



UNIVERSIDADE DE ÉVORA

Mestrado em Matemática Aplicada

**Avaliação da Abundância de Recursos
Pesqueiros Utilizando a Inferência Bayesiana**

João Filipe Gonçalves Monteiro

Dissertação apresentada para obtenção do grau de
mestre em Matemática Aplicada

Co-orientadora: **Doutora Manuela Azevedo Lebre**

Orientador: **Prof. Doutor Russel Gerardo Alpizar-Jara**

ÉVORA 2004

“Este trabalho não inclui as observações e críticas feitas pelo júri”

João Filipe Gonçalves Monteiro foi bolseiro de uma Bolsa de Investigação Científica financiado pela Fundação para a Ciência e Tecnologia (FCT) no âmbito do projecto *PELÁGICOS – PLE/13/00 (FCT) – Reforço da capacidade de investigação no domínio das oceanografias e biologia das pescas – no Departamento de Recursos Marinhos (DRM), do Instituto de Investigação das Pescas e do Mar, Portugal (IPIMAR).*

Orientadora no Projecto: Doutora Manuela Azevedo Lebre

À Cláudia
Aos Meus Pais
Ao Nene

Agradecimentos

Um trabalho de investigação, como este, tem a particularidade de ser, simultaneamente, solitário e irrealizável sem a colaboração de terceiros.

Por isso mesmo, e chegado a este momento, não posso deixar de me lembrar, com especial carinho, de todos aqueles que, das mais diferentes formas, me apoiaram e de a todos endereçar os meus sinceros e calorosos agradecimentos.

Começo pelo meu orientador de dissertação, o Senhor Professor Doutor Russel Alpizar-Jara, do Departamento de Matemática da Universidade de Évora, que sempre se mostrou disponível e de quem recebi apoio e oportunas sugestões, não só durante a realização desta tese, mas também ao longo de todo o tempo em que trabalhei sob a sua orientação.

Da mesma forma, endereço os meus agradecimentos à Doutora Manuela Azevedo, minha co-orientadora e orientadora no Projecto PELÁGICOS, pela dedicada atenção que me dispensou, pela brilhante transmissão de conhecimentos e pelo afectuoso incentivo.

Especial gratidão devo-a também ao coordenador do Projecto PELÁGICOS, o Professor Doutor Yorgos Stratoudakis, que me ofereceu a excepcional oportunidade de participar neste Projecto.

Ainda à minha amiga André Trindade, uma palavra de simpatia, pela encorajadora confiança que sempre me transmitiu.

Com respeitosa amizade, recordo também o apoio, a força e o estímulo do meu primo Zé B. Galvão, que me foram muito úteis durante a parte curricular do mestrado.

Não posso deixar de realçar ainda as valiosas ideias e pistas sugeridas para a organização deste trabalho, pelo meu amigo Njalo.

Uma palavra de reconhecimento também à minha amiga Gabriela, cujo apoio na revisão desta dissertação se mostrou imprescindível.

Agradeço ainda à minha grande amiga Soraya, pela amizade e pelo grande apoio que sempre me deu.

Igualmente a minha gratidão à Karine, ao Hermes, ao Evaldo e ao Djodji, que muito me apoiaram com a sua amizade.

Um muito obrigado ao meu irmão Nene, por tudo.

O meu mais especial agradecimento à Cláudia, pelo amor, carinho e atenção.

Finalmente, agradeço à minha família, em particular, aos meus Pais – mesmo à distância, nunca deixei de sentir o seu carinho e a sua atenção, especialmente nas horas mais difíceis.

Évora, Março de 2004

João F. G. Monteiro

No âmbito deste trabalho realizou-se a seguinte comunicação (poster):

Monteiro, J. F.; Azevedo, M. e Alpizar-Jara, R. (2003). Uma abordagem bayesiana na projecção da abundância de recursos pesqueiros. *XI Congresso Anual da Sociedade Portuguesa de Estatística*. Universidade de Algarve – Faculdade de Economia. Faro.

Índice

| | Página |
|---|--------|
| RESUMO | i |
| ABSTRACT | ii |
| Lista de Abreviaturas | iii |
| Lista de Figuras | v |
| Lista de Quadros | vii |
| | |
| CAPÍTULO I – INTRODUÇÃO | 1 |
| | |
| CAPÍTULO II – INFERÊNCIA BAYESIANA VIA MCMC | 9 |
| | |
| II.1. Inferência Bayesiana | 11 |
| II.1.1. Distribuições <i>a Priori</i> Não-Informativa | 16 |
| II.1.2. Distribuições <i>a Priori</i> Conjugadas | 18 |
| II.2. Métodos de Monte Carlo via Cadeias de Markov (MCMC) | 21 |
| II.2.1. Métodos de Monte Carlo Simples | 21 |
| II.2.1.1. Técnicas de Amostragem | 23 |
| II.2.2. MCMC | 24 |
| II.2.2.1. Cadeias de Markov | 25 |
| II.2.2.2. Algoritmo Metropolis–Hastings | 32 |
| II.2.2.3. Algoritmo de Gibbs | 34 |
| II.2.2.4. Comparação entre o Algoritmo de Gibbs e o Algoritmo M–H | 36 |
| II.2.2.5. Diagnósticos de Convergências | 37 |
| II.3. Inferência Bayesiana através do BUGS | 42 |

| | |
|---|------------|
| CAPÍTULO III – UMA ABORDAGEM BAYESIANA NA AVALIAÇÃO DA ABUNDÂNCIA DE RECURSOS PESQUEIROS | 47 |
| III.1. Avaliação da Abundância de Recursos Pesqueiros | 49 |
| III.2. Fase de Diagnóstico | 56 |
| III.2.1. Análise Histórica do Manancial (<i>VPA</i>) – Modelo <i>VPA7802</i> | 56 |
| III.2.2. Implementação no WinBUGS | 58 |
| III. 3. Fase de Prognóstico | 62 |
| III.3.1. Estimação do Recrutamento | 64 |
| III.3.2. Projecção da Captura e da Biomassa Desovante | 66 |
| CAPÍTULO IV – RESULTADOS E DISCUSSÃO | 67 |
| IV.1. Diagnóstico de Convergências | 69 |
| IV.2. Resultados e Discussão | 72 |
| CAPÍTULO V – CONCLUSÕES E PERSPECTIVAS FUTURAS | 87 |
| BIBLIOGRAFIA | 93 |
| GLOSSÁRIO | 101 |
| ANEXOS | 103 |
| Anexo A – Algumas Definições | 105 |
| Anexo B – Dados sobre a sardinha obtidos no <i>WG2003</i> | 106 |
| Anexo C – Código do Modelo <i>VPA7802</i> para <i>WinBUGS</i> e Resultados | 108 |
| Anexo D – Código do Modelo <i>PROJ</i> para <i>WinBUGS</i> e Resultados | 113 |

Avaliação da Abundância de Recursos Pesqueiros Utilizando a Inferência Bayesiana

Resumo

Neste trabalho apresenta-se um modelo estruturado por idades para a avaliação de recursos pesqueiros, tendo em conta duas fases: diagnóstico e prognóstico. Tomou-se a título de exemplo o stock Ibérico da sardinha (*Sardina pilchardus*). Na fase de diagnóstico estimam-se os parâmetros envolvidos na análise histórica do manancial, nomeadamente a estrutura em número do stock, a biomassa desovante anual, o coeficiente de mortalidade por pesca e os coeficientes de capturabilidade utilizando a Inferência bayesiana. As distribuições *a posteriori* dos estimadores foram obtidas através do amostrador de Gibbs implementado no WinBUGS e a convergência das cadeias foram analisadas com o software CODA. Na fase de prognóstico realizam-se projecções estocásticas, assumindo distribuições de probabilidades para o número de indivíduos em cada idade e o coeficiente de mortalidade por pesca, tendo em conta a informação obtida na fase de diagnóstico. O recrutamento foi projectado com base na fertilidade, definida em função de uma relação stock-recrutamento, para diferentes níveis da biomassa desovante.

Palavras chave: Amostrador de Gibbs; BUGS; CODA; Inferência Bayesiana; Monte Carlo via Cadeias de Markov; Sardinha; Variabilidade no recrutamento.

Fisheries stock assessment Using Bayesian Inference

Abstract

This work shows a model structured by age for fisheries stock assessment, considering two phases: diagnostic and prognostic. The Iberian stock of the sardine (*Sardina pilchardus*) was taken as an example. In the diagnosis phase, the involved parameters in the coorte analysis which have been estimated are: number of individuals at age, annual spawning stock biomass, coefficient of fishing mortality and the coefficients of catchabilities using the Bayesian Inference. Samplings from the joint posterior distribution of the estimator were accomplished using Gibbs sampling via BUGS and chains convergence were analysed with CODA software. Stochastic projection was done in the prognostic phase, assuming distributions of probabilities for the number of individuals in each age and the coefficient of fishing mortality, regarding the information gotten in the diagnosis phase. The recruitment was projected through fertility, defined in function of stock–recruitment relationship, for different levels of the spawning stock biomass.

Keywords: Bayesian Inference; BUGS; CODA; Gibbs Sampling; Markov Chain Monte Carlo; Recruitment Variability; Sardine.

Lista de Notação e Abreviaturas

- Notações tais como Σ , ∞ , etc. não serão descritas por razões óbvias.

$h(\theta | x)$ – Distribuição *a posteriori*;

$x = (x_1, x_2, \dots, x_n)$ – Amostra de dimensão n ;

χ – Espaço amostral;

X – Variável aleatória;

\Re – Conjunto dos números reais;

θ – Designação de um parâmetro ou de um vector de parâmetros;

Θ – Espaço paramétrico;

$h(\theta)$ – Distribuição *a priori*;

$f(x | \theta)$ – Função de verosimilhança;

\propto – Proporcionalidade em relação a θ ;

\sim – Segue a distribuição (referida);

$<$ – Tem o mesmo significado que o símbolo igual, "=";

$\{X_t\} = \{X(t, \omega)\} = \{X_t(\omega)\}$ – Processo estocástico;

$P_{ij}(s, t)$ – Probabilidade de transição entre os instantes s e t ;

$\underline{P}(s, t)$ – Matriz probabilidade de transição entre o instantes s e t ;

a – Representa a idade (começa na idade 0, recrutamento, e acaba na idade 6^+);

y – Representa o ano;

F_{ay} – Coeficiente de mortalidade por pesca na idade a no ano y ;

s_{ay} – Padrão relativo de exploração na idade a do ano y ;

F_y – Nível de pesca ao longo do ano y ;

\bar{F}_y – Coeficiente de mortalidade por pesca médio das idades 2 a 5 anos no ano y ;

M_{ay} – Coeficiente de mortalidade natural na idade a no ano y ;

Z_{ay} – Coeficiente de mortalidade total na idade a no ano y ;

N_{ay} – Número de indivíduos na idade a do ano y ;

\bar{N}_{ay} – Número médio de indivíduos na idade a do ano y ;

O_{ay} – Proporção de indivíduos maduros na idade a do ano y ;

w_{ay} – Peso médio dos indivíduos na idade a no ano y ;

B_y – Biomassa no ano y ;

BD_y – Biomassa desovante no ano y ;

C_{ay} – Captura na idade a e no ano y ;

q_{ay} – Coeficiente de capturabilidade na idade a do ano y ;

I_{ay} – Captura por unidade de esforço na idade a no ano y ;

λ_y – Fertilidade no ano y ;

\xrightarrow{p} – Convergência em distribuição;

$\xrightarrow{q.c.}$ – Convergência quase certamente;

▪ Abreviaturas

S – Stock;

LP – Longo Prazo;

CP – Curto Prazo;

MP – Médio Prazo;

VPA – *Virtual Population Assessment*;

$MCMC$ – Métodos de Monte Carlo via Cadeias de Markov;

MC – Métodos de Monte Carlo;

$v.a$ – Variável Aleatória;

DAG – *Directed Acrillic Graphs*;

$i.i.d$ – Independente e Identicamente Distribuídas;

mdc – Mínimo Divisor Comum;

R – Recrutamento;

$CPUE$ – Captura Por Unidade de Esforço;

$WG2003$ – *Working Group 2003*.

Lista de Figuras

| | |
|---|----|
| Figura 2.1: Esquema da inferência clássica, adaptado de Paulino <i>et al.</i> 2003. | 12 |
| Figura 2.2: Esquema da inferência bayesiana, adaptado de Paulino <i>et al.</i> 2003. | 15 |
| Figura 2.3: Ilustração gráfica (<i>DAG</i>) de um modelo hierárquico. Sejam as variáveis aleatórias X e Y tal que X depende de Y , $X Y$, onde, a e b são constantes e o sentido das setas indica a relação de dependência. Por exemplo: a distribuição <i>a priori</i> de X é $X \sim \text{dexp}(Y)$, que por sua vez depende de $Y \sim \text{dunif}(a, b)$ | 20 |
| Figura 2.4: Move-se do estado 0 ao estado 1 com probabilidade 1. Do estado 2 move-se tanto ao estado 0 como ou estado 1 com igual probabilidade $\frac{1}{2}$, e do estado 1 passa-se ao estado 2 com probabilidade $\frac{1}{3}$, caso contrário fica-se no estado 1. É necessário fazer um <i>loop</i> no estado 1 para somar a probabilidade 1, adaptado de Norris 1999. | 27 |
| Figura 3.1: Ciclo de vida de uma coorte, segundo (Cadima 2000). | 49 |
| Figura 3.2: Para uma análise histórica desde 1978 até ao ano 2002 é necessário fazer suposições para o recrutamento, em todos anos, e também relativamente aos indivíduos nas idades compreendidas entre 1 a 6+ do ano de partida, pois são desconhecidos. A primeira linha e a primeira coluna cinzenta correspondem ao número de indivíduos desconhecidos. A diagonal de cor acinzentada representa a evolução de uma coorte. | 57 |
| Figura 3.3: Ilustração gráfica, <i>DAG</i> , do Modelo <i>VPA7802</i> . As constantes e as variáveis aleatórias estão representadas por quadrados e elipses respectivamente. No <i>WinBUGS</i> , os <i>loop</i> começam a partir de 1, daí as classes de idade (a) variarem entre 1 e 7 e os anos (y) entre 1 e 25. A informação por amostragem está presente por C_{ay} e I_{ay} . As variáveis de interesse são F_{ay} e os q_{ay} . M é constante. | 62 |
| Figura 3.4: Estrutura do manancial ao início e ao final do ano y . Note-se que o final do ano y coincide com o início do ano $y + 1$, adaptado de Cadima 2000. | 63 |
| Figura 3.5: Comparação entre a forma dos modelos de stock-recrutamento de Beverton-Holt e Ricker. O eixo das abcissas representa o stock e o eixo das ordenadas representa o recrutamento resultante. | 64 |
| Figura 3.6: <i>DAG</i> do Modelo <i>PROJ</i> | 66 |
| Figura 4.1: Gráfico da análise do diagnóstico de Gelman e Rubin (1992) para duas cadeias, implementado no <i>WinBUGS</i> , para o nível de pesca (a) e recrutamento (b) no ano de 1997, isto é $y = 20$. Foram realizadas 50000 iterações. | 70 |
| Figura 4.2: Autocorrelação obtida com as 50000 iterações para o nível de pesca (a) e o recrutamento (b) em função de diferentes espaçamentos (<i>lag</i>). Para um espaçamento de 10 iterações, o valor da autocorrelação para o nível de pesca e recrutamento no ano de 1997 diminui para respectivamente 0.106 e 0.016. | 71 |
| Figura 4.3: Evolução da cadeia da distribuição <i>a posteriori</i> do coeficiente de mortalidade por pesca médio no ano 1978, para 30000 iterações, <i>burn in</i> de 1000 iterações e espaçamento de 10. | 71 |

Figura 4.4: Box-plot da distribuição *a posteriori* do padrão relativo de exploração (ano^{-1}) em cada idade, no caso em que é considerada constante ao longo dos anos, com distribuição *a priori* $\text{dunif}(0,2)$, para amostra de tamanho 4900. O centro de cada caixa corresponde a mediana e os limites superiores e inferiores correspondem aos quantis 97.5 % e 2.5% respectivamente. 73

Figura 4.5: Estimativas do padrão relativo de exploração (ano^{-1}) em cada idade no período de 1978 a 2002, obtidas com o Modelo $VPA7802$, para uma distribuição *a priori* $\text{dunif}(0,2)$, com uma amostra de tamanho 2900. 74

Figura 4.6: Estimativas da capturabilidade para cada campanha de investigação do mês Março [(a), (b), (c) e (d)] e de Novembro [(e), (f) e (g)] para o período de 1978 a 2003, obtidas com o Modelo $VPA7802$, quando o padrão relativo de exploração varia com o tempo. As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “.” representa as estimativas frequencistas (ICES 2004). 77

Figura 4.7: Estimativas do recrutamento (milhares de milhões) para o período de 1978 a 2003, obtidas com o Modelo $VPA7802$, quando o padrão relativo de exploração é constante ao longo dos anos (a) e quando varia com o tempo (b). As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “.” representa as estimativas frequencistas (ICES 2004). 80

Figura 4.8: Estimativas da biomassa desovante (t) para o período de 1978 a 2003, obtidas com o Modelo $VPA7802$, quando o padrão relativo de exploração é constante ao longo dos anos (a) e quando varia com o tempo (b). As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “.” representa as estimativas frequencistas (ICES 2004). 81

Figura 4.9: Estimativas do coeficiente de mortalidade por pesca médio (ano^{-1}) para o período de 1978 a 2003, obtidas com o Modelo $VPA7802$, quando o padrão relativo de exploração é constante ao longo dos anos (a) e quando varia com o tempo (b). As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “.” representa as estimativas frequencistas (ICES 2004). 83

Figura 4.10: Relação $S - R$ entre 1978 e 2002 para as estimativas obtidas no Modelo $VPA7802$ 84

Figura 4.11: Categorias para a biomassa desovante em função da fertilidade λ_j 85

Figura C.1: Estimativas da capturabilidade para a cada campanha de investigação portuguesa do mês Março (a), portuguesa de Novembro [(b), (c) e (d)] e espanhola de Março [(f), (g) e (h)] para o período de 1978 a 2003, obtidas com o Modelo $VPA7802$, quando o padrão relativo de exploração é constante ao longo dos anos. ... 112

Lista de Quadros

| | |
|--|-----|
| Quadro 4.1: Análise de diagnóstico de Geweke ($Z - score$), realizadas no CODA, para o nível de pesca nas idades 0 a 6+. Foram consideradas amostras de tamanho 50000 para as cadeias, com espaçamento de tamanho 1. Fracção para a primeira janela= 0.1. Fracção para a última janela =0.5. | 70 |
| Quadro B.1: Dados da captura (em milhões) por ano e classe de idades, C_{ay} | 106 |
| Quadro B.2: Dados da série da campanha portuguesa de Março (em milhões) por ano e idades, $I1_{ay}$ | 106 |
| Quadro B.3: Dados da série da campanha portuguesa de Novembro (em milhões) por ano e idades, $I2_{ay}$ | 106 |
| Quadro B.4: Dados da série da campanha espanhola de Março (em milhões) por ano e idades, $I3_{ay}$ | 107 |
| Quadro B.5: Dados da ogiva de maturação por ano e idades, O_{ay} | 107 |
| Quadro B.6: Dados do peso médio (em kg) por ano e idades, w_{ay} | 107 |
| Quadro B.7: Estimativas do recrutamento (milhares de milhões), biomassa desovante (milhões), a captura (milhões) e o coeficiente de mortalidade por pesca médio (ano ⁻¹) por ano obtidas pelo $WG2003$, R , BD_y , C_y , \bar{F}_y respectivamente. | 107 |

CAPÍTULO I

Introdução

O tema desta dissertação está relacionado com a análise da evolução da abundância dos recursos pesqueiros, recursos esses que, quando devidamente geridos, podem constituir uma fonte renovável ilimitada. A avaliação de recursos é uma área da investigação pesqueira destinada a apoiar a gestão, através da recomendação das medidas consideradas adequadas à conservação e sustentabilidade dos recursos.

Deve ter-se em conta que a avaliação de recursos pesqueiros é feita em duas fases distintas: *diagnóstico* e *prognóstico* (Azevedo, 2002).

Na fase de *diagnóstico*, o objectivo consiste em caracterizar o estado de exploração do recurso pesqueiro, através da análise histórica do manancial. Esse estudo é feito mediante a utilização de modelos matemáticos e estatísticos, por meio dos quais são estimados os parâmetros anuais do *stock*, S . No caso dos modelos estruturados por idade, trata-se de modelos complexos que envolvem a estimação de vários parâmetros, nomeadamente a estrutura anual da população em número, a sua biomassa total e reprodutora e, ainda, parâmetros que caracterizam o regime de exploração da pesca. A informação utilizada nesta fase obtém-se a partir de estudos sobre a biologia das espécies, sua abundância relativa e exploração pesqueira.

Na fase de *prognóstico* o objectivo consiste em analisar, através da evolução da abundância e captura do recurso, o efeito a *longo prazo*, LP , de estratégias de gestão e o efeito a *curto prazo*, CP , e a *médio prazo*, MP , de medidas de gestão.

Para as projecções a *CP* e *MP* é necessário atribuir um valor para o recrutamento anual (número de indivíduos que, em cada ano, entram pela primeira vez na área da pescaria) nos anos a projectar, o que só é possível se se ultrapassarem duas questões de difícil resolução. A primeira tem que ver com o valor a adoptar para o recrutamento e a segunda, com a determinação do modo de procedimento para a adopção desse mesmo valor.

A necessidade de incorporar o conhecimento existente sobre os parâmetros populacionais com interesse para a avaliação e de interpretar os resultados em termos probabilísticos, veio motivar a aplicação da inferência bayesiana na área das pescas.

Esta necessidade explica o grande desenvolvimento na aplicação da inferência bayesiana nas pescas a que se assistiu desde os finais dos anos 80 e, em particular, na última década. Podem dar-se como exemplos os trabalhos de (Bergh e Butterworth 1987, Kuikka 1999, Kuikka *et al.* 1999) sobre modelos estruturados por idade e incerteza nos parâmetros; (Meyer e Millar 1999, Millar e Meyer 1999a e 1999b) sobre a implementação de modelos de avaliação no *WinBUGS*¹; (McAllister e Ianelli 1997) sobre análise bayesiana de modelos de captura estruturados por idade; (Patterson 1999) sobre incerteza nos parâmetros através dos métodos de *Monte Carlo via Cadeias de Markov, MCMC* e, mais recentemente, (Nielsen 2000, Nielsen e Lewy 2002) sobre análise de coortes (*VPA*) bayesiana.

¹ Bayesian inference Using Gibbs Sampling (BUGS), <http://www.mrc-bsu.cam.ac.uk/bugs/>. Copyright© 1996–2003: Imperial College&MRC, UK.

Este desenvolvimento foi particularmente possível devido ao avanço computacional e à disponibilização de algoritmos, em especial *MCMC*, que vieram permitir ultrapassar a impossibilidade de resolução analítica da aplicação do teorema de Bayes a problemas mais complexos, como por exemplo a análise histórica do manancial.

Na análise bayesiana, toda a inferência procede a partir da distribuição *a posteriori*, $h(\theta | x)$. Contudo, na maior parte das vezes, a expressão analítica para $h(\theta | x)$ é muito complexa, envolvendo, em regra, o cálculo de integrais múltiplos de difícil resolução.

Este problema pode ser resolvido utilizando os métodos de *Monte Carlo* simples, *MC*. Todavia, nem sempre tal é possível, uma vez que a expressão para $h(\theta | x)$ pode ser muito complexa, a ponto de impossibilitar que se gerem amostras, cuja distribuição se aproxima da distribuição de interesse $h(\theta | x)$.

Os métodos mais utilizados hoje em dia, alternativos aos métodos de *MC*, são os métodos *MCMC*. O facto de existir *software*, como por exemplo o *WinBUGS*, para a implementação de *MCMC*, faz com que a utilização desses métodos seja prática e fácil, constituindo, assim, uma grande vantagem para a sua adopção. No caso do *WinBUGS*, o algoritmo usado para gerar as cadeias é o algoritmo de Gibbs que é um caso particular do algoritmo de Metropolis–Hastings.

O trabalho desenvolvido nesta dissertação tem como objectivo aplicar a teoria bayesiana na avaliação do manancial ou *stock* Ibérico de sardinha (*Sardina pilchardus*),

que constitui um dos principais recursos pesqueiros nacionais. A sardinha, um pequeno peixe pelágico de vida curta, crescimento rápido e sexualmente madura a partir de um ano de idade, é explorada ao longo da costa Ibérica por Portugal e Espanha, utilizando como arte de pesca o cerco. Esta arte de pesca, constituída por uma única rede com um tamanho de malha autorizado de 18 mm, captura o peixe envolvendo-o pelos lados e por baixo, impedindo a sua fuga. Nos últimos 20 anos as capturas totais anuais de sardinha (Portugal e Espanha) oscilaram entre 86000 e 217000 toneladas. As capturas realizadas pela frota portuguesa são superiores às de Espanha e representaram na década de 90 cerca de 70% do total capturado.

Este trabalho incide principalmente na fase de diagnóstico, onde se estimam os parâmetros relacionados nos modelos do tipo *VPA* para o período de 1978 a 2002 com uma abordagem bayesiana. Para tal, utilizou-se a informação disponível referente à sua biologia, captura e abundância relativa obtida nas séries de campanhas científicas realizadas entre 1978 e 2003 (ICES 2004) e realizou-se uma *VPA* bayesiana utilizando o software *WinBUGS*®, ver (Spiegelhalter *et al.* 1999).

O trabalho desenvolvido na presente tese relativo à fase de prognóstico dá especial atenção ao problema relacionado com o recrutamento. De facto, a existência de uma elevada variabilidade do recrutamento, em particular nas espécies pelágicas, e os vários factores externos à pesca (factores ambientais, competição, disponibilidade de

alimentos, etc.), tornam esta tarefa complexa, árdua e muito imprevisível, vejam-se (Cushing 1996, Azevedo 2002).

Tendo em conta as estimativas obtidas na fase de diagnóstico com a abordagem bayesiana, realizam-se projecções estocásticas da abundância e captura do stock Ibérico de sardinha.

Visando atingir os objectivos acima referidos, esta dissertação apresenta o seguinte esquema de trabalho:

No Capítulo II, expõem-se as definições e os fundamentos teóricos da inferência bayesiana via MCMC.

Para tal, apresentam-se os métodos de Monte Carlo via Cadeias de Markov, bem como, os principais algoritmos para gerar cadeias de markov, Metropolis-Hastings e Gibbs e os métodos de diagnóstico de convergência das cadeias. No fim do capítulo, faz-se um breve resumo sobre a implementação de um modelo no *WinBUGS*.

O Capítulo III é dedicado à aplicação da metodologia da inferência bayesiana na avaliação do stock Ibérico de sardinha. Para além de se referirem noções básicas fundamentais para a avaliação, é apresentada a informação disponível, as suposições básicas dos modelos, as distribuições *a priori* adoptadas para os parâmetros e, finalmente a implementação dos modelos no *WinBUGS*.

O Capítulo IV é reservado à apresentação, análise e discussão dos resultados sob o ponto de vista biológico e matemático.

Finalmente, no Capítulo V são apresentadas as conclusões e as perspectivas futuras.

O trabalho desenvolvido nesta dissertação está enquadrado no projecto PELAGICOS² financiado pela Fundação para a Ciência e Tecnologia (FCT).

² Projecto PELÁGICOS PLE/13/00 (FCT). <http://ipimar-iniap.ipimar.pt/pelagicos/index.html>

CAPÍTULO II

Inferência Bayesiana via MCMC

II.1. Inferência Bayesiana

Na *inferência clássica*³, a estatística faz-se de modo indutivo, isto é, parte-se do particular (**amostra**) para o geral (**população**), ver (Murteira 1996).

A amostra entende-se como observações provenientes de experiências repetidas em condições constantes ou aproximadamente constantes e é representada por $x = (x_1, \dots, x_n)$, onde n é a dimensão da amostra.

Ao conjunto de todas as amostras possíveis, χ , designa-se por espaço amostral, $x \in \chi$.

Para os clássicos é muito importante ter em conta que uma dada amostra é apenas um conjunto de dados entre muitos outros, que poderia ser obtido em condições semelhantes. Por outro lado, a população define-se como sendo o conjunto de todas as observações possíveis feitas em condições análogas.

Quando se faz inferência estatística, considera-se que a amostra (x_1, \dots, x_n) é uma observação das *variáveis aleatórias*⁴, *v.a.*, (X_1, \dots, X_n) com função distribuição F_0 .

Neste processo, a função distribuição, isto é, a lei que caracteriza os valores que as variáveis aleatórias assumem, não é perfeitamente conhecida.

³ À inferência clássica associam-se nomes como Karl Pearson, Ronald A. Fisher e Jerzy Neyman.

⁴ Define-se como variável aleatória real X uma representação simbólica de um espaço mensurável (\mathfrak{R}, B) e de medida de probabilidade P sobre (\mathfrak{R}, B) onde \mathfrak{R} é o conjunto dos números reais e B a família dos borelianos lineares e onde para todo o conjunto $A \subset \mathfrak{R}$, $A \in B$, $P(A)$ é a probabilidade do acontecimento $X \in A$. Especificar uma variável aleatória é introduzir um espaço de probabilidade (\mathfrak{R}, B, P) , ver (Murteira 1996).

As distribuições são representadas pelas funções de densidades, que podem depender ou não do parâmetro⁵ θ , $\theta \in \Theta$.

Posto isto, propõe-se uma família de distribuições F a que pertence F_0 e o modelo estatístico, e que se escreve do seguinte modo:

$$F = \{f(x | \theta) : \theta \in \Theta\}, x \in \mathcal{X}$$

Na Figura 2.1 tem-se um resumo do modo de procedimento da inferência clássica.

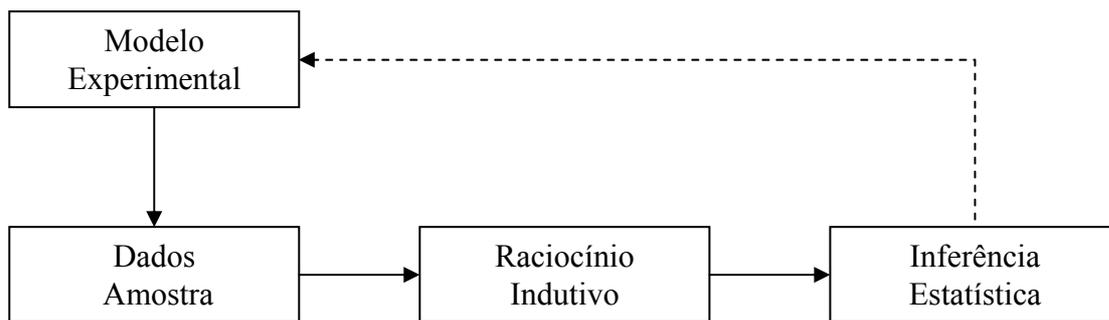


Figura 2.1: Esquema da inferência clássica, adaptado de Paulino *et al.* 2003.

Os métodos bayesianos podem ser interpretados como uma extensão da estatística clássica. Enquanto que na inferência clássica o parâmetro θ , representa um escalar ou um vector desconhecido, mas fixo, na inferência bayesiana este passa a ser uma *v.a* não observável.

⁵ Θ designa-se por espaço paramétrico.

Para os bayesianos, toda a **informação *a priori*** em relação à amostra ou experiência é demasiada importante para ser ignorada e deve ser quantificada em termos de probabilidade, que é expressa em termos de uma distribuição de probabilidades, $h(\theta)$, designada por **distribuição *a priori***.

Geralmente, a distribuição de probabilidade *a priori* é interpretada⁶ como uma medida do grau de credibilidade que uma dada pessoa, que procede à análise, atribui ao particular θ em questão. No caso contínuo, tem-se a função de densidade *a priori*, e $h(\theta)d\theta$ entende-se como sendo o grau de credibilidade que se atribui ao intervalo $(\theta, \theta + d\theta)$, $\theta \in \mathfrak{R}$.

Para além da informação anterior à experiência, considera-se ainda a **informação por amostragem**, $f(x | \theta)$, obtida através de experiências realizadas em condições idênticas. É de notar, que na inferência clássica toma-se em conta apenas a informação por amostragem, desprezando assim a informação *a priori*.

O teorema de Bayes, a seguir indicado, possibilita a actualização das credibilidades iniciais tendo em conta a informação por amostragem.

⁶ Esta é a interpretação subjectiva (dada pelos Bayesianos “subjectivos”), mas existe uma outra corrente designada por Bayesianos “objectivos” que faz uma interpretação lógica da função de probabilidade. As duas correntes contrapõem-se à interpretação frequentista da função de probabilidade dada pelos clássicos, ver (Murteira 1996).

Teorema 1 (de Bayes⁷): Sejam A_1, A_2, \dots uma partição do conjunto A . A e B são dois conjuntos arbitrários. Então para qualquer $i = 1, 2, \dots$ vem que:

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B | A_j)P(A_j)} \quad (2.1)$$

Sejam x uma realização particular da v.a. X , $f(x | \theta)$ um elemento da família F e $h(\theta)$ a distribuição *a priori*, tendo em conta o Teorema de Bayes, (2.1), resulta que a **distribuição *a posteriori***, $h(x | \theta)$, é expressa por:

$$h(\theta | x) = \frac{f(x | \theta)h(\theta)}{\int_{\Theta} f(x | \theta)h(\theta)d\theta} \propto f(x | \theta)h(\theta) \quad (2.2)$$

, onde “ \propto ” exprime a proporcionalidade em relação a θ .

É de referir que a relação de proporcionalidade expressa em (2.2) faz sentido uma vez que o denominador, $f(x) = \int_{\Theta} f(x|\theta)h(\theta)d\theta$, não depende de θ , ver (Paulino *et al.* 2003).

No contexto bayesiano a distribuição *a posteriori* é o elemento fulcral, pois constitui a base da inferência estatística.

⁷ A inferência Bayesiana deve-se ao inglês Thomas Bayes (1702–1761), que desenvolveu a teoria da probabilidade em 1764, posteriormente aceite por Laplace em 1781. Desde então, a teoria dele tem sido utilizado em diversas áreas, mas também tem gerado alguma controvérsia.

Na Figura 2.2 tem-se um resumo do modo de procedimento da inferência bayesiana.

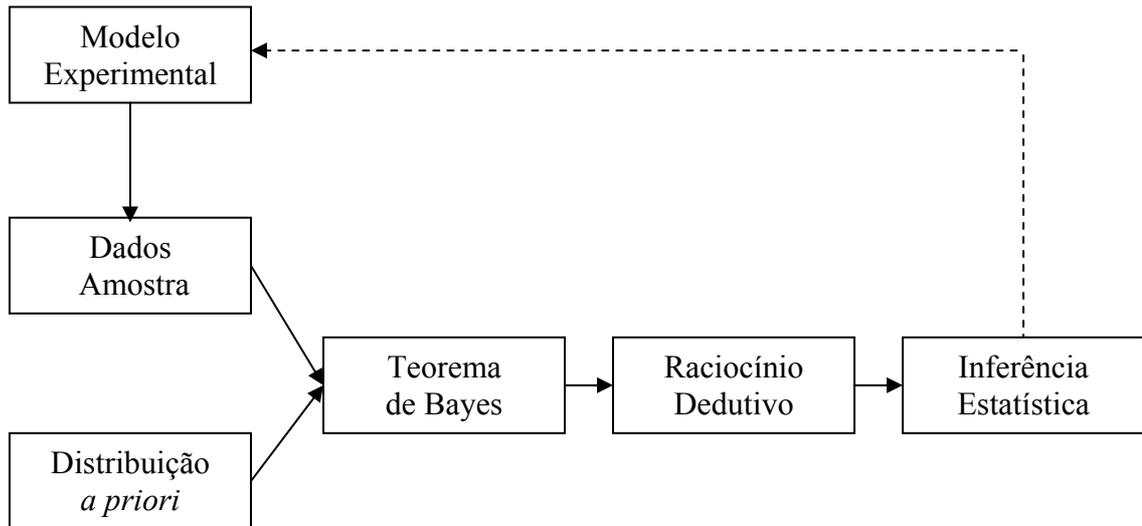


Figura 2.2: Esquema da inferência bayesiana, adaptado de Paulino *et al.* 2003.

Depois da escolha da distribuição *a priori* e da actualização por amostragem, obtém-se a distribuição *a posteriori* através da equação (2.2).

A sumarização da informação passa pela estimação pontual dos parâmetros, isto é, pelo cálculo da média *a posteriori*, da moda *a posteriori*, da mediana *a posteriori* e de outras medidas (de localização e de dispersão) de interesse. Seguindo a notação apresentada em Paulino *et al.* 2003, para $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, tem-se que:

- Média *a posteriori*: $\hat{\theta} = E[\theta | x]$, com

$$E[\theta_i | x] = \int \theta_i h(\theta | x) d\theta, \quad i = 1, \dots, k \quad (2.3)$$

- Moda *a posteriori*: $\hat{\theta}$, tal que

$$h(\hat{\theta} | \mathbf{x}) = \max_{\theta \in \Theta} h(\theta | \mathbf{x}) = \max_{\theta \in \Theta} \{h(\theta)L(\theta | \mathbf{x})\}$$

- Vector das medianas *a posteriori*: $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$

$$P\{\theta_i \geq \hat{\theta}_i | \mathbf{x}\} \geq 0.5 \quad \text{e} \quad P\{\theta_i \leq \hat{\theta}_i | \mathbf{x}\} \geq 0.5, \quad i = 1, \dots, k.$$

Pela equação (2.3) verifica-se que a operação integração desempenha um papel fundamental em inferência bayesiana. Contudo, nem sempre é possível a resolução do integral pelos métodos directos. Muitas vezes é impossível gerar amostras independentes com distribuição $h(\theta)$ através dos métodos de *MC*, uma vez que nem sempre se consegue obter uma expressão analítica para o integral do segundo membro da equação (2.3). Entretanto, durante as últimas décadas foram desenvolvidos vários algoritmos no sentido de obter a expressão para a distribuição *a posteriori*, a que se refere o algoritmo de Metropolis–Hastings e o algoritmo de Gibbs.

II.1.1. Distribuições *a Priori* Não-Informativas

Quando a informação *a priori* é inexistente, ou então, o conhecimento efectivo sobre θ é excessivamente vago, a distribuição *a priori* para o parâmetro será

minimamente informativa e neste caso é denominada por **distribuição *a priori* não-informativa**.

A maneira mais fácil e óbvia de definir a distribuição *a priori* não-informativa foi proposta por Bayes e Laplace, que sugeriram uma distribuição uniforme no espaço paramétrico Θ para o parâmetro θ . A ideia é que quando há pouca informação *a priori* deve-se adoptar a equiprobabilidade para os acontecimentos.

Se Θ for finito, isto é $\Theta = \{\theta_1, \theta_2, \dots, \theta_k\}$, não há nenhum problema com a distribuição *a priori* uniforme, e resulta que:

$$h(\theta) = \frac{1}{k}$$

Contudo, se Θ for infinito numerável, deixa de existir uma distribuição uniforme em Θ . Neste caso, $h(\theta)$ é designada por **distribuição *a priori* imprópria**. Se Θ for infinito não numerável e ilimitado, também não existe nenhuma distribuição em Θ , uma vez que $\int_{\Theta} h(\theta) d\theta$ é infinito, não satisfazendo o axioma de probabilidade total unitária.

Contudo, nem sempre a distribuição uniforme é invariante a transformações injectivas na representação formal da ignorância. Assim, para resolver este problema, Jeffreys propôs o uso da medida de informação de Fisher para $\theta \in \mathfrak{R}$, $I(\theta)$:

$$h(\theta) \propto \sqrt{|I(\theta)|},$$

onde:

$$I(\theta) = E \left[- \frac{\partial^2 \log f(x | \theta)}{\partial \theta^2} \Bigg|_{\theta} \right] \text{ e } \theta \in \mathfrak{R}, \text{ ver (Box e Tiao 1992).}$$

Jeffreys sugere que a distribuição *a priori* mais plausível para θ é aquela que apresenta maior informação de Fisher, predominando neste caso a informação por amostragem face à informação *a priori*, daí a denominação de distribuição *a priori* não-informativa.

II.1.2. Distribuições *a Priori* Conjugadas

Quando o investigador tem muita informação *a priori* sobre θ , levanta-se o problema de como quantificar essa informação de carácter subjectivo, traduzida por uma **distribuição *a priori* subjectiva**. Existem vários métodos para quantificar essa informação, ver (Paulino *et al.* 2003).

Em particular, no método preditivo de eliciação, o perito depara-se com o problema da escolha conveniente da distribuição *a priori*, $h(\theta)$, tal que $\int_{\Theta} f_T(t|\theta)h(\theta)d\theta$ se aproxime da distribuição $h_T(t)$ proposta por ele. É de notar, que o índice T representa uma estatística⁸ arbitrária e que t é um valor particular de T .

As distribuições *a priori* conjugadas, surgem então como resposta das dificuldades da eliciação das distribuições subjectivas *a priori*.

Definição 1 (Distribuição conjugada): Seja F a classe das funções de densidade de probabilidade $f(x|\theta)$. A classe Π das distribuições *a priori* é uma família para F se a

⁸ Uma estatística $T(X_1, X_2, \dots, X_n)$ é uma variável ou vector aleatório, função das n variáveis aleatórias da amostra casual, que cada amostra concreta (x_1, x_2, \dots, x_n) faz corresponder um número ou vector aleatório concreto $T(X_1, X_2, \dots, X_n)$, ver (Murteira 1996).

distribuição *a posteriori* está na classe Π para qualquer $f \in F$, todas as distribuições em Π , e qualquer $x \in \mathcal{X}$, ver (Casella e Berger 1990).

Do ponto de vista de inferência, as **distribuições conjugadas**, são muito úteis, uma vez que a determinação da distribuição *a posteriori*, quando a conjugada existe, resume-se a alterações nos parâmetros, sem necessidade de fazer cálculos adicionais, como se pode constatar no Exemplo 2.1.

Exemplo 2.1: Seja o modelo observável $dpois(\lambda)$, a distribuição de poisson, para a variável aleatória X e a distribuição *a priori* $dgamma(r, \mu)$, Gamma, para λ . Por (2.2), resulta que:

$$h(\lambda | x) \propto e^{-\lambda} \lambda^x \lambda^r \lambda^{r-1} e^{-\frac{\lambda}{\mu}} \propto \lambda^{r+x-1} e^{-\left(\frac{1}{\mu}+1\right)\lambda}, \lambda > 0$$

que é a densidade de uma $dgamma\left(r + x, \frac{1}{\mu} + 1\right)$. Portanto, a função distribuição Gamma é conjugada ao modelo Poisson.

Nos modelos com muitos parâmetros é comum a adoção de modelos hierárquicos. Assim, a incerteza nos *hiperparâmetros*⁹ é modelada assumindo vários *níveis*¹⁰ para a distribuição *a priori*. Uma das maneiras de representar os modelos hierárquicos é através dos DAG, **Directed Acyclic Graphs**, como se pode ver na Figura 2.3.

⁹ São parâmetros que por definição dependem de outros parâmetros diferentes, que por sua vez, podem depender ou não de outros.

¹⁰ Cada nível no modelo traduz-se pelo facto de uma variável aleatória depender ou não de outras variáveis aleatórias, ver Figura 2.3.

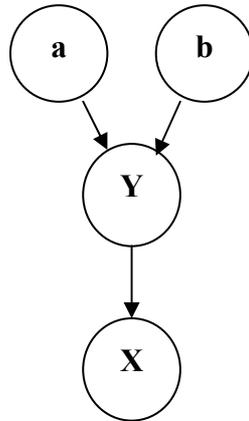


Figura 2.3: Ilustração gráfica (*DAG*) de um modelo hierárquico. Sejam as variáveis aleatórias X e Y tal que X depende de Y , $X | Y$, onde, a e b são constantes e o sentido das setas indica a relação de dependência. Por exemplo: a distribuição *a priori* de X é $X \sim \text{dexp}(Y)$, que por sua vez depende de $Y \sim \text{dunif}(a, b)$.

Geralmente, os modelos bayesianos hierárquicos apresentam apenas dois níveis. A distribuição *a priori* do primeiro nível é quase sempre uma distribuição conjugada do modelo amostral, quando tal é possível, no intuito de facilitar a implementação computacional. Entretanto, a distribuição do último nível é muitas vezes de carácter não-informativo, uma vez que se torna muito difícil descrever os hiperparâmetros quando o nível é superior a dois.

II.2. Métodos de Monte Carlo via Cadeias de Markov

No método de *MCMC*, tendo em conta o Teorema Ergódico, geram-se cadeias de markov através do algoritmo de Gibbs ou de Metropolis–Hastings, *M – H*, tais que sejam homogêneas irredutíveis ergódicas e portanto, aproximam-se da distribuição desejada.

É de referir que há outros algoritmos que aqui não foram abordados, como por exemplo, o algoritmo de **Metropolis–Hastings em blocos**, que é uma combinação dos dois algoritmos, Metropolis–Hastings e Gibbs, Gamerman (1999).

II.2.1. Métodos de Monte Carlo Simples

Os métodos de Monte Carlo simples são sem dúvida os métodos mais utilizados para resolver problemas que envolvem cálculos de integrais complexos, sendo muito úteis sob o ponto de vista de inferência bayesiana, pois permitem resolver integrais do tipo:

$$\int g(\theta)h(\theta | x)d\theta \tag{2.4}$$

O método consiste em gerar, caso seja possível, uma sequência $\theta_1, \theta_2, \dots, \theta_n$, suficientemente grande de amostras aleatórias independentes da distribuição *a posteriori*

$h(\theta | x)$, tal que à luz da lei dos grandes números, ver Teorema A.1 em anexo, o método de *MC* aproxima o integral do tipo (2.4) à média da amostra gerada, ou seja $\int g(\theta)h(\theta | x)d\theta \approx \hat{E}[g(\theta) | x]$, tal que:

$$\hat{E}[g(\theta) | x] = \frac{1}{n} \sum_{i=1}^n g(\theta_i) \quad (2.5)$$

É de notar que, $g(\theta_1), g(\theta_2), \dots, g(\theta_n)$ são independentes, pois $\theta_1, \theta_2, \dots, \theta_n$ também o são. Portanto, $\frac{1}{n} \sum_{i=1}^n g(\theta_i)$ converge quase certamente para $E[g(\theta) | x]$.

O teorema do limite central, Teorema A.2 em anexo, permite determinar qual é o número de amostras, n , necessárias que devem ser geradas para garantir uma determinada precisão, isto é, a estimativa para o erro padrão do método de Monte Carlo, ver (Paulino *et al.* 2003):

$$\frac{\sqrt{\sum_{i=1}^n \left[g(\theta_i) - \frac{1}{n} \sum_{i=1}^n g(\theta_i) \right]^2}}{\sqrt{n(n-1)}} \quad (2.6)$$

Da equação (2.6), verifica-se que a precisão da aproximação do método de *MC*, será tanto melhor, quanto maior for o tamanho da amostra. Além do mais, o método de *MC* constitui uma ferramenta poderosa para determinar aproximações de equações do tipo (2.4), pois, hoje em dia, existem computadores potentes, que podem gerar amostras com dimensões elevadas. A grande limitação do método de *MC* reside na dificuldade de gerar amostras das distribuição *a posteriori* $h(\theta | x)$ quando a expressão analítica é muito complexa.

II.2.1.1. Técnicas de Amostragem

As técnicas mais utilizadas para gerar amostras aleatórias *i.i.d* de uma determinada distribuição $h(\theta | x)$ são: o método por inversão e o método de rejeição.

Método de inversão

Seja X uma variável unidimensional com função densidade de probabilidade $f(x)$. A função distribuição de X é dada por $F(X \leq x)$, $\forall x \in \mathfrak{R}$ e define-se $F^{-1}(u) = \min\{x | F(x) \geq u\}$ como sendo função inversa da função distribuição. Geram-se amostras aleatórias *i.i.d* com distribuição uniforme no intervalo $(0,1)$, U_1, U_2, \dots e obtém-se a amostra desejada fazendo $X_1 = F^{-1}(U_1), X_2 = F^{-1}(U_2), \dots$:

1. $U \sim Unif(0,1)$;
2. $X = F^{-1}(U)$.

A eficácia do método é tanto maior quanto menor for o grau de dificuldade no cálculo de $F^{-1}(U)$.

Método de rejeição

O método de rejeição é utilizado quando a função densidade de probabilidade f é de difícil tratamento analítico, o que se traduz por lentidão do método de amostragem

por inversão. Nestes casos, toma-se como g_u uma nova função densidade de probabilidade, de fácil tratamento analítico, da variável aleatória X , e existe uma constante $K (K > 1)$, tal que:

$$f(x) \leq K g_u(x), \forall x$$

Geram-se amostras X de g_u e U de uma uniforme no intervalo $(0,1)$. Quando $KU \leq \frac{f(x)}{g_u(x)}$, X é aceite como um valor gerado de f :

1. $U \sim Unif(0,1)$;
2. Gera-se X de g_u ;
3. X é aceite como um valor gerado de $f \Leftrightarrow KU \leq \frac{f(x)}{g_u(x)}$

A escolha da constante K e da função densidade auxiliar g_u deve ser feita com muito cuidado, de tal maneira que $K g_u(x)$ seja o mais próximo de $f(x)$ e que g_u seja de fácil cálculo.

II.2.2. MCMC

As definições e os teoremas aqui apresentados têm como finalidade uma percepção dos algoritmos utilizados nos métodos de Monte Carlo via Cadeias de Markov.

Por questões de comodidade, a notação utilizada nas secções será diferente da estipulada nesta dissertação.

II.2.2.1. Cadeias de Markov

As cadeias de markov são *processos estocásticos*¹¹, onde se estuda a independência e o comportamento limite de uma determinada colecção de variáveis aleatórias.

Um processo de markov é um processo estocástico em que conhecido o presente, o futuro é independente do passado (para o efeito da previsão do futuro).

De uma maneira formal, diz-se que um processo é de markov, se:

$$P(X_t = x_j | X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_{n-1}} = x_{n-1}, X_{t_n} = x_n) = P(X_t = x_j | X_{t_n} = x_n)$$

onde X_t é o estado em que o sistema se encontra no instante t e $t_1 < t_2 < \dots < t_n$.

Definição 2 (Cadeias de Markov em tempo discreto): É um processo de markov com conjunto de índices discreto I , $I = \{0, 1, 2, \dots\}$, e com espaços de estados discreto S , $S = \{0, 1, 2, \dots\}$.

Seja X_t , com $t \in I$, uma cadeia de markov. Para caracterizar o processo, isto é, para obter as probabilidades em qualquer instante, basta conhecer:

- as probabilidades iniciais do processo, $\underline{p}^{(0)}$:

$$\underline{p}^{(0)} = [p_0^{(0)}, p_1^{(0)}, p_2^{(0)} \dots]$$

¹¹ Dado um espaço de probabilidade (Ω, A, P) e um conjunto arbitrário T , espaços de índices ou parâmetros, um processo estocástico é uma função real e finita $X(t, \omega)$, definida no produto cartesiano $T \times \Omega$ que para cada τ fixo $\tau \in T$, é mensurável de ω . Neste caso o processo estocástico representa um fenómeno aleatório que evolua no tempo, onde t ($t \in (0, +\infty)$) representa o tempo.

tal que $p_i^{(0)} = P[X_0 = i]$ representam as probabilidades de no instante 0 o processo estar no estado i , para $i \in S$.

– as probabilidades de transição a um passo entre um instante e outro, $p(m, m + 1)$.

Para, $s \leq t$, s e $t \in I$ e i e $j \in S$, define-se como sendo as probabilidades de transição entre s e t por:

$$P_{ij}(s, t) = P(s, i; t, j) = P[X_t = j | X_s = i]$$

Tais probabilidades podem ser organizadas na matriz de transição entre os instantes s e t :

$$\underline{P}(s, t) = \begin{bmatrix} p_{00}(s, t) & p_{01}(s, t) & p_{02}(s, t) & \vdots \\ p_{10}(s, t) & p_{11}(s, t) & p_{12}(s, t) & \vdots \\ p_{20}(s, t) & p_{21}(s, t) & p_{22}(s, t) & \vdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (2.7)$$

As equações de Chapman–Kolmogorov, estabelecem a relação entre as transições.

Sejam s , u e t tal que $s \leq u \leq t$:

$$P_{ij}(s, t) = \sum_{k \in S} P_{ik}(s, u) P_{kj}(u, t) \quad (2.8)$$

As equações de Chapman–kolmogorov, (2.8), deduzem-se usando a **propriedade de markov**¹² e representam-se na forma matricial por:

$$\underline{P}(s, t) = \underline{P}(s, u) \underline{P}(u, t) \quad (2.9)$$

¹² $P_{ij}(s, t) = P[X_t = j | X_s = i] = \sum_k P[X_t = j | X_s = i, X_u = k] = \sum_k P[X_t = j | X_u = k] P[X_u = k | X_s = i]$

A matriz (2.7), é uma matriz estocástica, isto é, os elementos são não negativos e a soma dos elementos de cada linha é igual a 1, que se pode traduzir por $\underline{P} \geq \underline{0}$ e $\underline{P}\underline{1} = \underline{1}$, onde $\underline{0}$ e $\underline{1}$ são, respectivamente, uma matriz cujos elementos são zeros e um vector coluna de "uns".

Sejam $p_j^{(n)} = P[X_n = j]$ as probabilidades no instante n . Tais probabilidades podem ser organizadas num vector linha:

$$\underline{p}^{(n)} = [p_0^{(n)}, p_1^{(n)}, p_2^{(n)}, \dots]$$

Assim, tem-se que

$$\underline{p}^{(n)} = \underline{p}^{(0)} \underline{P}(0, n) = \underline{p}^{(0)} \underline{P}(0, 1) \underline{P}(1, 2) \dots \underline{P}(n - 1, n) \tag{2.10}$$

Exemplo 2.2: A cadeia de markov apresentada na Figura 2.4 tem como matriz probabilidade de transição:

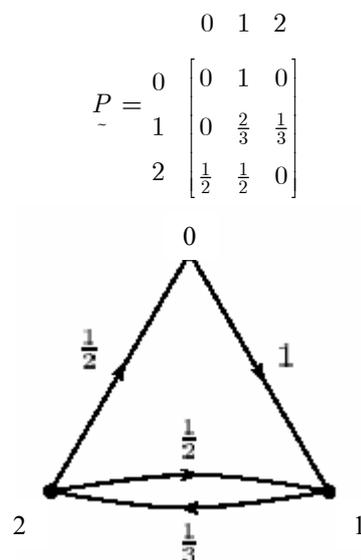


Figura 2.4: Move-se do estado 0 ao estado 1 com probabilidade 1. Do estado 2 move-se tanto ao estado 0 como ou estado 1 com igual probabilidade $\frac{1}{2}$, e do estado 1 passa-se ao estado 2 com probabilidade $\frac{1}{3}$, caso contrário fica-se no estado 1. É necessário fazer um *loop* no estado 1 para somar a probabilidade 1, adaptado de Norris 1999.

Caracterização das Cadeias de Markov

Uma **cadeia de markov** diz-se **homogênea** quando as probabilidades de transição são estacionárias, isto é, quando:

$$\underline{P}(s, t) = \underline{P}(s + \tau, t + \tau)$$

para $s \leq t$ e τ arbitrário. Considerando $\tau = -s$, tem-se que:

$$\underline{P}(s, t) = \underline{P}(0, t - s)$$

ou seja $\underline{P}(s, t)$ só depende de $t - s$. Definindo

$P_{ij}^{(n)} = P[X_n = j | X_0 = i] = P[X_{n+\tau} = j | X_\tau = i]$ como sendo as probabilidades de

transição em n passos, vem que:

$$\underline{P}^{(n)} = \underline{P}(0, n) = \underline{P}(0, 1)\underline{P}(1, 2)\dots\underline{P}(n - 1, n) = \underline{P} \cdot \underline{P} \cdot \underline{P} \dots \underline{P} = \underline{P}^n$$

Assim as equações (2.8), (2.9) e (2.10) passam a ser respectivamente (2.11), (2.12)

e (2.13):

$$P_{ij}^{(n+m)}(s, t) = \sum_{k \in S} P_{ik}^{(n)} P_{kj}^{(m)} \tag{2.11}$$

$$\underline{P}^{(n+m)} = \underline{P}^{(n)} \underline{P}^{(m)} \tag{2.12}$$

$$\underline{p}^{(n)} = \underline{p}^{(0)} \underline{P}^n \tag{2.13}$$

Por notação na equação (2.13) tem-se que: $i \neq j \Rightarrow P_{ij}^{(0)} = 0, \quad P_{ii}^{(0)} = 1$.

Antes de caracterizar os estados é necessário ter em conta algumas definições.

Definição 3: Seja $f_{ij}^{(n)} = P[X_n = j; X_{n-1}, \dots, X_1 \neq j \mid X_0 = i]$ a **probabilidade de transição em exactamente n passos** (ou de **retorno** caso $i = j$), onde n representa o tempo da primeira transição de i para j . Define-se $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ como sendo a **probabilidade de alguma vez transitar de i para j** e $\sum_{n=1}^{\infty} n f_{ij}^{(n)}$ representa o tempo médio de a primeira passagem de i para j (caso $i = j$, tempo médio de recorrência i).

Definição 4: Define-se como **período de um estado i** e representa-se por $d(i)$:

$$\forall n \geq 1, P_{ii}^{(n)} = 0, d(i) = 0 \text{ e } \exists n \geq 1, P_{ii}^{(n)} > 0, d(i) = \text{mdc}\{n \geq 1 : P_{ii}^{(n)} > 0\},$$

onde *mdc* representa o mínimo divisor comum.

Um **estado i** diz-se **aperiódico** se tem período 1. Uma **cadeia aperiódica** é aquela em que todos os estados têm período 1.

Diz-se que o **estado j** é **acessível** a partir do estado i se existir um n positivo tal que a probabilidade de transitar do estado i ao estado j em n passos for positivo, isto é:

$$\exists n \geq P_{ij}^{(n)} > 0$$

Dois **estados i e j** **comunicam-se** entre si se o estado i é acessível a partir de j e vice-versa. Um conjunto de estados que comunicam entre si forma uma **classe de equivalência**. Uma **cadeia** de markov homogénea é **irreduzível** se tem apenas uma classe de equivalência e **redutível** caso contrário.

Um estado i diz-se **recorrente** quando $f_{ii} = 1$, **recorrente nulo** se para além disso $\sum_{n=1}^{\infty} n f_{ij}^{(n)} = \infty$ e **recorrente positivo** caso $\sum_{n=1}^{\infty} n f_{ij}^{(n)} < \infty$, o que significa que o tempo para que a cadeia retorne ao mesmo estado tem média finita. O estado i diz-se **transiente** quando $f_{ii} < 1$. Quando todos os estados de uma cadeia forem recorrentes positivos, a cadeia diz-se **recorrente positiva**.

Um estado i diz-se **ergódico** quando é aperiódico e recorrente positivo e uma cadeia de Markov diz-se **ergódica** se for irredutível, aperiódica e recorrente positiva.

Distribuição Assimptótica das Cadeias de Markov

Teorema 2 (Estacionaridade): Numa classe recorrente positiva aperiódica, com estados $j = 0, 1, 2, \dots$, vem que $\sum_{i=0}^{\infty} \pi_i = 1$ e:

$$\lim_{n \rightarrow \infty} p_{jj}^{(n)} = \pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij} \quad (2.14)$$

Os π_i são únicos e obtêm-se resolvendo o sistema $\pi_i \geq 0$, $\sum_{i=0}^{\infty} \pi_i = 1$ e $\underline{\pi} = \underline{\pi} P$, ver (Karlin e Taylor 1975).

Diz-se nestas condições, que π_i é a distribuição limite ou estacionária (invariante) da cadeia.

Uma cadeia de Markov diz-se reversível se e só se satisfizer a propriedade de reversibilidade:

$$\pi_i p_{ij} = \pi_j p_{ji}, \forall i \neq j \quad (2.15)$$

É de notar, que a propriedade (2.15) implica as equações enunciadas em (2.14).

Teorema 3 (Ergodicidade): Para uma cadeia de markov homogénea ergódica, $\{\theta_j\}$, $j = 1, \dots, n$, e uma função qualquer $g(\theta)$, a média ergódica é dada por:

$$\bar{g}_n = \frac{1}{n} \sum_{j=1}^n g(\theta_j) \xrightarrow[n \rightarrow \infty]{q.c.} E_\pi[g(\theta)], \quad (2.16)$$

onde $E_\pi[g(\theta)]$ representa a média teórica da distribuição limite π .

A variância das médias amostrais \bar{g}_n é dada por:

$$Var[\bar{g}_n] = c \frac{Var_\pi[g(\theta)]}{n} \quad (2.17)$$

onde $c = 1$ caso a amostragem seja independente, com $\rho_j = 0, \forall j$, e

$$c = 1 + 2 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \rho_j \quad (2.18)$$

para médias ergódicas de uma cadeia onde ρ_j é a correlação de ordem j , $j = 1, \dots, n - 1$ da cadeia, ver (Gamerman 1999).

Portanto, quanto maior for a autocorrelação da cadeia, (2.18), maior será a variância, (2.17), menos fiável será a média ergódica, (2.16), uma vez que a precisão diminui.

II.2.2.2. Algoritmo Metropolis–Hastings

O algoritmo de Metropolis–Hastings teve como base os trabalhos desenvolvidos por Metropolis *et al.* em 1953 e mais tarde Hastings (1970).

O algoritmo Metropolis–Hastings tem como objectivo gerar uma sequência de números aleatórios $\{\theta_n\}$, com espaço de estados $\Theta \subset \mathfrak{R}^k$, de entre os quais serão aceites alguns, de acordo com a probabilidade de aceitação α , que conduzem a uma cadeia de markov com distribuição limite pretendida, $h(\theta | \mathbf{x})$.

Seja um sistema com estados x_i , de um espaço discreto qualquer, pretende-se obter uma amostra de uma distribuição que atribui a cada estado uma probabilidade $\pi_i = P[X = x_i] > 0$:

Considere uma cadeia de markov, tal que o estado x_i move ao estado x_j com uma proposta de transição q , onde a probabilidade de aceitação dos movimentos é $\alpha_{ij} = \alpha(x_i, x_j) = \min\{1, w(x_i, x_j)\}$, tal que $w_{ij} = w(x_i, x_j)$ é expressa por:

$$w_{ij} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}}$$

e a probabilidade de transição da cadeia de markov é dada por:

$$p_{ij} = q_{ij} \alpha_{ij}, \text{ para } i \neq j \text{ e } p_{ii} = 1 - \sum_{i \neq j} p_{ij} \quad (2.19)$$

Algoritmo Metropolis–Hastings:

- i) O contador de iterações inicia em $n = 1$ e assume-se um valor inicial $x^{(0)}$;
- ii) Gera-se um valor $x_j \sim q(x^{(n-1)}, \cdot)$ e um valor $u \sim U(0,1)$ independente de x_j ;
- iii) De acordo com o valor da probabilidade de aceitação,

$$\alpha_{ij} = \min\{1, w(x_i, x_j)\}, \quad (2.20)$$

escolhe-se o próximo estado $x^{(n)}$:

$$x^{(n)} = \begin{cases} x_j & \text{se } u \leq \alpha(x^{(n-1)}, x_j) \\ x_i & \text{se } u > \alpha(x^{(n-1)}, x_j) \end{cases}$$

iv) Passa-se ao estado $n + 1$. Volta-se ao passo *ii*) até que a convergência tenha sido atingida, ver (Gamerman 1999).

Nota-se, que a probabilidade de aceitação dada em (2.20) é apenas uma entre outras propostas por Hastings (1970). Existem também várias variações para o algoritmo de Metropolis–Hastings, mediante a função que se assume para a taxa de rejeição, ver (Neal 1993).

A cadeia de markov assim construída satisfaz a propriedade de reversibilidade, (2.15), logo satisfaz as condições de equilíbrio, (2.14), pois para taxas de rejeição baixas e $i \neq j$ tem-se que:

$$\pi_i p_{ij} = \pi_i q_{ij} \alpha_{ij} = \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \pi_j p_{ji}$$

Por outro lado, a unicidade da solução é garantida, escolhendo adequadamente a proposta de transição q_{ij} , de modo que forme uma cadeia de markov irreduzível e

aperiódica. Prova-se que, neste caso, a cadeia com as probabilidades de transição p_{ij} , definidas em (2.19), também será uma cadeia de Markov irredutível e aperiódica.

O algoritmo é relativamente fácil de implementar, pois, a distribuição de equilíbrio, aparece apenas na condição (2.20), sob a forma da razão π_i/π_j . Ora, basta conhecê-la a menos de uma constante.

II.2.2.3. Algoritmo de Gibbs

Primeiramente, o algoritmo de Gibbs foi desenvolvido, por Geman e Geman (1984), para a reconstrução de imagens e, mais tarde, Gelfand e Smith (1990) utilizaram-no para simular distribuições *a posteriori*. Este algoritmo é um caso particular do algoritmo Metropolis–Hastings.

O algoritmo gera amostras a partir de distribuições arbitrárias complexas e multidimensionais, através das distribuições condicionais completas de cada variável,

$$h(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k).$$

Algoritmo de Gibbs:

i) Seja

$$\theta^0 = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$$

um valor arbitrário inicial para θ .

ii) O novo valor de θ obtém-se da seguinte maneira:

$$\theta_1^{(1)} \text{ obtém-se de } h(\theta_1 | x, \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$$

$$\theta_2^{(1)} \text{ obtém-se de } h(\theta_2 | x, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$$

$$\theta_3^{(1)} \text{ obtém-se de } h(\theta_3 | x, \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(0)})$$

⋮

$$\theta_k^{(1)} \text{ obtém-se de } h(\theta_k | x, \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)})$$

O novo valor de θ escreve-se:

$$\theta^1 = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)})$$

iii) O vector θ^1 passa a ser o novo valor inicial e repete-se o esquema anterior para gerar o θ^2 .

iv) Realizam-se n iterações até que a convergência seja atingida, obtendo-se assim uma sequência de n vectores, $(\theta^1, \theta^2, \dots, \theta^n)$, que é uma cadeia de markov homogénea, pois, os valores da cadeia dependem apenas do valor da iteração anterior e a função de transição não varia com n e é dada por:

$$p(\theta^{(n-1)}, \theta^{(n)}) = \prod_{i=1}^k p(\theta_i^{(n)} | \theta_1^{(n-1)}, \theta_2^{(n-1)}, \dots, \theta_{i-1}^{(n-1)}, \theta_{i+1}^{(n-1)}, \dots, \theta_{k-1}^{(n-1)})$$

Quando n tende para infinito, o vector aleatório $(\theta^1, \theta^2, \dots, \theta^n)$ converge em distribuição para um vector aleatório com função de densidade conjunta $h(\theta | x)$. A cadeia assim construída, satisfaz o Teorema 3, e portanto, tem-se o valor esperado de $g(\theta)$, (2.16), em relação à distribuição $h(\theta | x)$.

II.2.2.4. Comparação entre o algoritmo de Gibbs e o algoritmo M-H

A alta correlação entre os parâmetros ou o enviesamento à direita da distribuição marginal do parâmetro θ , pode traduzir-se pela lentidão da convergência do algoritmo de Gibbs. Normalmente, este tipo de problema resolve-se por uma reparametrização, por exemplo, através de transformações lineares ou aplicando a função logaritmo. Contudo, não existe uma regra única a ser aplicada nas diferentes situações.

A escolha do núcleo de transição, q , no algoritmo M-H exige muito cuidado, pois se as taxas de aceitação forem muito baixas, serão necessárias muito mais iterações para que a convergência seja atingida utilizando o algoritmo Metropolis-Hastings.

Consequentemente, o sucesso do método de Metropolis-Hastings depende da probabilidade de aceitação não ser muito baixa e das distribuições propostas para q_{ij} serem fáceis de simular. Não é fácil encontrar propostas para q_{ij} , de modo a que a aproximação de $h(\theta | x)$ se faça sem erro.

No algoritmo de Gibbs as transições baseiam-se nas distribuições condicionais completas das componentes do vector θ . Deste modo, quanto menor for o erro entre q_i e π_i mais próximas de 1 serão as probabilidades de aceitação.

II.2.2.5. Diagnósticos de convergências

Independentemente do algoritmo, Metropolis–Hastings ou Gibbs, utilizado para gerar amostras que convergem para a função distribuição desejada, $h(\theta | x)$, levantam-se várias questões. Interessa saber, por exemplo, se existe ou não convergência e, caso exista, se é rápida ou lenta. Outro aspecto importante a considerar, é a existência, ou não, de independência entre os vectores gerados e como é que se obtém uma amostra *i.i.d* da distribuição *a posteriori*.

Alguns autores, como por exemplo Best e Cowles (1996), recomendam que a análise das cadeias deve ser feita, comparando o diagnóstico de varias técnicas, já que nenhuma dessas técnicas é muito fiável, por si só.

Apresentam-se a seguir uma descrição sucinta dos métodos propostos por alguns autores.

Gelfand *et al.* (1990)

Existem diversas técnicas para responder às questões acima apontada.

A maneira mais fácil, mas informal, de verificar a convergência foi proposta por Gelfand *et al.* (1990). Trata-se de um método *ad-hoc*, baseado em representações gráficas das estimativas das densidades e na monitorização da convergência das médias ergódicas dos parâmetros em causa. Consideram-se m cadeias paralelas e geram-se n

iterações, constrói-se então histogramas dos m valores gerados. Geram-se mais k valores e constrói-se outro histograma dos m valores, a partir das $n + k$. Comparam-se os histogramas, por exemplo sobrepondo-os, e se forem idênticos, então há convergência. Contudo, sabe-se que apesar de existir convergência até uma certa iteração j , não há garantia da convergência para além da iteração j .

Raftery e Lewis (1995) – uma cadeia

O método proposto por Raftery e Lewis (1995), para além de analisar a convergência de uma única cadeia, propõe ainda que um número M de iterações devem ser ignoradas, também designada por **período de *burn in***, N é o total de iterações a gerar para garantir a convergência da cadeia e, finalmente, os valores da cadeia a serem usadas para a análise devem ter um **espaçamento** de tamanho k .

Seja U uma função do parâmetro θ , os valores de M , N e k são determinados de modo a garantir que $P(U \leq u | x)$ é estimada com uma probabilidade d e um erro $\pm r$, isto é, procura-se o quantil q tal que:

$$P\{|P(U \leq u | x) - q| \leq r\} \geq d \quad (2.21)$$

Para tal constrói-se uma sequência Z_t formada por zeros e uns em função de $U(\theta)$ e do quantil empírico u :

$$Z_t = \begin{cases} 1 & \text{se } U \leq u \\ 0 & \text{se } U > u \end{cases}$$

A sequência Z_t não constitui uma cadeia de Markov. Assim, são escolhidos alguns valores de Z_t com o menor espaçamento possível k de tal forma que a nova sequência $Z_t^{(k)}$ seja uma cadeia de Markov de segunda ordem com matriz de transição \underline{P} :

$$\underline{P} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

e distribuição estacionária:

$$\underline{\pi} = (\pi_0 \quad \pi_1) = \left(\frac{\beta}{\alpha + \beta} \quad \frac{\alpha}{\alpha + \beta} \right)$$

onde $\pi_0 = 1 - \pi_1 = P(U \leq u | \mathbf{x})$.

Admitindo-se que o número de iterações necessárias para que a cadeia atinja o estado de equilíbrio, isto é o período *de burn in*, acontece antes que $|P(Z_m^{(k)} = i | Z_0^{(k)} = j) - \pi_j| < \varepsilon$, para $i, j = 0, 1$, então o período de *burn in* para a cadeia $Z_t^{(k)}$ acaba quando:

$$m \geq \frac{\log\left(\frac{(\alpha + \beta)\varepsilon}{\max(\alpha, \beta)}\right)}{\log(1 - \alpha - \beta)}$$

e admite-se para a cadeia inicial Z_t um período de *burn in* de tamanho $N = mk$.

Resta então saber qual o número total de iterações para garantir as condições impostas em (2.20).

Seja $\bar{Z}_t^{(k)} = \frac{1}{n} \sum_{t=1}^n Z_t^{(k)}$ uma estimativa de $P(U \leq u | x)$ e tendo em conta que $Z_t^{(k)}$

é aproximadamente normal com média q e variância $\frac{\alpha\beta(2-\alpha-\beta)}{n(\alpha+\beta)^2}$, a cadeia $Z_t^{(k)}$ satisfaz a

condição (2.21) quando:

$$n \geq \frac{(2 - \alpha - \beta)\alpha\beta}{(\alpha + \beta)^3 r^2} \left(\Phi\left(\frac{d+1}{2}\right) \right)^2$$

e a cadeia original Z_t converge para a distribuição original, $h(\theta | x)$, para $N = nk$

iterações.

Geweke (1992) – uma cadeia

Este método de diagnóstico de convergência baseia-se, sobretudo, nas técnicas usadas em séries temporais.

Pretende-se estimar o valor esperado da função $g(\theta)$ do parâmetro θ . O método de Geweke é utilizado para uma única cadeia, tendo como ideia base a comparação da média, g_B , das primeiras iterações, n_B , com a média, g_E , das últimas iterações, n_E . Espera-se que as médias sejam iguais quando há convergência.

O teste é feito admitindo-se a hipótese segundo a qual para cadeias suficientemente grandes a diferença das duas médias dividida pelo desvio padrão assintótico das diferenças, é normalmente distribuído com média 0 e desvio padrão 1, ver Gamerman (1996):

$$Z_{score} = \frac{g_B - g_E}{\frac{S_B^2}{n_B} + \frac{S_E^2}{n_E}} \xrightarrow{n \rightarrow \infty} N(0,1)$$

onde s_B^2 e s_E^2 são estimativas independentes das variâncias assintóticas de g_B , e g_E ,

A hipótese de convergência é rejeitada para valores de Z_{score} distantes de 0. No entanto, valores próximos de zero não significam necessariamente a existência de convergência, pelo que, a análise deve ser feita com algum cuidado, por exemplo, devem ser testadas diferentes funções para g .

Gelman e Rubin (1992) – cadeias múltiplas

O método desenvolvido por Gelman e Rubin (1992) compara a trajectória de duas ou mais cadeias para pontos iniciais diferentes, e avalia, considerando apenas a segunda metade do total das iterações, se estas cadeias são ou não idênticas.

Este método baseia-se em técnicas de análise de variâncias e sugere que a convergência existe quando a variância entre as cadeias for menor do que a variância dentro das cadeias.

Seja m o número de cadeias a gerar e $U(\theta)$ uma função real do parâmetro, θ , que se pretende estimar. Define-se U_{ij} como sendo o valor da função U na i -ésima cadeia da j -ésima iteração, não esquecendo que se considera apenas a segunda metade das iterações. Tendo em conta que a média e a variância para cada cadeia são dadas

respectivamente por $U_{i\cdot}$ e S_i^2 , define-se W como sendo a média das variâncias dentro das cadeias:

$$W = \frac{1}{m} \sum_{i=1}^m S_i^2, \text{ onde } S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (U_{ij} - \bar{U}_{i\cdot})^2 \text{ e } \bar{U}_{i\cdot} = \frac{1}{n} \sum_{j=1}^n U_{ij}$$

e B representa a variância entre as médias das cadeias:

$$\sigma^2 = \frac{n-1}{n} W + \frac{1}{n} B, \text{ onde } B = \frac{n}{m-1} \sum_{i=1}^m (U_{i\cdot} - U_{\cdot\cdot})^2 \text{ e } U_{\cdot\cdot} = \frac{1}{m} \sum_{i=1}^m U_{i\cdot}$$

A convergência ou não da cadeia é medida através do factor de redução da escala definido por:

$$\sqrt{\widehat{R}} = \sqrt{\frac{\hat{\sigma}^2}{W}}$$

O método sugere que, quando o número de iterações for suficientemente grande, para valores do factor de redução da escala próximos de 1, o período de aquecimento, *burn in*, tende a terminar e as cadeias convergem.

II.7. Inferência Bayesiana através do BUGS

WinBUGS é uma versão interactiva do software Bayesian inference Using Gibbs Sampling, BUGS. BUGS é um software de livre acesso, www.mrc-bsu.cam.ac.uk/bugs, que permite fazer uma análise Bayesiana de modelos estocásticos complexos utilizando o amostrador de Gibbs.

Uma das vantagens deste software é a flexibilidade, pois a implementação de um modelo qualquer faz-se em poucas linhas de códigos e estas podem ser alteradas a qualquer altura sem ter de se reconstruir o modelo todo, ver (Spiegelhalter *et al.* 1996 e 1999).

A amostragem é feita utilizando o amostrador de Gibbs, tendo em conta a seguinte hierarquia:

1. Se a função condicional completa, (2.2), for discreta, é utilizada a amostragem por inversão;
2. Quando são escolhidas distribuições *a priori* conjugadas, a amostragem é feita pelo método directo.
3. O amostrador de rejeição adaptativo é utilizado quando a função condicional completa é log-côncavo¹³, ver (Gilks e Wild 1992);
4. Para funções truncadas utiliza-se o amostrador por blocos, ver (Neal 1997);
5. Caso nenhuma das condições acima referidas sejam satisfeitas o amostrador utilizado é o algoritmo *passeio aleatório de Metropolis-Hastings*¹⁴.

Um modelo pode ser implementado no *WinBUGS* de duas maneiras: utilizando o modo gráfico, *Doodle*, ou através do modo código do BUGS. Apesar de o modo gráfico ser mais fácil e prático de trabalhar, não tem incorporado todas as funções que existem no modo código.

¹³ A função densidade $h(x)$ é designada por log-côncavo se: $\frac{d^2 \log h(x)}{dx^2} < 0$.

¹⁴ É um caso particular do algoritmo de Metropolis-Hastings, quando a proposta de transição depende apenas da distância entre X_i e X_j , isto é, $q(X_j | X_i) = q(|X_j - X_i|)$.

No modo *Doodle*, as quantidades são representadas por nódulos, *nodes*, que, por sua vez, são independentes de todos os outros *nodes* com exceção dos *nodes* descendentes do próprio.

Existem três tipos de *nodes*:

- constantes, que são representadas por quadrados;
- determinísticos, funções lógicas de outros *nodes*;
- estocásticos, são variáveis aleatórias com uma determinada função de distribuição e que são representadas por elipses.

Uma seta simples, que representa uma dependência estocástica, indica que o *node* de chegada depende do *node* de partida, enquanto que uma seta grossa representa uma função lógica: uma igualdade ou então uma desigualdade.

No modo linha de comandos não há necessidade de declarar as variáveis uma a uma, pois o modelo é declarado entre chavetas, $\{\dots\}$, de tal modo que as quantidades à esquerda de \sim representam variáveis estocásticas, enquanto que uma quantidade à esquerda de $< -$ representa uma constante. Onde \sim quer dizer "distribuído por" e $< -$ significa "igual a".

É de salientar que o BUGS define as distribuições, por exemplo a distribuição normal, em função da média e da precisão em vez do desvio padrão. Daí a necessidade de se criar uma variável auxiliar τ , tal que $\tau = \frac{1}{\sigma^2}$, onde σ representa o desvio padrão, ver (Spiegelhalter *et al.* 1999).

Para além dos valores de entrada que definem as constantes no modelo, é necessário definir valores iniciais, *inits*, para as variáveis aleatórias para que o algoritmo de reamostragem se possa iniciar.

Os dados de entrada, *data*, e de inicialização, *inits*, são apresentados no formato do software S-PLUS^{©15}.

Os ciclos iniciam-se em 1 e não a partir do zero. Quando a distribuição *a priori* é imprópria, há necessidade de truncá-la, de tal modo que o valor inicial esteja dentro do intervalo de valores possíveis.

Para melhor compreensão e análise das cadeias, existe o software CODA^{©16} que complementa o BUGS.

CODA[©] é também um software de livre acesso, *ftp.mtc-bsu.cam.ac.uk*, que trabalha no ambiente do software S-PLUS[©], no qual é possível fazer um estudo exaustivo das cadeias, através da análise quanto à convergência, ao período de *burn in* e ao espaçamento necessário entre as iterações para garantir a independência e ausência de autocorrelação na cadeia, ver (Best e Cowles 1996).

Os métodos para análise das cadeias, desenvolvidos por Rafery e Lewis (1992), Geweke (1992), Gelman e Rubin (1992) e entre outros, estão implementados no CODA.

¹⁵ S-PLUS 2000 professional Release 3, Copyright© 1988–2000 MathSoft, Inc. All Right reserved.

¹⁶ Convergence Diagnosis and Output Analysis. Copyright© 1995 MRC Biostatistics Unit.

CAPÍTULO III

**Uma Abordagem Bayesiana na
Avaliação da Abundância de
Recursos Pesqueiros**

III.1. Avaliação dos Recursos Pesqueiros

A avaliação de um recurso pesqueiro compreende uma análise histórica da evolução da abundância da população de peixes, tendo em vista a caracterização do seu estado de exploração; a realização de projecções a curto, médio e a longo prazo do **stock**, S , e da **captura**, C ; e ainda a determinação dos pontos de referência biológica.

Com essa avaliação, pretende-se estimar os efeitos que as diferentes estratégias de exploração pesqueira exercem, a curto e a longo prazo, nas capturas e na população de peixes, com o objectivo de definir os níveis de pesca que possibilitem as maiores capturas a longo prazo, sem colocar em risco a conservação do recurso.

Para esta análise importa compreender o ciclo de vida de um grupo de peixes nascidos de uma mesma época de desova, ou seja, da **coorte**, ver Figura 3.1.

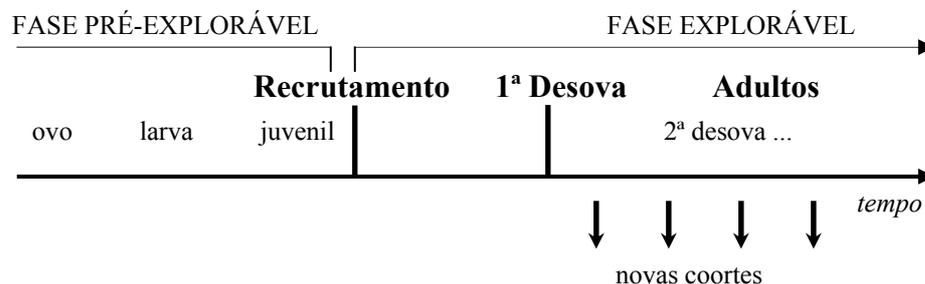


Figura 3.1: Ciclo de vida de uma coorte, segundo (Cadima 2000).

O período da fase pré-explorável é determinante no ciclo de vida de cada coorte, pois a abundância acarreta fortes variações nesta etapa devido a factores naturais, tais como ventos, correntes, temperaturas e salinidade, ver (Cushing 1996).

Ao conjunto dos indivíduos da mesma idade que, num determinado ano, entram, pela primeira vez, na fase explorável chama-se **recrutamento**, R .

Como os indivíduos de uma coorte não desovam todos, pela primeira vez, na mesma idade, o número dos que já desovaram, pelo menos uma vez, aumenta com o passar dos anos.

A curva que representa a proporção desses indivíduos designa-se por **ogiva de maturação**, O .

Os indivíduos de uma coorte morrem em consequência de factores naturais, o que se resume pelo **coeficiente de mortalidade natural**, M , e da actividade da pesca, traduzida pelo **coeficiente de mortalidade por pesca**, F .

Em cada ano, a população de peixes é constituída pelos sobreviventes de várias coortes, designando-se esse conjunto por **manancial** ou **stock**.

Os modelos de avaliação estruturados por idade são mais realistas, permitindo analisar a evolução de cada coorte de forma independente. É natural que a evolução de uma espécie, nos primeiros anos de idade, seja diferente nos últimos anos de idade. Por exemplo, os peixes podem só estar disponíveis para a pesca a partir de uma certa idade, sendo assim, tanto a mortalidade por pesca, como a mortalidade natural variam em

função da idade, vejam-se (Fournier e Archibald 1982, Pope e Shepherd 1982, Megrey 1989, McAllister e Ianelli 1997, Fournier *et al.* 1998).

A mortalidade dos indivíduos, nas diferentes idades a , $a = 1, \dots, A$, causada pela pesca durante o ano y , $y = 1, \dots, Y$, pode resultar de efeitos multiplicativos do **nível de pesca**, isto é, da intensidade da pesca ao longo do ano y , F_y , e do **padrão relativo de exploração**, que depende das propriedades selectivas do tipo de arte utilizado e varia com a idade e com o ano, s_{ay} , (3.1):

$$F_{ay} = s_{ay} \times F_y \quad (3.1)$$

O **coeficiente de mortalidade total** na idade a , no ano y , Z_{ay} , resulta da soma dos coeficientes de mortalidade, natural e por pesca, donde $Z_{ay} = F_{ay} + M_{ay}$. A **taxa de sobrevivência** dos indivíduos pode então expressar-se por $e^{-Z_{ay}}$.

Seja N_{ay} , o número de peixes na idade a , no início do ano y , por convenção referida a 1 de Janeiro, a evolução do número de peixes pode ser traduzida pela seguinte equação diferencial:

$$\frac{dN_{ay}}{dy} = -Z_{ay} \times N_{ay} \quad (3.2)$$

Sejam y_0 e y_1 , $y_1 > y_0$, assumindo-se que a pesca é feita durante um ano, que, por hipótese $y_1 - y_0 = 1$ e considerando-se que o coeficiente de mortalidade total é constante ao longo do ano, da equação (3.2) resulta que:

$$\begin{aligned}
 \frac{dN_{ay}}{N_{ay}} = -Z_a &\Leftrightarrow \int_{y_0}^{y_1} \frac{dN_{ay}}{N_{ay}} dy = \int_{y_0}^{y_1} -Z_a dy \\
 \Leftrightarrow \ln(N_{ay}) \Big|_{y_0}^{y_1} &= -Z_a y \Big|_{y_0}^{y_1} \\
 \Leftrightarrow \ln(N_{ay_1}) - \ln(N_{ay_0}) &= -Z_a \times (y_1 - y_0) \\
 \Leftrightarrow \ln(N_{ay_1})^{y_1 - y_0 = 1} &= -Z_a + \ln(N_{ay_0}) \\
 \Leftrightarrow \exp(\ln(N_{ay_1})) &= \exp(-Z_a + \ln(N_{ay_0})) \\
 \Leftrightarrow N_{ay_1} &= N_{ay_0} \times \exp(-Z_a)
 \end{aligned}$$

□

Tendo em atenção que y_0 e y_1 são arbitrários, para cada idade a , resulta, para todo o y , a seguinte equação matemática, designada por **equação de stock** (Beverton e Holt 1957):

$$N_{a+1y+1} = N_{ay} \times e^{-(F_{ay} + M_{ay})} \quad (3.3)$$

De notar que, o número de sobreviventes em cada idade, para cada ano, pode ser calculado se se conhecerem os coeficientes¹⁷ de mortalidade natural e por pesca e o recrutamento anual. Posto isto, é possível analisar a evolução de uma coorte ao longo do tempo, que também é conhecida por *Virtual Population Analysis*, *VPA*.

Conhecida a estrutura da população, em cada ano, e o peso médio dos indivíduos, em cada classe de idade, w_{ay} , é possível analisar a evolução anual em peso do stock, **biomassa**, B_y , através da equação (3.4):

$$B_y = \sum_{a=1}^A N_{ay} \times w_{ay} \quad (3.4)$$

¹⁷ Tendo em conta a expressão (3.2), deduz-se que a unidade dos coeficientes de mortalidade por pesca e mortalidade natural é expressa em ano^{-1} , pois tem-se: $\frac{[1]}{[N_{ay}]} \times \frac{[dN_{ay}]}{[dy]} = [-Z_{ay}]$ e $\frac{[1]}{[n^o]} \times \frac{[n^o]}{[tempo]} = +[Z_{ay}]$, donde $[Z_{ay}] = tempo^{-1}$. Nota-se que $[x]$ representa o maior valor inteiro de x .

e a evolução anual da **biomassa desovante**, BD_y , que representa os indivíduos em peso, no ano y , que já desovaram pelo menos uma vez, desde que conhecida a ogiva de maturação:

$$BD_y = \sum_{a=1}^A O_{ay} \times w_{ay} \times N_{ay} \quad (3.5)$$

A captura em cada idade a , ao longo do ano y , C_{ay} , define-se através da **equação de captura de Baranov (1918)**:

$$C_{ay} = \underbrace{\frac{F_{ay}}{Z_{ay}}}_{\mathbf{C1}} \times N_{ay} \times \underbrace{(1 - e^{-Z_{ay}})}_{\mathbf{C2}} \quad (3.6)$$

$\underbrace{\hspace{10em}}_{\mathbf{C3}}$

em que **C1** representa a proporção de mortes devido à pesca, **C2**, a fracção de peixes que morrerá ao longo do ano e **C3**, o número de mortes verificadas ao longo do ano.

Para além da informação da captura, também interessa conhecer a **captura por unidade de esforço** de pesca em cada idade a , ao longo do ano y , $CPUE$ ou I_{ay} , que pode ser considerada proporcional ao número médio de sobreviventes, em cada idade, a , ao longo do ano y , $\bar{N}_{ay} = N_{ay} \times \frac{1 - e^{-Z_{ay}}}{Z_{ay}}$, através do **coeficiente de capturabilidade** q_{ay} :

$$I_{ay} = q_{ay} \times \bar{N}_{ay} \quad (3.7)$$

A informação dada por $CPUE$, complementa a informação fornecida pela captura, já que esta não incorpora o esforço de cada embarcação.

Os modelos apresentados para a captura, (3.6), para o $CPUE$, (3.7), e para a evolução em número e em peso do stock, (3.3) e (3.4), são determinísticos, uma vez que não admitem incerteza nos parâmetros.

Na inferência Bayesiana, tudo o que é desconhecido ou não observável deve ser quantificado em termos probabilísticos.

Todo o conhecimento *a priori* sobre os parâmetros relacionados com a dinâmica do recurso pesqueiro (por exemplo, através do historial do recurso ou de estudos já feitos por peritos) pode ser formalizado e actualizado através da informação recolhida por amostragem. Esta mais valia não é possível com uma abordagem frequentista.

Para a compreensão da aplicação prática da análise bayesiana ao stock ibérico de sardinha, *Sardina pilchardus*, é necessário ter em conta a informação disponível para este stock (ICES 2004) e alguns pressupostos sobre a dinâmica do recurso.

Relativamente à captura, existe informação para o período de 1978 a 2002, estruturada em classes de idades (ver Quadro B.1 no Anexo B).

A sardinha está estruturada em sete classes de idades, a , desde a idade 0 até à idade 6^+ , $a = 1, \dots, 6^+$. Na última classe de idade agruparam-se todos os indivíduos com idade igual ou superior a 6 anos, pois a identificação da idade exacta para os indivíduos mais velhos é imprecisa.

Em relação ao *CPUE*, existem dados sobre três séries que foram obtidos com campanhas de investigação realizadas ao longo da costa portuguesa e espanhola (vejam-se os Quadros B.2, B.3 e B.4 no Anexo B): a série portuguesa de Março (1996 até 2002), $I1_{ay}$, a série portuguesa de Novembro (1984–1987, 1992 e 1997–2001), $I2_{ay}$, e a série espanhola de Março (1987–1988, 1990–1993 e 1996–2002), $I3_{ay}$.

A informação relativa ao peso médio individual e à proporção de indivíduos maduros em classe de idade (ogiva de maturação) está disponível para o período de 1978 a 2002 (vejam-se os Quadro B.5 e B.6 no Anexo B respectivamente).

Considera-se que o coeficiente de mortalidade natural é igual a 0.33 (Pestana 1989). Este valor corresponde a uma longevidade média entre 9 e 12 anos para probabilidades de sobrevivência de 5% e 2%, respectivamente.

Para além da informação sobre a pesca e a biologia da sardinha, acima mencionada, consideram-se os seguintes **pressupostos**:

- [1]. A população é fechada, isto é, não há emigração nem imigração dos indivíduos.
- [2]. O coeficiente de mortalidade natural é constante em todas as idades e em todos os anos;
- [3]. O coeficiente de capturabilidade de cada campanha é constante em cada idade.

III.2. Fase de Diagnóstico

III.2.1. Análise Histórica do Manancial (VPA) – Modelo VPA7802

A seguir, proceder-se-á à estimação dos parâmetros populacionais do stock ibérico de sardinha com abordagem bayesiana, nomeadamente, a estrutura em número do stock, N_{ay} , a biomassa desovante anual, BD_y , o coeficiente de mortalidade por pesca, F_{ay} , e os coeficientes de capturabilidade, q_{ay} .

Nota-se que no modelo usado, análise de coortes ou *VPA (Virtual Population Analysis)*, doravante designado por Modelo *VPA7803*, são válidos os pressupostos [1]–[3].

Análises exploratórias dos dados históricos da captura e do *CPUE* sugeriram alterações temporais (princípio da década de 90) no padrão relativo de exploração da pescaria, com uma diminuição nas idades mais jovens e um aumento nas idades mais velhas (Azevedo 1999, Skagen 2003a). Dado que esta alteração não é evidenciada em todas as subáreas da costa Ibérica onde é explorada a sardinha, consideraram-se os dois cenários possíveis para o padrão relativo de exploração, s_{ay} , durante o período de 1978 a 2002, nomeadamente:

- s_{ay} constante por idade ao longo do tempo;
- s_{ay} variável com a idade e com o tempo.

Na abordagem deste trabalho, a primeira etapa consiste em calcular o número de indivíduos para o ano $y + 1$ em todas as idades $a + 1$, através da equação de stock, (3.3).

Em particular, na aplicação prática, a ultima idade é projectada da seguinte maneira:

$$N_{6^+,y+1} = N_{5y} \times e^{-(F_{5y}+M_{5y})} + N_{6^+y} \times e^{-(F_{6^+y}+M_{6^+y})}$$

pois a última idade é agrupada.

Para tal, na primeira etapa, é necessário saber o número de indivíduos no ano y , em todas as idades, $a = 1, \dots, 6^+$, e também, o valor do coeficiente de mortalidade total, ver Figura 3.2.

| $a \setminus y$ | 1978 | 1979 | 1979 | 1980 | 1981 | 1982 | 1983 | ... | 2002 |
|-----------------|------|------|--------------|--------------|------|------|------|-----|------|
| 0 | | | | | | | | | |
| 1 | | | | | | | | | |
| 2 | | | $N_{2,1979}$ | | | | | | |
| 3 | | | | $N_{3,1980}$ | | | | | |
| 4 | | | | | | | | | |
| 5 | | | | | | | | | |
| 6 ⁺ | | | | | | | | | |



Figura 3.2: Para uma análise histórica desde 1978 até ao ano 2002 é necessário fazer suposições para o recrutamento, em todos anos, e também relativamente aos indivíduos nas idades compreendidas entre 1 a 6⁺ do ano de partida, pois são desconhecidos. A primeira linha e a primeira coluna cinzenta correspondem ao número de indivíduos desconhecidos. A diagonal de cor acinzentada representa a evolução de uma coorte.

A segunda etapa consiste em reconstruir a matriz da captura, para todos os anos em todas as idades, C_{ay} , e comparada com a captura observada.

Uma vez calculado o número de indivíduos e o coeficiente de mortalidade total, em todas as idades, para todos os anos, isto é N_{ay} e Z_{ay} , $a = 1, \dots, A$ e $y = 1, \dots, Y$, obtém-se os C_{ay} , para $a = 1, \dots, A$ e $y = 1, \dots, Y$, através da equação de captura, (3.6).

Atendendo-se à equação (3.7) e à primeira fase do *VPA*, é suficiente que se conheça o valor do coeficiente de capturabilidade em todos os anos para todas as idades, q_{ay} para $a = 1, \dots, A$ e $y = 1, \dots, Y$, para se ter os valores de I_{ay} para $a = 1, \dots, A$ e $y = 1, \dots, Y$.

Para estimar a biomassa desovante anual usou-se a expressão (3.5). Resumiu-se a informação relativa ao coeficiente de mortalidade por pesca de cada ano y , no seu valor médio entre as idades compreendidas entre 2 a 5 anos, \bar{F}_y , por serem as idades em que a sardinha se encontra completamente recrutada à pescaria.

III.2.2. Implementação no *WinBUGS*

A implementação bayesiana do modelo requer a especificação das distribuições *a priori* para todas as quantidades não observáveis, incluindo os hiperparâmetros, bem como a identificação da informação por amostragem mediante a especificação de um modelo para função de verosimilhança.

Informação por amostragem

A informação por amostragem disponível já foi referida na secção III.1. As séries de campanhas não foram realizadas no início do ano civil, pelo que é necessário introduzir uma correcção aos valores de $CPUE$, para o início de cada ano (1 de Janeiro).

Por exemplo, para as campanhas realizadas anualmente no mês de Março far-se-á a seguinte correcção:

$$I_{ay} = q_{ay} \times \frac{75}{365} e^{-Z_{ay}} \times \bar{N}_{ay}$$

onde $75/365$ representa a fracção do número de dias entre 1 de Janeiro e 15 de Março relativamente ao ano.

As funções de verosimilhança para a captura e o $CPUE$ assumem-se lognormal. Assim, o logaritmo dos dados das capturas segue uma distribuição normal de média dada pelo logaritmo da equação (3.6) e precisão τ_{ayC} . Assume-se que o logaritmo dos dados de $CPUE$ segue uma distribuição normal de média dada pelo logaritmo da equação (3.7) e precisão τ_{ayI} .

Nota 3.1: Sejam X uma *v.a.*, tal que a variável segue uma distribuição Log Normal¹⁸:

$$X \sim \text{dlnorm}(\mu, \tau)$$

$$f(x | \mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \frac{1}{x} e^{-\frac{(\ln x - \mu)^2 \tau}{2}}; x > 0, -\infty < \mu < \infty, \tau = \frac{1}{\text{Var}[X]} = \frac{1}{\sigma^2}, \sigma^2 > 0$$

então $Y = \ln(X)$ onde

$$Y \sim \text{dnorm}(\mu, \tau)$$

$$f(y | \mu, \tau) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{(y - \mu)^2 \tau}{2}}; y > 0, -\infty < \mu < \infty, \tau = \frac{1}{\text{Var}[X]} = \frac{1}{\sigma^2}, \sigma^2 > 0$$

e $E[X] = E[e^{\ln(X)}] = E[e^Y] = e^{\mu + \frac{\sigma^2}{2}}, \text{Var}[X] = e^{2(\mu + \sigma^2)} - e^{2\mu + \sigma^2}$

Distribuição a priori

A informação existente sobre o recrutamento, o número de indivíduos nas idades $a = 2, \dots, 7$ no ano de partida, $y = 1$ (1978), sobre o coeficiente de mortalidade por pesca e sobre a capturabilidade é vaga. Sendo assim, a distribuição *a priori* para estes parâmetros é de carácter não informativo, pelo que se assume, em todos os casos, uma distribuição *a priori* uniforme.

O coeficiente de mortalidade por pesca é separável, (3.1). O padrão relativo de exploração, s_a , o nível de pesca, F_y , e os coeficientes de capturabilidade de cada uma das séries de campanhas, $q_{1_{ay}}$ (portuguesa de Março), $q_{2_{ay}}$ (portuguesa de Novembro) e

¹⁸ É de referir que se adoptou para definir as distribuições a notação utilizada no *WinBUGS*.

$q3_{ay}$ (espanhola de Março), assumem valores positivos para todo $a = 1, \dots, 7$ e $y = 1, \dots, 25$

e a informação *a priori* indica que não tomam valores superiores a 2 (ICES 2004):

$$F[a, y] < -s[a, y] \times F[y]$$

$$s[a, y] \sim \text{dunif}(0, 2)$$

$$F[y] \sim \text{dunif}(0, 2)$$

$$q1[a, y] \sim \text{dunif}(0, 2) \quad q2[a, y] \sim \text{dunif}(0, 2) \quad q3[a, y] \sim \text{dunif}(0, 2)$$

Tem-se uma ideia da ordem de grandeza da abundância da sardinha ibérica, na ordem dos milhares de milhões (ICES 2004). Assim, tomou-se como recrutamento anual máximo, N_{1y} , com $y = 1, \dots, 25$, e o número de sardinhas no ano de partida, N_{a1} , $a = 2, \dots, 7$, na ordem dos 120000 mil milhões de sardinhas, tendo em conta as estimativas obtidas em (ICES 2004). Deste modo temos que:

$$U1[y] \sim \text{dunif}(0, 1)$$

$$U2[a] \sim \text{dunif}(0, 1)$$

$$N[1, y] < -U1[y] * 12.0E10$$

$$N[a, 1] < -U2[a] * 12.0E10$$

Finalmente, é de referir o facto de se desconher a precisão da captura (τC) e das séries dos *CPUE* (τI), embora se considere que o seu coeficiente de variação possa ser da ordem de 60%, o que é se reflecte no valor dos parâmetros da distribuição *a priori*:

$$\tau C[a, y] < -1 / \text{var } C$$

$$\tau I[a, y] < -1 / \text{var } I$$

$$\text{var } C \sim \text{dunif}(0, 1)$$

$$\text{var } I \sim \text{dunif}(0, 1)$$

Apresenta-se na Figura 3.3, o Modelo *VPA7802* que será implementado no *WinBUGS*, cujo o código pode ser consultado no Anexo C.

index: y from: 1 up to: 25

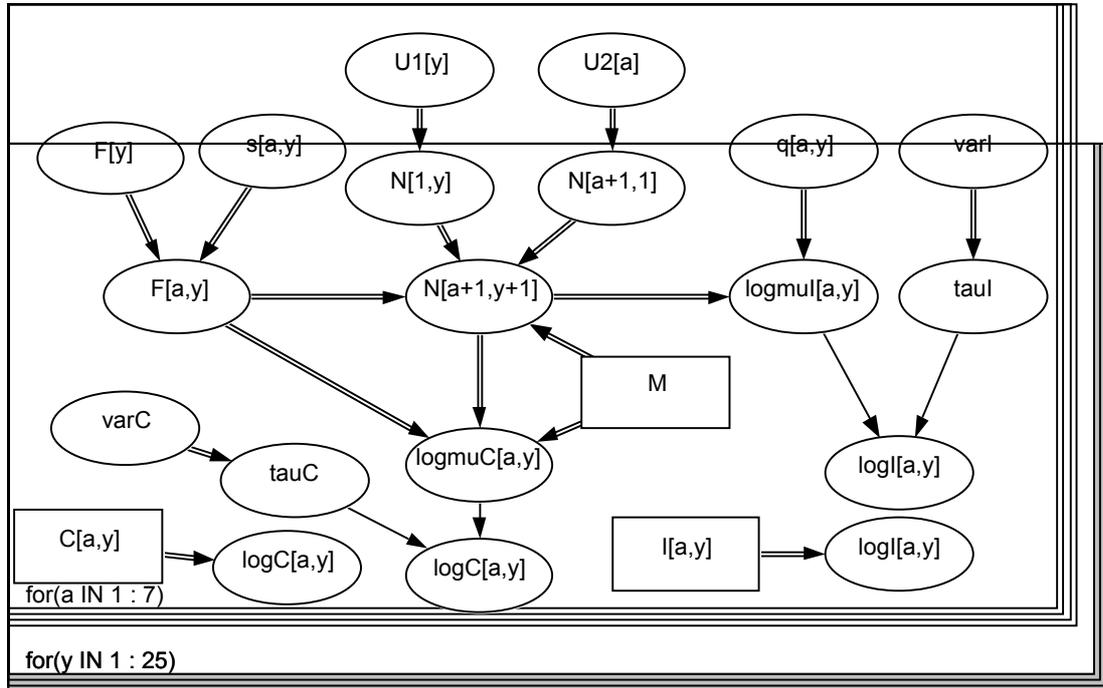


Figura 3.3: Ilustração gráfica, *DAG*, do Modelo *VPA7802*. As constantes e as variáveis aleatórias estão representadas por quadrados e elipses respectivamente. No *WinBUGS*, os *loop* começam a partir de 1, daí as classes de idade (a) variarem entre 1 e 7 e os anos (y) entre 1 e 25. A informação por amostragem está presente por C_{ay} e I_{ay} . As variáveis de interesse são F_{ay} e os q_{ay} . M é constante.

III.3. Fase de Prognóstico

Uma vez conhecida a estrutura da população ao início de um ano, digamos y , isto é, o número de indivíduos em cada coorte, N_{ay} , a ogiva de maturação, o peso médio e o coeficiente de mortalidade natural, é possível através da equação de stock, ver (3.3), projectar a população ao final desse ano, o que coincide com as estimativas da estrutura do manancial ao início do ano $y + 1$. Este procedimento pode ser feito assumindo diferentes níveis para F_{ay} . Entretanto, note-se que é sempre desconhecido o recrutamento para o ano $y + 1$, ver (Figura 3.4).

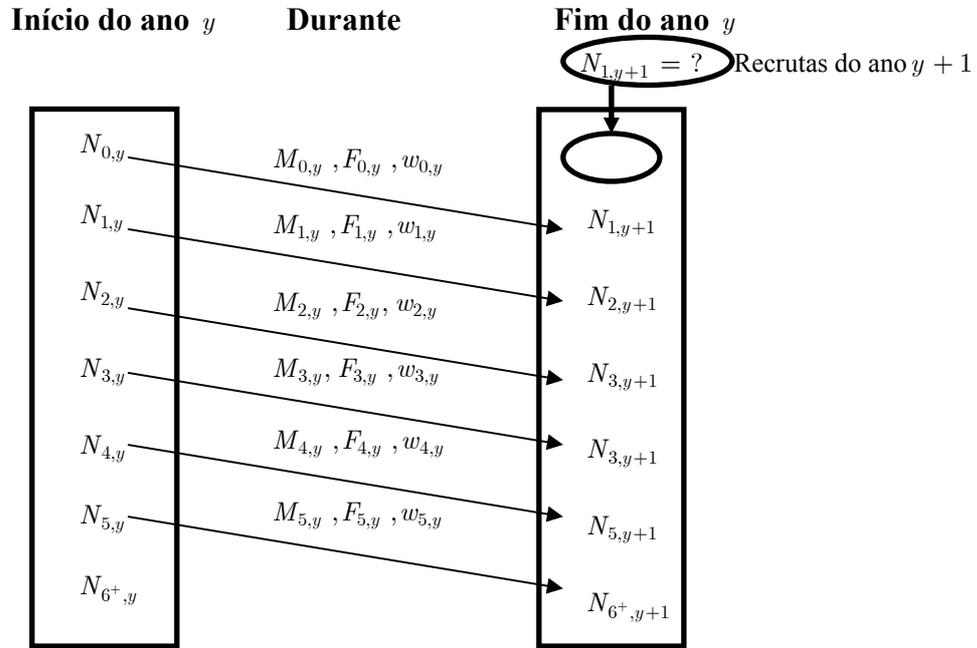


Figura 3.4: Estrutura do manancial ao início e ao final do ano y . Note-se que o final do ano y coincide com o início do ano $y + 1$, adaptado de Cadima 2000.

Para fazer projecções (predições) utilizando uma abordagem bayesiana é necessário ter informação por amostragem e definir distribuições *a priori*, ver (Paulino *et al.* 2003). Porém, no caso das projecções da abundância dos recursos pesqueiros, não existe informação por amostragem disponível para os anos a projectar, o que impede a utilização deste tipo de abordagem.

Assim, na fase de prognóstico realizam-se projecções estocásticas, atribuindo distribuições de probabilidades aos parâmetros tendo em conta a informação obtida na fase de diagnóstico com a abordagem bayesiana.

III.3.1. Estimação do Recrutamento

Para projectar as características do manancial ao início do ano $y + 2$ é necessário adoptar um valor para o recrutamento do ano $y + 1$ e assim sucessivamente, ver Figura 3.3. Portanto, a escolha de um valor para o recrutamento terá com certeza um impacto muito forte na estrutura do manancial ao longo dos anos, mas menor impacto nas projecções a curto prazo.

Existem modelos matemáticos que estabelecem uma relação entre o stock (em número ou em peso) e o recrutamento resultante. Os modelos mais usados nas pescas são os modelos de Beverton–Holt (1957) e de Ricker (1954), ver (Figura 3.5).

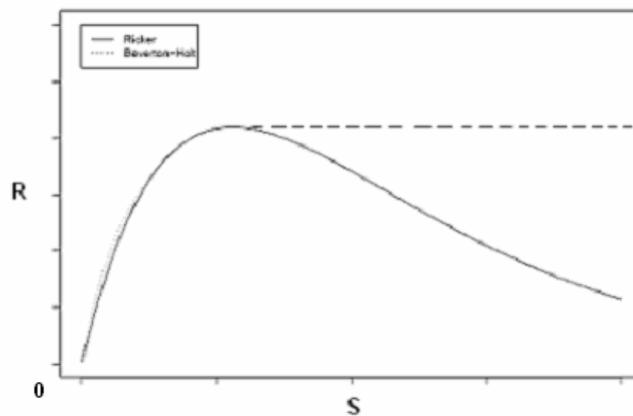


Figura 3.5: Comparação entre a forma dos modelos de stock–recrutamento de Beverton–Holt e Ricker. O eixo das abcissas representa o stock e o eixo das ordenadas representa o recrutamento resultante.

Em ambos os modelos o recrutamento aumenta com o aumento da biomassa até um valor máximo. No modelo de Beverton–Holt, que tem um recrutamento resultante

assimptótico, o recrutamento mantém-se constante a partir desse máximo, enquanto que no modelo de Ricker o recrutamento diminui com o aumento da biomassa.

Em consequência de uma elevada variabilidade do recrutamento, em particular para as espécies pelágicas, não se verifica, na prática um bom ajustamento aos dados dos modelos de $S - R$ referidos (Azevedo *et al.* 2002). Face a esta dificuldade, considera-se na presente dissertação uma abordagem alternativa para a estimação do recrutamento anual, designado por Modelo *PROJ*, com base na fertilidade da sardinha, isto é, no número de recrutas por quilograma de adulto no ano y , λ_y :

$$\lambda_y = \frac{R_y}{BD_y} \quad (3.8)$$

Tendo em conta que a fertilidade pode variar com os níveis de biomassa desovante, a abordagem adoptada para estimar o recrutamento consiste, numa primeira fase em categorizar os níveis de biomassa desovante de modo a que a variabilidade da fertilidade em cada categoria c seja a menor possível.

Para cada categoria da biomassa desovante calculou-se a fertilidade média, $mu1_c$, e a respectiva precisão, se_c .

Assim, no Modelo *PROJ* considera-se que a fertilidade, λ_c , é uma variável aleatória com a seguinte distribuição de probabilidade:

$$\lambda_c \sim dnorm(mu1_c, se_c)$$

O recrutamento em cada ano y das projecções é, então, estimado do seguinte modo, ver Figura 3.6:

- projecta-se o número de sobreviventes ao final do ano y ;
- calcula-se a biomassa desovante, BD_y , e identifica-se a respectiva categoria c a que corresponderá a *v.a* λ_c ;
- estima-se o recrutamento no ano y , \hat{R}_y , com base na relação:

$$\hat{R}_y = \hat{\lambda}_c \times BD_y \quad (3.9)$$

name: R type: logical link: identity
 value: BD*lambda

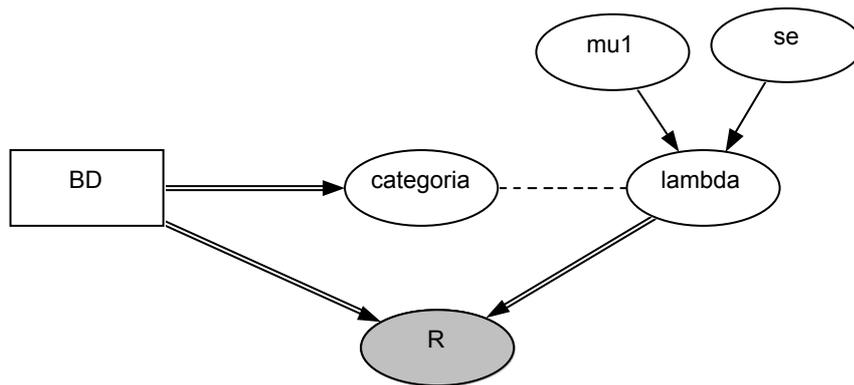


Figura 3.6: DAG do Modelo PROJ.

III.3.2. Projecções da Captura e da Biomassa Desovante

Tendo em conta os pressupostos [1]–[3] apresentados na fase de diagnóstico, projecta-se a população para o ano $y + 1$ e estimado o recrutamento, passamos a ter a estrutura completa do manancial. Pelo que, também podemos projectar a captura e biomassa desovante para o ano $y + 1$.

CAPÍTULO IV

Resultados e Discussão

IV.1. Diagnóstico de Convergências

Para a análise de convergência foram utilizadas duas cadeias com pontos iniciais diferentes, um gerado pelo *WinBUGS* e outro como sendo o valor esperado da distribuição *a priori* correspondente ao parâmetro. As análises foram realizadas num computador com um processador *Pentium 4* com velocidade de 1.4 GHz e memória de 512 MB. A partir da mesma semente, 314159.

A convergência das cadeias foi verificada utilizando o método Geweke (1992) implementado no CODA. No Quadro 4.1 averigua-se que há fortes indícios que a cadeia para o parâmetro nível de pesca em cada ano converge, F_y , pois os valores de *Z - score* são relativamente próximos de 0, isto é, há pouca diferença entre a média dos valores obtida com a distribuição *a posterior* do nível de pesca nos 10% das primeiras iterações com a média obtida com 50% das últimas iterações. No Capítulo II vimos que a convergência existe quando o *Z - score* é normalmente distribuído com média 0 e variância 1 (Best e Cowles 1996). A mesma tendência foi conferida para as outras cadeias respeitante aos outros parâmetros.

Para reforçar a ideia que de facto existe convergência, foram realizadas teste de convergência de Gelman e Rubin (1992), implementado no *WinBUGS*, para todas as cadeias. A Figura 4.1 mostra que existe convergência e que tal acontece ao fim de poucas iterações, o que indica que o período de *burn in* é relativamente curto. Adoptou-se um período de *burn in* de 1000 iterações.

Quadro 4.1: Análise de diagnóstico de Geweke (Z – score), realizadas no CODA, para o nível de pesca nas idades 0 a 6+. Foram consideradas amostras de tamanho 50000 para as cadeias, com espaçamento de tamanho 1. Fracção para a primeira janela= 0.1. Fracção para a última janela =0.5.

| VARIABLE | Fyear |
|-----------|-------|
| Fyear[1] | -1.52 |
| Fyear[2] | -0.74 |
| Fyear[3] | -0.84 |
| Fyear[4] | -2.63 |
| Fyear[5] | 0.57 |
| Fyear[6] | 0.66 |
| Fyear[7] | -0.38 |
| Fyear[8] | -0.38 |
| Fyear[9] | -1.05 |
| Fyear[10] | 0.07 |
| Fyear[11] | -0.35 |
| Fyear[12] | -1.22 |
| Fyear[13] | -0.41 |
| Fyear[14] | -0.92 |
| Fyear[15] | -0.39 |
| Fyear[16] | -0.78 |
| Fyear[17] | -0.68 |
| Fyear[18] | -1.7 |
| Fyear[19] | -0.58 |
| Fyear[20] | 0.22 |
| Fyear[21] | -0.48 |
| Fyear[22] | 0.09 |
| Fyear[23] | -1.5 |
| Fyear[24] | -0.15 |
| Fyear[25] | -1.09 |

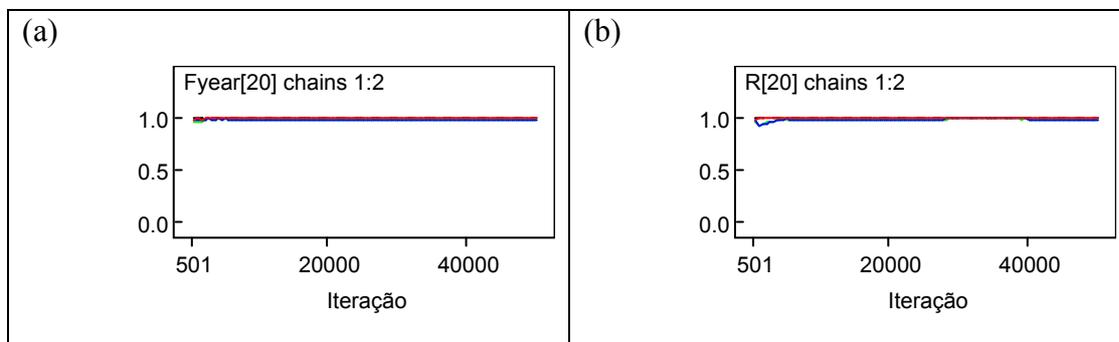


Figura 4.1: Gráfico da análise do diagnóstico de Gelman e Rubin (1992) para duas cadeias, implementado no *WinBUGS*, para o nível de pesca (a) e recrutamento (b) no ano de 1997, isto é $y = 20$. Foram realizadas 50000 iterações.

O diagnóstico de autocorrelação foi realizado para todas as cadeias, indicando que um espaçamento de tamanho 10 entre iterações diminuía a autocorrelação para valores próximos de zero, tal como ilustrado na Figura 4.2 para o nível de pesca e o recrutamento no ano de 1997 ($y = 20$).

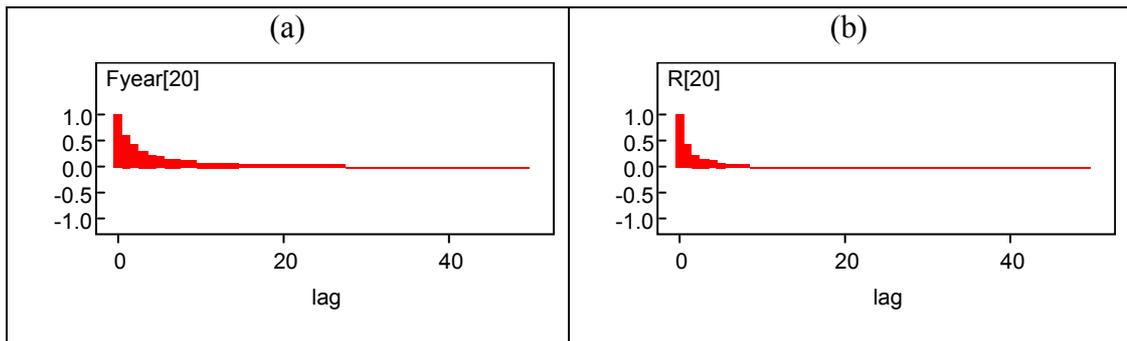


Figura 4.2: Autocorrelação obtida com as 50000 iterações para o nível de pesca (a) e o recrutamento (b) em função de diferentes espaçamentos (*lag*). Para um espaçamento de 10 iterações, o valor da autocorrelação para o nível de pesca e recrutamento no ano de 1997 diminui para respectivamente 0.106 e 0.016.

A Figura 4.3 mostra a cadeia resultante para o coeficiente de mortalidade por pesca médio no ano 1978 a título ilustrativo da convergência dos parâmetros.

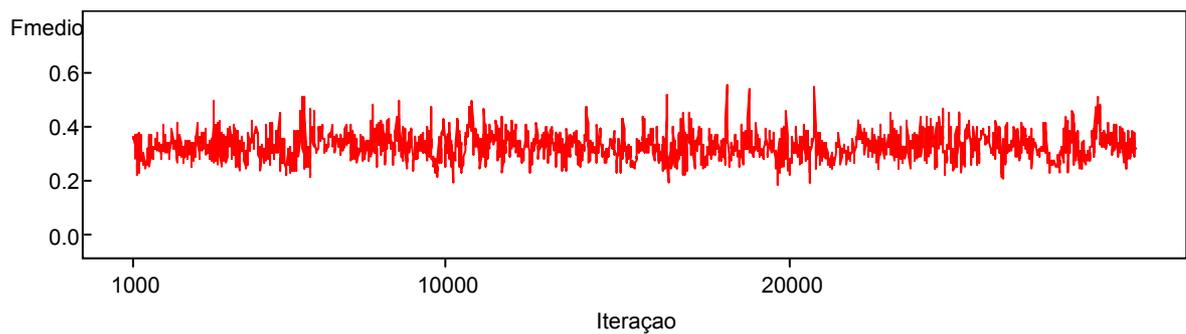


Figura 4.3: Evolução da cadeia da distribuição *a posteriori* do coeficiente de mortalidade por pesca médio no ano 1978, para 30000 iterações, *burn in* de 1000 iterações e espaçamento de 10.

IV.2. Resultados e Discussão

Neste capítulo apresentam-se e discutem-se os resultados obtidos com a *VPA* bayesiana, comparando-se ainda com os resultados obtidos com a *VPA* frequentista (método AMCI¹⁹, (Skagen 2003b)) usada na última avaliação do manancial Ibérico de sardinha, realizada em 2003, ver (ICES 2004).

Utilizando o método de *MCMC* implementado no *WinBUGS*, geraram-se amostras suficientemente grandes a partir da distribuição *a posteriori* e tomou-se como estimativa pontual a mediana e um intervalo de credibilidade de 95%.

Fase de Diagnóstico

Padrão relativo de exploração e capturabilidade

A Figura 4.4 apresenta a estimativa da distribuição *a posteriori* do padrão relativo de exploração em cada idade obtidas para o cenário em que se considerou s_{ay} constante com o tempo. Verifica-se que o padrão relativo de exploração apresenta uma tendência crescente com a idade, indicando que a partir da idade 3 os indivíduos se encontram completamente recrutados à pesca ($s_a > 50\%$). O padrão relativo de exploração

¹⁹ Assessment Model Combining Information from varies sources.

apresenta a forma de uma logística, tal como seria de esperar numa pescaria em que a arte de pesca utilizada é única.

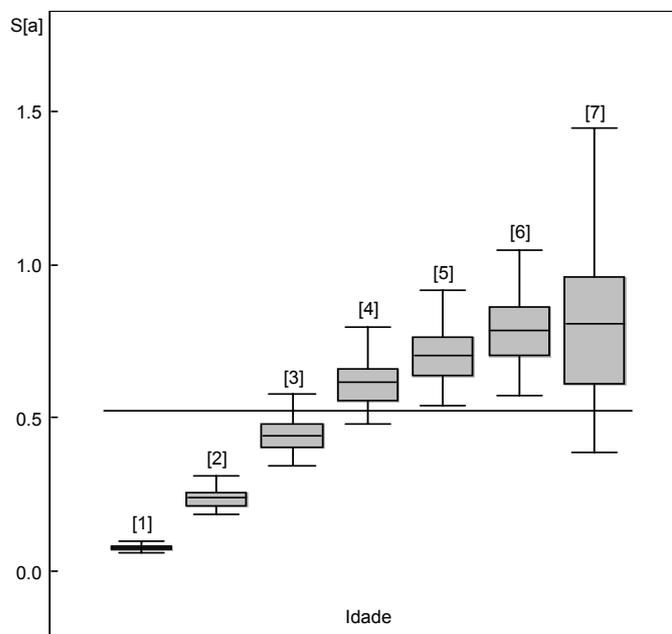


Figura 4.4: Box-plot da distribuição *a posteriori* do padrão relativo de exploração (ano^{-1}) em cada idade, no caso em que é considerada constante ao longo dos anos, com distribuição *a priori* $\text{dunif}(0,2)$, para amostra de tamanho 4900. O centro de cada caixa corresponde a mediana e os limites superiores e inferiores correspondem aos quantis 97.5 % e 2.5% respectivamente.

A Figura 4.5 apresenta a estimativa da distribuição *a posteriori* do padrão relativo de exploração obtidas para o cenário em que se considerou s_{ay} variável com o tempo para os diferentes grupos de idades.

O padrão relativo de exploração apresenta maior variabilidade nas idades mais jovens durante todo o período analisado, em particular nas idades 1 e 2. Nota-se que nas idades 1 e 2 a variabilidade de s_{ay} é superior no período que antecede o ano de 1994, com um cv na ordem dos 57% e 38% respectivamente contra um cv de 48% e 31% respectivamente.

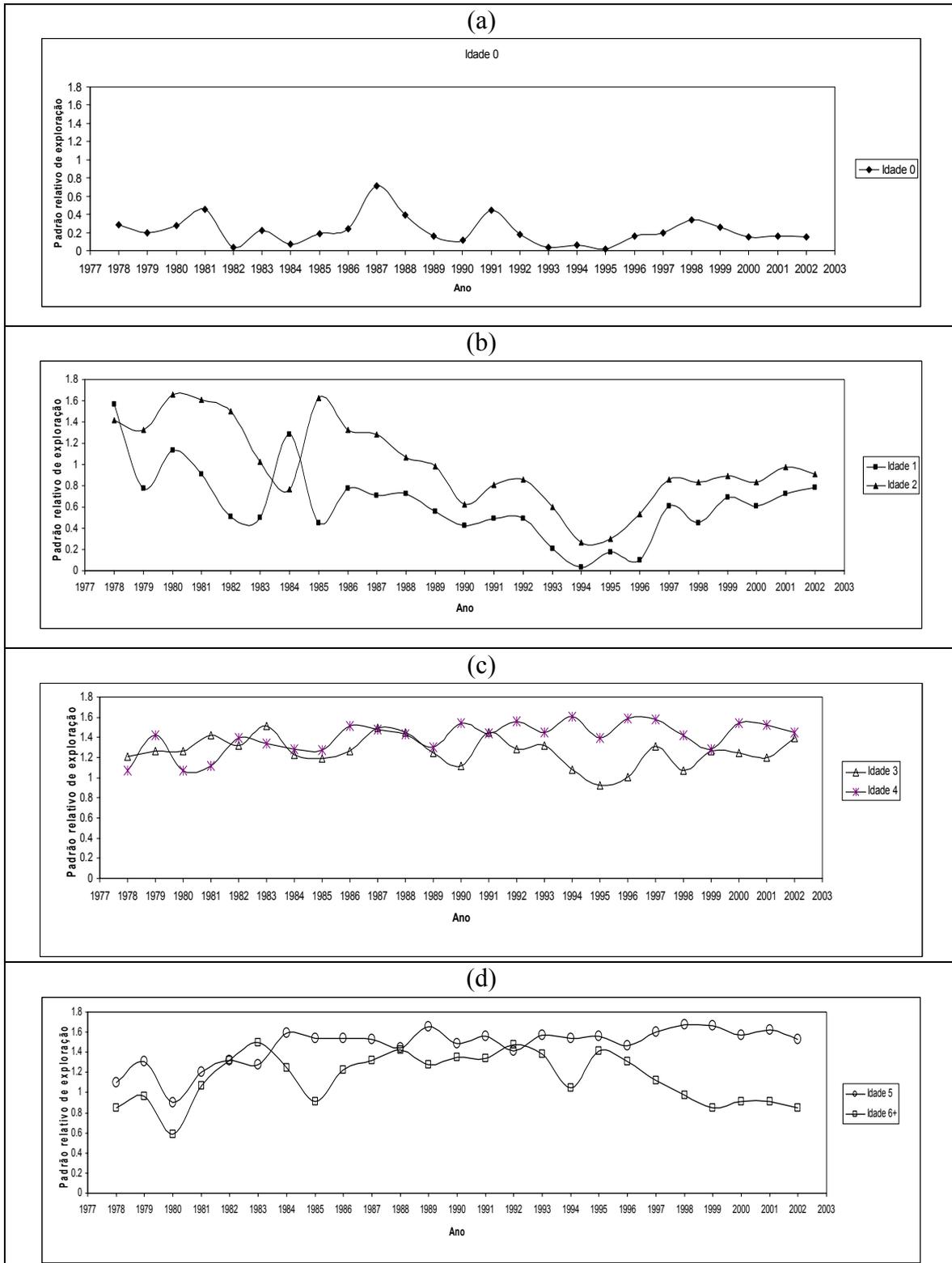


Figura 4.5: Estimativas do padrão relativo de exploração (ano⁻¹) em cada idade no período de 1978 a 2002, obtidas com o Modelo *VPA7802*, para uma distribuição *a priori* *dumif*(0,2), com uma amostra de tamanho 2900.

Entretanto, na idade 0 a variabilidade é mais acentuada entre 1978 e 1991. Porém, no grupo das idades mais velhas, verifica-se que a variabilidade do padrão relativo de exploração é superior na idade 6+, ver (Figura 4.5).

Para além deste aspecto, evidenciam-se também tendências diferentes no padrão relativo de exploração em algumas idades, destacando-se:

- Uma tendência semelhante nas idades 1 e 2, sendo decrescente de 1978 a 1994 e crescente até 2002, ver (Figura 4.5 (b));
- Uma tendência crescente entre 1980 e 1984 nas idades mais velhas, isto é, 5 e 6+, mantendo-se constante na idade 5 até ao final do período, enquanto que na idade 6+ existe uma maior variabilidade entre 1985 e 1995 e uma nítida tendência decrescente a partir desse ano, ver (Figura 4.5 (d)).

Existem indícios de que houve alteração no padrão relativo de exploração. Tendo em conta que não houve mudanças na arte e nos locais de pesca durante o período analisado, os resultados permitem colocar a hipótese de existência de alterações na mortalidade natural, M , e/ou migrações da sardinha, com um padrão distinto: aumento de M e/ou emigração entre 1978 e 1994 e diminuição de M e/ou imigração entre 1995 e 2002 nas idades 1 e 2; aumento de M e/ou emigração a partir do ano 1995 de indivíduos na idade 6+. No caso dos indivíduos mais velhos é mais plausível a hipótese de

aumento de mortalidade natural, uma vez que a partir de 1995 haverá poucos sobreviventes do forte recrutamento ocorrido em 1983. O recrutamento de 1983 foi estimado como sendo o mais forte da série temporal, o que se mostra mais adiante.

As estimativas obtidas com o Modelo VPA7802 para o padrão relativo de exploração são precisas nos dois cenários considerados, apresentando um *desvio padrão*²⁰ e o erro de Monte Carlo máximos de 0.6 e 0.04, respectivamente, consultar quadros C.4 e C.10 em anexo.

A distribuição *a posteriori* da capturabilidade para cada campanha de investigação e período correspondente ao cenário de s_{ay} variável com o tempo é resumida na Figura 4.6, dado que as estimativas e precisão da capturabilidade são idênticas para ambos os cenários como se mostra nas Figura 4.6 e C.1 e nos quadros C.5 e C.9 em anexo. A precisão das estimativas é menor para campanhas portuguesas e nas campanhas espanholas diminui com a idade.

A capturabilidade à idade das campanhas portuguesas é superior às das espanholas.

²⁰ A estimativa para o erro de Monte Carlo para a média é dado por $\frac{\sigma}{\sqrt{n}}$, onde σ representa uma estimativa do desvio padrão da média e n representa o total das iterações.

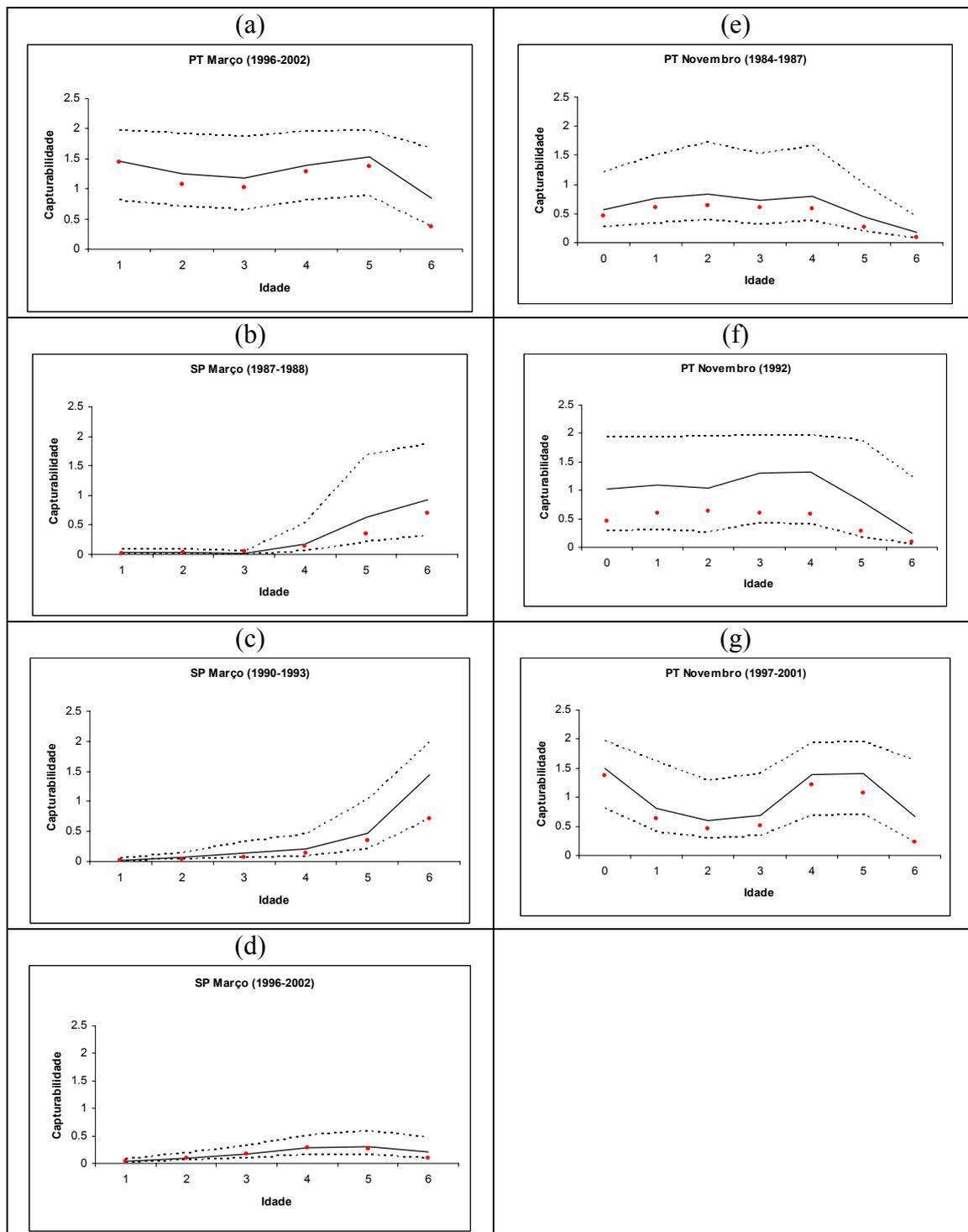


Figura 4.6: Estimativas da capturabilidade para cada campanha de investigação do mês Março [(a), (b), (c) e (d)] e de Novembro [(e), (f) e (g)] para o período de 1978 a 2003, obtidas com o Modelo *VPA7802*, quando o padrão relativo de exploração varia com o tempo. As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “•” representa as estimativas frequencistas (ICES 2004).

As estimativas para a capturabilidade à idade das campanhas portuguesas indicam uma tendência idêntica nas campanhas de Março e de Novembro (no período de 1984–1987 e em 1992). Nestes anos a capturabilidade é praticamente constante até a idade 5 na campanha de Março e até a idade 4 na campanha de Novembro, diminuindo em ambas para as idades seguintes, indicando uma menor disponibilidade/acessibilidade das sardinhas mais velhas na costa portuguesa.

Estima-se uma mudança substancial no padrão da capturabilidade à idade para a campanha de Novembro nos anos mais recentes, indicando uma menor acessibilidade das sardinhas nas idades mais jovens (1, 2 e 3) a partir de meados da década de 90. Nota-se que o facto desta mudança de padrão não se observar na campanha de Março, realizada no mesmo período, indica que a mudança não é temporal (não muda com o ano) mas sim estacional (muda com a estação da campanha de investigação).

No caso das campanhas espanholas a capturabilidade é reduzida nas idades mais jovens e nos períodos de 1987–1988 e 1990–1993 aumenta da idade 5 para a idade 6+, enquanto que a partir de meados da década de 90 (1996 a 2002) esta diminui.

Recrutamento, Biomassa desovante e Mortalidade por pesca

As estimativas obtidas com o Modelo *VPA7802*, para o recrutamento (Figura 4.7), biomassa desovante (Figura 4.8) e o coeficiente de mortalidade por pesca médio das idades 2 a 5 anos, \bar{F}_y , (Figura 4.9) são mais precisas no cenário em que se deixa variar temporalmente o padrão relativo de exploração (consultar também Quadros C.1–C.10 em anexo) com a excepção do recrutamento em 2002. Embora não haja grandes discrepâncias entre os dois cenários em termos de tendência temporal (há apenas algumas discrepâncias para \bar{F}_y , que se referirão mais adiante) existem diferenças na escala dos valores.

Dada a maior precisão e o facto de se terem detectado alterações temporais no padrão relativo de exploração, adoptam-se as estimativas obtidas com o cenário em que s_{ay} é variável com o tempo como sendo mais realistas para analisar a evolução histórica do manancial.

As estimativas da distribuição *a posteriori* do recrutamento indicam que os recrutamentos máximos ocorreram nos anos de 1983 e 2000. Houve no entanto recrutamentos fortes entre 1978 e 1980, no ano de 1987 e entre 1991 e 1992. Destaca-se um período alargado de 7 anos, entre 1993 e 1999, em que os recrutamentos foram sempre baixos. Existe evidência de uma mudança no padrão de recrutamento ao longo da série de anos: até 1993 os recrutamentos baixos ocorriam em períodos curtos de cerca de 3 anos e alternavam com picos de recrutamento.

Quando se comparam as estimativas bayesianas com as frequentistas verifica-se que as estimativas frequentistas do recrutamento coincidem com o limite inferior do intervalo de credibilidade a 95% praticamente ao longo da série temporal analisada, sendo apenas excepção os anos de 2000 e 2001 em que coincidem com as estimativas pontuais, ver (Figura 4.7 (b)).

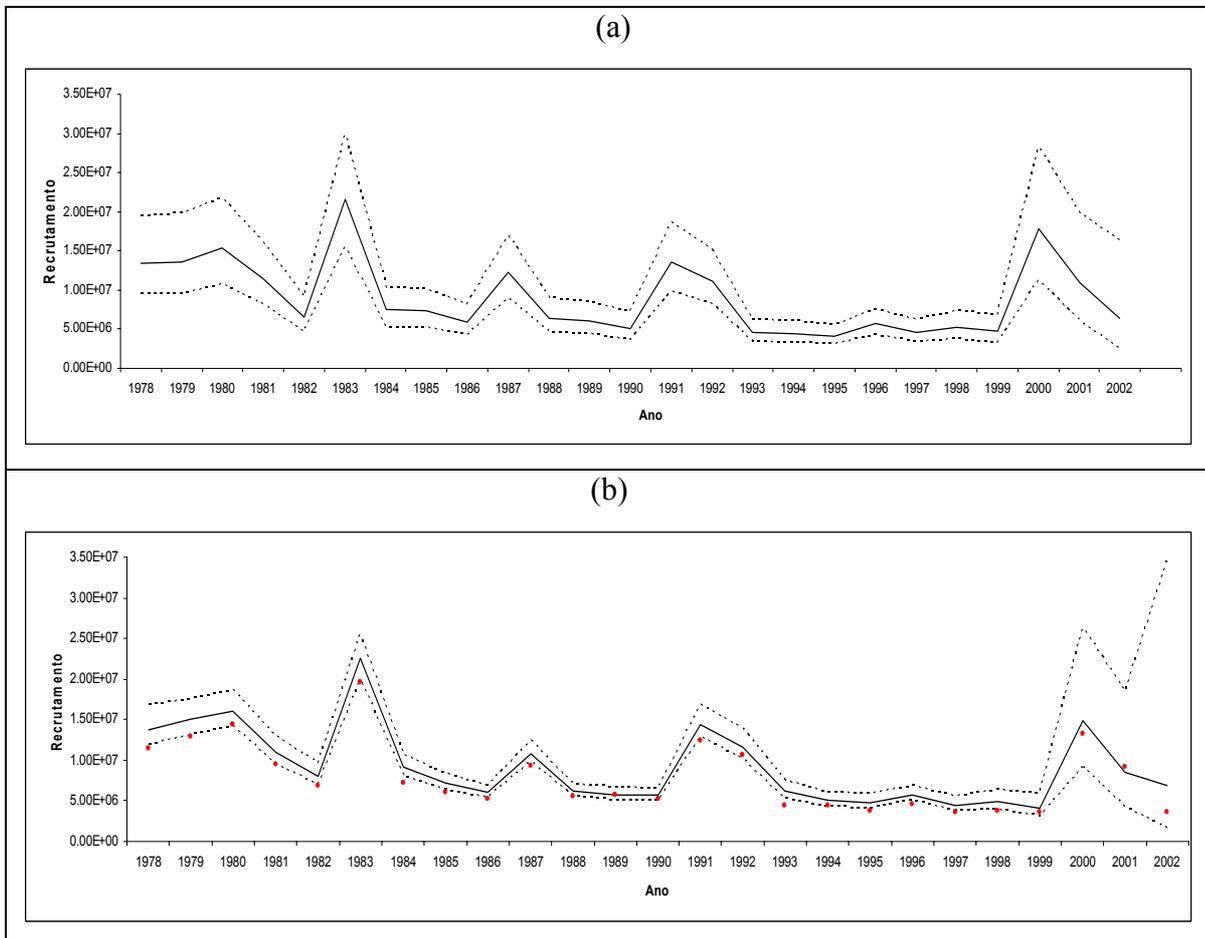


Figura 4.7: Estimativas do recrutamento (milhares de milhões) para o período de 1978 a 2003, obtidas com o Modelo *VPA7802*, quando o padrão relativo de exploração é constante ao longo dos anos (a) e quando varia com o tempo (b). As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “•” representa as estimativas frequentistas (ICES 2004).

As estimativas da distribuição *a posteriori* da biomassa desovante evidenciam uma certa regularidade na sua evolução. Existem dois períodos, uma entre 1978 a 1990 e outra entre 1991 a 2000, em que a biomassa desovante aumenta até a metade desse período e depois diminui. A biomassa desovante atingiu níveis superiores no primeiro período em virtude de terem ocorrido quatro fortes recrutamentos entre 1978 e 1983.

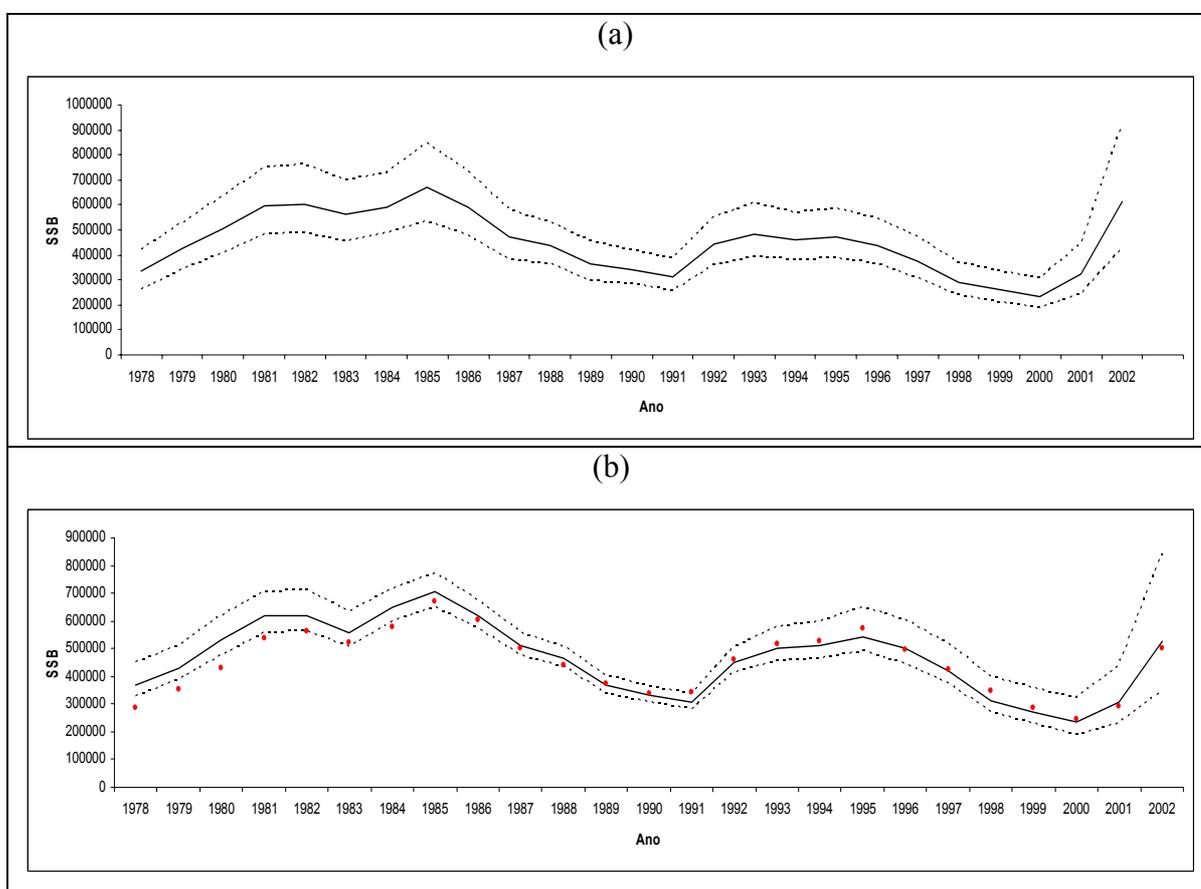


Figura 4.8: Estimativas da biomassa desovante (t) para o período de 1978 a 2003, obtidas com o Modelo *VPA7802*, quando o padrão relativo de exploração é constante ao longo dos anos (a) e quando varia com o tempo (b). As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “.” representa as estimativas frequencistas (ICES 2004).

As estimativas frequencistas para a biomassa desovante estão fora (inferiores) do intervalo de credibilidade entre 1978 e 1981 e coincidem com as estimativas bayesianas nos anos 1986–1990 e 1996–2003, ver (Figura 4.8 (b)). Nos restantes anos anteriores a 1990, isto é, entre 1982 e 1985 as estimativas frequencistas são próximas dos valores do limite inferior do intervalo de credibilidade, enquanto que entre 1991 e 1995 estas são superiores as estimativas bayesianas.

Quando se comparam as estimativas para o \bar{F}_y nos dois cenários considerados para o padrão relativo de exploração, verifica-se que para além do valor das estimativas ser menor no caso do cenário em que s_{ay} é variável com o tempo, observa-se que entre 1984 e 1990 a tendência crescente do \bar{F}_y é mais pronunciada no caso do cenário em que s_{ay} é variável com o tempo e que no período entre 1992 e 1994 a evolução do \bar{F}_y é diferente, aumentando neste cenário, enquanto no outro cenário para o s_{ay} há uma diminuição pronunciada ao longo desses anos, ver (Figura 4.9).

As estimativas para a mortalidade por pesca (Figura 4.9 (b)) indicam um aumento entre 1978 e 1990 (0.33 e 0.46 respectivamente) e após um período de flutuações até 1995 o \bar{F}_y mantém-se estável em redor de 0.40 até 1998, decrescendo nos anos seguintes para o valor de 0.23 em 2002, isto é, para níveis inferiores ao de mortalidade natural.

Comparando com as estimativas frequentistas observa-se que estas só coincidem com a estimativa pontual bayesiana nos anos mais recentes (1998–2002). De facto, entre 1982 e 1990 as estimativas frequentistas estão entre a estimativa pontual bayesiana e o limite inferior do intervalo de credibilidade e entre 1990 e 1996 estão fora (inferiores) do intervalo de credibilidade.

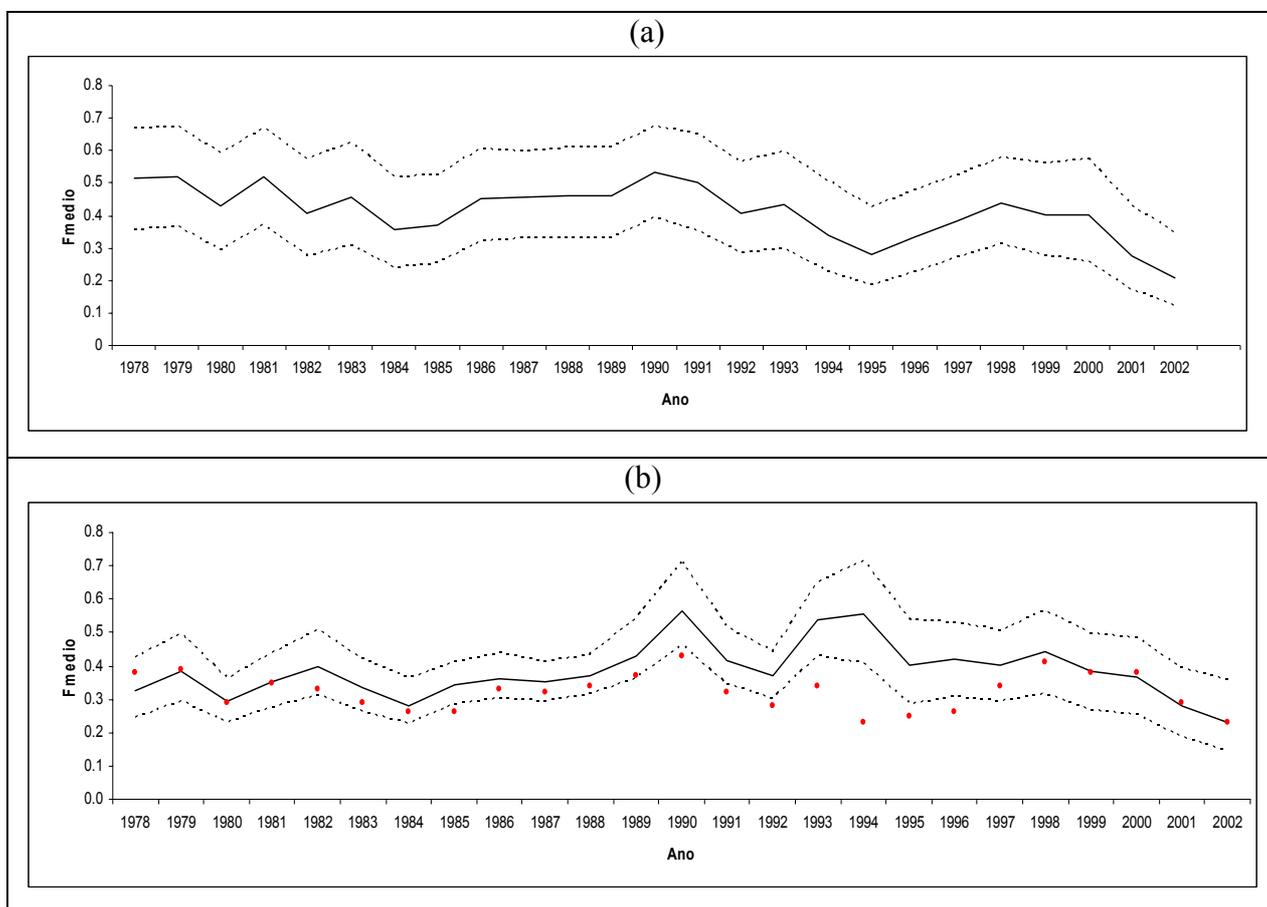


Figura 4.9: Estimativas do coeficiente de mortalidade por pesca médio (ano^{-1}) para o período de 1978 a 2003, obtidas com o Modelo *VPA7802*, quando o padrão relativo de exploração é constante ao longo dos anos (a) e quando varia com o tempo (b). As linhas contínuas e a tracejado representam respectivamente a mediana *a posteriori* e os quantis 2.5% e 97.5%. O símbolo “•” representa as estimativas frequentistas (ICES 2004).

Fase de Prognóstico

A projecção a curto prazo é feita, tendo em conta os resultados obtidos no Modelo *VPA7802* para o cenário em que se considerou s_{ay} variável com o tempo para os diferentes grupos de idades, por razões já apontadas.

A Figura 4.10 apresenta a relação stock–recrutamento entre 1978 e 2002 para as estimativas *a posteriori* da biomassa desovante e do recrutamento obtidas na fase de diagnóstico. Nota-se que um ajuste do modelo de Beverton–Holt ou de Ricker aos dados, não seria muito adequado.

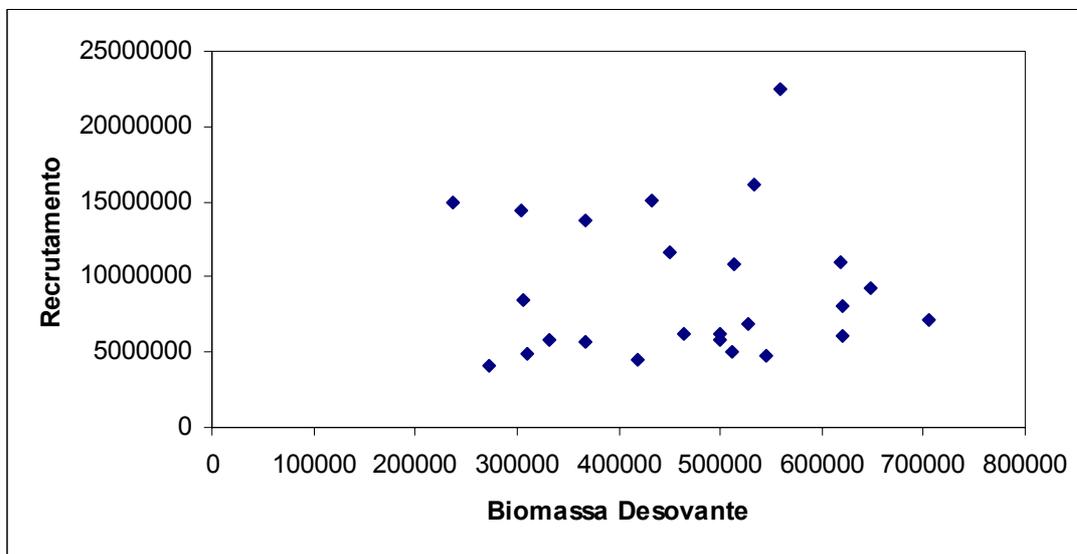


Figura 4.10: Relação $S - R$ entre 1978 e 2002 para as estimativas obtidas no Modelo *VPA7802*.

Na Figura 4.11 apresenta-se, a título de exemplo, uma possível escolha das categorias para a biomassa desovante em função da fertilidade. Posto isto, procede-se à projecção do recrutamento.

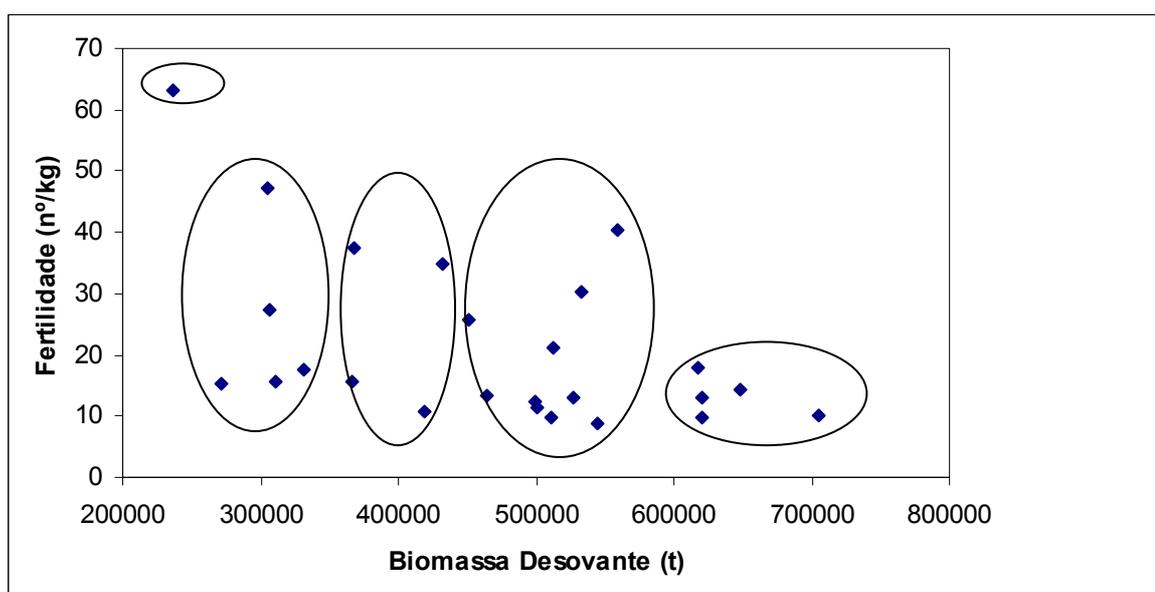


Figura 4.11: Categorias para a biomassa desovante em função da fertilidade λ_y .

A partir das estimativas do número de sardinhas em cada idade ao início de 2002 e do coeficiente de mortalidade por pesca em cada idade do ano 2002, a população é projectada deterministicamente para o final de 2002, o que coincide com o início de 2003, pois o objectivo é estimar a captura e a biomassa desovante do ano 2003, uma vez que a estrutura da população para o ano 2002 já é conhecida. O recrutamento é projectado utilizando o Modelo *PROJ*, tendo em conta que a ogiva de maturação e peso médio são dados pelas estimativas de 2002.

Assumiu-se que o coeficiente de mortalidade por pesca segue uma Log Normal. Assim, temos que o logaritmo do coeficiente de mortalidade por pesca em cada idade segue a distribuição Normal com média dada pelo logaritmo da estimativa obtida de F em cada idade, $\mu_a = \ln(F_{a,2002})$, para $a = 1, \dots, 7$, e desvio padrão calculado de modo que a obter o mesmo coeficiente de variação da variável original.

Admitiu-se que o número de indivíduos em cada idade é normalmente distribuído com média dada pelo número de indivíduos ao início de 2003 e um coeficiente de variação igual ao estimado no Modelo *VPA7802*.

Conhecido o número de indivíduos em cada idade ao início de 2003 e definidas as distribuições de probabilidade para N e F , a população é projectada ao final de 2003 utilizando o método de Monte Carlo. A biomassa desovante e a captura em peso são calculados considerando a ogiva de maturação, o peso médio dos indivíduos e o peso médio para a captura como sendo a média aritmética dos respectivos valores dos últimos três anos.

O Modelo *PROJ* foi implementado no S-PLUS, ver código em (Anexo D).

CAPÍTULO V

Conclusões e Perspectivas Futuras

Neste trabalho propôs-se um modelo estocástico estruturado por idades para a análise histórica do manancial utilizando a inferência bayesiana para estimar os parâmetros mais importantes que caracterizam o estado do recurso pesqueiro e posteriormente projectar o recrutamento com base na fertilidade da sardinha.

Com a análise bayesiana é possível incorporar o conhecimento *a priori* que se tem sobre as variáveis não observáveis e com a distribuição *a posteriori* tem-se uma interpretação probabilística da estimativa obtida para o nível de pesca, o padrão relativo de exploração e dos coeficientes de capturabilidade, o que faz com que esta abordagem seja eficaz na resolução de problemas complexos que envolvem modelos matemáticos com vários parâmetros.

À eficácia da inferência bayesiana, associa-se os métodos de Monte Carlo via Cadeias de Markov, pois, são muito simples e úteis na resolução de problemas relacionados com a monitorização de recursos pesqueiros. Contudo, o estudo da evolução da abundância dos recursos pesqueiros, baseada em modelos estruturados por idade, utilizando o teorema de Bayes, é muito custoso sob o ponto de vista computacional. Mesmo em computadores pessoais, razoavelmente potentes, a análise do Modelo *VPA7802* foi feita em, pelo menos, dois dias.

A evolução dos recursos pesqueiros é um fenómeno natural muito complicado que depende de vários factores, entre os quais os naturais e a acção do homem. Para formulação de um modelo matemático e descrição deste processo são necessários vários

parâmetros. A complexidade deste fenómeno faz com que seja difícil extrair muita informação sobre um determinado parâmetro, de forma isolada.

Assim, dado a natureza da informação que se têm dos parâmetros, as distribuições *a priori* não informativas desempenham um papel muito importante na avaliação dos recursos pesqueiros com a abordagem bayesiana. Atribuiu-se a mesma probabilidade de acontecimento a todos os valores possíveis de um determinado parâmetro, assumindo uma distribuição uniforme, que depois de actualizada com a informação por amostragem obtém-se uma estimativa do mesmo parâmetro.

Feita a análise histórica do stock Ibérico da sardinha, conclui-se que há fortes indícios de que houve alteração no padrão relativo de exploração, apresentando, neste caso, tendências diferentes nalgumas idades.

O recrutamento apresenta uma mudança no padrão, ao longo da série de anos. Existem dois períodos, um entre 1978 a 1990 e outro entre 1991 a 2000, em que a biomassa desovante aumenta até à metade desses períodos e depois diminui.

Contudo, não se pode deixar de salientar que existem ainda alguns aspectos por explorar. Por exemplo, a análise de sensibilidade das distribuições *a priori*, que se traduz na adopção de diferentes distribuições *a priori* para os parâmetros, entre os quais o coeficiente de mortalidade por pesca.

Outro ponto de interesse seria um estudo sobre a correlação, por exemplo, entre o recrutamento e a captura ou a biomassa.

Pretende-se posteriormente desenvolver um método que permita avaliar o mínimo de conhecimento que se tem sobre os parâmetros com o intuito de obter distribuições *a priori* não informativa, pois a distribuição uniforme apesar de útil, é uma maneira muito simplista de traduzir a informação.

Outro facto a ter em conta, é tentar explorar novos modelos, menos simplistas para o coeficiente de mortalidade natural, considerando a hipótese de que população não é fechada.

O trabalho desenvolvido na fase de prognóstico da avaliação de recurso pesqueiro ainda está numa fase embrionária, pois a atribuição das categorias dos níveis da biomassa desovante foi *ad hoc*. Futuramente, poder-se-á desenvolver um método de optimização, atribuindo, por exemplo, o número das categorias de modo que o coeficiente de variação de cada categoria seja mínima.

Por último, é imperioso que se realce que o trabalho realizado no âmbito deste projecto constituiu uma enorme satisfação, tendo permitido o aprofundamento e a aquisição de novos conhecimentos, para além de ter transmitido o incentivo e o estímulo, necessários e essenciais para que se continue, com dedicação e empenho, a árdua tarefa que é a investigação científica.

BIBLIOGRAFIA

- [1]. **Azevedo, M.** (2002). Enfoque de precaución en pesca: concepto y aplicación em El criterio de precaución en la gestión de los recursos pesqueros (Maguire, J-J. e Azevedo, M. eds.). *Centro de Investigación Económica y Financiera-Fundación Caixa Galicia*. Editorial Galaxia. Documentos de Economía 17.
- [2]. **Azevedo, M.** (1999). Exploratory data analysis for Iberian Sardine (*Sardina pilchardus*) Working document for the ICES working Group on the Assessment of Mackerel, Horse Mackerel, Sardine, and Anchovy. ICES, Copenhagen, 14-23 September 1999. ICES CM 2000/ACFM: 5.
- [3]. **Azevedo, M., Morgado, C. e Cardador, F.** (2002). Are there general patterns in SSB-R relations and F-SSB trajectories that can be used as guides for establishing PA referente points? Report of the Study Group on the Further development of the Precautionary Approach to Fishery Management. Lisbon, 4-8 March 2002. ICES CM 2002/ACFM: 10:72-86.
- [4]. **Baranov, F. I.** (1918). On the question of de biological basis of fisheries. *Nauchn. Issled. Ikhtiologicheskii Inst. Izv.* 1, 81-128. Traduzido em espanhol por Granados, R. R., 1965. Inst. Nac. Inv. Biologico Pesqueras, 2, 70p.
- [5]. **Bergh, M. O. e Butterworth, D. S.** (1987). Towards rational harvesting of the South African anchovy considering survey imprecision and recruitment variability. *S. Afr. J. mar. Sci.* 5:937-951.

-
- [6]. **Best, N. e Cowles, M. K.** (1996). CODA[®] – *Convergence Diagnosis and Output Analysis Software for Gibbs sampling output*, version 0.3. MRC Biostatistics Unit, Cambridge, UK. www.mrc-bsu.cam.ac.uk/bugs.
- [7]. **Beverton, R. J. H. e Holt, S. J.** (1957). *On the dynamics of exploited fish population*. Fish. Invest. Ser. II, Mar. Fish. G.B. Minist Agric. Fish. Food, **19**.
- [8]. **Box, G. E. P. e Tiao, G. C.** (1992). *Bayesian inference in statistical analysis*. Wiley – Interscience.
- [9]. **Cadima, E. L.** (2000). *Manual de avaliação de recursos pesqueiros*. DANIDA.
- [10]. **Casella, G. e R. L. Berger** (1990). *Statistical inference*. Duxbury, California.
- [11]. **Cushing, D.H.** (1996). *Towards a science of recruitment in fish populations*. In: Excellence in Ecology, Book 7, Ecology Institut, Oldendorf/Scuhe, Germany.
- [12]. **Fournier, D. e Archibald, C. P.** (1982). A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* **39**:1195–1207.
- [13]. **Fournier, A. D., Hampton, J. e Sibert, J. R.** (1998). MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Can. J. Fish. Aquat. Sci.* **55**:2105–2116.
- [14]. **Gamerman, D.** (1999). Notas do mini curso: Métodos computacionais em inferência bayesiana. Instituto de matemática – UFRJ. <http://acd.ufrj.br/~dani/teaching/5ebeb.ps>.

-
- [15]. **Gamerman, D.** (1996). *Simulação estocástica via cadeias de markov*. Associação Brasileira de Estatística.
- [16]. **Gelfand, E. e Simth, A. F. M.** (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**:398–409.
- [17]. **Gelfand, A. E., Hills, S. E., Racine-Poon, A. e Smith, A. F. M.** (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**:972–985.
- [18]. **Gelman, A. e Rubin, D. B.** (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**:457–511.
- [19]. **Geman, S. e Geman, D.** (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**:721–741.
- [20]. **Geweke, J.** (1992). *Evaluating the accuracy of the sampling-based approaches to calculating posterior moments*. Bayesian Statistics 4, (ed. Berger, J. M., Dawid, A. P. e Smith, A. F. M.). Clarendon Press, Oxford, UK.
- [21]. **Gilks, W. R. e Wild, P.** (1992). Adaptive Rejection Sampling from Log-concave Density Functions. *Aplied Statistics*. **41**:337–348.
- [22]. **Hastings, W. K.** (1970), Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**:97–109.

-
- [23]. **ICES** (2004). Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine, and Anchovy. ICES, Copenhagen, 9-18 September 2003. ICES CM 2004/ACFM: 08.
- [24]. **Karlin, S. e Taylor, H. M.** (1975). *A First Course in Stochastic Processes*. 2nd ed., Academic Press. New York.
- [25]. **Kuikka, S.** (1999). Parameter and structural uncertainties in Bayesian decision analysis. ICES CM 1999/T: 04.
- [26]. **Kuikka, S., Hildén, M., Gislason, H., Hansson, S., Sparholt H. e Varis, O.** (1999). Modeling environmentally driven uncertainties in Baltic cod (*Gadus morhua*) Management by Bayesian influence diagrams. *Can. J. Fish. Aquat. Sci.* **56**: 629-641.
- [27]. **McAllister, M.K. e Ianelli, J. N.** (1997). Bayesian stock assessment using catch-age data and the sampling - importance resampling algorithm. *Can. J. Fish. Aquat. Sci.* **54**: 284-300.
- [28]. **Megrey, B. A.** (1989). Review and comparison of age-structured stock assessment models from theoretical and applied points of view. In Mathematical analysis of fish stock dynamics. Edited by E.F. Edwards and B.A. Megrey. *American Fisheries Society Symposium* 6. pp. 8-48.
- [29]. **Metropolis, Rosenbluth, N. A, Rosenbluth, M., Teller, A. e Teller, E.** (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**:1087-1092.

-
- [30]. **Meyer, R. e Millar, R. B.** (1999). BUGS in Bayesian stock assessment. *Can. J. Fish. Aquat. Sci.* **56**: 1078–1087.
- [31]. **Millar, R. B, e Meyer, R.** (1999a). Bayesian stock assessment using a state–space implementation of the delay difference model. *Can. J. Fish. Aquat. Sci.* **56**: 37–52.
- [32]. **Millar, R. B, e Meyer, R.** (1999b). Bayesian state–space modelling of age–structured data: fitting a model is just the beginning. *Can. J. Fish. Aquat. Sci.* **57**: 43–50.
- [33]. **Murteira, B. J. F.** (1996). *Probabilidades e estatísticas. Vol. I e II*, 2ª Ed. Revista McGraw – Hill, Portugal.
- [34]. **Neal, R. M.** (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods*, Technical Report CRG–TR–93–1, Dept. of Computer Science, University of Toronto, 144 pages.
- [35]. **Neal, R. M.** (1997). *Markov chain Monte Carlo methods based on 'slicing' the density function*, Technical Report No. 9722, Dept. of Statistics, University of Toronto, 27 pages.
- [36]. **Nielsen, A.** (2000). *Fish Stock Assessment using Markov Chain Monte Carlo*. University of Copenhagen. Tese de Mestrado.
- [37]. **Nielsen, A. e Lewy, P.** (2002). Comparison of the frequentist properties of Bayes and maximum likelihood estimators in an age–structured fish stock assessment model. *Can. J. Fish. Aquat. Sci.* **59**:136–143.

-
- [38]. **Norris, J.** (1999). *Markov Chains*. Cambridge University. USA. p xiv..
- [39]. **Patterson, K. R.** (1999). Evaluating uncertainty in harvest control law catches using Bayesian Markov Chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. *Can. J. Fish. Aquat. Sci.* **56**:208–221.
- [40]. **Paulino, C. D., Turkman, M. A. A. e Murteira, B.** (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian.
- [41]. **Pestana, G.** (1989). *Manancial Ibérico–Atlântico de sardinha (*Sardina pilchardus*, Walb.). Sua avaliação e medidas de gestão*. Tese para Investigador auxiliar. IPIMAR (mimeo).
- [42]. **Pope, J. G. e Shepherd, J. G.** (1982). A simple method for the consistent interpretation of catch-at-age data. *J. Cons. Int. Exp. Mer.* **40**: 176–184.
- [43]. **Raftery, A. E. e Lewis, S. M.** (1995). *Implementing MCMC. In Markov Chain Monte Carlo in Practice*. 115–130. London. Chapman and Hall.
- [44]. **Ricker, W. E.** (1954). Stock and recruitment. *J. Fish. Res. Bd. Can.* **11**:559–623.
- [45]. **Skagen, D. W.** (2003a). Some analyses of the sardine assessment data. Working document for the ICES working Group on the Assessment of Mackerel, Horse Mackerel, Sardine, and Anchovy. ICES, Copenhagen, 9–18 September 2003. ICES CM 2004/ACFM:08.

- [46]. **Skagen, D. W.** (2003b). AMCI (version 2.3) – Assessment model combining information from various sources: Model description, instructions for installation and running file formats (disponível através do autor: dankert@imr.no).
- [47]. **Spiegelhalter, D. J., Thomas, A., Best, N., e Gilks, W. R.** (1996). *BUGS®– Bayesian inference Using Gibbs Sampling – Manual*, Version 0.5, (version ii). www.mrc-bsu.cam.ac.uk/bugs/.
- [48]. **Spiegelhalter, D. J., Thomas, A. e Best, N.** (1999). *WinBUGS® Version 1.2 – User Manual*. MRC Biostatistics Unit. www.mrc-bsu.cam.ac.uk/bugs/.

Moradas da Internet consultadas:

<http://acd.ufrj.br/~dani>

<http://www.citeseer.nj.nec.com>

<http://www.cs.utoronto.ca/~radford>

<http://www.ices.com>

<http://www.mrc-bsu.cam.ac.uk/bugs/>

Software utilizado:

CODA® – Convergence Diagnosis and Output Analysis. Copyright© 1995 MRC Biostatistics Unit. <http://www.mrc-bsu.cam.ac.uk/bugs/>.

WinBUGS® – Bugs Bayesian inference Using Gibbs Sampling (BUGS). Copyright 1996–2003: Imperial College&MRC, UK, <http://www.mrc-bsu.cam.ac.uk/bugs/>.

S-PLUS® – 2000 professional Release 3, MathSoft, Inc. Copyright 1988–2000 All Right reserved.

GLOSSÁRIO

Biomassa – Peso de um indivíduo ou de um grupo de indivíduos contemporâneos de um manancial;

Biomassa Desovante – Parte do stock que já desovou pelo menos uma única vez;

Captura em número (C) – Número de indivíduos capturados.

Coefficiente de capturabilidade (q) – Fração da biomassa que é capturada por unidade de esforço de pesca;

Coefficiente de mortalidade natural (M) – Coeficiente instantâneo relativo da variação do número de sobreviventes dos que morrem devido a todas as causas excepto as devidas à pesca;

Coefficiente de mortalidade por pesca (F) – Coeficiente instantâneo relativo da variação do número de sobreviventes dos que morrem devido à pesca;

Coorte – Conjunto dos indivíduos de um recurso pesqueiro nascidos de uma mesma época de desova;

Stock/Manancial (S) – Conjunto dos sobreviventes das coortes de um recurso pesqueiro, num certo instante ou período de tempo. Pode referir-se à biomassa ou ao número de indivíduos;

Modelos de produção – Modelos que consideram a biomassa do manancial na sua globalidade, i.e., não tomam em consideração a estrutura do manancial por idades ou tamanhos. Estes modelos só se aplicam a análises que consideram mudanças do nível de pescas e, portanto, não permitem analisar os efeitos nas capturas e biomassas de mudanças do padrão relativo de exploração.

Modelos estruturais – Modelos que consideram a estrutura do manancial por idades ou tamanhos. Estes modelos permitem analisar os efeitos nas capturas e biomassas, de mudanças do nível de pesca e também do padrão relativo de exploração.

Recrutamento (R) – Total dos indivíduos na idade zero.

ANEXOS

Anexo A – Alguns Definições e Teoremas

Definição (Convergência em Distribuição) A.1: Uma sequência de variáveis aleatórias, X_1, X_2, \dots , converge em distribuição para uma variável X , se:

$$\lim_{n \rightarrow \infty} F_{X_n}(X) = F_X(X), \forall X, \text{ onde } F_X(X) \text{ é contínua.}$$

(Casella e Berger 1990)

Definição (Convergência quase certa) A.2: Uma sequência de variáveis aleatórias, X_1, X_2, \dots , converge quase certamente para uma variável X , se:

$$\forall \varepsilon > 0 : P\left(\lim_{n \rightarrow \infty} |X_n - X| < \varepsilon\right) = 1$$

(Casella e Berger 1990)

Teorema A.1 (Lei forte dos grandes números): Sejam $\theta_1, \theta_2, \dots$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, *i.i.d.*, com esperança matemática $E[\theta_i]$ e variância $Var[\theta_i] < \infty$, e seja $\bar{\theta}_n = \frac{1}{n} \sum_{i=1}^n \theta_i$, então:

$$\forall \varepsilon > 0 : P\left(\lim_{n \rightarrow \infty} |\bar{\theta}_n - E[\theta_i]| < \varepsilon\right) = 1$$

ou seja $\bar{\theta}_n$ converge quase certamente, *q.c.*, para $E[\theta_i]$, $\frac{1}{n} \sum_{i=1}^n \theta_i \xrightarrow{q.c.} E[\theta_i]$.

(Casella e Berger 1990)

Teorema A.2 (Limite central): Sejam $\theta_1, \theta_2, \dots$ uma sequência de variáveis aleatórias *i.i.d.*, tal que $E[\theta_i^2] < \infty$, então:

$$\sqrt{n} \frac{\frac{1}{n} \sum_{i=1}^n g(\theta_i) - E[g(\theta_1)]}{\sqrt{Var[\theta_i]}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1).$$

(Casella e Berger 1990)

Anexo B – Dados sobre a sardinha obtidos no WG2003

Quadro B.1: Dados da captura (em milhões) por ano e classe de idades, C_{ay} .

| Idade\Ano | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 |
|-----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--------|--------|
| 0 | 869437 | 674489 | 856671 | 1025961 | 62000 | 1070000 | 118000 | 268000 | 304000 | 1437000 | 521000 | 248000 | 258000 |
| 1 | 2296646 | 1535557 | 2037400 | 1934838 | 795000 | 577000 | 3312000 | 564000 | 755000 | 543000 | 990000 | 566000 | 602000 |
| 2 | 946698 | 956132 | 1561971 | 1733725 | 1869000 | 857000 | 487000 | 2371000 | 1027000 | 667000 | 535000 | 909000 | 517000 |
| 3 | 295360 | 431466 | 378785 | 679001 | 709000 | 803000 | 502000 | 469000 | 919000 | 569000 | 439000 | 389000 | 707000 |
| 4 | 136661 | 189107 | 156922 | 195304 | 353000 | 324000 | 301000 | 294000 | 333000 | 535000 | 304000 | 221000 | 295000 |
| 5 | 41744 | 93185 | 47302 | 104545 | 131000 | 141000 | 179000 | 201000 | 196000 | 154000 | 292000 | 200000 | 151000 |
| 6+ | 16468 | 36038 | 30006 | 76466 | 129000 | 139000 | 117000 | 103000 | 167000 | 171000 | 189000 | 245000 | 248000 |
| Idade\Ano | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | |
| 0 | 1580579 | 498265 | 87808 | 120797 | 30512 | 277053 | 208570 | 449115 | 246016 | 489836 | 219973 | 106882 | |
| 1 | 477368 | 1001856 | 566221 | 60194 | 189147 | 101267 | 548594 | 366176 | 475225 | 354822 | 1172301 | 587354 | |
| 2 | 436081 | 451367 | 1081818 | 542163 | 280715 | 347690 | 453324 | 501585 | 361509 | 313972 | 256133 | 753897 | |
| 3 | 406886 | 340313 | 521458 | 1094442 | 829707 | 514741 | 391118 | 352485 | 339691 | 255523 | 195897 | 181381 | |
| 4 | 265762 | 186234 | 257209 | 272466 | 472880 | 652711 | 337282 | 233672 | 177170 | 194156 | 126389 | 112166 | |
| 5 | 74726 | 110932 | 113871 | 112635 | 70208 | 197235 | 225170 | 178735 | 105518 | 97693 | 75145 | 55650 | |
| 6+ | 105186 | 80579 | 120282 | 72091 | 64485 | 46607 | 70268 | 105884 | 72541 | 64373 | 49547 | 40219 | |

Quadro B.2: Dados da série da campanha portuguesa de Março (em milhões) por ano e idades, $I1_{ay}$.

| Idade\Ano | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|-----------|---------|---------|---------|---------|---------|----------|----------|
| 1 | 1624985 | 6344145 | 1636191 | 5711743 | 6581454 | 18684340 | 12407967 |
| 2 | 2082197 | 3238140 | 4014982 | 2552623 | 2169927 | 774490 | 6131089 |
| 3 | 2414528 | 1551784 | 2190882 | 1460677 | 1221678 | 515440 | 655527 |
| 4 | 2906008 | 1260213 | 1433972 | 844435 | 756681 | 337330 | 436980 |
| 5 | 386476 | 1360066 | 1185007 | 595713 | 531945 | 275530 | 231591 |
| 6+ | 11964 | 202795 | 979993 | 469137 | 613224 | 183680 | 265765 |

Quadro B.3: Dados da série da campanha portuguesa de Novembro (em milhões) por ano e idades, $I2_{ay}$.

| Idade\Ano | 1984 | 1985 | 1986 | 1987 | 1992 | 1997 | 1998 | 1999 | 2000 | 2001 |
|-----------|---------|---------|---------|---------|---------|---------|---------|---------|----------|---------|
| 0 | 2956621 | 2063177 | 2493102 | 3714540 | 6349072 | 2424702 | 8680376 | 3696787 | 30871080 | 8955265 |
| 1 | 5733231 | 2743525 | 1611895 | 2379377 | 5480539 | 1961202 | 1809393 | 798000 | 1615890 | 5394731 |
| 2 | 1152160 | 4548240 | 1669563 | 1343695 | 1157103 | 906448 | 1214608 | 646000 | 246620 | 694782 |
| 3 | 1036826 | 1083437 | 658385 | 928682 | 1002580 | 728899 | 823316 | 391121 | 89920 | 521626 |
| 4 | 528343 | 839215 | 322912 | 665600 | 437424 | 1040594 | 396247 | 4593424 | 121900 | 116260 |
| 5 | 76423 | 143789 | 127266 | 236473 | 108224 | 771805 | 367120 | 382447 | 93970 | 124615 |
| 6+ | 40140 | 69987 | 49634 | 79903 | 18772 | 322421 | 220416 | 164649 | 66460 | 49336 |

Quadro B.4: Dados da série da campanha espanhola de Março (em milhões) por ano e idades, $I_{3_{ay}}$.

| Idade\Ano | 1987 | 1988 | 1990 | 1991 | 1992 | 1993 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|-----------|--------|--------|--------|--------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 44000 | 224056 | 69072 | 25415 | 167959 | 238561 | 10639 | 56495 | 509838 | 214525 | 91656 | 975603 | 270396 |
| 2 | 36000 | 63832 | 56015 | 208127 | 77477 | 427333 | 54249 | 263095 | 103126 | 160375 | 285808 | 262883 | 760202 |
| 3 | 4000 | 73627 | 272946 | 163708 | 88392 | 135919 | 90547 | 125658 | 80396 | 134618 | 435440 | 186538 | 448599 |
| 4 | 390000 | 64156 | 53317 | 400984 | 30956 | 126078 | 350825 | 123331 | 33762 | 124313 | 242249 | 142929 | 651658 |
| 5 | 118000 | 848302 | 87541 | 62373 | 116886 | 145795 | 213842 | 65713 | 20590 | 28357 | 188879 | 98945 | 318591 |
| 6+ | 245000 | 885665 | 582299 | 574261 | 122791 | 1117949 | 24779 | 61002 | 25410 | 64013 | 68124 | 66062 | 163290 |

Quadro B.5: Dados da ogiva de maturação por ano e idades, O_{ay} .

| Idade\Ano | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|-----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65 | 0.23 | 0.60 | 0.74 | 0.79 | 0.47 | 0.80 | 0.73 | 0.83 | 0.73 | 0.72 | 0.62 | 0.26 | 0.39 | 0.49 |
| 2 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.83 | 0.81 | 0.91 | 0.91 | 0.93 | 0.89 | 0.98 | 0.89 | 0.92 | 0.92 | 0.91 | 0.91 | 0.9 | 0.94 |
| 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.91 | 0.88 | 0.96 | 0.95 | 0.94 | 0.96 | 0.97 | 0.92 | 0.95 | 0.96 | 0.99 | 0.95 | 0.96 | 0.97 |
| 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.92 | 0.89 | 0.97 | 0.98 | 0.97 | 0.96 | 0.99 | 0.96 | 0.97 | 0.99 | 1.00 | 0.95 | 0.99 | 0.98 |
| 5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.94 | 0.94 | 1.00 | 1.00 | 0.99 | 0.97 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 |
| 6+ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Quadro B.6: Dados do peso médio (em kg) por ano e idades, w_{ay} .

| Idade\Ano | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|-----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.03 | 0.02 | 0.03 | 0.03 | 0.04 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 2 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.05 | 0.05 | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 |
| 3 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.06 | 0.06 | 0.06 | 0.05 | 0.05 | 0.06 | 0.06 | 0.06 |
| 4 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.07 | 0.07 | 0.07 | 0.06 | 0.06 | 0.06 | 0.08 | 0.07 |
| 5 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.08 | 0.08 | 0.07 | 0.07 | 0.07 | 0.07 | 0.08 | 0.08 |
| 6+ | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |

Quadro B.7: Estimativas do recrutamento (milhares de milhões), biomassa desovante (milhões), a captura (milhões) e o coeficiente de mortalidade por pesca médio (ano⁻¹) por ano obtidas pelo $WG2003$, R , BD_y , C_y , \bar{F}_y respectivamente.

| | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 |
|-------------|----------|----------|----------|---------|---------|----------|---------|---------|---------|----------|---------|---------|---------|
| R | 11372576 | 12963996 | 14363770 | 9501528 | 6842104 | 19612910 | 7165749 | 6100579 | 5191591 | 9299334 | 5563235 | 5681586 | 5233848 |
| BD_y | 287689 | 352291 | 431608 | 535601 | 563487 | 522262 | 576961 | 670203 | 603668 | 500991 | 439950 | 373056 | 336944 |
| C_y | 173761 | 162454 | 204861 | 242574 | 214148 | 176636 | 215114 | 219928 | 192838 | 176283 | 157273 | 146539 | 142966 |
| \bar{F}_y | 0.38 | 0.39 | 0.29 | 0.35 | 0.33 | 0.29 | 0.26 | 0.26 | 0.33 | 0.32 | 0.34 | 0.37 | 0.43 |
| | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | |
| R | 12457198 | 10553737 | 4468554 | 4353234 | 3842821 | 4517620 | 3519468 | 3773028 | 3625930 | 13172605 | 9148660 | 3635335 | |
| BD_y | 342900 | 460031 | 519001 | 526404 | 574774 | 494939 | 426555 | 345729 | 287821 | 246289 | 293065 | 501795 | |
| C_y | 132785 | 131196 | 144949 | 138725 | 126755 | 115179 | 117250 | 112033 | 95793 | 87272 | 102903 | 101741 | |
| \bar{F}_y | 0.32 | 0.28 | 0.34 | 0.23 | 0.25 | 0.26 | 0.34 | 0.41 | 0.38 | 0.38 | 0.29 | 0.23 | |

Anexo C – Código do Modelo VPA7802 para WinBUGS e Resultados

As diferenças entre o modelo apresentado no modo *Doodle* (gráfico) e o modo linha de comandos devem-se as limitações técnicas do primeiro, ver (Spiegelhalter *et al.* 1999).

Caso 1. Modelo VPA7802 – Quando o padrão relativo de exploração é constante longo dos anos.

```

Model;
{
#informação por amostragem
  #Captura 1978-2002
    for(a in 1:A){
      for(y in 1:Y-1){
        logC[a,y]<-log(C[a,y])
        logC[a,y]~dnorm(logmu.c[a,y],tauC[a,y])
        logmu.c[a,y]<-log(F.s[a,y]/Z[a,y]*N[a,y]*(1-exp(-Z[a,y])))
      }
    }

  #CPUE
    #Cruzeiro Português de Março 1996-2003
      for(a in 2:A){
        for(y in 19:26){
          logC.s1[a,y]<-log(C.s1[a-1,y-18])
          logC.s1[a,y]~dnorm(logmuC.s1[a,y],taul[a,y])
          logmuC.s1[a,y]<-log(qs1[a,y]*exp(-Z[a,y]*75/365)*N[a,y])
        }
      }
    #Cruzeiro Português de Novembro 1984-1987, 1992, 1997-2001
      for(a in 1:7){
        for(y in 7:10){
          logC.s20[a,y]<-log(C.s20[a,y-6])
          logC.s20[a,y]~dnorm(logmuC.s20[a,y],taul[a,y])
          logmuC.s20[a,y]<-log(qs20[a,y]*exp(-Z[a,y]*315/365)*N[a,y])
        }
      }
    for(a in 1:7){
      logC.s21[a,15]<-log(C.s21[a])
      logC.s21[a,15]~dnorm(logmuC.s21[a,15],taul[a,15])
      logmuC.s21[a,15]<-log(qs21[a,15]*exp(-Z[a,15]*315/365)*N[a,15])
    }
    for(a in 1:7){
      for(y in 20:24){
        logC.s2[a,y]<-log(C.s2[a,y-19])
        logC.s2[a,y]~dnorm(logmuC.s2[a,y],taul[a,y])
        logmuC.s2[a,y]<-log(qs22[a,y]*exp(-Z[a,y]*315/365)*N[a,y])
      }
    }
  #Cruzeiro Espanhol de Março 1987-1988, 1990-1993, 1996-2002
    for(a in 2:7){

```

```

        for(y in 10:11){
            logC.s30[a,y]<-log(C.s30[a-1,y-9])
            logC.s30[a,y]~dnorm(logmuC.s30[a,y],taul[a,y])
            logmuC.s30[a,y]<-log(qs30[a,y]*exp(-Z[a,y]*75/365)*N[a,y])
        }
    }
for(a in 2:7){
    for(y in 13:16){
        logC.s31[a,y]<-log(C.s31[a-1,y-12])
        logC.s31[a,y]~dnorm(logmuC.s31[a,y],taul[a,y])
        logmuC.s31[a,y]<-log(qs31[a,y]*exp(-Z[a,y]*75/365)*N[a,y])
    }
}
for(a in 2:7){
    for(y in 19:25){
        logC.s3[a,y]<-log(C.s3[a-1,y-18])
        logC.s3[a,y]~dnorm(logmuC.s3[a,y],taul[a,y])
        logmuC.s3[a,y]<-log(qs32[a,y]*exp(-Z[a,y]*75/365)*N[a,y])
    }
}

#Informação a priori
#Coeficiente de Mortalidade Total
for(a in 1:A){
    for(y in 1:Y){
        Z[a,y]<-F.s[a,y]+M
        F.s[a,y]<-Fageyear[a,y]
    }
}

#Coeficiente de Mortalidade Natural
M<-0.33
#Coeficiente de Mortalidade por Pesca - F separável (componente padrão relativo de exploração, Sage, e nível de pesca,
Fyear)
for(a in 1:A){
    for(y in 1:Y){
        Fageyear[a,y]<-Sage[a]*Fyear[y]
    }
}
for(a in 1:A){
    for(y in 1:Y){
        Sage1[a]~dunif(0,2)
        Sage[a,y]<-Sage1[a]
    }
}
for(y in 1:Y){
    Fyear1[y]~dunif(0,1)
    Fyear[y]<-Fyear1[y]
}

#Coeficiente de Capturabilidade para cada Campanha
#Cruzeiro Português de Março 1996-2003
for(a in 2:A){
    for(y in 19:26){
        qs1[a,y]<-qs.sa1[a]
    }
}
#Cruzeiro Português de Novembro 1984-1987, 1992, 1997-2001
for(a in 1:A){
    for(y in 7:10){
        qs20[a,y]<-qs.sa20[a]
    }
}
for(a in 1:A){
    qs21[a,15]<-qs.sa21[a]
    for(y in 20:24){
        qs22[a,y]<-qs.sa22[a]
    }
}

```

```

    }
#Cruzeiro Espanhol de Março 1987-1988, 1990-1993, 1996-2002
  for(a in 2:A){
    for(y in 10:11){
      qs30[a,y]<-qs.sa30[a]
    }
  }
  for(a in 2:A){
    for(y in 13:16){
      qs31[a,y]<-qs.sa31[a]
    }
  }
  for(a in 2:A){
    for(y in 19:25){
      qs32[a,y]<-qs.sa32[a]
    }
  }
  for(a in 1:A){
    qs.sa1[a]~dunif(0,2)
    qs.sa20[a]~dunif(0,2)
    qs.sa21[a]~dunif(0,2)
    qs.sa22[a]~dunif(0,2)
    qs.sa30[a]~dunif(0,2)
    qs.sa31[a]~dunif(0,2)
    qs.sa32[a]~dunif(0,2)
  }
#Matriz N
  for(y in 1:Y){
    N.year1[y]~dunif(0,1)
    N[1,y]<-N.year1[y]*12.0E10
  }
  N.age[1]<-N.year1[1]
  for(a in 1:A-1){
    N.age1[a]~dunif(0,1)
    N.age[a+1]<-N.age1[a]
    N[a+1,1]<-N.age[a+1]*12.0E10
  }
  for(y in 1:(Y)){
    R[y]<-N[1,y]
  }
  for(a in 2:A-1){
    for(y in 2:Y){
      N[a,y]<-N[a-1,y-1]*exp(-Z[a-1,y-1])
    }
  }
  for(y in 2:Y){
    N[7,y]<-N[6,y-1]*exp(-Z[6,y-1])+N[7,y-1]*exp(-Z[7,y-1])
  }

#Priors para taus
  varC~dunif(0,1)
  varI~dunif(0,1)
  for(a in 1:A){
    for(y in 1:Y){
      tauC[a,y]<-1/varC
      tauI[a,y]<-1/varI
    }
  }

#Algumas Funções
# F médio para cada ano y entre as idades 2-5
  for(y in 1:Y){
    Fsum[y]<-sum(F.s[3:6,y])
    Fbar[y]<-Fsum[y]/4
  }

#SSB
  for(a in 1:A){

```

```
        for(y in 1:Y){
            SSB[a,y]<-N[a,y]*w[a,y]*mat[a,y]
        }
    for(y in 1:Y){
        SSBY[y]<-sum(SSB[1:A,y])
    }
# Residuos para a Captura
    for(a in 1:A){
        for(y in 1:Y-1){
            C.esp[a,y]<-F.s[a,y]/Z[a,y]*N[a,y]*(1-exp(-Z[a,y]))
            res.C[a,y]<-C[a,y]-C.esp[a,y]
        }
    }
    for(y in 1:Y-1){
        C.espT[y]<-sum(C.esp[1:A,y])
    }
}
```

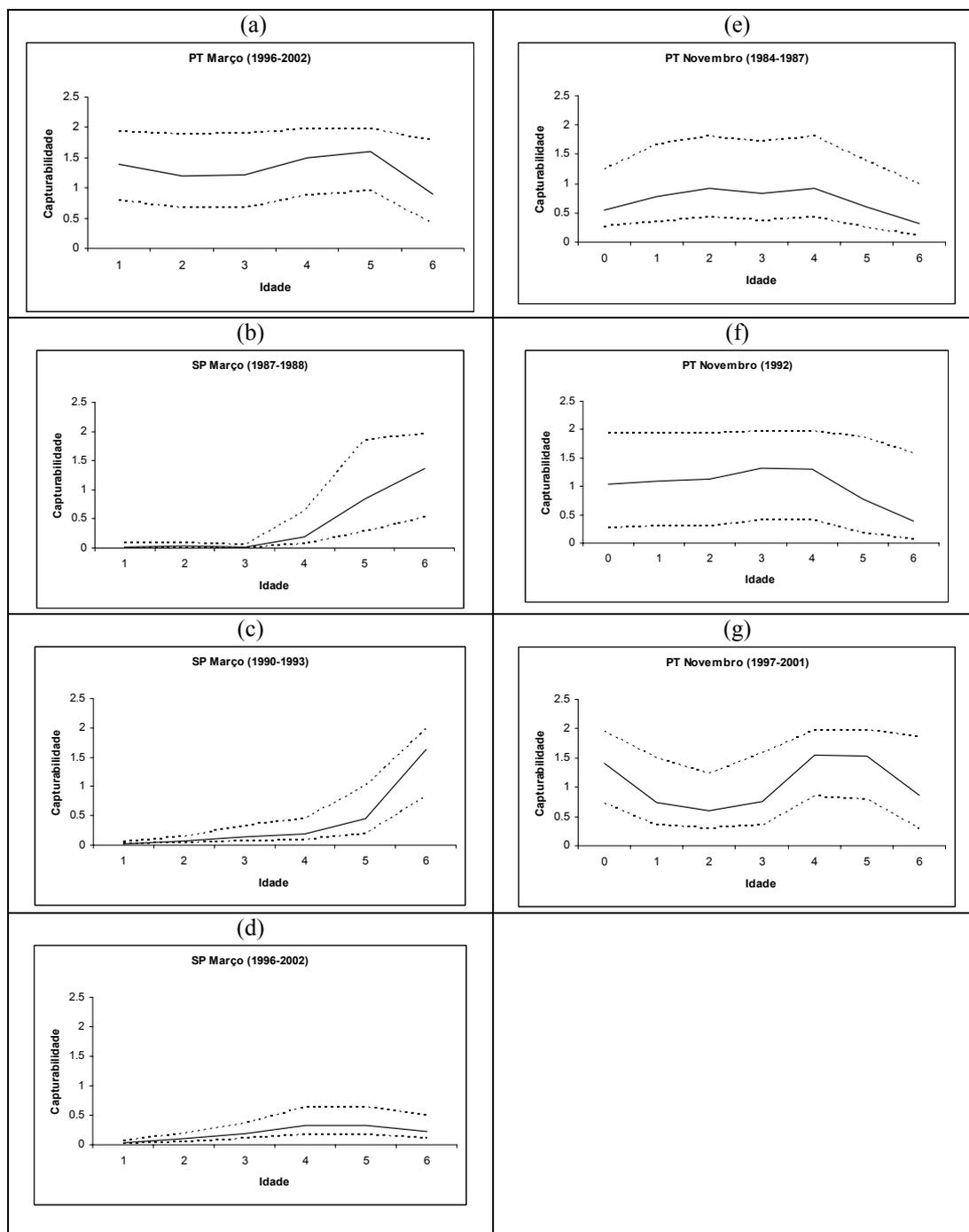


Figura C.1: Estimativas da capturabilidade para cada campanha de investigação portuguesa do mês Março (a), portuguesa de Novembro [(b), (c) e (d)] e espanhola de Março [(f), (g) e (h)] para o período de 1978 a 2003, obtidas com o Modelo *VPA7802*, quando o padrão relativo de exploração é constante ao longo dos anos.

Caso 2. Modelo *VPA7802* – Quando o padrão relativo de exploração varia ao longo dos anos.

Todo o resto do código mantém-se, com a excepção de:

```
for(a in 1:A){
  for(y in 1:Y){
    Sage1[a]~dunif(0,2)
    Sage[a,y]<-Sage1[a]
  }
}
```

que muda para

```
for(a in 1:A){
  for(y in 1:Y){
    Sage1[a,y]~dunif(0,2)
    Sage[a,y]<-Sage1[a,y]
  }
}
```

Anexo D – Código do Modelo *PROJ* para S-PLUS

```
function(data, seed,n)
{
  " ***** Modelo PROJ ***** "
  "Funcao reamostra em cada ano projectado F(i) dado sigma e N(i)"
  "Projecta N, C e a BD"
  "O Recrutamento em cada ano depende da fertilidade, lambda"
  "data=ficheiro dados com N,sdN,f,m,sdLF,om,w,wc"
  "n=numero de reamostragens"
  "assumido que : - lnF(i)~N(miu(lnF),sdLF)"
  "          - N~N(Ni,sdN)
  "          - lambda(i)~N(lambdac,sdambdac)"
  " ***** "
  " ***** matriz a ***** "
  gp <- nrow(data)
  vec <- matrix(1, nrow = gp - 1, ncol = 1)
  mat <- diag(vec, nrow = gp - 1, ncol = gp - 1)
  mat1 <- cbind(0, mat)
  a <- rbind(mat1, mat1[gp - 1, ])
  vec1 <- matrix(0, gp - 1, 1)
  vecR <- c()
  " ***** criacao matriz s ***** "
  z <- data$f + data$m
  s2 <- exp(- z)
  s1 <- s2 * a
  s <- as.matrix(s1)
  " ***** logaritmicacao dos Fs ***** "
  a1 <- log(data$f)
  a1 <- cbind(a1, data$sdLf)
  d1 <- data$N
  d1 <- cbind(d1, data$sdN)
  matout <- c()
}
```

```

" ***** loop reamostragem Fs e projeccoes ***** "
set.seed(seed)
for(i in 1:n) {
  b <- rnorm(gp, a1[, 1], a1[, 2])
  d <- rnorm(gp, d1[, 1], d1[, 2])
  N1 <- d
  f1 <- exp(b)
  z <- f1 + data$m
  s2 <- exp(-z)
  s1 <- s2 * a
  s <- as.matrix(s1)
  nstart <- sum(data$N)
  y1start <- (f1/z) * N1 * (1 - exp(-z)) * data$wc
  ystart <- sum(y1start)
  SSB1start <- N1 * data$Som * data$w
  SSBstart <- sum(SSB1start)
  nproj2 <- N1 %*% s
  "\tBD <- sum(nproj2 * data$Som)"
  BD <- SSBstart
  "**** categorias de BD e valores médios de lambda****"
  if(BD > 0 & BD < 259999)
    lambda <- 63.3
  else if(BD > 260000 & BD < 349999)
    lambda <- rnorm(1, 24.6, 13.5)
  else if(BD > 350000 & BD < 449999)
    lambda <- rnorm(1, 24.7, 13.5)
  else if(BD > 450000 & BD < 599999)
    lambda <- rnorm(1, 18.6, 10.5)
  else if(BD > 600000 & BD < 1000000)
    lambda <- rnorm(1, 13, 3.2)
  else "aviso"
  R <- BD * lambda
  vec2 <- rbind(R, vec1)
  vecR <- c(vec2)
  nproj <- nproj2 + vecR
  n <- sum(nproj)
  yproj <- (f1/z) * nproj * (1 - exp(-z)) * data$wc
  y <- sum(yproj)
  BDproj <- nproj * data$Som * data$w
  BDprojT <- sum(BDproj)
  vecoutn <- c()
  vecouty <- c()
  vecoutSSB <- c()
  vecres <- c(y, BD, R, yproj)
  matout <- rbind(matout, vecres)
}
matout
}

```