

**UNIVERSIDADE FEDERAL DO RIO GRANDE
PÓS-GRADUAÇÃO EM OCEANOGRAFIA BIOLÓGICA**

**MODELAGEM BAYESIANA PARA PREDIÇÃO ESPACIAL DOS
HABITATS ESSENCIAIS DO ARIOCÓ *Lutjanus synagris* NA COSTA
DO RIO GRANDE DO NORTE**

MARIE-CHRISTINE RUFENER

Dissertação apresentada ao Programa de
Pós-Graduação em Oceanografia Biológica
da Universidade Federal do Rio Grande
como requisito parcial à obtenção do título
de Mestre.

Orientação: Prof. Dr. Paul Gerhard Kinas
Orientação colaborativa: Prof. Dr. Marcelo Francisco de Nóbrega

**RIO GRANDE / RS
MARÇO 2016**

*Aos meus pais, Maria das Graças
e Jacques, e minha tia Nícia Maria
(in memoriam).*

AGRADECIMENTOS

Esse trabalho somente foi possível ser realizado mediante o incentivo, apoio e carinho de muitas pessoas que já se marcavam presentes em minha vida, e outras que cruzaram o meu caminho durante a minha estadia em Rio Grande. Devido a isso, gostaria de prestar meus sinceros agradecimentos a todos que me ajudaram a alcançar essa minha conquista.

Em primeiro lugar gostaria de agradecer aos meus pais por sempre estarem ao meu lado me dando apoio moral, intelectual, emocional e, mais importante, por sempre me oferecerem tanto carinho e amor. À minha irmã, que mesmo longe, também torce pelas minhas vitórias.

Ao meu namorado que, ao longo desses dois anos, se revelou ser um grande amigo. Obrigada pela paciência, pelo voto de confiança e por sempre ter me incentivando e apoiado nos momentos em que desacreditei desse trabalho.

Meus sinceros agradecimentos ao professor Kinas que, apesar de todas atividades, me aceitou orientar e sempre encontrou um tempinho livre para discutir comigo assuntos que diziam respeito ao meu trabalho. Agradeço, também, ao professor Marcelo não apenas pela concessão dos dados, como também por ter acreditado em mim e me incentivado a realizar o Mestrado aqui em Rio Grande.

À toda a equipe do Laboratório de Estatística Ambiental (LEA), por terem tornado os meus dias mais alegres principalmente quando batia a saudade da minha família e amigos em Natal. Meu especial agradecimento a Rayd pela amizade que surgiu, pelo apoio e também pelas nossas longas discussões filosóficas ao longo do nosso tão querido “beco do sono”. Não poderia também deixar de agradecer a Liana, Laura, Aline e Rayd pelas nossas ótimas conversas e risadas durante os nossos encontros.

Aos membros da banca Luciano Dalla Rosa e Lauro Madureira por terem oferecido ótimas sugestões para melhorar este trabalho. À María Pennino gostaria de agradecer em específico não só pela participação da banca, como também pelas inúmeras ajudas, incentivos e inspirações fornecidas durante este período e, que sem eles, metade deste trabalho não teria sido realizado.

Agradeço imensamente à Thaísa e ao Gabo não só pelas horas de diversão, mas também pela amizade sincera que surgiu ao longo desses dois anos. Meu

muitíssimo obrigada ofereço também ao Victor, Denis e Mário que se revelaram grandes companheiros durante o meu último semestre em Rio Grande.

Aos meus amigos de Natal, em especial Natasha, Elaine e Rodrigo. Mesmo estando tão ausente, obrigada por sempre terem me procurado e oferecido tanto carinho apesar de toda a distância. Obrigada também ao Vinícius pelas horas de discussão e ajuda sobre estatística, INLA, R, SIG e temas a fins.

Ao Programa de Pós-Graduação em Oceanografia Biológica por ter me aceitado como aluna.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa durante o período do meu Mestrado.

E, por fim, mas não menos importante, meu obrigado a todos os colegas dos mais diversos fóruns sobre o R por terem me ajudado a resolver prontamente alguns dos meus vários desafios relacionados ao programa.

*“A mente que se abre a uma nova ideia
jamais volta ao seu tamanho original.”*

Albert Einstein

*“An approximate answer to the right problem
is worth a good deal more than an exact
answer to an approximate problem.”*

John Tukey

SUMÁRIO

Resumo	8
Abstract	9
Lista de siglas	10
Contextualização	11
Capítulo 1: Introdução geral	15
1.2. Objetivos	18
Capítulo 2: Fundamentação teórica	19
2. Caracterização dos recursos.....	20
2.1. <i>Lutjanus synagris</i>	20
2.2. Modelagem estatística.....	23
2.3. Inferência Bayesiana.....	25
2.4. Modelos de distribuição de espécies.....	28
2.5. Estatística espacial.....	32
2.6. Modelagem de habitats essenciais via Geoestatística Bayesiana.....	34
2.6.1. Formulação hierárquica do modelo espacial.....	41
2.6.2. Inferência Bayesiana via INLA.....	42
Capítulo 3: Material & Métodos	44
3. Caracterização da área de estudo.....	45
3.1. Amostragem.....	47
3.1.1. Procedimentos laboratoriais.....	50
3.2. Variáveis ambientais.....	51
3.3. Análise dos dados.....	54
3.3.1. Determinação do tamanho médio de primeira maturação.....	54
3.3.2. Modelagem da abundância e estrutura de idades.....	57
3.3.3. Inferência, seleção e avaliação dos modelos.....	59
Capítulo 4: Bayesian spatial predictive models for data-poor fisheries	62
4.1. Abstract.....	63
4.2. Introduction.....	64
4.3. Material & Methods.....	67
4.3.1. Case study.....	67

4.3.2.	Study area and data survey.....	68
4.3.3.	Environmental predictors	70
4.3.4.	Bayesian estimation of mean length at first maturity	71
4.3.5.	Essential fish habitat modeling through Bayesian Geostatistics.....	74
4.3.6.	Modeling abundance and age groupings	77
4.3.7.	Model fitting.....	78
4.3.8.	Model selection and evaluation.....	80
4.4.	Results.....	81
4.4.1.	Length at first maturity & age grouping	81
4.4.2.	Data overview	83
4.4.3.	Model selection.....	83
4.4.4.	Prediction of essential fish habitat for the lane snapper.....	88
4.5.	Discussion	90
4.5.1.	Biological discussion: size at first maturtiy	90
4.5.2.	Model discussion.....	90
4.5.3.	Ecological discussion.....	92
4.6.	Concluding remarks	94
4.7.	Acknowledgments	96
Capítulo 5:	Conclusão geral	97
Referências bibliográficas	100
Anexo I: Códigos	115

RESUMO

Compreender a distribuição espacial e identificar um conjunto de variáveis ambientais que influenciam na abundância de espécies de peixes sobreexploradas constituem aspectos primordiais para a implementação de medidas de pescas sustentáveis e, dessa forma, garantir a sua conservação. Tendo em vista que muitos dos recursos pesqueiros tradicionalmente explorados na região nordeste do Brasil já se encontram sobreexplorados ou em exploração plena, o objetivo central desta Dissertação consistiu em desenvolver modelos hierárquicos Bayesianos espaciais a fim de quantificar e mapear habitats essenciais para a abundância e grupos de idades do ariocó (*Lutjanus synagris*), atualmente categorizado como ameaçado. Durante um período de dois anos foram realizadas mensalmente pescarias experimentais por intermédio das frotas artesanais que operam com redes de emalhar de fundo ao longo da costa do Rio Grande do Norte. A fim de delimitar o estoque em grupo de indivíduos jovens e adultos, foi utilizado uma regressão logística Bayesiana para definir o comprimento de primeira maturação (L_{50}). O conjunto de dados foi, posteriormente, ajustado mediante os métodos INLA (do inglês *Integrated Nested Laplace Approximations*) e SPDE (do inglês *Stochastic Partial Differential Equations*), os quais constituem novas metodologias tanto para a inferência Bayesiana quanto para a estatística espacial. Dentre as 126 estações amostrais o ariocó foi registrado em 83 estações, somando uma captura de 259,808 kg. Dos 606 indivíduos capturados, 101 foram categorizados como jovens e 505 como adultos. Os resultados provenientes dos modelos e dos mapas preditivos confirmaram que a abundância e a distribuição etária do ariocó possuem dependência espacial, são agregados distintamente ao longo da costa do Rio Grande do Norte, e tem suas abundâncias principalmente afetadas pela clorofila-a, rugosidade do leito marinho e distância da costa. Sendo assim, os resultados buscaram consolidar um recente modelo geoestatístico Bayesiano introduzido às ciências pesqueiras, destacando o seu potencial no estabelecimento de medidas mais confiáveis para futuras medidas de conservação e gestão de espécies de peixes vulneráveis.

Palavras-chave: Habitats essenciais para peixes; ecologia pesqueira; modelos de distribuição de espécies; inferência Bayesiana; INLA

ABSTRACT

Understanding the spatial distribution and identifying environmental variables that drive endangered fish species abundance are key factors to implement sustainable fishery management strategies, and thus ensure their conservation. Considering that many of the fishery resources traditionally exploited in the northeastern Brazil are already overexploited or, at least exploited, the central objective of this dissertation was to develop hierarchical Bayesian spatial models in order to quantify and map sensitive habitats for the abundance and age groupings of the lane snapper (*Lutjanus synagris*), currently categorized as threatened. During two years experimental fisheries were monthly conducted by fishing vessels of the artisanal fleets which operate with bottom gillnets along the Rio Grande do Norte coast. In order to delimit the lane snapper stocks into young and adult groupings, we used a Bayesian logistic regression to define length at first maturity (L_{50}). Data were posteriorly fitted through Integrated Nested Laplace Approximations (INLA) and the Stochastic Partial Differential Equations (SPDE), which constitute new methodologies for both Bayesian inference and spatial statistics. Among the 126 sampling stations, lane snapper was recorded at 83 stations with a total of 259.808 kg captured weight. Of the 606 individuals caught, 101 were categorized as juveniles and 505 as adults. The results provided from the models and predictive maps confirmed that the abundance, number of juveniles and adults of *L. synagris*, are spatially correlated, with a patchy distribution along the Rio Grande do Norte coast and are mainly affected by environmental predictors such as distance to coast, chlorophyll-a concentrations and sea floor rugosity. By means of our results we intended to consolidate a recently introduced Bayesian Geostatistical model into fisheries science, highlighting its potential for establishing more reliable measures for the conservation and management of vulnerable fish species.

Key words: Essential fish habitat; fisheries ecology; species distribution models; Bayesian inference; INLA

LISTA DE SIGLAS

AIC	Akaike information criterion
ANN	Artificial neural network
CART	Classification and regression trees
CB	Corrente do Brasil
CE	Climatic envelopes
CNB	Corrente norte brasileira
CPO	Conditional predictive ordinate
CSE	Corrente sul-equatorial
DIC	Deviance information criterion
EAFM	Ecosystem approach to fisheries management
EFH	Essential fish habitat
GAM	Generalized additive models
GLM	Generalized linear models
GMRF	Gaussian markov random field
GRF	Gaussian random field
HBM	Hierarchical Bayesian models
INLA	Integrated nested Laplace approximations
MCMC	Markov chain monte carlo
ME	Maximum entropy
SDM	Species distribution modeling
SPDE	Stochastic partial differential equations
STAR	Structured additive regression models
ZCIT	Zona de convergência intertropical
ZEE	Zona econômica exclusiva

CONTEXTUALIZAÇÃO

Os ambientes marinhos comportam a maior fronteira para pesquisa e exploração de recursos naturais por parte do homem (MMA, 2006). Dentre os ecossistemas marinhos, sabe-se que o domínio da plataforma continental sustenta a maior fração de produção mundial dos recursos renováveis, sendo que 95% dessas pescarias provêm das adjacências costeiras (Caddy, 1997; McLean & Tsyban, 2001). Os ambientes costeiros ocupam uma das interfaces mais dinâmicas na Terra, tendo em vista estabelecerem fronteira entre o ambiente terrestre e marinho e, em consequência disso, constituírem habitats de elevada biodiversidade e produtividade (McLean & Tsyban, 2001; Pinet, 2008). Apesar disto, esses ambientes também são os que estão mais passíveis à pressões de origem antrópica, levando populações de importantes recursos pesqueiros a níveis reduzidos de abundância e, em alguns casos, à ameaça de extinção.

Considerando as últimas cinco décadas, a produção do pescado mundial tem crescido em taxas aceleradas com uma oferta de pescado superestimada em relação à demanda populacional (FAO, 2014). Ainda segundo a FAO (2014), este desenvolvimento acelerado é fruto de um conjunto de fatores, os quais estão principalmente relacionados ao crescimento populacional, urbanização e facilitação da distribuição do pescado.

Muito embora os recursos biológicos de zonas costeiras apresentem grande capacidade de produção, as atividades humanas aliadas a diversos estressores antrópicos têm impactado de forma significativa diversas populações marinhas (Hilborn *et al.*, 2003). Com o advento do desenvolvimento tecnológico proporcionado principalmente após a Segunda Guerra Mundial, o poder de pesca e, consequentemente o impacto da pesca, sofreu um grande salto. Contudo, dado ao colapso completo de importantes recursos pesqueiros, tais como a sardinha californiana na década de 50, o arenque do atlântico ao final da década de 60 e a anchova Peruana na década de 70, fez-se perceber que os recursos marinhos, embora renováveis, eram limitados em sua capacidade de reposição (Jennings *et al.*, 2001). Esses colapsos serviram como sinais de advertência de que as práticas de pesca em muitas regiões do mundo eram insustentáveis, devendo-se assim iniciar a aplicação de medidas de manejo que visam a sustentabilidade dos recursos pesqueiros (Jennings *et al.*, 2001).

Jackson *et al.* (2001) relataram que o colapso de alguns estoques se deve majoritariamente à sobrepesca. Em função disto, a integridade das teias tróficas pode ficar seriamente ameaçada uma vez que os ecossistemas marinhos dependem de ligações tróficas específicas para manter a sua estabilidade, resiliência e persistência (Jackson *et al.*, 2001; Worm *et al.*, 2002; Myers & Worm, 2003; Myers *et al.*, 2007). A fim de mitigar esses problemas, a FAO tem sugerido desde 2003 a adoção de uma abordagem ecossistêmica para a gestão das pescarias (EAFM, do inglês *Ecosystem Approach to Fisheries Management*), aonde as interações entre os componentes biológicos, físicos e antrôpicos devem ser consideradas (FAO, 2003; Pikitch *et al.*, 2004; Marasco *et al.*, 2007; Cotter *et al.*, 2009).

O principal propósito da EAFM reside na busca da minimização dos impactos ocasionados pela pesca e, desta forma, preservar a integridade biológica dos ecossistemas marinhos. Adicionalmente, segundo Pennino (2013a), a EAFM deve (i) reduzir os riscos de mudanças irreversíveis nas assembléias de peixes e processos ecossistêmicos; (ii) gerar conhecimento dos processos ecológicos com o intuito de entender suas causas e consequências frente às ações humanas; (iii) evitar a degradação dos sistemas ecológicos; e (iii) manter benefícios sócio-econômicos de longa duração sem comprometer o ecossistema.

Os ambientes marinhos são conhecidos pela sua dinâmica espaço-temporal no que diz respeito a variabilidade de processos físico-químicos, fato este que reflete posteriormente também nos sistemas biológicos. Esse dinamismo suporta o conceito fundamental da EAFM inicialmente proposto por Walters (1986), o qual alerta que o manejo deveria ser considerado como um processo adaptativo, integrando os conhecimentos derivados da ciência e manejo a fim de subsidiar informações que auxiliem na sustentabilidade e conservação de espécies ameaçadas.

É diante deste contexto também que surge o termo de modelos de distribuição de espécies (SDM, do inglês *Species Distribution Modeling*), os quais estabelecem conexão direta com os modelos de habitats essenciais para peixes (EFH, do inglês *Essential Fish Habitat*) quando se trata de espécies marinhas. A identificação de EFHs está diretamente relacionada a áreas marinhas que promovem aos peixes habitats mais favoráveis para reprodução, alimentação e desenvolvimento completo, e tem por objetivo a conservação da biodiversidade e o manejo de pescarias sustentáveis (Benaka, 1999; Valavanis *et al.*, 2008).

Tradicionalmente os EFHs são acessados via modelos estatísticos que avaliam a relação entre a presença da espécie com diversas variáveis ambientais. O desenvolvimento constante de novas metodologias aliadas aos avanços dos Sistemas de Informação Geográfica (SIG) tem resultado em potentes modelos, capazes de estimarem e predizerem áreas preferencialmente ocupadas por uma espécie. Embora se conheça a existência de uma variabilidade espacial intrínseca à cada espécie, parte considerável dos SDMs/EFHs ignoram o componente espacial na estruturação de seus modelos (Valavanis, 2004; Valavanis, 2008). Além disso, quando avaliados sob enfoque Bayesiano, é constatado que o número de trabalhos que dizem respeito a modelos preditivos se tornam mais escassos ainda. Ignorar a incerteza dos dados e o componente espacial (quando presente), pode resultar em estimações e predições enviesadas de modo que área de ocorrência de uma espécie ficará à mercê da super ou subestimação. Recentemente Muñoz *et al.* (2013) introduziram à comunidade científica uma nova abordagem para modelos EFHs, a qual é fundamentada na estatística espacial e com inferência efetuada sob o prisma Bayesiano.

Baseando-se, portanto, nas informações previamente expostas, o presente trabalho teve por propósito aplicar essa nova abordagem de modelo EFH a uma espécie de peixe economicamente importante no Brasil e cuja população já vem dando sinais de sobreexploração.

Para melhor entendimento dos processos analíticos empregados no artigo, esta dissertação foi estruturada em capítulos. Sendo assim, o primeiro capítulo introduzirá brevemente a situação da pesca no Brasil, em específico no nordeste, e as motivações que levaram a realizar este trabalho. O capítulo 2, por sua vez, reúne algumas fundamentações teóricas básicas sobre a temática da modelagem, inferência Bayesiana, estatística espacial e modelos de distribuição de espécies, cada qual compõe tema central deste manuscrito. O capítulo 3 versará sobre os materiais e métodos empregados neste trabalho, dando-se maior ênfase na metodologia utilizada para definir os modelos de EFHs. O capítulo 4 reuniu, em forma de artigo, alguns tópicos dos capítulos antecedentes, aplicados ao estudo de caso do ariocó (*Lutjanus synagris*). O capítulo 5 finaliza esta dissertação expondo as principais conclusões. Por fim, o manuscrito também dispõe de um anexo contendo um tutorial reunindo comandos básicos que foram utilizados para o desenvolvimento dos modelos Bayesianos via INLA (do inglês *Integrated Nested*

Laplace Approximations), bem como suas extensões espaciais adaptadas pelo módulo SPDE, todos implementados exclusivamente no ambiente R (R Development Core Team, 2013).

A leitura sequencial de cada capítulo é opcional. Aos leitores cujo foco principal não se restrinja aos aspectos estatísticos, recomendo pular a leitura dos capítulos 2 e 3 e retomar a partir do quarto capítulo.

Capítulo 1

INTRODUÇÃO GERAL

1. Pesca marinha no Brasil com ênfase na região nordeste

Semelhante ao cenário mundial de sobreexploração e consequente declínio de algumas populações marinhas, quando se avalia as tendências históricas observa-se que os recursos pesqueiros tradicionalmente explorados na Zona Econômica Exclusiva (ZEE) do Brasil também se encontram sobreexplorados ou em exploração plena (Isaac *et al.*, 2006). Até meados da década de 80 as pesquisas pesqueiras no país estavam primordialmente voltadas a recursos pesqueiros individuais, focando em sua biologia, dinâmica populacional, ecologia e dimensionamento de sua abundância (Isaac *et al.*, 2006). Entretanto, no que diz respeito à abordagens ecossistêmicas, bem como sobre aspectos tecnológicos e sócio-econômicos, percebe-se uma lacuna de conhecimento. Desta forma as medidas de manejo respondiam basicamente ao desenvolvimento da pesca sem, contudo, dar muita ênfase à sustentabilidade da produção pesqueira.

Com a assinatura em 1982 e a ratificação em 1988 à Convenção das Nações Unidas sobre o Direito do Mar, o Brasil passou a assumir compromissos tanto nacionais quanto internacionais e, pela primeira vez, o governo brasileiro se viu obrigado a adotar políticas mais incisivas no que diz respeito aos recursos pesqueiros que habitam em suas águas (Isaac *et al.*, 2006; MMA, 2006). Dentre os principais compromissos assumidos pelo país, todos apoiados no uso sustentável, destacam-se aqueles associados à conservação, exploração, gestão e aproveitamento dos recursos vivos da ZEE (MMA, 2006).

Constata-se que a partir da década de 90 a produção proveniente da pesca extrativa marinha se manteve estagnada, oscilando entre 400 e 500 mil toneladas (Haimovici *et al.*, 2006). De acordo com o Ministério da Pesca e Aquicultura (MPA, 2013), em 2011 a pesca marinha continuou sendo a principal fonte de produção de pescado nacional, somando-se 553.670 t. Apesar de se ter registrado uma queda de aproximadamente 5% em relação ao ano de 2010, a região Nordeste registrou para este mesmo ano a maior produção de pescado marinho, respondendo por 31,7% da produção nacional (MMA, 2013).

Diferentemente da região sul-sudeste, onde predomina a pesca industrial, no nordeste a pesca artesanal é o principal provedor de pescado (Lessa *et al.*, 2009a). Segundo Lessa *et al.* (2009a), no nordeste a presença de espécies de elevado valor

comercial se encontram em baixas densidades. Essas baixas abundâncias refletem diretamente às águas oligotróficas que percolam essa área, bem como aos baixos valores de biomassa primária (Rossi-Wongtschowski *et al.*, 2006). No entanto, as ilhas oceânicas e os montes submarinhos registrados ao longo da costa do nordeste constituem exceções para essa baixa produtividade. Tendo em vista que essas feições provocam a ascensão de águas profundas e ricas em nutrientes, promove-se uma elevação de biomassa primária a qual permite sustentar maiores teias tróficas tanto em número quanto em peso (Lalli & Parsons, 1997; Travassos *et al.*, 1999; Rossi-Wongtschowski *et al.*, 2006).

No que diz respeito à contribuição individual por unidade de federação do nordeste, é verificado que a Bahia, Ceará, Maranhão e Rio Grande do Norte representam os maiores produtores de pescado marinho da região (MMA, 2013). Dentro os principais recursos capturados, constam algumas espécies de lagostas (e.x., lagosta vermelha – *Panulirus argus*; e lagosta verde – *P. laevicauda*) e em sua grande maioria espécies de peixes, tais como ariocó (*Lutjanus synagris*), guaiúba (*Ocyurus chrysurus*), a serra (*Scomberomorus brasiliensis*) e o dourado (*Coryphaena hippurus*) (MMA, 2006; Lessa *et al.*, 2009a).

Em função do seu elevado valor econômico, o ariocó particularmente representa um dos recursos tradicionalmente mais explorados no nordeste e, em específico no Rio Grande do Norte. Muito embora essa espécie apresenta elevadas abundâncias, Lessa *et al.* (2009b) já tem alertado para o seu declínio populacional. No entanto, ainda se sabe pouco a respeito da sua dinâmica populacional, e variação no espaço-tempo frente às oscilações ambientais. Desta maneira, se desejarmos recuperar o estoque do ariocó e explorá-lo de forma sustentável, fica evidente a necessidade em se estudar a sua relação com o ambiente aplicando os conceitos da EAFM.

1.2. Objetivos

Munidos do conjunto das informações previamente expostas, o presente estudo teve por objetivo central desenvolver modelos hierárquicos Bayesianos para predizer a distribuição espacial de *L. synagris* ao longo da costa do Rio Grande do Norte.

Como objetivo específico, o estudo se propôs a:

- ◆ Estimar o comprimento médio de primeira maturação gonadal de *L. synagris* para delimitar o estoque em grupos de idades (jovens e adultos).
- ◆ Identificar o conjunto de preditores ambientais que influenciam na variação espacial da CPUE e nos grupos de idades de *L. synagris*.
- ◆ Gerar mapas cartográficos para as CPUEs e grupos de idades de *L. synagris*.
- ◆ Identificar potenciais habitats essenciais para *L. synagris* ao longo da costa do Rio Grande do Norte.

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

2. Caracterização dos recursos

2.1. *Lutjanus synagris*

O *Lutjanus synagris*, popularmente conhecido como ariocó, é uma espécie pertencente à família Lutjanidae sendo facilmente identificado por caracteres morfológicos. A espécie apresenta coloração avermelhada com tons de oliva e pálido na região mais dorsal e ventral, respectivamente. Além disso, estrias amarelas longitudinais na parte inferior do corpo e diagonais na parte superior são típicas de *L. synagris*, assim como a presença de uma mancha escura difusa localizada sob a nadadeira dorsal (Fig. 1) (Garcia Jr. *et al.*, 2010).



Figura 1: Exemplar de indivíduo adulto do ariocó capturado na bacia Potiguar (Fonte: Garcia Jr. *et al.*, 2010).

O ariocó é largamente distribuído ao longo de toda a região tropical do Atlântico ocidental desde a Carolina do Norte nos (EUA) até a região Sudeste do Brasil, incluindo o Golfo do México, as Bahamas e as Antilhas (McEachran & Fechhelm, 2005). São encontrados associados a uma variedade de tipos de substratos, ocorrendo frequentemente nas proximidades de recifes de coral e em vegetações sobre fundos arenosos desde águas costeiras até profundidades de 400 m (Carpenter, 2002).

Estudos de dieta conduzidos por Duarte & Garcia (1999) na Colômbia e por Fonseca (2009) no Banco dos Abrolhos, Brasil, classificaram o ariocó como uma espécie generalista e carnívora oportunista. No que diz respeito à maturação sexual, de acordo com o estudo efetuado por Sousa-Júnior *et al.* (2008) na costa do Ceará, a maturação gonadal ocorre por volta dos 24 cm, com o período de desova

apresentando dois picos: um de janeiro a abril (mais intenso) e outro de agosto a dezembro (menos intenso).

Dado a seu valor comercial, esta espécie constitui um dos principais suportes das pescarias artesanais no Nordeste do Brasil, na Venezuela e na região do mar do Caribe (Luckhurst *et al.*, 2000; Gómez *et al.*, 2001; Rezende *et al.*, 2003). Muito das informações que conhecemos a respeito da idade e crescimento, biologia reprodutiva, e caracterização ecológica e populacional de *L. synagris* são obtidas por meio de estudos realizados no hemisfério norte (Claro, 1981a/b e 1982; Manooch & Mason, 1984; Manickchand-Dass, 1987; Torres-Lara & Chávez, 1987; Torres-Larra & Salas-Márquez, 1990; Acosta & Appeldoorn, 1992; Barros *et al.*, 1996; Rivera-Arriaga *et al.*, 1996; Luckhurst *et al.*, 2000).

No Brasil as primeiras investigações sobre a espécie em questão foram realizadas no Estado do Ceará, por intermédio dos estudos de idade e crescimento (Alegria & Menezes, 1970) e de fecundidade (Gesteira & Rocha, 1976). A partir do ano 2000 verifica-se uma explosão de artigos voltados ao estudo biológico do ariocó, citando-se como exemplo os trabalhos efetuados por: Lima, 2004a; Abreu, 2005; Sousa-Júnior *et al.*, 2008; Fonseca, 2009; Freitas *et al.*, 2011; Cavalcante *et al.*, 2012; Oliveira *et al.*, 2012; Freitas *et al.*, 2014.

De acordo com Lessa *et al.* (2009b), as capturas do ariocó no nordeste do Brasil são registradas desde o final da década de 70, tendo sido observado um aumento significativo nos seus desembarques nos últimos anos. Entre 1959 e 2013 houve um aumento no peso capturado de 1.000 t para 3.154 t, com uma captura máxima de 4.550 t em 1975 (FAO, 2013) (Fig. 2). Considerando a região do Atlântico sudoeste, o desembarque aumentou de 506 t para 1.689 t em 1995 e 2013, respectivamente (FAO, 2013) (Fig. 2).

Ao atentarmos somente para a região nordeste do Brasil, percebe-se que segundo os dados disponibilizados pelo IBAMA (2005 a 2007), houve um aumento consecutivo do peso desembarcado de *L. synagris* entre os anos de 2005 a 2007 (2005: 1.775 t; 2006: 1.841 t; 2007: 2.030 t) (Fig. 3). Os estados do Ceará, Bahia e Rio Grande do Norte foram os que mais contribuíram para esses desembarques, sendo registrado uma captura total de 834 t durante este período somente para o Rio Grande do Norte, o que corresponde a aproximadamente 15% do peso desembarcado para o ariocó no Nordeste (Fig. 3).

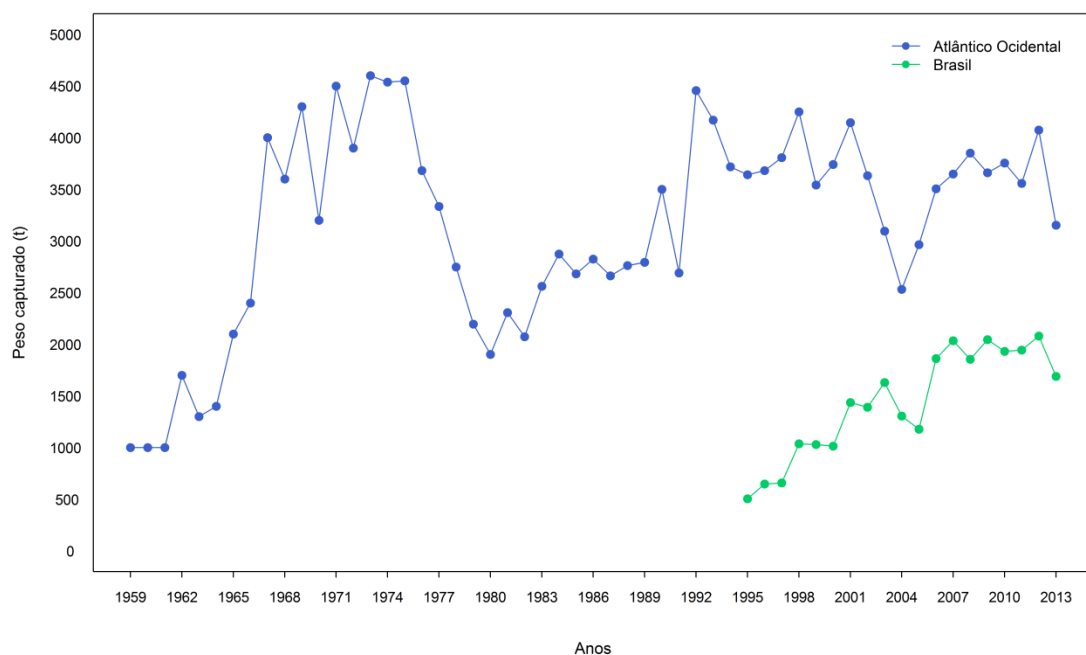


Figura 2: Peso capturado de *L. synagris* para todo o Atlântico ocidental entre os anos de 1959 a 2013 e para o Brasil entre 1995 a 2013.

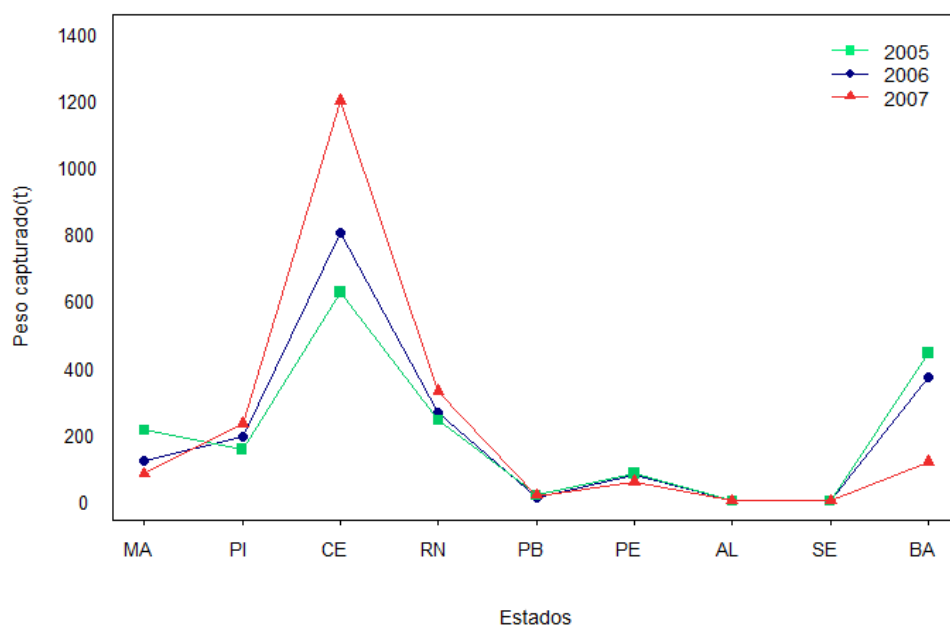


Figura 3: Peso capturado de *L. synagris* na região nordeste entre 2005 e 2007.

2.2. Modelagem estatística

Entende-se por modelo como sendo uma forma simplificada de descrever qualitativamente uma realidade complexa. Como afirmado por Parent & Rivot (2012), fazemos uso de modelos para sintetizar um conhecimento e especular sobre um determinado sistema. De forma bem genérica, um modelo pode ser decomposto em duas partes: uma parte determinística que é resumida em modelos matemáticos; e uma parte estocástica (aleatória), a qual consagra ao modelo a parte estatística (Kéry, 2010).

Hilborn & Mangel (1997), inspirados na expressão de Schnute (1987), declararam que nas ciências ambientais, mais especificamente na ecologia, o processo de modelagem envolve um verdadeiro trabalho de *detetive ecológico*, cujo objetivo consiste na busca constante de um modelo que mais se aproxime da realidade. No que diz respeito especificamente aos modelos aplicados aos estudos da dinâmica populacional de peixes, Schnute (1987) afirmou que, embora não consigamos avaliar todas as populações de uma determinada espécie de forma absoluta, podemos acessar isoladamente pequenas informações que, juntas, fazem parte de uma história maior.

Nesse sentido fica claro que os dados provenientes do âmbito marinho (e.x.: aspectos ecológicos de uma população) dificilmente poderão ser integralmente incorporados em um único modelo, haja vista serem extremamente complexos e dinâmicos em termos espaço-temporais. Em função disso, qualquer modelo é obrigado a simplificar todo o processo que está sendo avaliado, fato este que muitas vezes pode acarretar na deturpação dos dados e, conseqüentemente, na interpretação enviesada dos resultados.

Uma maneira pela qual podemos contornar parte deste problema é aceitar que certo grau de incerteza deva ser incorporado ao longo do ciclo que perfaz o processo de modelagem (Lemos, 2010). A figura 4 denota um ciclo que resume todo este processo, o qual conecta quatro principais etapas que contemplam a modelagem: *dados*, *conhecimento* (tais como os sumários fornecidos por modelos empíricos), *entendimento* (i.e., a inferência, em termos conceituais, sobre os processos subjacentes) e, por fim, o *processamento do modelo* (tradução de processos ambientais em equações matemáticas).

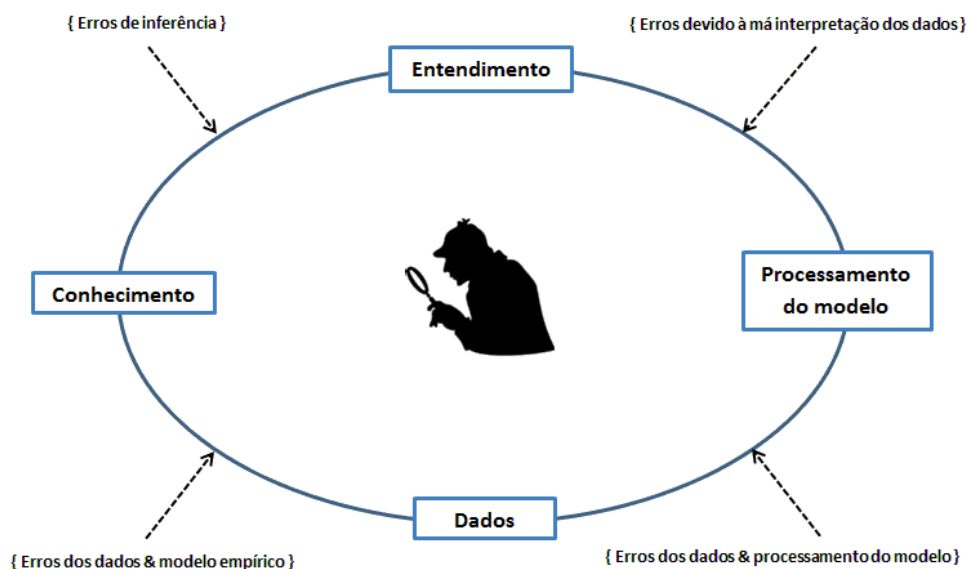


Figura 4: Esquema destacando as etapas de modelagem (adaptado de Lemos (2010) e inspirado em Hilborn & Mangel (1997)).

A partir da estrutura unificadora estabelecida entre as etapas da modelagem, o intercâmbio de informações advindos dos dados e dos modelos fornecerá um panorama geral de todo o procedimento, permitindo, assim, realizar declarações probabilísticas sobre os dados, parâmetros e complexidade dos modelos de modo a minimizar essas incertezas e, por consequência, a se aproximar da realidade que nos cerca.

À vista disso, como pesquisadores, o nosso trabalho é montar e reunir essas informações da mesma maneira que um detetive monta suas pistas através de uma imagem coerente. De acordo com os objetivos ressaltados no item 1.2, o propósito central deste trabalho não é construir uma única história, e sim fornecer outras vias de informações sobre *L. synagris* que possam dar mais suporte a “história maior” da espécie. Sendo assim, no presente estudo o trabalho de detetive ecológico foi efetuado pontualmente, no sentido que as investigações foram realizadas de maneira a minimizar os erros inerentes ao processo de modelagem.

2.3. Inferência Bayesiana

Considerando que a ideia fundamental dos modelos estatísticos é a descrição quali-quantitativa da realidade, e, reconhecendo que o processo de amostragem e modelagem dos dados contém diversas fontes de incertezas, é conveniente pensar que essas incertezas poderiam ser representadas matematicamente por meio de probabilidades (Mäntyniemi, 2006). Na área da Estatística, a teoria da probabilidade é essencial tendo em vista que é por meio dela que toda a inferência sobre os parâmetros de um modelo é realizada.

Dado às diferentes conceituações de probabilidade, o ramo da inferência estatística atualmente é dividido em duas vertentes: a inferência frequentista (ou clássica) e a inferência Bayesiana. Discussões filosóficas a respeito de ambas as abordagens vão além do escopo desta dissertação. Contudo, à título introdutório será brevemente discutido as principais diferenças existentes entre essas duas vertentes, destacando também as razões pelas quais se optou em fazer uso da inferência Bayesiana para análise dos dados.

De acordo com Hall (2011), a inferência frequentista considera a probabilidade em termos de frequências relativas de um evento contido em um grande número de ensaios, o qual, por sua vez, é definido no contexto de experimentos puramente aleatórios e bem definidos. A inferência Bayesiana, em contraste, fundamenta a probabilidade com base no grau de credibilidade individual para qualquer tipo de evento, mesmo quando um processo aleatório não está necessariamente envolvido. Mais especificamente, a probabilidade quantifica a plausibilidade atribuída a uma proposição cuja veracidade é incerta à luz dos conhecimentos disponíveis (Kinas & Andrade, 2010).

Embora a subjetividade na definição Bayesiana de probabilidades seja alvo de críticas, ela simboliza a pedra angular de toda a inferência Bayesiana, pois é justamente através dela que podemos incorporar informações pretéritas (*priori*) antes que o experimento seja efetivamente conduzido. Segundo Kinas & Andrade (2007), a probabilidade subjetiva sempre representa uma consequência das informações disponíveis e não meramente uma quantificação arbitrária. Além disso, esses mesmo autores ressaltam que um julgamento científico sobre a escolha dessas

probabilidades à priori é tão necessário quanto às decisões que se toma a respeito da escolha dos modelos mais adequados para cada tipo de dado.

Em termos conceituais a inferência Bayesiana é mais simples tendo em vista que todos os questionamentos impostos pelo pesquisador poderem ser respondidos a partir da análise da distribuição posterior via Teorema de Bayes (também conhecido por Teorema da probabilidade inversa) (Kinas & Andradem, 2010). Esse Teorema é diretamente derivado da probabilidade condicional, a qual é definida por:

$$p(\theta|Y) = \frac{p(\theta,Y)}{p(Y)}$$

Onde $p(\theta|Y)$ é distribuição dos parâmetros dadas as observações; $p(\theta,Y)$ é a distribuição conjunta dos parâmetros e das observações; e $p(Y)$ é a distribuição marginal das observações.

Prosseguindo com esse mesmo raciocínio, sabe-se que de acordo com o Teorema da Probabilidade Total, teremos:

$$p(\theta,Y) = p(Y|\theta) p(\theta)$$

Em que $p(Y|\theta)$ representa a distribuição das observações dado os parâmetros; e $p(\theta)$ é a distribuição a *priori* dos parâmetros.

Dado às equações anteriores, podemos então reescrever o teorema de Bayes conforme:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}$$

Sabendo que a distribuição marginal das observações $p(Y)$ independe do vetor de parâmetros θ , uma vez que ela representa uma constante normalizadora que visa reter $p(\theta|Y)$ dentro do intervalo de $[0,1]$, teremos a densidade a *posteriori* representada por:

$$p(\theta|Y) \propto p(Y|\theta) p(\theta)$$

Neste caso, a inferência Bayesiana pode ser decomposta em três componentes básicos:

- 1) A probabilidade à priori de um conjunto de parâmetros ($p(\theta)$), a qual usa a probabilidade como forma de quantificar a incerteza sobre θ antes de considerar os dados;
- 2) Os dados coletados ($p(Y|\theta)$) expressos pela função de verossimilhança, em que todas as variáveis estão relacionadas em um modelo de probabilidade específico;
- 3) A distribuição posterior ($p(\theta|Y)$), a qual expressa a incerteza dos parâmetros θ após considerar o produto da distribuição à priori e a verossimilhança.

Em vista disso, percebe-se que a distribuição posterior denota a forma mais completa para se expressar o estado do conhecimento sobre determinado fenômeno investigado (Kinas & Andrade, 2010). A essência da inferência Bayesiana reside exatamente no dinamismo interativo existente entre as probabilidades *a priori* e as funções de verossimilhança, as quais acarretarão conjuntamente na realocação da credibilidade denotada na distribuição posterior. Sob o paradigma Bayesiano, os dados podem ser acumulados e utilizados para atualizar as informações geradas através da análise de dados previamente disponíveis. Nesse sentido é possível inferir que a distribuição posterior de hoje pode se tornar a distribuição à priori de amanhã, fato este que não poderia ser realizado de acordo com o contexto clássico da inferência estatística.

Segundo Jaynes (2003), é nesse ponto também em que se revela uma das principais distinções entre as duas vertentes inferenciais, pois as probabilidades mudam conforme mudamos o nosso estado de conhecimento; as frequências, não. Além disso, percebe-se que certos questionamentos não podem ser respondidos conforme a definição frequentista de probabilidade. Em manejo de recursos pesqueiros, por exemplo, questões como *“qual a probabilidade de que os atuais níveis de captura de um determinado recurso pesqueiro continue sustentável”* e *“qual a probabilidade da área A apresentar maior potencial de conservação que a área B”* somente podem ser respondidos mediante o uso da inferência Bayesiana.

Sabendo que a incerteza é inerente em todas as atividades científicas, a sua inclusão no processo de tomada de decisão não é apenas desejável, e sim, essencial. A abordagem Bayesiana na inferência estatística e na tomada de decisão tem crescido consideravelmente nas últimas duas décadas não só na modelagem de dados ambientais, como também nas ciências pesqueiras (Hilborn *et al.*, 1993; Ludwig *et al.*, 1993). Conforme destacado por Ludwig *et al.* (1993), ignorar as incertezas presentes nos processos ecológicos já resultou repetidamente em catástrofes ambientais. Entretanto, o uso de abordagens Bayesianas tem demonstrado ser uma poderosa ferramenta estatística principalmente para quem trabalha na área ambiental, uma vez que por meio dela é possível incorporar todos os tipos de incertezas nas decisões de conservação e manejo mais pertinentes e, conseqüentemente, evitar o acontecimento catástrofes ambientais que podem, inclusive, serem irreversíveis.

Conforme mencionado no início deste tópico, foi visado apenas uma breve introdução sobre a estatística Bayesiana e destacar algumas vantagens em usá-la no manejo e conservação de recursos pesqueiros. Para maiores detalhamentos a respeito da incorporação da inferência Bayesiana no contexto ecológico, e pesqueiro recomendo a leitura dos artigos de Dennis (1996), Ellison (1996 e 2004), Clark (2005) e Kinas & Andrade (2007). No quesito teórico, indico a leitura dos livros de McCarthy (2007), Kinas & Andrade (2010) e Kruschke (2014).

2.4. Modelos de distribuição de espécies

Muito embora as ações humanas sejam o principal gatilho para a maior parte da degradação do meio ambiente, atualmente a sociedade vem tomando consciencia sobre a necessidade de conservar os seus recursos naturais, que são a fonte de sustento. Em função das mudanças ambientais, tem-se constatado que a distribuição da maioria das espécies é afetada, fato este que está constantemente em pauta na ecologia, biogeografia e biologia da conservação (Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005). Cada vez mais se percebe a necessidade em usar medições locais para avaliar as mudanças no ambiente, sejam elas regionais ou globais (Miller *et al.*, 2004). Dado a esse advento, a estatística tem

se tornando uma aliada fiel aos cientistas ambientais, uma vez que a extrapolação de dados ambientais para o espaço certamente será realizada mediante modelos estatísticos (Miller *et al.*, 2004; Peters *et al.*, 2004).

Neste sentido, os modelos de distribuição de espécies (SDM, do inglês *Species Distribution Modeling*), também conhecidos por modelos de distribuição geográfica, modelos de nicho ecológico, modelos bioclimáticos, entre outros, representam um exemplo desta realidade. De acordo com esses modelos, dados referentes à presença/abundância de uma espécie são extrapolados no espaço e tempo por intermédio de modelos estatísticos (Franklin, 2010). Em linhas gerais, os SDMs consistem em algum algoritmo que relaciona registros de ocorrência de uma espécie a um conjunto de variáveis ambientais e, desta forma, constrói-se uma representação espacial das condições que sejam mais favoráveis à sua sobrevivência (Giannini *et al.*, 2012) (Fig. 5).

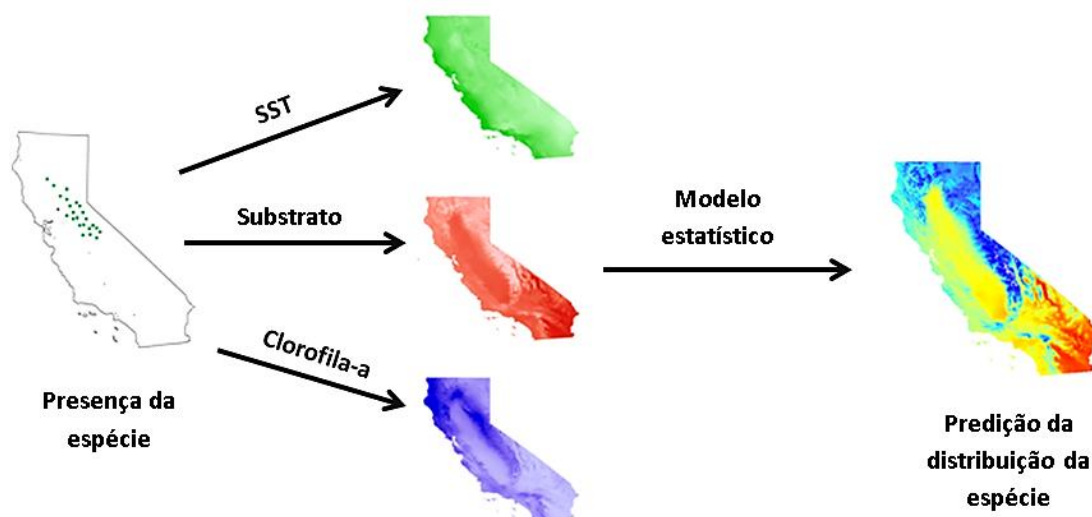


Figura 5: Esquema ilustrativo destacando os processos envolvidos durante a modelagem de distribuição de espécies¹.

Tendo em vista que a natureza é demasiadamente complexa para ser predita acuradamente em todos os aspectos que dizem respeito ao espaço e tempo, Guisan & Zimmermann (2000) ressaltam que os modelos estatísticos aplicados à ecologia

¹ Figura extraída e adaptada da página : <https://scienceasaverb.wordpress.com/2010/10/29/introduction-to-ecological-niche-modeling-environmental-niche-modeling-species-distribution-modeling-part-two-what-is-niche-modeling/> (acessa em 16.10.2015)

sempre irão falhar em relação a um dos três pilares que os sustenta: precisão, generalidade e realidade (Fig. 6). Segundo esses autores, três grupos de modelos são derivados de cada um desses pilares: o grupo I está principalmente voltado à generalidade e precisão e envolvem modelos analíticos, à exemplo dos modelos logísticos que quantificam o crescimento populacional; o grupo II foca na generalidade e realidade, sendo conhecidos por modelos mecanicistas, fisiológicos ou de processos; por fim, o grupo III desenvolve modelos voltados à precisão e realidade, que é onde os modelos estatísticos e, conseqüentemente os SDMs, se enquadram.

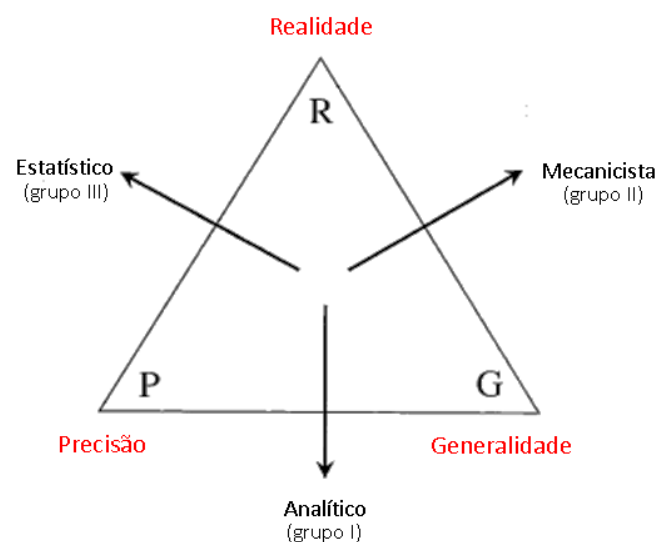


Figura 6: Tipos de modelos ecológicos com destaque para suas respectivas penalizações (Adaptado de Guisan & Zimmermann (2000)).

Nas últimas três décadas têm-se dispensado um grande esforço para desenvolver modelos estatísticos robustos, que sejam capazes de incluir cenários cada vez mais condizentes com a realidade no que diz respeito às relações espécie-ambiente. Com base nisso, Elith *et al.* (2006) propuseram uma classificação de SDMs em dois grupos principais, baseando-se exclusivamente nos dados que alimentam esses modelos. No primeiro grupo estão incluídos os modelos que utilizam apenas os dados de presença. Já no segundo grupo estão inseridos o

modelos que empregam tanto dados de presença quanto de ausência, cujo intuito é reduzir os erros de falsos positivos.

Dentre esses modelos, pode-se citar como exemplo as redes neurais artificiais (ANN, do inglês *Artificial Neural Networks*), entropia máxima (ME, do inglês *Maximum Entropy*), envelopes climáticos (CE, do inglês *Climatic Envelopes*), árvores de regressão e classificação (CART, do inglês *Classification and Regression Trees*) e modelos de regressão, tais como os modelos lineares generalizados (GLM, do inglês *Generalized Linear Models*) e modelos aditivos generalizados (GAM, do inglês *Generalized Additive Models*). Para maiores detalhamentos a respeito de cada um desses exemplos metodológicos aconselho como leitura indispensável os trabalhos de revisão de Guisan & Zimmermann (2000), Guisan & Thuiller (2005), Elith & Leathwick (2009), e o livro de Franklin (2010).

Embora essas aplicações sejam consideradas robustas, comumente elas englobam modelos explanatórios que visam apenas verificar a relação entre a variável resposta (e.x., presença ou ausência de uma espécie) e um conjunto de preditores ambientais (e.x., batimetria, concentração de clorofila-a quando lidamos com o ambiente marinho) sem, contudo, considerar explicitamente o componente espacial (Latimer *et al.*, 2006; Ciannelli *et al.*, 2008; Elith & Leathwick, 2009). Em função disto, esses modelos podem resultar em uma fraca caracterização das respostas das espécies frente aos fatores ambientais, uma vez que podem subestimar o grau de incerteza associados aos seus preditores (Latimer *et al.*, 2006).

A incorporação de um componente espacial, também conhecido pelo termo de autocorrelação espacial, ainda tem sido um assunto tratado timidamente em modelos preditivos. Entretanto, percebe-se que à medida em que se deseja aumentar o realismo desses modelos, faz-se necessário considerar esse tipo de autocorrelação, haja vista a maioria dos dados ecológicos possuírem algum tipo de dependência espacial (Dormann, 2007). A próxima seção irá tratar brevemente da área da estatística que trata exclusivamente esta questão, destacando a sua relevância em pesquisas pesqueiras.

2.5. Estatística espacial

O padrão de distribuição individual dentro de uma unidade de estoque pode ser de três tipos: uniforme, aleatório ou agregado (King, 2007). Tendo em vista que o ambiente marinho é um domínio extremamente heterogêneo, e que uma determinada espécie marinha apresenta limitações biológicas e ecológicas, é comum observar que os recursos biológicos, à exemplo da maioria dos peixes, apresentam uma distribuição agregada (King, 2007). Isso denota claramente que os organismos marinhos têm sua distribuição dependente do espaço, podendo, assim, serem descritos conforme a primeira lei da Geografia definida por Tobler (1970): “Tudo está relacionado entre si, porém coisas próximas estão mais relacionadas entre si do que coisas distantes” (Nishida & Chen, 2004; Kneib *et al*, 2008).

O fato dos recursos biológicos estarem relacionadas espacialmente, evidencia a violação de um dos principais pressupostos anunciados pela estatística clássica, a qual afirma que os dados devem ser independentes e identicamente distribuídos entre si (Plant, 2012). A correlação espacial deve, portanto, ser considerada em modelos preditivos, tendo em vista que as espécies geralmente estão sujeitas a fatores ambientais similares (Nóbrega, 2008; Muñoz *et al.*, 2013). Além disso, esses tipos de modelos tornam-se extremamente poderosos, porque eles promovem não só a estimação dos processos que influenciam a distribuição das espécies, como também permitem a possibilidade de prever as suas ocorrências em áreas não amostradas (Chakraborty *et al.*, 2010). Segundo Nóbrega (2008), é justamente devido a essa dependência espacial dos organismos que surge um campo de ação propício para a atuação da estatística espacial.

A modelagem espacial é uma aquisição recente na literatura estatística e tem seu ponto de referência marcado no livro de Cressie (1993). Segundo esse autor, a estatística espacial permite fornecer um maior realismo ao processo estudado e vem se ampliando por vastas áreas do conhecimento, como a Geologia, Climatologia, Epidemiologia, Ecologia, Astronomia ou qualquer disciplina que trate de dados coletados em diferentes localizações espaciais (Cressie, 1993).

Em modelos de regressão tradicionalmente empregados para avaliar a distribuição das espécies, assume-se que os dados observados sejam representados pelo valor esperado e este, por sua vez, esteja associado a um ruído

que é expresso pela variância (σ^2). Quando tratamos de autocorrelação espacial, essa variância é composta por um segundo componente, o qual especifica uma estrutura de covariância entre todos os pontos amostrados. Maiores detalhamentos a respeito desta função de covariância serão expostos na próxima seção. Por hora focar-se-á somente na fundamentação teórica envolvida no campo da estatística espacial.

Em linhas gerais, a estatística espacial trata de dados georreferenciados, também conhecidos por dados espacialmente estruturados, que possuem caráter estocástico e envolvem dois principais componentes: uma área \mathcal{D} na qual os dados são coletados, e um vetor Y representado por uma ou mais variáveis mensuradas em uma posição S georreferenciada $\{Y(s_1), Y(s_2), \dots, Y(s_n)\}$, tal que:

$$\{Y(s): s \in \mathcal{D}\}, \mathcal{D} \subset \mathbb{R}^d$$

Aonde d representa a dimensão (1, 2 ou 3) em que o experimento foi desenvolvido.

Como os valores de Y nas diferentes localizações \mathcal{D} representam vetores aleatórios, o conjunto desses vetores constituem uma realização do processo espacial. Dependendo de como \mathcal{D} e $Y(S_n)$ são considerados, isto é, fixos ou aleatórios, costuma-se dividir os dados espaciais em três tipos (Cressie, 1993; Banerjee *et al.*, 2004):

- ♦ Dados ponto-referenciados: Também conhecido por geoestatística, esse tipo de dado assume que $Y(S_n)$ é um vetor aleatório em uma localização $s \in \mathbb{R}^d$, em que s varia de forma contínua ao longo de um \mathcal{D} fixo. O objetivo principal desse tipo de análise é a reconstrução dessa superfície contínua (também conhecida por campo latente, o qual representa a porção não observada) tendo por base um conjunto finito de dados observados (pontos georreferenciados). Ex.: Concentração de poluentes, Precipitação pluviométrica e abundância de uma espécie.
- ♦ Dados de área: $Y(S_n)$ e \mathcal{D} também são considerados como aleatório e fixo, respectivamente. Contudo, o espaço \mathcal{D} agora é segmentado em um número finito de áreas cujos limites são bem definidos. Em outras palavras, pode-se pensar em \mathcal{D} como sendo um processo contínuo

separado por zonas. O propósito desse tipo de análise reside na modelagem do padrão espacial dos dados e no estabelecimento de associações com outras variáveis. Ex.: Estatísticas de saúde e dados de censos.

- ♦ Dados de padrões pontuais: Diferentemente dos casos anteriores, agora tanto $Y(S_n)$ quanto \mathcal{D} representam variáveis aleatórias. Neste caso, \mathcal{D} é considerado um processo pontual ou um subconjunto de \mathbb{R}^d , ao passo que $Y(S_n)$ é um vetor aleatório na posição $S_n \in \mathcal{D}$. O principal propósito deste tipo de análise consiste no entendimento do padrão espacial (unifome, agregado ou aleatório) do processo estudado. Ex.: Distribuição de espécies vegetais e localização de epicentros.

2.6. Modelagem de habitats essenciais via Geoestatística Bayesiana

Tradicionalmente a avaliação da abundância espacial de peixes é realizada mediante técnicas de regressão ou Geoestatística (Ciannelli *et al.*, 2008). Embora nos procedimentos de regressão seja possível incorporar as coordenadas como covariáveis, esse tipo de modelo apenas avalia a tendência da média baseado nas coordenadas e, em função disso, possui unicamente efeitos fixos. A Geoestatística, por sua vez, permite construir modelos espacialmente estruturados com efeitos aleatórios e, desta forma, avaliar também diretamente a correlação espacial entre as coordenadas (Ciannelli *et al.*, 2008, Krainski & Lindgren, 2015).

O termo *Geoestatística* foi introduzido por Matheron (1963) e emergiu efetivamente na década de 80 da fusão interdisciplinar entre a Geologia, Engenharia de mineração, Estatística e Matemática (Cressie, 1993; Diggle & Ribeiro, 2007). Conforme exposto no item 2.4, a Geoestatística é caracterizada por um processo estocástico $\{Y(s): s \in \mathcal{D}\}$, onde \mathcal{D} consiste em um subconjunto de um espaço d -dimensional fixo, e $Y(s)$ comumente constitui uma distribuição Gaussiana multivariada que representa uma realização parcial do processo estocástico

(Banerjee *et al.*, 2004). Esse processo pode ser descrito por um modelo básico que segue a seguinte estruturação:

$$Y(s) = \mu(s) + w(s) + \varepsilon(s)$$

$$g(\mu(s)) = x^T(s)\beta$$

Esta estrutura comporta uma média $g(\mu)$ que está ligada a um preditor estruturado $Y(s)$ que acomoda os efeitos das covariáveis de modo aditivo, bem como resíduos que englobam tanto processos puramente espaciais ($w(s)$) quanto processos independentes ($\varepsilon(s)$), caracteristicamente conhecidos por *efeito pepita* (τ^2) na literatura Geoestatística.

Ao modelarmos tanto a aleatoriedade contida nos dados observados quanto as incertezas associadas aos parâmetros do modelo, é conveniente estruturá-lo de forma hierárquica e sob contexto Bayesiano, tendo em vista que isso permite incorporar tal estrutura naturalmente, resultando em consequência, em uma inferência estatística mais robusta (Cressie *et al.*, 2009). Quando tratamos de modelos conforme expresso anteriormente, é importante mencionar que estes modelos hierárquicos podem ser enquadrados dentro dos modelos de regressão estruturados aditivamente (STAR, do inglês *Structured Additive Regression models*) (Rue *et al.*, 2009). Ademais, de acordo com Rue *et al.* (2009), os modelos geoestatísticos hierárquicos Bayesianos podem ser tratados como um caso particular dos modelos STAR, denominados de modelos Gaussianos latentes, os quais atribuem prioris Gaussianas à todos os componentes do preditor aditivo.

Nesses modelos, se assumirmos que $Y(s)$ é contínuo ao longo do espaço D , então teremos um campo aleatório Gaussiano (GRF, do inglês *Gaussian random field*) indexado continuamente (Lindgren *et al.*, 2011). Em outras palavras, esse campo representa explicitamente todos os processos estocásticos espaciais que podem exercer algum efeito sobre o fenômeno investigado.

Na maioria dos casos o GRF é especificado por meio de duas funções: uma que engloba a média $\mu(\cdot)$ denotada por $\mu = (\mu(s_i))$; e outra referente à covariância $C(\cdot, \cdot)$ cuja matriz é descrita na forma de $\Sigma = (C(s_i, s_j))$. A função de covariância é frequentemente tratada como uma função que depende apenas da distância absoluta entre dois pontos amostrados $\|s_i - s_j\|$ (Cressie, 1993; Banerjee *et al.*, 2004;

Diggle & Ribeiro, 2007). Neste caso, o processo espacial Gaussiano é considerado *estacionário* se sua média é constante ao longo de toda a área estudada $\mu(s_n) = \mu$, e *isotrópico* se a função de covariância depender apenas da distância Euclidiana entre os pontos amostrados.

De acordo com Diggle & Ribeiro (2007), a forma mais prática para se avaliar o comportamento empírico da estacionariedade da covariância consiste em tratar a correlação entre dois pontos $\rho(s_i, s_j)$ como uma função inversa da distância $h = \|s_i - s_j\|$, isto é, a correlação entre os pontos amostrados aumenta à medida que diminui a distância entre eles e vice-versa. Esse comportamento pode ser descrito por muitos modelos, à exemplo do Exponencial, Linear, Gaussiano, Esférico e Matérn (para maiores detalhes, vide Baerjee *et al.*, 2004). Contudo, dentre esses modelos, àqueles pertencentes à família Matérn são os mais populares na modelagem dos GRFs uma vez que eles oferecem maior flexibilidade à suavização do processo espacial, o qual é crucial durante os procedimentos de interpolação. A função de correlação de Matérn pode ser expressa conforme:

$$\rho(h) = \frac{(s_v h)^v K_v(s_v h)}{2^{v-1} \Gamma(v)}$$

Em que h é a distância entre dois pontos, $K_v(\cdot)$ é a função modificada de Bessel de ordem $v > 0$, onde v é um parâmetro que confere forma e, consequentemente, determina a suavização do processo estudado, s_v representa um parâmetro de escala e $\Gamma(v)$ é uma função gama para v .

Dependendo do tamanho de v , a função de Matérn pode assumir diferentes formas pertencentes à distintas funções de correlação, tais como a Gaussiana quando $v \rightarrow \infty$ e a exponencial quando $v = 0.5$ (Paciorek & Schervish, 2006). A função de correlação também pode ser reparametrizada em termos do alcance ($r = \sqrt{8v}/s_v$)

em que:

$$\rho(h) = \frac{1}{2^{v-1} \Gamma(v)} \left(\frac{\sqrt{(8v)h}}{r} \right)^v K_v \left(\frac{\sqrt{(8v)h}}{r} \right)$$

Esta reparametrização acaba tendo um significado mais intuitivo, uma vez que o parâmetro r pode ser interpretado como sendo a distância mínima em que dois pontos amostrais não estão mais correlacionados.

Embora os GRFs sejam convenientes no ponto de vista analítico para realizar esse tipo de inferência, o poder computacional ainda tem sido um fator limitante no que diz respeito a fatorização das densas matrizes ($n \times n$) de covariâncias que são cúbicas em sua dimensão $\mathcal{O}(n^3)$ (Lindgren *et al.*, 2011). Devido a isto, procura-se contornar este problema particionando a área de estudo em uma retícula de pequenos quadrados - conhecido por *grid* na terminologia geoestatística - aonde serão projetadas os resultados provenientes da inferência espacial.

Quando se modela um GRF com uma função de covariância do tipo Matérn, é comum recorrer à aproximação do Campo Aleatório Markoviano Gaussiano (GMRF, do inglês *Gaussian Markov Random Field*) para evitar o problema do grande “ n ” (Simpson *et al.*, 2012a e 2012b). Enquanto que o GRF é indexado continuamente, um GMRF é um GRF indexado de forma discreta (Lindgren *et al.*, 2011). Neste aspecto, assume-se que um GMRF seja representado por um vetor aleatório de dimensão finita seguindo uma distribuição Gaussiana multivariada (Rue & Held, 2005).

O termo Markoviano surge nesse contexto quando os GMRFs discretizados se restringem a atender uma das propriedades intrínsecas de Markov, que é a da independência condicional. Isto é, se assumirmos um vetor aleatório denotado por $z = (z_1, z_2, z_3)^T$, então z_1 e z_2 serão considerados condicionalmente independentes dado z_3 se, para um valor conhecido de z_3 , não será fornecido informação adicional a respeito da distribuição de z_1 ao considerar z_2 . Matematicamente, essa propriedade pode ser facilmente representada por:

$$\pi(z) = \pi(z_1 | z_3) \pi(z_2 | z_3) \pi(z_3)$$

Em termos espaciais, diz-se que em um GMRF a distribuição condicional total $\pi(z_i | z_{-i})$, $i = 1, \dots, n$, depende apenas de um conjunto finito de vizinhos ∂_i para cada local i (Fig. 7).

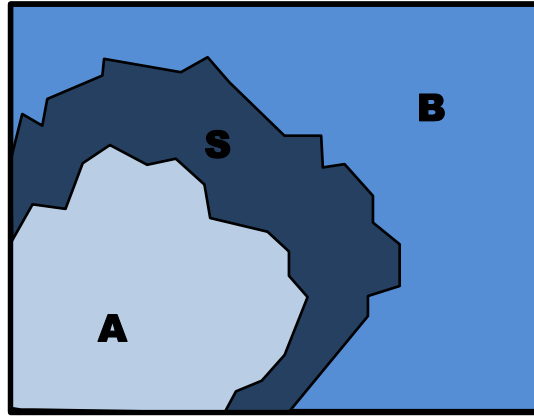


Figura 7: Esquema ilustrativo do pressuposto Markoviano espacial. Se para qualquer conjunto S , $\{x(s): s \in A\}$ for independente de $\{x(s): s \in B\}$ dado $\{x(s): s \in S\}$, então o campo $x(s)$ cumpre o pressuposto Markoviano espacial (Adaptado de Simpson *et al.*, 2012a).

Segundo Rue *et al.* (2009), a propriedade Markoviana é convenientemente codificada na matriz de precisão (inverso da covariância) Q : $Q_{ij} = 0$ se z_i e z_j forem independentes e condicionados a z_{ij} . Logo, a matriz que antes era extremamente densa, torna-se esparsa devido ao grande número de zeros que a contempla devido a propriedade markoviana e, por consequência, a fatorização da matriz Q é realizada a um custo de $\mathcal{O}(n^{3/2})$ para um GMRF bidimensional (Rue *et al.*, 2009; Lindgren *et al.*, 2011). Esse é o ponto chave que leva a adoção de um GMFR, haja vista se economizar enormemente o tempo computacional para calcular essa matriz, bem como apresentar vantagem quando se estrutura um modelo de forma hierárquica.

Diversas metodologias têm sido desenvolvidas para realizar a aproximação de um GMRF quando se está modelando um GRF com covariância de Matérn. Porém, a tendência dos últimos anos é fazer uso de uma abordagem alternativa proposta em 2011 por Lindgren *et al.*, a qual usa uma aproximação estocástica fraca para uma equação diferencial estocástica parcial (SPDE, do inglês *Stochastic Partial Differential Equations*). Neste caso, esses autores partem do pressuposto de que um GRF $x(u)$ com função de covariância de Matérn é uma solução linear fracionada da SPDE conforme denotado por:

$$(k^2 - \Delta)^{\alpha/2} (\tau x(s)) = W(s),$$

$$s \in \mathbb{R}^d, \alpha = v + \frac{d}{2}, k > 0, v > 0$$

Em que Δ é um operador Laplaciano; k é o parâmetro espacial de escala; α e τ controlam a suavidade e a variância do efeito espacial, respectivamente; e $W(s)$ é um ruído gaussiano branco (ruído puro).

Pela equação anterior pode-se constatar que a função de correlação não é modelada diretamente. Ao contrário dos procedimentos rotineiramente adotados pela estatística espacial, nesse novo procedimento a função de covariância é ignorada, circunstância esta que torna a inferência e predição computacionalmente muito mais viável (Simpson *et al.*, 2012b). Além disto, a aproximação via SPDE divide o campo latente em malhas triangulares e cria, desta forma, uma malha de índice aonde a predição do fenômeno estudado será projetada (Fig. 8).

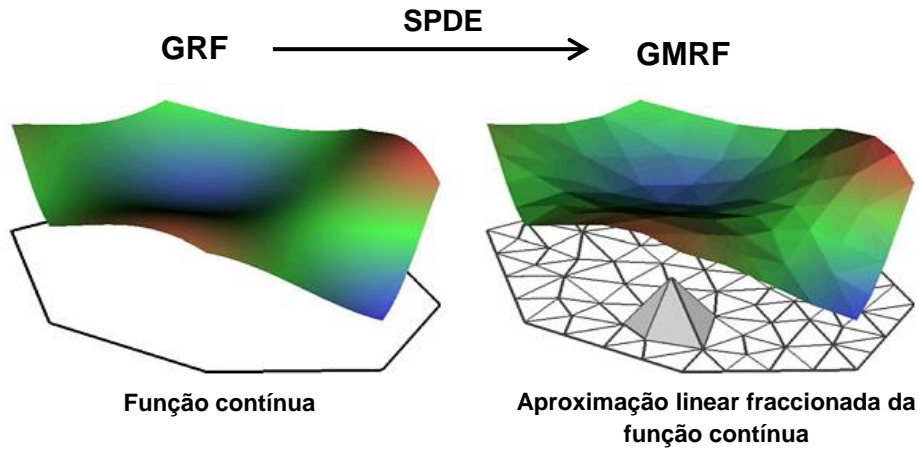


Figura 8: Ilustração da aproximação de uma função linear fracionada em uma malha triangular (Adaptado de Simpson *et al.*, 2012a).

Segundo Muñoz *et al.* (2013), a triangulação possui duas vantagens em relação a um grid regular: (1) economiza o tempo computacional, visto que os locais de previsão são menos numerosos que aqueles adotados em uma malha regular; e (2) a triangulação costuma ser mais densa em locais amostrados, fornecendo assim informações mais precisas.

Esse passo é crucial durante a construção de todo o modelo espacial, devendo-se prestar muita atenção na escolha da triangulação certa, tendo em vista que ela será utilizada para predição final. Lindgren *et al.* (2011) ressaltam que é fundamental a malha apresentar triângulos que tenham tamanhos e formas regulares. Além disso, a triangulação precisa ser expandida espacialmente com o intuito de evitar o efeito de borda, já que ele causa inflação da variância nas bordas.

A função *inla.mesh.2d* do pacote R-INLA (<http://www.r-inla.org>) dispõe diversos argumentos que possibilitam especificar a triangulação. Os argumentos *loc* (ou *loc.domain* em casos em que se tem uma extensão de um domínio) e *max.edge* são indispensáveis durante a elaboração da malha, tendo em vista que eles especificam a matriz dos pontos amostrais e o comprimento máximo dos lados dos triângulos, respectivamente. Outros argumentos comumente utilizados em conjunto com o *max.edge* são os argumentos *min.angle* e *cutoff*, os quais permitem controlar tanto o tamanho quanto a forma dos triângulos. Caso a área de estudo possua algum domínio em específico, à exemplo de ilhas ou linhas de costa que delimitam a região costeira, pode-se fazer uso do argumento *boundary*. Este argumento é composto por uma lista que especifica as delimitações desses domínios (também chamados polígonos) e desta forma a malha final será construída mediante a exclusão do polígono especificado. Para maiores detalhes sobre a criação da malha recomendo a leitura do módulo disponibilizado por Krainski (2015). Este módulo, contudo, está em constante atualização devendo-se, portanto, acessar periodicamente o site <http://www.r-inla.org/examples/tutorials/spde-tutorial> para acompanhar as novas atualizações.

A figura 9 demonstra algumas malhas construídas com especificações distintas cujo resultado se reflete no número de triangulações criadas. Para as predições deste trabalho utilizou-se a malha representada na figura 9C, visto que as formas e tamanhos dos triângulos foram relativamente regulares.

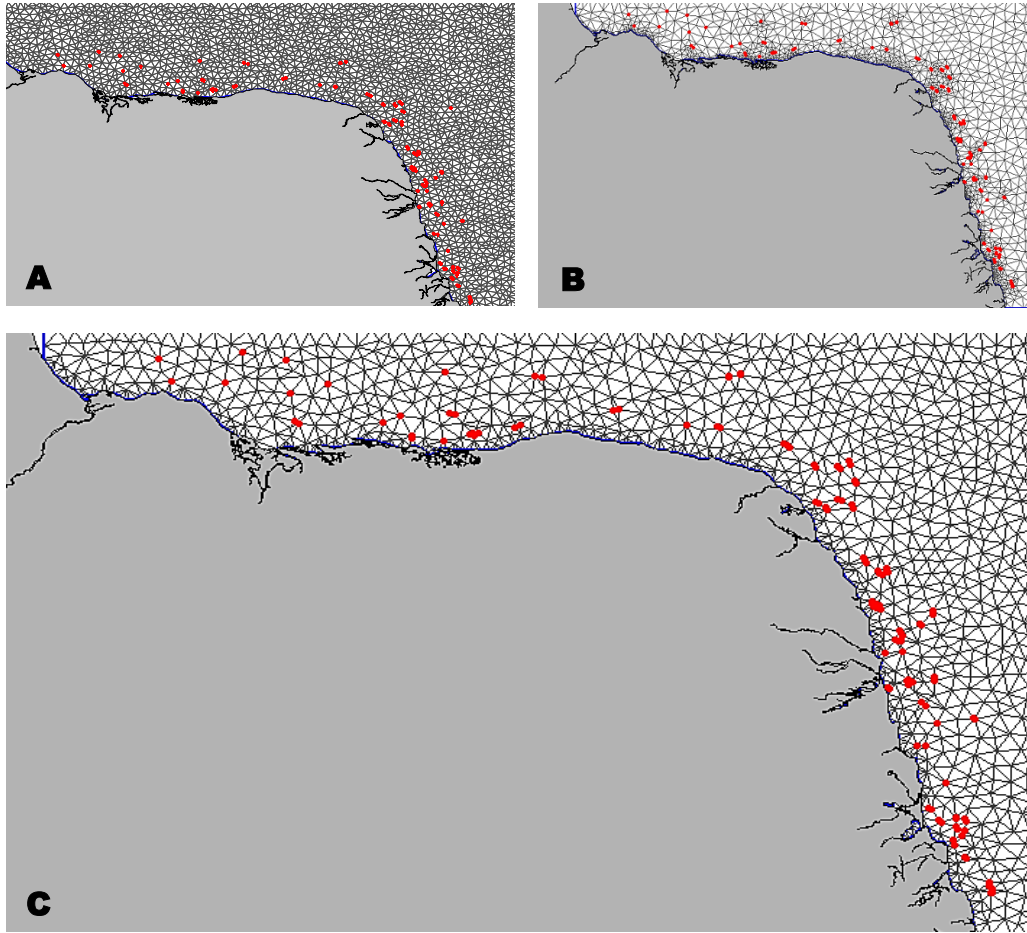


Figura 9: Malhas triangulares com diferentes restrições (A - 2.831 triângulos; B - 8.466 triângulos; C - 12.740 triângulos).

2.6.1. Formulação hierárquica do modelo espacial

Um modelo espacial Bayesiano é tipicamente definido em três estágios hierárquicos:

$$\begin{aligned} y_i | x_i, w_i, \theta_1 &\sim \pi(y_i | x_i, w_i, \theta_1) \\ w_i | \theta_2 &\sim \pi(w_i | \theta_2) \\ \theta &\sim \pi(\theta) \end{aligned}$$

O primeiro estágio diz respeito ao processo que origina os dados e pode ser expresso de forma similar aos GLMs. Nesse sentido, uma variável resposta Y_i (ex.: CPUE, presença, n° de indivíduos, etc.) possui uma distribuição F pertencente à família

exponencial com média definida por $E(y_i) = \mu_i$, a qual se conecta linearmente com M covariáveis preditoras X_m através da função de ligação $g(\cdot)$, tal que:

$$Y_i = F(\mu_i) \quad ; \quad i = 1, 2, \dots, n$$

$$g(\mu_i) = \eta_i = \beta_0 + \sum_m^M \beta_m X_m + W_i$$

Onde η_i é preditor linear da variável resposta, β_0 é o intercepto; β_m é um vetor dos regressores que quantifica o efeito das covariáveis e W_i é o efeito aleatório estruturado espacialmente.

O segundo estágio diz respeito à especificação do campo latente e que, no contexto espacial, é representado por W_i . Esse efeito espacial pode ser modelado de acordo com um GMRF, o qual possui distribuição Gaussiana com média $\mu = 0$ e uma matriz de covariância $Q(\kappa, \tau)$ com estrutura de Matérn. Por fim, o terceiro estágio é composto por todos os parâmetros desconhecidos do modelo ($\theta = \theta_1 + \theta_2$), também chamados por *hiperparâmetros*, e podem ser resumidos pelo vetor $\theta = \{\beta_0, \beta_M, \kappa, \tau, \Phi, \sigma_\varepsilon^2\}$.

2.6.2. Inferência Bayesiana via INLA

De acordo com Wikle (2003), ao seguirmos com uma abordagem Bayesiana deseja-se descobrir a distribuição posterior do processo ecológico, bem como os parâmetros dos modelos que são atualizados pelos dados. Como já denotado na seção 1.3.1, a distribuição posterior pode ser obtida conforme o Teorema de Bayes, aqui denotado por:

$$\begin{aligned} \pi(\text{processo}, \text{parâmetros} | \text{dados}) \\ \propto \pi[\text{dados} | \text{processo}, \text{parâmetros}] \cdot \pi[\text{processo} | \text{parâmetros}] \\ \cdot \pi[\text{parâmetros}] \end{aligned}$$

Usualmente a distribuição posterior é obtida por meio de simulações estocásticas, à exemplo dos métodos MCMC. Entretanto, recentemente Rue *et al.* (2009) propuseram uma nova abordagem, conhecida pelo acrônimo INLA (do inglês

Integrated Nested Laplace Approximations). Diferentemente dos métodos MCMC, o INLA substitui as simulações por aproximações determinísticas, que além de serem muito mais precisas, são consideravelmente mais rápidas no quesito computacional (Rue *et al.*, 2009).

No caso dos modelos hierárquicos Bayesianos espaciais, o principal objetivo da abordagem INLA é a obtenção das distribuições marginais a posteriori das suas variáveis latentes e dos seus hiperparâmetros. Podemos aplicar facilmente a expressão anterior se resumirmos todos os parâmetros das variáveis latentes no vetor $\theta = (\beta_0, \beta_M, \xi)$, e todos os hiperparâmetros no vetor $\psi = (\Phi, \kappa, \tau, \sigma_\varepsilon^2)$, em que ξ é o GMRF, Φ é o parâmetro da distribuição da variável resposta, κ e τ são os parâmetros da função de correlação de Matérn e σ_ε^2 é ruído branco, também conhecido por efeito pepita na terminologia geoestatística. Teremos, assim, uma distribuição posterior conjunta denotada por:

$$\pi(\theta, \psi | y) \propto \pi(y | \theta, \psi) \cdot \pi(\theta | \psi) \cdot \pi(\psi)$$

A obtenção da distribuição marginal posterior para cada uma das variáveis latentes e dos hiperparâmetros é obtida por aproximações numéricas. O cálculo dessas aproximações são constituídos por três etapas consecutivas, cada qual está contida no próprio nome *INLA*. Em suma, na primeira etapa são estimadas as distribuições posteriores conjunta dos hiperparâmetros (θ) via aproximações de Laplace. Em seguida, cálculos aproximados para as distribuições a posteriori do campo latente dado os dados (y) e hiperparâmetros (θ) são computados; daí o termo aninhado (*nested*, em inglês) haja vista estas aproximações estarem aninhadas uma dentro da outra. A última etapa, por fim, combina as duas etapas antecedentes por meio de uma integração numérica a fim de obter a variável latente marginal – daí o termo integrado (*integrated*, em inglês). Maiores detalhamentos teóricos podem ser verificados no trabalho de Rue *et al.* (2009).

Capítulo 3

MATERIAL & MÉTODOS

3. Caracterização da área de estudo

Situado no Nordeste do Brasil, o Estado do Rio Grande do Norte está localizado em uma importante zona de transição da linha da costa que muda abruptamente sua direção de sentido sul-norte para leste-oeste (Garcia Jr. *et al.*, 2015). Posicionada no extremo leste da Margem Continental Brasileira, a Bacia Potiguar ocupa principalmente o Estado do Rio Grande do Norte e uma fração minoritária do Estado do Ceará (Tabosa, 2006b). Topograficamente, de acordo com Amaro & Ferreira (2012) a plataforma continental da Bacia Potiguar é relativamente estreita e possui uma largura de apenas 40 a 45 km com no máximo 60 m de profundidade.

Em termos gerais, o Rio Grande do Norte possui uma extensão territorial de 53.077,3 km², e é acompanhado por uma faixa litorânea composta por aproximadamente 410 km de praias (IDEMA, 2002). Segundo Vital (2005), a costa norte-rio-grandense pode ser dividida em duas áreas: uma área que compreende o litoral oriental, localizada entre os municípios de Rio do Fogo e Baía Formosa (divisa com o Estado da Paraíba) e o litoral norte, situado entre os municípios de Tibau (divisa com o Estado do Ceará) e Touros. Essa segregação é dada principalmente em função da localização geográfica do Estado potiguar, bem como devido às diferenças tectônicas e climáticas existentes entre ambos os setores. Essas divergências, por sua vez, influenciam diretamente no padrão de circulação oceânica e no regime de direção dos ventos, os quais irão conjuntamente modelar o litoral do Rio Grande do Norte (Vital, 2005).

O Litoral Oriental possui 166 km de extensão e representa 41 % de todo o litoral do estado. Nesta porção, o clima é classificado como tropical quente úmido e sub-úmido (Nimer, 1989). A precipitação pluviométrica oscila entre 800 a 1200 mm por ano, com o período chuvoso concentrando-se entre os meses de fevereiro a julho (Vital, 2005). Os ventos alísios e úmidos provêm predominantemente do sudeste e apresentam velocidade variando entre 3,8 a 5 m/s (Vital, 2005). Banhada continuamente por arenitos de praia (*beach rocks*) grandes, a região apresenta três pequenos estuários com poucas ramificações, bem como numerosos recifes de corais pequenos (Garcia Jr. *et al.*, 2015).

O Litoral Norte, em contra partida, representa a maior porção do litoral norte-rio-grandense (59%) com 244 km de extensão. Nesta área o clima é tropical quente e seco ou semi-árido, expondo duas estações pluviométrica bem definidas: um período de seca (maio a dezembro), resultante do afastamento da Zona de Convergência Intertropical (ZCIT) da costa que provoca ausência de chuvas e o aparecimento de ventos mais fortes; e uma estação chuvosa (janeiro a abril), estando esta associada ao deslocamento da ZCIT para o sul e no abrandamento da força dos ventos (Nimer, 1989; Tabosa, 2006a). A pluviometria é considerada irregular visto oscilar entre 601 a 854 mm por ano e, em casos extremos, podendo chegar até 2.238 mm por ano (Tabosa, 2006a). Os ventos provêm preponderadamente de leste-nordeste, com uma velocidade média anual de 6,2 m/s. No período de agosto a dezembro os ventos tendem a ser mais fortes, podendo atingir até 9 m/s no município de Macau (Vital, 2005). Como já denotado por Garcia Jr. *et al.* (2015), diferentemente da costa leste, esta porção do litoral possui três grandes complexos estuarinos fortemente ramificados. Além disso, também estão presentes alguns poucos arenitos de praia e recifes de corais.

No que diz respeito às correntes de superfície da plataforma nordestina, a circulação superficial é prevalecida pela Corrente Sul-Equatorial (CSE) a qual flui de leste a oeste no Oceano Atlântico. Nas imediações do talude continental próximo ao Estado da Bahia (entre 12° S e 14° S), esta corrente se bifurca e dá origem à Corrente do Brasil (CB) que flui para o sul com uma velocidade que varia de 20-60 cm/s, e à corrente Norte Brasileira (CNB) dirigida para o norte e oeste com velocidades entre 55-60 cm/s (França *et al.*, 1976; Stramma *et al.*, 1990). A princípio a velocidade de ambas as correntes são caracteristicamente lentas, variando em torno de 20-60 cm/s. Todavia, dado ao encontro da CNB e o ramo central da CSE nas imediações de 4°S, 36°W (Fig. 10), é observado um aumento significativo da CNB para 55-60 cm/s (Lumpkin & Garzoli, 2005).

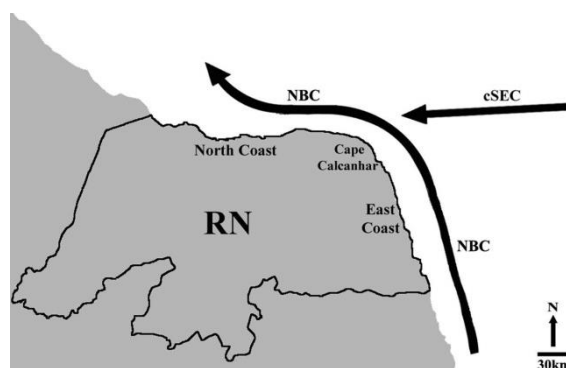


Figura 10: Representação esquemática da corrente Norte Brasileira (NBC) que flui pela costa do Rio Grande do Norte, destacando seu encontro com a corrente Sul-Equatorial (eSEC) (Adaptado de Garcia Jr. *et al.* (2015)).

3.1. Amostragem

Os dados do presente estudo resultaram da fusão de dois projetos intitulados por “*Viabilidade e sustentabilidade da utilização de redes de emalhar na pesca artesanal da costa norte do Rio Grande do Norte*” e “*Exploração pesqueira na costa leste do Rio Grande do Norte (Brasil): estrutura, funcionamento e sustentabilidade*”. Tendo em vista que um dos objetivos partilhados por ambos os projetos era o de dispor um panorama geral sobre a comunidade icitofaunística que é capturada pela frota artesanal, cabe-se destacar que a espécie *L. synagris* constituiu apenas uma fração de toda a ictiofauna capturada.

De julho de 2012 a junho de 2014 foram realizadas mensalmente de 2 a 3 pescarias experimentais por intermédio das frotas artesanais que operam com redes de emalhar de fundo ao longo da costa do Rio Grande do Norte (Fig. 11A e B). Durante todo esse período, somaram-se 126 pescarias (Fig. 12). Com o intuito de tornar os dados independentes da pesca, o delineamento experimental visou preencher toda a área de estudo com distâncias regulares entre os lances de rede. Destaca-se que, apesar do uso da frota artesanal, o termo “independência da pesca” foi definido no sentido que as pescarias não foram enviesadas pelo conhecimento do pescador.

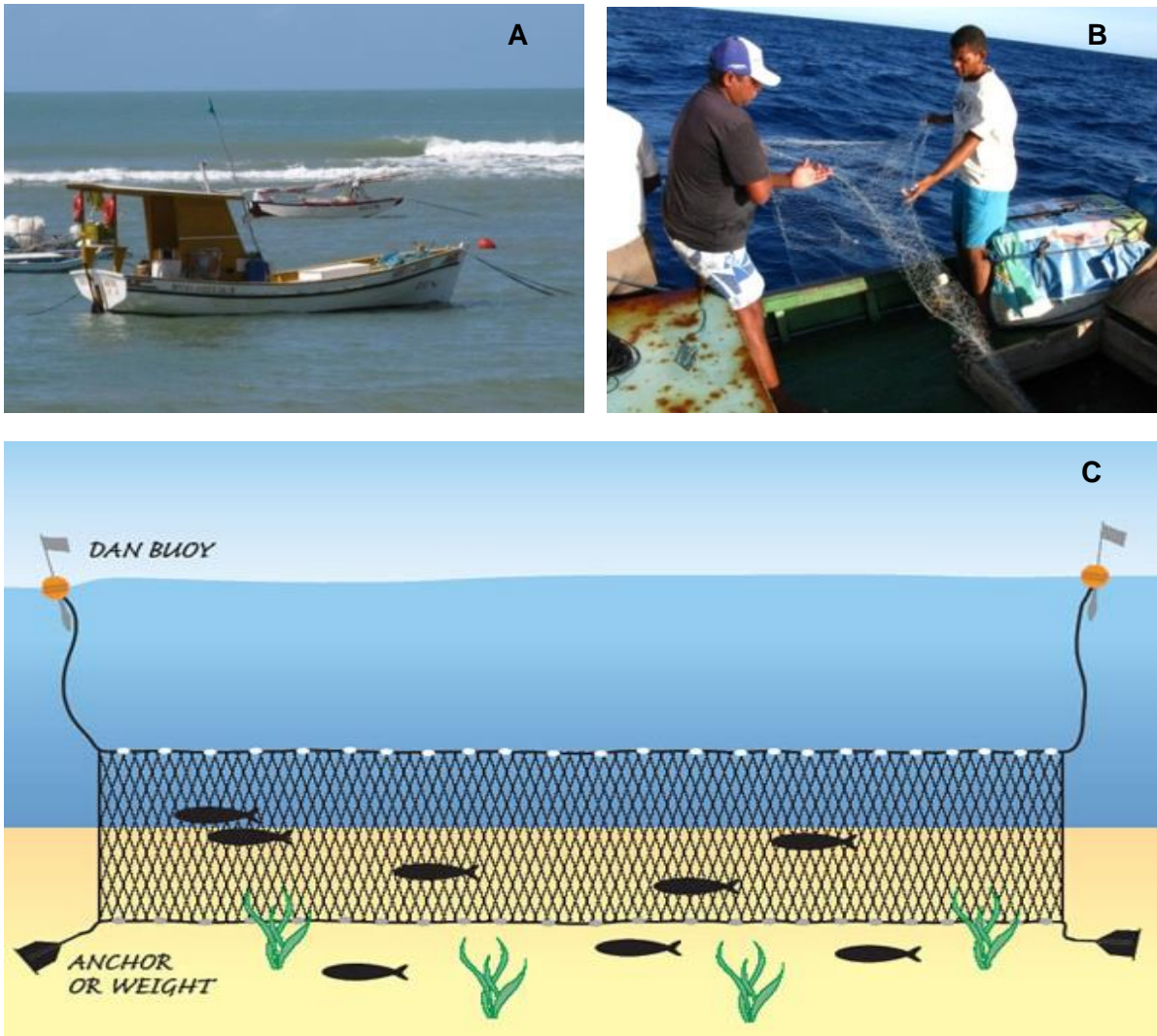


Figura 11: Exemplo de uma embarcação da frota de emalhar de fundo que opera no Rio Grande do Norte (A), lançamento de rede (B) e ilustração esquemática de uma rede de emalhar de fundo (C) (Fotos: Marcelo Nóbrega; Ilustração: Ailloud, 2010).

Além da coleta biológica, isto é, recursos capturados, também foram registradas as seguintes informações para cada estação de coleta: profundidade, temperatura, longitude, latitude, distância da costa, tipo de substrato, estação do ano, comprimento e altura da rede e tempo de imersão da rede na água

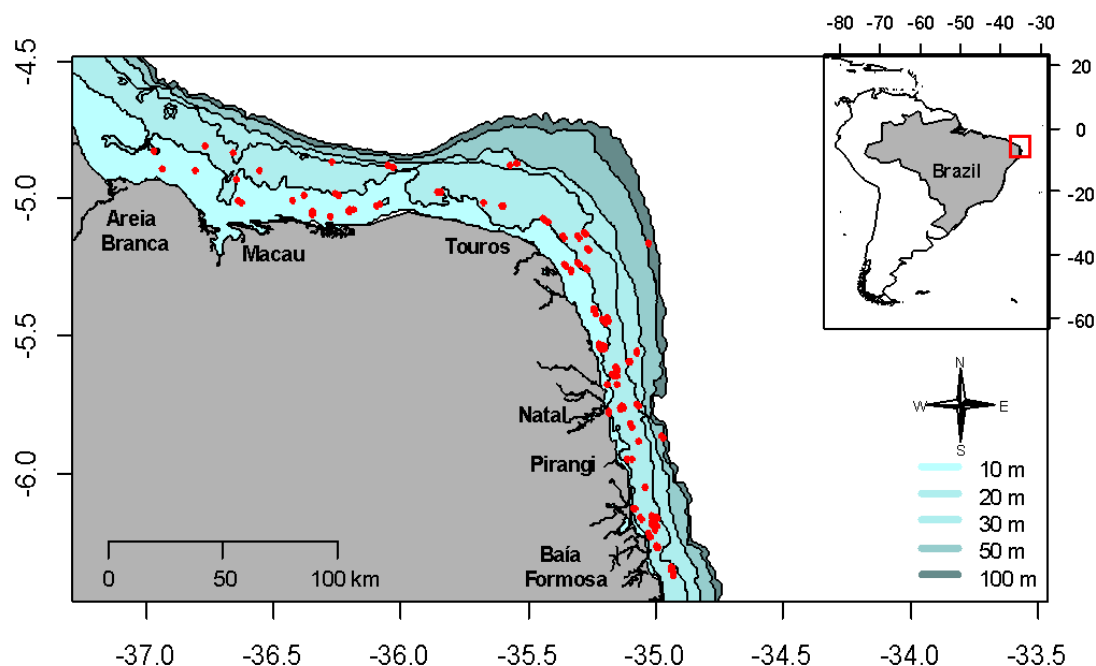


Figura 12: Área de estudo com destaque para os pontos amostrados.

As redes de emalhe de fundo empregadas pela frota artesanal norte-rio-grandense eram todas confeccionadas com nylon de monofilamento e fixadas no fundo do mar por meio de pesos de chumbo distribuídos uniformemente ao longo da rede (Fig. 11C). Os comprimentos, alturas, tamanho das malhas e tempo de submersão das redes variaram conforme a embarcação acompanhada.

Concomitantemente às pescarias, todos os indivíduos capturados foram pesados (g) com uma balança de precisão e medidos com um ictiômetro em termos de seu comprimento total (CT; cm) e zoológico (CZ; cm) (Fig. 13). Ao final de cada uma das pescarias, uma subamostra representativa dos recursos capturados foi selecionada aleatoriamente e levada para o Laboratório de Biologia Pesqueira (LABIPE) da Universidade Federal do Rio Grande do Norte (UFRN), aonde foram realizados análises biométricas.

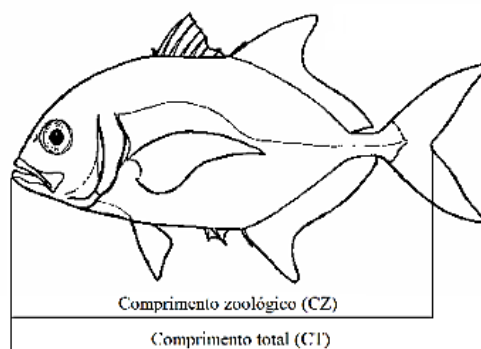


Figura 13: Ilustração das medidas morfológicas que foram tomadas para os recursos capturados (Fonte: Nóbrega, 2008).

3.1.1. Procedimentos laboratoriais

Para cada espécime constituinte das subamostras, foram determinados o sexo e os estados de maturação a fim de, posteriormente, subsidiar informações necessárias para determinação do comprimento de primeira maturação. Cada indivíduo foi dissecado ventro-longitudinalmente para a retirada das gônadas, a partir das quais foram estabelecidas a sexagem e a identificação macroscópica do estado de maturação. Esta avaliação foi realizada mediante a classificação proposta por Vazzoler (1996), a qual consiste em cinco estágios (Tabela I) fundamentados em características secundárias tais como o grau de turgidez, tamanho, presença de vasos sanguíneos e coloração das gônadas.

Tabela I: Classificação gonadal segundo Vazzoler (1996).

Estado de maturação	Classificação
A	Imaturo (ou virgem)
B	Em maturação
C	Maduro
D	Desovado
E	Repouso

3.2. Variáveis ambientais

Uma vez que estamos lidando com modelos de distribuição de espécies, um passo importante é definir possíveis preditores ambientais relacionados com a biologia da(s) espécie(s) estudada(s). Para o presente estudo foram selecionados seis potenciais preditores os quais incluíram: temperatura superficial do mar (TSM, °C), distância da costa (km), batimetria (m), rugosidade (índice), declividade do leito marinho (°), e concentração de clorofila-a (mg/m³). A distância da costa, batimetria e parâmetros derivados da própria batimetria tais como a rugosidade e a declividade, foram extraídas a partir do banco de dados marinhos fornecidos pelo MARSPEC, o qual está disponível pelo site <http://www.marspec.org>. O MARSPEC é constituído por um conjunto de imagens (rasters) climáticas e geofísicas relacionadas ao ambiente marinho, cuja resolução espacial é de 30 arco-segundos (Sbrocco & Barber, 2013). Tendo em vista que as variáveis contidas nesse banco de dados são úteis para organismos que possuem hábitos pelágicos associados a superfície bem como de águas rasas, estas foram consideradas apropriadas para o presente estudo uma vez que *L. synagris* apresenta hábitos tipicamente costeiros.

A distância da costa e batimetria são comumente utilizadas em modelos EFHs, uma vez que ambas estão relacionadas com a distribuição vertical e horizontal dos peixes marinhos (Valavanis, 2002) (Fig. 14F e D). A rugosidade e a declividade, por sua vez, frequentemente são utilizadas como indicadores da morfologia do leito marinho. Tendo em vista que dentro do banco de dados fornecidos pelo MARSPEC apenas a declividade está presente, derivou-se rugosidade do próprio *raster* de batimetria por meio das ferramentas *focal statistics* e *raster calculator* fornecidos pelo programa ArcGIS 10. Este passo tomou por base o método descrito por Janness (2004), o qual designa a posição de um determinado ponto em relação aos pontos vizinhos, e com os valores variando entre 0 e 1.

Este índice, portanto, foi utilizado para indicar a complexidade do leito marinho em que valores pequenos tendem a corresponder a substratos inconsolidados, à exemplo de areia e lama, e valores mais altos geralmente estão associados à substratos consolidados, como é o caso de pedras e cascalhos (Fig. 14C). No que diz respeito à declividade, tem-se um parâmetro que descreve a taxa de variação na elevação entre valores vizinhos, com valores variando entre 0 e 90 graus.

Analogamente à rugosidade, este preditor foi utilizado como uma aproximação do tipo de substrato, cujos menores valores correspondem a fundos planos e maiores valores a fundos rochosos (Fig. 14A).

A TSM e a clorofila-a foram selecionadas haja vista ambas estarem relacionadas à produtividade marinha. Além disso, essas duas variáveis são potenciais indicadoras de frentes térmicas e de produtividade (Valavanis *et al.*, 2008). Sabe-se que a clorofila-a, em específico, está diretamente relacionada à produtividade marinha. Embora a produtividade primária dependa de diversos outros fatores (e.x., intensidade luminosa), a clorofila-a pode ser utilizada como aproximação da produtividade primária visto que sua variabilidade pode modificar condições tróficas e, conseqüentemente, também modificar o hábitat de algumas espécies (Katara *et al.*, 2008; Pennino *et al.*, 2013b). Para o estudo tratado nesta dissertação foram utilizados *rasters* de médias mensais da clorofila-a fornecidos pela NASA, disponível em <http://www.neo.sci.gsfc.nasa.gov> (Fig. 14B). Em seguida, para fins de suavização da superfície, os valores extraídos foram interpolados por meio de um kriging ordinário via ArcGIS 10. Finalmente, a climatologia da TSM também foi proveniente do repositório do MARSPEC (Fig. 14E).

Considerando que algumas imagens derivaram de fontes diferentes, teve-se o cuidado especial de assegurar que cada um dos preditores apresentassem a mesma resolução espacial (30 arco-segundos). Deve-se mencionar, também, que para fins de estimação dos modelos, foram utilizados os dados coletados *in situ* conforme mencionado previamente na seção 3.2. No caso de variáveis que não foram registradas durante as amostragens (declividade, rugosidade e clorofila-a), extraiu-se seus valores a partir de suas respectivas imagens.

À parte dos tratamentos das imagens realizados no ArcGIS, todos os procedimentos restantes foram realizados no programa R (R Development Core Team, 2013) mediante os seguintes pacotes: *raster* (Hijmans, 2015), *rasterVis* (Lamigueiro & Hijmans, 2014), *maptools* (Bivand & Lewin-Koh, 2015), *gridExtra* (Auguie, 2012), *lattice* (Deepayan, 2008) e *fields* (Nychka *et al.*, 2015). Todos os preditores ambientais foram padronizados a fim de reduzir a multicolinearidade, bem como para permitir futuras comparações entre cada uma deles (Kinas & Andrade, 2010). A multicolinearidade foi verificada por meio do coeficiente de correlação de Pearson (ou de Spearman, quando necessário). Por fim, outras análises

exploratórias dos dados foram realizadas mediante o protocolo sugerido por Zuur *et al.* (2010), no qual se analisa basicamente a presença de outliers, a normalidade e a homocedasticidade.

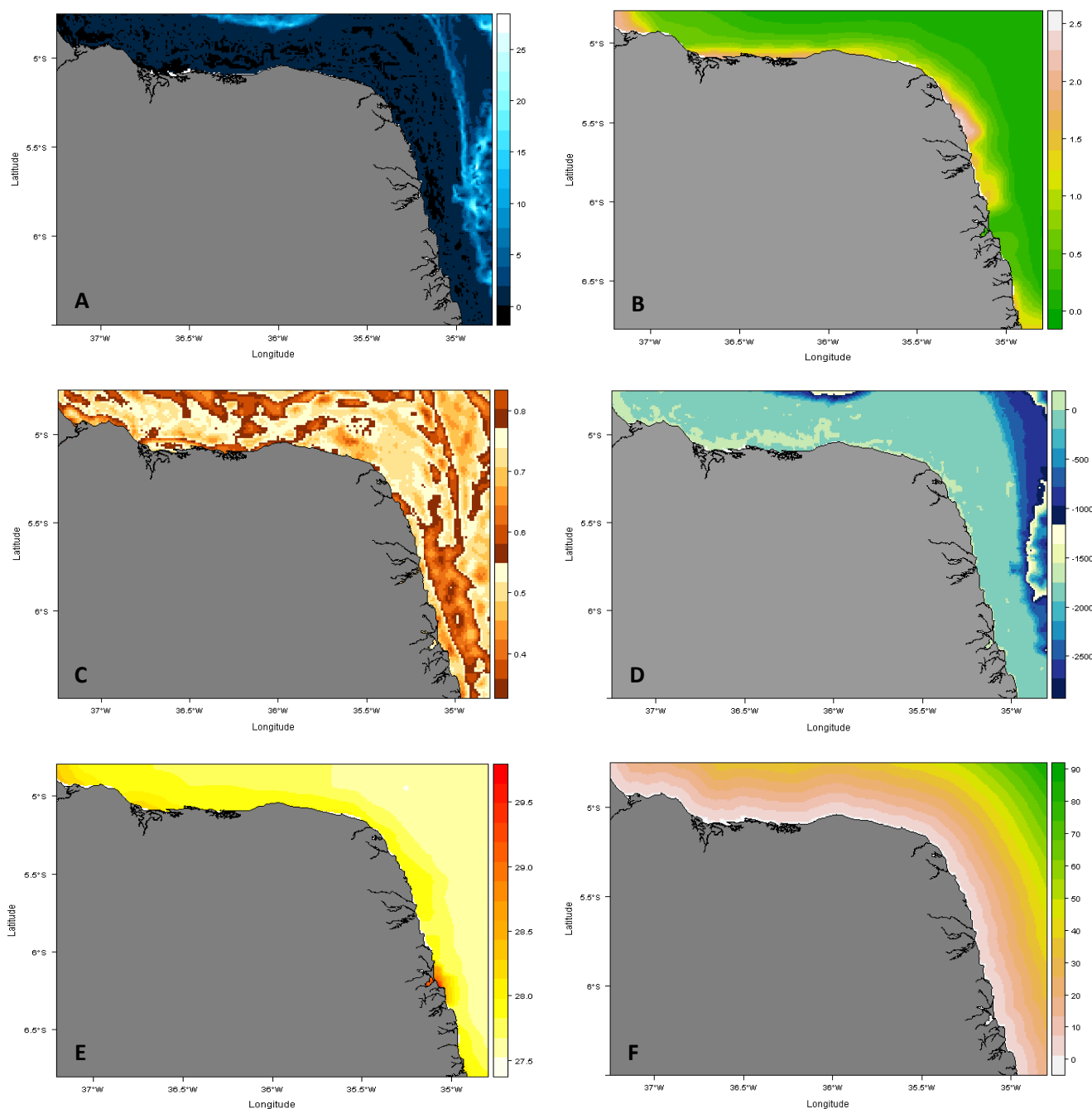


Figura 14: Mapas dos preditores lineares para a região do Rio Grande do Norte onde (A) é a declividade, (B) a concentração de clorofila-a, (C) a rugosidade, (D) a batimetria, (E) a TSM e (F) a distância da costa.

3.3. Análise dos dados

3.3.1. Determinação do tamanho médio de primeira maturação

Sabe-se que parâmetros populacionais constituem dados essenciais para controlar suas dinâmicas, haja vista eles serem indicativos do nível de exploração (King, 2007). Dentre esses parâmetros destaca-se o parâmetro de tamanho de primeira maturação gonadal (L_{50}), o qual corresponde ao comprimento médio em que 50% dos indivíduos já desovaram pelo menos uma vez (Vazzoler, 1996). De acordo com Fonteles-Filho (1989) a determinação do L_{50} possui três principais objetivos:

- 1) Delimitar o comprimento entre os estoques jovem e adulto;
- 2) Determinar o limite do estoque reprodutor quando não se tem acesso aos dados de distribuição do comprimento das fêmeas em estágio de desova;
- 3) Estabelecer um tamanho mínimo de captura para administrar racionalmente os estoques de populações comercialmente importantes;

Dentre os objetivos destacados anteriormente, será dado maior ênfase ao primeiro tópico visto que este é de maior interesse para os propósitos deste trabalho (vide os objetivos específicos no item 1.2). Conforme já abordado no item 2.2.1, o primeiro passo para a estimação do L_{50} consiste na sexagem e identificação do estado de maturação de cada indivíduo. Das subamostras trazidas em laboratório, contabilizaram-se 89 indivíduos de *L. synagris* (42 fêmeas e 47 machos) (Tabela II).

Tabela II: Número de indivíduos por estados de maturação, sexo e espécie.

Espécie Estado de maturação	<i>L. synagris</i>	
	♀	♂
A	0	03
B	16	19
C	19	21
D	07	04
E	0	0
Total	42	47

O cálculo de L_{50} comumente é realizado por meio de uma regressão logística, cujos parâmetros (β_0 e β_1) são tradicionalmente estimados via método de máxima verossimilhança (King, 2007; Karna & Panda, 2011). Contudo, quando a amostra é pequena, esse método produz estimativas enviesadas e os limites de confiança não podem ser calculados (Peduzzi *et al.*, 1996). Dado a essa falha imposta pelo método frequentista, inferências Bayesianas têm sido cada vez mais acessadas pela ciência pesqueira haja vista não ser tão limitante quanto à restrição do tamanho amostral (Doll & Lauer, 2013). Portanto, considerando que as subamostras de *L. synagris* foram relativamente pequenas foi optado pela utilização de métodos Bayesianos para delimitar o seu estoque em jovem e adulto.

Inicialmente foram estabelecidas k classes de comprimento para o conjunto de dados, em cada qual foi registrado o número total de indivíduos (n_i) e o número total de indivíduos maduros (Y_i). O estado de maturação foi modelado segundo uma regressão logística conforme proposto por Kinas & Andrade (2010) (Tab. III), cuja probabilidade de um indivíduo i estar maduro assume uma distribuição binomial,

$$Y_i \sim \text{Binomial}(n_i, p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(x_i - \bar{x})$$

$$\bar{x} = \frac{\sum x_i}{k} = \frac{(x_1 + x_2 + \dots + x_k)}{k}$$

onde p_i denota a probabilidade de que um indivíduo i -ésima classe de comprimento esteja sexualmente maduro; $g(p_i)$ representa a função de ligação *logit*; β_0 é o *logit* da probabilidade de que um indivíduo com comprimento igual à média \bar{x} se encontre sexualmente maduro; β_1 o incremento médio no *logit* de p_i para cada centímetro adicionado ao comprimento; e x_i denota o comprimento individual médio da i -ésima classe de comprimento.

Tabela III: Exemplo da estruturação dos dados para o cálculo do L_{50} via regressão logística.

Classe de tamanho	Nº total de indivíduos	Nº de indivíduos maduros	Comprimento médio
$i = 1$	n_1	Y_1	x_1
$i = 2$	n_2	Y_2	x_2
$i = 3$	n_4	Y_3	x_3
.	.	.	.
.	.	.	.
.	.	.	.
$i = k$	n_k	Y_k	x_k

As estimações dos L_{50} foram realizadas a partir da equação $[(-\beta_0/\beta_1)+\bar{x}]$. Usualmente o cálculo do L_{50} é fundamentado em indivíduos que estejam no mínimo no estado B da classificação de maturação. Entretanto, dado ao pequeno n subamostral, os L_{50} deste estudo basearam-se em indivíduos que estiveram no mínimo no estado C a fim de melhorar o ajuste da curva de Ogiva.

Considerando que os parâmetros da regressão logística foram estimados mediante inferência Bayesiana, assumiram-se distribuições Normais com prioris não informativas ($\beta_0 \sim \mathcal{N}(0, 0.01)$; $\beta_1 \sim \mathcal{N}(0, 0.01)$) para cada um deles. As distribuições de probabilidade posteriores foram simuladas por intermédio do método de Monte Carlo com Cadeias de Markov (MCMC, do inglês *Markov Chain Monte Carlo*), via função *R2jags* (Su & Yajima, 2014) pela interface do RStudio versão 0.98 (R Development Core Team, 2013).

Para cada um dos casos considerados, isto é, parâmetros derivados dos sexos separados e dos sexos agrupados, foram rodadas três cadeias com 60.000 iterações com descarte das primeiras 20.000 e retenção de 6.000 iterações com intervalo estabelecido em 20 passos. A convergência das cadeias foi avaliada segundo métodos gráficos convencionais, tais como gráfico das iterações e de autocorrelação, e por métodos matemáticos, à exemplo do valor de \hat{R} , o qual denota a variabilidade observada dentro e entre as cadeias (Maiores detalhamentos podem ser acessados no cap. 14 de Kruschke (2014)).

Por fim, um teste de hipótese Bayesiano foi utilizado para avaliar se houve diferenças significativas entre as distribuições posteriores dos L_{50} dos machos e dos

L_{50} das fêmeas, seguindo as especificações propostas por Berry (1995). Para isso foi obtida a distribuição posterior da diferença $d = L_{50\text{machos}} - L_{50\text{fêmeas}}$, a partir da qual se pôde acessar a probabilidade deste valor ser igual a zero. Para a distribuição posterior d foi estabelecido um intervalo de credibilidade de 95% ($ICr_{95\%}$). Nos casos em que este intervalo não cobriu o zero, obtemos respaldo suficiente para “rejeitar” a hipótese nula de igualdade $L_{50\text{machos}} = L_{50\text{fêmeas}}$. Portanto, baseando-se no resultado proveniente do teste de hipótese, o L_{50} estimado foi utilizado como ponto de corte para delimitar as amostras totais de *L. synagris* em número de indivíduos jovens e adultos.

3.3.2. Modelagem da abundância e estrutura de idades

Foram utilizados modelos Bayesianos hierárquicos espaciais e GMRFs para estimar e prever a abundância (caso 1), número de adultos (caso 2) e número de jovens (caso 3) de *L. synagris* em relação a diversos preditores ambientais. Como índice de abundância foi utilizada a captura-por-unidade-de-esforço (CPUE), a qual foi definida como sendo a captura total do ariocó (g) ponderada pelo esforço de pesca. O esforço, por sua vez, foi estabelecido pela multiplicação da área da rede de emalhe (m^2) e pelo seu tempo de submersão (h).

Dados pesqueiros como a CPUE frequentemente são caracterizados por uma distribuição assimétrica e com um número elevado de observações nulas. Neste sentido, ignorando os zeros, é comum recorrer à distribuições do tipo Gama ou log-Normal para modelar a CPUE uma vez que elas atendem a esses pressupostos básicos (Venables & Ripley, 2004). Ambas as distribuições costumam apresentar bom ajuste, entretanto como o melhor ajuste é particular para cada tipo de dado, foi resolvido testar diferentes modelos para ambas as distribuições. Como nenhuma dessas distribuições pode conter o valor zero, foi adicionado uma constante referente à 10% da mediana da CPUE a fim de contornar essa limitação.

No que diz respeito aos modelos de contagem para os adultos e jovens, foi esperado observar um elevado número de valores nulos haja vista eles possivelmente apresentarem uma segregação espacial. Desta forma foram testados quatro diferentes distribuições que tipicamente podem ser utilizadas nesse contexto:

poisson, poisson zero-inflacionada, binomial negativa e binomial negativa zero-inflacionada. Vale mencionar que o R-INLA promove três formas diferentes para modelar dados zero-inflacionados, os quais recebem os nomes de tipo 0, 1 e 2. Esses tipos diferem na maneira em que os zeros são definidos entre zeros estruturados (falsos verdadeiros) e zeros devido à falta de detecção fruto de um erro experimental ou erro do observador (falsos negativos).

Em sumo, o modelo tipo 0 é um modelo em dois estágios, também conhecido por *hurdle model*, o qual é truncado em zero de forma que a variável resposta só pode ser positiva. Nesse tipo de modelo se assume que o zero surgiu de um único processo binomial, cuja função de verossimilhança de Y_i é condicionada para todo $Y > 0$; dessa forma, todos os zeros são unicamente zeros estruturados. O modelo tipo 1, por sua vez, consiste apenas em um único estágio, conhecido por modelos de mistura, e modela os zeros como se fossem provenientes de dois processos distintos: o processo binomial e o processo de contagem. Dessa forma, a diferença fundamental com os *hurdle models* reside ao fato de que o processo de contagem também pode produzir valores nulos e, assim, a distribuição não é truncada em zero. Em contra partida, os modelos tipo 2 constituem uma extensão do tipo 1, o qual incorpora um outro hiperparâmetro na sua formulação que dá maior ênfase aos valores nulos. Uma vez que os dados deste estudo foram independentes da pesca, acredita-se que as amostragens sem registros do ariocó poderiam potencialmente indicar que ele não ocorre nessas áreas e, dessa maneira, os zeros seriam apenas zeros estruturados. Consequentemente, para os dados de contagens deste estudo foi utilizado apenas o modelo do tipo 0.

Como já previamente ressaltado no item 2.5.2, todos esses modelos podem ser estruturados de forma similar a um GLM. Dessa maneira, cada caso de modelos descritos anteriormente podem ser resumidos conforme:

Caso 1: $Y = \log(\text{CPUE})$

$$\begin{array}{ccc}
 Y_i \sim \text{Gamma}(\mu_i) & & Y_i \sim \log - \text{Normal}(\mu_i) \\
 g(\mu_i) = \log(\mu_i) = \beta_0 + \sum_m^M \beta_m X_m + W_i & \text{OU} & g(\mu_i) = \mu = \beta_0 + \sum_m^M \beta_m X_m + W_i
 \end{array}$$

Caso 2 e 3: $Y = \log(\text{N}^\circ \text{ de indivíduos adultos ou jovens})$

$$Y_i \sim \text{Poisson}(\mu_i) \quad \text{ou} \quad Y_i \sim \text{Binomial negativa}(\mu_i)$$
$$g(\mu_i) = \log(\mu_i) = \beta_0 + \sum_m^M \beta_m X_m + \text{offset}(\text{esforço}) + W_i \quad g(\mu_i) = \log(\mu_i) = \beta_0 + \sum_m^M \beta_m X_m + \text{offset}(\text{esforço}) + W_i$$

Aonde $g(\mu_i)$ é a função de ligação, β_0 é o intercepto, β_M é um vetor dos coeficientes de regressão que quantifica o efeito de alguns preditores X_M na variável resposta Y_i , W_i é o efeito espacial e *offset* foi especificamente utilizado nos modelos de contagens a fim de contemplar os diferentes esforços de pesca.

3.3.3. Inferência, seleção e avaliação dos modelos

Uma vez que o modelo foi definido, o próximo passo consiste na estimação de seus parâmetros. Como estamos lidando com uma abordagem Bayesiana, as estimações dos parâmetros e predição das variáveis observáveis são acessadas por meio da distribuição posterior, resultantes da aplicação do Teorema de Bayes. Sendo assim, distribuições à priori padrões foram especificadas para todos os parâmetros conforme recomendando por Held *et al.* (2010), os quais já se encontram implementados no R-INLA. Essas prioris constituem aproximações de prioris não informativas e objetivam exercer pouca influência sobre a distribuição posterior. Dessa forma os resultados ficarão a marcê dos próprios dados e, conseqüentemente, serão semelhantes às inferências por abordagem frequentista nos casos em que comparações são possíveis.

Para o componente espacial foi utilizado a metodologia SPDE conforme definida na seção anterior. Prioris Gaussianas com média zero e matriz de covariância dependente de cada hiperparâmetro foram assumidas seguindo as recomendações de Lindgren & Rue (2013) e que também já se encontram implementadas no R-INLA. A distribuição posterior marginal de todos os parâmetros foram calculadas por meio do procedimento padrão do INLA, o qual utiliza as aproximações simplificadas e integração numérica para obter distribuições posteriores marginais dos hiperparâmetros.

No que diz respeito à seleção dos modelos, todos foram realizados mediante método manual da inserção sequencial de variáveis (*forward stepwise*). O critério de informação de deviança (DIC, do inglês *Deviance Information Criterion*) foi utilizado para acessar a qualidade de ajuste dos modelos, o qual guarda semelhança com o critério de informação de Akaike (AIC, do inglês Akaike Information Criterion), porém é mais adequado quando lidamos com modelos hierárquicos Bayesianos (Spiegelhalter, 2002). Em suma, o DIC é dado pela soma da média posterior da deviança (\bar{D}) e pelo número efetivo de parâmetros (p_D), os quais quantificam o ajuste do modelo e avaliam a sua complexidade, respectivamente.

Para medir o poder preditivo dos modelos, foi utilizado o CPO (do inglês *Conditional Predictive Ordinate*) seguindo as recomendações dos trabalhos de Muñoz *et al.* (2013) e Pennino *et al.* (2014). Denotado pela relação $CPO_i = p(y_i | y_{-i})$, essa medida é definida por uma validação cruzada da densidade preditiva para uma dada observação y_i , onde y_{-i} denota os dados sem a i 'ésima observação. Ross & Held (2011), entretanto, indicaram o uso da versão logaritmizada para propósitos preditivos, sendo definido por: $LCPO = -(n^{-1}) \sum_i \log(CPO_i)$. Como o R-INLA não oferece o LCPO em sua saída padrão, esta medida deve ser computada manualmente de acordo com o seguinte código: `[-mean(log(modelcpocpo))]`. Valores mais baixos tanto para DIC quanto LCPO são indicadores de melhor ajuste e qualidade preditiva, respectivamente.

Como as vezes o LCPO e o DIC podem não concordar conjuntamente no mesmo modelo, o melhor modelo foi selecionado seguindo três passos consecutivos. Neste sentido, considerando que diversos modelos foram testados de acordo com diferentes distribuições de probabilidade, o primeiro passo da seleção dos modelos consistiu na avaliação do LCPO entre os modelos com diferentes distribuições de probabilidade. Em seguida, nos casos em que os valores de LCPO foram similares entre dois ou mais modelos com diferentes distribuições, usou-se o DIC como segundo critério de seleção com o intuito de definir a distribuição de probabilidade que melhor se adequou aos dados. Por fim, após a seleção de diversos modelos, foi avaliado modelos que contiveram apenas preditores relevantes, isto é, aqueles cujo intervalo de credibilidade a 95 % ($CI_{95\%}$) não cobriu o zero. Como o principal objetivo consistiu na qualidade preditiva, não faria sentido selecionar modelos contendo alguns preditores irrelevantes em sua estrutura, mesmo se o LCPO e DIC fossem os

menores. Dessa maneira, o melhor modelo foi escolhido como aquele contendo apenas preditores relevantes.

Uma vez selecionado, os modelos foram então avaliados quanto às suas qualidades de ajustes, tais como como gráficos de dispersão entre os valores preditos e valores observados e gráficos dos quantis-quantis residuais. Como se espera observar uma relação linear entre os valores observados e preditos, utilizou-se o coeficiente de correlação de Pearson (ρ) para avaliar esta relação, bem como testado ao nível de 5% de significância. Modelos com $\rho \geq 0.7$ e p-valor ≤ 0.05 foram considerados aceitáveis e consequentemente utilizados para a predição final.

Capítulo 4

ARTIGO: *Bayesian spatial predictive models for data-poor fisheries*

Submetido em: *Ecological Modelling*

RUFENER, M. C.^{1,2}; KINAS, P. G.¹; NÓBREGA, M. F.² & LINS, J. O.²

¹ Laboratório de Estatística Ambiental, Instituto de Matemática, Estatística e Física, Universidade Federal do Rio Grande, Rio Grande, Brasil.

² Laboratório de Biologia Pesqueira, Departamento de Oceanografia e Limnologia, Universidade Federal do Rio Grande do Norte, Natal, Brasil.

4.1. Abstract

Understanding the spatial distribution and identifying environmental variables that drive endangered fish species abundance are key factors to implement sustainable fishery management strategies. In the present study we proposed Bayesian hierarchical spatial models to quantify and map sensitive habitats for juveniles, adults and overall abundance of the vulnerable lane snapper (*Lutjanus synagris*) present in the northeastern Brazil. Data were collected by fishery-independent gillnet surveys, and fitted through the Integrated Nested Laplace Approximations (INLA) and the Stochastic Partial Differential Equations (SPDE) tools, both implemented in the R environment by the R-INLA library (<http://www.r-inla.org>). Our results confirmed that the abundance, juveniles and adults of *L. synagris* are spatially correlated, have patchy distributions along the Rio Grande do Norte coast and are mainly affected by environmental predictors such as distance to coast, chlorophyll-a concentrations and sea floor rugosity. By means of our results we intended to consolidate a recently introduced Bayesian Geostatistical model into fisheries science, highlighting its potential for establishing more reliable measures for the conservation and management of vulnerable fish species even when data are sparse.

Key words: Bayesian Geostatistical models; Integrated Nested Laplace Approximations; Stochastic Partial Differential Equations; Essential fish habitats; Fisheries ecology

4.2. Introduction

Fisheries action may induce many changes in fish stocks which are commonly related to abundance, size and spatial distribution of these resources (Haddon, 2001). Additionally, fish stocks are also affected by environmental fluctuations (King, 2007). Therefore, detecting the main environmental factors which drive natural abundance fluctuation and understanding how these associations vary over space and time are key concepts not only in ecology but also in fisheries sciences (Ross et al., 2012).

Knowing the links between the fish stocks dynamics and its surrounding environment is therefore of fundamental importance in order to improve sustainable fisheries management and conservation strategies (Babcock et al., 2005; Valavanis et al., 2008). The way in which we may access the species-environment relationships are commonly treated in the specialized literature as Species Distribution Models (SDM), or more specifically, as Essential Fish Habitat (EFH) models which concerns a subset of the former when dealing with marine species. According to Valavanis et al. (2008), EFH's constitute areas which promote the fishes more favorable habitats for spawning, feeding, or growth to maturity, and are frequently analyzed in context of several statistical tools to evaluate the species distribution in respect to some environmental variables.

Over the past three decades a huge effort has been spent in the development of powerful statistical models to include more realistic scenarios in regards to the species-environmental relationships. Artificial Neural Networks (ANN, e.g., SPECIES), Maximum Entropy (ME, e.g., MAXENT), Climatic Envelops (CE, e.g., BIOCLIM), Classification and Regression Trees (CART, e.g., BIOMOD) and regression models such as Generalized Linear and Additive (Mixed) Models (GLM/GLMM/GAM/GAMM) are some of the many methodological examples that have been proposed for modelling the species distribution (for instance, see Franklin, 2010 and Guisan and Thuiller, 2005 for a review and detailed version of SDM themes).

Although these applications are considered robust, they usually address only explanatory models that simply aim to verify the relationship between the response variable (e.g. presence or abundance of the species) and some environmental

predictors (e.g. bathymetry and chlorophyll-a concentration in marine context), however, without considering explicitly the spatial component (Ciannelli et al., 2008; Elith and Leathwick, 2009; Latimer et al., 2006). In light of this, such models may result in a poor characterization of the species response to environmental factors by underestimating the degree of uncertainty in its predictors (Latimer et al., 2006).

Considering that the marine environment is an extremely heterogeneous domain, and that a given marine species has biological and ecological constraints, it is commonly noted that biological resources like most fish species present a gregarious distribution. Therefore, fishery data usually tend to be spatially autocorrelated and may be described by the first law of geography formulated by Tobler (1970): “Everything is related to everything else, but near things are more related than distant things” (Kneib et al., 2008; Nishida and Chen, 2004; Plant, 2012).

Hence, ignoring spatial autocorrelation violates the main assumption of classical inference, which assumes that the data are independent and identically distributed, and thus can lead to biased estimates (Plant, 2012). Spatial correlation should therefore be considered since species are generally subject to similar environmental factors (Muñoz et al., 2013). Such spatial models allow to construct powerful predictive models that not only provide the estimation of the processes that influenced species distribution, but also promote the possibility of predicting their occurrence in unsampled areas (Chakraborty et al., 2010).

Additionally, it is advantageous to incorporate Bayesian inference in predictive models, given that it is possible can integrate all types of uncertainties using exclusively the probability as its metric. Combining the uncertainty in the data (expressed by the likelihood) with extra-data information (expressed by prior distributions) using Bayes' theorem, posterior probability distributions are built for all model parameters, which in turn represent the most complete form to express the up-to-date knowledge on an investigated phenomenon (Kinas and Andrade, 2010).

Hierarchical Bayesian Models (HBMs) are very suitable for such situations, as they allow to introduce sequentially the uncertainties associated to the entire fishery phenomenon, as well as a spatial random effect, also known as Gaussian random field (GRF) (Cosandey-Godin et al., 2015). Traditionally HBMs relied on simulation techniques such as Markov Chain Monte Carlo (MCMC). However, as model complexity increases, it reveals that the computational time required to approximate

the posterior distributions becomes almost unfeasible. Recently a new Bayesian statistical methodology to obtain posterior distributions in HBMs has been proposed by Rue et al. (2009), called Integrated Nested Laplace Approximations (INLA), and which is currently implemented in the R environment by the R-INLA package (<http://www.r-inla.org>).

Rather than using stochastic simulation techniques, INLA uses numerical approximations by means of the Laplace operator, which revealed to be much faster, flexible and accurate than MCMC whenever applicable (Rue et al., 2009). Thanks to Illian et al. (2013) and Muñoz et al. (2013), spatio-temporal models were introduced to the ecological community in order to fit point process and point-referenced data through the INLA approach. In regards to marine ecology and fisheries research, more studies have slowly emerged since then, using exclusively INLA for spatial and temporal purposes (Cosandey-Godin et al., 2015; Muñoz et al., 2013; Paradinas et al., 2015; Pennino et al., 2013; Pennino et al., 2014; Quiroz et al., 2015; Roos et al., 2015).

However, most of these studies relied on “data-rich” fisheries, where certainly any quantitative method would have had good performance. On the other hand, in developing countries such as Brazil, fisheries tend to be poorly documented and inadequately managed due to lack of research fundings for monitoring and analysis (Honey et al., 2010). Thereby, conventional fisheries stock assessment tools demand big data sets and may be not applicable in “data-poor” situations. In this sense, flexible statistical tools are highly recommended as they may also perform effectively, despite the uncertainty and limited information (Bentley, 2014; Pilling et al., 2008).

In order to expand the use of such tools in data limited fisheries, we will demonstrate the flexible and usefulness of Bayesian models in some important fisheries issues, such as delimiting fish stocks into age groupings and evaluating the spatial distribution of these age groupings. Specifically, our main objective is: (i) predict abundance and age groupings of the target lane snapper (*Lutjanus synagris*) along a fraction of the northeastern coast of Brazil; and (ii) identify environmental drivers that affect lane snapper’s abundance and so, provide important insights of its spatial distribution.

This paper is organized as follows: firstly we describe the importance of our study subject, how the main dataset was obtained and how we achieve age groupings

using Bayesian logistic regression. Thereafter we describe briefly the theory behind Bayesian hierarchical spatial models and its application to EFH's, and discuss the entire modeling procedures used in this study. Finally we expose and discuss our results, outlining opportunities for future spatial fisheries management.

4.3. Material & Methods

4.3.1. Case study

As a case study, we modeled the spatial occurrence of *Lutjanus synagris* (Linnaeus, 1758), popularly known as lane snapper and which is considered one of the most important fishing resource caught within the Lutjanidae family (Luckhurst et al., 2000). Lane snappers inhabit a variety of habitats from coastal waters to depths up to 400 m, often occurring in coral reefs and vegetation on sandy bottoms (Carpenter, 2002), and are widely distributed throughout the tropical region of the western Atlantic, from North Carolina to southeastern Brazil, including the Gulf of Mexico and the Caribbean Sea (McEachran and Fechhelm, 2005). According to Duarte and Garcia (1999) they are classified as a generalist species as well as an opportunistic carnivore.

Given its high commercial and recreational value, this species is one of the mainstays of artisanal fisheries not only in Venezuela and the Caribbean Sea, but also in northeastern Brazil (Gómez et al., 2001; Luckhurst et al., 2000; Rezende et al., 2003). According to Lessa et al. (2009), catches of lane snapper in northeastern Brazil are recorded since the late 70's. Moreover, among the northeastern states of Brazil, Rio Grande do Norte contributed with approximately 15% of lane snappers landings between 2005 and 2007 (IBAMA, 2007a, 2007b and 2008). Despite their high abundance, this species is suffering strong fishing pressure in northeastern Brazil which, in turn, is leading to its decline since the past few decades (Lessa et al., 2009).

In general, most information about *L. synagris* relies on its biology, whereas its population dynamics and spatial distribution remains poorly understood (Cavalcante et al., 2012; Freitas et al., 2011 and 2014; Lessa et al., 2009; Oliveira et al., 2012). In this sense, understanding their spatial distribution and identifying environmental

variables that drive their abundance are key factors to implement sustainable fishery management strategies.

4.3.2. Study area and data survey

Situated in the northeast of Brazil, the Rio Grande do Norte state is located in an important coastline transition zone which abruptly changes its direction from south-north to east-west. Between July 2012 and June 2014, about three experimental fisheries were monthly conducted by fishing vessels of the artisanal fleets which operate with bottom gillnets along the Rio Grande do Norte coast. Throughout this period, 126 fishing events were reported whose depths ranged from 5 to 50 meters (Fig. 1). Besides biological sampling, extra information was also recorded for each fishery including geolocation (latitude and longitude), bathymetry (m), sea surface temperature (°C), distance to coast (km), season of the year (spring, summer, autumn and winter), month (from January to December), gill net length (m) and height (m), as well as its soaking time (h). In order to have fishery-independent data, the experimental design aimed to cover the entire study area with regular distances between the fishing bids. Note that despite the use of artisanal fleets to collect data, fishery independence makes reference to the spatial distribution of fishing effort that was defined by the researchers rather than left to the discretion of fishers preferences.

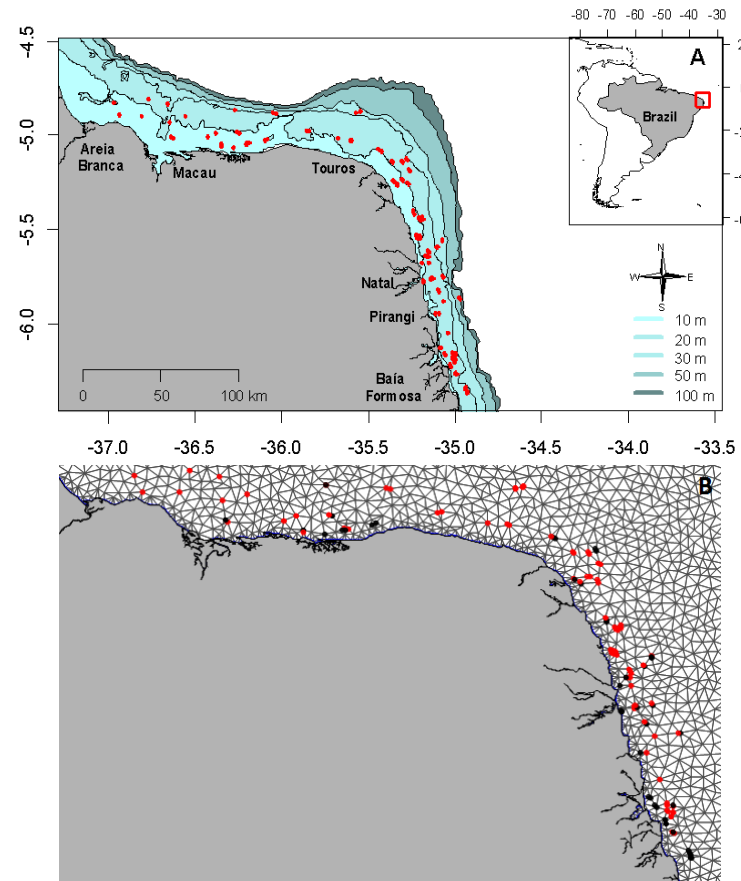


Figure 15: (a) Study area highlighting the sample stations (red dots) ; (b) Triangulation used to calculate the GMRF for the SPDE approach. The dots are referring if the species was present (red) or absent (black) during the sample stations.

All captured individuals were weighted (g) and measured for its total length (TL, cm). At the end of each fishing event, a representative subsample was randomly selected and taken to the laboratory. Sex and macroscopic maturation states were determined for each specimen according to Vazzoler (1996).

4.3.3. Environmental predictors

Besides month and season of the year, to evaluate lane snappers habitat suitability with respect its abundance and age groupings, we used six potentially predictors thought to be related to its biology, including sea surface temperature (SST, °C), distance to coast (km), bathymetry (m), rugosity (index), slope of the seabed (°) and chlorophyll-a concentration (mg/m³). Distance to coast, bathymetry, and related derived parameters such as rugosity and seabed slope, were extracted from the ocean climate layers for marine spatial ecology (MARSPEC) database, which is available at <http://www.marspec.org>. MARSPEC constitutes a high resolution dataset for marine climatic and geophysical GIS layers where the spatial resolution is 30 arc-second (Sbrocco and Barber, 2013). Since the variables contained in this database are useful for marine shallow-water and surface-associated pelagic organisms, we consider them applicable to our study given that the focal species is typically coastal.

Distance to coast and bathymetry are commonly used in fish species distribution models since they are related to its horizontal and vertical range, respectively (Valavanis, 2002). Moreover, rugosity and slope are generally used as indicative of seabed morphology. Rugosity was derived from bathymetry layer using the *focal statistics* and *raster calculator* tools from ArcGIS 10. We used the method described by Jenness (2004), which designates a position of a specific point relative to its local neighborhood and whose values ranged from 0 to 1. Thus, this index was used to indicate seafloor complexity, where small values of rugosity correspond to soft substrata such as sand and mud, while higher values indicate hard substrata like rocks and gravels. The slope described the rate of change in elevation between neighboring values and ranged from 0 to 90 degrees. This parameter was also used as a proxy of substrate type as lower values correspond to flat bottoms and higher values relate to rocky ledges.

SST and chlorophyll-a concentration are commonly related to marine productivity and both tend to decrease in areas further offshore. Besides this, both are potential indicators of thermal and productivity-enhancing fronts (Valavanis et al., 2008). In particular, chlorophyll-a concentrations are directly related to primary production. As stated by Pennino et al. (2013), it is clear that primary productivity also depends on several other factors, such as light and light penetration. Nevertheless, chlorophyll-a

concentrations can be used as a proxy of primary productivity since its variability may modify trophic conditions and thus shift the species habitat (Katara et al., 2008). Chlorophyll-a concentration was extracted from the AQUA/MODIS satellite monthly mean data provided by the Nasa Earth Observations (NEO, at <http://www.neo.sci.gsfc.nasa.gov>), and was interpolated for smoothness purposes by means of ordinary kriging via ArcGIS 10. SST was extracted from the MARSPEC climatology.

Care was taken to ensure that the spatial resolution (30 arc-second) was the same for each raster layer as they were extracted from different sources. It should be noted that for estimation purposes, we used the *in situ* collected data set to the estimation of the relationship between the response variable and predictors, in order to be more accurate. Predictors that were not registered during the fisheries (slope, rugosity and chlorophyll-a concentration) were extracted from their respective raster layers. Apart from the ArcGIS raster layer treatment, the remaining procedures were conducted using exclusively the R statistical software (R Development Core Team, 2015), by means of the following packages: *raster* (Hijmans, 2015), *rasterVis* (Lamigueiro and Hijmans, 2014), *maptools* (Bivand and Lewin-Koh, 2015), *gridExtra* (Auguie, 2012), *lattice* (Deepayan, 2008) and *fields* (Nychka et al., 2015).

All environmental predictors were standardized in order to reduce correlation among model coefficients, and to enable comparison of relative weights between variables (Kinas and Andrade, 2010). Multicollinearity was checked using Pearson's (or Spearman's, when necessary) correlation index. Other typical data exploration procedures were routinely done according to Zuur et al. (2010).

4.3.4. Bayesian estimation of mean length at first maturity

Population parameters constitute essential data for monitoring fisheries dynamics since they are potential exploitation indicators (King, 2007). Among these parameters, the size at first maturation (L_{50}) stands out, which corresponds to the average length at which 50% of the individuals reach sexual maturity (Vazzoler, 1996). According to Fonteles-Filho (1989), one of the main objectives of determining L_{50} is to delimit the young (mean length $< L_{50}$) and adult stocks (mean length $\geq L_{50}$), which constitutes our first step for species-environmental modeling.

Of the subsamples taken to the laboratory, 89 individuals of lane snapper were analyzed (Tab. I). The L_{50} is commonly calculated using a logistic regression, whose parameters (β_0 and β_1) are traditionally estimated by Maximum Likelihood method (Karna and Panda, 2011). However, when the sample is small, this method gives biased estimates and confidence limits can not be adequately calculated (Peduzzi et al., 1996). Since our biological sample was small, we decided instead to estimate our logistic parameters through Bayesian inference which is not as restrictive as the frequentist framework when concerning sample size, and provide more realistic estimations when dealing with environmental sources like ours (Doll and Lauer, 2013).

Table I: Number of individuals by sex and maturation stage.

Maturation stage	Female	Male
A	0	3
B	16	19
C	19	21
D	7	4
E	0	0
Total	42	47

Initially, k length classes were established for the set of data, where the total number of individuals in size class i (n_i) and the number of mature individuals (Y_i) were registered for each class. The maturity stage was modeled according to a logistic regression as described by Kinas and Andrade (2010), whose probability of an individual of length-class i (for $i = 1, 2, \dots, k$) being mature assumes a binomial distribution, as:

$$Y_i \sim \text{Binomial}(n_i, p_i)$$

$$\log(p_i/1 - p_i) = \beta_0 + \beta_1(x_i - \bar{x})$$

$$\bar{x} = \sum x_i/k = (x_1 + x_2 + \dots + x_k)/k$$

where p_i denotes the probability of an individual of the i -th length-class be sexually mature; $g(p_i)$ represents the logit link function; β_0 is the logit probability that an individual with the mean length \bar{x} is sexually mature; β_1 is the average increase in the logit of p_i for each centimeter added to length and x_i denotes the average length of size class i .

Estimates of L_{50} were provided from the equation $[(-\beta_0/\beta_1)+\bar{x}]$. Usually the estimation of L_{50} is grounded in individuals who are at least in the maturing stage (stage B – Tab. I). However, given the small subsample size for this species, the L_{50} of the present study was based primarily on individuals who were at least mature (stage C – Tab. I) in order to improve a better fit of the logistic curve.

Whereas the parameters of the logistic regression were estimated by a Bayesian framework, Normal vague prior distributions have been assumed for the mean and precision parameters ($\beta_0 \sim \mathcal{N}(0, 0.01)$; $\beta_1 \sim \mathcal{N}(0, 0.01)$). The posterior probability distributions were simulated by means of Markov Chain Monte Carlo (MCMC) methods using the *R2jags* package (Su and Yajima, 2014). We simulated three different situations, where the parameters were derived from grouped sexes (females + males - case 1) or separated sexes (males only - case 2; females only – case 3). For each of these cases we ran three MCMC chains simultaneously for a total of 60 000 iterations, where the first 20 000 iterations were discarded as a burn-in period and each 20th step was stored (*thinning*) in order to reduce autocorrelation.

The chains convergence was assessed using conventional graphical methods such as iteration and autocorrelation graphs, and by mathematical methods as the \hat{R} value which denotes the variability within and between chains (Gelman, 1996). Finally, a Bayesian hypothesis test was performed to assess whether there were significant differences between the posterior distributions of female L_{50} and male L_{50} . For this purpose, the posterior distribution of the difference $d = L_{50\text{male}} - L_{50\text{female}}$ was obtained and evaluated with respect to the convergence of $d = 0$. The posterior mean was taken as estimate of L_{50} , and used as a cutoff point to delimit the lane snapper stocks into young and adult groupings.

4.3.5. Essential fish habitat modeling through Bayesian Geostatistics

This section describe an emerging spatial modeling framework that can be applied for fisheries data like catch-per-unit-effort (CPUE) and count events such as age groupings expressed by number of juveniles/adults, mainly due by the SPDE approach implemented in the R-INLA package.

The present data were spatially structured, i.e., samples $Y(s_i)$ ($i = 1, \dots, n$) taken at continuous geolocations $Y(s)$ within a defined spatial domain D , where $\{Y(s): s \in D\}$ in which D consists as a subset of a d -dimensional fixed space and $Y(s)$ as a multivariate Gaussian distribution representing a partial realization of the stochastic process (Banerjee et al., 2004). Since we intended to recover areas at unsampled locations, it seemed very suitable to model our data through point-referenced spatial models, known also as Geostatistical models (Banerjee et al., 2004; Diggle and Ribeiro, 2007).

When dealing with both the randomness in observed data and the uncertainties associated with model parameters, it is straightforward to construct the models in a Bayesian hierarchical framework as they allow for a natural way to incorporate such structure and, furthermore, results in more robust statistical inference (Cressie et al., 2009). Moreover, according to Rue et al. (2009) Bayesian hierarchical geostatistical models can be treated as a particular case of Structured Additive Regression models (STAR), namely Gaussian latent models which assign Gaussian priors to all components of the additive predictor. In such models, if we assumes that $Y(s)$ is continuous along the spatial domain D then we have a continuously indexed Gaussian random field (GRF) (Lindgren et al., 2011). In other words, this field represents all spatially explicit processes that may have an effect on the investigated phenomenon.

The GRFs are mostly specified by two functions: one that deals with the mean value as denoted by: $\mu = (\mu(s_i))$; and another which incorporates the covariance structure between sample stations: $\Sigma = (C(s_i, s_j))$. According to Diggle and Ribeiro (2007), the most practical form to evaluate the empirical behaviour of the covariance function consists in treating the correlation between two points $\rho(s_i, s_j)$ as an inverse function of distance $h = ||s_i - s_j||$, i.e., the correlation between sample stations increases with decreasing distances and vice-versa. Despite a wide range of covariance functions (see Banerjee et al., 2004 and Cressie, 1993 for some

examples), the Matérn covariance function is the most popular for GRF modeling as it offers a major smoothness flexibility which is critical during the interpolation process.

Although GRFs are convenient from an analytical point of view to perform this kind of inference, the computational power has been a limiting factor with respect to factorization of the dense covariance matrices which are cubic in their dimension $\mathcal{O}(n^3)$ (Lindgren et al., 2011). Therefore, this problem is often avoided by partitioning the study area into small grids where the results provided from the spatial inference will be projected. When modeling a GRF with a Matérn covariance function, it is common and computationally efficient, to approach it by a Gaussian Markov Random Field (GMRF). Thus, while the GRF is continuously indexed, a GMRF is nothing more than a GRF indexed in a discretized way (Lindgren et al., 2011). Hence, it is assumed that a GMRF is represented by a finite-dimensional random vector following a multivariate Gaussian distribution (Rue and Held, 2005).

The Markov term arises in this context when the discretized GRFs are restricted to meet one of Markov's intrinsic properties, which is conditional independence. That is, if we assume a random vector denoted by $z = (z_1, z_2, z_3)^T$, then z_1 and z_2 will be considered conditionally independent given z_3 if, for a known value of z_3 , no additional information will be provided regarding the distribution of z_1 by considering z_2 (Rue and Held, 2005). In spatial context, it is assumed that for a GMRF the total conditional distribution $\pi(z_i | z_{-i})$, $i = 1, \dots, n$, depends only on a finite set of neighbors ∂_i for each i .

Moreover, according to Rue et al. (2009), the Markov property is conveniently encoded in the precision matrix (inverse of the covariance matrix) Q : $Q_{ij} = 0$ if z_i and z_j are independent and conditioned to z_{-ij} . Therefore, the GRFs matrix which was extremely dense, becomes sparse due to the large number of zeros that comprises the Markov property and, consequently, the matrix Q factorization is held at a cost of $\mathcal{O}(n^{3/2})$ for a bidimensional GMRF (Lindgren et al., 2011; Rue et al., 2009). Thus, this is the key point leading to an adoption of a GMRF since it saves greatly the computational time to calculate such matrices as well as presenting advantage when we are structuring a hierarchical model.

It is also at this stage that the Stochastic Partial Differential Equations (SPDE) arise. Specifically, Lindgren et al. (2011) proposed an alternative approach to provide a GMRF approximation to continuous GRF with Matérn covariance function, which is mainly due by using an approximate stochastic weak solution to the SPDE. The approximation by means of SPDE divides the latent field into triangular meshes and, thus, creates an index mesh where the prediction of the spatial phenomenon will be projected. According to Muñoz et al. (2013), the triangulation has two main advantages over a regular grid: (1) it saves greatly computational time since the prediction sites usually are lesser than those from regular grids, and (2) it uses to be denser in sampled sites providing more accurate information.

The mesh construction step is crucial during the entire modeling procedure, and extra attention should be paid when choosing the right triangulation given that it would be used for final prediction. Lindgren et al. (2011) point out that it is important that the triangles should have regular sizes and shapes. Furthermore, triangulation need to be spatially extended in order to avoid boundary effect as it usually causes variance inflation at the edges. The function *inla.mesh.2d* from R-INLA provides several arguments that can specify the triangulation. The use of the *loc* (or *loc.domain* in case for a domain extent) and *max.edge* arguments are mandatory during the mesh construction, as they define the geolocation matrix and the maximum allowed triangle edge lengths, respectively. In conjunction with the *max.edge* argument, *min.angle* and *cutoff* are commonly used which, together, allow to control both size and shape of the triangles. Moreover, if the survey area has some particular domain, such as islands or coastlines that delimit the coastal area, we may also include the *boundary* argument. For our mesh construction we used all former arguments, except the *min.angle* argument (Fig. 14B). More examples and detailed information about this function can be found in the tutorial by Krainski (2015) tutorial, which is routinely updated.

Once the predictions have been done at all mesh vertices, results are linearly interpolated to the entire domain, which is also known as Bayesian kriging (Muñoz et al., 2013). The great bonus here is that at the end of this process, for each vertex we can obtain, following a Bayesian reasoning, the predictive posterior distribution of the response. This means that unknown parameter values or predictions bounded

between the 0.025 and 0.975 quantiles of the marginal posterior distribution define the 95% probability interval for the true values.

4.3.6. Modeling abundance and age groupings

We used Bayesian hierarchical spatial models and GMRFs in order to estimate and predict abundance (case 1), number of adults (case 2) and number of juveniles (case 3) of *L. synagris* with respect to several environmental predictors. As abundance index we used catch-per-unit-effort (CPUE), which was defined as lane snapper's total catch (g) weighted by the fishing effort. Fishing effort was defined as gillnet area (m²) multiplied by soaking time (h).

Similar to a GLM approach, the response variable Y_i possesses a probability distribution that belongs to the exponential family and which is linked to a structured additive predictor η_i through a link function $g(\cdot)$ such that $g(\mu_i) = \eta_i$, and where $\mu_i = E(Y_i)$ is the expected value of the response variable:

$$\eta_i = \beta_0 + \sum_{m=1}^M \beta_m X_{m_i} + \sum_{l=1}^L f_l(v_{li}) + \text{offset}(i) \quad ; \quad i = 1, \dots, n$$

Where η_i is the linear predictor either for CPUE or count data (n° of adults/juveniles), β_0 is a scalar representing the intercept, β_M is a vector of the regressors coefficients which quantify the effect of some predictors X_m on the response, and $f_l(\cdot)$ are unknown functions defined for a set of covariates v_l . The $f(\cdot)$ terms can be used either to relax the linearity of the covariates (e.g. smooth effects), or to introduce random effects (e.g. spatial and/or temporal effect) (Rue et al., 2009). In the present study we used the $f(\cdot)$ terms particularly to introduce the spatially structured random and to test smooth effects for some predictors. Also, an offset was used specifically for the count data models in order to account for the different fishing efforts.

Fisheries data like CPUE are often characterized by a right-skewed distribution with non-negative numbers that contains a large amount of observed zeros. In this case, when ignoring zeros, lognormal and gamma distributions are commonly used to model CPUE as they meet these basic requirements (Venables and Dichmont,

2004). Usually both distributions have good fit, but since the better fit is particular to each data set, we decided to test several models regarding both distributions. Because neither of these two distributions can contain zeros, we added a constant relative to 10% of the CPUEs median in order to overcome this limitation.

With respect to the count models for adults and juveniles, we expected to observe a high amount of zeros as there could be spatial segregation. Thus, four different distributions were tested: Poisson (P), zero-inflated Poisson (ZIP), Negative Binomial (NB) and zero-inflated negative-binomial (ZINB). It is noteworthy that R-INLA provides three different forms of zero-inflated models, namely types 0, 1 and 2. These types differ among themselves in the way zeros are defined between structured zeros (true negative) and zeros due to lack of detection as a result of the from experimental design or observer error (false negative).

In a nutshell, type 0 is a two-stage model, also called *hurdle model*, which is truncated at zero such that the response can only be positive. In such model it is assumed that the zero came only from a single binomial process, and where the likelihood of Y_i is conditioned for all $Y > 0$; hence, all zeros are structured zeroes only. Type 1, in turn, are a one-stage model, also known by *mixture model*, which models the zeros as coming from two different processes: the binomial and the count process. Therefore, the key difference with a hurdle model is that the count process from type 1 can also produce zeros and thus the distribution is not zero truncated. On the other hand, type 2 is an extension of 1 which introduces another hyperparameter in its formulation that gives greater emphasis to non-zero values. Since our data were fishery-independent, we believe that sample stations without any records of the present species could potentially indicate that it does not occur there, and thus our zeros are structured zeros. Consequently, for our examples we only used the type 0 models.

4.3.7. Model fitting

Once a model is stated, its parameters have to be estimated. As we are dealing with a Bayesian framework, parameters estimates and predictions were achieved in the form of posterior distributions throughout the INLA approach. The main objective

of INLA is to obtain the marginal posterior distributions of the latent field as well as the hyperparameters of the hierarchical model. If we summarize all parameters of the latent field by the vector $\theta = (\beta_0, \beta_M, \xi)$, where ξ is the GMRF, and all hyperparameters by the vector $\psi = (\Phi, \tau, \kappa, \sigma_\varepsilon^2)$, where Φ is the parameter of the response distribution, κ and τ are the Matérn covariance function parameters and σ_ε^2 is the pure error term (also known as nugget effect in geostatistical terminology), then according to Bayes theorem we can denote their joint posterior distribution as:

$$\pi(\theta, \psi | y) \propto \pi(y | \theta, \psi) \cdot \pi(\theta | \psi) \cdot \pi(\psi)$$

With the INLA methodology, all marginal posterior distributions are obtained by numerical approximations, which can be divided in three main tasks and which are all contained in the name INLA. In a nutshell, first Laplace approximations of the joint posterior distributions of the hyperparameters are estimated. Then, approximate computations for the conditional marginal posterior distributions of the latent field given the data and hyperparameters are computed; hence, the term Nested because these approximations are nested within one another due to the hierarchical model structure. Finally, the third stage combines the two previous steps by using numerical integration in order to obtain the marginal latent variable – thus the term Integrated. For more detailed information about the INLA procedure, please refer to Rue et al. (2009).

As priors, default distributions have been assigned for all parameter as recommended by Held et al. (2010), and which are already implemented in R-INLA. These priors are approximations of non-informative priors, designed to have little influence on the posterior distribution. Hence, results are mostly driven by the data and thus are similar to the frequentist approach for those cases where comparison are possible. For the spatial component we used the SPDE approach as defined in the previous section. We have assumed prior Gaussian distributions with mean zero and covariance matrix dependent to each hyperparameter as recommended by Lindgren and Rue (2013) and also implemented in R-INLA by default.

4.3.8. Model selection and evaluation

For all models, variable selection proceeded by a manually forward stepwise entry. In order to compare the goodness-of-fit of these models, we used the Deviance Information Criterion (DIC) which is equivalent to the Akaike Information Criterion (AIC) but better suited for HBMs (Spiegelhalter, 2002), and which is also directly computed by R-INLA. Briefly, DIC is given as the sum of the posterior mean deviance (\bar{D}) which quantifies model fit, and the effective number of parameters (p_D), which evaluates model complexity.

Additionally, as used by Muñoz et al. (2013) and Pennino et al. (2014), we used the Conditional Predictive Ordinate (CPO) for predictive power measure, where $CPO_i = p(y_i | y_{-i})$ (Geisser, 1993) and which is also available in R-INLA. This measure is defined as a cross-validated (or “leave-one-out”) predictive density at a given observation y_i , where y_{-i} denotes the data without the i ’th observation. However, Roos and Held (2011) indicated its logarithmic score for predictive purposes, which is defined as: $LCPO = -(n^{-1}) \sum_i \log(CPO_i)$. As R-INLA does not offer the LCPO in its default output, we had to compute it manually setting the following code: `[-mean(log(modelcpocpo))]`. Lower values for both DIC and LCPO are indicative of better fit and predictive power, respectively.

While sometimes LCPO and DIC may not agree on the same model, we selected the best model following three main steps. In these sense, considering that several model were tested regarding different probability distributions, the first step of our model selection relied on the LCPO evaluation between the different model distributions. Secondly, in cases where the LCPO values were similar between two or more model distributions, we used the DIC as second selection criterion in order to define the probability distribution that better suited our data. After having selected a set of models, we checked for models that only contained relevant predictors, i. e., those whose 95% credible interval ($CI_{95\%}$) did not covered zero. As our main objectives consisted in a model’s predictive quality, it did not made sense to select models containing some irrelevant predictors in its structure even if the LCPO and DIC would be smaller. Thus, the best model was finally chosen as the one which had only relevant predictors.

Selected models were then evaluated for goodness-of-fit according to graphical checks, such as observed versus predicted scatter plots and residuals quantile-quantil plots. Since linearity is expected between the observed and predicted values, Pearson's correlation coefficient (ρ) was calculated and tested. Models with $\rho \geq 0.7$ and p-value ≤ 0.05 were considered acceptable and thus used for final prediction.

4.4. Results

4.4.1. Length at first maturity & age grouping

Graphical evaluation of the logistic models revealed good convergence of the chains for all cases. Figure 2 shows the joint distribution of the logistic regression parameters for grouped and separated sexes. Despite all three cases had a reasonable fit, grouped sexes revealed a slightly better fit as the joint distribution of model's parameters was almost entirely concentrated within the right upper quadrant (Fig. 2).

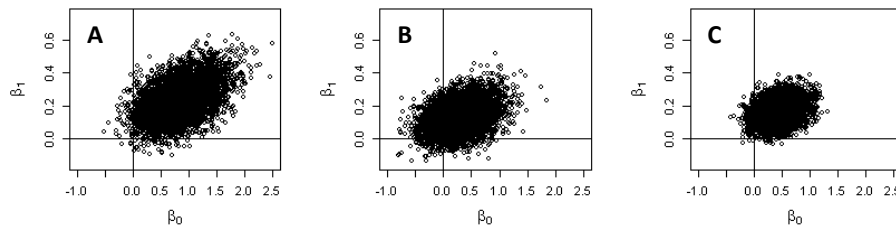


Figure 2: Posterior joint distribution of logistic regression parameters for females (A), males (B) and grouped sexes (C).

Table II summarizes some basic statistic summaries of the posterior distribution for the logistic models. A good convergence was also observed by the numerical outputs, whose \hat{R} values for each parameter were all situated around 1.0. With respect the L_{50} estimation, both mean and median values either for females and males were very similar (Table II) (Fig. 3A and B).

The Bayesian hypothesis test pointed out that there were no differences between the L_{50} of males and females (Fig. 3D). Based upon this information we decided to use the L_{50} estimation derived from the grouped sexes model, from which we used its

median value (25.17 cm) to split the lane snapper's stock into juveniles and adults for the models of the next section.

Table II: Summary of parameters and their associated statistics resulted from the logistic model. (Sd = standard deviation ; $CI_{95\%}$ = 95% credible interval).

Sex	Parameters	Mean	Median	Sd	$CI_{95\%}$		\hat{R}
					$Q_{0.025}$	$Q_{0.975}$	
Females	β_0	0.835	0.820	0.387	0.112	1.633	1.001
	β_1	0.230	0.227	0.103	0.041	0.450	1.001
	L_{50}	25.332	25.870	12.757	16.835	29.002	1.063
	Deviance	16.728	16.163	2.004	14.749	22.280	1.001
Males	β_0	0.306	0.297	0.331	-0.322	0.978	1.001
	β_1	0.128	0.125	0.080	-0.022	0.291	1.001
	L_{50}	25.375	25.072	77.786	10.546	37.669	1.068
	Deviance	24.368	23.758	2.111	22.37	29.831	1.001
Females & Males	β_0	0.457	0.456	0.238	-0.010	0.927	1.002
	β_1	0.161	0.159	0.061	0.044	0.285	1.002
	L_{50}	24.653	25.174	5.996	18.701	28.094	1.038
	Deviance	35.062	34.443	2.026	33.075	40.667	1.001

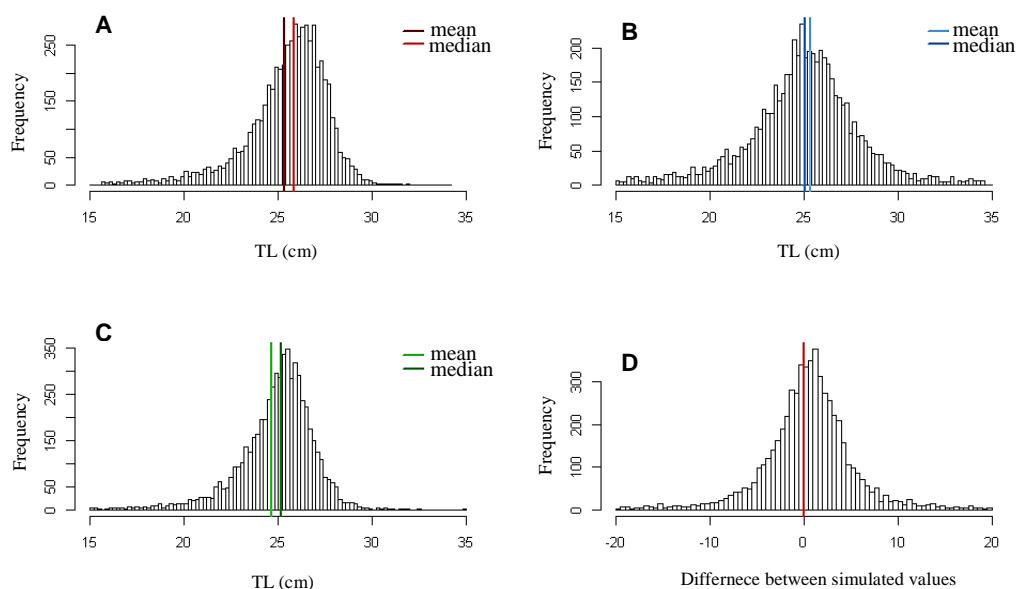


Figure 3: Posterior distribution histograms of females (A), males (B), females & males (C) and the difference between males and females (D) total length (TL) simulations.

4.4.2. Data overview

Among the 126 sample stations, lane snapper occurred at 83 (66.14 %) whose distance to coast ranged from 2.8 to 41 km (mean = 11.58 km; sd = 7.86) and depths from 5.1 to 52.7 m (mean = 16.8 m; sd = 8.85). The total weight caught during this period was 259.808 kg, whereas CPUE ranged from 0 to 6.12 g/m²*h (mean = 0.35 g/m²*h; sd = 0.85). According to the adopted L₅₀ we counted a total of 101 juveniles and 505 adults individuals, thereby evidencing that of the catches were almost composed by adult individuals (83.3%). Whereas adults occurred in almost all sample stations (81), juveniles occurred in 36 sample station only. Total length of all specimen varied from 18.5 cm to 46 cm (mean = 30.1 cm; sd = 3.92).

4.4.3. Model selection

Among all environmental predictors, only bathymetry and distance to coast were highly correlated ($\rho > 0.7$). As distance to coast usually performed a better fit, we used it rather than bathymetry for modeling purposes. Several models with different probability distributions have been tested according to data nature. We also tested models that included quadratic terms and/or smoothing effects which, however, did not show any relevant fit improvement.

Table III summarizes the most relevant models regarding different combinations of environmental predictors for lane snappers CPUE. LCPO and DIC did not differ greatly within models of each distribution. Despite this, DIC differed significantly between models with different distributions, where Gamma models performed better fit. Thus, models with Gamma distribution indicated always the better fit and predictive quality, whose best model included only distance to coast, chlorophyll-a concentration and the spatial effect (model 8, Table III).

Table III: Model comparison for the CPUE where the best model is highlighted in bold. Predictors acronyms are: DC = Distance to coast, R = Rugosity, CHLa = Chlorophyll-a concentration, SST = Sea Surface Temperature, S = Slope, SE = Season, W = Spatial effect. Relevant environmental predictors are indicated by *.

	Model	Log-Normal		Gamma	
		DIC	LCPO	DIC	LCPO
1	1 + DC* + W	460.5	1.8729	-165.2	1.7124
2	1 + DC* + R + W	460.0	1.8720	-166.3	1.7022
3	1 + DC* + R + CHLa* + W	461.1	1.8745	-165.2	1.6781
4	1 + DC* + R + CHLa* + SST + W	461.7	1.8779	-164.7	1.7335
5	1 + DC* + R + CHLa* + SST + S + W	462.9	1.8884	-164.7	1.8138
6	1 + DC* + R + CHLa* + SST + S + SE + W	458.2	1.8652	-166.6	1.4774
7	1 + CHLa + W	462.8	1.8858	-164.2	1.7794
8	1 + DC* + CHLa* + W	461.2	1.8732	-164.4	1.6674
9	1 + DC* + CHLa*	486.5	1.9290	-120.0	NA
10	1 + W	462.3	1.8827	-164.5	1.7658

With respect to models for adult and for juvenile individuals, both agree on the same distribution when LCPO and DIC were analyzed sequentially (Tables IV and V, respectively). Despite models with Poisson distributions had always the lowest DIC, LCPO values were almost the highest for all tested cases. ZINB models presented always the highest DIC values, although LCPO were more or less similar to those of the NB. With respect to ZIP models, DIC values were very similar to those of the NB but failed when the LCPO was taken into account. When evaluating first the predictive quality and after the model fit, NB models presented the most reasonable results for both cases. Thus, the best selected model for adults count data included distance to coast, chlorophyll-a concentration and the spatial effect (model 8, Table IV), whereas for juveniles count data we selected the second model which contained distance to coast, rugosity and the spatial effect (Table V). Posterior summary statistics for all parameter of each selected model are shown in Table VI.

Table IV: Model comparison for adults count data where the best model is highlighted in bold. Predictors acronyms are the same as stated in table VI and NA suggests inappropriate models. Relevant environmental predictors are indicated by *.

Model		P		ZIP		NB		ZINB	
		DIC	LCPO	DIC	LCPO	DIC	LCPO	DIC	LCPO
1	1 + DC* + W	537.3	5.6107	543.9	4.0918	544.5	2.777	579.5	2.7059
2	1 + DC* + R + W	519.2	5.6457	543.6	4.1055	541.3	2.7634	580.4	2.8877
3	1 + DC* + R + CHLa* + W	519.6	5.684	543.4	4.1164	542	2.7256	580.1	2.3182
4	1 + DC* + R + CHLa* + SST + W	519.8	5.7941	543.7	4.143	541.9	2.7423	579.6	NA
5	1 + DC* + R + CHLa* + SST + S + W	517.6	5.6463	543.5	4.2025	542	2.6907	582.0	2.8568
6	1 + DC* + R + CHLa* + SST + S + SE + W	517.8	5.6012	544.3	4.2696	540.6	2.7237	585.9	NA
7	1 + CHLa + W	537.6	5.6789	544.7	4.1004	545.5	2.7172	576.0	2.9112
8	1 + DC* + CHLa* + W	538.1	5.6458	543.9	4.1105	544.5	2.7917	579.7	2.518
9	1 + DC* + CHLa*	1172.2	4.858	1021.5	4.2685	573	2.2742	787.2	3.1202
10	1 + W	539.3	5.9838	553.6	4.0077	555.2	2.7795	579.3	2.8547

Table V: Model comparison for juveniles count data where the best model is highlighted in bold. Predictors acronyms are the same as stated in table VI and NA suggests inappropriate models. Relevant environmental predictors are indicated by *.

Model		P		ZIP		NB		ZINB	
		DIC	LCPO	DIC	LCPO	DIC	LCPO	DIC	LCPO
1	1 + DC* + W	251.6	6.4929	268.3	2.1429	262.2	1.0598	385.5	1.5264
2	1 + DC* + R* + W	240.4	7.1448	269.1	7.7364	261.3	1.0533	385.1	1.5265
3	1 + DC* + R* + CHLa + W	240.4	7.3213	269.2	13.302	261.5	1.0591	385.2	1.5265
4	1 + DC* + R* + CHLa + SST + W	240.5	7.4566	269.2	18.8071	260.9	1.0554	385.1	1.5262
5	1 + DC* + R* + CHLa + SST + S + W	230.4	7.4465	269.1	46.4482	256.9	1.1636	385.1	1.5263
6	1 + DC* + R* + CHLa + SST + S + SE + W	216.5	7.5211	- Inf.	NA	242.1	2.8561	385.4	1.5266
7	1 + S + SST* + W	231.1	7.0302	267.0	18.7967	256.2	1.0516	385.0	1.5252
8	1 + R + S + SST* + W	230.1	7.2151	268.2	29.8947	255.7	1.0553	385.6	1.5284
9	1 + DC* + R*	419.0	1.7027	300.8	1.2395	280.7	1.1115	385.9	1.5279
10	1 + W	250.2	6.8047	271.3	2.2400	256.0	1.0351	278.9	NA

Table VI: Summary of the marginal posterior distribution for model parameters provided by the selected model for each considered case. The hyperparameters ϕ_G and ϕ_{NB} are the precision parameters of the gamma and negative binomial observations, respectively, and τ and κ represents the variance and scaling parameter of the spatial effect, respectively.

Response variable	Parameters	Mean	Sd	CI _{95%}	
				Q _{0.025}	Q _{0.975}
CPUE (model 8)	(Intercept)	-1.873	0.223	-2.324	-1.442
	DC	0.859	0.254	0.353	1.356
	CHLa	0.529	0.293	-0.058	1.105
	τ	-5.142	0.465	-6.075	-4.248
	κ	3.776	0.379	3.046	4.536
	ρ (km)	0.069	0.026	0.025	0.122
	Φ_G	0.983	0.156	0.709	1.321
Adults (model 8)	(Intercept)	-8.179	0.264	-8.733	-7.684
	DC	0.951	0.274	0.417	1.503
	CHLa	0.606	0.319	-0.017	1.249
	τ	-5.168	0.53	-6.21	-4.128
	κ	3.758	0.413	2.948	4.57
	ρ (km)	0.072	0.03	0.023	0.132
	Φ_{NB}	1.323	0.49	0.613	1.093
Juveniles (model 2)	(Intercept)	-10.413	0.718	-11.975	-9.075
	DC	0.57	0.356	-0.115	1.299
	R	0.44	0.35	-0.2112	1.735
	τ	-4.607	0.806	-6.048	-2.903
	κ	3.042	0.617	1.766	4.182
	ρ (km)	0.192	0.156	0.027	0.488
	Φ_{NB}	0.863	0.592	0.18	2.386

Figure 4 denotes 2 types of models fit evaluation. Predicted versus observed CPUE, number of adults and number of juveniles are displayed in the left panels of Figure 4. As it can be noticed, for all three cases the predicted versus observed values were positively and significantly correlated ($0.74 < \rho < 0.88$, $p < 0.05$) (Fig. 4 A, C and D). Q-Q plots shown reasonable normal distribution for the residuals of each selected model, despite some outliers were presented for juveniles and adults (right panels of Fig. 4 B, D and E). The CPUE model was able to predict fairly non-zero values while values around zero were highly overestimated (Fig. 4 A). Thus, together with the respective Q-Q plot both figures clearly suggests that our data is

split into two data-sets, one regarding the zeros and another regarding non-zero values. For adults and juveniles, both models were able to predict with more accuracy small to medium values (Fig. 4 C and D).

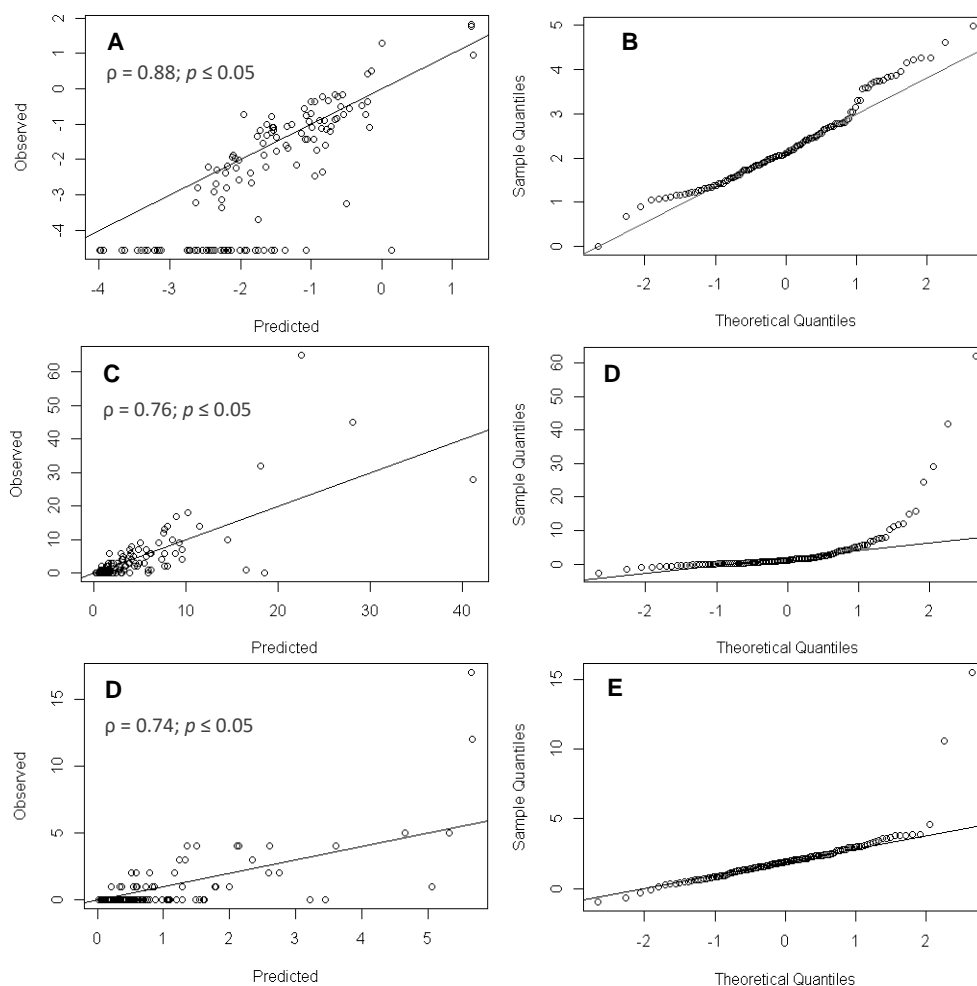


Figure 4: Observed versus predicted values for log(CPUE) (A), number of adults (B) and number of juveniles (C). Right panels shows quantile-quantil plots for each evaluated case.

4.4.4. Prediction of essential fish habitat for the lane snapper

When combining both the fixed effects and the model output from the random field, we can easily display the mean and standard deviation for the posterior mean of the linear predictor as well for the random field itself. Figure 5 displays mean and standard deviation prediction maps for each considered situation. According to figure 5 A and C, both CPUE as adults of *L. synagris* have similar patchy distributions along the entire coast as well as slightly constant densities along the northern shore. This is expected given that the largest amount of the captured stock was composed of adults. Higher densities were also found towards the slope (northeastern region of Figures 5 A and C). Juveniles, on the other hand, showed two big aggregation spots along the east coast and some higher patch toward the slope too (Fig. 5 E). The standard deviation for all linear predictors was small, uniform and behaved in similar manner for the entire domain (Fig. 5 B, D and F). Thus, this supports that the high densities observed specifically for CPUE and for adults toward northeast are real potential indicators of offshore area concentrations, since the standard deviations for the same region were small.

The smaller panels in each figure reflects the spatial random effect and indicates the intrinsic variability of the lane snappers distribution when removing all covariables. It may reflect other hidden factors that were not accounted in the models, such as biotic processes (e.g, competition and predation) and abiotic processes (e.g, sea surface salinity and currents). Again, for both CPUE and adults the mean spatial component pattern was similar between each other, where the existence of a higher number of aggregation spots along the coast can be noticed (Fig. 5 A and C). Therefore, through these panels we can realize that it seems that the present species actually is more widely distributed since these maps were not so consistent when contrasted with their respective mean linear predictor maps. When dealing with the spatial random effect of juveniles, we can observe that, besides the northeastern coast, it was consistent with its respective prediction map (Fig. 5 E).

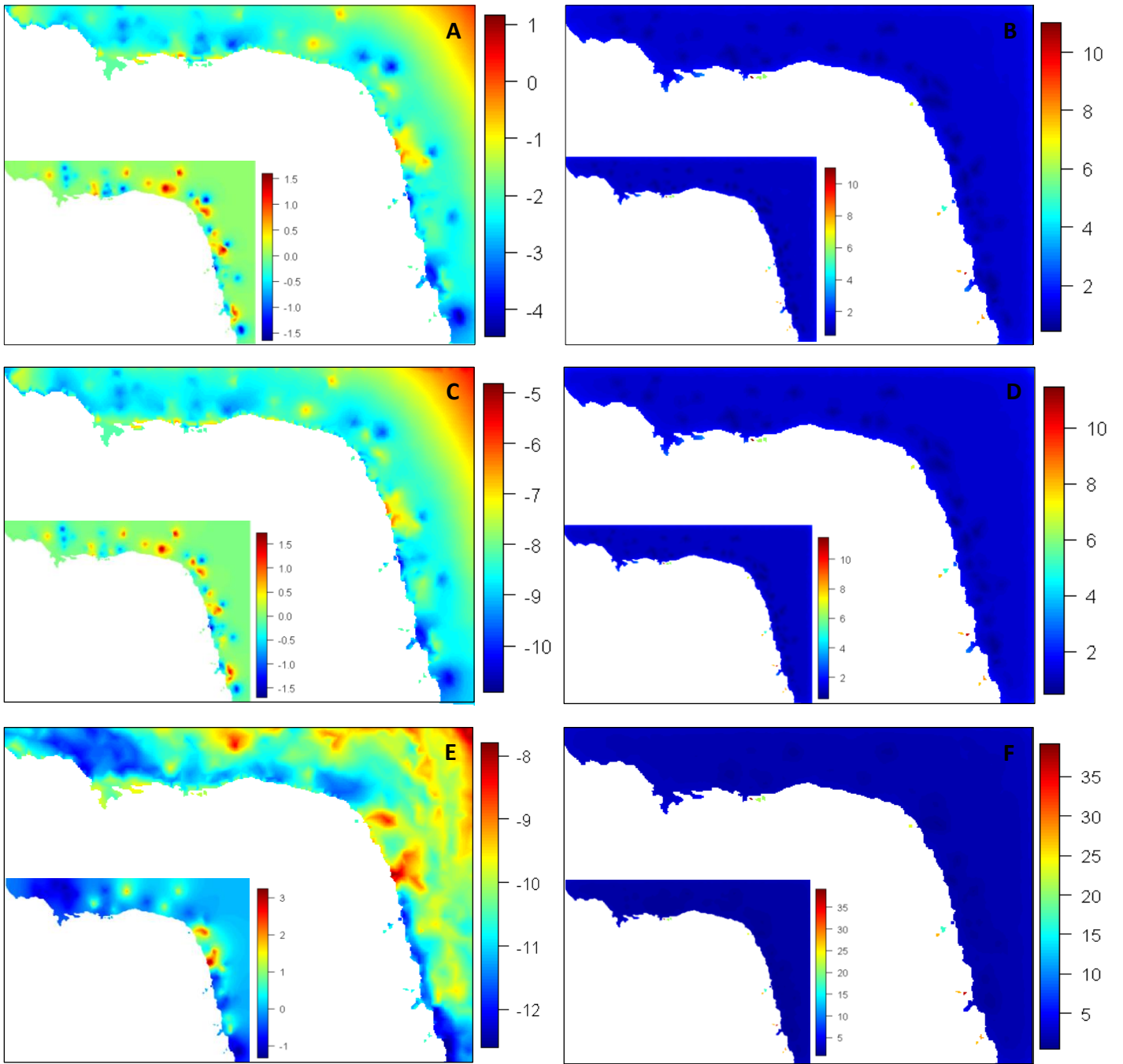


Figure 5: Posterior mean (left panels) and standard deviation (right panels) of the predictive distribution of lane snappers CPUE (A,B), number of adults (C,D) and number of juveniles (E,F). Small panels represent the mean and standard deviation of the spatial random effect for their respective response variable. It is worth mentioning that the scales of each map are log-scaled.

4.5. Discussion

4.5.1. Biological discussion: size at first maturity

The first part of our result aimed to estimate the size of first maturity for *L. synagris*. Although differences in this parameter among sexes are commonly reported for lane snappers from different regions (Aiken, 2001; Figuerola et al., 1998; Freitas et al., 2014; Luckhurst et al., 2000; Manickchand-Dass, 1987), our results indicated otherwise. Thus an overall size for both sexes was estimated at lengths of 25.17 cm. Our estimation agreed with those from Cavalcante et al. (2012) also conducted in the Rio Grande do Norte state, which reported mean sizes of 25.7 cm for grouped sexes.

4.5.2. Model discussion

The Bayesian spatial modeling analysis yielded some clear results with respect to the environmental preferences of adults, juveniles and their overall abundance (CPUE). Despite most models fitted well to the data, some few models for each response variable suggested that they were inappropriate (producing by *NA* and *-Inf.* in Tables III-V). Moreover, when attempted to the count models (Tables IV and V), it seemed that number of adults and juveniles always fitted better (lower DICs) according Poisson or its Zero-inflated version when compared to the alternative models belonging to other probability distributions. Nevertheless, these models also displayed the worst predictive measures (higher LCPOs). This could be explained since extreme low values of CPO are indicative of outliers and influential observations. Thus, the lower CPO values the higher LCPO values which, in turn, reveals that these model parameters were biased and consequently inappropriate for prediction purposes.

Regarding the models for CPUE, we concluded that besides the spatial effect, only distance to coast and chlorophyll-a concentration were statistically relevant to explain the variability in CPUE. Table VI reveals statistical evidence of a positive relationship between the response and both fixed effects, indicating therefore that the CPUE of *L. synagris* increases toward offshore areas and higher chlorophyll-a

concentrations. The remaining parameters are the hyperparameters that specify the spatial effect in the model.

R-INLA usually provides the simplest internal representation of these parameters, namely $\log(\tau) = \theta_1$ and $\log(\kappa) = \theta_2$. However, it is more natural to construct these parameters as a function of the spatial correlation range (ρ), since τ and κ have jointly influence on the marginal variances of the resulting spatial field. For practical purposes we exposed all hyperparameters, but we mainly focused on ρ as it has more intuitive interpretation. In this way, the mean value of (ρ) was 0.069 which represents the distance at which correlation is reduced to approximately 0.13. The posterior mean of the precision for the gamma observations (Φ) was 0.983. It is worth noting that the selected CPUE model suggested that our data would have probably been better fitted, if we had conditioned the zero values. Thus, if we had modeled our CPUE according a Gamma hurdle model as done by Quiroz et al. (2015) with Peruvian anchovies biomass, then we maybe could have had more precision estimations and predictions.

With respect to the count models applied for number of adults and juveniles, we were able to affirm that they respond differently to distinct environmental predictors. Likewise CPUE, only distance to coast and chlorophyll-a concentration were statistically relevant to explain the variation in the number of adults (Table IV). Furthermore, Table VI revealed that all fixed effects had also positive coefficients. Thus, the same idea could be applied for this response variable, i.e., number of adults tend to increase toward offshore areas and higher chlorophyll-a concentrations. Number of juveniles, instead, showed a different response pattern. Distance to coast was also relevant and, although rugosity was almost relevant, we still considered it since it has ecological importance for this species (Table VI). According to Table VI it was found that number of juveniles also surprisingly increases toward offshore areas, as well as at higher rugosity index.

Concerning the hyperparameters from both count models, it was noted that juveniles showed slightly smaller values for mean variance and scaling parameter when compared to adults (Table VI). The mean spatial correlation range (ρ) for adults and juveniles were 0.072 and 0.192, respectively (Table VI). Contrasting these numbers with that from CPUE (0.069), it seems that despite weak spatial correlation are displayed in all cases, juveniles apparently are spatially more correlated.

Moreover, when dealing with negative binomial models we can evaluate if the data are spatially clustered, which in turn reflects in its spatial autocorrelation. Under these circumstances, the precision parameter (Φ) of such models is also known as the overdispersion parameter, and is related to the variance $\sigma^2 = \mu(1 + \mu/n)$. As the sample n becomes large the variance approaches the mean, and thus we fall back to the Poisson distribution where variance equals the mean. On the other hand, smaller values of n are indicative of overdispersion ($\Phi > 0$). According to our estimates for the both remaining cases, we have evidence that number of adults as well of juveniles was clustered over the spatial field since this parameter was relevantly greater than zero (Table VI).

It is worth mentioning that for each selected model discussed in this section, we also tested them without the spatial effect (see models 9 in Tables III-V) as they had small spatial correlation ranges. It could be seen that these models used to have higher DICs and LCPOs when compared to the remaining models, which in turn reveals that the spatial effect, despite small, was indeed relevant in all three cases.

4.5.3. Ecological discussion

Despite biological studies such as age, growth, feeding habits and reproduction are widely available for *L. synagris*, to date it is still poorly known what are their ecological preferences. Through the HBMs we were able to investigate the relationship between the life stages of *L. synagris* and environmental predictors, and therefore quantify and define suitable habitat according to relevant predictors.

As discussed in the previous section, we observed that the adults had the same environmental preferences as CPUE. The fact that the captured stock for this species was basically composed by individuals of this age group, showed that the CPUE was primordially driven by adults. This also explains why the spatial distribution pattern of CPUE and adults was nearly identical. Because of this, the discussion of this section will primarily focus on juvenile and adult individuals, since the former reflect almost the same pattern observed in the CPUE.

As already noted by Rodríguez and Páramo (2012) in Colombia, the predictive maps of the present study also revealed a clear spatial segregation between juveniles and adults along the Rio Grande do Norte coast. Despite both groups

seemed to prefer offshore areas, they had particular preferences, such as higher chlorophyll-a concentrations for adults and higher rugosity index in the case of juvenile individuals. Thus, these results were somewhat surprising since a reverse pattern have been expected with respect these particular requirements.

The literature usually reports that juveniles of lane snapper typically occur over mud bottoms near marine estuarine and seagrass beds and, therefore near shallow coastal systems, whereas adults tend to prefer consolidated bottoms in towards offshore areas (Claro and Lindeman, 2004; Doncel and Paramo, 2010; Franks and VanderKooy, 2000; Mikulas and Rooker, 2008; Pimentel and Joyeux, 2010; Rodríguez and Páramo, 2012). In this sense, the occurrence of a greater number of juveniles toward offshore areas and, consequently in deeper waters, was also an unexpected result. Although Manickchand-Dass (1987) has also reported an inverted distribution pattern for juveniles and adults with respect depth, we believe that our results probably reflect more to an erroneous definition of juveniles in this work.

As the catches were practically composed only by adults, this study revealed that the fishing gear used in this work had a clear size selectivity for the *L. synagris* populations. Among the few catches of juveniles, it was observed that they already had relatively large sizes, whose smallest measured size was 18.5 cm. Whereas the terminology *juveniles* is usually applied to individuals with sizes shorter than 10 cm (Franks and VanderKooy, 2000; Lindeman et al., 1998; Mikulas and Rooker, 2008; Pimentel and Joyeux, 2010), we concluded that individuals treated as juveniles in this study should be better regarded as sub-adults. Thus, what we observe in fact is the spatial segregation among adult and sub-adult stages of *L. synagris*.

Although both age-groups showed preference toward offshore areas, we identified some hot spots closer to the coast. In the case of number of adults, the predictive map highlighted three main hot spots, which curiously correspond to reef complexes present in the Rio Grande do Norte coast. Also, when we attempted to its respective spatial random effect map, we noted that it indicated several isolated areas with higher occurrence mainly along the east coast. As previously commented, the spatial random effect is a reflection of environmental variables that are important to the investigated ecological process which, however, were not included in the model. Even though the models did not show relevance for rugosity and/ or slope for the adults, both used here as a proxy for substrate type, and given that on the east

coast of Rio Grande do Norte many reefs and sand banks are recorded (Vital et al., 2010), the hot spots identified by the spatial effect strengthened the idea that adults of *L. synagris* seemed to aggregate preferably around areas marked by consolidated bottoms.

The aggregation of adults around reef regions is also indirectly related to their preferences for higher chlorophyll-a concentrations, which constitutes a proxy to primary productivity. It is well known that in marine environments, higher chlorophyll-a concentrations occur mainly in coastal areas and in upwelling fronts, due to higher nutrient availability and sunlight incidence. Reefs, in particular, are areas of high primary productivity, and thus allow to support higher trophic food webs from which adults of *L. synagris*, among many other fishery resources, find favorable means to sustain (Crossland et al., 1991).

With regard the sub-adult aggregations, it was found that besides aggregating on the same hot spots as adults do, they also seemed to occur preferentially along the continental shelf break, which is clearly marked on the predictive map (Fig. 5E). According Vital et al. (2010), this region represents a transition zone between consolidated (high rugosity index) and unconsolidated (low rugosity index) bottoms, which, in turn, allow us to characterize it as a marine ecotone. Furthermore, Feitoza et al. (2005) reported the presence of *L. synagris* over some deep reef systems (45-70 m depth) along the coast of the Rio Grande do Norte, Paraíba and Pernambuco states. Although deep reefs are poorly mapped in the Rio Grande do Norte state and, considering that the continental shelf of this area is very narrow and shallow (mean width of 40 km and maximum depth of 60 m) (Vital et al., 2010), it seemed reasonable to suppose that these areas might contain deep reef systems or, at least, sand banks, as they showed high rugosity index.

4.6. Concluding remarks

A set of statistical approaches were used in order to extend the scope of a data limited fishery. By means of Bayesian models and Geographic Information Systems (GIS) schemes we provided some novel insights of the potential spatial distribution for abundance and different life stages of the vulnerable lane snapper. However, our

case study attempted to present a new and emerging SDM tool, rather to focusing solely on traditional biological and ecological discussions.

Marine ecosystems constitute dynamic areas, where the replication of fisheries experiments are almost impossible and are also vulnerable to several error sources associated with observations, sampling procedures, model structure and parameters. In this sense it is convenient to apply Bayesian inference into fisheries modeling procedures, since the posterior distribution itself is a dynamic process initially shaped according to our prior beliefs, which, however, evolves as we acquire more data. Additionally, when dealing with spatial analysis, fisheries ecology mostly uses the spatial component as a fixed effect rather than a random effect, and thus ignores explicitly the spatial autocorrelation which is inherent in fisheries.

The Bayesian hierarchical spatial models proposed in this study are extremely powerful and suit perfectly in fisheries science, by quantifying both the spatial magnitude and the different sources of uncertainty. Spatial predictions, therefore, are much more accurate and consistent with reality. Furthermore, as we are dealing within a Bayesian framework, posterior distribution predictions maps like ours enables the decision makers to take much more efficient fishery management and conservation measures since some decisions, such as those defining the probability of the most favorable areas to conserve, would be antagonistic within traditional frequentist approach to probabilities.

However, it is noteworthy that the models presented in this study are limited not only in space but also in time. Thus, the fitted and predicted models revealed only a snapshot of the ecological process. Since fisheries are dynamic in space and time, we encourage the use of both effects as they may improve even more realism. Bayesian spatio-temporal models were already applied in some important fisheries issues, such as discard and by-catch problems (Cosandey-Godin et al., 2015; Pennino et al., 2014). Moreover, when dealing with fishery-dependent data, it is important to account for distinct sources of biases as those are related to fishers behaviour, and which has also been recently introduced in models performed by Roos et al. (2015).

Our study, demonstrated once more how easy it is nowadays to account for random effects in HBMs when performed through the INLA methodology. R-INLA revolutionized the way we may perform easily Bayesian inference, since its interface

is similar to the conventional *glm* function in R and thus we do not have to write minusciously the models as done in BUGS and JAGS.

Besides this, R-INLA is continuously evolving and greatly extending the scope of Bayesian models for applied scientists. Using R-INLA enables us to fit complex models at considerable lesser time and with more accuracy when comparing to standard MCMC methods, although MCMC is still a slightly more flexible approach.

Finally, it is worth mentioning that R-INLA can also be used to fit models that does not have necessarily spatial and/or temporal components in its structure. Rue et al. (2009) provide some other examples, which include generalized linear and additive (mixed) models, dynamic linear models, survival models, spline smoothing and semiparametric regressions, among many others. In this way, we also motivate the use of R-INLA when performing ecological Bayesian modeling with conventional GLMs or GAMs, since the package's interface displays similarity with already existing *glm* and *gam* tools in R. At least, we have shown that, even for “data-poor” fisheries, this modeling approach has good performance.

4.7. Acknowledgments

Authors would like to gratefully thank María Grazia Pennino for her valuable technical support in the usage of R-INLA. The first author is also grateful to the Brazilian National Research Council (CNPq), which provided financial support during her M. Sc. research. Finally, all authors wish to thank to all fishermen, researchers and (under)graduate students which contributed with both field and laboratory works, and therefore enabled this study.

Capítulo 5

CONCLUSÃO GERAL

O presente trabalho combinou técnicas de sistemas de informação geográfica (SIG) e geoestatística Bayesiana com o intuito de promover novas informações a respeito da distribuição espacial da abundância e grupos de idades do ariocó. Constataram-se preferências diferenciadas quanto aos preditores ambientais para os grupos de jovens e adultos, fato este que refletiu em ocupações espaciais distintas. Muito mais que discussões bioecológicas tradicionais, este trabalho se propôs a apresentar uma nova e emergente ferramenta estatística que visa auxiliar não só uma definição mais precisa de medidas prioritárias para conservação de espécies ameaçadas, como também incentivar o uso da inferência Bayesiana na análise de dados ecológicos.

Os ecossistemas marinhos constituem áreas dinâmicas, aonde a replicação de experimentos pesqueiros raramente é possível de ser realizada, além de estar passível a diversas fontes de erros tais como aquelas associadas às observações, metodologia de coleta, estrutura do modelo e parâmetros. Nesse sentido foi demonstrado o quão conveniente se torna aplicar a inferência Bayesiana nos procedimentos de modelagem de dados pesqueiros, tendo em vista que a própria distribuição posterior constitui um processo dinâmico no sentido em que se torna possível atualizá-la à medida em que se adquire novas informações a respeito do processo estudado.

Os modelos espaciais Bayesianos hierárquicos propostos no presente estudo são extremamente potentes e demonstraram se adequar perfeitamente nas ciências pesqueiras, uma vez que eles permitiram quantificar tanto a magnitude do efeito espacial como também as diferentes fontes de incertezas. Em função disso as previsões espaciais foram muito mais acuradas e consistentes com a realidade. Além disso, como estamos tratando com uma abordagem Bayesiana, os mapas preditivos da distribuição posterior como os apresentados neste trabalho permitem aos tomadores de decisão decidirem por medidas de manejo e conservação muito mais eficientes, haja vista estes resultados permitirem definir, por exemplo, a probabilidade das áreas mais favoráveis para conservação, o que seria impossível de realizar dentro da definição frequentista de probabilidades.

Contudo, cabe ressaltar que os modelos aqui apresentados são limitados não só no quesito espacial como também temporal. Sendo assim, os modelos ajustados e preditos revelaram apenas uma fotografia instantânea do processo ecológico.

Considerando que os recursos pesqueiros são dinâmicos no espaço e tempo, destaca-se a importância em incorporar ambos os efeitos nos modelos uma vez que conjuntamente eles podem promover mais realismo ao processo estudado. Além disso, fica claro também que os modelos fizeram apenas o uso de uma fração minoritária de preditores ambientais potencialmente importantes para espécie, ignorando possíveis variáveis bióticas à exemplo de indicadores de competição e predação. Portanto, é evidente que mesmo incorporando o efeito espacial, os modelos foram relativamente simples no quesito ecológico, ressaltando a necessidade de incorporar outros componentes para melhor entendimento da bioecologia do ariocó.

REFERÊNCIAS BIBLIOGRÁFICAS

~ A ~

ABREU, M. C. 2005. Distribuição espacial e temporal do ariocó, *Lutjanus synagris* (Linneus, 1758), no estuário dos rios Piraquê-Açu e Piraquê-Mirim (ES) e seu crescimento em ambiente estuarino. Trabalho de Conclusão de Curso, Universidade Federal do Espírito Santo, Brasil, 61 p.

ACOSTA, A. & APPELDOORN, R. S. 1992. Estimation of growth, mortality and yield per recruit for *Lutjanus synagris* (Linnaeus) in Puerto Rico. *Bull. Mar. Sci.*, 50: 282-291.

AILLOUD, L. 2010. < <http://nw08.american.edu/~vconn/seafood/techniques.htm> > acesso em março de 2015.

AIKEN, K. A., 2001. Aspects of reproduction, age and growth of the lane snapper, *Lutjanus synagris* (Linnaeus, 1758) in Jamaican coastal waters. *Proc. Gulf Carib. Fish. Inst.*, 52: 116-134.

ALEGRÍA, J. R. C. & MENEZES, M. F. 1970. Edad y crecimiento del ariacó, *Lutjanus synagris* (Linnaeus), en el nordeste del Brasil. *Arq. Ciên. Mar.*, 10: 65-68.

AMARO, V. E & FERREIRA, A. T. S. 2012. Imagens do AQUA-MODIS aplicadas à estimativa dos valores de clorofila-a e da temperatura da superfície do mar na zona offshore da Bacia Potiguar no Rio Grande do Norte. *Ver. Geol.*, 25: 128-142.

AUGUIE, B. 2012. GridExtra: functions in grid graphics. R package version 0.9.1, <http://CRAN.R-project.org/package=gridExtra>.

~ B ~

BABCOK, E. A.; PIKITCH, E. K.; MCALLISTER, M. K.; APOSTOLAKI, P. & SANTORA, C. 2005. A perspective on the use of spatialized indicators for ecosystem-based fishery management through spatial zoning. *ICES Journal of Marine Science* 62: 469-476.

BANERJEE, S; CARLIN, B. P. & GELFAND, A. E. 2004. Hierarchical modeling and analysis for spatial data, New York, CRC Press, 451p.

BARROS, M. J; CORREA, J. D. & MANJARRÉS, M. 1996. Analisis biológico pesquero del pargo rayado *Lutjanus synagris* (Linnaeus, 1758). *Bol. Cient.*, 4: 79-105.

BENAKA, L. R. 1999. Fish habitat: essential fish habitat and rehabilitation. American Fisheries Society, Connecticut, 459 p.

BENTLEY, N. 2014. Data and time poverty in fisheries estimation: potential approaches and solutions. *ICES J. Mar. Sci.*, 72: 186-193.

BERRY, D. A. 1995. Statistics: A Bayesian perspective. Duxbury Press, 518p.

BIVAND, R. & LWEIN-KOH, N. 2015. Maptools: tools for reading and handling spatial objects. R package version 0.8-34, <http://CRAN.R-project.org/package=maptools>.

~ C ~

CADDY, J. 1997. In review of the state of world fisheries resources: Marine Fisheries. *FAO Fish. Tech. Doc.*, 920 p.

CARPENTER, K. 2002. The living marine resources of the Central Western Atlantic. Volume 3: Bony fishes part 2 (Opistogonathidae to Molidae), sea turtles and marine mammals. Roma, FAO, 2127p.

CAVALCANTE, L. F. M; OLIVEIRA, M. R. & CHELLAPPA, S. 2012. Aspectos reprodutivos do ariacó, *Lutjanus synagris* nas águas costeiras do Rio Grande do Norte. *Bio. Amaz.*, 2: 45-50.

CHAKRABORTY, A; GELFAND, A. E; WILSON, A. M; LATIMER, A. M & SILANDER, J. A. 2010. Modelling large scale species abundance with latent spatial processes. *An. Appl. Stat.*, 4: 1403-1429.

CIANNELLI, L.; FAUCHALD, P.; CHAN, K.S.; AGOSTINI, V. N. & DINGSOR, G. E. 2008. Spatial fisheries ecology: recent progress and future prospects. *J. Mar. Syst.*, 71: 223-236.

CLARK, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecol. Let.*, 8: 2-14.

CLARO, R. 1981a. Ecología y ciclo de vida de la bíaiba, *Lutjanus synagris* (Linnaeus), en la plataforma cubana. II - biología pesquera, *Rep. Invest. Inst. Oceanol.*, 177: 1-53.

CLARO, R. 1981b. Ecología y ciclo de vida de la bíaiba, *Lutjanus synagris* (Linnaeus), en la plataforma cubana. III - nutrición, *Cien. Biol.*, 6: 93-110.

CLARO, R. 1982. Ecología y ciclo de vida de la biajaiba, *Lutjanus synagris* (Linnaeus), en la plataforma cubana. IV – reproducción. *Rep. Invest. Inst. Oceanol.*, 5: 1-37.

CLARO, R. & LINDEMAN, K. C. 2004. Biología y manejo de los pargos (Lutjanidae) en el Atlántico Occidental. Instituto de Oceanología, CITMA, La Habana, Cuba. 472p.

COSANDEY-GODIN, A.; KRAINSKI, E. T.; WORM, B. & FLEMMING, J. M. 2015. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. *Can. J. Fish. Aquat. Sci.*, 72: 1-12.

COTTER, J.; PETITGAS, P.; ABELLA, A; APOSTOLAKI, P; MESNIL, B; POLITOU, C; RIVOIRARD, J; ROCHET, M. J; SPEDICATO, M. T; TRENKEL, V. M. & WOILLEZ, M. 2009. Towards an ecosystem approach to fisheries management (EAFM) when trawl surveys provide the main source of information. *Aqua. Liv. Res.*, 22: 243-254.

CRESSIE, N. A. C. 1993. Statistics for spatial data – Revised edition. New York, Wiley Series, 904p.

CRESSIE, N. A. C.; CLARK, J. S.; Ver Hoef, J. M. & WIKLE, C. K. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.*, 19: 553-570.

CROSSLAND, C. J.; HATCHER, B. G. & SMITH, S. V. 1991. Role of coral reefs in global ocean production. *Coral Reefs*, 10: 55-64.

~ D ~

DEEPAYAN, S. 2008. Lattice: multivariate data visualization with R. New York, Springer, 296p.

DENNIS, B. 1996. Should ecologists become Bayesians? *Ecol. Appl.*, 6: 1095-1103.

DIGGLE, P. J. & RIBEIRO Jr., P. J. 2007. Model-based Geostatistics. New York, Springer, 228p.

DOLL, J. C. & Lauer, N. T. E. 2013. Bayesian estimation of age and length at 50% maturity. *T. A. Fish. Soc.*, 142: 1012-1024.

DORMANN, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global. Ecol. Biogeogr.*, 16: 129-138.

DONCEL, O. & PARAMO, J. 2010. Hábitos alimenticios del pargo rayado, *Lutjanus synagris* (Perciformes: Lutjanidae), en la zona norte del Caribe colombiano. *Lat. Am. J. Aquat. Res.*, 38: 413-426.

DUARTE, L. O. & GARCIA, C. B. 1999. Diet of the Lane Snapper, *Lutjanus synagris* (Lutjanidae), in the Gulf of Salamanca, Colombia. *Caribb. J. Sci.*, 35: 54-69.

~ E ~

ELITH, J.; GRAHAM, C. J.; ANDERSON, R. P.; DUDÍK, M.; FERRIER, S.; GUISAN, A.; HIJMANS, R. J.; HUTTNAM, F.; LEATHWICK, J. R.; LEHRNMANN A.; LI, J.; LOHRNANN, L. G.; LOISELLE, B. A.; MANION, G.; MORITZ, C.; NAKAMURA, M.; NAKATAWA, Y.; OVERTON, J. M.; PETERSON, A. T.; PHILLIPS, S. J.; RICHARDSON, K.; SCACHETTI-PEREIRA, R.; SCHAPIRE, R. E.; SOBERÓN, J.; WILLIAMS, S.; WISZ, M. S. & ZIMMERMANN, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

ELITH, J. & LEATHWICK, J. R. 2009. Species distribution models: ecological and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.*, 40: 677-977.

ELLISON, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecol. Appl.*, 6: 1036-1046.

ELLISON, A. M. 2004. Bayesian inference in ecology. *Ecol. Let.*, 7: 509-520.

~ F ~

FAO. 2003. The ecosystem approach to fisheries. FAO Technical guidelines for responsible fisheries, 4.

FAO. 2013. FISHTAT PLUS 2.3: Fishery Information, Data and Statistics, time series from aquaculture (quantities and values) captures (quantities); World Wide Web Eletronic Publications; www.fao.org/fi/statist/FISOFT/FISHPLUS.asp.

FEITOZA, B. M.; ROSA, R. S. & ROCHA, L. A. 2005. Ecology and zoogeography of deep-reef fishes in northeastern Brazil. *Bul. Mar. Sci.*, 76: 725-742.

FIGUEROLA, M.; MATOS-CARABALLO, D.; TORRES, W.; CRESWELL, R. L., 1998. Maturation and reproductive seasonality of four reef fish species in Puerto Rico. *Proc. Gulf. Carib. Fish. Inst.* 50: 938-968.

FONSECA, J. F. 2009. Estudo da dieta do *Lutjanus synagris* (Linnaeus, 1758) e *Ocyurus chrysurus* (Block, 1791), Teleostei: Perciformes: Lutjanidae, no banco dos Abrolhos, Bahia, Brasil e pesca das principais espécies de lutjanídeos e serranídeos

na região. Dissertação de Mestrado, Universidade Estadual de São Paulo, Brasil, 146 p.

FONTELES-FILHO, A. A. 1989. Recursos pesqueiros: Biologia e dinâmica populacional. Fortaleza, Imprensa Oficial do Ceará, 296p.

FRANÇA, A. M. C.; COUTINHO, P. N. & SUMMERHAYES, C. P. 1976. Sedimentos superficiais da margem continental nordeste brasileira. *R. Bras. Geo.*, 6: 147-164.

FRANKLIN, J. 2010. Mapping species distributions: spatial inference and prediction. New York, Cambridge University Press, 336p.

FRANKS, J. S. & VANDERKOOY, K. E. 2000. Feeding habits of juvenile lane snapper *Lutjanus synagris* from Mississippi coastal waters, with comments on the diet of gray snapper *Lutjanus griseus*. *Gulf. Carib. Res.*, 12: 11-17.

FREITAS, M. O.; MOURA, R. L.; FRANCINI-FILHO, R. B. & MINTE-VERA, C. V. 2011. Spawning patterns of commercially important reef fish (Lutjanidae and Serranidae) in the tropical western South Atlantic. *Sci. Mar.*, 75: 135-146.

FREITAS, M. O.; ROCHA, G. R.; CHAVES, P. T. C. & MOURA, R. L. 2014. Reproductive biology of the lane snapper, *Lutjanus synagris*, and recommendations for its management on the Abrolhos Shelf, Brazil. *J. Mar. Biol. Assoc. UK*, 1: 1-10.

~ G ~

GARCIA, J. Jr.; MENDES, L. F.; SAMPAIO, C. L. S. & LINS, J. E. 2010. Biodiversidade marinha da Bacia Potiguar: ictiofauna. Rio de Janeiro, Série Livros 38, 195p.

GARCIA, J. Jr.; Nóbrega, M. F. & Oliveira, J. E. L. 2015. Coastal fishes of Rio Grande do Norte, northeastern Brazil, with new records. *Check list* 11: 1-24.

GEISSER, S. 1993. Predictive inference. Chapman & Hall/CRC. 240p.

GELMAN, A.; 1996. Inference and monitoring convergence, in: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. (Eds), Markov chain monte carlo in practice. Chapman and Hall/CRC, Florida, pp. 131-143.

GESTEIRA, T. C. V. & ROCHA, C. A. S. 1976. Estudo sobre a fecundidade do ariocó, *Lutjanus synagris* (Linnaeus), da costa do estado do Ceará (Brasil). *Arg. Ciên. Mar.*, 16: 19-22..

GIANNINI, T. C.; SIQUEIRA, M. F.; ACOSTA, A. L. & BARRETO, F. C. C. 2012. Desafios atuais da modelagem preditiva de distribuição de espécies. *Rodriguésia* 63: 733-749.

GÓMEZ, G.; GUZMÁN, R. & CHACÓN, R. 2001. Parámetros reproductivos y poblacionales de *Lutjanus synagris* en el Golfo de Paria, Venezuela. *Zootec. Trop.*, 19: 335-357.

GUISAN, A. & THUILLER, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.*, 8: 993-1009.

GUISAN, A. & ZIMMERMANN, N. E. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.*, 135: 147-186.

~ H ~

HADDON, M. 2001. Modelling and quantitative methods in fisheries, first ed. Chapman & Hall/CRC, New York, 424 p.

HAIMOVICI, M.; CERGOLE, M. C.; LESSA, R. P.; MADUREIRA, L. S. P.; JABLONSKI, S. & ROSSI-WONGTSCHOWSKI, C. L. D. B. 2006. Panorama nacional. In: MMA (ed.). Programa REVIZEE: avaliação do potencial sustentável de recursos vivos na zona econômica exclusiva do Brasil: relatório executivo. MMA, Brasília, cap. 2: 77-120.

HAIMOVICI, M. 2007. A prospecção pesqueira e abundância de estoques marinhos no Brasil nas décadas de 1960 a 1990: levantamento de dados e avaliação crítica. MMA, Brasília, 330 p.

HALL, B. Bayesian inference. 2011. <<http://www.idg.pl/mirrors/CRAN/web/packages/LaplacesDemon/vignettes/BayesianInference.pdf>> acesso em novembro de 2015.

HELD, L.; SCHRÖDLE, B. & RUE, H. 2010. Posterior and cross-validators predictive checks: a comparison on MCMC and INLA. *In* Statistical modelling and regression structures, pp. 111-131.

HIJMANS, R. J. 2015. Raster: geographic data analysis and modeling. R package version 2.3-40, <http://CRAN.R-project.org/package=raster>.

HILBORN, R.; PIKITCH, E. K. & FRANCIS, R. C. 1993. Current trends in including risk and uncertainty in stock assessment and harvest decision. *Can. J. Fish. Aquat. Sci.*, 50: 874-880.

HILBORN, R. & MANGEL, M. 1997. The ecological detective: confronting models with data. New Jersey, Princeton University Press, 315p.

HILBORN, R.; BRANCH, T. A.; ERNST, B.; MAGNUSSON, A.; MINTE-VERA, C. V.; SCHEURELL, M. D. & VALERO, J. L. 2003. State of the world's fisheries. *Annu. Rev. Envir. and Res.*, 28: 359-399.

HONEY, K. T.; MOXLEY, J. H. & FUJITA, R. M. 2010. From rags to fishes: Data-poor methods for fishery managers. *In*: STARR, R. M.; CULVER, C. S.; POMEROY, C.; MCMILLAN, S.; BARNES, T. & ASTELINE-NEILSON, D (ed.). Managing Data-poor fisheries: case studies, model and solutions. San Diego, California Sea Grant College Program, 1: 155-180.

~ I ~

IBAMA, 2007a. Estatística da Pesca 2005 Brasil: grandes regiões e unidades da federação. Brasília, 108p.

IBAMA, 2007b. Estatística da Pesca 2007 Brasil: grandes regiões e unidades da federação. Brasília, 113p.

IBAMA, 2008. Estatística da Pesca 2006 Brasil: grandes regiões e unidades da federação. Brasília, 180p.

IDEMA. 2002. Secretaria de Estado do Planejamento e das Finanças do Rio Grande do Norte. Perfil do Estado do Rio Grande do Norte, Natal, 85p.

ILLIAN, J. B; MARTINO, S ; SORBYE, S. H; GALLEGU-FERNÁNDEZ, J. B; ZUNZUNEGUI, M; ESQUIVIAS, M. P & TRAVIS, M. J. 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. *Meth. Ecol. Evol.* 4: 305-315.

ISAAC, V. J.; MARTINS, A. S.; HAIMOVICI, M. & ANDRIGUETTO, J. M. 2006. A pesca marinha e estuarina do Brasil no início do século XXI: recursos, tecnologias, aspectos socioeconômicos e institucionais. Editora Universitária UFPA, Belém, 188 p.

~ J ~

JENNESS, J. S., 2004. Calculating landscape surface area from digital elevation models. *Wild. Soc. Bull.* 32: 829-839.

JENNINGS, S.; KAISER, M. J. & REYNOLDS, J. D. 2001. Marine fisheries ecology. Blackwell Publishing, Oxford, 417 p.

JAYNES, E. T. 2003. Probability Theory – The Logic of Science. Cambridge, Cambridge University Press, 727p.

JACKSON, J. B. C.; KIRBY, M. X.; BERGER, W. H.; BJORN DAL, K. A.; BOTSFORD, L. W.; BOURQUE, B. J.; BRADBURY, R. H.; COOKE, R.;

ERLANDSON, J.; ESTES, J. A.; HUGHES, T. P.; KIDWELL, S.; LANGE, C. B.; LENIHAN, H. S.; PANDOLFI, J. M.; PETERSON, C. H.; STENECK, R. S.; TEGNER, M. J. & WARNER, R. R. 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science*, 293: 629-638.

~ K ~

KARNA, S. K. & PANDA, S. 2011. Growth estimation and length at maturity of a commercially importante fish species i. e., *Daysciaena albida* (Boroga) in Chilika Lagoon. *India. Euro. J. Exp. Bio.*, 1: 84-91.

KATARA, I.; ILLIAN, J.; PIERCE, G. J.; SCOTT, B. & WANG, J., 2008. Atmospheric forcing on chlorophyll concentration in the Mediterranean. *Hydrobiologia*. 612: 33-48.

KÉRY, M. 2010. Intoduction to WinBUGS for ecologists: a bayesian approach to regression, ANOVA, mixed models and related analyses. Burlington, Academic Press, 302p.

KINAS, P. G. & ANDRADE, H. A. 2007. Bayesian statistics for fishery stock assessment and management: a synthesis. *PANAMJAS*, 2: 103-112.

KINAS, P. G. & ANDRADE, H. A. 2010. Introdução à Análise Bayesiana (com R). Porto Alegre, maisQnada, 240p.

KING, M. 2007. Fisheries biology: assessment and management. Queensland, Blackwell Publishing, 382p.

KNEIB, T.; MÜLLER, J. & HOTHORN, T. 2008. Spatial smoothing techniques for the assessment of habitat suitability. *Environ. Ecol. Stat.*, 15: 343-364.

KRAINSKI, E. T & LINDGREN, F. 2014. The R-INLA tutorial: SPDE models. < <http://www.math.ntnu.no/inla/r-inla.org/tutorials/spde/spde-tutorial.pdf> > acesso em março de 2015.

KRUSCHKE, J. K. 2014. Doing Bayesian data analysis: a tutorial with R, JAGS and Stand. London, Elsevier, 759p.

~ L ~

LALLI, C. M. & PARSONS, T. R. 1997. Biological oceanography: an introduction. Vancouver, Elsevier, 314p.

LAMIGUEIRO, O. P. & HIJMANS, R. 2014. RasterVis. R package version 0.31, doi: 10.5281/zenodo.12394.

LATIMER, A. M; WU, S; GELFAND, A. E & SILANDER, J. A. J. 2006. Building statistical models to analyze species distributions. *Ecol. Appl.*, 16: 33-50.

LEMOS, R. S. T. 2010. Hierarchical bayesian methods for marine sciences: analysis of climate variability and fish abundance. Tese de Doutorado, Universidade de Lisboa, Portugal, 182p.

LESSA, R.; BEZERRA, Jr. J. L. & NÓBREGA, M. F. 2009a. Dinâmica das frotas pesqueiras da região nordeste do Brasil. Fortaleza, Martins & Cordeiro Ltda, 164p.

LESSA, R.; BEZERRA, Jr. J. L. & NÓBREGA, M. F. 2009b. Dinâmica de populações e avaliação dos estoques dos recursos pesqueiros da região Nordeste. Fortaleza, Martins & Cordeiro Ltda, 304p.

LIMA, W. B. 2004a. Idade e crescimento do ariocó *Lutjanus synagris* Linnaeus, 1758 (Teleostei: Lutjanidae) da costa Norte da Bahia- Brasil. Trabalho de Conclusão de Curso, Universidade Federal do Espírito Santo, Brasil, 39p.

LINDEMAN, K.C . ; DIAZ, G. A.; SERAFY, J. E. & AULT, J. S. 1998. A spatial framework for assessing cross shelf habitat use among newly settled grunts and snappers. *Proc. Gulf Carib. Fish. Inst.*, 50: 385-416.

LINDGREN, F.; RUE, H. & LINDSTRÖM, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion). *J. R. Stat. Soc. Ser. B.*, 73: 423-498.

LINDGREN, F. ; RUE, H. 2013. Bayesian spatial and spatio-temporal modelling with R-INLA. *J. Stat. Sof.*, 1-26

LUCKHURST, B. E.; DEAN, M. J. & REICHERT, M. 2000. Age, growth and reproduction of the lane snapper *Lutjanus synagris* (Pisces: Lutjanidae) at Bermuda. *Mar. Ecol. Prog. Ser.*, 203: 255-261.

LUDWIG, D. ; HILBORN, R. & WALTERS, C. 1993. Uncertainty, resource exploitation, and conservation: lessons from history. *Science*, 260: 17-36

LUMPKIN, R. & GARZOLI, S. L. 2005. Near-surface circulation in the Tropical Atlantic Ocean. *Deep-Sea Res. pt 1*, 52: 495-518.

~ M ~

MANICKCHAND-DASS, S. 1980. Reproduction, age and growth of lane snapper, *Lutjanus synagris* (Linnaeus), in Trinidad, West Indies. *Bull. Mari. Sci.*, 40: 411- 429.

MANOOCH, C. S. & MASON, D. 1984. Age, growth and mortality of lane snapper from Southern Florida. *Northeast Gulf. Sci.*, 7: 109-115.

MÄNTYNIEMI, S. 2006. Bayesian fisheries stock assessment: integrating and updating knowledge. Dissertação de Mestrado, Universidade de Helsinki, Finlândia, 24p.

MARASCO, R. J.; GOODMAN, D.; GRIMES, C. B.; LAWSON, P. W.; PUNT, A. E. & QUINN, T. J. I. I. 2007. Ecosystem-based fisheries management: some practical suggestions. *Can. J. Fish. Aquat. Sci.*, 64: 928-939.

MATHERON, G. 1963. Principles of geostatistics. *Econ. Geol.*, 58: 1246-1266.

MCCARTHY, M. A. 2007. Bayesian methods for Ecology. Cambridge, Cambridge University Press, 306p.

MCEACHRAN, J. & FECHHELM, J. 2005. Fishes of the Gulf of Mexico Vol. 2. Texas, University of Texas Press, 1004p.

MCLEAN, R. R. F. & TSYBAN, A. 2001. Coastal zones and marine ecosystems. In: MCCARTHY, J. J.; CANZIANI, O. F.; LEARY, N. A.; DOKKEN, D. J. & WHITE, K. S. (ed.). Climate change 2001: impacts, adaptation, and vulnerability. Cambridge University Press, Chap 6: 343-379.

MIKULAS, J. J. & ROOKER, J. R. 2008. Habitat use, growth, and mortality of post-settlement lane snapper (*Lutjanus synagris*) on natural banks in the northwestern Gulf of Mexico. *Fish. Res.*, 93: 77-84.

MILLER, J. R.; TURNER, M. G.; SMITHWICK, E. A. H.; DENT, C. L. & STANLEY, E. H. 2004. Spatial extrapolation: the science of predicting ecological patterns and processes. *BioScience*, 54: 310-320.

MINISTÉRIO DO MEIO AMBIENTE (MMA). 2006. Programa REVIZEE: avaliação do potencial sustentável de recursos vivos na zona econômica: relatório executivo. Brasília, MMA, 280p.

MUÑOZ F.; PENNINO, M. G.; CONSEA, D.; LÓPEZ-QUÍLEZ, A. & BELLIDO, J. M. 2013. Estimating and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. *Stoch. Environ. Res. Risk. Assess.*, 27: 1171-1180.

MYERS, R. A.; WORM, B. 2003. Rapid worldwide depletion of predatory fish communities. *Nature*, 423: 280-283.

MYERS, R. A.; BAUM, J. K.; SHEPHERD, T. D.; POWERS, S. P. & PETERSON, C. H. 2007. Cascading effects of the loss of apex predatory sharks from a coastal ocean. *Science*, 315: 1846-1850.

~ N ~

NIMER, E. 1989. Climatologia do Brasil. Rio de Janeiro, IBGE, 422p.

NISHIDA, T. & CHEN, D. G. 2004. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data. *Fish. Res.*, 70: 265-274.

NÓBREGA, M. F. 2008. Modelagem espacial e temporal para avaliação de estoques de *Ocyurus chrisurus* e *Coryphaena hippurus* da região nordeste do Brasil. Tese de Doutorado, Universidade Federal do Rio Grande, Brasil, 206p.

NYCHKA, D.; FURRER, R. & SAIN, S. 2015. Fields: tools for spatial data. R package version 8.2-1, <http://CRAN.R-project.org/package=fields>.

~ O ~

OLIVEIRA, M. R.; COSTA, E. F. S.; ARAÚJO, A. S.; PESSOA, E. K. R.; CARVALHO, M. M.; CAVALCANTE, L. F. M. & CHELLAPPA, S. 2012. Sex ratio and length-weight relationship for five marine fish species from Brazil. *J. Mar. Biol. & Oc.*, 1: 1-3.

~ P ~

PACIOREK, C. J. & SCHERVISH, M. J. 2006. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17: 483-506.

PARADINAS, I.; CONSEA, D.; PENNINO, M. G.; MUÑOZ, F.; FERNÁNDEZ, A. M.; LÓPEZ-QUÍLEZ, A. & BELLIDO, J. M. 2015. Bayesian spatio-temporal approach to identifying fish nurseries by validating persistence areas. *Mar. Ecol. Prog. Ser.*, 528: 245-255.

PARENT, É. & RIVOT, É. 2012. Introduction to hierarchical bayesian modeling for ecological data. North Carolina, CRC Press, 402p.

PEDUZZI, P.; CONTACTO, J.; KEMPER, E.; HOLFORD, T. R. & FEINSTEIN, A. R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *J. Cli. Epi.*, 49: 1373-1379.

PENNINO, M. G. 2013a. Implementing ecosystem approach to fisheries management: advances and new tools. Tese de Doutorado, Universidade de Valência, Espanha, 201p.

PENNINO, M. G.; MUÑOZ, F.; CONSEA, D.; LÓPEZ-QUÍLEZ, A. & BELLIDO, J. M. 2013b. Modeling sensitive elasmobranch habitats. *J. Sea Res.*, 83: 209-218.

PENNINO, M. G; MUÑOZ, F; CONSEA, D; LÓPEZ-QUÍLEZ, A & BELLIDO, J. M. 2014. Bayesian spatio-temporal discard model in a demersal trawl fishery. *J. Sea Res.*, 90: 44-53.

PETERS, D. C. P.; HERRICK, J. E.; URBAN, D. L.; GARDNER, R. H. & BRESHEARS, D. D. 2004. Strategies for ecological extrapolation. *Oikos*, 106: 627-636.

PILLING, G. M.; APOSTOLAKI, P.; FAILLER, P.; FLOROS, C.; LARGE, P. A.; MORALES-NIN, B.; REGLERO, P.; STERGIO, K. I. & TSIKLIRAS, A. C. 2008. Assessment and management of data-poor fisheries. *In*: PAYNE, A.; COTTER, J. & POTTER, T. (ed.). *Advances in fisheries science: 50 years on from Beverton and Holt*. CEFAS, Blackwell Publishing, pp: 280-305.

PIMENTEL, C. R. & JOYEUX, J.-C. 2010. Diet and food partitioning between juveniles of mutton *Lutjanus analis*, dog *Lutjanus jocu* and lane *Lutjanus synagris* snappers (Perciformes : Lutjanidae) in a mangrove-fringed estuarine environment. *J. Fish Biol.*, 76 : 2299-2317.

PLANT, R. E. 2012. *Spatial data analysis in ecology and agriculture using R*. Florida, CRC Press, 648p.

PIKITCH, E. K.; SANTORA, C.; BABCOCK, E. A.; BAKUN, A.; BONFIL, R.; CONOVER, D. O.; DAYTON, P.; DOUKAKIS, P.; FLUHARTY, D.; HENEMAN, B.; HOUDE, E. D.; LINK, J.; LIVINGSTON, P. A.; MANGEL, M.; MCALLISTER, M. K.; POPE, J. & SAINSBURY, K. J. 2004. Ecosystem-based fishery management. *Science*, 305: 346-347

PINET, P. R. 2008. *Invitation to oceanography*. Massachusetts, Jones and Bartlett Publishers, 626p.

~ Q ~

QUIROZ, Z. C.; PRATES, M. O. & RUE, H. 2015. A Bayesian approach to estimate the biomass of anchovies off the coast of Perú. *Biometrics*, 71: 208-217.

~ R ~

R DEVELOPMENT CORE TEAM. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, <http://R-project.org/>

REZENDE, S. M.; FERREIRA, B. P. & FRÉDOU, T. 2003. A pesca de lutjanídeos no nordeste do Brasil: Histórico das pescarias, características das espécies e relevância para o manejo. *Bol. Téc. Ci. do CEPENE*, 11: 257-270.

RIVERA-ARRIAGA, E.; LARA-DOMINGUEZ, A. L.; RAMOS-MIRANDA, J.; SANCHEZ-GIL, P. & YANEZ-ARANCIBIA, A. 1996. Ecology and population dynamics of *Lutjanus synagris* on Campeche Bank. In: ARREGUÍN-SÁNCHEZ, F.; MUNRO, J. L.; BALGOS, M. C. & PAULY, D. (ed). Biology, fisheries and culture of tropical groupers and snappers. Mexico, EPOMEX, Chap. 2: 11-18.

RODRÍGUEZ, A. & PÁRAMO, J. 2012. Distribución espacial del pargo rayado *Lutjanus synagris* (Pisces: Lutjanidae) y su relación con las variables ambientales en el caribe Colombiano. *Actual. Biol.*, 34: 55-66.

ROOS, M. & HELD, L. 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis*, 6: 259-278.

ROOS, N. C.; CARVALHO, A. R.; LOPES, F. M. & PENNINO, M. G. 2015. Modeling sensitive parrotfish (Labridae: Scarini) habitats along the Brazilian coast. *Mar. Env. Res.*, 110 : 92-100.

ROSS B. E.; HOOTEN, M. B. & KOONS, D. N. 2012. An accessible method for implementing hierarchical models with spatio-temporal abundance data. *PLoS ONE* 7(11): e49395.

ROSSI-WONGTSCHOWSKI, C. L. D. B.; VALENTIN, J. L.; JABLONSKI, S.; AMARAL, A. C. Z.; HAZIN, F. H. V. & EL-ROBRINI, M. 2006. O ambiente marinho. In: MMA (ed.). Programa REVIZEE: avaliação do potencial sustentável de recursos vivos na zona econômica exclusiva do Brasil: relatório executivo. Brasília, MMA, chap. 1: 21-75.

RUE, H. & HELD, L. 2005. Gaussian Markov random fields. Theory and applications. Florida, Chapman & Hall. 280p.

RUE, H. & MARTINO, S. & CHOPIN, N. 2009. Approximate bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Statist. Soc. B.*, 71: 319-392.

~ S ~

SBROCCO, E. J. & BARBER, P. H. 2013. MARSPEC: Ocean climate layers for marine spatial ecology. *Ecology* 94: 979.

SCHNUTE, J. T. 1987. Data, uncertainty, model ambiguity, and model identification. *Nat. Resour. Mod.*, 2: 159-212.

SIMPSON, D.; LINDGREN, F. & RUE, H.. 2012a. Think continuous: Markovian Gaussian models in spatial statistics. *Sp. Stat.*, 1: 16-29.

SIMPSON, D.; LINDGREN, F. & RUE, H. 2012b. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, 23: 65-74.

SOUSA-JÚNIOR, V. B.; SILVA, J. R. F. & SALLES, R. 2008. Análise ovariana do ariacó, *Lutjanus synagris* (Actinopterygii: Lutjanidae), e considerações sobre sua reprodução no estado do Ceará. *Arq. Ciên. Mar.*, 41: 90-97.

SPIEGELHALTER, D. J; BEST, N. J; CARLIN, B. P & Van Der Linde, A. 2002. Bayesian measure of model complexity and fit. *J. Roy. Stat. Soc.*, 64: 583-639.

STRAMMA, L.; IKEDA, Y. & PETERSON, R. G. 1990. Geostrophic transport in the Brazil Current region north of 20° S. *Deep-Sea Res.*, 37: 1875-1886.

SU, Y. S. & YAJIMA, M. 2014. R2jags: A package for Running jags from R. R package version 0.04-02, <http://CRAN.R-project.org/package=R2jags>.

~ T ~

TABOSA, W. F. 2006a. Morfologia, hidrodinâmica e sedimentologia da plataforma continental brasileira adjacente a São Bento do Norte e Caiçara do Norte – RN/NE – Brasil. Tese de Doutorado, Universidade Federal do Rio Grande do Norte, Brasil, 159p.

TABOSA, W. F. 2006b. Monitoramento costeiro das praias de São Bento do Norte e Caiçara do Norte-RN: Implicações para o polo petrolífero de Guamaré. Dissertação de Mestrado em Geodinâmica e Geofísica, Universidade Federal do Rio Grande do Norte, Brasil, 112p.

TOBLER, W. F. 1970. A computer movie simulating urban growth in the Detroit region. *Econ. Geo.*, 46: 234-240.

TORRES-LARA, R. & CHÁVEZ, E. 1987. Evaluación y diagnóstico de la pesquería de la rubia (*Lutjanus synagris*) en el Estado de Yucatán. *Cienc. Mar.*, 13: 7- 29.

TORRES-LARA, R. T. & SALAS-MÁRQUEZ, S. 1990. Crecimiento y mortalidad de la rubia *Lutjanus synagris* de las costas de Yucatán durante las temporadas de pesca 1983-1985. *An. Inst. Cienc. Mar y Limnol. Univ. Nal. Autón. México*, 17: 205-214.

TRAVASSOS, P.; HAZIN, F. H. V.; ZAGAGLIA, J.; ADVÍNCULA, R.; SCHÖBER, J. 1999. Thermohaline structure around seamounts and islands off North-Eastern Brazil. *Arch. Fish. Mar. Res.*, 3: 211-222.

~ W ~

WALTERS, C. 1986. Adaptive management of renewable resources. New York, MacMillan Publishing Company, 374 p.

WIKLE, C. K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84: 1382-1394.

WORM, B.; LOTZE, H. K.; HILLEBRAND, H. & SOMMER, U. 2002. Consumer versus resource control of species diversity and ecosystem functioning. *Nature*, 417: 848-851.

~ V ~

VALAVANIS, V. D. 2002. Geographic information system in oceanography and fisheries. London, CRC Press, 209p.

VALAVANIS, V. D. 2008. Essential fish habitat mapping in the mediterranean. Creta, Hydrobiologia, 300p.

VALAVANIS, V. D.; PIERCE, G. J.; ZUUR, A. F.; PALIALEXIS, A.; SAVELIEV, A.; KATARA, I. & WANG, J. 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia*, 612: 5-20.

VAZZOLER, A. 1996. Biologia da reprodução de peixes Teleósteos: teoria e prática. Ceará, EDUEM, 169p.

VENABLES, W. N & DICHMONT, C. M. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fish. Res.*, 70: 319-337.

VITAL, H. 2005. Erosão e progradação do litoral do Rio Grande do Norte. In: MUEHE, D (ed). Erosão e progradação do litoral brasileiro. Brasília, MMA, chap. 5: 156-172.

VITAL, H.; GOMES, M. P.; TABOSA, W. F.; FARZÃO, E. P.; SANTOS, C. L. A. S. & JÚNIOR, J. S. P. 2009. Characterization of the brazilian continental shelf adjacent to Rio Grande do Norte state, NE, Brazil. *Braz. J. Oc.*, 58: 43-54.

~ Z ~

ZUUR, A. F.; IENO, E. N. & ELPHICK, C. S. 2010. A protocol for data exploration to avoid common statistical problems. *Meth. Ecol. Evol.*, 1: 3-14.

ANEXO I

CÓDIGOS

```
#####
#
#                               Modelagem de EFHs via INLA/SPDE
#
#####

### Marie-Christine Rufener
### Data: 10/2015
### machris55@hotmail.com

rm(list = ls(all = TRUE))

#Carregando bibliotecas
library(raster)
library(rasterVis)
library(mapttools)
library(INLA)
library(gridExtra)
library(lattice)
library(fields)

# Carregar os dados
lutjanus<- read.csv("C:/Users/Marie Cristine/Dropbox/Mestrado/Planilhas e scripts/Scripts
INLA/Lutjanus/lutjanus corrigido .csv", sep=";")
str(lutjanus)
attach(lutjanus)
names(lutjanus)

# Criar matriz de coordenadas
coords<-as.matrix(lutjanus[,11:10]) # coluna de longitude antecede SEMPRE a latitude!

#####
#           Carregar imagens satelitais (rasters) de preditores ambientais
#####

# Observação: atentar para que os rasters estejam sempre na mesma resolução

# Sea Surface Temperature (SST)
sst<-raster("../temperatura.tif") # indicar diretório
res(sst)

# Especificar a área de estudo
ext <- extent(-37.25,-34.8,-6.5,-4.75)

# Extrair os valores de temperatura para os pontos amostrados
temperatura<-extract(sst,coords)
temperatura

# Repetir os últimos passos para o n° desejável de preditores ambientais

#-----
```

```

#~~~~~
## Passo à passo da construção, especificação e execução do modelo espacial
#~~~~~

#####
## Passo 1: Construção do mesh ###
#####

## Aqui o mesh foi construído tomando por base o polígono da costa. Os próximos passos
## serão referentes a importação do polígono e delimitação da zona marítima com base
## nesse polígono

# Importando o shapefile da costa
br.sp<-readShapePoly("../costa")

# Criar um polígono (retângulo) em torno do shapefile
box.sel <- matrix(c(-37.25,-34.8,-34.8,-37.25,-37.25,
-6.5,-6.5,-4.75,-4.75,-6.5), ncol=2) #especificar lat/long da área
sel.box <- SpatialPolygons(list(Polygons(list(Polygon(
box.sel, FALSE)), '0')), proj4string=br.sp@proj4string)

par(mar=c(3,2,1,1))
plot(sel.box)
plot(br.sp, add=TRUE)
for (i in 1:length(br.sp@polygons[[1]]@Polygons)) {
  pi <- br.sp@polygons[[1]]@Polygons[[i]]
  coo.sel <- seq(1,nrow(pi@coords),300)
  text(pi@coords[coo.sel, 1], pi@coords[coo.sel, 2], paste(coo.sel))
  text(pi@labpt[1], pi@labpt[2], paste(i))
}

### Definir o polígono completo com base no polígono da costa. Para isso todo o polígono
### da costa será identificado por meio de números e depois uma linha será percorrida por
### todo o polígono com o intuito de demarcar a sua exclusão do polígono geral criado
### anteriormente
coo1 <- br.sp@polygons[[1]]@Polygons[[3]]@coords
pl1 <- rbind(c(-37.25,-6.5), c(coo1[5180,1:2]), # Inserir as long/lat mínimas...e especificar
coo1[5180:890, ], c(-37.25,-6.5)) # a amplitude dos números no polígono

lines(pl1, col=4,lwd=1)

bound.pol <- SpatialPolygons(list(Polygons(
list(Polygon(box.sel, FALSE),
Polygon(pl1, TRUE)), '0')))
plot(bound.pol, col=gray(.9))

bound.seg <- inla.sp2segment(bound.pol)

# Criar a malha triangular (mesh) do modelo espacial. Atentar para os argumentos extras
# presentes na função inla.mesh.2d
?inla.mesh.2d

(mesh <- inla.mesh.2d(boundary=bound.seg, max.edge=0.064,cutoff=0.0125,loc=coords))$n

plot(mesh,main="")
plot(br.sp,add=T,col="grey70")
points(coords,pch=20,col=2)

```

```

# Identificar as coordenadas de todos os vértices dos triângulos da malha - necessário
# para definir o pontos para predição
mesh.points<- SpatialPoints(
  coords = cbind(mesh$loc[,1], mesh$loc[,2]),
  proj4string = CRS("+proj=longlat +datum=WGS84 +no_defs") # definir a projeção correta
)
points(mesh.points)

#####
##          Passo 2: Criar a matriz de projeção          ##
#####

# O modelo SPDE é definido no mesh; Precisamos especificar o preditor linear para a variável
# resposta. O ponto chave é que temos um campo aleatório que é modelado em cima dos
# vértices e a variável resposta em n localizações. Portanto, é necessário definir como o
# campo aleatório e os outros componentes do modelo serão conectados à variável resposta.
# No módulo INLA-SPDE esse passo é realizado mediante a matriz de projeção A.
# Brevemente, a matriz A localiza os pontos no mesh. A função de correlação espacial
# também é definida no mesh. Teremos, portanto, (mesh$n) pontos. Os valores nesses
# pontos constituem o campo espacial aleatório. O problema, contudo, é que temos um
# menor n° de pontos amostrais do que o número de vértices na malha. Logo, é necessário
# manter o controle de quais elementos do campo espacial aleatório e quais dos dados
# observados (bem como das covariáveis) são conectados entre si, função esta desempenhada
# pela matriz A.

# Construção da matriz de estimação
A.est<- inla.spde.make.A(mesh, loc=coords)

# Construção da matriz de predição
A.pred = inla.spde.make.A(mesh, loc=matrix(c(mesh$loc[,1],mesh$loc[,2]),ncol=2))

dim(A.est)
dim(A.pred)

#####
##          Passo 3: Criar o modelo SPDE com função de correlação Matérn          ##
#####

spde <- inla.spde2.matern(mesh, alpha=2)
#alpha=parâmetro de suavização da matriz de correlação

#####
##          Passo 4: Gerar a matriz de estimação e predição          ##
#####

#~~ Estimação

# Vamos definir um data frame contendo um intercepto e as covariáveis que serão usadas no
modelo

# Montando o data frame da estimação
N<-NROW(cpue) #indicar aqui a variável resposta
E<-data.frame(Intercept=rep(1,N),
  temp=temp padr) # Inserir quantas variáveis desejáveis. O nome a esquerda (temp)
# é utilizado como identificador, isto é, o modelo será montado com base nesse nome

# Montando o stack para as observações
stk.est<-inla.stack(data=list(y=cpue), # indicar a variável resposta no "y"

```

```

A=list(A.est, 1),
effects=list(spatial=1:spde$n.spde,
            E),
tag="est")
# A.est é o nome da matriz de estimação construída anteriormente.
# Em "effects" indicar o modelo SPDE e a matriz de estimação. o argumento "tag"
# serve para identificar se o stack é de estimação ou predição. Pode inserir o nome que
# preferir. Contudo após esse passo manter ele padronizado até o final do script.

#~~~~~

#~~ Predição

# Aqui vamos extrair os valores do(s) preditor(es) para cada ponto do mesh
Temperatura=extract(sst,mesh.points) # indicar o nome do raster do preditor (sst)

# Montando o data frame da predição
P<-data.frame(Intercept=1,
              TEMPERATURA=Temperatura) # Usar mesmo procedimento do data frame da
estimação

# Stack para predição; contém um conjunto de efeitos (variáveis), matrizes de predição
# e confere NA para a variável resposta
stk.pred<-inla.stack(data=list(y=rep(NA,spde$n.spde)),
                    A=list(A.pred,1),
                    effects=list(spatial=1:spde$n.spde,
                                P),
                    tag='pred')
# Repare que o stack é de predição, portanto identificar corretamente o tag (pred)

# Agora vamos juntar ambos os stacks (stack dos dados e da predição)
stk=inla.stack(stk.est, stk.pred)

#####
## Modelo
#####
# Aqui vamos especificar o modelo desejado. Podemos rodar a estimação juntamente ao
# processo de predição. Contudo, se quisermos, poderíamos fazer isso separadamente. Aqui
# fizemos os dois procedimentos ao mesmo tempo. Isto foi feito mediante a indicação do
# stack conjunto (stk, comando anterior) na linha: data=inla.stack.data. Caso se queira fazer
# separadamente, trocar apenas por stk.est/stk.pred

form<-log(y)~0+Intercept+temp+f(spatial,model=spde) # repare que os nomes no modelo
# são exatamente iguais aos especificado na construção do data frame da estimação
res<- inla(form1,
          family="gaussian",
          data=inla.stack.data(stk),
          control.predictor=list(A=inla.stack.A(stk), compute=TRUE), #indicar stack de
estimação
          control.compute = list(return.marginals=TRUE, dic=TRUE, cpo=TRUE),
          control.fixed = list(expand.factor.strategy='inla')) # este comando é utilizado caso
# se tenha algum fator no modelo

summary(res)
plot(res)
#~~~~~

```

```
#####
## Resultados ##
#####

# Para acessar a distribuição posterior marginal do GMRF nos locais específicos,
# é necessário extrair o índice do stack completo usando o tag adequado.
(inddp<-inla.stack.index(stk,tag="pred")$data)

# Sumário da distribuição posterior do GMRF
round(res$summary.linear.predictor[inddp,],4)

# Extrair os parâmetros do GMRF
result = inla.spde2.result(res, "spatial", spde)

# Valor médio de kappa
inla.emarginal(function(x)x,result$marginals.kappa[[1]])

# valor médio de tau
inla.emarginal(function(x)x,result$marginals.variance.nominal[[1]])

### rho (km) e intervalo de credibilidade a 95% (distância em que rho=0.1)
(rangecor<-c(mean=inla.emarginal(function(x) x, result$marginals.range[[1]]),
  q=inla.hpdmarginal(0.95, result$marginals.range[[1]]))c(1,2,3))

### média e desvio padrão do campo latente
inla.emarginal(function(x) x, result$marginals.range.nominal[[1]])
range.nom.marg = result$marginals.range.nominal[[1]]
range.nom.marg
range.nom.m1 = inla.emarginal(function(x) x, range.nom.marg)
range.nom.m1
range.nom.m2 = inla.emarginal(function(x) x^2, range.nom.marg)
range.nom.m2
range.nom.stdev = sqrt(range.nom.m2 - range.nom.m1^2)
range.nom.stdev
range.nom.quantiles = inla.qmarginal(c(0.025, 0.5, 0.975), range.nom.marg)
range.nom.quantiles

# Extrair a distribuição marginal e HPD a 95% da variável resposta
(ind2<-inla.stack.index(stk,"est")$data)
round(res$summary.fitted.values[ind2,])
mr<-res$marginals.fitted.values[ind2]
inla.hpdmarginal(0.95,res[[2]])

#~~~~~

#####
# Passo 5: Predição
#####

#####
## Projeção do efeito espacial/preditor linear em um grid ##
#####

pgrid<-inla.mesh.projector(mesh,dims=c(600,450))

prd.m<-inla.mesh.project(pgrid,res$summary.ran$spat$mean) # média do GMRF
prd.s<-inla.mesh.project(pgrid,res$summary.ran$spat$s) # desvio padrão do GMRF

# Extrair o índice de predição
```



```

igr <- inla.stack.index(stk, "pred")$data

ms.m<-inla.mesh.project(pgrid,(res$summary.fitted.values[igr,1])) # média do preditor linear
sd.s<-inla.mesh.project(pgrid,(res$summary.fitted.values[igr,2])) # desv. padrão do preditor
linear

### Projetar média/D.P do GMRF/preditor linear
grid.arrange(levelplot(prd.m, col.regions=topo.colors(99),
  main="latent field mean", xlab=" ", ylab=" ",
  scales=list(draw=F)),
  levelplot(prd.s,col.regions=topo.colors(99),main="latent field SD",
  xlab=" ", ylab=" ", scales=list(draw=F)),
  levelplot(exp(ms.m),
  xlab="", ylab="",main="response mean",
  col.regions=tim.colors(99),scales=list(draw=F)),
  levelplot(sd.s,
  xlab=" ", ylab=" ", main="response SD",
  col.regions=topo.colors(99),scales=list(draw=F)),
  nrow=2)

#~~~~~

# Extra:
# Função que automatiza um modelo padrão com diferentes distribuições de probabilidades.
# À título de exemplo, vamos especificar modelos poisson, binomial negativa e suas
# versões zero-inflacionadas

fams <- c(paste(c(" ", rep('zeroinflated', 1)),
  'poisson', c(" ", 0), sep=""),
  paste(c(" ", rep('zeroinflated', 1)),
  'nbinomial', c(" ", 0), sep="))
names(fams) <- gsub('ial', " ", gsub('zeroinflated', " ", fams))
fams

# Fórmula
formal <- y ~ 0+Intercept+temp+f(spatial,model=spde)+offset(log(esf))

dics.a1 <- sapply(fams, function(fam) {
  cat(fam, '\n') ### só para acompanhar o progresso...
  modal<-inla(formal,
    family=fam,
    data=inla.stack.data(stk.est),
    control.predictor=list(A=inla.stack.A(stk.est), compute=TRUE),
    control.compute = list(return.marginals=TRUE, dic=TRUE, cpo=TRUE),
    control.fixed = list(expand.factor.strategy='inla'))
  c(modal$dic$dic, -mean(log(modal$cpo$cpo)))
})

```