UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Estimation of Pagellus bogaraveo biologic parameters- Application of Bayesian state space models

Érica Sofia Cláudio Ribeiro

Mestrado em Bioestatística

Dissertação orientada por:
Professor Doutor Rui Martins
Doutora Ivone Figueiredo

2024

Agradecimentos

Gostaria de expressar a minha imensa gratidão a todas as pessoas incríveis que tornaram a realização deste projeto possível. A todos os que apoiaram não só tecnicamente, mas também através de encorajamento e motivação, o meu mais profundo obrigada.

Aos meus orientadores, Doutora Ivone Figueiredo e Professor Rui Martins, agradeço pela orientação e pelos conselhos ao longo desta jornada académica. Um especial obrigada pela paciência para responder a todas as minhas questões.

Ao IPMA e ao CEAUL, o meu sincero agradecimento por me terem proporcionado o primeiro contacto com o tema deste projeto. Em particular à doutora Inês Farias, pela disponibilidade em reunir e providenciar os recursos necessários. E à professora Lisete Sousa, que como docente e também no CEAUL, me apresentou importantes oportunidades.

A todos os docentes que fizeram parte do meu percurso, agradeço não só pelo conhecimento transmitido, mas também por terem despertado o meu interesse na estatística, particularmente na bioestatística. Acredito que me deram ferramentas valiosas não só para a elaboração desta tese, mas também para o meu futuro profissional.

Agradeço imensamente aos meus amigos e familiares por serem o meu maior suporte nos momentos mais difíceis, que não foram poucos. O apoio incondicional que recebi, especialmente dos meus pais, que foram meu suporte quando eu estava a flutuar nas possibilidades da investigação, é algo que valorizo profundamente. À minha melhor amiga Eva, mesmo diante do meu natural pessimismo, conseguiu arrancar-me um sorriso com as nossas peripécias da vida, que me fizeram encarar este projeto de outra forma.

Por fim, um agradecimento especial à minha irmã, que mesmo sem interesse no tema da tese e sobrecarregada com os seus próprios projetos, leu o meu texto inúmeras vezes para me ajudar a estruturar ideias e aprimorar a escrita. Sem ela, teria sido muito mais difícil encontrar o rumo.

Resumo

O goraz (*Pagellus bogaraveo*), é um peixe comum nas águas do Oceano Atlântico oriental e Mar Mediterrâneo, com um especial interesse comercial em Portugal. No entanto, apesar da sua importância comercial, a avaliação do stock do goraz consiste num grande desafio devido à escassez de dados e à complexidade dos seus processos biológicos. Apesar destes desafios, a avaliação do estado do stock e a estimativa dos parâmetros biológicos da espécie são essenciais para antecipar cenários de alteração do stock e ajustar as regras de exploração de modo a garantir a sustentabilidade. Estudos genéticos revelaram diferenças entre populações de goraz nas regiões do Atlântico e do Mediterrâneo, destacando a importância da compreensão da espécie que se encontra na área de Portugal continental especificamente.

Esta espécie possui um ciclo de vida e dinâmica populacional complexas, que a tornam sensíveis à sobrepesca. O ciclo de vida inclui uma sequência de processos desde o estado de ovo a adulto, que são influenciados por fatores externos como mudanças ambientais. Esses processos incluem migrações ontogenéticas nas quais as fêmeas vão até as áreas costeiras desovar, os ovos eclodem nessa área e mais tarde os indivíduos movimentam-se para águas mais profundas à medida que crescem. O goraz é também uma espécie hermafrodita, o que significa que pode mudar de macho para fêmea ao longo da vida. Estas características evidenciam a complexidade do goraz, e revelam a importância de estudar esses aspetos para a gestão e conservação da espécie.

Apesar destes conhecimentos, os poucos dados existentes são um grande desafio para melhor conhecer a espécie e estimar os seus parâmetros biológicos e abundância. Dessa forma esta tese tem como principal objetivo desenvolver e implementar um modelo estatístico que permita estimar os parâmetros biológicos e abundância, personalizado para a população no território continental português.

A metodologia utilizada neste estudo baseou-se em modelos de espaço de estados (MEE) com uma abordagem bayesiana. Esses modelos permitem estimar os parâmetros biológicos dos processos dinâmicos da população, incluindo Reprodução e Recrutamento, Mortalidade e Sobrevivência e Crescimento, integrando conhecimentos prévios e acomodando não só a variabilidade dos estados, mas também a dos dados. O modelo foi desenvolvido utilizando o pacote NIMBLE em R, aplicando a dados fornecidos pelo Instituto Português do Mar e da Atmosfera (IPMA). Os dados incluem 8 anos de atividade pesqueira (2014-2021), o número de indivíduos capturados e o seu tamanho em centímetros.

Reconhecendo as limitações dos modelos determinísticos para lidar com desafios proeminentes das populações de peixes, como variabilidade e incerteza, os MEE com abordagem bayesiana emergem como uma alternativa poderosa graças às suas características. O modelo divide o sistema em dois processos representados como séries temporais estocásticas em paralelo: o processo de estados e o processo observacional. O processo de estados descreve as mudanças de um estado de tempo para o próximo, neste modelo irá descrever as dinâmicas populacionais do goraz ao longo de um ano, que foram caracterizados através do uso de dez classes de tamanho dos indivíduos para melhor performance do modelo. O processo observacional contém os dados, e estes são conectados aos estados não observados e desconhecidos, neste caso conecta os dados providenciados pelo IPMA aos estados da população

descritos no processo de estados. Essa estrutura permite que o MEE estime a abundância e os parâmetros biológicos da espécie, considerando simultaneamente duas fontes de variação: variação do processo e erro de observação, fontes que no caso de espécies de peixes tem elevada importância dado o desconhecimento existente por falta de observação direta.

A abordagem bayesiana proporciona uma perspetiva distinta em relação a uma abordagem frequentista, considerando os parâmetros como variáveis aleatórias, desta forma é possível não só informar o modelo com conhecimento prévio sobre os parâmetros através das prioris, mas também considerar e quantificar as incertezas. Optou-se pelo uso de prioris informativas, numa tentativa de melhor desempenho do modelo.

Após o ajustamento do modelo aos dados, este é analisado através da comparação dos valores dos dados com a média a posteriori das estimativas obtidas, os valores são próximos, indicando que o modelo captura as tendências subjacentes nos dados e foram obtidas as estimativas pretendidas.

O modelo apresenta diferentes níveis de convergência em diferentes subprocessos e níveis de hierarquia do modelo, enquanto os valores estimados são biologicamente credíveis. Em todos os processos os hiperparâmetros apresentam valores de medidas de convergência que indicam que ocorreu convergência. Apesar dessa semelhança apenas no processo de Mortalidade e Sobrevivência é que os hiperparâmetros possuem uma distribuição posteriori diferente da priori, o que pode estar relacionado com prioris demasiado restritas nos outros processos.

Quanto aos parâmetros dos outros níveis de hierarquia, na Reprodução e Recrutamento houve uma homogeneidade na convergência dos parâmetros para valores biologicamente plausíveis, no processo de Crescimento embora a distribuição obtida para cada parâmetro seja credível, graças à grande presença de zero as medidas de convergência desses parâmetros mostram algumas discrepâncias, no entanto é no subprocesso da Mortalidade e Sobrevivência que existe um maior discrepância nos valores das medidas de convergência, em que todos os parâmetros associados as classes de tamanho 2,3 e 10 demonstram ter falhado na convergência, que mesmo após consideração de detalhes da estrutura dos dados e do modelo não foi encontrada uma resposta plausível.

Quanto à abundância total do goraz na costa continental portuguesa, os parâmetros apresentam intervalos de credibilidade estreitos, e bons valores nas medidas de análise de convergência. Os valores estimados demonstram que existe um grande declínio na população a partir de 2018, o que corrobora com estudos previamente realizados.

Após esta avaliação do modelo que foi estruturado utilizando 10 classes de tamanho, decidiu-se adaptar os dados a um modelo reformulado com 5 classes de tamanho. Existiu uma simplificação na estrutura do modelo, e uma redução de aproximadamente 50% dos parâmetros a estimar. No entanto, após o ajuste desse novo modelo aos dados, a avaliação de convergência demonstrou piores medidas de convergência na maioria dos parâmetros, principalmente nos níveis de hierarquia mais elevados. Essa ocorrência pode ser explicada por falta de ajustamento das cinco classes ao ciclo de vida do goraz, isto é, uma vez que maioria das equações que definem as dinâmicas populacionais no modelo dependem do tamanho central da classe, um modelo com classes de menor amplitude pode ter uma melhor performance. Sobretudo no processo de crescimento, onde o uso de 5 classes de tamanho faz com que a probabilidade de os peixes alterarem de classe de tamanho de um ano para o outro seja baixa, acabando por existir uma acumulação de indivíduos nas primeiras classes levando a um mau funcionamento do modelo.

De forma a validar os resultados obtidos, as médias à posteriori foram comparadas com resultados obtidos noutros estudos sobre o goraz. O modelo de dez classes de tamanho revela pequenas disparidades que se mantêm dentro do intervalo esperado quando comparados com os estudos de outros investigadores. Por outro lado, o modelo de cinco classes de tamanho apresentou variações mais evidentes. Ao comparar

Portugal continental com outras regiões, não foram encontradas evidências convincentes de desvios significativos nos valores dos parâmetros biológicos.

As implicações deste estudo são significativas para a gestão e conservação da espécie em Portugal. Ao melhorar a compreensão da dinâmica populacional do goraz, os gestores de pesca podem tomar decisões mais informadas sobre políticas de exploração sustentável e medidas de conservação para garantir a saúde a longo prazo desta espécie.

Em suma, esta pesquisa contribui não apenas para o avanço do conhecimento científico sobre a ecologia e biologia do goraz, mas também para a promoção da gestão sustentável dos recursos pesqueiros em Portugal, que tratando-se de uma área de extrema importância, no futuro seria relevante executar o modelo com condições previamente descritas, usando um conjunto de dados maior do International Council for the Exploration of the Sea (ICES), para analisar se os valores estimados mudam e se a convergência melhora. Incorporando mais dados, o modelo pode capturar mais padrões e fornecer estimativas mais robustas. Nesta tese, prioris informativas foram usadas para informar melhor o modelo sobre crenças anteriores, no entanto, o uso de prioris vagas poderia levar o modelo a convergir para valores diferentes, e seria valioso analisar essas diferenças realizando a repetição do estudo atual usando diferentes distribuições para os hiperparâmetros.

Palavras chave: Goraz, Modelo Estatístico, Modelo Espaço de Estados, Inferência Bayesiana, Dinâmicas Populacionais

Abstract

The blackspot seabream (*Pagellus bogaraveo*) is a relevant species with commercial importance. However, it is challenging to assess its stock due to scarce data and its complex biological dynamics. This thesis aims to address some of the knowledge gaps by developing a tailored statistical model, specific for mainland Portugal. By integrating previous knowledge, accommodating data variability, and considering its complex life cycle, the model aims to estimate population parameters using a dataset provided by IPMA. Utilizing State-Space Models with a Bayesian approach, implemented in R with the NIMBLE package, the model allows for the estimation of biological parameters of the population dynamics, while considering the dynamic interplay between observable data and hidden processes. Despite challenges posed by scarce data, the model provides acceptable estimates for most parameters. Particularly, it is effective in estimation of total abundance. Convergence analysis of parameters revealed varying levels of convergence across different sub-processes and hierarchical levels. Most hyperparameters showed overlap between prior and posterior distributions. Furthermore, comparison with other models and exploration of parameter convergence further validated the model's performance. This study contributes to a better understanding of the blackspot seabream population dynamics, which is crucial for effective fisheries management and conservation strategies in mainland Portugal.

Keywords: Blackspot seabream, Statistical modeling, State Space Model, Bayesian inference, Population dynamics

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In a world marked by the increasing need to preserve biodiversity and ensure the sustainability of marine resources, the application of statistical models emerges as a crucial tool for understanding and managing ecosystem complexities. These models play a vital role in our understanding of intricate systems marked by uncertainty, unknown states, and enigmas, addressing complex issues like population ecology (Koller and Friedman, 2009).

Within the realm of ecology, statistics play a pivotal role in studying relationships between organisms and their environments, aiding conservation efforts and sustainable resource management (Guillot et al., 2022; King, 2014). When considering population ecology, a key sub-discipline focusing on individual species, a quantitative approach becomes essential. This would allow a more precise and comprehensive understanding of the dynamics and interactions governing populations, providing a foundation for deeper insights and informed decision-making. Here, ecological data must be modeled and analyzed to make inferences about each species, leading to the emergence of population dynamic models that offer insights into species characteristics and predict population development (King et al., 2009).

This thesis focuses on *Pagellus bogaraveo* (Brünnich, 1768), herein blackspot seabream. This species has complex reproductive and life cycles. The estimation of biological parameters and of the abundance of this species, in other geographical locations, has proven to be challenging due to its particularities (Lorance, 2011). Additionally, the fishery data in mainland Portugal is scarce, adding to the challenge and relevance of developing this model. Considering the importance and utility of modeling species, this thesis' focus is to build a statistical model for blackspot seabream in mainland Portugal.

The information derived from this model could be valuable for management decision, explaining past events, predicting future scenarios, such as population collapses, and providing essential insights into the species' evolution over time (Newman et al., 2009). Species dynamics are related to aspects with high uncertainty, which increases when dealing with species, due to the observation of the states becomes indirect and uncertainty gains a main role. Therefore, statistical models' importance and utility become more evident. Statistical models serve as vital instruments for understanding hidden states and uncertain parameters.

The management of fisheries resources plays a critical role in maintaining the ecological balance of our oceans and ensuring a sustainable food supply for the growing global population (Newman et al., 2009). In this pursuit, State Space Models (SSM) have emerged as a powerful and versatile tool, offering a dynamic framework to analyze and understand complex systems with inherent uncertainties. These models allow for the explicit modeling of the underlying dynamics of a system and also the uncertainties associated with the observed data. This capability makes SSM valuable in ecological studies, where they enable a more detailed and comprehensive analysis of species dynamics while accounting for various

sources of uncertainty.

The blackspot seabream is a demersal species belonging to the Perciformes order and Sparidae family (Martins and Carneiro, 2018; Micale et al., 2011). Its distinctive biological characteristics and complex life cycle, comprehensively described in Chapter 2, make it exceptionally susceptible to the adverse effects of overfishing (Lorance, 2011). However, incomplete knowledge of its spatial dynamics contributes to the uncertainty about the current state of the stock (M. Pinho et al., 2014). Understanding the causes associated with blackspot seabream fluctuation of abundance in relation to fishing is crucial, and provide valuable insights for stock management decisions. The complexity of its biological dynamics and limited data availability make this task particularly challenging. Therefore, predictive studies regarding population dynamics, despite being highly relevant, are extremely challenging.

Modeling population dynamics is a relevant topic that can have different levels of complexity, depending on the population dynamics and the questions under consideration. The construction of a model involves numerous possibilities, and the key to success lies in clearly defining its purpose to address the precise research questions under consideration.

In this thesis, a tailored statistical model has been developed to estimate the abundance of blackspot seabream and its biological parameters in mainland Portugal. This model adopts a State Space Model framework with Bayesian inference through the Markov Chain Monte Carlo algorithms, offering a robust methodology for capturing the intricate dynamics of the population.

The primary objective of this research is to formulate a stock assessment model that delves into the dynamic interplay between observable data and hidden processes associated to the blackspot seabream population dynamics. To achieve this, Bayesian inference, Markov Chain Monte Carlo algorithms, and tools such as the NIMBLE package (version 1.0.1) are employed in R (version 4.3.1). These statistical techniques, described in Chapter 4, facilitate the extraction of inferences about the species, addressing and overcoming challenges presented by limited data and complex biological dynamics. The overarching aim is to provide a comprehensive understanding of the blackspot seabream population, contributing valuable insights for effective fisheries management and conservation efforts in the region.

In particular, this thesis's objectives are:

- Understand population dynamics of blackspot seabream;
- Review models applied to population dynamics within the context of fish ecology;
- Design and develop a model tailored for blackspot seabream population dynamics;
- Evaluate the model under different scenarios and datasets;
- Analyze the estimates obtained and comparing them with values from other authors.

This thesis is organized into several chapters, each serving a specific purpose, including the following items:

- Chapter 2 - Literature Review of *Pagellus bogaraveo*: Investigate the physical characteristics influencing blackspot seabream population dynamics; Analyze the life cycle stages and their implications for population dynamics; Explore spatial distribution patterns and the environmental factors shaping them; Examination of historical overview of early studies on blackspot seabream; Identification of knowledge gaps that the present thesis aims to address;
- Chapter 3 - Dataset used: Presentation of the dataset used; Analysis of these data;
- Chapter 4 - Methodology and Resources: Presentation of historical evolution of fish model methods; Presentation of State Space Models as an alternative methodology; Bayesian inference using the Markov Chain of Monte Carlo (MCMC) algorithms;

- Chapter 5 - The Model developed: Description of the model developed; Detailed presentation of equation used; Introduction to prior distributions used; Respective references;
- Chapter 6 - Analysis of Results: Analysis of the results obtain from the model presented and other alternatives; Comparison with results from other authors;
- Chapter 7- Conclusion and future possible studies.

Chapter 2

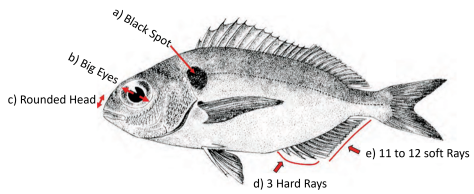
Literature review of *Pagellus bogaraveo*

The blackspot seabream was first described by Brünnich in 1768 (Brünnich,1768), is a demersal species of the Perciformes order and Sparidae family (Martins and Carneiro, 2018). This species is predominantly found in the eastern Atlantic Ocean and the Mediterranean Sea (Micale et al., 2011). Notably, countries with coastlines along the Mediterranean, including Spain, France, Italy, and Greece, have shown considerable fishing interest in it, due to its esteemed culinary value and popularity. Concerns about the population's status emerged toward the end of the 20^{th} century, when landings decreased (Lorance, 2011).

This chapter provides an overview of the blackspot seabream's biological characteristics. It includes physical features that are relevant to distinguishing this species from others, essential knowledge for fisheries and biologists when collecting data, as the data used in this thesis. Next, an overview of the species dynamics, that are fundamental to stock assessment, and central to the aims of this thesis that rely on available information about life cycle stages and biological processes, is presented. Finally, the spatial predominance of the species and its connection to genetic differences are explained.

2.1 Physical characteristics

The blackspot seabream can be distinguished from other species by its physical characteristics, as illustrated in Figure 2.1. It has a rounded head profile and relatively large eyes compared to the length of its snout, the dorsal scales end between the posterior margin and the middle of the eye. It possesses pointed teeth at the front and molar-like teeth at the back, displays a grayish-reddish coloration and often exhibits a prominent black spot at the origin of its lateral line, although this spot may be absent in young individuals (Martins and Carneiro, 2018). Blackspot seabream grows to at least 70 cm standard length (Lorance, 2011), it is important to acknowledge that fish may undergo unintended modifications, such as tail loss. Consequently, two fish length measurements exist: fork-length and total-length.



(a) Description of physical characteristics

(b) Illustration of appearance

Figure 2.1: Physical characteristics of blackspot seabream adult phase. These images are adapted from Martins and Carneiro, 2018

2.2 Life Cycle

While physical characteristics enable differentiation of blackspot seabream from other species, the most distinctive aspect lies in its biological dynamics.

The life cycle is a sequence of biological processes through witch the species passes during its development from an egg into an adult, going through larvae and juvenile stages. Blackspot seabream exhibits complex biological dynamics (Krug, 1990, 1998; Lorance, 2011). Biological processes include growth, sex change, reproduction and survival. Transitions between life stages can be influenced by external factors, adding another layer of complexity (Wingfield, 2013).

The life cycle of the blackspot seabream, illustrated on Figure 2.2, includes external fertilization of eggs, fertilized eggs develop into larvae stage, lasting approximately 40 days, based on species of the same genus. During this stage, blackspot seabream drifts in shallow waters columns, with low survival rate, very dependent on environmental conditions such wind direction (Araújo et al., 2016; M. Pinho et al., 2014; Stockley et al., 2005; Teixeira, 2013).

Larvae undergo growth and survival processes until reaching the juvenile stage, when individual obtain 22cm fork length (Krug, 1998; Micale et al., 2011). The initial years of these individuals are spent in coastal areas, then juveniles migrate to deeper waters, while continuing to grow, during witch they recruit to exploited population (Araújo et al., 2016; M. Pinho et al., 2014; Sobrino and Gil, 2001; Teixeira, 2013).

When reaching the adult stage, blackspot seabream migrates further to deeper water. This means that larger, and older, fish are found in deeper waters than juveniles, excluding when they are spawning (Lorance, 2011).

Blackspot seabream is a protandric hermaphrodite species, i.e., it changes from male to female (Krug, 1986, 1990; Lorance, 2011; Sobrino and Gil, 2001). The majority of individuals are firstly functional males, and then they undergo a sex change, becoming functional females. The change of sex occurs between 27 and 34 cm, with males dominating intervals between 21-26 cm fork length, and females are more abundant when the fork length is bigger than 35 cm (Krug, 1986).

The ratio of females and their maturation status is contingent upon population abundance, where a decrease in species abundance corresponds to an elevated proportion of females (Krug, 1998; Lorance, 2011).

In the adult stage, fish became able to reproduce. Maturation involves an increase and the development of adult characteristics, including gonad development. Fecundity (female fecundity, referring to the number of ripening eggs before spawning) is directly related to length (Krug, 1986, 1990; Micale et al., 2011). According to studies conducted in the Azores, males attain maturity at around 27 cm fork length, while females achieve it at approximately 29 cm fork length (Krug, 1986; Lorance, 2011; Micale et al., 2011). Mature adults migrate to coastal areas and release their sexual products (eggs and sperm) during the spawning process (Krug, 1986).

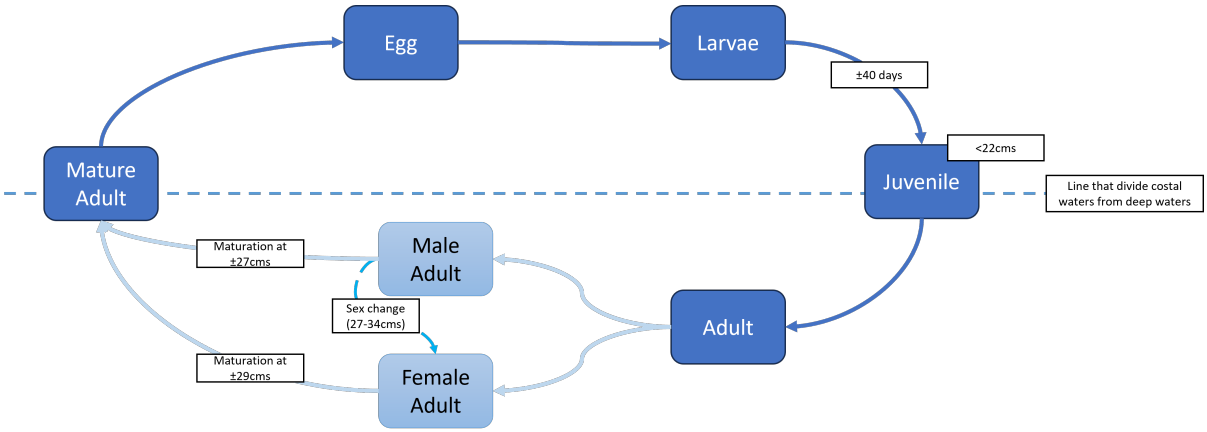


Figure 2.2: Blackspot seabream life cycle. Sequential stages and pivotal events from Egg to Adult. These stages include Egg, Larvae, Juvenile and Adult, encompassing notable transitions such as sex-change, growth, maturation and migrations.

Biological dynamics of blackspot seabream are complex, including ontogenetic migrations, discontinuous habitat, hermaphroditism, slow growth and seasonal development of gonads (Krug, 1990, 1998; Lorance, 2011; Micale et al., 2011). Due to its particular biology, blackspot seabream may be particularly sensitive to overfishing (Lorance, 2011).

2.3 Spatial distribution

Blackspot seabream has also a complex spatial distribution. It is primarily found in the eastern Atlantic Ocean and the Mediterranean Sea (Micale et al., 2011).

In Figure 2.3, The Food and Agriculture Organization (FAO) Major Fishing Area 27, encompassing the Atlantic Northeast and its subareas (represented by red traces), is depicted along with the corresponding ecoregion as defined by the International Council for the Exploration of the Sea (ICES). This visual representation aids in understanding the following description of blackspot seabream spatial distribution (FAO, n.d.; International Council for the Exploration of the Sea (ICES), 2023).

Understanding this is fundamental for sustainable fisheries management (e.g. preventing overexploitation in specific areas) and scientific research (e.g. studying population dynamics).

This thesis, with a specific focus in mainland Portugal, places major importance on unraveling the spatial distribution of the blackspot seabream. The emphasis lies in gaining a nuanced understanding of how this population segregates and distinguishes itself from counterparts in different geographical locations, as well as comprehending the intricate dynamics of their interactions.

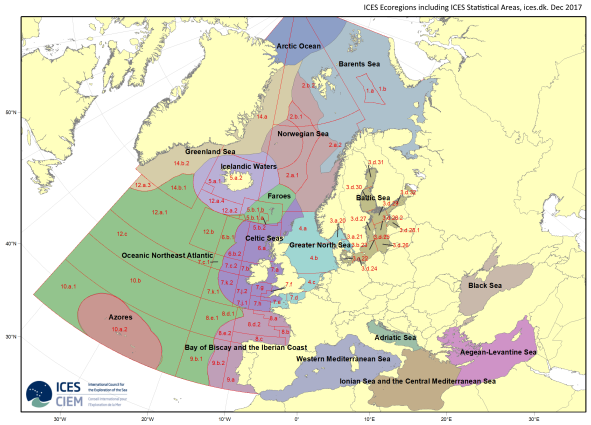


Figure 2.3: ICES fishing regions: Image collected from ICES documents ("ICES - Maps", n.d.)

The stock structure of blackspot seabream in ICES area is considered unknown. Despite that and for assessment purposes, ICES considers three assessment units, Subareas 6 to 10 (6-9 being Atlantic Iberian waters and 10 Azores region). According to recent reports by ICES those components consider migrations between subareas 6, 7, and 8 to the southern part of division 9.a and genetic differences between populations of division 10.a.2 and 9.a (ICES, 2022).

In Subareas 6, 7, and 8, blackspot seabream was heavily exploited by the French and Spanish. Since the mid-1970's, the fishery targeting this species in these regions has significantly declined, which was associated to a depletion of the stock. In recent years, species catches have mainly been bycatch, meaning that this species has not been directly targeted and catches happen while targeting other species. Reflecting the critical state of this stock, ICES advice from 2020 is that a precautionary approach with zero catch recommendations for 2021 and 2022 (ICES, 2022).

In Subarea 9, the majority of the landings come from Portugal ($\pm$55%) and Spain ($\pm$45%). In mainland Portugal, blackspot seabream is mainly caught as bycatch. The ICES advice of 2020 was "that when the precautionary approach is applied, catches should be no more than 119 tonnes in each of the years 2021 and 2022." (ICES, 2022).

In Subarea 10.a.2, the Azores have a historical fishery for blackspot seabream, however landings have fluctuated over the years (M. Pinho et al., 2014; M. R. Pinho and Menezes, 2009). In recent years landing have decreased, ICES advice a maximum of 610 tonnes catches for area 10 in 2022 (ICES, 2022). A minimum landing size of 33cm has been adopted by European Union (EU) for this species.

The management strategies for blackspot seabream differ among regions. Understanding the spatial distribution of the species is highly relevant for the development of species dynamic model.

Genetic studies on blackspot seabream populations, revealed limited genetic differentiation between populations in the Mediterranean Sea and those found along the Atlantic continental shelf (Robalo et al., 2021). Recent genetic inquiries have shed light on distinct genetic variations between blackspot seabream populations in the Azores (ICES Subdivision 10.a.2) and those in mainland Portugal within ICES Division 9.a (ICES, 2022). These studies contribute for the understanding of blackspot seabream stock dynamics. Specifically, they highlight the presence of genetic differences between populations in the Azores and mainland Portugal, indicating unique dynamics within these regions.

In conclusion, the spatial distribution of blackspot seabream presents unique challenges for sustainable management, with significant differences between subareas. Understanding these discrepancies underscores the complexity of blackspot seabream stock dynamics, a crucial aspect for fisheries management and conservation efforts.

2.4 Early studies

Early studies on blackspot seabream provide valuable insights into various aspects of this species, turning out to be virtuous supplies to this thesis.

Helena Krug has been a big contributor to blackspot seabream studies, primarily using data collected of the Azores population. Her work spans various topics, covering all processes and stages of blackspot seabream's life cycle, including age and growth, and reproduction (Krug, 1986, 1989, 1990, 1998). The equations derived from her discoveries have been relevant to modeling blackspot seabream population dynamics, forming the basis for subsequent studies, including the present thesis. The specific application of these equations in this thesis will be detailed in the Chapter 5.

In 2011, Pascal Lorence made significant contributions by reconstructing a time-series of landings from the Bay of Biscay and analyzing size compositions within these landings. Additionally, a dynamic model was applied to gain a better understanding of blackspot seabream population dynamics (Lorance, 2011).

Genetic studies, played a crucial role for understanding the stock structure of blackspot seabream populations (Robalo et al., 2021; Stockley et al., 2005). Genetic differences in different geographical areas offer a comprehensive view of genetic variations essential for understanding genetic diversity and population structure (Robalo et al., 2021; Stockley et al., 2005).

In 2019 IPMA presented a scientific and technique report on species maturation, providing detailed information on maturaty macroscopic observations (Farias and Figueiredo, 2019).

In addition to those studies, ICES at the working group on the biology and assessment of deep-sea fisheries resources (WGDEEP), provides in annual reports descriptive analyzes of blackspot seabream populations based on landing data from various regions. These reports can offer valuable historical context and insights into broader trends in fisheries. Data from mainland Portugal was recently been reported independently.

Although these earlier studies have made substantial contributions to the field, they also pinpoint areas where further research is imperative. This becomes even more evident when data sources are scarce or where findings from previous studies remain inconclusive, reinforcing the importance of this thesis.

This thesis, unlike the previously presented studies, will focus on the population of mainland Portugal and estimate parameters of various processes of the biological dynamics of the blackspot seabream simultaneously. To achieve this, it also employs a different framework from other studies,using SSM with bayesian approach, as presented in Chapter 4.

Chapter 3

The data

In this chapter, the data available and used in this thesis is presented. This is relevant as it will allow a better understanding of how the statistical methods can be applied to the available data.

The dataset used in this thesis, provided by the IPMA, includes the number of individuals caught between 2014 and 2021 by total length classes of one centimeter range (19cm-60cm), as outlined in Table 3.1.

Table 3.1: Dataset Excerpt: First and last six rows of the dataset provided by IPMA

Length	**Year**							
(cm)	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
19	0	84	0	0	0	0	0	0
20	0	336	0	0	0	0	0	0
21	0	839	0	0	0	0	0	121
22	0	611	214	362	0	0	0	182
23	210	366	641	0	273	0	556	303
24	914	204	855	432	1211	41	2779	91
				(...)				
55	92	211	293	0	0	35	0	0
56	0	235	506	175	256	0	0	366
57	0	471	0	67	0	0	0	0
58	0	0	0	0	0	0	0	0
59	0	235	0	0	0	0	0	0
60	92	0	0	67	0	65	0	0

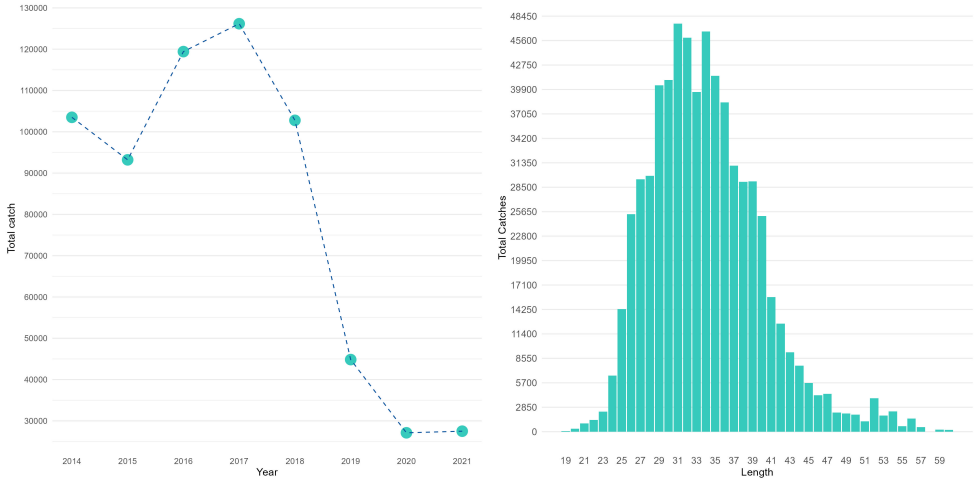
Most biological features of fish are related to length, making it a good measure to organize and split the sample. Length is a direct measure, making it easier to use, in comparison to age which requires indirect measures (Blomstedt et al., 2015).

The dataset is restricted to Peniche fishing area, and it is important to emphasize that the primary fishing activity did not exclusively target blackspot seabream, (see Chapter 2). To better understand the data, a descriptive analysis is presented in the next section.

3.1 Descriptive Analysis

An overview of the number of fishes caught by year is shown in Figure 3.1 (a). There is a substantial decline in total catches in 2019 when compared to previous years. From 2014 to 2018, the total catch

consistently exceeded 93207 (the total catch of 2015), whereas from 2019 forward, it did not surpass the 44841 (the total catch of 2019). Blackspot seabream catches varied along years (Figure 3.1 (b)), highest frequencies occur between 30 cm to 40 cm. This preliminary analysis aligns with the studies mentioned in the previous Chapter2, which reported landings decreasing in recent years and more adult individuals caught by fisheries (ICES, 2022; Lorance, 2011).



(a) Annual total catch of blackspot seabream in Peniche

(b) Length frequency distribution of blackspot seabream from catches

Figure 3.1: Graphs that present the distribution of catches by year and length: In the left Figure, each point represents the total number of fish caught each year. In the right Figure, each bar represents the total number of fish caught at each size

Length of exploited population is presented in Figure 3.2. In 2021 length of specimen caught were higher as evident in the boxplots presented in Figure 3.2. In 2020, characterized by lower values (Figure 3.2), the only blue point below the 33 cm line is the year 2020, with a mean of approximately 30.81 cm (being 33cm the minimum landing size mentioned in Chapter 2).

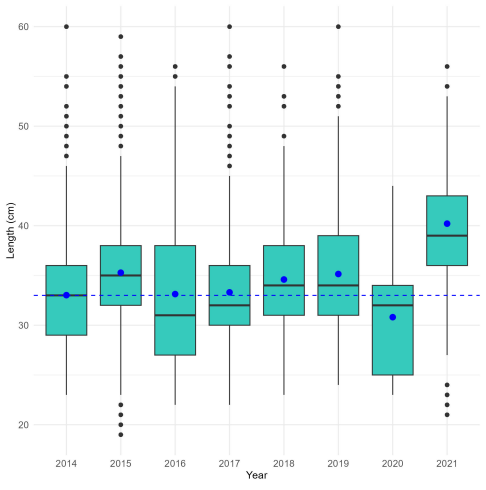


Figure 3.2: Boxplots presenting the distribution of fish length in catches by year: Each box represents the size distribution of catches for each respective year, with blue points indicating the mean size for each year. The dashed blue line represents the 33cm length.

When comparing Figure 3.2 with a similar graph showed by ICES, 2022, where data from blackspot seabream fishery in the Strait of Gibraltar a noticeable distinction arises: the data utilized in this thesis,

pertaining to mainland Portugal, exhibits lower values than those of in the Strait of Gibraltar. This can indicate differences between the two population that could be explored.

In this project, the use of size-class, which involves grouping lengths into distinct categories, was adopted. The use of size-classes introduces a level of flexibility and simplifies the modeling process by reducing the number of parameters. The definition of these size-classes took into consideration the data collected, the life cycle described in the previous Chapter 2, and considerations for model convenience. Ten size-classes were defined, starting with an initial length of 11 cm, which marks the point when the fish becomes part of the population, approximately one year after birth. Each size-class encompasses 5 distinct lengths. These intervals with uniform amplitudes play a crucial role in defining the subsequent model and applying most equations, as it is going to be presented in the following Chapter 5. Table 3.2 provides a representation of all the size-classes along with descriptions of the most pertinent features associated with each class.

Table 3.2: Size-classes adopted in this study: This table presents 10 size-classes, each defined by length class intervals of 5 cm, along with descriptions of characteristics associated with each class.

Size-Class	Interval	Description
Class 1	[11,16[	**Recruits**: Newly born fish are expected to be recruit to this class a year after being born. Those individual are not targeted by fisheries.
Class 2	[16,21[	**Juveniles**: Live near the coast and are not caught in fishery
Class 3	[21,26[	**Pre-sex-change adults**: Adult fish before undergoing sex change, predominantly male.
Class 4	[26,31[	**Sex-change period**: Transition period where sex change occurs, with an increasing number of females and maturation.
Class 5	[31,36[	**Higher maturity**: High proportion of mature individuals, with a significant increase in catches.
Class 6	[36,41[	**Mature females**: Mature females with fecundity higher than in the previous.
Class 7	[41,46[	**Older mature females**: Similar to mature females but with increased fecundity due to age/length.
Class 8	[46,51[	**High fecundity**: Individuals in this class have high fecundity, but catches decrease.
Class 9	[51,56[	**Older individuals**: Similar behavior to the previous class, but catches decrease.
Class 10	[56,60[	**Rarely caught older individuals**: High natural death rates with catches being rare or not observable. For proper functioning of the model this class encompasses 5 distinct lengths and there is no higher size-class, if individuals get expected length higher than 60cm the expected outcome is to be considered to belong to this size-class.

These size-classes offer a structured framework for analysis, taking into account the life cycle of the species and facilitating the modeling process for subsequent chapters. Figure 3.3 shows that across all years, catches tend to cluster around the middle size-classes, a pattern consistent with the observations in Graph (b) of Figure 3.1. Specifically, the modal size-class is [31,36[in all years, except for 2016 and 2021, where the modal classes are [26,31[and [36,41[respectively. Additionally, a striking disparity in the number of catches can be observed, especially, in the last three years. Between 2019 and 2021, when

compared to the previous years a considerable reduction in catch quantities was found.

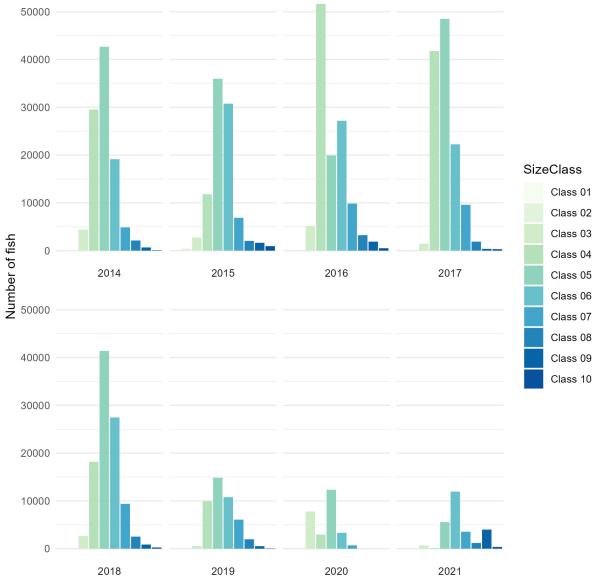


Figure 3.3: Graphs representing catches by size-class on different years: Each graph offers a detailed depiction of the division of catches by size-class

Figure 3.4, using a color scale, provides a distinctive insight into the catch distribution, disregarding the total number of individuals captured. In 2021, with the bar appearing darker, meaning that, there was a predominance of catches in higher size-classes. In contrast, the year 2020 appears lighter in comparison with the previous years that looked balanced in terms of distribution.

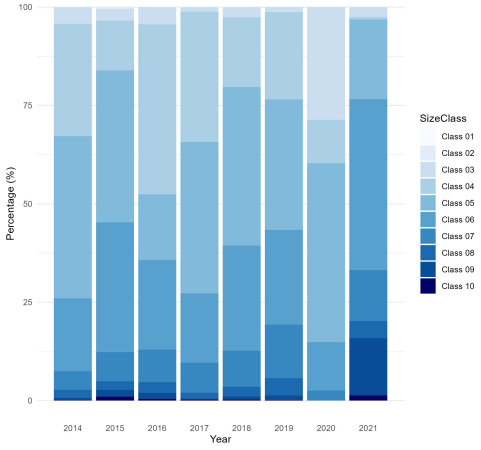


Figure 3.4: Graph of size-classes distribution over the years: Each bar represents 100% of catches for a given year, with colors dividing the bar to represent the distribution of size-classes across each year

Chapter 4

Methodology and resources

This chapter provides a comprehensive understanding of fish modeling and introduce State Space Models (SSM) as an approach to stock assessment modeling. It also explores methods for inference and relevant software tools. The details presented here are selected based on their relevance to the primary objectives of this thesis and the dataset discussed in Chapter 3.

Mathematical models utilize data from scientific surveys looking for insights into the dynamics of fish populations. Estimation of key parameters, such as population size, recruitment, natural mortality, and fishing mortality, improves the knowledge about the species and allows decision-makers to formulate well-informed regulations regarding catch limits, fishing seasons, and other governing measures. Fish stock assessment face multifarious challenges, including limitations associated with data quality and uncertainties.

First development in fish stock assessment was by Baranov catch Equation (4.1). The equation describe the population dynamics, it includes two key mortalities: fishing (F) and natural mortality (M), applied to an initial cohort of N_0 fishes. This equation allows to calculate the number of fish caught (C) at a specific time-step (T). Although this equation laid the foundation for early studies and algorithms, its deterministic nature limited its ability to account for uncertainty measures or predict future stock sizes (Aeberhard et al., 2018).

$$C = \frac{F}{F+M}(1 - e^{-(F+M)T})N_0 \tag{4.1}$$

Later, researchers have developed more complex equations and models incorporating additional parameters and biological processes. Two seminal books, Hilborn and Walters, 1992 and Quinn and Deriso, 1999, have been foundational references for fish modeling, offering comprehensive insights into a diverse array of methods and models used to characterize fish population dynamics (Newman et al., 2009).

In the context of fish dynamics studies, the distinction between process variation and observation error is crucial because it acknowledges the considerable amount of information that remains unknown or uncertain in fisheries research. The omission of this distinction in modeling approaches can lead to significant information loss when dealing with fishing-related data. However, due to limitations in computational methods, the first fish models do not address this topic (Newman et al., 2009).

Recent developments in computational resources have enabled studies to consider such information, consequently, it becomes imperative to take these factors into account when developing a more comprehensive model.

4.1 State Space Models

Deterministic models exhibit inherent limitations, particularly in handling prominent challenges of fish populations as: variability, quantifying uncertainty, or provide forecasts. Addressing uncertainty in fisheries modeling becomes imperative due to imprecise fishing data, observed data to unknown quantities, and incomplete understanding of system dynamics (Aeberhard et al., 2018; Thomas et al., 2005). Acknowledging these limitations and the demand for a probabilistic approach, State Space Models (SSM) within a Bayesian framework emerge, among others, as a powerful alternative (Durbin and Koopman, 2001).

Under SSM, the system is divided into two processes represented as stochastic time series running in parallel: the state process and the observation process. The state process describes the unobserved information, while the observation process contains the observed information, connecting it to the unobserved and unknown states (Kitagawa and Gersch, 1996; Newman et al., 2009). The SSMs allows to estimate the abundance and biological parameters of the species, maintaining consistency in population dynamics while, simultaneously, considering two critical sources of variation: process variation and observation error. This variation is a crucial topic when evaluating fish species, so SSM have emerged as a flexible and robust alternative to the methods previously mentioned (Aeberhard et al., 2018; Auger-Méthé et al., 2020; Newman et al., 2009; Thomas et al., 2005).

Since the late 1980s, early SSM applications, especially in fishes, like in Mendelssohn, 1988 and Sullivan, 1992, opened the way for multiplied adoption of the method. Recent advancements in computing efficiency have further enhanced SSM, enriching them with innovative approaches.

This thesis seeks to explore and apply State Space Models as a versatile and potent tool for assessing blackspot seabream stock.

4.1.1 Mathematical definition of State Space Models

SSM are hierarchical, being characterized by two parallel time series processes: the state process and the observation process (Auger-Méthé et al., 2020).

The state process, typically unobserved and hidden, describes the changes from a state N_t to N_{t+1}, where t stands for time-step, $t \geq 1$ and to simplify notation, it is here assumed discrete evenly spaced time indices. The state N_{t+1} only depends on the previous state (N_t), constituting a first-order Markov chain (Newman et al., 2009; Thomas et al., 2005). A first order Markov chain is a stochastic process $N_t : t \in T$ where the distribution of N_{t+1}, knowing all previous values $N_1, ..., N_t$, only depends on N_t. Considering the state space S a first order Markov chain can be define as in Equation (4.2).

$$P(N_{t+1} = s_{t+1} | N_1 = s_1; N_2 = s_2; ...; N_t = s_t) = P(N_{t+1} = s_{t+1} | N_t = s_t) \tag{4.2}$$

These state transitions, from a state N_t to a N_{t+1}, are encapsulated by a conditional function, denoted the probability density function (pdf) as g_t where $\theta \in \Theta$ the vector of all model parameters such as the state process distribution is defined in Equation (4.3).

$$g_t(N_t | N_{t-1}; \theta) \tag{4.3}$$

This function is often challenging to analyze due to be a representation of a sequence of stochastic sub-processes, instead it can be expressed as a product, implying distinct sub-processes with individual probability distributions and its specific parameters as defined in Equation (4.4) (Newman et al., 2009).

$$g_t(N_{t+1}|N_t;\theta) = \prod_i g_{t,i}(N_{t+1,i}|N_{t,i};\theta_i) \tag{4.4}$$

The decomposing of the state process into sub-processes facilitates the description and analysis of specific aspects. The sub-processes produce an essential tool when studying population dynamics, dividing complex systems as a life cycle, which, consequently, makes them easier to model (Newman et al., 2009; Thomas et al., 2005).

The state vector is generally unknown or uncertain. Therefore, to make inferences regarding the state model, it is necessary to consider the data that will inform it. Conversely, the observation process is the bridge between the data collected, y_t, the observation vector, and the state vector N_t. The observational model, as the state process, is modeled by a conditional pdf denoted f_t defined in Equation (4.5).

$$f_t(y_t|N_t;\theta) \tag{4.5}$$

These processes are modeled separately at different levels of the model hierarchy, assuming that states are autocorrelated and observations are independent. This dependence is accounted for in the states. Figure 4.1 illustrates how observation independence is achieved (Auger-Méthé et al., 2020).

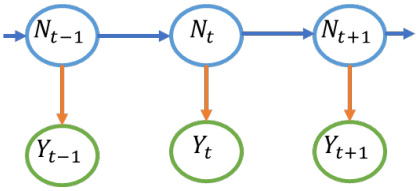


Figure 4.1: Scheme depicting the correlation between non-observational and observational states: Blue circles represent the non-observable states N_t, green circles represent the observational states, blue arrows represent the correlation between non-observational states Y_t, and orange arrows represent the correlation relationships between non-observable and observable states

Combining these processes with an initial state g_0, that can be assumed known, viewed as an unknown constant, or can be described as a stochastic process, defined in Equation (4.6) (Kitagawa and Gersch, 1996).

$$g_0(N_0|\theta) \tag{4.6}$$

The mathematical description of SSM can be done using the three probability density functions, that represent the hierarchical structure of SSM. The top level of hierarchy that links the data to the unknown states, is represented by the observation state Equation (4.5), then the process distribution Equation (4.3) representing a second level where the underlying dynamics of the states are defined, and the initial state Equation (4.6) (Aeberhard et al., 2018).

In summary, the mathematical foundations of SSM provide a framework that captures the interplay between unobservable states and observed data, with a keen focus on addressing uncertainty and variability.

4.2 Bayesian approach applied to State Space Models

The Bayesian approach provides a distinctive perspective compared to frequentist methods. In frequencist approaches each parameter is treated as fixed unknown constants. In contrast, the Bayesian

approach considers parameters as random variables, relying on its posterior density to make inference about it. This fundamental difference enables Bayesian methods to comprehensively account for and quantify uncertainties (Auger-Méthé et al., 2020).

In Bayesian inference for SSM, the models account for process variation and observational uncertainty while also leverages prior distributions to include existing knowledge about parameters, independently of the data used in the model, that were not taken into account in non-Bayesian models (Mäntyniemi et al., 2015).

After defining the questions and hypothesis of a research about $\theta \in \Theta$, the use of a Bayesian approach involves the following steps (van de Schoot et al., 2021):

- Formalize prior distributions, which reflect using a probability distribution, initial beliefs or previous knowledge about θ before seeing the data, denoted as $p(\theta)$.
- Determinate the Likelihood function, which consist in defining the probability distribution of the data (y) conditional to the given parameter vector θ, that is used as constants, $p(y|\theta)$.
- Obtain the posterior distribution, which contains the updated belief about the parameter given the observed data, being the probability distribution of θ as a function of both the specified prior and the likelihood function, $p(\theta|y)$.

This approach adds to the complexity of the SSM previous described, adding two new equations that encompass the bayesian approach. A parameter function that describes the stochasticity in θ where Γ is a hyperparameter, as defined in Equation (4.7). And the second function, the prior function, includes the initial beliefs about those hyperparameters, as defined in Equation (4.8) (Newman et al., 2014).

$$h(\theta|\Gamma) \tag{4.7}$$

$$\pi(\Gamma) \tag{4.8}$$

The SSM formulation now includes those two new functions ((4.7) and (4.8)) that give extra deeper levels to the hierarchical structure of SSM (Aeberhard et al., 2018). Using those five equations (4.3),(4.6),(4.5)(4.8)and (4.7) states and parameters posterior can be described assuming that prior distribution to the unknown state $N_{1:T}$ was defined in Equation (4.9) (Newman et al., 2014).

$$\pi(N_{1:T}, \Gamma|y_{1:T}) \propto f(y_{1:T}|N_{1:T}, \Gamma) \cdot g_0(N_0|\theta) \cdot \prod_{t=1}^{T}[f_t(y_t|N_t, \theta) \cdot g_t(N_t|N_{t-1}, \theta)] \cdot h(\theta|\Gamma) \cdot \pi(\Gamma) \tag{4.9}$$

Priors have a high relevance on a Bayesian approach, reducing the uncertainty about θ, by reflecting previous information about it. They are often characterized by hyperparameters, parameters of the prior distribution (for example mean and variance for a normal prior) (van de Schoot et al., 2021). A non-informative prior function that reflects complete uncertainty about the parameters prior can be used to maximize the influence of the data on the posterior, although they still contribute to shaping the outcomes. Instead, to help ensure that the data inform the model and that the posterior is well behaved, Gelfand and Sahu, 1999 suggested using a weakly informative function, that has information about the parameter but is not as precise as an informative prior (Auger-Méthé et al., 2020).

This Bayesian framework is going to give in this thesis an iterative, refined understanding of parameters, to make more informed inferences about the population dynamics of blackspot seabream. All prior

used and state definitions are presented later in Chapter 5.

4.3 Markov Chain Monte Carlo Algorithms

As discussed in the preceding section, Bayesian inference primarily centers on the posterior distribution. However, this distribution of interest often requires the integration of high-dimensional functions, making it challenging to calculate analytically and computationally complex. Methods based on Monte Carlo methods, like Sequential Importance Sampling (SIS) or Markov chain Monte Carlo (MCMC) attempt to simulate direct draws from the posterior distribution. This is an important feature to address the complexity problem, as detailed in the following section (Ehlers, 2003; Kitagawa and Gersch, 1996; Newman et al., 2009; Roberts and Rosenthal, 2004; Turkman and Paulino, 2015; Walsh, 2004).

The MCMC stands out as one of the most widely used algorithms in the Bayesian statistical community (Roberts and Rosenthal, 2004) and the chosen method for the purpose of Bayesian inference on SSM. It is considered by, Newman et al., 2009, the "gold standard" for cases where data are informative to priors. Therefore, it is the chosen method for this thesis and is introduced in the following sections.

4.3.1 Simple Monte Carlo

To understand MCMC its basics should be understood to simplify a simpler notation is used. Here, $g(\theta)$ is a mathematical function, and the parameter θ has density function $p(\theta)$. Starting with the simple Monte Carlo method, the goal is to calculate an integral writing it as the expected value, as defined in Equation (4.10) (Ehlers, 2003; Turkman and Paulino, 2015).

$$I = \int g(\theta)p(\theta)d\theta = E[g(\theta)] \tag{4.10}$$

Then a simple algorithm to solve this is (Ehlers, 2003):

1. Generate a samples for θ, (θ_1, ...θ_n), using the distribution of $p(\theta)$.
2. Calculate $g(\theta_1)$,...,$g(\theta_n)$.
3. Calculate the average value as $\overline{g} = \frac{1}{n}\sum_{i=1}^{n} g(\theta_i)$
4. Finally $\hat{I} = \overline{g}$

Due to the complexity of the probability distribution $p(\theta)$ obtaining samples became challenging. To solve this, methods like MCMC are used (Walsh, 2004).

4.3.2 Sampling Algorithm on Markov Chain Monte Carlo

Markov chain Monte Carlo serves as a method to generate the samples. It is an iterative method, which means that uses iterative simulation techniques based on Markov chains, making the generated values of the sample not independent (Ehlers, 2003; Roberts and Rosenthal, 2004).

This method implies that the chain is not only a Markov chain described on Equation (4.2) but also that the chain fulfill the following conditions, given X_t ($t \in T$) a Markov chain with state space S and m-step probabilities $p_{ij}^{(m)} = P(X_{t+m} = s_j | X_t = s_i)$, $s_i, s_j \in S$:

- Homogeneity, the transition probabilities are independent of the time. Meaning, for all $P(X_{t+1} = s_j | X_t = s_i) = P(X_{\gamma+1} = s_j | X_\gamma = s_i)$ for all $s_i, s_j \in S$ and all $t, \gamma \in T$;

- Irreducibility, any state can be reached from any other state, i.e., $\forall s_i, s_j \in S \; \exists m \leq 1 : p_{ij}^{(m)} > 0$;
- Aperiodic, which means that there are no absorbing states. In other words, the chain is not forced into some cycle of fixed length between certain states.

Gibbs sampler or Metropolis-Hastings algorithm are the most used algorithms inside the MCMC method, and they are presented in this thesis. They differ in the way that they approach sampling, but the main idea is always to start from an initial value, generate a sample, and use it to update the previous distribution/knowledge. Repeating this sampling and updating until convergence to a stationary distribution, the desired posterior distribution (Ehlers, 2003).

4.3.2.1 Metropolis-Hastings's algorithm

Metropolis-Hastings's Algorithm is commonly used, due to its effectiveness even when there is little information about the distribution of interest.

Defining θ^t as the state at the moment t, $q(\cdot|\theta)$ the proposed distribution and π as the distribution of interest. This algorithm can be resumed in five steps:

1. Start iterations with a initial value for the parameter, $\theta^{(0)}$;
2. Generate a new value θ' from $q(\cdot|\theta)$;
3. Calculate the probability of acceptance of the new value, $\alpha(\theta, \theta')$ defined as $\alpha(\theta, \theta') = min(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)})$ and generate $u \sim U(0,1)$;
4. If $u \leq \alpha$ the new value is accepted and $\theta^{t+1} = \theta'$, on the other hand if $u > \alpha$ the value is reject and $\theta^{t+1} = \theta$;
5. Then a new iteration begins back to step two.

Note that the new value θ^{t+1} only depend on θ^t and not $(\theta^{(0)}, ..., \theta^{t-1})$. Due to its feature to calculating the probabilities of a new value being accepted it is not necessary for π to being entirely known, only the target ratio $\frac{\pi(\theta')}{\pi(\theta)}$., which is fundamental to the Bayesian approach where the posterior distribution is often not entirely known (Ehlers, 2003).

4.3.2.2 Gibbs sampler

In the previous algorithm, depending on the accepting probability, the value could be the same for some iterations if new value were rejected. This does not happen in Gibbs, that is a special case of the previously presented method (Walsh, 2004), where the value is always moving to a new one, i.e. the probability of acceptance is always one.

The d^{th} component of the Gibbs sampler is defined such that θ_d^t leaves all components besides d unchanged and replace the d component by a draw from the full conditional distribution of π conditional on all the other components (Roberts and Rosenthal, 2004).

The algorithm can be resumed in the following steps:

1. Star iteration and set initial values such as $\theta^{(0)} = (\theta_1^{(0)}, ...\theta_d^{(0)})'$;

2. Obtain a new value of $\theta^{(t)}$ from$\theta^{(t-1)}$ from successive generation of values

$$\theta_1^{(t)} \sim \pi(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, ..., \theta_d^{(t-1)}),$$

$$\theta_1^{(t)} \sim \pi(\theta_1|\theta_3^{(t-1)}, \theta_4^{(t-1)}, ..., \theta_d^{(t-1)}),$$

$$...,$$

$$\theta_d^{(t)} \sim \pi(\theta_d|\theta_1^{(t-1)}, \theta_2^{(t-1)}, ..., \theta_{d-1}^{(t-1)})$$

;

3. Then a new iteration begins back to step two.

4.3.3 NIMBLE Package

To implement the SSM using MCMC in this thesis, the R package NIMBLE is employed. NIMBLE stands for Numerical Inference for Statistical Models for Bayesian and Likelihood Estimation it offers flexible means to construct and analyze hierarchical and computationally-intensive statistical models, primarily through a Bayesian approach. This package empowers the project with the ability to employ MCMC to generate new samples and conduct Bayesian inference. The package uses BUGS language to create flexible models and uses C++ to compile the model (de Valpine et al., 2017, 2023a, 2023b).

The first task in implementing the model on NIMBLE is to write the code to the model, where all nodes, variables and their relationships are defined along with the prior distributions. It incorporates the biological aspects described in Chapter 2, considering NIMBLE specifications and language, as explained in Chapter 5. Before starting the MCMC all constants, data and initial values are set.

For each stochastic node, a sampler is assigned, as described in the previous section of this chapter. The default sampler algorithms are assigned as described in Section 7.2.1.1 of the NIMBLE user manual (de Valpine et al., 2023b).

These steps result in obtaining posterior samples and posterior summary statistics. These outputs are useful for analyzing results and making inferences about parameters and states.

In conclusion, in this thesis, in order to implement the SSM using MCMC on the NIMBLE package, the first step involves specifying prior distributions and initial state vectors. Then, on each iteration, a candidate value for each parameter is generated at time t. Subsequently, in a hierarchical manner, the state vectors are sequentially generated, assuming that all components of the state vector at time t are generated simultaneously (Newman et al., 2009). At the end of the iteration, all the states and parameters are updated.

4.4 Convergence of MCMC

Convergence is a critical aspect of MCMC methods. As an iterative technique, it is important for the chain approach its stationary distribution, and the samples of θ are samples from $p(\theta)$. In other way, when the model's states and parameters, henceforth referred to collectively as parameters, reach a stage that can be considered a sample from the posterior distribution.

In MCMC, achieving convergence often requires running long simulations with a sufficiently high number of iterations. The more iterations, the higher the likelihood of the algorithm reaching a representative stationary distribution of the posterior, in other words, reaching the equilibrium state of the chain where the probability distribution if the parameters no longer changes with new iterations. This is

a challenge, as it can be time consuming and require a high computational effort. However, there is no fixed number for the right amount of iterations, and it depends on the complexity of the model (Aeberhard et al., 2018; Newman et al., 2009).

Running multiple chains can be useful to analyze convergence in a more accurate way, and to test different initial values. So typically, more than one chain runs with different initial values. Presenting new challenges, convergence within and between chains and getting the right number of chains, which make some authors believe that, sometimes, one chain is enough (Vehtari et al., 2021). In this thesis, given the level of uncertainty about the parameters, multiple chains were used, this way was possible to explore the effect of different initial values on model results. To maintain computational efficiency five chains were executed.

The first iterations are often referred to as the burn-in period, during this phase the samples may not accurately represent the stationary distribution of interest because the chain has not yet converged. Typically, those samples are disregarded until the algorithm reaches convergence. Determining the appropriate burn-in period is challenging and influenced by factors such as poor initial values or vague prior distributions (Walsh, 2004).

After the burn-in period, considered B iterations, being N the number of iterations that is the sum of the Burn-in B, and the retain samples M, so $N = M + B$, as defined in Equation (4.11).

$$\underbrace{\theta_1, \theta_2, \ldots \theta_B,}_{Burn-in} \quad \underbrace{\theta_{B+1}, \theta_{b+2}, \ldots \theta_N,}_{\text{M Samples from the posterior}} \tag{4.11}$$

A problem affecting model convergence is auto-correlation, causing slow mixing, where consecutive samples are correlated rather than independent, leading to inefficient sampling and biased parameter estimates. One strategy to help with it is thinning the output, i.e. after the burn-in period storing only every i^{th} point (Walsh, 2004), so the number of samples of the posterior distribution after thinning is $\frac{M}{I}$, where I is the sampling interval used in the thinning, as defined in Equation (4.12).

$$\underbrace{\theta_1, \theta_2, \ldots \theta_B,}_{Burn-in} \quad \underbrace{\overbrace{\theta_{B+1}, \theta_{B+2}, \ldots \theta_N,}\ \theta_{B+1}, \theta_{B+1+i}, \theta_{B+1+2i}, \ldots}_{\text{Sample of the posterior distribution after thin}} \tag{4.12}$$

The convergence of the MCMC algorithms can be analyzed using graphs and statistical measures. The R packages *CODA* and *MCMCVis* are used to produce trace and density plots and calculate the potential scale reduction factor ($\hat{R}$) and the effective sample size (ESS) (defined below), for each parameter. The analysis can be found in the Chapter 6.

The trace plot represents the evolution of the specified parameter as a time series. The aim is for the plot to present the convergence of the parameter to a specific value. Getting the convergence point, the parameter will oscillate between the neighborhood of that value, showing a well mixed chain. When there's minimal overlap between traces of distinct chains, suggesting divergent convergence paths and poor mixing between chains. These indications may flag potential issues and that the output of the MCMC algorithm is not a representative sample of the posterior distribution (Plummer et al., 2005). In this thesis, the trace plot function plotted the last 5000 iterations of each chain (Casey Youngflesh, 2023).

Density plots represent the estimated density of the posterior distribution for each parameter. This type of graph is very useful to have a better understanding of the posterior distribution and allow the comparison between the prior and posterior distribution.

In models like the one developed in this thesis, with a relatively large number of parameters, as

explained in Chapter 5, the use of summary statistics became simpler than analyzing plots for each parameter and are relatively quick manner to reveal potential problems in the model (Vehtari et al., 2021).

The potential scale reduction factor, $\hat{R}$ (Brooks and Gelman, 1998; Gelman and Rubin, 1992), is calculated for each parameter a value that is the square root of the ratio of two estimators for the target variance σ^2. Being σ^2 the variance of the target posterior distribution of each parameter. The aim is that all chains are at equilibrium, making $\hat{R}$ near 1, declaring convergence. If the chains do not converge to a common distribution, $\hat{R}$ is going to be greater than one. According to Gelman and Rubin, 1992 the ideal scenario would be sample until all parameters have $\hat{R}$ below 1.1.(Stan Development Team, 2024; Vats and Knudson, 2020; Vehtari et al., 2021). It is considered that $\hat{R}$ can fail to diagnose poor mixing, so this measure is not enough to make a convergence diagnostic and should not be used as a default rule. Instead, should be analyzed in parallel with other statistics, such as the ESS (Vehtari et al., 2021).

The effective sample size answers the question, "How many independent samples from the posterior distribution contain the same amount of information?" (Plummer et al., 2005; Vats and Knudson, 2020). So, the higher the ESS the better the convergence, Vehtari et al., 2021 suggests considering as a standard recommendation rely on the $\hat{R}$ measure if ESS value is at least 50 if only one chain runs or at least 400 otherwise. Properly managing the burn-in period and the total number of samples is crucial for obtaining a high ESS by ensuring that the samples used for analysis are representative of the target distribution and have lower autocorrelation.

Nevertheless, in this thesis, both methods, graphical and summary statistics, were analyzed in order to have a better view of if and where potential problems appear.

Chapter 5

The Model

This chapter outlines the application of the methodologies of Chapter 4 to the blackspot seabream data presented in Chapter 3.

The methodologies previously presented offer a powerful framework for modeling complex systems, in particularly for blackspot seabream in mainland Portugal. Its specific features give the possibility of dividing the state process from the observational process. The state process of the model used here describes the biological characteristics of the species as described in Chapter 2. The complexity of population dynamics and the life cycle of blackspot seabream, renders the state process function, g_t, hard to formulate and analyze. To accommodate the complexity of the state process, this is decomposed into the sub-processes: Growth, Survival and Mortality, and Reproduction and Recruitment. The initial sub-process, shapes the population distribution, affecting the fish distribution between size-classes. The remaining sub-processes influence the population composition, causing the maintenance, removal, or introduction of individuals within the existing population.

Each sub-process is characterized by a function, which is, in turn, associated with a stochastic process. The g_t function is thus a set of several stochastic processes as defined in Equation (5.1)(Newman et al., 2009; Whitlock et al., 2018). The strategy adopted allows the use of a simpler framework, facilitating the implementation of a complex model and granting the analysis of the state vector after each sub-process. It also allows for a detailed analyses of the population state at each time-step. In modelling procedure the sub-processes will be sequential, though in nature, some processes happen simultaneously (Newman et al., 2009).

$$g_t(N_t|N_{t-1},\theta) = g_{3,t}(g_{2,t}(g_{1,t}(N_{t-1},\theta),\theta),\theta) \tag{5.1}$$

Where $g_{1,t}$, $g_{2,t}$ e $g_{3,t}$ are the pdfs representing the processes of Reproduction and Recruitment, Survival and Mortality and Growth. And θ is the parameters vector.

The observational process links the model to data on blackspot seabream catches in mainland Portugal (Chapter 3). This data informs the model about population parameters and states.

Model initialization requires priors distributions and initial values. The priors, as explained in Chapter 4, are critical and challenging, particularly in data-scarce scenarios, and are informed by previous studies. Selecting appropriate initial values can significantly enhance the efficiency of the model, because well-chosen initial values can help the model converge more quickly (de Valpine et al., 2023b). Prior research offers valuable guidelines for these selections. Following a suggestion made in Walsh, 2004, the initial values where defined close to the center of the prior distribution, as it can facilitate faster convergence.

The models' structure is based on the general population dynamics model in Mäntyniemi et al., 2015;

Mäntyniemi, Haapasaari, et al., 2013 and Blomstedt et al., 2015. It is based on a size-class structured population model derived from the blackspot seabream life cycle information presented in the dedicated Chapter 2 and the size-classes presented in Chapter 3.

5.1 Population Dynamics Processes

Considering that the total number of fish in the beginning of the year $t+1$, N_{t+1}, is equal to the number of fish that survive year t ($N_t S_t$, where S_t is the proportion of survival the time-step t) plus the recruits entering the population at the beginning of year $t+1$ (R_{t+1}), that were born at year t.

$$N_{t+1} = N_t S_t + R_{t+1} \tag{5.2}$$

However, as explained in Mäntyniemi et al., 2015, as data consists in catches, they provide more information on temporal changes in population size rather than its absolute abundance. As a result, states of the model are going to have a high posterior correlation with the general population abundance and size of the population, of each time-step. To address these challenges and enhance the convergence of the MCMC, a reparameterization is used. The reparameterization involves a new N^* parameter, representing the overall abundance in the first year, and redefines all subsequent parameters on a relative scale instead of using absolute numbers. This approach helps in dealing with the high posterior correlations and improves the efficiency of the MCMC algorithm by stabilizing the parameter estimation process and reducing the complexity of the model (Mäntyniemi, Haapasaari, et al., 2013).

Having a total number of individuals on the first time-step ($t = 1$) defined as N^*, and setting $N_1 = 1$, it is possible to describe the total number of individuals on time-step t, on relative scale to the first time-step. The absolute population size at time-step t can be obtained as the product $N^* N_t$. With this new parameterization the Equation (5.2) is the same but R_{t+1} is, as N_t, no longer the absolute number, it is the proportional number.

Fish population dimensions are usually unknown, so for lack of previous information and inability to arrange a stable estimated value, N^* is set as a constant equal to 1×10^7.

Instead of working with the number of fish in each size-class, the model works with vectors that represent the distribution of the population at time-step t on the ten size-classes. The vector $\Phi_t = (\phi_{t,1}, \phi_{t,2}, ..., \phi_{t,10})$ defines the state of the population at the beginning of the time-step t, in terms of relative size-class frequencies (Mäntyniemi et al., 2015). To give some variation this vector follows a Dirichlet distribution.

The forthcoming sections will expand upon the varied assumptions and notations underpinning each sub-process in the model. To facilitate this classification, a set of notations is introduced on Table 5.1. A description of all the parameters used in the model and its notation used in code are in the Table A.1 in Annex A.

Table 5.1: Initial notation used along the model

Notation	Meaning
i and j	Size-class (see Table 3.2)
t	Time-step in years. $t \in \{1,..,8\}$ corresponding to the collection period spanning eight years from 2014 to 2021
I	Breakpoints for the Size-class i, where I_i and I_{i+1} are, respectively, the lower and upper bounds for the Size-class i. (see Table 3.2)
l_i	Midpoint of size-class i (see Table 3.2)

5.1.1 Reproduction and Recruitment

This process is a complex one for blackspot seabream, as explained in Chapter 2. The reproduction of the species depends on several parameters, such as the number of females and their stage of maturity.

Beginning with the Recruits, that are considered to be individual born at t that join the general population, at size-class 1, at $t+1$. The proportional number of recruits, R_{t+1}, defined by the Equation (5.3), is calculated based on the total number of survival eggs produced by the mature females on year t. Being E_t the proportional total number of eggs produced at the time-step t and π^E the survival probability of the egg until recruitment (Mäntyniemi et al., 2015).

$$R_{t+1} = E_t \pi^E \tag{5.3}$$

Note that, it is assumed that $R_1 = 0$, so there is no recruitment on the first time-step. And to simplify the model π^E was set constant, with a value of 1×10^{-6}, based on values presented in Mäntyniemi, Haapasaari, et al., 2013 and Mäntyniemi, Uusitalo, et al., 2013.

The value of E_t is calculated multiplying the number of eggs produced by size-class (Egg_i) by the relative number of individuals in that size-class $(\phi_{t,i}N_t)$ as defined in Equation (5.4).

$$E_t = N_t \sum_{i=1}^{Class} Egg_i \phi_{t,i} \tag{5.4}$$

The spawning capacity female of each size-class, defined in Equation (5.5) is given by the fecundity of the mature females on each size-class, Fec_i, having in account the probability of being a mature female on each size-class, $\pi_i^F \pi_i^M$, where π_i^F and pi_i^M are the probability of being female and being mature in each size-class respectively.

$$Egg_i = \pi_i^F \pi_i^M Fec_i \tag{5.5}$$

Although, some previous studies belief that the proportion of mature females change over the years depending on the abundance, to simplify the model and reduce computational effort these parameters do not change over the years. It is assumed that the probability of being female in each size-class is defined as in Equation (5.6), based on equations presented in Krug, 1998, and that γ_i^F follows a Normal distribution with a mean value such as: $E[\gamma_i^F] = \alpha_F + \beta_F l_i$, and a standard deviation σ_F.

$$\pi_i^F = \frac{e^{\gamma_i^F}}{1 + e^{\gamma_i^F}} \tag{5.6}$$

Re-parameterizing, using the length where there is 50% of individuals are male and 50% are female,

$L_F^{(50)} = \frac{-\alpha_F}{\beta_F}$ it is possible to define prior distributions for $L_F^{(50)}$ and for β_F, which controls the curve's softness. After re-parameterizing $E[\gamma_i^F] = -L_F^{(50)}\beta_F + \beta_F l_i$ (Krug, 1998).

A similar approach is used to define the probability of being mature in each size-class, based on Equation 1 of Krug, 1998, defined at Equation (5.7). Where γ_i^M follows a Normal distribution with a mean value such as: $E[\gamma_i^M] = \alpha_M + \beta_M l_i$, and standard deviation σ_M. Re-parameterizing, in terms of the length where there are 50% of individual mature and 50% immature, $L_M^{(50)} = \frac{-\alpha_M}{\beta_M}$ setting prior distributions for $L_M^{(50)}$ and for β_M, which controls the softness curve. After re-parameterize $E[\gamma_i^M] = -L_M^{(50)}\beta_M + \beta_M l_i$ (Krug, 1998).

$$\pi_i^M = \frac{e^{\gamma_i^M}}{1 + e^{\gamma_i^M}} \tag{5.7}$$

The fecundity is considered to follows a Log-Normal distribution, facilitating some calculus, once the logarithm of the variable is normal distributed and ensuring that the fecundity can not be negative. The mean value of fecundity is based on Krug, 1986 equation that defines fecundity of blackspot seabream, $E[Fec_i] = \frac{l_i^{\beta_{Fec}}}{\alpha_{Fec}}$. And the standard deviation of the distribution is defined using the variation coefficient CV_{Fec}.

All hyperparameters related to the Reproduction and Recruitment process described above, are discriminated on Table 5.2 as their prior distributions and references.

Table 5.2: Hyperparameters from the Reproduction and Recruitment process

Parameter	**Prior**	**Mean***	**Sd***	**Reference**
$L_F^{(50)}$	Normal	29	1	Based on values of Krug, 1986
β_F	Normal	0.2	0.06	Based on values Table 4 of Krug, 1998
σ_F and σ_M	Gamma	1.4	0.1	(Arbitrarily)
$L_M^{(50)}$	Normal	34	1	Based on values of Table 1 on (Krug, 1998)
β_M	Normal	0.5	0.3	Based on values on Table 1 of Krug, 1998
α_{Fec}	Normal	9.3	1	Based on values of Krug, 1986
β_{Fec}	Normal	6.9	1	Based on values of Krug, 1986
CV_{Fec}	Uniform	0.5	0.29	(Arbitrarily)

*These are rounded values

5.1.2 Survival and Mortality

Individuals are going to be subject to the Survival and Mortality process. Here there are different possible out-comings: surviving (S), or death, being the last one divided in two, being catch (C) and natural (D) death. Each individual can experience only one of those outcomes.

The probabilities of each outcome are constant over each time-step and all individuals in the same size-class have the same probability of each event. For each time-step and each size-class, the vector of probabilities $(\pi_{t,i}^S, \pi_{t,i}^C, \pi_{t,i}^D)$ can be defined using the Baranov equations as they are in Equation (5.8), where $Z_{t,i}$, the total mortality rate, is the sum of fishing mortality and natural mortality $(Z_{t,i} = F_{t,i} + M_{t,i})$ (Blomstedt et al., 2015).

$$\pi_{t,i}^S = e^{-Z_{t,i}}$$

$$\pi^{C}_{t,i} = \frac{F_{t,i}}{Z_{t,i}}(1 - e^{-Z_{t,i}}) \tag{5.8}$$

$$\pi^{D}_{t,i} = \frac{M_i}{Z_{t,i}}(1 - e^{-Z_{t,i}})$$

The natural mortality is assumed to follow a Log-Normal distribution, which ensures that those values are positive, where the standard deviation is defined using the variation coefficient ,CV_D, and the expected value of natural mortality, depending on the length, is based on Charnov defined in Equation (3) (Charnov et al., 2012), $E[M_i] = k(\frac{l_i}{L_\infty})^{-1.5}$, being k and L_∞ growth parameters presented in Table 5.4. Note that for better model performance, the natural mortality in the first size-class is fixed to 0.25.

Fishing mortality is given by the multiplication of maximum instantaneous fishing mortality at year t, F^{max}_t, and selectivity of the fishing gear for each size-class, f_i, as it is defined in Equation (5.9).

$$F_{t,i} = F^{max}_t f_i \tag{5.9}$$

The maximum fishing mortality is assumed to vary over each time-step, F^{max}_t follows a Log-Normal distribution with a mean value $\mu_{F^{max}}$ and where the standard deviation is defined using the variation coefficient $CV_{F^{max}}$.

The selectivity of the fishing gear is defined similarly to the female and mature probability, as defined in Equation (5.10). Where γ^f_i follows a Normal distribution with a mean value based on Czerwinski et al., 2009 such as $E[\gamma^f_i] = \alpha_f + \beta_f l_i$, and a standard deviation σ_f. Re-parameterizing, in terms of the length where there is 50% of selectivity, $L^{(50)}_f = \frac{-\alpha_f}{\beta_f}$ where it is define prior distributions for $L^{(50)}_f$ and for β_f, which controls the softness of the selection curve. After re-parameterize $E[\gamma^f_i] = -\beta_f L^{(50)}_f + \beta_f l_i$.

$$f_i = \frac{e^{\gamma^f_i}}{1 + e^{\gamma^f_i}} \tag{5.10}$$

The proportion of survivors in each year (S_t) follows a Beta distribution, with the expected value being the sum of proportional survivors from each size-class, such as $E[S_t] = \sum_{i=1}^{10} \pi^S_{t,i}\phi_{t,i}$ (Mäntyniemi et al., 2015).

The size-class distribution of the survivors on time-step t is represented by the vector $\Phi^S_t = (\phi^S_{t,1}, ..., \phi^S_{t,10})$, and once again, to give some variation, this vector follows a Dirichlet distribution.

All hyperparameters related to the Survival and Mortality process described above, are discriminated on Table 5.3 as their prior distributions and references.

Table 5.3: Hyperparameters from the Survival and Mortality process

Parameter	**Prior**	**Mean***	**Sd***	**Reference**
CV^D and $CV_{F^{max}}$	Uniform	0.5	0.29	(Arbitrarily)
$\mu_{F^{max}}$	Log-Normal	0.4	0.25	Based on values of Blomstedt et al., 2015
$L^{(50)}_f$	Normal	31	1	Based on values from Table 3 of Czerwinski et al., 2009
β_f	Normal	0.35	0.1	Based on values from Table 3 of Czerwinski et al., 2009
σ_f	Gamma	1.4	0.1	(Arbitrarily)

*These are rounded values

5.1.3 Growth

The growth process consists in moving from a size-class to another. In each time-step, the age distribution of the survivors Φ_t^G is given by the multiplication between the vector of survivors Φ_t^S and the probability of moving to a size-class j, knowing that in the previous time-step, it was on size-class i ($g_{i,j}$). This probability, defined in Equation (5.11), rely on the expected fish length at year $t+1$, knowing that it was in size-class i on year t (μ_L^i), where σ_L is the correspondent standard deviation is fixed to 1, and μ_L^i is defined using the Von Bertalanffy curve as $\mu_L^i = (L_\infty - l_i)(1-\exp(-k)) + l_i$. For simplicity, this parameter is assumed to be constant over the years. To add more variation, g follows a Dirichlet distribution.

$$g_{i,j} = \frac{\Phi(\frac{I_{j+1}-\mu_L^i}{\sigma_L}) - \Phi(\frac{I_j-\mu_L^i}{\sigma_L})}{\Phi(\frac{I_{m+1}-\mu_L^i}{\sigma_L}) - \Phi(\frac{I_1-\mu_L^i}{\sigma_L})} \tag{5.11}$$

The vector $\Phi_t^G = (\phi_{t,1}^G, \phi_{t,2}^G, ..., \phi_{t,i}^G)$ represents the state of the population at the end of the year t, after the population got through all the sub-process. This vector is used to update the Φ_t vector presented on the beginning as the size-class distribution, at the being each time-step.

All hyperparameters related to the Growth processes described above are discriminated on Table 5.4 as their prior distributions and references.

Table 5.4: Hyperparameters from the Growth process

Parameter	**Prior**	**Mean***	**Sd***	**Reference**
k	Uniforme	0.12	0.02	Based on values from Table 5 of Lorance, 2011 and Erzini et al., 2006
L_∞	Normal	57	3	Based on values from Table 5 of Lorance, 2011 and Erzini et al., 2006

*These are rounded values

5.2 Observation Model

The observation model establishes the connection between observed data and the state process. The data is assumed to be observed with errors, as described before, so it follows a distribution with support based on the estimate values.

Two sources inform the observation model: total catch per size-class in numbers and the size-class distribution of catches. Although the fundamental processes remain consistent across both approaches, each data type is linked to the model in distinct manners.

Firstly, the observed total catch, $N_{t,i}^O$, follows a Normal distribution, centered around $N_{t,i}^C$, the estimated number of fish caught by the model (defined in Equation (5.12)), and with standard deviation of σ_O (Mäntyniemi, Haapasaari, et al., 2013). Where σ_O was set constant in each size-class $\sigma_O = (10, 10, 1000, 1000, 10000, 10000, 1000, 1000, 100)$. Equation (5.13) defines this part of the observational model.

$$N_{t,i}^C = \pi_{t,i}^C \phi_{t,i} N_t N^* \tag{5.12}$$

$$N_{t,i}^O \sim Normal(N_{t,i}^C; \sigma_i^O) \tag{5.13}$$

Secondly, as defined in Equation (5.14), the observed size-class distribution of catches ,ϕ_t^O, follows a Multinomial distribution with probabilities π_t^C, and the total sample size adjusted to 10000 fish caught per year.

$$\phi_t^O \sim Multinomial(\pi_t^C; 10000) \tag{5.14}$$

The Figure 5.1 is a simplified illustration of how the model behave between two time-steps N_{t-1} and N_t having in attention the processes mentioned before. The code developed based on the model described can be found in the repository created for the purpose of this thesis[1].

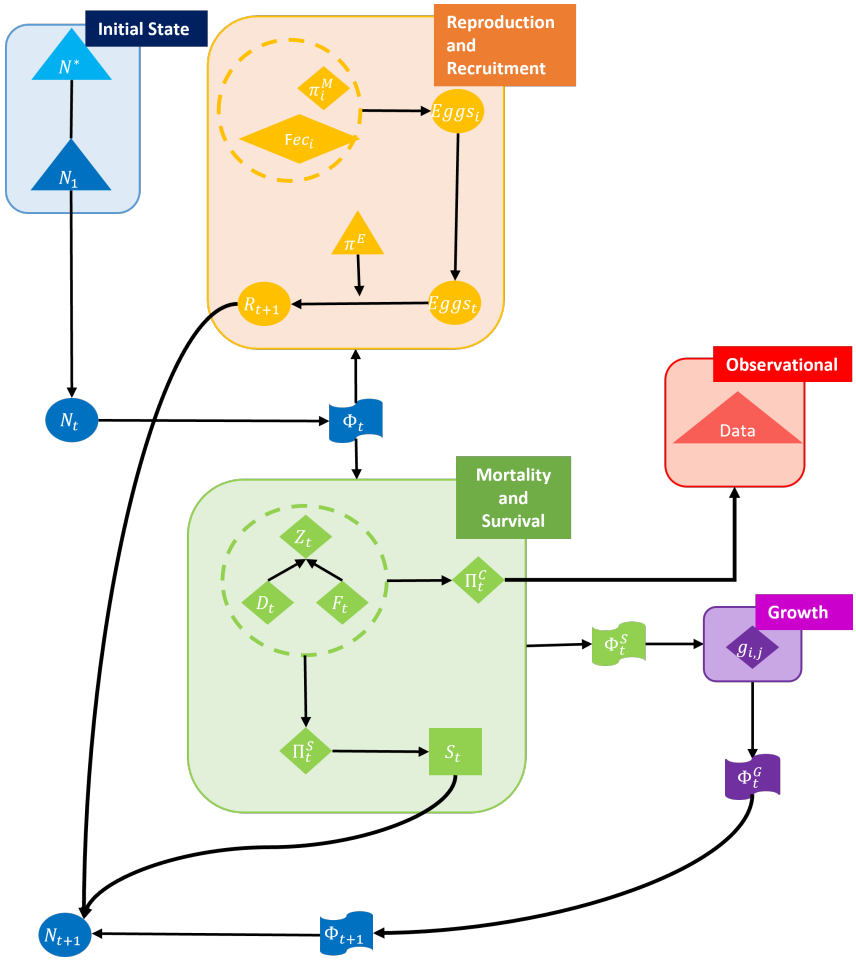


Figure 5.1: Diagram illustrating the developed model: Each rectangle represents a sub-process of the model, with blue indicating the initial state, in yellow representing reproduction/recruitment processes, green denoting survival/mortality sub-processes, purple indicating growth, and red representing observational process. Flag icons represent state vectors, triangles symbolize constants, circles depict states, and diamonds represent probabilities and taxes. All notation used in this diagram is presented in the above text and Table A.1 in Annex A

[1]GitHub repository:Pagellus bogaraveo - Bayesian SSM

Chapter 6

Analyzes of Results

In this section, the results obtained undergo analysis for convergence and biological acceptance. As mentioned before, all results were obtained using the R package NIMBLE (de Valpine et al., 2023b).

The initial number of iterations was 11000000, with a burn-in period of 1000000, and a thin of 1000. It took 10000 samples per chain, with a total of 5 chains executed. Each successive chain initiated with the final value of the preceding chain. The computational execution lasted approximately 62 hours ($\pm$2.5 days).

The same order used in Chapter 2 and Chapter 5 is maintained to simplify analyses. For each process, summary statistics tables, trace plots, and density plots, generated using the *MCMCvis* package, are analysed (Casey Youngflesh, 2023).

6.1 Reproduction and Recruitment

The hyperparameter results, presented in Table 6.1, align well with biological expectations. Specifically, the values are consistent with those previously reported for this species, and no anomalous or illogical values were observed.

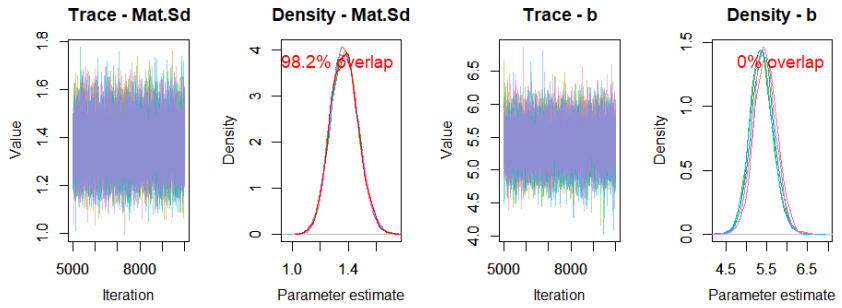
A comparison with the initial assumptions presented in Table 5.2 indicates that all hyperparameter results, except for β_{Fec}, are closer to the initial beliefs, suggesting a posterior distribution overlapping the prior distribution for most parameters. Having in mind the previous mentioned measures that evaluate convergence, all have $\hat{R}$ near 1, all smaller than 1.1, and high values of ESS ($\geq$ 15130).

Table 6.1: Summary of results obtained for hyperparameters from the Reproduction and Recruitment process. Displays measures considering the five chains: the mean, quantiles (2.5%, 50% and 97.5%), Gelman-Rubin convergence statistic ($\hat{R}$), and the number of effective samples (ESS)

Parameter	**Mean**	**Sd**	**2.5%**	**50%**	**97.5%**	$\hat{R}$	**ESS**
$L_F^{(50)}$	29.024	1.008	27.052	29.025	30.999	1	50000
β_F	0.193	0.062	0.072	0.194	0.314	1	48428
$L_M^{(50)}$	34.01	1.009	32.04	34.004	35.981	1	49954
β_M	0.466	0.304	-0.068	0.459	1.076	1	15439
α_{Fec}	9.761	0.968	7.871	9.758	11.669	1	44020
β_{Fec}	5.382	0.323	4.722	5.392	5.984	1.02	15130
σ_F	1.372	0.098	1.186	1.371	1.571	1	49558
σ_M	1.374	0.098	1.188	1.372	1.571	1	50345
CV_{Fec}	0.547	0.287	0.031	0.572	0.981	1.01	35815

By observing trace and density plots, obtained using the *MCMCtrace* function of the MCMCvis package. Reproduction and Recruitment hyperparameters have converged, with trace plots showing well-mixed chains oscillating around a specific neighborhood. On the density plots, only β_{Fec} exhibits noteworthy differences between its prior and posterior distributions. All graphs can be found in the repository created for the purpose of this thesis[1].

A representative portion of the graphical representations mentioned above (see Figure 6.1) showcases an example from the hyperparameter σ_M, which exhibited $\hat{R}$ equal to 1 and had the highest ESS of all hyperparameters in Table 6.1. Conversely, the parameter β_{Fec} that had the poorest measures of convergence and diverged more from initial beliefs. In both trace plots, values stabilized around certain ranges and the chains were well-mixed. However, the density plots, reveal that while σ_M presents an overlap between the prior and posterior distributions of 98.2%, β_{Fec} shows no overlap (0%).



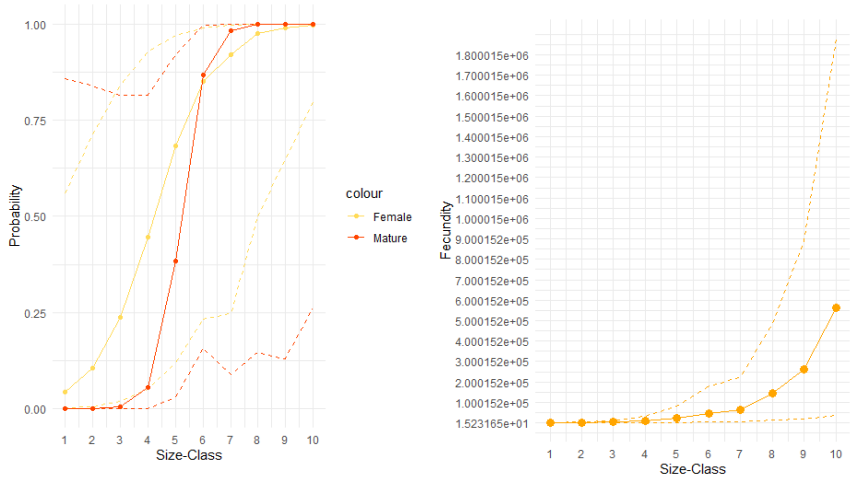
(a) Trace and density plots of parameter σ_M (b) Trace and density plots of parameter β_{Fec}

Figure 6.1: Trace and density plots from the Reproduction and Recruitment process' hyperparameters. The Graph (a) illustrate the trace and density plots of σ_M and in Graph (b) the trace and density plot of β_{Fec}. The red line on the density plot represents the prior distribution, while other colors, in both trace and density plots, denote different chains. On the right density plot the red line is not visible given the 0% overlap between prior and posterior distribution

[1]GitHub repository:Pagellus bogaraveo - Bayesian SSM

These hyperparameters are used to determine the distribution of other parameters within this process, constituting a second level of hierarchy, which includes Fec_i, γ_i^F, and γ_i^M. The summary statistics of these parameters are presented on Table B.1 in Annex B, and their respective trace and density plots are presented in the repository[1]. All these parameters demonstrate biological plausibility, supported by convergence measures and trace and density plots that exhibit similar patterns to the hyperparameters. All parameters exhibit $\hat{R}$ values near 1 ($\hat{R} \in [1; 1.11]$) and possess a high number of effective sample sizes (ESS $\in [3546; 50514]$).

Analysis of the probability of being female and mature in each size-class, depicted in Graph (a) of Figure 6.2, reveals expected behavior, as explain in Chapter 2. It is noteworthy that the credibility interval (95%) for the probability of a female be mature, it remains wide across all size-classes, while the female probability narrows as the size-classes advance, indicating more precise estimates of the probability of being female in higher size-classes. Fecundity in each size-class follows the anticipated pattern, ascending through the different size-classes, as illustrated in Graph (b) of Figure 6.2. Initially low in the first 5 size-classes, the mean fecundity value doubles upon reaching size-class 6, with increments becoming clearer thereafter. The dashed lines representing the credibility intervals suggest that as fish size (or size-class) increases, uncertainty also tends to increase.



(a) Graph depicting the probability of being female (yellow) and being mature (red) within each size-class

(b) Graph depicting the fecundity within each size-class

Figure 6.2: Graphs illustrating reproductive parameters (π^F, π^M and Fec) across different size-classes: In both Graphs (a) and (b), the points represent the posterior mean value and the dashed lines the credibility intervals of 95%

Lastly, Figure 6.3 presents the estimated proportional number of recruits for each year (R_t). The mean trend over time exhibits fluctuations, with years 7, 8 and 9 having lower mean values but never reaching zero, and particularly notable peaks in years 2 and 4, where the proportional number of recruits is higher compared to other years, with wider credibility intervals. This is further analyzed and discussed in Chapter 7.

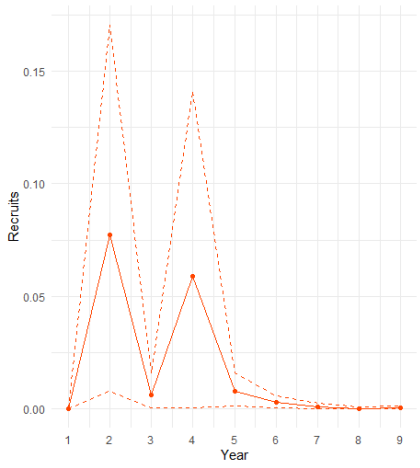


Figure 6.3: Graph showing the proportional number of recruits by year: the points represent the posterior mean value and the dashed lines the credibility intervals of 95%

6.2 Survival and Mortality

Regarding the Survival and Mortality process, the summary of the hyperparameters obtained from the model are presented in Table 6.2. Observing the results in this process, there are some differences between the prior and posterior distributions, even though they still biological accepted, meaning that no anomalous or illogical values were observed. One example of it is the hyperparameter β_f, that shows a posterior distribution with a much lower mean value, and lower standard deviation.

When analyzing the convergence measures, the hyperparameters $\mu_{F_{max}}$, $CV_{F_{max}}$ and CV_D show a $\hat{R}$ equal to 1.1 or 1, and a high number of ESS, even though CV_D has less than the other two, it still high. The other tree hyperparameters β_f, $L_f^{(50)}$ and σ_f show higher $\hat{R}$ ($\hat{R} \in [1.18; 2.36]$), leading to believe in some convergence issue, despite their high ESS values (ESS $\in [41578; 48802]$).

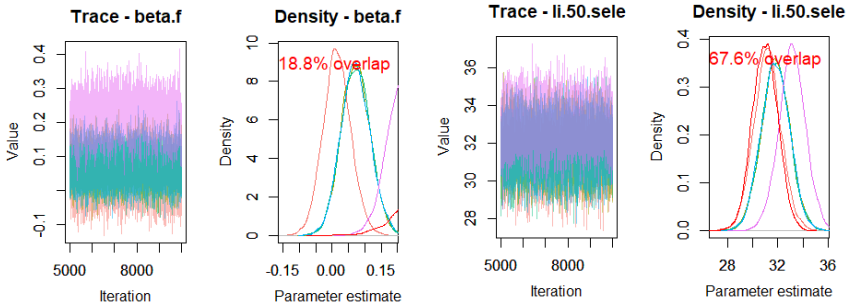
Table 6.2: Summary of results obtained for hyperparameters from the Survival and Mortality process. Displays measures considering the five chains: the mean, quantiles (2.5%, 50% and 97.5%), Gelman-Rubin convergence statistic ($\hat{R}$), and the number of effective samples (ESS)

Parameter	**Mean**	**Sd**	**2.5%**	**50%**	**97.5%**	$\hat{R}$	**ESS**
$\mu_{F^{max}}$	0.97	0.192	0.649	0.951	1.399	1	40392
$CV_{F^{max}}$	0.973	0.026	0.904	0.98	0.999	1	49264
CV^D	0.2	0.077	0.087	0.187	0.383	1.11	12087
β_f	0.091	0.08	-0.035	0.075	0.274	2.35	41578
$L_f^{(50)}$	31.956	1.274	29.537	31.923	34.51	1.27	48802
σ_f	1.964	0.103	1.766	1.963	2.168	1.18	44995

When analyzing the trace and density plots of the hyperparameters, all presented in the repository [1], it is possible to confirm the previous information. All chains in the trace plots show oscillations, smooth, overlapping traces with no obvious trends, and density plots show fusion between chains, except for β_f, as illustrated in Graph (a) of the Figure 6.4, which exhibits varying results in different chains, there is no good representation of mixing between chains, and the different color are more identifiable. Furthermore, in contrast with what happen in the Reproduction and Recruitment process, all hyperparameters exhibit

[1]GitHub repository:Pagellus bogaraveo - Bayesian SSM

low overlap ($\leq 33.8\%$) between the posterior and the prior distribution, indicating a departure from prior beliefs, as in Graph (a) of Figure 6.4, $L_f^{(50)}$ is an exception, it shows an overlap with the prior distribution of 67.6%, as depicted in Graph (b) of Figure 6.4.



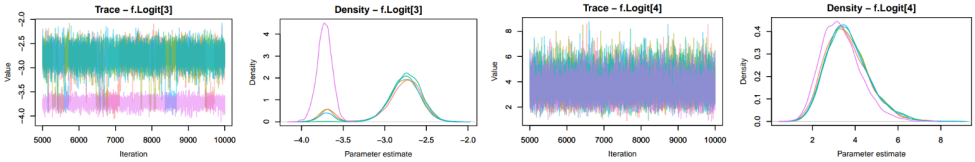
(a) Trace and density plots of parameter β_f (b) Trace and density plots of parameter $L_f^{(50)}$

Figure 6.4: Trace and density plots from the Survival and Mortality process' hyperparameters. The Graph (a) illustrate the trace and density plots of β_f and Graph (b) the trace plot of $L_f^{(50)}$. The red line on the density plot represents the prior distribution, while other colors, in both trace and density plots, denote different chains

These hyperparameters are used to estimate other Survival and Mortality parameters, including F_t^{max}, M_i, and γ_i^f. The summary statistics for these parameters are included in Table B.2 of Annex B, and all trace and density plots are also provided in the repository [1].

The convergence of these parameters differs from that of the Reproduction and Recruitment process, with some parameters displaying poorer convergence diagnosis, characterized by higher $\hat{R}$ values and lower ESS. Specifically, the parameters associated with size-classes 2, 3, and 10 (γ_2^f, γ_3^f, γ_{10}^f, M_2, M_3, and M_{10}), as well as those pertaining to years 7 and 8 (F_7^{max} and F_8^{max}) but, not quite as severe. In contrast, all other parameters appear to demonstrate convergence, as they exhibit lower $\hat{R}$ values ($\hat{R} \in [1;1.05]$) and higher ESS ($ESS \in [10128;45960]$).

Graph (a) of Figure 6.5 depicts the parameter γ_3^f, which serves as a clear example of poor mixing between chains, despite apparent convergence within each individual chain. This phenomenon is also observed in the other mentioned parameters (γ_2^f, γ_3^f, γ_{10}^f, M_2, M_3, and M_{10}, F_7^{max}, and F_8^{max}). However, this issue is not observed in the other parameters, which trace and density plots, exhibit well-mixed chains, as illustrated in Graph (b) of Figure 6.5.



(a) Trace and density plots of parameter γ_3^f (b) Trace and density plots of γ_4^f

Figure 6.5: Trace and density plot from the Survival and Mortality process. The Graph (a) illustrate the trace and density plots of γ_3^f and Graph (b) the trace plot of γ_4^f. Difference colors, in both trace and density plots, denote different chains

[1]GitHub repository:Pagellus bogaraveo - Bayesian SSM

To explore the probability of each of the three outcomes in each size-class by year (natural mortality, fishing, and survival), Figure 6.6 illustrates the mean value of each probability for each of the three outcomes evaluated across different size-classes and years. Upon analyzing the differences between years, it becomes evident that the various outcomes exhibit similar behavior in the initial size-classes across the years. However, they display distinct patterns across the other size-classes in different years. In $t = 7$ and $t = 8$ the fishing mortality mean value, at the middle size-classes, reach higher probability values than in the other years.

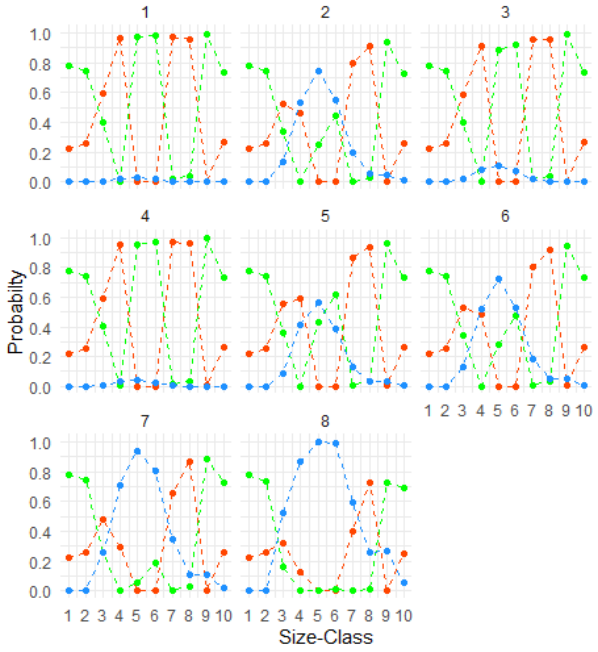


Figure 6.6: Graph illustrating how the probability of the three outcomes changes across size-classes in each year: Each circle represents the posterior mean value of each probability. The red line represents natural mortality, green represents survival, and blue represents fishing mortality

6.3 Growth

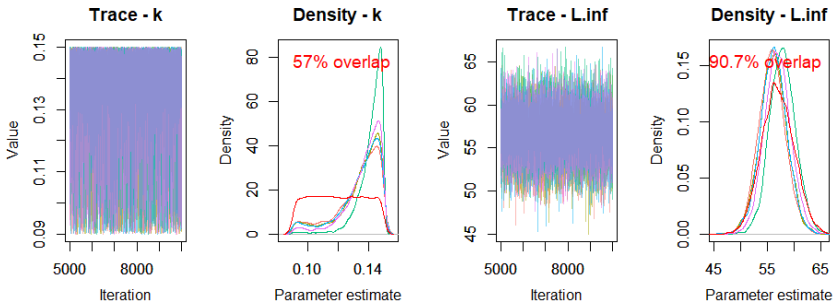
The summary of the Growth hyperparameters obtained from the model is presented in Table 6.3. k shows a higher mean value and smaller standard deviation, while L_∞, with smaller differences, shows a lower mean value and standard deviation. Both hyperparameters show $\hat{R}$ values close to 1 and high ESS values, representing good convergence measures.

Table 6.3: Summary of results obtained for hyperparameters from the Growth process. Displays measures considering the five chains: the mean, quantiles (2.5%, 50% and 97.5%), Gelman-Rubin convergence statistic ($\hat{R}$), and the number of effective samples (ESS)

Parameter	**Mean**	**Sd**	**2.5%**	**50%**	**97.5%**	$\hat{R}$	**ESS**
k	0.135	0.014	0.094	0.139	0.15	1.05	39610
L_∞	56.478	2.603	51.353	56.46	61.644	1.06	40222

The convergence can also be observed in the trace and density plots depicted in Figure 6.7, on Graph (a), it is apparent that k, originated from a vague prior, converged to a posterior distribution with a higher

mean value and smaller standard deviation. Meanwhile, L_∞ demonstrates convergence and appears to be well-mixed, with the posterior distribution overlapping with the prior distribution by 90.7%.



(a) Trace and density plots of k (b) Trace and density plots of L_∞

Figure 6.7: Trace and density plots of Growth parameters (k and L_∞):The red line on the density plots represents the prior distribution, while other colors, in both trace and density plots denote different chains.

The hyperparameters were used to estimate the distribution of the probability of changing size-class. The trace and density plots for each parameter, are included in the repository [1]. However, analyzing Table B.3 of Annex B, it is notable that most parameters display $\hat{R}$ values close to 1, and some slightly higher ($\hat{R} \in [1;2.78]$). High values of ESS (ESS $\in [34711;50197]$) where also seen.

Based on the obtained values, the Table 6.4 presents the mean value of the size-classes change probabilities, rounded to 3 decimal places. Observing this table, it becomes apparent that up to the 4^{th} size-class (inclusive), the highest probability is for a fish in size-class i at year t to move to size-class $i+1$ at year $t+1$. From the 5^{th} size-class forward, it becomes more likely for a fish to remain in the same class.

Table 6.4: Probability of size-class changes: Each cell represents the posterior mean value of the probability of transitioning (rounded with 3 decimal places) to the size-class indicated by the row if in the previous year the individual was in the size-class indicated by the column

From (at $t-1$) / To (at t)	**Class 01**	**Class 02**	**Class 03**	**Class 04**	**Class 05**	**Class 06**	**Class 07**	**Class 08**	**Class 09**	**Class 10**
Class 01	0.05	0.002	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.002
Class 02	0.748	0.062	0.001	0.004	0.033	0.018	0.001	0.001	0.002	0.001
Class 03	0.035	0.923	0.141	0.025	0.043	0.022	0.002	0.001	0.001	0.001
Class 04	0.023	0.003	0.847	0.245	0.045	0.023	0.001	0.001	0.002	0.002
Class 05	0.022	0.002	0.002	0.597	0.345	0.023	0.001	0.001	0.002	0.001
Class 06	0.021	0.001	0.001	0.022	0.342	0.599	0.002	0.001	0.002	0.001
Class 07	0.027	0.001	0.002	0.028	0.052	0.247	0.879	0.003	0.002	0.002
Class 08	0.023	0.002	0.001	0.025	0.045	0.024	0.11	0.954	0.012	0.001
Class 09	0.026	0.002	0.002	0.026	0.048	0.026	0.001	0.034	0.968	0.055
Class 10	0.025	0.002	0.002	0.026	0.046	0.015	0.001	0.001	0.009	0.934

[1]GitHub repository:Pagellus bogaraveo - Bayesian SSM

6.4 General Population Dynamics

To analyze the estimated number of fish abundances for each year in proportion to the first year (N_t), Graph 6.8 displays the mean value and credibility intervals of that parameter. It is evident that the proportional number of fish decreases constantly, with a smaller slope from year 5 (2018), where N reaches values near zero (between 0.002 and 0.045). The narrow dashed lines suggest that the estimated values are precise.

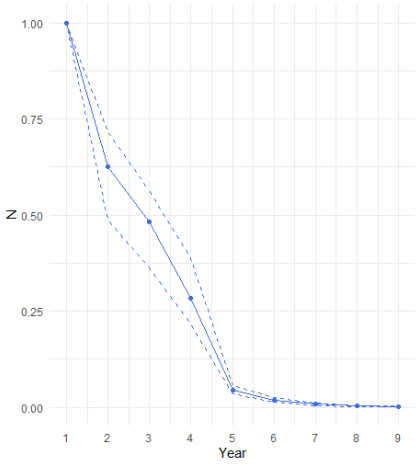


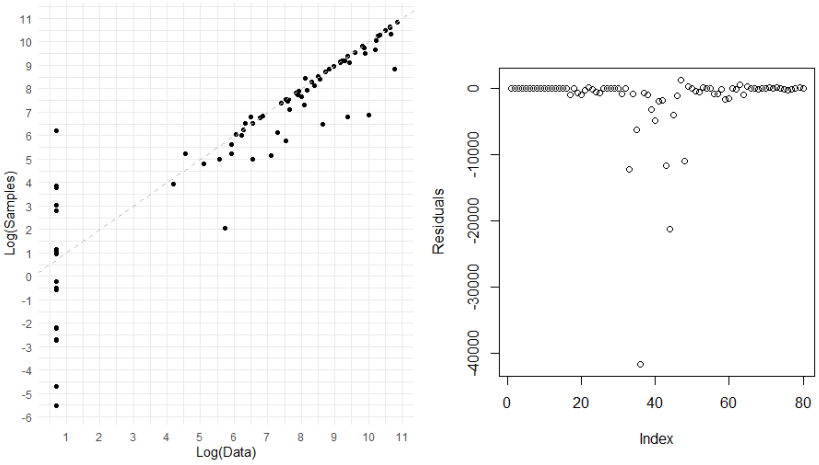
Figure 6.8: Proportional total number of fish by year: the points represent the posterior mean value and the dashed lines the credibility intervals of 95%

However, as mentioned before, fishery data can provide more information about changes in population dynamics than general abundance. Therefore, when analyzing the vector $\Phi_t = (\phi_{t,1}, ..., \phi_{t,10})$ presented in Figure B.1 in Annex B, that display the 50% and 95% credibility intervals and the median value, is evident that the distribution of individuals over the years varies. There are differences in the way fish are divided by size-classes in different years, however, there is no specific graphical pattern that would make possible to predict the distribution in next time-step.

6.5 Model Adjustment

An analysis of the adequacy of the model for the data is presented to understand how well the model's estimated values adjust to the observed data. Graph (a) of Figure 6.9 illustrates the goodness of fit of the model to the data, the alignment of the model's estimates with the observed data is evident from the proximity of the points to the dashed diagonal line, which represents equality between the observed data and the model mean value estimates. There are some points that are slightly more distant from the dashed line, but in general, most points align around it. When the data values are low, it seems to have bigger differences between the data and the mean estimated values, this is explained by the use of the log-transformation.

Graph (b) of Figure 6.9, illustrates the residuals on the original scale, the differences between the observed data and the model's estimated values, the majority of the residuals have values near zero. The few residuals that deviate more significantly from zero all have negative values, suggesting that the model underestimates these values.



(a) Scatter plot Log-Transformed mean of Estimate Samples vs. Data Values

(b) Scatter plot of residuals (Mean of Estimate Samples- Data values)

Figure 6.9: Model adjustment: comparison between mean of sampled values and data values

6.6 Five Size-Classes Model

The results from the presented model are compared with those arising from an alternative. Given the results previously presented, some modifications were introduced, these variations in the model were conceived with the aim of reducing the number of parameters, simplifying the model architecture, and improving performance through more optimal convergence of parameters situated at higher hierarchical levels.

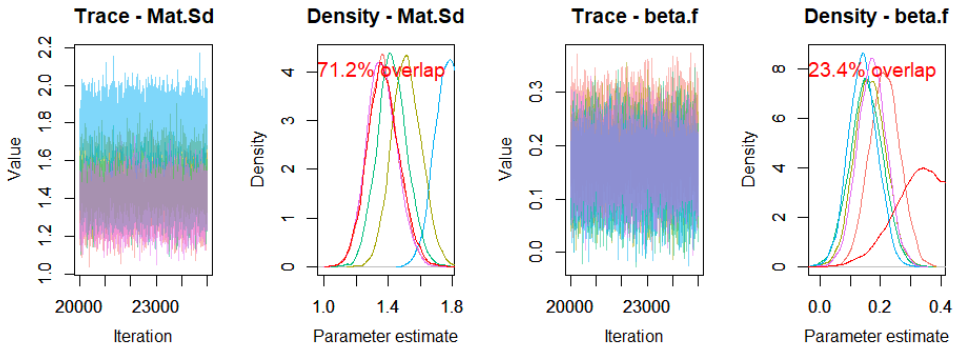
The modification applied to the base model involved reducing the number of size-classes from 10 to 5. This change resulted in a reduction of approximately 50% of the parameters, as all parameters previously calculated by size-class are now halved and the computational time taken was only 38 hours. The five size-classes were obtained by pairing the original ten size-classes while maintaining uniform amplitude intervals, crucial for the model's equations implementation, and still considering the life cycle of blackspot seabream:

- Size-Class 1: old size-classes 1 and 2, includes recruits and juveniles usually not caught in fisheries;
- Size-Class 2: old size-classes 3 and 4, includes adults on the sex-change process with low tax of maturation;
- Size-Class 3: old size-classes 5 and 6, includes adults with high proportion of mature females;
- Size-Class 4: old size-classes 7 and 8, includes adults with high proportion of mature females and higher fecundity;
- Size-Class 5: old size-classes 9 and 10, includes higher natural death and rare catches.

Analyzing the Reproduction and Recruitment process, as expected, the mean and quantile values change with the reduction in the size-classes, but they remain biologically acceptable. However, when comparing the convergence measures results, the five-classes model generally exhibits higher $\hat{R}$ values, with some reaching exceedingly high numbers, particularly in the non-hyperparameters, with the highest being γ_5, reaching 1605.69. However, there was a higher ESS for the hyperparameters ($ESS \in [78547; 126165]$) but a much lower ESS for the other parameters ($ESS \in [8; 52]$). These measures

indicate poorer convergence in this second model, as evident when analyzing the Reproduction and Recruitment trace and density plots, as exemplified in Graph (a) of Figure 6.10, where the same parameter shown in Graph (a) of Figure 6.1 are represented, additionally the chains do not mix as effectively as before.

When analyzing the mortality parameters, similar differences in the convergence measures were observed, almost all parameters exhibit higher $\hat{R}$ values, except for β_f and $L_f^{(50)}$, which, as illustrated in Graph (b) of Figure 6.10, demonstrate better chain mixing compared to the ten size-classes model.



(a) Trace and density plots of parameter σ_M (b) Trace and density plots of parameter β_{Fec}

Figure 6.10: Trace and density plots from the Reproduction and Recruitment process' hyperparameters: Five size-classes model. The Graph (a) illustrate the trace and density plots of σ_M and Graph (b), the trace and density plot of β_{Fec}. The red line on the density plot represents the prior distribution, while other colors, in both trace and density plots, denote different chains.

The Growth process exhibits a different behavior compared to the previous process, all parameters demonstrate improved convergence measures while maintaining biological consistency. Now, all parameters have $\hat{R} \geq 1.07$ and $ESS \geq 106071$. Table 6.5 shows that using five size-class the probability of being in the same size-classes in years t and $t-1$ gets higher, which is what is expected because this means that when using wider size-classes the fish spends more time in that size-class.

Table 6.5: Probability of size-class changes (Five Size-classes Model): Each cell represents the mean value of the probability of transitioning (rounded with 3 decimal places) to the size-class indicated by the row if in the previous year the individual was in the size-class indicated by the column

From (at $t-1$) / To (at t)	**Class 01**	**Class 02**	**Class 03**	**Class 04**	**Class 05**
Class 01	0.059	0.001	0.001	0.001	0.001
Class 02	0.241	0.992	0.002	0.002	0.002
Class 03	0.244	0.004	0.994	0.001	0.001
Class 04	0.254	0.002	0.002	0.994	0.004
Class 05	0.201	0.002	0.002	0.001	0.992

Analyzing the general population dynamics, particularly the proportional total number of fish by year, notable differences between the two models were found. Figure 6.11 illustrates these distinctions. Unlike

the model with ten size-classes (see Figure 6.8), the mean estimates in the five size-classes model do not exhibit significant declines, and in the final year, the values are not as close to zero. However, the credibility intervals are considerably wider, indicating that the estimates are less precise compared to the model with ten size-classes.

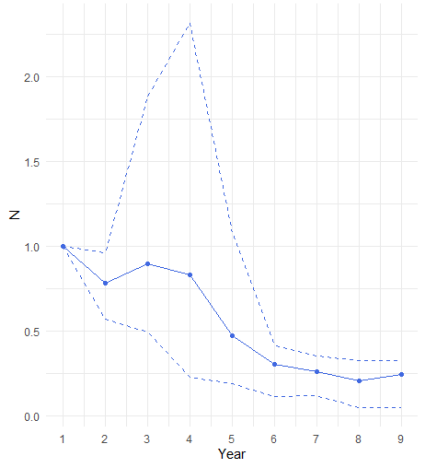


Figure 6.11: Proportional total number of fish by year (Five Size-classes Model):the points represent the posterior mean value and the dashed lines the credibility intervals of 95%

When analyzing the model fit, there is an unexpected result, the model appears to have an almost perfect fit, with all the residuals very close to zero, except for one point with a residual around -300, as shown in Figure 6.12.

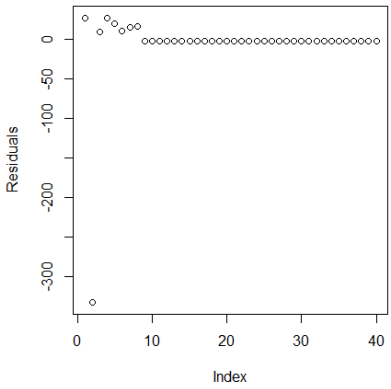


Figure 6.12: Scatter plot of Residuals (mean Estimate Samples- Data values): Five Size-classes Model

In summary, the 5 size-classes model although also be a candidate to model the population dynamics of the blackspot seabream, exhibit different performance than the 10 size-classes model. When comparing the Widely Applicable Information Criterion (WAIC), a measure commonly used to assess the goodness of fit of Bayesian models (Turkman and Paulino, 2015), the ten size-classes model exhibits a WAIC value of 40420.08, whereas the five size-classes model shows a value more than 5000000 times greater (220595728033), which means that the balance between the model complexity and fit to the data reveals that the 10 size-classes model is the best of the two models for a given dataset.

In this thesis, the five size-classes model was used to implement supplementary datasets obtained from ICES, to compare and validate the results obtained by using different datasets from different areas.

However, given the extended time frame covered by these datasets, the number of parameters increases, and even with the five size-classes model the computational effort is too high and it was not possible to get results. After ± 26 days running, the model can not even start the sampling because it is still checking model calculations and has already stored $\geq 60.67 GiB$.

6.7 Comparison with Published Data

In this section, the results obtained are compared with those obtained by other authors, this comparison aims to understand the differences between the results and to determine whether any disparities can be attributed to differences in modeling approaches, datasets, or study areas.

Table 6.6 displays only the parameters results that were transversal in other papers, looking at the β_M parameter, its values vary across the studies, the rounded posterior mean value obtained with the ten size-classes model is equal to the second estimation of Krug, 1998. The mean value obtained with the five size-classes model is lower from all values presented, being more than 6 times less. The parameter $L_M^{(50)}$, there is less variance compering to β_M, and none of the values presented seem to be extremely distant from each other. Analyzing the fecundity parameters, it is difficult to make a clear comparison given the few values to compare, but it is possible to see that there are some differences between studies.

The growth parameters vary across the different studies, and while the area of Bay of Biscay shows lower values of L_∞ and higher values of k compared with the other areas, there is no other specific pattern noteworthy.

Overall, the 10 size-classes provided similar estimation to others published. The advantages and disadvantages will be discussed in further detail in Chapter 7.

Table 6.6: Comparison of results with different authors

References	**Area**	β_M	$L_M^{(50)}$	α_{Fec}	β_{Fec}	k	L_∞
Present Study (10 size-classes)	Mainland portugal	0.466	34.01	9.761	5.382	0.135	56.478
Present Study (5 size-classes)	Mainland portugal	0.066	34.196	6.139	6.351	0.143	60.101
Krug, 1998	Azores	0.413	34.45				
		0.466	33.93				
		0.794	32.31				
Krug, 1986	Azores			9.2935	6.9194		
Krug, 1989	Azores					0.117	58.5
						0.102	57.45
Castro Uranga, 1990	Bay of Biscay					0.174	54.2
Menezes et al., 2001	Azores					0.136	63.8
Sobrino and Gil, 2001	Strait of gilbrata					0.169	58
Ramos, 1967	Bay of Biscay					0.174	53.9
Sánchez, 1983	Bay of Biscay					0.209	51.56
Lorance, 2011	Carabian					0.128	53.86
Alcazar et al., 1987	Bay of Biscay					0.196	48.06

Chapter 7

Conclusion

As previously mentioned, others studies have analyzed the blackspot seabream. A consensus among all those studies was that, despite the commercial interest and of being a well known species, it is also a species with knowledge gaps. Its complex biological dynamics pose challenges for parameters estimation, and knowing the species abundance its particularly difficult (Lorance, 2011).

In this thesis, it was proposed to address some of these gaps for the species found in mainland Portugal, by employing a tailored model that integrates: previous knowledge about the species, accommodates data variability, accounts for its complex life cycle, and uses the dataset provided by IPMA.

After a review on population dynamics models and the advantages of SSM, to accommodate all those specifications, a special model was developed and coded in R using the NIMBLE, facilitating parameter estimation via Bayesian inference.

A SSM methodology was applied, allowing to run the model using two parallel time series: one for the state, where the population dynamics were considered. And one observational model, linked to the data, provided by IPMA, the population dynamics. The times series for the state included events of population dynamics that were divided into 3 major sub-processes (Reproduction and Recruitment; Survival and Mortality; Growth). It is also important to note that the data used in the observational model included 8 years of data describing the numbers and length of caught fish by year. Together, the features of the state and observational models provided a solid base for developing the model described here including previous knowledge about the species while granting a way to account for observational error.

The model developed and described in this thesis uses 10 size-classes to better operation. To define the state process, especially each sub-process of the population dynamics, an exhaustive review of the blackspot seabream was conducted. Several equations and insights into the species were identified, such as equations to calculate probability of being a mature female related to length took from Krug, 1998. These were incorporated into the tailored model, and adjusted to the model specification due to its complexity. One of these adjustment that is relevant to enumerate were the processes that change over time were modified to only change over size-classes. An example of this was the probability of being a mature female, which biologically varies over time. However, in the model, it was considered constant over time to reduce the number of parameters and improve the efficiency of the model. Although this information was omitted, the anticipated changes over the years are not expected to vary enough over the year to make major differences in the estimation results.

After running the developed model with the IPMA dataset, all these parameters demonstrate biological plausibility, because they were in the range of values previously reported and no anomalous or illogical values were observed (e.g:Krug, 1998; Lorance, 2011). Additionally, when comparing the data values

with the posterior mean of the estimates obtained, those were similar, indicating that the model accurately captures the underlying trends in the data and, therefore, it can be considered a good fit.

The results indicate different levels of convergence across different sub-processes and hierarchical levels. In the Reproduction and Recruitment process, all hyperparameters and parameters demonstrate high-quality convergence measures ($\hat{R} \leq 1.11$ and $ESS \geq 3546$), with most hyperparameters showing overlap between prior and posterior distributions.

Regarding the Growth process, both hyperparameters demonstrate convergence ($\hat{R} \leq 1.06$ and $ESS \geq 39610$). Biologically, it is impossible for a fish to decrease in size over time. However, for the model to function effectively, the probability of transitioning from size-class i at year t to size-class j, where $j < i$ at year $t+1$, is sometimes low but not zero. Due to this calculation method, which includes a high number of values near zero, the convergence of these parameters presents peculiar pattern in the convergence analysis.

In the Survival and Mortality process, while hyperparameters also exhibit high-quality convergence measures, except β_f, ($\hat{R} \leq 1.27$ and $ESS \geq 12087$), they diverge from the initial beliefs. No specific explanation could be identified for the poorer convergence measures of β_f, an analysis with different priors could help to get further insight. At the second hierarchical level, some parameters show poorer convergence, more specifically the parameters regarding size-classes 2,3 and 10. This could possibility be explained by some data structure or model structure, but even after those considerations, no clear justification could be found. The analysis of the five size-classes model would help to understand if this occurrence could be replicated.

The model produces estimates with narrow credibility intervals for biological parameters of interest, particularly in the estimation of total abundance (on a proportional scale to the first year). However, the significant decrease in population abundance deserves a careful consideration. With the model results showing a big decline from $N_1 = 1$ (1×10^7 in absolute numbers) in year 1 (2014) to N_9 approximately 0.002 (2×10^4 in absolute numbers) by the beginning of year 9 (2022) in population abundance. It might be interesting to explore whether this decline is related to the exclusion of fishing effort in the model. Since the results in this thesis were obtain using only catches from fisheries data without accounting for fishing effort, it would be valuable to investigate if including this information, when available, affects the total abundance estimates. This aspect should be kept in mind for future studies.

Additionally, the model provides estimates for the population distribution over size-classes, although some of these show less precision and do not show any specific pattern that allows identification of changes in the population distribution over time.

A simplified model was applied, which used 5 size-classes, this model showed improved adjustments on the model computational performance. However, most parameters showed a poorer convergence, and less precise estimates, that remained within biologically possible values. One possible explanation could be the suitability of the model's structure to the life cycle of the species. When working with 10 size-classes, there are instances where the probability of transitioning between size-classes is higher than remaining in the same size-class. However, with 5 size-classes, it is more probable for individuals to remain in the same size-class, leading to an increase in the distribution of individuals in the initial size-class which makes the model have worst performance.

Careful consideration, of whether the trade-offs are worthwhile to simplify the model, is essential when selecting the appropriate size-class structure for modeling purposes. However, while the ten size-classes model appears to better fulfill the aim of this thesis, a simpler model is essential when dealing with denser datasets. In this thesis, despite some limitations, the five size-classes model was applied to compare the obtained results with results using larger datasets from ICES, which included a wider

time-frame and data from other regions. The computational effort, even with the five size-classes model, required for those datasets proved to be impractical, due to time constrictions and software complications. Consequently, it was not possible to compare the results of the model using different datasets with the model developed in this thesis. Due to it, the results of the model developed here were compared with the results of relevant publications. The results generated by the ten size-classes model reveal minimal disparities that remain within the expected range when considering the findings of other researchers. While the five-size classes model exhibited noticeable variations. Comparing mainland Portugal with other regions, no compelling evidence emerged to suggest significant deviations in biological parameter values.

In the future, would be relevant to run the model with the conditions previous described, using larger dataset from ICES, to analyze if the results change and if convergence improves. By incorporating more data, the model may be able to capture more patterns and provide more robust estimates. Next future research could include repetition of the current study using different prior distributions for the hyperparameters. In this thesis, informative priors were used to better inform the model about prior beliefs, however, using vague priors could lead the model to converge to different values, and it would be valuable to analyze these differences. Other interesting approach to future work would be conducting the study using simulated data, that allows for controlled experimentation and evaluation of the model's performance under various scenarios, providing insights into its strengths and limitations in different conditions.

In conclusion, this study could contribute to the understanding of the dynamics of the blackspot seabream population in mainland Portugal through the design, development, and application of a tailored statistical model. Integrating existing knowledge of the species and employing a state-space modeling approach with Bayesian inference, produced a model capable to estimate biological parameters crucial to understanding complex population dynamics and for effective management and conservation efforts. However, the exploration of statistical models to assess species with complex population dynamics, such as the blackspot seabream, remains an ongoing subject.

References

Aeberhard, W. H., Mills Flemming, J., & Nielsen, A. (2018). Review of state-space models for fisheries science. *Annual Review of Statistics and Its Application*, *5*(1), 215–235. https://doi.org/10.1146/annurev-statistics-031017-100427

Alcazar, J., Carrasco, J., Llera, E., Menéndez, M., Ortea Rato, J., & Vizcaíno, A. (1987). Recursos pesqueros de asturias: Aportación al estudio del besugo en el principado de asturias. *Recursos Pesqueros de Asturias*, *4*, 1–88.

Araújo, G., Moura, T., & Figueiredo, I. (2016). Notes on *Pagellus bogaraveo* in the portuguese continental waters (ices division ixa). Working Document for the ICES Working Group on Biology and Assessment of Deep-Sea Fisheries Resources.

Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A., Leos-Barajas, V., Flemming, J., Nielsen, A., Petris, G., & Thomas, L. (2020). An introduction to state-space modeling of ecological time series. https://arxiv.org/abs/2002.02001

Blomstedt, P., Vanhatalo, J., Ulmestrand, M., Gårdmark, A., & Mäntyniemi, S. (2015). A bayesian length-based population dynamics model for northern shrimp (pandalus borealis). *arXiv*. https://arxiv.org/abs/1509.08774

Brooks, S., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graphi. Stat.*, *7*, 434–455. https://doi.org/10.1080/10618600.1998.10474787

Casey Youngflesh, T. H., Christian Che-Castaldo. (2023). Tools to visualize, manipulate, and summarize mcmc output. https://cran.r-project.org/web/packages/MCMCvis/MCMCvis.pdf

Castro Uranga, R. (1990). Biología y pesca del besugo (pagellus bogaraveo b.)informes técnicos 30. *Publicaciones del Gobierno*, 42.

Charnov, E. L., Gislason, H., & Pope, J. G. (2012). Evolutionary assembly rules for fish life histories. *Fish and Fisheries*, *14*, 213–224. https://doi.org/10.1111/j.1467-2979.2012.00467.x

Czerwinski, I. A., Erzini, K., Gutiérrez-Estrada, J. C., & Hernando, J. A. (2009). Deep water longline selectivity for black spot seabream (pagellus bogaraveo) in the strait of gibraltar. *Fisheries Science*, *75*(2), 285–294. https://doi.org/10.1007/s12562-009-0071-7

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodrìguez, A., Temple Lang, D., & Paganin, S. (2023a). *NIMBLE user manual* [R package manual version 1.0.1]. Version 1.0.1. https://doi.org/10.5281/zenodo.1211190

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodrìguez, A., Temple Lang, D., & Paganin, S. (2023b). *NIMBLE: MCMC, particle filtering, and programmable hierarchical modeling* [R package version 1.0.1]. Version 1.0.1. https://doi.org/10.5281/zenodo.1211190

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with

NIMBLE. *Journal of Computational and Graphical Statistics*, *26*, 403–413. https://doi.org/10.1080/10618600.2016.1172487

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*.

Ehlers, R. S. (2003). *Introdução à inferência bayesiana*.

Erzini, K., Salgado, M., & Castro, M. (2006). Dynamics of black spot sea bream (pagellus bogaraveo) mean length: Evaluating the influence of life history parameters, recruitment, size selectivity and exploitation rates. *Journal of Applied Ichthyology*, *22*(3), 183–188. https://doi.org/10.1111/j.1439-0426.2006.00702.x

FAO. (n.d.). Main water areas [Accessed on 21/09/2023]. Coordinating Working Party on Fishery Statistics (CWP). https://www.fao.org/cwp-on-fishery-statistics/handbook/general-concepts/main-water-areas/en/

Farias, I., & Figueiredo, I. (2019). *Guia de identificação macroscópica de estados de maturação de goraz,* Pagellus bogaraveo *(brünnich, 1768)* (tech. rep. No. 24). Relat. Cient. Téc. IPMA.

Gelfand, A. E., & Sahu, S. K. (1999). Identifiability, improper priors, and gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, *94*(445), 247–253. https://doi.org/10.2307/2669699

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. Retrieved April 30, 2024, from http://www.jstor.org/stable/2246093

Guillot, G., Arab, A., Illian, J., & Dray, S. (2022). Editorial: Advances in statistical ecology: New methods and software. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.828919

Hilborn, R., & Walters, C. J. (1992). *Quantitative fisheries stock assessment: Choice, dynamics and uncertainty*. Springer New York, NY. https://doi.org/10.1007/978-1-4615-3598-0

ICES. (2022). Working Group on the Biology and Assessment of Deep-sea Fisheries Resources (WGDEEP). *4*(40), 995. https://doi.org/10.17895/ices.pub.20037233

Ices - maps [Accessed on 21/09/2023]. (n.d.). https://www.ices.dk/data/documents/forms/allitems.aspx?rootfolder=/data/documents/maps&folderctid=0x012000acf6fba45737584389ad23dd43bb914c

International Council for the Exploration of the Sea (ICES). (2023). Definition and rationale for ices ecoregions. https://doi.org/10.17895/ices.advice.23634480.v1

King, R. (2014). Statistical ecology. *Annual Review of Statistics and Its Application*, *1*(1), 401–426. https://doi.org/10.1146/annurev-statistics-022513-115633

King, R., Morgan, B., Gimenez, O., & Brooks, S. (2009). *Bayesian analysis for population ecology*. CRC Press. https://books.google.pt/books?id=mjJ-fpMNG5EC

Kitagawa, G., & Gersch, W. (1996). Smoothness priors analysis of time series. *116*. https://doi.org/10.1007/978-1-4612-0761-0

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*.

Krug, H. (1986). Reproduction of the blackspot seabream, *Pagellus bogaraveo* (brünnich, 1768) in azorean waters.

Krug, H. (1989). Krug, h. m. 1989. the azorean blackspot seabream, pagellus bogaraveo (brünnich, 1768) (teleostei: Sparidae): Age and growth. *Cybium: international journal of ichthyology*, *13*(4), 347–355.

Krug, H. (1990). The azorean blackspot seabream, pagellus bogaraveo (brünnich, 1768) (teleostei, sparidae). reproductive cycle, hermaphrodism, maturity and fecundity. *Cybium: international journal of ichthyology*, *14*(2), 151–159. https://doi.org/10.26028/cybium/1990-142-007

Krug, H. (1998). Variation in the reproductive cycle of the blackspot seabream, pagellus bogaraveo (brünnich, 1768) in the azores. *ARQUIPÉLAGO. Life and Marine Sciences*, *16*, 37–47.

Lorance, P. (2011). History and dynamics of the overexploitation of the blackspot sea bream (pagellus bogaraveo) in the bay of biscay. *ICES Journal of Marine Science*, *68*(2), 290–301. https://doi.org/10.1093/icesjms/fsq072

Mäntyniemi, S., Haapasaari, P., Kuikka, S., Parmanne, R., Lehtiniemi, M., & Kaitaranta, J. (2013). Incorporating stakeholders' knowledge to stock assessment: Central baltic herring. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(4), 591–599. https://doi.org/10.1139/cjfas-2012-0316

Mäntyniemi, S., Uusitalo, L., Peltonen, H., Haapasaari, P., & Kuikka, S. (2013). Integrated, age-structured, length-based stock assessment model with uncertain process variances, structural uncertainty, and environmental covariates: Case of central baltic herring. *Canadian Journal of Fisheries and Aquatic ciences*, *70*(9), 1317–1326. https://doi.org/10.1139/cjfas-2012-0315

Mäntyniemi, S., Whitlock, R., Perälä, T., Blomstedt, P., Vanhatalo, J., Rincón, M., Kuparinen, A., Pulkkinen, H., & Kuikka, S. (2015). General state-space population dynamics model for bayesian stock assessment. *ICES Journal of Marine Science*, *72*(8), 2209–2222. https://doi.org/10.1093/icesjms/fsv117

Martins, R., & Carneiro, M. (2018). *Manual prático de identificação de peixes ósseos da costa continental portuguesa - principais características diagnosticantes*.

Mendelssohn, R. (1988). Some problems in estimating population sizes from catch-at-age data. *Fishery Bulletin*, *86*(4).

Menezes, G., Rogers, A., Krug, H., Mendonça, A., Stockley, B. M., Isidro, E., & Pinho, M. R. (2001). Seasonal changes in biological and ecological traits of demersal and deep-water fish species in the azores [Final report, draft, DG XIV/C/1- study contract 97-081], 184.

Micale, V., Genovese, L., Cristina Guerrera, M., Laurà, R., Maricchiolo, G., & Muglia, U. (2011). The reproductive biology of pagellus bogaraveo, a new candidate species for aquaculture. *The Open Marine Biology Journal*, *5*(1). https://doi.org/10.2174/1874450801105010042

Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D. J., Besbeas, P., Gimenez, O., & Thomas, L. (2014). *Modelling population dynamics: Model formulation, fitting and assessment using state-space methods*. Springer New York. https://doi.org/10.1007/978-1-4939-0977-3

Newman, K. B., Fernández, C., Thomas, L., & Buckland, S. T. (2009). Monte carlo inference for state-space models of wild animal populations. *Biometrics*, *65*(2), 572–583. https://doi.org/10.1111/j.1541-0420.2008.01073.x

Pinho, M., Diogo, H., Carvalho, J., & Pereira, J. G. (2014). Harvesting juveniles of blackspot sea bream (Pagellus bogaraveo) in the azores (northeast atlantic): Biological implications, management, and life cycle considerations. *ICES Journal of Marine Science*, *71*(9), 2448–2456. https://doi.org/10.1093/icesjms/fsu089

Pinho, M. R., & Menezes, G. (2009). Pescaria de demersais dos açores. *Boletim do Núcleo Cultural da Horta*, *18*, 85–102.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2005). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, *6*.

Quinn, T. J., & Deriso, R. B. (1999). *Quantitative fish dynamics*. Oxford University Press.

Ramos, O., F.and Cendrero. (1967). Notes on the age and growth of pagellus cantabricus (asso) of northern spain. *ICES Document CM1967/G: 3. 8 pp*.

Robalo, J. I., Farias, I., Francisco, S. M., Avellaneda, K., Castilho, R., & Figueiredo, I. (2021). Genetic population structure of the blackspot seabream (*Pagellus bogaraveo*): Contribution of mtdna control region to fisheries management. *Mitochondrial DNA. Part A*, *32*(4), 115–119. https://doi.org/10.1080/24701394.2021.1882445

Roberts, G. O., & Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, *1*(none), 20–71. https://doi.org/10.1214/154957804100000024

Sánchez, F. (1983). Biology and fishery of the red sea-bream (Pagellus bogaraveo B.) in vi, vii and viii subareas of ices. *ICES C.M.*, *1983*(G:38), 15.

Sobrino, I., & Gil, J. (2001). Studies on age determination and growth pattern of the red (blackspot) seabream [*Pagellus bogaraveo* (brünnich, 1768)] from the strait of gibraltar (ices ixa/sw spain): Application of the species migratory pattern. *Centro Oceanográfico de Cádiz*.

Stan Development Team. (2024). Stan reference manual. *Version 2.34*.

Stockley, B., Menezes, G. M., Pinho, M. R., & Rogers, A. D. (2005). Genetic population structure in the black-spot sea bream (*Pagellus bogaraveo* brünnich, 1768) from the ne atlantic. *Marine Biology*, *146*(4), 793–804. https://doi.org/10.1007/s00227-004-1479-3

Sullivan, P. J. (1992). A kalman filter approach to catch-at-length analysis. *Biometrics*, *48*(1), 237–257.

Teixeira, J. P. N. (2013). *Recruitment dynamics and early life history of the blackspot seabream, pagellus bogaraveo (perciformes: Sparidae)* [Doctoral dissertation, Universidade dos Açores Departamento de Oceanografia e Pescas].

Thomas, L., Buckland, S. T., Newman, K. B., & Harwood, J. (2005). A unified framework for modelling wildlife population dynamics. *Australian & New Zealand Journal of Statistics*, *47*(1), 19–34. https://doi.org/10.1111/j.1467-842X.2005.00369.x

Turkman, M. A. A., & Paulino, C. D. (2015). *Estatística bayesiana computacional: Uma introdução* [[1], VIII, 196 p. : il. ; 21 cm]. Sociedade Portuguesa de Estatística.

van de Schoot, R., Depaoli, S., King, R., et al. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1–26. https://doi.org/10.1038/s43586-020-00001-2

Vats, D., & Knudson, C. (2020). Revisiting the gelman-rubin diagnostic. http://arxiv.org/abs/1812.09384

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of mcmc (with discussion). *Bayesian Analysis*, *16*(2). https://doi.org/10.1214/20-BA1221

Walsh, B. (2004). Markov chain monte carlo and gibbs sampling. *Lecture Notes for EEB 581, version 26, April*.

Whitlock, R., Mäntyniemi, S., Palm, S., Koljonen, M.-L., Dannewitz, J., & Östergren, J. (2018). Integrating genetic analysis of mixed populations with a spatially explicit population dynamics model. *Methods in Ecology and Evolution*, *9*(4), 1017–1035. https://doi.org/10.1111/2041-210X.12946

Wingfield, J. C. (2013). Ecological processes and the ecology of stress: The impacts of abiotic environmental factors. *Functional Ecology*, *27*(1), 37–44. https://doi.org/10.1111/1365-2435.12039

Appendix A

Table A.1: Notation of the model and respective Code Notation

Notation	Meaning	Code Notation
N^*	Total number of individual on the first time-step	N0
N_t	Total number of individual on time-step t on relative scale to the first time-step	N[t]
$\Phi_t = (\phi_{t,1}, ..., \phi_{t,10})$	State of the population at the beginning of a time-step t in terms of relative size-class frequencies	Phi[t,i]
S_t	Proportion that survives time-step t to the next time-step	S[t]
R_t	Recruits on time-step t on relative scale to the first time-step	R[t]
π^E	Probability of survival from the egg stage until recruitment	Egg_surv
E_t	Number of eggs spawned on time-step t on relative scale to the first time-step	Eggs[t]
Egg_i	Number of eggs spawned by size-class i	eggs_class[i]
π_i^F	Probability of being female belonging to size-class i	Fem[i]
π_i^M	Probability of being mature belonging to size-class i	Mat[i]
γ_i^F	Auxiliary parameter to calculate π_i^F	Fem.Logit[i]
$L_F^{(50)}$	The length at 50% female	li.50.female
β_F	Parameter that controls the curve's softness	beta.F
σ_F	Standard deviation of the distribution of γ_i^F	Fem.Sd
γ_i^M	Auxiliary parameter to calculate π_i^M	Mat.Logit[i]
$L_M^{(50)}$	The length at 50% maturity	li.50.mature
β_M	Parameter that controls the curve's softness	beta.M
σ_M	Standard deviation of the distribution of γ_i^M	Mat.Sd
Fec_i	Fecundity of a individual on size-class i	Fec[i]
β_{Fec}	Parameter that controls the curve's softness	b
$L_f^{(50)}$	The length at 50% selectivity	li.50.sele
CV_{Fec}	Variation coefficient of fecundity	Fec.Cv

Continuation of Table A.1:

Notation	Meaning	Code Notation
$\pi^C_{t,i}$	Probability of a size-class i fish being caught on time-step t	Pi.C
$\pi^D_{t,i}$	Probability of a size-class i fish suffers natural death on time-step t	Pi.D
$\pi^S_{t,i}$	Probability of a size-class i survive time-step t	Pi.S
$Z_{t,i}$	Total mortality rate	Z[t,i]
$F_{t,i}$	Fishing mortality rate	Fish[t,i]
f_i	Selectivity of the gear for size-class i	Sele[i]
γ^f_i	Auxiliary parameter to calculate f_i	f.Logit[i]
σ_f	Standard deviation of the distribution of γ^f_i	f.Sd
F^{max}_t	Maximum instantaneous fishing mortality on time-step t	F.max[t]
$\mu^{F^{max}}$	Mean maximum instantaneous fishing mortality	F.max.Mu
$CV_{F^{max}}$	Variation coefficient of $\mu^{F^{max}}$	F.max.Cv
M_i	Natural mortality rate	Mort[i]
$\Phi^S_t = (\phi^S_{t,1}, ..., \phi^S_{t,10})$	Size-class distribution of survivors on each time-step	PhiS[t,i]
k	Parameter from the Von Bertalanffy curve	k
L_∞	Parameter from the Von Bertalanffy curve	L.inf
σ_L	Standard deviation of the distribution of the growth process	L.Sigma
$g_{i,j}$	Probability of moving to a size-class i at year $t+1$, when at time-step t was in size-class j	g[i,j]
$\Phi^G_t = (\phi^G_{t,1}, ..., \phi^G_{t,10})$	Size-class distribution of survivors after growing on each time-step	PhiG[t,i]

Appendix B

Table B.1: Summary of Others Reproduction and Recruitment parameters

Parameter	Mean	Sd	2.5%	50%	97.5%	$\hat{R}$	ESS
Fec_1	134.756	148.777	15.232	98.012	481.738	1.04	10253
Fec_2	780.368	892.401	81.383	569.886	2849.442	1.04	9253
Fec_3	2979.429	3677.714	287.07	2146.808	11031.204	1.06	6984
Fec_4	8494.123	9814.49	813.014	6171.073	31285.537	1.05	7366
Fec_5	21327.812	29859.93	1773.052	15020.464	80765.052	1.11	6856
Fec_6	47410.643	78232.01	3765.376	32368.955	179095.606	1.11	5818
Fec_7	64954.715	87929.451	6603.204	48889.503	219306.703	1.06	4463
Fec_8	145659.659	200222.024	12298.829	107747.469	488508.275	1.05	4322
Fec_9	260580.103	389076.097	20058.377	188016.964	886045.598	1.05	3546
Fec_{10}	565098.703	665870.471	38181.622	412695.733	1878115.052	1.06	3812
γ_1^F	-3.102	1.704	-6.475	-3.099	0.23	1	50000
γ_2^F	-2.138	1.556	-5.19	-2.132	0.901	1	50296
γ_3^F	-1.163	1.442	-3.974	-1.166	1.658	1	49449
γ_4^F	-0.215	1.389	-2.94	-0.215	2.531	1	50514
γ_5^F	0.767	1.414	-1.992	0.765	3.528	1	49537
γ_6^F	1.746	1.492	-1.191	1.746	4.641	1	49947
γ_7^F	2.437	1.792	-1.117	2.48	5.867	1	48189
γ_8^F	3.628	1.848	-0.007	3.632	7.243	1	50125
γ_9^F	4.622	2.043	0.608	4.628	8.637	1	49519
γ_{10}^F	5.639	2.217	1.362	5.622	10.072	1	47917
γ_1^M	-9.802	6.556	-23.018	-9.616	1.792	1	15874
γ_2^M	-7.458	5.087	-17.795	-7.317	1.648	1	16322
γ_3^M	-5.129	3.658	-12.54	-5.022	1.474	1	17061
γ_4^M	-2.829	2.349	-7.68	-2.744	1.482	1	20878
γ_5^M	-0.479	1.517	-3.485	-0.46	2.457	1	48236
γ_6^M	1.875	1.916	-1.682	1.8	5.801	1	32120
γ_7^M	3.978	3.338	-2.324	4.026	10.532	1	16393
γ_8^M	6.498	4.554	-1.759	6.393	15.631	1	16903
γ_9^M	8.84	5.985	-1.913	8.694	20.878	1	16104
γ_{10}^M	11.274	7.328	-1.042	10.999	26.175	1	15629

Table B.2: Summary of Others Survival and Mortality parameter

Parameter	Mean	Sd	2.5%	50%	97.5%	$\hat{R}$	ESS
F_1^{max}	0.127	0.014	0.101	0.127	0.157	1.01	26701
F_2^{max}	5.928	0.599	4.853	5.892	7.214	1.01	19474
F_3^{max}	0.664	1.557	0.24	0.32	7.465	1.04	14107
F_4^{max}	0.205	0.044	0.129	0.201	0.304	1.01	10128
F_5^{max}	3.594	0.41	2.872	3.566	4.473	1.02	25323
F_6^{max}	5.54	0.659	4.37	5.501	6.947	1.05	13299
F_7^{max}	12.666	2.591	8.082	12.03	18.089	1.7	6048
F_8^{max}	35.513	4.607	27.199	35.58	44.666	1.31	7197
γ_1^f	-14.472	0.48	-15.566	-14.418	-13.696	1.03	49249
γ_2^f	-7.043	3.323	-8.995	-8.595	1.637	6.14	14989
γ_3^f	-3.025	0.501	-3.84	-2.841	-2.219	2	4944
γ_4^f	3.583	0.992	1.933	3.478	5.808	1.01	45960
γ_5^f	-1.19	0.169	-1.432	-1.21	-0.61	1.03	21976
γ_6^f	-1.851	0.107	-2.054	-1.853	-1.62	1.02	18418
γ_7^f	-1.566	0.169	-1.893	-1.568	-1.228	1.02	24727
γ_8^f	-3.351	0.18	-3.718	-3.347	-3.012	1	38134
γ_9^f	-4.711	0.09	-4.893	-4.709	-4.536	1.01	18001
γ_{10}^f	-4.093	4.811	-6.682	-6.441	7.882	7	16459
M_1	0.25	0	0.25	0.25	0.25		0
M_2	452.174	1209.519	0	0.029	4643.637	2.22	7588
M_3	1.416	1.017	0	1.777	3.004	1.96	1915
M_4	4.945	0.425	4.01	4.944	5.783	1.02	25876
M_5	0.001	0.002	0	0	0.005	1.01	34531
M_6	0.001	0.002	0	0	0.006	1.01	37141
M_7	4.127	0.503	3.2	4.102	5.178	1.03	22628
M_8	3.43	0.572	2.337	3.421	4.586	1	38227
M_9	0.006	0.013	0	0.001	0.043	1.01	37518
M_{10}	128.539	259.648	0	0.021	728.602	9.87	9104

Table B.3: Summary of Others Growth parameters

Parameter	Mean	Sd	2.5%	50%	97.5%	$\hat{R}$	ESS
$g_{1,1}$	0.05	0.162	0	0	0.682	1.05	38063
$g_{2,1}$	0.002	0.029	0	0	0	1.04	49609
$g_{3,1}$	0.002	0.028	0	0	0	1.01	47803
$g_{4,1}$	0.001	0.024	0	0	0	1.04	47449
$g_{5,1}$	0.001	0.024	0	0	0	1.07	45850
$g_{6,1}$	0.001	0.022	0	0	0	1.05	49590
$g_{7,1}$	0.002	0.028	0	0	0	1.02	50000
$g_{8,1}$	0.001	0.026	0	0	0	1.02	48948
$g_{9,1}$	0.001	0.024	0	0	0	1	47548
$g_{10,1}$	0.002	0.029	0	0	0	1	50000
$g_{1,2}$	0.748	0.375	0.057	1	1	2.54	40889
$g_{2,2}$	0.062	0.185	0	0	0.779	1.03	39406
$g_{3,2}$	0.001	0.025	0	0	0	1.02	44215
$g_{4,2}$	0.004	0.023	0	0	0.039	1.09	45776
$g_{5,2}$	0.033	0.052	0	0	0.146	1.84	43641
$g_{6,2}$	0.018	0.047	0	0	0.136	1.76	41834
$g_{7,2}$	0.001	0.024	0	0	0	1.01	47845
$g_{8,2}$	0.001	0.026	0	0	0	1.01	46120
$g_{9,2}$	0.002	0.035	0	0	0	1.03	34711
$g_{10,2}$	0.001	0.027	0	0	0	1.01	45929
$g_{1,3}$	0.035	0.095	0	0	0.221	1.12	40736
$g_{2,3}$	0.923	0.201	0.173	1	1	1.02	39544
$g_{3,3}$	0.141	0.262	0	0.002	0.938	1.02	39581
$g_{4,3}$	0.025	0.056	0	0	0.175	2.55	50000

Continuation of Table B.3:

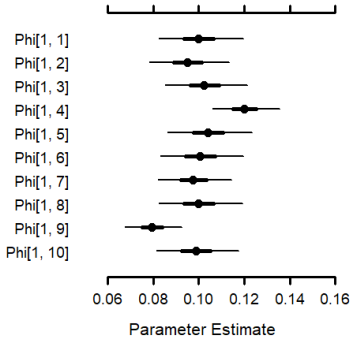
Parameter	**Mean**	**Sd**	**2.5%**	**50%**	**97.5%**	$\hat{R}$	**ESS**
$g_{5,3}$	0.043	0.06	0	0	0.176	2.41	45691
$g_{6,3}$	0.022	0.052	0	0	0.161	2.08	44355
$g_{7,3}$	0.002	0.03	0	0	0	1.02	45669
$g_{8,3}$	0.001	0.025	0	0	0	1.01	48916
$g_{9,3}$	0.001	0.025	0	0	0	1.02	44718
$g_{10,3}$	0.001	0.025	0	0	0	1.01	48294
$g_{1,4}$	0.023	0.055	0	0	0.162	1.94	43328
$g_{2,4}$	0.003	0.041	0	0	0	1.01	49617
$g_{3,4}$	0.847	0.27	0.053	0.996	1	1.02	39344
$g_{4,4}$	0.245	0.306	0	0.112	0.988	1.07	38142
$g_{5,4}$	0.045	0.064	0	0	0.187	2.34	44479
$g_{6,4}$	0.023	0.052	0	0	0.166	2.4	48427
$g_{7,4}$	0.001	0.025	0	0	0	1.01	47981
$g_{8,4}$	0.001	0.026	0	0	0	1.02	49466
$g_{9,4}$	0.002	0.029	0	0	0	1.01	46969
$g_{10,4}$	0.002	0.029	0	0	0	1.01	48714
$g_{1,5}$	0.022	0.052	0	0	0.155	2.14	46178
$g_{2,5}$	0.002	0.029	0	0	0	1	48093
$g_{3,5}$	0.002	0.028	0	0	0	1.01	48536
$g_{4,5}$	0.597	0.383	0.01	0.724	1	1.46	40994
$g_{5,5}$	0.345	0.335	0.001	0.167	0.999	1.33	40897
$g_{6,5}$	0.023	0.053	0	0	0.166	2.32	46604
$g_{7,5}$	0.001	0.024	0	0	0	1.03	48133
$g_{8,5}$	0.001	0.024	0	0	0	1.03	47780
$g_{9,5}$	0.002	0.03	0	0	0	1	48353
$g_{10,5}$	0.001	0.024	0	0	0	1.02	48787

Continuation of Table B.3:

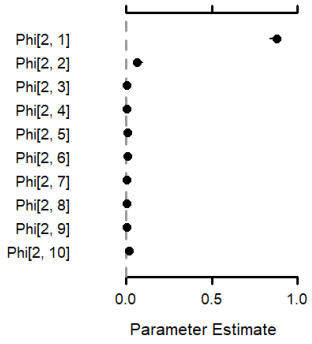
Parameter	**Mean**	**Sd**	**2.5%**	**50%**	**97.5%**	$\hat{R}$	**ESS**
$g_{1,6}$	0.021	0.05	0	0	0.148	1.83	46224
$g_{2,6}$	0.001	0.025	0	0	0	1.04	47385
$g_{3,6}$	0.001	0.026	0	0	0	1.01	48772
$g_{4,6}$	0.022	0.052	0	0	0.163	2.16	48645
$g_{5,6}$	0.342	0.341	0	0.152	0.999	1.41	41248
$g_{6,6}$	0.599	0.373	0.026	0.704	1	1.44	41799
$g_{7,6}$	0.002	0.028	0	0	0	1.02	50000
$g_{8,6}$	0.001	0.025	0	0	0	1	48985
$g_{9,6}$	0.002	0.03	0	0	0	1.01	49136
$g_{10,6}$	0.001	0.025	0	0	0	1.01	46040
$g_{1,7}$	0.027	0.063	0	0	0.184	1.99	49487
$g_{2,7}$	0.001	0.027	0	0	0	1.01	48643
$g_{3,7}$	0.002	0.031	0	0	0	1.02	49050
$g_{4,7}$	0.028	0.063	0	0	0.201	2.78	48103
$g_{5,7}$	0.052	0.071	0	0	0.21	2.78	46233
$g_{6,7}$	0.247	0.294	0	0.129	0.971	1.07	40555
$g_{7,7}$	0.879	0.235	0.127	0.998	1	1.01	44945
$g_{8,7}$	0.003	0.038	0	0	0	1	48539
$g_{9,7}$	0.002	0.027	0	0	0	1.02	49256
$g_{10,7}$	0.002	0.029	0	0	0	1.05	49303
$g_{1,8}$	0.023	0.054	0	0	0.162	2.18	45784
$g_{2,8}$	0.002	0.028	0	0	0	1.02	46905
$g_{3,8}$	0.001	0.027	0	0	0	1.02	46833
$g_{4,8}$	0.025	0.057	0	0	0.178	2.43	48510
$g_{5,8}$	0.045	0.064	0	0	0.186	2.34	45469

Continuation of Table B.3:

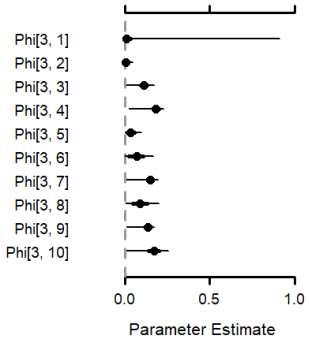
Parameter	**Mean**	**Sd**	**2.5%**	**50%**	**97.5%**	$\hat{R}$	**ESS**
$g_{6,8}$	0.024	0.056	0	0	0.173	2.19	45503
$g_{7,8}$	0.11	0.226	0	0.001	0.854	1.02	44677
$g_{8,8}$	0.954	0.15	0.4	1	1	1.01	50000
$g_{9,8}$	0.012	0.079	0	0	0.129	1.01	46881
$g_{10,8}$	0.001	0.026	0	0	0	1.02	48729
$g_{1,9}$	0.026	0.063	0	0	0.172	1.67	41084
$g_{2,9}$	0.002	0.03	0	0	0	1.01	47739
$g_{3,9}$	0.002	0.031	0	0	0	1.04	46697
$g_{4,9}$	0.026	0.059	0	0	0.187	2.47	49650
$g_{5,9}$	0.048	0.067	0	0	0.194	2.59	50007
$g_{6,9}$	0.026	0.059	0	0	0.187	2.68	47523
$g_{7,9}$	0.001	0.025	0	0	0	1	47375
$g_{8,9}$	0.034	0.13	0	0	0.499	1.01	50000
$g_{9,9}$	0.968	0.126	0.521	1	1	1	46584
$g_{10,9}$	0.055	0.161	0	0	0.653	1.01	49270
$g_{1,10}$	0.025	0.062	0	0	0.169	1.63	44218
$g_{2,10}$	0.002	0.029	0	0	0	1.03	49194
$g_{3,10}$	0.002	0.03	0	0	0	1.02	47614
$g_{4,10}$	0.026	0.058	0	0	0.181	2.31	48053
$g_{5,10}$	0.046	0.065	0	0	0.188	2.37	40792
$g_{6,10}$	0.015	0.04	0	0	0.119	1.83	42689
$g_{7,10}$	0.001	0.025	0	0	0	1.01	43835
$g_{8,10}$	0.001	0.025	0	0	0	1.01	48698
$g_{9,10}$	0.009	0.065	0	0	0.057	1.01	50000
$g_{10,10}$	0.934	0.176	0.291	1	1	1.01	50197



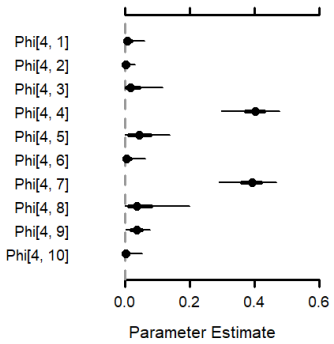
(a) Φ_1



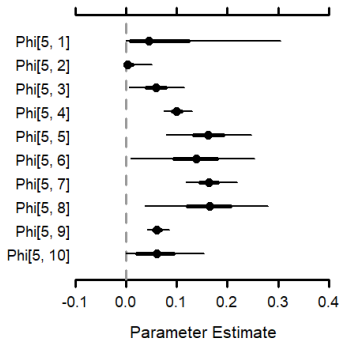
(b) Φ_2



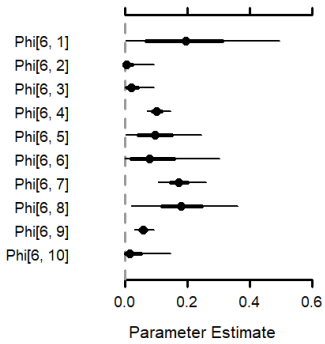
(c) Φ_3



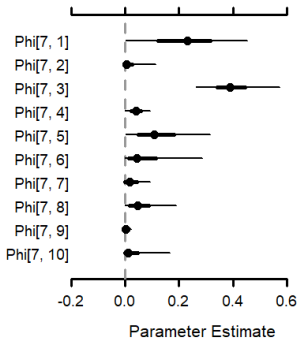
(d) Φ_4



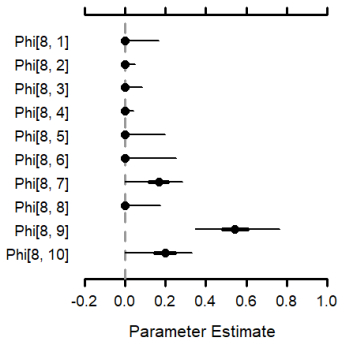
(e) Φ_5



(f) Φ_6



(g) Φ_7



(h) Φ_8

Figure B.1: Caterpillar plots of Population distribution by size-class in each year ($\Phi_t = (\phi_{t,1}, ..., \phi_{t,10})$):Each graph represents a different year.Points represent posterior medians. Thick and thin lines represent 50% and 95% credibility intervals, respectively