

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

Modelagem estatística de dados pesqueiros

Fernando de Pol Mayer

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba
2019**

Fernando de Pol Mayer
Oceanógrafo

Modelagem estatística de dados pesqueiros

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **PAULO JUSTINIANO RIBEIRO JR.**

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2019

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Mayer, Fernando de Pol

Modelagem estatística de dados pesqueiros / Fernando de Pol Mayer.
— versão revisada de acordo com a resolução CoPGr 6018 de 2011. —
Piracicaba, 2019 .

113 p.

Tese (Doutorado) — USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Estatística ecológica 2. Modelagem estatística 3. Modelos espaço-temporais 4. Dados pesqueiros . I. Título.

AGRADECIMENTOS

Aos meus pais, por terem me apoiado em todas as etapas que passei.

À Ana, por toda paciência, carinho e compreensão nessa fase.

À todos os professores do Programa de Estatística e Experimentação Agronômica da ESALQ, por todos os valiosos ensinamentos passados. Também à Solange por toda a ajuda e atenção de sempre.

Ao meu orientador (e agora colega), Prof. Paulo Justiniano, por ter aceito essa empreitada.

Aos Professores Miguel Petrere, Roseli Leandro e Clarice Demétrio, pelas valiosas observações feitas na qualificação.

Ao Prof. Paul Kinas, que, mesmo indiretamente, fez parte deste projeto.

À todos os amigos que fiz em Piracicaba, especialmente aos três que tenho uma dívida eterna: Rick, Erasnilson e Douglas. E também ao Julio pela parceria nos estudos.

Ao meu colega e amigo Rodrigo Sant'Ana, pelas discussões e ajuda de sempre.

Aos meus colegas de laboratório, especialmente Wagner, Walmes e Elias, por toda a ajuda (direta e indireta) no desenvolvimento da tese.

EPÍGRAFE

“It appears to be a quite general principle that, whenever there is a randomized way of doing something, then there is a nonrandomized way that delivers better performance but requires more thought.”

— E. T. Jaynes, 2003

SUMÁRIO

Resumo	7
Abstract	8
1 Introdução	9
Referências	11
2 Modelagem da estrutura temporal de capturas incidentais em pescarias comerciais através de modelos hierárquicos bayesianos	13
Resumo	13
2.1 Introdução	13
2.2 Metodologia	15
2.2.1 Dados	15
2.2.2 Definição dos modelos	15
2.2.3 Seleção de modelos	18
2.3 Resultados e discussão	19
2.4 Conclusões	24
Referências	26
3 Modelos dinâmicos de biomassa com estruturas de dependência espaço-temporal	29
Resumo	29
3.1 Introdução	29
3.2 Fundamentação teórica	32
3.2.1 Dinâmica populacional de recursos pesqueiros	32
3.2.2 Modelos dinâmicos de biomassa	34
3.2.3 Definição dos modelos	35
3.2.4 Métodos de estimativa	38
3.2.4.1 Métodos de equilíbrio	38
3.2.4.2 Métodos de regressão linear	39
3.2.4.3 Modelos com erros nas variáveis	41
3.2.4.4 Modelos de espaço de estados	42
3.3 Metodologia	44
3.3.1 Modelos dinâmicos de biomassa espaço-temporais	44
3.3.2 Método de estimativa	45
3.3.3 Modelos considerados	48
3.3.3.1 Primeiro estágio	48
3.3.3.2 Segundo estágio	50
3.3.3.3 Prioris e hiperprioris	53
3.3.4 Simulação	53
3.3.4.1 Simulação 1	56
3.3.4.2 Simulação 2	56
3.3.4.3 Medidas de adequação	57
3.4 Resultados e discussão	57
3.4.1 Simulação 1	57
3.4.1.1 Primeiro estágio	58
3.4.1.2 Segundo estágio	59
3.4.2 Simulação 2	62
3.5 Conclusões	70
Referências	70

4 Modelos Multivariados de Covariância Linear Generalizada (McGLM) para a avaliação do efeito de hookpods em pescarias comerciais	75
Resumo	75
4.1 Introdução	75
4.2 Metodologia	76
4.2.1 Dados	76
4.2.2 Modelos	78
4.3 Resultados e discussão	80
4.3.1 Análise exploratória	80
4.3.2 Modelo	81
4.4 Conclusões	84
Referências	84
5 Considerações finais	87
Apêndices	89
A Demonstrações	89
B Figuras	91
C Tabelas	102

RESUMO

Modelagem estatística de dados pesqueiros

A flutuação na abundância de estoques pesqueiros é tema de investigação científica desde o final do século XIX. A ciência pesqueira se estabeleceu para que se pudesse compreender os impactos da exploração da pesca em recursos naturais renováveis, e, principalmente, para evitar colapsos e preservar as populações exploradas. A principal fonte de dados para investigações pesqueiras são provenientes da própria pesca. Um dos problemas associados à esse fato, é que os pescadores procuram sempre por áreas de alto rendimento, o que faz com que os dados de captura analisados sejam de uma amostra não aleatória. Além disso, entre as diversas espécies que são capturadas, algumas poucas são consideradas alvo da pescaria, enquanto que as demais são chamadas de capturas incidentais. Tudo isso faz com que dados pesqueiros possuam características particulares, como assimetria, excesso de zeros e sobredispersão. Tradicionalmente, estes dados são analisados ignorando-se a estrutura de dependência espacial e temporal, reconhecidamente importante devido a natureza das informações. O objetivo deste trabalho foi então utilizar abordagens estatísticas recentes, que são capazes de lidar com as características desafiadoras de dados pesqueiros. Inicialmente, foram utilizados modelos hierárquicos bayesianos para a modelagem da estrutura temporal da captura de uma espécie. Para esta modelagem, foi utilizado a aproximação de Laplace encaixada e integrada (INLA), um método recente na literatura para realizar inferência no contexto bayesiano através de aproximação, ao contrário dos tradicionais métodos por amostragem. Esta alternativa é computacionalmente mais eficiente, e se mostrou extremamente viável para a análise de dados pesqueiros. Ao final, pode-se concluir que a dependência temporal é de fato importante, e não pode ser ignorada nesse tipo de dado. Seguindo a mesma ideia, foi também desenvolvido um método de estimação em dois estágios para a obtenção dos parâmetros de modelos dinâmicos de biomassa. A estimação em dois estágios permite a consideração de toda a estrutura de dependência espaço-temporal presente nos dados pesqueiros. De fato, os resultados mostraram que quando estas estruturas são levadas em consideração, as estimativas dos parâmetros de interesse são não viesadas e consistentes. Abordagens multivariadas seriam escolhas naturais para a análise da captura de diversas espécies. No entanto, a maioria destas técnicas envolve a suposição de normalidade, o que praticamente torna inviável seu uso em dados de captura. Os Modelos Multivariados de Covariância Linear Generalizada (McGLM) são uma alternativa recente, pois permitem considerar que as respostas possam assumir qualquer distribuição da família exponencial de dispersão. Isso faz com que dados que possuem dependência, excesso de zeros e sobredispersão possam ser considerados de maneira natural. Neste trabalho, os McGLMs foram utilizados para se verificar a influência de um dispositivo que evita a captura incidental de aves, na captura de duas espécies consideradas alvo da pescaria. Com isso, pode-se concluir que o artefato, comprovadamente eficiente na redução da captura de aves, não diminui a captura das espécies desejadas.

Palavras-chave: Estatística ecológica, Modelagem estatística, Modelos espaço-temporais, Dados pesqueiros

ABSTRACT

Statistical modelling of fisheries data

Fluctuation in abundance of fish stocks is scientifically studied since the end of the 19th century. Fisheries science was established to understand impacts of fisheries in renewable natural resources, and mainly to avoid overfishing and to preserve explored populations. The primary data source for fisheries investigation is from fisheries itself. One of the problems associated with this fact is that fishermen are always looking for high productivity areas, which makes catch data a non-random sample. In addition, among many species captured, only a few are considered targets, while the rest are called bycatch. As a consequence, fisheries data presents particular features, such as asymmetry, excess zeros, and overdispersion. It is well known that catches are highly dependent in time and space, however, this spatiotemporal structure is traditionally ignored when analyzing catch data. The aim of this work was then to apply newly developed statistical approaches, capable of handling the challenging features of fisheries data. Initially, the temporal structure of a bycatch species was modeled through Bayesian hierarchical models. The Integrated Nested Laplace Approximation (INLA), a recent method that uses approximation rather than sampling, in the Bayesian framework, was used. This approach is computationally more efficient and turned out to be a quite viable method to analyze fisheries data. As a conclusion, we showed that the temporal dependence is indeed important, and cannot be ignored when analyzing this kind of data. Following the same reasoning, a two-stage approach was developed for the estimation of parameters from biomass dynamic models. This two-stage approach allows the inclusion of spatial and temporal effects, inherently present in fisheries data. Results showed that when these structures are considered, parameter estimates are unbiased and consistent. Multivariate statistical methods should be the default approach when analyzing catch data from several species. However, most of these techniques are based on the normality assumption, which makes it of little use for catch data. The recently developed Multivariate Covariance Generalized Linear Models (McGLM) assumes that response variables could follow any distribution from the exponential dispersion family. This flexibility allows that data with both overdispersion and excess zeros may be modeled in a natural way. In this work, McGLMs were used to assess the influence of a device used to prevent seabird bycatch, in two target species captures. As a conclusion, we showed that the device does not interfere in the catch of targeted species.

Keywords: Ecological statistics, Statistical modelling, Spatio-temporal models, Fisheries data

1 INTRODUÇÃO

A exploração de recursos pesqueiros é uma das atividades mais antigas praticadas pelo ser humano, tanto para própria subsistência quanto como uma alternativa de proteína animal. A pesca irrestrita no passado gerou uma série de impactos nas populações exploradas, desde alterações no comportamento das espécies e modificação nas estruturas tróficas, até a extinção total. Com isso, surgiu também a preocupação em se entender os impactos da pesca e como evitar que as populações exploradas entrem em colapso, dando início ao que hoje é chamado de “ciência pesqueira”.

A preocupação com a flutuação anual do rendimento de pescarias teve início efetivamente no final dos anos 1850, quando o governo norueguês contratou cientistas para descobrir o motivo das oscilações anuais na captura do bacalhau. Nos anos seguintes, cientistas pesqueiros começaram a atuar também em grande parte dos países da Europa, devido à crescente preocupação na variação dos estoques (Jennings et al., 2001). Houve um grande consenso de que pesquisas eram necessárias para se tentar quantificar os efeitos do impacto da pesca nas populações naturais.

As embarcações pesqueiras aumentaram rapidamente em número e eficiência nos anos anteriores à Primeira Guerra Mundial, e os peixes capturados estavam se tornando cada vez menores e menos abundantes. Durante o período da Guerra, o esforço pesqueiro diminuiu consideravelmente devido às atividades militares, e quando a atividade pesqueira retornou no final da Guerra em 1918, a captura voltou a ser composta por peixes maiores e com taxas mais elevadas. Assim, ficou claro que a redução do esforço pesqueiro tinha permitido a recuperação dos estoques explotados, e a ciência pesqueira passou a focar no entendimento destes efeitos (Gulland, 1983).

Apenas no final dos anos 1950, cientistas começaram a desenvolver modelos para descrever a dinâmica de populações pesqueiras e as respostas à pesca. Nos anos subsequentes, a ciência pesqueira desempenhou um papel fundamental no campo da Ecologia, ao descrever porque as populações flutuam, e como a estrutura das populações variam no espaço, tempo, e em resposta ao aumento das taxas de mortalidade imposta por humanos.

No final do século XX, houve uma diversificação na ciência pesqueira, uma vez que houve um aumento na preocupação sobre o colapso de estoques pesqueiros, a incerteza que é inerente ao manejo pesqueiro, e o reconhecimento de que os objetivos de manejo não deveriam ser baseados somente nos estoques pesqueiros. Atualmente, o manejo de uma pescaria também envolve a atuação de sociólogos e economistas, com a intenção de levar em consideração o impacto nas vidas e no rendimento das pessoas envolvidas (Hilborn e Walters, 1992). Também há uma grande preocupação dos efeitos da pesca no ambiente marinho e em outras espécies.

Uma característica única associada à ciência pesqueira é o fato de que raramente trabalha-se com dados provenientes de uma amostragem aleatória. Os dados utilizados para se conhecer e avaliar uma população geralmente são aqueles de desembarques das embarcações de pesca. Os pescadores jamais fazem uma viagem ao acaso, e tendem a se deslocar para os lugares sabidamente com maior abundância das espécies de interesse. Isso faz com que a “amostragem” de dados pesqueiros seja altamente viesada, o que promove ainda mais incerteza em qualquer tipo de inferência. Desenhos experimentais para evitar esse tipo de viés só são possíveis através de cruzeiros científicos pesqueiros. Apenas alguns poucos países no mundo possuem a capacidade financeira e a cultura científica para realizar estes cruzeiros com a frequência necessária. Desnecessário dizer que em países como o Brasil, este tipo de empreendimento científico seja raro ou completamente inexistente. Sendo assim, a realidade da maioria dos cientistas pesqueiros no mundo é a de lidar com os dados viesados provenientes da pesca.

Os dados provenientes da pesca possuem também características desafiadoras. Os pescadores geralmente procuram capturar uma, ou algumas poucas, espécies. No entanto, a seletividade dos artefatos não é perfeita, e diversas outras espécies são também capturadas. A captura de espécies que não são alvo

é chamada de “captura incidental”, e podem ou não serem aproveitadas pelos pescadores. Por exemplo, a captura incidental de outros peixes de menor valor comercial ainda pode ser aproveitada, mas a captura de aves marinhas, por exemplo, gera prejuízos tanto para os pescadores quanto para o meio ambiente. Dados de captura de espécies que não são alvo, ou de interesse secundário, geralmente possuem uma grande quantidade de zeros, o que pode gerar uma série de dificuldades do ponto de vista estatístico. Além disso, muitas espécies formam cardumes, e os dados de capturas são compostos por muitos zeros e alguns valores elevados, o que pode gerar a chamada “sobredispersão” (Hinde e Demétrio, 1998), quando a variância é maior do que a média (em contraste com a suposição de que média e variância sejam iguais). Isto ainda faz com que a distribuição dos dados de captura seja altamente assimétrica, e raramente as técnicas tradicionais estatísticas que assumem normalidade podem ser aplicadas.

Soma-se a isso o fato de que as capturas ocorrem em diferentes áreas, e que possivelmente mudam conforme a época do ano, devido à migração de algumas espécies. Portanto, não é difícil reconhecer que as capturas em uma determinada área e em um determinado tempo sejam dependentes das áreas adjacentes e das capturas em tempos anteriores. Os métodos tradicionais de análise de dados de captura geralmente ignoram a dependência espaço-temporal, e assumem independência entre as observações. Walters (2003), cita dois erros comumente cometidos na análise de dados de captura. O primeiro deles é o uso incorreto de estimadores do tipo razão, como é o caso da Captura Por Unidade de Esforço (CPUE). O autor comenta que somar as capturas no tempo e espaço, e dividir pelo esforço total só seria funcional em áreas e intervalos de tempo pequenos, e se a amostragem fosse aleatória. Como na maioria dos casos, captura e esforço são somas anuais ao longo de estratos espaciais, não são dados os devidos pesos para áreas com maior esforço. O segundo erro, e possivelmente mais grave, segundo o autor, é ignorar as áreas ocupadas pelo estoque e que não foram exploradas pela pesca. As alternativas seriam então: (i) ignorar estas áreas completamente; (ii) usar a média de CPUE de áreas vizinhas; e (iii) estimar valores com base em modelos adequados (*e.g.* interpolação).

Para quaisquer processos que ocorram no espaço e no tempo, assumir que a agregação de dados ao longo do tempo (processo espacial) ou ao longo do espaço (processo temporal) é suficiente para descrever tal processo, é cada vez mais irreal (Wikle e Hooten, 2010). Na ciência pesqueira, já é reconhecida a importância de se considerar a estrutura de dependência espacial e temporal (Thorson et al., 2015), embora as aplicações, na prática, ainda sejam modestas. Por esse motivo, é fundamental que o conhecimento da dinâmica temporal, espacial, e espaço-temporal das populações seja explorado.

Os modelos dinâmicos de biomassa são uma importante ferramenta para o manejo pesqueiro. Estes modelos necessitam apenas de dados de captura e esforço por ano, e fornecem informações importantes sobre a situação de um estoque, que podem então ser utilizadas para o controle e gerenciamento da pesca. No entanto, a estimativa dos parâmetros destes modelos aparentemente simples não é uma tarefa fácil devido à sua natureza não linear. Além disso, os dados anuais são agregados por área, portanto toda a informação sobre a estrutura espacial do estoque é perdida. Dessa forma, é de fundamental importância que sejam investigados métodos de estimativa para estes modelos, que permitam a consideração da estrutura de dependência espaço-temporal, inerente aos dados pesqueiros.

Outro tipo de modelagem particularmente útil para a ciência pesqueira, mas que ainda é pouco explorada, é a análise multivariada. Como a captura ocorre conjuntamente para diversas espécies, a análise de fatores que interferem nas capturas também deve considerar as espécies em conjunto. Um dos possíveis motivos pelo qual a análise multivariada não é muito utilizada, é o já mencionado fato de que as capturas frequentemente possuem uma distribuição assimétrica, possivelmente com excesso de zeros e/ou sobredispersão. No entanto, a maioria dos métodos multivariados requer a suposição de normalidade, o que claramente inviabiliza muitas das aplicações. Portanto, a utilização de métodos multivariados que possam acomodar estas diferentes características dos dados pesqueiros também deve ser explorada.

Sendo assim, o objetivo deste trabalho é explorar e utilizar técnicas de modelagem modernas na

análise de dados pesqueiros. A ideia é mostrar que a suposição de observações independentes no tempo e no espaço não é sustentável, a partir da aplicação de modelos que permitem incorporar estruturas de dependência. Também é apresentado um método alternativo de estimativa para os parâmetros de modelos dinâmicos de biomassa, que permitem considerar a dependência espaço-temporal dos dados. Além disso, pretende-se mostrar também que análises multivariadas considerando distribuições diferentes da normal são viáveis atualmente, e que, em muitos casos, a modelagem deve ser feita considerando as diversas espécies envolvidas como variável resposta, ao invés de apenas uma.

A tese está estruturada da seguinte forma. No Capítulo 2 é apresentada uma análise da estrutura temporal das capturas de uma espécie de agulhão em uma pescaria com espinhel de superfície no Atlântico Sul. As capturas de agulhões nesse tipo de pescaria geralmente são consideradas incidentais. O conhecimento inicial da distribuição temporal destas espécies pode ajudar no seu manejo. Diferentes estruturas temporais foram analisadas, através de modelos hierárquicos bayesianos. No capítulo 3 é desenvolvido um método alternativo para o cálculo dos parâmetros de modelos dinâmicos de biomassa. Esse nova metodologia faz uso da linearização destes modelos, para que seja viável a consideração da estrutura espacial da população. O Capítulo 4 usa Modelos Multivariados de Covariância Linear Generalizada (McGLM) para avaliar a influência de um dispositivo que ajuda a evitar capturas incidentais de aves. Este dispositivo, chamado de *hookpod* é comprovadamente eficiente para diminuir a captura de aves em pescarias com espinhel de superfície. No entanto, o objetivo do trabalho foi verificar se ele poderia influenciar negativamente na captura de duas espécies de atuns, consideradas alvos da pescaria.

Referências

- Gulland, J. A. (1983). *Fish stock assessment: a manual of basic methods*. John Wiley & Sons.
- Hilborn, R. e Walters, C. J. (1992). *Quantitative fisheries stock assessment: choice, dynamics & uncertainty*. Chapman & Hall, New York.
- Hinde, J. e Demétrio, C. G. (1998). Overdispersion: models and estimation. *Computational Statistics & Data Analysis*, 27(2):151–170.
- Jennings, S., Kaiser, M. J., e Reynolds, J. D. (2001). *Marine fisheries ecology*. Blackwell Publishing, Malden.
- Thorson, J. T., Skaug, H. J., Kristensen, K., Shelton, A. O., Ward, E. J., Harms, J. H., e Benante, J. A. (2015). The importance of spatial models for estimating the strength of density dependence. *Ecology*, 96(5):1202–1212.
- Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(12):1433–1436.
- Wikle, C. K. e Hooten, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST*, 19(3):417–451.

2 MODELAGEM DA ESTRUTURA TEMPORAL DE CAPTURAS INCIDENTAIS EM PESCARIAS COMERCIAIS ATRAVÉS DE MODELOS HIERÁRQUICOS BAYESIANOS

Resumo

Em pescarias comerciais, frequentemente ocorrem capturas de espécies que não são o alvo principal, denominadas de “captura incidental”. No oceano Atlântico, o agulhão-negro (*Makaira nigricans*) é capturado incidentalmente por embarcações que utilizam o espinhel de superfície como artefato de pesca, tendo como alvo os atuns e o espadarte. A análise dos padrões temporais de captura desta espécie em um contexto histórico, pode auxiliar nas decisões para o manejo pesqueiro. O objetivo deste trabalho é construir, avaliar e comparar modelos que descrevam esta estrutura de dependência temporal das capturas do agulhão-negro. Para isso, foram utilizados modelos hierárquicos bayesianos com estrutura gaussiana latente para modelar a captura (em número) dos agulhões capturados no Atlântico sul pela frota do Japão no período 1970–2009. Foram considerados componentes de tendência temporal, assumindo os modelos de passeio aleatório (RW) e auto regressivos (AR), além de um componente sazonal para os trimestres. Os parâmetros foram estimados através de inferência bayesiana, utilizando o método de aproximações de Laplace encaixadas e integradas (INLA). A distribuição mais apropriada aos dados foi a binomial negativa, já que observou-se sobredispersão em relação à distribuição de Poisson. O modelo temporal mais adequado foi o auto regressivo de ordem 10, AR(10). O efeito da tendência temporal foi mais pronunciado do que o efeito da sazonalidade. A inferência realizada através do método INLA demonstrou ser uma alternativa efetiva para o uso em grandes bases de dados, como a utilizada neste trabalho.

Palavras-chave: Captura incidental; Agulhão-negro; Modelos temporais; Inferência bayesiana.

2.1 Introdução

Pescarias comerciais são atividades extrativistas, praticadas em todos os oceanos. Há grandes variações quanto às espécies capturadas e ao ambiente em que se encontram, desde pequenos animais bentônicos (moluscos e camarões) até grandes peixes pelágicos (atuns e tubarões). O impacto gerado pela extração de organismos pela pesca vai desde a alteração do ambiente físico (*e.g.* redes de arrasto passando sobre fundos coralinos) até a depleção dos estoques pesqueiros.

Uma das modalidades de pesca de maior importância econômica é aquela praticada com espinhéis de superfície, que tem como alvo a captura de atuns e espécies afins em regiões oceânicas. No Atlântico, estas pescarias são realizadas principalmente por frotas de países como Estados Unidos, Espanha e Japão. Este tipo de espinhel pelágico consiste de uma extensa linha principal suspensa por boias. Nesta linha principal são conectadas diversas linhas secundárias, as quais tem na extremidade um único anzol. Um espinhel tradicional possui em torno de 60 km de extensão, por onde são distribuídos cerca de 1100 anzóis. Apesar deste aparente alto “poder” de pesca, os espinhéis são artefatos de pesca passivos, e portanto são considerados como um dos métodos de pesca mais “sustentáveis” atualmente.

Em pescarias oceânicas, como as de espinhel de superfície, muitos dos estoques explotados são compartilhados por diversos países. Isso se deve à ampla distribuição e a capacidade de migração dos grandes peixes pelágicos (Sund et al., 1981). Sendo assim, torna-se inviável que cada país tenha suas próprias medidas de manejo. Portanto, para que medidas reguladoras fossem elaboradas e compartilhadas entre as diferentes nações que possuem alguma frota atuando sobre atuns e espécies afins no oceano Atlântico, criou-se a *International Commission for the Conservation of Atlantic Tunas* (ICCAT). Esta comissão tem como finalidade: (a) requerer e organizar as informações pesqueiras de todos os países membros; (b) promover reuniões para a avaliação dos estoques; e (c) recomendar medidas de manejo para as espécies de atuns e afins.

Em pescarias comerciais, frequentemente ocorrem capturas de espécies que não são o alvo principal, denominadas de “captura incidental”. No caso da pescaria com espinhel de superfície no oceano Atlântico, onde os alvos são os atuns e o espadarte, outras espécies de peixes são também atraídas pelas iscas nos anzóis e capturadas de forma não intencional. Este é o caso de várias espécies de tubarões e de agulhões, que possuem populações frágeis e diversas espécies ameaçadas de extinção.

A redução das mortalidades dos agulhões tem sido tomada como meta internacional devido à tendência atual de declínio da abundância dos estoques (ICCAT, 2007). A gravidade do problema fez com que, em 2000, fosse criado um Plano de Recuperação das Populações de Agulhões. Este Plano de Recuperação estabelece duas medidas principais: um limite anual de captura para algumas espécies, e a devolução ao mar de todos os exemplares que ainda estiverem vivos no momento da captura. Mesmo com esta regulamentação, a captura anual dos últimos anos vêm se mantendo acima das capturas máximas recomendadas (ICCAT, 2012). Além disso, a liberação de indivíduos capturados vivos é ainda uma questão em aberto pois não existem informações suficientes sobre as estimativas da proporção de capturados vivos/mortos, e nem da frequência com que os vivos são liberados. No Brasil a pescaria de agulhões é regulamentada por uma instrução normativa (IN 12, de 14 de julho de 2005), que tenta proibir a captura destas espécies, estabelecendo que os agulhões capturados vivos devem ser devolvidos, e os mortos devem ser recolhidos e destinados à doação. No entanto, ainda não existem formas de se evitar a captura incidental destas espécies.

A análise dos padrões de distribuição temporal destas espécies é fundamental para o conhecimento do estado atual dos estoques de agulhões. A identificação de padrões temporais em uma série histórica pode auxiliar no entendimento do comportamento sazonal destas espécies, podendo auxiliar nas tomadas de decisão para o manejo pesqueiro. Por exemplo, podem-se identificar os períodos de maior vulnerabilidade destas espécies para a captura, e optar-se por reduzir o esforço pesqueiro nestas épocas. Nesse sentido, os modelos para séries temporais são ferramentas que podem auxiliar na identificação de padrões que colaborem para o gerenciamento de estoques pesqueiros.

Normalmente, a modelagem da captura de uma determinada espécie é realizada sem qualquer suposição sobre a estrutura ou a dependência das observações no tempo. No entanto, sabe-se por exemplo que as capturas de um ano são influenciadas, entre muitos outros fatores, pelas capturas realizadas no ano anterior. Biologicamente, a abundância de uma população no tempo t é influenciada pela abundância no tempo $t - 1$. Sabe-se ainda, que diferentes espécies tendem a ocupar uma determinada área por diversos motivos, como por exemplo a disponibilidade de comida e a temperatura da água, dependendo da época do ano. Por esses motivos, a dependência temporal deve ser levada em consideração na modelagem de dados de populações biológicas.

O objetivo deste trabalho é construir, avaliar, e comparar modelos que descrevem a estrutura de dependência temporal das capturas do agulhão-negro (*Makaira nigricans*), capturado pela frota de espinhel de superfície que atuam no oceano Atlântico sul. O agulhão-negro é uma espécie de população frágil, capturada de maneira incidental pelos barcos espinheiros, e que faz parte dos planos nacionais e internacionais de manejo pesqueiro. Esta espécie se distribui principalmente em águas temperadas e tropicais do oceano Atlântico, sendo considerada a mais oceânica entre todas as espécies de peixes de bico. Possui peso médio de 150 kg, mas pode chegar a 580 kg e mais de 3 metros de comprimento (de Sylva e Breder, 1997).

Os modelos analisados foram construídos considerando diferentes estruturas para a dependência temporal. A intenção é entender o comportamento temporal das capturas do agulhão-negro, com a finalidade de que este conhecimento possa servir de embasamento para análises futuras, além de poder auxiliar nas tomadas de decisão no manejo desta pescaria no Atlântico sul. Em termos de manejo, as variações sazonais na captura de um determinado recurso podem auxiliar na compreensão de épocas com maior vulnerabilidade do estoque à atividade de pesca e, com isto, permitir que o processo de decisão

considere a adoção de medidas de mitigação e redução de capturas focadas nestes períodos. Além disso, variações temporais e sazonais na captura podem ainda refletir, indiretamente, respostas do recurso às taxas de remoções sofridas, seja por interface direta da atividade de pesca, seja por componentes indiretas resultantes de processos não avaliados no manejo, mas que estão atuando sobre o estoque de alguma maneira (*e.g.* variações climáticas, sucesso nos padrões reprodutivos).

2.2 Metodologia

2.2.1 Dados

Os dados de captura em número do agulhão-negro são provenientes do banco de dados *Task 2* da *International Commission for the Conservation of Atlantic Tunas* (ICCAT). Os países signatários desta comissão tem o direito de participar de todo o processo de debate e deliberação sobre novas medidas de manejo. Em contrapartida é requisitado que todas as nações reportem anualmente informações sobre as diversas pescarias de atuns e afins que ocorrem sob sua jurisdição. A base de dados analisada (assim como outras informações) é de domínio público, e pode ser consultada no endereço www.iccat.int.

A princípio, foram selecionadas as entradas correspondentes a todas as frotas que atuaram sobre o estoque sul do agulhão-negro (latitudes ao sul de 5°N). Através de uma breve análise do número de informações por ano, trimestre e frota, constatou-se que apenas a frota do Japão apresenta dados suficientes para uma análise temporal da captura dos agulhões. As demais frotas apresentaram informações dispersas ao longo do tempo, e muitas vezes com longos intervalos anuais sem dados. Por esse motivo, optou-se por utilizar os dados referentes apenas à frota do Japão, a partir do ano de 1970 até 2009, uma vez que essa foi a série temporal utilizada na última avaliação de estoque do agulhão-negro (ICCAT, 2012).

Cada linha da base de dados contém informações como a posição geográfica, o trimestre, o esforço de pesca (número de anzóis) e a captura. Com o objetivo de se analisar a série temporal das capturas do estoque sul do agulhão-negro, os dados de captura e esforço foram agregados por ano e trimestre. Com isso, foram obtidas 160 linhas de informações, a partir do primeiro trimestre de 1970 até o quarto trimestre de 2009.

2.2.2 Definição dos modelos

Para analisar o padrão temporal das capturas do agulhão-negro no Atlântico sul, foram construídos modelos hierárquicos bayesianos com estrutura gaussiana latente para representar os efeitos temporais e sazonais. Nesta abordagem, assume-se que a variável resposta “captura” (em forma de contagem do número de peixes), Y_{it} , com $i = 1, \dots, n$ trimestres ($n = 4$) e $t = 1, \dots, T$ unidades de tempo (combinação de ano e trimestre, $T = 160$), segue uma distribuição de probabilidade apropriada para contagens. Neste trabalho foram analisados e comparados os modelos assumindo as distribuições de Poisson e binomial negativa.

A distribuição de Poisson é a escolha natural normalmente utilizada para modelar dados em forma de contagens. Sua função de probabilidade é dada por

$$P[Y = y] = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, \dots$$

onde o parâmetro $\lambda > 0$ é o valor médio das contagens. A esperança e a variância são dadas por

$$E(Y) = \mu = \lambda \quad \text{e} \quad \text{Var}(Y) = \sigma^2 = \lambda$$

onde nota-se que uma propriedade importante desta distribuição é que a esperança e a variância são iguais. Na prática, esta propriedade muitas vezes não é verificada, devido à uma série de fatores que

fazem com que a variância das contagens seja menor ou maior do que a média, fenômenos denominados de subdispersão e sobredispersão, respectivamente.

Uma forma de contornar o problema da sobredispersão observada em dados amostrais, é utilizar a distribuição binomial negativa. A parametrização desta distribuição adotada aqui é aquela que considera $Y \sim \text{Pois}(\lambda)$, e λ também possui uma distribuição de probabilidade. Se λ segue uma distribuição gama com parâmetros $\delta > 0$ e $\beta > 0$, de forma que $E(\lambda) = \delta/\beta$ e $\text{Var}(\lambda) = \delta/\beta^2$, então a função de probabilidade de Y será a binomial negativa dada por

$$P[Y = y] = \binom{y + \delta - 1}{y} \pi^\delta (1 - \pi)^{y}, \quad y = 0, 1, \dots \quad (2.1)$$

onde $\pi = \beta/(\beta + 1)$ é a probabilidade de sucesso do evento de interesse, e δ é o número de sucessos, que não necessariamente precisa ser inteiro (ver desenvolvimento no Apêndice A). Dessa forma, a esperança e a variância de Y são dadas por

$$E(Y) = \mu = \delta \frac{1 - \pi}{\pi} \quad \text{e} \quad \text{Var}(Y) = \sigma^2 = \mu \left(1 + \frac{\mu}{\delta}\right).$$

Note que, nesse caso, a variância sempre excede a média, já que $\delta > 0$. Por esse motivo, δ também é chamado de parâmetro de sobredispersão, e a distribuição binomial negativa é mais indicada quando a variância é maior do que a média em dados de contagem.

Quando ambas distribuições dependem de variáveis explicativas, elas estarão funcionalmente associadas com a média da variável resposta através de um preditor linear η_{it} . A esperança da variável resposta é associada com o preditor linear através de uma função de ligação $g(\cdot)$ conhecida, monótona e diferenciável

$$g(\mu_{it}) = f_{it} \exp(\eta_{it}).$$

A função de ligação canônica (que lineariza a relação entre a média da variável resposta e o preditor linear) para os modelos Poisson e binomial negativo é a logarítmica, portanto

$$\ln(\mu_{it}) = \ln(f_{it}) + \eta_{it}$$

onde f_{it} é o intervalo de observação para as contagens, um valor conhecido porém não necessariamente constante, conhecido como *offset*. Quando este intervalo não é constante, então é necessária a sua inclusão no preditor linear, para que as contagens sejam consideradas apropriadamente. Neste trabalho foi considerado o número de anzóis (esforço de pesca) como *offset* no modelo, uma vez que as capturas foram observadas em diferentes cenários e com números de anzóis desiguais.

O modelo latente completo como definido na expressão anterior, possui o preditor linear dado por

$$\eta_{it} = \alpha + \rho_i + \gamma_t. \quad (2.2)$$

O parâmetro α representa o nível médio do processo (intercepto), ρ_i é o componente sazonal, e γ_t é o componente de tendência temporal. Estes dois componentes podem ser estruturados ou não estruturados. Quando não assumem nenhuma relação temporal são ditos não estruturados, e assume-se que as observações sejam independentes e identicamente distribuídas (iid). Quando assume-se que as observações possuem alguma dependência temporal, estes componentes podem ser estruturados de acordo com algum processo de dependência no tempo.

O modelo formulado em (2.2) é completamente especificado quando definem-se as distribuições *a priori* para os dois componentes, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)'$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_T)'$, que definem blocos de efeitos aleatórios em η_{it} . Assume-se como *priori*, que estas distribuições são normais multivariadas com vetor de médias igual a zero e matriz de precisão $\tau \mathbf{K}$, onde τ é um escalar desconhecido e \mathbf{K} é a matriz que define a estrutura de dependência temporal. Nesta abordagem, τ_ρ e τ_γ são parâmetros de precisão associados aos

efeitos sazonal e temporal, respectivamente, e que indicam a relevância de cada um dos termos de efeitos aleatórios. Como a precisão é definida como o inverso da variância, espera-se que termos importantes no modelo possuam uma baixa precisão, indicando uma alta variabilidade, e portanto, uma importância maior do termo.

Diferentes matrizes \mathbf{K} podem ser especificadas, descrevendo a estrutura de dependência temporal em cada componente do modelo. Para o termo de componente sazonal ρ , um modelo com periodicidade m para o vetor $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)$, $n > m$, é obtido assumindo que as somas $\rho_1 + \rho_{i+1} + \dots + \rho_{i+m-1}$ são independentes e possuem distribuição normal com média zero e parâmetro de precisão τ_ρ . Neste trabalho, adotou-se a periodicidade trimestral, portanto $m = 4$.

Para o termo de componente de tendência temporal, γ , diversas estruturas de dependência da matriz \mathbf{K} foram avaliadas. Inicialmente, pode-se considerar que uma observação no tempo t é influenciada apenas pela observação no tempo anterior ($t - 1$). Esse processo é denominado de passeio aleatório, ou *random walk*, de primeira ordem, e denotado por RW(1). Considerando o vetor gaussiano $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_T)$, o modelo RW(1) é definido por

$$\Delta\gamma_t = \gamma_{t+1} - \gamma_t, \quad t = 1, \dots, T - 1 \quad (2.3)$$

onde $\Delta\gamma_t \sim N(0, \tau_\gamma^{-1})$ e também é chamado de diferença de primeira ordem entre duas observações consecutivas. Pode-se ainda definir passeios aleatórios de ordens superiores. Por exemplo, um processo *random walk* de ordem 2, RW(2), é obtido a partir da diferença de segunda ordem, ou seja, a diferença entre diferenças de primeira ordem consecutivas (Natário e Knorr-Held, 2003)

$$\begin{aligned} \Delta^2\gamma_t &= \Delta(\Delta\gamma_t) \\ &= \Delta(\gamma_{t+1} - \gamma_t) \\ &= (\gamma_{t+1} - \gamma_t) - (\gamma_t - \gamma_{t-1}) \\ &= \gamma_{t+1} - 2\gamma_t + \gamma_{t-1}, \quad t = 2, \dots, T - 1 \end{aligned} \quad (2.4)$$

onde $\Delta^2\gamma_t \sim N(0, \tau_\gamma^{-1})$. Sendo assim, note que, dados os valores observados em todos os tempos, tanto as observações no tempo ($t - 1$), quanto as observações no tempo ($t + 1$), ajudam a prever as observações no tempo t . De fato, o termo $\gamma_{t+1} - 2\gamma_t + \gamma_{t-1}$ pode ser interpretado como uma estimativa da derivada de segunda ordem de uma função no tempo contínuo $\gamma(t)$ no ponto $t = i$, usando valores de $\gamma(t)$ nos pontos $t = i + 1$, i , e $i - 1$ (Lindgren e Rue, 2008). Portanto, na prática, a diferença entre os modelos RW(1) e RW(2), é que este último deverá apresentar um comportamento mais suavizado do que o primeiro (Fahrmeir e Lang, 2001).

Outro processo possível para o termo de componente de tendência temporal é considerar que as observações em um ano são dependentes somente das observações nos anos anteriores. Nesse caso, um modelo sugerido é o auto regressivo de ordem p , AR(p), o qual especifica que a variável resposta depende linearmente dos seus valores anteriores, gerando uma estrutura Markoviana. Os modelos AR(p) são definidos para o mesmo vetor gaussiano $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_T)$, como

$$\gamma_t = \phi_1 \gamma_{t-1} + \phi_2 \gamma_{t-2} + \dots + \phi_p \gamma_{t-p} + \epsilon_t, \quad (2.5)$$

para $t = p, \dots, n$, com parâmetros ϕ_1, \dots, ϕ_p e ϵ_t representando o ruído branco com $E(\epsilon_t) = 0$, $\text{Var}(\epsilon_t) = \sigma_\gamma^2 = \tau_\gamma^{-1}$, e $\text{Cov}(\epsilon_i, \epsilon_j) = 0, \forall i \neq j$. Isso implica que o ruído branco possui média zero, variância constante e que não existe correlação entre eles. Sendo assim, para um processo AR(1) com ϕ positivo, apenas a observação de um ano anterior do processo e o termo de ruído branco contribuem para a observação atual. Se ϕ está próximo de zero, então o processo se aproxima apenas do ruído branco.

Para se determinar a ordem p do processo autoregressivo para a captura de agulhões, foram ajustados todos os modelos AR(p) para $p = 1, \dots, 10$. Além disso, os modelos AR(p) foram comparados

com os modelos RW(1) e RW(2) para determinar qual destes processos é mais plausível para explicar a variabilidade temporal da captura do agulhão-negro. A comparação e escolha do modelo mais apropriado foi realizada conforme critérios de seleção de modelos apresentados na próxima seção.

Para completar a especificação do modelo, é necessário especificar as distribuições *a priori* para todos os parâmetros envolvidos. Seguindo a proposta de Natário e Knorr-Held (2003) e Schrödle et al. (2011), assume-se que os parâmetros de precisão, τ_ρ e τ_γ para os efeitos sazonal (ρ) e temporal (γ), associados à matriz estrutural \mathbf{K} , possuem distribuição *a priori* $G(1, 0.00005)$, onde $G(a, b)$ é a distribuição gama com esperança a/b e variância a/b^2 . Para o parâmetro α , que determina a média geral do processo, assumiu-se uma distribuição *a priori* vaga, ou seja, uma uniforme no intervalo $(-\infty, \infty)$.

As estruturas assumidas para os efeitos temporais definem campos aleatórios gaussianos markovianos (*Gaussian Markov Random Fields*, GMRF) (Rue e Held, 2005). A grande dimensão do vetor de efeitos aleatórios pode tornar a inferência baseada em simulação, como os métodos de Monte Carlo via Cadeias de Markov (MCMC), computacionalmente ineficientes ou até mesmo inviáveis. Por esse motivo, foi utilizado um método baseado em aproximações numéricas para inferência bayesiana em modelos gaussianos latentes altamente estruturados, definido inicialmente por Rue et al. (2009). A metodologia denominada INLA (*Integrated Nested Laplace Approximation*) faz uso de métodos de integração numérica para se obterem as distribuições marginais *a posteriori*, e contorna os problemas de elevado tempo computacional e de convergência em métodos como o MCMC. As funções para o cálculo das distribuições posteriores estão implementadas no ambiente estatístico R (R Core Team, 2018), através do pacote **INLA** (Lindgren e Rue, 2015), disponível em www.r-inla.org.

2.2.3 Seleção de modelos

Ao todo, foram ajustados 14 modelos para cada distribuição (Poisson e binomial negativa): o modelo nulo (apenas com o intercepto), um modelo que considera que não existe nenhum tipo de relação temporal entre as observações (*i.e.* observações são independentes e identicamente distribuídas, iid), os modelos de passeio aleatório RW(1) e RW(2) como definidos em (2.3) e (2.4), e os modelos AR(p), $p = 1, \dots, 10$ como definidos em (2.5). Em todos estes modelos também foi considerado o efeito sazonal para os trimestres.

Para a comparação destes modelos, foram utilizados quatro critérios em conjunto: a redução na deviance, o número efetivo de parâmetros, o critério de informação da deviance (DIC), e a ordenada preditiva condicional (*Conditioning Predictive Ordinate*, CPO). Como medida de diagnóstico do modelo selecionado foi utilizada a transformação integral de probabilidade (*Probability Integral Transform*, PIT).

A distribuição posterior da estatística deviance é dada por

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y}|\boldsymbol{\theta}) + 2 \log h(\mathbf{y})$$

onde $f(\mathbf{y}|\boldsymbol{\theta})$ é a função de verossimilhança para o vetor de dados observado \mathbf{y} , dado o vetor de parâmetros $\boldsymbol{\theta}$, e $h(\mathbf{y})$ é uma função de padronização que depende apenas dos dados (não interferindo no processo de seleção de modelos). Dessa forma, uma medida de bondade de ajuste do modelo pode ser resumida pela esperança da distribuição posterior da deviance, $\bar{D} = E_{\boldsymbol{\theta}|\mathbf{y}}[D]$, e uma medida de complexidade do modelo pode ser representada pelo número efetivo de parâmetros p_D ,

$$p_D = E_{\boldsymbol{\theta}|\mathbf{y}}[D] - D(E_{\boldsymbol{\theta}|\mathbf{y}}[\boldsymbol{\theta}]) = \bar{D} - D(\bar{\boldsymbol{\theta}})$$

ou seja, a diferença entre a esperança de distribuição posterior da deviance e a deviance da distribuição posterior no seu ponto médio (Spiegelhalter et al., 2002). De maneira geral, o número efetivo de parâmetros deve estar entre zero e o número total de parâmetros do modelo (Carlin e Louis, 2009). Como a deviance é uma função dos dados, então o número efetivo de parâmetros também é uma variável aleatória, e é razoável supor que esta medida também dependa dos dados (Gelman et al., 2014). Dessa forma,

dados mais dispersos podem apresentar valores de p_D maiores. Para mais detalhes veja Spiegelhalter et al. (2002) e Carlin e Louis (2009, p. 211). O critério de informação da deviance (DIC) é então definido como

$$\text{DIC} = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta})$$

que considera tanto os valores da deviance quanto do número efetivo de parâmetros (Spiegelhalter et al., 2002). Assim, valores de DIC menores indicam modelos mais bem ajustados. Note que valores individuais de DIC não possuem uma interpretação prática. Apenas a diferença entre DICs de diferentes modelos são de interesse (Carlin e Louis, 2009).

A ordenada preditiva condicional (CPO) é definida por

$$f(y_i|\mathbf{y}_{(i)}) = \frac{f(\mathbf{y})}{f(\mathbf{y}_{(i)})} = \int f(y_i|\boldsymbol{\theta}, \mathbf{y}_{(i)})p(\boldsymbol{\theta}|\mathbf{y}_{(i)})d\boldsymbol{\theta}$$

que define o valor da verossimilhança para cada observação i (y_i), dado o restante das observações sem i ($\mathbf{y}_{(i)}$). Note que esta medida pode ser utilizada como uma forma de validação cruzada. A média do logaritmo da CPO, $\text{LCPO} = -\frac{1}{160} \sum_{i=1}^{160} \log(\text{CPO}_i)$, é uma medida utilizada como indicativo do poder de predição de um modelo (Beguin et al., 2012). Nesse caso, modelos com valores menores de LCPO possuem um maior capacidade preditiva (Held et al., 2010). Uma característica importante desta medida é que ela pode ser utilizada para comparação entre modelos sem a necessidade de que estes modelos sejam aninhados ou mesmo relacionados de alguma forma (Schrödle et al., 2011).

A transformação integral de probabilidade (PIT) é definida como $\text{PIT}_i = P(Y_i \leq y_i|\mathbf{y}_{(i)})$, onde Y_i é uma replicação da observação y_i (Gelman et al., 2014, p. 145). Esta é uma medida utilizada para se avaliar a adequação da suposição da distribuição assumida para a variável resposta. No caso de distribuições contínuas, um histograma com os valores de PIT deve mostrar uma distribuição que se aproxima de uma uniforme no intervalo [0,1] (Schrödle e Held, 2011). No entanto, para distribuições discretas, como as assumidas aqui, essa suposição não é mais válida. Czado et al. (2009) propuseram então um ajuste para os valores de PIT quando são utilizadas distribuições discretas. Portanto, o PIT ajustado $\text{PIT}_{\text{a}i} = \text{PIT}_i + 0.5P(Y_i = y_i|\mathbf{y}_{(i)})$, proposto por aqueles autores, foi utilizado neste trabalho. Dessa forma, permanece a interpretação de que se a distribuição dos valores de PITa for próxima de uma uniforme, a variável resposta assumida é adequada (Schrödle et al., 2011).

2.3 Resultados e discussão

A comparação entre os modelos ajustados para as distribuições de Poisson e binomial negativa está demonstrada na Tabela 2.1. Nota-se que em comparação, os valores de deviance e DIC para os modelos ajustados com a distribuição de Poisson são menores do que as mesmas medidas obtidas para os modelos ajustados com a distribuição binomial negativa. Os valores ligeiramente menores de LCPO também indicam que os modelos de Poisson possuem uma melhor capacidade preditiva. Através destas comparações, os modelos de Poisson parecem ser mais adequados. No entanto, chama a atenção o elevado número efetivo de parâmetros para os modelos de Poisson. Considerando que o conjunto de dados analisado é composto de uma combinação de todos os trimestres para os anos 1970–2009, existem $40 \times 4 = 160$ observações. Em todos os modelos ajustados considerando a distribuição de Poisson, o número efetivo de parâmetros foi sempre superior a 154, ou seja, o modelo não contribui para a redução dos dados. De fato, foi verificado através dos valores preditos de um destes modelos que houve praticamente a reprodução das observações originais. Uma possível explicação para esse fato é que, devido à variabilidade dos dados, a distribuição de Poisson (que possui a propriedade de que a média é igual a variância) não foi capaz de capturar esta alta variabilidade, e uma grande quantidade de parâmetros foi necessária para que o modelo pudesse ser ajustado, promovendo um superajustamento.

Tabela 2.1. Deviance residual, número efetivo de parâmetros (p_D), critério de informação da deviance (DIC), e logaritmo da ordenada preditiva condicional (LCPO) para os diferentes modelos temporais ajustados, considerando as distribuições de poisson e binomial negativa.

Modelo	Poisson				Binomial negativa			
	Deviance	p_D	DIC	LCPO	Deviance	p_D	DIC	LCPO
NULL	62603,61	3,46	62607,07	180,77	2395,67	1,00	2397,31	7,49
IID	1435,52	158,71	1594,24	6,16	2382,14	2,89	2387,00	7,46
RW1	1441,00	156,85	1597,88	6,11	2293,40	20,17	2314,81	7,22
RW2	1440,62	156,22	1596,88	6,08	2299,58	11,87	2313,01	7,22
AR1	1439,21	157,25	1596,51	6,10	2291,96	22,60	2315,72	7,23
AR2	1438,90	157,23	1596,21	6,09	2293,95	19,36	2314,87	7,22
AR3	1437,67	155,66	1593,50	6,06	2165,25	67,22	2231,34	7,77
AR4	1441,08	155,85	1597,13	6,06	2181,77	55,81	2239,43	7,41
AR5	1443,05	155,30	1598,60	6,05	2188,75	54,12	2243,28	7,40
AR6	1443,28	155,24	1598,79	6,06	2171,57	60,87	2231,18	7,54
AR7	1443,38	155,23	1598,96	6,06	2201,38	44,90	2248,89	7,08
AR8	1438,51	154,92	1593,78	5,99	2201,20	44,28	2248,70	7,07
AR9	1438,61	154,82	1593,82	5,99	2190,62	48,09	2237,99	7,17
AR10	1438,35	154,79	1593,52	5,99	2152,47	56,90	2216,03	7,02

As distribuições dos valores da transformação integral de probabilidade ajustados (PITA) para cada modelo considerando a distribuição de Poisson estavam mais próximas de uma distribuição simétrica do que de uma uniforme, como seria esperado se a suposição de que a variável resposta segue uma distribuição de Poisson fosse adequada. Em geral, os histogramas de PITA para os modelos considerando a distribuição binomial negativa apresentaram distribuições mais próximas de uma uniforme no intervalo [0,1]. Isso indica, portanto, que a distribuição mais adequada para a variável resposta seria de fato a binomial negativa. Portanto, mesmo que os valores de DIC e LCPO tenham sido menores para os modelos que consideraram a distribuição de Poisson, as evidências indicam que a distribuição binomial negativa é mais adequada, tanto pelo menor número efetivo de parâmetros, quanto pela distribuição dos valores de PITA.

Por estes motivos, desconsideraram-se os modelos de Poisson, e os modelos ajustados com a distribuição binomial negativa foram analisados individualmente. Entre estes modelos, percebe-se inicialmente uma diferença considerável dos valores da deviance para os modelos nulo e não estruturado (iid), ou seja, qualquer outro modelo que considere alguma relação temporal (RW ou AR) para os dados é mais apropriado (Tabela 2.1). Este primeiro resultado já indica a necessidade de se considerar os efeitos de dependência temporal para os dados de captura do agulhão-negro no Atlântico sul.

Entre os modelos estruturados temporalmente, nota-se pelos valores da deviance e DIC, que, em geral, os modelos que consideram a estrutura auto regressiva (AR) são mais bem ajustados do que os modelos estruturados através da estrutura de passeios aleatórios (RW). Os valores de LCPO para os modelos RW são, em geral, maiores do que os mesmos valores para a maioria dos modelos AR, o que indica um menor poder preditivo dos modelos com estrutura de passeio aleatório. Entre os modelos AR, aquele que apresentou os menores valores da deviance e DIC foi o AR(10). Este mesmo modelo também possui o menor valor de LCPO, sendo, portanto, aquele com maior poder de predição. Para ajudar na decisão sobre o modelo a ser selecionado, na Figura 2.1 são apresentados todos os modelos com seus respectivos valores de deviance, DIC e número efetivo de parâmetros. Em uma situação ideal, o melhor modelo entre um conjunto de possíveis modelos, é aquele que apresenta o menor valor nestes três critérios em conjunto. Através desta figura, observa-se que o modelo AR(10) é o que está mais próximo desta situação ideal, uma vez que possui os menores valores de deviance e DIC entre todos os modelos, e ainda possui o menor número de parâmetros entre os modelos mais próximos em termos de deviance e DIC. Por estes resultados, há uma tendência para a escolha do modelo AR(10).

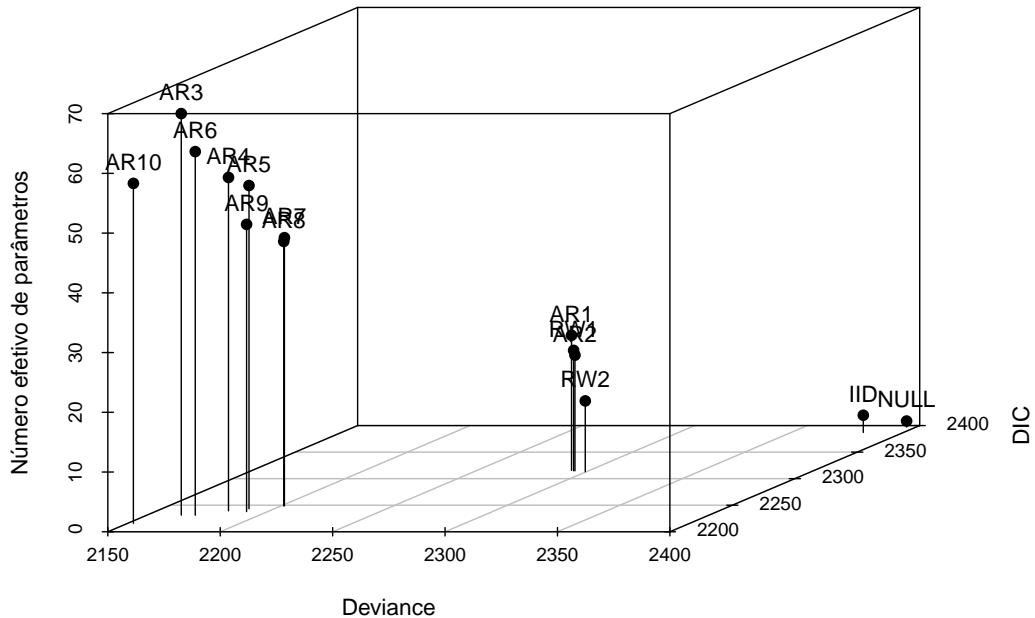


Figura 2.1. Gráfico em três dimensões com os valores de deviance, DIC, e número efetivo de parâmetros para todos os modelos considerados, assumindo a distribuição binomial negativa para a variável resposta.

A Tabela 2.2 apresenta a mediana da distribuição posterior para cada efeito aleatório estimado em todos os modelos considerados (por conveniência, os valores para as autocorrelações parciais para todos os modelos foram suprimidas, mas serão analisadas para o modelo escolhido). Nesse caso, espera-se que valores elevados para o parâmetro de precisão sejam de pouca importância. Uma vez que a precisão é o inverso da variância, componentes com elevado valor de precisão apresentam pouca variabilidade, indicando, portanto, que seus valores estão próximos da média, que por construção do modelo é zero. Ou seja, componentes com alta precisão (baixa variabilidade), tem pouco efeito sobre a variável resposta. Nota-se pela Tabela 2.2 que a precisão para o parâmetro do componente de tendência temporal (γ) é parecida na maioria dos modelos, com exceção dos modelos IID, RW1 e RW2, que apresentaram valores elevados para este componente. Já para o parâmetro do componente de sazonalidade (ρ), os valores medianos da precisão são mais variados. Analisando especificamente os resultados para o modelo AR(10), que mostrou ser o mais adequado pelos critérios anteriores, nota-se que a precisão para o termo de tendência temporal é a segunda menor entre todos os modelos. Já para o componente de sazonalidade, a precisão não é tão alta comparativamente. No entanto, parâmetro de sobredispersão (δ) para o modelo AR(10) apresentou o valor mediano mais alto entre todos os modelos. Dessa forma, optou-se por selecionar o modelo AR(10) para se fazerem previsões e análises de efeitos dos parâmetros sazonais e temporais.

As distribuições posteriores para os efeitos aleatórios (componentes de tendência temporal, sazonalidade e autocorrelações parciais) do modelo AR(10), e do parâmetro de sobredispersão da distribuição binomial negativa estão apresentados na Figura 2.2. Percebe-se inicialmente que a distribuição posterior para o parâmetro de precisão do componente de tendência temporal não contém o zero, e seu valor mediano é baixo, indicando que existe uma grande influência temporal nas capturas do agulhão-negro no Atlântico sul. Da mesma forma, a distribuição posterior para o parâmetro de sazonalidade não inclui

Tabela 2.2. Mediana da distribuição posterior do intercepto (α), dos efeitos aleatórios (componente de tendência temporal, γ , e componente sazonal, ρ), e do parâmetro de sobredispersão (δ) da distribuição binomial negativa, para os modelos considerados.

Modelo	Efeitos			
	α	δ	γ	ρ
NULL	-9,34	1,29		
IID	-9,37	1,40	13293,40	30,83
RW1	-9,57	2,28	50,37	12,90
RW2	-9,56	2,20	4837,28	12,99
AR1	-9,54	2,30	2,87	12,73
AR2	-9,54	2,29	2,86	12,79
AR3	-9,60	4,78	1,07	34,74
AR4	-9,61	4,45	1,15	36,82
AR5	-9,61	4,14	1,21	24,50
AR6	-9,61	3,99	1,32	49,47
AR7	-9,61	3,99	1,30	32,61
AR8	-9,61	4,04	1,32	22,14
AR9	-9,61	3,98	1,32	29,04
AR10	-9,66	5,48	1,10	19,08

o zero, mas sua distribuição exibe grande assimetria à direita. De acordo com a Tabela 2.2, a mediana desta distribuição posterior é 19,08, que associado com a grande assimetria, indica que o termo sazonal não possui grande importância para explicar as capturas do agulhão-negro. Em relação ao parâmetro de sobredispersão da distribuição binomial negativa, nota-se que sua distribuição posterior não inclui o zero, e possui valor mediano de 5,48 (ver Tabela 2.2), confirmando a presença de sobredispersão em relação à distribuição de Poisson, mesmo com a inclusão dos efeitos aleatórios.

As distribuições posteriores para os parâmetros de autocorrelação parcial estimadas pelo modelo AR(10) (Figura 2.2) mostram que, em geral, estes termos possuem baixa precisão (valores próximos do zero), ou seja, são importantes para a explicação do modelo. Nota-se que quatro distribuições posteriores, entre as dez das FACP's, incluem o zero: FACP2, FACP7, FACP8 e FACP9. Para ficar mais claro a interpretação destes parâmetros, a Figura 2.3 mostra apenas o valor mediano destas distribuições posteriores, e indica aqueles cujo intervalo de credibilidade inclui (linhas pontilhadas) ou não (linhas sólidas) o valor zero. Por esta figura, percebe-se que a dependência temporal é positiva até a quarta diferença de tempo, ficando negativa na quinta e novamente positiva na sexta. A partir desse ponto as FACP's estimadas são não significativas. Apenas na décima diferença de tempo a FACP volta a ser significativa e é negativa. O fato da distribuição posterior da FACP10 não incluir o zero, mostra que é necessário considerar uma longa dependência temporal para explicar a variabilidade das capturas do agulhão-negro ao longo dos anos.

O fato desta espécie possuir uma alta longevidade (em torno de 20 anos, Andrews et al., 2017) pode ser um motivo para explicar essa longa dependência. Espécies longevas apresentam um crescimento mais lento, e, por consequência, uma maior idade de primeira maturação sexual. Alguns estudos (*e.g.*, de Sylva e Breder, 1997) mencionam que o tamanho de primeira maturação se dá em torno dos 230 cm. Considerando que seu tamanho máximo pode chegar a mais de 350 cm (Goodyear, 2015), é possível deduzir, de forma relativa, que o recrutamento leve alguns poucos anos para ocorrer, fazendo com que as capturas sejam dependentes dos recrutamentos em tempos mais distantes.

A Figura 2.4 apresenta os efeitos temporais (tendência temporal) e sazonais, com bandas de credibilidade de 95%. Em relação aos efeitos temporais, observa-se claramente o padrão cíclico de capturas, conforme o trimestre do ano (Figura 2.4A). Nota-se uma queda nas capturas do agulhão-negro até o primeiro trimestre de 1979. De maneira geral, o período entre o primeiro trimestre de 1976 e o mesmo para 1979, foi o único que apresentou intervalos de credibilidade que não incluem o zero, mos-

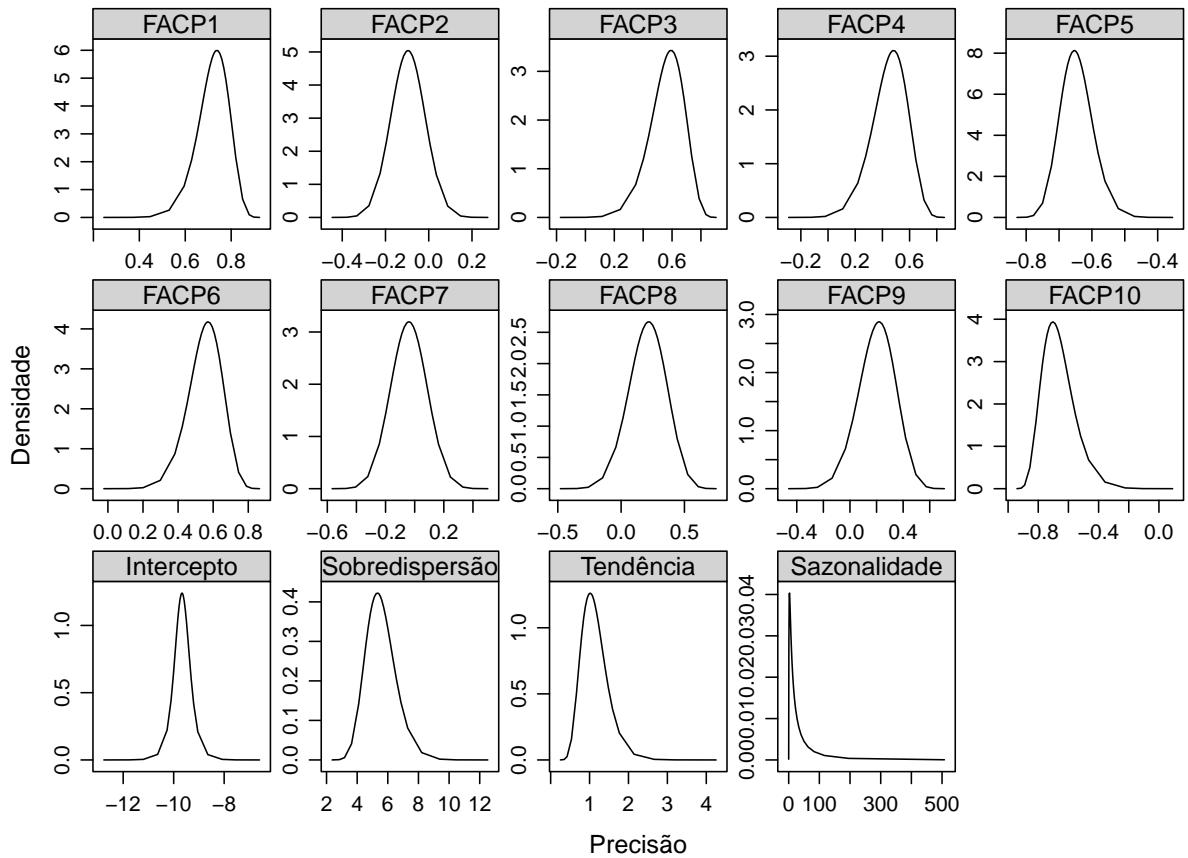


Figura 2.2. Distribuições posteriores para o efeito fixo (intercepto), e aleatórios (tendência, sazonalidade e funções de autocorrelação parcial, FACP) do modelo proposto, além da distribuição posterior para o parâmetro de sobredispersão da distribuição binomial negativa.

trando assim que esta foi uma época onde as capturas do agulhão-negro eram relativamente mais baixas. Este fato pode estar associado à uma combinação de fatores como por exemplo a estratégia de pesca da frota japonesa naquele período (*e.g.* tipo e profundidade de atuação do anzol, horário de lançamento do equipamento), que fazia com que menos agulhões fossem capturados, ou à uma redução natural da população dos agulhões. Passada essa fase, todos os efeitos para os demais períodos incluem o zero no intervalo de credibilidade, indicando portanto que não houveram alterações significativas na captura do agulhão-negro para todo o período analisado. Sem poder tirar maiores conclusões, nota-se apenas que a partir de 1985 os efeitos temporais são positivos (apenas o limite inferior do intervalo de credibilidade está próximo do zero), demonstrando capturas mais elevadas. A partir do ano 2000, há uma pequena diminuição e uma oscilação em torno do zero até o final de 2009.

Os efeitos sazonais para os trimestres mostram que as capturas são maiores no primeiro e no quarto trimestres (Figura 2.4B), que são os períodos mais quentes no hemisfério austral. Normalmente nas épocas mais quentes ocorre a agregação de indivíduos desta espécie para fins de reprodução, o que torna os indivíduos mais vulneráveis à pesca. No segundo e no terceiro trimestres os efeitos sazonais são negativos. No entanto, maiores conclusões não podem ser feitas pois em todos os casos, os intervalos de credibilidade para as distribuições posteriores contém o valor zero.

A Figura 2.5 mostra a série temporal analisada, com a captura total (em número) de agulhões-negros, realizada pela frota do Japão para o período de 1970 a 2009. Sobreposta à esta série, estão os valores de captura estimados (mediana) pelo modelo AR(10), com as respectivas bandas de confiança.

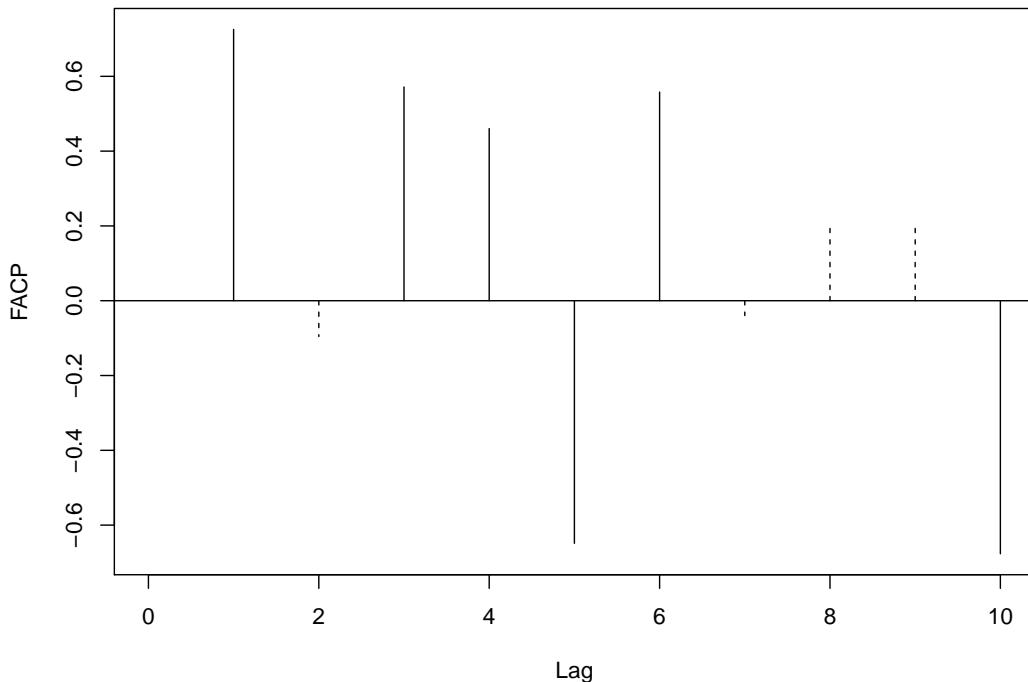


Figura 2.3. Mediana da distribuição posterior das funções de autocorrelação parcial (FACP) obtidas pelo modelo AR(10). As linhas sólidas indicam que o intervalo de credibilidade de 95% para a distribuição não inclui o zero. As linhas pontilhadas indicam que o mesmo intervalo contém o zero.

Percebe-se o bom ajuste do modelo, que de maneira geral tende a seguir o comportamento dos dados observados. Através desta série, pode-se notar que as capturas eram relativamente baixas até meados de 1981. A partir desse ano, houve um aumento considerável das capturas, embora sempre de maneira oscilatória, até aproximadamente o final dos anos 1990. A partir do início da década de 2000, as capturas foram reduzidas, embora não chegando aos patamares mais baixos observados no início da série. No final da década de 2000, há uma pequena tendência de aumento das capturas, mas ainda com uma grande oscilação.

2.4 Conclusões

A classe de modelos proposta para se analisar as capturas do agulhão-negro no Atlântico sul, permite explorar diversas possibilidades para a estrutura de dependência temporal entre as observações. Através da metodologia utilizada, é possível adquirir uma maior flexibilidade na modelagem de séries temporais. Por exemplo, é possível a utilização de distribuições apropriadas para a variável resposta, como é o caso das contagens com sobredispersão neste caso. Além disso, também é possível ajustar diversos modelos com estruturas temporais diferentes e fazer a comparação e seleção através de abordagens baseadas em critérios gerais de comparação e seleção de modelos.

Através da análise de comparação dos modelos, notou-se que o modelo de Poisson, que é comumente utilizado para se modelar dados de contagens, não foi capaz de fornecer uma descrição parcimoniosa dos dados. Pelo contrário, o modelo de Poisson necessitou de um número de parâmetros muito próximo ao número de informações na amostra, sendo ineficaz no objetivo de resumir os dados. Por outro lado, a modelagem utilizando a distribuição binomial negativa para a variável resposta foi claramente superior,

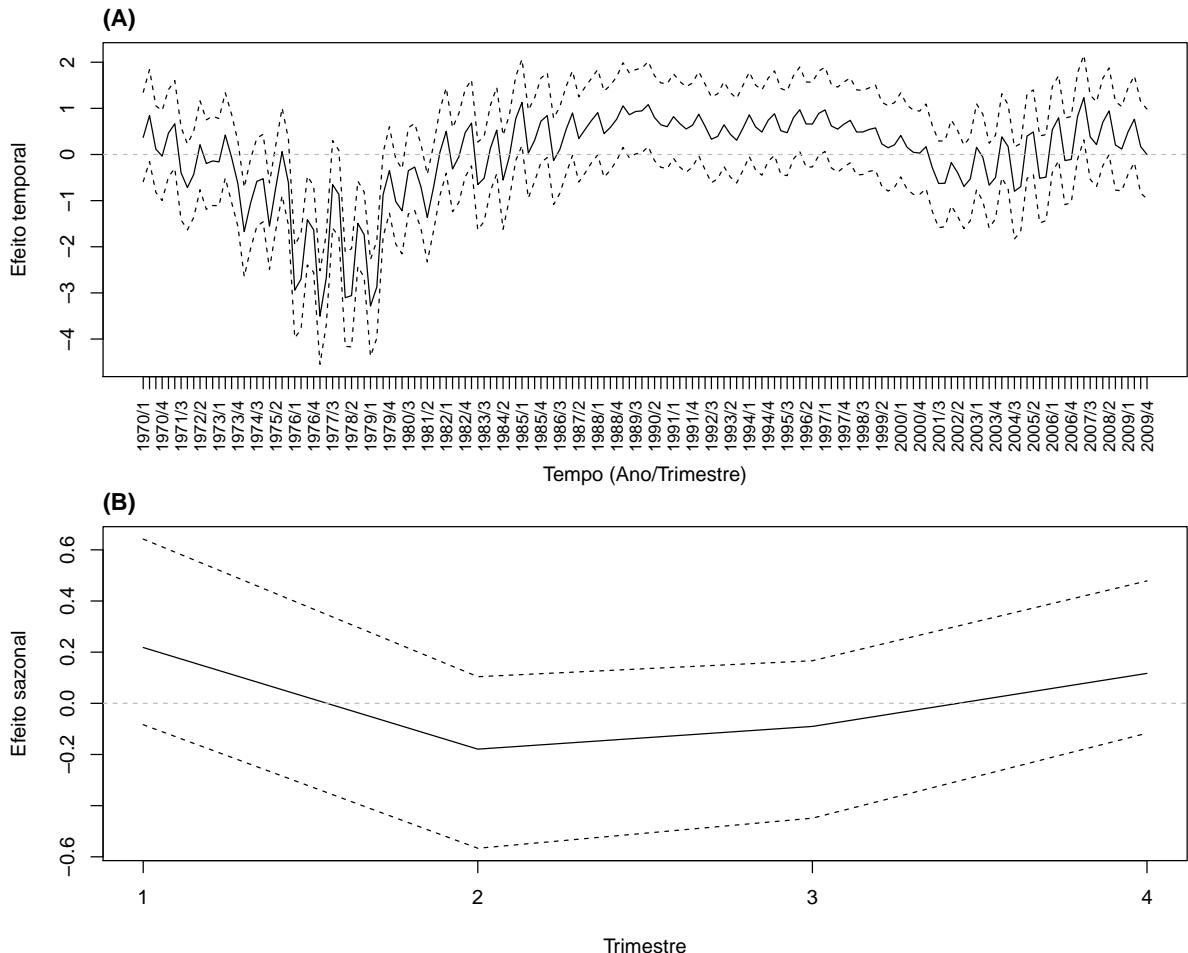


Figura 2.4. Mediana e intervalo de credibilidade de 95% das distribuições posteriores para os efeitos de tendência temporal (A), e sazonalidade (B).

fornecendo boas estimativas com um número reduzido de parâmetros. O fato da distribuição binomial negativa se adequar melhor aos dados analisados provavelmente é devido às características da pesca de agulhão-negro no Atlântico sul. Por se tratar de uma espécie considerada oriunda de uma população pequena em relação às outras espécies exploradas comercialmente, a captura do agulhão-negro é pouco frequente. Isto faz com que em um grande número de anzóis, poucos indivíduos são capturados, o que gera um excesso de zeros, e que pode ser um dos motivos da elevada sobredispersão observada.

Em relação às estruturas temporais analisadas, verificou-se que os modelos que consideram um processo autoregressivo tiveram um melhor desempenho quando comparados com os modelos que consideram um processo de passeio aleatório. Isto sugere que as capturas de agulhões em um determinado tempo t podem ser melhor explicadas pelas capturas nos tempos $t - p$, e não tão bem descritas simultaneamente pelas capturas nos tempos anteriores ($t - 1$) e posteriores ($t + 1$). Ainda, neste caso, observou-se que as capturas são dependentes no tempo até a ordem $p = 10$, visto que o modelo indicado como o melhor entre os possíveis candidatos foi o AR(10). Para fins de gerenciamento pesqueiro, este resultado sugere que qualquer medida de manejo a ser adotada deve levar em consideração a série histórica de capturas do agulhão-negro.

A análise dos efeitos de tendência temporal e sazonal mostrou que o efeito temporal é mais importante para descrever as capturas dos agulhões, como pode ser observado pelo baixo valor da precisão (mediana) para este componente. O efeito sazonal, embora menos pronunciado e pouco significativo, apontou evidências de que as capturas são maiores nos meses mais quentes do ano no Atlântico sul.

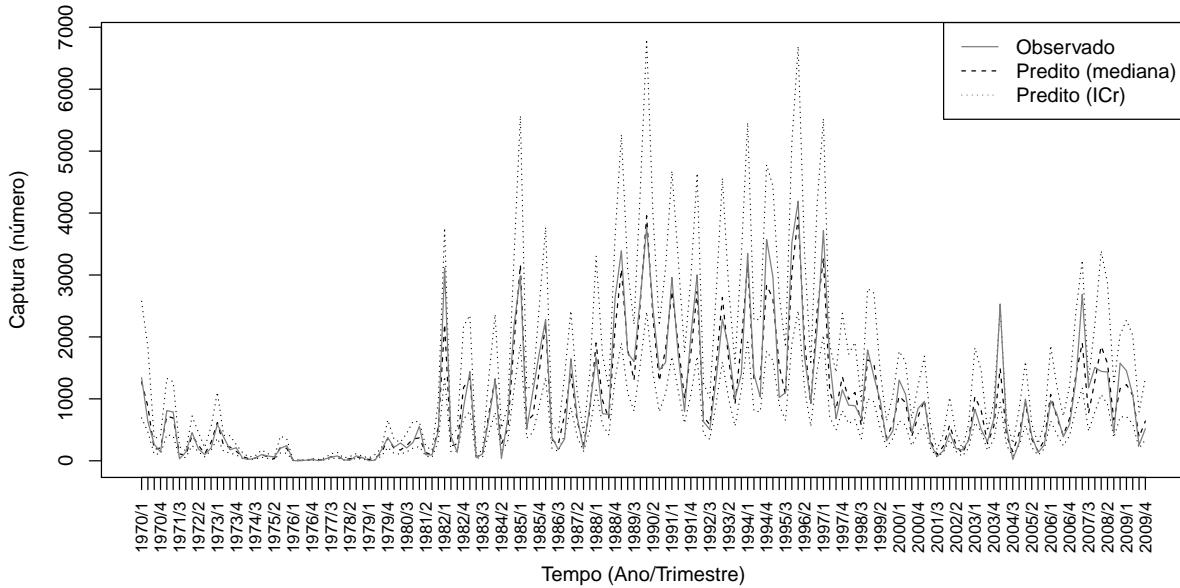


Figura 2.5. Número de agulhões-negros capturados no Atlântico sul pela frota japonesa, no período 1970–2009. A linha tracejada representa a mediana, e as linhas pontilhadas o intervalo de credibilidade (ICr) de 95% para a distribuição posterior dos valores ajustados pelo modelo analisado.

A análise de dados de pescarias comerciais normalmente envolve a padronização da captura como a primeira etapa para análises subsequentes. Normalmente são utilizados modelos lineares generalizados (MLGs) para essa tarefa, onde, no entanto, a dependência temporal raramente é levada em consideração através de estruturas adequadas. A utilização de estruturas de dependência temporal representa uma alternativa mais acurada para a modelagem deste tipo de dados, uma vez que as capturas em pescarias comerciais são dependentes no tempo, já que atuam sobre populações biológicas que são afetadas pela pesca e dependem do tempo para sua recuperação e volta ao equilíbrio natural.

A realização de inferência bayesiana através do método INLA representa uma alternativa viável e computacionalmente mais eficiente do que os tradicionais métodos de Monte Carlo via cadeias de Markov (MCMC). Esta metodologia resolve o problema de tempo computacional elevado na inferência via MCMC, permitindo a comparação, relativamente rápida, de diversas alternativas de modelagem, que testam diferentes hipóteses sobre padrões do processo em estudo. Especialmente em dados de pescarias comerciais, como os analisados neste trabalho, onde há uma grande quantidade de informação, e possivelmente também um grande número de covariáveis, este método pode ser uma opção mais plausível para os cálculos necessários na inferência bayesiana e na análise e seleção de modelos.

Referências

- Andrews, A. H., Humphreys, R. L., e Sampaga, J. D. (2017). Blue marlin (*Makaira nigricans*) longevity estimates confirmed with bomb radiocarbon dating. *Canadian Journal of Fisheries and Aquatic Sciences*, 75(1):17–25.
- Beguin, J., Martino, S., Rue, H., e Cumming, S. G. (2012). Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution*, 3(5):921–929.
- Carlin, B. P. e Louis, T. A. (2009). *Bayesian methods for data analysis*. Chapman & Hall/CRC, Boca Raton.

- Czado, C., Gneiting, T., e Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–61.
- de Sylva, D. P. e Breder, P. R. (1997). Reproduction, gonad histology, and spawning cycles of north atlantic billfishes (Istiophoridae). *Bulletin of Marine Science*, 60(3):668–697.
- Fahrmeir, L. e Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society Series C*, 50(2):201–220.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., e Rubin, D. B. (2014). *Bayesian data analysis*. Chapman and Hall/CRC.
- Goodyear, C. P. (2015). Understanding maximum size in the catch: Atlantic blue marlin as an example. *Transactions of the American Fisheries Society*, 144(2):274–282.
- Held, L., Schrödle, B., e Rue, H. (2010). Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures*, pages 91–110. Physica-Verlag HD, Heidelberg.
- ICCAT (2007). Report of the 2006 ICCAT billfish stock assessment. *Col. Vol Sci. Pap. ICCAT*, 60:1431–1546.
- ICCAT (2012). Report of the 2011 blue marlin stock assessment and white marlin data preparatory meeting. *Col. Vol Sci. Pap. ICCAT*, 68:1273–1386.
- Lindgren, F. e Rue, H. (2008). On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics*, 35(1):691–700.
- Lindgren, F. e Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25.
- Natário, I. e Knorr-Held, L. (2003). Non-parametric ecological regression and spatial variation. *Biometrical Journal*, 45(6):670–688.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rue, H. e Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman & Hall/CRC, Boca Raton.
- Rue, H., Martino, S., e Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- Schrödle, B. e Held, L. (2011). Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6):725–734.
- Schrödle, B., Held, L., Riebler, A., e Danuser, J. (2011). Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(2):261–279.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., e Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Sund, P. N., Blackburn, M., e Williams, F. (1981). Tunas and their environment in the pacific ocean: a review. *Mar. Biol. Ann. Rev.*, 19:443–512.

3 MODELOS DINÂMICOS DE BIOMASSA COM ESTRUTURAS DE DEPENDÊNCIA ESPAÇO-TEMPORAL

Resumo

Os modelos dinâmicos de biomassa tradicionalmente são utilizados como ferramenta para o manejo pesqueiro. Normalmente, a estimação dos parâmetros destes modelos é realizada através de métodos de otimização não linear ou de amostragem. Os dados necessários são apenas a captura e o esforço pesqueiro anual. Com isso, toda a informação sobre a estrutura espacial do estoque explorado é perdida. O objetivo deste trabalho foi desenvolver um método de estimação alternativo, que permitisse incluir a estrutura de dependência espacial e temporal, inerente aos dados pesqueiros. Com isso, foi desenvolvido um método de estimação em dois estágios, que permite modelar os dados no espaço e no tempo, além de ser computacionalmente eficiente. No primeiro estágio é realizada a modelagem da captura, já considerando diversas alternativas de dependência espaço-temporal. Os dados de captura preditos pelos modelos no primeiro estágio são então utilizados como variáveis nos modelos do segundo estágio. Neste segundo estágio, uma forma linear de um modelo dinâmico de biomassa fornece as estimativas dos parâmetros envolvidos, levando também em consideração estruturas de dependências espaço-temporais. Dados simulados de um modelo dinâmico de biomassa com estrutura de dependência espaço-temporal foram utilizados para avaliação do método proposto. Os resultados indicaram que o método de estimação em dois estágios fornece estimativas não viesadas e consistentes, desde que sejam de fato consideradas a dependência espaço-temporal dos dados.

Palavras-chave: Modelos dinâmicos; Modelos espaço-temporais; Inferência bayesiana.

3.1 Introdução

Tradicionalmente, um dos maiores objetivos de cientistas pesqueiros, é encontrar o chamado “Rendimento Máximo Sustentável” (RMS), ou seja, o valor máximo de captura que pode ser retirada de uma população, com a intenção de manter estável o balanço entre o aumento populacional e a remoção pela pesca, evitando assim o colapso que poderia ser causado pela captura excessiva. Em teoria, esse balanço é atingido quando a parcela da população removida pela pesca é aquela tida como “excedente” nessa equação de equilíbrio, ou seja, haveria sempre uma parte da população disponível, capaz de “recompor” aquela parcela removida (Jennings et al., 2001).

Existem diversos modelos na literatura pesqueira para se tentar encontrar de alguma forma o RMS de uma pescaria. A grande maioria destas abordagens considera modelos de dois estágios: um para descrever o processo da dinâmica populacional e outro para fazer a ligação entre os dados observados de uma pescaria (*e.g.* captura, esforço) ao modelo do processo. Normalmente são consideradas as variações anuais na abundância de uma população, e por isso, os dados necessários para o ajuste destes modelos são relativamente simples de serem obtidos, pois são os totais de captura e esforço de pesca ao longo de uma série temporal anual. Modelos como os descritos aqui, são conhecidos como “modelos dinâmicos de biomassa” (ou modelos DB, para simplificação) (Gulland, 1983; Hilborn e Walters, 1992; Quinn e Deriso, 1999).

Inicialmente, o ajuste destes modelos considerava situações de equilíbrio, ou seja, a suposição de que toda a biomassa que está sendo retirada pela pesca está sendo reposta pelo aumento populacional (Gulland, 1983). Claramente esta não é uma suposição razoável, e logo se desenvolveram métodos para situações que não assumiam equilíbrio. Estes métodos, denominados de “erros nas variáveis”, permitiam considerar erros associados à dinâmica do processo, ou associados ao modelo observational, ou em ambos os modelos, mas com a necessidade de uma restrição de que a razão entre as variâncias deveria ser

conhecida (Ludwig et al., 1988; Schnute, 1994). Por muitos anos, esses foram os procedimentos padrões para se ajustar os modelos DB, tanto no contexto clássico como no bayesiano.

Dada a crescente preocupação em se acomodar as incertezas nos modelos do processo e observacional, os modelos DB começaram a ser considerados dentro do contexto de modelos de espaço de estados. Estes modelos estão entre as ferramentas mais poderosas para modelos dinâmicos e de previsão (Fahrmeir e Tutz, 1994). Esta abordagem tem se tornado popular no contexto de manejo de estoques pesqueiros, pois ela permite modelar realisticamente os erros associados tanto ao modelo do processo (de dinâmica da população) quanto ao modelo observacional. O filtro de Kalman (Kalman, 1960), um caso particular de modelos de espaço de estado, foi inicialmente utilizado devido a sua estrutura relativamente simples. No entanto, as suposições para este modelo de funções lineares para o processo, e erros com distribuição normal, provaram ser não muito realistas no contexto de dinâmica de biomassa. Isso havia limitado o número de modelos que poderiam ser ajustados dentro do paradigma clássico/frequentista. Nesse sentido, a abordagem bayesiana dentro do contexto de modelos de espaço de estados fez que as suposições do filtro de Kalman não fossem mais necessárias, permitindo modelar funções não lineares para o processo e considerar erros não normais de uma forma natural.

Embora estes modelos sejam muito utilizados na prática, e a estrutura dos dados necessária seja muito simples, uma desvantagem é que toda a informação relacionada à dinâmica populacional intra-anual é ignorada. Por exemplo, a dependência espacial e temporal entre as capturas de diversas áreas e em épocas do ano diferentes são descartadas ao se considerar apenas o total anual.

Walters (2003), em um artigo convenientemente intitulado *Folly and fantasy in the analysis of spatial catch rate data*, cita dois erros comumente cometidos na análise de dados de captura. O primeiro deles é o uso incorreto de estimadores do tipo razão, como é o caso da Captura Por Unidade de Esforço (CPUE). O autor comenta que somar as capturas no tempo e espaço, e dividir pelo esforço total só seria funcional em áreas e intervalos de tempo pequenos, e se a amostragem fosse aleatória. Como na maioria dos casos, captura e esforço são somas anuais ao longo de estratos espaciais, não são dados os devidos pesos para áreas com maior esforço. O segundo erro, e possivelmente mais grave, segundo o autor, é ignorar as áreas ocupadas pelo estoque e que não foram exploradas pela pescaria. As alternativas seriam então: (i) ignorar estas áreas completamente; (ii) usar a média de CPUE de áreas vizinhas; e (iii) estimar valores com base em modelos adequados (*e.g.* interpolação).

Os modelos DB tradicionais apresentados no contexto de modelos de espaço de estados já representaram um avanço quando comparados aos tradicionais modelos com erros nas variáveis utilizados anteriormente, no sentido de incorporar os erros simultaneamente nas equações do processo e observacional. No entanto, a utilização de um índice de biomassa agregado por ano, faz com que toda a complexidade que envolve tanto a dinâmica da população em cada ano, quanto a incerteza associada às observações coletadas, seja ignorada. Embora esta abordagem tenha sido utilizada com relativo sucesso nas últimas décadas, não há, até então, qualquer estudo que tenha investigado o impacto dessa simplificação nas estimativas de biomassa populacional. Segundo a lógica de pensamento descrita por Walters (2003), a utilização dos dados agregados por ano em modelos dinâmicos de biomassa está fadada aos dois erros comuns descritos pelo autor: não considerar os devidos pesos para áreas com maior esforço, e ignorar as áreas não exploradas pela pescaria em um determinado ano.

Ignorar as complexidades espaciais de um estoque pesqueiro e dos dados coletados pode ser desastroso para um gerenciamento pesqueiro sustentável. Por esse motivo, considera-se que deve ser feita pelo menos uma avaliação comparativa entre os modelos DB de espaço de estados utilizados atualmente (com dados agregados por ano), com modelos DB de espaço de estados que considerem a estrutura de dependência espacial da dinâmica populacional e dos dados observados. Acredita-se que a consideração dessa estrutura deve levar à uma descrição mais realista do processo de dinâmica de biomassa, levando assim à estimativas de parâmetros populacionais mais confiáveis para subsidiar o manejo pesqueiro.

Recentemente, modelos de manejo pesqueiro com estrutura espacial explícita tem sido desenvolvidos, em uma tentativa de se considerar a escala espacial de populações naturais. Uma destas abordagens foi realizada por Carruthers et al. (2011), que derivaram um modelo DB espacial para modelar a dependência espacial entre estoques de uma mesma população. Este tipo de modelo exige dados de diversas pescarias e também de marcação e recaptura de indivíduos. Booth (2000), Goethel et al. (2011), e Jiao et al. (2016), desenvolveram modelos de manejo pesqueiro (modelos de estoque-recrutamento e de captura por idade) tentando incorporar de alguma forma a estrutura espacial das populações. Em todos os trabalhos mencionados, o requerimento de dados vai muito além da captura e esforço por área e tempo. Alguns destes modelos exigem dados de marcação de indivíduos (como já mencionado), ou de captura por comprimento/idade, ou ambos. Portanto, a viabilidade de aplicação destes modelos é restrita a poucas pescarias de onde se podem conseguir estas informações, já que a logística para se obter esse tipo de informação é complexa e o custo é bem maior.

Nesse sentido, o desenvolvimento de modelos DB que possam incorporar a estrutura espacial poderia trazer modelos mais realistas, e possivelmente sem custo adicional, uma vez que dados de captura e esforço de pesca por área e período são informações mínimas obtidas por qualquer programa de coleta de dados pesqueiros.

Dado o grau de complexidade para o ajuste de modelos dinâmicos de biomassa em sua forma tradicional não linear, a inclusão de efeitos espaciais parece ser ainda mais desafiadora. Os métodos de estimação destes modelos atualmente envolvem simulação estocástica ou algoritmos complexos de otimização de funções não lineares. Em ambos os casos, o tempo computacional para o ajuste dos modelos pode ser elevado e de difícil convergência. Por esse motivo, existem poucos estudos que tentam incorporar estruturas mais complexas, como a dependência temporal. Nesse sentido, alguns trabalhos recentes estão indo na direção contrária, tentando encontrar formas alternativas de estimação destes modelos que sejam mais simples. Uma forma de se obter maior simplicidade é a linearização das equações dinâmicas ou a estimação em dois estágios. Por exemplo, Zhang e Smith (2011) desenvolveram um método de estimação em dois estágios. No primeiro estágio a captura é modelada e obtém-se um índice anual, que então é utilizado em um modelo de regressão linear simples, mas proveniente de uma transformação do modelo dinâmico, para a obtenção dos parâmetros populacionais. Rankin e Lemos (2015) fazem uma linearização direta das equações dinâmicas, que são então utilizadas em um processo de inferência bayesiano para obtenção dos parâmetros. Com isso, já existe um bom ganho na questão de tempo computacional. Ambos trabalhos concluem que estas formas mais simples de estimação podem gerar resultados similares aos métodos mais complexos de modelagem não linear. Além disso, a forma linear ou a estimação em dois estágios abre caminho para que estruturas de dependência espaço-temporais possam ser então consideradas.

Este trabalho tem como finalidade, portanto, explorar uma forma alternativa de estimação em dois estágios dos parâmetros dos modelos dinâmicos de biomassa. No primeiro estágio a captura é modelada já se considerando diversas estruturas de dependência espaço-temporal. Os valores preditos nesse primeiro estágio são então utilizados como variáveis no modelo do segundo estágio. Neste segundo estágio, uma versão dos modelos dinâmicos de biomassa foi linearizado, permitindo que os parâmetros envolvidos na dinâmica do processo sejam estimados de maneira rápida e eficiente através de métodos de regressão linear. No entanto, neste segundo estágio também foram consideradas diversas estruturas de efeitos aleatórios espaciais, temporais e espaço-temporais, permitindo assim a obtenção de estimativas dos parâmetros considerando que também sejam dinâmicos (*i.e.* podem variar no tempo e espaço). Com isso, foram utilizadas as técnicas para modelos lineares mistos, considerando efeitos aleatórios espaço-temporais. A inferência foi realizada no contexto bayesiano, através de técnicas de aproximação, que são mais eficientes computacionalmente do que as técnicas tradicionais de inferência bayesiana por meio de simulação. O objetivo é, portanto, desenvolver um método para estimação dos parâmetros dos modelos

dinâmicos de biomassa, que seja eficiente em termos computacionais, mas que leve em consideração a dependência espacial dos dados.

Inicialmente será apresentada uma fundamentação teórica sobre os modelos populacionais em geral, e mais especificamente sobre o uso destes modelos em populações exploradas pela pesca. Os principais métodos de estimação disponíveis também serão apresentados. Na sequência será apresentado então o modelo proposto e a forma de estimativa em dois estágios. Também será descrita a simulação de dados gerados por um modelo dinâmico de biomassa com estrutura de dependência espaço-temporal. Por fim, os resultados do método desenvolvido aplicados aos dados simulados são apresentados, com uma breve discussão e finalmente a conclusão.

3.2 Fundamentação teórica

3.2.1 Dinâmica populacional de recursos pesqueiros

A revisão desta seção foi baseada principalmente em Jennings et al. (2001).

A preocupação com a flutuação anual do rendimento de pescarias teve início efetivamente no final dos anos 1850, quando o governo Norueguês contratou cientistas para descobrir o motivo das oscilações anuais na captura do bacalhau (*Gadus morhua*) (Angelini e Moloney, 2007). Como a economia daquele país era fortemente baseada na exportação do bacalhau salgado, se a produção falhasse haveria um grande prejuízo econômico.

Nos anos seguintes, cientistas pesqueiros começaram a atuar também em alguns países da Europa, devido à crescente preocupação na variação dos estoques. Houve um grande consenso de que pesquisas eram necessárias para se tentar quantificar os efeitos do impacto da pesca nas populações naturais.

No final do século XIX, cientistas já haviam desenvolvido técnicas para determinar a idade de peixes e métodos de marcação para identificar padrões de migração, mas ainda não conseguiam explicar porque as capturas variavam. Apenas no começo do século XX, através dos trabalhos de Johan Hjort, foi evidenciado que a abundância de uma classe etária de determinada espécie era determinada nos primeiros meses de vida, e que a renovação dos estoques pesqueiros não se dava por uma produção anual constante de juvenis, mas sim por uma produção anual altamente irregular, com poucos peixes sobrevivendo a maioria dos anos, e muitos peixes sobrevivendo poucos anos. Portanto, as mudanças na taxa de sobrevivência é que seriam responsáveis pelas flutuações na captura.

As embarcações pesqueiras aumentaram rapidamente em número e eficiência nos anos anteriores à Primeira Guerra Mundial, e os peixes capturados estavam se tornando cada vez menores e menos abundantes. Durante o período da Guerra, o esforço pesqueiro diminuiu consideravelmente devido às atividades militares, e quando a atividade pesqueira retornou no final da Guerra em 1918, a captura voltou a ser composta por peixes maiores e com taxas mais elevadas (Angelini e Moloney, 2007). Assim, ficou claro que a redução do esforço pesqueiro tinha permitido a recuperação dos estoques explorados, e a ciência pesqueira passou a focar no entendimento destes efeitos (Gulland, 1983).

Apenas no final dos anos 1950, cientistas começaram a desenvolver modelos para descrever a dinâmica de populações pesqueiras e as respostas à pesca. Nos anos subsequentes, a ciência pesqueira desempenhou um papel fundamental no campo da Ecologia, ao descrever porque as populações flutuam, e como a estrutura das populações variam no espaço, tempo, e em resposta ao aumento das taxas de mortalidade imposta por humanos.

No final do século XX, houve uma diversificação na ciência pesqueira, uma vez que houve um aumento na preocupação sobre o colapso de estoques pesqueiros, a incerteza que é inerente ao manejo pesqueiro, e o reconhecimento de que os objetivos de manejo não deveriam ser baseados somente nos estoques pesqueiros. Atualmente, o manejo de uma pescaria também envolve a atuação de sociólogos e

economistas, com a intenção de levar em consideração o impacto nas vidas e no rendimento das pessoas envolvidas (Hilborn e Walters, 1992). Também há uma grande preocupação dos efeitos da pesca no ambiente marinho e em outras espécies.

Há uma grande variação na distribuição geográfica de populações pesqueiras. Alguns grandes peixes pelágicos possuem distribuição circunglobal ao longo de áreas tropicais e temperadas, enquanto que alguns peixes recifais são endêmicos em determinadas ilhas ou ecossistemas marinhos específicos (*e.g.* Grande Barreira de Coral). A adequação de um ambiente para uma população é determinada, provavelmente, por fatores denso-independentes, bem como por fatores fisiológicos.

Ao longo da distribuição latitudinal de uma espécie, indivíduos raramente acasalam ao acaso com membros da mesma população. Em vez disso, eles formam populações que são conhecidas como estoques. Um estoque é delimitado por indivíduos que possuem as mesmas taxas de natalidade e mortalidade. Por isso, para uma mesma população pode haver mais de um estoque, e cada um responde de maneira independente aos efeitos da pesca. O conceito de estoque é fundamental na ciência pesqueira já que cada estoque é considerado como uma unidade populacional sobre a qual será exercido o manejo (Gulland, 1983).

Algumas espécies exploradas pela pesca permanecem em um único local durante a vida adulta. No entanto, a grande maioria das espécies realiza movimentos em diferentes escalas ao longo de toda vida. Estes movimentos são extremamente importantes no contexto pesqueiro: eles podem levar indivíduos para dentro ou fora de áreas protegidas, cruzar fronteiras marinhas internacionais, ou fazer com que fiquem particularmente vulneráveis à pesca quando se reúnem para desovar, por exemplo. Um exemplo de migração é o movimento repetitivo entre áreas de alimentação e desova. O que diferencia a migração de movimentos esporádicos (*e.g.* para defesa ou exploração) é a persistência dos movimentos e a grande duração. A distribuição do esforço pesqueiro apresenta grandes variações sazonais como resultado das migrações das espécies exploradas.

Muitas migrações dependem da assistência de correntes, e espécies migratórias podem usar uma variedade de dispositivos químicos, físicos e visuais. Uma das primeiras tentativas de se estabelecer uma regra geral para este processo foi proposta por Harden-Jones (1968), na forma de um “triângulo” (Figura 3.1). Nele estariam estabelecidas áreas de desova, áreas de criação e áreas de distribuição do estoque adulto ou áreas de alimentação. Indivíduos em maturação se deslocam “contra a corrente” (contranatante) para as áreas de desova, e retornam “a favor da corrente” (denatante) às áreas de alimentação. Ovos e larvas são transportados passivamente pelas correntes (denatante) para as áreas de criação. Os recrutas se deslocam às áreas de alimentação onde se incorporam ao estoque adulto.

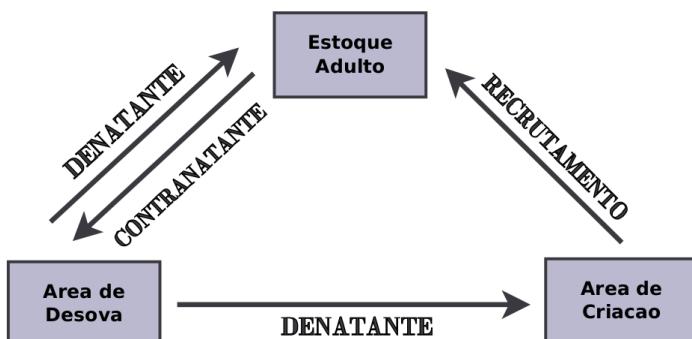


Figura 3.1. O triângulo de Harden-Jones é um modelo geral para explicar a dinâmica de movimentação de populações de peixes marinhos.

As flutuações na abundância das espécies marinhas no tempo e no espaço são afetadas principalmente por: (i) processos físicos e biológicos que afetam a produção e sobrevivência de ovos e larvas;

(ii) crescimento e mortalidade durante as fases juvenis e adultas; e (iii) processos comportamentais como migração e denso-dependência.

Variações na abundância de populações naturais já existiam muito antes da pesca atuar sobre alguns recursos. Por exemplo, através de amostras de escamas de sardinha (*Sardinops sagax*) e anchova (*Engraulis mordax*) do Pacífico, retiradas de sedimentos anaeróbicos na costa da Califórnia mostraram que a abundância flutuou drasticamente nos últimos 2000 anos (Soutar e Isaacs, 1969). Portanto, flutuações na abundância são imprevisíveis mesmo na ausência de pescarias, o que gera grandes incertezas associadas ao manejo pesqueiro (Hilborn e Walters, 1992). Dessa forma, o reconhecimento do comportamento dinâmico de um estoque pesqueiro é fundamental para se obter cenários realistas de administração pesqueira.

Quando as técnicas para se determinar a idade de peixes foram desenvolvidas, cientistas notaram que algumas poucas classes de idades eram responsáveis pela maioria da biomassa capturada. Os indivíduos das classes abundantes, geralmente, dominariam as capturas por muitos anos, antes de uma nova classe abundante recrutar para a pescaria. O recrutamento é definido como o número de indivíduos que atingem um determinado estágio específico do ciclo de vida (e.g. metamorfose, assentamento larval, ou disponibilidade para uma pescaria). O recrutamento depende de muitos fatores, como a abundância e a distribuição da população adulta madura, o número de ovos e larvas produzidos, e a consequente sobrevivência. A maioria das espécies gera muitos ovos pequenos, e mais de 99% da mortalidade ocorre entre as fases larval e juvenil. Esse é um importante período no ciclo de vida de uma população, pois pequenas variações nas taxas de mortalidade podem ter um impacto acentuado na abundância subsequente. O fornecimento variável de recrutas determina a estrutura da população, a força das interações em comunidades multi-específicas, e a biomassa de indivíduos que podem ser capturados pela pesca.

Biologicamente, portanto, a abundância (N , em número de indivíduos) de uma população no tempo $t + 1$ é influenciada pela abundância no tempo t , adicionada do balanço entre nascimento (B) e mortalidade (D), imigração (I) e emigração (E), definindo o modelo mais simples de aumento populacional. No entanto, em populações exploradas pela pesca, a remoção de indivíduos pela captura (Y) ainda deve ser considerada, levando ao seguinte modelo geral

$$N_{t+1} = N_t + (B - D) + (I - E) - Y. \quad (3.1)$$

Como a pesca normalmente atua sob um determinado estoque, então o número de indivíduos que migram entre estoques pode ser considerado desprezível. Com isso, a diferença entre imigração (I) e emigração (E) na equação (3.1) pode ser considerada nula, e este modelo torna-se ainda mais simples ao considerar apenas o balanço entre natalidade e mortalidade, além da captura na presença de pescaria. No entanto, não é uma tarefa fácil conhecer as diferentes taxas de nascimentos e mortalidades de uma população. É ainda mais desafiador predizer como as mudanças nesses parâmetros podem afetar a população no futuro, pois isso dependerá das inúmeras interações dentre e entre espécies, além das variações ambientais possíveis.

3.2.2 Modelos dinâmicos de biomassa

Os modelos dinâmicos de biomassa (modelos DB) são utilizados para se encontrar a maior taxa de mortalidade que pode ser compensada pelo aumento populacional, normalmente medida em mudanças na biomassa populacional por unidade de tempo. Estes modelos podem ser bons pontos de partida pois eles se utilizam da lógica por trás do aumento populacional denso-dependente (Quinn e Deriso, 1999). Os modelos DB usam dados que foram agregados ao longo de diversas classes etárias. Estes modelos aparecem na literatura sob uma série de sinônimos, como “modelos de produção”, “modelos de excedente de produção”, ou “modelos de excedente de rendimento”.

Em termos gerais, modelos DB não necessitam de detalhamentos sobre as características biológicas da população estudada (Gulland, 1983). Sua aplicação visa determinar um nível ótimo de esforço

capaz de produzir um rendimento máximo dentro de um cenário de sustentabilidade da pescaria e do estoque explorado (Gulland, 1983; Hilborn e Walters, 1992). Estes modelos são relativamente simples porque eles reúnem os efeitos gerais de recrutamento, crescimento e mortalidade (todos os aspectos de produção) em uma única função de produção. O estoque é considerado como uma grande biomassa, ou seja, as estruturas de comprimento, idade, e diferenças sexuais e outros aspectos são completamente ignorados.

Para estimar os parâmetros destes modelos, é necessário apenas uma série temporal de um índice relativo de abundância (geralmente a CPUE), e a captura correspondente. Uma alternativa comum seria utilizar modelos mais complexos considerando a estrutura de idade do estoque (*e.g.* análise de populações virtuais) (Quinn e Deriso, 1999). Outra alternativa, utilizada quando é difícil ou impossível determinar a idade, são os modelos baseados no comprimento (Sullivan, 1992). No entanto, Ludwig e Walters (1985) e Ludwig e Walters (1989) mostraram que os modelos DB podem prover estimativas tão boas quanto as dos modelos mais complexos (com estrutura etária ou de comprimento), e muitas vezes até melhores, e com informações estruturalmente mais simples. Por outro lado, Hilborn (1979), em um estudo comparativo com diferentes séries de dados em modelos DB, observou que frequentemente os dados eram muito homogêneos, havia uma falta de contraste, e portanto eram pouco informativos sobre a dinâmica da população considerada. A falta de contraste nos dados significa que as informações sobre captura e esforço estão disponíveis apenas para uma fração limitada do estoque. No entanto, esta falta de contraste também pode levar a resultados incorretos em modelos baseados em idade ou comprimento.

3.2.3 Definição dos modelos

Suponha que o tamanho de uma população se desenvolve seguindo um modelo logístico tradicional, ou seja, considerando tempo contínuo, a taxa de aumento da população em termos de abundância (N) seria

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right) = rN - \frac{r}{K}N^2, \quad (3.2)$$

em que r é a taxa de crescimento instantâneo (per capita) da população (também chamado de “taxa intrínseca de crescimento” ou “parâmetro Malthusiano”), e K é a capacidade de suporte do ambiente. O parâmetro K representa o tamanho máximo suportável de uma população, determinado por uma variedade de recursos limitantes como disponibilidade de alimento, limitações de espaço, e competições inter e intraespecíficas. Portanto, K é um parâmetro estritamente relacionado com a densidade-dependência de uma população. A equação (3.2) foi inicialmente introduzida na Ecologia por PF Verhulst em 1837, e é a mais simples entre as equações que descrevem o aumento populacional limitado, servindo de base para outros modelos usados em Ecologia.

É possível notar que para um tamanho populacional próximo de zero, a taxa de aumento populacional instantânea, $r(1 - N/K)$, é aproximadamente igual a r , e diminui para zero quando $N = K$. A expressão (3.2) é uma equação diferencial de primeira-ordem, quadrática e homogênea, e pode ser resolvida pelo método de frações parciais. Dessa forma, se for escrita como

$$\int_{N_0}^{N(t)} \frac{dN}{N} + \int_{N_0}^{N(t)} \frac{dN}{K - N} = \int_0^t r dt,$$

sua solução será

$$N(t) = \frac{N_0 e^{rt}}{1 - \frac{N_0}{K} + \frac{N_0}{K} e^{rt}}, \quad (3.3)$$

que é chamada de “lei logística de aumento populacional”. O numerador na expressão acima é a própria lei de aumento populacional exponencial (que descreve o aumento sem a presença de fatores limitantes), e o denominador atua como um termo de regulação desse aumento. Note que quando $t = 0$, $N(0) \rightarrow N_0$, e

quando $t \rightarrow \infty$, $N(t) \rightarrow K$. Dessa forma, o comportamento dessa curva logística é sigmoidal para valores de $N < K$, constante para $N = K$, e exponencialmente decrescente para $N > K$. Isso mostra que, independente do tamanho N , a população sempre tende a atingir sua capacidade de suporte. Quando a população é pequena, ela tende a aumentar rapidamente. A população atinge a taxa máxima de aumento quando $N = K/2$, e posteriormente o aumento desacelera à medida que a população se aproxima de K , devido aos processos denso-dependentes (Quinn e Deriso, 1999).

Na equação (3.2), $\frac{dN}{dt}$ representa a taxa instantânea de aumento de uma população, pois é a derivada de uma equação contínua no tempo. Se considerarmos o tempo de forma discreta, ou seja, a abundância da população é quantificada em um determinado tempo específico t , então a abundância N é representada por N_t , e a versão discreta da equação (3.2) é

$$N_{t+1} - N_t = rN_t \left(1 - \frac{N_t}{K}\right) = rN_t - \frac{r}{K}N_t^2. \quad (3.4)$$

Dessa forma, $N_{t+1} - N_t$ representa a taxa de aumento discreto de uma população. A equação que define a trajetória de aumento da população no caso discreto, análoga a equação (3.3) no caso contínuo, é então simplificada para

$$N_{t+1} = N_t + rN_t \left(1 - \frac{N_t}{K}\right). \quad (3.5)$$

A ideia por trás dos modelos DB é então a de se “aproveitar” dessa taxa máxima de aumento populacional que ocorre quando $N = K/2$. Ou seja, o número de indivíduos quando a população está na metade da sua capacidade de suporte seria então o Rendimento Máximo Sustentável (RMS) da pescaria, pois dessa forma, cerca de metade dos indivíduos estaria disponível para recompor a população, com taxas de crescimento elevadas. Claramente essa é uma suposição teórica muito forte, e na prática, sempre deve ser considerado um rendimento abaixo do RMS para evitar uma sobre-exploração do recurso, e ainda deve-se considerar as flutuações naturais da abundância de uma população. Além disso, mesmo se uma pescaria respeitasse o RMS, ainda assim ele poderia variar de ano para ano, devido, por exemplo, às condições de produtividade dos oceanos, eventos climáticos, e mudanças nas taxas de sobrevivência de juvenis e recrutamento.

A maioria dos modelos de dinâmica populacional de recursos pesqueiros é focado na captura. Dada a produtividade latente de uma população, é de interesse determinar qual o excedente disponível para captura, após a recomposição da população através do recrutamento e crescimento, e diminuição após mortalidade natural. Por esse motivo, ao se lidar com a dinâmica das capturas, é importante usar o conceito de estoque unitário (Gulland, 1983), que é uma parcela da população que pode ser gerenciada como uma unidade discreta (*i.e.* é independente de outros estoques).

Em pescarias comerciais, o tamanho do estoque e quantidade de captura são geralmente tratadas com unidades de peso. Portanto, o desenvolvimento a seguir considera a biomassa do estoque como o peso de todos os indivíduos no estoque, e o rendimento como o peso de todos os indivíduos capturados em um determinado momento. Assim, seja $B = B(t)$ a biomassa do estoque no tempo t , e $C = C(t)$, o rendimento acumulado pela captura até o tempo t . Estas quantidades são análogas à abundância, N , e à captura Y , apresentadas previamente, através da transformação usando o peso médio dos indivíduos, \bar{W} , e portanto são representadas por

$$B = N\bar{W} \quad \text{e} \quad C = Y\bar{W}.$$

Fazendo uma analogia com o modelo de aumento populacional geral apresentado em (3.1), e considerando um estoque fechado ($I - E = 0$), obtemos o seguinte modelo para a dinâmica de biomassa de uma população pesqueira

$$B_{t+1} = B_t + (B - D) - C.$$

Vamos considerar também que a população terá aumento em biomassa dado pelo aumento de peso individual (G), e pelo peso incorporado com novos recrutas (R), sendo assim, $B = R + G$. Também vamos considerar que a redução da biomassa da população será consequência da mortalidade natural (M), que pode ser causada por diversos fatores, e assim, $D = M$. Além disso, vamos assumir que todas estas quantidades variam ao longo do tempo (discreto) t . Portanto, o modelo geral para descrever a dinâmica de populações pesqueiras no tempo será

$$B_{t+1} = B_t + [(R_t + G_t) - M_t] - C_t. \quad (3.6)$$

Dessa forma, na ausência de pesca, $C_t = 0$, e denominando a produção de biomassa por $P(B_t) = R_t + G_t$, temos

$$B_{t+1} = B_t + P(B_t) - M_t,$$

e, portanto, haverá um excedente de produção quando $P(B_t) > M_t$. O termo “excedente de produção” geralmente se refere à diferença entre a produção e a mortalidade natural. Por isso, vamos denominar por excedente de produção a função $G(B_t)$ composta por

$$\begin{aligned} G(B_t) &= P(B_t) - M_t \\ &= (R_t + G_t) - M_t. \end{aligned}$$

Sendo assim, o chamado modelo dinâmico de biomassa (ou modelo de excedente de produção) possui forma geral dada por

$$B_{t+1} = B_t + G(B_t) - C_t. \quad (3.7)$$

Note que a produção excedente representa a quantidade de biomassa adicionada à população na ausência de pescaria. Portanto, esse excedente seria a quantidade de biomassa que poderia ser removida pela pesca, C_t , mantendo a biomassa da população constante. Nesse caso, se o rendimento da pescaria for exatamente igual ao excedente de produção, $C_t = G(B_t)$, então este valor é chamado de Rendimento Máximo Sustentável (RMS). Como consequência desse rendimento, a biomassa no tempo $t + 1$ será igual a biomassa no tempo t , $B_{t+1} = B_t$, e portanto se mantém em equilíbrio (Hilborn e Walters, 1992; Quinn e Deriso, 1999).

Existem alguns modelos propostos na literatura para descrever o excedente de produção. Schaefer (1954) foi o primeiro a usar a curva logística (similar ao modelo (3.5)) para descrever a dinâmica da biomassa do linguado do Pacífico. O chamado modelo de Schaefer para o excedente de produção é então

$$G(B_t) = rB_t \left(1 - \frac{B_t}{K}\right). \quad (3.8)$$

Uma propriedade deste modelo é que a produção máxima ocorre quando $B = K/2 = B_{RMS}$. Independente do tamanho da população, seria possível remover o excedente de produção e manter o estoque nas mesmas condições anteriores à produção e à pescaria. Uma estratégia óbvia de manejo, seguindo esta lógica, seria então manter o estoque em um tamanho que maximizaria o excedente de produção, e por consequência, o rendimento. Isto significaria que seria necessário então diminuir o tamanho do estoque para que ele se torne mais produtivo (Haddon, 2001). Obviamente, existem uma série de problemas relacionados com esta visão simplificada da produção pesqueira. O principal deles seria ter que assumir a perigosa suposição de que a população está em equilíbrio.

Uma alternativa ao modelo de Schaefer, foi proposta por Fox (1970). O chamado modelo de Fox é denotado por

$$G(B_t) = rB_t \left(1 - \ln \frac{B_t}{K}\right), \quad (3.9)$$

e possui como característica o ponto de inflexão da curva logística (*i.e.* o tamanho da população que produz o aumento máximo) deslocado para a esquerda. De fato, este ponto está situado a menos da

metade da capacidade de suporte da população, e portanto a taxa máxima de aumento e a B_{RMS} são também deslocadas para a esquerda. Embora essa seja uma alternativa mais “segura” ao modelo de Schaefer, o modelo continua sendo inflexível em relação à não permitir variar o ponto de inflexão da curva.

Para contornar esse problema, Pella e Tomlinson (1969) propuseram um modelo mais geral, com um parâmetro adicional, que permite que a biomassa no ponto B_{RMS} seja deslocada para a esquerda ou direita. A parametrização do modelo de Pella-Tomlinson é

$$G(B_t) = rB_t - \frac{r}{K}B_t^m. \quad (3.10)$$

Nesta versão do modelo de excedente de produção, se $m = 2$, o modelo é equivalente ao modelo de Schaefer, e no limite quando $m \rightarrow 1$, o modelo corresponde ao de Fox. Se $m < 2$, o máximo de produção deste modelo será deslocado para a esquerda ($B_{RMS} < K/2$), já se $m > 2$, o máximo estará à direita ($B_{RMS} > K/2$). Portanto, o modelo de Pella-Tomlinson permite uma maior flexibilidade no formato das curvas de produção.

Os três modelos apresentados acima correspondem à descrição do processo da dinâmica da biomassa de um estoque pesqueiro. Note que todos eles são função da biomassa da população (B), e de outros parâmetros desconhecidos (r, K, m). Como dificilmente temos acesso aos valores de biomassa, então é necessário fazer alguma ligação com dados que podem ser observados a partir de informações da pescaria. O chamado modelo observational é o que faz essa ligação entre o modelo do processo e os dados coletados. Tipicamente, o modelo observational é representado por

$$U_t = qB_t, \quad (3.11)$$

onde U_t é um índice relativo de abundância, e q é o coeficiente de capturabilidade. Na prática, o índice utilizado é a Captura Por Unidade de Esforço (CPUE), calculada como o total de captura dividida pelo total do esforço no tempo t . O coeficiente de capturabilidade q é geralmente desconhecido, e pode ser interpretado como a fração da população que é removida pela pesca (Gulland, 1983), e atua como uma constante de proporcionalidade entre o índice relativo e a própria abundância, ou seja, $U_t \propto B_t$ para q constante.

3.2.4 Métodos de estimação

3.2.4.1 Métodos de equilíbrio

Os métodos de equilíbrio são baseados na suposição de que para cada nível de esforço de pesca, existe um rendimento sustentável. Assume-se que o estoque está em um determinado nível de equilíbrio, produzindo uma determinada quantia de produção excedente. Se o esforço de pesca muda, é assumido que a biomassa do estoque move imediatamente para uma biomassa estável diferente, com uma produção excedente associada. Essa suposição é evidentemente errada, pois ignora completamente a diferença entre os níveis de captura e o tempo que a população leva para responder às condições alteradas (Haddon, 2001).

No geral, a suposição é a de que o rendimento da pescaria é sempre proveniente de excedente de produção de uma população em equilíbrio. A partir dessa suposição, então $B_{t+1} = B_t = \text{constante}$ para todos os anos, e a equação (3.11) é exata. Com isso, e considerando o modelo de Schaefer (3.8), por exemplo, o rendimento máximo da pescaria ocorre exatamente quando $B = K/2$. Usando a CPUE como índice de abundância, então o esforço ótimo (E_{RMS}) pode ser deduzido, ou seja, o nível de esforço (e.g. número de anzóis) necessário para se obter o RMS quando $B = K/2$ (Quinn e Deriso, 1999).

Este método de estimativa foi utilizado inicialmente por ser a solução mais óbvia para a estimativa dos parâmetros. No entanto, se um estoque pesqueiro está em declínio, então os métodos de

equilíbrio superestimam consistentemente o RMS (Haddon, 2001). Seu uso no passado claramente contribuiu para um grande número de colapsos de pescarias (Larkin, 1977), e não é mais recomendado atualmente (Hilborn e Walters, 1992).

3.2.4.2 Métodos de regressão linear

Uma forma alternativa aos métodos de equilíbrio, envolve a transformação das equações dinâmicas do modelo em uma equação linear. Esse método não assume que o estoque está em equilíbrio e leva em consideração a dinâmica do processo (Hilborn e Walters, 1992). Além disso, os parâmetros de interesse, (r , q , e K) são estimados simultaneamente por meio dos métodos tradicionais de regressão linear simples (Quinn e Deriso, 1999).

Walters e Hilborn (1976) e Hilborn (1979) foram os primeiros a formular equações recursivas simples para a estimativa dos parâmetros. Considere inicialmente o modelo dinâmico de biomassa (3.7) com a função de produção dada pelo modelo de Schaefer (3.8)

$$B_{t+1} = B_t + rB_t \left(1 - \frac{B_t}{K}\right) - C_t. \quad (3.12)$$

O modelo observacional, ou seja, aquele que relaciona o processo não observado com a parte observável é

$$C_t = qE_t B_t, \quad (3.13)$$

ou seja, a captura C_t é uma função linear do coeficiente de capturabilidade (q), do esforço de pesca (E_t), e da biomassa (B_t). Se definirmos a Captura Por Unidade de Esforço (CPUE, U_t) como sendo a fração de captura realizada com uma unidade de esforço, então chegamos à seguinte relação

$$\begin{aligned} C_t &= qE_t B_t \\ \frac{C_t}{E_t} &\equiv U_t = qB_t. \end{aligned}$$

Dessa simples relação podemos concluir que a CPUE é um índice proporcional à abundância da população, $U_t \propto B_t$, e que, supondo q conhecido, a biomassa pode ser calculada como

$$B_t = \frac{U_t}{q}. \quad (3.14)$$

Ao usar a definição (3.13) e rearranjando os termos da equação (3.12), temos

$$B_{t+1} = B_t + rB_t - \frac{r}{K}B_t^2 - qE_t B_t,$$

e, substituindo (3.14), chegamos à seguinte relação

$$\frac{U_{t+1} - U_t}{U_t} = r - \frac{r}{qK}U_t - qE_t \quad (3.15)$$

(ver desenvolvimento no Apêndice A). Claramente, a equação (3.15) pode ser escrita como um modelo de regressão linear da seguinte forma

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}$$

onde

$$\begin{aligned} Y_t &= \frac{U_{t+1} - U_t}{U_t}, & X_{1t} &= U_t, & X_{2t} &= E_t, \\ \beta_0 &= r, & \beta_1 &= -\frac{r}{qK}, & \beta_2 &= -q. \end{aligned}$$

Note que, se $t = 1, \dots, T$ representa as observações em cada tempo, então, nessa equação, são utilizadas $T - 1$ observações, pois aqui $t = 1, \dots, T - 1$.

Integrando a versão contínua do modelo de Schaefer, do tempo t ao tempo $t + 1$, Schnute (1977) chegou no seguinte resultado

$$\ln\left(\frac{U_{t+1}}{U_t}\right) = r - \frac{r}{qK} \left(\frac{U_{t+1} + U_t}{2}\right) - qE_t \left(\frac{E_{t+1} + E_t}{2}\right),$$

claramente análogo à equação (3.15).

Em qualquer uma das formas, estes modelos na forma de regressão linear são altamente atrativos pela simplicidade. Para a estimativa dos parâmetros populacionais r e K , e do coeficiente de capturabilidade q são necessários apenas informações de captura e esforço por unidade de tempo (geralmente ano). A estimativa pode então ser realizada pelos métodos de mínimos quadrados, e toda teoria decorrente desse método pode ser aplicada (*e.g.* obtenção de intervalos de confiança para as estimativas dos parâmetros).

No entanto, Uhler (1980), em um vasto estudo de simulação comparando os modelos acima e algumas de suas variações, mostrou que, em geral, as estimativas obtidas pelo método de regressão são viesadas. Praticamente ao mesmo tempo, Hilborn (1979) realizou um estudo de simulação similar, e concluiu que, de fato, as estimativas obtidas por esse método tendem a ser viesadas. No entanto, Hilborn (1979) atribui o viés das estimativas à falta de contraste nos dados, e não ao fato de que o método pudesse ser falho ou muito simples.

Uhler (1980) mostra que, quando considerados erros estocásticos nas equações dinâmicas, ou seja,

$$\begin{aligned} B_{t+1} &= B_t + rB_t \left(1 - \frac{B_t}{K}\right) - C_t + u_t \\ C_t &= qE_t B_t + v_t, \end{aligned}$$

com $u_t \sim N(0, \sigma_u^2)$ e $v_t \sim N(0, \sigma_v^2)$, o modelo de regressão a ser considerado é

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \phi_t$$

onde

$$\phi_t = -\frac{2b}{c}w_t - \frac{1+a}{U_t}w_t + \frac{b}{cU_t}w_t^2 + \frac{c}{U_t}u_t + \frac{1}{U_t}w_{t+1}$$

e $a = r$, $b = r/qK$, $c = q$, e $w_t = v_t/E_t$. Como

$$U_t = qB_t + w_t,$$

ao usar U_t como um *proxy* para B_t , temos uma equação com “erros nas variáveis” clássico¹. O problema de erros nas variáveis (ou erro de medida) ocorre quando usamos uma variável explicativa em um modelo de regressão linear, considerando que foi medida exata (sem erros), mas na verdade existe uma incerteza associada às medidas que não estão sendo consideradas (Carroll et al., 2006). Por esse motivo, as estimativas obtidas em modelos de regressão com erros nas variáveis tendem a ser viesadas. Além disso $\phi_t = \phi(w_{t+1}, w_t, u_t, U_t)$, e, portanto, quando $w_t \neq 0$, os erros serão autocorrelacionados. Mesmo quando $w_t = 0$, as estimativas podem ser viesadas devido à estrutura de dependência temporal dos dados (Uhler, 1980).

Por outro lado, Hilborn (1979) argumenta que as falhas em obter estimativas razoáveis para essa classe de modelos é devido à falta de contraste entre os dados de captura e esforço. Isso significa que a série temporal dos dados deve conter valores altos e baixos de esforço, tentando capturar ao máximo as flutuações na abundância através da captura. O mesmo autor ainda contesta o fato de que o modelo de

¹Outra forma é o erro nas variáveis de Berkson.

Schaefer era inapropriado para modelos populacionais complexos, ou ainda de que o método de regressão era muito simples.

Ludwig e Walters (1985) sugerem que modelos simples de produção como estes devem ser preferidos ao invés de modelos mais complexos que consideram a estrutura etária da população, uma vez que os modelos mais simples fazem menos suposições sobre os dados.

Os modelos DB na sua forma linearizada foram importantes na década de 1980, quando a capacidade computacional era limitada para o ajuste de modelos dinâmicos não lineares complexos. Como as estimativas dos parâmetros são obtidas de maneira analítica, através do método de mínimos quadrados, esta era uma forma conveniente de obtenção dos parâmetros de um modelo relativamente complexo. No entanto, devido à incerteza sobre a qualidade das estimativas, e também à crescente capacidade computacional, este método passou a ser de interesse secundário, e os estudos passaram a ser focados nas formas de estimação não linear para estes modelos.

3.2.4.3 Modelos com erros nas variáveis

Desde os anos 1980, o desenvolvimento de modelos DB tem sido focado nos modelos com erros nas variáveis (Punt, 2003). Três classes principais desses modelos surgiram. Conceitualmente, estas três classes são casos particulares do modelo DB genérico:

$$B_{t+1} = (B_t + G(B_t) - C_t)e^{\epsilon_t} \quad (3.16a)$$

$$U_t = qB_t e^{\eta_t}. \quad (3.16b)$$

A diferença destes modelos para aqueles apresentados até então, é que estes modelos podem incorporar termos de erro multiplicativo tanto na equação do processo quanto na equação das observações. Até então, os modelos apresentados não admitiam termos aleatórios, ou seja, eram puramente determinísticos, o que claramente não é uma suposição razoável. Portanto, aqui, ϵ_t é erro do processo para o ano t , onde geralmente assume-se $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, e η_t é o erro devido às observações no ano t , que comumente é assumido $\eta_t \sim N(0, \sigma_\eta^2)$.

As três classes de modelos que foram desenvolvidas a partir desse modelo geral são:

- (i) Modelo com erros nas observações: $\sigma_\epsilon = 0$
- (ii) Modelo com erros no processo: $\sigma_\eta = 0$
- (iii) Modelo com erros totais: também conhecidos como modelos com erros nas variáveis. Nesse caso, assume-se que $\sigma_\epsilon/\sigma_\eta = C$, onde C é uma constante conhecida.

Dessa forma, para qualquer modelo considerado, a verossimilhança para o índice de abundância U_t , dada a biomassa do ano t é

$$L(U_t|q, B_t, \sigma_\eta) = \frac{1}{\sqrt{2\pi}\sigma_\eta U_t} \exp \left[-\frac{1}{2} \left(\frac{\ln(U_t) - \ln(qB_t)}{\sigma_\eta} \right)^2 \right]. \quad (3.17)$$

Os modelos com erro no processo assumem que os dados de captura e esforço foram observados sem erro, e que todo erro é atribuído para a relação funcional entre a taxa de aumento e o tamanho da população (Hilborn e Walters, 1992; Polacheck et al., 1993; Quinn e Deriso, 1999). Inicialmente, os modelos com erros no processo eram mais utilizados, pois para algumas escolhas de $G(B_t)$, o modelo poderia ser linearizado, e a estimação dos parâmetros poderia ser realizada com técnicas tradicionais de regressão múltipla, ao invés de métodos para modelos não lineares (Schnute, 1977). No entanto, estudos de simulação (*e.g.* Polacheck et al., 1993) mostraram que os modelos com erro no processo tendem a prover estimativas menos precisas do que os modelos com erros nas observações apenas.

O modelo com erros nas observações, assume que o modelo para o processo é correto, e que todo erro ocorre na relação entre a medida de abundância relativa e a abundância verdadeira. Assumindo o erro lognormal na equação (3.16b), as estimativas dos parâmetros (r, q, K) são obtidas através da minimização da soma dos logaritmos da equação (3.17), em todos os anos onde o índice de abundância está disponível. A série temporal de biomassa é estimada projetando-se a biomassa no começo de série de capturas (geralmente assumindo-se que $B_1 = K$) para frente, usando a equação (3.16a), sob a suposição de uma dinâmica determinística, *i.e.*, $\{\epsilon_t = 0; t = 1, 2, \dots, n\}$.

O modelo que assume erros tanto nas observações quanto no processo (modelo com erro nas variáveis ou erro total) foi formalizado inicialmente por Ludwig et al. (1988), e formou a base teórica atual para o desenvolvimento destes modelos. As estimativas de (r, q, K) são obtidas através da minimização de uma função negativa do logaritmo da verossimilhança que inclui a contribuição de ambos os erros, que são tratados como parâmetros “nuisance”. Assim, o vetor de parâmetros é $\{r, q, K, [\epsilon_t; t = 1, 2, \dots, n], \sigma_\epsilon, \sigma_\eta\}$, e o logaritmo da função de verossimilhança fica

$$-\ln L \propto n \ln \sigma_\eta + \frac{1}{2\sigma_\eta^2} \sum_t [\ln U_t - \ln(qB_t)]^2 + n \ln \sigma_\epsilon + \frac{1}{2\sigma_\epsilon^2} \sum_t \epsilon_t^2. \quad (3.18)$$

Ainda é necessário especificar a relação entre σ_ϵ e σ_η , de forma que ela seja constante. Essa suposição é arbitrária, e geralmente $\sigma_\epsilon = \sigma_\eta$ (Punt, 2003).

Estudos de simulação mostraram que os modelos com erros nas observações forneciam estimativas viesadas e imprecisas na presença de erros no processo (Punt, 1995), o que levanta a questão se abordagens que consideram os erros tanto no processo quanto nas observações podem de fato fornecer estimativas mais precisas.

3.2.4.4 Modelos de espaço de estados

Além da abordagem de Ludwig et al. (1988) para se considerar a presença de erros nos modelos do processo e observacional, o filtro de Kalman, e de maneira mais geral, modelos de espaço de estado, têm sido também utilizados como uma alternativa.

Estas duas abordagens diferem das anteriores de erros nas variáveis, no sentido que a biomassa no ano $t+1$ é determinada pela distribuição da biomassa no ano t , através da previsão da distribuição da biomassa no ano $t+1$ usando um modelo de dinâmica populacional, e então atualizando esta distribuição baseado na índice de abundância do ano $t+1$ (Punt, 2003). Dada uma distribuição para a biomassa inicial, este processo de previsão e atualização pode ser aplicado recursivamente para se estimar as distribuições de biomassa para cada ano de interesse. Portanto, ao contrário dos modelos com erros nas variáveis, o filtro de Kalman e os modelos de espaço de estados integram ao longo da biomassa populacional desconhecida e/ou dos erros do processo, como é apropriado, uma vez que os erros do processo são efeitos aleatórios em um modelo de séries temporais não linear (Punt, 2003).

A descrição geral de modelos de espaços de estado a seguir é baseada nos trabalhos de Kitagawa (1987), Schnute (1994), e de Valpine e Hilborn (2005), e usa a notação baseada no texto dos últimos autores. Seja o verdadeiro estado da população em um determinado tempo t denotado por um escalar ou vetor \mathbf{X}_t , que pode ser por exemplo o tamanho da população. Seja $\mathbf{X}^t = (\mathbf{X}_1, \dots, \mathbf{X}_t)$, e T o tempo final observado, então a série temporal completa é \mathbf{X}^T . O modelo estocástico para a verdadeira dinâmica populacional é

$$\mathbf{X}_t = F(\mathbf{X}_{t-1}, \nu_{t-1}) \quad (3.19)$$

onde ν_{t-1} é uma variável aleatória associada ao erro do modelo para o processo. A função $F(\cdot)$ pode assumir diversos formatos, mas aqui assumiremos que ela será um modelo dinâmico de biomassa, que de maneira geral é o modelo (3.7).

Considere agora a observação no tempo t como um escalar ou vetor \mathbf{Y}_t , por exemplo, a captura. Seja $\mathbf{Y}^t = (\mathbf{Y}_1, \dots, \mathbf{Y}_t)$. O modelo estocástico para as observações é então

$$\mathbf{Y}_t = G(\mathbf{X}_t, \epsilon_t) \quad (3.20)$$

onde ϵ_t é uma variável aleatória para acomodar o erro observacional. Estes modelos também poderiam incluir variáveis de controle, como descrito em Schnute (1994), mas para simplificação elas são excluídas da notação em um primeiro momento. Os parâmetros a serem estimados das funções $F(\cdot)$ e $G(\cdot)$ serão denotados por Θ .

No contexto de modelos de espaço de estados, os estados e as observações são variáveis aleatórias conjuntamente distribuídas, e a verossimilhança é a distribuição marginal das observações, integrada sob todas as possíveis trajetórias de estado. Formalmente,

$$\begin{aligned} L(\Theta|\mathbf{Y}^T) &= P(\mathbf{Y}^T|\Theta) \\ &= \int P(\mathbf{Y}^T, \mathbf{X}^T|\Theta) d\mathbf{X}^T \\ &= \int P(\mathbf{Y}^T|\mathbf{X}^T, \Theta) P(\mathbf{X}^T|\Theta) d\mathbf{X}^T \end{aligned} \quad (3.21)$$

onde $P(\mathbf{Y}^T, \mathbf{X}^T|\Theta)$ é a densidade de probabilidade dos estados e observações dado os parâmetros, calculada usando os modelos $F(\cdot)$ e $G(\cdot)$. A equação (3.21) corresponde à definição clássica de verossimilhança, na qual a teoria assintótica usual se aplica. A integral de alta dimensão em \mathbf{X}^T não pode ser resolvida analiticamente, exceto em casos mais simples, como no filtro de Kalman, onde $F(\cdot)$ e $G(\cdot)$ são lineares com erros aditivos gaussianos. Pela abordagem de verossimilhança tradicional, o próximo passo seria maximizar $L(\Theta|\mathbf{Y}^T)$ sob Θ , e estimar regiões de confiança, normalmente aproximadas usando a teoria assintótica.

O filtro de Kalman (Kalman, 1960), que foi inicialmente utilizado dentro do contexto de espaços de estado para modelos dinâmicos de biomassa (Schnute, 1994; Sullivan, 1992) para incorporar os erros de processo e observacional, é um caso particular dos modelos (3.19) e (3.20), onde as funções $F(\cdot)$ e $G(\cdot)$ são lineares, e os erros são normalmente distribuídos (Meinhold e Singpurwalla, 1983). Claramente estas suposições limitam a aplicabilidade deste método ao contexto pesqueiro, e alguns autores (*e.g.* Reed e Simons (1996)), de fato, salientam que essa abordagem sacrifica muito do realismo no contexto de dinâmica populacional.

Os modelos com erros nas variáveis, apresentados anteriormente, também podem ser elaborados com uma versão de espaço de estados. O modelo com erros nas variáveis (EV) pode ser elaborado como um modelo de espaço de estados considerando que o estado \mathbf{X}^T é único, ou seja, possui uma única trajetória possível. Dessa forma, a verossimilhança para o modelo EV (L^{EV}) é definida como a probabilidade das observações e de uma trajetória específica, ou seja,

$$\begin{aligned} L^{EV}(\Theta, \mathbf{X}^T|\mathbf{Y}^T) &= P(\mathbf{Y}^T, \mathbf{X}^T|\Theta) \\ &= P(\mathbf{Y}^T|\mathbf{X}^T, \Theta) P(\mathbf{X}^T|\Theta). \end{aligned} \quad (3.22)$$

Note que essa expressão iguala o integrando da verossimilhança da equação (3.21). O ajuste desse modelo seria então realizado através da maximização de $L^{EV}(\Theta, \mathbf{X}^T|\mathbf{Y}^T)$ sob Θ e \mathbf{X}^T . Assim como mencionado anteriormente, é comum se assumir uma restrição conhecida entre as variâncias do processo e das observações, geralmente na forma de uma razão conhecida entre estas duas variâncias.

Os modelos EV possuem história complicada de justificativa e aplicação. Este método surgiu como um aperfeiçoamento sobre os modelos que simplesmente ignoravam os erros no processo ou nas observações (Ludwig e Walters, 1981; Ludwig et al., 1988). Naquele contexto, assumir que a razão entre as duas variâncias deve ser uma constante conhecida era melhor que simplesmente omitir um tipo de

erro. Como demonstrado, o modelo EV maximiza a verossimilhança conjunta dos estados da população e dos parâmetros, sem a integração nos espaços de estado. Há muito tempo já se reconhece que os modelos EV não usam verossimilhanças tradicionais (Schnute, 1994; Millar e Meyer, 2000b), e podem fornecer estimativas viesadas (Polacheck et al., 1993). No entanto, a verossimilhança de modelos EV tem sido tratada algumas vezes como se a teoria assintótica de verossimilhança fosse aplicável, como, por exemplo, estimando intervalos de confiança usando as derivadas de segunda ordem da verossimilhança conjunta para estimar a matriz de informação de Fisher (Ludwig e Walters, 1981), ou usando critérios de informação para a seleção de modelos.

Ao comparar as estimativas de um modelo linear autoregressivo simples feitas com o modelo EV e com a verossimilhança clássica, de Valpine e Hilborn (2005) obtiveram estimativas viesadas para os parâmetros, e intervalos de confiança incorretos para o modelo EV. Em uma extensa revisão do conceito de modelos de espaço de estados, Schnute (1994) comparou os modelos EV com o filtro de Kalman. Este autor ainda colocou os modelos EV sob o paradigma bayesiano, e mesmo assim, levantou sérias questões sobre sua utilidade prática.

Em um contexto bayesiano para a estimativa de Θ nos modelos EV, a distribuição posterior para os estados e os parâmetros conjuntamente seria dada por

$$\pi(\Theta, \mathbf{X}^T | \mathbf{Y}^T) = \frac{P(\mathbf{Y}^T | \mathbf{X}^T, \Theta) P(\mathbf{X}^T | \Theta) P(\Theta)}{P(\mathbf{Y}^T)}. \quad (3.23)$$

Já a posterior para a estimativa de Θ nos modelos de estado de espaço gerais seria dada por

$$\pi(\Theta | \mathbf{Y}^T) = \frac{P(\mathbf{Y}^T | \Theta) P(\Theta)}{P(\mathbf{Y}^T)} \quad (3.24)$$

que é obtida pela integração da expressão (3.23) na dimensão dos estados, ou seja,

$$\pi(\Theta | \mathbf{Y}^T) = \int \pi(\Theta, \mathbf{X}^T | \mathbf{Y}^T) d\mathbf{X}^T. \quad (3.25)$$

Em uma implementação de Monte Carlo, obtém-se uma amostra de (Θ, \mathbf{X}^T) de $\pi(\Theta, \mathbf{X}^T | \mathbf{Y}^T)$. Removendo a dimensão de \mathbf{X}^T , representaria uma aproximação de Monte Carlo para a integral (3.25), então os valores remanescentes de Θ seriam uma amostra de $\pi(\Theta | \mathbf{Y}^T)$.

Dada a prática geral (mas não necessariamente correta) de se aplicar a teoria de verossimilhança clássica aos modelos EV, convém ressaltar que algumas das propriedades mais úteis dos contextos clássicos e bayesianos, se aplicam somente às equações (3.21) e (3.24), respectivamente, e não às equações (3.22) e (3.23). Por exemplo, no limite quando $T \rightarrow \infty$, $\pi(\Theta | \mathbf{Y}^T)$ é assintoticamente normal (com covariância que pode ser estimada usando a matriz de informação de Fisher a partir da matriz Hessiana de derivadas de segunda ordem), e contém mais peso sob o verdadeiro Θ . Já o mesmo não se aplica para $\pi(\Theta, \mathbf{X}^T | \mathbf{Y}^T)$, com respeito aos verdadeiros valores de Θ ou \mathbf{X}^T (de Valpine e Hilborn, 2005).

A abordagem bayesiana e os modelos de espaços de estado tem sido utilizados com bastante frequência na literatura recente (Sullivan, 1992; Schnute, 1994; Kinas, 1996; Millar e Meyer, 2000b; Meyer e Millar, 1999; de Valpine, 2002; de Valpine e Hastings, 2002; de Valpine, 2004; de Valpine e Hilborn, 2005; Sant'Ana et al., 2017), e estas duas metodologias são denominadas por alguns autores (*e.g.* Hilborn, 1992) como o futuro da metodologia para o manejo pesqueiro.

3.3 Metodologia

3.3.1 Modelos dinâmicos de biomassa espaço-temporais

Os modelos dinâmicos de biomassa, como apresentados na seção 3.2.2, consideram naturalmente a evolução temporal do processo de geração de biomassa de uma população afetada pela pesca. Também

como mencionado naquela seção, os dados para o ajuste destes modelos se resumem ao valor total da captura e do esforço por ano para um determinado estoque.

Neste trabalho será considerado uma forma mais geral deste modelo, que considera também a estrutura espacial da pescaria em todos os anos. Uma forma geral de definir o modelo dinâmico de biomassa considerando a função de produção dada pelo modelo de Schaefer (3.8) é

$$B_{(t+1),i} = \left[B_{ti} + r B_{ti} \left(1 - \frac{B_{ti}}{K} \right) - C_{ti} \right] e^{\epsilon_{ti}} \quad (3.26a)$$

$$C_{ti} = q E_{ti} B_{ti} e^{\eta_{ti}}, \quad (3.26b)$$

onde $t = 1, \dots, T$ anos, e $i = 1, \dots, n$ áreas. Note que os erros são multiplicativos e podem assumir estruturas simples como iid ou qualquer outra especificação mais complexa.

O modelo acima, apesar de considerar a estrutura espacial na dinâmica do processo, ainda possui uma forte suposição de que os parâmetros envolvidos (r , q , K) são constantes para todos os anos e todas as áreas. No entanto, sabe-se, por exemplo, que o coeficiente de capturabilidade (q) depende de uma série de fatores, tanto ambientais (e.g. área, época do ano), quanto operacionais (e.g. tipo de artefato de pesca, experiência dos pescadores). Por isso, é natural considerar que esse parâmetro possua diferentes valores em diferentes áreas e anos. Além disso, os parâmetros r (taxa de crescimento populacional per capita) e K (capacidade de suporte da população) também devem variar ao longo dos anos devido à flutuações ambientais naturais, e, principalmente, em populações exploradas pela pesca. Ainda se considerarmos que áreas distintas, se pensadas como unidades discretas, podem conter indivíduos em diferentes estágios de vida (cf. triângulo de Harden-Jones, Figura 3.1), é muito provável que estes parâmetros também possam variar conforme a área.

Dessa forma, é natural assumir que os parâmetros (r , q , K) também possuem uma distribuição de probabilidade, e que estas distribuições podem assumir diferentes estruturas espaço-temporais. Sendo assim, o modelo mais geral considerado neste trabalho é dado pelas seguintes equações

$$B_{(t+1),i} = \left[B_{ti} + r_{ti} B_{ti} \left(1 - \frac{B_{ti}}{K_{ti}} \right) - C_{ti} \right] e^{\epsilon_{ti}} \quad (3.27a)$$

$$C_{ti} = q_{ti} E_{ti} B_{ti} e^{\eta_{ti}}, \quad (3.27b)$$

onde r_{ti} , q_{ti} , e K_{ti} são também considerados variáveis aleatórias, e, portanto, possuem alguma distribuição de probabilidade. Um caso mais simples, por exemplo, é assumir que $\ln r_{ti} \sim N(r, \sigma_r^2)$ (o mesmo com os outros parâmetros), e, assim, pode-se dizer que as equações (3.26a) e (3.26b) do modelo anterior (onde os parâmetros são fixos) são casos particulares das equações (3.27a) e (3.27b), respectivamente.

No entanto, como será visto na seção 3.3.4 abaixo, a grande vantagem do modelo especificado dessa forma é a grande flexibilidade que os parâmetros possuem, possibilitando estimativas mais condizentes com a complexa realidade dos processos dinâmicos. Ao se considerar que os parâmetros possuem uma distribuição de probabilidade é natural que a abordagem bayesiana seja considerada para o processo de estimação.

3.3.2 Método de estimação

Como mencionado na seção 3.2.4, existem várias abordagens para a estimação dos modelos dinâmicos de biomassa. Atualmente a forma mais utilizada e com maior flexibilidade para considerar erros tanto na equação do processo quanto da observação são os modelos de espaço de estado, principalmente no contexto bayesiano (cf. seção 3.2.4.4). Embora seja uma alternativa viável atualmente, a abordagem bayesiana geralmente envolve um processo de simulação estocástica, onde, uma vez que um modelo tenha sido especificado, são realizadas milhares de amostragens desse modelo com a intenção de aproximar a distribuição posterior dos parâmetros. Esse processo de simulação geralmente é feito pelos método de

Monte Carlo via cadeias de Markov (MCMC) (*e.g.* Millar e Meyer, 2000b) ou amostragem por importância (*e.g.* Kinas, 1996). Os *softwares* geralmente utilizados para estas implementações são o BUGS (Lunn et al., 2009) ou o JAGS (Plummer, 2003). Recentemente, uma alternativa é o *software* STAN (Carpenter et al., 2017), que implementa o chamado método de Monte Carlo Hamiltoniano, que promete ser mais eficiente em termos computacionais do que os anteriores.

Mesmo com todo o desempenho e capacidade computacional atual, os métodos de simulação por amostragem podem ser computacionalmente ineficientes para modelos complexos, como são os modelos dinâmicos de biomassa. Mesmo para os modelos DB tradicionais (aqueles que não consideram estruturas espaço-temporais) e um série temporal modesta (20-30 anos), o processo de estimação pode levar de alguns minutos a algumas horas para convergir. Se for então considerado o modelo proposto (com estrutura espaço-temporal), a complexidade do modelo aumenta, e por consequência também o tempo computacional. Além disso, a implementação de modelos com as estruturas propostas nestes *softwares* não é trivial.

No contexto de estimação por máxima verossimilhança, os modelos DB tradicionais podem ser implementados por *softwares* como AD Model Builder (*Automatic Differentiation Model Builder*, Fournier et al., 2012), e, mais recentemente, com o TMB (*Template Model Builder*, Kristensen et al., 2016). Ambos utilizam o método matemático de diferenciação automática para obtenção das derivadas de primeira e segunda ordem das funções de verossimilhança, e ainda utilizam a aproximação de Laplace para solução de integrais complexas. No TMB, por exemplo, o modelo é escrito como um programa em C++, que, após compilado, pode ser passado para alguma rotina de otimização como Nelder-Mead ou BFGS. A inclusão de estruturas espaço-temporais é possível no TMB, mas o custo computacional no processo de otimização também é alto.

Seja no contexto bayesiano ou verosimilhancista, uma desvantagem dos *softwares* disponíveis é que todos necessitam de algum tipo de “afinamento” (*tuning*) em seus algoritmos. Por exemplo, todos os métodos requerem valores iniciais dos parâmetros, e modelos não lineares complexos podem ser altamente sensíveis à esses valores. No caso bayesiano, alguma reparametrização do modelo pode ser necessária (*e.g.* Millar e Meyer, 2000b), enquanto que no caso verosimilhancista, os ajustes dependem do método de otimização utilizado.

No contexto bayesiano, uma alternativa atrativa é o uso de aproximações analíticas para as distribuições posteriores. A aproximação de Laplace encaixada e integrada (*Integrated Nested Laplace Approximation*, INLA), proposta por Rue et al. (2009), tem provado ser capaz de ajustar uma variedade de modelos complexos com precisão e grande rapidez computacional. Inicialmente desenvolvido como um programa *standalone*, o INLA ganhou popularidade com a disponibilização do pacote **INLA** (Lindgren e Rue, 2015) para a linguagem R (R Core Team, 2018). A metodologia INLA utiliza um algoritmo determinístico para inferência bayesiana, ao invés de métodos baseados em simulação como o MCMC. O INLA foi desenvolvido especificamente para a classe de modelos gaussianos latentes, que cobre uma vasta família de modelos, incluindo, *e.g.*, regressão linear simples, GLMs, GLMMs, e GAMs. Comparado aos métodos de MCMC, o INLA fornece resultados acurados em um tempo computacional menor (Blangiardo e Cameletti, 2015).

Apesar da grande flexibilidade de modelos contemplados pelo INLA, modelos de espaço de estados (ou modelos dinâmicos) não possuíam uma forma trivial de serem implementados. No entanto Ruiz-Cárdenas et al. (2012) desenvolveram um método para o ajuste direto desse tipo de modelo no INLA, através de algumas extensões como a “aumentação de dados” (*data augmentation*). Estes autores mostraram que diversos modelos dinâmicos, como o filtro de Kalman e modelos dinâmicos espaço-temporais poderiam ser ajustados pelo INLA com apenas algumas adaptações. Apesar disso, a classe de modelos capaz de ser acomodada por estas extensões não inclui modelos que envolvem termos quadráticos, como é o caso dos modelos dinâmicos de biomassa, e, em geral, modelos de aumento populacional. O motivo é

que modelos dinâmicos com termos quadráticos não podem ser escritos na forma de um campo aleatório gaussiano markoviano (*Gaussian Markov Random Field* (GMRF), Rue e Held, 2005), fundamental para a estimação via INLA.

Com o objetivo de desenvolver uma forma de estimar os parâmetros de modelos dinâmicos de biomassa espaço-temporais (equações (3.27a) e (3.27b)), utilizando a abordagem mais natural que seria a bayesiana, de forma rápida e sem a necessidade de ajustes em algoritmos, foi adotado então o método de regressão linear, similar ao descrito na seção 3.2.4.2. Para fins de demonstração, considere as equações dinâmicas com estrutura espacial, mas sem os termos de erros estocásticos e com os parâmetros r , q , e K fixos, ou seja,

$$\begin{aligned} B_{(t+1),i} &= B_{ti} + rB_{ti} \left(1 - \frac{B_{ti}}{K}\right) - C_{ti} \\ C_{ti} &= qE_{ti}B_{ti}. \end{aligned}$$

Seguindo a mesma ideia proposta por Hilborn (1979), podemos mostrar que a CPUE por ano e área é proporcional à biomassa por ano e área, ou seja,

$$\begin{aligned} C_{ti} &= qE_{ti}B_{ti} \\ \frac{C_{ti}}{E_{ti}} &\equiv U_{ti} = qB_{ti}. \end{aligned}$$

Isolando a biomassa nessa equação, temos

$$B_{ti} = \frac{U_{ti}}{q}. \quad (3.28)$$

Usando o fato de que $C_{ti} = qE_{ti}B_{ti}$, e rearranjando os termos na equação do processo,

$$B_{(t+1),i} = B_{ti} + rB_{ti} - \frac{r}{K}B_{ti}^2 - qE_{ti}B_{ti}. \quad (3.29)$$

Substituindo (3.28) em (3.29), chegamos à seguinte relação

$$\frac{U_{(t+1),i} - U_{ti}}{U_{ti}} = r - \frac{r}{qK}U_{ti} - qE_{ti} \quad (3.30)$$

(ver desenvolvimento no Apêndice A). Claramente (3.30) pode ser escrita como um modelo de regressão linear da seguinte forma

$$Y_{ti} = \beta_0 + \beta_1 X_{1ti} + \beta_2 X_{2ti}$$

onde

$$\begin{aligned} Y_{ti} &= \frac{U_{(t+1),i} - U_{ti}}{U_{ti}}, & X_{1ti} &= U_{ti}, & X_{2ti} &= E_{ti}, \\ \beta_0 &= r, & \beta_1 &= -\frac{r}{qK}, & \beta_2 &= -q. \end{aligned}$$

Note que, se $t = 1, \dots, T$ representa as observações em cada tempo, então, nessa equação, são utilizadas $T - 1$ observações em cada área, pois aqui $t = 1, \dots, T - 1$. Se o número de áreas é n , então são utilizadas $(T - 1)n$ observações nessa equação.

A equação (3.30), sendo um modelo de regressão linear, pode ser facilmente ajustada através do INLA. A primeira vantagem é que é uma forma rápida de ajuste e obtenção dos parâmetros de interesse, além de requerer apenas dados de captura e esforço por ano e área. A segunda vantagem é que os parâmetros desse modelo podem ser tratados como efeitos aleatórios, que podem considerar então diferentes estruturas espaciais, temporais e/ou espaço-temporais. Com isso, é possível obter a distribuição posterior para cada um dos parâmetros, assim como seus efeitos temporais e espaciais específicos.

Portanto, a inferência para os parâmetros do modelo dinâmico dado pelas equações (3.27a) e (3.27b) pode ser realizada, pelo menos de forma aproximada, pela equação linear dada em (3.30), através

do INLA. A diferença principal até aqui dos modelos propostos por Hilborn (1979) e Schnute (1977) é a consideração explícita da estrutura espacial. De qualquer maneira, a simples utilização de dados temporais estruturados no espaço não deve fornecer vantagens, a não ser que a estrutura de dependência espacial ou espaço-temporal seja levada em consideração. Por isso, algumas formas alternativas deste modelo, e que consideram estruturas espaciais e temporais são apresentadas na próxima seção.

3.3.3 Modelos considerados

Como mencionado na seção anterior, a estimativa dos parâmetros de interesse r , q , e K do modelo dinâmico de biomassa espaço-temporal definido em (3.27a) e (3.27b) será realizado através da equação linearizada (3.30), por meio de inferência bayesiana utilizando o INLA. Para isso, será adotada também uma abordagem em dois estágios. No primeiro estágio será realizada a modelagem dos dados brutos de captura e esforço por ano e área. O objetivo desta primeira etapa é obter a predição dos dados de captura livre de efeitos que não sejam devidos apenas à flutuação da biomassa da população. O segundo estágio consiste então em calcular a CPUE predita, através da captura predita no primeiro estágio, e utilizá-la na equação linear (3.30).

3.3.3.1 Primeiro estágio

O primeiro estágio consiste na modelagem da captura observada e possui dois objetivos: (i) fazer com que a relação $U_{ti} \propto B_{ti}$ seja de fato proporcional, e (ii) minimizar o erro observacional para que o uso de U_{ti} na equação do segundo estágio não cause o chamado “erro nas variáveis”.

Aplicando o logaritmo em ambos os lados da equação observacional (3.27b) tem-se

$$\ln C_{ti} = \ln q_{ti} + \ln E_{ti} + \ln B_{ti} + \eta_{ti}, \quad (3.31)$$

onde as variáveis observadas são $\ln C_{ti}$ e $\ln E_{ti}$, enquanto que $\ln q_{ti}$ é um parâmetro desconhecido, $\ln B_{ti}$ é uma variável latente, e η_{ti} é um termo de erro, que no caso mais simples pode ser especificado como $\eta_{ti} \sim N(0, \sigma_\eta^2)$. Como o coeficiente da capturabilidade q_{ti} é o parâmetro de proporcionalidade entre a CPUE e a biomassa ($U_{ti} = q_{ti}B_{ti} \Rightarrow U_{ti} \propto B_{ti}$), então sabe-se que toda a variabilidade observada na captura (por consequência na CPUE) que não é devida exclusivamente à variações na biomassa, são atribuídas a esse parâmetro. Por exemplo, variações na biomassa entre áreas e épocas do ano podem ser captadas pelo coeficiente de capturabilidade. Por isso também a importância de não considerar esse parâmetro como fixo, como normalmente é assumido. Dessa forma, considerando também que a biomassa é uma variável latente, podemos escrever (3.31) como um modelo de efeitos aleatórios

$$\ln C_{ti} = b_0 + \beta_1 \ln E_{ti} + u_i + v_i + \gamma_t + \phi_t + \delta_{ti} \quad (3.32)$$

onde b_0 é o intercepto, que representa a média geral de $\ln C_{ti}$ para todas as áreas e todos os anos, β_1 é um efeito fixo relacionado ao esforço de pesca, v_i é o efeito específico de área modelado como iid, $v_i \sim N(0, \tau_v^{-1})$, ϕ_t é o efeito específico de ano, também iid, $\phi_t \sim N(0, \tau_\phi^{-1})$.²

O termo u_i é também é um efeito específico de área, mas estruturado espacialmente. Diversas estruturas podem ser consideradas dependendo do tipo de dado espacial considerado (*e.g.* dados no espaço contínuo ou discreto). Neste trabalho o foco serão dados de área (ou *lattice*), pois são os mais comumente encontrados em bases de dados de pesca. Normalmente os dados de captura e esforço em uma determinada coordenada são agregados em uma área maior com uma resolução específica (*e.g.* $1^\circ \times 1^\circ$ ou $5^\circ \times 5^\circ$). Sendo assim, um dos modelos mais utilizados para esse tipo de dados é o condicional autoregressivo intrínseco (iCAR), proposto inicialmente por Besag (1974). Considere o vetor $\mathbf{u} = u_1, \dots, u_n$, onde cada

²De maneira geral, τ é um parâmetro de precisão, definido como $\tau = 1/\sigma^2$.

uma das n áreas é caracterizada por um conjunto de vizinhos $\mathcal{N}(i)$, então vamos assumir que u_i é a seguinte variável aleatória

$$u_i | \mathbf{u}_{-i} \sim N \left(\mu_i + \frac{1}{\mathcal{N}_i} \sum_{j=1}^n a_{ij} (u_j - \mu_j), \frac{\tau_u^{-1}}{\mathcal{N}_i} \right) \quad (3.33)$$

onde $u_i | \mathbf{u}_{-i}$ representa o valor u_i condicionado a todos os outros valores com exceção de u_i , μ_i é a média da área i , e $\mathcal{N}_i = \#\mathcal{N}(i)$, i.e. é o número de vizinhos da área i . A variável a_{ij} é 1 se as áreas i e j são vizinhas, e 0 caso contrário ($a_{ii} = 0$). Note que, através desta especificação, quanto maior for o número de vizinhos de uma área, mais informação haverá sobre aquele efeito aleatório, e consequentemente a variância será menor. O parâmetro de variância $\sigma_u^2 = \tau_u^{-1}$ controla a quantidade de variação entre efeitos aleatórios estruturados (Blangiardo e Cameletti, 2015).

Junto com o efeito aleatório não estruturado v_i em (3.32), a especificação dada em (3.33) dá origem ao chamado modelo BYM (*Besag-York-Mollie*), proposto por Besag et al. (1991). A matriz de covariância deste modelo não é positiva definida (Rue e Held, 2005), portanto não existe uma distribuição conjunta apropriada para \mathbf{u} . Para fazer com que a distribuição conjunta seja própria, assume-se a restrição $\sum_{i=1}^n u_i = 0$ (Schrödle e Held, 2011).

Ainda em relação a 3.32, o termo γ_t representa um efeito temporal estruturado, modelado dinamicamente através de uma estrutura de passeio aleatório (*Random Walk*, RW) de ordem 1 ou 2. Sendo assim, a distribuição condicional de γ_t pode ser

$$\gamma_t | \gamma_{t-1} \sim N(\gamma_{t-1}, \tau_\gamma^{-1}) \quad \text{RW de ordem 1} \quad (3.34a)$$

$$\gamma_t | \gamma_{t-1}, \gamma_{t-2} \sim N(2\gamma_{t-1} + \gamma_{t-2}, \tau_\gamma^{-1}) \quad \text{RW de ordem 2.} \quad (3.34b)$$

O termo δ_{ti} representa a interação espaço-temporal, que pode assumir diferentes formas dependendo dos termos que estão interagindo. De maneira geral, o vetor de parâmetros $\boldsymbol{\delta}$ segue uma distribuição Normal com matriz de precisão $\mathbf{Q}_\delta = \tau_\delta \mathbf{R}_\delta$. τ_δ é um parâmetro de precisão desconhecido e \mathbf{R}_δ é uma matriz estruturada, identificando o tipo de dependência temporal e/ou espaço-temporal entre os elementos de $\boldsymbol{\delta}$ (Blangiardo e Cameletti, 2015). A matriz \mathbf{R}_δ pode ser fatorada como o produto Kronecker das matrizes estruturadas que definem cada termo principal. Com isso, existem quatro tipos de interação espaço-temporal possíveis, definidas inicialmente por Knorr-Held (2000), e apresentadas aqui na Tabela 3.1.

Tabela 3.1. Tipos de interação espaço-temporal. Os termos de interação se referem ao modelo (3.32).

Interação	Termos	\mathbf{R}_δ	Descrição
I	v_i e ϕ_t	$\mathbf{R}_v \otimes \mathbf{R}_\phi$	Interação espaço-temporal não estruturada.
II	v_i e γ_t	$\mathbf{R}_v \otimes \mathbf{R}_\gamma$	Cada área possui uma estrutura autoregressiva no tempo, independente das outras áreas.
III	ϕ_t e u_i	$\mathbf{R}_\phi \otimes \mathbf{R}_u$	Cada ano possui uma estrutura no espaço, independente de outros anos.
IV	u_i e γ_t	$\mathbf{R}_u \otimes \mathbf{R}_\gamma$	Cada área possui uma estrutura autoregressiva no tempo, que depende também do padrão temporal das outras áreas.

O objetivo da modelagem da captura neste primeiro estágio é portanto obter previsões que levem em conta todas as possíveis formas de interação espacial e temporal, de forma a fazer com que a relação entre CPUE e biomassa seja de fato proporcional. Para investigar quais das possíveis interações descreve melhor essa relação, foram ajustados sete diferentes modelos aos dados. A Tabela 3.2 apresenta os modelos considerados para a modelagem da captura no primeiro estágio. Em todos os modelos, o termo espacial estruturado (u_i) considerado foi o iCAR (equação 3.33), e o termo temporal estruturado (γ_t) foi modelado considerando-se um passeio aleatório de ordem 2 (equação 3.34b).

Tabela 3.2. Estrutura e descrição dos modelos considerados no primeiro estágio. Em todos os casos a variável resposta é $\ln C_{ti}$.

Identificação	Predictor linear	Descrição
1 m1	$\beta_0 + \beta_1 \ln E_{ti}$	Modelo de efeitos fixos.
2 m.pt	$b_0 + \beta_1 \ln E_{ti} + u_i + v_i + (\beta + \delta_i)t$	Modelo com efeitos espaciais, tendência temporal global β e efeito aleatório δ_i que identifica a interação entre tempo e espaço.
3 m.npt	$b_0 + \beta_1 \ln E_{ti} + u_i + v_i + \gamma_t + \phi_t$	Modelo com efeitos espaciais e temporais, sem interação.
4 m.t1i	$b_0 + \beta_1 \ln E_{ti} + u_i + v_i + \gamma_t + \phi_t + \delta_{ti}$	Modelo com interação do tipo I.
5 m.t2i	$b_0 + \beta_1 \ln E_{ti} + u_i + v_i + \gamma_t + \phi_t + \delta_{ti}$	Modelo com interação do tipo II.
6 m.t3i	$b_0 + \beta_1 \ln E_{ti} + u_i + v_i + \gamma_t + \phi_t + \delta_{ti}$	Modelo com interação do tipo III.
7 m.t4i	$b_0 + \beta_1 \ln E_{ti} + u_i + v_i + \gamma_t + \phi_t + \delta_{ti}$	Modelo com interação do tipo IV.

Para a comparação destes modelos, foram utilizados alguns critérios em conjunto: a redução na deviance, o critério de informação da deviance (*Deviance Information Criterion*, DIC) e o número efetivo de parâmetros p_D (Spiegelhalter et al., 2002), o critério de informação amplamente aplicável (*Widely applicable Information Criterion*, WAIC) e o número efetivo de parâmetros p_{WAIC} (Watanabe, 2012). Como medida do poder de predição do modelo, foi utilizada a soma do logaritmo da ordenada preditiva condicional (*Conditioning Predictive Ordinate* (CPO), Carlin e Louis, 2009). Dado este conjunto de medidas de diagnóstico, modelos com valores de DIC e WAIC menores são considerados melhores (no sentido de que explicam melhor a variável resposta), enquanto que modelos com valores maiores de LCPO possuem uma maior capacidade preditiva. O número efetivo de parâmetros ajuda a identificar modelos com sobreajuste, ou seja, quando p_D e/ou p_{WAIC} se aproximam do número de observações na amostra (Carlin e Louis, 2009). Modelos com esta característica não auxiliam na redução da informação já contida na amostra, portanto não possuem muita utilidade.

Em todos os casos foram obtidas as distribuições posteriores das predições para cada observação. Como a variável resposta está na escala logarítmica, as distribuições posteriores foram transformadas por meio da função exponencial. As médias de cada distribuição posterior transformada foram então utilizadas como valor predito de captura C_{ti} na escala original. Seja \hat{y}_{ti} a média da distribuição posterior transformada da predição de cada observação, então

$$\hat{C}_{ti} = \hat{y}_{ti}$$

é o valor predito de captura no ano t e área i .

3.3.3.2 Segundo estágio

O segundo estágio de estimação tem a finalidade de estimar os parâmetros de interesse (r, q, K), a partir da equação linearizada (3.30) do processo dinâmico. A ideia é então a de utilizar os valores de captura preditos no primeiro estágio, supostamente livres de viés observacional, e utilizar esses valores para ajustar o modelo linear.

Para fins de comparação, os dados agregados por ano também serão utilizados para a obtenção das estimativas. Dessa forma será possível comparar se existe de fato vantagem ao se considerar os dados com estrutura espacial. Além disso, também serão utilizados os dados brutos (simulados/observados) neste segundo estágio, tanto para o modelo para dados agregados quanto para o modelo para dados desagregados. Dessa forma será também possível verificar se os valores preditos realmente podem interferir nas estimativas dos parâmetros no segundo estágio.

A seguir serão detalhados os modelos considerados tanto para os dados agregados quanto para os dados desagregados.

Dados agregados Os dados agregados são a soma das capturas de todas as áreas para cada ano. Esta é a configuração padrão considerada atualmente para a aplicação dos modelos dinâmicos de biomassa. Os dados agregados por ano são utilizados em todos os métodos de estimação descritos na seção 3.2.4, portanto serão utilizados aqui para fins de comparação.

O modelo linear utilizado para dados agregados é

$$\frac{\hat{U}_{t+1} - \hat{U}_t}{\hat{U}_t} = r - \frac{r}{qK} \hat{U}_t - qE_t + \nu_t \quad (3.35)$$

onde ν_t é um termo de erro aleatório, e

$$\hat{U}_t = \frac{\sum_{i=1}^n \hat{C}_{ti}}{\sum_{i=1}^n E_{ti}},$$

ou seja, a CPUE predita por ano é a razão entre a soma anual da captura predita por ano e área e o esforço somado da mesma forma.

O modelo (3.35) pode então ser escrito como um modelo de regressão linear

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \nu_t \quad (3.36)$$

onde

$$\begin{aligned} Y_t &= \frac{\hat{U}_{t+1} - \hat{U}_t}{\hat{U}_t}, & X_{1t} &= \hat{U}_t, & X_{2t} &= E_t, \\ \beta_0 &= r, & \beta_1 &= -\frac{r}{qK}, & \beta_2 &= -q. \end{aligned}$$

Dessa forma, após a obtenção das distribuições posteriores de β_0 , β_1 e β_2 , as distribuições posteriores dos parâmetros r , q , e K podem ser obtidas por meio de

$$r = \beta_0, \quad q = -\beta_2, \quad K = \frac{\beta_0}{\beta_1 \beta_2}.$$

Com isso, toda a informação sobre a distribuição dos parâmetros pode ser obtida com medidas resumo destas posteriores.³

Além de considerar o modelo (3.36), também foram ajustados modelos que levam em consideração a estrutura de dependência temporal dos dados agregados. Para cada um dos parâmetros, um termo de efeito aleatório foi adicionado. O efeito considerado aqui foi um passeio aleatório de ordem 2 (RW2), e, com isso, foram ajustados quatro modelos com aumento de complexidade entre eles, conforme descrito na Tabela 3.3.

Tabela 3.3. Estrutura e descrição dos modelos considerados no segundo estágio para dados agregados.

	Identificação	Predictor linear	Descrição
1	a1	$\beta_0, \beta_1, \beta_2$	Modelo de efeitos fixos.
2	a2	$\beta_{0,t}, \beta_1, \beta_2$	Efeito RW2 em β_0
3	a3	$\beta_{0,t}, \beta_{1,t}, \beta_2$	Efeito RW2 em β_0 e β_1
4	a4	$\beta_{0,t}, \beta_{1,t}, \beta_{2,t}$	Efeito RW2 em β_0 , β_1 e β_2

Convém ressaltar também que cada um destes quatro modelos foi ajustado para cada um dos sete modelos considerados no estágio 1 (cf. Tabela 3.2). Sendo mais específico, o valor predito em cada um dos sete modelos foi utilizado para o ajuste dos quatro modelos considerados aqui. Com isso, um total de $7 \times 4 = 28$ modelos foram ajustados. Vale lembrar também que os dados brutos (*i.e.* a CPUE simulada/observada) foram utilizados igualmente para o ajuste destes modelos, portanto resultando em um total de $28 + 4 = 32$ modelos considerados para os dados agregados.

³Por padrão, no INLA, são obtidas as distribuições posteriores marginais dos parâmetros. Como mostrado acima, o parâmetro K é uma função dos demais parâmetros envolvidos, e, por isso, se faz necessária a obtenção da distribuição conjunta de todos os parâmetros. No INLA, a obtenção desta posterior conjunta (aproximada) pode ser obtida por simulação de Monte Carlo, através da função `inla.posterior.sample()`. Os detalhes técnicos desse procedimento fogem do âmbito deste trabalho.

Dados desagregados Os dados desagregados são os dados que consideram as observações provenientes de todas as áreas em cada ano. Espera-se que a consideração da estrutura espacial neste segundo estágio resulte em estimativas mais precisas dos parâmetros de interesse através do modelo dinâmico linearizado.

Como demonstrado na equação (3.30), o modelo linear para os dados desagregados é

$$\frac{\hat{U}_{(t+1),i} - \hat{U}_{ti}}{\hat{U}_{ti}} = r - \frac{r}{qK} \hat{U}_{ti} - qE_{ti} + \nu_{ti} \quad (3.37)$$

onde ν_{ti} é um termo de erro aleatório, e

$$\hat{U}_{ti} = \frac{\hat{C}_{ti}}{E_{ti}},$$

ou seja, a CPUE predita é obtida simplesmente pela divisão entre a captura predita por ano e área no primeiro estágio, dividida pelo esforço observado.

Claramente o modelo 3.37 pode ser escrito como

$$Y_{ti} = \beta_0 + \beta_1 X_{1ti} + \beta_2 X_{2ti} + \nu_{ti} \quad (3.38)$$

onde

$$\begin{aligned} Y_{ti} &= \frac{\hat{U}_{(t+1),i} - \hat{U}_{ti}}{\hat{U}_{ti}}, & X_{1ti} &= \hat{U}_{ti}, & X_{2ti} &= E_{ti}, \\ \beta_0 &= r, & \beta_1 &= -\frac{r}{qK}, & \beta_2 &= -q. \end{aligned}$$

Assim como para o caso dos dados agregados, as distribuições posteriores dos parâmetros r , q , e K podem ser obtidas por

$$r = \beta_0, \quad q = -\beta_2, \quad K = \frac{\beta_0}{\beta_1 \beta_2}.$$

O modelo (3.38) utiliza agora informações provenientes de área e ano. Com isso, diversas estruturas de dependência puderam ser testadas, dando origem aos modelos descritos na Tabela 3.4. Estes modelos consideram inicialmente que os parâmetros possuem dependência apenas temporal, através de um passeio aleatório de ordem 2 (RW2), ou apenas espacial, através da especificação de efeitos iCAR somados a efeitos iid (modelo BYM, cf. seção 3.3.3.1). Os modelos mais complexos consideram uma estrutura de dependência espaço-temporal nos parâmetros da regressão. Novamente, os modelos foram construídos em crescente grau de complexidade.

Tabela 3.4. Estrutura e descrição dos modelos considerados no segundo estágio para dados desagregados.

	Identificação	Predictor linear	Descrição
1	d1	$\beta_0, \beta_1, \beta_2$	Modelo de efeitos fixos.
2	d2	$\beta_{0,t}, \beta_1, \beta_2$	Efeito RW2 em β_0
3	d3	$\beta_{0,t}, \beta_{1,t}, \beta_2$	Efeito RW2 em β_0 e β_1
4	d4	$\beta_{0,t}, \beta_{1,t}, \beta_{2,t}$	Efeito RW2 em β_0 , β_1 e β_2
5	d5	$\beta_{0,i}, \beta_1, \beta_2$	Efeito BYM em β_0
6	d6	$\beta_{0,i}, \beta_{1,i}, \beta_2$	Efeito BYM em β_0 e β_1
7	d7	$\beta_{0,i}, \beta_{1,i}, \beta_{2,i}$	Efeito BYM em β_0 , β_1 e β_2
8	d8	$\beta_{0,ti}, \beta_1, \beta_2$	Interação espaço-temporal em β_0
9	d9	$\beta_{0,ti}, \beta_{1,ti}, \beta_2$	Interação espaço-temporal em β_0 e β_1
10	d10	$\beta_{0,ti}, \beta_{1,ti}, \beta_{2,ti}$	Interação espaço-temporal em β_0 , β_1 e β_2

Assim como foi feito com os dados agregados, a predição de cada um dos sete modelos considerados no primeiro estágio foi utilizada aqui para o ajuste de cada um dos dez modelos para dados desagregados. Ou seja, nesta etapa, foram ajustados $7 \times 10 = 70$ modelos. Considere ainda que os dados brutos desagregados também foram utilizados para o ajuste destes dez modelos. Dessa forma, $70 + 10 = 80$ modelos foram considerados no total.

Levando em consideração todos os modelos propostos, no primeiro estágio são 7 modelos ajustados, no segundo estágio são 32 para os dados agregados, e 80 para os dados desagregados. Com isso, $7 + 32 + 80 = 119$ modelos foram analisados neste trabalho.

3.3.3.3 Prioris e hiperprioris

Para completar a definição dos modelos no contexto bayesiano, é necessário especificar as distribuições *a priori* (prioris) para os parâmetros envolvidos, e as hiperprioris para os parâmetros de precisão.

Considerando o modelo completo do primeiro estágio (equação (3.32)), as prioris para os efeitos fixos são $b_0 \sim N(0, 0^{-1})$ (equivalente a $U(-\infty, \infty)$), e $\beta_1 \sim N(0, 0.001^{-1})$. As prioris para os efeitos aleatórios ($u_i, v_i, \gamma_t, \phi_t, \delta_{ti}$) foram definidas na Seção 3.3.3.1. De maneira geral, cada um destes termos possui distribuição normal com média zero e matriz de precisão $\mathbf{Q} = \tau \mathbf{R}$, onde τ é a precisão, e \mathbf{R} é a matriz que define a estrutura do efeito aleatório. Neste caso, \mathbf{R}_u é estruturada no espaço conforme o modelo autoregressivo intrínseco (iCAR, equação 3.33), e $\mathbf{R}_v = \mathbf{I}_n$, onde \mathbf{I}_n é a matriz identidade de ordem n . A matriz \mathbf{R}_γ é estruturada no tempo seguindo um modelo de passeio aleatório de ordem 2 (RW2, equação 3.34b), enquanto que $\mathbf{R}_\phi = \mathbf{I}_T$. A matriz \mathbf{R}_δ poderá assumir diferentes estruturas dependendo dos termos considerados na interação. Estas estruturas estão definidas na Tabela 3.1.

Foi considerado que os parâmetros de precisão $\tau_v, \tau_\gamma, \tau_\phi$, e τ_δ , dos efeitos v_i, γ_t, ϕ_t , e δ_{ti} , respectivamente, seguem uma distribuição $G(1, 0.00005)$, onde $G(a, b)$ é a distribuição gama parametrizada de forma que sua esperança é a/b e sua variância é a/b^2 . A distribuição *a priori* para a precisão do efeito iCAR, u_i , foi especificada através de uma priori com complexidade penalizada (*Penalized Complexity prior* (PC prior), Simpson et al., 2017), que é uma abordagem nova para a construção de prioris. Dessa forma, considere que $\sigma_u = 1/\sqrt{\tau_u}$ é o desvio padrão do efeito aleatório. Uma priori PC para σ_u é determinada definindo-se os parâmetros (z, α) de forma que

$$\text{Prob}(\sigma_u > z) = \alpha, \quad z > 0, \quad 0 < \alpha < 1.$$

A vantagem da especificação dessa forma é que pode-se definir prioris não informativas e informativas baseando-se apenas no comportamento do desvio padrão. Nesse trabalho foi utilizado $(z, \alpha) = (0.1, 0.01)$.

Para os modelos considerados no segundo estágio de estimativa, definidos nas tabelas 3.3 para dados agregados, e 3.4 para dados desagregados, as prioris e hiperprioris foram definidas de forma similar. De maneira mais específica, pode-se escrever o modelo mais complexo (a4) da Tabela 3.3 da seguinte forma

$$Y_t = \beta_{0,t} + \beta_{1,t}X_{1t} + \beta_{2,t}X_{2t} + \nu_t$$

onde todos os parâmetros possuem um efeito aleatório RW2. Sendo assim, $\beta_{0,t} = b_0 + w_{0t}$, $\beta_{1,t} = b_1 + w_{1t}$, $\beta_{2,t} = b_2 + w_{2t}$, onde b_0, b_1 , e b_2 são efeitos fixos, enquanto que w_{0t}, w_{1t} , e w_{2t} são os efeitos aleatórios. Assim, assume-se que as distribuições *a priori* para os efeitos fixos são $b_0 \sim N(0, 0^{-1})$, $b_j \sim N(0, 0.001^{-1})$, $j = 1, 2$. Neste caso específico, as prioris para $w_{jt}, j = 1, 2, 3$, são normais com média zero e matriz de precisão $\tau_w \mathbf{R}_w$, onde \mathbf{R}_w é estruturada seguindo o modelo RW2, e $\tau_w \sim G(1, 0.00005)$. De maneira similar, $\nu_t \sim N(0, \tau_\nu^{-1})$ com $\tau_\nu \sim G(1, 0.00005)$. Todos os demais modelos, inclusive aqueles para dados desagregados (Tabela 3.4), seguem a mesma especificação de prioris e hiperprioris, mudando apenas a estrutura da matriz \mathbf{R}_w no caso de efeitos BYM e interação.

3.3.4 Simulação

O método de estimativa dos parâmetros dos modelos (3.27a) e (3.27b) foi apresentado na seção 3.3.2, e mais especificamente, as equações consideradas para dados agregados e desagregados foram apresentadas na seção 3.3.3.2. Para verificar a adequação do método de estimativa à dados gerados pelas

equações dinâmicas do processo de geração de biomassa, foi realizado inicialmente um estudo de simulação. O objetivo da simulação é especificar valores para os parâmetros de interesse (ou distribuições para estes parâmetros) e gerar dados de captura, com base em toda a dinâmica e estocasticidade que este modelo apresenta. Com os dados gerados, o método de estimativa será utilizado para se tentar estimar os parâmetros conhecidos da simulação.

Para gerar dados de captura com dependência espaço-temporal, inicialmente foram definidas 25 áreas no oceano, cada uma com $5^\circ \times 5^\circ$ de dimensão. As áreas foram definidas em um grid regular com 5 linhas e 5 colunas. Os dados provenientes de áreas consideradas dessa forma, e em unidades discretas são denominados de “dados de área” (ou *lattice*). Sendo assim, qualquer observação realizada dentro de uma área específica será atribuída àquela área (independente de onde foi realizada dentro da área). Essa primeira suposição sobre a estrutura espacial dos dados foi feita pois esse é o tipo de informação mais comum em bases de dados de pesca. A resolução da área pode ser maior ou menor, mas dados de pesca pontuais são difíceis de serem determinados, principalmente quando são utilizados equipamentos de pesca como o espinhel de superfície, por exemplo, que ocupam várias milhas e se estendem por grandes regiões. O grid regular considerado pode ser visto na Figura 3.2.

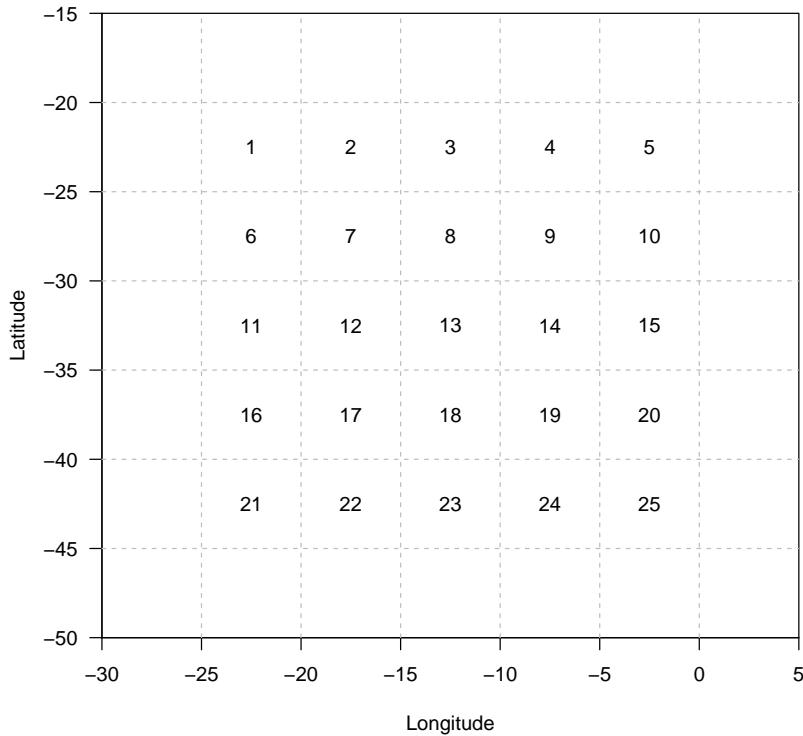


Figura 3.2. Definição de 25 áreas genéricas no oceano. Cada quadrado possui $5^\circ \times 5^\circ$ de dimensão.

Para utilizar os modelos (3.27a) e (3.27b) como geradores dos dados, foi aplicado o logaritmo em ambos os lados das duas equações, chegando ao processo equivalente

$$\ln B_{(t+1),i} = \ln \left[B_{ti} + r_{ti}B_{ti} \left(1 - \frac{B_{ti}}{K_{ti}} \right) - C_{ti} \right] + \epsilon_{ti} \quad (3.39a)$$

$$\ln C_{ti} = \ln q_{ti} + \ln E_{ti} + \ln B_{ti} + \eta_{ti}. \quad (3.39b)$$

Esse procedimento foi realizado por conveniência numérica, uma vez que agora todos os termos de erro são aditivos e poderão ser considerados como tendo distribuição Normal. Sendo assim, os erros do processo

e das observações foram definidos como

$$\epsilon_{ti} \sim N(0, \tau_\epsilon^{-1} \mathbf{R}^-) \quad \text{e} \quad \eta_{ti} \sim N(0, \tau_\eta^{-1} \mathbf{R}^-).$$

Da mesma forma, os parâmetros r_{ti} , q_{ti} e K_{ti} foram também considerados como tendo distribuição Normal com médias r , q e K , respectivamente, e estrutura de precisão espaço-temporal, ou seja,

$$\ln r_{ti} \sim N(r, \tau_r^{-1} \mathbf{R}^-) \quad \ln q_{ti} \sim N(q, \tau_q^{-1} \mathbf{R}^-) \quad \ln K_{ti} \sim N(K, \tau_K^{-1} \mathbf{R}^-).$$

Para gerar efeitos espaço-temporais através desta especificação, considere genericamente um vetor de efeitos temporais $\mathbf{u}_t = \{u_{t1}, \dots, u_{tn}\}$ para cada uma das n áreas, e $\mathbf{u}_t \sim N(0, \tau_u^{-1} \mathbf{R}^-)$. A evolução temporal de \mathbf{u}_t pode ser escrita como um processo autoregressivo de primeira ordem, ou seja,

$$\begin{aligned} \mathbf{u}_t &= \rho \mathbf{u}_{t-1} + (1 - \rho^2) \mathbf{w}_t \\ \mathbf{w}_t &\sim N(0, \tau_u^{-1} \mathbf{R}^-). \end{aligned}$$

O parâmetro ρ define o grau de dependência temporal entre as observações, o parâmetro de precisão τ_u determina o grau de dependência espacial, e a matriz \mathbf{R}^- é a inversa generalizada da matriz de vizinhança \mathbf{R} definida por

$$R_{ij} = \begin{cases} \mathcal{N}_i, & \text{se } i = j \\ -1, & \text{se } i \sim j \\ 0, & \text{caso contrário} \end{cases}$$

onde \mathcal{N}_i é o número de vizinhos da área i , e $i \sim j$ denota se as áreas i e j são vizinhas. Essa construção leva a uma dependência espaço-temporal dos efeitos em \mathbf{u}_t . Tanto os erros do processo e observacional quanto os efeitos dos parâmetros foram simuladas dessa forma.

Foi então considerada uma série temporal de $T = 50$ anos, com parâmetro de dependência temporal $\rho = 0.8$. A precisão de cada termo foi definida como $\tau_\epsilon = \tau_\eta = 1000$ para os erros do processo e observacional, e $\tau_r = \tau_q = \tau_K = 50$. A média dos parâmetros de interesse foram fixadas em $r = 0.4$, $q = 0.01$, e $K = 1000$. Uma observação importante é que o valor da capacidade de suporte K é definido aqui “por área”, ou seja, permite-se que cada área tenha em média determinado valor, somado ao efeito aleatório específico de cada área/ano. Isso significa que, em média, a capacidade de suporte total da população considerada (em um ano), para todas as áreas é 25000. Para dar início ao processo de simulação, também é necessário especificar um valor inicial de biomassa (B_0). Esse valor foi considerado como sendo 90% da capacidade de suporte, $B_0 = 0.9K$. Isso implica a suposição de que esta população ainda não teria sido explorada pela pesca (ou foi pouco explorada).

O último elemento necessário para gerar os dados de captura é a definição do esforço de pesca. A trajetória do esforço foi dividida em cinco etapas, conforme o modelo geral de desenvolvimento de uma pescaria descrito por Hilborn e Walters (1992). A fase inicial é a de desenvolvimento da pescaria, onde o esforço ainda é baixo. Posteriormente mais barcos entram na pescaria, e o esforço aumenta gradativamente até atingir um nível máximo, onde permanece por um período. Depois disso vem a fase de declínio da pescaria, onde o esforço gradativamente começa a diminuir até voltar ao mesmo nível de esforço inicial, onde se mantém até o final da série. A trajetória do esforço agregada por ano pode ser vista na Figura 3.3. O esforço foi determinado genericamente como “unidades de esforço”, e os valores foram definidos com mínimo de 5 e máximo de 40 unidades de esforço por área. É importante também salientar que apesar do esforço ter sido determinado como fixo, através dessa trajetória, o esforço quando determinado por área também teve um efeito aleatório espaço-temporal especificado.

O processo de simulação especificado acima foi utilizado para então gerar dados de captura com dependência espaço-temporal, proveniente de um processo dinâmico de evolução da biomassa latente de

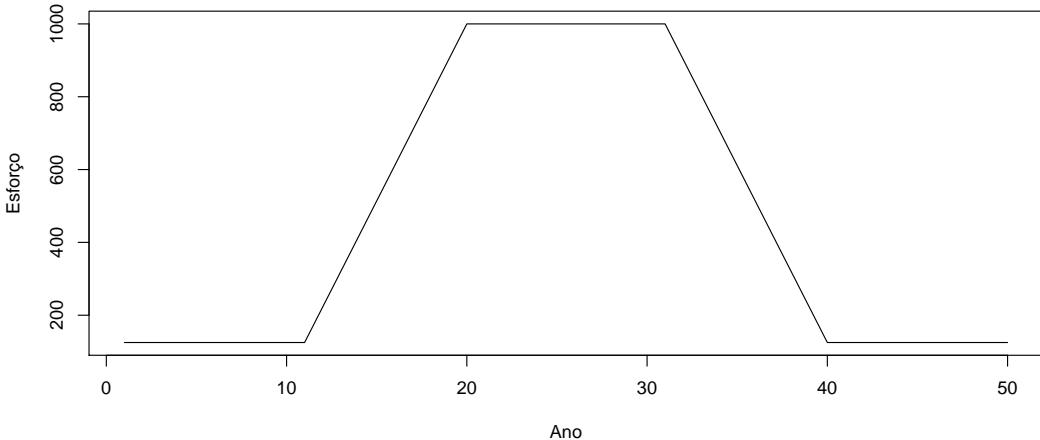


Figura 3.3. Série temporal do esforço de pesca considerado no processo de simulação (somado por ano).

uma população. Com os dados de captura e esforço por ano e área, o método de estimação definido na seção 3.3.3 foi aplicado, com a finalidade de se verificar se as estimativas dadas pelos modelos considerados seriam próximas aos valores especificados na simulação. Para isso, dois processos diferentes, mas complementares, de simulação foram executados, conforme descrito abaixo.

3.3.4.1 Simulação 1

O primeiro processo de simulação consistiu em simular uma única base de dados (\mathbf{y}). Todos os modelos especificados na seção 3.3.3 foram ajustados a essa base de dados. Para cada modelo ajustado, foram obtidas as distribuições posteriores, $\pi(\boldsymbol{\theta}|\mathbf{y})$, de cada um dos parâmetros de interesse $\boldsymbol{\theta} = \{r, q, K\}$. A partir destas distribuições posteriores foram simuladas 10000 amostras, e então, com base nos 10000 valores amostrados para cada parâmetro foram obtidas medidas resumo como média, quartis e desvio padrão.

Esse processo de simulação da posterior dado \mathbf{y} fixado tem o objetivo de avaliar a distribuição de cada um dos parâmetros em $\boldsymbol{\theta}$, que poderiam ter gerado \mathbf{y} (Gelman et al., 2014). Com isso é possível verificar se as estimativas pontuais $\hat{\boldsymbol{\theta}}(\mathbf{y})$ dos parâmetros, obtidas nas posteriores, estão próximas dos valores simulados, além de verificar eles estão dentro do intervalo de credibilidade. Neste cenário foram ajustados todos os 119 modelos considerados no estudo.

3.3.4.2 Simulação 2

O segundo processo de simulação consistiu em simular 100 bases de dados diferentes, e para cada uma destas bases de dados, todos os modelos propostos na seção 3.3.3 foram ajustados. Para cada um dos modelos, as distribuições posteriores dos parâmetros $\boldsymbol{\theta} = \{r, q, K\}$ foram obtidas, e, a partir destas posteriores, foram simuladas 1000 amostras. Com exceção do fato de que foram geradas 100 bases de dados ao invés de uma apenas, o processo até aqui é similar ao anterior. No entanto, o objetivo desta simulação é agora calcular a média das 100 distribuições posteriores simuladas para cada parâmetro e calcular medidas de viés.

O objetivo aqui é, portanto, contrário do anterior. Deseja-se verificar se, sob diferentes bases de dados, as estimativas pontuais dos parâmetros, como calculadas pela média da distribuição posterior, são acurados. Apesar de parecer um processo frequentista, Carlin e Louis (2009) argumentam que o uso de

uma estimativa pontual $\hat{\theta}(\mathbf{y})$ como a média da posterior, por exemplo, pode ser usado como medida de acurácia no contexto bayesiano. Neste cenário, os 119 modelos considerados no estudo foram ajustados para 100 bases de dados diferentes, portanto, ao total, foram $119 \times 100 = 11900$ modelos.

3.3.4.3 Medidas de adequação

A verificação da adequação das estimativas dos parâmetros obtidas em cada um dos modelos na simulação 1, foi realizada dentro do contexto completamente bayesiano. Dessa forma, as medidas resumo da distribuição posterior simulada para cada parâmetro foram comparadas aos valores reais dos parâmetros utilizados na simulação. Isso significa que foi verificado a proximidade da média (estimativa pontual) da posterior com o valor médio verdadeiro, e se este valor está ou não presente no intervalo de credibilidade de 95%.

No caso da simulação 2, o interesse estava em calcular o viés das estimativas pontuais dos parâmetros, obtidos para diferentes bases de dados, em relação à média verdadeira. Considerando a média da distribuição posterior como estimativa pontual de cada parâmetro, $\hat{\theta}(\mathbf{y})$, foram calculados o viés

$$\text{Viés}[\hat{\theta}(\mathbf{y})] = E[\hat{\theta}(\mathbf{y})] - \theta$$

o viés relativo

$$\text{Viés}_r[\hat{\theta}(\mathbf{y})] = \frac{E[\hat{\theta}(\mathbf{y})] - \theta}{\theta}$$

o erro quadrático médio (EQM)

$$\text{EQM}[\hat{\theta}(\mathbf{y})] = E[(\hat{\theta}(\mathbf{y}) - \theta)^2]$$

e a raiz quadrada do EQM, $\text{REQM}[\hat{\theta}(\mathbf{y})] = \sqrt{\text{EQM}}[...]$. Também foi calculada a taxa de cobertura, definida como a fração de vezes que o valor verdadeiro do parâmetro esteve dentro do intervalo de credibilidade de 95% de cada uma das 100 distribuições posteriores obtidas no processo de simulação para cada parâmetro.

3.4 Resultados e discussão

3.4.1 Simulação 1

Na simulação 1 foi gerada uma única base de dados, com o objetivo de se obterem as estimativas dos parâmetros r , q e K por meio de cada um dos modelos descritos na seção 3.3.3. As séries temporais agregadas por área de biomassa, captura e CPUE simuladas podem ser vistas na Figura 3.4. Pode-se notar que o comportamento destas variáveis ao longo do tempo segue um padrão esperado, de acordo com o esforço determinado. Note que a biomassa inicial está próxima da capacidade de suporte da população (como definido na simulação $K = 25000$ considerando todas as áreas, e $B_0 = 0.9K$). Quando o esforço aumenta, a captura aumenta, fazendo com que a biomassa seja reduzida. Nos períodos intermediários, mesmo com o elevado esforço, a captura tende a diminuir pois a biomassa está mais baixa. Com a redução do esforço, a biomassa apresenta uma tendência de aumento em direção à capacidade de suporte. A série temporal da CPUE tende a seguir o padrão da série da biomassa, indicando proporcionalidade.

Nas Figuras B.1, B.2 e B.3 (ver Apêndice B), o comportamento das séries temporais da biomassa, captura e CPUE podem ser analisadas dentro de cada uma das 25 áreas consideradas (cf. Figura 3.2). Pelo padrão temporal destas variáveis, pode-se perceber que as tendências são similares em áreas vizinhas. Isso mostra claramente a dependência espaço-temporal dos dados simulados.

Nas Figuras B.4, B.5, B.6 e B.7 são apresentados mapas com os valores simulados em cada ano e área, de biomassa, esforço, captura e CPUE, respectivamente. De maneira geral, é possível notar o padrão geral de diminuição da biomassa nos períodos intermediários da série, representados pelas áreas mais claras. O mesmo padrão, agora inverso, pode ser notado nos valores de esforço, que são maiores nos

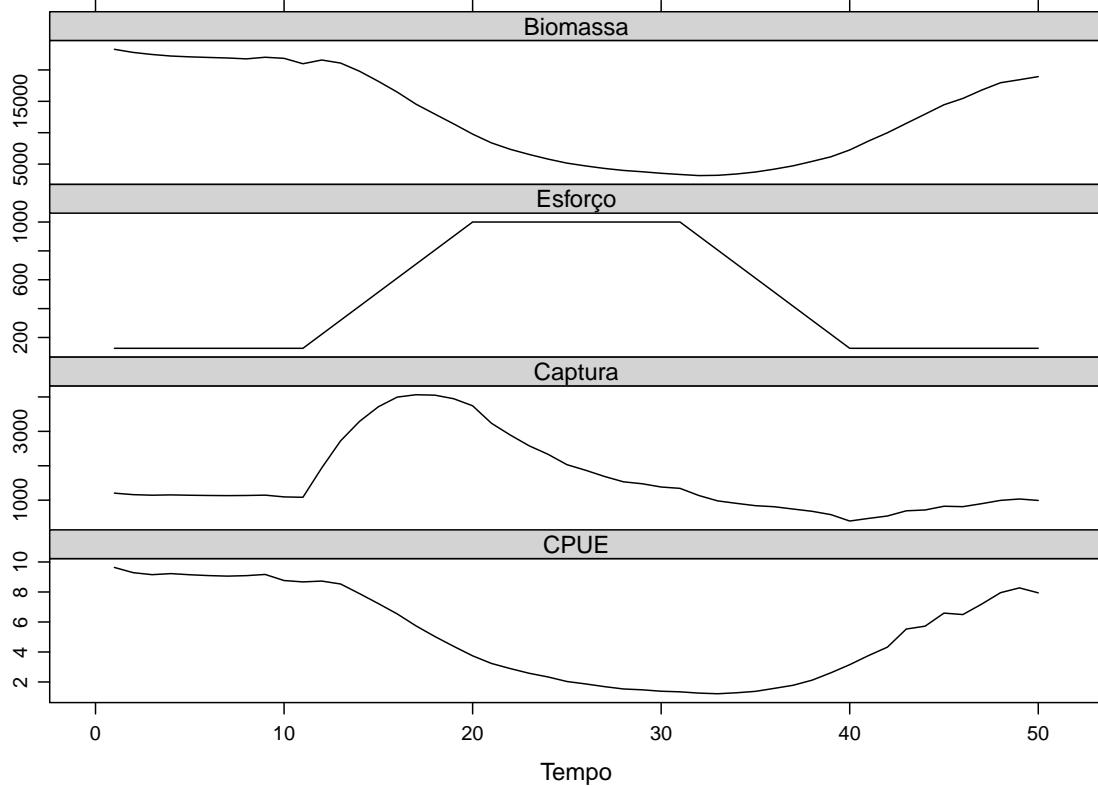


Figura 3.4. Séries temporais (agregadas por área) da biomassa, esforço de pesca, captura e CPUE simuladas.

anos intermediários. Novamente é possível observar o padrão de dependência espaço-temporal em todas as variáveis simuladas. O mapa da CPUE, por exemplo, mostra que áreas mais próximas tendem a ser mais parecidas dentro de cada ano. Além disso, pode ser notado que áreas com CPUEs altas em um ano repetem basicamente o mesmo padrão nos outros anos. Padrões como estes podem ser observados em todos os mapas.

3.4.1.1 Primeiro estágio

Com os dados simulados, foi realizado então o ajuste dos sete diferentes modelos especificados na seção 3.3.3.1. Os resultados das medidas de comparação entre estes modelos está na Tabela 3.5. De acordo com os critérios de informação DIC e WAIC, os modelos que se destacam como tendo um melhor ajuste são aqueles que consideram interações espaço-temporais do tipo II (`m.i2`) e do tipo IV (`m.i4`), embora o número efetivo de parâmetros, de acordo com ambos critérios seja relativamente alto. Entre os dois modelos, aquele que apresenta a maior capacidade de predição, de acordo com o CPO é o modelo mais complexo, com interação espaço-temporal do tipo IV. Portanto, de acordo com estes critérios, este seria o modelo mais adequado para se obterem as previsões a serem utilizadas nos modelos do segundo estágio.

Os gráficos dos resíduos padronizados (Figura 3.5) mostram claramente que os modelos mais simples não são capazes de capturar a dependência temporal dos dados. Isso mostra a necessidade de consideração de estruturas mais complexas. Os modelos que consideraram as interações do tipo II, III, e IV apresentaram os resíduos mais concentrados ao redor de zero. Chama a atenção o padrão de que os resíduos do meio da série temporal são maiores que os demais. Isso pode ser explicado pelo fato de que o

Tabela 3.5. Deviance residual, critério de informação da deviance (DIC), número efetivo de parâmetros (p_D), critério de informação amplamente aplicável (WAIC) e número efetivo de parâmetros (p_{WAIC}), e logaritmo da ordenada preditiva condicional (LCPO) para os diferentes modelos considerados no primeiro estágio.

Modelo	Deviance	DIC	p_D	WAIC	p_{WAIC}	LCPO
m1	3117,57	3120,62	3,05	3121,51	3,92	-1560,75
m.tp	2336,30	2365,05	28,75	2370,39	32,95	-1185,25
m.tnp	1816,40	1857,02	40,61	1858,75	41,03	-929,43
m.i1	1815,51	1857,01	41,50	1858,66	41,79	-929,38
m.i2	-2559,20	-2034,46	524,73	-2038,17	410,21	940,04
m.i3	104,48	930,83	826,35	881,35	579,01	-805,08
m.i4	-3052,95	-2452,69	600,27	-2454,75	461,22	1099,84

esforço é maior nestes períodos, e, por consequência, as capturas devem apresentar maior variabilidade.

3.4.1.2 Segundo estágio

Inicialmente, os modelos descritos na seção 3.3.3.2, tanto para dados agregados quanto para os desagregados, foram ajustados considerando-se os dados brutos. Portanto os próprios valores simulados foram utilizados nos modelos, sem nenhum tipo de tratamento para isso. Esse procedimento tem a finalidade de servir como uma espécie de “controle” para comparação com os mesmos modelos ajustados utilizando-se os dados de captura preditos pelos modelos no primeiro estágio. Dessa forma, o resultado da qualidade do ajuste dos modelos descritos nas tabelas 3.3 e 3.4 para os dados brutos, podem ser visualizados na Tabela C.1 (ver Apêndice C). Através destes resultados é possível notar claramente a grande diferença entre os valores de DIC e WAIC ao comparar os modelos para dados agregados com aqueles para dados desagregados. Estes últimos apresentaram uma maior qualidade de ajuste de acordo com estes critérios. O mesmo padrão pode ser notado também ao se analisar a capacidade preditiva de

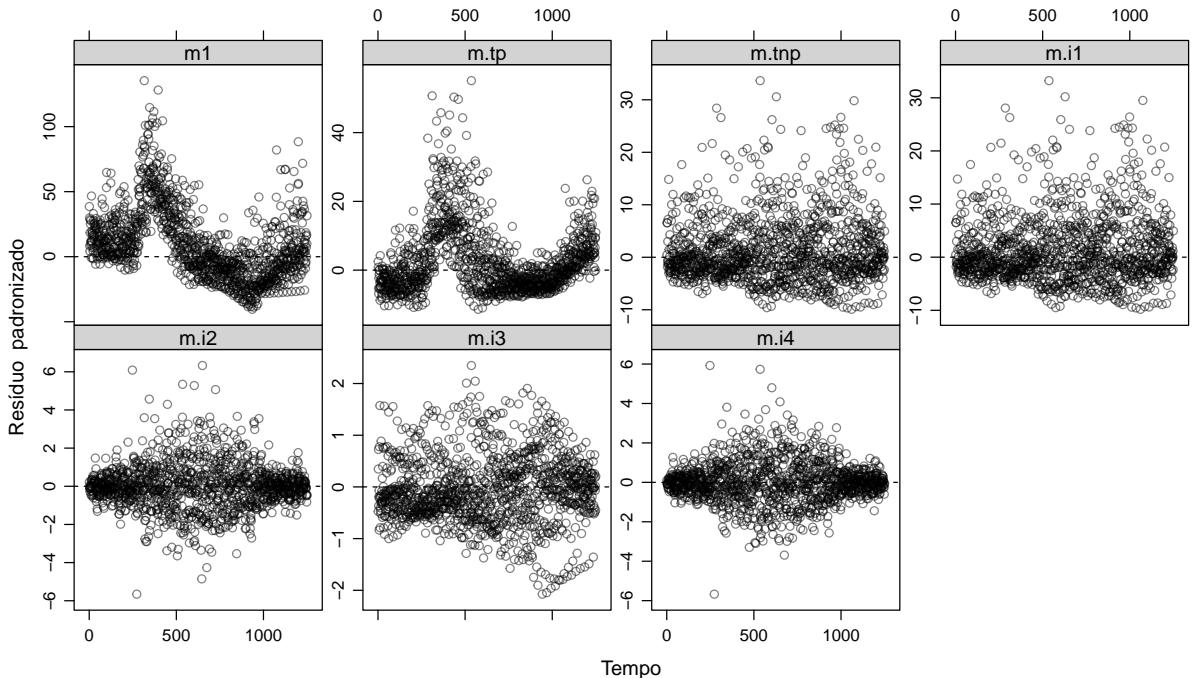


Figura 3.5. Gráficos dos resíduos padronizados para os modelos considerados no primeiro estágio.

cada modelo através do CPO. Os maiores valores de CPO são obtidos ao se utilizar os modelos para dados desagregados. Entre todos os modelos, o que considera interação espaço-temporal em β_0 (**d8**) foi o que apresentou menores valores de DIC e WAIC, e maior valor de CPO.

Analizando-se um resumo da distribuição posterior marginal de cada parâmetro através dos intervalos de credibilidade de 95%, pode-se notar que os dados brutos agregados geram estimativas não viesadas apenas para os parâmetros r e K (Figura 3.6). Chama a atenção o fato de que os intervalos obtidos para o parâmetro q , independente do modelo utilizado, são sempre viesados negativamente, e nenhum deles contém o verdadeiro valor médio simulado (0,01). O modelo mais complexo considerado (**a4**, com efeito RW2 em todos os parâmetros) se destaca por apresentar intervalos de credibilidade muito maiores do que os outros modelos, cobrindo inclusive estimativas negativas dos parâmetros, o que não é realista. De fato, este modelo apresentou o pior desempenho em todos os critérios apresentados na Tabela C.1. A obtenção de estimativas viesadas e até mesmo negativas para estes parâmetros, através do método da regressão linear utilizando-se dados brutos agregados é bem conhecida na literatura (*e.g.* Hilborn, 1979; Uhler, 1980). Esse foi um dos motivos pelo qual este método deixou de ser estudado e utilizado, em troca de métodos mais robustos para modelos não lineares. Hilborn (1979) argumenta que, de fato, estes modelos quando aplicados nestas condições, podem apresentar resultados insatisfatórios. No entanto, este mesmo autor concluiu que se os dados de captura e esforço possuem um bom contraste⁴, as estimativas geradas por esse método serão razoáveis. No entanto, mesmo com dados simulados com um bom contraste (como definido por Hilborn (1979)), a estimativa do parâmetro q obtida neste estudo foi viesada, embora os outros parâmetros (r e K) tenham sido estimados de forma satisfatória.

Por outro lado, quando considerados os dados brutos desagregados, ou seja, com informação da estrutura de dependência espacial, os resultados são claramente mais robustos (Figura 3.7). Nesse caso, considerando os dez diferentes modelos ajustados, os intervalos de credibilidade das posteriores tendem sempre a ficar próximos aos valores médios verdadeiros. Pode-se notar que os modelos mais simples, que consideram nenhum tipo de efeito ou apenas efeitos temporais (modelos **d1** ao **d4**), e os modelos mais complexos, com interação espaço-temporal (modelos **d9** e **d10**) apresentam sempre algum tipo de viés para alguns dos parâmetros. Os modelos intermediários, que consideram apenas estruturas de dependência espacial, ou espaço-temporal apenas no β_0 , foram aqueles que tiveram melhores resultados em termos de vício (não viesados), e que o intervalo de credibilidade inclui o parâmetro médio verdadeiro. De fato, estes também foram os modelos com melhor desempenho nos critérios avaliados na Tabela C.1. Isto mostra que o fato de se considerar os dados espaciais nesse tipo de modelo já provê um grande avanço em termos de qualidade das estimativas, mesmo utilizando-se os dados brutos. No entanto, fica claro também que o simples fato de se utilizar os dados espaciais não garante de imediato o melhor resultado. Como foi mostrado, a consideração da dependência espacial entre as observações é importante e deve ser levada em consideração no ajuste destes modelos.

A Tabela C.2 apresenta os resultados dos critérios para avaliação dos mesmos modelos apresentados nas tabelas para dados agregados (Tabela 3.3) e para dados brutos (Tabela 3.4). No entanto, os dados de CPUE utilizados no modelo linear foram obtidos através da predição de cada um dos sete modelos considerados no primeiro estágio (Tabela 3.2). Independente de qual modelo foi utilizado para se obter a predição, os modelos para dados agregados apresentaram sempre um desempenho inferior quando comparados aos modelos para dados desagregados, considerando os critérios de adequação de ajuste (DIC e WAIC) e de poder de predição (CPO).

Entre os modelos para dados agregados, em geral, o modelo mais complexo (**a4**) apresentou as

⁴Hilborn (1979) declara que as falhas em se obterem boas estimativas através deste método é devido a falta de contraste na série temporal utilizada de captura e esforço, e não pelo fato de que o método possa ser muito simples. O autor define que dados com bom contraste são aqueles que provenientes dos diversos estágios que a população tenha passado, em termos de variação de biomassa. Além disso, os dados também devem ter sido obtidos nas fases de exploração e recuperação de uma pesca.

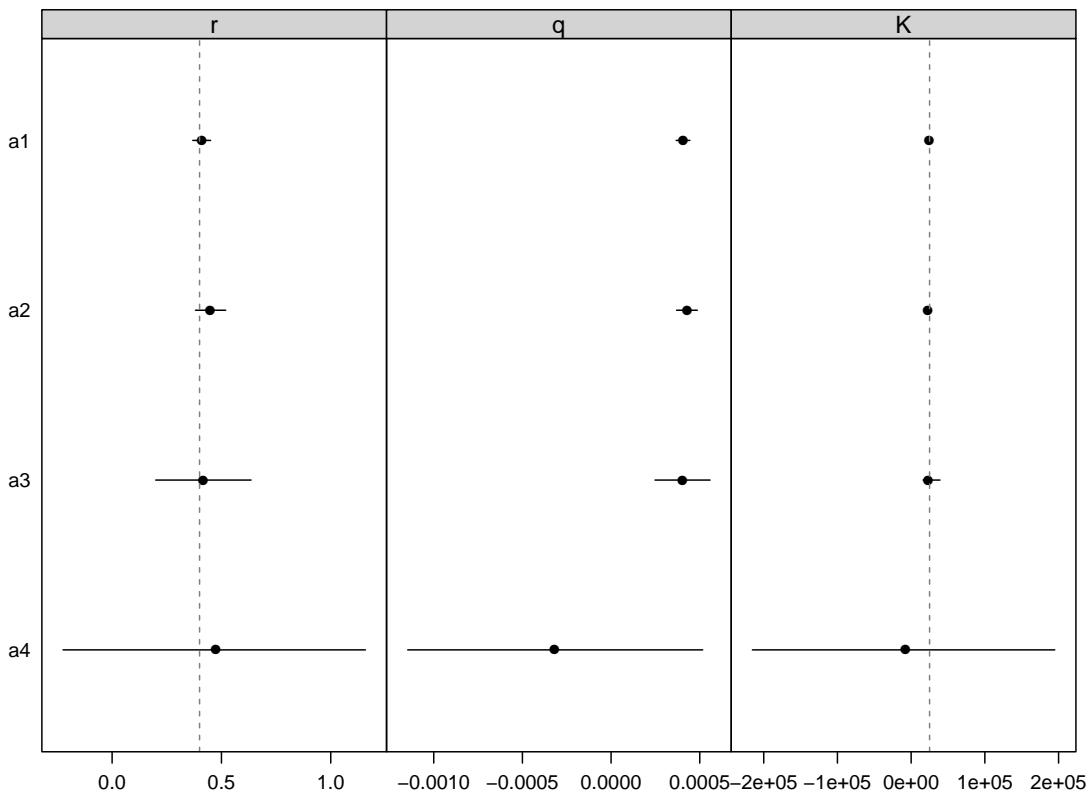


Figura 3.6. Mediana (pontos) e intervalos de credibilidade (95%, linhas escuras horizontais) das distribuições posteriores dos parâmetros para os modelos considerando os dados brutos agregados por ano. A linha pontilhada vertical representa o valor verdadeiro médio.

piores medidas de diagnóstico, independente do modelo utilizado para predição. Esse é o mesmo padrão já observado para os dados brutos agregados. Entre os demais modelos não há um padrão claro sobre qual das previsões poderia gerar melhores resultados. No entanto, quando analisados os intervalos de credibilidade das distribuições posteriores dos parâmetros de todos estes modelos (Figura 3.8), fica claro que aqueles que utilizaram as previsões de captura dos modelos mais simples possuem algum tipo de viés e maior incerteza. Outro fato importante é que, assim como no caso dos dados brutos, as estimativas do parâmetro q para todos os modelos são sempre viesadas negativamente. Isso mostra que, mesmo ao se utilizar dados provenientes de uma padronização prévia, os modelos para dados agregados tendem sempre a fornecer estimativas viesadas para este parâmetro.

Entre os modelos para dados desagregados, aqueles que utilizaram as previsões dos modelos mais simples do primeiro estágio, apresentam em geral valores menores de DIC e WAIC, e maiores de CPO (Tabela C.2). No entanto, ao se analisarem os intervalos de credibilidade das distribuições posteriores dos parâmetros para estes modelos (Figura 3.9), percebe-se claramente que qualquer um dos dez modelos utilizados com as previsões dos modelos $m1$ (somente esforço como covariável) e $m.tp$ (espacial com tendência temporal global) fornecem estimativas sistematicamente viesadas, onde os intervalos de credibilidade não contém nenhum dos parâmetros. É importante mencionar também que os intervalos de alguns modelos não aparecem nessa figura pois apresentaram valores muito extremos e tiveram que ser removidos para não prejudicar a visualização dos outros intervalos. Quando foram utilizadas as previsões dos modelos com interação espaço-temporal dos tipos II, III e IV, os intervalos de credibilidade dos modelos do segundo estágio foram em geral melhores. Considerando estes casos, as melhores estimativas

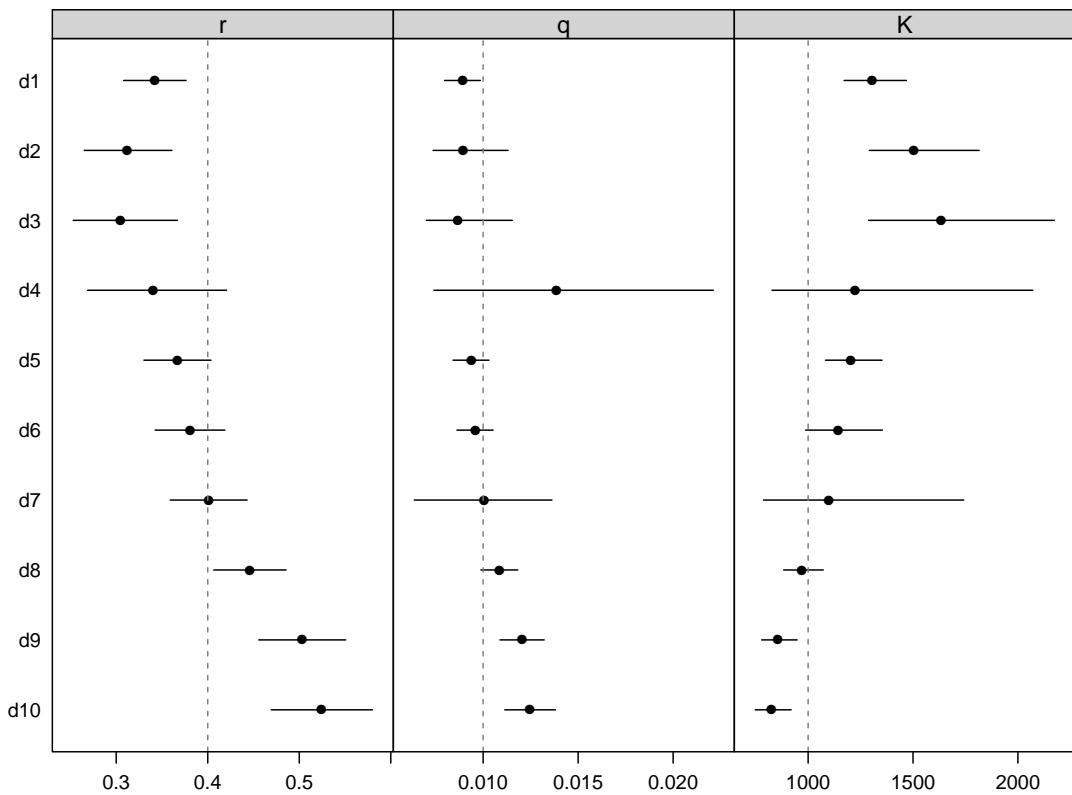


Figura 3.7. Mediana (pontos) e intervalos de credibilidade (95%, linhas escuras horizontais) das distribuições posteriores dos parâmetros para os modelos considerando os dados brutos desagregados. A linha pontilhada vertical representa o valor verdadeiro médio.

(menor viés e maior precisão) dos três parâmetros em geral, são provenientes dos modelos de complexidade intermediária, como aqueles que consideram apenas efeitos aleatórios espaciais nos parâmetros. Os modelos do segundo estágio mais complexos, com interações espaço-temporais, apresentam geralmente menores valores de DIC e WAIC, principalmente quando são utilizadas as previsões dos modelos mais complexos do primeiro estágio (Tabela C.2). No entanto, como pode ser visto pelos intervalos de credibilidade das posteriores destes modelos (Figura 3.9), estes critérios não implicam necessariamente que estes modelos apresentarão boas estimativas para os parâmetros. Nota-se que estes modelos com valores menores de DIC e WAIC apresentam também um maior número de parâmetros efetivos. Isso indica que pode estar ocorrendo um superajuste nestes modelos, ou seja, os valores preditos são praticamente idênticos aos observados.

Quando comparados os intervalos de credibilidade entre modelos que utilizam dados brutos e dados preditos, em geral, aqueles que usam os dados desagregados preditos tendem a ser mais precisos. Mais especificamente, os modelos que consideram efeitos aleatórios espaciais, e que utilizam dados preditos dos modelos mais complexos do primeiro estágio apresentaram uma performance melhor. Já os modelos para dados agregados, independentemente do uso de valores brutos ou preditos, apresentaram um sério problema de viés na estimativa do parâmetro q .

3.4.2 Simulação 2

No segundo cenário de simulação, foram geradas 100 bases de dados diferentes, conforme descrito na seção 3.3.4.2. Para cada base de dados, todo o processo de estimação realizado na simulação 1

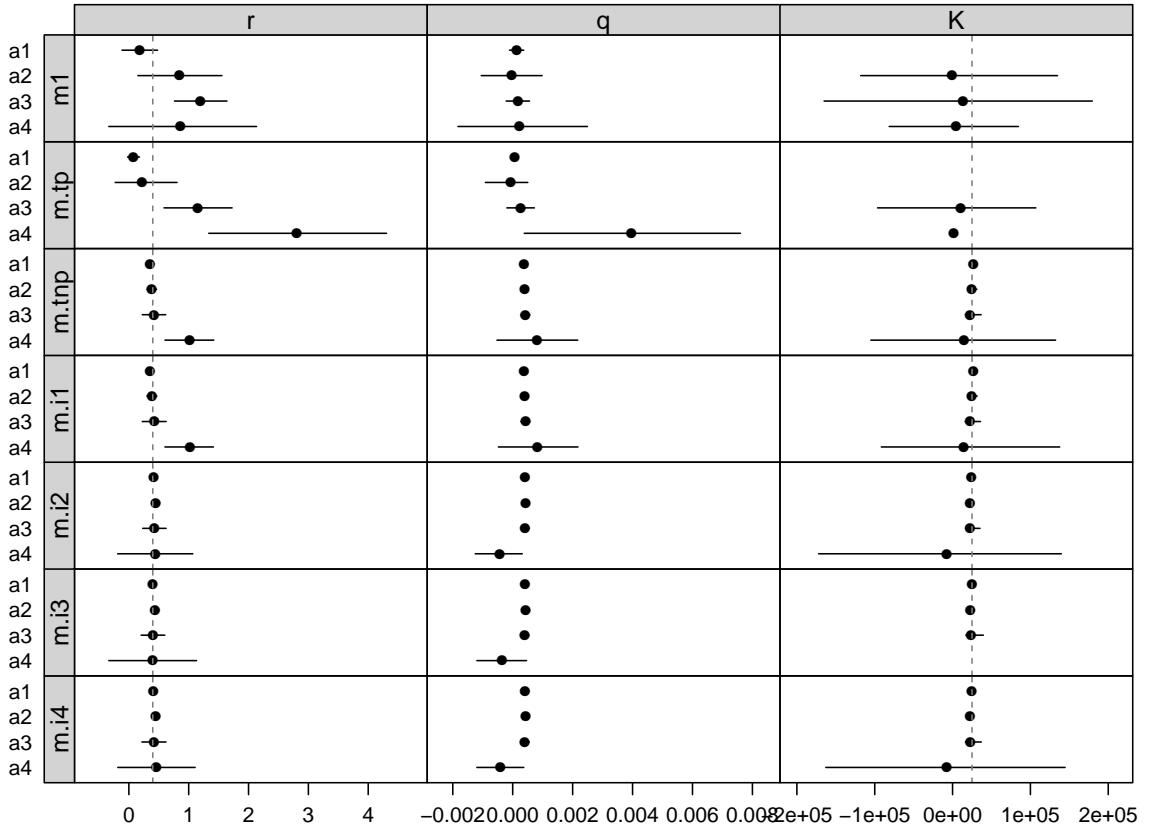


Figura 3.8. Mediana (pontos) e intervalos de credibilidade (95%, linhas escuras horizontais) das distribuições posteriores dos parâmetros para os modelos considerando os dados preditos, agregados por ano, de cada um dos modelos do primeiro estágio. A linha pontilhada vertical representa o valor verdadeiro médio.

foi realizado. No entanto o objetivo agora é verificar a consistência das estimativas, ou seja, diante de diferentes cenários de pesca, verificar se o comportamento observado na primeira simulação (para uma base de dados) irá se repetir.

Assim como antes, os dados brutos também foram utilizados para o ajuste dos modelos. A Figura 3.10 mostra o viés relativo de todos os parâmetros para os modelos que consideraram os dados brutos agregados. O primeiro fato a notar é que as estimativas do parâmetro q são consistentemente viesadas negativamente, com viés próximo a 100% em todos os casos. O modelo mais complexo (a4) forneceu estimativas viesadas também para os outros parâmetros, enquanto que os outros modelos estimaram muito bem r e K . A Figura B.8 mostra as taxas de cobertura dos parâmetros estimados, conforme definido por intervalos de credibilidade de 95%. Nota-se novamente que a distribuição posterior do parâmetro q de todos os modelos não cobre o valor verdadeiro nenhuma vez. Os modelos intermediários possuem taxa de cobertura alta para os parâmetros q e K . O modelo mais complexo, apesar de fornecer estimativa viesada para o parâmetro K , possui uma taxa de cobertura de 100% para esse parâmetro. Isso ocorre provavelmente porque a distribuição posterior deste parâmetro para este modelo é tão ampla, que mesmo que a estimativa pontual seja viesada, a distribuição irá conter o verdadeiro valor (cf. Figura 3.6). No Apêndice C, a Tabela C.3 mostra as estimativas pontuais e outras medidas de viés para todos os parâmetros dos modelos para dados brutos.

Considerando os dados brutos desagregados, nota-se, em geral, que estes modelos fornecem estimativas com viés baixo, e, ao contrário dos modelos para dados agregados, o parâmetro q consegue ser

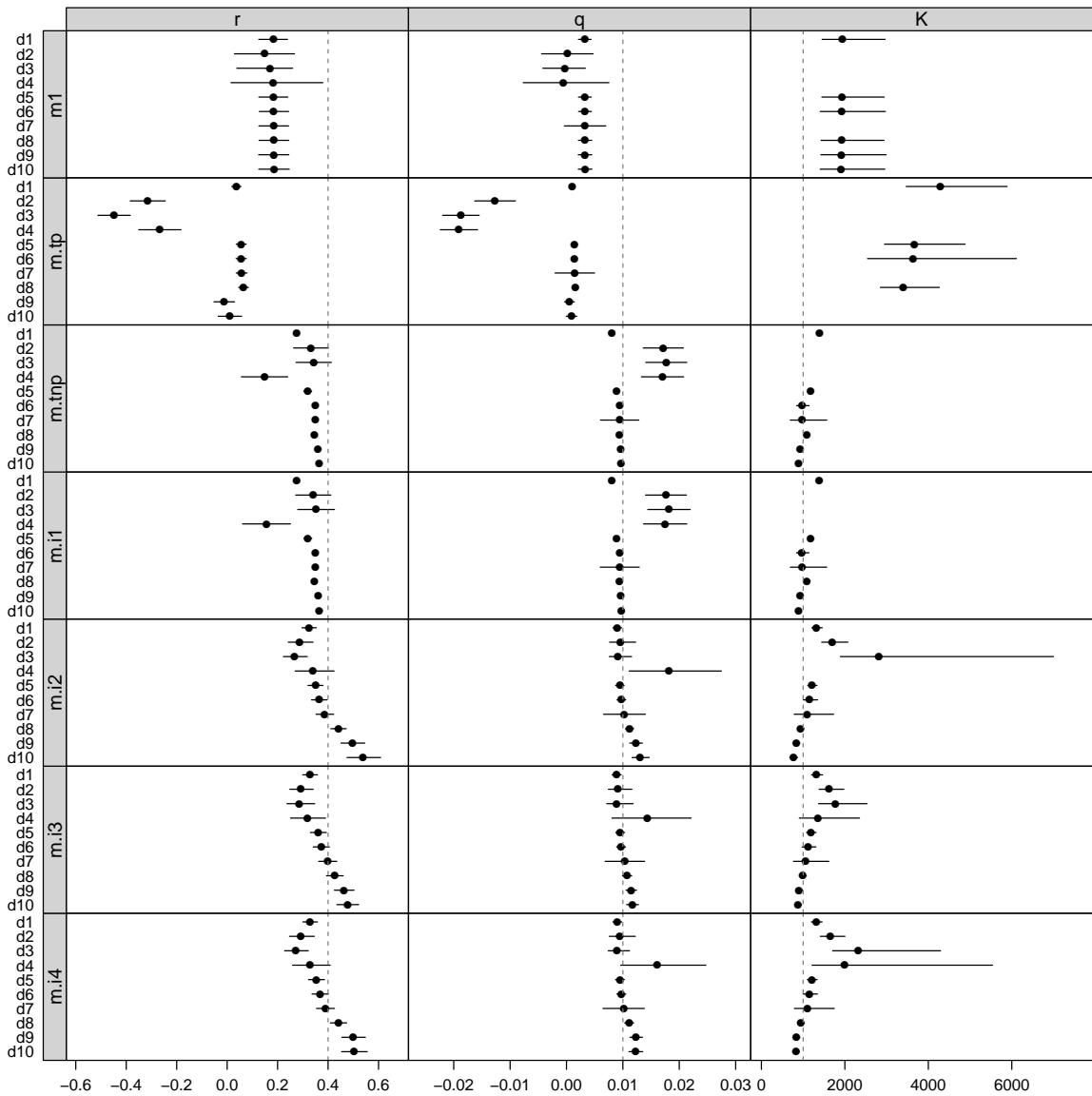


Figura 3.9. Mediana (pontos) e intervalos de credibilidade (95%, linhas escuras horizontais) das distribuições posteriores dos parâmetros para os modelos considerando os dados preditos desagregados, de cada um dos modelos do primeiro estágio. A linha pontilhada vertical representa o valor verdadeiro médio.

estimado próximo ao valor verdadeiro para todos os modelos (Figura 3.11). Aparentemente, os modelos mais complexos fornecem estimativas menos viesadas do que os modelos com estruturas mais simples. Em particular, destacam-se os modelos que consideram de alguma forma apenas estrutura espacial (modelos d5, d6 e d7), e o modelo d8 que admite uma estrutura de interação espaço-temporal. No entanto, ao se analisar a taxa de cobertura dos parâmetros para estes modelos (Figura B.9), percebe-se que apenas os modelos d6 e d7 são consistentes, ou seja, possuem uma taxa de cobertura maior do que 80% para todos os parâmetros. Tanto os modelos mais simples quanto os mais complexos apresentam taxas de cobertura baixa em todos os casos. Novamente percebe-se que a inclusão da estrutura espacial nos modelos de regressão para estimativa dos parâmetros do modelo dinâmico de biomassa é uma alternativa viável. Os resultados mostram que os modelos que consideram a dependência espacial são de fato adequados, e ressaltam a importância de se considerar as observações no espaço. Mesmo ao se utilizarem os dados brutos nos modelos do segundo estágio, as estimativas são precisas quando a estrutura espacial é considerada.

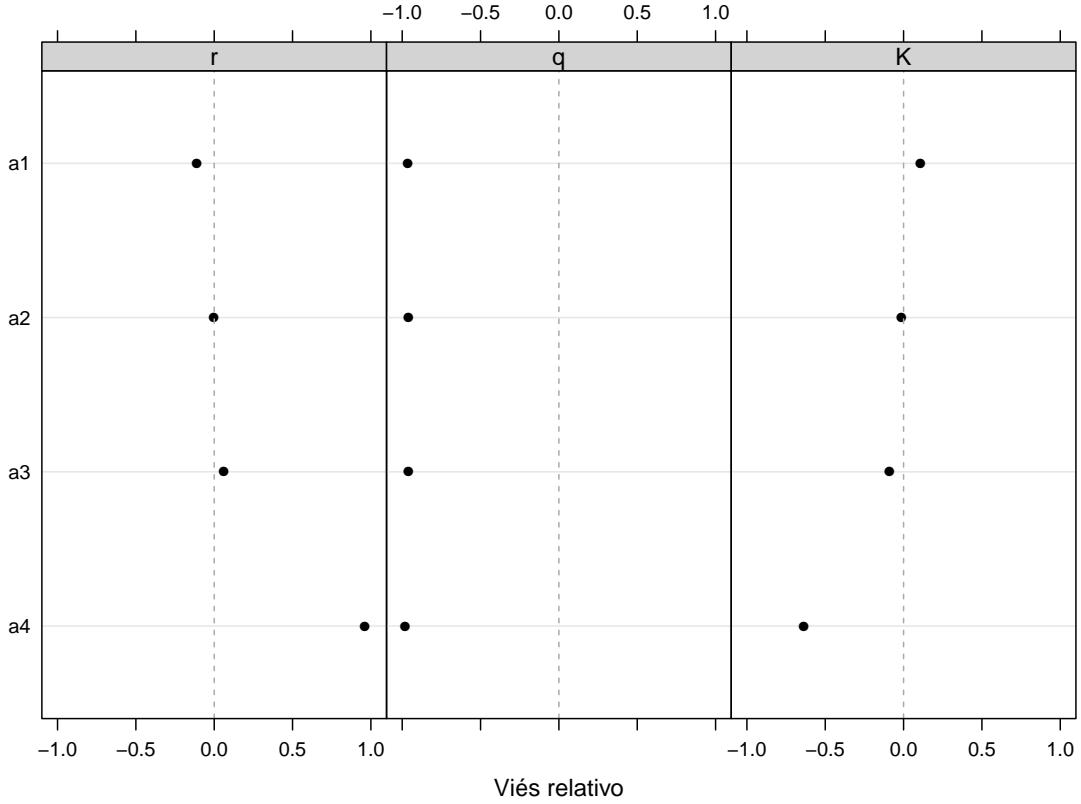


Figura 3.10. Viés relativo das estimativas dos parâmetros por meio dos modelos para dados brutos agregados.

Quando os dados preditos pelos modelos do primeiro estágio são utilizados no ajuste dos modelos do segundo estágio para dados agregados, nota-se novamente que não há nenhuma vantagem em relação à estimativa do parâmetro q (Figura 3.12). Todas as combinações de modelo fornecem estimativas viesadas negativamente. Em relação aos outros parâmetros, quando são utilizadas as previsões dos dois primeiros modelos do primeiro estágio, as estimativas são consistentemente viesadas. Os parâmetros q e K são estimados com menor viés quando são utilizadas as previsões dos modelos mais complexos do primeiro estágio e os modelos intermediários do segundo. No entanto, quando analisada a taxa de cobertura para estes modelos (Figura B.10), nota-se que os modelos intermediários são mais consistentes quando são utilizadas as previsões dos dois modelos mais complexos do primeiro estágio. A alta taxa de cobertura dos modelos que utilizaram as previsões dos modelos mais simples se deve novamente ao fato de que, mesmo a estimativa pontual sendo viesada, a distribuição posterior dos parâmetros é tão ampla que acaba cobrindo também os valores verdadeiros. A Tabela C.4 no Apêndice C apresenta a média e outras medidas de viés para todos os parâmetros estimados com todos os modelos considerados para dados preditos agregados e desagregados.

O viés relativo dos parâmetros estimados com os modelos para dados desagregados, utilizando os valores preditos por cada um dos modelos do primeiro estágio estão na Figura 3.13. Novamente, o primeiro contraste com os modelos para dados agregados é o fato de que agora o parâmetro q é capaz de ser estimado com relativa precisão, pelo menos para a maioria dos modelos. Nota-se o grande viés dos parâmetros quando são utilizadas as previsões dos dois primeiros modelos do primeiro estágio. Conforme a complexidade dos modelos de previsão vai aumentando, melhoram também as estimativas, principalmente para os modelos do segundo estágio intermediários, que consideram a estrutura de dependência

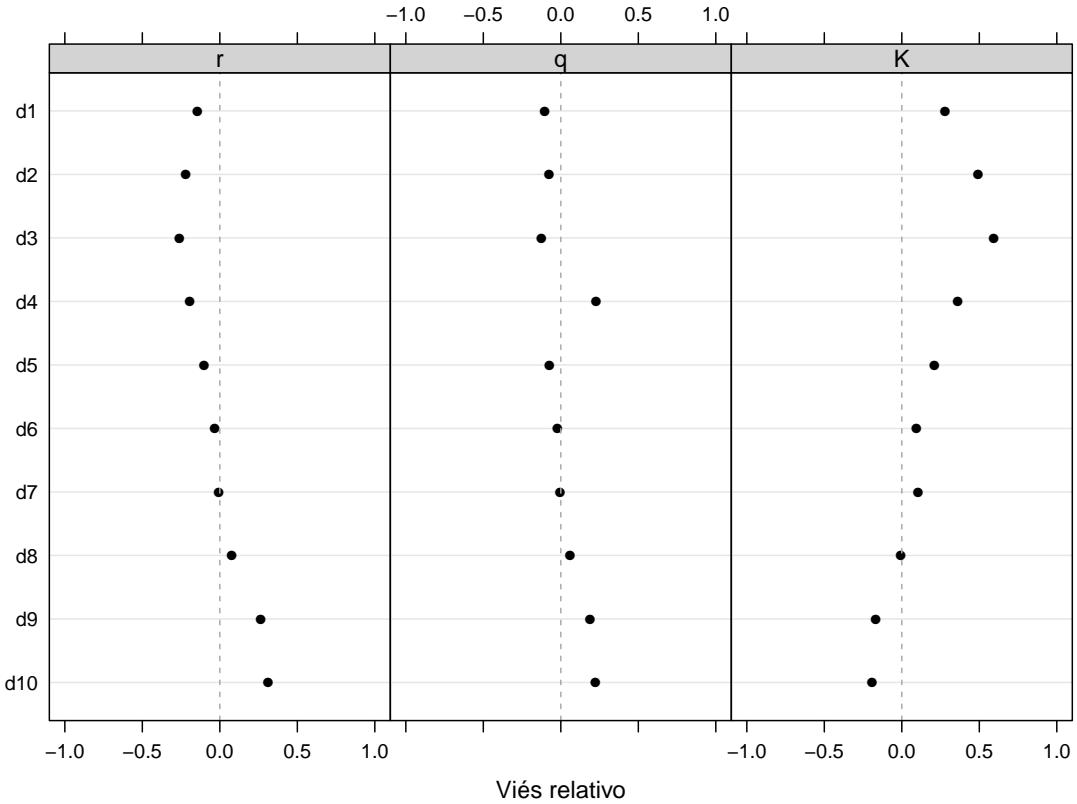


Figura 3.11. Viés relativo das estimativas dos parâmetros por meio dos modelos para dados brutos desagregados.

espacial dos dados. É importante notar que, mesmo quando são utilizadas as previsões destes modelos mais complexos, estimativas viesadas podem ser obtidas, dependendo do modelo considerado no segundo estágio. Isso fica claro pelo padrão de viés crescente para os parâmetros r e q , e decrescente para o parâmetro K , no sentido dos modelos menos complexos para os mais complexos do segundo estágio. Esse fato mostra que, mesmo que seja considerado o modelo de previsão do primeiro estágio com as melhores propriedades (*e.g.* boa adequação aos dados, grande poder de previsão, resíduos aleatórios), se o modelo do segundo estágio não for adequado, as estimativas dos parâmetros tenderão a ser de fato viesadas. Considerando esse fato, e analisando-se a taxa de cobertura dos parâmetros (Figura B.11), nota-se que a combinação de modelos que possui as estimativas menos viesadas e com maior taxa de cobertura são aquelas provenientes dos modelos que consideram a estrutura espacial no segundo estágio, e utilizam os valores preditos pelos modelos com interação do tipo III e IV do primeiro estágio. Mais especificamente, o modelo que considera um efeito aleatório espacial BYM nos parâmetros β_0 , β_1 e β_2 (**d7**) da equação (3.38), e que utiliza as previsões dos modelos **m.i3** (interação do tipo III) ou **m.i4** (interação do tipo IV) do primeiro estágio, é o que fornece as estimativas menos viesadas e mais consistentes (com maior taxa de cobertura). A taxa de cobertura dos parâmetros obtidos com os dois modelos mais simples do primeiro estágio sempre foram próximas de zero. Já a taxa de cobertura para os parâmetros estimados com as previsões dos modelos intermediários geralmente ficou abaixo dos 80%, mesmo que as estimativas pontuais não sejam fortemente viesadas.

Os dois cenários de simulação desenvolvidos apresentaram resultados concordantes. No primeiro cenário chegou-se à conclusão que os modelos considerando os dados agregados sempre fornecem estimativas viesadas, particularmente para o parâmetro q . Esse fato ficou claro quando analisadas as

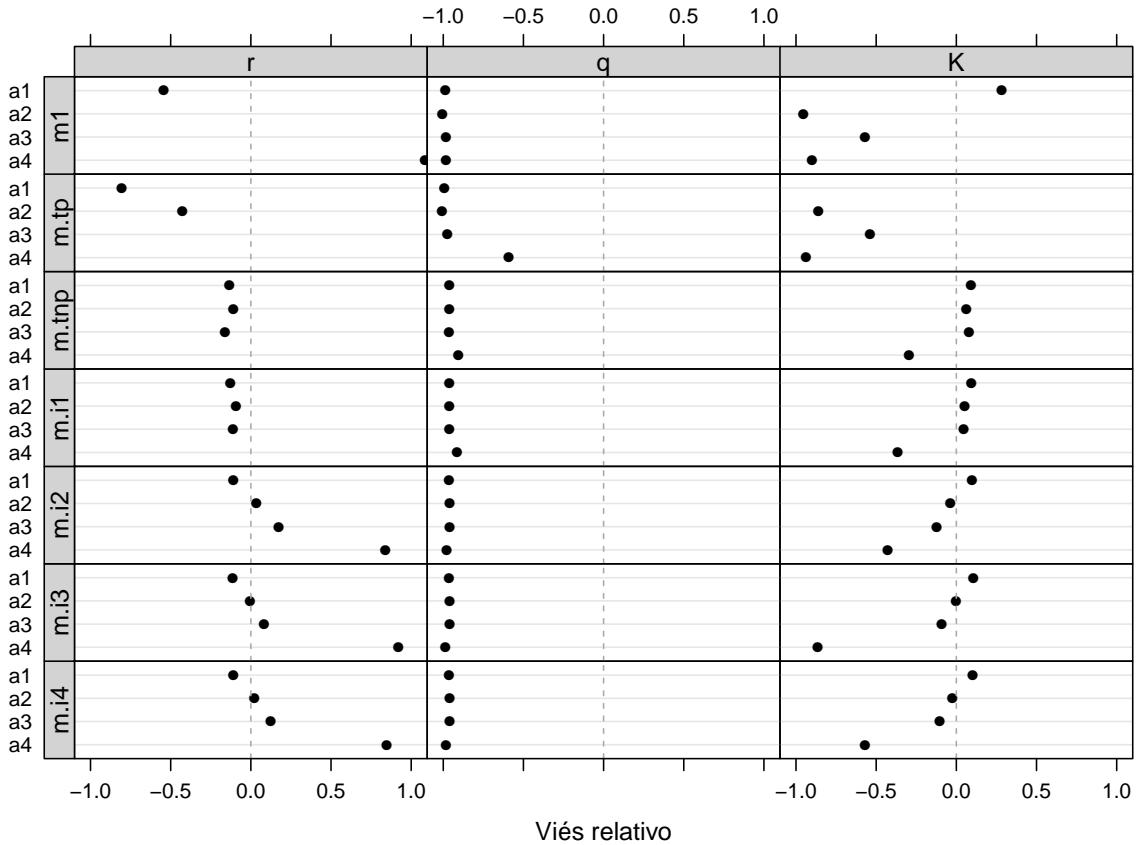


Figura 3.12. Viés relativo das estimativas dos parâmetros por meio dos modelos para dados agregados, considerando a predição de cada um dos modelos do primeiro estágio.

medidas de viés e as taxas de cobertura no segundo cenário simulado. Da mesma forma, em ambos os cenários, mostrou-se que a consideração da estrutura espacial dos dados representa um ganho no sentido de que as estimativas dos parâmetros podem ser obtidas com razoável precisão, mesmo quando os dados brutos (desagregados) são utilizados. No entanto, quando a captura é modelada em um primeiro estágio considerando-se também as interações espaço-temporais, e as previsões desses modelos são utilizadas no segundo estágio (regressão linear), então as estimativas podem ser ainda mais consistentes. No entanto, um modelo adequado, como aquele que considera apenas a dependência espacial dos dados, deve ser utilizado no segundo estágio para que as estimativas sejam de fato não viesadas e consistentes.

Diversos autores salientam a importância de se considerar a estrutura de dependência espacial ao se analisar dados de populações em geral, mas especialmente dados sobre populações provenientes da pesca (Hilborn e Walters, 1992; Quinn e Deriso, 1999; Walters, 2003; Thorson et al., 2015). No entanto, modelos dinâmicos que combinam erros de processo e observacional produzem dificuldades computacionais significativas (Dennis et al., 2006). Por esse motivo, estudos recentes têm tentado alguma forma de linearização destes modelos, com a finalidade de se obterem estimativas mais precisas e de maneira mais rápida. Thorson (2018) salienta que os estudos atuais sobre modelos de manejo pesqueiro devem se concentrar em utilizar mais conceitos estatísticos concretos (como modelos de efeitos aleatórios espaço-temporais), do que se preocupar com métodos que envolvem procedimentos mais *ad hoc*, como a solução de modelos não lineares complexos através de métodos iterativos, por exemplo.

Zhang e Smith (2011) desenvolveram um método de estimação em dois estágios para os parâmetros do modelo dinâmico de biomassa. A ideia dos autores inicialmente é muito similar à apresentada

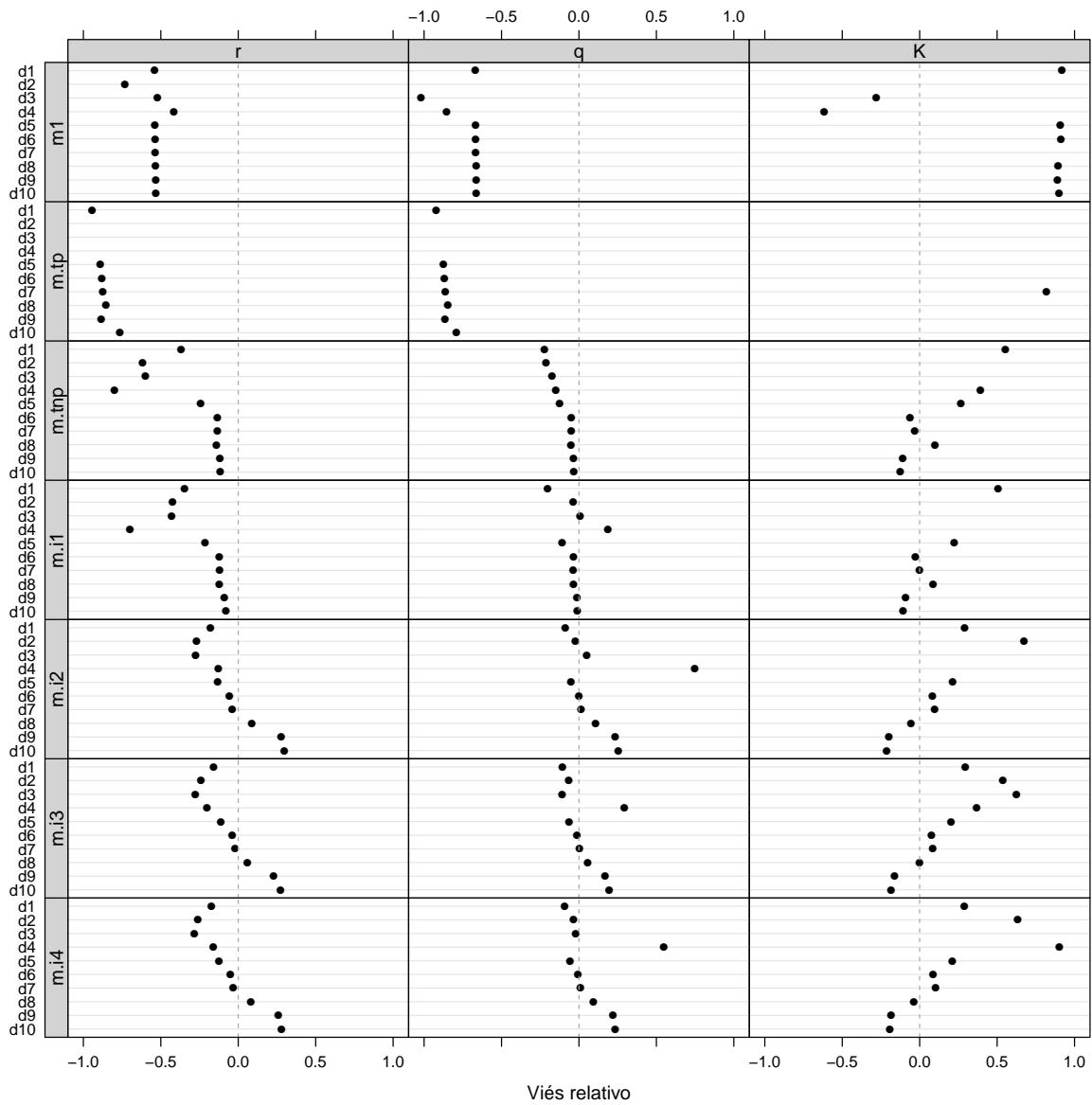


Figura 3.13. Viés relativo das estimativas dos parâmetros por meio dos modelos para dados desagregados, considerando a predição de cada um dos modelos do primeiro estágio.

neste trabalho. Em um primeiro estágio os dados brutos de captura são modelados levando-se em conta diversos fatores como ano, área, e frota. No entanto, o modelo considera que todos estes fatores são variáveis categóricas, e o modelo é similar a uma Análise de Variância (ANOVA) com efeitos fixos. Um índice de ano é extraído desse modelo, e utilizado no segundo estágio em um modelo de regressão similar à equação (3.36), para dados agregados. Através desse método os autores concluíram que as estimativas obtidas por eles era superior ao método tradicional de regressão, usando apenas dados brutos. No entanto, as estimativas fornecidas por essa abordagem se mostraram viesadas, e os autores ainda recomendam que algum procedimento posterior para correção de viés deveria ser adotado (por exemplo o método *Simulation-Extrapolation*, SIMEX). Nessa abordagem, os dados brutos detalhados de uma pesca podem ser utilizados, mas a estrutura de dependência espacial e/ou temporal não é levada em consideração. Além disso, o modelo de regressão considerado no segundo estágio utiliza apenas os dados preditos agregados por ano, então toda a informação detalhada acaba sendo ignorada.

Utilizando uma abordagem diferente, Rankin e Lemos (2015) desenvolveram uma forma linear do modelo de produção de biomassa de Schaefer. Dessa forma, as duas equações do modelo dinâmico de biomassa passam a ser lineares. Os autores utilizaram apenas dados anuais de captura e esforço para ajustar o modelo via MCMC. Os dados foram os mesmos utilizados por diversos outros autores em diferentes contextos, e as estimativas obtidas por este método se mostraram compatíveis quando comparadas com as estimativas obtidas com a forma não linear dos mesmos modelos, também ajustados via MCMC. A vantagem então é que o tempo computacional é drasticamente reduzido quando o modelo de produção é utilizado na forma linear.

Todo esse movimento atual vai, portanto, na direção contrária do que vinha sendo o foco de desenvolvimento em modelos dinâmicos de biomassa, que era o desenvolvimento de algoritmos e métodos de estimação cada vez mais complexos para realizar o ajuste do modelo na forma não linear. Tanto Zhang e Smith (2011), quanto Rankin e Lemos (2015), argumentam que a linearização destes modelos (ou a estimação em dois estágios) tem grande potencial para serem utilizadas e desenvolvidas. De fato, o ponto central da linearização em ambos os trabalhos é justamente visando a viabilização destes modelos para serem utilizados considerando estruturas espaciais, e recomendam que estudos devem ser realizados nesse sentido.

O método de estimação em dois estágios apresentado neste trabalho contempla, portanto, a utilização explícita das estruturas de dependência espaciais e temporais dos dados de captura e esforço para a obtenção de parâmetros populacionais de um estoque. Além disso, a inferência bayesiana permite que a grande incerteza associada aos parâmetros seja resumida na forma das distribuições posteriores. Existe ainda a vantagem de que informações prévias sobre os parâmetros podem ser incorporadas na forma de *prioris*. O método de aproximação via INLA traz ainda a vantagem de que modelos espaço-temporais complexos podem ser ajustados rapidamente, fazendo com que diversos modelos possam ser comparados. Através dos modelos do segundo estágio com estrutura espacial apresentados neste trabalho, além da distribuição posterior dos parâmetros, também é possível se obter os efeitos específicos de área, provenientes da parte aleatória. Por exemplo, a distribuição espacial do parâmetro q (coeficiente de capturabilidade) possui aplicações práticas diretas para o manejo pesqueiro, uma vez que áreas com maiores valores de q podem significar que a população naquela área seja mais vulnerável, e assim, alguma medida de manejo pode ser adotada. A distribuição espacial do parâmetro r também pode ser explorada de maneira similar. Áreas onde a população apresenta taxas de crescimento menor podem ser protegidas da pesca. Outra vantagem dos modelos apresentados aqui visa solucionar um dos pontos críticos apontados por Walters (2003), que é a aplicação de modelos de manejo pesqueiro (como os modelos dinâmicos de biomassa) ignorando áreas ocupadas pelo estoque, mas que não foram exploradas pela pesca. Áreas de onde não se tem informação pesqueira, mas que sabe-se que são ocupadas pelo estoque, podem ser facilmente integradas aos modelos apresentados, como parte natural do processo de estimação.

A partir deste trabalho, diversas linhas de estudo podem ser derivadas. A primeira, e mais óbvia, é a aplicação desta metodologia à dados reais de pesca, fazendo ainda algum tipo de validação. A utilização de dados de captura em número, ao invés de peso como usado aqui, também tem grande potencial para ser explorada, tanto do ponto de vista estatístico (comparando diferentes distribuições para dados de contagem), quanto do ponto de vista de modelagem ecológica (modelos populacionais para a abundância em número). Embora diversas estruturas espaciais e temporais tenham sido consideradas, existem ainda outras alternativas que também podem ser estudadas. Uma possibilidade imediata seria considerar também nos modelos um efeito sazonal, permitindo a utilização de dados ainda mais informativos sobre a tendência temporal. Em um nível de detalhamento ainda maior, poderiam ser considerados modelos para microdados no primeiro estágio, ou seja, dados na forma mais bruta, considerando, por exemplo, capturas por embarcação. No entanto, devido ao formato diferencial entre observações consecutivas no segundo estágio, alguma agregação teria que ser feita neste caso. Ao utilizar microdados, ainda

existe a possibilidade de se explorar a modelagem considerando distribuições para dados de contagem, possivelmente com sobredispersão e excesso de zeros, como é comum na maioria dos casos.

3.5 Conclusões

Inicialmente considera-se que o processo de simulação desenvolvido tenha sido satisfatório para o objetivo do trabalho, e poderá ser utilizado para estudos futuros. A simulação de dados a partir de modelos populacionais, por mais simples que sejam, já é um desafio por si só, uma vez que estes modelos podem apresentar um comportamento caótico, e são muito sensíveis dependendo da combinação de parâmetros utilizada (May, 1976). A inclusão da estrutura espacial neste tipo de modelo de dinâmica populacional é ainda mais desafiadora pois são muitos efeitos que devem ser controlados ao mesmo tempo. Por isso, conclui-se que o método de simulação desenvolvido poderá servir para o desenvolvimento contínuo de modelos que consideram a dinâmica espaço-temporal de populações exploradas pela pesca.

O método da regressão linear para a estimativa dos parâmetros do modelo dinâmico de biomassa deixou de ser utilizado e estudado quando alguns trabalhos apontaram que as estimativas poderiam ser viesadas, e também que a aparente simplicidade do modelo não seria capaz de considerar toda a complexidade que envolve os modelos populacionais. De fato, foi mostrado que quando se utilizam os dados de captura e esforço anuais (agregados por área), o método da regressão forneceu estimativas viesadas para o parâmetro q , embora os parâmetros r e K tenham sido estimados razoavelmente bem para estes casos. Mesmo quando estruturas de dependência temporal foram consideradas para estes dados agregados, não houve ganho na diminuição do viés. No entanto, pôde ser mostrado que o fato de se considerarem os dados espaciais, e não apenas anuais, apresenta um grande avanço no sentido de se obterem estimativas não viesadas e consistentes através do método da regressão. Apesar disso, ficou claro também que apenas utilizar os dados espaciais sem considerar a estrutura de dependência entre eles, não é uma abordagem satisfatória. O ideal então é adotar algum tipo de estrutura que permita modelar a dependência espacial, naturalmente observada em dados pesqueiros. Mesmo quando foram considerados os dados brutos, o método da regressão com dependência espacial já apresentou resultados bastante razoáveis. Quando considerado ainda a modelagem dos dados brutos em um primeiro estágio, e o subsequente uso dos valores preditos no segundo estágio da regressão, os resultados mostraram-se ainda mais precisos.

Além disso, o uso de um método aproximado para inferência bayesiana, como é o caso do INLA utilizado aqui, permite com que diversos modelos possam ser ajustados em questão de segundos ou minutos, o que é o oposto dos métodos de inferência bayesiana baseados em simulação, onde mesmo modelos simples podem levar horas para convergir. Isso permite com que o desenvolvimento de novas abordagens que envolvam inferência bayesiana em modelos populacionais complexos seja feito de forma mais rápida e direta. Salienta-se ainda que poucos softwares possuem a capacidade de ajuste direto de modelos espaço-temporais complexos como os apresentados aqui.

Por fim, conclui-se então que o uso do método de dois estágios para estimativa de parâmetros de modelos populacionais complexos pode ser utilizado, desde que se considere a estrutura espacial dos dados.

Referências

- Angelini, R. e Moloney, C. L. (2007). Fisheries, ecology and modelling: an historical perspective. *Pan-American Journal of Aquatic Sciences*, 2(2):75–85.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society. Series B*, 36(2):192–225.

- Besag, J., York, J., e Mollie, A. (1991). Bayesian image-restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.
- Blangiardo, M. e Cameletti, M. (2015). *Spatial and spatio-temporal bayesian models with R-INLA*. John Wiley & Sons, Ltd., Chichester, UK.
- Booth, A. J. (2000). Incorporating the spatial component of fisheries data into stock assessment models. *ICES Journal of Marine Science*, 57:858–865.
- Carlin, B. P. e Louis, T. A. (2009). *Bayesian methods for data analysis*. Chapman & Hall/CRC, Boca Raton.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., e Riddell, A. (2017). Stan: a probabilistic programming language. *Journal of Statistical Software*, 76(1).
- Carroll, R. J., Ruppert, D., Stefanski, L. A., e Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*. Chapman & Hall/CRC, Boca Raton.
- Carruthers, T. R., McAllister, M. K., e Taylor, N. G. (2011). Spatial surplus production modeling of Atlantic tunas and billfish. *Ecological Applications*, 21(7):2734–2755.
- de Valpine, P. (2002). Review of methods for fitting time-series models with process and observation error and likelihood calculations for nonlinear, non-Gaussian state-space models. *Bulletin of Marine Science*, 70(2):455–471.
- de Valpine, P. (2004). Monte Carlo state-space likelihoods by weighted posterior kernel density estimation. *Journal of the American Statistical Association*, 99(466):523–536.
- de Valpine, P. e Hastings, A. (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs*, 72(1):57–76.
- de Valpine, P. e Hilborn, R. (2005). State-space likelihoods for nonlinear fisheries time-series. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(9):1937–1952.
- Dennis, B., Ponciano, J., e Lele, S. (2006). Estimation density dependence, process noise, and observation error. *Ecological Monographs*, 76(3):323–341.
- Fahrmeir, L. e Tutz, G. (1994). *Multivariate statistical modelling based on Generalized Linear Models*. Springer Series in Statistics. Springer, New York.
- Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., e Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.
- Fox, W. W. (1970). An exponential surplus-yield model for optimizing exploited fish populations. *Transactions of the American Fisheries Society*, 99:80–88.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., e Rubin, D. B. (2014). *Bayesian data analysis*. Chapman and Hall/CRC.
- Goethel, D. R., Quinn, T. J., e Cadrin, S. X. (2011). Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. *Reviews in Fisheries Science*, 19(2):119–136.
- Gulland, J. A. (1983). *Fish stock assessment: a manual of basic methods*. John Wiley & Sons.

- Haddon, M. (2001). *Modelling and quantitative methods in fisheries*. Chapman & Hall, Boca Raton.
- Harden-Jones, F. R. (1968). *Fish migration*. Edward Arnold, London.
- Hilborn, R. (1979). Comparison of fisheries control systems that utilize catch and effort data. *Journal of the Fisheries Research Board of Canada*, 36(12):1477–1489.
- Hilborn, R. (1992). Current and future trends in fisheries stock assessment and management. *South African Journal of Marine Science*, 12:975–988.
- Hilborn, R. e Walters, C. J. (1992). *Quantitative fisheries stock assessment: choice, dynamics & uncertainty*. Chapman & Hall, New York.
- Jennings, S., Kaiser, M. J., e Reynolds, J. D. (2001). *Marine fisheries ecology*. Blackwell Publishing, Malden.
- Jiao, Y., O'Reilly, R., Smith, E., e Orth, D. (2016). Integrating spatial synchrony/asynchrony of population distribution into stock assessment models: a spatial hierarchical Bayesian statistical catch-at-age approach. *ICES Journal of Marine Science*, 76(4):1–14.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:34–45.
- Kinas, P. G. (1996). Bayesian fishery stock assessment and decision making using adaptive importance sampling. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(APRIL 2011):414–423.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling on nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1063.
- Knorr-Held, L. (2000). Bayesian modelling of inseperable space-time variation in disease risk. *Statistics in Medicine*, 19:2555–2567.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., e Bell, B. M. (2016). TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5):1–21.
- Larkin, P. A. (1977). An epitaph for the concept of maximum sustained yield. *Transactions of the American Fisheries Society*, 106(1):1–11.
- Lindgren, F. e Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25.
- Ludwig, D. e Walters, C. J. (1981). Measurement errors and uncertainty in parameter estimates for stock and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6):711–720.
- Ludwig, D. e Walters, C. J. (1985). Are age-structured models appropriate for catch-effort data? *Canadian Journal of Fisheries and Aquatic Sciences*, 42(6):1066–1072.
- Ludwig, D. e Walters, C. J. (1989). A robust method for parameter estimation from catch and effort data. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(1):137–144.
- Ludwig, D., Walters, C. J., e Cooke, J. G. (1988). Comparison of two models and two estimation methods for catch and effort data. *Natural Resources Modeling*, 2:457–498.
- Lunn, D., Spiegelhalter, D., Thomas, A., e Best, N. (2009). The BUGS project: evolution, critique and future directions. *Statistics in Medicine*, 28(25):3049–3067.

- May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459–467.
- Meinhold, R. J. e Singpurwalla, N. D. (1983). Understanding the Kalman filter. *The American Statistician*, 37(2):123–127.
- Meyer, R. e Millar, R. B. (1999). Bayesian stock assessment using a state-space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences*, 56(1):37–52.
- Millar, R. B. e Meyer, R. (2000b). Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *Journal of the Royal Statistical Society Series C Applied Statistics*, 49(3):327–342.
- Pella, J. J. e Tomlinson, P. K. (1969). A generalized stock production model. *Bulletin of the Inter-American Tropical Tuna Commission*, 13:421–458.
- Plummer, M. (2003). JAGS: a program for analysis of bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, pages March 20–22, Vienna, Austria. ISSN 1609–395X.
- Polacheck, T., Hilborn, R., e Punt, A. E. (1993). Fitting surplus production models: comparing methods and measuring uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(12):2597–2607.
- Punt, A. (1995). The performance of a production-model management procedure. *Fisheries Research*, 21(3-4):349–374.
- Punt, A. E. (2003). Extending production models to include process error in the population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(10):1217–1228.
- Quinn, T. J. e Deriso, R. B. (1999). *Quantitative fish dynamics*. Oxford University Press, Oxford.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rankin, P. S. e Lemos, R. T. (2015). An alternative surplus production model. *Ecological Modelling*, 313:109–126.
- Reed, W. J. e Simons, C. M. (1996). Analyzing catch-effort data by means of the Kalman filter. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(10):2157–2166.
- Rue, H. e Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman & Hall/CRC, Boca Raton.
- Rue, H., Martino, S., e Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- Ruiz-Cárdenas, R., Krainski, E. T., e Rue, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations - INLA. *Computational Statistics & Data Analysis*, 56(6):1808–1828.
- Sant’Ana, R., Kinas, P. G., Miranda, L. V. D., Schwingel, P. R., Castello, J. P., e Vieira, J. P. (2017). Bayesian state-space models with multiple CPUE data: the case of a mullet fishery. *Scientia Marina*, 81(3).
- Schaefer, M. B. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bulletin of the Inter-American Tropical Tuna Commission*, 1(2):25–56.

- Schnute, J. (1977). Improved estimates from the Schaefer production model: theoretical considerations. *Journal of the Fisheries Research Board of Canada*, 34(5):583–603.
- Schnute, J. T. (1994). A general framework for developing sequential fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(8):1676–1688.
- Schrödle, B. e Held, L. (2011). Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6):725–734.
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., e Sørbye, S. H. (2017). Penalising model component complexity: a principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.
- Soutar, A. e Isaacs, J. D. (1969). History of fish populations inferred from fish scales in Anaerobic sediments off California. *Calif. Mar. Res. Comm. CalCOFI*, 13(63-70):8.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., e Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.
- Sullivan, P. J. (1992). A Kalman filter approach to catch-at-length analysis. *Biometrics*, 48(1):237–257.
- Thorson, J. T. (2018). Perspective: let's simplify stock assessment by replacing tuning algorithms with statistics. *Fisheries Research*, (No prelo).
- Thorson, J. T., Skaug, H. J., Kristensen, K., Shelton, A. O., Ward, E. J., Harms, J. H., e Benante, J. A. (2015). The importance of spatial models for estimating the strength of density dependence. *Ecology*, 96(5):1202–1212.
- Uhler, R. S. (1980). Least squares regression estimates of the Schaefer production model: some monte carlo simulation results. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(8):1284–1294.
- Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(12):1433–1436.
- Walters, C. J. e Hilborn, R. (1976). Adaptive control of fishing systems. *Journal of the Fisheries Research Board of Canada*, 33(1):145–159.
- Watanabe, S. (2012). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, 14:867–897.
- Zhang, J. e Smith, M. D. (2011). Estimation of a generalized fishery model: a two-stage approach. *Review of Economics and Statistics*, 93(2):690–699.

4 MODELOS MULTIVARIADOS DE COVARIÂNCIA LINEAR GENERALIZADA (McGLM) PARA A AVALIAÇÃO DO EFEITO DE HOOKPODS EM PESCARIAS COMERCIAIS

Resumo

A captura de aves em pescarias comerciais gera prejuízos tanto ambientais, como econômicos. Um dispositivo chamado *hookpod* foi desenvolvido para ser uma alternativa aos chumbos normais, e com a intenção de ajudar a diminuir esse tipo de captura incidental. A eficácia dos *hookpods* já é comprovada na diminuição das capturas de aves. No entanto, ainda existiam dúvidas se este dispositivo não poderia diminuir também a captura de espécies de peixes alvo da pescaria, como os atuns. Neste trabalho foram analisados dados de um experimento onde foram utilizados chumbos normais e *hookpods*, com a intenção de se avaliar se o uso do dispositivo pode interferir nas capturas das espécies alvo. As capturas de duas das principais espécies de atuns foram utilizadas como resposta, e alguns outros fatores, além do tratamento (*hookpods*), como variáveis explicativas. Para modelar essa resposta bivariada, foram utilizados os Modelos de Covariância Linear Generalizada (McGLM). Essa classe de modelos permite uma série de flexibilizações em relação aos GLMs tradicionais, como lidar com dados não independentes, a modelagem de múltiplas respostas, e a seleção de modelos automática. A análise dos dados através desta abordagem permitiu verificar que os *hookpods* não interferem na captura das espécies alvo, tanto conjuntamente, quanto individualmente.

Palavras-chave: Captura incidental; Albacora-lage; Albacora-branca; Modelos multivariados; Modelos de Covariância Linear Generalizada

4.1 Introdução

A coleta de dados experimentais ou observacionais geralmente é composta de uma variável resposta e uma ou mais variáveis explicativas. No entanto, frequentemente são observadas mais de uma variável resposta, que podem ou não ser correlacionadas. No caso univariado, os métodos estatísticos para análise são bem estabelecidos. Já no caso multivariado, pode ocorrer uma das duas opções: (a) análise separada de cada variável resposta, ou (b) análise por métodos estatísticos multivariados. No caso da análise individual de várias respostas, o problema é que nem sempre existem evidências suficientes individualmente, e além disso, as respostas podem ter influências umas às outras, o que não está sendo levado em consideração. Quando se opta por uma análise multivariada, os métodos disponíveis são bem mais restritos, e muitas vezes difíceis de serem determinados para alguns problemas.

Quando diversas variáveis resposta são medidas, a análise multivariada é de fato mais apropriada, pois se existe correlação entre as variáveis, essa informação deve ser explorada. Os modelos multivariados são mais poderosos em detectar efeitos de fatores quando as respostas são correlacionadas. Além disso, pode-se obter uma descrição dos resultados mais compacta e uma conclusão mais global. No entanto, nem sempre é óbvio como fazer uma análise multivariada ou mesmo interpretar seus resultados. Além disso, a maioria dos métodos são baseados na suposição de que as respostas seguem uma distribuição normal multivariada.

A abordagem padrão quando se deseja avaliar a influência de um ou mais fatores em uma resposta é a de modelos lineares. No caso multivariado, um modelo linear multivariado pode ser empregado de maneira similar. Em ambos os casos, quando se deseja fazer inferência sobre os parâmetros estimados, as suposições são as mesmas. Estas suposições são de que as observações são provenientes de uma distribuição normal com variância constante σ^2 no caso univariado, e de que as respostas no caso multivariado

seguem uma distribuição normal multivariada com matriz de covariância (simétrica e positiva definida) Σ , onde as entradas fora da diagonal principal indicam a covariância (ou correlação) entre as respostas (Johnson e Wichern, 2008).

No caso univariado, os Modelos Lineares Generalizados (GLM) (Nelder e Wedderburn, 1972) são uma forma de flexibilizar tanto a suposição de normalidade quanto a de variância constante das observações. Nessa abordagem, dados de contagem e proporções, por exemplo, podem ser modelados com distribuições apropriadas, eliminando a suposição de normalidade. Já no caso multivariado, até então não existia uma abordagem similar nesse sentido.

Atualmente a análise multivariada de dados não normais envolve uma considerável gama de abordagens, indo de Equações de Estimação Generalizadas (GEE) (Zeger et al., 1988) à modelos lineares generalizados mistos (GLMM). No entanto, cada abordagem permite modelar apenas tipos específicos de dependência ou correlação, sem pertencer à uma abordagem mais geral. Além disso, os *softwares* disponíveis atualmente apresentam uma capacidade limitada em termos de complexidade e dimensão dos dados.

Com a intenção de obter uma abordagem tão flexível quanto os GLMs para o caso univariado, Bonat e Jørgensen (2016) propuseram uma classe de modelos chamada de Modelos Multivariados de Covariância Linear Generalizada (McGLM). Segundo a ideia de Pourahmadi (1999), estes modelos são especificados por funções de ligação e preditores lineares para um vetor de médias e matriz de covariância respectivamente. Essa metodologia permite uma abordagem unificada para dados multivariados correlacionados, levando em consideração respostas com diferentes distribuições e diferentes estruturas de covariância. Os modelos são ajustados por meio de quase-verossimilhança e funções de estimativa de Pearson, baseados apenas em suposições de segundo momento.

Neste trabalho, os McGLMs serão utilizados para a análise de dados de um experimento para avaliar se a utilização de um artefato de proteção de captura incidental de aves (*hookpods*) interfere na captura das espécies-alvo de uma pescaria. Em pescarias de espinhel de superfície, a captura incidental de aves causa sérios problemas, tanto ambientais (como a morte dos animais), quanto econômicos (com a inutilização de diversos anzóis). Nesse sentido, os *hookpods* foram desenvolvidos para serem acoplados aos anzóis, evitando a captura das aves, que geralmente ocorre até os primeiros 10 metros de profundidade. Ao passar de 10-15 metros, o dispositivo se solta do anzol e permanece conectado à linha funcionando e substituindo chumbos normais. A eficiência para evitar a captura das aves já é comprovada (Sullivan et al., 2018), portanto o objetivo deste trabalho é então avaliar se o uso deste dispositivo pode interferir na captura das espécies de interesse comercial, como os atuns e agulhões.

4.2 Metodologia

4.2.1 Dados

Os dados foram obtidos por meio de observadores de bordo do Projeto Albatroz. Entre os meses de junho e outubro de 2013, os observadores participaram de seis viagens de pesca em barcos espinheiros da frota nacional. Em cada lançamento do espinhel, uma parte dos anzóis era lançada com o chumbo normal (controle) e outra parte com os *hookpods* (tratamento) (ver Figura 4.1). O peso do próprio *hookpod* serve como substituto do chumbo normal. Para maiores detalhes sobre o funcionamento deste dispositivo, veja Sullivan et al. (2018). Diversas variáveis foram também coletadas, como profundidade, direção do vento e temperatura da água. As variáveis resposta são então a captura das diversas espécies de peixes, principalmente atuns, agulhões e tubarões. Todas as respostas são em forma de contagens do número de peixes capturados com um determinado número de anzóis.

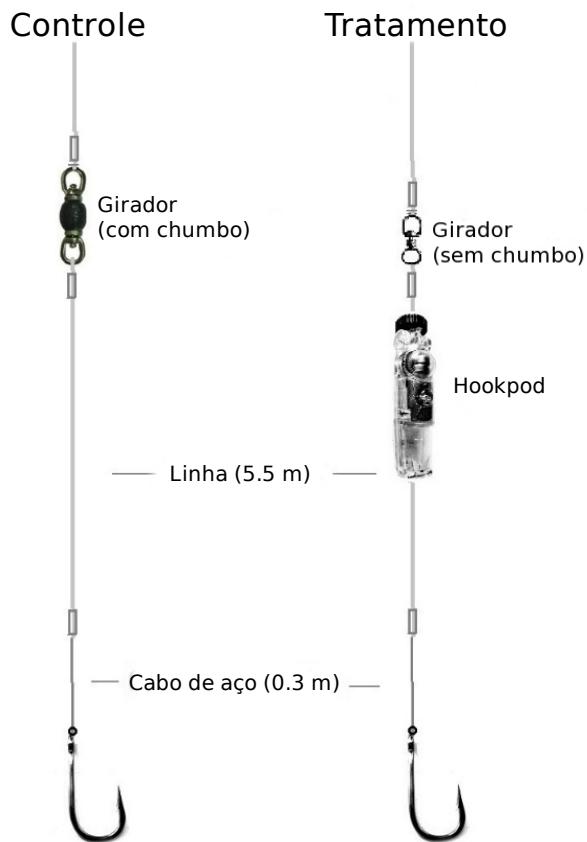


Figura 4.1. Figura esquemática mostrando as linhas do espinhel utilizadas no experimento. No controle foi utilizado o girador padrão com chumbo, enquanto que nas linhas com o tratamento, foi utilizado o *hookpod* e um girador sem chumbo. (Imagem adaptada de D. Gianuca, Projeto Albatroz).

De particular interesse nesse trabalho são as espécies de atuns, que possuem grande valor comercial e são destinadas quase que em sua totalidade para exportação. Das quatro espécies de atuns capturadas, apenas duas possuíam dados suficientes para análise. Portanto, aqui foram consideradas como respostas a captura da albacora-lage (*Thunnus albacares*) e da albacora-branca (*Thunnus alalunga*). Os diversos fatores coletados foram utilizados como variáveis explicativas, assim como o tratamento (uso de chumbo ou *hookpod*). A descrição de todas as variáveis disponíveis pode ser consultada na Tabela 4.1.

Tabela 4.1. Descrição das variáveis utilizadas no estudo.

Tipo	Variável	Descrição
Explicativa	periodo	Período do dia em que foram lançados os anzóis.
	profundidade	Profundidade do local de lançamento (metros).
	beauford	Escala Beauford de vento (varia de 0 a 12, com 0 sendo mar calmo, e 12 sendo furacão).
	dir.vento	Direção do vento.
	temp.agua	Temperatura da água, em graus Celsius.
	tratamento	Se foi utilizado o <i>hookpod</i> (tratamento) ou chumbo normal (controle).
	bloco	Ordem de lançamento do tratamento (ou controle).
	n.anzois	Número de anzóis lançados (<i>offset</i>).
Resposta	t.alalunga.cap	Captura (em número) de albacora-branca (<i>Thunnus alalunga</i>).
	t.albacares.cap	Captura (em número) de albacora-lage (<i>Thunnus albacares</i>).

4.2.2 Modelos

Os Modelos Multivariados de Covariância Linear Generalizada (McGLM) serão apresentados como uma extensão dos GLMs. Seja \mathbf{Y} um vetor de respostas $N \times 1$, \mathbf{X} uma matriz de delineamento $N \times k$, e $\boldsymbol{\beta}$ um vetor de parâmetros $k \times 1$. Um GLM pode ser escrito da seguinte forma

$$\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta}) \quad (4.1a)$$

$$\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma} = V(\boldsymbol{\mu}; p)^{\frac{1}{2}} (\tau_0 \mathbf{I}) V(\boldsymbol{\mu}; p)^{\frac{1}{2}} \quad (4.1b)$$

onde g é uma função de ligação monótona e diferenciável, $V(\boldsymbol{\mu}; p) = \text{diag}(\vartheta(\boldsymbol{\mu}; p))$ é uma matriz diagonal cujas entradas são dadas pela função de variância $\vartheta(\boldsymbol{\mu}; p)$ aplicada à cada elemento do vetor $\boldsymbol{\mu}$. Finalmente, p e τ_0 são os parâmetros de potência e dispersão, respectivamente, e \mathbf{I} é a matriz identidade $N \times N$. Note que o modelo linear normal padrão é obtido quando a função de ligação é a identidade e a função de variância é constante.

A grande flexibilidade dos GLMs é devida ao uso de apenas duas funções distintas, a função de ligação e a função de variância. Esta última possui grande importância nos GLMs pois diferentes escolhas implicam em diferentes suposições sobre a distribuição da variável resposta. As funções de potência

$$\vartheta(\boldsymbol{\mu}; p) = \mu^p$$

onde $p \in (-\infty, 0] \cup [1, \infty)$, são escolhas frequentes e caracterizam a família de distribuições Tweedie, cujos casos particulares mais importantes são a normal ($p = 0$), Poisson ($p = 1$), gama ($p = 2$), e Gaussiana inversa ($p = 3$) (Jørgensen, 1997). Para uma variável aleatória Z com distribuição Tweedie, escrevemos $Z \sim \text{Tw}_p(\mu, \phi)$, onde $\phi > 0$ é um parâmetro de dispersão.

Apesar de toda a flexibilidade, os GLMs lidam apenas com dados univariados independentes, e a função de variância é assumida como conhecida. Os objetivos do McGLM são então: (a) lidar com dados não independentes, (b) possibilitar a modelagem de múltiplas respostas, e (c) calcular o parâmetro de potência, que funciona como uma seleção de modelos automática.

Para dados de contagem, a família Tweedie pode ser estendida para a família Poisson-Tweedie (El-Shaarawi et al., 2011), $\text{PTw}_p(\mu, \phi)$, que surge a partir da seguinte especificação hierárquica (Bonat et al., 2018)

$$\begin{aligned} Y|Z &\sim \text{Poisson}(Z) \\ Z &\sim \text{Tw}_p(\mu, \phi) \end{aligned}$$

assumindo que $p \geq 1$ para garantir que Z é não negativa. Em geral, a densidade da distribuição Poisson-Tweedie não possui solução fechada. No entanto, os dois primeiros momentos (média e variância) podem ser facilmente obtidos. Jørgensen e Kokonendji (2016) mostraram que, se $Y \sim \text{PTw}_p(\mu, \phi)$, então $\mathbb{E}(Y) = \mu$ e $\text{Var}(Y) = \mu + \phi\mu^p$. Esse fato motiva a especificação de um modelo baseado apenas em suposições de momentos de segunda-ordem (Bonat et al., 2018). Sendo assim, a função de variância para esta família de distribuições é

$$\vartheta(\boldsymbol{\mu}; p) = \mu + \phi\mu^p,$$

em que modelos importantes para dados de contagem são casos particulares, como Hermite ($p = 0$), Neyman tipo A ($p = 1$), binomial negativo ($p = 2$), e Poisson-inversa Gaussiana ($p = 3$). Quando utiliza-se a família Poisson-Tweedie, a matriz $\boldsymbol{\Sigma}$ na equação (4.1b) adquire a forma especial

$$\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\mu}) + V(\boldsymbol{\mu}; p)^{\frac{1}{2}} (\tau_0 \mathbf{I}) V(\boldsymbol{\mu}; p)^{\frac{1}{2}}$$

que é capaz de acomodar naturalmente dados com sub/sobredispersão e inflacionados de zero.

Ao trocar a matriz identidade na equação (4.1b), por uma matriz não-diagonal $\Omega(\boldsymbol{\tau})$, obtém-se

$$\text{Var}(\mathbf{Y}) = \Sigma = V(\boldsymbol{\mu}; p)^{\frac{1}{2}} (\Omega(\boldsymbol{\tau})) V(\boldsymbol{\mu}; p)^{\frac{1}{2}}$$

que é similar ao conceito de *working correlation matrix* no contexto de GEE (Liang e Zeger, 1986). No entanto, a proposta é modelar $\Omega(\boldsymbol{\tau})$ como uma combinação linear de matrizes conhecidas

$$h(\Omega(\boldsymbol{\tau})) = \tau_0 Z_0 + \cdots + \tau_D Z_D,$$

onde h é uma função de ligação para as covariâncias, Z_d com $d = 0, \dots, D$ são matrizes conhecidas, e $\boldsymbol{\tau} = (\tau_0, \dots, \tau_D)$ é um vetor de parâmetros de dispersão $(D+1) \times 1$. Essa estrutura é análoga ao preditor linear para a média, e é chamada de preditor linear matricial. Ao incorporar esse preditor linear matricial na equação (4.1b), obtemos o chamado Modelo de Covariância Linear Generalizado (cGLM). Casos especiais dessa abordagem incluem os modelos mistos generalizados, modelos de médias móveis e modelos condicionais autorregressivos.

A abordagem acima, apesar de ser mais flexível por permitir modelar a covariância entre as observações, é capaz de lidar apenas com respostas univariadas. Agora considere $\mathbf{Y}_{N \times R} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_R\}$ uma matriz de variáveis respostas, e seja $\mathbf{M}_{N \times R} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_R\}$ a respectiva matriz de valores esperados. Para indicar que cada resposta \mathbf{Y}_R possui sua própria matriz de covariância, considere

$$\Sigma_r = V_r(\boldsymbol{\mu}; p)^{\frac{1}{2}} \Omega_r(\boldsymbol{\tau}) V_r(\boldsymbol{\mu}; p)^{\frac{1}{2}}$$

como a matriz de covariâncias dentro de cada resposta. Considere também a matriz Σ_b como sendo a matriz de correlações entre as variáveis resposta. Com isso, o Modelo Multivariado de Covariância Linear Generalizada (McGLM) pode ser definido como

$$\begin{aligned} E(\mathbf{Y}) &= \mathbf{M} = \{g_1^{-1}(\mathbf{X}_1 \boldsymbol{\beta}_1), \dots, g_R^{-1}(\mathbf{X}_R \boldsymbol{\beta}_R)\} \\ \text{Var}(\mathbf{Y}) &= \mathbf{C} = \Sigma_R \stackrel{G}{\otimes} \Sigma_b \end{aligned}$$

onde

$$\Sigma_R \stackrel{G}{\otimes} \Sigma_b = \text{Bdiag}(\tilde{\Sigma}_1, \dots, \tilde{\Sigma}_R)(\Sigma_b \otimes \mathbf{I})\text{Bdiag}(\tilde{\Sigma}_1^\top, \dots, \tilde{\Sigma}_R^\top)$$

é o produto de Kronecker generalizado (Martinez-Beneito, 2013). A matriz $\tilde{\Sigma}_r$ é a matriz triangular inferior da decomposição de Cholesky de Σ_r , e o operador $\text{Bdiag}(\cdot)$ denota uma matriz bloco diagonal.

Os GLMs podem ser vistos como casos particulares dos McGLMs quando a função de ligação é a identidade, $\mathbf{X}_r = \mathbf{X}$ para todo r , a função de variância é constante, e $\Sigma_r = \tau_{0r} \mathbf{I}$ para todo r .

A especificação do modelo através de momentos de segunda-ordem motiva a dividir os parâmetros em dois subconjuntos $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$. Nesta notação, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_R^\top)^\top$ é um vetor $K \times 1$ de parâmetros de regressão, e $\boldsymbol{\lambda} = (\rho_1, \dots, \rho_{R(R-1)/2}, p_1, \dots, p_R, \boldsymbol{\tau}_1^\top, \dots, \boldsymbol{\tau}_R^\top)^\top$ um vetor $Q \times 1$ com os parâmetros de dispersão. As estimativas dos parâmetros de regressão são realizadas pelo método de quase-escore, e os parâmetros de dispersão pelas funções de estimação de Pearson. A formulação do modelo Poisson-Tweedie é exatamente a mesma daquela para os modelos quase-binomial e quase-Poisson no contexto de GLMs (Bonat et al., 2018). O método de estimação considera então $\mathcal{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_R^\top)^\top$ o vetor “empilhado” $NR \times 1$ da matriz de variáveis respostas $\mathbf{Y}_{N \times R}$. De maneira similar, $\mathcal{M} = (\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_R^\top)^\top$ é o vetor empilhado $NR \times 1$ da matriz de valores esperados $\mathbf{M}_{N \times R}$. Os detalhes do método estão bem descritos em Jørgensen e Knudsen (2004), Bonat e Jørgensen (2016) e Bonat et al. (2018).

Testes de hipótese linear multivariada podem ser realizadas da seguinte forma geral

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{c} \quad \text{vs} \quad H_1 : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{c}$$

onde \mathbf{L} é uma matriz conhecida de contrastes $JR \times K$. A estatística de teste de Wald é dada por

$$W_s = (\mathbf{L}\boldsymbol{\beta})^\top (\mathbf{L}\mathbf{J}_{\boldsymbol{\beta}}^{-1}\mathbf{L})^{-1} (\mathbf{L}\boldsymbol{\beta}),$$

onde $J_{\beta} = \mathbf{D}^T \mathbf{C}^{-1} \mathbf{D}$, e J denota o número de coeficientes de regressão sob teste. Assintoticamente, W_s possui uma distribuição χ^2 com JR graus de liberdade. Nesse caso, W_s é uma generalização direta da estatística de Hotelling-Lawley para dados multivariados não Gaussianos.

O ajuste de McGLMs podem ser realizados pelo pacote **mcglm** (Bonat, 2018) disponível para o R (R Core Team, 2018). Nesse pacote são implementadas diversas funções auxiliares para a construção das matrizes de covariância, além da função principal, **mcglm()** que possui sintaxe similar às demais funções de funcionalidades similares no R. O pacote estima os parâmetros através de dois algoritmos diferentes, o *modified chaser*, e o *reciprocal likelihood*.

4.3 Resultados e discussão

4.3.1 Análise exploratória

A distribuição das capturas das duas espécies de interesse pode ser vista na Figura 4.2 (A e B). Para a albacora-lage, nota-se uma grande dispersão dos valores de captura, sendo aproximadamente 19% de zeros e capturas de até 26 indivíduos. Já para a albacora-branca a quantidade de zeros é de aproximadamente 43%, e ocorreram no máximo a captura de 11 indivíduos por lance de pesca. A Figura 4.2 (C) mostra que não existe grande associação entre as capturas das duas espécies. O coeficiente de correlação entre as capturas foi de -0.04 (p -valor = 0.791), indicando que não existem evidências de que as capturas sejam correlacionadas.

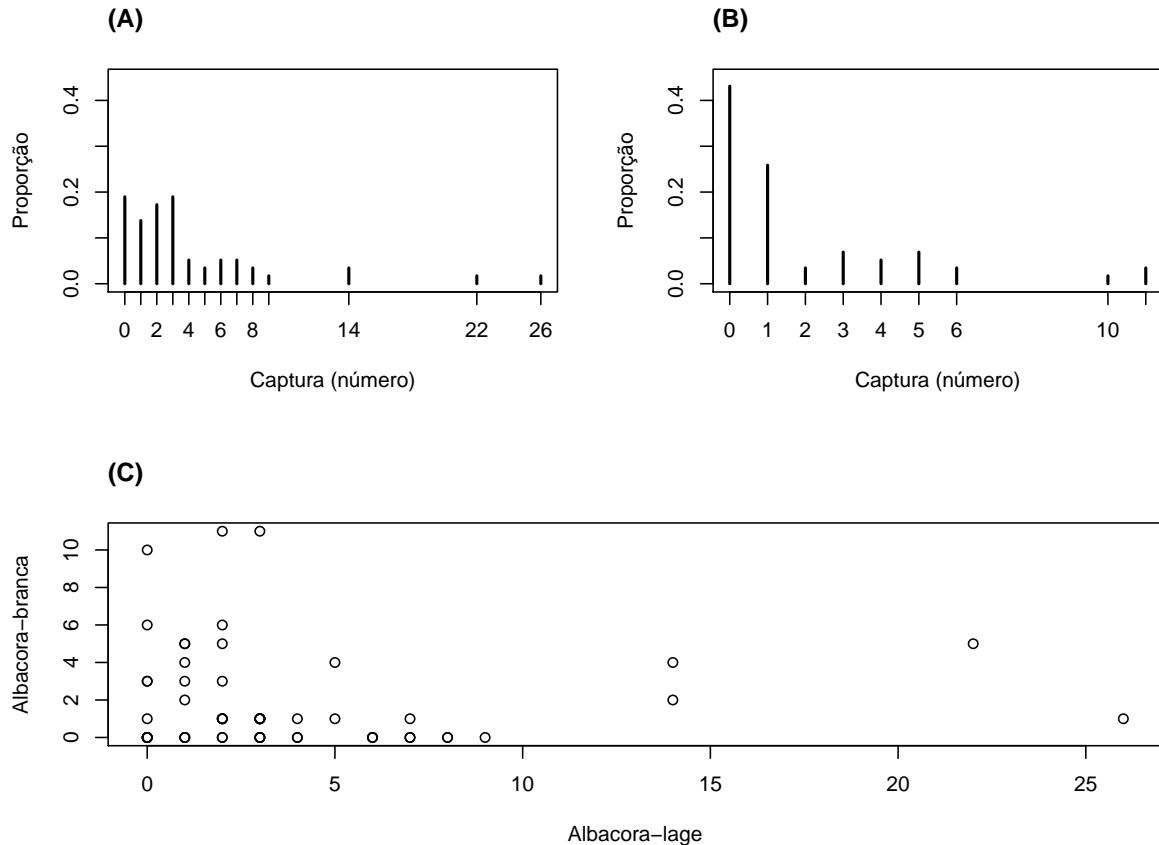


Figura 4.2. Distribuição das capturas de (A) albacora-lage, e (B) albacora-branca. (C) Associação entre as capturas de albacora-lage e albacora-branca (número de indivíduos de cada espécie por lance de pesca).

Considerando que o objetivo do trabalho é avaliar a eficiência do uso dos *hookpods*, a Figura 4.3 mostra os boxplots das capturas das duas espécies, separadas por controle (chumbo normal) e tratamento (*hookpods*). No caso da albacora-branca, nota-se que o uso dos *hookpods* aparentemente diminuiu as capturas. No entanto, algumas capturas altas também foram observadas. Já para a albacora-lage, as capturas foram relativamente maiores com os *hookpods*, mas também observam-se valores altos no controle.

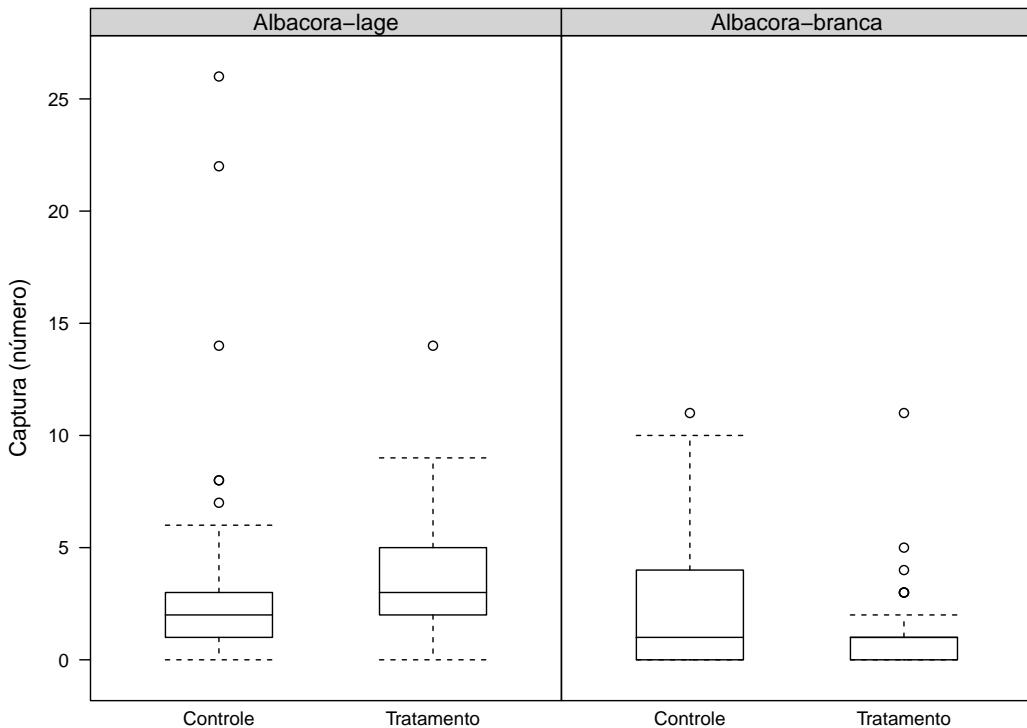


Figura 4.3. Boxplots das capturas de albacora-branca e albacora-lage, de acordo com uso do chumbo normal (controle), e dos hookpods (tratamento).

4.3.2 Modelo

Para o ajuste do modelo, inicialmente foram consideradas todas as variáveis explicativas descritas na Tabela 4.1. No entanto, após uma análise preliminar, verificou-se que a variável **beauford** (escala Beauford de vento) era não significativa, por isso não foi considerada.

Como assumiu-se independência entre as observações de cada variável resposta, as matrizes de covariância para cada uma delas foram especificadas como identidades de ordem n . Para a relação funcional entre a média e a variância de cada resposta, assumiu-se a distribuição Poisson-Tweedie, ou seja, $E(Y_i) = \mu_i$, e $\text{Var}(Y_i) = \mu_i + \phi\mu_i^p$, o que determina o primeiro e segundo momento respectivamente. O parâmetro de potência p da função de variância também foi estimado. A função de ligação entre cada resposta e os preditores lineares utilizada foi a logarítmica. O número de anzóis foi utilizado como um *offset* para especificar o intervalo (variável) onde as contagens foram realizadas. O modelo de covariância multivariado foi especificado através da função **mcglm()** do pacote **mcglm**.

Uma forma de conferir a convergência do modelo é através dos gráficos da Figura 4.4. Percebe-se que em todos os casos o algoritmo convergiu com 53 iterações.

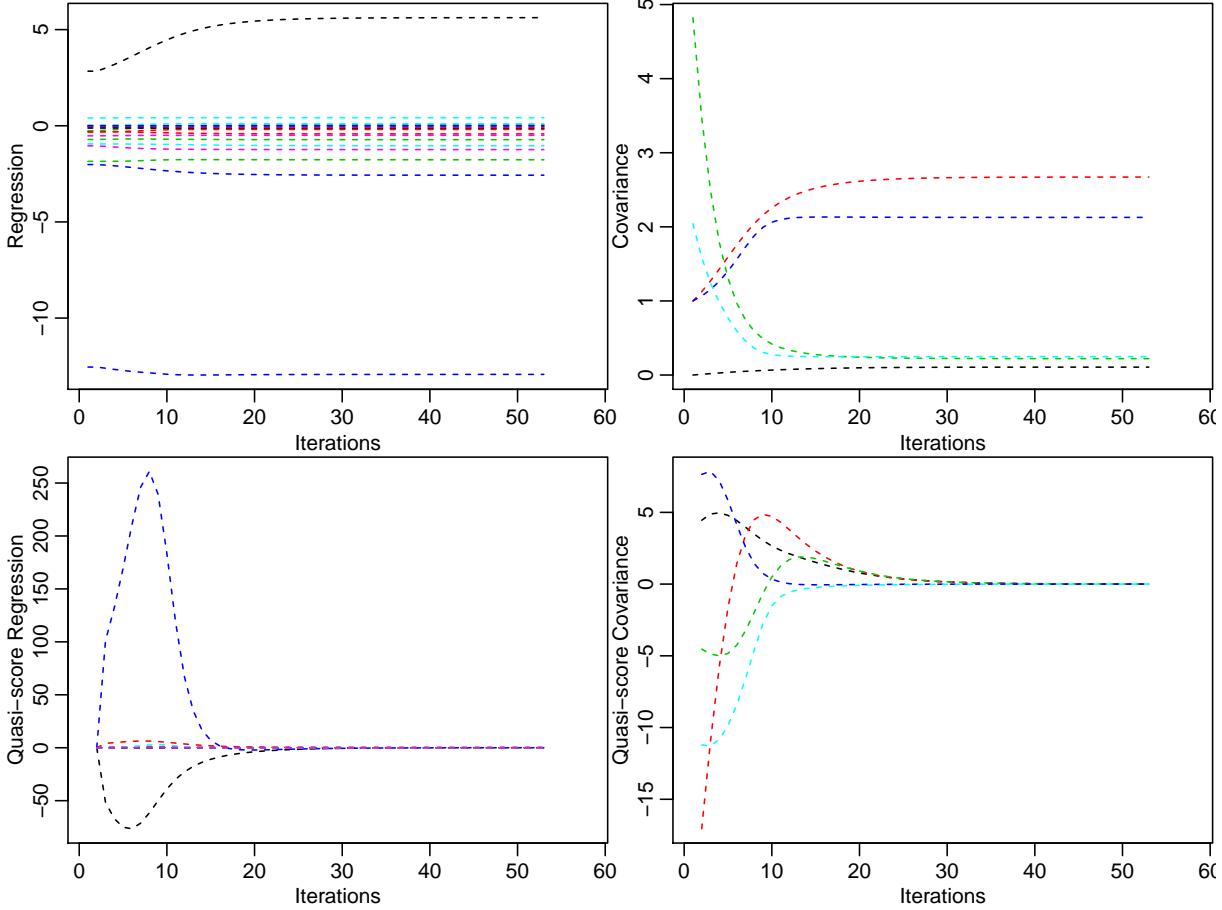


Figura 4.4. Diagnósticos de convergência do algoritmo.

As estimativas do parâmetro de dispersão ϕ são relevantes pois indicam se existe sub/equi/super-dispersão em cada variável. Por exemplo, se o parâmetro de dispersão for zero, então a variância se iguala à esperança e os dados podem ser descritos por uma distribuição Poisson (equidisperso). Ao ajustar os modelos com os parâmetros de potência à serem estimados, os erros padrão associados às estimativas dos parâmetros de dispersão podem ser inflacionados. Por isso, nestes casos, é aconselhável realizar um teste condicional dos parâmetros de dispersão, dado os valores das estimativas do parâmetro de potência. O resultado deste teste condicional está na Tabela 4.2. Ambas as estimativas são positivas, o que indica que existe uma aparente sobredispersão nos dados. No primeiro caso, para a espécie albacora-lage, o erro padrão é relativamente alto, e a estimativa é marginalmente significativa. Para a espécie albacora-branca, o erro padrão condicional é menor, e a estimativa do parâmetro de dispersão se mostrou significativa. Dessa forma, pode-se concluir que a captura de ambas as espécies possuem sobredisparso.

Tabela 4.2. Estimativas dos parâmetros de dispersão para albacora-lage (ϕ_1) e albacora-branca (ϕ_2).

	Estimativa	Erro padrão	Z	$\Pr(> z)$
ϕ_1	0,22	0,19	1,17	0,24
ϕ_2	0,25	0,07	3,65	0,00

Pelo fato de se saber que existe sobredisparso, a estimativa do parâmetro de potência p é de extrema importância pois pode indicar a distribuição aproximada das observações. As estimativas desse parâmetro são mostradas na Tabela 4.3. Como pode ser observado, a estimativa deste parâmetro para a espécie albacora-lage foi de 2,67, embora não tenha sido possível estimar seu erro padrão, e consequentemente seu nível de significância. De qualquer maneira, pode-se dizer que a distribuição dos

dados se aproxima de uma Poisson inversa Gaussiana ($p = 3$). Já a estimativa do parâmetro de potência para a albacora-branca foi de 2,13, significativamente diferente de zero, o que indica que a distribuição dos dados para esta espécie é próxima de uma binomial negativa ($p = 2$).

Tabela 4.3. Estimativas dos parâmetros de potência para albacora-lage (p_1) e albacora-branca (p_2).

	Estimativa	Erro padrão	Z	Pr($> z $)
p_1	2,67			
p_2	2,13	0,84	2,55	0,01

A correlação entre a captura das duas espécies pode agora ser estimada considerando-se todos os fatores incluídos no modelo. Essa correlação “ajustada”, ou corrigida pelos fatores está na Tabela 4.4. Nota-se que a correlação é positiva e não significativa. Por isso não existem evidências suficientes para supor que existe correlação entre a captura das duas espécies.

Tabela 4.4. Correlação ajustada entre as capturas das espécies albacora-lage e albacora-branca.

	Estimativa	Erro padrão	Z	Pr($> z $)
ρ_{12}	0,11	0,13	0,82	0,41

Um teste de significância conjunto pode ser obtido através de uma tabela de análise de variância multivariada (MANOVA), com os testes de Hotelling-Lawley que segue uma distribuição qui-quadrado (Tabela 4.5). Os resultados mostram que todas as variáveis são conjuntamente significativas, com exceção de bloco e tratamento. Por ser uma variável de controle apenas, bloco foi mantido no modelo. Como o objetivo era avaliar a eficiência entre controle e tratamento, pode-se concluir que não existem diferenças nas capturas de ambas espécies quando se utiliza o chumbo normal ou o *hookpod*, pelo menos conjuntamente. Esse era o resultado esperado, pois o objetivo do *hookpod* é diminuir a captura incidental de aves e não afetar a captura de espécies alvo.

Tabela 4.5. Análise de Variância Multivariada (MANOVA) conjunta para os termos considerados no modelo.

Efeitos	GL	Hotelling-Lawley	χ^2	Pr($> \chi^2$)
Intercept	2	0,819	47,508	0,000
bloco	2	0,004	0,255	0,880
periodo	2	0,095	5,509	0,064
dir.vento	6	0,582	33,745	0,000
profundidade	2	0,136	7,880	0,019
temp.agua	2	0,643	37,279	0,000
tratamento	2	0,061	3,562	0,169

A análise individual dos efeitos estimados de cada fator em relação à cada uma das variáveis resposta pode ser vista na Tabela 4.6. No caso da albacora-lage, nota-se que não existe efeito na captura pelo uso do *hookpod*. Em contraste, as capturas de albacora-branca parecem ter sido influenciadas negativamente pela presença dos *hookpods*, embora a significância seja baixa.

Um modelo alternativo foi ajustado, com a variável **tratamento** excluída. A pseudo-verossimilhança do modelo completo e deste modelo reduzido foram calculadas, e a diferença entre elas utilizada para a realização de um teste aproximado de razão de verossimilhança. O teste resultou em um *p*-valor de 0,869, indicando que conjuntamente o tratamento não é significativo, ou seja, o uso do *hookpod* não interfere nas capturas das espécies consideradas.

Tabela 4.6. Efeitos individuais de cada fator considerado no modelo, para as espécies albacora-lage e albacora-branca.

Efeitos	Albacora-lage				Albacora-branca			
	Estimativa	EP	Z	Pr(> z)	Estimativa	EP	Z	Pr(> z)
(Intercept)	5,621	1,927	2,917	0,004	-12,931	2,169	-5,961	0,000
bloco2	-0,072	0,255	-0,282	0,778	0,104	0,268	0,389	0,697
periodonoite	-0,725	0,309	-2,347	0,019	-0,079	0,360	-0,221	0,825
dir.ventone	-2,571	0,583	-4,411	0,000	-0,147	0,461	-0,318	0,750
dir.ventose	-1,041	0,534	-1,950	0,051	-0,194	0,437	-0,443	0,658
dir.ventosw	-1,240	0,590	-2,102	0,036	-1,769	0,705	-2,511	0,012
profundidade	-0,000	0,001	-0,283	0,777	0,002	0,001	2,751	0,006
temp.agua	-0,424	0,095	-4,461	0,000	0,417	0,111	3,753	0,000
tratamentopods	0,050	0,255	0,196	0,845	-0,497	0,269	-1,849	0,064

4.4 Conclusões

Os Modelos Multivariados de Covariância Linear Generalizada (McGLM) demonstraram ser uma alternativa adequada para a modelagem de dados de captura. Normalmente dados de captura de diferentes espécies são altamente correlacionados, devido ao fato de que muitas espécies possuem hábitos de vida similares, e são capturadas pelo mesmo artefato. O McGLM é capaz de medir a correlação entre as espécies, considerando todos os outros fatores que podem influenciar nessa associação.

Dados de captura frequentemente possuem sobredispersão e uma grande proporção de zeros. Esta abordagem é flexível o suficiente para acomodar ambas estas situações. Além disso, a estimativa do parâmetro de potência p , que define a distribuição mais adequada para cada variável resposta, torna essa classe de modelos ainda mais flexível, pois não fica-se restrito às tradicionais distribuições para dados de contagem, como a Poisson e a binomial negativa.

Através das estimativas dos parâmetros de dispersão, verificou-se que os dados de ambas espécies possuem sobredispersão. Já as distribuições mais adequadas, de acordo com as estimativas do parâmetro de potência p , foram a Poisson inversa Gaussiana para a albacora-lage, e a binomial negativa para a albacora-branca.

A possibilidade de realização de uma Análise de Variância Multivariada (MANOVA) conjunta faz com que os resultados sejam particularmente úteis na prática. Nesse trabalho, esse teste permitiu verificar que os *hookpods* (tratamento) não interferem na captura conjunta das duas espécies analisadas. Esse resultado foi corroborado quando foram analisados os efeitos individuais dos *hookpods* sobre a captura de cada uma das espécies. Em ambos os casos, o uso dos *hookpods* não demonstrou nenhum efeito significativo na captura das duas espécies.

O uso dos *hookpods* já é comprovadamente eficaz para evitar a captura incidental de aves em pescarias com espinhel de superfície (Sullivan et al., 2018). No entanto, havia uma preocupação por parte dos pescadores de que este artefato também poderia diminuir a captura das espécies alvo da pescaria, diminuindo assim sua rentabilidade. Os resultados obtidos neste estudo, através da análise dos dados de um experimento especificamente planejado para verificar esse questionamento, mostraram que o uso dos *hookpods* não interferem na captura das espécies alvo.

Referências

- Bonat, W. H. (2018). Multiple response variables regression models in R: the mcglm package. *Journal of Statistical Software*, 84(4).
- Bonat, W. H. e Jørgensen, B. (2016). Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(5):649–675.

- Bonat, W. H., Jørgensen, B., Kokonendji, C. C., Hinde, J., e Demétrio, C. G. (2018). Extended Poisson–Tweedie: properties and regression models for count data. *Statistical Modelling*, 18(1):24–49.
- El-Shaarawi, A. H., Zhu, R., e Joe, H. (2011). Modelling species abundance using the Poisson-Tweedie family. *Environmetrics*, 22(2):152–164.
- Johnson, R. A. e Wichern, D. W. (2008). *Applied multivariate statistical analysis*. Pearson, London.
- Jørgensen, B. (1997). *The theory of dispersion models*. Chapman & Hall.
- Jørgensen, B. e Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, 31(1):93–114.
- Jørgensen, B. e Kokonendji, C. C. (2016). Discrete dispersion models and their Tweedie asymptotics. *ASTA Advances in Statistical Analysis*, 100(1):43–78.
- Liang, L. e Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.
- Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3):539–553.
- Nelder, J. A. e Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3).
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, 86(3):677–690.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sullivan, B. J., Kibel, B., Kibel, P., Yates, O., Potts, J. M., Ingham, B., Domingo, A., Gianuca, D., Jiménez, S., Lebepe, B., Maree, B. A., Neves, T., Peppes, F., Rasehlomi, T., Silva-Costa, A., e Wanless, R. M. (2018). At-sea trialling of the Hookpod: a “one-stop” mitigation solution for seabird bycatch in pelagic longline fisheries. *Animal Conservation*, 21(2):159–167.
- Zeger, S. L., Liang, K.-Y., e Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44(4):1049.

5 CONSIDERAÇÕES FINAIS

Este trabalho pretendeu mostrar novas abordagens para a modelagem estatística de dados pesqueiros. Este tipo de dado geralmente apresenta desafios do ponto de vista estatístico, como a dependência espaço-temporal, excesso de zeros e sobredispersão. Frequentemente, muitas das suposições necessárias são ignoradas quando são utilizadas as técnicas tradicionais difundidas na ciência pesqueira. Com isso, mostrou-se que atualmente estas suposições podem ser consideradas, através de técnicas modernas em termos de modelagem estatística.

No Capítulo 2 foi mostrado que a estrutura temporal das capturas deve ser levada em consideração. Para isso, foram ajustados uma série de modelos através do INLA, uma abordagem relativamente nova para realização de inferência no contexto bayesiano. Por ser um método que utiliza aproximação, ele possibilita que diversos modelos sejam ajustados e comparados rapidamente, uma tarefa que seria mais lenta tanto do ponto de vista de programação dos modelos, quanto do ponto de vista de tempo computacional, quando comparado aos métodos por amostragem. Sendo assim, mostrou-se que o INLA é uma ferramenta extremamente viável para a modelagem de capturas considerando a estrutura de dependência temporal.

No Capítulo 3, uma nova abordagem para a estimativa dos parâmetros de modelos dinâmicos de biomassa foi desenvolvida. A estimação em dois estágios via INLA proposta possui duas vantagens principais. A primeira delas é o rápido tempo computacional para se ajustar modelos complexos. A segunda vantagem é que este método permite com que toda a estrutura de dependência espaço-temporal dos dados de pesca seja levada em consideração. Os resultados mostraram que, de fato, ao se considerarem estruturas espaciais e temporais, as estimativas obtidas são não viesadas e consistentes. Os resultados derivados das estimativas dos modelos dinâmicos de biomassa através deste método podem ser de grande utilidade no gerenciamento de estoques pesqueiros. Por exemplo, a distribuição espacial de parâmetros populacionais, como a capacidade de suporte, e de parâmetros associados à pesca, como o coeficiente de capturabilidade, podem ser úteis para a identificação de áreas onde as populações são mais vulneráveis. Associado às épocas do ano, estas informações podem servir de base para o fechamento de áreas para a pesca, por exemplo.

Uma abordagem multivariada para avaliar a influência de fatores na captura de espécies de atum foi utilizada no Capítulo 4. Apesar de parecer uma abordagem natural para dados de captura de várias espécies, análises multivariadas são pouco utilizadas neste tipo de dado, devido principalmente às suposições de normalidade na maioria dos métodos disponíveis. Neste trabalho foi mostrado que os Modelos Multivariados de Covariância Linear Generalizada (McGLM) são uma alternativa nova e viável para esse tipo de dado. Essa classe de modelos permite a modelagem da estrutura de covariância entre múltiplas respostas, mas, principalmente, flexibiliza a suposição de normalidade, ao considerar que as respostas podem assumir qualquer distribuição da família exponencial de dispersão. A implementação destes modelos através do pacote **mcglm** facilita e torna viável seu uso em diversos estudos envolvendo dados pesqueiros.

Finalmente, todos os códigos e demais produtos gerados nesta tese podem ser consultados no endereço <http://leg.ufpr.br/~fernandomayer/tese>.

APÊNDICES

A Demonstrações

Demonstração de (2.1) A parametrização da binomial negativa adotada é aquela que surge tradicionalmente quando considera-se $Y \sim \text{Pois}(\lambda)$, e $\lambda > 0$ segue uma distribuição gama com parâmetros $\delta > 0$ e $\beta > 0$, de forma que $E(\lambda) = \delta/\beta$ e $\text{Var}(\lambda) = \delta/\beta^2$. Ou seja,

$$P[Y = y|\lambda] = f(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, 2, \dots$$

e

$$f(\lambda; \delta, \beta) = \frac{\beta^\delta}{\Gamma(\delta)} \lambda^{\delta-1} e^{-\beta\lambda}, \quad \lambda > 0.$$

Como

$$\begin{aligned} f(y|\lambda) &= \frac{f(y, \lambda)}{f(\lambda)} \\ f(y, \lambda) &= f(y|\lambda)f(\lambda) \end{aligned}$$

é a distribuição conjunta, basta integrá-la em λ para obter a distribuição marginal de Y , ou seja,

$$f(y) = \int f(y, \lambda) d\lambda.$$

Sendo assim,

$$\begin{aligned} f(y) &= \int_0^\infty f(y, \lambda) d\lambda \\ &= \int_0^\infty f(y|\lambda)f(\lambda) d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \frac{\beta^\delta}{\Gamma(\delta)} \lambda^{\delta-1} e^{-\beta\lambda} d\lambda \\ &= \frac{\beta^\delta}{\Gamma(\delta)y!} \int_0^\infty \lambda^{y+\delta-1} e^{-\lambda(\beta+1)} d\lambda \\ &= \frac{\beta^\delta}{\Gamma(\delta)y!} \frac{\Gamma(y+\delta)}{(\beta+1)^{y+\delta}} \\ &= \frac{\Gamma(y+\delta)}{\Gamma(\delta)y!} \frac{\beta^\delta}{(\beta+1)^\delta} \frac{1}{(\beta+1)^y} \\ &= \frac{\Gamma(y+\delta)}{\Gamma(\delta)y!} \left(\frac{\beta}{\beta+1} \right)^\delta \left(\frac{1}{\beta+1} \right)^y \\ &= \frac{\Gamma(y+\delta)}{\Gamma(\delta)y!} \pi^\delta (1-\pi)^y \end{aligned}$$

onde $\pi = \beta/(\beta+1)$. Como $y! = \Gamma(y+1)$, então

$$f(y) = \frac{\Gamma(y+\delta)}{\Gamma(\delta)\Gamma(y+1)} \pi^\delta (1-\pi)^y.$$

Note que

$$\frac{\Gamma(y+\delta)}{\Gamma(\delta)\Gamma(y+1)} = \frac{(y+\delta-1)!}{(\delta-1)!y!} = \binom{y+\delta-1}{y},$$

portanto

$$f(y) = \binom{y+\delta-1}{y} \pi^\delta (1-\pi)^y \blacksquare$$

Demonstração de (3.15)

$$\begin{aligned}
 B_{t+1} &= B_t + rB_t - \frac{r}{K}B_t^2 - qE_t B_t & B_t &= \frac{U_t}{q} \\
 \frac{U_{t+1}}{q} &= \frac{U_t}{q} + r\frac{U_t}{q} - \frac{r}{K}\frac{U_t^2}{q^2} - qE_t \frac{U_t}{q} & U_t &= \frac{C_t}{E_t} \\
 U_{t+1} &= U_t + rU_t - \frac{r}{qK}U_t^2 - qE_t U_t & \div U_t \\
 U_{t+1} - U_t &= rU_t - \frac{r}{qK}U_t^2 - qC_t & \\
 \frac{U_{t+1} - U_t}{U_t} &= r - \frac{r}{qK}U_t - qE_t & \text{pois } \frac{C_t}{U_t} = E_t
 \end{aligned}$$

■

Demonstração de (3.30)

$$\begin{aligned}
 B_{(t+1),i} &= B_{ti} + rB_{ti} - \frac{r}{K}B_{ti}^2 - qE_{ti} B_{ti} & B_{ti} &= \frac{U_{ti}}{q} \\
 \frac{U_{(t+1),i}}{q} &= \frac{U_{ti}}{q} + r\frac{U_{ti}}{q} - \frac{r}{K}\frac{U_{ti}^2}{q^2} - qE_{ti} \frac{U_{ti}}{q} & U_{ti} &= \frac{C_{ti}}{E_{ti}} \\
 U_{(t+1),i} &= U_{ti} + rU_{ti} - \frac{r}{qK}U_{ti}^2 - qE_{ti} U_{ti} & \div U_{ti} \\
 U_{(t+1),i} - U_{ti} &= rU_{ti} - \frac{r}{qK}U_{ti}^2 - qC_{ti} & \\
 \frac{U_{(t+1),i} - U_{ti}}{U_{ti}} &= r - \frac{r}{qK}U_{ti} - qE_{ti} & \text{pois } \frac{C_{ti}}{U_{ti}} = E_{ti}
 \end{aligned}$$

■

B Figuras

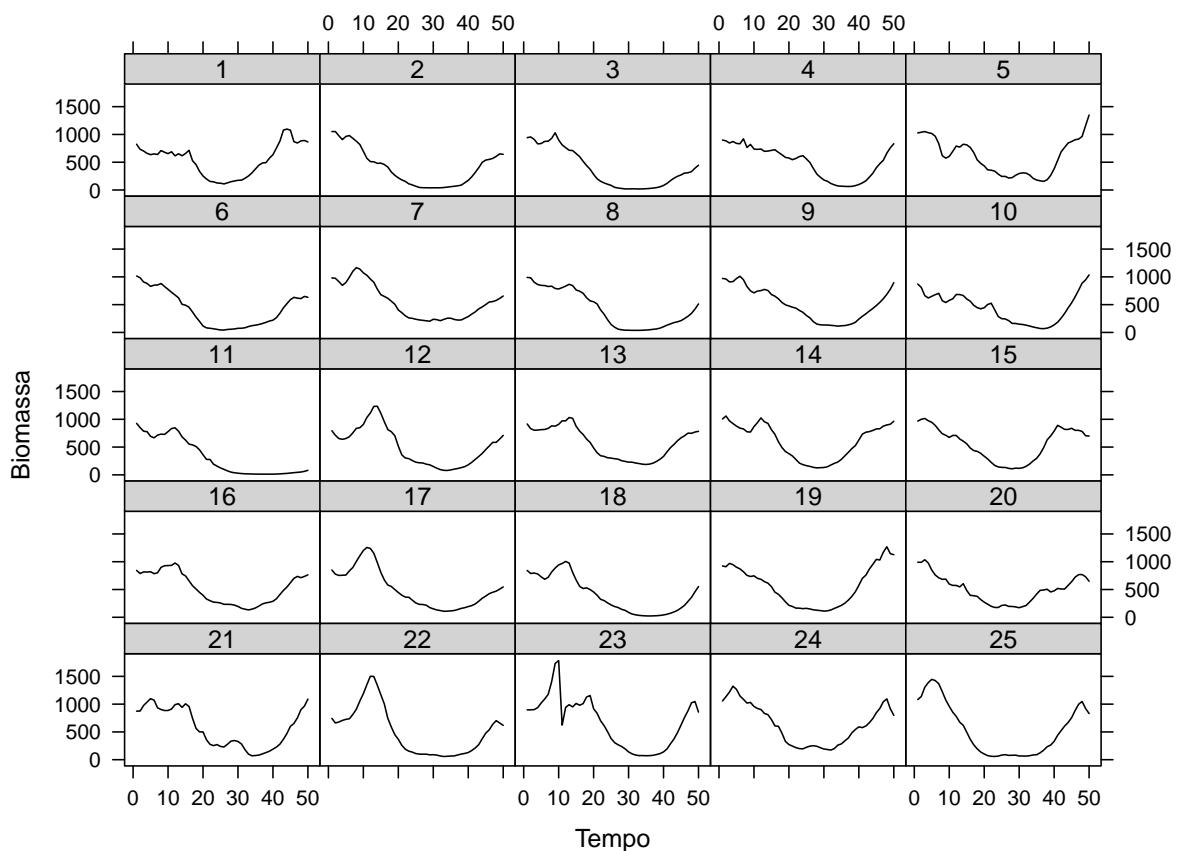


Figura B.1. Séries temporais da biomassa simulada por área.

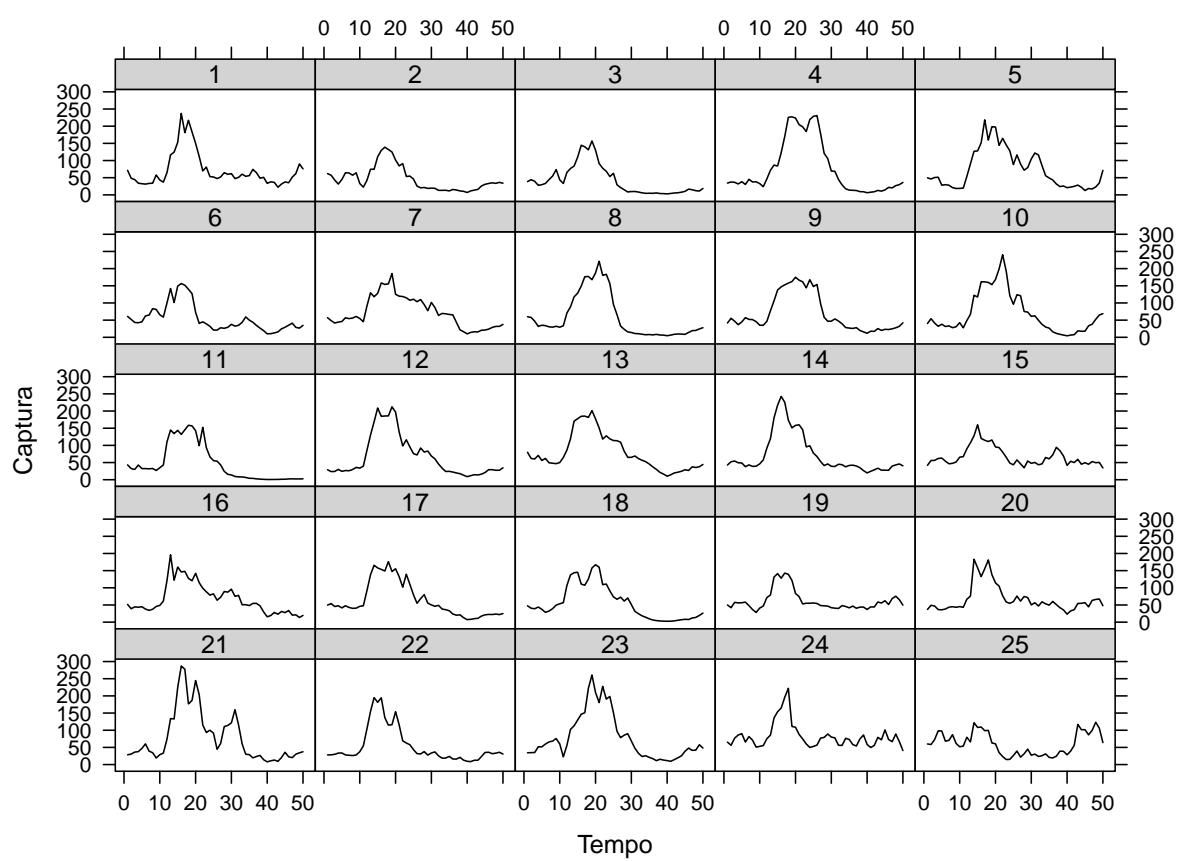


Figura B.2. Séries temporais da captura simulada por área.

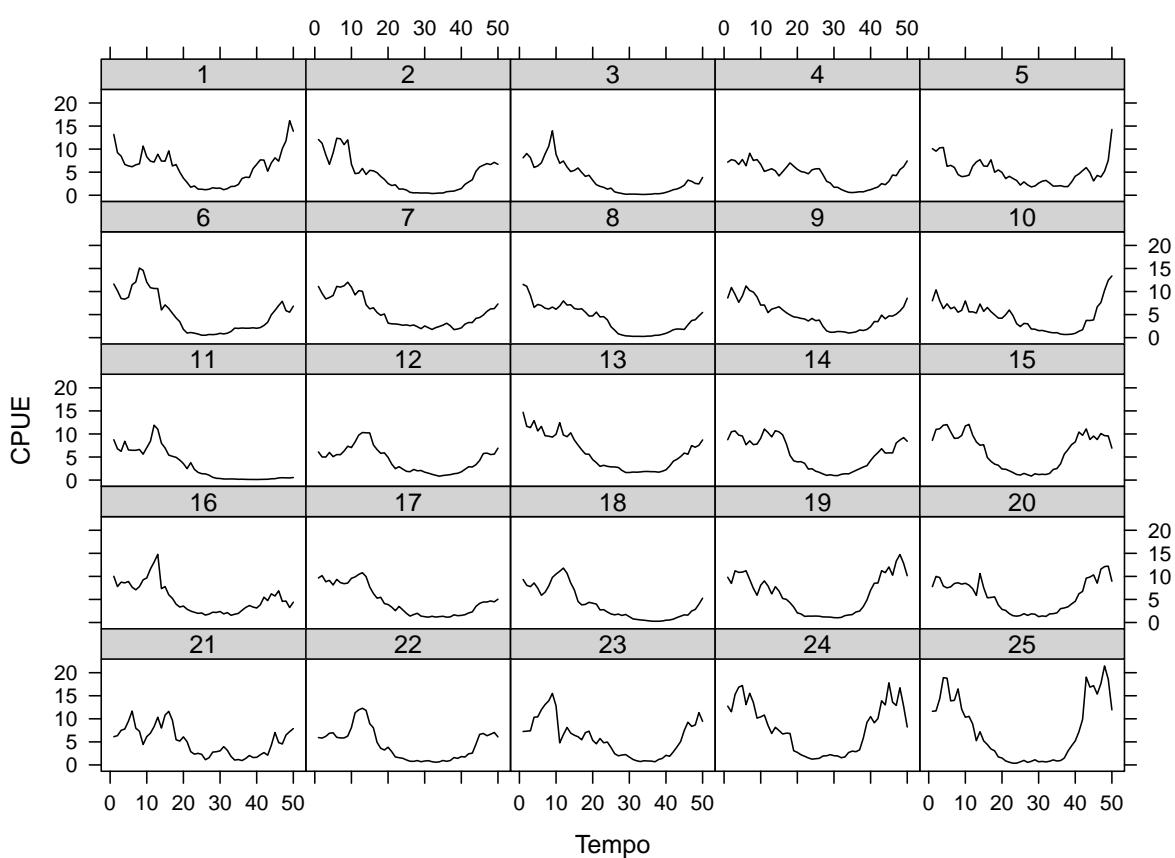


Figura B.3. Séries temporais da CPUE simulada por área.

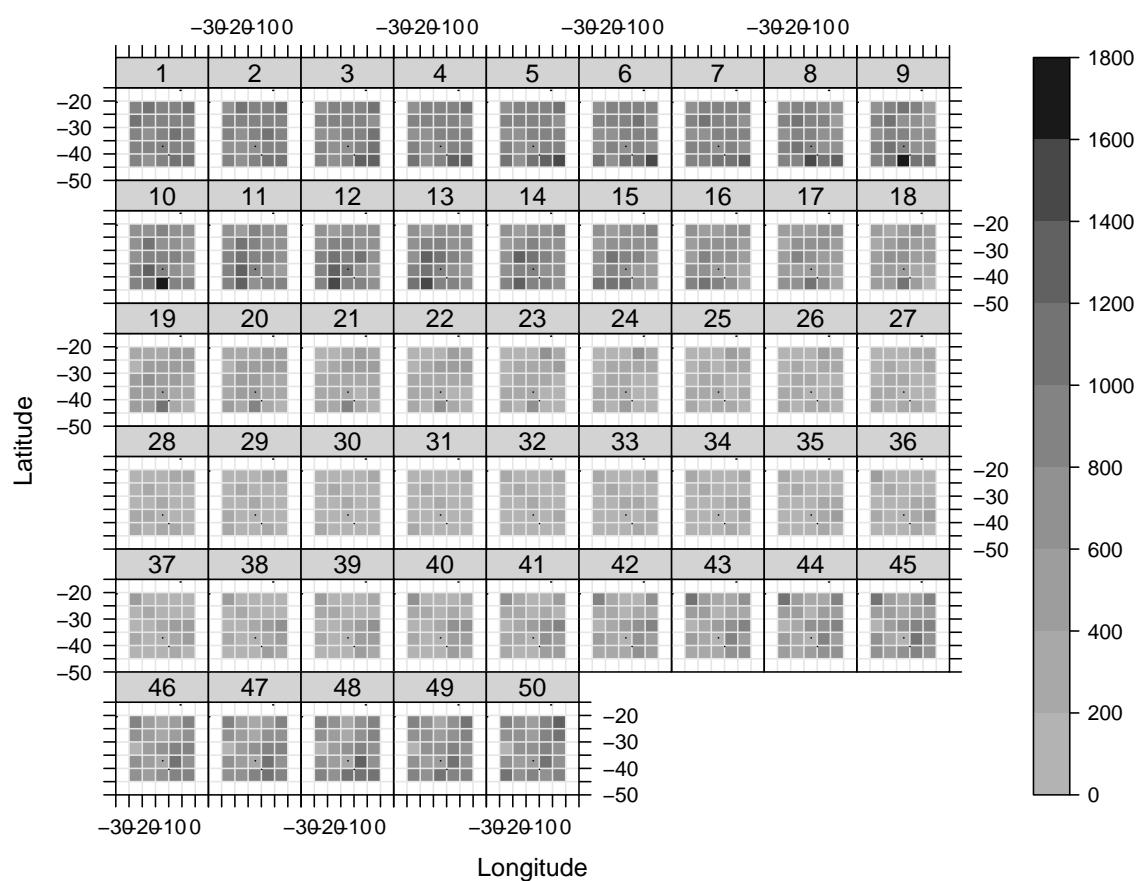


Figura B.4. Mapas anuais com a distribuição espacial da biomassa simulada.

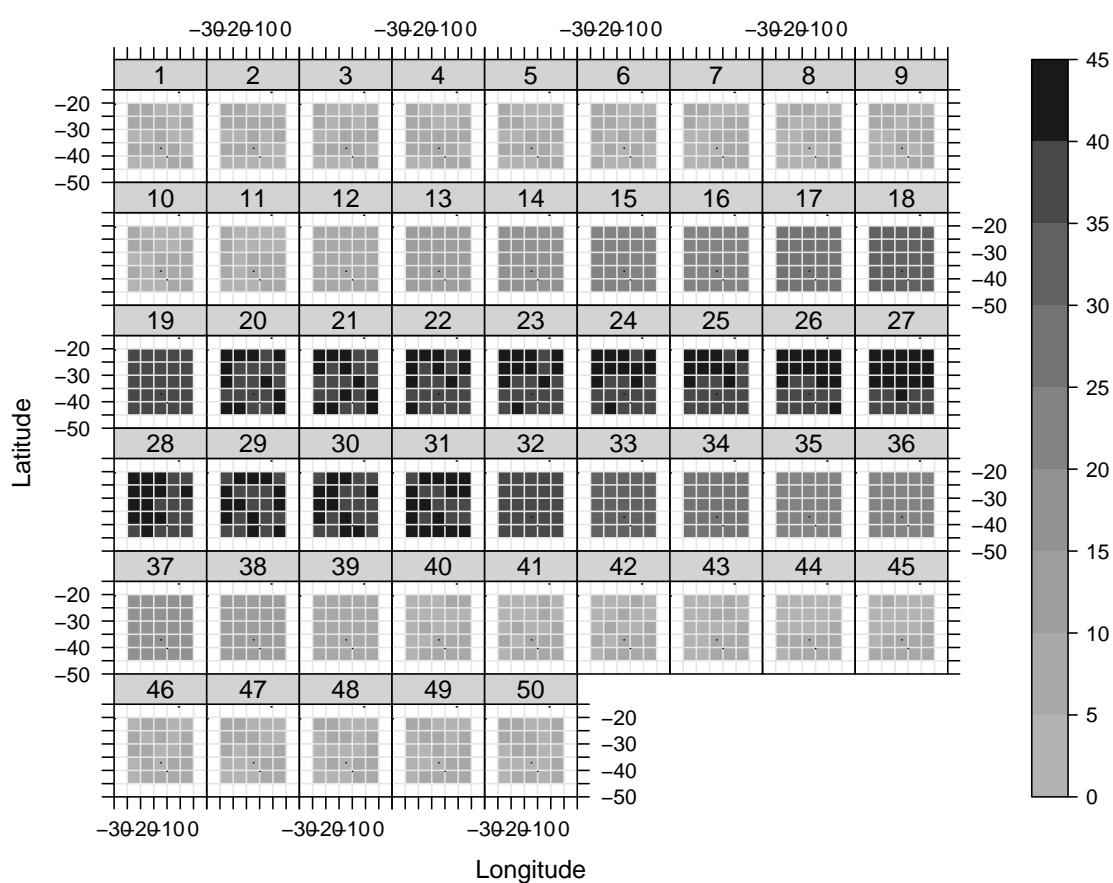


Figura B.5. Mapas anuais com a distribuição espacial do esforço simulado.

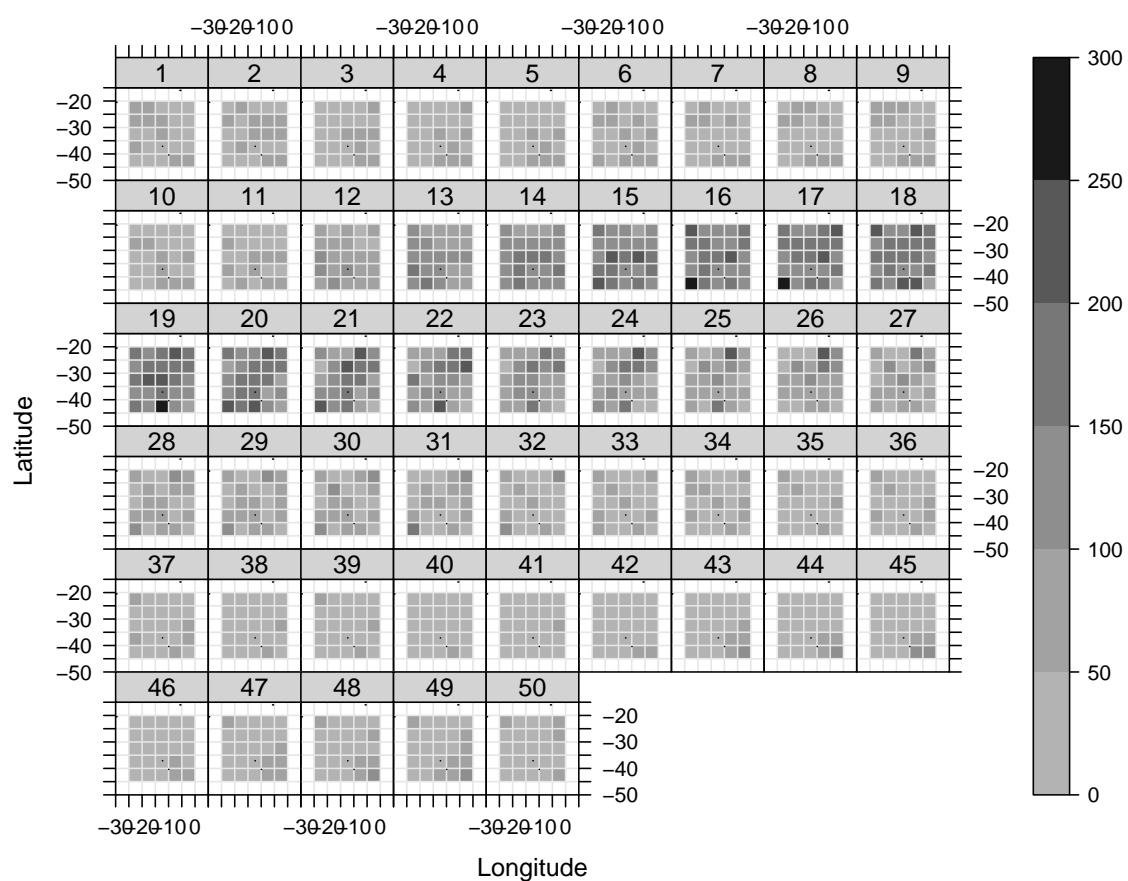


Figura B.6. Mapas anuais com a distribuição espacial da captura simulada.

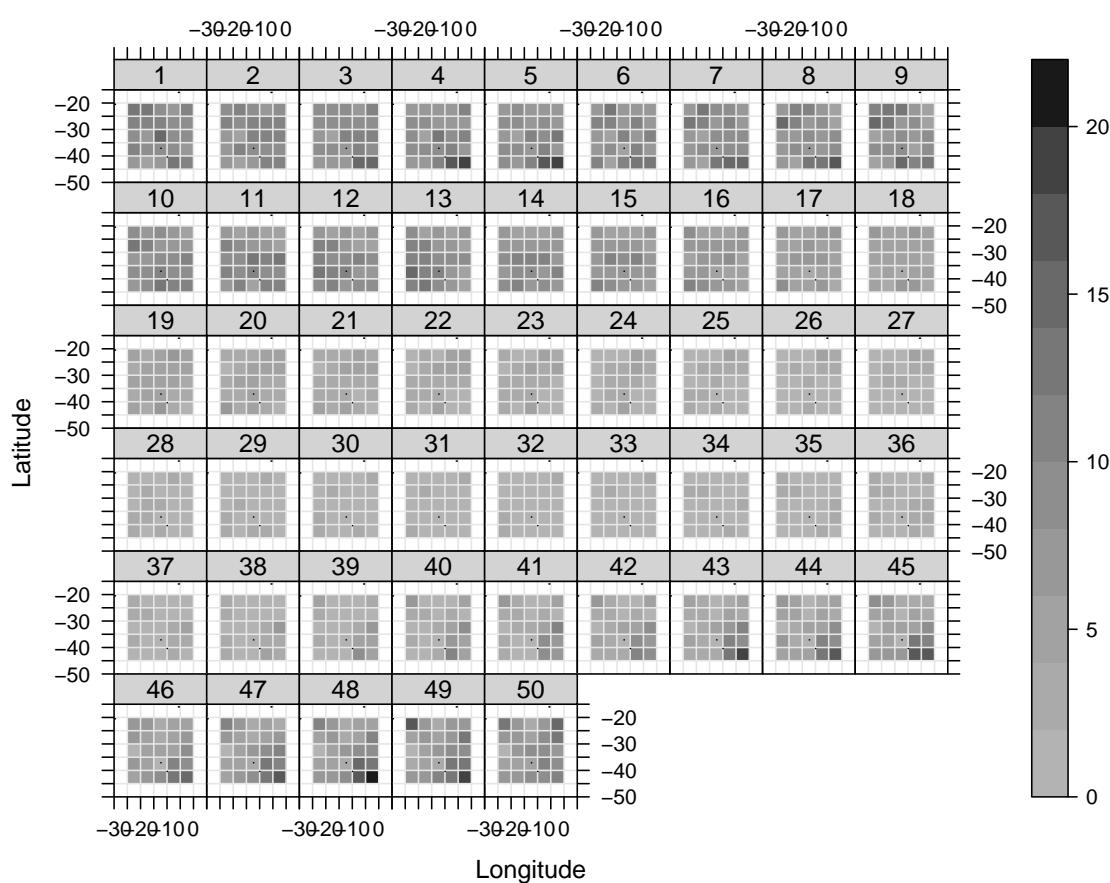


Figura B.7. Mapas anuais com a distribuição espacial da CPUE simulada.

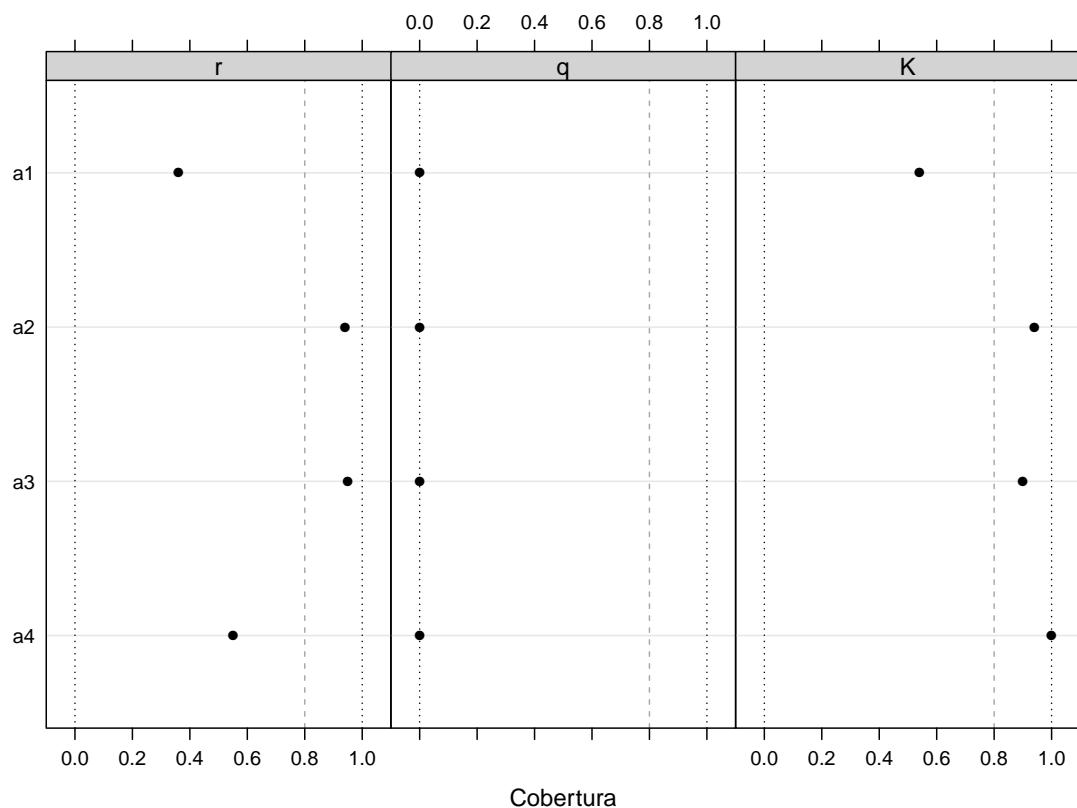


Figura B.8. Nível de cobertura das estimativas dos parâmetros por meio dos modelos para dados brutos agregados, considerando um intervalo de credibilidade de 95%. A linha tracejada vertical indica o limiar de 80% de cobertura, enquanto que as linhas pontilhadas indicam 0% e 100%.

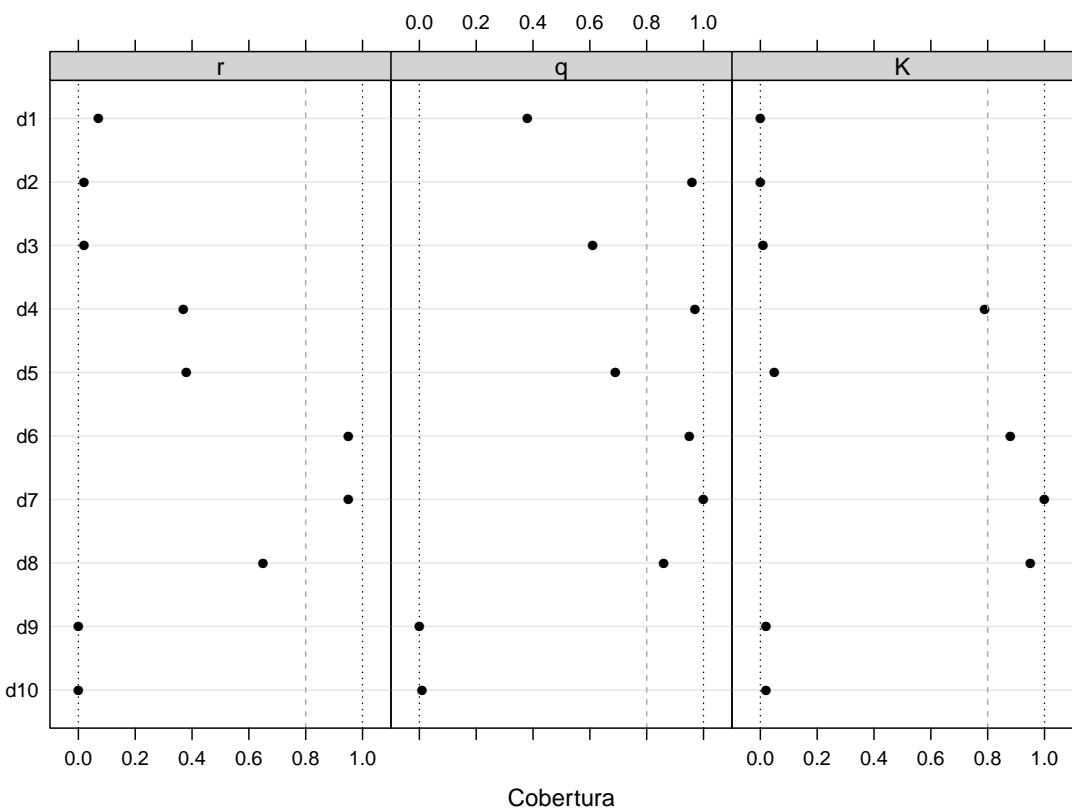


Figura B.9. Nível de cobertura das estimativas dos parâmetros por meio dos modelos para dados brutos desagregados, considerando um intervalo de credibilidade de 95%. A linha tracejada vertical indica o limiar de 80% de cobertura, enquanto que as linhas pontilhadas indicam 0% e 100%.

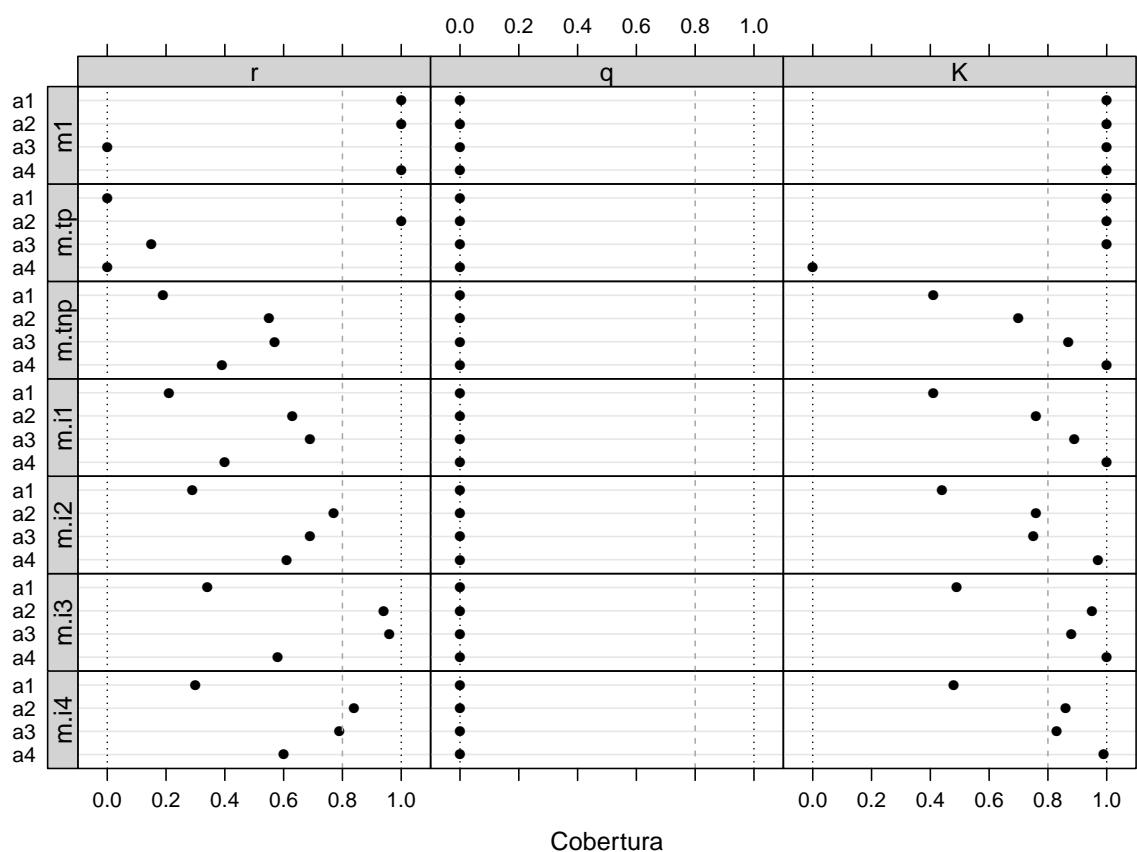


Figura B.10. Nível de cobertura das estimativas dos parâmetros por meio dos modelos para dados preditos agregados, considerando um intervalo de credibilidade de 95%. A linha tracejada vertical indica o limiar de 80% de cobertura, enquanto que as linhas pontilhadas indicam 0% e 100%.

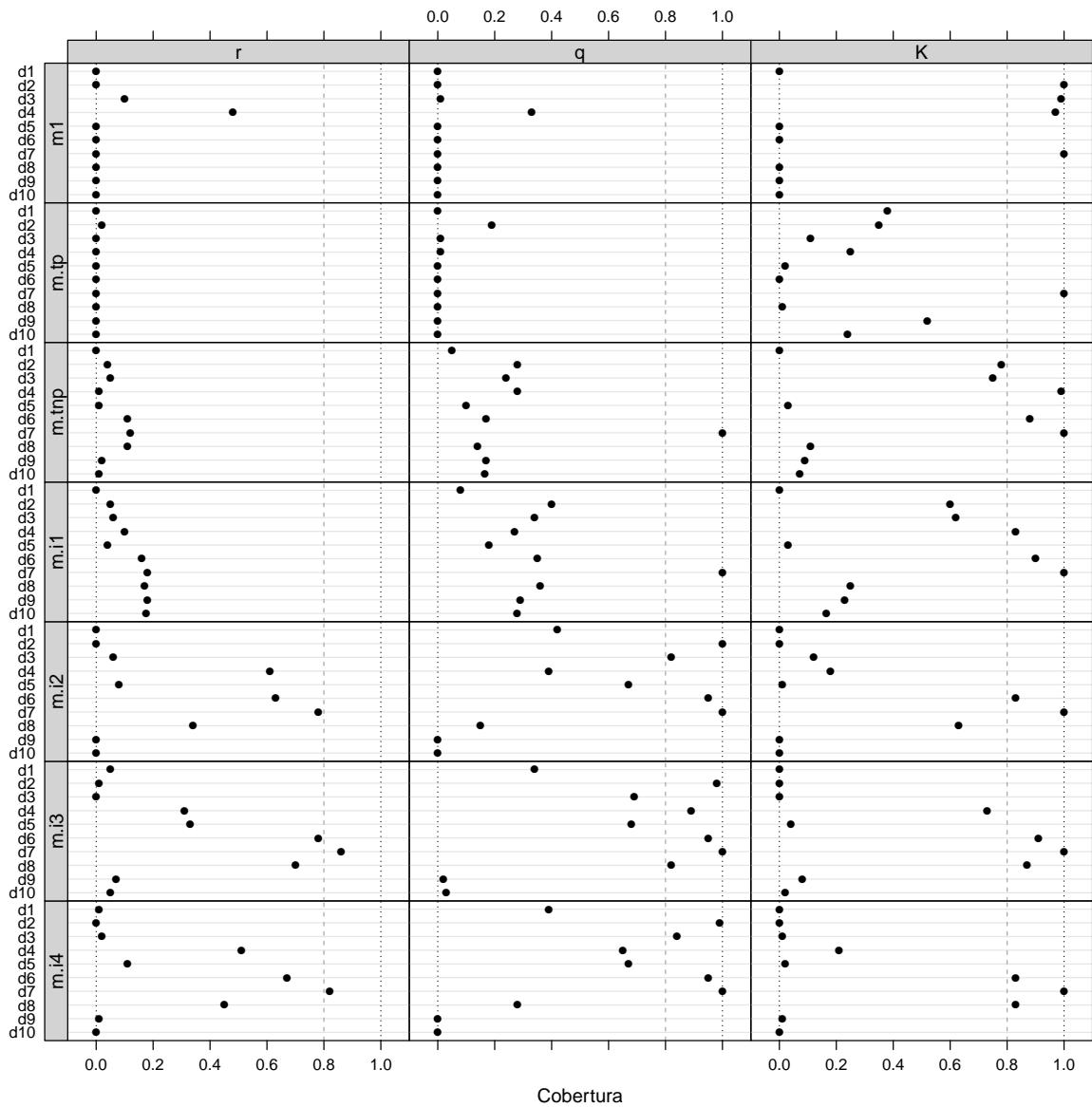


Figura B.11. Nível de cobertura das estimativas dos parâmetros por meio dos modelos para dados preditos desagregados, considerando um intervalo de credibilidade de 95%. A linha tracejada vertical indica o limiar de 80% de cobertura, enquanto que as linhas pontilhadas indicam 0% e 100%.

C Tabelas

Tabela C.1. Deviance residual, critério de informação da deviance (DIC), número efetivo de parâmetros (p_D), critério de informação amplamente aplicável (WAIC) e número efetivo de parâmetros (p_{WAIC}), e logaritmo da ordenada preditiva condicional (LCPO) para os diferentes modelos considerados para dados brutos no segundo estágio.

Dado	Modelo	Deviance	DIC	p_D	WAIC	p_{WAIC}	LCPO
Agregado	a1	-191,63	-187,27	4,36	-185,14	5,76	92,41
	a2	-193,45	-187,43	6,02	-184,96	7,40	92,19
	a3	-192,51	-183,96	8,55	-181,24	9,58	90,03
	a4	-196,01	-165,33	30,68	-169,55	20,37	75,14
Desagregado	d1	-434,62	-430,40	4,22	-429,57	5,02	214,79
	d2	-457,18	-449,79	7,38	-449,03	8,07	224,51
	d3	-459,67	-449,75	9,92	-449,57	9,97	224,79
	d4	-459,84	-445,82	14,02	-445,37	14,27	222,68
	d5	-471,79	-453,68	18,11	-452,01	19,46	226,01
	d6	-467,83	-433,72	34,11	-434,41	32,37	217,14
	d7	-477,78	-420,89	56,89	-419,39	55,72	209,53
	d8	-572,75	-486,07	86,68	-482,42	84,07	240,44
	d9	-591,51	-460,43	131,08	-458,66	119,30	225,33
	d10	-642,76	-453,80	188,96	-448,09	170,84	215,98

Tabela C.2. Deviance residual, critério de informação da deviance (DIC), número efetivo de parâmetros (p_D), critério de informação amplamente aplicável (WAIC) e número efetivo de parâmetros (p_{WAIC}), e logaritmo da ordenada preditiva condicional (LCPO) para os diferentes modelos considerados para dados preditos no segundo estágio.

Dado	Predição	Modelo	Deviance	DIC	p_D	WAIC	p_{WAIC}	LCPO
Agregado	m1	a1	-88,38	-84,23	4,15	-80,13	7,16	39,78
		a2	-302,99	-246,67	56,33	-246,88	44,02	80,10
		a3	-365,54	-318,63	46,91	-327,13	29,19	80,51
		a4	-196,88	-166,20	30,69	-164,03	24,59	62,29
	m.tp	a1	-92,57	-88,42	4,15	-86,02	5,71	42,79
		a2	-169,72	-153,21	16,52	-138,70	25,47	62,67
		a3	-342,38	-293,95	48,42	-294,65	36,35	50,08
		a4	-348,13	-296,88	51,25	-304,73	33,32	43,04
	m.tnp	a1	-164,33	-160,08	4,26	-152,73	9,61	75,65
		a2	-164,09	-158,36	5,74	-150,16	11,35	74,14
		a3	-162,53	-154,57	7,96	-145,52	13,63	72,20
		a4	-184,23	-155,03	29,21	-154,33	22,77	60,59
	m.i1	a1	-163,35	-159,10	4,25	-151,74	9,61	75,16
		a2	-163,13	-157,40	5,73	-149,19	11,35	73,66
		a3	-161,59	-153,64	7,95	-144,58	13,64	71,73
		a4	-182,87	-153,76	29,11	-153,06	22,70	60,16
	m.i2	a1	-204,88	-200,44	4,44	-197,90	6,16	98,76
		a2	-207,67	-201,47	6,20	-198,44	7,99	98,85
		a3	-207,25	-198,35	8,89	-194,73	10,51	96,51
		a4	-217,40	-185,07	32,34	-189,58	21,37	82,83
	m.i3	a1	-197,34	-192,95	4,39	-189,53	6,80	94,49
		a2	-199,25	-193,17	6,08	-189,47	8,39	94,32
		a3	-197,76	-189,14	8,63	-185,29	10,46	92,04
		a4	-198,67	-167,88	30,79	-172,13	20,42	76,42
	m.i4	a1	-201,96	-197,54	4,42	-195,25	5,95	97,45
		a2	-204,43	-198,26	6,16	-195,48	7,77	97,40
		a3	-203,87	-195,07	8,80	-191,73	10,23	95,06
		a4	-211,29	-179,46	31,82	-183,95	21,01	80,88
Desagregado	m1	d1	-2283,09	-2278,23	4,86	-2275,27	7,76	1137,63
		d2	-7271,65	-7174,88	96,76	-7167,88	96,36	3583,22
		d3	-7323,44	-7222,33	101,11	-7211,11	103,68	3604,70
		d4	-7331,70	-7228,45	103,25	-7218,65	104,35	3608,41
		d5	-2269,99	-2251,86	18,13	-2248,78	20,83	1124,39
		d6	-2254,32	-2220,59	33,73	-2221,71	31,72	1110,82
		d7	-2235,61	-2183,60	52,01	-2197,48	36,94	1098,70
		d8	-2233,77	-2180,04	53,74	-2180,78	50,88	1090,30
		d9	-2181,46	-2077,93	103,53	-2082,91	91,84	1040,84
		d10	-2128,00	-1974,09	153,91	-1896,99	199,89	940,62
	m.tp	d1	-2332,17	-2327,27	4,90	-2325,60	6,51	1162,80
		d2	-6811,51	-6728,45	83,06	-6723,26	82,81	3361,20

Continua na próxima página

Dado	Predição	Modelo	Deviance	DIC	p_D	WAIC	p_{WAIC}	LCPO
m.tnp	d3	-7104,42	-7000,39	104,03	-6988,46	106,70	3493,14	
	d4	-7204,16	-7094,80	109,36	-7081,64	112,35	3539,58	
	d5	-2379,88	-2356,91	22,97	-2355,15	24,23	1177,58	
	d6	-2361,64	-2321,15	40,49	-2323,55	36,68	1161,69	
	d7	-2344,41	-2285,62	58,79	-2299,19	43,43	1149,47	
	d8	-2357,33	-2303,47	53,87	-2302,99	52,08	1151,38	
	d9	-2625,83	-2388,70	237,12	-2317,96	254,82	1125,95	
	d10	-2724,41	-2428,90	295,51	-2315,55	329,10	1099,87	
	d1	-3244,83	-3238,98	5,85	-3237,81	6,97	1618,91	
	d2	-6800,42	-6717,61	82,81	-6711,56	83,29	3355,33	
m.i1	d3	-6802,56	-6713,52	89,04	-6705,73	90,20	3352,26	
	d4	-6864,47	-6770,16	94,31	-6760,42	96,49	3379,47	
	d5	-3645,67	-3616,17	29,50	-3613,61	31,25	1806,79	
	d6	-3994,37	-3948,98	45,40	-3945,20	47,26	1972,50	
	d7	-3978,80	-3914,32	64,48	-3913,42	62,22	1956,40	
	d8	-3916,04	-3822,66	93,39	-3828,38	82,00	1913,63	
	d9	-3955,77	-3849,27	106,50	-3848,80	98,87	1923,52	
	d10	-3914,19	-3758,94	155,25	-3631,02	235,02	1784,82	
	d1	-3230,06	-3224,23	5,83	-3223,04	6,97	1611,52	
	d2	-6756,43	-6674,66	81,77	-6668,69	82,31	3333,91	
m.i2	d3	-6758,48	-6670,44	88,04	-6662,77	89,22	3330,81	
	d4	-6819,19	-6726,19	93,00	-6716,46	95,33	3357,51	
	d5	-3625,87	-3596,42	29,44	-3593,84	31,22	1796,91	
	d6	-3968,06	-3922,83	45,23	-3919,09	47,06	1959,45	
	d7	-3952,51	-3888,20	64,31	-3887,54	61,85	1943,46	
	d8	-3890,14	-3797,24	92,90	-3802,90	81,63	1900,90	
	d9	-3929,65	-3823,20	106,44	-3822,56	98,98	1910,40	
	d10	-3888,33	-3733,26	155,07	-3604,28	235,63	1771,03	
	d1	-972,13	-967,81	4,32	-965,78	6,28	482,89	
	d2	-1034,13	-1025,27	8,87	-1022,56	11,40	511,28	
m.i3	d3	-1046,85	-1034,49	12,36	-1031,94	14,60	516,01	
	d4	-1059,77	-1041,84	17,93	-1040,22	19,10	520,14	
	d5	-1022,98	-1002,93	20,05	-1000,27	22,24	500,18	
	d6	-1027,44	-990,02	37,42	-987,75	38,24	493,79	
	d7	-1051,50	-991,12	60,38	-989,88	58,64	494,71	
	d8	-1343,60	-1212,96	130,64	-1190,10	137,28	591,32	
	d9	-1468,73	-1266,60	202,13	-1223,47	209,24	595,51	
	d10	-1595,56	-1331,40	264,16	-1284,62	259,51	614,23	
	d1	-832,80	-828,51	4,29	-827,51	5,25	413,76	
	d2	-867,59	-859,15	8,43	-857,99	9,49	428,99	
Continua na próxima página	d3	-869,48	-858,42	11,06	-857,38	11,92	428,70	
	d4	-870,25	-855,30	14,95	-854,59	15,42	427,29	
	d5	-892,69	-871,74	20,95	-869,95	22,32	434,97	
	d6	-889,91	-854,04	35,87	-853,67	35,03	426,76	
	d7	-906,38	-846,37	60,01	-844,83	58,63	422,22	

Dado	Predição	Modelo	Deviance	DIC	p_D	WAIC	p_{WAIC}	LCPO
m.i4	d8		-987,78	-906,23	81,56	-898,12	83,47	448,32
	d9		-1007,33	-878,32	129,00	-869,03	123,89	430,03
	d10		-1031,65	-849,85	181,80	-840,58	167,46	411,48
	d1		-843,72	-839,43	4,29	-837,85	5,81	418,93
	d2		-895,13	-886,45	8,68	-884,39	10,59	442,20
	d3		-902,26	-890,87	11,39	-889,02	13,00	444,53
	d4		-909,70	-893,35	16,35	-892,11	17,24	446,07
	d5		-891,52	-871,88	19,64	-869,63	21,48	434,83
	d6		-893,47	-856,75	36,72	-855,09	37,02	427,47
	d7		-913,28	-853,85	59,43	-852,90	57,52	426,24
	d8		-1144,52	-1025,78	118,74	-1007,37	123,81	500,98
	d9		-1225,45	-1045,54	179,90	-1015,02	182,22	495,84
	d10		-1278,13	-1050,41	227,72	-1032,54	210,19	501,42

Tabela C.3. Estimativa pontual (média) e medidas de viés e cobertura dos parâmetros, calculados para os modelos em que foram utilizados os dados brutos (agregados e desagregados). Valores de referência: $r = 0,4$, $q = 0,01$, $K = 25000$ (dados agregados) e $K = 1000$ (dados desagregados).

Dado	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
Agregado	a1	r	0,36	-0,04	-0,11	0,05	0,36
		q	0,00	-0,01	-0,96	0,01	0,00
		K	27508,53	2508,53	0,11	2885,75	0,54
	a2	r	0,41	0,01	-0,00	0,06	0,94
		q	0,00	-0,01	-0,96	0,01	0,00
		K	24558,79	-441,21	-0,01	2398,45	0,94
	a3	r	0,42	0,02	0,06	0,10	0,95
		q	0,00	-0,01	-0,96	0,01	0,00
		K	23126,64	-1873,36	-0,09	8453,66	0,90
	a4	r	0,80	0,40	0,96	0,45	0,55
		q	0,00	-0,01	-0,98	0,01	0,00
		K	27075,97	2075,97	-0,64	153706,75	1,00
Desagregado	d1	r	0,34	-0,06	-0,14	0,06	0,07
		q	0,01	-0,00	-0,10	0,00	0,38
		K	1300,65	300,65	0,28	316,29	0,00
	d2	r	0,31	-0,09	-0,22	0,09	0,02
		q	0,01	-0,00	-0,07	0,00	0,96
		K	1525,17	525,17	0,49	555,68	0,00
	d3	r	0,30	-0,10	-0,26	0,11	0,02
		q	0,01	-0,00	-0,13	0,00	0,61
		K	1637,28	637,28	0,59	674,01	0,01
	d4	r	0,32	-0,08	-0,19	0,08	0,37
		q	0,01	0,00	0,23	0,00	0,97
		K	1426,33	426,33	0,36	562,62	0,79
	d5	r	0,36	-0,04	-0,10	0,05	0,38
		q	0,01	-0,00	-0,07	0,00	0,69
		K	1231,44	231,44	0,21	248,34	0,05
	d6	r	0,39	-0,01	-0,03	0,02	0,95
		q	0,01	-0,00	-0,02	0,00	0,95
		K	1101,81	101,81	0,09	117,75	0,88
	d7	r	0,40	-0,00	-0,01	0,02	0,95
		q	0,01	-0,00	-0,01	0,00	1,00
		K	1109,00	109,00	0,11	124,97	1,00
	d8	r	0,43	0,03	0,08	0,03	0,65
		q	0,01	0,00	0,06	0,00	0,86
		K	996,64	-3,36	-0,01	44,51	0,95
	d9	r	0,51	0,11	0,26	0,11	0,00
		q	0,01	0,00	0,19	0,00	0,00
		K	829,91	-170,09	-0,17	174,82	0,02
	d10	r	0,53	0,13	0,31	0,13	0,00
		q	0,01	0,00	0,22	0,00	0,01
		K	805,11	-194,89	-0,19	198,74	0,02

Tabela C.4. Estimativa pontual (média) e medidas de viés e cobertura dos parâmetros, calculados para os modelos em que foram utilizados os dados preditos (agregados e desagregados) pelos diversos modelos considerados no primeiro estágio. Valores de referência: $r = 0,4$, $q = 0,01$, $K = 25000$ (dados agregados) e $K = 1000$ (dados desagregados).

Dado	Pred.	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
Agregado	m1	a1	r	0,18	-0,22	-0,54	0,22	1,00
			q	0,00	-0,01	-0,99	0,01	0,00
			K	-417758,42	-442758,42	0,28	4013343,07	1,00
	a2	a2	r	0,88	0,48	1,19	0,48	1,00
			q	-0,00	-0,01	-1,01	0,01	0,00
			K	-1138,29	-26138,29	-0,95	61514,04	1,00
	a3	a3	r	1,20	0,80	2,01	0,80	0,00
			q	0,00	-0,01	-0,98	0,01	0,00
			K	27353,23	2353,23	-0,57	220734,42	1,00
	a4	a4	r	0,83	0,43	1,09	0,44	1,00
			q	0,00	-0,01	-0,98	0,01	0,00
			K	5711,51	-19288,49	-0,90	41092,94	1,00
m.tp	a1	a1	r	0,08	-0,32	-0,81	0,32	0,00
			q	0,00	-0,01	-0,99	0,01	0,00
			K	231461,88	206461,88	2,26	2299887,51	1,00
	a2	a2	r	0,23	-0,17	-0,43	0,17	1,00
			q	-0,00	-0,01	-1,01	0,01	0,00
			K	-25093,47	-50093,47	-0,86	234493,13	1,00
	a3	a3	r	1,08	0,68	1,91	0,71	0,15
			q	0,00	-0,01	-0,97	0,01	0,00
			K	16898,92	-8101,08	-0,54	218206,07	1,00
	a4	a4	r	2,87	2,47	6,21	2,48	0,00
			q	0,00	-0,01	-0,59	0,01	0,00
			K	1622,98	-23377,02	-0,94	23429,59	0,00
m.tnp	a1	a1	r	0,35	-0,05	-0,13	0,06	0,19
			q	0,00	-0,01	-0,96	0,01	0,00
			K	27685,02	2685,02	0,09	3499,69	0,41
	a2	a2	r	0,35	-0,05	-0,11	0,10	0,55
			q	0,00	-0,01	-0,96	0,01	0,00
			K	31275,87	6275,87	0,06	25111,18	0,70
	a3	a3	r	0,37	-0,03	-0,16	0,17	0,57
			q	0,00	-0,01	-0,96	0,01	0,00
			K	27723,67	2723,67	0,08	16551,90	0,87
	a4	a4	r	0,84	0,44	1,28	0,51	0,39
			q	0,00	-0,01	-0,91	0,01	0,00
			K	24925,40	-74,60	-0,29	155969,07	1,00
m.i1	a1	r	0,35	-0,05	-0,13	0,06	0,21	
		q	0,00	-0,01	-0,96	0,01	0,00	
		K	27563,53	2563,53	0,09	3094,69	0,41	
	a2	r	0,37	-0,03	-0,09	0,08	0,63	

Continua na próxima página

Dado	Pred.	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
			q	0,00	-0,01	-0,96	0,01	0,00
			K	27915,63	2915,63	0,05	11466,79	0,76
		a3	r	0,37	-0,03	-0,11	0,13	0,69
			q	0,00	-0,01	-0,96	0,01	0,00
			K	24800,28	-199,72	0,05	14967,70	0,89
		a4	r	0,83	0,43	1,21	0,49	0,40
			q	0,00	-0,01	-0,91	0,01	0,00
			K	35730,33	10730,33	-0,37	196343,84	1,00
m.i2	a1		r	0,36	-0,04	-0,11	0,05	0,29
			q	0,00	-0,01	-0,96	0,01	0,00
			K	27293,21	2293,21	0,10	2699,39	0,44
		a2	r	0,46	0,06	0,04	0,16	0,77
			q	0,00	-0,01	-0,96	0,01	0,00
			K	23553,22	-1446,78	-0,04	4502,82	0,76
		a3	r	0,54	0,14	0,17	0,27	0,69
			q	0,00	-0,01	-0,96	0,01	0,00
			K	21990,04	-3009,96	-0,12	9964,77	0,75
		a4	r	0,75	0,35	0,84	0,41	0,61
			q	0,00	-0,01	-0,98	0,01	0,00
			K	-5440,70	-30440,70	-0,43	170117,03	0,97
m.i3	a1		r	0,36	-0,04	-0,11	0,05	0,34
			q	0,00	-0,01	-0,96	0,01	0,00
			K	27555,29	2555,29	0,11	3002,66	0,49
		a2	r	0,41	0,01	-0,00	0,05	0,94
			q	0,00	-0,01	-0,96	0,01	0,00
			K	24712,90	-287,10	-0,00	2294,88	0,95
		a3	r	0,42	0,02	0,08	0,09	0,96
			q	0,00	-0,01	-0,96	0,01	0,00
			K	24421,81	-578,19	-0,09	9165,14	0,88
		a4	r	0,78	0,38	0,92	0,44	0,58
			q	0,00	-0,01	-0,99	0,01	0,00
			K	-16466,00	-41466,00	-0,86	226139,12	1,00
m.i4	a1		r	0,36	-0,04	-0,11	0,05	0,30
			q	0,00	-0,01	-0,96	0,01	0,00
			K	27365,00	2365,00	0,10	2759,16	0,48
		a2	r	0,44	0,04	0,02	0,13	0,84
			q	0,00	-0,01	-0,96	0,01	0,00
			K	23821,66	-1178,34	-0,03	3876,88	0,86
		a3	r	0,48	0,08	0,12	0,19	0,79
			q	0,00	-0,01	-0,96	0,01	0,00
			K	24268,50	-731,50	-0,10	21613,93	0,83
		a4	r	0,76	0,36	0,85	0,42	0,60
			q	0,00	-0,01	-0,98	0,01	0,00
			K	-31123,52	-56123,52	-0,57	305234,26	0,99
Desagregado	m1	d1	r	0,18	-0,22	-0,54	0,22	0,00

Continua na próxima página

Dado	Pred.	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
			q	0,00	-0,01	-0,67	0,01	0,00
			K	1907,81	907,81	0,92	912,21	0,00
d2	r		0,11	-0,29	-0,73	0,29	0,00	
			q	-0,00	-0,01	-1,13	0,01	0,00
			K	280,15	-719,85	-1,38	6907,68	1,00
d3	r		0,19	-0,21	-0,52	0,23	0,10	
			q	-0,00	-0,01	-1,02	0,01	0,01
			K	-1043,88	-2043,88	-0,28	29054,96	0,99
d4	r		0,24	-0,16	-0,42	0,21	0,48	
			q	0,00	-0,01	-0,85	0,01	0,33
			K	166,45	-833,55	-0,62	7683,28	0,97
d5	r		0,18	-0,22	-0,54	0,22	0,00	
			q	0,00	-0,01	-0,67	0,01	0,00
			K	1903,21	903,21	0,91	907,47	0,00
d6	r		0,19	-0,21	-0,54	0,21	0,00	
			q	0,00	-0,01	-0,67	0,01	0,00
			K	1903,91	903,91	0,91	908,32	0,00
d7	r		0,19	-0,21	-0,54	0,21	0,00	
			q	0,00	-0,01	-0,67	0,01	0,00
			K	3639,59	2639,59	1,26	7484,99	1,00
d8	r		0,19	-0,21	-0,53	0,21	0,00	
			q	0,00	-0,01	-0,66	0,01	0,00
			K	1891,00	891,00	0,89	895,05	0,00
d9	r		0,19	-0,21	-0,53	0,21	0,00	
			q	0,00	-0,01	-0,66	0,01	0,00
			K	1882,89	882,89	0,89	887,33	0,00
d10	r		0,19	-0,21	-0,53	0,21	0,00	
			q	0,00	-0,01	-0,66	0,01	0,00
			K	1887,99	887,99	0,90	892,34	0,00
m.tp	d1	r	0,02	-0,38	-0,94	0,38	0,00	
		q	0,00	-0,01	-0,92	0,01	0,00	
		K	3671,06	2671,06	3,20	14737,79	0,38	
	d2	r	-0,10	-0,50	-1,33	0,53	0,02	
		q	-0,00	-0,01	-1,31	0,02	0,19	
		K	7581,19	6581,19	-4,06	129357,28	0,35	
	d3	r	-0,29	-0,69	-1,74	0,70	0,00	
		q	-0,01	-0,02	-2,22	0,02	0,01	
		K	-1898,62	-2898,62	-2,73	3334,29	0,11	
	d4	r	-0,14	-0,54	-1,40	0,55	0,00	
		q	-0,01	-0,02	-2,23	0,02	0,01	
		K	-30,96	-1030,96	-1,76	4565,55	0,25	
	d5	r	0,04	-0,36	-0,89	0,36	0,00	
		q	0,00	-0,01	-0,88	0,01	0,00	
		K	3601,73	2601,73	2,80	3320,15	0,02	
	d6	r	0,05	-0,35	-0,88	0,35	0,00	

Continua na próxima página

Dado	Pred.	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
			q	0,00	-0,01	-0,87	0,01	0,00
			K	3744,27	2744,27	2,76	2750,62	0,00
d7		r		0,05	-0,35	-0,87	0,35	0,00
			q	0,00	-0,01	-0,86	0,01	0,00
			K	-3078,67	-4078,67	0,82	42543,27	1,00
d8		r		0,06	-0,34	-0,85	0,34	0,00
			q	0,00	-0,01	-0,85	0,01	0,00
			K	3431,04	2431,04	2,42	2434,66	0,01
d9		r		0,05	-0,35	-0,89	0,35	0,00
			q	0,00	-0,01	-0,86	0,01	0,00
			K	-34629,84	-35629,84	1,63	483538,86	0,52
d10		r		0,10	-0,30	-0,76	0,31	0,00
			q	0,00	-0,01	-0,79	0,01	0,00
			K	77410,41	76410,41	1,93	884864,03	0,24
m.tnp	d1	r		0,25	-0,15	-0,37	0,15	0,00
			q	0,01	-0,00	-0,22	0,00	0,05
			K	1577,12	577,12	0,55	615,74	0,00
d2		r		0,29	-0,11	-0,62	0,50	0,04
			q	0,01	0,00	-0,21	0,03	0,28
			K	38602,33	37602,33	80,23	2285673,48	0,78
d3		r		0,33	-0,07	-0,60	0,54	0,05
			q	0,02	0,01	-0,17	0,03	0,24
			K	-456494,51	-457494,51	28,88	5784194,31	0,75
d4		r		0,21	-0,19	-0,80	0,45	0,01
			q	0,02	0,01	-0,15	0,04	0,28
			K	-824107,87	-825107,87	0,39	7659178,01	0,99
d5		r		0,30	-0,10	-0,24	0,10	0,01
			q	0,01	-0,00	-0,12	0,00	0,10
			K	1278,94	278,94	0,27	306,83	0,03
d6		r		0,34	-0,06	-0,14	0,06	0,11
			q	0,01	-0,00	-0,05	0,00	0,17
			K	919,96	-80,04	-0,06	107,60	0,88
d7		r		0,34	-0,06	-0,13	0,06	0,12
			q	0,01	-0,00	-0,05	0,00	1,00
			K	952,92	-47,08	-0,03	88,06	1,00
d8		r		0,34	-0,06	-0,14	0,07	0,11
			q	0,01	-0,00	-0,05	0,00	0,14
			K	1114,63	114,63	0,10	144,72	0,11
d9		r		0,36	-0,04	-0,12	0,06	0,02
			q	0,01	-0,00	-0,04	0,00	0,17
			K	881,09	-118,91	-0,11	167,48	0,09
d10		r		0,36	-0,04	-0,12	0,06	0,01
			q	0,01	-0,00	-0,03	0,00	0,16
			K	883,54	-116,46	-0,13	209,34	0,07
m.i1	d1	r		0,27	-0,13	-0,35	0,13	0,00

Continua na próxima página

Dado	Pred.	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
			q	0,01	-0,00	-0,20	0,00	0,08
			K	1496,87	496,87	0,51	533,83	0,00
d2		r		0,40	0,00	-0,42	0,55	0,05
			q	0,02	0,01	-0,04	0,03	0,40
			K	7554,81	6554,81	2,14	619177,16	0,60
d3		r		0,49	0,09	-0,43	0,69	0,06
			q	0,02	0,01	0,01	0,04	0,34
			K	54958,55	53958,55	6,42	479736,59	0,62
d4		r		0,54	0,14	-0,70	0,87	0,10
			q	0,03	0,02	0,19	0,05	0,27
			K	12169,87	11169,87	1,40	260377,50	0,83
d5		r		0,32	-0,08	-0,21	0,09	0,04
			q	0,01	-0,00	-0,11	0,00	0,18
			K	1238,09	238,09	0,22	260,62	0,03
d6		r		0,35	-0,05	-0,12	0,06	0,16
			q	0,01	-0,00	-0,03	0,00	0,35
			K	964,54	-35,46	-0,03	92,95	0,90
d7		r		0,35	-0,05	-0,12	0,06	0,18
			q	0,01	-0,00	-0,04	0,00	1,00
			K	995,71	-4,29	-0,00	86,30	1,00
d8		r		0,36	-0,04	-0,12	0,06	0,17
			q	0,01	-0,00	-0,03	0,00	0,36
			K	1087,77	87,77	0,09	115,31	0,25
d9		r		0,38	-0,02	-0,09	0,06	0,18
			q	0,01	0,00	-0,01	0,00	0,29
			K	900,11	-99,89	-0,09	145,22	0,23
d10		r		0,38	-0,02	-0,08	0,06	0,18
			q	0,01	0,00	-0,01	0,00	0,28
			K	882,16	-117,84	-0,11	158,99	0,16
m.i2	d1	r		0,33	-0,07	-0,18	0,08	0,00
			q	0,01	-0,00	-0,09	0,00	0,42
			K	1307,11	307,11	0,29	322,38	0,00
d2		r		0,29	-0,11	-0,27	0,11	0,00
			q	0,01	-0,00	-0,02	0,00	1,00
			K	1704,04	704,04	0,67	738,40	0,00
d3		r		0,29	-0,11	-0,28	0,11	0,06
			q	0,01	0,00	0,05	0,00	0,82
			K	3377,13	2377,13	1,72	7556,37	0,12
d4		r		0,37	-0,03	-0,13	0,28	0,61
			q	0,02	0,01	0,75	0,02	0,39
			K	-2785,31	-3785,31	1,33	53493,94	0,18
d5		r		0,34	-0,06	-0,13	0,06	0,08
			q	0,01	-0,00	-0,05	0,00	0,67
			K	1227,30	227,30	0,21	244,81	0,01
d6		r		0,38	-0,02	-0,06	0,03	0,63

Continua na próxima página

Dado	Pred.	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
			q	0,01	-0,00	0,00	0,00	0,95
			K	1094,82	94,82	0,08	112,05	0,83
d7		r		0,38	-0,02	-0,04	0,02	0,78
			q	0,01	0,00	0,01	0,00	1,00
			K	1107,44	107,44	0,10	123,92	1,00
d8		r		0,43	0,03	0,09	0,04	0,34
			q	0,01	0,00	0,11	0,00	0,15
			K	949,01	-50,99	-0,05	60,90	0,63
d9		r		0,51	0,11	0,28	0,12	0,00
			q	0,01	0,00	0,23	0,00	0,00
			K	798,53	-201,47	-0,20	206,16	0,00
d10		r		0,53	0,13	0,30	0,14	0,00
			q	0,01	0,00	0,26	0,00	0,00
			K	778,74	-221,26	-0,21	229,46	0,00
m.i3	d1	r		0,33	-0,07	-0,16	0,07	0,05
			q	0,01	-0,00	-0,11	0,00	0,34
			K	1322,06	322,06	0,29	343,10	0,00
	d2	r		0,30	-0,10	-0,24	0,10	0,01
			q	0,01	-0,00	-0,07	0,00	0,98
			K	1664,45	664,45	0,54	796,00	0,00
	d3	r		0,29	-0,11	-0,28	0,11	0,00
			q	0,01	-0,00	-0,11	0,00	0,69
			K	1738,03	738,03	0,63	828,85	0,00
	d4	r		0,32	-0,08	-0,20	0,09	0,31
			q	0,01	0,00	0,29	0,00	0,89
			K	1576,39	576,39	0,37	1470,91	0,73
	d5	r		0,35	-0,05	-0,11	0,05	0,33
			q	0,01	-0,00	-0,06	0,00	0,68
			K	1223,21	223,21	0,20	240,62	0,04
	d6	r		0,38	-0,02	-0,04	0,03	0,78
			q	0,01	-0,00	-0,01	0,00	0,95
			K	1080,41	80,41	0,08	103,82	0,91
	d7	r		0,39	-0,01	-0,02	0,02	0,86
			q	0,01	0,00	0,00	0,00	1,00
			K	1091,63	91,63	0,08	110,81	1,00
	d8	r		0,42	0,02	0,06	0,03	0,70
			q	0,01	0,00	0,06	0,00	0,82
			K	1009,27	9,27	-0,00	54,99	0,87
	d9	r		0,49	0,09	0,23	0,10	0,07
			q	0,01	0,00	0,17	0,00	0,02
			K	843,69	-156,31	-0,16	164,01	0,08
	d10	r		0,51	0,11	0,27	0,12	0,05
			q	0,01	0,00	0,20	0,00	0,03
			K	815,32	-184,68	-0,18	191,15	0,02
m.i4	d1	r		0,33	-0,07	-0,17	0,07	0,01

Continua na próxima página

Dado	Pred.	Mod.	Par.	Média	Viés	Viés rel.	REQM	Cob.
			q	0,01	-0,00	-0,09	0,00	0,39
			K	1306,39	306,39	0,29	321,55	0,00
d2	r		0,30	-0,10	-0,26	0,11	0,00	
			q	0,01	-0,00	-0,03	0,00	0,99
			K	1657,81	657,81	0,63	690,30	0,00
d3	r		0,29	-0,11	-0,28	0,11	0,02	
			q	0,01	-0,00	-0,02	0,00	0,84
			K	1832,92	832,92	1,22	7075,73	0,01
d4	r		0,34	-0,06	-0,16	0,07	0,51	
			q	0,02	0,01	0,55	0,01	0,65
			K	2381,79	1381,79	0,90	2057,55	0,21
d5	r		0,35	-0,05	-0,12	0,06	0,11	
			q	0,01	-0,00	-0,06	0,00	0,67
			K	1229,96	229,96	0,21	247,00	0,02
d6	r		0,38	-0,02	-0,05	0,03	0,67	
			q	0,01	-0,00	-0,01	0,00	0,95
			K	1098,34	98,34	0,09	114,80	0,83
d7	r		0,39	-0,01	-0,03	0,02	0,82	
			q	0,01	0,00	0,01	0,00	1,00
			K	1108,98	108,98	0,10	124,23	1,00
d8	r		0,43	0,03	0,08	0,04	0,45	
			q	0,01	0,00	0,09	0,00	0,28
			K	963,31	-36,69	-0,04	51,53	0,83
d9	r		0,51	0,11	0,26	0,11	0,01	
			q	0,01	0,00	0,22	0,00	0,00
			K	816,20	-183,80	-0,18	188,27	0,01
d10	r		0,52	0,12	0,28	0,12	0,00	
			q	0,01	0,00	0,23	0,00	0,00
			K	801,72	-198,28	-0,19	202,95	0,00