

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

**Estimação do índice de abundância de um estoque pesqueiro  
com estrutura de correlação espacial: uma abordagem  
bayesiana**

Júlio César Pereira

Tese apresentada para obtenção do título de Doutor em  
Agronomia. Área de concentração: Estatística e Experi-  
mentação Agronômica.

Piracicaba  
2009

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Júlio César Pereira  
Matemático

**Estimação do índice de abundância de um estoque pesqueiro  
com estrutura de correlação espacial: uma abordagem  
bayesiana**

Orientadora:

Prof<sup>a</sup>. Dr<sup>a</sup>. **ROSELI APARECIDA LEANDRO**

Tese apresentada para obtenção do título de Doutor em  
Agronomia. Área de concentração: Estatística e Experi-  
mentação Agronômica.

Piracicaba  
2009

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

Pereira, Júlio César

Estimação do índice de abundância de um estoque pesqueiro com estrutura de correlação espacial: uma abordagem bayesiana / Júlio César Pereira. - - Piracicaba, 2009. 119 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2009. Bibliografia.

1. Geostatística 2. Inferência bayesiana 3. Pesca 4. Recursos pesqueiros - Simulação  
I. Título

CDD 639.2

P436e

**“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”**

## Dedicatória

À minha amada esposa  
**Adriana Aparecida Alves da Silva Pereira**  
ofereço.

**A Deus,**  
dedico.

## Agradecimentos

À Profa. Dra. Roseli Aparecida Leandro, pela orientação, amizade, ensinamentos, sugestões, paciência e apoio, que foram fundamentais para a realização deste trabalho.

Aos dois grandes cientistas que tive a oportunidade de conhecer e a felicidade de trabalhar em parceria, Profa. Dra. Alexandra M. Schmidt da UFRJ e ao Prof. Dr. Miguel Petrere Jr. (o cientista pilarense) da UNESP - Rio Claro, por sempre demonstrarem boa vontade em co-orientar este trabalho, fazendo-o por amor à ciência. Agradeço, também, pelos ensinamentos, pela paciência e pela amizade.

Ao Prof. Dr. Paulo Justiniano Ribeiro Jr. e à Profa. Dra. Clarice Garcia Borges Demétrio, pelo apoio e pelas valiosas sugestões.

Ao Prof. Dr. Acácio Ribeiro G. Tomás do Instituto de Pesca, por ceder os dados.

Ao Prof. Dr. Valdemar Antonio Demétrio da ESALQ/USP e ao Iata Anderson, doutorando em geologia pela UNESP - Rio Claro, por me ajudarem a transformar coordenadas geodésicas em UTM.

Ao Joaquim Henriques Vianna Neto, Lucimary Afonso dos Santos, Juliana Cespedes, Paulo Travassos, Júlia Pereira, Gonzalo, Henrique e a todos aqueles que de alguma forma contribuíram.

A todos os amigos do mestrado e doutorado, em especial ao Vanderly Janeiro e ao Renato Rodrigues Silva, pelas dicas, discussões e troca de experiências.

À minha esposa Adriana, por ceder grande parte do tempo que poderíamos ter passado juntos para que eu pudesse me dedicar ao doutorado, pelo amor, apoio, compreensão e paciência.

A toda minha família, por me incentivar constantemente.

Aos amigos, Leandro (Milhause), Adenilson, Salomão, Lilo e Ticiane (Gulozeima), que me acolheram nas repúblicas por onde passei, por terem sido muito gentis. E aos amigos Fábio e Tunico, que há muito tempo estão sempre por perto.

Aos professores e funcionários de Departamento de Ciências Exatas da ESALQ/USP que propiciaram condições para a realização deste trabalho.

À coordenação acadêmica da UFSCAR, campus de Sorocaba, por conceder o afastamento parcial das minhas atividades docentes para que eu pudesse concluir o doutorado e por disponibilizar o laboratório de informática onde foram realizadas parte das simulações. Em especial, agradeço ao Prof. Dr. Isaias Torres, coordenador do curso de Engenharia de Produção, ao Prof. Dr. Walde-mar Marques, coordenador acadêmico, e ao Prof. Dr. Miguel Angel Aires Borrás, coordenador administrativo.

Ao CNPq pelo apoio financeiro.

## SUMÁRIO

|   |    |
|---|----|
| RESUMO . . . . .  | 7  |
| ABSTRACT . . . . .  | 8  |
| 1 INTRODUÇÃO . . . . .  | 9  |
| Referências . . . . .   | 12 |
| 2 COMPARAÇÃO DOS MÉTODOS DE ESTIMAÇÃO DE TRÊS ÍNDICES DE<br>ABUNDÂNCIA DE UM ESTOQUE PESQUEIRO . . . . .                      | 15 |
| Resumo . . . . .  | 15 |
| Abstract . . . . .  | 15 |
| 2.1 Introdução . . . . .  | 16 |
| 2.2 Desenvolvimento . . . . .   | 18 |
| 2.2.1 Conceitos básicos de geoestatística . . . . .   | 20 |
| 2.2.2 Modelo de co-regionalização linear . . . . .  | 23 |
| 2.2.3 Metodologia . . . . .   | 24 |
| 2.2.3.1 Distribuição <i>a priori</i> para os parâmetros . . . . .   | 25 |
| 2.2.3.2 Função de verossimilhança . . . . .   | 26 |
| 2.2.3.3 Distribuições <i>a posteriori</i> para os parâmetros . . . . .  | 26 |
| 2.2.3.4 Distribuição preditiva . . . . .  | 27 |
| 2.2.4 Estudo de Simulação . . . . .   | 28 |
| 2.2.5 Resultados e discussão . . . . .  | 33 |
| 2.2.6 Aplicação . . . . .   | 45 |
| 2.3 Considerações Finais . . . . .  | 53 |
| Referências . . . . .   | 54 |
| 3 COMPARAÇÃO ENTRE MODELOS GEOESTATÍSTICOS UNIVARIADO E BI-<br>VARIADO NA ESTIMAÇÃO DA CPUE: UM ESTUDO DE SIMULAÇÃO . . . . . | 58 |
| Resumo . . . . .  | 58 |
| Abstract . . . . .  | 58 |
| 3.1 Introdução . . . . .  | 59 |
| 3.2 Desenvolvimento . . . . .   | 61 |
| 3.2.1 Aplicações da geoestatística à pesca . . . . .  | 63 |
| 3.2.2 Metodologia . . . . .   | 64 |
| 3.2.2.1 Modelo bivariado para esforço e captura . . . . .   | 64 |
| 3.2.2.2 Modelo univariado para a razão entre captura e esforço . . . . .  | 67 |
| 3.2.3 Estudo de simulação . . . . .   | 68 |
| 3.2.4 Resultados e Discussão . . . . .  | 71 |
| 3.2.5 Aplicação . . . . .   | 76 |

3.3 Considerações Finais . . . . . 84  
Referências . . . . . 84  
APÊNDICE . . . . . 88  
ANEXO . . . . . 106



## RESUMO

### **Estimação do índice de abundância de um estoque pesqueiro com estrutura de correlação espacial: uma abordagem bayesiana**

Dados de captura e esforço estão entre as informações mais importantes a serem obtidas na pesca. Muitas vezes, ao se fazer a avaliação de um estoque pesqueiro, são utilizados índices de abundância baseados na captura por unidade de esforço. Como esses índices são utilizados para a tomada de decisão sobre a regulamentação e distribuição da atividade pesqueira, é importante conhecer quais índices se aplicam melhor em diferentes cenários, bem como a melhor maneira de estimá-los. Na literatura, alguns trabalhos tratam as diferentes localizações geográficas onde ocorrem as pescarias como covariável, considerando que esse é todo o efeito espacial presente nos dados. Entretanto, é possível que ainda haja efeitos remanescentes de segunda ordem. Assim, seria importante utilizar modelos que permitam capturar essa estrutura de correlação espacial. Outros trabalhos, que utilizam métodos geoestatísticos, consideram a razão entre a captura e o esforço como sendo a única variável resposta. Nesse tipo de abordagem, não se considera a relação que pode existir entre a variável esforço em uma localização  $s$  qualquer e a captura em outra localização  $s'$  e vice-versa (a covariância cruzada). Esses fatos motivaram a elaboração dos dois artigos apresentados nesta tese. No primeiro artigo foram comparadas estimativas de três índices de abundância relativa, considerando-se dois métodos distintos: 1. usando apenas dados observados nos locais de ocorrência de pesca; 2. usando o método proposto neste artigo, que consiste em se obter estimativas após a interpolação da captura e do esforço nos locais não observados, através de um modelo ajustado aos dados. Para a comparação dos métodos de estimação e das estimativas dos três índices, foram simulados dados de esforço e captura em diferentes cenários. O estudo de simulação realizado mostrou que o segundo método de estimação melhora as estimativas dos índices, principalmente quando se tem correlação entre esforço e captura e correlação espacial. Mostrou ainda, que as estimativas do índice  $CPUE_1$ , apresentam, em geral, melhores resultados para os diferentes cenários analisados. No segundo artigo, foi realizado um estudo de simulação para comparar estimativas de um índice de abundância relativa, baseado na captura por unidade de esforço considerando-se o ajuste de um modelo geoestatístico univariado para a razão entre captura e esforço, e um bivariado, em que são modelados conjuntamente a captura e o esforço. As estimativas obtidas após o ajuste dos dois modelos apresentaram resultados muito próximos, indicando que não há vantagem em usar o modelo bivariado na estimação do índice, para os cenários analisados.

Palavras-chave: Estudo de simulação; Captura por unidade de esforço; Pesca comercial; Geoestatística; Modelo de co-regionalização linear

## ABSTRACT

### **Estimation of the indice of abundance of fish stocks considering a structure of spatial correlation: a bayesian approach**

Catch and effort data are among the most important information to be obtained in fishing. Often when making an assessment of fish stocks, are used indices of abundance based on catch per unit effort. Because these indices are used for decision on regulation and distribution of fishing activity, it is important to know which indices are better applied in different sceneries and the best way to estimate them. Some studies in the literature deal with different geographical locations where occur the fisheries as a covariate, considering that this is the whole spacial effect in the data. However, it is possible that there are still remaining second-round effects. It would be important use models that allow to capture spatial structure of correlation. Other works, using geostatistical methods, consider the ratio between catch and effort as the only response variable. In this approach, it is not considered the relationship that exist between the variable effort in a location  $s$  and the catch in any other location  $s'$  and vice versa (the cross-covariance). These facts motivated the development of the two articles in this thesis. In the first article we compared estimates of three indices of relative abundance, considering two distinct methods: 1. using only data observed in places where the fishing occur, 2. using the proposed method in this article that is to get estimates after the capture and interpolation efforts in places not observed, through a model fitted to data. To compare the estimation methods and the estimates of the three indices were simulated data of effort and catch on different scenarios. The simulation study showed that the second estimation method improves the estimates of the indices, especially when there is correlation between effort and catch and spatial correlation, and that estimates of the indice  $CPUE_1$ , present, in general, better results for different scenarios analyzed. In the second article was done a simulation study to compare estimates of an indice of abundance based on the catch per unit effort considering the fit of a geostatistical univariate model for the ratio between catch and effort and a bivariate model, where the catch and effort are modeled together. The estimates obtained after the two fitted models showed very similar results, indicating that there is no advantage in using the bivariate model in the estimation of the indice to analyzed scenarios.

Keywords: Simulation; Catch per unit effort; Commercial fishing; Geostatistics; Linear model of coregionalization

# 1 INTRODUÇÃO

A pesca, uma das atividades econômicas mais antigas do homem, possui enorme relevância na alimentação de vários povos (FAVERET FILHO; SIQUEIRA, 1997). Uma revisão da situação da pesca marítima no mundo, realizada pela Organização das Nações Unidas para Agricultura e Alimentação - FAO, mostrou que a taxa de crescimento relativa anual das capturas desembarcadas vem diminuindo significativamente desde 1950. Essa é uma indicação de que a produção máxima mundial dos recursos marinhos convencionais, sob o regime de exploração atual, está próxima de ser alcançada. Além disso, os estudos sugerem que a maioria dos estoques pesqueiros marinhos necessita de medidas de manejo para ser recuperada (FAO, 1997).

Dados de captura e esforço estão entre as informações mais importantes a serem obtidas na pesca. Muitas vezes, ao se fazer a avaliação de um estoque pesqueiro, a captura por unidade de esforço (CPUE) é utilizada como índice da abundância do estoque (FAO, 2006). O termo CPUE é utilizado para designar a quantidade capturada por unidade de esforço de pesca, como, por exemplo, número de peixes por anzol ao mês. Os dados de CPUE podem estar disponíveis por embarcação, por unidade de tempo (mês, ano, etc.), por unidade de área.

A captura pode ser medida em peso (kg) ou em número de peixes. O esforço de pesca é a quantidade de operações ou de tempo de operação das artes de pesca numa determinada pescaria, durante um período determinado. A unidade do esforço pode ser, por exemplo, dias no mar ou dias de pesca, ou de forma mais precisa, número de anzóis por unidade de tempo (NADAL - EGEEA, 1996).

Diferentes índices de abundância de um estoque pesqueiro podem ser estabelecidos usando o esforço de pesca e a captura. Um índice usual para uma dada região, que será denotado por  $CPUE_1$ , é dado pela média das razões entre captura e o esforço de pesca. Outro índice muito utilizado, denotado por  $CPUE_2$ , é dado pela razão entre o total das capturas e o total do esforço para toda a região. Um terceiro índice, denotado por  $CPUE_3$ , é dado pela razão entre a soma das capturas ponderadas pelos esforços e a soma dos quadrados dos esforços. Petrere Jr., Giacomini e De Marco Jr. (2007) realizaram um estudo de simulação com o propósito de comparar os três índices, usando diferentes distribuições de probabilidade e diferentes variâncias. As estimativas para os índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , obtidas a partir de uma amostra da captura e do esforço em uma região considerada para análise, são estimativas do tipo razão (COCHRAN, 1963; SNEDECOR; COCHRAN, 1967).

Devido às dificuldades em se fazer uma amostragem probabilística em grandes regiões oceânicas é necessário, geralmente, usar dados de pesca comercial para se avaliar um estoque pesqueiro. Registros detalhados da pesca, em que são fornecidas as coordenadas geográficas de

onde ocorreram as pescarias, permitem que se faça uma análise espacial da pesca. É importante observar que os dados de pesca comercial resultam de uma amostragem não probabilística, ou seja, os locais onde são realizadas as pescarias não são escolhidos segundo um delineamento amostral. Em geral são escolhidos a partir de algum conhecimento prévio do pescador a respeito do estoque e levando-se em conta outros fatores, como, por exemplo, o custo de deslocamento até determinado local (HILBORN; WALTERS, 1992). Dessa maneira, deve-se ter cuidado ao tirar conclusões quando se analisa dados de pesca comercial. Por exemplo, se a pesca é realizada em regiões com maior concentração de peixes, os índices de abundância tendem a relatar uma densidade maior que a densidade populacional (QUINN; DERISO, 1999). Esse fenômeno é chamado de hiperestabilidade (HILBORN; WALTERS, 1992). Porém, esse tipo de problema não é o foco deste trabalho.

Uma outra situação que ocorre na pesca, para uma dada região, é que às vezes a pesca não é realizada em todos os quadratins (sub-regiões delimitadas por paralelos e meridianos). Nessa situação, há que se fazer alguma suposição do que teria ocorrido nesses quadratins não observados. Nesse caso, estimar cada um dos três índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , usando somente os dados observados de captura e esforço, pode não ser a melhor opção. Segundo Walters (2003), se for possível mostrar que existe uma estrutura de correlação espacial nas taxas de captura, pode-se usar estatística espacial para fazer a interpolação nos quadratins onde não ocorreram pescas, antes de se estimarem esses índices. Além disso, é possível usar covariáveis, tais como, frota, ano, temperatura da água, para fornecer estimativas da tendência espacial (WALTERS, 2003; NISHIDA; CHEN, 2004).

Como muitas vezes algum índice de abundância do estoque é usado para a tomada de decisão sobre a regulamentação e distribuição da atividade pesqueira, é importante conhecer qual dos três índices ( $CPUE_1$ ,  $CPUE_2$  ou  $CPUE_3$ ) se aplica melhor em diferentes cenários, bem como a melhor maneira de estimá-los. Na literatura, alguns trabalhos tratam as diferentes localizações geográficas onde ocorrem as pescarias como covariável (LARGE, 1992; ABUABARA, 1996; MATSUNAGA; HOSONO; SHONO, 2006), considerando que esse é todo o efeito espacial presente nos dados. Entretanto, é possível que ainda haja efeitos espaciais remanescentes de segunda ordem. Assim, seria natural permitir aos modelos que considerem uma estrutura de correlação espacial. Outros trabalhos, que utilizam métodos geoestatísticos, consideram a razão entre a captura e o esforço como sendo a única variável resposta (VIGNAUX, 1996; STELZENMÜLLER; MAYNOU; MARTÍN, 2007). Nesse tipo de abordagem, não se considera a relação que pode existir entre a variável esforço em uma localização  $s$  qualquer e a captura em outra localização  $s'$  e vice-versa (a covariância cruzada). Esses fatos motivaram a elaboração dos dois artigos apresentados nos próximos capítulos.

No primeiro artigo é feito um estudo de simulação considerando um modelo conjunto para a captura e o esforço, ou seja, foram consideradas como variáveis respostas a captura e o esforço de pesca e como covariável a temperatura. O modelo adotado contém uma componente aleatória que

descreve a dependência espacial, a relação entre captura e esforço e a covariância cruzada, ou seja, a relação entre a captura na localização  $s$  e o esforço na localização  $s'$  e vice-versa. A motivação para considerar um modelo conjunto para a captura e o esforço está associada a algumas características que são inerentes a certos conjuntos de dados de pesca, a saber: existência de correlação espacial nos dados observados (SWAIN; WADE, 2004; WALTERS; MARTELL, 2004) e existência de uma relação entre captura e esforço. Na prática, os dados observados são bivariados (esforço ( $E$ ) e captura ( $C$ )) (ICCAT, 2007) e não univariado (taxa de captura  $C/E$ ), como usualmente é tratado na literatura; o esforço não é fixo, uma vez que ele pode depender, por exemplo, do valor comercial da espécie alvo, período do ano, condições climáticas, temperatura da superfície oceânica, percepção dos pescadores em observar um cardume ou não, informações de outros pescadores (SANCHÉZ; MAYNOU; DEMESTRE, 2004; WALTERS, 2004; HILBORN; WALTERS, 1992).

Os objetivos desse artigo são i) Comparar as estimativas dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  obtidos por dois métodos distintos. O primeiro método consiste, simplesmente, em substituir os dados de esforço e captura, observados nos quadratins onde as pescarias foram realizadas, no cálculo dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  propostos na literatura. O segundo consiste em se obter estimativas desses índices considerando-se os dados de captura e esforço nos quadratins onde a pesca foi realizada e os dados obtidos por interpolação nos quadratins da região nos quais a pescaria não foi realizada. A interpolação foi realizada usando-se: um modelo geoestatístico bivariado proposto por Gelfand et al. (2004), em que o esforço e a captura são modelados conjuntamente; um modelo geoestatístico em que o esforço é considerado conhecido; e um modelo bivariado em que esforço e captura são independentes no espaço. ii) Comparar entre si os três índices de abundância  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ .

Este artigo busca dar respaldo à escolha do índice de abundância relativa ( $CPUE_1$ ,  $CPUE_2$  ou  $CPUE_3$ ) mais apropriado para diferentes cenários, e fornecer um método de estimação que pode melhorar as estimativas desses índices para dados de pesca de um estoque pesqueiro espacialmente estruturado. Dessa forma, espera contribuir para que sejam evitados equívocos na tomada de decisão e no gerenciamento de estoques pesqueiros.

O segundo artigo foi motivado pelo fato de existirem trabalhos na literatura, como por exemplo, Large (1992), que propõem modelos para a razão  $C/E$ , ou seja, tratam  $C/E$  como variável resposta. Decidiu-se comparar, através de um estudo de simulação, a abordagem conjunta proposta no primeiro artigo, com uma abordagem univariada, em que é proposto um modelo geoestatístico univariado para a razão  $C/E$ , na estimação da  $CPUE$ . Ao propor um modelo geoestatístico univariado para a variável  $C/E$ , só é possível estimar a  $CPUE_1$ , de forma que a comparação entre a abordagem conjunta e univariada é realizada sobre a estimação da  $CPUE_1$ . Assim, este artigo contribui na escolha da melhor forma de modelar (modelagem conjunta ou univariada) os dados de esforço e captura, quando o objetivo é a estimação da  $CPUE_1$ .

O ajuste dos modelos são feitos usando a abordagem bayesiana. Dessa maneira, descreve-se a incerteza associada aos parâmetros do modelo através de uma distribuição de probabilidade *a priori*. A informação *a priori* é atualizada pela informação proveniente dos dados, resultando na distribuição de probabilidade *a posteriori*, que contém toda a informação atualizada dos parâmetros. As inferências sobre os parâmetros são feitas através da caracterização dessa distribuição (BOX; TIAO, 1992).

## Referências

ABUABARA, M. A. P. **Análise de dados de captura e esforço da pesca com espinhel de atum no Oceano Atlântico de 1974 a 1990**. 1996. 155 p. Tese (Doutorado em Engenharia) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. 1996.

BOX, G. E. P.; TIAO, G. C. **Bayesian Inference in Statistical Analysis**. New York: Wiley, 1992. 588 p.

COCHRAN, G. W. **Sampling techniques**. 2.ed. London: John Wiley & Sons, 1963. 555 p.

FAO, **Stock assessment for fishery management**: a framework guide to the stock assessment tools of the fisheries management science programme (FMSP). Roma, 2006. 261 p. (FAO. Fisheries Technical Paper, 487).

FAO. **Review of state world fishery resources**: marine fisheries. Roma, 1997. (FAO Fisheries Circular, 920). Disponível em: <<http://www.fao.org/DOCREP/003/W4248E/W4248E00.HTM>> . Acesso em: 27 jun. 2008.

FAVERET FILHO, P.; SIQUEIRA, S. H. Panorama da pesca marítima no Mundo e no Brasil.

**BNDES Setorial**, n. 5, mar. 1997. Disponível em:

<<http://www.bndes.gov.br/conhecimento/Bnset/rspesca.pdf>>. Acesso em: 27 jun. 2008.

HILBORN, R.; WALTERS, C.J. **Quantitative Fisheries Stock Assessment**: Choice, Dynamics and Uncertainty. New York: Chapman and Hall, 1992. 570 p.

COMISSÃO INTERNACIONAL PARA A CONSERVAÇÃO DOS ATUNS DO ATLÂNTICO. Banco de dados. Disponível em: <<http://www.iccat.int/accesingdb.htm>>. Acesso em: 17 abril 2007.

LARGE, P.A. Use of a multiplicative model to estimate relative abundance from commercial CPUE data. **ICES Journal of Marine Science**, Oxford, v.49, p.253-261, aug. 1992.

MATSUNAGA, H.; HOSONO, T.; SHONO H. **Analysis of longline CPUE of major pelagic shark species collected by Japanese research and training vessels in the Pacific Ocean**. National Research Institute of Far Seas Fisheries, FRA, JAPAN, 2006. 8 p. Relatório Técnico.

NADAL-EGEA, J. A. **Esfuerzo y Captura: Tecnología y Sobreexplotación de Recursos Marinos Vivos**. México: El Colegio de México, 1996. 476 p.

NISHIDA, T.; CHEN, D, G. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data. **Fisheries research**, v. 70, p. 265-274, 2004.

PETREIRE JR, M.; GIACOMINI, H.C.; DE MARCO JR. P. **Catch-per-unit effort: which estimator is best?** Rio Claro: UNESP, Departamento de Ecologia, 2006. 17 p. Relatório Técnico.

QUINN, T. J.II; DERISO. R. B. **Quantitative Fish Dynamics**. Oxford: Oxford University Press, 1999. 542 p.

SANCHEZ, P.; MAYNOU, F.; DEMESTRE, M. Modelling catch, effort and price in a juvenile *Eledone cirrhosa* fishery over a 10 year period. **Fisheries Research**, Barcelona, v.68, p.319-327, 2004.

SWAIN, D.P.; WADE, E.J. Spatial distribution of catch and effort in a fishery for snow crab (*Chionoecetes opilio*): tests of predictions of the ideal free distribution. **Canadian Journal of Fisheries and Aquatic Sciences**. Toronto, v.60, 897-909, sept. 2003.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical Methods**. 6.ed. Ames: The Iowa state University Press, 1967. 593 p.

STELZENMULLER, V.; MAYNOU, F.; MARTÍN, P. Spatial assessment of benefits of a coastal Mediterranean Marine protected area. **Biological conservation**. Barcelona, v. 36, p. 571-583, 2007.

VIGNAUX, M. Analysis of spatial structure in fish distribution using commercial catch and effort data from the New Zealand hoki fishery. **Canadian Journal of Fisheries and Aquatic Sciences**. Toronto, v. 53, p. 963-973, 1996.

WALTERS, C. Folly and fantasy in the analysis of spatial catch rate data. **Canadian Journal of Fisheries and Aquatic Sciences**. Toronto, v.60, 1433-1436, Dec. 2003.

WALTERS, C. J.; MARTELL, S. J. D. **Fisheries ecology and management**. Princeton: Princeton University Press, 2004. 448 p.



## 2 COMPARAÇÃO DOS MÉTODOS DE ESTIMAÇÃO DE TRÊS ÍNDICES DE ABUNDÂNCIA DE UM ESTOQUE PESQUEIRO

### Resumo

Muitas vezes, ao se fazer a avaliação de um estoque pesqueiro, são utilizados índices baseados na captura por unidade de esforço (CPUE) para estimar a abundância relativa do estoque. Neste trabalho, foi realizado um estudo de simulação para comparar dois métodos de estimação de três índices de abundância relativa propostos na literatura, e em cada método foram comparados os três índices. Foram gerados dados bivariados compostos por esforço e captura, em diferentes cenários, para estimar três índices de abundância relativa, denotados por  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ . O primeiro método consiste, simplesmente, em substituir os dados de esforço e captura, observados nos quadratins onde as pescarias foram realizadas, no cálculo dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  propostos na literatura. O segundo consiste em se obter estimativas desses índices considerando-se os dados de captura e esforço nos quadratins onde a pesca foi realizada e os dados obtidos, por interpolação, nos quadratins da região nos quais a pescaria não foi realizada. A interpolação espacial foi realizada usando-se: um modelo geoestatístico bivariado em que o esforço e a captura são modelados conjuntamente; um modelo geoestatístico em que o esforço é considerado conhecido; e um modelo bivariado em que esforço e captura são independentes no espaço. O ajuste dos modelos é feito utilizando-se uma abordagem bayesiana. O estudo de simulação realizado mostrou que o segundo método de estimação melhora as estimativas dos índices, principalmente quando se tem correlação entre esforço e captura e correlação espacial. E em cada método, as estimativas do índice  $CPUE_1$ , apresentam, em geral, melhores resultados para os diferentes cenários analisados. Uma aplicação da metodologia foi feita utilizando dados de pesca de arrasto do camarão rosa *Farfantepenaeus brasiliensis* and *F. paulensis* desembarcado nas regiões Sudeste e Sul do Brasil no ano de 1998.

Palavras-chave: Estudo de Simulação; Geoestatística; Captura por Unidade de Esforço; Abordagem Bayesiana; Co-regionalização linear

### Abstract

Often when you make the assessment of a fishing stocks, indices are used based on the catch per unit effort (CPUE) to estimate the relative abundance of stock. This work was done a study of simulation to compare two estimation methods of three indices of relative abundance proposed in the literature and in each method were compared the three indices. Were generated bivariate data composed by effort and catch, in different scenarios, to estimate three relative abundance indices, denoted by  $CPUE_1$ ,  $CPUE_2$  and  $CPUE_3$ . The first method is simply to replace the data effort and catch, observed in quadrat where the fisheries were made in calculating the indice  $CPUE_1$ ,  $CPUE_2$  and  $CPUE_3$  proposed in the literature. The second is to get estimates of these indices considering the data of catch and effort in quadrat where fishing was conducted and the data obtained by interpolation in quadrat of the region in which the fishery was not performed. The spatial interpolation

was performed by using: a geostatistical model in which the bivariate catch and effort are jointly modeled, a geostatistical model where the effort is known and considered a bivariate model in that effort and catch are independent in the space. The fit of the models are made using a Bayesian approach. The simulation study showed that the second method estimation improves the estimates of the indices, especially when it has correlation between effort and catch and spatial correlation. In each method, estimates of the indice  $CPUE_1$ , present in general, better results for the different analyzed scenarios. An application of the methodology was done using data from the fishing trawling for pink shrimp *Farfantepenaeus brasiliensis* and *F. paulensis* landed in the Southeast and Southern Brazil in 1998.

Keywords: Simulation; Geostatistics; Catch per unit effort; Bayesian approach, linear coregionalization

## 2.1 Introdução

Os recursos pesqueiros são componentes fundamentais para a alimentação e geração de empregos em muitos países; um outro aspecto que os tornam importantes é o seu caráter auto-renovável. Isso significa que, se um recurso pesqueiro for bem gerido, a sua duração é praticamente ilimitada (CADIMA, 2000).

Usualmente, dados de captura e esforço, resultantes de pesca comercial, são usados para avaliar um estoque pesqueiro. Dos dados de captura e esforço são obtidos índices de abundância relativa que fornecem informações sobre a tendência de um estoque ao longo do tempo. Espera-se que mudanças nos índices reflitam mudanças no estoque pesqueiro.

Pode-se definir, para uma determinada região, três índices de abundância relativa, baseados na captura por unidade de esforço, que são utilizados na literatura, a saber:

1.  $CPUE_1 = \frac{1}{n} \sum_{i=1}^n \frac{C_i}{E_i}$ ;
2.  $CPUE_2 = \frac{\sum_{i=1}^n C_i}{\sum_{i=1}^n E_i}$ ;
3.  $CPUE_3 = \frac{\sum_{i=1}^n C_i E_i}{\sum_{i=1}^n E_i^2}$ ,

em que  $C_i$  representa a captura no quadratim  $i$  e  $E_i$  o respectivo esforço. O índice  $CPUE_1$  é a média das razões entre captura e esforço, o índice  $CPUE_2$  é a razão entre os totais da captura e do esforço e  $CPUE_3$  é um estimador razão apresentado em Snedecor e Cochran (1967).

Para se obter esses índices, geralmente, não se tem os dados de captura e esforço para todos os quadratins (sub-regiões delimitadas por paralelos e meridianos), ou seja, a pesca não é realizada em toda a região estudada. Nessa situação, uma forma de estimá-los é, simplesmente, usar os dados de esforço e captura observados. No entanto, essa pode não ser a melhor opção. Segundo Walters

(2003), se for possível mostrar que existe uma estrutura de correlação espacial nos dados de razão entre captura e esforço, pode-se usar a estatística espacial para fazer a interpolação nos quadratins onde não ocorreram pescas, antes de se calcular qualquer índice de abundância. Além disso, é possível usar covariáveis, tais como, frota, ano, temperatura da água, para fornecer estimativas da tendência espacial (WALTERS, 2003; NISHIDA; CHEN, 2004).

Considerando o que sugere Walters (2003) em relação ao uso das técnicas de estatística espacial e o uso de covariáveis, e considerando, também, que, na prática, os dados observados são bivariados (esforço e captura) (ICCAT, 2007), além de outras características que são inerentes a certos conjuntos de dados de pesca, tais como existência de correlação espacial, relação entre captura e esforço, esforço aleatório, adotou-se um modelo de co-regionalização linear (GELFAND ET AL., 2004), sobre o qual foi realizado o estudo de simulação. Esse modelo, além de permitir descrever a estrutura de covariância entre as variáveis esforço e captura e a correlação espacial existente entre elas, modela a estrutura de covariância cruzada, isto é, a covariância entre o esforço em uma localização  $s_i$  qualquer e a captura em uma localização  $s_j$  qualquer, e vice-versa. Ou seja, as observações  $(E_i, C_i)$  são tratadas como uma amostra de um processo espacial bivariado. Ajustando o modelo proposto, é possível fazer a interpolação da captura e do esforço para os quadratins que não foram observados.

Uma vez que os índices de abundância são usados na avaliação de estoques pesqueiros, é relevante um estudo para averiguar quais dos três índices (CPUE<sub>1</sub>, CPUE<sub>2</sub> ou CPUE<sub>3</sub>) devem ser utilizados em diferentes situações. Sobretudo, é essencial o uso de métodos que estimem cada um dos índices de forma precisa.

Petriere Jr., Giacomini e De Marco Jr. (2007) realizaram um estudo de simulação com o propósito de comparar as CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub>. Porém, esses autores não realizaram estudos para esses índices na presença de correlação espacial. Além disso, na pesca não têm sido empregados modelos geoestatísticos multivariados na análise de dados de esforço e captura; desse modo, não são consideradas as covariâncias cruzadas espaciais entre essas variáveis.

Neste trabalho, foi realizado um estudo de simulação com o objetivo de comparar:

- i) dois métodos de estimação de três índices de abundância relativa propostos na literatura;
- ii) em cada método, comparar os três índices: CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub>.

O primeiro método consiste, simplesmente, em substituir os dados de esforço e captura, observados nos quadratins onde as pescarias foram realizadas, no cálculo dos índices CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub> propostos na literatura. O segundo consiste em se obter estimativas desses índices considerando-se os dados de captura e esforço nos quadratins onde a pesca foi realizada e os dados obtidos, por interpolação, nos quadratins da região nos quais a pescaria não foi realizada. A interpolação (ou krigagem) foi realizada usando-se: um modelo geoestatístico bivariado proposto

por Gelfand et al. (2004), em que o esforço e a captura são modelados conjuntamente; um modelo geoestatístico em que o esforço é considerado conhecido; e um modelo bivariado em que esforço e captura são independentes no espaço.

Os cenários considerados foram:

- (a) baixa correlação entre captura e esforço, retratando a alta incerteza do pescador, e considerando que os dados apresentam pouca correlação espacial;
- (b) forte correlação entre captura e esforço, retratando uma menor incerteza do pescador, e considerando que os dados apresentam pouca correlação espacial;
- (c) baixa correlação entre captura e esforço e considerando que os dados apresentam forte correlação espacial;
- (d) forte correlação entre captura e esforço, e considerando que os dados apresentam forte correlação espacial;

## 2.2 Desenvolvimento

Dados com padrões específicos de distribuição espacial podem ser encontrados nas áreas agrônômica, florestal, biológica, ambiental, da saúde, educacional, dentre outras (VIOLA, 2007). O padrão espacial dos dados pode ser classificado em três tipos básicos: aleatório, agregado ou regular (FIGURA 2.1)(REIS, 1998). Na pesca, geralmente é fornecido um ponto de referência dentro do quadratim onde ocorreram as pescarias (ICCAT, 2007). Se a pesca é realizada em todos os quadratins de uma região, e se o centro de cada quadratim for usado como referência, isso induz a uma configuração espacial regular. Geralmente, a pesca não é realizada em todos os quadratins, e a configuração que se mostra mais realista para a distribuição espacial da pesca é o padrão agregado (ANGANUZZI, 2004; HILBORN; WALTERS, 1992).

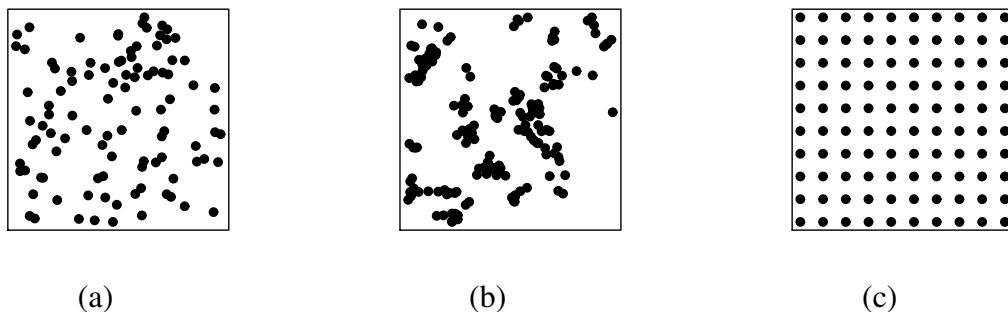


Figura 2.1 – Padrão espacial aleatório (a), agregado (b) e regular (c)

Muitos trabalhos da área de pesca têm modelado a CPUE usando diversos tipos de covariáveis. Uma importante covariável que aparece em vários trabalhos, e que está correlacionada com a captura por unidade de esforço, com a captura e com o esforço é a temperatura da superfície do mar. Dow et al. (1975), analisando dados da pesca da *Homarus americanus* encontraram forte associação da captura e do esforço com a temperatura da superfície oceânica. Posteriormente, Dow (1980) apresenta um modelo de regressão linear para desembarque da *Homarus americanus* em função da temperatura da superfície oceânica, em que o desembarque e a temperatura estão positivamente correlacionados. Evans, Lockwood e Evans (1995) encontraram outro tipo de relação, uma relação quadrática entre a CPUE das espécies (*Panulirus argus* e *Panulirus guttatus*) e a temperatura média anual da superfície oceânica. Fonteneau (1995) realizou uma análise comparativa das pescarias de atuns tropicais, com pescarias de superfície, realizadas nos três oceanos (Atlântico, Pacífico e Índico), relacionando as tendências nas capturas da albacora-laje (*Thunnus albacares*), albacora-bandolim (*Thunnus obesus*) e bonito-listrado (*Katsuwonus pelamis*) com a temperatura da superfície do mar. Sua conclusão é de que a maioria das capturas é obtida em áreas com predominância de águas quentes na superfície do mar (por exemplo, com uma temperatura média acima de 20°C), sendo este padrão térmico semelhante em todos os oceanos. Entretanto, enquanto o bonito-listrado é geralmente capturado em águas relativamente frias (em torno de 20°C), a albacora-laje é capturada, sobretudo, em águas mais quentes (acima de 25°C). Lima, Lin e Meneses (2000) também perceberam a existência de uma correlação positiva entre a temperatura e a abundância ou disponibilidade do bonito-listrado na costa das regiões Sudeste e Sul do Brasil. Goodyear (2003), analisando dados da CPUE da *Makaira nigricans* e da CPUE da *Tetrapturus albidus*, contrastou a distribuição resultante das taxas de captura dessas espécies com a temperatura média mensal da superfície oceânica, para examinar a possível associação entre temperatura e CPUE. Seus resultados indicam que a temperatura é um importante determinante na distribuição das duas espécies no oceano Atlântico. Enfim, a captura e o esforço de pesca se relacionam com a temperatura da superfície do mar de diferentes formas, para diferentes espécies, e nos diferentes locais.

Outro aspecto importante a ser considerado quando se analisa dados de pesca é a distribuição de probabilidade das variáveis envolvidas. As técnicas de geoestatística apresentadas a seguir, assumem que as variáveis a serem modeladas seguem distribuição normal. Segundo Schmidt, Nobre e Ferreira (2003), dados ambientais raramente seguem distribuição normal. Esses autores, ao analisarem dados de precipitação pluviométrica, utilizam a raiz cúbica da quantidade de chuva para obter a normalidade. Com dados de pesca não é diferente; muitas vezes, dados de esforço e captura indicam que essas variáveis não seguem distribuição normal. Isso é notado nos trabalhos de Abuabara (1996) e Sánchez, Maynou e Demestre (2004). Em geral, as distribuições dessas variáveis são assimétricas, e em muitos casos a transformação logarítmica é suficiente para corrigir a falta de normalidade.

A seguir, faz-se uma breve revisão dos conceitos de geoestatística e do modelo de co-regionalização linear necessários para o desenvolvimento deste trabalho.

### 2.2.1 Conceitos básicos de geoestatística

Segundo O'Hagan (1994), um processo gaussiano univariado pode ser definido do seguinte modo: seja  $Y(\cdot)$  uma função assumindo valores  $Y(s)$ ,  $s \in G \subset \mathbb{R}^g$ , geralmente,  $g = 1, 2$ , ou  $3$ , diz-se que  $Y(\cdot)$  segue um processo gaussiano com uma função de médias  $m(\cdot)$  e função de covariância  $C(\cdot, \cdot)$ , se para quaisquer  $s_1, s_2, \dots, s_n \in G$  e qualquer  $n = 2, 3, \dots$ , a distribuição conjunta de  $Y(s_1), Y(s_2), \dots, Y(s_n)$  é normal multivariada com parâmetros  $\mathbf{E}(Y(s_j)) = m(s_j)$  e  $cov(Y(s_i), Y(s_j)) = C(s_i, s_j)$ . Notação:  $Y(\cdot) \sim PG(m(\cdot), C(\cdot, \cdot))$ .

Na literatura de geoestatística, os processos gaussianos univariados são muito utilizados. Geralmente, tem-se um processo,  $\{Y(s), s \in G\}$ , em que  $G$  é um subconjunto fixo de um espaço euclidiano de dimensão 2 ou 3. Observada a variável  $Y(\cdot)$  para um conjunto finito de localizações  $s_1, s_2, \dots, s_n$  tem-se uma realização parcial do processo. Baseado nas  $n$  observações, tem-se interesse em fazer inferência sobre o processo  $Y(\cdot)$  e prever o processo para novas localizações de interesse (não observadas) (SCHMIDT; NOBRE; FERREIRA, 2003).

Quando se usa um processo gaussiano para modelar um fenômeno, um grande desafio é a especificação de uma função de correlação válida (SCHMIDT, CONCEIÇÃO; MOREIRA, 2007).

As duas famílias de funções de correlação mais usadas na literatura são:

#### (1) Exponencial Potência

$$\rho(d, \phi, \kappa) = \exp(-(\phi d)^\kappa),$$

em que  $\phi > 0$ ,  $\kappa \in (0, 2]$  e  $d > 0$  é a distância euclidiana entre duas localizações quaisquer  $s$  e  $s'$ . Quando  $\kappa = 1$ , tem-se o caso particular da função de correlação exponencial,  $\rho(d, \phi) = \rho(d, \phi, 1)$  e  $\kappa = 2$ , resulta na função de correlação gaussiana.

#### (2) Matérn

$$\rho(d, \phi, \kappa) = \frac{1}{2^{\kappa-1} \Gamma(\kappa)} (d\phi)^\kappa \kappa_\kappa(d\phi);$$

em que  $\phi > 0$ ,  $\kappa > 0$  e  $\kappa_\kappa$  é a função de Bessel modificada do terceiro tipo de ordem  $\kappa$ .

A especificação da família de função de correlação é de grande importância, pois em processos gaussianos, a suavidade do processo está diretamente relacionada à diferenciabilidade de sua estrutura de covariância (DIGGLE; RIBEIRO JR, 2007).

Um processo gaussiano  $Y(\cdot)$  é dito estacionário se a média de  $Y(s)$  é a mesma para todo  $s \in G$ , a variância é a mesma para todo  $s \in G$  e a correlação entre  $Y(s)$  e  $Y(s')$  depende somente

de  $s - s'$ . O Processo gaussiano é dito estacionário e isotrópico (homogêneo), se a função de correlação depende somente da distância euclidiana  $d = \|s - s'\|$ , entre as localizações em  $G$ .

Outro tipo de estacionariedade é a intrínseca. Assumindo que  $\mathbf{E}[Y(s+h) - Y(s)] = 0$ ,  $\forall s, s+h \in G$  e se  $\mathbf{E}[Y(s+h) - Y(s)]^2 = \text{var}(Y(s+h) - Y(s)) = 2\gamma(h)$ , em que, depende apenas de  $h$ , então o processo  $Y(\cdot)$  é dito intrinsecamente estacionário. A função  $2\gamma(h)$  é chamada de variograma, e  $\gamma(h)$  é chamada de semi-variograma (BANERJEE et al., 2004). Quando o processo é homogêneo denota-se o semi-variograma por  $\gamma(d)$ .

Na terminologia geoestatística,  $\tau^2 = \lim_{d \rightarrow 0^+} \gamma(d)$  é chamado de *efeito pepita* e  $\tau^2 + \sigma^2 = \lim_{d \rightarrow \infty} \gamma(d)$  é chamado de *patamar*. Quando  $\gamma(d)$  atinge o patamar para um valor finito  $d'$ , o valor  $d'$  é chamado de *alcance*. Na Figura 2.2 são ilustrados, através de um semivariograma, o efeito pepita, o alcance e o patamar. O alcance consiste na máxima distância, para a qual os pontos amostrais podem ser considerados dependentes.

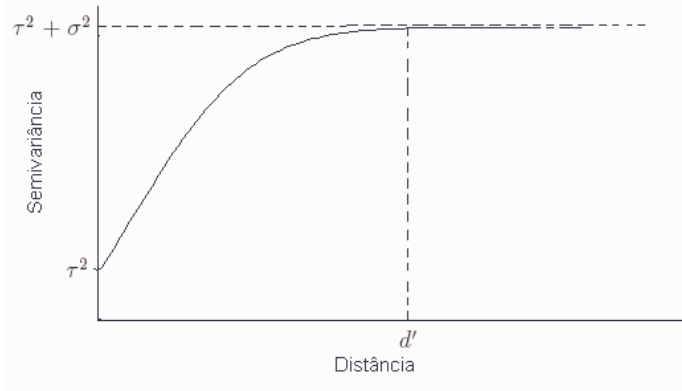


Figura 2.2 – Semivariograma ilustrando o efeito pepita, o alcance e o patamar

Freqüentemente, esse valor de distância é infinito. Por isso é utilizado o alcance efetivo. O alcance efetivo, denotado por  $d^*$ , é o valor de  $d$  para o qual o valor da correlação cai para valores pequenos, por exemplo, 0,05 (SCHMIDT; SANSÓ, 2006). Na Figura 2.3 é apresentado o gráfico de uma função de correlação, ilustrando o alcance efetivo.

A predição ou interpolação de  $Y(\cdot)$  em um ponto não observado,  $s_u \in G$  é chamada na literatura geoestatística de *krigagem*. Supondo que se deseja fazer predição (interpolação) para um conjunto de  $K$  locais não observados,  $S_u = \{s_{u1}, s_{u2}, \dots, s_{uK}\}$ , pode-se formar um vetor da variável nos pontos não observados  $\mathbf{Y}_u = (Y(s_{u1}), Y(s_{u2}), \dots, Y(s_{uK}))'$  e obter a distribuição preditiva de  $(\mathbf{Y}_u | \mathbf{Y})$  dada por

$$p(\mathbf{Y}_u | \mathbf{Y}) = \int_{\boldsymbol{\theta}} p(\mathbf{Y}_u | \mathbf{Y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{Y}) d\boldsymbol{\theta} \quad (2.1)$$

em que  $p(\mathbf{Y}_u | \mathbf{Y}, \boldsymbol{\theta})$  é a distribuição condicional de  $\mathbf{Y}_u$  dados  $\mathbf{Y}$  e  $\boldsymbol{\theta}$  e  $p(\boldsymbol{\theta} | \mathbf{Y})$  é a distribuição *a posteriori* de  $\boldsymbol{\theta}$ .

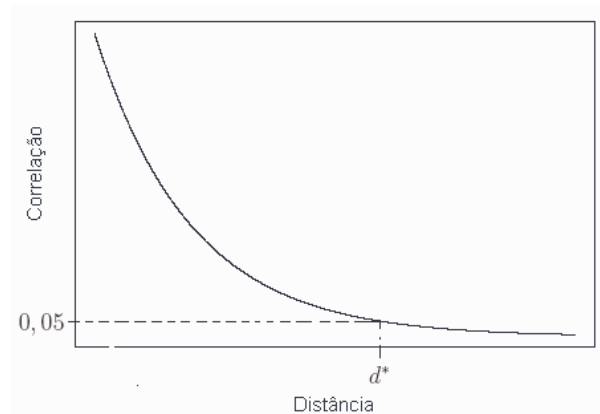


Figura 2.3 – Função de correlação ilustrando o alcance efetivo

Da teoria da distribuição normal multivariada (JOHNSON;WICHERN, 2002) segue que

$$(\mathbf{Y}_u | \mathbf{Y}, \boldsymbol{\theta}) \sim N_k(\boldsymbol{\mu}_u + \boldsymbol{\Psi}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}); \boldsymbol{\Sigma}_u - \boldsymbol{\Psi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}), \quad (2.2)$$

pois dado que  $\mathbf{Y}$  segue um processo gaussiano, tem-se que a distribuição conjunta de  $\mathbf{Y}$  e  $\mathbf{Y}_u$ , condicionada a  $\boldsymbol{\theta}$ , é dada por

$$\begin{pmatrix} \mathbf{Y}_u \\ \mathbf{Y} \end{pmatrix} | \boldsymbol{\theta} \sim N_{n+K} \left( \begin{pmatrix} \boldsymbol{\mu}_u \\ \boldsymbol{\mu} \end{pmatrix}; \begin{pmatrix} \boldsymbol{\Sigma}_u & \boldsymbol{\Psi}' \\ \boldsymbol{\Psi} & \boldsymbol{\Sigma} \end{pmatrix} \right) \quad (2.3)$$

$\boldsymbol{\mu}_u$  é um vetor de dimensão  $K$  com as médias das respectivas localizações não observadas;  $\boldsymbol{\mu}$  é um vetor contendo as médias dos  $n$  pontos observados;  $\boldsymbol{\Sigma}_u$  é uma matriz de dimensão  $K$  e cada elemento representa a covariância entre os pontos não observados. Cada linha da matriz  $\boldsymbol{\Psi}$ ,  $n \times K$ , representa a covariância entre a  $i$ -ésima localização observada e a  $j$ -ésima não observada,  $i = 1, \dots, n$  e  $j = 1, \dots, k$ .  $\boldsymbol{\Sigma}$  é a matriz de covariância, de dimensão  $n \times n$ , entre as localizações observadas.

A distribuição *a posteriori*  $p(\boldsymbol{\theta} | \mathbf{Y})$  é analiticamente intratável. Entretanto, avanços recentes em técnicas de simulação tais como métodos Monte Carlo com Cadeias de Markov (Markov Chain Monte Carlo, MCMC) (GAMERMAN, LOPES, 2006) tornaram possível a simulação de qualquer distribuição *a posteriori* que seja conhecida a menos da constante normalizadora. Além disso, a distribuição *a posteriori* de qualquer função explícita dos parâmetros do modelo pode ser obtida. Sendo assim, métodos de Monte Carlo podem ser usados para se obter aproximações da integral da eq. (2.1) (BANERJEE ET AL., 2004). Para cada iteração  $t$ ,  $t = 1, \dots, T$ , obtém-se amostras *a posteriori*  $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(T)}$  da distribuição *a posteriori*  $p(\boldsymbol{\theta} | \mathbf{Y})$  e substituindo-se esses valores na eq. (2.2), amostras da distribuição  $p(\mathbf{Y}_u | \mathbf{Y})$  podem ser obtidas através da aproximação dada pela eq. (2.4)

$$p(\mathbf{Y}_u | \mathbf{Y}) \approx \frac{1}{T} \sum_{t=1}^T p(\mathbf{Y}_u | \boldsymbol{\theta}^{(t)}). \quad (2.4)$$



Pode-se utilizar a média amostral de  $\mathbf{Y}_u$  como uma estimativa para o valor esperado de  $\mathbf{Y}_u \mid \mathbf{Y}$ .

Em muitos casos, conjuntos de dados espacialmente referenciados apresentam medidas multivariadas obtidas ao longo de uma região geográfica. As observações multivariadas provenientes de  $n$  localizações geo-referenciadas podem ser tratadas como um processo estocástico parcialmente observado. Ou seja, considera-se que os dados consistem em vetores  $\mathbf{Y}(s_i)$  de dimensão  $p \times 1$ , observados em  $n$  localizações espaciais,  $s_i, i = 1, 2, \dots, n$ , de uma determinada região, a qual é denotada por  $G$ . Assim,  $\mathbf{Y} = (\mathbf{Y}(s_1), \mathbf{Y}(s_2), \dots, \mathbf{Y}(s_n))'$  forma uma amostra aleatória parcial do processo estocástico  $\{\mathbf{Y}(s) : s \in G\}$  em que  $G \in \mathbb{R}^g$  e  $\mathbf{Y}(s_i) = (Y_1(s_i), \dots, Y_p(s_i))'$ .

Um modelo para o processo  $\mathbf{Y}(\cdot)$  pode ser descrito pela soma de três componentes, isto é

$$\mathbf{Y}(s) = \mathbf{X}(s)\boldsymbol{\beta} + \mathbf{v}(s) + \boldsymbol{\epsilon}(s), \quad (2.5)$$

em que  $\mathbf{X}(\cdot)$  é uma matriz  $p \times pq$  contendo  $q$  possíveis covariáveis,  $\boldsymbol{\beta}$  é um vetor de dimensão  $pq$ ;  $\mathbf{v}(s)$  é um componente aleatório com uma matriz de covariância  $p \times p$  que descreve a estrutura de covariância existente entre: i)  $Y_j(s)$  e  $Y_{j'}(s)$ ; ii)  $Y_j(s)$  e  $Y_j(s')$  e iii)  $Y_j(s')$  e  $Y_j'(s)$ ;  $\boldsymbol{\epsilon}(s)$  é um vetor de ruído branco normalmente distribuído, com matriz de covariâncias,  $\mathbf{D}$ ,  $p \times p$ , diagonal tal que  $D_{jj} = \tau_j^2$ .  $\boldsymbol{\epsilon}(\cdot)$  descreve a variação de microescala.

### 2.2.2 Modelo de co-regionalização linear

Gelfand et al. (2004) propõem um modelo de co-regionalização linear para a componente  $\mathbf{v}(s)$ . O modelo proposto, de fácil interpretação e computacionalmente tratável, resulta em uma estrutura de covariância flexível, em que os alcances associados às componentes de  $\mathbf{v}(s)$  não são necessariamente os mesmos. Esses autores assumem que

$$\mathbf{v}(s) = \mathbf{A}\mathbf{w}(s),$$

em que:  $\mathbf{A}$  é  $p \times p$ , de posto completo, triangular inferior e as componentes de  $\mathbf{w}(s)$ ,  $w_j(s), j = 1, \dots, p$  são processos espaciais independentes, mas não identicamente distribuídos. Os  $w_j(s)$  são processos gaussianos com média 0, variância 1 e função de correlação estacionária  $\rho_j(h)$ .

Desta forma, a matriz de covariância cruzada associada a  $\mathbf{v}$  pode ser escrita como

$$\boldsymbol{\Sigma}_{\mathbf{v}(s), \mathbf{v}(s')} \equiv C(s - s') = \sum_{j=1}^2 \rho_j(s - s') \mathbf{T}_j, \quad (2.6)$$

em que  $\mathbf{T}_j = \mathbf{a}_j \mathbf{a}_j'$ , com  $\mathbf{a}_j$  sendo a  $j$ -ésima coluna de  $\mathbf{A}$ ;  $\mathbf{T}_j$  tem posto 1 e  $\sum_j \mathbf{T}_j = \mathbf{T}$ .

Se em (2.6) forem introduzidas funções de correlação, monótonas e isotrópicas, haverá um alcance associado a cada componente,  $v_j(s), j = 1, \dots, p$ . Tem-se que o alcance efetivo é o valor

de  $d$  para o qual o valor da correlação cai a valores pequenos, por exemplo, 0,05. Para  $p = 2$ , o alcance para  $v_1(s)$  resolve a equação

$$\rho_1(d) = 0,05, \quad (2.7)$$

enquanto que o alcance para  $v_2(s)$  soluciona a correlação ponderada dada por

$$\frac{a_{21}^2 \rho_1(d) + a_{22}^2 \rho_2(d)}{a_{21}^2 + a_{22}^2} = 0,05. \quad (2.8)$$

Para a obtenção da função de verossimilhança utilizando-se o modelo de co-regionalização linear, considera-se o modelo (2.5) como um modelo hierárquico. No primeiro estágio dados  $\{\beta_j, j = 1, 2, \dots, p\}$  e  $\{v(s_i)\}$  os  $Y(s_i), i = 1, 2, \dots, n$  são, condicionalmente, independentes com  $Y(s_i) | \beta, v(s_i) \sim N(\mathbf{X}(s_i)\beta + v(s_i), \mathbf{D})$ , para cada  $s_i$ . No segundo estágio a distribuição conjunta de  $\mathbf{v} = (v(s_1), v(s_2), \dots, v(s_n))$  é  $N(\mathbf{0}, \sum_{j=1}^p \mathbf{R}_j \otimes \mathbf{T}_j)$ , em que  $\mathbf{T}_j$  é uma matriz de dimensão  $p \times p$ ,  $\mathbf{R}_j$  tem dimensão  $n \times n$ , com  $(R_j)_{ii'} = \rho_j(s_i - s_{i'})$ . Concatenando os  $Y(s_i)$  num vetor  $\mathbf{Y}_{np \times 1}$  e, do mesmo modo, os elementos  $\mu(s_i) = \mathbf{X}(s_i)\beta$  em  $\mu$ , a densidade marginal de  $\mathbf{Y}$  é dada por

$$\mathbf{Y} \sim N(\mu_{np}, \sum_{j=1}^2 \rho_j(s - s')\mathbf{T}_j + \mathbf{I}_n \otimes \mathbf{D}). \quad (2.9)$$

O modelo de co-regionalização linear pode ser reparametrizado segundo uma abordagem condicional. Essa abordagem tem a vantagem computacional, uma vez que, ao invés de se ter uma matriz de covariâncias  $pn \times pn$ , tem-se  $p$  matrizes de covariâncias  $n \times n$ . Gelfand et al. (2004) mostram que as especificações condicional e incondicional do modelo (2.5) são equivalentes se todos os processos, exceto um, não tiverem componentes com ruído branco. Outra possível vantagem da especificação condicional é a interpretação do modelo. Se há uma ordem cronológica natural, ou algum tipo de causalidade nos eventos, então, isso pode determinar uma ordem natural para o condicionamento e, portanto, podem surgir especificações naturais. Na abordagem condicional a função de verossimilhança é dada pelo produto das verossimilhanças de cada variável envolvida no processo.

### 2.2.3 Metodologia

Considerando-se, aqui, dados bivariados compostos por esforço e captura e a temperatura da superfície do mar como covariável para as duas variáveis. Visto que existe uma ordem cronológica natural entre as variáveis esforço e captura, adotou-se a especificação condicional do modelo de co-regionalização linear. Assim, modela-se primeiro o logaritmo do esforço de pesca ( $Y_1$ ), e em seguida o logaritmo da captura dado o logaritmo do esforço ( $Y_2 | Y_1$ ), pois assume-se que os dados do esforço e da captura seguem distribuição normal após a transformação logarítmica:

$$\begin{aligned}
Y_1(s) &= \beta_{01} + \beta_{11}temp(s) + \sigma_1 w_1(s) \\
Y_2(s) | Y_1(s) &= \beta_{02} + \beta_{12}temp(s) + \alpha Y_1(s) + \sigma_2 w_2(s) + \tau_2 u_2(s),
\end{aligned} \tag{2.10}$$

em que  $temp(s)$  representa a temperatura na localização  $s$ ,  $w_1(s)$  e  $w_2(s)$  são processos espaciais gaussianos com média zero e variância 1, independentes, mas não identicamente distribuídos e  $u_2(s)$  tem distribuição  $N(0, 1)$ . A função de correlação utilizada foi a exponencial  $\rho(d, \phi)$ , em que  $d$  é a distância entre dois pontos  $s, s'$  quaisquer. Nela, controla-se através do parâmetro  $\phi$  o quão rapidamente a correlação decai para zero.

Ao usar a especificação condicional, tem-se a restrição de que o modelo para  $Y_1(s)$  não deve ter componente com ruído branco para que exista a equivalência. Por outro lado, o uso da abordagem condicional traz a vantagem computacional, uma vez que ao invés de se ter uma matriz de covariâncias  $2n \times 2n$ , tem-se duas matrizes de covariâncias  $n \times n$  (GELFAND ET AL., 2004).

Dada uma localização  $s$  qualquer, a correlação, entre  $Y_1(s)$  (logaritmo do esforço) e  $Y_2(s)$  (logaritmo da captura) é dada por  $\rho_{Y_1, Y_2} = \frac{T_{12}}{\sqrt{T_{11}T_{22}}}$ , em que  $T_{i,j}$  são os elementos da matriz  $\mathbf{T}$  dada na eq. (2.6). Usando a reparametrização da abordagem condicional, obtém-se

$$\rho_{Y_1, Y_2} = \frac{\alpha \sigma_1^2}{\sqrt{\sigma_1^2(\alpha^2 \sigma_1^2 + \sigma_2^2)}}. \tag{2.11}$$

### 2.2.3.1 Distribuição *a priori* para os parâmetros

Ao se utilizar a abordagem bayesiana, há necessidade de especificar a distribuição *a priori* para o vetor  $\theta$  de parâmetros do modelo. Considerando-se independência entre os parâmetros, a *priori* para  $\theta$  é dada pelo produto das *prioris* individuais.

Como é usual na literatura, aos parâmetros  $\sigma_1^2, \sigma_2^2$  e  $\tau_2^2$ , atribuiu-se densidades *a priori* gama invertida, isto é,  $\sigma_1^2 \sim IG(a_{\sigma_1}, b_{\sigma_1}), \sigma_2^2 \sim IG(a_{\sigma_2}, b_{\sigma_2})$  e  $\tau_2^2 \sim IG(a_{\tau_2}, b_{\tau_2})$ . Os hiperparâmetros foram obtidos usando a estratégia adotada por Gelfand et al. (2004), que consiste em: fixar os valores esperados dos parâmetros nas estimativas de mínimos quadrados dos dados observados, obtidas sob a hipótese de independência espacial; e considerar variância infinita *a priori* para esses parâmetros. Para os parâmetros  $\phi_1$  e  $\phi_2$ , da função de correlação exponencial, para as variáveis esforço e captura, assumiu-se a mesma densidade *a priori* gama,  $\phi_1 \sim G(a_{\phi_1}, b_{\phi_1}), \phi_2 \sim G(a_{\phi_2}, b_{\phi_2})$ , e os hiperparâmetros foram obtidos resolvendo-se as equações  $\mathbf{E}(\phi_j) = \frac{6}{max.dist} = 0,05$  e  $var(\phi_j) = 20, j = 1, 2$  (variância grande), em que  $max.dist$  é a máxima distância entre as localizações observadas, cujo valor é aproximadamente 12. Essa *priori* reflete o fato de se esperar que para distâncias maiores do que  $d_0 = \frac{max.dist}{2}$  a correlação espacial é aproximadamente zero (SCHMIDT; GELFAND, 2003; BANERJEE ET AL., 2004; PAEZ ET AL., 2005). Atribuiu-se aos

vetores de parâmetros  $\beta_1$  e  $\beta_2$  a densidade *a priori* normal com matriz de covariâncias  $\sigma^2 I$ , sendo que o valor de  $\sigma^2_{\beta}$  foi fixado em 100 (*priori vaga*). A parametrização das distribuições *a priori* estão de acordo com Gelman et al. (2004).

### 2.2.3.2 Função de verossimilhança

Considerando-se o modelo dado pelas eq. (2.10) a função de verossimilhança pode ser escrita como o produto da verossimilhança para  $Y_1$  e da verossimilhança para  $Y_2$ . A verossimilhança para  $Y_1$  é dada por

$$L(\beta_1, \phi, \sigma_1^2 | \mathbf{y}_1) = (2\pi)^{-\frac{n}{2}} |\sigma_1^2 \mathbf{R}_1|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_1 - \mathbf{X}_1 \beta_1]^T [\sigma_1^2 \mathbf{R}_1]^{-1} [\mathbf{y}_1 - \mathbf{X}_1 \beta_1] \right\} \quad (2.12)$$

em que  $\mathbf{X}_1$  é a matriz de covariáveis para  $Y_1(s)$  e  $\beta_1$  é o vetor de coeficientes das covariáveis.

A função de verossimilhança para  $Y_2$  é dada por

$$L(\beta_2, \phi, \sigma_2^2, \tau_2^2 | \mathbf{y}_2, \mathbf{y}_1) = (2\pi)^{-\frac{n}{2}} |\sigma_2^2 \mathbf{R}_2 + \tau_2^2 I|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_2 - \mathbf{X}_2 \beta_2]^T [\sigma_2^2 \mathbf{R}_2 + \tau_2^2 I]^{-1} [\mathbf{y}_2 - \mathbf{X}_2 \beta_2] \right\} \quad (2.13)$$

em que  $\mathbf{X}_2$  é a matriz de covariáveis para  $Y_2$  e  $\beta_2$  é o vetor de coeficientes das covariáveis.

### 2.2.3.3 Distribuições *a posteriori* para os parâmetros

No contexto bayesiano, as inferências são feitas através da distribuição *a posteriori* dos parâmetros. Devido a independência das verossimilhanças dadas nas equações (2.12) e (2.13) e dos parâmetros *a priori*, a distribuição *a posteriori* conjunta para os parâmetros do modelo (2.10) é dada pelo produto das distribuições

$$\pi(\beta_1, \phi_1, \sigma_1^2 | \mathbf{y}_1) \propto L(\beta_1, \phi_1, \sigma_1^2 | \mathbf{y}_1) \pi(\beta_1) \pi(\phi_1) \pi(\sigma_1^2) \quad (2.14)$$

e

$$\pi(\beta_2, \phi_2, \sigma_2^2, \tau_2^2 | \mathbf{y}_2, \mathbf{y}_1) \propto L(\beta_2, \phi_2, \sigma_2^2, \tau_2^2 | \mathbf{y}_2, \mathbf{y}_1) \pi(\beta_2) \pi(\phi_2) \pi(\sigma_2^2) \pi(\tau_2^2), \quad (2.15)$$

sendo que as funções de verossimilhanças são dadas pelas equações (2.12) e (2.13) e as distribuições *a priori* especificadas anteriormente.

Como as distribuições *a posteriori* (2.14) e (2.15) não têm forma analítica fechada, foram utilizados métodos MCMC (GAMERMAN, LOPES, 2006) para a obtenção de uma amostra da distribuição *a posteriori* conjunta dos parâmetros. Mais especificamente, nestes casos pode-se utilizar o amostrador de Gibbs com passos de Metropolis-Hastings, uma vez que alguns parâmetros possuem distribuição condicional completa *a posteriori* conhecida e outros não. A seguir, são apresentadas as distribuições condicionais completas *a posteriori* para os parâmetros do modelo (2.10), necessárias para a implementação computacional dos algoritmos MCMC.

1.  $\beta_1 \mid \sigma_1^2, \phi_1, \mathbf{y}_1 \sim N(\mathbf{B}\mathbf{b}, \mathbf{B})$ ,  
em que  $\mathbf{B} = [\mathbf{X}'_1(\sigma_1^2\mathbf{R}_1)^{-1}\mathbf{X}_1 + \Sigma_{\beta_1}^{-1}]^{-1}$  e  $\mathbf{b} = \mathbf{X}'_1(\sigma_1^2\mathbf{R}_1)^{-1}\mathbf{y}_1 + \Sigma_{\beta_1}^{-1}\boldsymbol{\mu}_{\beta_1}$ ;
2.  $\sigma_1^2 \mid \beta_1, \phi_1, \mathbf{y}_1 \sim IG(\frac{n}{2} + a_{\sigma_1}, c_{\sigma_1})$ ,  
em que  $c_{\sigma_1} = \frac{1}{2}[\mathbf{y}_1 - \mathbf{X}_1\beta_1]^T R_1^{-1}[\mathbf{y}_1 - \mathbf{X}_1\beta_1]$ ;
3.  $\pi(\phi_1 \mid \beta_1, \sigma_1^2, \mathbf{y}_1) \propto |\mathbf{R}_1|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_1 - \mathbf{X}_1\beta_1]^T [\sigma_1^2\mathbf{R}_1]^{-1}[\mathbf{y}_1 - \mathbf{X}_1\beta_1] - b_{\phi_1}\phi_1\right\} \phi_1^{a_{\phi_1}-1}$ ;
4.  $\beta_2 \mid \sigma_2^2, \alpha, \tau_2^2, \phi_2, \mathbf{y}_1, \mathbf{y}_2 \sim N(\mathbf{B}_2\mathbf{b}_2, \mathbf{B}_2)$ ,  
em que  $\mathbf{B}_2 = [\mathbf{X}'_2(\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I})^{-1}\mathbf{X}_2 + \Sigma_{\beta_2}^{-1}]^{-1}$  e  $\mathbf{b}_2 = \mathbf{X}'_2(\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I})^{-1}\mathbf{y}_2 + \Sigma_{\beta_2}^{-1}\boldsymbol{\mu}_{\beta_2}$ ;
5.  $\pi(\phi_2 \mid \beta_2, \sigma_2^2, \tau_2^2, \mathbf{y}_1, \mathbf{y}_2) \propto$   
 $|\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\beta_2]^T [\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\beta_2] - b_{\phi_2}\phi_2\right\} \phi_2^{a_{\phi_2}-1}$ ;
6.  $\pi(\sigma_2^2 \mid \beta_{02}, \tau_2^2, \phi_2, \mathbf{y}_1, \mathbf{y}_2) \propto$   
 $|\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\beta_2]^T [\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\beta_2] - \frac{b_{\sigma_2}}{\sigma_2^2}\right\} (\sigma_2^2)^{-a_{\sigma_2}-1}$ ;
7.  $\pi(\tau_2^2 \mid \beta_2, \sigma_2^2, \phi_2, \mathbf{y}_1, \mathbf{y}_2) \propto$   
 $|\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\beta_2]^T [\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\beta_2] - \frac{b_{\tau_2}}{\tau_2^2}\right\} (\tau_2^2)^{-a_{\tau_2}-1}$ .

Para se obter amostra da distribuição *a posteriori* conjunta dos parâmetros utilizou-se o programa computacional *WinBugs* (SPIEGELHATER ET AL., 2002). Foram utilizadas cadeias de tamanho 55000 em que as primeiras 5000 foram descartadas (*burn-in*). A partir daí armazenou-se as observações de 50 em 50 (*thinning*) para minimizar problemas de autocorrelação, obtendo-se uma amostra de tamanho 1000. A verificação da convergência das cadeias foi realizada pela análise gráfica do seu traço.

### 2.2.3.4 Distribuição preditiva

A partir da distribuição *a posteriori* dos parâmetros  $\boldsymbol{\theta}$  e de propriedades da distribuição normal multivariada, pode-se obter a distribuição preditiva de captura e esforço para localizações não observadas (eq. 2.1). Fez-se, assim, a interpolação da captura e esforço para as localizações não observadas, cujos valores são dados pela média obtida de amostras da distribuição preditiva dadas pela equação (2.4).

### 2.2.4 Estudo de Simulação

Para simular dados bivariados,  $\mathbf{Y} = (Y_1, Y_2)$ , com a finalidade de avaliar os métodos de estimação propostos e comparar em cada método os índices CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub>, foi realizado um estudo de simulação. Para tanto, foi criada uma grade regular com 100 pontos (10 × 10) para representar uma região (hipotética) do oceano. Nesses pontos, foram simulados valores para temperatura da superfície oceânica, que foi usada como covariável, considerando que existe um gradiente na direção norte-sul. Usando o Atlântico Sul como referência, considerou-se um gradiente em que os valores de temperatura aumentam conforme o deslocamento no sentido de sul para norte (LIMA; LIN; MENEZES, 2000; GOODYEAR, 2003). Dessa forma, foram simulados dados de temperatura variando entre 10 e 25 °C.

Aos pontos da grade regular foram associados valores do logaritmo do esforço ( $Y_1$ ) e do logaritmo da captura dado o logaritmo do esforço ( $Y_2 | Y_1$ ), esses valores foram simulados diretamente do modelo dado pela eq. (2.10). Quando são simulados dados do modelo (2.10), obtém-se dados de captura e esforço transformados pela função logarítmica, uma vez que é assumido neste trabalho que os dados do esforço e de captura seguem distribuição normal após a transformação logarítmica. Assim para se obter dados simulados de esforço e captura basta aplicar a função exponencial nos dados simulados obtidos.

Os dados do logaritmo do esforço e do logaritmo da captura foram simulados considerando os seguintes cenários:

a) baixa correlação entre o logaritmo da captura e logaritmo do esforço, isto é,  $\rho_{Y_1, Y_2} = 0,3$ , retratando a alta incerteza do pescador, e considerando que os dados apresentam pouca correlação espacial  $\phi_1 = 6$ ,  $\phi_2 = 5,5$ , o que leva a um alcance efetivo de 0,5 unidades de distância para o esforço (valor obtido resolvendo a equação (2.7)) e 0,54 para a captura (obtido resolvendo-se a equação (2.8) pelo método de Newton). Para se obter  $\rho_{Y_1, Y_2} = 0,3$  (eq. (2.11)), foram fixados  $\alpha = 0,28$ ,  $\sigma_1^2 = 2,3$  e  $\sigma_2^2 = 1,8$ . Os demais parâmetros usados na simulação foram  $\tau_2^2 = 0,1$ ,  $\beta_{01} = 2$ ,  $\beta_{11} = 0,25$ ,  $\beta_{02} = 0$ ,  $\beta_{12} = 0,35$ .

b) Correlação entre captura e esforço de  $\rho_{Y_1, Y_2} = 0,8$ , retratando uma menor incerteza do pescador, e considerando que os dados apresentam pouca correlação espacial  $\phi_1 = 6$ ,  $\phi_2 = 5,5$ , o que leva a um alcance efetivo de 0,5 unidades de distância para o esforço e 0,52 para a captura. Para se obter  $\rho_{Y_1, Y_2} = 0,8$ , foram fixados  $\alpha = 1,2$ ,  $\sigma_1^2 = 2,3$  e  $\sigma_2^2 = 1,8$ . Os demais parâmetros usados na simulação foram os mesmos do cenário a).

c) Baixa correlação entre captura e esforço  $\rho_{Y_1, Y_2} = 0,3$  e considerando que os dados apresentam forte correlação espacial  $\phi_1 = 0,45$  (para esforço) e  $\phi_2 = 0,5$  (para captura), o que leva a um alcance efetivo de 6,67 unidades de distância para o esforço e 6,05 para a captura. Os demais parâmetros usados na simulação foram  $\sigma_1^2 = 2,3$ ,  $\sigma_2^2 = 1,8$ ,  $\tau_2^2 = 0,1$ ,  $\alpha = 0,28$ ,  $\beta_{01} = 2$ ,  $\beta_{11} = 0,25$ ,  $\beta_{02} = 0$ ,  $\beta_{12} = 0,15$ .

d) Correlação entre captura e esforço de  $\rho_{Y_1, Y_2} = 0,8$ , e considerando que os dados apresentam forte correlação espacial  $\phi_1 = 0,45$  (para esforço) e  $\phi_2 = 0,5$  (para captura), o que leva a um alcance efetivo de 6,67 unidades de distância para o esforço e 6,43 para a captura. O parâmetro  $\alpha$  foi fixado em 1,2 e os demais parâmetros foram os mesmos do cenário c).

Para a simulação dos dados utilizou-se o pacote *geoR* (RIBEIRO JR.; DIGGLE, 2001) do ambiente estatístico R (R Development Core Team, 2007).

Na Figura 2.4 estão ilustradas as relações entre captura e esforço para cada um dos cenários em que foram simulados os dados.

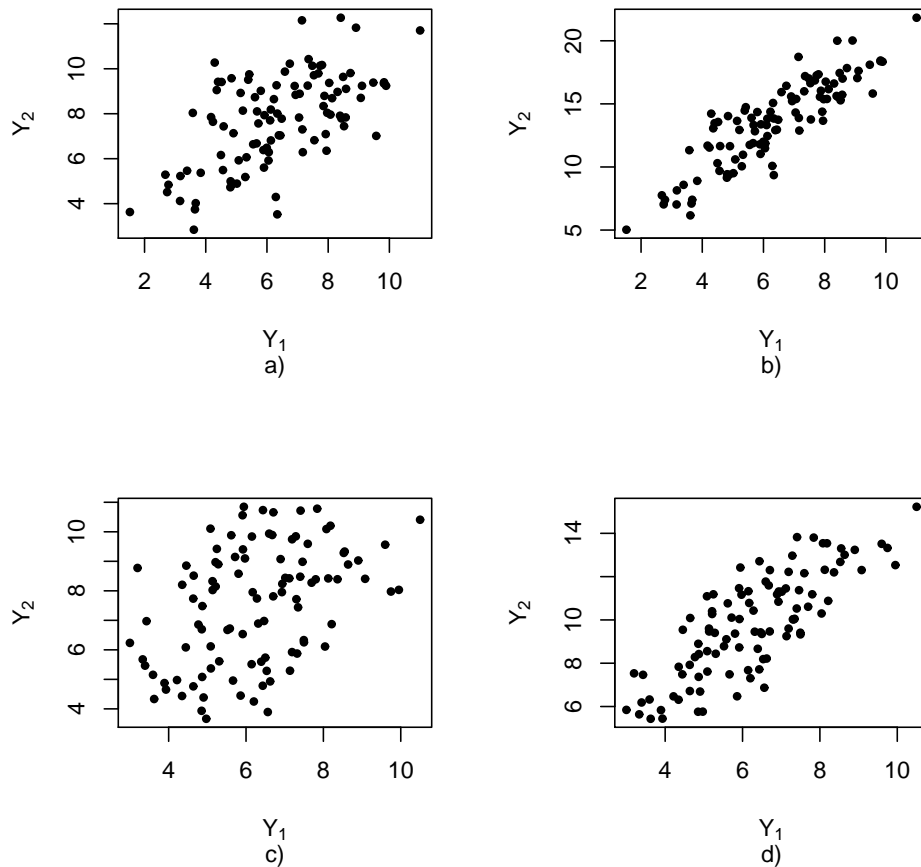


Figura 2.4 – Relação entre captura e esforço para uma realização dos cenários: a), b), c) e d)

Em seguida, foi gerado dentro da região em estudo (grade de 100 pontos) um processo de Thomas (REIS, 1998), resultando em um padrão espacial agregado, para representar os locais de ocorrência de pesca. Para tanto, usou-se a função *rThomas()* do pacote *spatstat* (BADDELEY; TURNER, 2005) do ambiente estatístico R (R Development Core Team, 2007). O processo de Thomas consiste em gerar um processo pontual de Poisson uniforme, cujos pontos gerados são

chamados de pais. Em seguida, cada ponto pai é substituído por um conglomerado de pontos (filhos). O número de pontos por conglomerado segue um processo Poisson com uma determinada média, e a posição de cada ponto filho é determinada pela adição, à localização do evento pai, de um vetor normal bivariado composto de duas variáveis independentes com média zero e uma determinada variância. Os pontos gerados por esse processo foram aproximados aos pontos da grade ( $10 \times 10$ ), obtendo-se um total de 85 pontos (FIGURA 2.5). Essas 85 localizações e os valores de temperatura, captura e esforço associados a elas foram considerados como dados observados, os quais foram usados para ajuste do modelo.

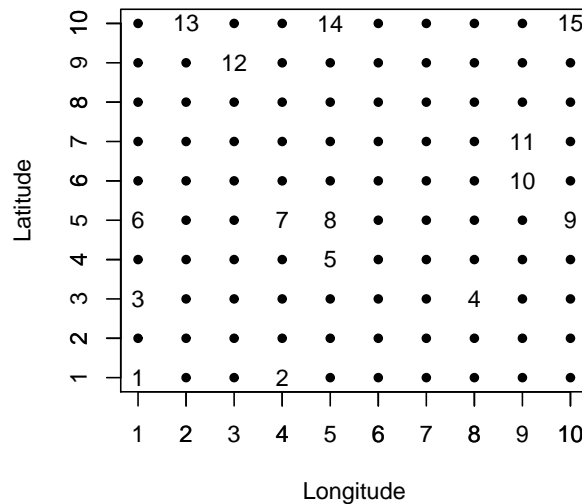


Figura 2.5 – Localizações observadas "•" e localizações não observadas enumeradas

Para cada um dos cenários a), b), c) e d), foram simulados 50 conjuntos de dados. Para cada um desses conjuntos considerou-se, como descrito anteriormente, 85 localizações da região como sendo localizações observadas (*localizações medidas* como são chamadas na literatura geostatística), restando 15 localizações que foram consideradas como não observadas (*não medidas*). Para cada um dos conjuntos de dados foi ajustado o modelo proposto dado pela eq. (2.10). No ANEXO A é apresentado o programa que foi utilizado para gerar os dados no cenário d), e para rodar o programa *WinBUGS 1.4.3* a partir do programa *R 2.5.1*, para a obtenção das amostras da distribuição *a posteriori* conjunta dos parâmetros do modelo (2.10). No ANEXO B é apresentado o código em *WinBUGS 1.4.3* para a obtenção das amostras das distribuições *a posteriori* dos parâmetros do modelo (2.10).

Foi ajustado, também, um modelo em que o esforço é considerado conhecido. Dessa forma,



ajusta-se apenas o modelo para a captura ( $Y_2$ ), dado por

$$Y_2(s) = \beta_0 + \beta_1 temp(s) + \alpha_0 y_1(s) + \sigma w(s) + \tau u(s), \quad (2.16)$$

em que  $temp(s)$  representa a temperatura na localização  $s$ ,  $w(s)$  é um processo gaussiano com média zero e variância 1 e função de correlação exponencial e  $u(s)$  tem distribuição  $N(0, 1)$ .

Ajustou-se, ainda, um modelo que considera que o esforço e a captura são correlacionados entre si, mas ambos independentes no espaço. O modelo é descrito pelas equações (2.17). Este modelo foi ajustado para verificar o efeito causado na estimação dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , quando não se considera a correlação espacial.

$$\begin{aligned} Y_1(s) &= \beta'_{01} + \beta'_{11} temp(s) + \sigma'_1 u_1(s) \\ Y_2(s) | Y_1(s) &= \beta'_{02} + \beta'_{12} temp(s) + \alpha' Y_1(s) + \tau'_2 u_2(s), \end{aligned} \quad (2.17)$$

em que  $u_1(s)$  e  $u_2(s)$  são independentes e seguem distribuição  $N(0, 1)$ . O modelo (2.17) é um caso particular do modelo (2.10), ou seja, tomando a função de correlação  $\rho(s - s') = \begin{cases} 0 & \text{se } s \neq s' \\ 1 & \text{se } s = s' \end{cases}$  para o processo  $Y_1(s)$  no modelo (2.10) e fazendo  $\sigma_2 = 0$ , obtém-se o modelo (2.17). Já foi observado que existe uma equivalência, através de uma reparametrização, entre a especificação condicional dada pela eq. (2.10) e sua especificação não condicional na forma da eq. (2.5). Como o modelo (2.17) é um caso particular do modelo (2.10), nota-se que há uma especificação não condicional equivalente para o modelo (2.17). Isto é, o modelo (2.17) pode ser escrito como

$$\begin{aligned} Y_1(s) &= \beta'_{01} + \beta'_{11} temp(s) + v_1(s) \\ Y_2(s) &= \tilde{\beta}'_{02} + \tilde{\beta}'_{12} temp(s) + v_2(s) + \epsilon_2(s), \end{aligned}$$

em que  $\tilde{\beta}'_{02} = \beta'_{02} + \alpha' \beta'_{01}$ ,  $\tilde{\beta}'_{12} = \beta'_{12} + \alpha' \beta'_{11}$ ,  $v_1(s) = \sigma'_1 u_1(s)$ ,  $v_2(s) = \alpha' \sigma'_1 u_1(s)$  e  $\epsilon_2(s) = \tau'_2 u_2(s)$ . Nota-se que, para o modelo (2.17), a covariância entre duas localizações distintas  $s$  e  $s'$  é nula, isto é,  $cov(\mathbf{Y}(s), \mathbf{Y}(s')) = \mathbf{0}_{2 \times 2}$ ,  $\mathbf{Y}(s) = (Y_1(s), Y_2(s))'$ , enquanto que  $cov(\mathbf{Y}(s), \mathbf{Y}(s)) = \begin{pmatrix} \sigma_1'^2 & \alpha' \sigma_1'^2 \\ \alpha' \sigma_1'^2 & \alpha'^2 \sigma_1'^2 + \tau_2'^2 \end{pmatrix}$ .

Após o ajuste de cada modelo, obteve-se, da distribuição preditiva, os valores de captura e esforço para os locais não observados. Foram, então, calculadas as  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  para cada conjunto de dados simulados, por dois métodos, denominados amostral e ajustado:

1. usando apenas o conjunto de dados observados nas 85 localizações. As estimativas obtidas por esse método foram chamadas de estimativas amostrais e representadas por  $CPUE_{1A}$ ,  $CPUE_{2A}$ ,  $CPUE_{3A}$ .

Para se obter uma estimativa amostral de cada índice, substituiu-se os dados de captura e esforço provenientes de uma amostra na fórmula de cada índice. Assim,  $CPUE_{1A} = \frac{1}{m} \sum_{i=1}^m \frac{C_i}{E_i}$ , em que  $m$  é o tamanho da amostra, é um preditor para o índice  $CPUE_1$ .

2. Após fazer a interpolação, usando o conjunto de dados composto pelos valores observados e pelos valores preditos nos locais não observados. Na abordagem bayesiana, tem-se uma amostra *a posteriori* para os índices, ou seja, para cada iteração do método MCMC, obtém-se uma estimativa para cada um dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ . Para comparar as estimativas obtidas por esse método com as estimativas amostrais, considerou-se uma estimativa pontual, dada pela mediana da amostra *a posteriori*. As estimativas obtidas por esse método foram chamadas de estimativas ajustadas e representadas por  $CPUE_{1B}$ ,  $CPUE_{2B}$ ,  $CPUE_{3B}$ .

Para se obter uma estimativa ajustada de cada índice, formou-se um conjunto de dados composto por  $m$  pares de captura e esforço, provenientes de uma amostra observada e por  $K$  pares de captura e esforço preditos (para os locais não observados) pelo modelo que foi ajustado. Obteve-se, dessa maneira, um conjunto de  $m + K = n$  pares de captura e esforço. Substituindo-se esse conjunto na fórmula de cada índice, obtém-se uma estimativa ajustada para cada índice. Assim, por exemplo,  $CPUE_{1B} = \frac{1}{m + K} \sum_{i=1}^{m+K} \frac{C_i}{E_i}$ , é um preditor para o índice  $CPUE_1$ .

Foram obtidos também, para cada conjunto de dados simulados, os valores verdadeiros dos índices ( $CPUE_{1V}$ ,  $CPUE_{2V}$ ,  $CPUE_{3V}$ ). Para se obter esses valores, simplesmente substituiu-se o conjunto de dados simulados para toda região ( $10 \times 10$ ) nas fórmulas de cada índice.

A medida usada para comparar os métodos de estimação dos índices foi o erro quadrático médio (EQM). Denotando por  $CPUE_{1V_l}$ ,  $CPUE_{2V_l}$ ,  $CPUE_{3V_l}$  os valores verdadeiros das  $CPUE_1$ ,  $CPUE_2$ ,  $CPUE_3$  do  $l$ -ésimo conjunto de dados simulados,  $CPUE_{1A_l}$ ,  $CPUE_{2A_l}$ ,  $CPUE_{3A_l}$  as estimativas amostrais e  $CPUE_{1B_l}$ ,  $CPUE_{2B_l}$ ,  $CPUE_{3B_l}$  as estimativas ajustadas,  $l = 1, \dots, 50$ , os EQM's são dados por:

$$EQM_A = \frac{1}{50} \sum_{l=1}^{50} (CPUE_{jA_l} - CPUE_{jV_l})^2, \quad j = 1, 2, 3, \text{ para as estimativas amostrais};$$

$$EQM_B = \frac{1}{50} \sum_{l=1}^{50} (CPUE_{jB_l} - CPUE_{jV_l})^2, \quad j = 1, 2, 3, \text{ para as estimativas ajustadas}.$$

O método que apresenta os menores EQM's é considerado melhor.

No Anexo C é apresentado o programa utilizado para fazer as predições de esforço e captura nos locais não observados, para obter as estimativas amostrais e ajustadas dos índices e para calcular os EQM's.

A fim de comparar os três índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  entre si, a medida utilizada foi a soma dos desvios absolutos relativos (SDAR) dada por

$$SDAR = \sum_{l=1}^{50} \frac{|CPUE_{est_l} - CPUE_{V_l}|}{CPUE_{V_l}} \quad (2.18)$$

em que  $CPUE_{est_l}$  representa a estimativa de um índice na simulação  $l$  e  $CPUE_{V_l}$ , o valor verdadeiro do índice na simulação  $l$ . Cada parcela da soma dada na eq. (2.18) representa o desvio da estimativa em relação ao valor verdadeiro do índice que está sendo estimado. Assim, esse critério não é afetado pela possível diferença de escala entre os índices. Segundo esse critério, quanto menor for o valor de SDAR mais preciso é o índice.

### 2.2.5 Resultados e discussão

Nesta seção, são apresentados os resultados referentes ao estudo de simulação realizado neste trabalho. Porém, antes de tirar conclusões sobre os resultados obtidos, é necessário verificar a convergência das realizações das distribuições *a posteriori* obtidas através de métodos MCMC.

Na Figura 2.6 apresentam-se os traços de duas cadeias, referentes à distribuição *a posteriori* dos parâmetros do modelo (2.10), geradas a partir de valores iniciais distintos, para um dos cinquenta conjuntos de dados simulados sob o cenário d). Verifica-se através dos gráficos que há indícios de convergência.

A Figura 2.7 ilustra a capacidade do processo de estimação em localizar os parâmetros do modelo (2.10) usados na simulação dos dados. A ilustração refere-se a uma simulação sob o cenário d). Observa-se que os histogramas das amostras *a posteriori*, para todos os parâmetros, contêm os valores verdadeiros, representados pelas retas em vermelho. Para os modelos (2.16) e (2.17), não é possível verificar a capacidade do processo de estimação, visto que os dados não foram gerados por esses modelos.

No Apêndice A, são apresentadas as funções de autocorrelação das cadeias *a posteriori* para os parâmetros do modelo (2.10), ajustados para dados simulados no cenário d). São apresentados, também, os intervalos de credibilidade de 95% para todos os parâmetros para cada simulação do cenário d). Observa-se que praticamente todos os intervalos cobrem o verdadeiro valor do parâmetro.

No Apêndice B, são apresentados os traços de duas cadeias *a posteriori* geradas a partir de valores iniciais distintos para os parâmetros do modelo (2.10), para os cenários a) e c). Os gráficos apresentados permitem verificar que existem indícios de convergência das cadeias.

A Figura 2.8 mostra um resumo das previsões feitas para o logaritmo do esforço e o logaritmo da captura, usando o modelo (2.10) para a localização designada pelo número 11 na Figura 2.5. Já a Figura 2.9 mostra o resumo das previsões para a localização designada pelo número 15 na Figura 2.5. Como esperado, as previsões feitas para os locais não observados da região em estudo tendem a ser melhores para os pontos que não estão nas fronteiras da região. Os pontos de fronteira têm menos pontos vizinhos para fornecer informação sobre os níveis das variáveis, e os pontos que não se localizam nas fronteiras têm mais vizinhos e, conseqüentemente, tendem a ser melhor estimados. Os pontos 15 e 11 da Figura 2.5 foram tomados como exemplo para ilustrar esse fato. O

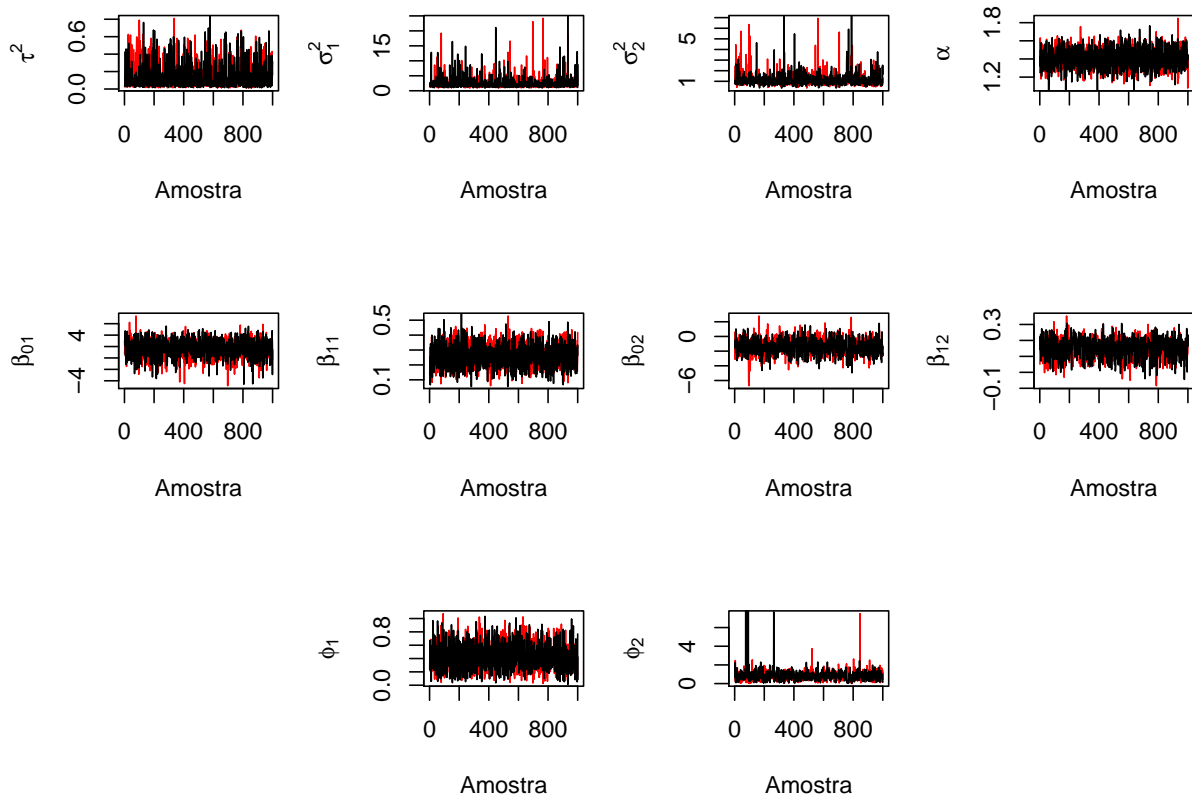


Figura 2.6 – Traço de duas cadeias *a posteriori* dos parâmetros sob o cenário d), modelo (2.10), representadas pelas cores vermelha e preta, partindo de valores iniciais distintos

ponto 15 é um ponto de fronteira, onde se tem poucos vizinhos próximos para fornecer informações. Consequentemente as estimativas do esforço e da captura nesse ponto são piores do que no ponto 11, como pode ser visto na Tabela 2.1 e nas Figuras 2.8 e 2.9.

A Figura 2.10 apresenta os gráficos dos valores dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , para os 50 conjuntos de dados gerados sob o cenário d), versus as estimativas amostrais e as medianas das estimativas ajustadas considerando o modelo (2.10). Nessa figura, quanto mais próximos da reta estiverem os pontos, melhor será o método de estimação. Talvez, pelo fato dos índices abrangerem uma grande faixa de valores, não seja possível distinguir visualmente qual o melhor método (amostral ou ajustado). Porém, os erros quadráticos médios que serão mostrados a seguir permitem tal comparação.

Na Figura 2.11 são apresentados individualmente alguns histogramas das estimativas ajustadas, considerando-se o modelo (2.10) sob o cenário d), juntamente com as respectivas medianas e também as estimativas amostrais. Essa figura ilustra o fato de que, para o método de estimação proposto

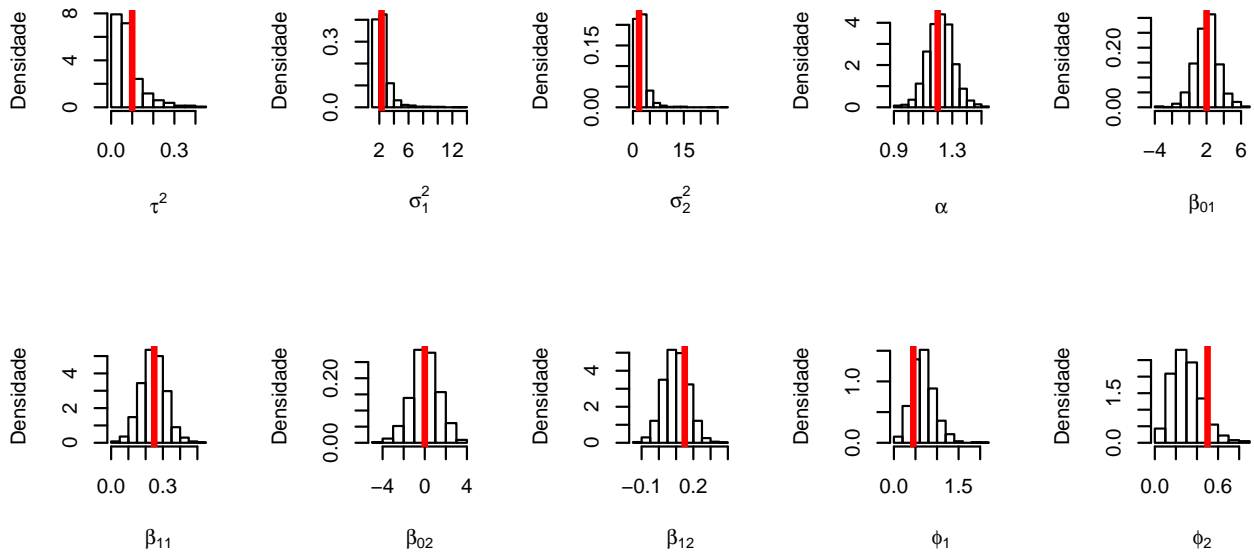
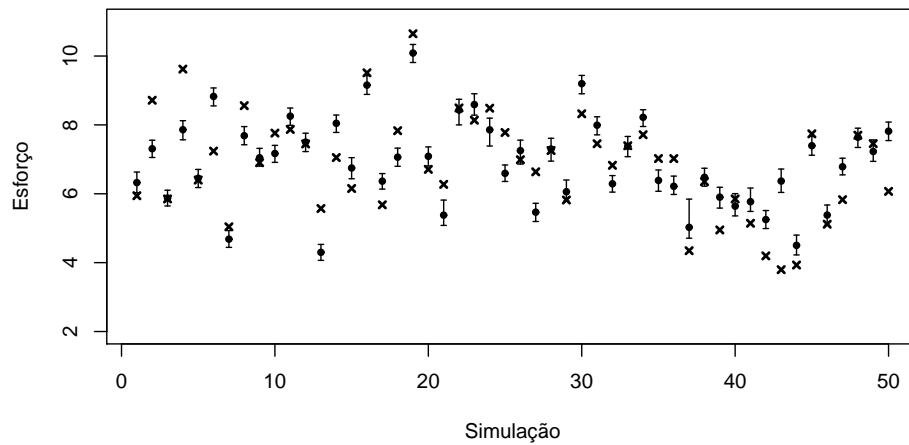


Figura 2.7 – Histogramas das amostras *a posteriori* dos parâmetros e valores usados na simulação dos dados (reta em vermelho), sob o cenário d), modelo (2.10)

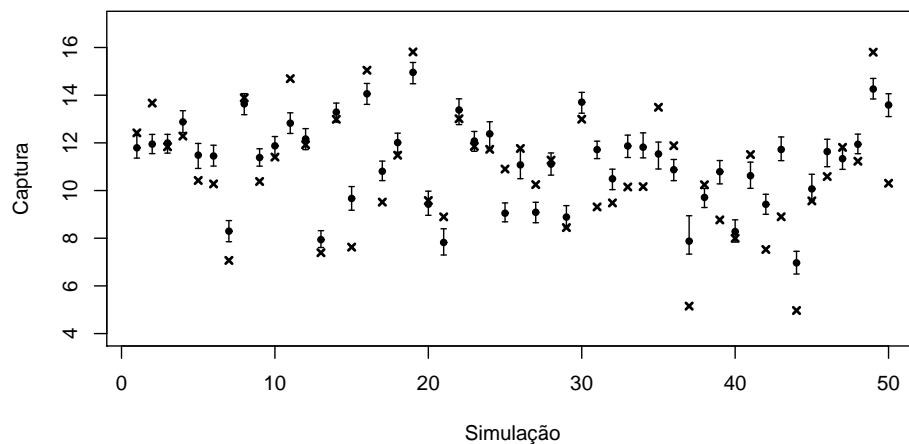
Tabela 2.1 – Erro quadrático médio para as previsões do logaritmo do esforço e da captura, cenário d), nos locais não observados

| Local | Esforço | Captura |
|-------|---------|---------|
| 1     | 1,3398  | 2,2419  |
| 2     | 0,9833  | 2,1748  |
| 3     | 0,9380  | 2,2707  |
| 4     | 0,8362  | 2,2656  |
| 5     | 0,8544  | 1,8020  |
| 6     | 0,9448  | 2,3737  |
| 7     | 0,9126  | 2,0383  |
| 8     | 0,7509  | 2,4897  |
| 9     | 0,8976  | 2,3796  |
| 10    | 0,7289  | 1,4818  |
| 11    | 0,6981  | 1,7847  |
| 12    | 0,9813  | 2,7792  |
| 13    | 0,8487  | 2,4491  |
| 14    | 0,8434  | 1,9880  |
| 15    | 1,5840  | 3,0321  |

neste trabalho (método ajustado), tem-se uma amostra de estimativas para cada um dos índices, ou seja, através dos histogramas tem-se uma descrição da incerteza associada. Além disso, resumos de interesse como, por exemplo, intervalos de credibilidade, podem ser obtidos.



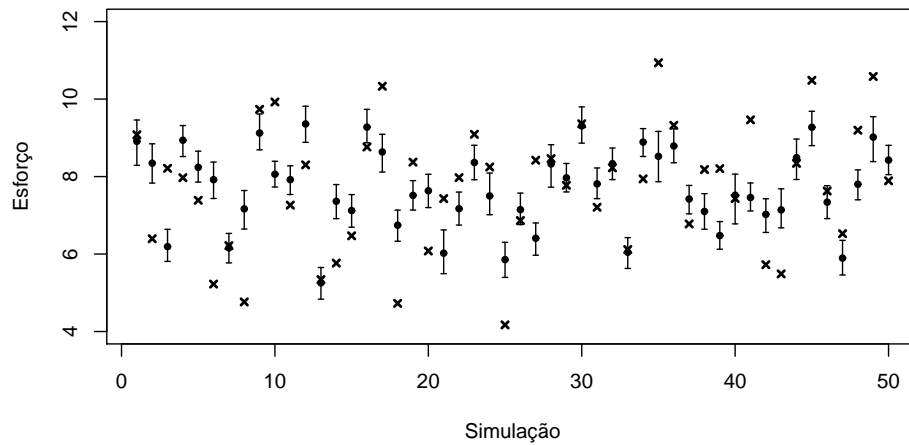
(a)



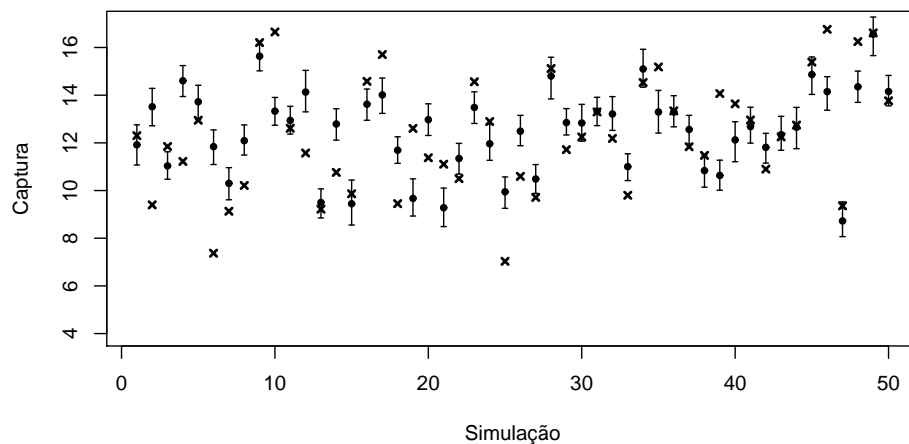
(b)

Figura 2.8 – Intervalo 95% de predição dos valores do logaritmo do esforço (a) e do logaritmo da captura (b) juntamente com a mediana dos valores preditos “•” e os valores verdadeiros “×” da localização de número 11 (Figura 2.5) para cada amostra. Cenário d), modelo (2.10)

Olhando para os histogramas das estimativas ajustadas dos índices, observa-se que suas amplitudes são relativamente pequenas. Ora, todas as 1000 estimativas que compõem cada histograma são calculadas pelos conjuntos de dados compostos por 85 pares de esforço e captura observados e 15 pares preditos. Os 85 pares de esforço e captura observados são comuns a todas as estimativas. Somente os 15 pares mudam de uma estimativa para outra (a cada iteração do MCMC). Ou seja, todas as estimativas que compõem cada histograma têm uma parte em comum. Isso faz com que



(a)



(b)

Figura 2.9 – Intervalo 95% de predição dos valores do logaritmo do esforço (a) e do logaritmo da captura (b) juntamente com a mediana dos valores preditos “•” e os valores verdadeiros “×” da localização de número 15 (Figura 2.5) para cada simulação. Cenário d), modelo (2.10)

as estimativas não se distanciem muito umas das outras, resultando em histogramas com amplitude pequena.

Para as estimativas amostrais, também é possível obter uma medida de incerteza através dos intervalos de confiança. No entanto, os intervalos de confiança para as estimativas amostrais não podem ser comparados com intervalos de credibilidade, que podem ser obtidos para estimativas ajustadas; ambos têm interpretações distintas. No Apêndice C, são apresentadas as medianas das

estimativas ajustadas usando o modelo (2.10) e as estimativas amostrais, juntamente com os intervalos de confiança *bootstrap* 95% para as estimativas amostrais.

Entre as diferentes simulações, observa-se (Figura 2.10) uma dispersão relativamente alta. Isso se deve, pelo menos em parte, ao fato de que foram simulados dados do logaritmo do esforço e da captura, e antes de se obter as estimativas dos índices foi aplicada a exponencial a esses dados, para voltar à escala original.

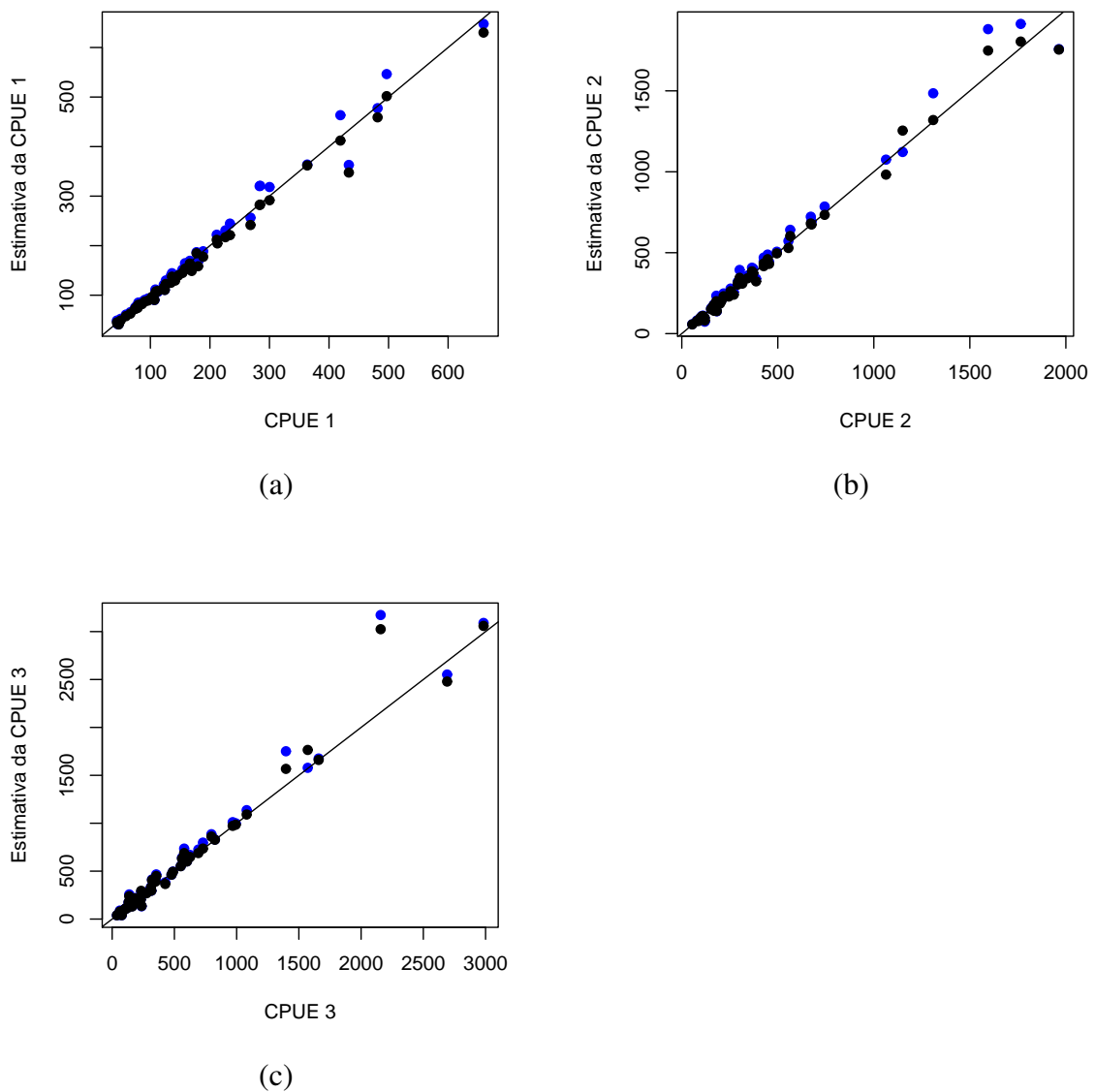


Figura 2.10 – Valores verdadeiros da  $CPUE_1$  (a),  $CPUE_2$  (b) e  $CPUE_3$  (c), versus mediana das estimativas ajustadas (em preto), amostrais (em azul). Cenário d), modelo (2.10)

Os EQM's resultantes das estimativas ajustadas usando o modelo (2.10) no estudo de simulação



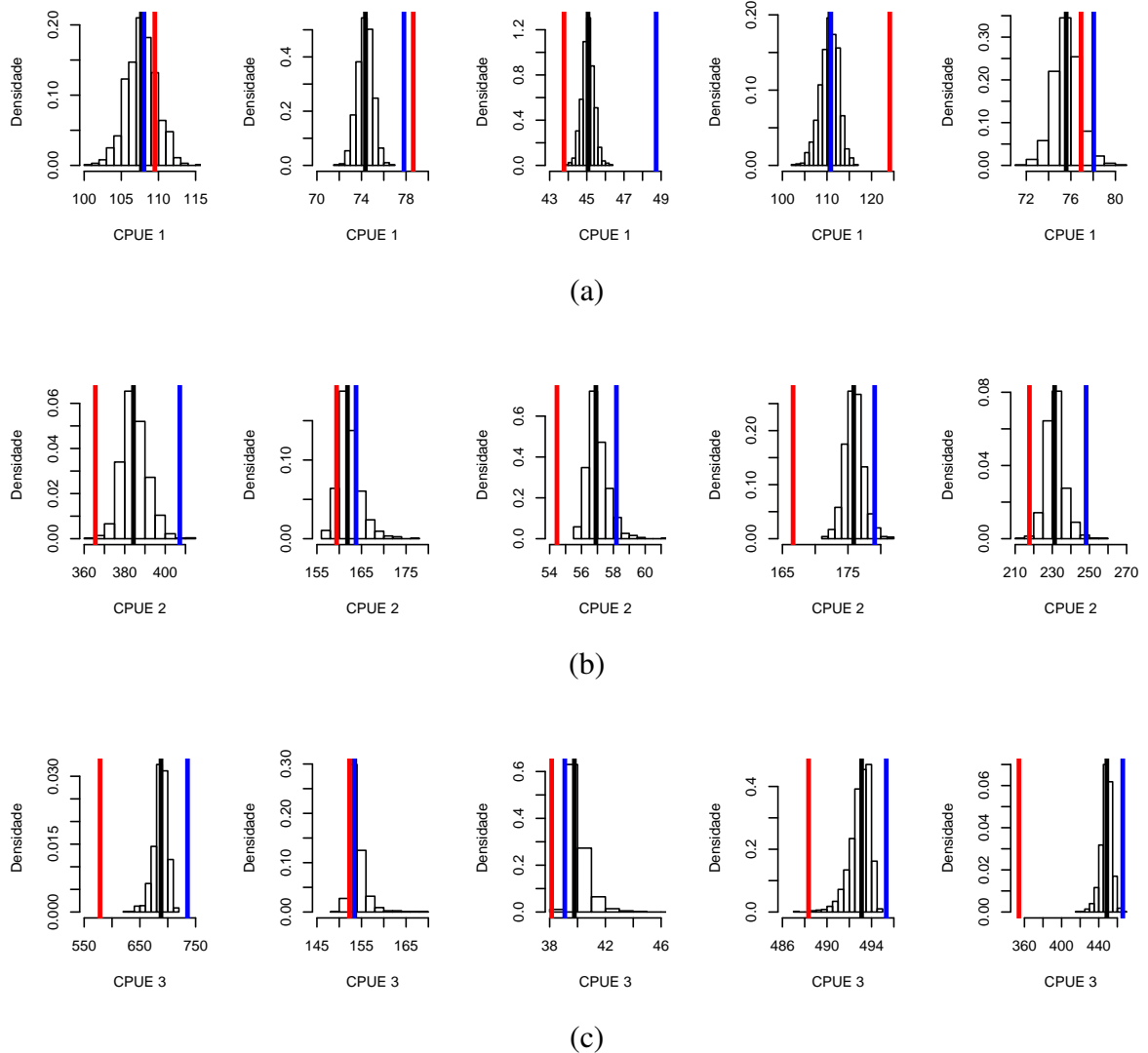


Figura 2.11 – Histogramas das estimativas ajustadas referentes a 5 das 50 amostras da  $CPUE_1$  (a),  $CPUE_2$  (b),  $CPUE_3$  (c), com a respectiva mediana (em preto), estimativa amostral (em azul) e valor verdadeiro (em vermelho). Modelo (2.10), cenário d)

são apresentados na Tabela 2.2. A tabela apresenta, também, os EQM's das estimativas amostrais, obtidas para cada um dos índices em cada um dos cenários. Comparando os resultados obtidos pelos dois métodos de estimação (amostral e ajustado), observa-se que o método ajustado apresentou os menores EQM's para os três índices no cenário d). Nos cenários b) e c) o método ajustado apresentou os menores EQM's para as  $CPUE_2$  e  $CPUE_3$  e no cenário a) apenas para a  $CPUE_2$ .

Os resultados obtidos eram esperados, pois no cenário a) praticamente não se tem correlação espacial. Assim era esperado que os quadratins vizinhos contribuíssem muito pouco para explicar o que ocorreria nos quadratins não observados. Além disso, a correlação entre esforço e captura

é baixa, ou seja, uma variável contribui muito pouco para explicar a outra (a informação de que se dispõe para fazer a interpolação é, praticamente, apenas da covariável). Dessa forma, não se esperava ganho significativo em fazer a interpolação nos quadratins não observados, uma vez que para esses quadratins se tem pouca informação para explicar os níveis das variáveis. Assim, para este cenário, usar os valores preditos, além dos valores observados (método ajustado), não contribui para melhorar as estimativas dos índices.

Os cenários b) e c) são situações intermediárias. No cenário b) a correlação espacial é baixa, mas a correlação entre captura e esforço é forte. Enquanto que no cenário c) a correlação entre captura e esforço é baixa, mas tem-se forte correlação espacial. Assim, para esses cenários não é surpreendente que o método ajustado não tenha sido melhor para todos os índices.

No cenário d) tem-se forte correlação espacial. Dessa forma, espera-se grande contribuição dos quadratins vizinhos para explicar o que ocorre nos quadratins não observados. Além disso, a forte correlação entre as variáveis esforço e captura faz com que uma variável nos quadratins observados ajude a explicar a outra variável nos quadratins vizinhos não observados. Neste cenário, era esperado um ganho, ao fazer a interpolação nos quadratins não observados usando o modelo (2.10). De fato, ter usado os valores preditos, além dos valores observados, para calcular os índices (método ajustado), contribuiu para melhorar as estimativas dos índices, tendo como critério o EQM. Pode-se afirmar que, para este cenário, as estimativas ajustadas dos três índices são melhores que as estimativas amostrais, uma vez que apresentam sempre os menores EQM's (Tabela 2.2).

Tabela 2.2 – Erro Quadrático Médio das estimativas ajustadas, usando o modelo (2.10) e amostrais das  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  para cada um dos cenários.

| Cenário | $CPUE_1$ |          | $CPUE_2$  |           | $CPUE_3$   |            |
|---------|----------|----------|-----------|-----------|------------|------------|
|         | Ajustada | Amostral | Ajustada  | Amostral  | Ajustada   | Amostral   |
| a)      | 35,1504  | 32,6330  | 2,4261    | 2,6006    | 2,0064     | 1,9894     |
| b)      | 473,4891 | 426,4459 | 2095,0472 | 4544,9838 | 11038,7357 | 11902,9450 |
| c)      | 5,2285   | 3,5032   | 1,8794    | 3,6860    | 5,6303     | 6,7847     |
| d)      | 242,7253 | 283,2965 | 2070,7258 | 4375,5495 | 18994,0142 | 26041,8815 |

Na Tabela 2.3 são apresentados os resultados das somas dos desvios absolutos relativos (SDAR), critério dado pela eq. (2.18). Esses resultados permitem comparar os três índices de abundância  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ .

Em todos os cenários, a  $CPUE_3$  foi a que apresentou os maiores valores da SDAR, tanto para as estimativas amostrais, quanto para as estimativas ajustadas. Se o pesquisador optar por usar a  $CPUE_3$  como índice de abundância, ao estimar o seu índice, ele não vai conseguir estimar tão bem quanto estimaria as  $CPUE_1$  ou  $CPUE_2$  se estas tivessem sido escolhidas, para qualquer um dos cenários a), b), c) e d).

Tabela 2.3 – Soma dos desvios absolutos relativos das estimativas ajustadas (modelo (2.10)) e amostrais das CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub> para cada um dos cenários.

| Cenário | Ajustada          |                   |                   | Amostral          |                   |                   |
|---------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|         | CPUE <sub>1</sub> | CPUE <sub>2</sub> | CPUE <sub>3</sub> | CPUE <sub>1</sub> | CPUE <sub>2</sub> | CPUE <sub>3</sub> |
| a)      | 5,8063            | 6,7768            | 34,6395           | 5,2275            | 7,1229            | 34,3925           |
| b)      | 4,0389            | 3,9539            | 6,3407            | 3,7867            | 4,8877            | 6,5419            |
| c)      | 3,9811            | 3,9168            | 8,4449            | 3,5011            | 5,1038            | 9,1513            |
| d)      | 2,4598            | 3,0322            | 5,4389            | 2,3549            | 4,1312            | 6,3204            |

Comparando a CPUE<sub>1</sub> com a CPUE<sub>2</sub>, observa-se na Tabela 2.3, que para as estimativas amostrais, a CPUE<sub>1</sub> apresenta sempre os menores valores da SDAR. Considerando as estimativas ajustadas, a CPUE<sub>1</sub> apresenta menores valores da SDAR do que a CPUE<sub>2</sub> nos cenários a) e d). Nos cenários b) e c) a CPUE<sub>2</sub> apresenta valores menores da SDAR, apesar dos valores serem muito próximos dos valores da SDAR da CPUE<sub>1</sub>. Assim, se o pesquisador pretende usar um dos dois índices CPUE<sub>1</sub> ou CPUE<sub>2</sub>, segundo o critério da SDAR, é aconselhável utilizar a CPUE<sub>1</sub> para qualquer um dos 4 cenários, se for estimar apenas com a amostra. Se ele pretende obter estimativas ajustadas, ou seja, fazer a predição nos locais não observados, antes de obter as estimativas é aconselhável utilizar a CPUE<sub>1</sub> para os cenários a) e d) e a CPUE<sub>2</sub> para os cenários b) e c), apesar de não perder muito se usar a CPUE<sub>1</sub> para esses dois cenários também.

Nos estudos de simulação realizados por Petreire Jr, Giacomini e De Marco Jr (2007), os autores observam que o desempenho dos índices (CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub>) dependem da variância e não da distribuição do erro. Porém, esses autores não analisaram os índices na presença de correlação espacial. Aqui os resultados obtidos não evidenciam uma associação entre a estrutura de covariância (os diferentes cenários) e os índices CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub>. Em geral a CPUE<sub>1</sub> é a que apresenta os menores valores da soma dos desvios absolutos relativos.

Na Tabela 2.4 são apresentados os EQM's das estimativas resultantes do ajuste do modelo em que o esforço é considerado conhecido (modelo (2.16)).

Tabela 2.4 – Erro Quadrático Médio das estimativas ajustadas usando o modelo (2.16) (modelo com esforço conhecido) e amostrais das CPUE<sub>1</sub>, CPUE<sub>2</sub> e CPUE<sub>3</sub> para cada um dos cenários.

| Cenário | CPUE <sub>1</sub> |          | CPUE <sub>2</sub> |           | CPUE <sub>3</sub> |            |
|---------|-------------------|----------|-------------------|-----------|-------------------|------------|
|         | Ajustada          | Amostral | Ajustada          | Amostral  | Ajustada          | Amostral   |
| a)      | 29,5145           | 32,6330  | 1,7895            | 2,6006    | 0,3149            | 1,9894     |
| b)      | 451,4268          | 426,4459 | 2427,6271         | 4544,9838 | 2989,5891         | 11902,9450 |
| c)      | 3,1127            | 3,5032   | 1,4844            | 3,6860    | 0,9247            | 6,7847     |
| d)      | 230,4046          | 283,2965 | 3258,3424         | 4375,5495 | 15172,1750        | 26041,8815 |

Os EQM's das estimativas amostrais apresentados na Tabela 2.4 não são influenciados pelo modelo que está sendo ajustado, visto que essas estimativas são obtidas apenas da amostra que está sendo considerada. Por essa razão, os EQM's das estimativas amostrais apresentadas na Tabela 2.4 e na Tabela 2.2 são os mesmos.

O EQM das estimativas ajustadas da  $CPUE_2$  para o cenário d), obtidas do ajuste do modelo em que o esforço é conhecido, foi de 3258,3424 (Tabela 2.4). Este valor é maior do que o EQM da  $CPUE_2$  do cenário d) para o modelo (2.10), cujo valor foi de 2070,7252 (Tabela 2.2). O fato ocorreu devido a duas simulações que produziram estimativas aberrantes. Excluindo-se essas duas estimativas aberrantes e calculando-se novamente o EQM, obtém-se 617,948 para o modelo com esforço conhecido e 783,3903 para o modelo (2.10), o que mostra que o modelo com esforço conhecido estima melhor os três índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  no cenário d) do que o modelo em que o esforço é aleatório (modelo (2.10)).

No cenário b) ocorre situação semelhante à do cenário d). Comparando-se os EQM's das estimativas ajustadas das Tabelas 2.4 e 2.2, observa-se que o EQM da  $CPUE_2$  do modelo em que o esforço é conhecido (Tabela 2.4) é maior do que o EQM do modelo (2.10) (Tabela 2.2). Isso também ocorre devido a duas simulações que produzem estimativas aberrantes. Calculando novamente os EQM's sem as estimativas aberrantes, obtém-se 1204,1860 para a  $CPUE_2$  do modelo com esforço conhecido e 1556,7810 para a  $CPUE_2$  do modelo (2.10). Assim, o modelo com esforço conhecido estima melhor os três índices no cenário b) do que o modelo (2.10).

Nos cenários a) e c) os EQM's das estimativas ajustadas para o modelo com esforço conhecido (Tabela 2.4) são sempre menores que para o modelo (2.10) (Tabela 2.2).

Pode-se afirmar de maneira geral que, para os quatro cenários analisados, as estimativas obtidas do ajuste do modelo com esforço conhecido são melhores do que as estimativas obtidas do ajuste do modelo (2.10). O resultado era esperado, uma vez que no ajuste do modelo com esforço conhecido não há incerteza associada ao esforço, para os locais onde não ocorreram pescas (locais não observados). Como não há incerteza associada ao esforço, isso ajuda a minimizar a incerteza associada à predição da captura para os locais não observados. Dessa forma, esse modelo produz melhores estimativas dos três índices do que o modelo (2.10), pois no modelo com esforço aleatório, há incerteza associada à predição do esforço e da captura. Além disso, a incerteza associada à predição do esforço contribui para aumentar a incerteza associada à predição da captura.

Deve-se notar que na pesca comercial, apesar dos planos de gestão muitas vezes limitarem o esforço de pesca para uma determinada região e/ou um determinado período de tempo, visando à exploração sustentável dos estoques, o esforço exercido em cada quadratim não é controlado, ou seja, é aleatório. Assim, o modelo que considera o esforço conhecido não é muito realista. Porém, em pesca experimental, em que o pesquisador decide a quantidade de esforço de pesca que será exercida em cada quadratim, esse modelo faz mais sentido.

Outro fato importante a ser notado é que as estimativas das  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , obtidas após o ajuste do modelo em que o esforço é conhecido (estimativas ajustadas), apresentam sempre os menores EQM's do que as estimativas amostrais. No modelo ajustado, considera-se que o esforço seja conhecido; assim, para os locais não observados, apenas a captura está sendo predita. No cenário d), para prever a captura, tem-se a informação referente à covariável (temperatura), tem-se a correlação espacial, de forma que a captura nos quadratins observados, fornece informação sobre a captura nos quadratins não observados tem-se, também, a correlação entre captura e esforço, ou seja, como o esforço é conhecido, ele fornece informação para prever a captura nos locais em que ela não foi observada. Além disso, o esforço nos locais observados ajuda a explicar a captura nos locais não observados (covariância cruzada). Dessa forma, há bastante informação para a predição da captura. Conseqüentemente, as  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  são melhor estimadas, após o ajuste do modelo. No cenário b), tem-se a correlação entre captura e esforço, além da covariável, para ajudar a explicar a captura nos locais não observados. E no cenário c), tem-se a correlação espacial, além da covariável, para auxiliar na predição da captura. Assim, nesses dois cenários, também, as  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  são melhor estimadas após o ajuste do modelo, do que simplesmente as estimativas obtidas da amostra.

No cenário a), tem-se a informação da covariável para explicar a captura nos locais não observados. Apesar de se ter baixa correlação entre captura e esforço e baixa correlação espacial, a captura está sendo predita suficientemente bem, de forma que as estimativas ajustadas são melhores do que as estimativas amostrais. O fato de não se ter incerteza associada ao esforço, faz com que se tenha menos incerteza associada à predição da captura. Por isso, e pela informação fornecida pela covariável, mesmo nessa situação extrema (baixa correlação entre captura e esforço e baixa correlação espacial), as  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  são melhor estimadas após o ajuste do modelo, do que simplesmente usando a amostra.

Os EQM's das estimativas obtidas após o ajuste do modelo em que tanto a captura quanto o esforço são considerados independentes no espaço (modelo (2.17)), são apresentados na Tabela 2.5.

Tabela 2.5 – Erro Quadrático Médio das estimativas ajustadas (modelo com esforço e captura independentes no espaço) e amostrais das  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  para cada um dos cenários.

| Cenário | $CPUE_1$ |          | $CPUE_2$  |           | $CPUE_3$   |            |
|---------|----------|----------|-----------|-----------|------------|------------|
|         | Ajustada | Amostral | Ajustada  | Amostral  | Ajustada   | Amostral   |
| a)      | 35,4609  | 32,6330  | 2,4108    | 2,6006    | 2,0086     | 1,9894     |
| b)      | 475,3316 | 426,4459 | 2068,0036 | 4544,9838 | 11079,3396 | 11902,9450 |
| c)      | 7,7926   | 3,5032   | 2,7029    | 3,6860    | 6,6656     | 6,7847     |
| d)      | 397,1175 | 283,2965 | 2962,4262 | 4375,5495 | 19993,4331 | 26041,8815 |

No cenário a) os dados foram simulados com baixa correlação espacial e baixa correlação en-

tre esforço e captura. Observa-se que o EQM da estimativa ajustada da  $CPUE_1$  da Tabela 2.5 (cenário a)) é maior do que o EQM da Tabela 2.2 para o mesmo cenário. Já para as estimativas ajustadas das  $CPUE_2$  e  $CPUE_3$  ocorre o contrário. Entretanto, os valores dos EQM's para este cenário (Tabela 2.5) são semelhantes aos valores dos EQM's para o mesmo cenário apresentados na Tabela 2.2 (estimativas ajustadas usando o modelo (2.10)). Apesar do modelo (2.10) levar em conta a correlação espacial, não há vantagem em usar esse modelo neste cenário. Isso se deve ao fato da correlação espacial ser muito baixa para este cenário.

No cenário b) observa-se, também, que os EQM's das estimativas ajustadas das  $CPUE_1$  e  $CPUE_3$  da Tabela 2.5 são maiores do que os EQM's da Tabela 2.2 e que o EQM da  $CPUE_2$  é maior na Tabela 2.2. Neste cenário, tem-se baixa correlação espacial e forte correlação entre captura e esforço. Porém, os dois modelos consideram a correlação entre a captura e o esforço. Assim, de fato, não era esperado que um dos modelos fosse sempre melhor que o outro neste cenário.

Para o cenário c) os EQM's das estimativas ajustadas da Tabela 2.5 são maiores do que os EQM's apresentados na Tabela 2.2 resultantes do ajuste do modelo 2.10. O resultado está de acordo com o previsto, pois o modelo (2.17), cujos resultados são apresentados na Tabela 2.5, não considera a correlação espacial, a qual se faz presente neste cenário.

Comparando os resultados apresentados nas Tabelas 2.2 e 2.5 para o cenário d), nota-se que os EQM's apresentados na Tabela 2.2 são sempre menores. De fato, quando se usa o modelo (2.10), cujos EQM's são apresentados na Tabela 2.2, tem-se mais informação para fazer a predição da captura e do esforço nos locais não observados, do que quando se usa o modelo (2.17). Usando o modelo (2.10) leva-se em conta, além da correlação entre captura e esforço, a correlação espacial e a relação existente entre esforço nos locais observados, com captura nos locais não observados.

Enfim, o modelo (2.10) se mostrou melhor do que o modelo (2.17), em que a captura e o esforço são considerados independentes no espaço, na estimação das  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  para os cenários c) e d), onde se tem correlação espacial, a qual é negligenciada quando se ajusta o modelo (2.17).

### 2.2.6 Aplicação

Uma aplicação da metodologia foi feita utilizando dados de pesca de arrasto do camarão rosa *Farfantepenaeus brasiliensis* and *F. paulensis* desembarcado nas regiões Sudeste e Sul do Brasil no ano de 1998.

O conjunto de dados é composto pelas variáveis esforço em número de lances e captura em kg. Tem-se um total de 65 pares de valores constituídos por esforço e captura. Para cada par tem-se um ponto de referência dado em graus de latitude e longitude, indicando um quadratim de  $10' \times 10'$ , onde ocorreu a pesca. Os valores das profundidades mínimas e máximas por local de pesca também foram disponibilizados e poderiam ser considerados como covariáveis, porém não foram considerados, pois não apresentaram qualquer relação com as variáveis esforço e captura. As coordenadas dadas em graus de latitude e longitude foram transformadas utilizando a projeção Universal Transversa de Mercator (UTM), antes da análise geoestatística dos dados. Dessa forma, pode-se calcular a distância euclideana entre os diferentes locais de pesca.

A Figura 2.12 apresenta os locais onde ocorreram as pescarias do camarão rosa e as intensidades das variáveis esforço e captura por local de pesca.

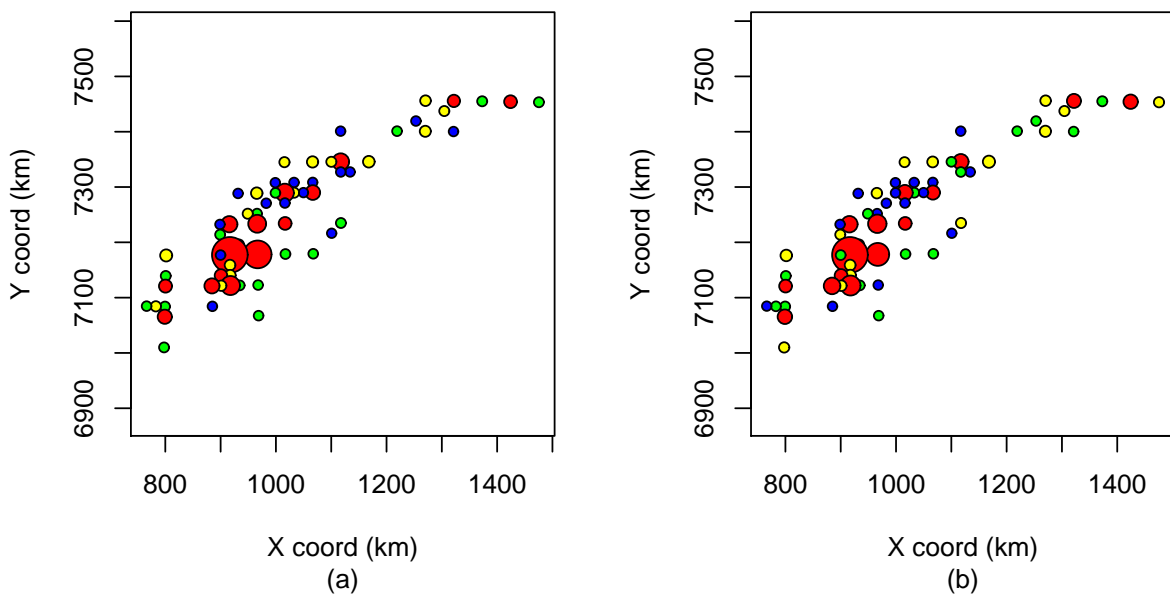


Figura 2.12 – Locais onde ocorreram as pescarias do camarão rosa, os valores das variáveis esforço (a) e captura (b) observadas são proporcionais aos diâmetros dos círculos e diferentes cores representam os diferentes quartis

Nota-se nas Figuras 2.13 e 2.14 que as distribuições do esforço e da captura são assimétricas

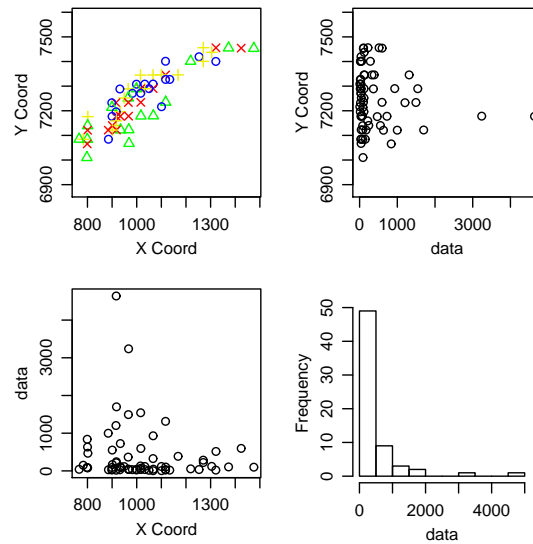


Figura 2.13 – Gráficos de dispersão e histograma para o esforço, os diferentes símbolos representam os quartis da distribuição dos dados

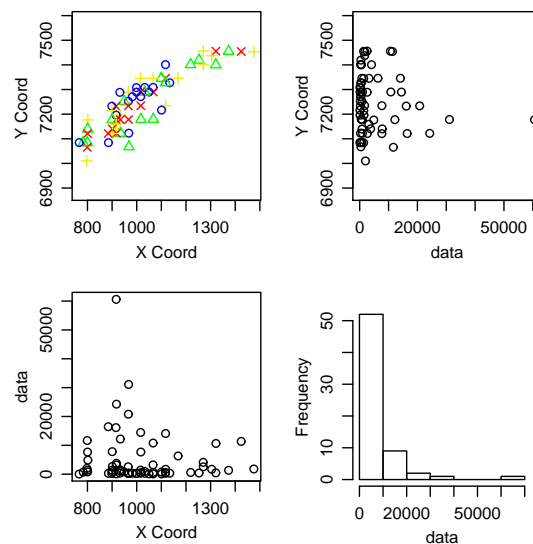


Figura 2.14 – Gráficos de dispersão e histograma para a captura. Os diferentes símbolos representam os quartis da distribuição dos dados

e aparentemente não apresentam tendências com as coordenadas. Foram construídos os gráficos que descrevem o valor estimado para o parâmetro da transformação Box Cox, apresentados na Figura 2.15. Esses gráficos mostram a necessidade da transformação logarítmica das variáveis esforço e captura.

As Figuras 2.16 e 2.17 apresentam gráficos de dispersão e os histogramas das variáveis esforço e captura, após a transformação logarítmica. Os histogramas das variáveis após a transformação



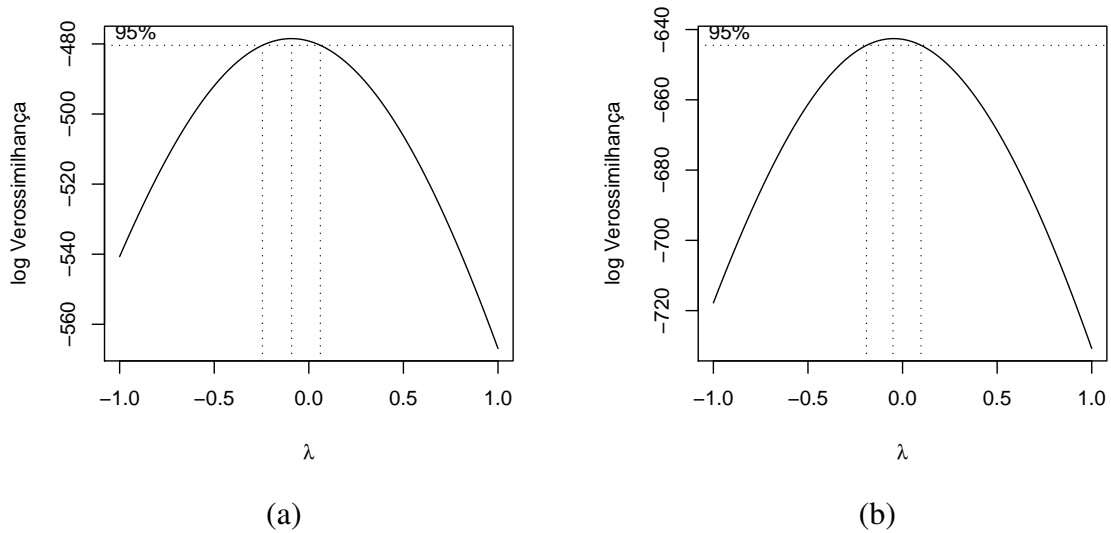


Figura 2.15 – Gráficos para a estimação do parâmetro da transformação Box Cox das variáveis esforço (a) e captura (b)

não apresentam mais assimetrias. Para testar a normalidade, foi aplicado o teste de Shapiro-Wilk, o qual indicou normalidade para o logaritmo do esforço ( $W = 0,9703$ ,  $p = 0,1198$ ) e para o logaritmo da captura ( $W = 0,9756$ ,  $p = 0,2254$ ).

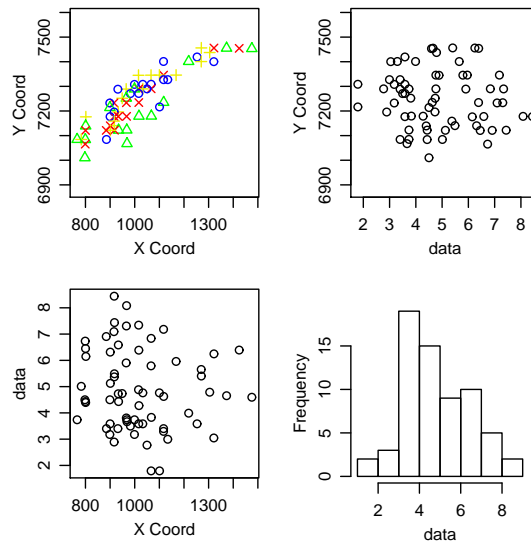


Figura 2.16 – Gráficos de dispersão e histograma para o logaritmo do esforço

A Figura 2.18 mostra, através do gráfico de dispersão, a forte correlação linear existente entre o log do esforço e o log da captura, cujo valor estimado da correlação é de 0,9615.

A fim de verificar a presença ou não de correlação espacial nos dados, construiu-se os semi-

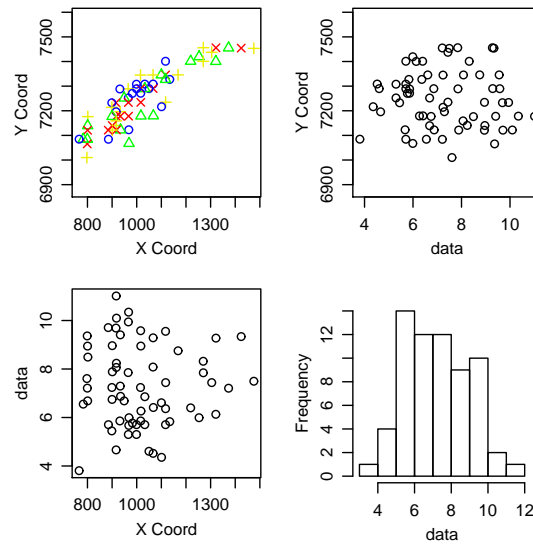


Figura 2.17 – Gráficos de dispersão e histograma para o logaritmo da captura

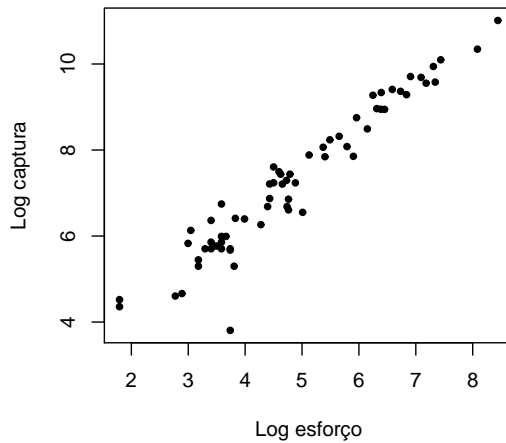


Figura 2.18 – Diagrama de dispersão para as variáveis log-esforço e log-captura

variogramas juntamente com o gráfico de envelope, para as variáveis log-esforço e log-captura (Figura 2.19). A região do envelope representa a região sem dependência espacial, e como os envelopes contruídos para os semivariogramas do logaritmo do esforço e do logaritmo da captura contêm todos os pontos dos semivariogramas, não há evidências de correlação espacial.

Por essa análise descritiva, os dados da pesca do camarão rosa estão de acordo com o cenário b) do estudo de simulação, onde se tem forte correlação entre o logaritmo do esforço e o logaritmo da captura, mas não se tem correlação espacial. Observa-se, também, que os dados de esforço e de captura seguem distribuição normal após a transformação logarítmica, o que está de acordo com as

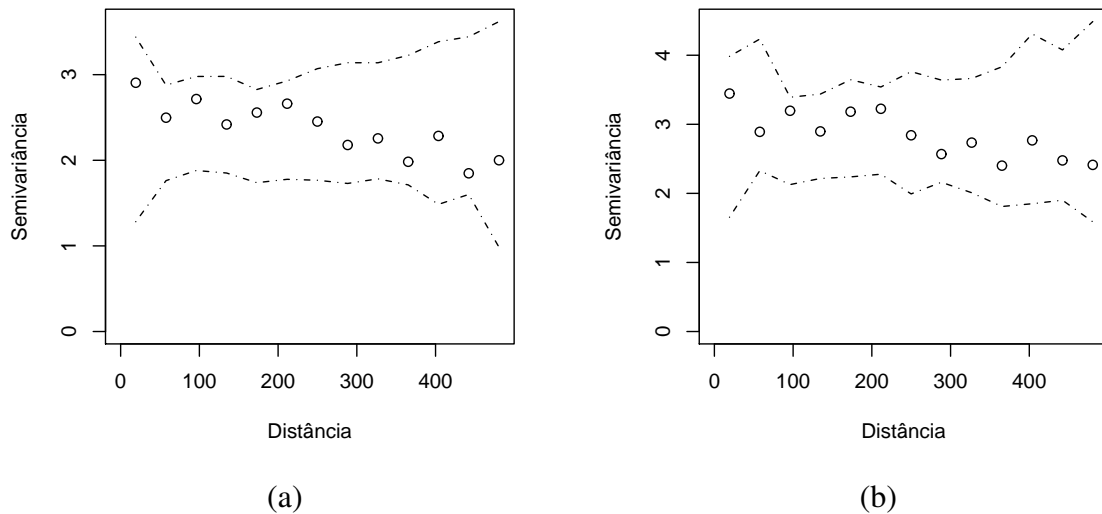


Figura 2.19 – Envelope e semivariograma das variáveis log-esforço (a) e log-captura (b)

suposições feitas no estudo de simulação.

Aos dados foram ajustados os modelos usados no estudo de simulação (modelos 2.10, 2.16 e 2.17), sem considerar, porém, a covariável temperatura, a qual não estava disponível para esses dados. Os modelos foram ajustados segundo a abordagem bayesiana, da mesma forma como foram ajustados no estudo de simulação.

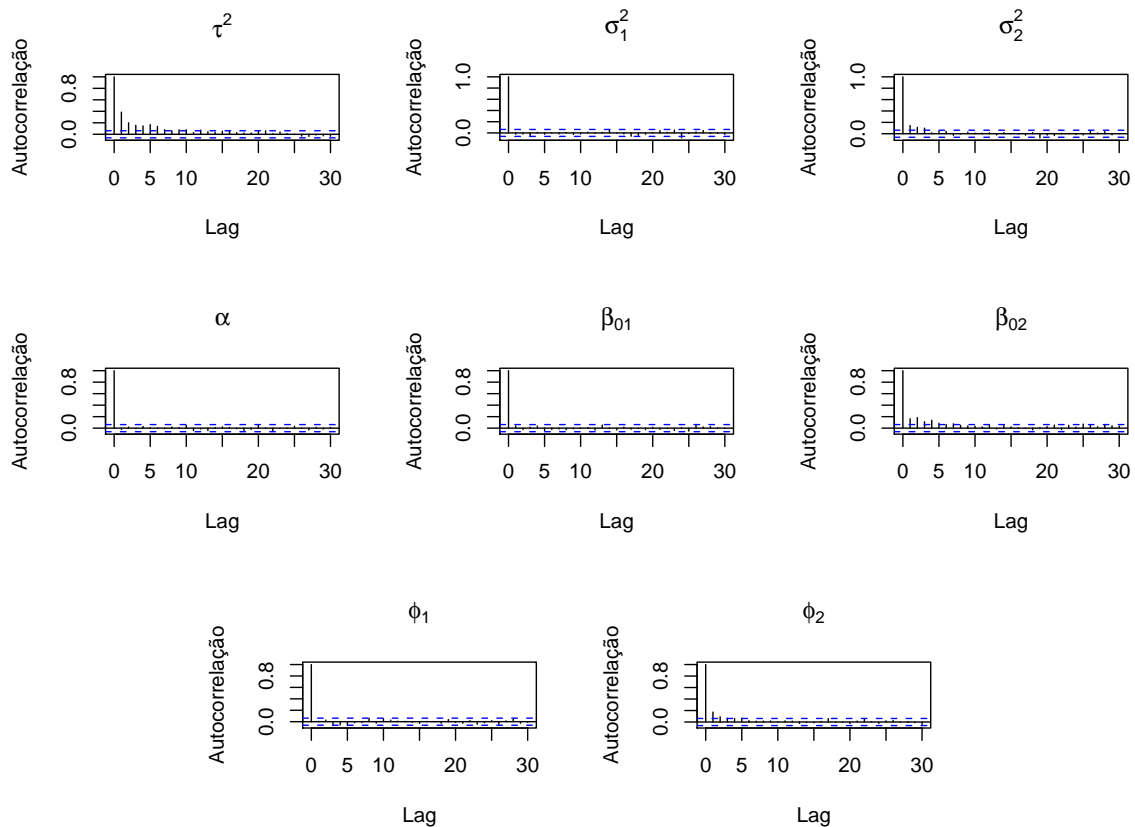
Após o ajuste de cada modelo foram obtidas as estimativas amostrais dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , ou seja, estas estimativas foram obtidas simplesmente utilizando-se os valores de captura e esforço das 65 localizações. Foram obtidas, também, as distribuições preditivas dos índices.

Na Figura 2.20 são apresentadas as funções de autocorrelação das cadeias da distribuição *a posteriori* dos parâmetros do modelo (2.10). E na Figura 2.21 são apresentados os traços de duas cadeias, para os parâmetros do modelo (2.10) ajustado aos dados da pesca do camarão rosa. As cadeias foram geradas a partir de valores iniciais distintos e permitem verificar que existem indícios de convergência das cadeias. As cadeias *a posteriori* obtidas para os parâmetros dos modelos (2.16) e (2.17) não foram apresentadas, mas também mostram indícios de convergência. Na Tabela 2.6 é apresentado um resumo da distribuição *a posteriori* de todos os parâmetros do modelo (2.10).

Na Tabela 2.7 é apresentado um resumo das distribuições preditivas dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  e a Figura 2.22 apresenta os histogramas das distribuições preditivas e suas medianas, obtidas após o ajuste dos modelos (2.10), (2.16) e (2.17), juntamente com as estimativas amostrais dos índices obtidas com base nos 65 pontos observados. Nessa figura, observa-se que as estimativas amostrais dos três índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  estão próximas das medianas das

Tabela 2.6 – Resumo da distribuição *a posteriori* dos parâmetros do modelo (2.10)

| Parâmetro    | Média   | 2,5%    | Mediana | 97,5%   |
|--------------|---------|---------|---------|---------|
| $\tau_2^2$   | 0,1111  | 0,0250  | 0,1036  | 0,2334  |
| $\sigma_1^2$ | 2,3964  | 1,7093  | 2,3635  | 3,2608  |
| $\sigma_2^2$ | 0,1546  | 0,0338  | 0,1343  | 0,4158  |
| $\alpha$     | 1,0547  | 0,9896  | 1,0535  | 1,1270  |
| $\beta_{01}$ | 4,8515  | 4,4650  | 4,8515  | 5,2190  |
| $\beta_{02}$ | 2,2135  | 1,6970  | 2,2175  | 2,7451  |
| $\phi_1$     | 30,7141 | 10,1382 | 24,9300 | 81,5277 |
| $\phi_2$     | 3,0112  | 0,0159  | 1,9870  | 11,9320 |

Figura 2.20 – Autocorrelação das cadeias da distribuição *a posteriori*, para os parâmetros do modelo 2.10 ajustado aos dados da pesca do camarão rosa

distribuições preditivas. Porém, para as estimativas amostrais tem-se apenas uma estimativa pontual dos três índices, enquanto que após o ajuste dos modelos (2.10), (2.16) e (2.17), tem-se uma estimativa pontual dada pela mediana e uma medida da incerteza dos índices através dos intervalos de predição 95%. Como já foi destacado, o ajuste de um modelo para a obtenção de estimativas dos índices leva a vantagem por permitir que se faça a predição do esforço e da captura nos locais não

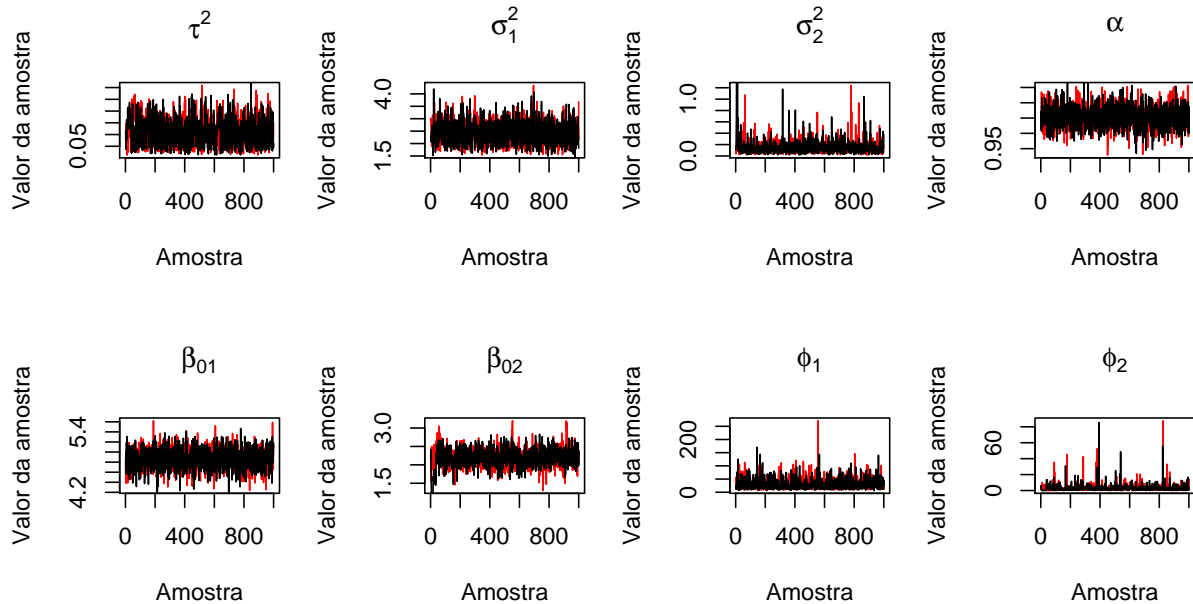


Figura 2.21 – Traço de duas cadeias *a posteriori*, representadas pelas cores vermelha e preta, partindo de valores iniciais distintos, para os parâmetros do modelo 2.10 ajustado aos dados da pesca do camarão rosa

observados e então pode-se, assim, obter estimativas para os índices considerando-se, também, valores de esforço e captura preditos em pontos não observados. Dos modelos ajustados aos dados do camarão rosa, para cada índice, a distribuição preditiva dos três modelos estão centradas em valores muito próximos. Entretanto, o modelo (2.17) foi o que apresentou uma distribuição preditiva mais concentrada. Para esses dados, que apresentam forte correlação entre esforço e captura e que não apresentam correlação espacial, não há vantagem em usar o modelo (2.10), segundo os resultados do estudo de simulação. Além disso, o modelo (2.10) apresentou uma distribuição preditiva mais dispersa que o modelo (2.17). Portanto, neste caso, o modelo (2.17) se mostra o mais adequado, por ser mais simples que o modelo (2.10), por levar em conta a correlação existente entre esforço e captura e por apresentar uma distribuição preditiva mais concentrada.

De acordo com o estudo de simulação, o modelo mais adequado a ser ajustado aos dados do camarão rosa é o modelo (2.17); sendo assim, o melhor índice de abundância relativa a ser usado neste caso é a  $CPUE_1$ , cuja mediana da preditiva é de 11,2183 com intervalo de predição (10,0167; 12,7400).

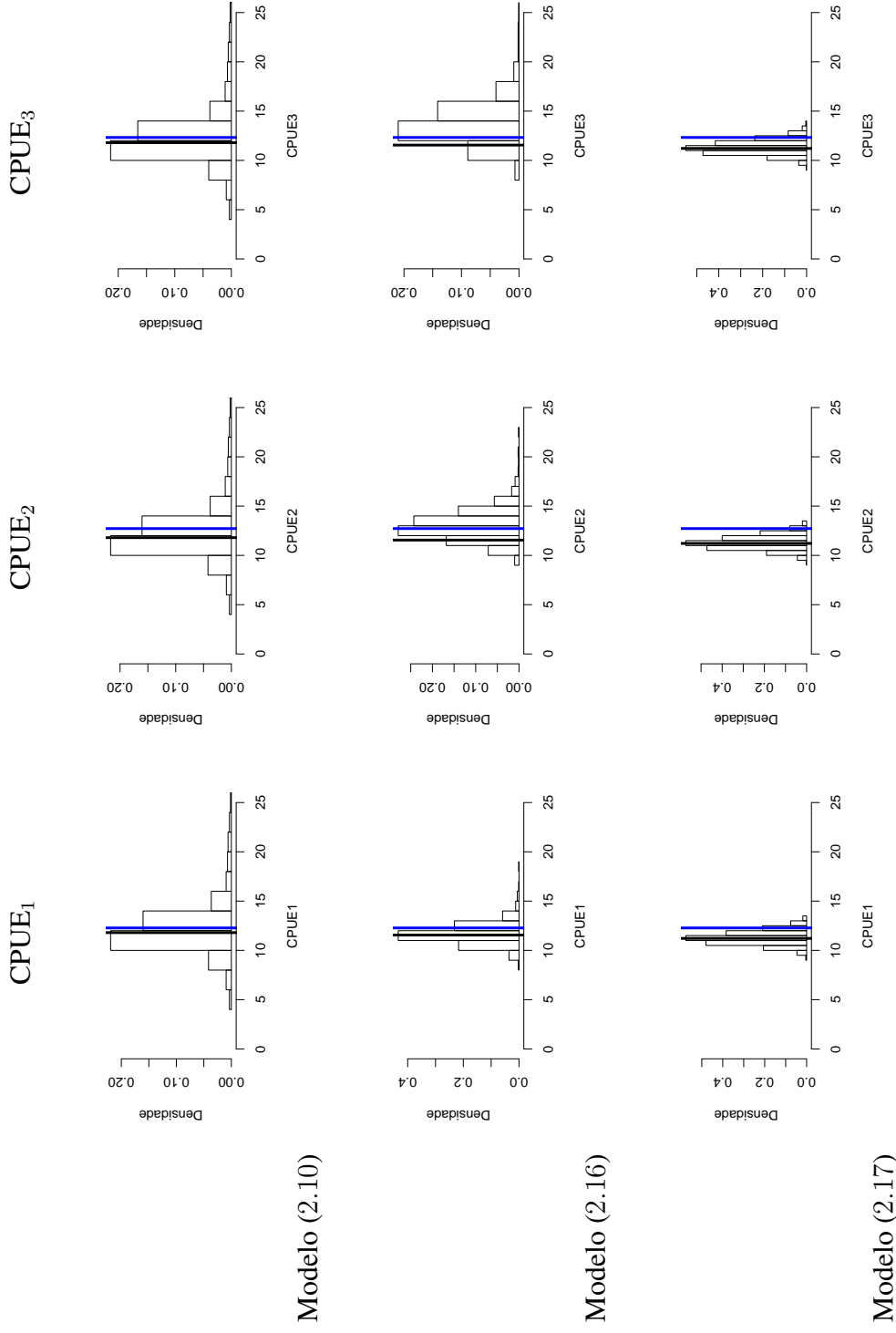


Figura 2.22 – Histogramas das distribuições preditivas dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , considerando os modelos (2.10), (2.16) e (2.17), juntamente com respectiva mediana (em preto) e a estimativa amostral de cada índice (em azul)

Tabela 2.7 – Resumo das distribuições preditivas dos índices  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$  obtidas após o ajuste dos modelos (2.10), (2.16) e (2.17)

|               |          | Média   | 2,5%    | Mediana | 97,5%   |
|---------------|----------|---------|---------|---------|---------|
| Modelo (2.10) | $CPUE_1$ | 12,2161 | 8,2283  | 11,8662 | 19,8507 |
|               | $CPUE_2$ | 12,2478 | 8,2812  | 11,8992 | 19,9759 |
|               | $CPUE_3$ | 12,2792 | 8,2893  | 11,9314 | 20,0599 |
| Modelo (2.16) | $CPUE_1$ | 11,6389 | 9,8905  | 11,5333 | 13,8886 |
|               | $CPUE_2$ | 13,0013 | 10,3725 | 12,8670 | 16,1207 |
|               | $CPUE_3$ | 13,6167 | 10,3749 | 13,4310 | 17,7611 |
| Modelo (2.17) | $CPUE_1$ | 11,2648 | 10,0167 | 11,2183 | 12,7400 |
|               | $CPUE_2$ | 11,2948 | 10,0444 | 11,2423 | 12,7761 |
|               | $CPUE_3$ | 11,3247 | 10,0667 | 11,2754 | 12,8385 |

### 2.3 Considerações Finais

O estudo de simulação realizado permitiu verificar que:

- (i) O método ajustado é melhor que o método amostral para o modelo (2.10), principalmente no cenário d);
- (ii) O método ajustado é melhor que o amostral para o modelo (2.16) em todos os cenários estudados;
- (iii) Dos modelos ajustados, o modelo (2.16) apresenta melhores estimativas dos três índices do que o modelo (2.10). Mas, para pesca comercial, o modelo (2.16) não é realista;
- (iv) Da comparação entre os modelos (2.10) e (2.17) quanto à estimação dos índices, verificou-se que as estimativas das  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ , obtidas após o ajuste do modelo (2.10), são melhores do que as estimativas obtidas após o ajuste do modelo (2.17), nos cenários c) e d), nos quais há presença de correlação espacial;
- (v) Dos três índices de abundância as estimativas da  $CPUE_1$  foram, em geral, as que apresentaram os menores desvios relativos para os cenários analisados.

Como já mencionado anteriormente, quando se emprega um índice de abundância que não apresenta as melhores propriedades e/ou, se o índice escolhido é mal estimado, isso pode levar a decisões equivocadas no manejo de estoques pesqueiros. Se um índice adotado for subestimado, isso indicará que o estoque está sub-explorado, quando na verdade pode não estar, o que implicará

em não tomar atitudes para conservação do recurso pesqueiro. Por outro lado, se o índice for superestimado, isso indicará sobre-pesca, o que poderá implicar na determinação da diminuição do esforço de pesca por parte dos gestores. Conseqüentemente, essa decisão afetaria trabalhadores do setor e consumidores do pescado e seu derivados.

Portanto, este trabalho mostra que a  $CPUE_1$  é o índice mais indicado nos diferentes cenários analisados e que o método de estimação ajustado melhora as estimativas dos índices em alguns cenários e dessa forma contribui para evitar equívocos na tomada de decisão no gerenciamento de estoques pesqueiros.

## Referências

ABUABARA, M. A. P. **Análise de dados de captura e esforço da pesca com espinhel de atum no Oceano Atlântico de 1974 a 1990**. 1996. 155 p. Tese (Doutorado em Engenharia) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. 1996.

ANGANUZZI, A. A. An aggregate model of effort distribution for the eastern Pacific tuna fishery. **Fisheries and Aquaculture Department**. Roma: FAO, 2004.

BADDELEY, A.; TURNER, R. Spatstat: an R package for analyzing spatial point patterns. **Journal of Statistical Software**, v. 12, p.1-42, 2005. Disponível em <[www.jstatsoft.org](http://www.jstatsoft.org)>. Acesso em: 8 de julho de 2008.

BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. **Hierarchical Modeling and Analysis for Spatial Data**. London: Chapman & Hall/CRC, 2004. 452 p.

CADIMA, E., L. **Manual de avaliação de recursos pesqueiros**. Roma: FAO, 2000. 162p.

COMISSÃO INTERNACIONAL PARA A CONSERVAÇÃO DOS ATUNS DO ATLÂNTICO. Banco de dados. Disponível em: <<http://www.iccat.int/accesingdb.htm>>. Acesso em: 17 abril 2007.

DIGGLE, P. J.; RIBEIRO JR. P. J. **Model-based Geostatistics**. Lancaster: Springer Series in Statistics, 2007. 230 p.

DOW, R. L.; BELL, F. W.; and HARRIMAN, D. M. **Bioeconomic relationships for the Maine lobster fishery with consideration of alternative management schemes**. NDAA Technical Report NMFS SSRF 683. 44 p. 1975.



DOW, R. L. The clawed lobster fisheries. In: COBB, J. S.; PHILLIPS, B. F. **The Biology and management of lobsters**. New York: Academic Press, 1980. v. 2. p. 265-316.

EVANS, C. R.; LOCKWOOD, A. P. M.; EVANS, A. J. Associations between sea temperature, catch per unit of fishing effort, and yield in the Bermuda spiny lobster fishery, 1975-89. **Marine & Freshwater Research** , Melbourne, v. 46, p. 809-818, 1995.

FONTENEAU, A. Preliminary comparative overview of the environment and the tuna fisheries catching yellowfin skipjack and bigeye, and operating in the Atlantic, Indian and Pacific oceans. Doc. SCRS/95/110-ICCAT . 1995. 16p.

GAMERMAN. D.; LOPES. H. F. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. 2.ed. London: Chapman & Hall, 2006. 323 p.

GELFAND, A.E.; SCHMIDT, A.M.; BANERJEE, S.; SIRMANS, C.F. Nonstationary Multivariate Process Modeling through Spatially Varying Coregionalization (with discussion). **Test**, Valencia, v. 13, n. 2, p. 263-312, Dec. 2004.

GOODYEAR, C. P. Spatio-temporal distribution of longline catch per unit effort, sea surface temperature and Atlantic marlin. **Marine and Freshwater Research**, New York, v. 54, p. 409-417, 2003.

HILBORN, R.; WALTERS, C.J. **Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty**. New York: Chapman and Hall, 1992. 570 p.

JOHNSON. R. A.; WICHERN. D. W. **Applied Multiariate Statistical Analysis**. 5.ed. New Jersey: Prentice-Hall, 2002. 767 p.

LIMA, J. H. M.; LIN, C. F.; MENESES, A. A. S. As pescarias brasileiras de bonito-listrado com vara e isca-viva, no sudeste e sul do Brasil, no período de 1980 a 1998. **Boletim Técnico Científico. CEPENE**, Tamandaré, v. 8, n. 1, p. 1-92, 2000.

O'HANGAN, A. **Kendall's Advanced Theory of Statistics: Bayesian Inference**, London: John Wiley & Sons, 1994. volume 2B, 448 p.

PAEZ, M. S.; GAMERMAN, D.; De OLIVEIRA, V. Interpolation performance of a spatio-temporal model with spatially varying Coefficients: Application to PM10 Concentration in Rio de Janeiro. **Environmental and Ecological Statistics**, New York, v. 12, p. 169-193, 2005.

PETREIRE JR, M.; GIACOMINI, H.C.; DE MARCO JR. P. **Catch-per-unit effort**: which estimator is best? Rio Claro: UNESP, Departamento de Ecologia, 2007. 17 p. Relatório Técnico.

REIS, I. A. **Testando a aleatoriedade espacial de uma configuração de pontos**. 1998. 43 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Minas Gerais, Belo Horizonte, 1998.

RIBEIRO JR., P.J; DIGGLE, P.J. geoR: A package for geostatistical analysis. **R-NEWS**, Vol 1, No 2, 15-18. 2001.

SANCHEZ, P.; MAYNOU, F.; DEMESTRE, M. Modelling catch, effort and price in a juvenile Eledone cirrhosa fishery over a 10 year period. **Fisheries Research**, Barcelona, v.68, p.319-327, 2004.

SCHMIDT, A. M.; CONCEIÇÃO, M. F. G.; MOREIRA, G. A. Investigating the sensitivity of Gaussian Processes to the choice of their correlation function and prior specifications. **Journal of Statistical Computation and Simulation**, New York, v. 78, p. 681-699, 2008.

SCHMIDT, A.M.; GELFAND, A. E. A bayesian coregionalization approach for multivariate pollutant data. **Journal of Geophysics Research - Atmospheres**, v. 108, no. D24, 8783. 2003.

SCHMIDT, A. M.; NOBRE, A. A.; FERREIRA, G. S. Alguns Aspectos da Modelagem de Dados Espacialmente Referenciados. **Revista brasileira de estatística**, Brasil, v. 63, n. 220, p. 59-88, 2003.

SCHMIDT, A. M.; SANSÓ, B. **Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço-Temporais**, Caxambu: ABE, 2006. 151 p.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical Methods**. 6.ed. Ames: The Iowa state University Press, 1967. 593 p.

SPIEGELHATER, D. J.; THOMAS, A; BEST, N.G. WinBugs version 1.4 user manual. Technical Report. Cambridge: Medical Research Council Biostatistics, 2002.

VIOLA, D. N. **Detecção e modelagem de padrão espacial em dados binários e de contagem.** 2007. 118 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

WALTERS, C. Folly and fantasy in the analysis of spatial catch rate data. **Canadian Journal of Fisheries and Aquatic Sciences**, Toronto, v.60, 1433-1436, Dec. 2003.

### **3 COMPARAÇÃO ENTRE MODELOS GEOESTATÍSTICOS UNIVARIADO E BIVARIADO NA ESTIMAÇÃO DA CPUE: UM ESTUDO DE SIMULAÇÃO**

#### **Resumo**

Dados de captura e esforço de pesca são utilizados para se obter índices de abundância relativa de um estoque pesqueiro. Recentemente, modelos geoestatísticos vêm sendo utilizados na análise de dados de captura e esforço de pesca. Modelos geoestatísticos multivariados têm sido utilizados em diversas áreas, mas não têm sido aplicados a dados de pesca. Neste artigo, foi realizado um estudo de simulação para comparar estimativas de um índice de abundância relativa baseado na captura por unidade de esforço, considerando-se o ajuste de um modelo geoestatístico univariado para a razão entre captura e esforço e um bivariado, em que são modelados conjuntamente a captura e o esforço. Foram considerados quatro cenários resultantes das combinações de duas intensidades de correlação espacial e duas de correlação entre esforço e captura. As estimativas obtidas após o ajuste dos dois modelos apresentaram resultados muito próximos, indicando que não há vantagem em usar o modelo bivariado na estimação do índice, para os cenários considerados. Uma aplicação da metodologia foi realizada utilizando dados de pesca de arrasto do camarão rosa *Farfantepenaeus brasiliensis* and *F. paulensis* desembarcado nas regiões Sudeste e Sul do Brasil no ano de 1998.

Palavras-chave: Estudo de Simulação; Inferência Bayesiana; Geoestatística; Captura por Unidade de Esforço; Estoque Pesqueiro

#### **Abstract**

Catch and effort data are used to obtain indices of relative abundance of a stock fishery. Recently, geostatistical models have been used in the analysis on catch and effort data. Multivariate geostatistical models have been used in several areas, but has not been applied to fisheries data. In this article was done a simulation study to compare the estimates of an indice of relative abundance based on catch per unit of effort considering the fit of a univariate geostatistical model for the ratio between catch and effort and a bivariate, which catch and effort are shaped together. We considered four scenarios resulting from combinations of two levels of spatial correlation and two levels of correlation between two effort and catch. Estimates obtained after fitting of the two models showed very similar results, indicating that no advantage in using the bivariate model in the estimation of the indices for scenarios considered. An application of the methodology was performed using data from the trawl fishery pink shrimp *Farfantepenaeus brasiliensis* and *F. paulensis* landed in the Southeast and Southern Brazil in 1998.

Keywords: Simulation, Bayesian inference; Geostatistics; Catch per unit effort, Fish stocks

### 3.1 Introdução

Na gestão de recursos pesqueiros, é fundamental considerar características biológicas, mas também são importantes os aspectos sociais e econômicos. Para garantir a sustentabilidade da pesca, deve-se buscar um equilíbrio entre a utilização do recurso em proveito da sociedade e a sua conservação (CADIMA, 1997).

Registros da pesca comercial fornecem bases de dados que muitas vezes são utilizados no auxílio do manejo de estoques (VIGNAUX, 1996; WALTERS, 2003). Os dados provenientes da pesca comercial em geral fornecem a captura - biomassa ou número de indivíduos retirados de um estoque pela pesca; e esforço de pesca - intensidade das operações de pesca, que podem ser medidas, por exemplo, em dias no mar ou dias de pesca ou, de forma mais precisa pelo número de anzóis por unidade de tempo (NADAL - EGEA, 1996). Nos registros da pesca comercial, usualmente são fornecidos também os quadratins (sub-regiões delimitadas por paralelos e meridianos) onde ocorreram as pescarias. É fornecido um ponto de referência do quadratim, dado em graus de latitude e longitude, indicando a ocorrência de pesca (ICCAT, 2007).

Segundo Vignaux (1996), dados de esforço e captura são usados na avaliação de estoques pesqueiros. Conjuntos de dados dessa natureza são usados, geralmente, para se obter um índice de abundância relativa de um estoque.

Dada uma região habitada por alguma espécie e dividida em quadratins, se estiverem disponíveis dados de esforço e de captura para toda a região, pode-se definir três índices de abundância relativa baseados na captura por unidade de esforço (CPUE). Esses índices foram denotados por Petrere Jr; Giacomini e De Marco Jr (2007) por  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ . O índice  $CPUE_1$  é dado pela média das razões entre a captura e o esforço nos diferentes quadratins;  $CPUE_2$  é dado pela razão entre o total das capturas e o total do esforço para toda a região, e  $CPUE_3$  é dado pela razão entre a soma das capturas ponderadas pelos esforços e a soma dos quadrados dos esforços.

Muitos autores têm relatado a autocorrelação espacial nos dados de pesca. Nesses casos, a estatística espacial, em particular a geoestatística, vem sendo empregada com diversas finalidades, tais como: otimizar delineamentos amostrais (PETITGAS, 2001, JARDIM; RIBEIRO JR, 2005); estimar biomassa ou abundância, levando-se em conta a existência de estruturas espaciais (MAYNOU, 1998; RIVOIRARD ET AL., 2000), o que pode resultar em estimativas mais acuradas da biomassa ou abundância (RUFINO ET AL., 2006); e prever a captura ou a razão entre captura e esforço para a obtenção de mapas dessas quantidades para uma determinada área (STELZENMÜLLER; MAYNOU; MARTÍN, 2007).

Modelos geoestatísticos multivariados, recentemente vêm sendo utilizados em diversas áreas. São utilizados para modelar: dados de poluentes na atmosfera (SCHMIDT; GELFAND, 2003); preço de venda e lucro advindo de imóveis comerciais (GELFAND ET AL., 2004); concentração

de argila e incremento anual volumétrico de espécies florestais (BOGNOLA, ET AL., 2007). Na pesca, porém, não têm sido empregados os modelos geoestatísticos multivariados na análise de dados de esforço e captura; desse modo, não são consideradas as covariâncias cruzadas espaciais entre essas variáveis.

Quando se analisa dados de pesca, para se obter um índice de abundância geral (para avaliar a tendência ao longo do tempo) em um estoque espacialmente estruturado, é necessário ter dados de captura e esforço para todos os quadratins que cobrem o estoque. Porém, geralmente não se tem dados para todos os quadratins (WALTERS, 2003). Em vez de se estimar o índice de abundância usando somente os dados observados, uma alternativa é sugerida por Walters (2003). O autor sugere que, se existirem dados referentes a um mesmo período de tempo de quadratins vizinhos e se for possível mostrar que existe estrutura de correlação espacial nas taxas de captura, pode-se usar métodos de estatística espacial para fazer a interpolação para os quadratins não observados. O autor sugere ainda usar alguma covariável, como por exemplo a temperatura da água, para auxiliar no processo de interpolação. Observa-se que o autor sugere o uso de métodos de estatística espacial para fazer a interpolação das taxas de captura, ou seja, da razão entre captura e esforço. Mas, uma outra possibilidade, que não tem sido proposta na literatura, é modelar conjuntamente a captura e o esforço, usando um modelo geoestatístico bivariado, em que a temperatura poderia ser considerada, também como covariável. Ajustando-se um modelo geoestatístico bivariado, pode-se fazer a interpolação do esforço e da captura, nos locais não observados, para depois calcular um índice de abundância para o estoque.

Diante dessas duas possibilidades, neste trabalho foi realizado um estudo de simulação. Para tanto, foram adotados dois modelos geoestatísticos: um univariado para a razão entre captura e esforço e outro bivariado, proposto por Gelfand et al. (2004) em que o esforço e a captura são modelados conjuntamente. Após o ajuste do modelo univariado foram feitas as interpolações para os quadratins não observados. Fazendo a interpolação da razão entre a captura e o esforço, dos três índices definidos acima é possível estimar apenas a  $CPUE_1$ . O modelo bivariado também foi ajustado, interpolações foram feitas para as variáveis captura e esforço, e em seguida foi estimado o índice de abundância  $CPUE_1$ .

O principal objetivo deste trabalho foi comparar, em diferentes cenários, através de um estudo de simulação: a) o comportamento das estimativas da  $CPUE_1$  obtidas após o ajuste dos dois modelos e b) comparar as estimativas da  $CPUE_1$  obtidas após o ajuste do modelo geoestatístico univariado com as estimativas amostrais que são obtidas apenas substituindo-se os dados observados no cálculo do índice  $CPUE_1$ , sem fazer previsões para locais não observados.

### 3.2 Desenvolvimento

A estatística espacial pode ser dividida em três grandes áreas, de acordo com os tipos de observações associadas ao espaço em que elas são observadas. As três grandes áreas da estatística espacial são a 1) geoestatística, 2) dados de área e 3) processos pontuais (CRESSIE, 1993). A seguir será feita uma breve revisão sobre a geoestatística. Uma revisão sobre dados de área pode ser encontrada em Reis E. (1998), e uma discussão sobre processos pontuais pode ser encontrada em Diggle (2003).

Dados geoestatísticos são medidas de uma variável que é contínua na região de estudo, tomadas em localizações referenciadas espacialmente. O objetivo da geoestatística é modelar a distribuição espacial dos valores de uma variável em uma região de estudo, decompondo esta estrutura espacial em dois componentes: a tendência média de larga escala (variação de primeira ordem) e a dependência espacial de pequena escala (variação de segunda ordem). Esta última pode ser modelada como correlação espacial, em uma função da distância entre as localizações amostrais, onde as localizações mais próximas no espaço geralmente têm valores mais similares de dados. A principal finalidade da análise de fenômenos espacialmente contínuos é usar tal modelo para obter boas predições dos valores em pontos onde a variável não foi amostrada (REIS E., 1998).

Na geoestatística, considera-se que o conjunto de observações  $\{y(s) : s \in G\}$  é uma realização parcial do processo aleatório  $\{Y(s) : s \in G\}$ , em que  $G$  é um subconjunto fixo do  $\mathbb{R}^g$ , geralmente  $g = 1, 2$  ou  $3$ . Assumindo-se que  $\mu(s) = \mathbf{E}[Y(s)]$  existe para todo  $s \in G$ ,  $\mu(s)$  é chamada de tendência do processo espacial. Supondo-se que a variância do processo existe para todo  $s \in G$ , o processo é chamado intrinsecamente estacionário se  $\mathbf{E}(Y(s+h) - Y(s)) = 0$  e  $V(Y(s+h) - Y(s)) = 2\gamma(h)$ , para quaisquer  $s, s+h \in G$ . O processo  $Y(\cdot)$  é chamado estacionário de segunda ordem ou fracamente estacionário se  $\mu(s)$  é constante para todo  $s \in G$ , isto é,  $\mu(s) = \mu, \forall s \in G$  e a covariância entre dois pontos quaisquer em  $G$  é função apenas da diferença entre as duas localizações, ou seja,  $cov[Y(s), Y(s')] = c(s - s'), \forall s, s' \in G$ . A quantidade  $2\gamma(\cdot)$  é chamada de variograma e a função  $c(\cdot)$  é chamada de covariograma. Se  $2\gamma(s - s')$  depender somente da distância euclidiana entre  $s$  e  $s'$ , o processo  $Y(\cdot)$  é então chamado de isotrópico. Um processo intrinsecamente estacionário e isotrópico é chamado de homogêneo, caso contrário, o processo é heterogêneo.

Em geral, assume-se que a variável aleatória  $Y(\cdot)$  segue um processo gaussiano. Mais especificamente, a quantidade aleatória  $Y(\cdot)$ , que assume valores  $y(s)$ ,  $s \in G$ , segue um processo gaussiano (PG) com média  $\mu(\cdot)$  e função de covariância  $c(\cdot, \cdot)$ , se para quaisquer  $s_1, s_2, \dots, s_n$ , e qualquer  $n = 1, 2, \dots$ , a distribuição conjunta de  $Y(s_1), \dots, Y(s_n)$  é uma normal multivariada com parâmetros dados por  $\mathbf{E}[Y(s_i)] = \mu(s_i)$  e  $cov[Y(s_i), Y(s_j)] = c(s_i, s_j)$  (O'HANGAN, 1994). O processo gaussiano é denotado por  $Y(\cdot) \sim PG(\mu(\cdot), c(\cdot, \cdot))$ .

Quando um processo é homogêneo, sua variância é constante ao longo de  $G$ , isto é,  $V(Y(s)) = \sigma^2, \forall s \in G$ . Assim, a função de covariância de  $Y(\cdot)$ , pode ser escrita como  $c(s, s') = \sigma^2 \rho(\|s - s'\|, \phi^*)$ , sendo  $\rho(\|s - s'\|, \phi^*)$  uma função de correlação válida em  $\mathbb{R}^g$ , que depende possivelmente de um vetor de parâmetros  $\phi$ . Nota-se que, sendo o processo homogêneo, a estrutura de covariância do processo  $Y(\cdot)$ , pode ser modelada apenas através dos parâmetros  $\sigma^2$  e  $\phi^*$  (SCHMIDT; NOBRE; FERREIRA, 2003). Existem na literatura muitas funções de correlação que levam a estruturas de covariâncias válidas. Entre elas, as mais usadas são a família exponencial potência e a família Matérn. Alguns dos principais modelos paramétricos de função de correlação podem ser encontrados em Diggle; Ribeiro-Jr (2007).

De forma geral, pode-se supor que a tendência do processo  $Y(\cdot)$  não é constante ao longo da região  $G$ , mas sim uma combinação linear de funções conhecidas  $\mathbf{f}(x) = [f_1(x), \dots, f_r(x)]'$ ,  $z \in G$ , isto é,  $\mu(s) = \boldsymbol{\beta}' \mathbf{f}(s)$ , em que  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_r]'$ . As quantidades  $f_j(\cdot)$ ,  $j = 1, \dots, r$  representam covariáveis que possivelmente explicam o nível de  $Y(\cdot)$ . O processo de interesse  $Y(\cdot)$  pode ser descrito pela soma de três componentes, isto é:

$$Y(s) = \boldsymbol{\beta}' \mathbf{f}(s) + Z(s) + \epsilon(s), \quad (3.1)$$

em que a primeira componente representa a tendência do processo, a segunda é um efeito aleatório com estrutura de correlação espacial  $Z(s)$ , e a última é um ruído branco,  $\epsilon(s)$ , com distribuição normal, média 0 e variância  $\tau^2$ , isto é,  $\epsilon(s) \sim N(0, \tau^2)$ . Para  $Z(\cdot)$  assume-se que  $(Z(s) | \sigma^2, \phi^*) \sim PG(0, \sigma^2 \rho(\|s - s'\|, \phi^*))$ , sendo que  $\sigma^2$  é a variância do processo  $Z(\cdot)$  e  $\rho(\cdot; \phi^*)$  representa a função de correlação que depende do parâmetro  $\phi^*$ . A componente de ruído  $\epsilon(s)$ , descreve um erro de medida chamado na literatura geoestatística de efeito pepita (CRESSIE, 1993).

Freqüentemente, ao invés de se ter uma única variável observada em alguns pontos de uma região  $G$ , tem-se observações de diversas variáveis que são obtidas nos mesmos pontos de amostragem. Neste caso, sendo as diversas variáveis correlacionadas espacialmente, mas não correlacionadas entre si, elas podem ser modeladas independentemente usando modelos geoestatísticos como o da eq. (3.1). No caso das diversas variáveis apresentarem, além da correlação espacial, correlação entre si, é interessante buscar modelos que levem em consideração, além da posição espacial dos pontos, a relação entre as variáveis de interesse. Dessa forma, os modelos geoestatísticos univariados podem ser estendidos para o caso multivariado, visando modelar as covariâncias simples e cruzadas espaciais entre as variáveis envolvidas, que não são contempladas quando se adota um modelo univariado para cada variável.

Assumindo-se que os vetores de observações nos pontos amostrados  $\mathbf{Y}(s_i), i = 1, 2, \dots, n$  formam uma realização parcial do processo estocástico  $\{\mathbf{Y}(s) : s \in G\}$ ,  $G \subset \mathbb{R}^g$  e que o vetor aleatório  $\mathbf{Y}(\cdot)$  segue um processo gaussiano, pode-se propor um modelo para esses dados descrito por

$$\mathbf{Y}(s) = \mathbf{X}(s)\boldsymbol{\beta} + \mathbf{v}(s) + \boldsymbol{\epsilon}(s) \quad (3.2)$$



em que  $v(s)$  segue uma distribuição normal  $p$ -variada com vetor de médias  $\mathbf{0}$  e matriz de covariâncias  $p \times p$ ,  $\epsilon(s)$  é um vetor de ruído branco com distribuição normal, com matriz de covariâncias  $\mathbf{D}$ , de dimensão  $p$ , diagonal, tal que  $D_{jj} = \tau_j^2$ . A componente  $\mathbf{X}(s)\beta$  representa a tendência do processo, em que  $\mathbf{X}(\cdot)$  é uma matriz  $p \times q$  contendo as possíveis covariáveis.

Para o modelo (3.2), a covariância cruzada é representada por  $cov[\mathbf{Y}(s), \mathbf{Y}(s')]$ , uma matriz  $p \times p$ , que segundo Banerjee et al. (2004), não precisa ser simétrica, isto é,  $cov[Y_j(s), Y_{j'}(s')]$  não precisa ser igual a  $cov[Y_{j'}(s), Y_j(s')]$ . Schmidt; Sansó (2006) discutem a modelagem da estrutura de covariância de modelos geoestatísticos multivariados, e a utilização de estruturas de covariâncias separáveis e modelos de co-regionalização linear, que podem assumir formas mais flexíveis e resultar em estruturas de covariância não separáveis e não estacionárias.

### 3.2.1 Aplicações da geoestatística à pesca

Segundo Petitgas (2001), desde a década de 1990 tem aumentado o interesse, por parte dos cientistas pesqueiros, em usar ferramentas geoestatísticas para a análise de dados de pesca. A geoestatística tem sido empregada na ciência pesqueira com diversas finalidades, tais como: otimizar delineamentos amostrais (PETITGAS, 1996, JARDIM; RIBEIRO JR, 2005); estimar biomassa ou abundância, levando-se em conta a existência de estruturas espaciais (MAYNOU, 1998; RIVOIRARD ET AL., 2000), o que pode resultar em estimativas mais acuradas da biomassa ou abundância (RUFINO ET AL., 2006); prever a captura ou a razão entre captura e esforço para a obtenção de mapas dessas quantidades para uma determinada área (STELZENMÜLLER; MAYNOU; MARTÍN, 2007).

Dados de captura e esforço de pesca são usualmente empregados na avaliação de estoques pesqueiros. São geralmente usados para se obter um índice de abundância relativa de um estoque (VIGNAUX, 1996; NISHIDA; CHEN, 2004). Os dados de captura e esforço muitas vezes são afetados por diferenças no poder de captura, causadas por variações nas características das embarcações, dos equipamentos de pesca, estação da pesca e vários fatores ambientais (HILBORN; WALTERS, 1992). Para se proceder a uma análise livre da influência desses fatores, geralmente os dados são padronizados. Existem várias abordagens para a padronização dos dados de pesca; a forma mais comum é o uso de modelos lineares generalizados (MAYNOU; DEMESTRE; SANCHEZ, 2003; MAUNDER; PUNT, 2004).

Segundo Stelzenmüller; Maynou e Martín (2007) a heterogeneidade espacial, isto é, a variação entre as diferentes áreas de pesca é às vezes levada em consideração na padronização dos dados. Entretanto, a padronização não elimina a influência da autocorrelação espacial. A autocorrelação espacial é causada pela estrutura espacial dos dados de esforço e captura, isto é, pela não independência de amostras vizinhas. Por isso, muitos autores, por exemplo Vignaux (1996), Stelzenmüller; Maynou e Martín (2007), após a padronização dos dados utilizam técnicas geoestatísticas

para a análise dos dados.

Alguns autores têm usado a geoestatística na análise de dados de captura por unidade de esforço. Nishida e Chen (2004), reconhecendo que geralmente a suposição de independência dos dados observados de captura por unidade de esforço não é válida por causa da correlação espacial, incorporam a correlação espacial no modelo linear generalizado, que chamaram de “spatial-GLM”. Ajustaram esse modelo a um conjunto de dados, em que testaram algumas funções de correlação e verificaram que, para os dados analisados, a função de correlação gaussiana foi a que melhor se ajustou. Os autores enfatizam a vantagem do modelo proposto por eles, sobre modelo linear generalizado sem incorporar a correlação espacial, principalmente quando os dados de captura por unidade de esforço apresentam forte correlação espacial.

Stelzenmüller; Maynou e Martín (2007), realizaram uma análise de dados combinando técnicas geoestatísticas e sistemas de informação geográfica (SIG). Esses autores usam dados padronizados de captura por unidade de esforço, ajustam funções de correlação e realizam a predição da razão entre captura e esforço, para a obtenção de mapas de captura por unidade de esforço para toda a área analisada.

Pode-se notar que os autores citados usam modelos geoestatísticos para contornar problemas de correlação espacial nos dados. Esses autores tratam a razão entre captura e esforço como variável resposta, nenhum deles trata a captura e o esforço conjuntamente, em que ambos poderiam ser considerados como variáveis respostas. Recentemente, em diversas áreas, como ciências florestais, geologia, ciências ambientais, têm sido usados modelos geoestatísticos multivariados. Porém, não foram encontrados trabalhos em pesca, que proponham uma modelagem geoestatística conjunta para esforço e captura, em que sejam contempladas em um modelo, além da correlação espacial, a relação existente entre captura e esforço e a covariância cruzada existente.

## 3.2.2 Metodologia

### 3.2.2.1 Modelo bivariado para esforço e captura

Com o propósito de comparar, através do estudo de simulação, o comportamento das estimativas da  $CPUE_1$  obtidas após o ajuste do modelo bivariado, em que são modelados conjuntamente a captura ( $C$ ) e o esforço ( $E$ ) e do modelo univariado para a razão entre captura e esforço ( $C/E$ ), considerou-se o modelo (3.2), com o vetor  $\mathbf{Y}(s)$  de dimensão  $p = 2$ , para a modelagem conjunta, em que  $Y_1 = \log E$  e  $Y_2 = \log C$ . Dessa forma, assumiu-se que o esforço e a captura após a transformação logarítmica são normalmente distribuídos e que as observações  $(\log E(s), \log C(s))'$ ,  $s \in G \subset \mathbb{R}^2$  ao longo da região  $G$  formam uma realização parcial de um processo gaussiano.

O modelo utilizado para a estrutura de covariância da componente  $\mathbf{v}(s)$  da eq. (3.2), foi o

modelo de co-regionalização linear proposto por Gelfand et al. (2004)

$$\mathbf{v}(s) = \mathbf{A}\mathbf{w}(s),$$

em que  $\mathbf{A}$  é uma matriz de posto completo, triangular inferior,  $w_j(s)$ ,  $j = 1, 2$  são independentes, mas não identicamente distribuídos. Dessa maneira, a matriz de covariância cruzada associada a  $\mathbf{Y}(s)$  é

$$\Sigma_{\mathbf{v}(s), \mathbf{v}(s')} = \sum_{j=1}^2 \rho_j(s - s') \mathbf{T}_j$$

em que  $\mathbf{T}_j = \mathbf{a}_j \mathbf{a}_j'$ , com  $\mathbf{a}_j$  sendo a  $j$ -ésima coluna de  $\mathbf{A}$ ;  $\mathbf{T}_j$  tem posto 1 e  $\sum_{j=1}^2 \mathbf{T}_j = \mathbf{T}$ ,  $\mathbf{T} = \mathbf{A}\mathbf{A}'$ .

Para este modelo, tem-se um alcance associado a cada componente de  $\mathbf{Y}(\cdot)$  se forem usadas funções de correlação monótonas e isotrópicas. A função de correlação usada foi a exponencial, tanto para o esforço quanto para a captura  $\rho_j(d) = \exp\{-|\phi_j d|\}$ ,  $j = 1, 2$ , em que  $d$  é a distância euclidiana entre dois pontos  $s, s'$  quaisquer.

A verossimilhança do modelo (3.2), pode ser obtida da seguinte forma: olhando para o modelo (3.2), como um modelo hierárquico, no primeiro estágio dados  $\{\beta_j, j = 1, 2\}$  e  $\{\mathbf{v}(s_i)\}$  os  $\mathbf{Y}(s_i)$ ,  $i = 1, 2, \dots, n$  são, condicionalmente a  $\mathbf{v}(s)$ , independentes com  $\mathbf{Y}(s_i) | \beta, \mathbf{v}(s_i) \sim \mathbf{N}(\mathbf{X}(s_i)\beta + \mathbf{v}(s_i), \mathbf{D})$ , para cada  $s_i$ . No segundo estágio, a distribuição conjunta de  $\mathbf{v} = (\mathbf{v}(s_1), \mathbf{v}(s_2), \dots, \mathbf{v}(s_n))$  é  $\mathbf{N}(\mathbf{0}, \sum_{j=1}^2 \mathbf{R}_j \otimes \mathbf{T}_j)$ , em que  $\mathbf{T}_j$  é uma matriz de dimensão  $2 \times 2$ ,  $\mathbf{R}_j$  tem dimensão  $n \times n$ , com  $(R_j)_{ii'} = \rho_j(s_i - s_{i'})$ . Concatenando os  $\mathbf{Y}(s_i)$  em um vetor  $\mathbf{Y}_{2n \times 1}$  e do mesmo modo os elementos  $\mu(s_i) = \mathbf{X}(s_i)\beta$  em  $\mu$ , a densidade marginal de  $\mathbf{Y}$  é dada por

$$\mathbf{Y} \sim \mathbf{N}(\mu_{2n}, \sum_{j=1}^2 \rho_j(s - s') \mathbf{T}_j + \mathbf{I}_n \otimes \mathbf{D}) \quad (3.3)$$

Porém, o modelo de co-regionalização linear para o processo  $\mathbf{v}(s) = \mathbf{A}\mathbf{w}(s)$  pode ser reparametrizado segundo uma abordagem condicional. Gelfand et al. (2004) mostram que existe uma equivalência entre o modelo (3.2) escrito na forma incondicional e condicional, desde que todos os processos exceto um sejam puramente espaciais. Assim, o modelo para duas variáveis, considerando uma covariável, temperatura (*temp*), escrito na forma condicional é dado por

$$\begin{aligned} Y_1(s) &= \beta_{01} + \beta_{11} \text{temp}(s) + \sigma_1 w_1(s) \\ Y_2(s) | Y_1(s) &= \beta_{02} + \beta_{12} \text{temp}(s) + \alpha Y_1(s) + \sigma_2 w_2(s) + \tau_2 u_2(s). \end{aligned} \quad (3.4)$$

Usando a abordagem condicional, a verossimilhança pode ser escrita como  $L(\mathbf{Y} | \theta) = L(\mathbf{Y}_1 | \theta_1) L(\mathbf{Y}_2 | \mathbf{Y}_1, \theta_2)$ , em que  $\theta$  é o vetor de parâmetros do modelo,  $\theta_1 = (\beta_1, \sigma_1^2, \phi_1)'$  é o vetor de parâmetros referente à primeira equação do modelo (3.4) e  $\theta_2 = (\beta_2, \sigma_2^2, \phi_2, \tau_2^2)'$  referente à segunda equação. Assim,

$$L(\boldsymbol{\beta}_1, \phi_1, \sigma_1^2 \mid \mathbf{y}_1) = (2\pi)^{-\frac{n}{2}} |\sigma_1^2 \mathbf{R}_1|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1]^T [\sigma_1^2 \mathbf{R}_1]^{-1} [\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1] \right\} \quad (3.5)$$

em que  $\mathbf{X}_1$  é a matriz de covariáveis para  $Y_1(s)$  e  $\boldsymbol{\beta}_1$  é o vetor de coeficientes das covariáveis.

$$L(\boldsymbol{\beta}_2, \phi_2, \sigma_2^2, \tau_2^2 \mid \mathbf{y}_2, \mathbf{y}_1) = (2\pi)^{-\frac{n}{2}} |\sigma_2^2 \mathbf{R}_2 + \tau_2^2 I|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_2 \mid \mathbf{y}_1 - \mathbf{X}_2 \boldsymbol{\beta}_2]^T [\sigma_2^2 \mathbf{R}_2 + \tau_2^2 I]^{-1} [\mathbf{y}_2 \mid \mathbf{y}_1 - \mathbf{X}_2 \boldsymbol{\beta}_2] \right\} \quad (3.6)$$

em que  $\mathbf{X}_2$  é a matriz de covariáveis para  $Y_2$  e  $\boldsymbol{\beta}_2$  é o vetor de coeficientes das covariáveis.

Utilizando a abordagem bayesiana, é necessário incorporar a incerteza relativa aos parâmetros de interesse. Assumindo-se independência, a distribuição *a priori* conjunta é dada por

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}_1) \pi(\boldsymbol{\theta}_2) = \pi(\boldsymbol{\beta}_1) \pi(\sigma_1^2) \pi(\phi_1) \pi(\boldsymbol{\beta}_2) \pi(\sigma_2^2) \pi(\phi_2) \pi(\tau_2^2).$$

Seguindo a literatura geoestatística, para os vetores  $\boldsymbol{\beta}_1$  e  $\boldsymbol{\beta}_2$  foram atribuídas *prioris* normais multivariadas com vetor de médias  $\mathbf{0}$  e matriz de covariâncias  $\mathbf{I} \sigma_{\boldsymbol{\beta}}^2$ , com  $\sigma_{\boldsymbol{\beta}}^2$  fixado em um valor grande,  $\sigma_{\boldsymbol{\beta}}^2 = 100$  (*priori vaga*). Aos parâmetros  $\sigma_1^2$  e  $\sigma_2^2$ , foram utilizadas *prioris* gama invertida, com média igual à estimativa de mínimos quadrados de um modelo espacialmente independente para cada variável,  $Y_1(s)$  e  $Y_2(s)$ , com variância infinita, isto é,  $\sigma_j^2 \sim IG(a_{\sigma_j}, b_{\sigma_j})$ , com  $a_{\sigma_j} = 2$ . Para  $\tau_2^2$  também, utilizou-se a *priori* gama inversa,  $\tau_2^2 \sim IG(a_{\tau_2}, b_{\tau_2})$ , com variância infinita ( $a_{\tau_2} = 2$ ) e média ( $b_{\tau_2}$ ) igual à estimativa de mínimos quadrados. E para os parâmetros  $\phi_1$  e  $\phi_2$  foram associadas *prioris* gama,  $\phi_j \sim G(a_{\phi_j}, b_{\phi_j})$ , com variância grande e média igual ao alcance efetivo associado à metade da máxima distância entre as localizações observadas, isto é,  $\rho \left( \mathbf{E}(\phi_j), \frac{\max.\text{dist}}{2} \right) = 0,05$ ,  $j = 1, 2$ , em que *max.dist* representa a máxima distância entre as localizações observadas (SCHMIDT; GELFAND, 2003; BANERJEE ET AL., 2004; PAEZ ET AL., 2005). A parametrização das distribuições *a priori* estão de acordo com Gelman et al. (2004).

Trabalhando-se com o modelo na forma condicional dado pela eq. (3.4) e com o uso das *prioris* especificadas anteriormente, a distribuição conjunta *a posteriori* para os parâmetros é dada por

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}_1, \mathbf{y}_2) = \pi(\boldsymbol{\theta}_1 \mid \mathbf{y}_1) \pi(\boldsymbol{\theta}_2 \mid \mathbf{y}_2),$$

em que

$$\pi(\boldsymbol{\theta}_1 \mid \mathbf{y}_1) = \pi(\boldsymbol{\beta}_1, \phi_1, \sigma_1^2 \mid \mathbf{y}_1) \propto L(\boldsymbol{\beta}_1, \phi_1, \sigma_1^2 \mid \mathbf{y}_1) \pi(\boldsymbol{\beta}_1) \pi(\phi_1) \pi(\sigma_1^2) \quad (3.7)$$

e

$$\pi(\boldsymbol{\theta}_2 \mid \mathbf{y}_2) = \pi(\boldsymbol{\beta}_2, \phi_2, \sigma_2^2, \tau_2^2 \mid \mathbf{y}_2, \mathbf{y}_1) \propto L(\boldsymbol{\beta}_2, \phi_2, \sigma_2^2, \tau_2^2 \mid \mathbf{y}_2, \mathbf{y}_1) \pi(\boldsymbol{\beta}_2) \pi(\phi_2) \pi(\sigma_2^2) \pi(\tau_2^2). \quad (3.8)$$

Como as distribuições *a posteriori* (3.7) e (3.8) não têm formas analíticas fechadas, resumos *a posteriori* de interesse são obtidos de amostras das distribuições *a posteriori* utilizando-se algoritmos MCMC, tais como amostrador de Gibbs e Metropolis-Hastings (GAMERMAN; LOPEZ, 2006). Para a implementação computacional desses algoritmos é necessário o conhecimento das distribuições condicionais completas *a posteriori* de todos os parâmetros de interesse presentes no modelo. No Apêndice D são apresentadas as distribuições condicionais completas de cada parâmetro do modelo (3.4).

### 3.2.2.2 Modelo univariado para a razão entre captura e esforço

O modelo univariado para a razão entre a captura ( $C$ ) e o esforço ( $E$ ) é dado por

$$Y(s) = \mathbf{X}(s)\boldsymbol{\beta} + \sigma w(s) + \tau_2 u(s) \quad (3.9)$$

em que,  $Y(s) = \log\left(\frac{C}{E}(s)\right)$ ,  $\mathbf{X}$  é a matriz contendo o efeito das covariáveis,  $w(\cdot)$  segue um processo gaussiano com média 0 e variância 1 e  $u(s) \sim N(0, 1)$ . A função de correlação utilizada foi a exponencial  $\rho(d) = \exp(-\phi d)$ . O parâmetro  $\phi$  desta função relata o quão rapidamente a correlação decai para zero.

Considerando-se que  $\mathbf{Y} = (Y(s_1), \dots, Y(s_n))$  forma uma amostra aleatória parcial do processo gaussiano  $\{Y(s) : s \in G\}$ , a distribuição conjunta de  $Y(s_1), \dots, Y(s_n)$  é normal  $n$ -variada, isto é,  $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{R} + \tau^2 \mathbf{I}_n)$ , em que  $R_{ii'} = \rho(s_i - s_{i'})$  e  $\mathbf{I}_n$  é a matriz identidade de ordem  $n$ . Aqui considerou-se, também, uma covariável, temperatura (*temp*), desta forma a  $i$ -ésima linha da matriz  $\mathbf{X}\boldsymbol{\beta}$  é dada por  $\beta_0 + \beta_1 \text{temp}(s_i)$ . A distribuição conjunta considerada como função dos parâmetros, dadas as observações, é a verossimilhança do modelo (3.9), que pode ser expressa por  $L(\boldsymbol{\beta}, \phi, \sigma^2, \tau^2 \mid \mathbf{y}) =$

$$(2\pi)^{-\frac{n}{2}} |\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]^T [\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}]^{-1} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}] \right\}. \quad (3.10)$$

Como é usual na literatura geoestatística, assumiu-se que os parâmetros do modelo (3.9) são independentes *a priori*, isto é,  $\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \phi) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\phi)$ . Da mesma forma que se fez para o modelo (3.4), para  $\sigma^2$  e  $\tau^2$  foram adotadas *prioris* informativas, ou seja, *prioris* gama inversa,  $\sigma^2 \sim IG(a_\sigma, b_\sigma)$ ,  $\tau^2 \sim IG(a_\tau, b_\tau)$ , sendo que suas médias *a priori* foram fixadas nas estimativas de mínimos quadrados dos dados observados com uma variância infinita. Para o parâmetro  $\phi$  da função de correlação exponencial foi adotada *priori* gama,  $\phi \sim G(a_\phi, b_\phi)$ , com variância *a priori* grande e média tal que quando a correlação é 0,05,  $\phi$  é igual à metade da distância máxima entre as localizações observadas. Ao parâmetro  $\boldsymbol{\beta} = (\beta_0, \beta_1)'$  foi atribuída uma *priori* normal bivariada com média  $\mathbf{0}$  e matriz de variâncias  $\sigma_{\boldsymbol{\beta}}^2 \mathbf{I}_2$ ,  $\sigma_{\boldsymbol{\beta}}^2 = 100$  (*priori* vaga).

Seguindo a abordagem bayesiana, a distribuição conjunta *a posteriori* é proporcional ao produto da verossimilhança pela *priori*, isto é,

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \phi \mid \mathbf{y}) &\propto |\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]^T [\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}]^{-1} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}] \right\} \\ &\times \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\phi) \end{aligned} \quad (3.11)$$

Como a distribuição *a posteriori* (3.11), também, não tem forma analítica fechada, torna-se necessário o uso de algoritmos MCMC para a obtenção de uma amostra da distribuição *a posteriori* conjunta dos parâmetros. As distribuições condicionais completas necessárias para a implementação dos algoritmos MCMC são apresentadas a seguir. Porém, cabe enfatizar que a implementação computacional dos modelos foi feita usando o programa WinBUGS (SPIEGELHATER ET AL., 2002).

1. Distribuição condicional completa *a posteriori* para  $\boldsymbol{\beta} = (\beta_0, \beta_1)'$  é normal bivariada

$$\boldsymbol{\beta} \mid \sigma^2, \tau^2, \phi, \mathbf{y} \sim N(\mathbf{B}\mathbf{b}, \mathbf{B}),$$

$$\text{em que } \mathbf{B} = (\mathbf{X}'[\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}]^{-1} \mathbf{X} + (\sigma^2 \boldsymbol{\beta} \mathbf{I})^{-1})^{-1} \text{ e } \mathbf{b} = \mathbf{X}'[\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}]\mathbf{y};$$

2.  $\pi(\phi \mid \boldsymbol{\beta}, \sigma^2, \tau^2, \mathbf{y}) \propto$

$$|\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]' [\sigma^2 \mathbf{R}_2 + \tau^2 \mathbf{I}]^{-1} [\mathbf{y}_2 - \mathbf{X}\boldsymbol{\beta}] - b_\phi \phi \right\} \phi^{a_\phi - 1};$$

3.  $\pi(\sigma^2 \mid \boldsymbol{\beta}, \tau^2, \phi, \mathbf{y}) \propto$

$$|\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]' [\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}]^{-1} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}] - \frac{b_\sigma}{\sigma^2} \right\} (\sigma^2)^{-a_\sigma - 1}$$

4.  $\pi(\tau^2 \mid \boldsymbol{\beta}, \sigma^2, \phi, \mathbf{y}) \propto$

$$|\sigma^2 \mathbf{R}_2 + \tau^2 \mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]' [\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}]^{-1} [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}] - \frac{b_\tau}{\tau^2} \right\} (\tau^2)^{-a_\tau - 1}.$$

### 3.2.3 Estudo de simulação

Foram simulados 50 conjuntos de dados do logaritmo do esforço ( $Y_1$ ) e do logaritmo da captura ( $Y_2$ ). Considerou-se dois valores diferentes para a correlação ( $\rho_{Y_1, Y_2}$ ) entre o logaritmo do esforço e da captura,  $\rho_{Y_1, Y_2} = 0,3$  baixa correlação, para simular a situação em que a incerteza do pescador é grande e  $\rho_{Y_1, Y_2} = 0,8$  retratando menor incerteza do pescador. Considerou-se também duas situações quanto à correlação espacial dos dados. Na primeira situação, considerou-se baixa correlação espacial para simular uma situação em que praticamente não existe correlação entre amostras vizinhas. Nessa situação, os valores dos alcances efetivos associados a cada uma das variáveis foram 0,5 unidades de distância para o esforço e 0,54 para a captura, ou seja, os parâmetros das funções

de correlação exponencial foram fixados em  $\phi_1 = 6$ ,  $\phi_2 = 5,5$  respectivamente. Na segunda situação, considerou-se forte correlação espacial, de forma que os alcances efetivos associados a cada uma das variáveis foram 6,67 unidades de distância para o esforço e 6,05 para a captura, isto é, os parâmetros das funções de correlação foram  $\phi_1 = 0,45$  (para esforço) e  $\phi_2 = 0,5$  (para captura). A combinação das diferentes intensidades de correlação entre captura e esforço e de correlação espacial leva a quatro cenários distintos, a saber

- (a) baixa correlação entre captura e esforço e baixa correlação espacial;
- (b) forte correlação entre captura e esforço e baixa correlação espacial;
- (c) baixa correlação entre captura e esforço e forte correlação espacial;
- (d) forte correlação entre captura e esforço e forte correlação espacial.

Os dados simulados representam realizações de processos gaussianos bivariados, ou seja, os dados foram simulados do modelo (3.4). Cada cenário especificado implica em um vetor de parâmetros diferente para o modelo (3.4). Foram simulados 50 conjuntos de dados para cada cenário.

Em cada cenário foram simulados dados em uma grade regular constituída de 100 pontos ( $10 \times 10$ ). Cada um desses pontos representa um ponto de referência dentro de um quadratim, onde teria ocorrido pesca, ou seja, onde teriam ocorridos esforço e captura. Deve-se notar que dada uma região habitada por um estoque, dividida em quadratins, nem sempre ocorre pesca em todos os quadratins, isto é, nem sempre são feitas "observações" de captura e esforço em todos os quadratins (WALTERS, 2003). Assim, os dados simulados em todos os quadratins representam o esforço e a captura que seriam observados, caso a pesca tivesse ocorrido em todos eles. Caso a pesca ocorresse em todos os quadratins, como a captura e o esforço são consideradas variáveis aleatórias em cada localização, poderiam apresentar diferentes realizações, por isso foram gerados 50 conjuntos de dados, para simular diferentes realizações dessas variáveis em cada localização.

Para cada conjunto de dados simulados, tem-se um total de 100 pares de esforço e captura. Com esses 100 pares foi calculado o índice de abundância do estoque, dado por  $CPUE_1 = \frac{1}{100} \sum_{i=1}^{100} \frac{C_i}{E_i}$ . O valor obtido é o valor verdadeiro do índice para a realização atual. Em seguida, para cada realização foi tomada uma amostra de tamanho 85 e outra de tamanho 76. Isto é, considerou-se duas situações, em uma delas 15% dos quadratins não foram amostrados (não houve pesca) e na outra 24%.

Os pontos considerados como amostrados foram pré-fixados, sendo os mesmos para as 50 realizações. Para isso, utilizou-se um processo de Thomas (REIS I., 1998) para representar uma situação real, em que geralmente o padrão espacial da pesca é agregado (ANGANUZZI, 2004). Porém, como em geral o local de pesca é referenciado por um único ponto em cada quadratim, as coordenadas dos pontos gerados pelo processo de Thomas foram aproximadas para as coordenadas

dos pontos mais próximos da grade ( $10 \times 10$ ). Na Figura 3.1 são mostrados os pontos considerados como observados nas duas situações, quando 15% e 24% dos quadratins não foram amostrados.

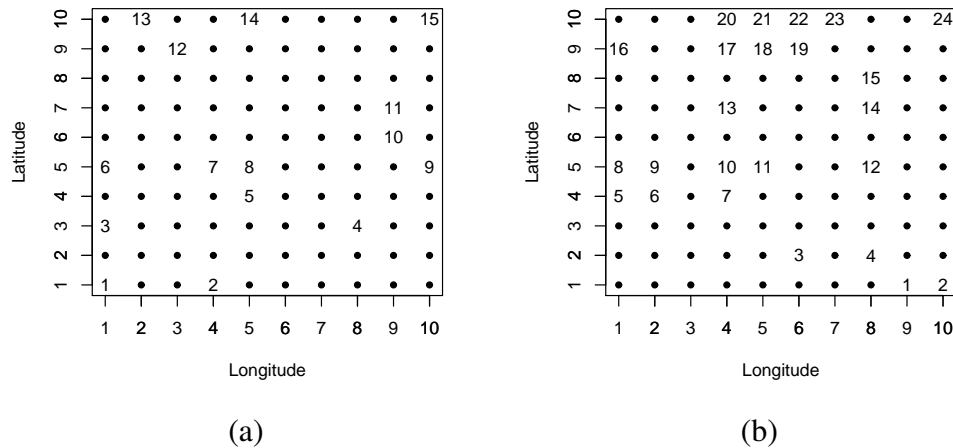


Figura 3.1 – Localizações observadas "•", e localizações não observadas enumeradas, considerando: (a) 15% dos pontos não amostrados e (b) 24% dos pontos não amostrados

Para cada um dos dois tamanhos de amostras considerados, em cada uma das 50 realizações foram ajustados os dois modelos:

1. o modelo (3.4), do qual os dados foram simulados, sendo as variáveis respostas o logaritmo do esforço e o logaritmo da captura e
2. o modelo (3.9). Neste caso, a variável resposta considerada foi  $Y = \log \frac{C}{E}$ , em que  $E$  representa o esforço e  $C$  a captura. Como os dados gerados são do logaritmo do esforço ( $Y_1$ ) e do logaritmo da captura ( $Y_2$ ), para poder ajustar o modelo (3.9) os dados simulados foram transformados da seguinte forma  $Y = \log \left( \frac{\exp(Y_2)}{\exp(Y_1)} \right)$ , em que  $Y_1$  e  $Y_2$  são os dados simulados, obtendo-se  $Y = \log \left( \frac{C}{E} \right)$ . O modelo foi ajustado para essa variável porque muitos trabalhos na literatura (NISHIDA; CHEN, 2004; MATSUNAGA; HOSONO; SHONO, 2006, entre outros), ao modelarem dados de esforço e captura, utilizam essa variável, o logaritmo da razão entre captura e esforço.

Foi utilizada a abordagem bayesiana para o ajuste dos modelos, de forma que a estimação dos parâmetros é feita através de métodos MCMC. Dessa forma, obtém-se uma amostra da distribuição *a posteriori* conjunta dos parâmetros. A partir da distribuição *a posteriori* e de propriedades da distribuição normal multivariada, foram obtidas amostras da distribuição preditiva da captura e



do esforço (quando se ajustou o modelo (3.4)) e amostras da distribuição preditiva da razão entre captura e esforço (quando se ajustou o modelo (3.9)) para os locais não observados.

Cada um dos conjuntos de valores preditos nos locais não observados juntou-se aos valores considerados como amostrados e, assim, obteve-se estimativas para o índice  $CPUE_1$ , obtendo-se uma amostra de estimativas para cada modelo ajustado. Considerou-se, então, uma estimativa pontual dada pela mediana da amostra. Obteve-se, assim, para cada uma das 50 realizações, uma estimativa pontual da  $CPUE_1$  obtida após o ajuste do modelo (3.4) e uma obtida após o ajuste do modelo (3.9). As estimativas obtidas após o ajuste dos modelos foram chamadas de estimativas ajustadas e as estimativas obtidas apenas de amostras dos dados simulados foram chamadas de estimativas amostrais.

As comparações das estimativas obtidas após o ajuste dos modelos (3.4) e (3.9) e, também das estimativas amostrais, foram feitas pelo erro quadrático médio (EQM). Para cada um dos 50 conjuntos de dados simulados é conhecido o valor verdadeiro do índice  $CPUE_1$ . Assim, para cada estimativa amostral ou cada estimativa obtida após o ajuste de um dos modelos ao conjunto de dados, foi possível obter o quadrado da diferença entre a estimativa da  $CPUE_1$  e seu valor verdadeiro. Tomou-se então a média dos 50 valores dados pelos quadrados dessas diferenças. Em cada cenário simulado e para cada tamanho de amostra considerado, foi obtido um valor de EQM para estimativas obtidas após o ajuste do modelo (3.4), um para estimativas obtidas após o ajuste do modelo (3.9) e um para a estimativa amostral. Esses EQM's foram comparados. O método cujas estimativas apresentaram menor EQM foi considerado melhor.

### 3.2.4 Resultados e Discussão

A Figura 3.2 apresenta os gráficos dos traços de duas cadeias *a posteriori*, geradas a partir de valores iniciais distintos, para os parâmetros do modelo (3.4), ajustado com tamanho de amostra 85, sob o cenário c). Os gráficos permitem verificar que existem indícios de convergência das cadeias. O mesmo pode ser notado na Figura 3.3, a qual se refere ao modelo (3.9), ajustado com tamanho de amostra 85 sob o cenário d). Para os ajustes dos modelos (3.4) e (3.9), com tamanho de amostra 76, também podem ser verificados indícios de convergência, sob o cenário c), através das Figuras 3.23 e 3.24 do Apêndice G.

Funções de autocorrelação e intervalos de credibilidade 95%, junto com os valores usados na simulação para os parâmetros do modelo (3.4) são mostrados no Apêndice E (para tamanho de amostra 85) e no Apêndice F para tamanho de amostra 76. Observa-se que praticamente todos os intervalos de credibilidade contêm os valores verdadeiros dos parâmetros. E para o modelo (3.9) são mostrados as funções de autocorrelação no Apêndice H. Para o modelo (3.9), não se tem os valores verdadeiros dos parâmetros, visto que os dados foram gerados do modelo (3.4). Assim, não é possível verificar se os intervalos 95% contêm os valores verdadeiros.

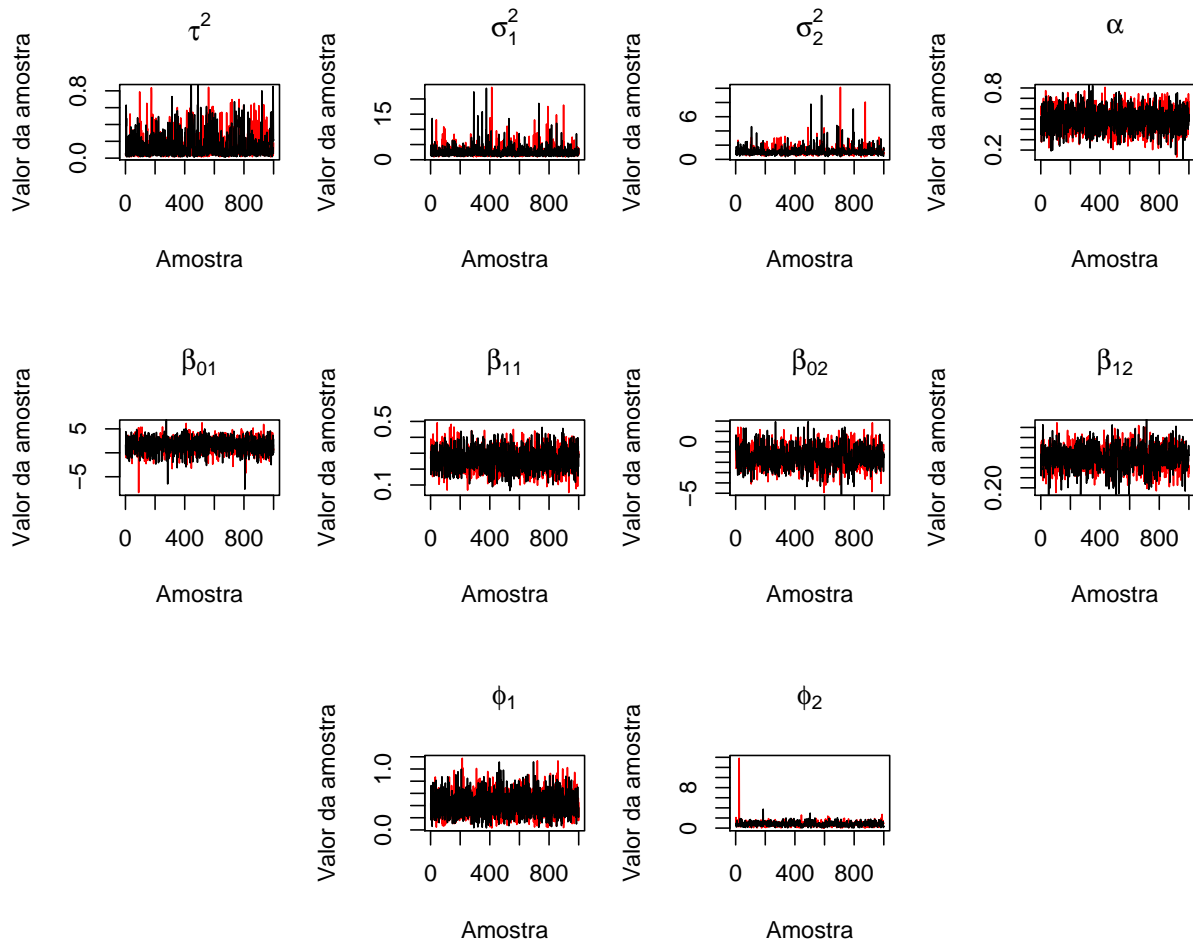


Figura 3.2 – Traço de duas cadeias *a posteriori* dos parâmetros sob o cenário c), representadas pelas cores vermelha e preta, partindo de valores iniciais distintos, modelo bivariado, ajuste com tamanho de amostra 85

Comparando os EQM's das estimativas da  $CPUE_1$ , obtidas após o ajuste do modelo univariado cuja variável resposta é o logaritmo da razão entre captura e esforço  $\frac{C}{E}$  (modelo (3.9)), com os EQM's das estimativas obtidas após o ajuste do modelo bivariado dado pelas equações (3.4), em que a variável resposta é o par  $(E, C)$ , observa-se na Tabela 3.1 que, apesar dos EQM's do modelo (3.4) serem menores para os 4 cenários, as diferenças entre os EQM's das estimativas obtidas após o ajuste dos dois modelos são muito pequenas. Isso pode ser notado também na Figura 3.26 do Apêndice I, a qual mostra que as estimativas obtidas após o ajuste dos dois modelos são bastante próximas. Em parte, o fato das estimativas obtidas após o ajuste dos dois modelos serem muito próximas se dá porque ambas as estimativas são obtidas por 85 pares de esforço e captura observados que são comuns, sendo somente 15 valores diferentes que são preditos pelos dois modelos. No entanto, os 15 valores preditos pelo modelo (3.4) não tornam as estimativas da  $CPUE_1$ , obtidas após

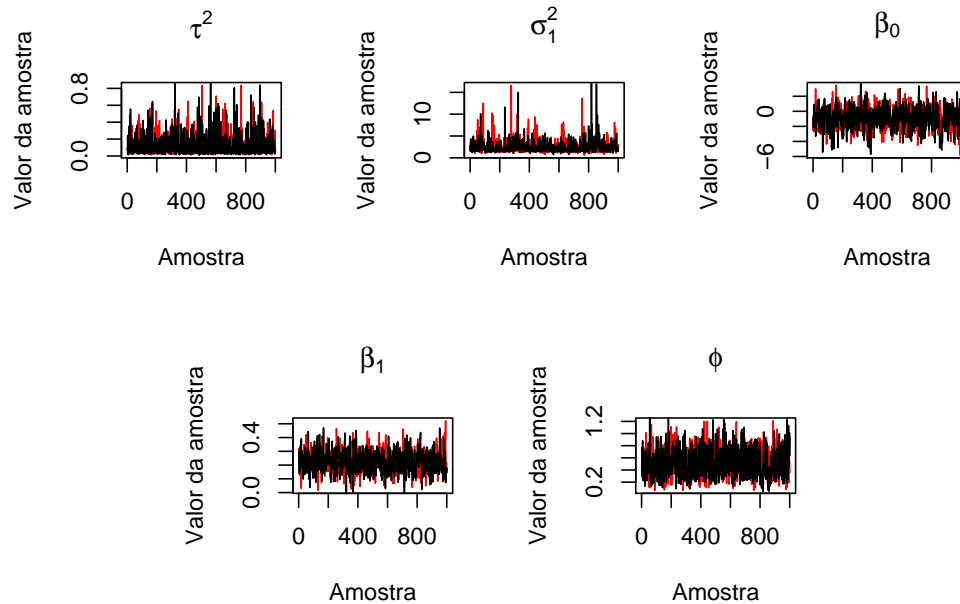


Figura 3.3 – Traço de duas cadeias *a posteriori* dos parâmetros sob o cenário d), representadas pelas cores vermelha e preta, partindo de valores iniciais distintos, modelo univariado, ajuste com amostra de tamanho 85

o ajuste desse modelo, melhores que as estimativas obtidas após o ajuste do modelo (3.9). Talvez as estimativas obtidas após o ajuste do modelo (3.4) não sejam melhores, pelo fato de que com esse modelo se faça predição de duas variáveis, esforço e captura, enquanto que com o modelo (3.9) se faz a predição de apenas uma variável. Dessa forma, tem-se a incerteza associada à predição do esforço e também da captura no caso do modelo (3.4), enquanto que no caso do modelo (3.9), tem-se a incerteza na predição de apenas uma variável (razão entre captura e esforço).

Tabela 3.1 – Erro Quadrático Médio das estimativas ajustada e amostral da  $CPUE_1$  para cada um dos cenários, usando os modelos bivariado e univariado, com tamanho de amostra 85

| Cenário | Ajustada   |           |          |
|---------|------------|-----------|----------|
|         | Univariado | Bivariado | Amostral |
| a)      | 35,4559    | 35,1504   | 32,6330  |
| b)      | 475,4903   | 473,4891  | 426,4459 |
| c)      | 5,2845     | 5,2285    | 3,5032   |
| d)      | 245,6568   | 242,7253  | 283,2965 |

Nota-se, também, que os intervalos de credibilidade 95% para as estimativas da  $CPUE_1$  (Figura 3.27 do Apêndice I e Tabela 3.2), têm amplitudes muito próximas para estimativas obtidas após o ajuste dos modelos (3.4) e (3.9)

Tabela 3.2 – Comprimento médio dos intervalos de credibilidade 95% da  $CPUE_1$  para cada um dos cenários

| Cenário | Univariado | Bivariado |
|---------|------------|-----------|
| a)      | 0,8665     | 0,8849    |
| b)      | 7,3706     | 8,9970    |
| c)      | 0,8080     | 0,8196    |
| d)      | 7,3711     | 7,1017    |

Na Figura 3.28 do Apêndice I é apresentado um gráfico com os valores verdadeiros da  $CPUE_1$ , versus as estimativas ajustadas para o modelo univariado e as estimativas amostrais, sob o cenário d). O gráfico mostra que as estimativas amostrais estão um pouco mais dispersas em relação aos valores verdadeiros da  $CPUE_1$  do que as estimativas ajustadas. Isto é, as estimativas ajustadas se concentram mais próximas da reta apresentada no gráfico. Isso pode ser confirmado pela Tabela 3.1, a qual mostra que para o cenário d), e somente para o cenário d), as estimativas ajustadas usando o modelo univariado são melhores que as amostrais. De fato, para esse cenário o resultado era esperado por se ter forte correlação espacial nos dados simulados, a qual é considerada na obtenção das estimativas ajustadas.

A Figura 3.29 do Apêndice I mostra alguns histogramas das estimativas ajustadas e suas respectivas medianas, considerando o modelo univariado, juntamente com a estimativa amostral e o valor verdadeiro da  $CPUE_1$ . Isso mostra que para as estimativas ajustadas tem-se, através do histograma, uma descrição da incerteza associada à estimação da  $CPUE_1$ . Além disso, intervalos de credibilidade para a  $CPUE_1$  podem ser obtidos diretamente das amostras das estimativas ajustadas.

Tabela 3.3 – Erro Quadrático Médio das estimativas ajustadas e amostrais da  $CPUE_1$  para cada um dos cenários, usando os modelos bivariado e univariado, com 24 pontos deixados para predição

| Cenário | Ajustada   |           |          |
|---------|------------|-----------|----------|
|         | Univariado | Bivariado | Amostral |
| a)      | 53,6231    | 53,5565   | 44,8562  |
| b)      | 1183,7940  | 918,5675  | 692,1623 |
| c)      | 24,8095    | 25,1901   | 17,0452  |
| d)      | 1397,8230  | 1322,8330 | 938,0567 |

Quando foram deixados mais pontos para se fazer a predição (24 pontos), ou seja, quando se ajustou os modelos usando tamanho de amostra 76, os EQM's obtidos das estimativas da  $CPUE_1$  (Tabela 3.3) aumentaram em relação aos EQM's obtidos das estimativas ajustadas com tamanho de amostra 85 (Tabela 3.1). Esse fato era esperado, visto que no primeiro caso, tem-se mais pontos para serem preditos, conseqüentemente maior incerteza na hora de fazer as predições.

Na Figura 3.32 são mostrados os intervalos de 95% de credibilidade para a  $CPUE_1$ , obtidos sob o cenário d). Nota-se que os intervalos de credibilidade obtidos para a  $CPUE_1$ , após o ajuste dos modelos univariado e bivariado com tamanho de amostra 76 (Figura 3.32), têm amplitudes semelhantes, mas são maiores que os intervalos obtidos após o ajuste desses modelos com tamanho de amostra 85 (Figura 3.27). De fato, as estimativas da  $CPUE_1$  que compõem o intervalo são calculadas com base em 76 valores amostrados (no caso do tamanho amostral 76), comuns a todas as estimativas e 24 valores preditos. No caso de tamanho amostral 85, as estimativas que compõem cada intervalo são calculadas com base em 85 valores comuns e 15 valores preditos. Dessa forma, as estimativas obtidas após o ajuste dos modelos com tamanho de amostra 85 são mais próximas entre si. Outro fato que contribui para que os intervalos obtidos após o ajuste dos modelos, com tamanho de amostra 76 sejam maiores, é a maior incerteza na predição dos pontos não observados, uma vez que se tem menos pontos vizinhos para ajudar a explicar a variável captura/esforço ou as variáveis captura e esforço nos locais não observados.

Na Figura 3.30 do Apêndice J são apresentadas os valores da  $CPUE_1$ , versus as estimativas ajustadas, com tamanho de amostra 76, usando os modelos univariado e bivariado, sob cenário d). Observa-se que, apesar das estimativas obtidas após o ajuste dos dois modelos estarem mais dispersas em relação à reta quando comparadas com a Figura 3.26, elas continuam próximas entre si, e consequentemente apresentam EQM's próximos (Tabela 3.3). Essa é uma indicação de que, para os cenários analisados, não há vantagem em usar o modelo bivariado em relação ao univariado quando o objetivo é obter uma estimativa da  $CPUE_1$ .

A Figura 3.31 do Apêndice J mostra os valores verdadeiros da  $CPUE_1$ , versus as estimativas ajustadas e as estimativas amostrais, com tamanho de amostra 76, usando o modelo univariado, sob cenário d). Comparando ainda as estimativas amostrais com as estimativas ajustadas, obtidas após o ajuste do modelo (3.9), nota-se que, para o tamanho de amostra 85, como já foi mencionado, as estimativas ajustadas são melhores para o cenário d). No entanto, com tamanho de amostra 76, nota-se na Figura 3.31 e na Tabela 3.3 que não há vantagem em usar estimativas ajustadas para a  $CPUE_1$ . Ou seja, os EQM's das estimativas ajustadas são maiores que os EQM's das estimativas amostrais. Usando o tamanho de amostra 76, para o ajuste do modelo (3.9), o número de pontos deixados para predição, 24, é grande o suficiente para que as predições feitas não sejam tão boas, e para que as estimativas ajustadas sejam melhores. Usar estimativas ajustadas pode ser útil para obter estimativas mais precisas da  $CPUE_1$ , porém é preciso cautela: principalmente quando se tem muitos pontos não observados, a má predição nos locais não observados pode levar a estimativas piores nesses casos.

### 3.2.5 Aplicação

Uma aplicação da metodologia foi feita utilizando dados de pesca de arrasto do camarão rosa *Farfantepenaeus brasiliensis* and *F. paulensis* desembarcado nas regiões Sudeste e Sul do Brasil no ano de 1998.

O conjunto de dados é composto pelas variáveis esforço em número de lances e captura em kg. Tem-se 65 pares de valores constituído por esforço e captura; para cada par tem-se um ponto de referência dado em graus de latitude e longitude, indicando um quadratim de  $10' \times 10'$ , onde ocorreu a pesca (Figura 3.4). Os valores das profundidades mínimas e máximas por local de pesca, também, foram disponibilizados e poderiam ser considerados como covariáveis, porém não foram considerados, pois não apresentaram qualquer relação com as variáveis esforço, captura e a razão entre captura e esforço. As coordenadas dadas em graus de latitude e longitude foram transformadas utilizando a projeção Universal Transversa de Mercator (UTM), antes da análise geoestatística dos dados. Dessa forma, pode-se calcular a distância euclideana entre os diferentes locais de pesca.

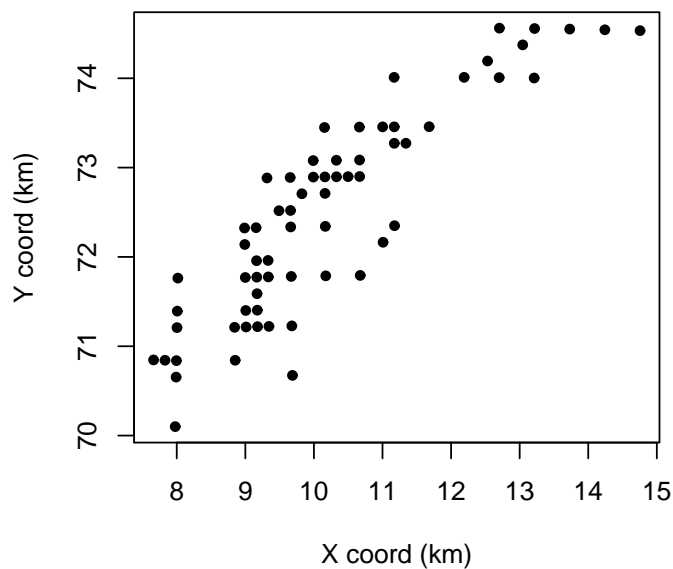


Figura 3.4 – Localização geográfica onde ocorreram as pescarias do camarão rosa

A seguir são apresentadas algumas análises descritivas dos dados considerando-se duas variáveis resposta (esforço e captura) e uma variável resposta (razão entre captura e esforço).

A fim de se obter normalidade para os dados, aplicou-se a transformação logarítmica nas variáveis esforço e captura. As Figuras 3.5 e 3.6 apresentam gráficos de dispersão e os histogramas das variáveis esforço e captura, após a transformação logarítmica. Nota-se que o logaritmo do esforço

e o logaritmo da captura não apresentam tendências com as coordenadas. Os histogramas das variáveis após a transformação não apresentam assimetrias. Para testar a normalidade, foi aplicado o teste de Shapiro-Wilk, o qual indicou normalidade para o logaritmo do esforço ( $W = 0.9703$ ,  $p = 0.1198$ ) e para o logaritmo da captura ( $W = 0.9756$ ,  $p = 0.2254$ ).

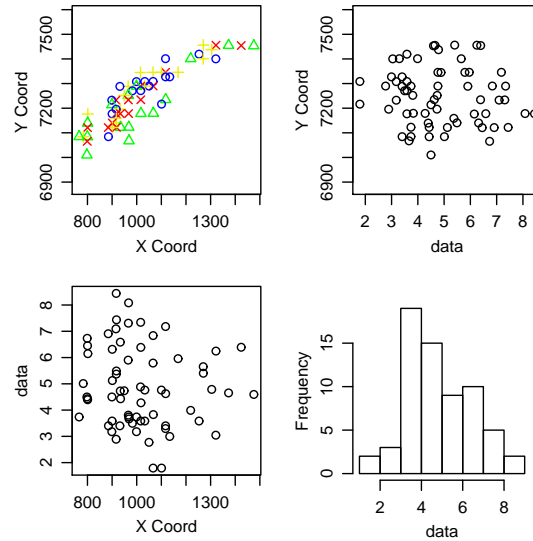


Figura 3.5 – Gráficos de dispersão e histograma para o logaritmo do esforço

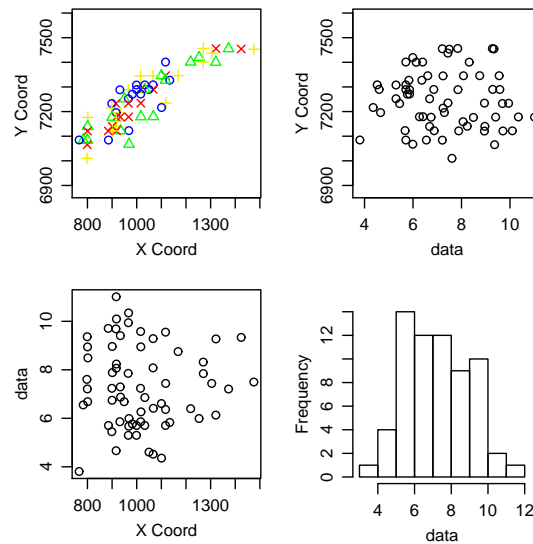


Figura 3.6 – Gráficos de dispersão e histograma para o logaritmo da captura

A Figura 3.7 mostra, através do gráfico de dispersão, a forte correlação linear existente entre o logaritmo do esforço e o logaritmo da captura, cujo valor estimado da correlação é de 0,9615.

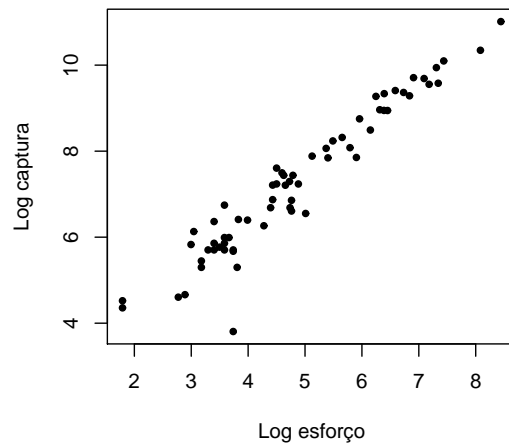


Figura 3.7 – Diagrama de dispersão para as variáveis log-esforço e log-captura

A fim de verificar a presença ou não de correlação espacial nos dados construiu-se os semivariogramas juntamente com o gráfico de envelope, para as variáveis logaritmo do esforço e logaritmo da captura (Figura 3.8). A região do envelope representa a região sem dependência espacial, e como os envelopes construídos para os semivariogramas do logaritmo do esforço e do logaritmo da captura contêm todos os pontos dos semivariogramas, não há evidências de correlação espacial.

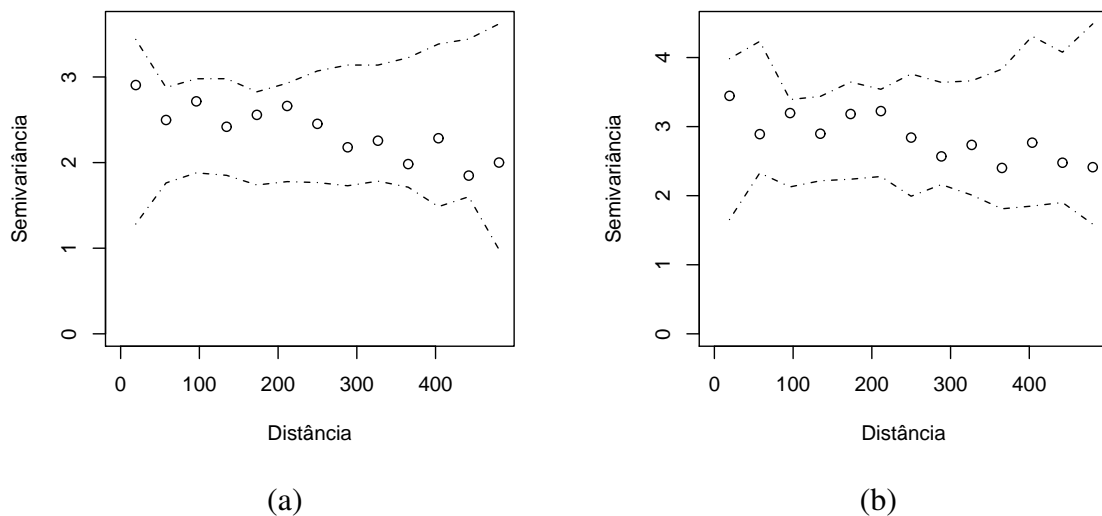


Figura 3.8 – Envelope e semivariograma das variáveis logaritmo do esforço (a) e logaritmo da captura (b)

Por essa análise descritiva, os dados da pesca do camarão rosa estão de acordo com o cenário



b) do estudo de simulação, onde se tem forte correlação entre o logaritmo do esforço e o logaritmo da captura, mas não se tem correlação espacial. Observa-se, também, que os dados de esforço e de captura seguem distribuição normal após a transformação logarítmica, o que está de acordo com as suposições feitas no estudo de simulação.

Considerando-se como variável resposta a razão entre a captura e o esforço, são apresentados na Figura 3.9 gráficos de dispersão e o histograma dessa variável. O histograma para essa variável é aparentemente simétrico e a variável aparentemente não apresenta tendência com as coordenadas. A fim de verificar se essa variável segue distribuição normal, aplicou-se o teste de Shapiro-Wilk, obtendo-se ( $W = 0,9896, p = 0.8634$ ), de forma que a variável pode ser considerada normalmente distribuída.

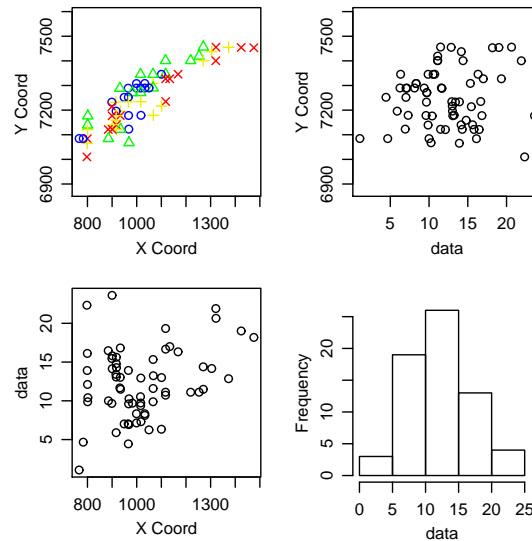


Figura 3.9 – Gráficos de dispersão e histograma para a variável captura/esforço

A variável dada pela razão entre captura e esforço também não apresentou evidências de correlação espacial, como mostra a Figura 3.10, em que o gráfico de envelope contém praticamente todos os pontos do semivariograma.

Os modelos foram ajustados com base nos dados observados nos pontos representados na Figura 3.4. Foram ajustados os modelos (3.4) e (3.9), sem considerar a covariável temperatura, uma vez que essa covariável não estava disponível para os dados da pesca do camarão rosa. Três outros modelos também foram ajustados: um modelo univariado eq. (3.12), cuja variável resposta é dada pela razão entre captura e esforço sem tomar o seu logaritmo.

$$\frac{C}{E}(s) = \beta + \sigma w(s) + \tau^2 u(s) \quad (3.12)$$

em que,  $\frac{C}{E}(s)$  representa a razão entre captura e esforço em uma localização  $s$  qualquer,  $w(\cdot)$  segue

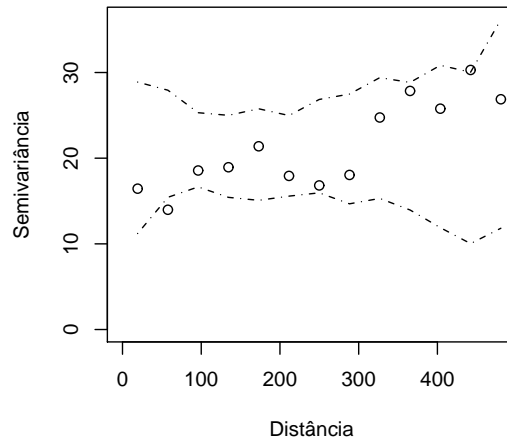


Figura 3.10 – Envelope e semivariograma da variável captura/esforço

um processo gaussiano com média 0 e variância 1 e  $u(s) \sim N(0, 1)$ . A função de correlação utilizada foi a exponencial  $\rho(d) = \exp(-\phi d)$ . Esse modelo foi ajustado, pois na análise descritiva da razão entre captura e esforço do camarão rosa, essa variável apresentou normalidade sem a transformação logarítmica. Como os dados da razão entre captura e esforço também não mostraram evidências de correlação espacial, foi ajustado um modelo sem a componente espacial dado pela eq. (3.13).

$$\frac{C}{E}(s) = \beta + \tau^2 u(s). \quad (3.13)$$

Ajustou-se ainda o modelo que considera a correlação entre esforço e captura, mas não leva em conta a correlação espacial, modelo (3.14):

$$\begin{aligned} Y_1(s) &= \beta'_{01} + \sigma'_1 u_1(s) \\ Y_2(s) | Y_1(s) &= \beta'_{02} + \alpha' Y_1(s) + \tau'_2 u_2(s). \end{aligned} \quad (3.14)$$

Como já descrito no Capítulo 1, esse modelo é um caso particular do modelo (3.4).

Após o ajuste de cada modelo foram obtidas as respectivas distribuições preditivas do índice CPUE<sub>1</sub>. Foi obtida também a estimativa amostral, utilizando-se as observações de captura e esforço nos pontos da Figura 3.4.

A Figura 3.11 mostra as funções de autocorrelação e a Figura 3.12 mostra os traços de duas cadeias *a posteriori* geradas a partir de valores iniciais distintos para os parâmetros do modelo (3.12). Por essas figuras, nota-se que existem indícios de convergência das cadeias. Foram omitidas essas figuras no caso dos parâmetros dos demais modelos ajustados, por apresentarem comportamentos similares, o que permite, também, constatar indícios de convergência.

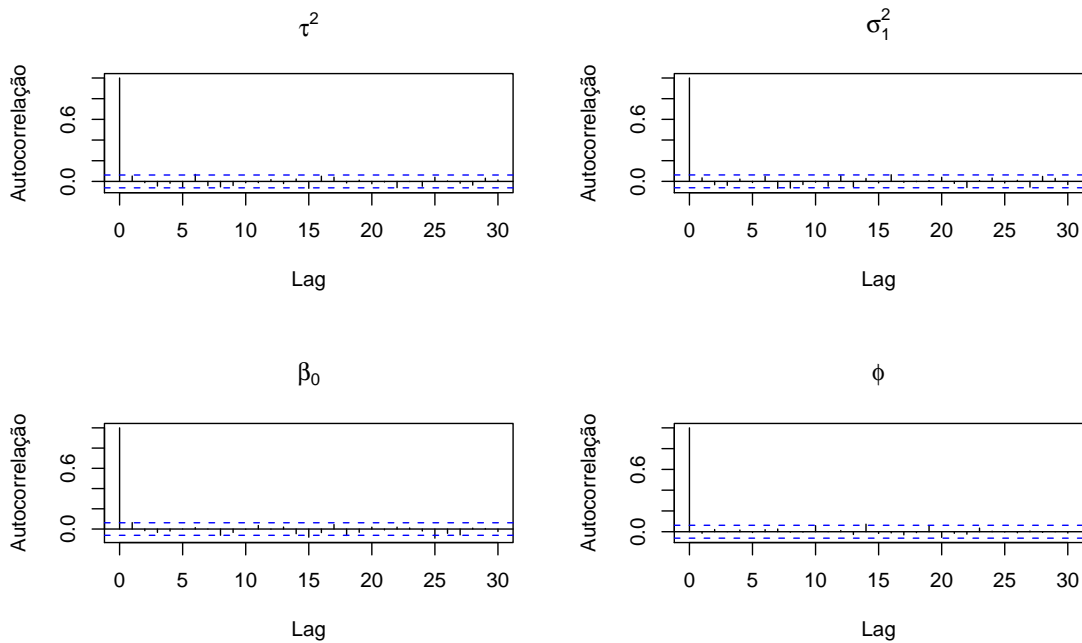


Figura 3.11 – Autocorrelação das cadeias *a posteriori*, para os parâmetros do modelo (3.9) ajustado aos dados da pesca do camarão rosa

Nas Tabelas 3.4, 3.5, 3.6, 3.7 e 3.8 são apresentados os resumos das distribuições *a posteriori* de todos os parâmetros dos modelos (3.4), (3.12), (3.9), (3.13) e (3.14) respectivamente.

Tabela 3.4 – Resumo da distribuição *a posteriori* dos parâmetros do modelo (3.4)

| Parâmetro    | Média   | 2,5%    | Mediana | 97,5%   |
|--------------|---------|---------|---------|---------|
| $\tau_2^2$   | 0,1111  | 0,0250  | 0,1036  | 0,2334  |
| $\sigma_1^2$ | 2,3964  | 1,7093  | 2,3635  | 3,2608  |
| $\sigma_2^2$ | 0,1546  | 0,0338  | 0,1343  | 0,4158  |
| $\alpha$     | 1,0547  | 0,9896  | 1,0535  | 1,1270  |
| $\beta_{01}$ | 4,8515  | 4,4650  | 4,8515  | 5,2190  |
| $\beta_{02}$ | 2,2135  | 1,6970  | 2,2175  | 2,7451  |
| $\phi_1$     | 30,7141 | 10,1382 | 24,9300 | 81,5277 |
| $\phi_2$     | 3,0112  | 0,0159  | 1,9870  | 11,9320 |

São apresentados na Figura 3.13 as distribuições preditivas da  $CPUE_1$ , obtidas após o ajuste dos modelos (3.4), (3.12), (3.9), (3.13) e (3.14), juntamente com a respectiva mediana e também a estimativa amostral. As distribuições preditivas da  $CPUE_1$  para os cinco modelos ajustados aos dados do camarão rosa estão centradas em pontos muito próximos, suas medianas estão muito próximas da estimativa amostral. O modelo (3.13) se mostrou o mais adequado para esses dados, uma vez que os dados de esforço e captura não apresentaram correlação espacial, e a razão entre

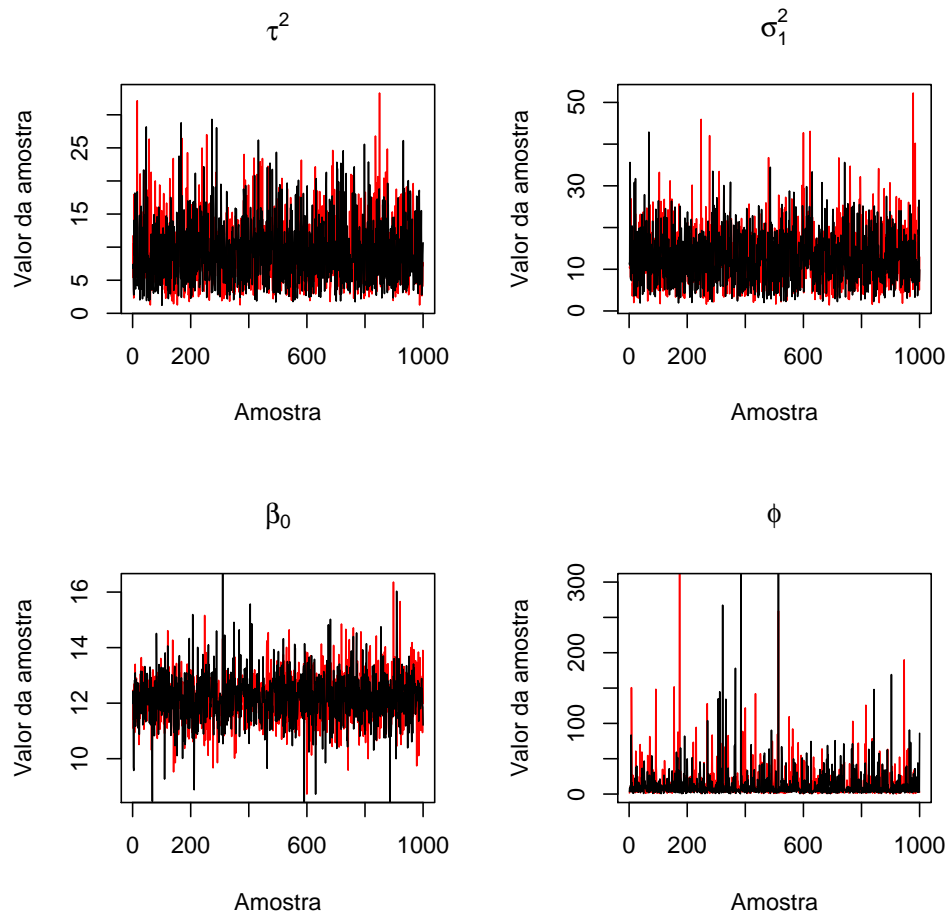


Figura 3.12 – Traço de duas cadeias *a posteriori*, representadas pelas cores vermelha e preta, partindo de valores iniciais distintos, para os parâmetros do modelo (3.9) ajustado aos dados da pesca do camarão rosa

Tabela 3.5 – Resumo da distribuição *a posteriori* dos parâmetros do modelo (3.12)

| Parâmetro    | Média   | 2,5%    | Mediana | 97,5%   |
|--------------|---------|---------|---------|---------|
| $\tau_2^2$   | 8,0965  | 2,4245  | 7,6307  | 16,7769 |
| $\sigma_1^2$ | 13,4593 | 3,9818  | 12,5055 | 28,0350 |
| $\beta_0$    | 12,5926 | 10,9300 | 12,5200 | 14,5605 |
| $\phi$       | 10,6072 | 0,9110  | 3,5725  | 76,3505 |

captura e esforço apresentou normalidade. Além disso, a distribuição preditiva da  $CPUE_1$ , após o ajuste do modelo (3.13), foi a que apresentou menor dispersão.

Tabela 3.6 – Resumo da distribuição *a posteriori* dos parâmetros do modelo (3.9)

| Parâmetro    | Média   | 2,5%   | Mediana | 97,5%   |
|--------------|---------|--------|---------|---------|
| $\tau_2^2$   | 80,0835 | 0,0220 | 0,0740  | 0,1949  |
| $\sigma_1^2$ | 0,1487  | 0,0381 | 0,1458  | 0,2991  |
| $\beta_0$    | 2,4450  | 2,2920 | 2,4410  | 2,6250  |
| $\phi$       | 8,5546  | 1,3719 | 4,7790  | 38,3972 |

Tabela 3.7 – Resumo da distribuição *a posteriori* dos parâmetros do modelo (3.13)

| Parâmetro  | Média   | 2,5%    | Mediana | 97,5%   |
|------------|---------|---------|---------|---------|
| $\tau_2^2$ | 20,6713 | 14,8469 | 20,3459 | 28,1533 |
| $\beta_0$  | 12,2209 | 11,1400 | 12,2100 | 13,3300 |

Tabela 3.8 – Resumo da distribuição *a posteriori* dos parâmetros do modelo (3.14)

| Parâmetro     | Média  | 2,5%   | Mediana | 97,5%  |
|---------------|--------|--------|---------|--------|
| $\tau_2'^2$   | 0,2198 | 0,1553 | 0,2135  | 0,3053 |
| $\sigma_1'^2$ | 2,3524 | 1,6781 | 2,2973  | 3,2693 |
| $\alpha'$     | 1,0585 | 0,9809 | 1,0587  | 1,1323 |
| $\beta_{01}'$ | 4,8578 | 4,4892 | 4,8487  | 5,2628 |
| $\beta_{02}'$ | 2,1334 | 1,7479 | 2,1351  | 2,5225 |

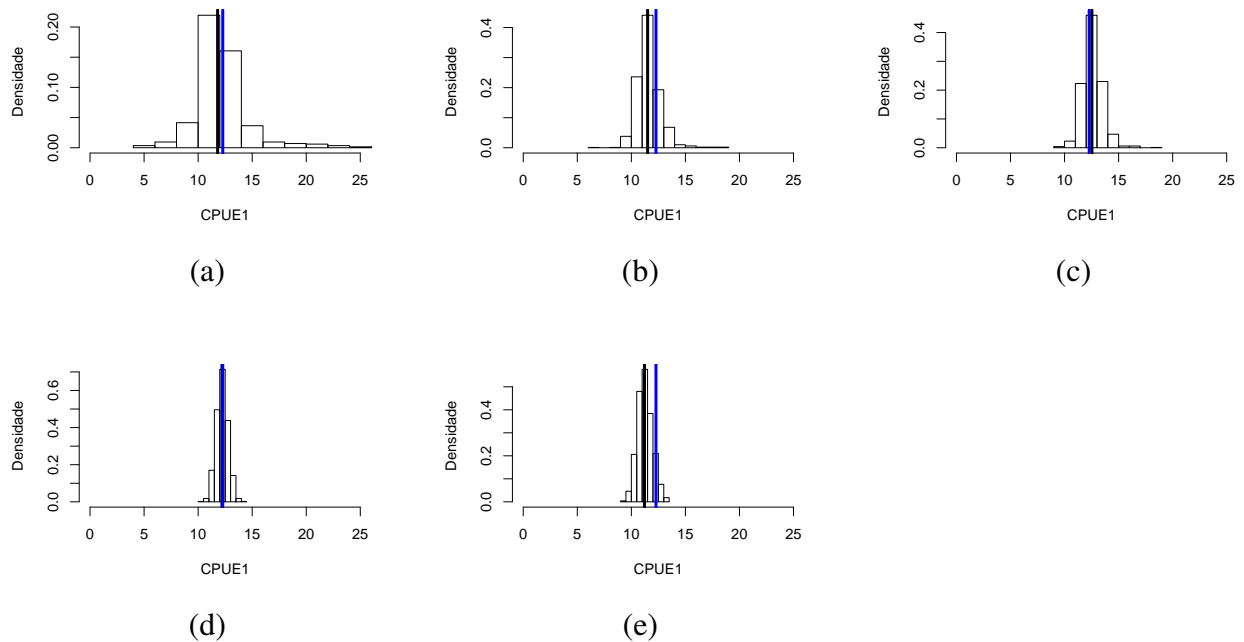


Figura 3.13 – Histogramas das distribuições preditivas do índice  $CPUE_1$ , considerando os modelos: (a) (3.4), (b) (3.9), (c) (3.12), (d) (3.13) e (e) (3.14), juntamente com a respectiva mediana (reta em preto) e a estimativa amostral (em azul)

### 3.3 Considerações Finais

O estudo de simulação realizado permitiu verificar que:

- a) As estimativas da  $CPUE_1$  obtidas após o ajuste dos modelos (3.4) e (3.9), em geral, apresentaram valores muito próximos, indicando que, para os cenários analisados, não há grande vantagem em usar o modelo (3.4) em relação ao modelo (3.9), quando o objetivo é a estimação da  $CPUE_1$ .
- b) As estimativas ajustadas usando o modelo (3.9), são melhores que as estimativas amostrais, no cenário d), quando 15% das observações foram deixadas para predição.
- c) Deixando mais pontos para predição, 24%, não há vantagem em usar estimativas ajustadas em relação às estimativas amostrais para a  $CPUE_1$ . Portanto, recomenda-se cautela ao usar estimativas ajustadas para a  $CPUE_1$  quando há muitos pontos a serem preditos na região em estudo.

Possíveis extensões deste trabalho são:

- (i) Usar modelos que considerem o tempo, visto que o interesse do pesquisador está em verificar o comportamento do índice de abundância ao longo do tempo;
- (ii) Usar modelos para tratar dados de esforço e captura inflacionados de zeros.

### Referências

- ANGANUZZI, A. A. **An aggregate model of effort distribution for the eastern Pacific tuna fishery**. Roma: FAO, Fisheries and Aquaculture Department, 2004. 8 p. Technical report.
- BOGNOLA, I.; RIBEIRO JR, P.; SILVA, E.; LINGNAU, C.; HIGA, A. Moelagem uni e bivariada da variabilidade espacial de rendimento de *pinus taeda* L. **Floresta**, Curitiba, v. 38, n. 2, p. 373-385 abr./jun. 2008.
- BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. **Hierarchical Modeling and Analysis for Spatial Data**. London: Chapman & Hall/CRC, 2004. 452 p.
- CADIMA, E., L. **Manual de avaliação de recursos pesqueiros**. Roma: FAO, 2000. 162p.
- COMISSÃO INTERNACIONAL PARA A CONSERVAÇÃO DOS ATUNS DO ATLÂNTICO. Banco de dados. Disponível em: <<http://www.iccat.int/accesingdb.htm>>. Acesso em: 17 abr. 2007.

CRESSIE, N.A.C. **Statistics for Spatial Data**. Revised Edition. New York: John Wiley and Sons, 1993. 900 p.

DIGGLE, P. J.; RIBEIRO JR. P. J. **Model-based Geostatistics**. Lancaster: Springer Series in Statistics, 2007. 230 p.

DIGGLE, P. J. **Statistical analysis of spatial point patterns**. 2.ed. London: Arnold, 2003. 159 p.

GAMERMAN. D.; LOPES. H. F. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. 2.ed. London: Chapman & Hall, 2006. 323 p.

GELMAN, A.; CARLIN, J.B.; STERN, H.S.; RUBIN, D.B. *Bayesian data analysis*. 2. ed. London: Chapman and Hall, 2003. 668 p.

GELFAND, A.E.; SCHMIDT, A.M.; BANERJEE, S.; SIRMANS, C.F. Nonstationary Multivariate Process Modeling through Spatially Varying Coregionalization (with discussion). **Test**, Valencia, v. 13, n. 2, p. 263-312, Dec. 2004.

HILBORN, R.; WALTERS, C.J. **Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty**. New York: Chapman and Hall, 1992. 570 p.

JARDIM, E.; RIBEIRO JR, P.J. Geoestatistical assessement of sampling designs for Portuguese botton trawl surveys (submetido), 2006.

MATSUNAGA, H.; HOSONO, T.; SHONO H. **Analysis of longline CPUE of major pelagic shark species collected by Japanese research and training vessels in the Pacific Ocean**. National Research Institute of Far Seas Fisheries, FRA, JAPAN, 2006. 8 p. Relatório Técnico.

MAUNDER, M. N.; PUNT, A. E. Standardizing catch and effort data: a review of recent approaches. **Fisheries Research**, Aberdeen , v. 70, p. 141-159, 2004.

MAYNOU, F. The application of geostatistics in mapping and assessment of demersal resources. *Nephrops norvegicus* (L.) in the northwestern Mediterranean: a case study. **Scientia Marina**, Barcelona, v. 62, p. 117-133, 1998.

MAYNOU, F.; DEMESTRE, M.; SANCHEZ, P. Analysis of catch per unit effort by multivariate analysis and generalised linear models for deep-water crustacean fisheries off Barcelona (NW Mediterranean). **Fisheries Research**, Aberdeen, v. 65, p. 257-269, 2003.

NADAL-EGEA, J. A. **Esfuerzo y Captura: Tecnología y Sobreexplotación de Recursos Marinos Vivos**. México: El Colegio de México, 1996. 476 p.

NISHIDA, T.; CHEN, D, G. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data. **Fisheries research**, Aberdeen, v. 70, p. 265-274, 2004.

O'HANGAN, A. **Kendall's Advanced Theory of Statistics: Bayesian Inference**, London: John Wiley & Sons, 1994. v. 2B, 448 p.

PAEZ, M. S.; GAMERMAN, D.; De OLIVEIRA, V. Interpolation performance of a spatio-temporal model with spatially varying Coefficients: Application to PM10 Concentration in Rio de Janeiro. **Environmental and Ecological Statistics**, New York, v. 12, p. 169-193, 2005.

PETITGAS, P. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. **Fish and fisheries**, Aberdeen, v. 2 , p. 231-249, 2001.

PETITGAS, P. Geostatistics and their applications to fisheries survey data. In: MEGREY, A.; MOKSNESS, E. **Computers in Fisheries Research**. London: Chapman & Hall, 1996. p. 113 – 142.

PETRERE JR, M.; GIACOMINI, H.C.; DE MARCO JR. P. **Catch-per-unit effort: which estimator is best?** Rio Claro: UNESP, Departamento de Ecologia, 2007. 17 p. Relatório Técnico.

REIS, E. A. **Análise espacial de taxas em pequenas áreas usando a metodologia Bayesiana empírica**. 1998. 80 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Minas Gerais, Belo Horizonte, 1998.

REIS, I. A. **Testando a aleatoriedade espacial de uma configuração de pontos**. 1998. 43 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Minas Gerais, Belo Horizonte, 1998.



RIVOIRARD, J.; SIMMONDS, E. J.; FOOTE, K. G.; FERNANDES, P. G.; BEZ, N. **Geostatistics for Estimating Fish Abundance**. Oxford: Blackwell Science, 2000. p. 206.

RUFINO, M.; STELZENMULLER, V.; MAYNOU, F.; ZAUKE, P. Assessing the performance of linear geostatistical tools applied to artificial fisheries data. **Fisheries research**, Aberdeen, v. 82, p. 263-279, jun, 2006.

SCHMIDT, A. M.; NOBRE, A. A.; FERREIRA, G. S. Alguns Aspectos da Modelagem de Dados Espacialmente Referenciados. **Revista brasileira de estatística**, Brasil, v. 63, n. 220, p. 59-88, 2003.

SCHMIDT, A. M.; SANSÓ, B. **Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço-Temporais**, Caxambu: ABE, 2006. 151 p.

SCHMIDT, A.M.; GELFAND, A. E. A bayesian coregionalization approach for multivariate pollutant data. **Journal of Geophysics Research - Atmospheres**, Washington, v. 108, no. D24, 8783. 2003.

SPIEGELHATER, D. J.; THOMAS, A; BEST, N.G. **WinBugs version 1.4 user manual**. **Technical Report**. Cambridge: Medical Research Council Biostatistics, 2002.

STELZENMULLER, V.; MAYNOU, F.; MARTÍN, P. Spatial assessment of benefits of a coastal Mediterranean Marine protected area. **Biological conservation**, New York, v. 36, p. 571-583, 2007.

VIGNAUX, M. Analysis of spatial structure in fish distribution using commercial catch and effort data from the New Zealand hoki fishery. **Canadian Journal of Fisheries and Aquatic Sciences**, Aberdeen, v. 53, p. 963-973, 1996.

WALTERS, C. Folly and fantasy in the analysis of spatial catch rate data. **Canadian Journal of Fisheries and Aquatic Sciences**, Toronto, v.60, p. 1433-1436, Dec. 2003.

## APÊNDICES

### APÊNDICE A - Funções de Autocorrelação Intervalos de Credibilidade

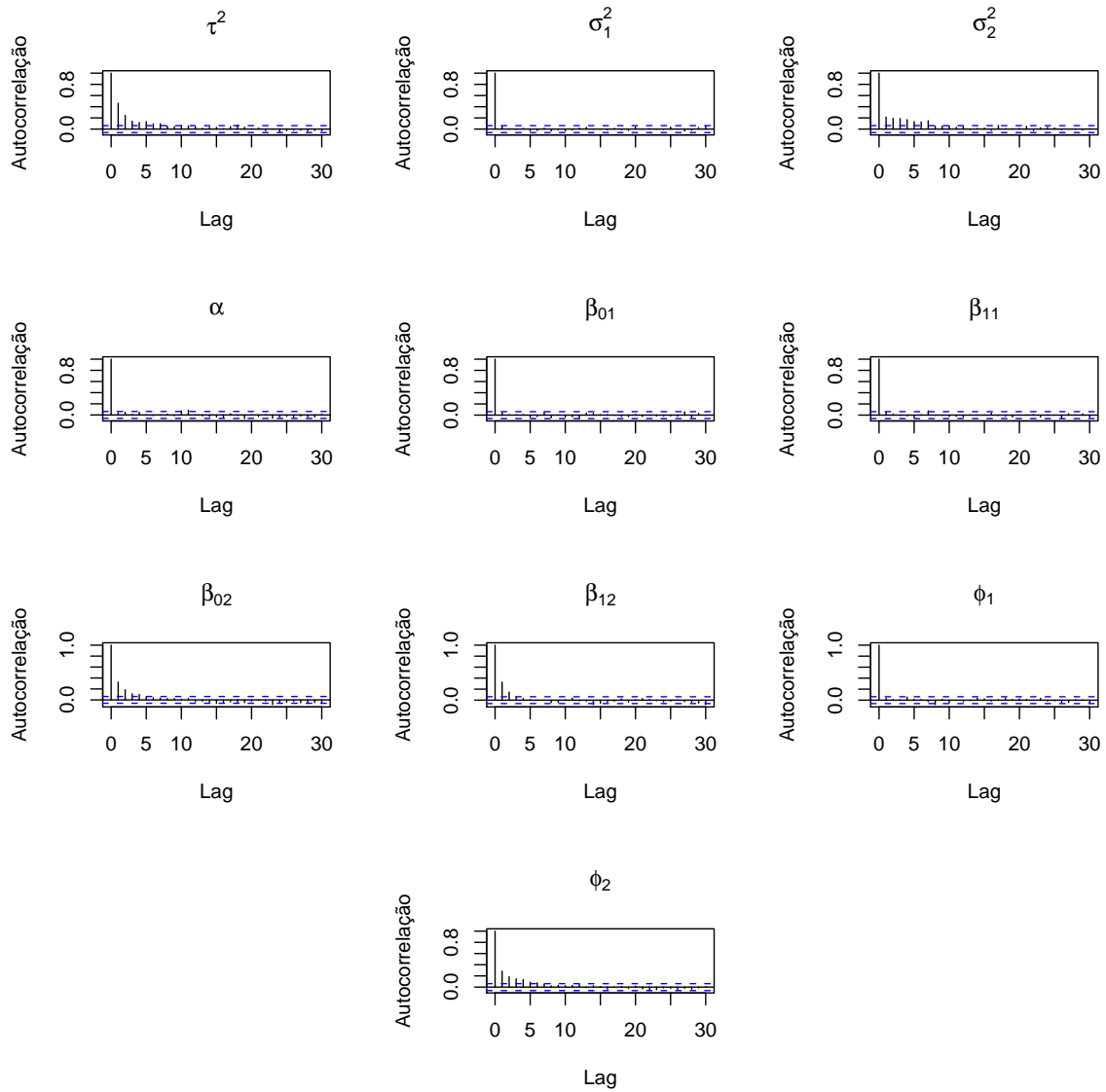


Figura 3.14 – Autocorrelação das cadeias *a posteriori* para os parâmetros usados na simulação dos dados, sob o cenário d)

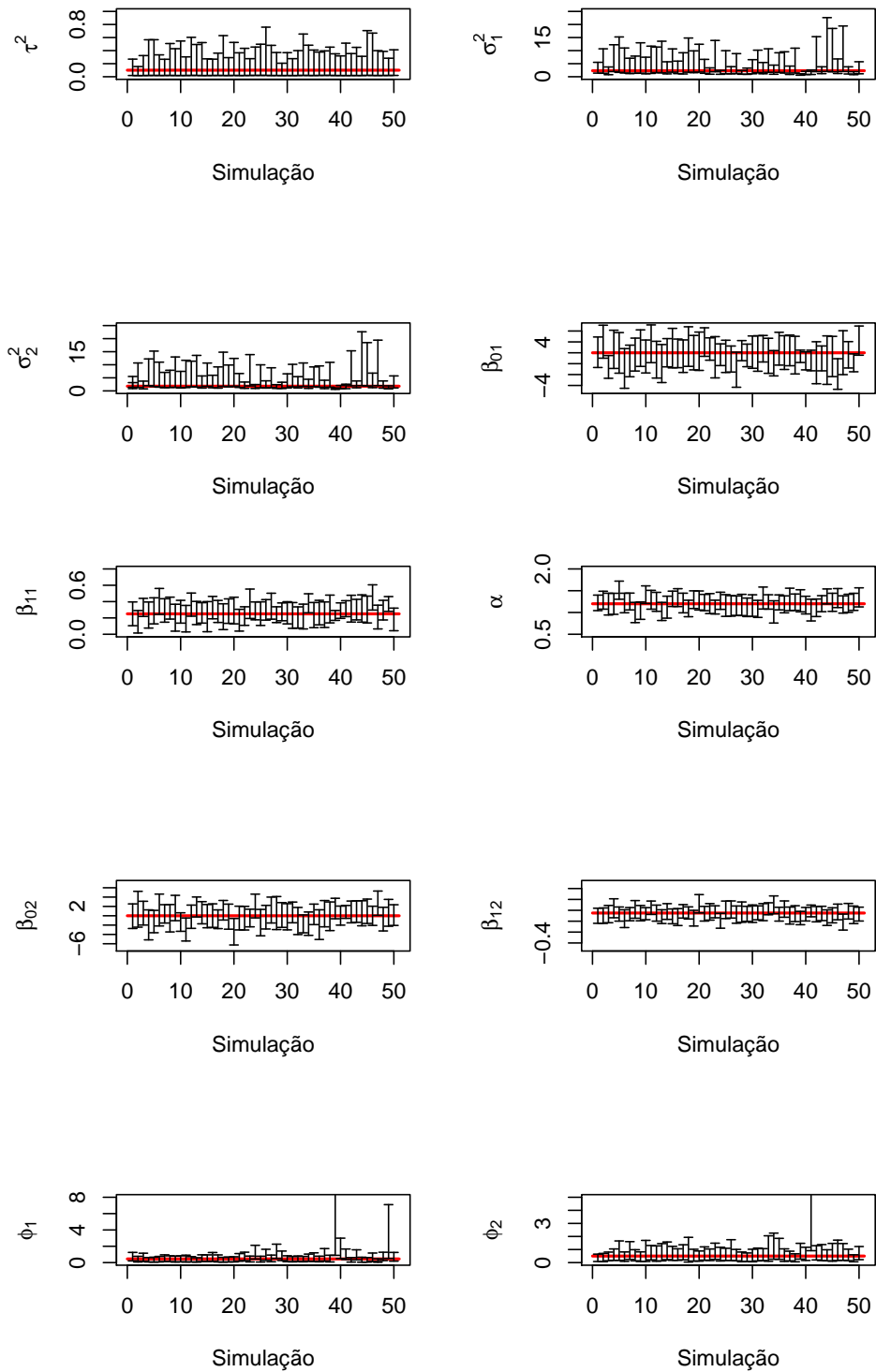


Figura 3.15 – Intervalo de credibilidade de 95% dos parâmetros e valores usados na simulação dos dados (reta em vermelho), sob o cenário d)

## APÊNDICE B

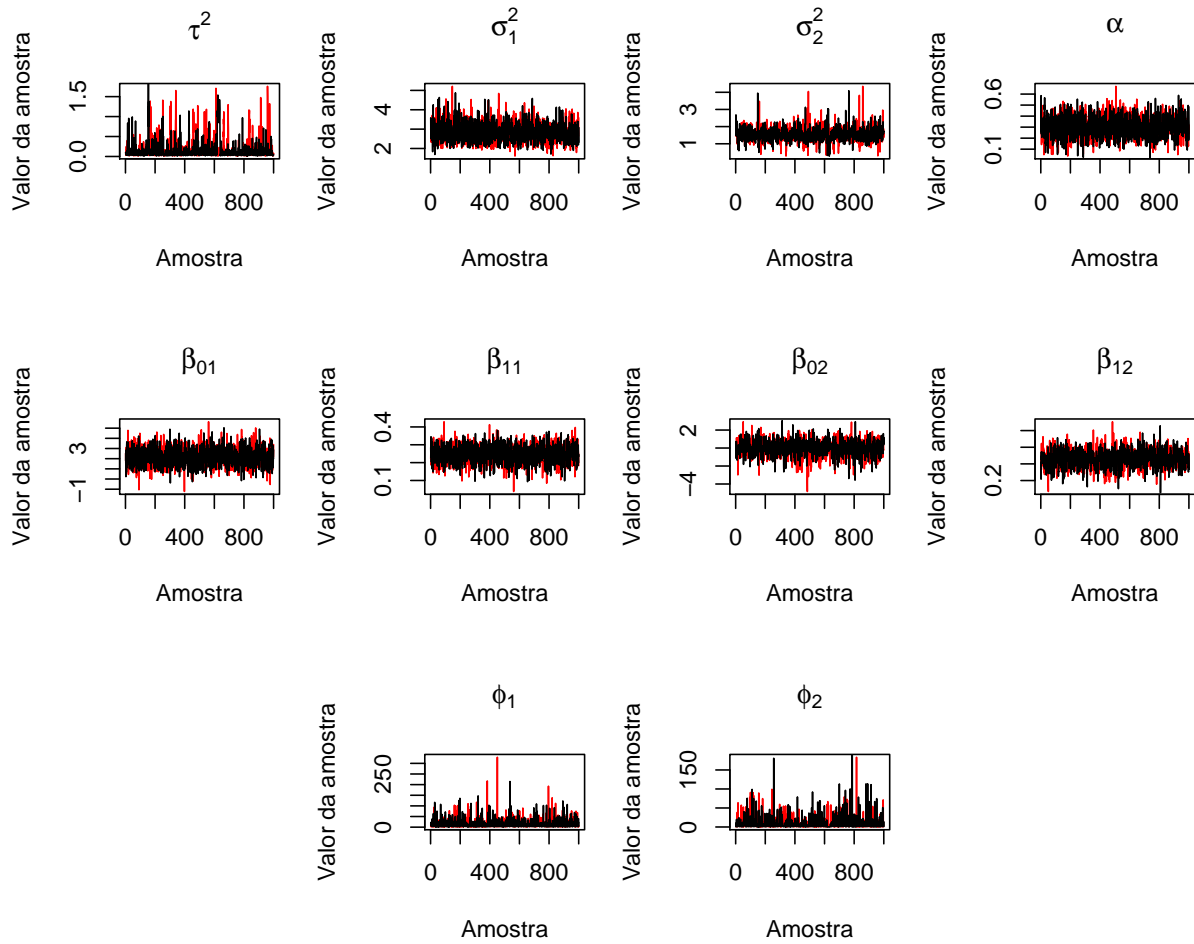


Figura 3.16 – Traço de duas cadeias *a posteriori* dos parâmetros sob o cenário a), representadas pelas cores vermelha e preta, partindo de valores iniciais distintos

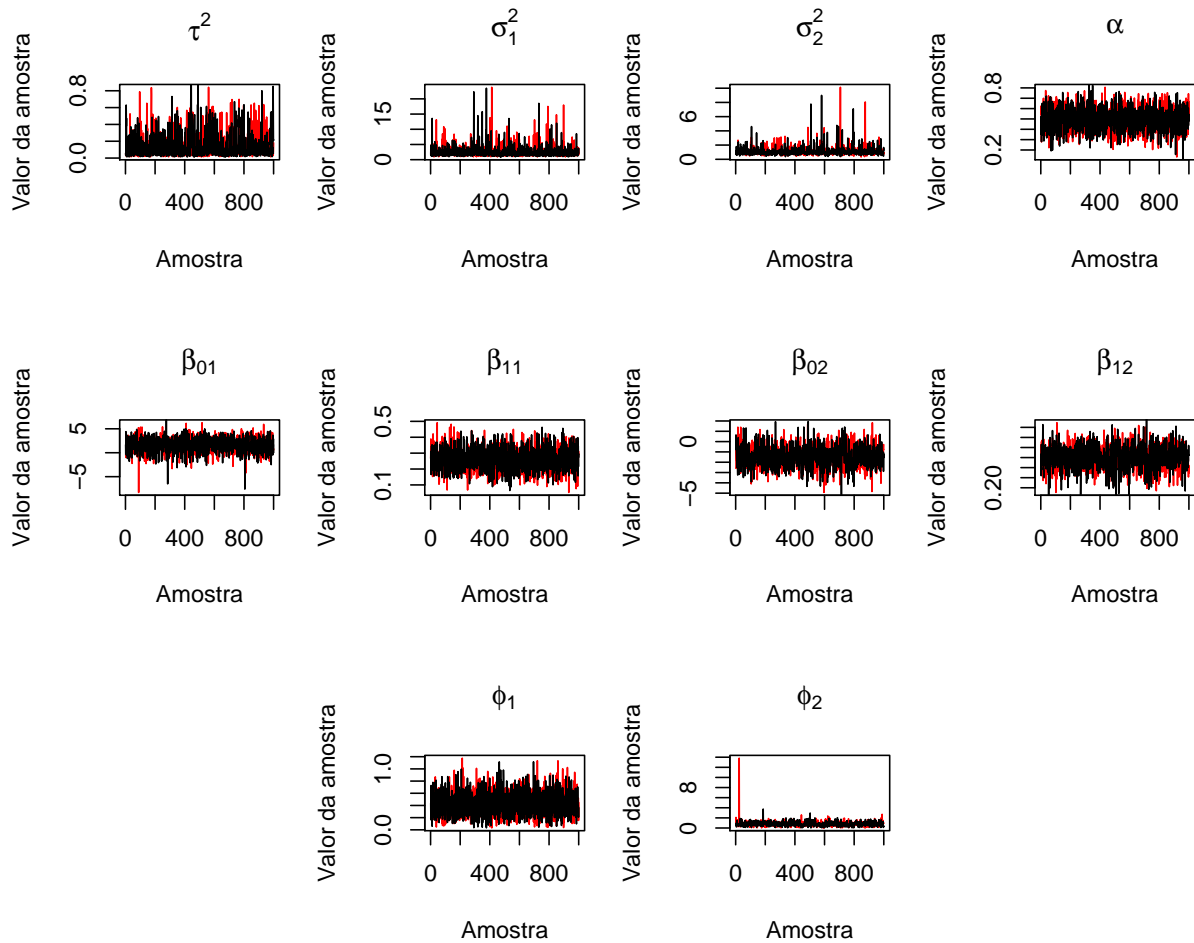


Figura 3.17 – Traço de duas cadeias *a posteriori* dos parâmetros sob o cenário c), representadas pelas cores vermelha e preta, partindo de valores iniciais distintos

## APÊNDICE C

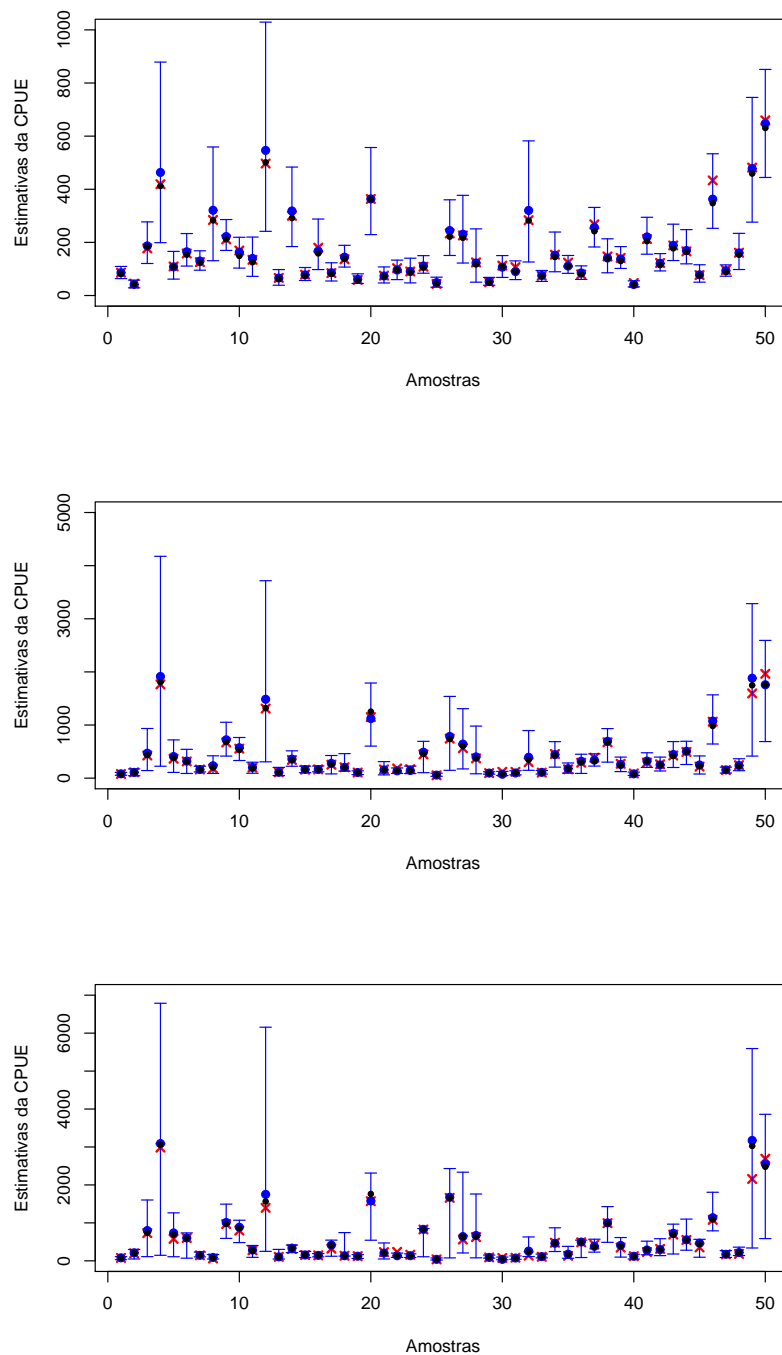


Figura 3.18 – Estimativas ajustadas (em preto), amostrais e intervalo bootstrap 95%(em azul) e valores verdadeiros (em vermelho) das  $CPUE_1$ ,  $CPUE_2$  e  $CPUE_3$ . Cenário d), modelo (3.4)

**APÊNDICE D - Distribuições condicionais completas *a posteriori* dos parâmetros do modelo condicional para esforço e captura**

1.  $\beta_1 \mid \sigma_1^2, \phi_1, \mathbf{y}_1 \sim N(\mathbf{B}\mathbf{b}, \mathbf{B}),$

em que  $\mathbf{B} = [\mathbf{X}'_1(\sigma_1^2\mathbf{R}_1)^{-1}\mathbf{X}_1 + \Sigma_{\beta_1}^{-1}]^{-1}$  e  $\mathbf{b} = \mathbf{X}'_1(\sigma_1^2\mathbf{R}_1)^{-1}\mathbf{y}_1 + \Sigma_{\beta_1}^{-1}\boldsymbol{\mu}_{\beta_1};$

2.  $\sigma_1^2 \mid \beta_1, \phi_1, \mathbf{y}_1 \sim IG(\frac{n}{2} + a_{\sigma_1}, c_{\sigma_1}),$

em que  $c_{\sigma_1} = \frac{1}{2}[\mathbf{y}_1 - \mathbf{X}_1\beta_1]^T \mathbf{R}_1^{-1}[\mathbf{y}_1 - \mathbf{X}_1\beta_1];$

3.  $\pi(\phi_1 \mid \beta_1, \sigma_1^2, \mathbf{y}_1) \propto |\mathbf{R}_1|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_1 - \mathbf{X}_1\beta_1]^T [\sigma_1^2\mathbf{R}_1]^{-1} [\mathbf{y}_1 - \mathbf{X}_1\beta_1] - b_{\phi_1}\phi_1 \right\} \phi_1^{a_{\phi_1}-1};$

4.  $\beta_2 \mid \sigma_2^2, \alpha, \tau_2^2, \phi_2, \mathbf{y}_1, \mathbf{y}_2 \sim N(\mathbf{B}_2\mathbf{b}_2, \mathbf{B}_2),$

em que  $\mathbf{B}_2 = [\mathbf{X}'_2(\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I})^{-1}\mathbf{X}_2 + \Sigma_{\beta_2}^{-1}]^{-1}$  e  $\mathbf{b}_2 = \mathbf{X}'_2(\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I})^{-1}\mathbf{y}_2 + \Sigma_{\beta_2}^{-1}\boldsymbol{\mu}_{\beta_2};$

5.  $\pi(\phi_2 \mid \beta_2, \sigma_2^2, \tau_2^2, \mathbf{y}_1, \mathbf{y}_2) \propto$

$|\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\beta_2]'[\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\beta_2] - b_{\phi_2}\phi_2 \right\} \phi_2^{a_{\phi_2}-1};$

6.  $\pi(\sigma_2^2 \mid \beta_{02}, \tau_2^2, \phi_2, \mathbf{y}_1, \mathbf{y}_2) \propto$

$|\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\beta_2]'[\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\beta_2] - \frac{b_{\sigma_2}}{\sigma_2^2} \right\} (\sigma_2^2)^{-a_{\sigma_2}-1};$

7.  $\pi(\tau_2^2 \mid \beta_2, \sigma_2^2, \phi_2, \mathbf{y}_1, \mathbf{y}_2) \propto$

$|\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}[\mathbf{y}_2 - \mathbf{X}_2\beta_2]'[\sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I}]^{-1}[\mathbf{y}_2 - \mathbf{X}_2\beta_2] - \frac{b_{\tau_2}}{\tau_2^2} \right\} (\tau_2^2)^{-a_{\tau_2}-1}.$



### APÊNDICE E - Funções de Autocorrelação Intervalos de Credibilidade

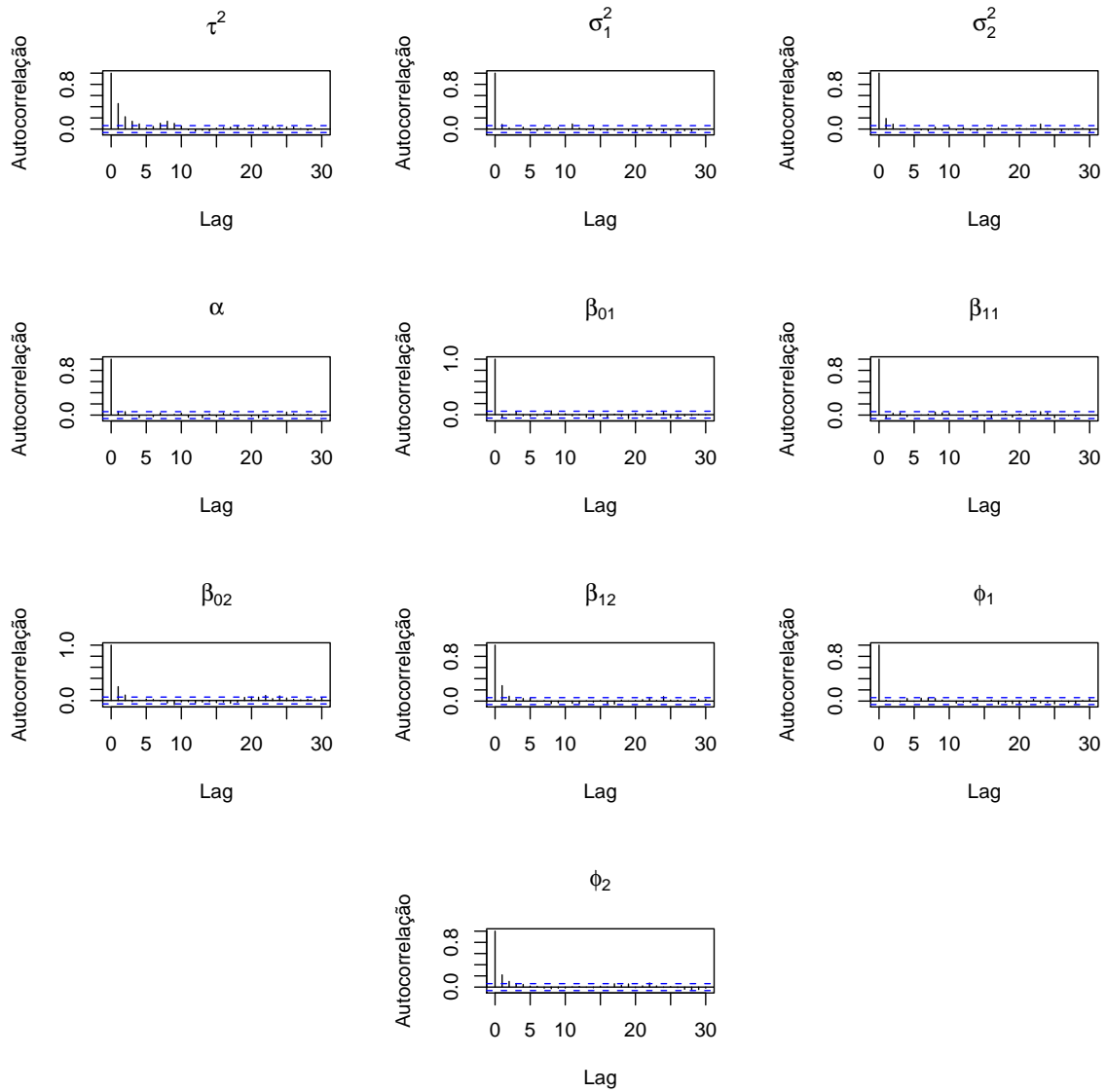


Figura 3.19 – Autocorrelação das cadeias *a posteriori* para os parâmetros do modelo bivariado, sob o cenário c, ajuste com amostra de tamanho 85

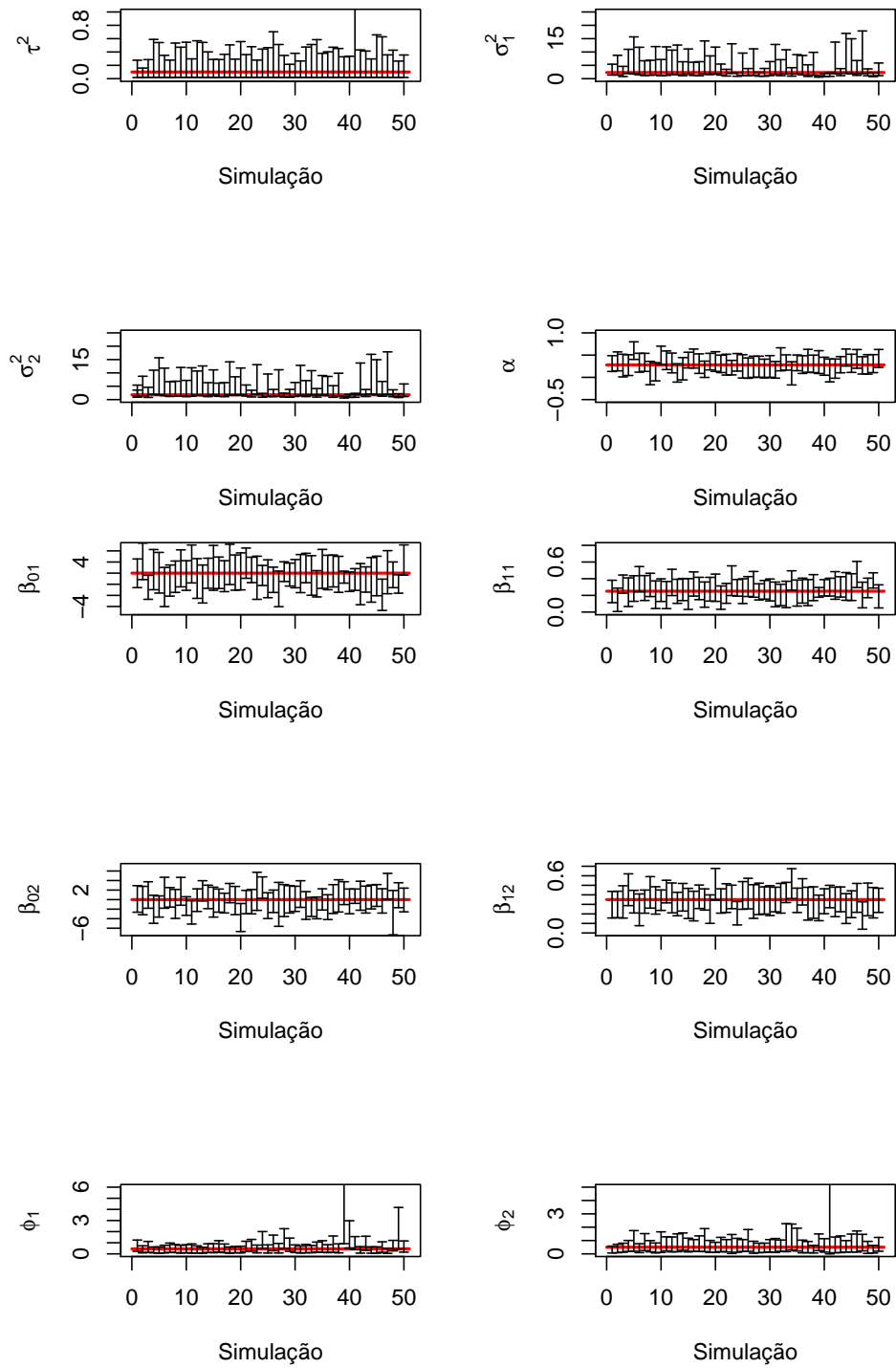


Figura 3.20 – Intervalo de credibilidade de 95% dos parâmetros e valores usados na simulação dos dados (reta em vermelho), sob o cenário c), modelo bivariado ajuste com amostra de tamanho 85

**APÊNDICE F - Funções de Autocorrelação Intervalos de Credibilidade)**

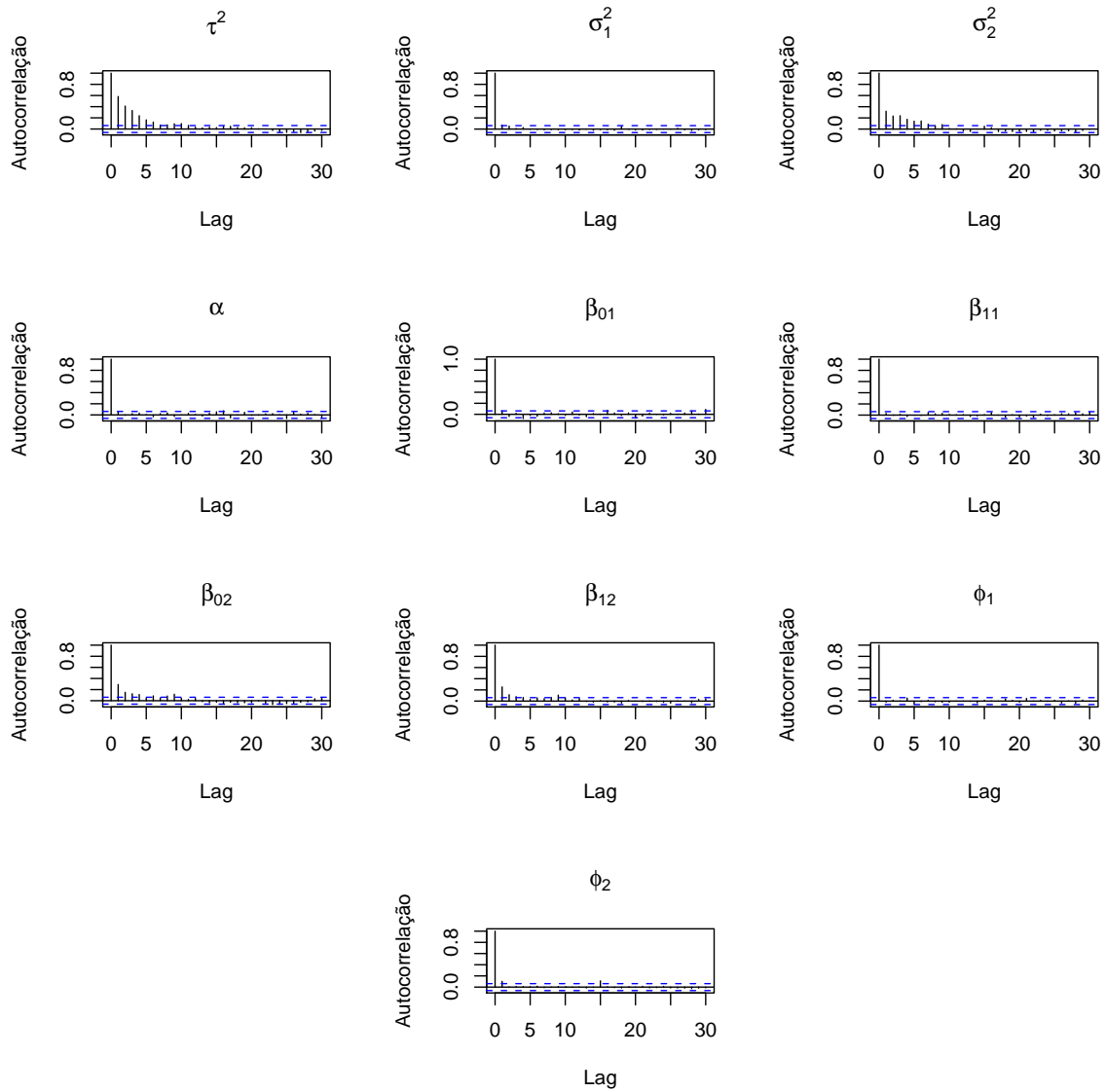


Figura 3.21 – Autocorrelação das cadeias *a posteriori* para os parâmetros do modelo bivariado, sob o cenário d, ajuste com amostra de tamanho 76

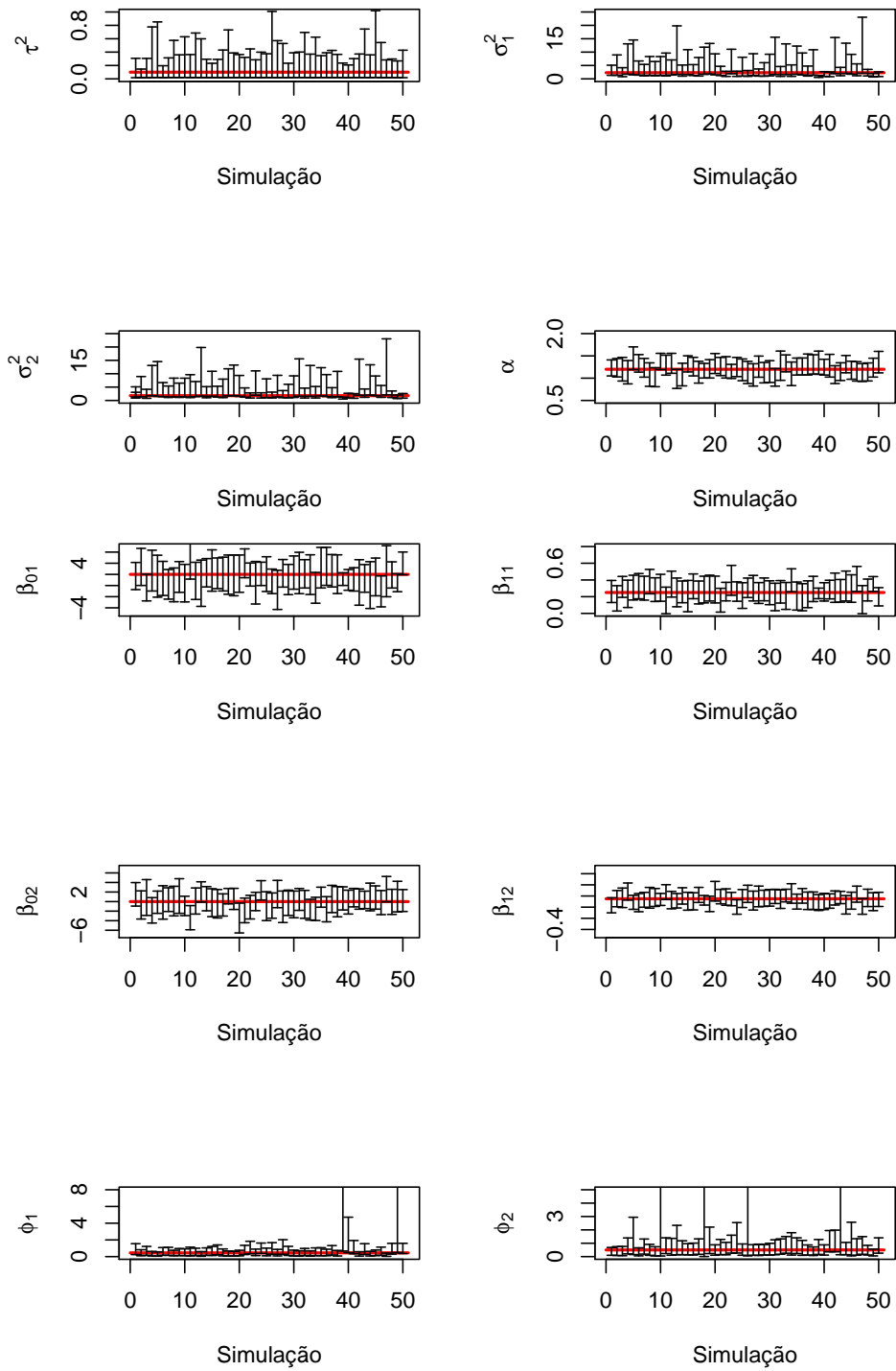


Figura 3.22 – Intervalo de credibilidade de 95% dos parâmetros e valores usados na simulação dos dados (reta em vermelho), sob o cenário d), modelo bivariado ajuste com amostra de tamanho 76

## APÊNDICE G

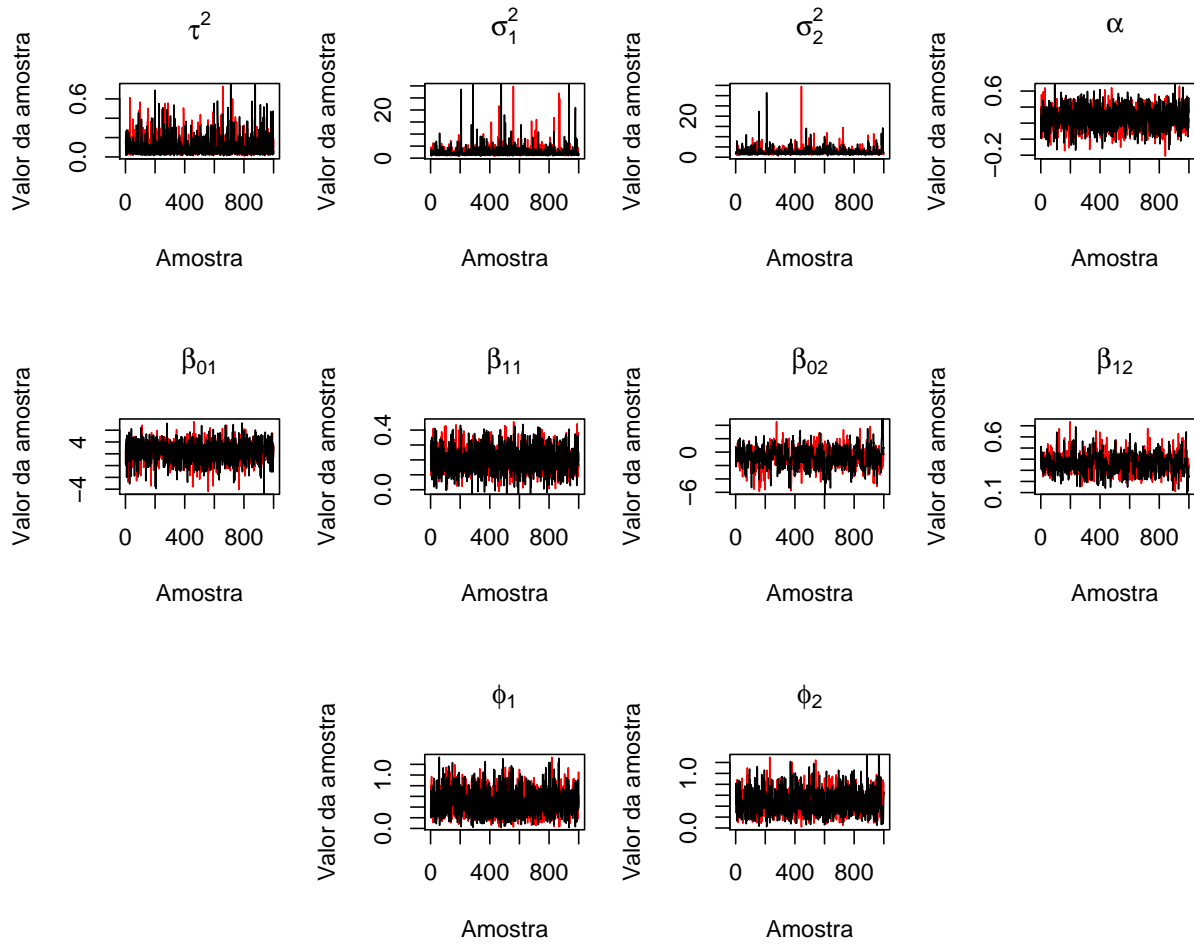


Figura 3.23 – Traço de duas cadeias *a posteriori* dos parâmetros sob o cenário c), representadas pelas cores vermelha e preta, partindo de valores iniciais distintos, modelo bivariado ajuste com amostra de tamanho 76

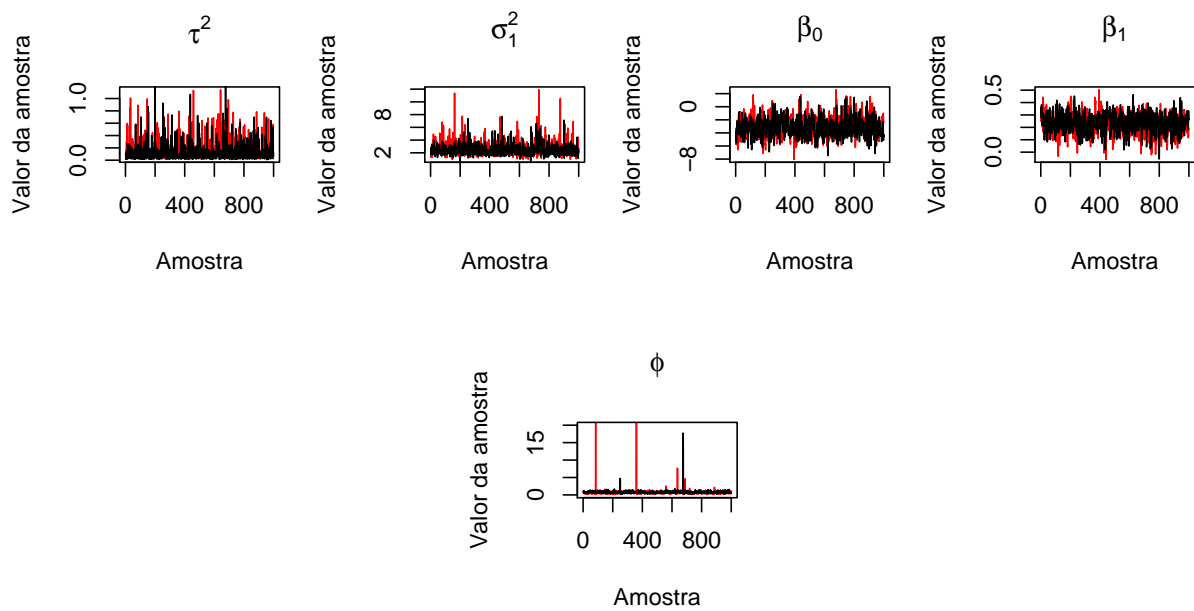


Figura 3.24 – Traço de duas cadeias *a posteriori* dos parâmetros sob o cenário c), representadas pelas cores vermelha e preta, partindo de valores iniciais distintos, modelo univariado, ajuste com amostra de tamanho 76

## APÊNDICE H

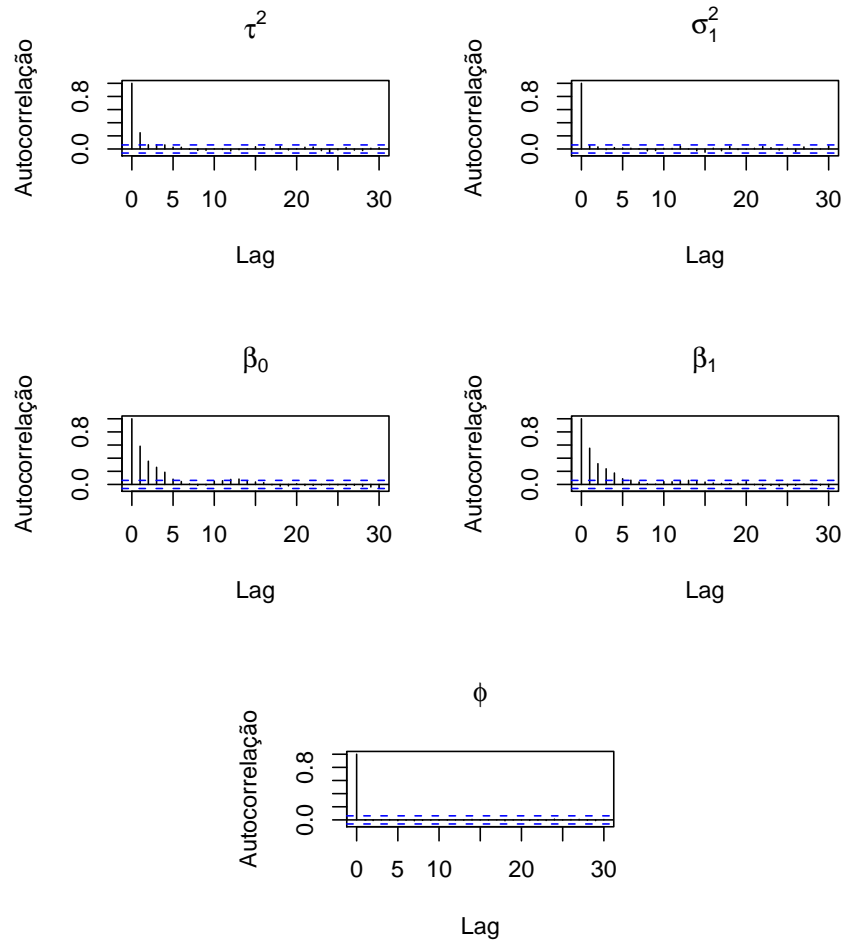


Figura 3.25 – Autocorrelação das cadeias *a posteriori* para os parâmetros do modelo univariado, sob o cenário c), ajuste com amostra de tamanho 76

## APÊNDICE I

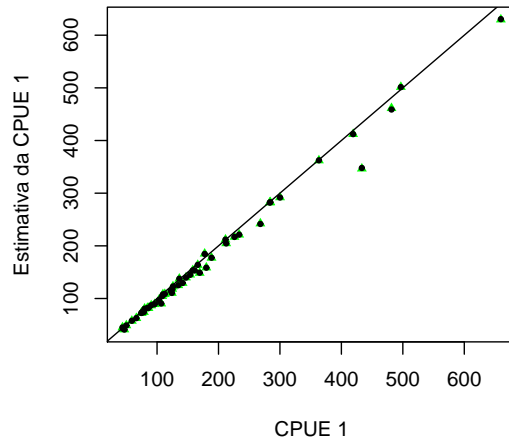


Figura 3.26 – Valores da  $CPUE_1$  versus estimativas ajustadas obtidas após o ajuste do modelo univariado (em verde) e bivariado (em preto), com tamanho de amostra,  $n = 85$ , sob o cenário d)

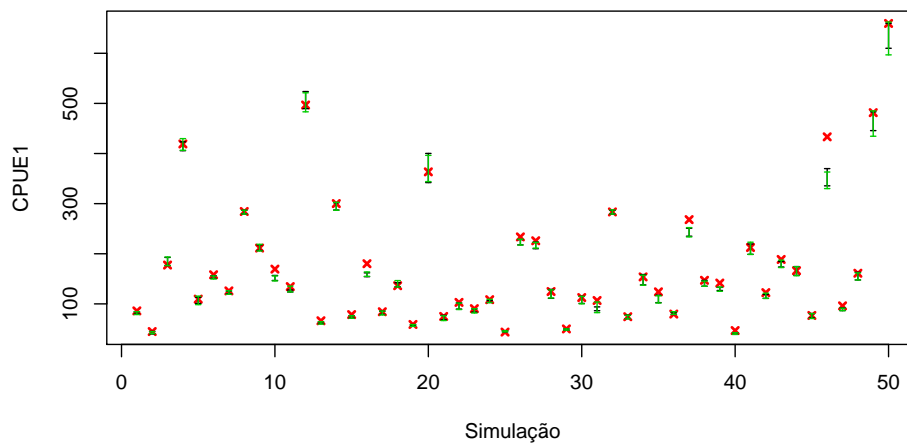


Figura 3.27 – Intervalos de credibilidade 95% da  $CPUE_1$  obtidas após o ajuste do modelo univariado (em verde) e bivariado (em preto), com tamanho de amostra,  $n = 85$ , sob o cenário d), e valores verdadeiros (em vermelho)



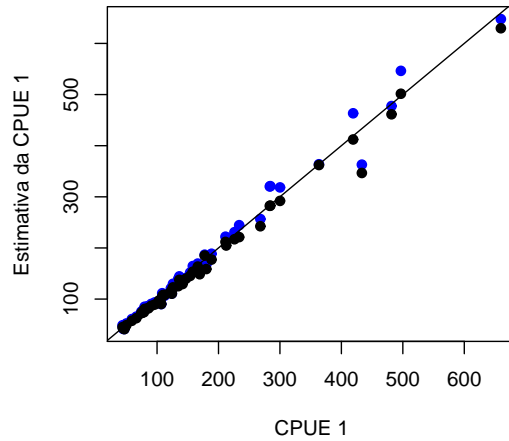


Figura 3.28 – Valores da  $CPUE_1$  versus estimativas amostrais (em azul) e estimativas ajustadas (em preto), considerando o modelo univariado, tamanho amostral  $n = 85$ , cenário d)

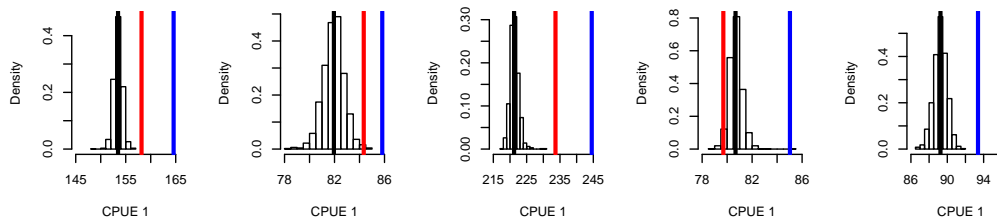


Figura 3.29 – Histogramas das estimativas ajustadas considerando o modelo univariado,  $n = 85$ , referente a 5 das 50 amostras da  $CPUE_1$  com a respectiva mediana (em preto), estimativa amostral (em azul) e valor verdadeiro (em vermelho), cenário d)

## APÊNDICE J

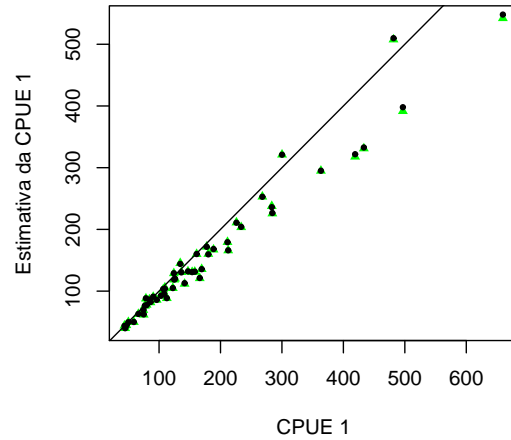


Figura 3.30 – Valores da  $CPUE_1$  versus estimativas obtidas após o ajuste do modelo univariado (em verde) e bivariado (em preto), com tamanho de amostra  $n = 76$ , sob o cenário d)

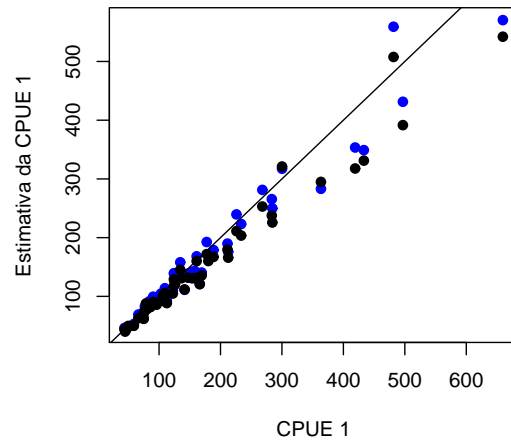


Figura 3.31 – Valores da  $CPUE_1$  versus estimativas ajustadas usando o modelo univariado (em preto) e estimativas amostrais (em azul), tamanho amostral  $n = 76$ , cenário d)

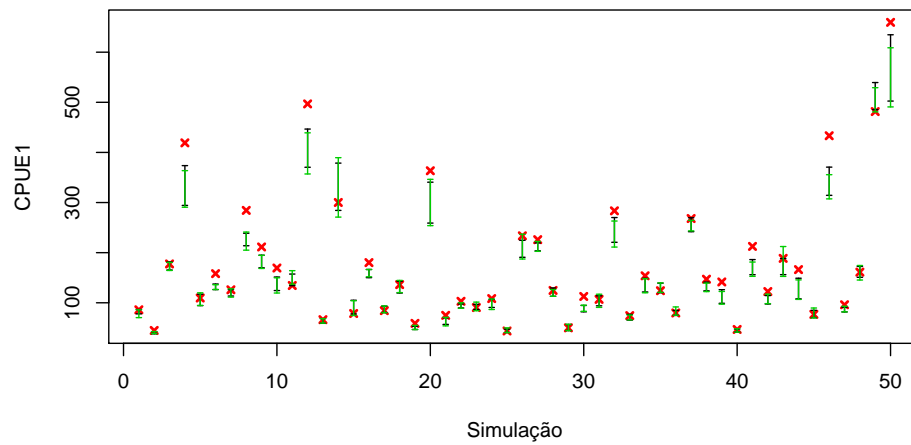


Figura 3.32 – Intervalos de credibilidade 95% da  $CPUE_1$  obtidas após o ajuste do modelo univariado (em verde) e bivariado (em preto), com tamanho de amostra,  $n = 76$ , sob o cenário d), e valores verdadeiros (em vermelho)

## ANEXOS

## ANEXO A - Programa usado para simular dados e rodar o winBUGS a partir do R

```
#####
## PROGRAMA USADO PARA SIMULAR DADOS E RODAR O WINBUGS A PARTIR DO R
#####
require(R2WinBUGS) require(geoR)
## Definindo um grid de 100 pontos
gs <- expand.grid((1:10), (1:10))
tam <- nrow(gs)

## Simulando temperatura
set.seed(2304) temp<- 10 + 1.3333*gs[,2] +
rnorm(length(gs[,2]),0,1) cord.cov<-cbind(gs,temp)

## Sorteando os locais de pesca (considerados observados) usando um
## processo de Thomas
require(spatstat)
set.seed(2304)
X <- rThomas(0.85, 0.3, 4,win =owin(c(1,10),c(1,10)))
cx<-round(X$x)
cy<-round(X$y)
coords.obs <- cbind(cx,cy)

## Referenciando todos os locais de um mesmo quadratim por um mesmo
## ponto coords.obsnr <- unique(coords.obs)

for(k in 1:50){

rm(simE)
rm(simC)

## Fixando os parâmetros a serem usados na simulação dos dados
phi1 <- 0.45
phi2 <- 0.5
alpha <- 1.2
sigma12 <- 2.3
sigma22 <- 1.8
```

```

##Simulando o logaritmo do esforço
beta01 <- 2
beta11 <- 0.25

## Média do log do esforço para ser usada na simulação
lEbarra <- beta01 + beta11*temp ## y1 puramente espacial
set.seed(340+k) # Para mudar a semente a cada simulação
simE<-grf(nsim=1,grid=gs,cov.pars=c(sigma12,1/phi1),cov.model=
  "powered.exponential",kappa=1,nug=0)
simE$data<-simE$data + lEbarra

## Simulando o logaritmo da captura dado o log do esforço
## Fixando mais parâmetros para simular log da captura dado o log do
## esforço
pep2 <- 0.1
beta02 <- 0
beta12 <- 0.15

## Média do log da captura para ser usada na simulação
lCbarra<-beta02+ beta12*temp + alpha*simE$data
set.seed(240+k) # Para mudar a semente a cada simulação
simC<-grf(nsim=1,grid=gs,cov.pars=c(sigma22,1/phi2),cov.model=
  "powered.exponential",kappa=1,nug=pep2)
simC$data<-simC$data + lCbarra

## Colunas: 1-long, 2-lat, 3-temperatura, 4-log Esforço, 5-log
##captura
exporta <- cbind(simE$coords[,1], simE$coords[,2], temp,
simE$data,simC$data)

## Atribuindo os valores da covariável e das variáveis aos locais
##considerados observados
capesf.obs <-NULL
linha <- NULL for(i in
1:length(coords.obsnr[,1])){

```

```

for(j in 1:length(exports[,1])){
  if(coords.obsnr[i,1]==exports[j,1] && coords.obsnr[i,2]==exports[j,2])
    {capesf.obs <- rbind(capesf.obs,exports[j, ]); linha<-c(linha,j)}
  }
}
coords.temp.pred <- exports[-linha,1:3]
coords.omit <- exports[-linha,]

## Número de pontos observados
N <- length(capesf.obs[,1])

## Número de pontos a serem preditos
M <- length(coords.temp.pred[,1])

## Coordenadas dos pontos observados
long <- capesf.obs[,1]
lat <- capesf.obs[,2]

## Temperatura observada
temp.obs <- capesf.obs[,3]

## Log do esforço observado
y1 <- capesf.obs[,4]

## Log da captura observada
y2 <- capesf.obs[,5]

## Coordenadas dos pontos a serem preditos
x.pred <- coords.temp.pred[,1] y.pred <- coords.temp.pred[,2]

## Valores da covariável nos pontos a serem preditos
temp.pred <- coords.temp.pred[,3]

## Valores verdadeiros de log do esforço e log da captura
## nos pontos a serem preditos
y1.omit <- coords.omit[,4] y2.omit <- coords.omit[,5]

```

```

## Preparando dados que serão usados para rodar o winBUGS
## Lista de dados que serão usados no winBUGS
capesf.data <- list ("N", "temp.obs", "y2", "y1", "lat", "long")

## Valores iniciais para o winBUGS gerar as cadeias para os
## parâmetros
capesf.inits <- list(list(tau=2, tau1=4, tau2=1, alpha=3, beta01=5,
beta11=1, beta02=1, beta12=-0.2, phi1=1, phi2=1))

## Parâmetros a serem estimados (usando o winBUGS)
capesf.parameters <- c("tau", "tau1", "tau2", "alpha", "beta01",
"beta11", "beta02", "beta12", "phi1", "phi2")

# Função usada para rodar o winBUGS a partir do R
capesf.sim <- bugs(capesf.data, capesf.inits, capesf.parameters,
"capesf.bug", n.chains=1, n.iter=55000, debug=F, n.burnin=0, n.thin=1)

## Dando thinning e burn-in após ter rodado o winBUGS
burn=5000 # Entre com o valor do burn-in desejado
tin=50 # Entre com o valor do thin

## tamanho total da cadeia
le=length(capesf.sim$sims.array[,1,"tau"])

## Armazenando amostras dos parâmetros após thinning e burn-in
cad1.pcp2<- 1/capesf.sim$sims.array[seq(burn+1,le,tin),1,"tau"]
cad1.sigma12<- 1/capesf.sim$sims.array[seq(burn+1,le,tin),1,"tau1"]
cad1.sigma22 <- 1/capesf.sim$sims.array[seq(burn+1,le,tin),1,"tau2"]
cad1.alpha<-capesf.sim$sims.array[seq(burn+1,le,tin),1,"alpha"]
cad1.beta01<-capesf.sim$sims.array[seq(burn+1,le,tin),1,"beta01"]
cad1.beta11<-capesf.sim$sims.array[seq(burn+1,le,tin),1,"beta11"]
cad1.beta02<-capesf.sim$sims.array[seq(burn+1,le,tin),1,"beta02"]
cad1.beta12<-capesf.sim$sims.array[seq(burn+1,le,tin),1,"beta12"]
cad1.phi1<-capesf.sim$sims.array[seq(burn+1,le,tin),1,"phi1"]
cad1.phi2<-capesf.sim$sims.array[seq(burn+1,le,tin),1,"phi2"]

```



```

## Verificando autocorrelação
par(mfrow=c(4,3))
acf(cad1.pep2,ylab="Autocorrelação",main = "pep2")
acf(cad1.sigma12,ylab="Autocorrelação",main = "sigma12")
acf(cad1.sigma22,ylab="Autocorrelação",main = "sigma22")
acf(cad1.alpha,ylab="Autocorrelação",main = "alpha")
acf(cad1.beta01,ylab="Autocorrelação",main = "beta01")
acf(cad1.beta11,ylab="Autocorrelação",main = "beta11")
acf(cad1.beta02,ylab="Autocorrelação",main = "beta02")
acf(cad1.beta12,ylab="Autocorrelação",main = "beta12")
acf(cad1.phi1,ylab="Autocorrelação",main = "phi1")
acf(cad1.phi2,ylab="Autocorrelação",main = "phi2")

cad1.tudo<-cbind(cad1.pep2,cad1.sigma12,cad1.sigma22,cad1.alpha,
cad1.beta01,cad1.beta11,cad1.beta02,cad1.beta12,cad1.phi1,cad1.phi2)

## Armazenando as cadeias as amostras a posteriori dos parâmetros em
##arquivo
write.table(cad1.tudo,file=paste("cad1tudo_",k,".txt"),append=FALSE,
quote=FALSE,sep=" ",eol="\n",na="NA",dec=".",row.names=FALSE,
col.names=TRUE,qmethod=c("escape","double"))

y1.verd <- exporta[,4]
y2.verd <- exporta[,5]

## Guardando os valores verdadeiros de esforço e captura, em 100
##localizações
lista.verd <- cbind(y1.verd, y2.verd)

## Guardando os valores observados, N = 85 localizações
lista.obs <- cbind(long, lat,temp.obs, y1, y2)

## Guardando os locais a serem preditos, M = 15 localizações
lista.pred <- cbind(x.pred, y.pred, temp.pred, y1.omit, y2.omit)

```

```
## Armazenando uma tabela em arquivo contendo os dados simulados
##para toda a região
write.table(lista.verd,file=paste("listaverd_",k,".txt"),append=FALSE,
  quote=FALSE,sep=" ",eol="\n",na="NA",dec=".",row.names=FALSE,
  col.names=TRUE,qmethod=c("escape","double"))
## Armazenando uma tabela em arquivo contendo os dados considerados
## como observados
write.table(lista.obs,file=paste("listaobs_",k,".txt"),append=FALSE,
  quote=FALSE,sep=" ",eol="\n",na="NA",dec=".",row.names=FALSE,
  col.names=TRUE,qmethod=c("escape","double"))
## Armazenando uma tabela em arquivo contendo os dados considerados
##não observados
write.table(lista.pred,file=paste("listapred_",k,".txt"),append=FALSE,
  quote=FALSE,sep=" ",eol="\n",na="NA",dec=".",row.names=FALSE,
  col.names=TRUE, qmethod=c("escape","double")) }
```

## **ANEXO B - Programa winBUGS para a obtenção de amostras da distribuição *a posteriori* conjunta dos parâmetros do modelo (2.10)**

```
#####
## Ajustando o Modelo Condicional para log-esforço e log-captura##
#####

model{

  for (i in 1:N){

    mu[i]<-beta01+beta11*temp.obs[i:N]
    y2[i]~dnorm(mu2[i],tau) # Obs. tau = 1/tau_2^2
    mu2[i]<-beta02+beta12*temp.obs[i:N]+alpha*y1[i]+S2[i]
    mu3[i]<-0
  }

  #kappa=1, função de correlação exponencial
  y1[1:N]~spatial.exp(mu[],lat[1:N],long[1:N],tau1,phi1,1)
  #processo Gaussiano
  S2[1:N]~spatial.exp(mu3[],lat[1:N],long[1:N],tau2,phi2,1)
```

```
pep2<-1/tau
sigma12<-1/tau1
sigma22<-1/tau2

## Prioris
tau ~dgamma(2,0.1)      # priori para pep2~IG(2,0.1)
tau1 ~dgamma(2,2.3)    # priori para sigma12~IG(2,2.3)
tau2 ~dgamma(2,1.8)    # priori para sigma22~IG(2,1.8)
alpha~dnorm(0,0.01)    # priori vaga var(alpha)=100
beta01~dnorm(0,0.01)
beta11~dnorm(0,0.01)
beta02~dnorm(0,0.01)
beta12~dnorm(0,0.01)
phi1~dgamma(0.0125,0.025)
phi2~dgamma(0.0125,0.025)

}
```

### ANEXO C - Programa para predição e estimativas dos índices

```
#####
## Programa para fazer a predição nos locais não observados, para ##
##obter estimativas amostrais e ajustadas dos índices e para ##
##calcular os EQM's ##
#####
require(geoR)

## Função que vai gerar uma amostra de esforço da Normal
##Multivariada
rmvnorm<-function(am,mu,S){ s<-chol(S)
am.star<-dim(S)[[1]]*am aux<-matrix(rnorm(am.star),dim(S)[[1]],am)
aux2<-mu+ t(s)%*%aux
t(aux2) }

## Fazendo predição
cpue1.v <- NULL
cpue2.v <- NULL
cpue3.v <- NULL

## Armazenando as medianas das cpue's
## obtidas para cada conjunto de dados

mdcpue1 <- NULL
mdcpue2 <- NULL
mdcpue3 <- NULL

## Armazenando as etimativas clássicas
## da cpue para cada conj. de dados
cpue1mp <- NULL
cpue2mp <- NULL
cpue3mp <- NULL

## Armazenando as ammostras das cpue's
## obtidas para cada conjunto de dados
amcpue1 = NULL
```

```

amcpue2 = NULL
amcpue3 = NULL

## Armazenando os valores de cap e esf. omitidos em cada simulação
y1.omit <- NULL
y2.omit <- NULL

## Armazenando o logaritmo do esforço e captura preditos
## As linhas são os valores do esf. nas localizações, as colunas são as
## iterações do MCMC e a profundidade são as 50 diferentes simulações.
gesf.marg <- array(NA,c(15,1000,50))
gcap.marg <- array(NA,c(15,1000,50))
for(i in 1:50){
  ## Entrar com os dados
#list.obs: long, lat, temp.obs, y1, y2
  lista.obs <- read.table(paste("listaobs_",i, ".txt", sep=""), head=T)
  #list.pred: x.pred, y.pred, temp.pred
  lista.pred <- read.table(paste("listapred_",i, ".txt", sep=""),
head=T)
#list.verd: y1.verd, y2.verd
  lista.verd <- read.table(paste("listaverd_",i, ".txt", sep=""),
head=T)

## Entrar com as cadeias referente a cada simulação
  cadl.tudo <- read.table(paste("cadltudo_",i, ".txt", sep=""), head=T)

## Extraindo os valores das listas
long <- lista.obs[ ,"long"]
lat <- lista.obs[ ,"lat"]
temp.obs <- lista.obs[ ,"temp.obs"]
y1 <- lista.obs[ ,"y1"]
y2 <- lista.obs[ ,"y2"]
N <- length(long)
x.pred <- lista.pred[ ,"x.pred"]
y.pred <- lista.pred[ ,"y.pred"]
temp.pred <- lista.pred[ ,"temp.pred"]

```

```

y1.omit <- cbind(y1.omit, lista.pred[ , "y1.omit"])
y2.omit <- cbind(y2.omit, lista.pred[ , "y2.omit"])
M <- length(x.pred)
y1.verd <- lista.verd[ , "y1.verd"]
y2.verd <- lista.verd[ , "y2.verd"]

cad1.pep2 <- cad1.tudo[, "cad1.pep2"]
cad1.sigma12 <- cad1.tudo[, "cad1.sigma12"]
cad1.sigma22 <- cad1.tudo[, "cad1.sigma22"]
cad1.alpha <- cad1.tudo[, "cad1.alpha"]
cad1.beta01 <- cad1.tudo[, "cad1.beta01"]
cad1.beta11 <- cad1.tudo[, "cad1.beta11"]
cad1.beta02 <- cad1.tudo[, "cad1.beta02"]
cad1.beta12 <- cad1.tudo[, "cad1.beta12"]
cad1.phi1 <- cad1.tudo[, "cad1.phi1"]
cad1.phi2 <- cad1.tudo[, "cad1.phi2"]
cpue1<-NULL
cpue2<-NULL
cpue3<-NULL

# matriz de distâncias geral
ma<-as.matrix(dist(cbind(c(long,x.pred), c(lat,y.pred))))
mcmc <- length(cad1.pep2)

for( m in 1:mcmc){
  # vetor de tamanho N de médias nos pontos observados
  mu.obse <- cad1.beta01[m]+cad1.beta11[m]*temp.obs
  # vetor de tamanho M de médias nos pontos não-observados
  mu.unobse <- cad1.beta01[m]+cad1.beta11[m]*temp.pred
  R1 <- exp(-cad1.phi1[m]*ma)
  #Construindo a matriz de covariâncias para esforço
  S1 <- cad1.sigma12[m]*R1
  # Matriz de covariâncias entre os pontos observados
  S1.obs <- S1[1:N,1:N]
  # Matriz de covariâncias entre os pontos não observados
  S1.unobs <- S1[ (N+1) : (N+M) , (N+1) : (N+M) ]
}

```

```

# Matriz de covariâncias entre os pontos observados e não observados
S1.obs.un <- S1[1:N, (N+1):(N+M)]
# Matriz de covariâncias entre os pontos não observados e observados
S1.un.obs <- S1[(N+1):(N+M), 1:N]
S1.obs.inv <- solve(S1.obs)
esf.pred <- rmvnorm(50, as.vector(mu.unobse + S1.un.obs %*%
  S1.obs.inv %*% (y1 - mu.obse)), S1.unobs - S1.un.obs%*%
  S1.obs.inv%*%S1.obs.un)
# tira a média por coluna, o resultado é um vetor_m das médias por
#local
esf.marg <- as.vector(apply(esf.pred,2,mean))
# guardando o esforço predito.
gesf.marg[,m,i] <- esf.marg

mu.obsc <- cad1.beta02[m]+cad1.beta12[m]*temp.obs + cad1.alpha[m]*y1
# Obs. mu.unobsc é uma matriz 15x50 cada coluna é uma amostra e
#cada linha é uma mesma localização
mu.unobsc <- cad1.beta02[m]+cad1.beta12[m]*temp.pred +
  cad1.alpha[m]*t(esf.pred)

R2 <- exp(-cad1.phi2[m]*ma)
##Construindo a matriz de covariancias para cap|esforço
S2 <- cad1.sigma22[m]*R2+diag(cad1.pep2[m],length(ma[,1]),
  length(ma[,1]))
# Matriz de covariâncias entre os pontos observados
S2.obs <- S2[1:N,1:N]
# Matriz de covariâncias entre os pontos não observados
S2.unobs <- S2[(N+1):(N+M), (N+1):(N+M)]
# Matriz de covariâncias entre os pontos observados e não observados
S2.obs.un <- S2[1:N, (N+1):(N+M)]
# Matriz de covariâncias entre os pontos não observados e observados
S2.un.obs <- S2[(N+1):(N+M), 1:N]
S2.obs.inv <- solve(S2.obs)
cap.pred <- rmvnorm(50, as.matrix(mu.unobsc) + as.vector(S2.un.obs
  %*% S2.obs.inv %*% (y2 - mu.obsc)), S2.unobs - S2.un.obs%*%
  S2.obs.inv%*%S2.obs.un)

```

```

# Calcula a cap marginal pelo método Monte Carlo (aprox. p/
#preditiva) em cada uma das 15 localizações.
cap.marg <- apply(cap.pred,2,mean)

# guardando o esforço predito.
gcap.marg[,m,i] <- cap.marg
# Usando
exp() no calculo dos estimadores
# porque os dados se referem ao log(esforço) e log(captura)
y1tudo <- c(y1,esf.marg)
y2tudo <- c(y2,cap.marg)
cpue1[m] <- mean(exp(y2tudo)/exp(y1tudo))
cpue2[m] <- sum(exp(y2tudo))/sum(exp(y1tudo))
cpue3[m] <- sum(exp(y2tudo)*exp(y1tudo))/sum(exp(y1tudo)^2)
} # Fecha o 2º for

amcpue1 = cbind(amcpue1, cpue1) amcpue2 = cbind(amcpue2, cpue2)
amcpue3 = cbind(amcpue3, cpue3)

mdcpue1[i] = median(cpue1)
mdcpue2[i] = median(cpue2)
mdcpue3[i] = median(cpue3)

## Obtendo estimativas amostris dos índices
cpue1mp[i] <- mean((exp(y2))/(exp(y1)))
cpue2mp[i] <- sum(exp(y2))/sum(exp(y1))
cpue3mp[i] <- sum(exp(y2)*exp(y1))/sum(exp(y1)^2)

## Calculando os valores verdadeiros dos índices
## y2.verd =captura, y1.verd = esforço
cpue1.v[i] <- mean((exp(y2.verd))/(exp(y1.verd)))
cpue2.v[i] <- sum(exp(y2.verd))/sum(exp(y1.verd))
cpue3.v[i] <- sum(exp(y2.verd)*exp(y1.verd))/sum(exp(y1.verd)^2)
} #fecha o 1º for

```



```
## Calculando o EQM para as etimativas dos índices
eqm.bayes1 <- mean((mdcpue1 - cpue1.v)^2)
eqm.mp1 <- mean((cpue1mp - cpue1.v)^2)
eqm.bayes2 <- mean((mdcpue2 - cpue2.v)^2)
eqm.mp2 <- mean((cpue2mp - cpue2.v)^2)
eqm.bayes3 <- mean((mdcpue3 - cpue3.v)^2)
eqm.mp3 <- mean((cpue3mp - cpue3.v)^2)
comp.eqm <- cbind(c(eqm.bayes1,eqm.bayes2,eqm.bayes3),
c(eqm.mp1,eqm.mp2,eqm.mp3))
```

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)