*.

Izmailov, P., Maddox, W. J., Kirichenko, P., Garipov, T., Vetrov, D., & Wilson, A. G. (2020). Subspace inference for bayesian deep learning. *Uncertainty in Artificial Intelligence*, 1169–1179.

Izmailov, P., Podoprikhin, D., Garipov, T., Vetrov, D., & Wilson, A. G. (2018). Averaging weights leads to wider optima and better generalization. *arXiv Preprint arXiv:1803.05407*.

Jabado, R. W., Kyne, P. M., Pollom, R. A., Ebert, D. A., Simpfendorfer, C. A., Ralph, G. M., Al Dhaheri, S. S., Akhilesh, K., Ali, K., Ali, M. H., et al. (2018). Troubled waters: Threats and extinction risk of the sharks, rays and chimaeras of the arabian sea and adjacent waters. *Fish and Fisheries*, *19*(6), 1043–1062.

Jiao, Y., Cortés, E., Andrews, K., & Guo, F. (2011). Poor-data and data-poor species stock assessment using a bayesian hierarchical approach. *Ecological Applications*, *21*(7), 2691–2708.

Jiao, Y., Hayes, C., & Cortés, E. (2009). Hierarchical bayesian approach for population dynamics modelling of fish complexes without species-specific data. *ICES Journal of Marine Science*, *66*(2), 367–377.

Jorgensen, S. J., Micheli, F., White, T. D., Van Houtan, K. S., Alfaro-Shigueto, J., Andrzejaczek, S., Arnoldi, N. S., Baum, J. K., Block, B., Britten, G. L., et al. (2022). Emergent research and priorities for shark and ray conservation. *Endangered Species Research*, *47*, 171–203.

Katsanevakis, S. (2006). Modelling fish growth: Model selection, multi-model inference and model selection uncertainty. *Fisheries Research*, *81*(2-3), 229–235.

Katsanevakis, S., & Maravelias, C. D. (2008). Modelling fish growth: Multi-model inference as a better alternative to a priori using von bertalanffy equation. *Fish and Fisheries*, *9*(2), 178–187.

Kingma, D. P., Salimans, T., & Welling, M. (2015). Variational dropout and the local reparameterization trick. *Advances in Neural Information Processing Systems*, *28*.

Kinney, M. J., Wells, R. J. D., & Kohin, S. (2016). Oxytetracycline age validation of an adult shortfin mako shark *Isurus oxyrinchus* after 6 years at liberty: Oxytetracycline age validation of *isurus oxyrinchus*. *J Fish Biol*, *89*(3), 1828–1833. https://doi.org/10.1111/jfb.13044

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, *114*(13), 3521–3526.

Kneebone, J., Natanson, L. J., Andrews, A. H., & Howell, W. H. (2008). Using bomb radiocarbon analyses to validate age and growth estimates for the tiger shark, Ga-

leocerdo cuvier, in the western North Atlantic. *Mar Biol*, *154*(3), 423–434. https://doi.org/10.1007/s00227-008-0934-y

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., et al. (2020). Captum: A unified and generic model interpretability library for pytorch. *arXiv Preprint arXiv:2009.07896*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, *30*.

Laplace, P. S. (1774). Mémoire de mathématique et de physique. *Tome Sixième*.

Lappalainen, A., Saks, L., Šuštar, M., Heikinheimo, O., Jürgens, K., Kokkonen, E., Kurkilahti, M., Verliin, A., & Vetemaa, M. (2016). Length at maturity as a potential indicator of fishing pressure effects on coastal pikeperch (sander lucioperca) stocks in the northern baltic sea. *Fisheries Research*, *174*, 47–57.

Laslett, G. M., Eveson, J. P., & Polacheck, T. (2002). A flexible maximum likelihood approach for fitting growth curves to tag recapture data. *Canadian Journal of Fisheries and Aquatic Sciences*, *59*(6), 976–986.

Last, P. R., & Stevens, J. D. (2009). *Sharks and rays of australia*.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, *2*.

Lee, S., Purushwalkam, S., Cogswell, M., Crandall, D., & Batra, D. (2015). Why m heads are better than one: Training a diverse ensemble of deep networks. *arXiv Preprint*

arXiv:1511.06314.

Lester, N., Shuter, B., & Abrams, P. (2004). Interpreting the von bertalanffy model of somatic growth in fishes: The cost of reproduction. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1548), 1625–1631.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*.

Lopez Quintero, F. O., Contreras-Reyes, J. E., Wiff, R., & Arellano-Valle, R. B. (2017). Flexible bayesian analysis of the von bertalanffy growth function with the use of a log-skew-t distribution. *Fishery Bulletin*, *115*(1).

Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv Preprint arXiv:1608.03983*.

Mackay, D. J. C. (1992). *Bayesian methods for adaptive models*. California Institute of Technology.

MacNeil, M. A., Chapman, D. D., Heupel, M., Simpfendorfer, C. A., Heithaus, M., Meekan, M., Harvey, E., Goetze, J., Kiszka, J., Bond, M. E., et al. (2020). Global status and conservation potential of reef sharks. *Nature*, *583*(7818), 801–806.

Maddox, W. J., Izmailov, P., Garipov, T., Vetrov, D. P., & Wilson, A. G. (2019). A simple baseline for bayesian uncertainty in deep learning. *Advances in Neural Information Processing Systems*, *32*.

Mandt, S., Hoffman, M. D., & Blei, D. M. (2017). Stochastic gradient descent as approximate bayesian inference. *arXiv Preprint arXiv:1704.04289*.

Manning, M., & Francis, M. (2005). Age and growth of blue shark (prionace glauca) from the new zealand exclusive economic zone. *New Zealand Fisheries Assessment Report*,

26(2005), 52.

Martens, J., & Grosse, R. (2015). Optimizing neural networks with kronecker-factored approximate curvature. *International Conference on Machine Learning*, 2408–2417.

Martinsen, I., Harbitz, A., & Bianchi, F. M. (2022). Age prediction by deep learning applied to greenland halibut (reinhardtius hippoglossoides) otolith images. *Plos One*, *17*(11), e0277244.

Massey, B. R., & Francis, M. (1989). Commercial catch composition and reproductive biology of rig (mustelus lenticulatus) from pegasus bay, canterbury, new zealand. *New Zealand Journal of Marine and Freshwater Research*, *23*(1), 113–120.

McElreath, R. (2020). *Statistical rethinking: A bayesian course with examples in r and stan*. CRC press.

McFarlane, G. A., McPhie, R. P., & King, J. R. (2010). *Distribution and Life History Parameters of Elasmobranch Species in British Columbia Waters*.

Mejía-Falla, P. A., Cortés, E., Navia, A. F., & Zapata, F. A. (2014). Age and growth of the round stingray urotrygon rogersi, a particularly fast-growing and short-lived elasmobranch. *PLoS One*, *9*(4), e96077.

Minte-Vera, C. V., Maunder, M. N., Casselman, J. M., & Campana, S. E. (2016). Growth functions that incorporate the cost of reproduction. *Fisheries Research*, *180*, 31–44. https://doi.org/10.1016/j.fishres.2015.10.023

Moe, B. J. (2015). *Estimating growth and mortality in elasmobranchs: Are we doing it correctly?*

Moen, E., Handegard, N. O., Allken, V., Albert, O. T., Harbitz, A., & Malde, K. (2018). Automatic interpretation of otoliths using deep learning. *PLoS One*, *13*(12), e0204713.

Montenegro, C., & Branco, M. (2016). Bayesian state-space approach to biomass dynamic

models with skewed and heavy-tailed error distributions. *Fisheries Research*, *181*, 48–62.

Moore, B., Ámar, Z., Schimel, A., Maolagáin, C., & Hoyle, S. (2021). Development of deep learning approaches for automating age estimation of hoki and snapper. *New Zealand Fisheries Assessment Report*, *69*(2021), 38.

Moore, B., Maclaren, J., Peat, C., Anjomrouz, M., Horn, P., & Hoyle, S. (2019). Feasibility of automating otolith ageing using CT scanning and machine learning. *New Zealand Fisheries Assessment Report*, *58*, 23.

Myers, S. C., Thorsen, A., Smoliński, S., Aanestad Godiksen, J., Malde, K., & Handegard, N. O. (2020). An efficient protocol and data set for automated otolith image analysis. *Geoscience Data Journal*, *7*(1), 80–88.

Nakano, H. (1994). Age, reproduction and migration of blue shark [prionace] in the north pacific ocean. *Bulletin-National Research Institute of Far Seas Fisheries (Japan)*.

Nakano, H., & Stevens, J. D. (2008). The biology and ecology of the blue shark, prionace glauca. *Sharks of the Open Ocean: Biology, Fisheries and Conservation*, 140–151.

Natanson, L. J. (n.d.). *Preliminary investigations into the age and growth of the short-fin mako, isurus oxyrinchus, white shark, carcharodon carcharias, and thresher shark, alopias vulpinus, in the western north atlantic ocean*.

Natanson, L. J. (1993). Effect of temperature on band deposition in the little skate, raja erinacea. *Copeia*, 199–206.

Natanson, L. J., & Cailliet, G. M. (1990). Vertebral growth zone deposition in pacific angel sharks. *Copeia*, 1133–1145.

Natanson, L. J., & Deacy, B. M. (2019). Using oxytetracycline validation for confirmation of changes in vertebral band-pair deposition rates with ontogeny in sandbar sharks

(Carcharhinus plumbeus) in the western North Atlantic Ocean. *FB*, *117*(1-2), 50–58. https://doi.org/10.7755/FB.117.1.6

Natanson, L. J., Kohler, N. E., Ardizzone, D., Cailliet, G. M., Wintner, S. P., & Mollet, H. F. (2006). Validated age and growth estimates for the shortfin mako, Isurus oxyrinchus, in the North Atlantic Ocean. *Environ Biol Fish*, *77*(3-4), 367–383. https://doi.org/10.1007/s10641-006-9127-z

Natanson, L. J., Mello, J. J., & Campana, S. E. (2002). *Validated age and growth of the porbeagle shark (lamna nasus) in the western north atlantic ocean.*

Natanson, L. J., Skomal, G. B., Hoffmann, S. L., Porter, M. E., Goldman, K. J., & Serra, D. (2018). Age and growth of sharks: Do vertebral band pairs record age? *Mar. Freshwater Res.*, *69*(9), 1440. https://doi.org/10.1071/MF17279

Natanson, L. J., Wintner, S. P., Johansson, F., Piercy, A., Campbell, P., De Maddalena, A., Gulak, S. J., Human, B., Fulgosi, F. C., Ebert, D. A., et al. (2008). Ontogenetic vertebral growth patterns in the basking shark cetorhinus maximus. *Marine Ecology Progress Series*, *361*, 267–278.

Neal, R. M. (2012). *Bayesian learning for neural networks* (Vol. 118). Springer Science & Business Media.

Nehmens, M. C., Varney, R. M., Janosik, A. M., & Ebert, D. A. (2021). An exploratory study of telomere length in the deep-sea shark, etmopterus granulosus. *Frontiers in Marine Science*, *8*, 642872.

Neiswanger, W., Wang, C., & Xing, E. (2013). Asymptotically exact, embarrassingly parallel MCMC. *arXiv Preprint arXiv:1311.4780*.

Nemeth, C., & Fearnhead, P. (2021). Stochastic gradient markov chain monte carlo. *Journal of the American Statistical Association*, *116*(533), 433–450.

Nielsen, J., Hedeholm, R. B., Heinemeier, J., Bushnell, P. G., Christiansen, J. S., Olsen, J., Ramsey, C. B., Brill, R. W., Simon, M., Steffensen, K. F., & Steffensen, J. F. (2016). Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (*Somniosus microcephalus*). *Science*, *353*(6300), 702–704. https://doi.org/10.1126/science.aaf1703

Nixon, J., Dusenberry, M. W., Zhang, L., Jerfel, G., & Tran, D. (2019). Measuring calibration in deep learning. *CVPR Workshops*, *2*.

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv Preprint arXiv:1511.08458*.

Ohnishi, S., Yamakawa, T., Okamura, H., & Akamine, T. (2012). *A note on the von bertalanffy growth function concerning the allocation of surplus energy to reproduction*.

Olmeda-de la Fuente, S. E., Rodríguez-Castro, J. H., Ramírez-de León, J. A., Caballero-Rico, F. C., Rodríguez-Olmeda, J. A., & Toledano-Toledano, F. (2022). Modeling the individual growth of the bonnethead shark sphyrna tiburo of the western gulf of mexico using the multimodel approach. *Fishes*, *7*(4), 157.

Ordoñez, A., Eikvil, L., Salberg, A.-B., Harbitz, A., & Elvarsson, B. ór. (2022). Automatic fish age determination across different otolith image labs using domain adaptation. *Fishes*, *7*(2), 71.

Osawa, K., Swaroop, S., Khan, M. E. E., Jain, A., Eschenhagen, R., Turner, R. E., & Yokota, R. (2019). Practical deep learning with bayesian principles. *Advances in Neural Information Processing Systems*, *32*.

Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J., Lakshminarayanan, B., & Snoek, J. (2019). Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. *Advances in Neural Information Processing Systems*, *32*.

Pacoureau, N., Rigby, C. L., Kyne, P. M., Sherley, R. B., Winker, H., Carlson, J. K., Fordham, S. V., Barreto, R., Fernando, D., Francis, M., et al. (2021). Half a century of global decline in oceanic sharks and rays. *Nature*, *589*(7843), 567–571.

Pardo, S. A., Cooper, A. B., & Dulvy, N. K. (2013). Avoiding fishy growth curves. *Methods in Ecology and Evolution*, *4*(4), 353–360.

Pardo, S. A., Kindsvater, H. K., Reynolds, J. D., & Dulvy, N. K. (2016). Maximum intrinsic rate of population increase in sharks, rays, and chimaeras: The importance of survival to maturity. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*(8), 1159–1163.

Parker, S. J., & Francis, M. (2012a). *Productivity of two species of deepwater sharks, Deania calcea and Centrophorus squamosus in New Zealand*. Ministry for Primary Industries.

Parker, S. J., & Francis, M. (2012b). *Productivity of two species of deepwater sharks, deania calcea and centrophorus squamosus in new zealand*. Ministry for Primary Industries Wellington, New Zealand.

Passerotti, M., Andrews, A., Carlson, J., Wintner, S., Goldman, K., & Natanson, L. (2014). Maximum age and missing time in the vertebrae of sand tiger shark (carcharias taurus): Validated lifespan from bomb radiocarbon dating in the western north atlantic and southwestern indian oceans. *Marine and Freshwater Research*, *65*(8), 674–687.

Pearce, T., Zaki, M., Brintrup, A., Anastassacos, N., & Neely, A. (2018). Uncertainty in neural networks: Bayesian ensembling. *Stat*, *1050*, 12.

Politikos, D. V., Petasis, G., Chatzispyrou, A., Mytilineou, C., & Anastasopoulou, A. (2021). Automating fish age estimation combining otolith images and deep learning: The role of multitask learning. *Fisheries Research*, *242*, 106033.

Politikos, D. V., Sykiniotis, N., Petasis, G., Dedousis, P., Ordoñez, A., Vabø, R., Anastasopoulou, A., Moen, E., Mytilineou, C., Salberg, A.-B., et al. (2022). DeepOtolith v1.

0: An open-source AI platform for automating fish age reading from otolith or scale images. *Fishes*, *7*(3), 121.

Pop, R., & Fulop, P. (2018). Deep ensemble bayesian active learning: Addressing the mode collapse issue in monte carlo dropout via ensembles. *arXiv Preprint arXiv:1811.03897*.

Porch, C. E., Wilson, C. A., & Nieland, D. L. (2002). *A new growth model for red drum (sciaenops ocellatus) that accommodates seasonal and ontogenic changes in growth rates*.

Pratt Jr, H. L. (1979). Reproduction in the blue shark, prionace glauca. *Fish. Bull.*, *77*, 445–470.

Pratt Jr, H. L., & Casey, J. G. (1983). Age and growth of the shortfin mako, isurus oxyrinchus, using four methods. *Canadian Journal of Fisheries and Aquatic Sciences*, *40*(11), 1944–1957.

Prince, J., Hordyk, A., Valencia, S. R., Loneragan, N., & Sainsbury, K. (2015). Revisiting the concept of beverton–holt life-history invariants with the aim of informing data-poor fisheries assessment. *ICES Journal of Marine Science*, *72*(1), 194–203.

Qi, Z., Khorram, S., & Li, F. (2019). Visualizing deep networks by optimizing with integrated gradients. *CVPR Workshops*, *2*, 1–4.

Quince, C., Shuter, B. J., Abrams, P. A., & Lester, N. P. (2008). Biphasic growth in fish II: Empirical assessment. *Journal of Theoretical Biology*, *254*(2), 207–214. https://doi.org/10.1016/j.jtbi.2008.05.030

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rahaman, R. et al. (2021). Uncertainty quantification and deep ensembles. *Advances in Neural Information Processing Systems*, *34*, 20063–20075.

Raoult, V., Howell, N., Zahra, D., Peddemors, V. M., Howard, D. L., Jonge, M. D. de,

Buchan, B. L., & Williamson, J. E. (2018). Localized zinc distribution in shark vertebrae suggests differential deposition during ontogeny and across vertebral structures. *PLoS One*, *13*(1), e0190927.

Ribot-Carballal, M. C., Galván-Magaña, F., & Quiñónez-Velázquez, C. (2005). Age and growth of the shortfin mako shark, Isurus oxyrinchus, from the western coast of Baja California Sur, Mexico. *Fisheries Research*, *76*(1), 14–21. https://doi.org/10.1016/j.fishres.2005.05.004

Ribot-Carballal, M., Galvan-Magaña, F., & Quiñonez-Velazquez, C. (2005). Age and growth of the shortfin mako shark, isurus oxyrinchus, from the western coast of baja california sur, mexico. *Fisheries Research*, *76*(1), 14–21.

Ricker, W. (1975). Computation and interpretation of biological statistics of fish populations. *Bull. Fish. Res. Bd. Can.*, *191*, 1–382.

Ricker, W. (1979). Growth rates and models. *Fish Physiology*, 677–744.

Ridewood, W., & MacBride, E. W. (1921). VIII.—on the calcification of the vertebral centra in sharks and rays. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *210*(372-381), 311–407.

Rigby, C. L., Wedding, B. B., Grauf, S., & Simpfendorfer, C. A. (2014). The utility of near infrared spectroscopy for age estimation of deepwater sharks. *Deep Sea Research Part I: Oceanographic Research Papers*, *94*, 184–194.

Rigby, C., Barreto, R., Carlson, J., Fernando, D., Fordham, S., Francis, M., Herman, K., Jabado, R., Liu, K.-M., Marshall, A., Pacoureau, N., Romanov, E., Sherley, R., & Winker, H. (2019a). *Lamna nasus-porbeagle. The IUCN red list of threatened species 2019*. https://doi.org/10.2305/IUCN.UK.2019-3.RLTS.T11200A500969.en

Rigby, C., Barreto, R., Carlson, J., Fernando, D., Fordham, S., Francis, M., Herman,

K., Jabado, R., Liu, K.-M., Marshall, A., Pacoureau, N., Romanov, E., Sherley, R., & Winker, H. (2019b). *Prionace glauca-blue shark. The IUCN red list of threatened species 2019*. https://doi.org/10.2305/IUCN.UK.2019-3.RLTS.T39381A2915850.en

Rigby, C., Barreto, R., Carlson, J., Fernando, D., Fordham, S., Francis, M., Jabado, R., Liu, K., Marshall, A., Pacoureau, N., et al. (2019). Isurus oxyrinchus. IUCN red list of threatened species. 2019, e. T39341A2903170. Doi: 10.2305/IUCN. UK. 2019-1. *RLTS. T39341A2903170. En*.

Ritter, H., Botev, A., & Barber, D. (2018a). A scalable laplace approximation for neural networks. *6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings*, *6*.

Ritter, H., Botev, A., & Barber, D. (2018b). Online structured laplace approximations for overcoming catastrophic forgetting. *Advances in Neural Information Processing Systems*, *31*.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 400–407.

Roberts, G. O., & Rosenthal, J. S. (2004). *General state space markov chains and MCMC algorithms*.

Robertson, S. G., & Morison, A. K. (1999). A trial of artificial neural networks for automatically estimating the age of fish. *Marine and Freshwater Research*, *50*(1), 73–82.

Robertson, S., & Morison, A. (2001). *Development of an aritificial neural network for automated age estimation*. Department of Natural Resources; Environment.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211–252.

Semba, Y., Nakano, H., & Aoki, I. (2009). Age and growth analysis of the shortfin mako, isurus oxyrinchus, in the western and central north pacific ocean. *Environmental Biology of Fishes*, *84*, 377–391.

Seoh, R. (2020). Qualitative analysis of monte carlo dropout. *arXiv Preprint arXiv:2007.01720*.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48.

Sigurardóttir, A. R., Sverrisson, ór, Jónsdóttir, A., Gudjónsdóttir, M., Elvarsson, B. ór, & Einarsson, H. (2023). Otolith age determination with a simple computer vision based few-shot learning method. *Ecological Informatics*, *76*, 102046.

Simard, P. Y., Steinkraus, D., Platt, J. C., et al. (2003). Best practices for convolutional neural networks applied to visual document analysis. *Icdar*, *3*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv Preprint arXiv:1409.1556*.

Skomal, G. B., & Natanson, L. J. (2003). *Age and growth of the blue shark (prionace glauca) in the north atlantic ocean*.

Slav, S. (2021). *Bayesian deep learning for large scale structure*.

Smart, J. J., & Grammer, G. L. (2021). Modernising fish and shark growth curves with Bayesian length-at-age models. *PLoS ONE*, *16*(2), e0246734. https://doi.org/10.1371/journal.pone.0246734

Soriano, M., Moreau, J., Hoenig, J. M., & Pauly, D. (1992). New Functions for the Analysis of Two-Phase Growth of Juvenile and Adult Fishes, with Application to Nile Perch. *Transactions of the American Fisheries Society*, *121*(4), 486–493. https://doi.org/10.1577/1548-8659(1992)121%3C0486:NFFTAO%3E2.3.CO;2

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Stevens, J., Bonfil, R., Dulvy, N. K., & Walker, P. (2000). The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems. *ICES Journal of Marine Science*, *57*(3), 476–494.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *International Conference on Machine Learning*, 3319–3328.

Takahashi, N., Kai, M., Semba, Y., Kanaiwa, M., Liu, K., Rodríguez-Madrigal, J., Tovar-Ávila, J., Kinney, M., & Taylor, J. (2017). Meta-analysis of growth curve for shortfin mako shark in the north pacific. *ISC/17/SharkWG-1/05*.

Takashima, Y., Takada, T., Matsuishi, T., & Kanno, Y. (2000). Validation of auto-counting method by NIH image using otoliths of white-spotted char salvelinus leucomaenis. *Fisheries Science*, *66*(3), 515–520.

Taylor, N. G., Walters, C. J., & Martell, S. J. (2005). A new likelihood for simultaneously estimating von bertalanffy growth parameters, gear selectivity, and natural and fishing mortality. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(1), 215–223.

Tracey, S. R., & Lyle, J. M. (2005). *Age validation, growth modeling, and mortality estimates for striped trumpeter (latris lineata) from southeastern australia: Making the most of patchy data*.

Trippel, E. A., & Harvey, H. H. (1991). Comparison of methods used to estimate age and length of fishes at sexual maturity using populations of white sucker (catostomus commersoni). *Canadian Journal of Fisheries and Aquatic Sciences*, *48*(8), 1446–1459.

Troadec, H. (1991). Frequency demodulation on otolith numerical images for the automa-

tion of fish age estimation. *Aquatic Living Resources*, *4*(4), 207–219.

Varghese, S. P., Unnikrishnan, N., Gulati, D. K., & Ayoob, A. (2017). Size, sex and reproductive biology of seven pelagic sharks in the eastern arabian sea. *Journal of the Marine Biological Association of the United Kingdom*, *97*(1), 181–196.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2015). Pareto smoothed importance sampling. *arXiv Preprint arXiv:1507.02646*.

Vincenzi, S., Jesensek, D., & Crivelli, A. J. (2020). Biological and statistical interpretation of size-at-age, mixed-effects models of growth. *Royal Society Open Science*, *7*(4), 192146.

Von Bertalanffy, L. (1938). A quantitative theory of organic growth (inquiries on growth laws. II). *Human Biology*, *10*(2), 181–213.

Von Bertalanffy, L. (1960). Principles and theory of growth. *Fundamental Aspects of Normal and Malignant Growth*, *493*, 137–259.

Walker, T. I. (1998). Can shark resources be harvested sustainably? A question revisited with a review of shark fisheries. *Marine and Freshwater Research*, *49*(7), 553–572.

Walker, T. I., Taylor, B. L., Hudson, R. J., & Cottier, J. P. (1998). The phenomenon of apparent change of growth rate in gummy shark (mustelus antarcticus) harvested off southern australia. *Fisheries Research*, *39*(2), 139–163.

Wang, H., & Yeung, D.-Y. (2020). A survey on bayesian deep learning. *ACM Computing Surveys (Csur)*, *53*(5), 1–37.

Wang, Y.-G., Thomas, M. R., & Somers, I. F. (1995). A maximum likelihood approach for estimating growth from tag–recapture data. *Canadian Journal of Fisheries and*

Aquatic Sciences, *52*(2), 252–259.

Weigmann, S. (2016). Annotated checklist of the living sharks, batoids and chimaeras (chondrichthyes) of the world, with a focus on biogeographical diversity. *Journal of Fish Biology*, *88*(3), 837–1037.

Wells, R., Smith, S. E., Kohin, S., Freund, E., Spear, N., & Ramon, D. A. (2013). *Age validation of juvenile shortfin mako (isurus oxyrinchus) tagged and marked with oxytetracycline off southern california*.

Wilson, A. G. (2020). The case for bayesian deep learning. *arXiv Preprint arXiv:2001.10995*.

Wilson, A. G., & Izmailov, P. (2020). Bayesian deep learning and a probabilistic perspective of generalization. *Advances in Neural Information Processing Systems*, *33*, 4697–4708.

Wilson, K., Honsey, A. E., Moe, B., & Venturelli, P. (2018). Growing the biphasic framework: Techniques and recommendations for fitting emerging growth models. *Methods in Ecology and Evolution*, *9*(4), 822–833.

Zhou, S., Martin, S., Fu, D., & Sharma, R. (2020). A bayesian hierarchical approach to estimate growth parameters from length data of narrow spread. *ICES Journal of Marine Science*, *77*(2), 613–623.

Zhu, X., Tallman, R. F., Howland, K. L., & Carmichael, T. J. (2016). Modeling spatiotemporal variabilities of length-at-age growth characteristics for slow-growing subarctic populations of lake whitefish, using hierarchical bayesian statistics. *Journal of Great Lakes Research*, *42*(2), 308–318.

Zou, D., Xu, P., & Gu, Q. (2019). Stochastic gradient hamiltonian monte carlo methods with recursive variance reduction. *Advances in Neural Information Processing Systems*, *32*.

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. *Graphics Gems*, 474–485.

Appendices

Appendix I: Growth Model Metrics

Table 16: PSIS-LOO metrics for *Centrophorus squamosus*.

(a) Males

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic L_∞	0.00	0.000	-748.0	18.09	19.99	2.621	1496	36.19
Biphasic k	-29.76	8.102	-777.8	18.27	43.77	7.519	1556	36.55
Common VBGM	-50.16	14.991	-798.2	17.38	17.79	3.249	1596	34.76
Original VBFM	-56.27	14.743	-804.3	18.17	34.41	5.148	1609	36.33

(b) Females

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic k	0.00	0.00	-1152	18.34	55.27	6.270	2304	36.68
Biphasic L_∞	-10.93	10.65	-1163	17.79	61.35	8.273	2326	35.59
Original VBFM	-82.65	14.58	-1235	17.85	42.47	3.048	2470	35.70
Common VBGM	-293.28	43.91	-1446	43.88	232.29	32.136	2891	87.76

Table 17: PSIS-LOO metrics for *Isurus oxyrinchus*.

(a) Males

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic k	0.0000	0.0000	-382.9	6.094	14.07	1.571	765.8	12.19
Biphasic L_∞	-0.4231	0.1674	-383.3	6.053	13.91	1.537	766.7	12.11
Original VBFM	-0.7734	1.6390	-383.7	6.068	13.72	1.422	767.4	12.14
Common VBGM	-0.7848	0.6601	-383.7	6.112	12.29	1.336	767.4	12.22

(b) Females

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic k	0.000	0.000	-1264	14.17	36.40	4.401	2527	28.34
Common VBGM	-2.075	2.078	-1266	13.98	30.84	3.634	2531	27.96
Original VBFM	-2.774	4.013	-1266	14.57	31.97	4.034	2533	29.14
Biphasic L_∞	-9.700	2.881	-1273	14.51	42.93	5.727	2546	29.02

Table 18: PSIS-LOO metrics for *Lamna nasus*.

(a) Males

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic L_∞	0.000	0.000	-769.8	15.11	19.60	2.840	1540	30.22
Biphasic k	-3.411	1.545	-773.2	15.35	20.83	3.321	1546	30.69
Common VBGM	-8.898	3.523	-778.7	14.09	14.17	2.154	1557	28.18
Original VBFM	-8.926	3.560	-778.8	13.95	14.39	2.016	1558	27.90

(b) Females

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic k	0.000	0.000	-627.5	10.97	17.24	2.396	1255	21.93
Biphasic L_∞	-4.023	2.106	-631.5	10.80	19.29	2.675	1263	21.61
Common VBGM	-10.861	5.148	-638.4	11.09	13.98	2.052	1277	22.18
Original VBFM	-12.441	4.987	-640.0	10.98	14.98	2.229	1280	21.97

Table 19: PSIS-LOO metrics for *Mustelus lenticulatus*.

(a) Males

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic L_∞	0.0000	0.0000	-608.6	10.34	16.52	3.961	1217	20.68
Biphasic k	-0.4607	0.3757	-609.0	10.41	16.17	3.768	1218	20.82
Common VBGM	-2.1474	1.5304	-610.7	10.79	14.17	2.931	1221	21.57
Original VBFM	-2.7434	1.9278	-611.3	10.93	13.72	2.911	1223	21.85

(b) Females

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Common VBGM	0.0000	0.0000	-358.5	7.826	13.69	2.159	717.0	15.65
Biphasic k	-0.2368	0.2002	-358.7	7.813	14.09	2.202	717.5	15.63
Original VBFM	-1.1103	0.5312	-359.6	7.769	14.19	2.234	719.2	15.54
Biphasic L_∞	-1.2674	1.6651	-359.8	8.482	15.54	3.179	719.5	16.96

Table 20: PSIS-LOO metrics for *Prionace glauca*.

(a) Males

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic k	0.000	0.000	-1415	16.08	27.98	3.930	2829	32.15
Biphasic L_∞	-3.351	2.458	-1418	15.68	27.04	3.768	2836	31.35
Original VBFM	-6.226	5.264	-1421	15.15	13.97	1.700	2842	30.30
Common VBGM	-6.523	5.247	-1421	15.17	13.82	1.700	2842	30.34

(b) Females

Model	ELPD Differences	SE Differences	ELPD LOO	SE ELPD LOO	P LOO	SE P LOO	LOOIC	SE LOOIC
Biphasic L_∞	0.000	0.000	-2576	29.27	28.53	3.811	5153	58.54
Biphasic k	-5.037	3.599	-2581	29.13	27.74	3.756	5163	58.25
Original VBFM	-8.563	5.812	-2585	28.84	18.50	3.454	5170	57.69
Common VBGM	-8.890	5.849	-2585	28.84	18.62	3.461	5171	57.68

Appendix II: Varying intercept estimates for growth models

Table 21: Varying intercept estimates for *Centrophorus squamosus*.

(a) Males

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
2007	-0.013	2.028	0.005	0.020	0.007	0.742	-0.407	1.443	-3.639	3.898
2008	-0.007	1.966	0.002	0.018	0.154	0.757	-0.709	1.466	-3.236	3.902
2009	-0.638	2.021	-0.001	0.019	0.294	0.806	0.195	1.823	4.894	10.664
2010	-0.329	2.194	-0.001	0.024	0.320	0.866	0.027	2.222	4.203	7.365
Chatham Rise	-39.377	11.393	0.045	0.069	2.545	1.654	-6.585	2.465	9.282	4.080
Campbell Plateau	9.889	14.651	0.113	0.455	-1.176	7.557	-1.328	7.907	-1.455	13.168
Auckland	-4.374	14.714	0.106	0.424	-0.161	7.324	0.603	8.207	-2.723	14.189
Puysegur	-6.893	10.111	0.002	0.053	3.374	1.378	1.645	1.473	-7.268	4.025

(b) Females

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
2007	-1.162	2.225	0.003	0.008	-0.237	1.095	-4.911	18.68	-1.624	4.468
2008	-0.406	2.016	0.004	0.007	-0.149	0.975	-6.464	21.64	-1.164	4.903
2009	-0.600	2.092	-0.003	0.007	0.092	1.051	-5.827	14.40	-7.360	4.791
2010	0.008	2.684	-0.001	0.007	0.193	1.310	-3.421	16.07	8.019	5.879
Chatham Rise	0.721	1.899	-0.029	0.033	-1.541	2.888	-25.754	18.52	6.048	4.991
Campbell Plateau	-0.420	2.650	0.579	1.443	-2.547	7.012	-32.777	115.95	-0.660	14.565
Auckland	-1.154	2.303	0.181	0.368	-2.720	7.380	-38.263	129.16	-1.011	17.381
Puysegur	0.013	1.891	0.031	0.042	3.556	2.235	-60.679	145.78	-4.693	9.551
Snares Shelf	-0.587	2.343	0.313	0.674	-3.366	6.854	-36.325	100.49	-1.536	15.441

Table 22: Varying intercept estimates for *Isurus oxyrinchus*.

(a) Males

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
1994	4.400	14.57	0.002	0.006	-0.465	1.060	-5.920	56.77	-22.352	121.8
2001	2.001	12.02	0.001	0.005	-0.061	1.128	-0.702	57.41	-11.917	118.0
2011	4.295	10.99	0.001	0.005	0.137	0.946	-2.616	57.25	4.934	127.7
2012	0.325	10.30	0.001	0.005	-0.013	0.777	3.635	56.04	-36.402	101.8
2013	-1.382	10.49	-0.001	0.005	0.675	0.909	3.881	66.68	-45.195	124.8
2014	-5.559	10.58	-0.003	0.005	0.386	0.947	-1.360	64.13	-31.245	114.0
2015	-2.412	11.51	-0.001	0.005	0.666	1.085	-0.271	59.81	20.128	129.7
Southland	3.350	13.62	0.000	0.005	0.554	1.235	1.176	52.94	-15.108	150.3
Challenger Plateau	2.301	11.14	0.001	0.005	0.004	0.848	-0.224	54.20	-55.639	127.3
Auckland West	2.737	12.16	0.001	0.005	0.483	0.979	-3.779	54.50	-20.259	135.9
Central East	2.575	13.78	0.001	0.006	-0.006	0.894	-2.356	51.29	-70.821	114.1
Auckland East	-7.743	14.41	-0.003	0.005	0.343	0.887	3.646	51.62	2.233	132.9
Egmont	-1.883	14.00	-0.001	0.006	0.139	1.384	1.918	52.47	7.018	145.0

(b) Females

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
1994	0.228	20.662	0.000	0.003	-0.468	1.027	0.474	44.47	48.327	147.64
1995	5.224	21.061	0.000	0.003	-0.044	1.076	1.128	47.47	-28.252	154.34
1996	1.364	18.547	0.001	0.003	-0.635	1.099	-1.405	44.30	17.521	165.01
1997	21.333	26.011	0.002	0.004	-0.555	1.491	-9.093	47.53	-9.021	147.91
1999	20.590	22.511	0.001	0.003	0.276	1.059	-6.990	46.81	6.597	168.88
2000	15.405	22.739	0.001	0.004	-0.282	1.509	-5.626	48.75	7.947	155.38
2001	14.691	19.068	0.001	0.003	-0.150	1.201	-4.591	42.74	-9.225	152.75
2002	-7.897	14.715	0.000	0.003	-0.205	0.886	-3.795	45.72	-117.477	112.59
2003	-10.106	14.494	0.000	0.003	-0.519	0.738	-0.696	45.16	-1.853	169.14
2004	-5.291	22.869	-0.001	0.003	0.411	1.118	1.583	46.06	-19.457	164.31
2011	-14.345	21.224	-0.001	0.003	0.632	1.357	2.263	47.16	-58.515	147.52
2012	2.923	18.691	0.000	0.003	0.753	0.957	-0.649	45.31	-53.658	158.36
2013	-3.302	16.934	-0.001	0.003	1.213	1.033	0.984	44.17	-110.685	144.70
2014	-8.775	24.502	-0.001	0.003	1.525	1.173	-8.913	45.72	47.852	100.43
2015	-10.328	20.707	-0.001	0.003	0.743	1.462	3.526	48.65	16.741	170.43
Southland	1.657	8.942	0.000	0.002	0.128	0.735	-6.735	58.77	-9.675	42.83
Challenger Plateau	-0.912	8.720	0.000	0.002	0.250	0.674	0.896	54.23	-4.796	43.33
Auckland West	-3.234	9.598	0.000	0.002	-0.242	0.662	-37.815	61.48	-15.876	46.17
Central East	8.920	14.456	0.000	0.002	0.545	0.813	-12.931	56.56	1.501	38.74
Auckland East	3.979	12.062	0.000	0.002	0.669	0.939	12.448	52.22	-14.532	55.71
Kermadec	-3.048	9.638	0.000	0.002	-0.256	0.759	-0.142	59.23	-0.934	48.85
Egmont	-0.640	11.323	0.000	0.002	-0.042	0.658	-2.213	58.76	3.789	44.66

Table 23: Varying intercept estimates for *Lamna oxyrinchus*.

(a) Males

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
1993	-0.532	6.379	-0.029	0.044	0.031	0.415	0.246	6.216	0.574	32.83
1995	-0.889	5.937	-0.001	0.018	0.036	0.364	0.073	4.806	-3.180	38.53
1996	-0.562	6.197	-0.028	0.042	0.034	0.422	0.175	4.488	-0.543	32.59
1997	-0.446	6.249	-0.012	0.034	0.012	0.413	0.140	5.363	-1.625	34.77
2000	0.936	4.932	0.013	0.032	-0.012	0.416	-0.074	4.064	0.957	38.95
2002	-2.144	3.913	0.003	0.016	-0.020	0.283	-0.084	5.434	6.016	33.61
2003	-1.035	3.744	-0.002	0.015	0.046	0.328	-0.024	4.770	-1.250	31.19
2011	2.099	4.772	0.009	0.017	-0.016	0.289	-0.156	3.935	1.276	33.63
2012	0.239	4.903	0.012	0.018	-0.037	0.270	-0.016	5.285	2.551	39.39
2013	0.169	4.663	0.016	0.020	-0.087	0.303	0.033	5.799	3.331	38.75
2014	2.274	4.764	0.019	0.021	0.035	0.323	-0.153	4.101	-1.975	41.86

(b) Females

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
1993	-2.747	13.967	-0.026	0.027	-0.034	1.060	0.797	7.498	-4.194	44.35
1995	1.842	7.734	0.009	0.013	-0.369	0.795	-0.501	6.965	-0.368	49.29
1996	-2.245	8.981	-0.010	0.011	0.400	0.748	0.695	6.201	-1.218	46.69
1997	-0.963	7.541	-0.004	0.010	0.167	0.665	0.279	7.128	11.786	53.29
1999	-1.155	7.314	-0.006	0.010	0.350	0.846	0.513	7.605	3.247	52.58
2001	-0.657	7.156	0.000	0.011	0.156	0.723	-0.227	6.701	-7.561	49.64
2002	1.059	5.154	0.001	0.007	0.132	0.453	-0.249	6.166	1.720	54.82
2003	0.285	5.221	0.006	0.009	-0.383	0.582	0.070	3.379	-12.054	8.47
2011	0.886	5.908	0.010	0.011	-0.265	0.721	0.053	6.961	-4.657	54.92
2012	0.995	6.819	0.004	0.008	0.005	0.450	-0.447	7.578	-0.880	49.43
2013	1.359	6.678	0.006	0.009	-0.228	0.534	-0.220	6.866	2.031	50.28
2014	1.428	6.318	0.012	0.011	-0.558	0.822	-0.247	6.862	14.071	46.66

Table 24: Varying intercept estimates for *Mustelus lenticulatus*.

(a) Males

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
1993	-2.126	7.345	0.007	0.069	-0.241	0.769	1.326	19.89	-9.626	48.54
1994	-0.536	6.352	0.000	0.036	0.023	0.461	-2.420	18.98	-0.508	36.11
1995	2.688	6.718	0.000	0.037	0.356	0.604	-1.947	19.67	-10.746	39.20
East Coast South Island	0.950	9.252	-0.006	0.037	0.079	0.466	-4.430	20.44	-5.721	33.18
West Coast South Island	0.172	6.617	0.003	0.035	-0.110	0.477	-0.512	18.50	-8.428	33.28
Porirua Harbour	-0.434	6.622	0.008	0.038	0.154	0.494	0.557	18.02	-3.178	37.01

(b) Females

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0
1993	-5.465	12.807	0.002	0.134	-0.651	0.891
1994	15.406	19.084	-0.112	0.115	-0.596	0.651
1995	9.500	14.086	-0.092	0.108	-0.189	0.539
East Coast South Island	-1.897	6.758	-0.011	0.040	-0.057	0.405
West Coast South Island	4.258	8.261	-0.016	0.044	-0.114	0.400
Porirua Harbour	2.364	7.563	-0.017	0.043	-0.028	0.401
Wellington Harbour	0.657	7.745	0.008	0.041	-0.399	0.611

Table 25: Varying intercept estimates for *Prionace glauca*.

(a) Males

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
1994	-4.846	14.84	-0.001	0.007	0.033	0.414	0.691	16.64	8.961	64.80
1995	9.669	14.93	0.001	0.007	0.185	0.361	-0.946	18.76	-0.459	55.08
1998	-4.381	17.86	-0.001	0.007	0.106	0.312	0.611	20.21	4.722	58.00
1999	-0.512	18.56	0.000	0.007	-0.003	0.261	-0.524	20.55	1.771	60.99
2002	-12.093	13.67	-0.003	0.006	-0.236	0.366	1.565	20.47	-23.200	53.01
2003	0.837	11.84	0.003	0.007	-0.166	0.411	-1.949	16.41	-1.505	42.09
2008	3.058	11.89	0.001	0.006	0.070	0.224	-0.917	16.81	12.494	58.45
2016	-9.703	12.70	0.001	0.006	-0.099	0.233	0.368	17.14	3.074	58.71

(b) Females

Factor Levels	Linf	Error-Linf	k	Error-k	t0	Error-t0	h	Error-h	th	Error-th
1994	9.884	20.556	0.017	0.023	-0.210	0.518	-4.067	19.76	3.771	54.540
1998	-8.144	18.166	-0.015	0.020	0.050	0.318	3.259	21.13	0.847	53.800
1999	3.796	13.589	0.012	0.020	-0.067	0.503	-0.170	19.48	14.382	59.837
2002	-4.398	15.365	-0.013	0.018	0.186	0.341	5.648	15.73	1.621	7.705
2003	4.109	9.736	0.008	0.015	-0.002	0.403	1.563	17.89	2.463	59.481
2008	-4.373	9.752	-0.006	0.013	-0.121	0.269	1.900	19.79	-9.015	58.016
2016	-2.713	9.953	-0.008	0.013	0.004	0.247	0.574	19.72	21.217	62.867