



Universidade do Minho
Escola de Ciências

Daniela de Jesus Pereira da Silva

Abundância da *Raja clavata* na costa
continental portuguesa: otimização
do desenho das campanhas de investiga
ção



Universidade do Minho
Escola de Ciências

Daniela de Jesus Pereira da Silva

Abundância da *Raja clavata* na costa
continental portuguesa: otimização
do desenho das campanhas de investiga-
ção

Dissertação de Mestrado
Mestrado em Estatística

Trabalho efetuado sob a orientação do
Doutora Professora Raquel Menezes
Doutora Ivone Figueiredo

Direitos de autor e Condições de utilização do trabalho por terceiros

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos. Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada. Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Agradecimentos

Os agradecimentos são dirigidos a todos aqueles que contribuíram para a realização deste projeto, nomeadamente:

Agradecimento à Professora Doutora Raquel Menezes e à Doutora Ivone Figueiredo pela orientação, pelas propostas desafiantes e pela partilha de conhecimento prestados ao longo do projeto.

Agradecimento a todos os docentes e colegas que me acompanharam por facilitarem o processo de aprendizagem.

Agradecimento ao Instituto Português do Mar e da Atmosfera por disponibilizar os dados analisados neste trabalho, parceiro do projeto FCT “PREFERENTIAL; PTDC/MAT-STA/28243/2017”.

Agradecimento à minha família, ao Paulo e amigos pelo apoio incondicional, por acreditarem em mim, nas minhas capacidades e por me incentivarem a lutar pelos meus objetivos.

Este trabalho foi financiado pela Fundação para a Ciência e Tecnologia/MCTES, através dos fundos nacionais (OE), no âmbito do Programa de financiamento a projetos de IC&DT em todos os domínios científicos.

Declaração de Integridade

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração. Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho

Resumo

Título: Abundância da *Raja clavata* na costa continental portuguesa: otimização do desenho das campanhas de investigação.

A diminuição de recursos naturais e as consequências adversas no ecossistema são aspetos muito importantes. Este estudo debruça-se sobre a modelação espacial de recursos marinhos, nomeadamente das espécies *Raja clavata* e *Merluccius merluccius*. A *Raja clavata* e *Merluccius merluccius* são peixes que co-ocorrem ao longo da costa continental portuguesa.

Os dados analisados dizem respeito a coordenadas geo-referenciadas de estações de pesca e para cada sabe-se o número e peso totais de *Raja clavata* capturada por hora. Os dados são provenientes de cruzeiros demersais realizados pelo IPMA entre 2013 e 2016. Adicionalmente, para 2015 e 2016, tem-se informação do número de elementos de *Merluccius merluccius* capturados por hora.

Com o objetivo de estimar os indicadores de abundância e biomassa da *Raja clavata* foram consideradas várias abordagens geoestatísticas, mais concretamente a suavização de *kernel* espacial, *kriging* com tendência externa e modelos de zeros inacionados. Nas duas últimas, a abundância de *Merluccius merluccius* pequena (comprimento < 26 cm) e a profundidade foram utilizadas como covariáveis. No método *kriging* com tendência externa, o tipo de fundo foi também considerado como variável explicativa. Os resultados indicam que os índices de abundância e biomassa de *Raja clavata* decrescem com a profundidade e com o aumento do indicador da abundância de juvenis de *Merluccius merluccius*. Relativamente ao tipo de fundo a média dos índices de abundância e biomassa são superiores em tipos de fundo de lama e lodo com areia.

De forma a desenhar futuras campanhas de investigação, é implementado um método de amostragem com base nas estimativas dos índices de abundância de *Raja clavata* e de *Merluccius merluccius*, obtidas de campanhas anteriores. Esta abordagem é avaliada através da comparação dos valores observados com os valores estimados por: correlações de *spearman* e o estimador da média, decorrente da amostragem aleatória estratificada, e a respetiva variância.

Palavras-chave: Aplicações de Pesca; Métodos *Kriging*; Interação Espaço-Tempo; Modelos de Zeros Inacionados; Amostragem Preferencial.

Abstract

Title: Abundance of *Raja clavata* in portuguese continental coast: optimization of design of investigation campaigns.

The depletion of natural resources and the adverse consequences on the ecosystem are very important aspects. This study deals with the spatial modeling of marine resources, namely the *Raja clavata* and *Merluccius merluccius*. *Raja clavata* and *Merluccius merluccius* are marine fishes co-occurring along the Portuguese continental coast.

The data analyzed concern the georeferenced coordinates of the fishing locations and for each the total number of *Raja clavata* specimens and their weight at each fishing haul. The data are derived from IPMA demersal cruises and cover the period 2013 to 2016. Additionally, for 2015 and 2016, the total number of *Merluccius merluccius* and small *Merluccius merluccius* specimens caught at each fishing haul is also available.

With the aim of estimating the abundance and biomass indicator of *Raja clavata* based on IPMA surveys several geostatistical approaches were applied, namely Spatial Kernel Smoothing, Kriging with External Drift and a Zero-Inflated Model. In the last two the small *Merluccius merluccius* abundance (length<26 cm) and depth were taken in consideration as covariates. In Kriging with External Drift, the seabed type was also used as explanatory variable. Results obtained indicate that *Raja clavata* abundance and biomass indicators decrease with depth and with the increase of small *Merluccius merluccius* abundance indicator. With respect to the seabed type, the mean of *Raja clavata* abundance and biomass indicators are higher in muddy and muddy sand.

In order to design future campaigns of investigation, it is applied a sampling method based on the estimates of *Raja clavata* and *Merluccius merluccius* abundance indicators. The proposal survey is assessed by comparing the observed values and estimated values by spearman correlation and the mean estimator, from stratified random sampling, and its variance.

Keywords: Fishery Applications; Kriging Methods; Space-Time Interaction; Zero-Inflated Models; Preferential Sampling.

Conteúdo

1	Introdução	1
1.1	Contextualização	1
1.2	Problema e Objetivos	3
1.3	Estado da Arte	3
1.4	Estrutura da Tese	5
2	Enquadramento Teórico	7
2.1	Modelos lineares generalizados	7
2.1.1	Família exponencial	8
2.2	Testes de Hipóteses	9
2.2.1	Teste para a homogeneidade de variâncias	10
2.2.2	Teste de interação espaço-tempo	10
2.3	Métodos de análise geoestatística	11
2.3.1	Requisitos	12
2.3.2	Variograma empírico	14
2.3.3	Estimação	16
2.3.4	Predição espacial	18
2.4	Uma abordagem geoestatística Bayesiana	21
2.4.1	Modelos de zeros inflacionados	21
2.5	Amostragem espacial	23
2.5.1	Métodos de estimação: <i>design-based vs. model-based</i>	24
2.5.2	Esquemas de amostragem	25
3	Os dados das campanhas demersais de 2013 a 2016	26
3.1	<i>Raja clavata</i> - Análise exploratória	27
3.1.1	Indicador de abundância	27
3.1.2	Indicador de biomassa	30
3.1.3	Comprimento médio	31
3.2	<i>Merluccius merluccius</i> - Análise Exploratória	32
3.2.1	Indicador de abundância dos adultos de pescada	33
3.2.2	Indicador da abundância dos juvenis de pescada	34

3.3	Análise exploratória das covariáveis	36
3.3.1	Profundidade	37
3.3.2	Tipo de fundo	37
3.3.3	Conclusões	38
4	Análise da distribuição espacial da <i>Raja clavata</i>	39
4.1	Análise da estrutura espacial pelo variograma individual	40
4.1.1	Indicador de abundância de <i>Raja clavata</i>	40
4.1.2	Indicador de biomassa de <i>Raja clavata</i>	42
4.1.3	Indicador da presença de <i>Raja clavata</i>	43
4.2	Análise da estrutura espacial pelo variograma médio	44
4.3	Interação espaço-tempo	46
4.3.1	Conclusões	46
5	Estimação da abundância da <i>Raja clavata</i> e da <i>Merluccius merluccius</i>	48
5.1	Abordagem Não Paramétrica	49
5.2	<i>Kriging</i> com Tendência Externa	50
5.2.1	A <i>Raja clavata</i>	50
5.2.2	A <i>Merluccius merluccius</i>	55
5.3	Modelos de zeros inflacionados	58
5.3.1	A <i>Raja clavata</i>	59
5.3.2	A <i>Merluccius merluccius</i>	65
6	Proposta e análise de plano de amostragem	69
6.1	Introdução	69
6.2	Proposta de plano de amostragem	71
6.2.1	Limitações do método proposto	72
6.3	Estudo de simulação	73
6.3.1	Geração de réplicas	73
6.3.2	Critérios de avaliação	74
6.3.3	Análise dos resultados	75
7	Conclusões e Trabalho futuro	78
	Anexos	83
A	Limite das zonas e setores	84
B	Análise exploratória	85
C	Resultados da Análise da distribuição espacial de <i>Raja clavata</i>	90

Lista de Figuras

1.1	Ilustração da arte de arrasto de fundo.	2
3.1	Localizações das realizações por campanha e por tipo de estação.	27
3.2	Localizações de observações nulas e não-nulas.	27
3.3	Número de elementos capturados por hora de raia-lenga por campanha. . .	28
3.4	Peso dos indivíduos capturados de raia-lenga por campanha.	31
3.5	Comprimento médio de raia-lenga capturada por campanha.	32
3.6	Número de indivíduos adultos de pescada capturados por hora por campanha.	33
3.7	Número de indivíduos juvenis de pescada capturados por hora por campanha.	35
3.8	Mapa da profundidade em metros na região de estudo.	37
3.9	Mapa do tipo de fundo na costa continental portuguesa (dados obtidos pela EMODnet).	38
4.1	Variogramas amostrais para o indicador de abundância de raia-lenga em cada campanha.	41
4.2	Superfícies preditas da distribuição espacial do indicador de abundância de raia-lenga em cada campanha.	42
4.3	Superfícies preditas da distribuição espacial do indicador de biomassa de raia-lenga em cada campanha.	43
4.4	Superfícies preditas da distribuição espacial da ocorrência de raia-lenga em cada campanha.	44
4.5	Abundância de raia-lenga capturada por hora. No painel da esquerda, variogramas empíricos individuais (linhas a cinzento) e médio (linha a vermelho); no painel da direita, variogramas empírico e teórico médios.	45
4.6	Biomassa da raia-lenga capturada por hora. No painel da esquerda, variogramas empíricos individuais (linhas a cinzento) e médio (linha a vermelho); no painel da direita, variogramas empírico e teórico médios.	46
5.1	Superfícies preditas por abordagem NP das distribuições dos indicadores de abundância (painel da esquerda) e biomassa (painel central) de raia-lenga e do indicador de abundância de pescada	50

5.2	Variogramas empírico e teórico dos resíduos associados à estimação da tendência dos valores do indicador de abundância de raia-lenga.	52
5.3	Superfície predita por KED da distribuição espacial do indicador de abundância de raia-lenga	53
5.4	Superfície de desvios padrão das estimativas por KED do indicador de abundância de raia-lenga	53
5.5	Variogramas empírico e teórico dos erros associados à estimação da tendência do indicador de biomassa de raia-lenga.	54
5.6	Superfície predita por KED da distribuição espacial do indicador de biomassa de raia-lenga	55
5.7	Superfície de desvios padrão das estimativas por KED do indicador de biomassa de raia-lenga	55
5.8	Variogramas empírico e teórico dos erros associados à estimação da tendência do indicador de abundância de pescada.	57
5.9	Superfície predita por KED da distribuição espacial do indicador de abundância de pescada	57
5.10	Superfície de desvios padrão das estimativas por KED do indicador de abundância de pescada	57
5.11	Histograma dos valores observados do número de elementos capturados por hora de raia-lenga.	60
5.12	Histograma dos valores observados não nulos do número de indivíduos capturados por hora de raia-lenga.	60
5.13	Superfícies preditas por ZIM das distribuições espaciais de: probabilidade de presença (painel da esquerda), abundância sob hipótese de presença da espécie (painel central) e da abundância de raia-lenga (painel da direita).	61
5.14	Superfícies por ZIM da média <i>à posteriori</i> (painel da esquerda) e do desvio padrão (painel da direita) dos efeitos espaciais do indicador de abundância de raia-lenga	62
5.15	Histograma dos valores observados do peso dos indivíduos capturados por hora de raia-lenga.	63
5.16	Histograma dos valores observados não nulos do peso dos elementos capturados por hora de raia-lenga.	63
5.17	Superfícies preditas por ZIM das distribuições espaciais de: probabilidade de presença (painel da esquerda), biomassa sob hipótese de presença da espécie (painel central) e do indicador de biomassa de raia-lenga (painel da esquerda).	64
5.18	Superfícies por ZIM da média <i>à posteriori</i> (painel da esquerda) e do desvio padrão (painel da direita) dos efeitos espaciais do indicador de biomassa de raia-lenga	65

5.19	Histograma dos valores observados do indicador de abundância de pescada.	66
5.20	Histograma dos valores observados não nulos do indicador de abundância de pescada.	66
5.21	Superfícies preditas por ZIM das distribuições espaciais de: probabilidade de presença (painel da esquerda), abundância sob hipótese de presença da espécie (painel central) e abundância de pescada (painel da esquerda). .	67
5.22	Superfícies por ZIM da média <i>à posteriori</i> (painel da esquerda) e do desvio padrão (painel da direita) dos efeitos espaciais do indicador de abundância de pescada	68
6.1	Mapa das estações fixas.	71
6.2	Mapa das estações fixas (identificadas por números) e das localizações observadas em 2015 (representadas por símbolos).	74
A.1	Mapa da costa continental portuguesa com os estratos (combinação da área geográfica e do grupo de profundidade: 0m-100m e 100m-200m). As áreas geográficas são representados pelas linhas horizontais e as linhas vermelha e azul são as linhas batimétricas dos 100 e 200 m, respetivamente.	84
C.1	Variogramas amostrais do indicador de biomassa de raia-lenga capturada por hora em cada campanha.	91
C.2	Variogramas amostrais da ocorrência de raia-lenga em cada campanha. . .	91

Lista de Tabelas

3.1	Sumário do número de indivíduos capturados por hora de raia-lenga por campanha.	28
3.2	Média e variância do número de indivíduos capturados por hora de raia-lenga em cada setor.	29
3.3	Sumário do peso dos elementos capturados de raia-lenga por hora.	30
3.4	Média e variância do peso dos elementos capturados por hora de raia-lenga em cada setor.	30
3.5	Sumário do número de indivíduos adultos de pescada capturados por hora.	33
3.6	Média e variância do número de indivíduos adultos capturados por hora capturada em cada setor.	34
3.7	Sumário do número de indivíduos juvenis de pescada capturados por hora por campanha.	35
3.8	Média e variância do número de indivíduos juvenis de pescada capturados por hora em cada setor.	36
3.9	Número de observações por tipo de fundo em cada campanha (entre parênteses estão apresentados o número de observações não nulas).	37
4.1	Estimativas dos parâmetros envolvidos nas estruturas do valor médio da abundância de raia-lenga.	40
4.2	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para a abundância de raia-lenga capturada por hora em cada campanha.	41
4.3	Sumário do número estandardizado de indivíduos capturados por hora de raia-lenga.	44
4.4	Sumário do peso estandardizado de elementos capturados por hora raia-lenga.	45
5.1	Estimativas dos coeficientes de regressão da tendência para o indicador de abundância de raia-lenga.	51
5.2	Estimativas dos parâmetros do variograma teórico dos resíduos relativos à estimação da tendência do indicador de abundância de raia-lenga.	52
5.3	Estimativas dos coeficientes de regressão da tendência do indicador de biomassa de raia-lenga.	54

5.4	Estimativas dos parâmetros do variograma teórico dos erros relativos à estimacão da tendência do indicador de biomassa de raia-lenga.	54
5.5	Estimativas dos coeficientes de regressão da tendência do indicador de abundância de pescada.	56
5.6	Estimativas dos parâmetros do variograma teórico dos erros relativos à estimacão da tendência do indicador de abundância de pescada.	56
5.7	Medidas das distribuições dos parâmetros de regressão do modelo de zero inflacionados para abundância de raia-lenga capturada por hora.	60
5.8	Medidas das distribuições dos parâmetros associados ao efeito espacial $W(\mathbf{s})$ para o modelo de zero inflacionados para abundância de raia-lenga capturada por hora.	61
5.9	Medidas das distribuições dos parâmetros de regressão do modelo de zero inflacionados para o indicador de biomassa de raia-lenga capturada por hora.	63
5.10	Medidas das distribuições dos parâmetros associados ao efeito espacial $W(\mathbf{s})$ para o modelo de zero inflacionados para o indicador de biomassa da raia-lenga capturada por hora.	64
5.11	Medidas das distribuições dos parâmetros de regressão do modelo de zero inflacionados para o indicador de abundância da pescada capturada por hora.	67
5.12	Medidas das distribuições dos parâmetros associados ao efeito espacial $W(\mathbf{s})$ para o modelo de zero inflacionados para o indicador de abundância de pescada.	67
6.1	Requisito do número de unidades amostrais por estrato.	70
6.2	Número de estações fixas por estrato (a preto) e número de estações em falta para cumprir o critério de representatividade dos estratos (a vermelho).	70
6.3	Campanha proposta pelo método de amostragem utilizado.	72
6.4	Sumário das correlações entre os valores observados e os valores estimados do indicador de abundância da raia-lenga em 2015.	75
6.5	Sumário das correlações entre os valores observados e os valores estimados do indicador de abundância da pescada em 2015.	75
6.6	Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de raia na campanha de 2015.	76
6.7	Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de raia-lenga estimada nas 100 campanhas simuladas.	76
6.8	Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de pescada na campanha de 2015.	76
6.9	Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de pescada estimada nas 100 campanhas simuladas.	76

B.1	Resultados do teste de homogeneidade da variância do número de elementos capturados por hora de raia-lenga em cada setor.	85
B.2	Resultados do teste de homogeneidade da variância do peso total de indivíduos capturados por hora de raia-lenga em cada setor.	87
B.3	Resultados do teste de homogeneidade da variância do número de elementos capturados por hora de pescada adulta em cada setor.	88
B.4	Resultados do teste de homogeneidade da variância do número de indivíduos capturados por hora de pescada pequena em cada setor.	89
C.1	Estimativas dos parâmetros envolvidos nas estruturas do valor médio do indicador de biomassa de raia-lenga.	90
C.2	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para o indicador de biomassa de raia-lenga em cada campanha.	90
C.3	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para a ocorrência de raia-lenga em cada campanha.	91

Lista de Abreviaturas

ARR	Setor Arrifana
AVE	Setor Aveiro
BER	Setor Berlengas
CAM	Setor Caminha
CTD	instrumento oceanográfico para medição de Condutividade, temperatura e profundidade (derivado do inglês <i>conductivity, temperature and depth</i>)
EMODnet	Rede Europeia de Observação e Dados Marinhos (do inglês <i>European Marine Observation and Data Network</i>)
FIG	Setor Figueira da Foz
GF	Campo Gaussiano (do inglês <i>Gaussian Field</i>)
GLM	Modelos Lineares Generalizados (do inglês <i>Generalized Linear Models</i>)
GMRF	Campo Aleatório de Markov Gaussiano (do inglês <i>Gaussian Markov Random Field</i>)
IPMA	Instituto Português do Mar e da Atmosfera
KLD	Divergência de <i>Kullback-Leibler</i> (do inglês <i>Kullback-Leibler Divergence</i>)
KED	<i>Kriging</i> com Tendência Externa (do inglês <i>Kriging with external drift</i>)
LIS	Setor Lisboa
MAT	Setor Matosinhos
MIL	Setor Milfontes

NP	Não Paramétrica (do inglês <i>Non-Parametric</i>)
POR	Setor Portimão
SAG	Setor Sagres
SIN	Setor Sines
SPDE	Equações Diferenciais Parciais Estocásticas (do inglês <i>Stochastic Partial Differential Equation</i>)
SRS	Amostragem Aleatória Estratificada (do inglês <i>Stratified Random Sampling</i>)
UTM	Transversa de Mercator Universal (do inglês <i>Universal Transverse Mercator</i>)
VSA	Setor Vila Real de Santo António
ZIM	Modelo de Zeros Inflacionados (do inglês <i>Zero-Inflated Model</i>)

Capítulo 1

Introdução

1.1 Contextualização

Nas últimas décadas, os problemas ambientais tornaram-se um assunto de interesse global, observando-se um crescimento da consciência das sociedades para com estes problemas [1]. Contudo, o meio ambiente continua a sofrer consequências dos impactos antropogénicos. O esgotamento dos recursos naturais e os impactos adversos no ecossistema, como a perda de biodiversidade, aumentam a lista de desafios que a humanidade enfrenta [2]. A perda de biodiversidade marinha é muitas vezes associada a atividades de pesca ilegais ou não adequadamente geridas. Para se poder monitorizar os recursos pesqueiros e reunir ferramentas para o aconselhamento do estado de exploração de diferentes stocks recorre-se frequentemente à análise de índices de abundância quer no tempo quer no espaço.

Com o objetivo de monitorizar a abundância dos principais recursos pesqueiros portugueses, o Instituto Português do Mar e da Atmosfera (IPMA) realiza desde 1979, campanhas de investigação. As campanhas demersais são exemplo de campanhas efetuadas pelo IPMA e cobrem a costa continental portuguesa desde Caminha até Vila Real de Santo António. Um dos principais objetivos destas campanhas é a estimação da abundância e distribuição dos juvenis de pescada e carapau na costa continental Portuguesa. No entanto, também são recolhidos dados sobre outras espécies, entre as quais a *R. clavata* estudada neste projeto, e vários parâmetros ambientais como a profundidade, salinidade e temperatura de superfície e de fundo, velocidade e direção do vento e tipo de fundo. A profundidade é medida com sonda, a temperatura e a salinidade (com profundidade associada) é medida com recurso a um CTD (Condutividade, Temperatura, Profundidade).

As campanhas demersais são realizadas com rede de arrasto de fundo. A arte de pesca arrasto de fundo, ilustrada na Figura 1.1, pode impactar o meio ambiente, mas a extensão e a duração desse impacto depende das condições locais [3].

O programa de amostragem adotado de 1989 até 2005 considera estações de pesca

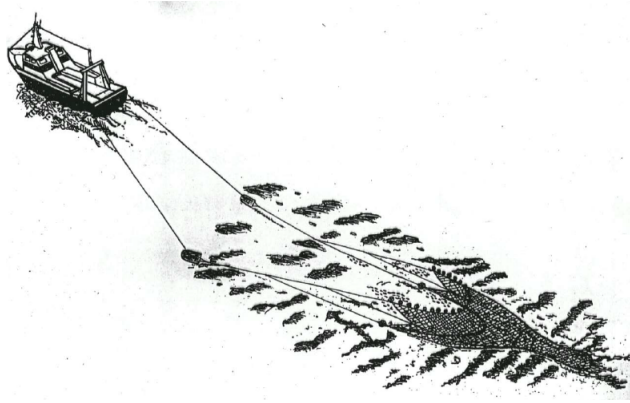


Figura 1.1: Ilustração da arte de arrasto de fundo.

relativas até à batimétrica dos 750 metros. Até à batimétrica dos 500 metros, as estações de pesca fixas são locais arrastáveis identificados nas campanhas demersais IPMA realizados entre 1979 e 1989. As restantes estações de pesca fixas basearam-se em informações resultantes de estudos realizados pelo IPMA. No total, o programa incluía 97 estações fixas e, para cada época, 30 estações de pesca são adicionadas, de forma a substituir localizações impossíveis de arrastar ou para o caso de haver disponibilidade de efetuar mais lances. Desde 2005, as campanhas demersais consideram estações até aos 500 metros de profundidade e são realizadas 66 estações fixas em conjunto com 30 estações de pesca aleatórias

As espécies em foco neste estudo são a raia-lenga, designada por *R. clavata*, e a pescada, de nome científico *Merluccius merluccius*. A *R. clavata* é uma espécie que habita no fundo do oceano, encontrada no Atlântico, na Islândia e na Noruega em direção a sul, até à África do Sul [4]. Esta espécie também é encontrada no Mediterrâneo, no oeste do Mar Negro e no sudoeste do Oceano Índico. A *M. merluccius* é uma das mais importantes unidades populacionais de peixes demersais nas águas europeias e, normalmente, é capturada em pescarias mistas em todo o Atlântico Nordeste [5]. Existem duas unidades populacionais de pescada nas águas da União Europeia: o stock do norte é encontrado no Mar do Norte e nas costas do Atlântico do Reino Unido, Irlanda e França e o stock do sul está localizado na costa atlântica de Espanha e Portugal.

A análise realizada ao longo desta dissertação irá incidir sobre indicadores de abundância e biomassa destas espécies, devido ao facto de estas serem capturadas ambas, nas campanhas demersais de investigação realizados pelo IPMA. Os dados disponibilizados pelo IPMA para o desenvolvimento deste trabalho incluem o número total de indivíduos capturados por hora, interpretado como um indicador de abundância da respetiva espécie. No caso particular da *R. clavata*, é também disponibilizado o peso total capturado por hora, que corresponde a um indicador de biomassa.

1.2 Problema e Objetivos

Este trabalho visa a análise comparativa dos indicadores de abundância da *R. clavata* e da *M. merluccius* obtidos nas campanhas demersais do IPMA e a proposta de um plano de amostragem que permita melhorar as estimativas da abundância a partir destas campanhas. As informações analisadas neste estudo referem-se a dados de abundância e biomassa de raia-lenga recolhidos em quatro campanhas, de 2013 a 2016. Para as campanhas de 2015 e 2016 também se possui informação do número de indivíduos capturados por hora de pescada.

Com os resultados da captura das espécies ao longo da costa de Portugal continental aplicaram-se metodologias estatísticas para dados geo-referenciados, como por exemplo os modelos geoestatísticos.

Neste trabalho avaliam-se igualmente o efeito de fatores ambientais (profundidade e tipo de substrato) na distribuição das duas espécies, procedendo-se igualmente à avaliação desses fatores na abundância de cada uma das espécies.

Face ao exposto, neste trabalho pretende-se proceder à estimação da distribuição espacial de cada uma das espécies e à sua sobreposição espacial. No âmbito deste trabalho construir-se-ão mapas da distribuição da *R. clavata* e da distribuição da pescada, com base nos valores observados e com auxílio da geoestatística. Estes mapas possibilitarão não são a identificação de áreas em que a ocorrência das espécies é maior, mas também a comparação das distribuições espaciais dos indicadores de abundância e biomassa da raia-lenga e da pescada. Assim, outro objetivo deste estudo é avaliar o grau de sobreposição espacial das distribuições espaciais dos indicadores de abundância de ambas as espécies. E, por último, pretende-se atualizar o desenho experimental para estações de pesca, tendo por base as distribuições estimadas de abundância de raia-lenga e pescada.

Em suma, este estudo surge da necessidade de se obter estimativas para a abundância de raia-lenga mais representativas da realidade, minimizando o impacto nas estimativas da abundância de pescada, na costa continental portuguesa. O objetivo final é o desenho de futuras campanhas de investigação que possam ser realizadas pelo IPMA.

1.3 Estado da Arte

A raia-lenga é uma espécie da família Rajidae e é considerada uma das espécies de raia mais abundante no Atlântico Nordeste, sendo a mais importante para as pescarias, do Mar do Norte até À costa sul da Ibéria. É abundante no Atlântico Nordeste [6], incluindo as águas portuguesas [7]. A *R. clavata* é um predador maioritariamente noturno, alimentando-se de crustáceos e peixes pequenos. A abundância desta espécie está fortemente correlacionada com a profundidade, o tipo de fundo do mar e a disponibilidade de presas [8]. Na costa continental portuguesa, o indicador de biomassa da *R. clavata*

apresenta, desde 1999, uma tendência crescente. O nível observado em 2017 representa o máximo histórico [9].

A pescada, *M. merluccius* é uma espécie muito procurada, que é amplamente distribuída a partir do Mar Negro, através do Estreito de Gibraltar e ao longo da costa atlântica desde Marrocos até à Noruega [10]. Desde 1998, a biomassa desovante da pescada tem vindo a aumentar. A mortalidade por pesca está a decrescer e desde 2010, o recrutamento da pescada tem-se mantido próximo do valor médio histórico [9]. Segundo Blaber (1997), o conhecimento sobre padrões de distribuição espacial de peixes e fatores influenciadores foi identificado como sendo importante para a implementação de estratégias de gestão de pesca [11].

Os principais objetivos da geoestatística são a modelação da tendência e da estrutura de correlação espacial, através do cálculo do variograma, e a predição em locais não observados [12]. A análise geoestatística engloba métodos clássicos de interpolação, como os métodos *kriging*, e métodos bayesianos, como por exemplo os modelos de zeros inflacionados (ZIM). Segundo Robertson (1987), a interpolação é considerada, na maioria dos estudos no âmbito da ecologia, uma metodologia primordial para a construção de mapas [13]. Por exemplo, Kondoh et al. (2013) adotaram *kriging* para interpolar sobre a abundância de uma espécie de veado, em localizações específicas de uma área de estudo [14]. Garcia-Soidan and Menezes (2012) propõe uma abordagem não-paramétrica baseada no método *kriging* da indicatriz, que permite a modelação de uma variável binária (variável que representa o sucesso ou insucesso de um determinado fenómeno) resultante de um processo aleatório. Neste estudo, a estrutura de correlação espacial foi analisada pelo variograma da indicatriz e, por fim, o mapa de predição é constituído pelas estimativas das probabilidades de se observar sucesso em cada localização [15]. Segundo Fernandes and Rivoirard (1999), a similaridade de estrutura de correlação ao longo de campanhas possibilita a estimação de uma estrutura de dependência média [16]. O método de *kriging* com tendência externa (KED), referido por exemplo em Webster and Oliver (2001), é apresentado como um método adequado para o caso em que a variável de interesse está correlacionada com variáveis auxiliares externas (covariáveis), as quais são importantes para a modelação da tendência inerente ao processo espacial [17]. Relativamente aos ZIM, estes são por exemplo adotados no trabalho de Ridout et al (1998), uma vez que os dados contêm um número elevado de zeros. Estes modelos complementam os modelos mais convencionais de super-dispersão, que se concentram em modelar corretamente a relação média-variância [18].

Outra metodologia estatística abordada ao longo deste trabalho é a amostragem espacial que consiste na seleção de unidades amostrais referenciadas geograficamente. Royle e Nychka (1998) descrevem uma abordagem, na qual o critério da amostragem é puramente geométrico, em vez de envolver um modelo geoestatístico assumido. De uma forma geral, esta abordagem favorece uma distribuição espacial regular dos locais de amostragem den-

tro das restrições impostas pela aplicação particular. Assim, por exemplo, ao considerar a adição ou exclusão de locais para ou de uma rede existente, os locais adicionais tenderiam a ser colocados de modo a preencher espaços vazios, enquanto os principais candidatos à remoção incluiriam um de um par de vizinhos próximos [19]. Hoef (2002) refere que a maior diferença entre a amostragem clássica e a geoestatística é que a amostragem clássica está associada a métodos *design-based*, enquanto a geoestatística está relacionada com os métodos *model-based*. Ambos os métodos produzem estimativas não-enviesadas, no entanto, no geral, os métodos geostatísticos são mais eficientes fornecendo estimativas de valores mais próximos dos valores observados [20]. Wang et al. (2004) utilizaram o método de amostragem espacial *model-based* de forma a otimizar os locais de amostragem numa área heterogénea. Assim, as localizações amostrais foram otimizadas pela minimização da variância do erro resultante da interpolação espacial pelo valor médio [21].

1.4 Estrutura da Tese

Tendo em conta o problema e os objetivos enunciados, ao longo da dissertação são integrados diversos métodos estatísticos, utilizados nas diferentes fases deste trabalho. Deste modo, no seguimento deste capítulo introdutório, o Capítulo 2 faz uma breve introdução às principais metodologias adotadas no desenvolvimento deste trabalho.

O Capítulo 3 apresenta os resultados da análise exploratória de dados por campanha de investigação de 2013 a 2016, mais especificamente das campanhas demersais que permitiram a recolha de informação da *R. clavata* e da pescada. Esta análise preliminar abrange os indicadores de abundância e biomassa destas espécies, obtidos respetivamente à custa do número total e peso total de indivíduos capturado por hora num determinado lance. Apresenta-se também uma análise de variância entre setores definidos ao longo da costa de Portugal continental. Termina-se com uma análise de possíveis covariáveis que possam explicar a distribuição destas espécies.

No Capítulo 4, prossegue-se com a análise exploratória dos dados recolhidos da *R. clavata*, tendo em conta que estes são geo-referenciados e que a sua dependência espacial deverá ser considerada. Inicialmente, faz-se uma análise da estrutura de correlação por campanha, que nos permite conhecer padrões espaciais quer via *kriging* ordinário, quer via *kriging* da indicatriz. Tendo em conta alguma similaridade da estrutura de correlação ao longo das quatro campanhas, é então investigada a possibilidade de estimar uma estrutura de dependência média.

O Capítulo 5 é dedicado à estimação dos indicadores de abundância da *R. clavata* e da pescada ao longo da costa portuguesa por metodologias espaciais distintas e, por sua vez, a interpretação dos parâmetros envolvidos na estrutura de correlação dos dados. É ainda analisada a influência das variáveis auxiliares através dos parâmetros de regressão e estudada a distribuição espacial dos indicadores de abundância das espécies. As metodo-

logias geoestatísticas utilizadas neste capítulo são a suavização de *kernel* espacial, *kriging* com tendência externa e modelos de zeros inflacionados.

No Capítulo 6, investigam-se desenhos amostrais para futuras campanhas de investigação, capazes de proporcionar estimativas mais robustas para a abundância da raia-lenga, sem prejuízo das estimativas para a pescada. O método de amostragem proposto, cumpre o requisito imposto pelo IPMA de amostragem aleatória estratificada (SRS), assumindo como fixas algumas estações de pesca definidas em campanhas anteriores. Adicionalmente, sugere-se a inclusão de cerca de 20% de novas estações, em áreas onde a ocorrência de raia-lenga é mais provável. Esta medida tem como objetivo mitigar o problema atual de zeros inflacionados nos dados recolhidos para a *R. clavata*.

No Capítulo 7, e último capítulo são apresentadas as principais conclusões decorrentes deste estudo e alguns comentários sobre possível trabalho que poderá ser realizado posteriormente.

Capítulo 2

Enquadramento Teórico

Neste capítulo apresenta-se o enquadramento teórico das metodologias utilizadas no que diz respeito a técnicas de análise exploratória e análise geoestatística.

Desta forma, é feita uma breve introdução às principais metodologias adotadas, nomeadamente modelos lineares generalizados (GLM), alguns testes de inferência tal como o da homogeneidade de variâncias, e métodos de *kriging*, que têm como objetivo a análise e a compreensão do comportamento dos dados. Relativamente à análise geoestatística, os respetivos modelos são indicados para dados recolhidos em diversas localizações espaciais, medindo a relação entre as observações amostradas em locais distintos.

As abordagens enunciadas neste capítulo foram aplicadas com auxílio do *software R*. Para a maioria das metodologias foi utilizado o pacote *base* do *software R*, com exceção de algumas abordagens em que o pacote é especificamente referido.

2.1 Modelos lineares generalizados

Os GLM são uma extensão do modelo linear clássico. Os GLM possibilitam utilizar outras distribuições para os erros e uma função de ligação para relacionar a média da variável resposta à combinação linear das variáveis explicativas. Através destes é possível modelar variáveis de interesse que assumem a forma de contagem, binárias e categóricas. Considerando que a variável resposta Z apresenta valor médio μ e variância σ^2 , os GLM assentam em três componentes: a componente aleatória, a componente sistemática e a função de ligação. A componente aleatória traduz-se pela distribuição da variável resposta Z condicionada aos valores das variáveis explicativas. A componente sistemática é ditada pelo preditor linear, que é representado por $\eta = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ onde p é o número de variáveis explicativas considerados. A terceira componente é a função de ligação $g(\mu)$, tal que $g(\mu) = \eta$.

2.1.1 Família exponencial

Os GLM são normalmente formalizados dentro da estrutura da família de distribuições exponencial [22]. A família exponencial inclui as distribuições Gaussiana, Gama, Poisson, etc. De facto, a família exponencial fornece uma notação que permite modelar variáveis discretas, contínuas, binárias e usualmente é descrita por:

$$f_z(z; \theta, \phi) = \exp\left\{\frac{z\theta - b(\theta)}{a(\phi)} + c(z, \phi)\right\} \quad (2.1)$$

onde θ representa o parâmetro característico da distribuição e ϕ o parâmetro de dispersão que normalmente é conhecido. As funções $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ são funções específicas de cada distribuição, onde $b(\cdot)$ é diferenciável. Para determinar uma expressão para o valor médio e para a variância dos dados, recorre-se à derivada do logaritmo da função densidade de probabilidade, $l(\theta; \phi, Z)$, usualmente denominada por função Score:

$$S(\theta) = \frac{\delta l(\theta; \phi, Z)}{\delta \theta} = \frac{y - b'(\theta)}{a(\phi)}. \quad (2.2)$$

Assim, sob determinadas condições de regularidade, verificam-se as igualdades $E[S(\theta)] = 0$ e $E[S^2(\theta)] = -E\left[\frac{\delta^2 l(\theta; \phi, Z)}{\delta \theta^2}\right]$ e, por sua vez,

$$\mu = b'(\theta)$$

$$Var(Z) = a^2(\phi)Var(S(\theta)) = a(\phi)b''(\theta) \quad (2.3)$$

Para ilustrar seguem alguns exemplos de distribuições em que as distribuições de Bernoulli e Gama serão abordadas no Capítulo 5. O caso particular da distribuição Tweedie será utilizado no Capítulo 4.

Distribuição de Bernoulli

Quando se pretende estudar a probabilidade de sucesso ou insucesso de um determinado acontecimento, a variável em estudo é binária. Neste caso, a variável aleatória real Z segue a lei de Bernoulli com parâmetro p e a sua função de probabilidade é dada por:

$$f_Z(z) = \begin{cases} p & \text{se } z = 1 \\ 1 - p & \text{se } z = 0 \\ 0 & \text{se } z \notin \{0, 1\} \end{cases} \quad (2.4)$$

onde $p \in]0, 1[$ representa a probabilidade de sucesso e $1 - p$ a probabilidade de insucesso. O valor esperado da variável aleatória é p e a respetiva variância corresponde a $p(1 - p)$.

Distribuição Gama

A distribuição gama é usada para modelar dados contínuos que são assimétricos à direita e maiores que 0. A sua função densidade é dada por:

$$f_Z(z) = \frac{(z/\beta)^{\alpha-1} e^{-z/\beta}}{\beta \Gamma(\alpha)}, \quad z, \alpha, \beta > 0 \quad (2.5)$$

onde $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$. O valor esperado e a variância dos dados são $\frac{\alpha}{\beta}$ e $\frac{\alpha}{\beta^2}$, respectivamente.

Distribuição Tweedie

A utilização da distribuição Tweedie, na generalidade, deve-se ao facto de se aproximar de algumas das distribuições mais conhecidas da família exponencial. Outra característica relevante é a boa acomodação de zeros, caso existam nos dados [23].

A variância da variável Z é dada por

$$Var(Z) = a(\phi)\mu^p \quad (2.6)$$

onde p é uma constante que assume os valores 0, 1, 2 e 3 perante as distribuições Gaussiana, Poisson, Gama e Gaussiana Inversa, respetivamente. A função $b(\theta)$ é representada da seguinte forma:

$$b(\theta) = \begin{cases} \frac{1}{2-p} ((1-p)\theta)^{\frac{2-p}{1-p}} & \text{se } p = 0, 3 \\ e^\theta & \text{se } p = 1 \\ -\log(-\theta) & \text{se } p = 2 \end{cases} \quad (2.7)$$

Deste modo, verifica-se que o comportamento da distribuição Tweedie difere conforme o valor da constante p .

2.2 Testes de Hipóteses

O conceito de teste de hipótese é habitualmente considerado como fundamental para a inferência estatística. A componente chave dos testes de hipóteses é permitir tirar conclusões sobre proposições distintas baseadas no valor de prova e no nível de significância considerado [24].

A formalização de testes de hipóteses envolve duas hipóteses. A hipótese a ser testada é chamada hipótese nula, denotada por H_0 , contra a hipótese alternativa, H_1 .

2.2.1 Teste para a homogeneidade de variâncias

O Teste de *Mood* é um teste não paramétrico para a homogeneidade das variâncias. Os testes não paramétricos de igualdade de variâncias são usados para testar hipóteses em que duas amostras de tamanhos m e n são provenientes de populações com características de dispersão iguais. Para calcular a estatística de teste, usa-se classificações em vez dos valores da amostra. A hipótese nula de igualdade de variâncias é $H_0 : \sigma_1^2 = \sigma_2^2$ e a hipótese alternativa é $H_1 : \sigma_1^2 \neq \sigma_2^2$, onde a desigualdade é válida.

A estatística de teste é definida da seguinte forma:

$$M = \sum_{i=1}^m \left(R_i - \frac{n + m + 1}{2} \right)^2 \quad (2.8)$$

onde M tem distribuição Normal de média 0 e desvio padrão 1. m, n são os tamanhos das amostras tal que $m \leq n$ e R_i é a classificação do i -ésimo valor da amostra com tamanho m . A discretização da distribuição estatística pode ser praticamente negligenciada quando $m, n > 20$. Segundo Gorbunova e Lemeshko (2012), a potência do Teste de *Mood* é superior à dos Testes *Ansari-Bradley*, *Siegel-Tukey*, *Capon* e *Klotz* [25].

O teste de *Mood* será utilizado no Capítulo 3, na comparação de variâncias dos indicadores de abundância e de biomassa de raia-lenga e de pescada ao longo da costa portuguesa, dos dados das campanhas demersais. A aplicação deste teste no *software R* foi realizada com auxílio do pacote *stats*.

2.2.2 Teste de interação espaço-tempo

Os testes clássicos para a interação espaço-temporal basicamente verificam se os eventos próximos no tempo também tendem a ser próximos no espaço [26]. Esta abordagem permite uma definição formal da hipótese nula de não existir interação espaço-tempo, expectável perante um processo com componentes espaciais e temporais independentes. A hipótese alternativa é definida como a existência de interação espaço-tempo. Considere-se um conjunto de dados definidos por $\{(\mathbf{s}_i, t_i) : i = 1, \dots, n\}$ onde \mathbf{s}_i representa as coordenadas espaciais e t_i os pontos temporais observados numa região \mathbf{W} durante um período $(0, T]$. Desta forma, $d_{ij}^s = \|\mathbf{s}_i - \mathbf{s}_j\|$ e $d_{ij}^t = |t_i - t_j|$.

Os testes aqui apresentados de interação espaço-tempo serão utilizados no Capítulo 4, para análise da inexistência de interações ao longo das diversas campanhas.

Teste *Knox*

As distâncias críticas no espaço (δ) e no tempo (τ) têm de ser especificadas para produzir uma categorização das distâncias em “próximo” ou “afastado”. A estatística de teste é então definida como o número de pares de eventos próximos no espaço e no tempo

de acordo com esses limites de distâncias:

$$T_{Knox} = \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i} \mathbb{1}(d_{ij}^s \leq \delta) \mathbb{1}(d_{ij}^t \leq \tau) \quad (2.9)$$

em que d_{ij}^s e d_{ij}^t correspondem a distâncias euclidianas tal que $d_{ij}^s = \|\mathbf{s}_i - \mathbf{s}_j\|$ e $d_{ij}^t = |t_i - t_j|$.

A distribuição de T_{Knox} , sob a hipótese nula de não existir interação espaço-temporal, é determinada recorrendo-se a diversas permutações de Monte Carlo. O teste de *Knox* é vantajoso pela sua simplicidade, contudo criticado pela subjetividade na especificação de δ e τ .

Este teste foi aplicado através da função *knox()* disponível no pacote *surveillance* do *software R*.

Teste de *Mantel*

Uma versão padronizada frequentemente utilizada da estatística de teste de *Mantel* é a correlação de Pearson entre as distâncias espaciais e temporais de todos os pares de eventos, ou seja,

$$T_{Mantel} = \frac{1}{n(n-1) - 2} \sum_{i=1}^n \sum_{j \neq i} \frac{d_{ij}^s - \bar{d}_s}{\hat{\sigma}_{d^s}} \frac{d_{ij}^t - \bar{d}_t}{\hat{\sigma}_{d^t}} \quad (2.10)$$

onde os símbolos \bar{d} e $\hat{\sigma}_d$ correspondem à média e desvio padrão amostrais de $n(n-1) - 2$ distâncias, respetivamente.

A distribuição de T_{Mantel} , sob a hipótese nula de não existir interação espaço-temporal, é determinada também à custa de permutações de Monte Carlo.

2.3 Métodos de análise geoestatística

Cada vez mais os investigadores são confrontados com a tarefa de analisar dados que, para além de serem multivariados e temporalmente correlacionados, estão também referenciados geograficamente [12].

A geoestatística envolve um conjunto de modelos e métodos que são designados para estudar variáveis geo-referenciadas que estão distribuídas no espaço (ou possivelmente espaço-tempo). Tais variáveis possuem um aspeto estruturado e aleatório e não podem simplesmente ser descritas por uma função regular de coordenadas [27]. O objeto de estudo é medido num conjunto finito de localizações, $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$, pertencentes a um certo domínio de área não nula, $D \subseteq \mathbb{R}^2$.

Cada observação $z_i = z(\mathbf{s}_i)$ é considerada uma realização de uma variável aleatória $Z_i = Z(\mathbf{s}_i)$, cuja distribuição é dada por um processo estocástico subjacente, espacial-

mente contínuo, $\{Z(\mathbf{s}), \mathbf{s} \in D\}$ denominado por Z . A distribuição da variável de interesse é definida por um processo estocástico, espacialmente contínuo, quando não existe a possibilidade de confusão sobre a restrição ao domínio do processo em todo o plano $\{Z(\mathbf{s}), \mathbf{s} \in \mathbb{R}^2\}$, habitualmente designado por campo aleatório.

É de salientar que as localizações de observação do processo necessitam de ser escolhidas conforme os planos de amostragem não preferenciais, ou seja a amostra deverá representar adequadamente toda a região de estudo D . Os principais objetivos da geostatística passam por modelar a tendência do processo e a respetiva estrutura de correlação, estimar parâmetros dos modelos pressupostos, realizar testes de seleção a modelos alternativos e efetuar predição em locais não observados.

2.3.1 Requisitos

Os processos espaciais subjacentes têm de verificar diversas hipóteses, nomeadamente que a média e a variância, representadas respetivamente por $E[Z(\mathbf{s})] = \mu(\mathbf{s})$ e $Var(Z(\mathbf{s})) = \sigma^2(\mathbf{s})$, são finitas.

Hipóteses relativas à estacionariedade do processo também têm de ser admitidas. Os processos espaciais fortemente estacionários exigem que, para todo $n \geq 1$, para qualquer conjunto de localizações $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\} \subset D$ e qualquer vetor $\mathbf{h} \in \mathbb{R}^2$ tal que $\{\mathbf{s}_1 + \mathbf{h}, \mathbf{s}_2 + \mathbf{h}, \dots, \mathbf{s}_n + \mathbf{h}\} \subset D$, as distribuições de $(Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))$ e $(Z(\mathbf{s}_1 + \mathbf{h}), Z(\mathbf{s}_2 + \mathbf{h}), \dots, Z(\mathbf{s}_n + \mathbf{h}))$ sejam idênticas.

Geralmente, basta considerar processos estacionários de segunda ordem. Um processo diz-se estacionário de segunda ordem ou fracamente estacionário caso se verifique

$$E[Z(\mathbf{s})] = \mu, \mathbf{s} \in D \quad (2.11)$$

e

$$Cov(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\mathbf{s}_1 - \mathbf{s}_2), \mathbf{s}_1, \mathbf{s}_2 \in D, \quad (2.12)$$

isto é, se a covariância depender apenas do vetor de separação $\mathbf{s}_1 - \mathbf{s}_2$. A função $C(\cdot)$ é a função de covariância ou covariograma.

A definição enunciada implica que o processo tenha variância constante, $\sigma^2(\mathbf{s}) = C(\mathbf{0})$ e que o covariograma seja simétrico, isto é, $C(\mathbf{h}) = C(-\mathbf{h})$. Caso $C(\mathbf{0}) \neq 0$, a segunda propriedade pode ser definida à custa da função de correlação, denotada por $\rho(\cdot)$,

$$Corr(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = \frac{Cov(Z(\mathbf{s}_1), Z(\mathbf{s}_2))}{C(\mathbf{0})} = \rho(\mathbf{s}_1 - \mathbf{s}_2), \quad (2.13)$$

onde $\rho(\mathbf{h}) = \rho(-\mathbf{h})$ e $\rho(\mathbf{0}) = 1$.

Contudo, a presença de tendência espacial é suficiente para se supor que a média não seja constante e que a ocorrência de casos em que o covariograma não seja dependente de

um vetor diferença para que não se admita a estacionariedade na variância. Assim, por vezes admite-se um tipo de estacionariedade mais fraco. Um processo diz-se intrinsecamente estacionário caso

$$E [Z(\mathbf{s}_1) - Z(\mathbf{s}_2)] = 0, \mathbf{s}_1, \mathbf{s}_2 \in D \quad (2.14)$$

e

$$\begin{aligned} Var(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) &= E [(Z(\mathbf{s}_1) - Z(\mathbf{s}_2))^2] \\ &= 2\gamma(\mathbf{s}_1 - \mathbf{s}_2), \mathbf{s}_1, \mathbf{s}_2 \in D \end{aligned} \quad (2.15)$$

À função $2\gamma(\cdot)$ chama-se variograma e a $\gamma(\cdot)$ semivariograma. Existem relações entre a estacionariedade de segunda ordem e a estacionariedade intrínseca e facilmente se verifica por

$$\begin{aligned} \gamma(\mathbf{s}_1 - \mathbf{s}_2) &= \frac{1}{2} Var(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) \\ &= \frac{1}{2} \{Var(Z(\mathbf{s}_1)) + Var(Z(\mathbf{s}_2)) - 2Cov(Z(\mathbf{s}_1), Z(\mathbf{s}_2))\} \end{aligned} \quad (2.16)$$

e pode ser reescrito através de

$$\gamma(\mathbf{s}_1 - \mathbf{s}_2) = \sigma^2 [1 - \rho(\mathbf{s}_1 - \mathbf{s}_2)] \quad (2.17)$$

De acordo com Carvalho e Natário (2008) o processo espacial pode ser decomposto em vários subprocessos: $Z(\mathbf{s}) = \mu(\mathbf{s}) + W(\mathbf{s}) + \epsilon(\mathbf{s})$ [12], cada um deles reflete o comportamento do fenómeno numa determinada escala:

- Uma primeira componente determinística $\mu(\cdot)$, normalmente denominada em geoestatística por tendência espacial, reflete o comportamento médio do processo e traduz a variabilidade em larga escala;
- Uma segunda componente $W(\cdot)$, de valor médio não nulo, estacionária de segunda ordem e contínua em média quadrática, modela a ligação espacial entre as observações e traduz a variabilidade em pequena e micro-escala¹;
- Uma terceira componente pretende representar os erros de medida e é modelada por um ruído branco $\epsilon(\cdot)$, independente da componente anterior, sob o pressuposto que, se fossem feitas várias observações do processo numa certa localização, estas flutuariam em torno do correspondente valor do processo.

Assim, os semivariogramas de processos estacionários de segunda ordem são genericamente representados por

$$\gamma(\mathbf{h}) = \tau^2 + \sigma^2(1 - \rho(\mathbf{h})), \quad (2.18)$$

¹Por vezes, esta parcela da decomposição é ainda decomposta em duas $W(\mathbf{s}) = W_1(\mathbf{s}) + W_2(\mathbf{s})$, estacionários de segunda e independentes, tal que $W_1(\cdot)$ e $W_2(\cdot)$ explicam a variação em pequena e micro-escala, respetivamente. Contudo, é impossível distinguir entre o efeito de $W_2(\cdot)$ e de $\epsilon(\cdot)$ quando existem observações repetidas nas mesmas de localizações.

em que $\rho(\cdot)$ é a função de correlação de um processo estacionário de segunda ordem, contínuo em média quadrática, e $\tau^2 = \text{Var}(\epsilon(\mathbf{s}))$ e σ^2 são parâmetros não negativos. Geralmente, à medida que a distância $\mathbf{s}_2 - \mathbf{s}_1$ aumenta, o variograma tende para um valor designado por *sill* e que representa a variância total do processo, $\tau^2 + \sigma^2$. Quando o variograma atinge este valor é porque existe uma distância, denominada por *range* ou raio de influência e denotada por ϕ , a partir da qual a associação entre as observações é nula. τ^2 , designado por efeito pepita, é o parâmetro representativo da variância do ruído.

Outro pressuposto requerido é a isotropia que se define normalmente à custa do variograma. Um processo espacial intrinsecamente estacionário é isotrópico, caso o variograma dependa apenas de \mathbf{h} através do seu comprimento $\|\mathbf{h}\|$, isto é, quando

$$\gamma(Z(\mathbf{s}), Z(\mathbf{s} + h)) = \gamma(\|\mathbf{h}\|)$$

para quaisquer $\mathbf{s} \in D$ e $\mathbf{h} \in \mathbb{R}^2$ tais que $\mathbf{s} + \mathbf{h} \in D$.

Um processo intrinsecamente estacionário isotrópico denomina-se homogêneo e um processo que não é isotrópico designa-se por anisotrópico.

2.3.2 Variograma empírico

Considerando que o processo subjacente aos dados é intrinsecamente estacionário e isotrópico, a associação entre as variáveis do processo pode ser modelada através de um variograma omnidirecional. É possível comparar a modelação do variograma à modelação paramétrica da distribuição teórica de uma variável aleatória. Deste modo, é criado um estimador do variograma com base na amostra, designado por variograma empírico. A partir deste último pode ser selecionada uma ou mais famílias paramétricas de variogramas para o eventual variograma teórico. A seleção das famílias paramétricas de variogramas é explicada nas Secções 4.1 e 5.2.

Matheron, em 1962, propôs através do método dos momentos o estimador para o semivariograma, conhecido por variograma empírico, apresentado por:

$$\hat{\gamma}(u) = \frac{1}{2|N(u)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(u)} (Z_i - Z_j)^2 \quad (2.19)$$

onde $N(u) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| = u\}$ e $|N(u)|$ é o respetivo cardinal [28]. No entanto, o variograma empírico acarreta alguns inconvenientes para poder ser um bom instrumento de estimação, tal como o facto de os pontos apresentarem tal dispersão que se torna impossível encontrar um padrão de comportamento do variograma.

Famílias paramétricas

Da mesma forma que acontece com as distribuições de variáveis aleatórias, entre os variogramas isotrópicos válidos também se definem classes cujos elementos podem ser representados por uma expressão analítica comum relativamente simples, variando num pequeno número de parâmetros. Estas classes designam-se por famílias paramétricas de variograma do tipo

$$P = \{\gamma : \gamma(\cdot) = \gamma(\cdot; \theta), \theta \in \Theta\}$$

onde $\gamma(\cdot; \theta)$ é uma função condicionalmente definida-negativa dependendo de valores dados no vetor de parâmetros θ . Uma função é condicionalmente definida-negativa se para todos os pesos w_1, \dots, w_N sujeitos a $\sum_{i=1}^n w_i = 0$ e localizações $\mathbf{s}_1, \dots, \mathbf{s}_n$ contém

$$\sum_{i=1}^N \sum_{j=1}^N w_i \gamma(\mathbf{s}_i, \mathbf{s}_j) w_j \leq 0.$$

Alguns modelos isotrópicos são o de Matérn, o Esférico, o Gaussiano e o Exponencial.

- Modelo de Matérn

$$\gamma(u) = \tau^2 + \sigma^2 \left[1 - \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{u}{\phi} \right)^{\nu} K_{\nu} \left(\frac{u}{\phi} \right) \right], \quad u \geq 0 \quad (2.20)$$

em que $K_{\nu}(\cdot)$ é a função de Bessel modificada de ordem ν , $\nu > 0$. A ordem é um parâmetro com um papel importante na regularidade das trajetórias do modelo. Caso $\nu = 0,5$ estar-se-á perante um modelo Exponencial e se $\nu \rightarrow \infty$ considerar-se-á um modelo Gaussiano. $\tau^2 + \sigma^2$ corresponde ao *sill* e ϕ ao raio de influência.

- Modelo Esférico

$$\gamma(u) = \begin{cases} \tau^2 + \sigma^2 & \text{se } u > \phi \\ \tau^2 + \sigma^2 \left[\frac{3u}{2\phi} - \frac{1}{2} \left(\frac{u}{\phi} \right)^3 \right] & \text{se } 0 < u \leq \phi \\ 0 & \text{se } u = 0 \end{cases} \quad (2.21)$$

em que $\tau^2 + \sigma^2$ corresponde ao *sill* e ϕ ao *range*. Este tem um comportamento linear perto da origem e é um dos modelos mais utilizados devido à facilidade com que se ajusta aos dados.

- Modelo Gaussiano

$$\gamma(u) = \tau^2 + \sigma^2 \left(1 - \exp \left(- \frac{u}{\phi} \right)^2 \right), \quad u \geq 0 \quad (2.22)$$

tal que o parâmetro $\tau^2 + \sigma^2$ corresponde ao *sill* e ϕ ao *range*, raio de influência.

- Modelo Exponencial

$$\gamma(u) = \tau^2 + \sigma^2 \left(1 - \exp \left(-\frac{u}{\phi} \right)^k \right), \quad u \geq 0 \quad (2.23)$$

onde $0 < k \leq 2$, os parâmetros $\tau^2 + \sigma^2$ e ϕ representam o *sill* e *range*, respetivamente. Tem um comportamento parabólico perto da origem.

2.3.3 Estimação

A estimação dos parâmetros envolvidos nas estruturas do valor médio e da associação espacial do modelo adotado para o processo estocástico é bastante relevante. Relativamente à estrutura de correlação espacial pretende-se estimar a variância dos dados σ^2 , a variância τ^2 do erro de medição, o valor do raio de influência ϕ , e outros parâmetros envolvidos no modelo teórico de variograma selecionado.

Estimação da tendência espacial

Considerando que o processo pode ser descrito de forma genérica por

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}) = \beta_0 + \sum_{j=1}^p \beta_j X_j(\mathbf{s}) + \delta(\mathbf{s}), \quad (2.24)$$

onde X_j , $j = 1, \dots, p$ correspondem a variáveis explicativas não aleatórias, que podem ou não depender da localização, e $\delta(\mathbf{s})$ é um processo de valor médio nulo e variância finita que poderá ou não ter correlação espacial.

O vetor β é habitualmente estimado por métodos de máxima verosimilhança ou o dos mínimos quadrados. A estimação dos parâmetros da estrutura de correlação do modelo é efetuada sobre a componente $\delta(\mathbf{s}) = Z(\mathbf{s}) - \mu(\mathbf{s})$ através do cálculo do variograma.

Estimação do variograma

O variograma é utilizado para fins exploratórios, mas também para a estimação dos parâmetros envolvidos na estrutura de associação espacial.

Um problema abordado no estudo de variogramas amostrais é o da anisotropia. O modelo de anisotropia mais comum é o da anisotropia geométrica, que ocorre quando uma estrutura de covariância estacionária é transformada por um diferencial de alongamento e de rotação dos eixos de coordenadas. Assim, a anisotropia geométrica é definida por dois parâmetros adicionais. Um modelo com anisotropia geométrica com coordenadas espaciais $\mathbf{s} = (s_1, s_2)$ pode ser convertido num modelo estacionário com coordenadas $\mathbf{s}' = (s'_1, s'_2)$,

pela transformação

$$(s'_1, s'_2) = (s_1, s_2) \begin{bmatrix} \cos(\psi_A) & -\text{sen}(\psi_A) \\ \text{sen}(\psi_A) & \cos(\psi_A) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & \psi_R^{-1} \end{bmatrix} \quad (2.25)$$

onde ψ_A e $\psi_R > 1$ dizem respeito ao ângulo e rácio da anisotropia [29].

Para estimar os parâmetros subjacentes à estrutura de associação espacial, pode utilizar-se o método da máxima verosimilhança ou o método dos mínimos quadrados. Ao aplicar estes métodos, supõe-se que a tendência espacial está identificada e também que se está sob a posse do variograma empírico da componente δ , cujo objetivo será estimar o vetor de parâmetros θ por:

- Método da máxima verosimilhança

A estimação por máxima verosimilhança é um método estatístico universalmente aceite com boas propriedades para amostras de grande dimensão. No contexto da Geoestatística, a implementação da máxima verosimilhança é apenas direta quando os dados são gerados por um modelo Gaussiano.

Supondo que $Z = (Z_1, Z_2, \dots, Z_n) \sim N(\mathbf{X}\beta, \Sigma(\theta))$ em que \mathbf{X} é a matriz $n \times p$ com as observações de p covariáveis, β o vetor de coeficientes de regressão e $\Sigma(\theta)$ a matriz de covariância que pode ser descrita por

$$\Sigma(\theta) = \tau^2 \mathbf{I} + \sigma^2 \mathbf{R}(\phi) \quad (2.26)$$

onde $\mathbf{R}(\phi)$ corresponde à matriz de correlação sendo $r_{ij} = \rho(\|s_i - s_j\|; \phi)$, $i, j = 1, \dots, n$ e \mathbf{I} a matriz identidade de dimensão $n \times n$. Por conseguinte, todos os parâmetros do modelo podem ser estimados maximizando a função de log-verosimilhança que é dada por:

$$L(\beta, \tau^2, \sigma^2, \phi) = -\frac{1}{2} \{n \log(2\pi) + \log(|\tau^2 \mathbf{I} + \sigma^2 \mathbf{R}(\phi)|) + (\mathbf{Z} - \mathbf{X}\beta)^t (\tau^2 \mathbf{I} + \sigma^2 \mathbf{R}(\phi))^{-1} (\mathbf{Z} - \mathbf{X}\beta)\} \quad (2.27)$$

- Método dos mínimos quadrados

Os critérios subjacentes ao método dos mínimos quadrados não são tão restritivos, dado que requer em menos pressupostos na distribuição de $Z(\mathbf{s})$. Este método pode ser descrito em notação matricial

$$\min_{\theta} \{(\hat{\gamma} - \gamma_{\theta})^T \mathbf{W}^{-1} (\hat{\gamma} - \gamma_{\theta})\} \quad (2.28)$$

onde $\hat{\gamma}$ representa um estimador empírico e γ_{θ} um modelo paramétrico válido, em que a forma é conhecida exceto para os parâmetros desconhecidos em θ . A matriz

\mathbf{W} corresponde à matriz de pesos tal que:

1. Se \mathbf{W} corresponde à matriz identidade, então estar-se-á perante o critério dos mínimos quadrados ordinários e o problema pode ser escrito por

$$\min_{\theta} \left\{ \sum_{i=1}^L (\hat{\gamma}(u_i) - \gamma_{\theta}(u_i))^2 \right\}, \quad (2.29)$$

em que L identifica o número de *lags* (isto é, distâncias) utilizadas na estimação. A simplicidade de aplicação é a principal vantagem deste método. Porém, não tem em consideração informação sobre a variância e covariância da distribuição do estimador empírico, nem do número de pares de pontos envolvidos na estimação do *lag* u .

2. se $\mathbf{W} = \mathbf{V}$ onde \mathbf{V} é a matriz covariância cujos elementos do tipo $V_{ij} = Cov|\hat{\gamma}(u_i), \hat{\gamma}(u_j)|$, então estar-se-á perante o critério de mínimos quadrados generalizados.

Este método tem como principal desvantagem o facto de ser demasiado complexo e depender da função desconhecida $\hat{\gamma}$.

3. Se \mathbf{D} representa a matriz diagonal onde $d_{ii} = Var|\hat{\gamma}(u_i)|$, então estar-se-á perante o critério dos mínimos quadrados pesados. Este é um caso em que é balanceada a eficiência do critério de mínimos quadrados generalizados e a simplicidade do critério de mínimos quadrados ordinários.

2.3.4 Predição espacial

Formalmente, na predição espacial pretende-se predizer, com base em n observações $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))' = (Z_1, \dots, Z_n)'$ de um campo aleatório, o valor $Z(\mathbf{s}_0)$ numa localização \mathbf{s}_0 .

Considerando que algumas medições na vizinhança do ponto \mathbf{s}_0 investigado, ou por vezes noutros locais, estão mais relacionadas com o valor verdadeiro nesse ponto, o procedimento mais adequado passa por adotar uma média pesada.

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) + \lambda_0 \quad (2.30)$$

onde i é o índice dos pontos amostrados, λ_0 pode ser 0 ou não tal que $\lambda_0 = \mu(1 - \sum_{i=1}^n \lambda_i)$, e o erro de predição é denotado por $\epsilon(\mathbf{s}_0) = \hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)$. De acordo com o método *kriging* escolher-se-á os pesos de predição, λ_i , que tornam o estimador centrado e de variância mínima. Esta é uma abordagem probabilística para interpolar entre pontos amostrados. Existem diferentes tipos de *kriging* dependendo das hipóteses do modelo, tais como *kriging* ordinário, *kriging* da indicatriz e *kriging* com tendência externa.

Os métodos clássicos de análise geoestatística foram realizados utilizando o *software R* com auxílio do pacote *RGeostats*, de forma a obter os resultados apresentados no Capítulo 4 e Secção 5.2.

***Kriging* ordinário**

kriging ordinário corresponde ao caso em que a média do processo é uma constante desconhecida [27]. A soma dos pesos de predição, $\sum \lambda_i$, deve ser 1 e $\lambda_0 = 0$. Isto pode assegurar que o erro tem valor médio zero independentemente da média desconhecida. Minimizar a variância sob a restrição da soma dos pesos resulta no seguinte sistema linear, que deve ser resolvido em ordem aos pesos de predição:

$$\begin{cases} \sum_j \lambda_j C_{ij} + \mu = C_{i0} & i = 1, \dots, n \\ \sum_j \lambda_j = 1 \end{cases} \quad (2.31)$$

onde $i, j = 1, \dots, n$ e μ corresponde a um parâmetro de Lagrange introduzido pela restrição. $C_{i0} = Cov(Z(\mathbf{s}_i), Z_0)$ e, por isso, o sistema pode ser escrito em termos do variograma em vez da covariância, substituindo μ por $-\mu$.

A variância do estimador pode ser descrita por

$$\sigma_K^2 = C_{00} - \sum_i C_{i0} - \mu \quad (2.32)$$

Este método pode ser utilizado, em particular, na estimação global para estimar a média sobre um domínio.

***Kriging* da indicatriz**

Como referido anteriormente, *kriging* ordinário permite a partir de um conjunto de dados geo-referenciados construir um mapa interpolado da variável de interesse em toda a região de estudo. Porém, pode-se estar interessado em estudar a probabilidade de a variável em estudo exceder ou não um determinado valor x . Uma abordagem para lidar com este problema designa-se *kriging* da indicatriz, que prova ser um método eficiente no que respeita à proporção de variância explicada [30].

A definição da variável binária a partir de uma variável numérica realizar-se-ia através da seguinte transformação:

$$I_{Z(\mathbf{s},x)} = I_{\{Z(\mathbf{s})>x\}} = \begin{cases} 1, & \text{se } Z(\mathbf{s}) > x \\ 0, & \text{caso contrário} \end{cases} \quad (2.33)$$

O uso do estimador do método dos mínimos quadrados da função indicatriz para a apro-

ximação do seu valor esperado é da seguinte forma:

$$\hat{I}_Z(\mathbf{s}, x) = \sum_{i=1}^n \lambda_i I_Z(\mathbf{s}_i, x) = \sum_{i=1}^n \lambda_i I_{\{Z(\mathbf{s}_i) > x\}} \quad (2.34)$$

onde $Z(\mathbf{s}_i)$ denota a variável observada na localização \mathbf{s}_i . Os parâmetros λ_i , dependentes de \mathbf{s} e x , podem ser obtidos por minimização do erro quadrático médio das predições

$$E \left[\left(I_Z(\mathbf{s}, x) - \hat{I}_Z(\mathbf{s}, x) \right)^2 \right]$$

sujeito a $E \left[\hat{I}_Z(\mathbf{s}, x) \right] = E \left[I_Z(\mathbf{s}, x) \right]$.

Na aplicação deste método no presente estudo, o valor de x é zero dado que se pretende estudar a probabilidade de ocorrência da espécie. Após a variável numérica ser transformada em binária, o *kriging* da indicatriz decorre seguindo-se o mesmo algoritmo do método *kriging* ordinário.

Kriging com Tendência Externa

Kriging com tendência externa é um método geoestatístico muito utilizado que permite estimar a variável de interesse tendo em conta variáveis explicativas amostradas [31]. Neste caso, a média de valores de variáveis auxiliares e é descrita da seguinte forma

$$\mu(\mathbf{s}) = \sum_{j=0}^p f_j(\mathbf{s}) \beta_j \quad (2.35)$$

tal que $\beta = (\beta_0, \dots, \beta_p)^T \in \mathbb{R}^{p+1}$ desconhecido e p representa o número de variáveis exploratórias consideradas. $f_j(\cdot)$ são funções das variáveis exploratórias conhecidas e $f_0(\cdot) \equiv 1$.

O estimador linear, $\hat{Z}(\mathbf{s}_0)$, da variável $Z(\mathbf{s}_0)$ numa localização \mathbf{s}_0 é escrito por:

$$\hat{Z}_0 = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) + \lambda_0 = \beta f_0 + \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)) \quad (2.36)$$

onde f_0 representa o vetor de valores das funções exploratórias no ponto \mathbf{s}_0 . De forma a que o erro apresente média zero, o somatório dos pesos de predição deve ser 1 e para que a variância dos mesmos seja minimizada então deve ser verificada a seguinte igualdade

$$\sum_{i=1}^n \lambda_i f_i = f_0$$

Os pesos e os parâmetros de Lagrange μ_l introduzidos para cumprir as restrições são a

solução do sistema

$$\begin{cases} \sum_{i=1}^n \lambda_i = 1 \\ \sum_{i=1}^n \lambda_i f_i = f_0 \\ \sum_{j=1}^n \lambda_j Cov(Z_i - \hat{Z}_i, Z_j - \hat{Z}_j) + \sum_{l=0}^p \mu_l f_i^l = Cov(Z_i - \hat{Z}_i, Z_0 - \hat{Z}_0) \end{cases} \quad (2.37)$$

2.4 Uma abordagem geoestatística Bayesiana

A modelação espacial preocupa-se essencialmente com três questões: especificação do modelo, estimação e inferência das estimativas dos parâmetros e predição. A abordagem Bayesiana pode abordar mais facilmente a especificação do modelo e, por sua vez, a inferência e predição [32]. Apesar desta dissertação se apoiar no paradigma frequentista, especificamente na abordagem *kriging* clássica, pretende-se também aplicar um modelo geoestatístico Bayesiano, tendo em conta a sua adequabilidade a este tipo de dados.

A estatística Bayesiana é baseada no teorema da probabilidade condicional de Bayes:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (2.38)$$

onde A e B são duas variáveis aleatórias. O teorema de probabilidade de Bayes apresentado na Expressão (2.38) pode também ser visto na estrutura de modelação estatística como:

$$p(\text{hipótese}|\text{dados}) = \frac{p(\text{dados}|\text{hipótese})p(\text{hipótese})}{p(\text{dados})} \quad (2.39)$$

em que tipicamente a hipótese expressa em termos dos parâmetros do modelo estatístico (θ). Ao contrário do que acontece na abordagem clássica, os parâmetros na estatística Bayesiana estão associados a distribuições de probabilidade, não sendo definidos apenas por estimativas pontuais. $p(\text{dados}|\text{hipótese})$ representa a verosimilhança do modelo e $p(\text{hipótese})$ a distribuição *à priori* dos parâmetros, ou seja, a informação que se possui sobre os parâmetros *à priori* dos dados que se está a analisar na modelo considerado.

A abordagem geoestatística Bayesiana adotada neste trabalho é suportada pelos ZIM, que surgem devido ao excesso de zeros, habitualmente em dados recolhidos em campanhas científicas [33].

2.4.1 Modelos de zeros inflacionados

Uma solução cada vez mais abordada para lidar com o elevado número de zeros é a distribuição de zeros inflacionados. Uma distribuição de zeros inflacionados é uma mistura de modelos de duas distribuições, a primeira para lidar com os zeros (ausência) e a segunda de forma a lidar com os valores positivos (abundância) [34]. A distribuição

de probabilidade da distribuição de zeros inflacionados pode ser definida por:

$$P(Z = z) = \begin{cases} p + (1 - p)P(Y = 0) & \text{se } z = 0 \\ (1 - p)P(Y = z) & \text{se } z > 0 \end{cases} \quad (2.40)$$

onde a variável aleatória Y segue uma distribuição de probabilidade. Quando $p \rightarrow 0$, o modelo de zeros inflacionados corresponde à distribuição de Y .

Alternativamente, a mistura de distribuições pode ser definida por:

$$T(\mathbf{s}) \sim \text{Bernoulli}(p(\mathbf{s})) \quad (2.41)$$

$$Y(\mathbf{s}) \sim \text{Gama}(a(\mathbf{s}), b(\mathbf{s})) \quad (2.42)$$

onde $T(\mathbf{s})$ representa a variável aleatória binária, na localização \mathbf{s} , cujos valores podem ser 0 ou 1 caso se observe valores nulos ou não, respetivamente. $Y(\mathbf{s})$ é uma variável aleatória que toma os valores positivos da variável de interesse $Z(\mathbf{s})$ caso, para a localização \mathbf{s} , o valor de $T(\mathbf{s})$ seja 1.

Os parâmetros envolvidos na Expressões (2.41) e (2.42) são modelados como funções das covariáveis, através das funções de ligação *logit* e *log*, à semelhança do referido na Secção 2.1, ficando respetivamente:

$$\begin{aligned} \text{logit}(p(\mathbf{s})) &= \alpha_1 + \sum_{j=1}^{p_1} \beta_{1,j} X_{1,j}(\mathbf{s}) + W(\mathbf{s}) \\ \log(a(\mathbf{s})/b(\mathbf{s})) &= \alpha_2 + \sum_{j=1}^{p_2} \beta_{2,j} X_{2,j}(\mathbf{s}) + kW(\mathbf{s}) \end{aligned} \quad (2.43)$$

onde $X_{i,j}(\mathbf{s}), i = 1, 2$ representa o valor da variável explicativa j na localização \mathbf{s} e $\alpha_i, \beta_{i,j}, i = 1, 2$ são os coeficientes de regressão. A função de ligação *logit* é descrita por $\text{logit}(p(\mathbf{s})) = \log\left(\frac{p(\mathbf{s})}{1-p(\mathbf{s})}\right)$ e as covariáveis utilizadas para modelar a presença podem ser diferentes das utilizadas para modelar a abundância.

O Campo Aleatório de Markov Gaussiano (GMRF), identificado por $W(\mathbf{s})$, é um processo intrinsecamente estacionário de média zero [35]. $W(\mathbf{s})$ é determinado a partir de um Campo Gaussiano (GF) contínuo, através de um método baseado em Equações Diferenciais Parciais Estocásticas (SPDE), proposto por Lindgren et al. (2011). A abordagem SPDE permite aproximar um campo contínuo espacial, representado por uma função de covariância de Matérn, por um campo de Markov. Esta aproximação permite ganhos significativos em termos de eficiência computacional, sendo bastante adequada para dados de elevada dimensão habituais em Geoestatística.

A parametrização é realizada em termos da variância dos dados σ^2 , do raio de influência ϕ e da variância associada aos erros de medição τ^2 . O parâmetro k é utilizado de forma a ajustar a escala da variável Y à da variável T .

De uma forma geral, verifica-se

$$[Z(\mathbf{s})] = [T(\mathbf{s})] [Y(\mathbf{s})|T(\mathbf{s})] \quad (2.44)$$

onde $[.]$ significa “distribuição de”.

Esta abordagem geoestatística Bayesiana foi realizada utilizando o pacote *inla* do *software R* em que os resultados estão apresentados no Capítulo 5, mais precisamente na Secção 5.3.

2.5 Amostragem espacial

A amostragem consiste em selecionar parte de uma população e observá-la com o objetivo de estimar uma ou mais características para a totalidade da mesma.

A amostragem espacial consiste na escolha da amostra dependente da posição geográfica das localizações [36].

Geralmente, um conjunto de dados geo-referenciados consiste num número n finito de localizações \mathbf{s}_i , $i = 1, \dots, n$ e medidas associadas Z_i . Caso se reconheça que as realizações da variável de interesse e as localizações sejam estocásticas, então o modelo para os dados é uma distribuição conjunta para as localizações e as respectivas medidas, que pode ser representado formalmente por $[S, Z]$ [29].

A inferência e a amostragem, na perspectiva espacial, estão divididas em seis principais passos [37]:

1. Clarificar os objetivos da amostragem: estimar um parâmetro populacional ou modelar uma tendência média do parâmetro;
2. Definição da população a ser amostrada e das unidades amostrais;
3. Escolha dos métodos de amostragem e da dimensão da amostra;
4. Definição do plano de amostragem que descreve onde, quando e como as unidades amostrais são escolhidas;
5. Recolha da amostra;
6. Análise da amostra, relativamente à aleatoriedade, tendência global das estimativas, erros ou dependência espacial, utilizando estatística espacial.

Segundo Soutinho (2014), as unidades amostrais, em amostragem espacial, podem ser classificadas em três categorias: pontos, linhas e segmentos. Pontos e segmentos são os mais usados em inquéritos para obtenção de estatísticas sobre o uso e ocupação do solo,

sendo a utilização de linhas menos habitual [36]. Outro aspeto inerente à amostragem espacial é a existência de associação espacial, onde, sob critérios geográficos, realizações próximas apresentam correlação mais elevada do que observações mais afastadas.

A amostragem clássica e a amostragem espacial podem ser distinguidas através da consideração ou rejeição da heterogeneidade espacial. A heterogeneidade espacial surge quando não se verifica a existência de variabilidade uniforme relativa aos valores da variável de interesse em cada localização da área populacional. Uma vez que na abordagem clássica, os dados têm de ser independentes e identicamente distribuídos, em contextos espaciais assume-se a homogeneidade espacial dos dados.

A teoria de amostragem clássica (Cochran, 1977) tende a enfatizar a virtude de algumas formas de amostragem aleatória para assegurar a estimação não enviesada subjacente às características da população, enquanto que a teoria de amostragem espacial (Matérn, 1986) mostra que, sob alguns pressupostos típicos da modelação, as propriedades espaciais são mais eficientemente estimadas quando as localizações dos dados formam um grelha regular [29] [38] [39].

2.5.1 Métodos de estimação: *design-based vs. model-based*

Duas abordagens em amostragem para a inferência de parâmetros populacionais são: os métodos *design-based*, baseados no desenho amostral, e os métodos *model-based*, baseados num modelo estatístico.

O método *design-based* é baseado na teoria clássica de amostragem, em que as unidades amostrais são selecionadas de acordo com probabilidades de inclusão e a inferência estatística é baseada no esquema de amostragem [40].

Os métodos *model-based*, como o próprio nome indica, é utilizado para descrever uma abordagem para problemas geoestatísticos baseados na aplicação de métodos formais estatísticos sob um modelo estocástico explicitamente assumido [41]. Segundo Brus (2010), na abordagem *model-based* não existem pressupostos no método para a seleção de localizações de amostragem, e tipicamente são selecionadas por amostragem seletiva. Na inferência estatística, é introduzido um modelo para a variação espacial, como por exemplo modelos *kiriging*.

Em ambas as abordagens o conceito de aleatoriedade é encarado de forma diferente [40]. Enquanto no método *design-based* a seleção das unidades amostrais é aleatória, na abordagem *model-based* a aleatoriedade é introduzida por um modelo de variação espacial.

De uma forma geral, ambas as abordagens abrangem metodologias que englobam técnicas de seleção de amostras e inferência. Outra característica comum a ambas é o facto de poderem incorporar informações e conhecimento prévios na amostragem [42].

2.5.2 Esquemas de amostragem

A amostragem aleatória, a SRS e a amostragem preferencial são alguns exemplos de esquemas de amostragem [43].

A amostragem aleatória é um método simples em que todas as localizações têm a mesma probabilidade de serem selecionadas e são independentes entre si. Normalmente, este esquema é pouco adotado por ser muito dispendioso, dado que a sua aplicação pode resultar em amostras dispersas geograficamente. No entanto, este tipo de esquema pode ser bastante útil se a população for pequena ou se a dispersão geográfica não constituir qualquer problema.

Existem situações em que é possível identificar subconjuntos dos elementos da população, que resultam na divisão da área populacional em k subáreas. Neste caso, estar-se-á perante a amostragem estratificada. Este esquema consiste em escolher aleatoriamente unidades amostrais de cada uma das k subáreas. A amostra é então composta pelas unidades amostrais selecionadas em cada subárea. De realçar que este tipo de amostragem é aplicável quando é possível identificar subáreas que variam entre si, mas que sejam homogéneas dentro si [44].

A amostragem não-preferencial assume que o processo de amostragem e o processo relativo à variável de interesse Z são independentes e a distribuição conjunta de S e Z é dada por $[S, Z] = [S][Z]$. Ao contrário, na amostragem preferencial a distribuição conjunta de S e Z pode ser escrita por $[S, Z] = [S][Z | S]$. Assim, o alvo inferencial implícito de uma análise geoestatística convencional, que apenas analisa os dados Z é a distribuição condicional $[Z | S]$, enquanto na amostragem não-preferencial, o alvo pretendido é a distribuição $[Z]$, e não há razão geralmente para supor que estas sejam iguais [29].

O presente trabalho, mais concretamente o estudo realizado no Capítulo 6, combina a SRS e a amostragem preferencial.

Capítulo 3

Os dados das campanhas demersais de 2013 a 2016

Neste estudo são utilizadas várias metodologias de análise estatística com o intuito de compreender o comportamento estatístico dos dados e também perceber as relações existentes entre diversos fatores, como por exemplo a profundidade e o tipo de fundo. O conhecimento estatístico dos dados, através da análise exploratória, constitui uma etapa fundamental para análise que deve preceder a modelação geoestatística [45].

Deste modo, neste capítulo são apresentados os dados geo-referenciados de abundância e biomassa por espécie e por estação de pesca, recolhidos nas campanhas demersais do IPMA, onde é utilizada arte de arrasto de fundo. Os resultados da análise exploratória são determinantes para a aplicação das metodologias apresentadas nos próximos capítulos. As campanhas demersais com arrasto de fundo IPMA têm como um dos alvos a avaliação da abundância e distribuição dos juvenis de pescada e diversos parâmetros ambientais. Cobrem a costa continental portuguesa desde Caminha até Vila Real de Santo António. Nestas campanhas verifica-se a coocorrência de raia-lenga.

Os dados de raia-lenga analisados dizem respeito a estações de pesca efetuadas nas campanhas demersais IPMA no período 2013-2016. Para cada estação de pesca dispõe-se das coordenadas geo-referenciadas de estações de pesca e o número e peso (em Kg) de indivíduos capturados de raia-lenga por hora de arrasto. O número total por hora de pescada em cada estação de pesca está disponível para campanhas demersais IPMA no período 2015 e 2016, em 101 estações de pesca. A par da informação relativa à abundância das duas espécies, dispõe-se de informação sobre tipo de fundo, a área geográfica (Figura A.1 do Apêndice A) associado a cada localização, e a profundidade a que cada lance de pesca foi iniciado e finalizado. Os dados relativos ao tipo de fundo foram obtidos através da rede de organizações, apoiadas pela política marítima integrada da União Europeia, Rede Europeia de Observação e Dados Marinhos (EMODnet).

Nas quatro campanhas, foram efetuados lances em 212 localizações distintas. Existe conhecimento relativo à abundância da *R. clavata* em nove setores, apresentados na Fi-

gura A.1 do Apêndice A, que perfazem as zonas norte e sudoeste da costa continental portuguesa.

3.1 *Raja clavata* - Análise exploratória

Esta secção é dedicada a uma análise exploratória das variáveis disponíveis para a *R. clavata*. Na Figura 3.1, apresentam-se as localizações dos lances efetuados no período em análise. De notar que algumas das localizações são comuns às quatro campanhas. Na Figura 3.2 distinguem-se as estações com capturas de raia-lenga e as estações sem captura desta espécie.

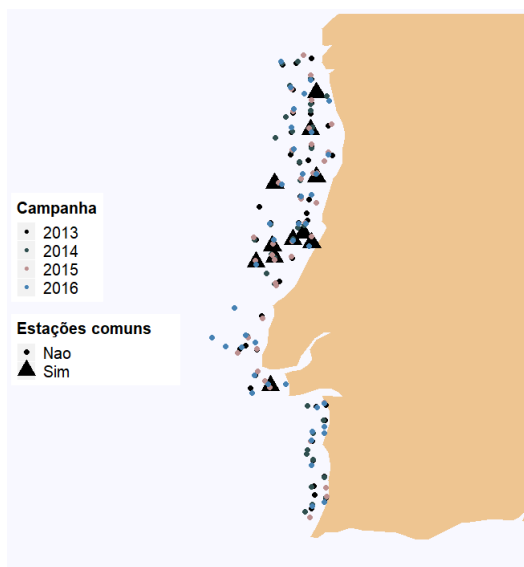


Figura 3.1: Localizações das realizações por campanha e por tipo de estação.

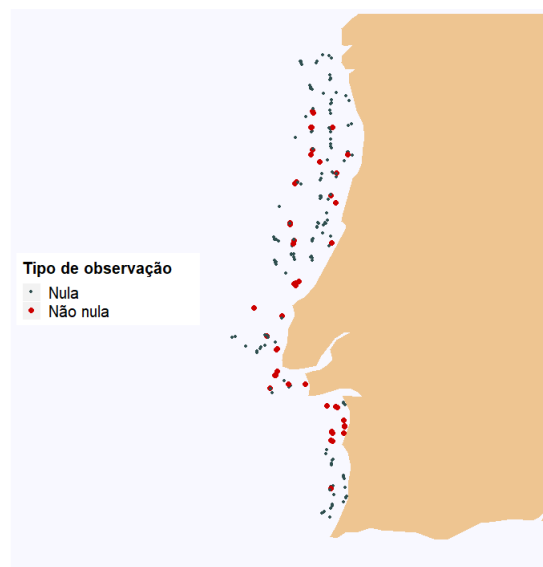


Figura 3.2: Localizações de observações nulas e não-nulas.

Em 2013, foram efetuados 57 lances e em 44 não se pescou qualquer elemento desta espécie. Em cerca de 81,6% dos 49 lances efetuados em 2014 não foi capturada raia-lenga. Das 54 localizações registadas na campanha de 2015, em apenas 17 existem capturas não nulas da espécie. Na campanha realizada, em 2016, verificou-se uma taxa de valores nulos de cerca de 77%. A Tabela 3.1 apresenta um sumário da percentagem de observações não-nulas ao longo das várias campanhas.

3.1.1 Indicador de abundância

A Figura 3.3 apresenta a distribuição espacial da abundância (número por hora) de *R. clavata* em cada uma das campanhas realizadas no período de 2013 a 2016. Da sua análise verifica-se maior ocorrência de raia-lenga na região de Lisboa (LIS) e arredores, em particular nas duas últimas campanhas.

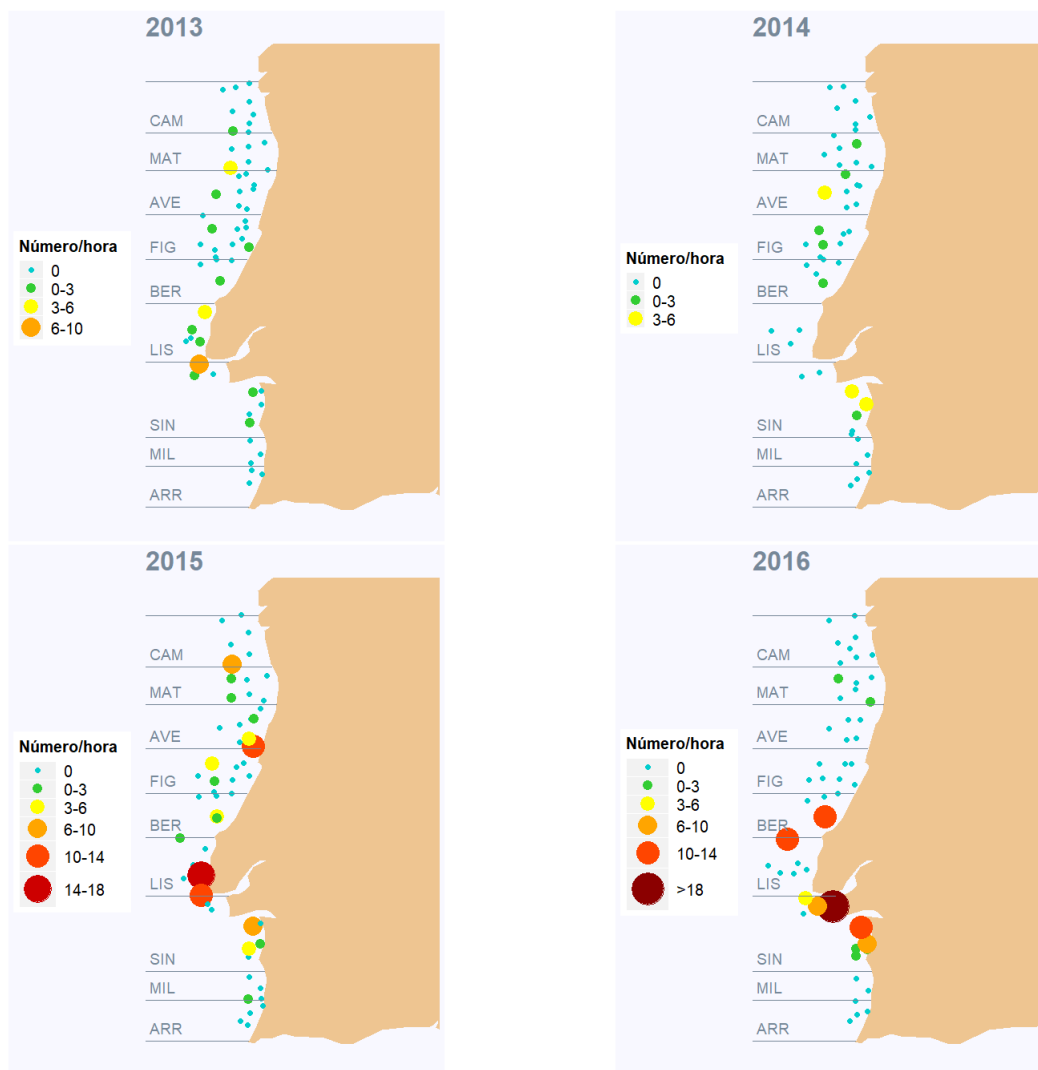


Figura 3.3: Número de elementos capturados por hora de raia-lenga por campanha.

Tabela 3.1: Sumário do número de indivíduos capturados por hora de raia-lenga por campanha.

Campanha	2013	2014	2015	2016	2013-2016
Mínimo	0	0	0	0	0
1º Quartil	0	0	0	0	0
Mediana	0	0	0	0	0
Média	0,69	0,57	1,70	1,73	1,18
3º Quartil	0	0	2	0	0
Máximo	8,18	6	18	23,33	23,33
Desvio Padrão	1,56	1,34	3,71	4,40	3,08
Coeficiente de Variação	227,55	234,24	218,02	253,67	262,00
Número de observações	57	49	54	52	212
Percentagem de observações não-nulas	22,81%	18,37%	31,48%	23,08%	24,06%

Na campanha referente a 2016 foi capturado o maior número de raia-lenga nos quatro anos de estudo.

As principais estatísticas descritivas do número total de raia-lenga capturada por hora encontram-se na Tabela 3.1. Pela consulta desta tabela verifica-se que em média, por cada lance, foram capturados menos de dois indivíduos da espécie; o valor mais baixo da média foi registado em 2014 (aproximadamente 0,6 indivíduos por hora) apesar da percentagem de estações não nulas não ter sido a mais elevada.

O número total de raia-lenga capturada por hora varia ao longo da costa. De um modo geral, os valores da variável número de indivíduos capturados por hora e a particularidade desta espécie ser pouco frequentemente encontrada nas campanhas demersais, já permitem antever dificuldades na modelação da abundância da raia-lenga.

A aplicação do Teste de *Mood* (apresentado na secção 2.2.1), neste estudo, surge para comparar a variância entre os nove setores representados na Figura A.1 do Apêndice A. De forma a que os resultados da aplicação do teste sejam consistentes, é necessário reunir um número razoável de observações em cada estrato e, para isso procedeu-se à agregação dos dados referentes às quatro campanhas demersais IPMA. Esta decisão, realizada por conveniência neste capítulo, será formalmente justificada no Capítulo 4. Pela consulta da Tabela 3.2 verifica-se que a zona de Sines (setor SIN) apresenta uma maior variabilidade ao longo dos quatro anos. Por outro lado, na zona de Arrifana (ARR) tendo em conta a média e a variância do número de elementos capturados de raia-lenga, confirma-se que nunca se pescou raia-lenga ao longo das várias campanhas. A aplicação do Teste de *Mood*, para

Tabela 3.2: Média e variância do número de indivíduos capturados por hora de raia-lenga em cada setor.

Medida	Setor								
	CAM	MAT	AVE	FIG	BER	LIS	SIN	MIL	ARR
Média	0,32	0,58	1,00	0,42	1,14	2,26	3,59	0,15	0,00
Variância	2,16	1,23	6,39	0,94	6,69	23,54	27,34	0,31	0,00
Número de observações	31	25	27	33	18	19	33	13	13

um nível de confiança de 95%, permitiu inferir que existe evidência para que o setor SIN apresente uma variabilidade diferente à dos setores MAT, CAM e FIG, que correspondem às zonas de Matosinhos (MAT), Caminha (CAM) e Figueira da Foz (FIG). A variância do número de indivíduos capturados de *R. clavata* por hora no setor correspondente à zona de Lisboa é significativamente diferente à da abundância capturada nos setores CAM e ARR, para um nível de confiança de 95%. O sumário dos resultados dos vários testes de homogeneidade de variâncias é apresentado na Tabela B.1 do Apêndice B.

3.1.2 Indicador de biomassa

O indicador de biomassa, tal como referido na Introdução desta dissertação, é dado pelo peso dos elementos capturados de raia-lenga (em Kg) por hora de arrasto na costa continental portuguesa.

A distribuição espacial dos indicadores de abundância e de biomassa (Figuras 3.3 e 3.4) são idênticas, confirmando-se que na zona de Lisboa e arredores existe uma maior ocorrência de raia-lenga e que o ano em que se verificou uma menor abundância/biomassa foi 2014.

As principais estatísticas descritivas para o indicador de biomassa de raia-lenga encontram-se na Tabela 3.3.

Tabela 3.3: Sumário do peso dos elementos capturados de raia-lenga por hora.

Campanha	2013	2014	2015	2016	2013-2016
Mínimo	0	0	0	0	0
1º Quartil	0	0	0	0	0
Mediana	0	0	0	0	0
Média	1,29	1,02	2,54	3,15	2,00
3º Quartil	0	0	1,99	0	0
Máximo	15,93	14,20	29,30	46,17	46,17
Desvio Padrão	3,38	2,86	5,81	8,15	5,49
Coefficiente de Variação	262,30	281,14	229,01	258,54	274,62
Número de observações	57	49	54	52	212

O valor médio do indicador de biomassa de raia-lenga capturada foi superior em 2016 e inferior em 2014. Em 2015, em cerca de 75% dos lances pescou-se até 1,99 Kg de raia-lenga, enquanto que nas restantes campanhas não se capturou qualquer elemento desta espécie em pelo menos 75% das localizações (ver Tabela 3.3).

Tabela 3.4: Média e variância do peso dos elementos capturados por hora de raia-lenga em cada setor.

Medida	Setor								
	CAM	MAT	AVE	FIG	BER	LIS	SIN	MIL	ARR
Média	0,62	1,20	2,42	0,70	1,26	2,56	6,52	0,01	0,00
Variância	7,63	6,68	38,60	5,05	12,85	31,05	97,18	$8,97 \times 10^{-4}$	0,00
Número de observações	31	25	27	33	18	19	33	13	13

De forma análoga ao efetuado para o indicador de abundância, prossegue-se de seguida com a análise de variância entre setores para o indicador de biomassa. Em conformidade

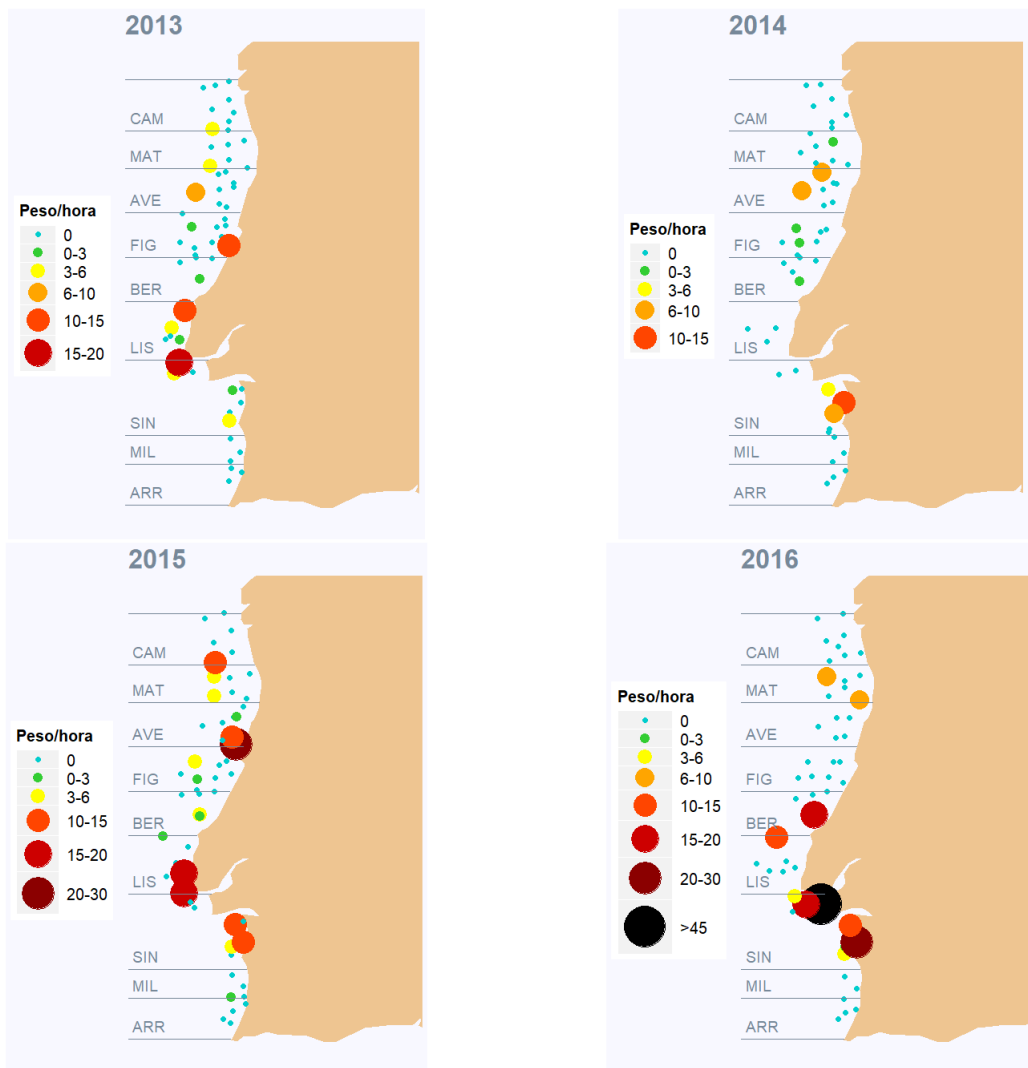


Figura 3.4: Peso dos indivíduos capturados de raia-lenga por campanha.

com os resultados da Tabela 3.2, verifica-se pela Tabela 3.4 que o setor onde se observou uma maior variabilidade do indicador de biomassa de raia-lenga foi associado à área de Sines (SIN) seguido dos setores de Aveiro (AVE) e Lisboa, respetivamente. Ao nível de 95% de confiança, verifica-se que os resultados das análises de variância por setor dos indicadores de abundância e biomassa da raia-lenga estão em concordância. Dado que a abundância e o indicador de biomassa estão relacionados, este facto seria expectável (detalhes apresentados na Tabela B.2 do Apêndice B).

3.1.3 Comprimento médio

O comprimento médio representa o valor médio do comprimento de todos os elementos da espécie *R. clavata* capturados em cada localização. A distribuição espacial por ano desta variável está representada na Figura 3.5.

Nos três últimos anos verifica-se que na zona de Sines o comprimento médio da espécie

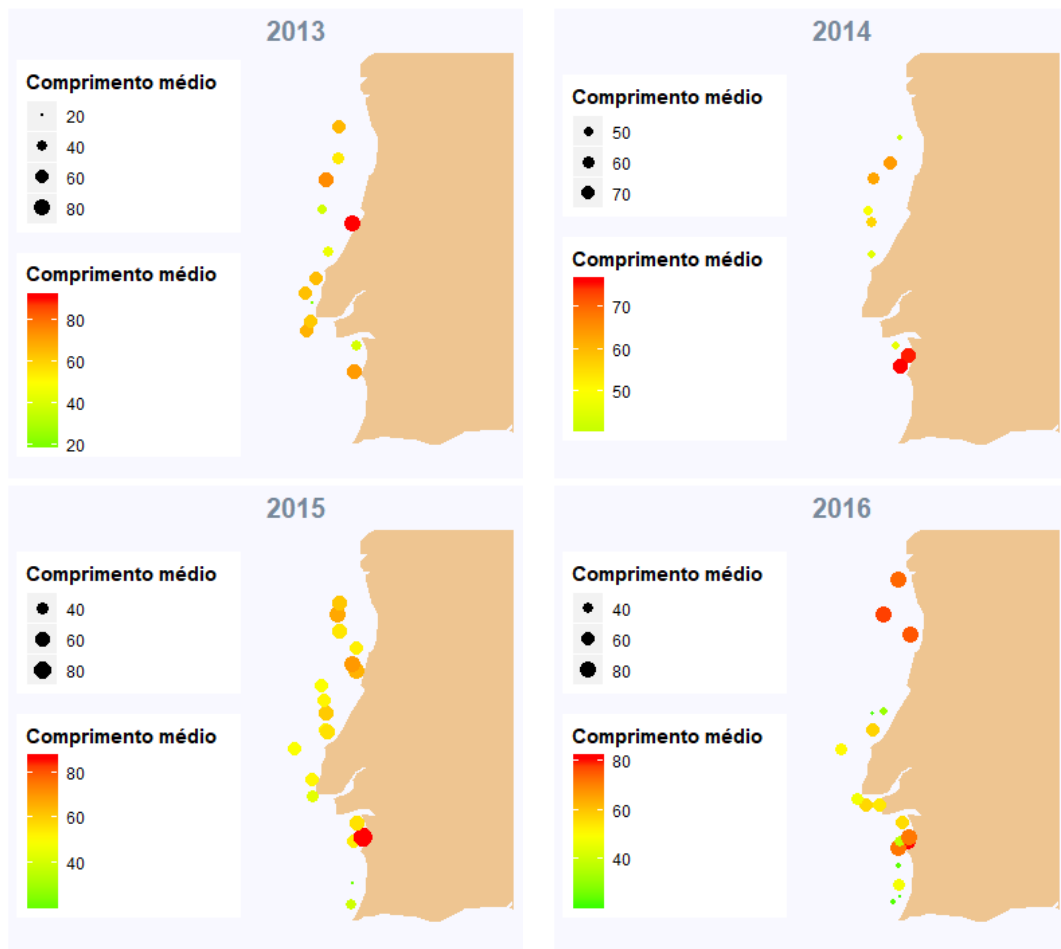


Figura 3.5: Comprimento médio de raia-lenga capturada por campanha.

tem apresentado valores mais altos.

Embora se tenha acesso a esta informação, esta não vai ser utilizada para a estimação dos indicadores de abundância dado que esta é obtida como consequência da abundância da espécie.

3.2 *Merluccius merluccius* - Análise Exploratória

As informações de pesca referem-se a campanhas demersais do IPMA de 2015 e 2016, e estão separados em dois grupos de comprimento total dos indivíduos: indivíduos com comprimento total inferior a 26 cm, isto é pescada pequena ou juvenis de pesca e indivíduos com comprimento superior a 26 cm, isto é indivíduos adultos.

Em 2015 foram realizados 53 lances em localizações distintas, dos quais em 50 se verificou a ocorrência de pesca. No mesmo ano, não se capturou pescada pequena em quatro das 54 localizações tentadas. Em 2016 cerca de 8% dos 48 lances realizados não se capturou pescada e em mais de 82% das 47 localizações observou-se a ocorrência de pescada pequena.

3.2.1 Indicador de abundância dos adultos de pescada

A Figura 3.6 contém os mapas da distribuição espacial do indicador de abundância de pescada em cada uma das campanhas demersais do IPMA. Em 2015 observou-se uma maior ocorrência de pescada nomeadamente na zona norte da costa, no entanto a diferença entre zonas, relativamente ao indicador de abundância, não é tão nítida na campanha de 2016.

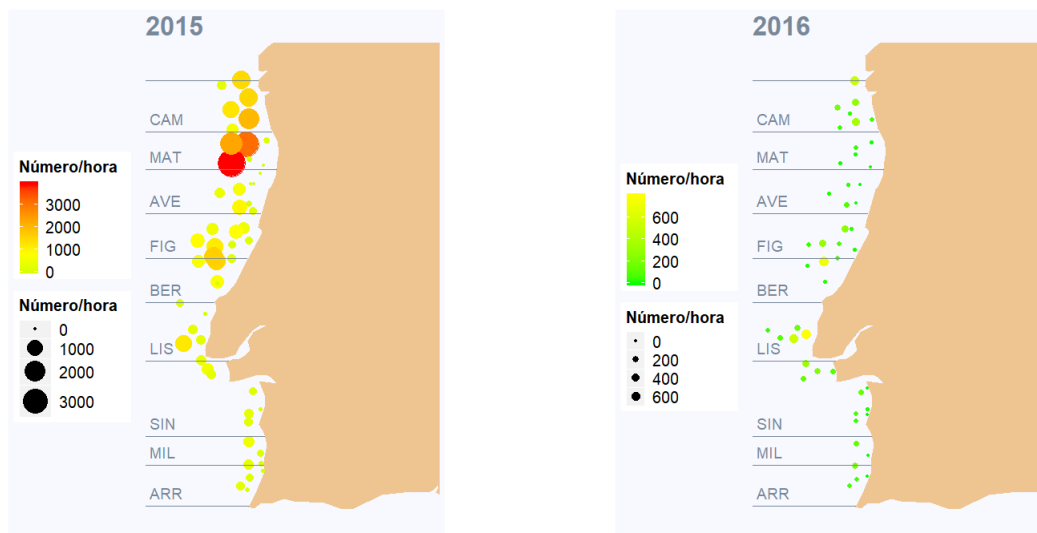


Figura 3.6: Número de indivíduos adultos de pescada capturados por hora por campanha.

Na Tabela 3.5 apresentam-se estatísticas descritivas do indicador de abundância. Em média, no ano de 2015, por cada lance capturaram-se cerca de 609 pescadas por hora. Porém, no ano seguinte este valor decresceu para menos de um quarto.

Tabela 3.5: Sumário do número de indivíduos adultos de pescada capturados por hora.

Campanha	2015	2016	2015-2016
Mínimo	0	0	0
1º Quartil	103,20	15,43	28,57
Mediana	272,00	56,00	136,00
Média	609,27	131,33	382,13
3º Quartil	770,09	164,00	450,00
Máximo	3933,46	795,40	3933,46
Desvio Padrão	811,84	184,89	645,23
Coefficiente de Variação	133,25	140,78	168,85
Número de observações	53	48	101

De forma a compreender melhor a variabilidade da abundância de pescada ao longo da costa, recorre-se tal como na Secção 3.1 a um teste de hipóteses para a homogeneidade

de variâncias, o Teste de *Mood*. Mais uma vez procede-se à agregação das campanhas de modo a obter um número razoável de observações por setor. Pela análise dos resultados

Tabela 3.6: Média e variância do número de indivíduos adultos capturados por hora capturada em cada setor.

Medida		Média	Variância	Número de observações
Setor	CAM	650,61	436692,60	13
	MAT	870,06	2237352,71	11
	AVE	174,21	86148,42	13
	FIG	453,35	233770,07	14
	BER	431,86	262062,56	9
	LIS	350,16	146139,99	10
	SIN	165,14	22753,99	17
	MIL	172,17	25626,87	7
	ARR	79,14	7235,81	7

sumarizados na Tabela 3.6, verifica-se que em média se capturou mais pescada na zona de Matosinhos (setor MAT), que por sua vez é também o setor que apresenta uma variabilidade superior. Ao nível de confiança de 95%, os resultados do teste para a igualdade de variâncias indicam que a variabilidade do setor MAT é significativamente diferente à da zona de Lisboa (setor LIS) e aos setores que compõem a zona norte exceto o setor de Aveiro. Informação mais detalhada sobre os resultados desta análise de variância é apresentada na Tabela B.3.

3.2.2 Indicador da abundância dos juvenis de pescada

A distribuição espacial da abundância de pescada pequena por ano está apresentada na Figura 3.7. Quando comparada com a distribuição espacial anual do número de indivíduos adultos (Figura 3.6), verifica-se que estas são idênticas. Tal como para a pescada (indivíduos adultos), a maior ocorrência de pescada pequena observou-se em 2015 nomeadamente na zona norte. Na campanha de 2015, a média do indicador de abundância de indivíduos juvenis em cada lance foi de cerca de 551 elementos, valor inferior à abundância média de pescada capturada em igual período (ver Tabela 3.7).

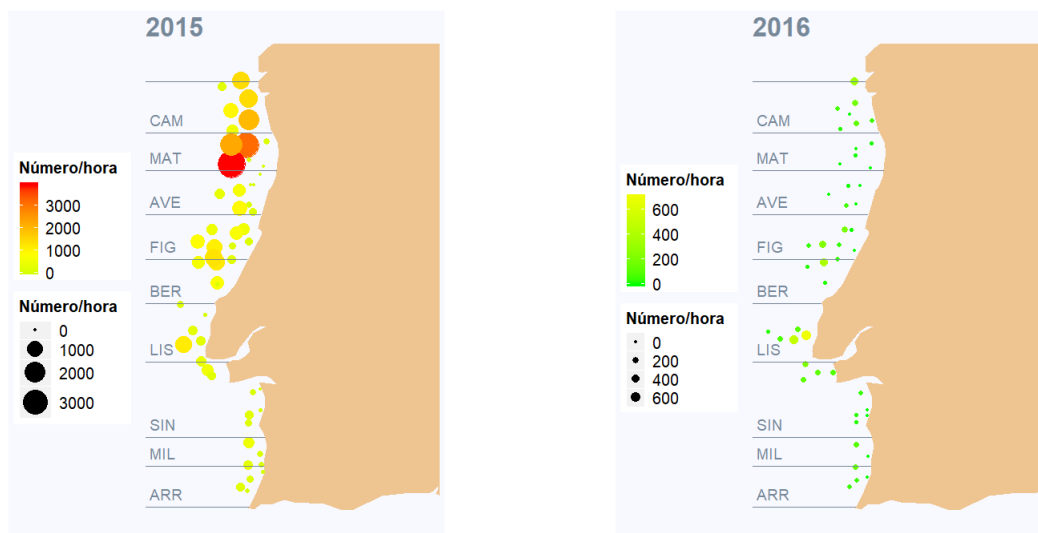


Figura 3.7: Número de indivíduos juvenis de pescada capturados por hora por campanha.

Tabela 3.7: Sumário do número de indivíduos juvenis de pescada capturados por hora por campanha.

Campanha	2015	2016	2015-2016
Mínimo	0	0	0
1º Quartil	71,250	4,500	15,652
Mediana	231,000	36,000	108,000
Média	551,129	88,814	335,993
3º Quartil	670,9815	113,000	336,000
Máximo	3903,456	705,055	3903,456
Desvio Padrão	784,404	141,212	623,687
Coefficiente de Variação	142,327	158,998	185,625
Número de observações	54	47	101

Apesar do valor máximo da abundância da pescada juvenil ser inferior ao do número total máximo de indivíduos adultos capturados, estes ocorrem no mesmo local.

Tabela 3.8: Média e variância do número de indivíduos juvenis de pescada capturados por hora em cada setor.

Medida		Média	Variância	Número de observações
Setor	CAM	551,55	428624,46	13
	MAT	854,81	2207069,93	11
	AVE	158,26	78167,82	13
	FIG	389,73	170841,43	14
	BER	341,74	161769,08	9
	LIS	317,30	130648,49	10
	SIN	123,77	20811,44	17
	MIL	149,71	23845,91	7
	ARR	64,00	5510,67	7

A par da análise da distribuição é importante realizar uma análise de variância entre setores, de forma a perceber se a variabilidade da abundância de pescada pequena difere ao longo da costa, tal como ocorre com a abundância de pescada adulta. Pela Tabela 3.8 verifica-se que o setor MAT apresenta maior variabilidade em relação aos restantes, tal como se verificou para a abundância de pescada capturada. Os resultados do Teste de *Mood*, apresentados na Tabela B.4, estão em conformidade com os resultados da aplicação deste teste aos dados do número de indivíduos adultos capturados de pescada. Este facto indica que, apesar das variabilidades do número de indivíduos adultos e do número de indivíduos juvenis serem diferentes ao longo dos setores, os resultados obtidos são semelhantes independentemente do comprimento da espécie.

3.3 Análise exploratória das covariáveis

As variáveis auxiliares para modelar os indicadores de abundância e biomassa da *R. clavata* consideradas nos Capítulo 5 são a profundidade, o tipo de fundo e a abundância de pescada pequena. A profundidade é também utilizada para modelar o indicador de abundância de pescada.

As estimativas das covariáveis profundidade e abundância de elementos de pescada juvenis na grelha de predição foram obtidas a partir de uma abordagem não paramétrica (NP), a suavização de *kernel* espacial, que será abordada no Capítulo 5. Relativamente à covariável de tipo de fundo possui-se informação em todos os pontos da grelha de predição utilizada na abordagem *kriging* com tendência externa.

3.3.1 Profundidade

A profundidade corresponde à distância, em metros, entre o nível do mar à superfície e o fundo oceânico. Das localizações referentes às estações de pesca onde se tencionou capturar elementos de raia-lenga, a profundidade varia entre 23 metros e 245 metros.

A distribuição espacial da profundidade na região de estudo está representada através do mapa da Figura 3.8.

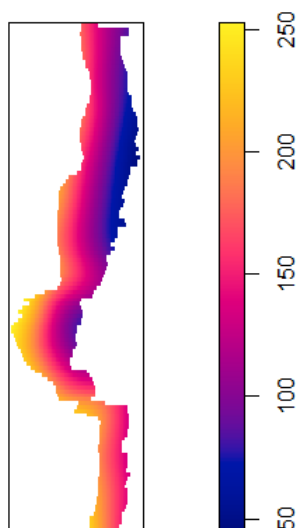


Figura 3.8: Mapa da profundidade em metros na região de estudo.

3.3.2 Tipo de fundo

A variável tipo de fundo indica qual o tipo de fundo oceânico em cada localização. Na Tabela 3.9 está apresentado o número de observações em cada tipo de fundo por campanha, e quantas destas são observações não-nulas. O tipo de fundo “lodo com areia” é aquele em que, em todas as campanhas, o número de valores não-nulos é maior a par do número de observações.

Tabela 3.9: Número de observações por tipo de fundo em cada campanha (entre parênteses estão apresentados o número de observações não nulas).

Campanha	Tipo de fundo					
	Mistura de Sedimentos	Lama	Lodo com Areia	Rocha	Areia	Areia Barrenta
2013	9 (2)	3 (1)	18 (5)	1 (1)	21 (3)	5 (1)
2014	7 (0)	3 (0)	23 (6)	3 (1)	11 (2)	2 (0)
2015	12 (2)	3 (0)	15 (8)	4 (2)	14 (4)	6 (1)
2016	8 (3)	3 (2)	17 (4)	7 (1)	15 (2)	2 (0)
2013-2016	36 (7)	12 (3)	73 (23)	15 (5)	61 (11)	15 (2)

A Figura 3.9 contém o mapa do tipo de fundo na costa continental portuguesa. Pela consulta deste, não é visível nenhum padrão espacial do tipo de fundo ao longo da costa. A vermelho destacam-se as zonas onde não se possui informação sobre esta característica do fundo oceânico na costa continental portuguesa.

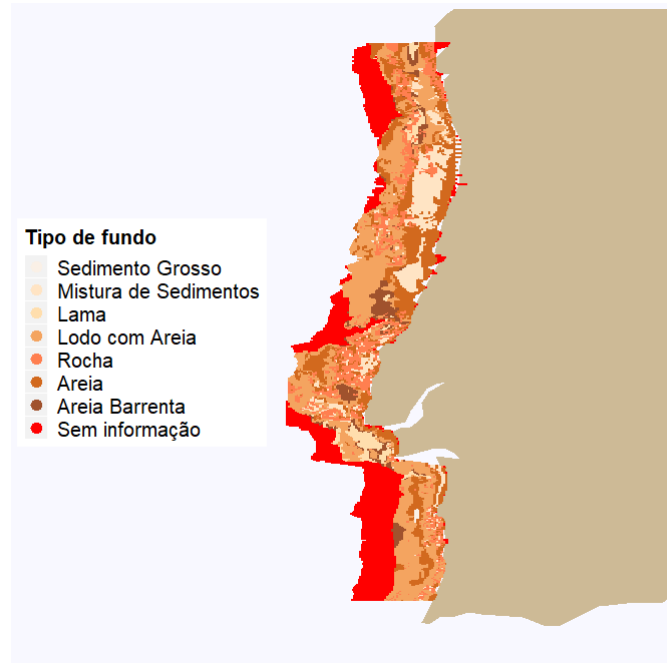


Figura 3.9: Mapa do tipo de fundo na costa continental portuguesa (dados obtidos pela EMODnet).

3.3.3 Conclusões

A análise realizada no Capítulo 3 possibilitou verificar que a abundância da raia-lenga e da pescada são de ordens diferentes. Enquanto que num lance se capturou mais de 3000 pescadas, apenas se conseguiu pescar, no máximo, cerca de 23 raias-lenga por hora. É também relevante perceber que a variabilidade de abundância de ambas as espécies é diferente ao longo da costa, facto verificado pela análise de variância entre setores. Adicionalmente, o pequeno número de observações não nulas para a *R. clavata* faz antever dificuldades na modelação da abundância desta espécie, que deverão ser tidas em conta no desenvolvimento do trabalho dos capítulos seguintes.

Capítulo 4

Análise da distribuição espacial da *Raja clavata*

O estudo geoestatístico dos indicadores de abundância e de biomassa implica o conhecimento das respectivas medidas de continuidade espacial com o intuito de caracterizar a sua distribuição e/ou variabilidade espacial. Considerando que $Z(\mathbf{s})$ é uma variável aleatória que apresenta uma única realização do indicador para cada $\mathbf{s} \in \mathbb{R}^2$, dado que trabalhar com realizações espaciais impossibilita a repetição de várias observações numa mesma localização. Contudo, para ser possível a inferência a partir deste tipo de amostras, existem requisitos que devem ser cumpridos: a estacionaridade e a isotropia. Após a sua verificação, é imperativo estudar a estrutura espacial das variáveis aleatórias pelo cálculo do variograma empírico. Através dos métodos *kriging* ordinário ou *kriging* da indicatriz e depois de ajustar um modelo teórico ao variograma empírico, realiza-se a interpolação da variável a toda a região de estudo. Para o cálculo das distâncias subjacentes ao cálculo do variograma, as coordenadas geográficas foram convertidas em coordenadas UTM. Transversa de Mercator Universal (UTM) é um sistema de coordenadas geográficas que é usado para identificar localizações em metros. No entanto, uma vez que a região em estudo é grande, as coordenadas a utilizar são apresentadas em Km.

Na Secção 4.1, as distribuições espaciais da variável nas quatro campanhas são analisadas separadamente, de forma a estudar a estrutura espacial associada a cada campanha através do respetivo variograma. Na Secção 4.2, a estrutura espacial é analisada através de um variograma médio comum às quatro campanhas, determinado, para cada distância, pela média dos variogramas individuais ponderada pelo número de pares. Esta análise permite compreender se a agregação das campanhas é plausível uma vez que o número de observações por ano é reduzido. Esta hipótese acaba por ser corroborada pela análise de interação espaço-tempo realizada na Secção 4.3.

4.1 Análise da estrutura espacial pelo variograma individual

Aplica-se o método de *kriging* ordinário aos valores de indicadores de abundância e de biomassa da *R. clavata*. O método de *kriging* de indicatriz é aplicado aos dados de presença/ausência de raia-lenga.

Para o indicador de abundância de raia-lenga, a apresentação sobre a aplicação do procedimento referido é mais detalhada, incluindo uma breve explicação do procedimento estatístico do indicador em causa e apresentando os mapas de predição e respetiva interpretação. A apresentação das Tabelas, que contêm as estimativas dos vários parâmetros, e dos gráficos dos variogramas empíricos e teóricos relativos ao peso total por hora, e à presença de raia-lenga, é feita no Apêndice C.

Os variogramas teóricos são ajustados aos variogramas empíricos tendo em conta várias componentes referentes à estrutura de correlação espacial (Modelo Esférico, Modelo Gaussiano e Modelo Exponencial). Deste modo, o variograma teórico será definido pela combinação linear destas componentes que minimiza a distância entre os variogramas empírico e teórico, pelo método dos mínimos quadrados.

4.1.1 Indicador de abundância de *Raja clavata*

A análise geoestatística inicia-se com a estimação do valor esperado, ou tendência, do processo estocástico espacial $E[Z(\mathbf{s})] = \mu(\mathbf{s})$. A tendência espacial $\mu(\mathbf{s})$ para todas as campanhas é modelada através de um modelo linear generalizado, onde se assume que a variável resposta segue uma distribuição Tweedie com função ligação $\log(\mu(\mathbf{s})) = \beta_0 + \beta_1 \times \text{Latitude}(\mathbf{s})$, dado que a amplitude relativa à longitude é reduzida. As estimativas destes parâmetros β_i , $i = 0, 1$, o respetivo desvio padrão $s(\beta_i)$ e o valor de prova associado $P(\beta_i)$ estão apresentados na Tabela 4.1.

Tabela 4.1: Estimativas dos parâmetros envolvidos nas estruturas do valor médio da abundância de raia-lenga.

Campanha	Tendência					
	$\hat{\beta}_0$	$\hat{s}(\beta_0)$	$P(\hat{\beta}_0)$	$\hat{\beta}_1$	$\hat{s}(\beta_1)$	$P(\hat{\beta}_1)$
2013	19,61	9,47	0,04	-0,005	0,002	0,04
2014	11,61	9,81	0,24	-0,003	0,002	0,22
2015	7,05	8,68	0,42	-0,001	0,002	0,46
2016	42,67	10,53	0,000	-0,010	0,002	0,000

Para o nível de confiança de 95%, apenas os parâmetros associados à latitude das campanhas de 2013 e 2016 são significativos, pelo que para as restantes não se verifica a existência de tendência espacial. Uma vez estimadas as componentes relativas à tendência espacial e à anisotropia, procede-se à análise da correlação espacial através do cálculo dos

variogramas empíricos, que para a primeira e última campanhas em estudos foi determinado à custa dos resíduos: $\delta(\mathbf{s}) = Z(\mathbf{s}) - \hat{\mu}(\mathbf{s})$.

Na Tabela 4.2 estão apresentados os valores dos parâmetros considerados no cálculo dos variogramas, mais concretamente as estimativas da variância do indicador de abundância σ^2 , da variância associada aos erros de medição τ^2 e do raio de influência ϕ . A Figura 4.1 contém os variogramas associados a cada campanha.

Tabela 4.2: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para a abundância de raia-lenga capturada por hora em cada campanha.

Campanha	Componente	Parâmetros		
		$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
2013	Esférica	0	1,38	16,86
	Gaussiana	0	1,21	126,46
2014	Esférica	0	1,52	56,16
	Gaussiana	0	0,44	48,85
2015	Exponencial	0	8,55	14,89
	Esférica	0	9,41	170,63
2016	Gaussiana	6,94	19,88	75,32

Na interpretação dos resultados destaca-se a ordem de grandeza distinta para a variância total do processo estocástico espacial nas duas primeiras campanhas comparativamente às duas últimas. Tal implica que a estimação da variograma médio, proposta na secção 4.2, só fará sentido após a standardização da variância.

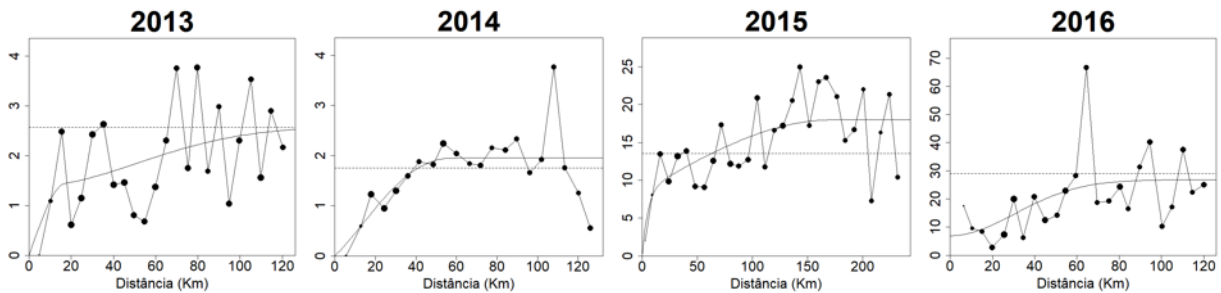


Figura 4.1: Variogramas amostrais para o indicador de abundância de raia-lenga em cada campanha.

Após a determinação dos variogramas teóricos ajustados aos variogramas empíricos por método dos mínimos quadrados, o próximo objetivo passa por estimar a abundância por hora de raia-lenga em localizações não observadas. A interpolação da abundância de raia-lenga para a costa continental portuguesa é efetuada através de *kriging* ordinário, cujas superfícies preditas se encontram na Figura 4.2.

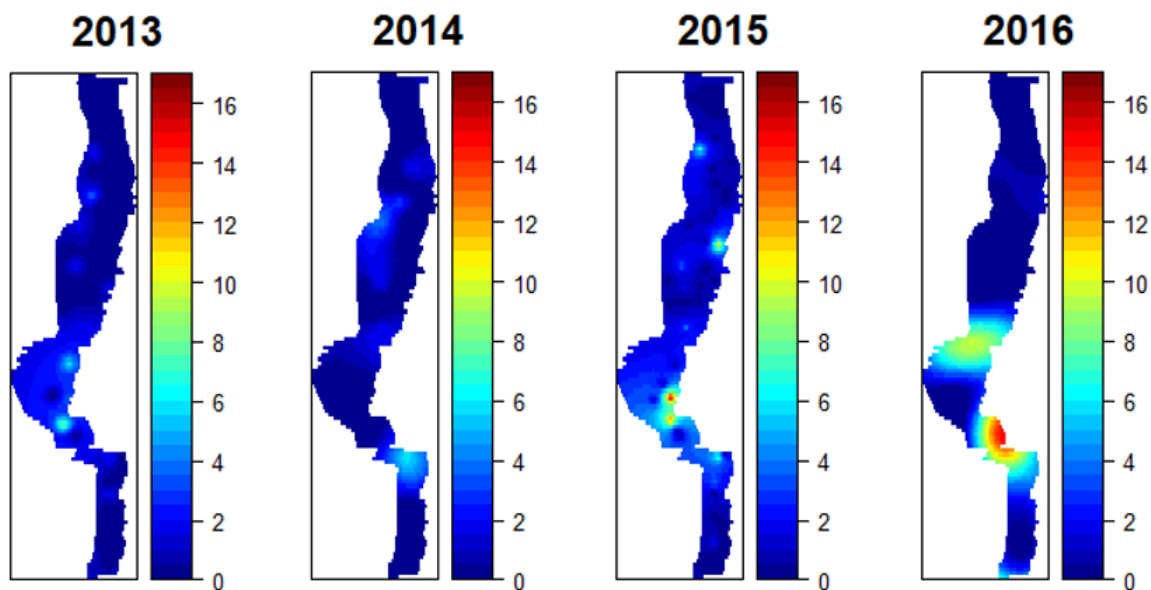


Figura 4.2: Superfícies previstas da distribuição espacial do indicador de abundância de raia-lenga em cada campanha.

Através da análise da Figura 4.2 verifica-se que na zona de Lisboa e arredores a abundância de raia-lenga é maior, embora este facto seja mais evidente para as campanhas realizadas em 2015 e 2016.

4.1.2 Indicador de biomassa de *Raja clavata*

Os processos espaciais relativos ao indicador de biomassa são estacionários na média para todas as campanhas à exceção da última, dado que a latitude apenas parece influenciar significativamente, ao nível de confiança de 95%, o indicador de biomassa observado no ano de 2016 (resultados da Tabela C.1). Os parâmetros estimados para o cálculo do variograma teórico ajustado pelo variograma empírico estão apresentados na Tabela C.2. A representação gráfica dos respetivos variogramas surgem na Figura C.1. Neste caso, a variância total do processo estocástico espacial associado à campanha de 2015 é a que apresenta um valor superior. De um modo geral, pela análise das estimativas para ϕ , apresentadas nas Tabelas 4.2 e C.2, pode-se concluir que a correlação espacial para o indicador de abundância permanece por mais alguns quilómetros do que para o da biomassa.

As superfícies previstas da distribuição espacial do indicador de biomassa da raia-lenga para cada ano estão apresentadas na Figura 4.3. A zona de Lisboa e arredores é das zonas com maior ocorrência desta espécie, o que está em conformidade com os resultados relativos ao indicador de abundância. No entanto, na campanha referente ao ano de 2015, verifica-se que a zona da Figueira da Foz apresenta valores do peso de elementos capturados por hora mais elevados.

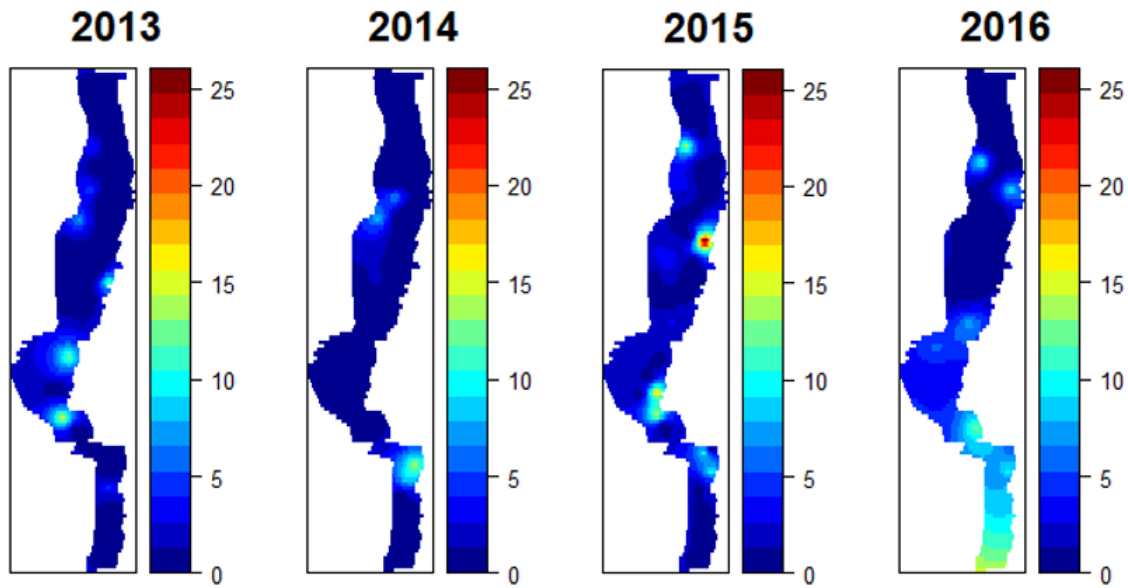


Figura 4.3: Superfícies previstas da distribuição espacial do indicador de biomassa de raia-lenga em cada campanha.

4.1.3 Indicador da presença de *Raja clavata*

Dado o elevado número de localizações em que não se capturou raia-lenga, analisou-se a ocorrência espacial. Considerando que os requisitos da estacionariedade na média e na variância estão verificadas, foram determinados os parâmetros subjacentes aos variogramas teóricos ajustados aos variogramas empíricos associados a cada campanha. A probabilidade de ocorrência de raia-lenga na região em estudo é realizada através de *Kriging* da Indicatriz.

A variável de interesse para as campanhas de 2013, 2014, 2015 e 2016 incluem 44, 40, 37 e 40 observações nulas e 13, 9, 17 e 12 não nulas, respectivamente. As estimativas dos parâmetros dos variogramas empíricos estão apresentados na Tabela C.3 e a representação dos variogramas empíricos e teóricos encontram-se na Figura C.2.

Na Figura 4.4 estão representadas as superfícies previstas da distribuição espacial da probabilidade de ocorrência de raia-lenga. A campanha de 2016 é aquela que apresenta valores de probabilidades mais extremos, isto é, verifica-se a existência de regiões associadas a probabilidade muito baixa mas também áreas em que a probabilidade de ocorrência é elevada. Em 2015, a superfície estimada é a mais suave das várias campanhas, sendo esta a única campanha que apresenta uma estimativa não-nula para o efeito pepita τ^2 (ver Tabela C.3). O tipo de superfície estimada é coerente com o número de observações nulas ser o menor, e o de não-nulas ser o maior, ao longo dos vários anos.

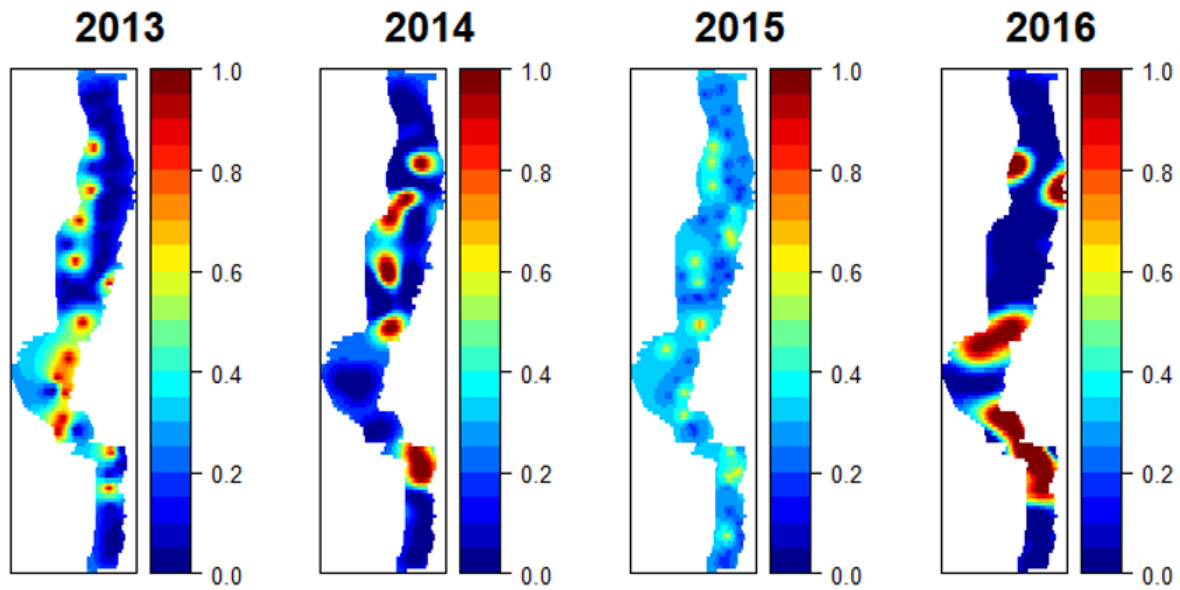


Figura 4.4: Superfícies preditas da distribuição espacial da ocorrência de raia-lenga em cada campanha.

4.2 Análise da estrutura espacial pelo variograma médio

O variograma médio é determinado, para cada distância, pela média dos variogramas individuais ponderado pelo número de pares. A variância dos dados é uma medida considerada no cálculo do variograma, como mencionado anteriormente para os variogramas de cada campanha. No entanto, esta medida pode ser diferente para os variogramas de diferentes campanhas e, por isso, os variogramas podem ser normalizados ou estandardizados (ou equivalentemente, a variável de estudo em cada campanha é estandardizada pelo seu desvio padrão). Nas Tabelas 4.3 e 4.4 estão apresentados, respectivamente, os resumos das principais estatísticas descritivas calculadas para o número e peso estandardizados dos indivíduos capturados de raia-lenga por campanha.

Tabela 4.3: Sumário do número estandardizado de indivíduos capturados por hora de raia-lenga.

Campanha	2013	2014	2015	2016	2013-2016
Mínimo	0	0	0	0	0
1° Quartil	0	0	0	0	0
Mediana	0	0	0	0	0
Média	0,44	0,43	0,46	0,40	0,44
3° Quartil	0	0	0,54	0	0
Máximo	5,28	4,53	4,89	5,36	5,36

Os variogramas empíricos e teórico relativos à variável número de raia-lenga capturada

Tabela 4.4: Sumário do peso estandardizado de elementos capturados por hora raia-lenga.

Campanha	2013	2014	2015	2016	2013-2016
Mínimo	0	0	0	0	0
1° Quartil	0	0	0	0	0
Mediana	0	0	0	0	0
Média	0,39	0,36	0,44	0,39	0,40
3° Quartil	0	0	0,35	0	0
Máximo	4,76	5,01	5,10	5,72	5,72

estandardizado encontram-se representados na Figura 4.5, respetivamente. Enquanto que na Figura 4.6 estão os variogramas resultantes da análise de associação espacial do peso de raia-lenga capturada.

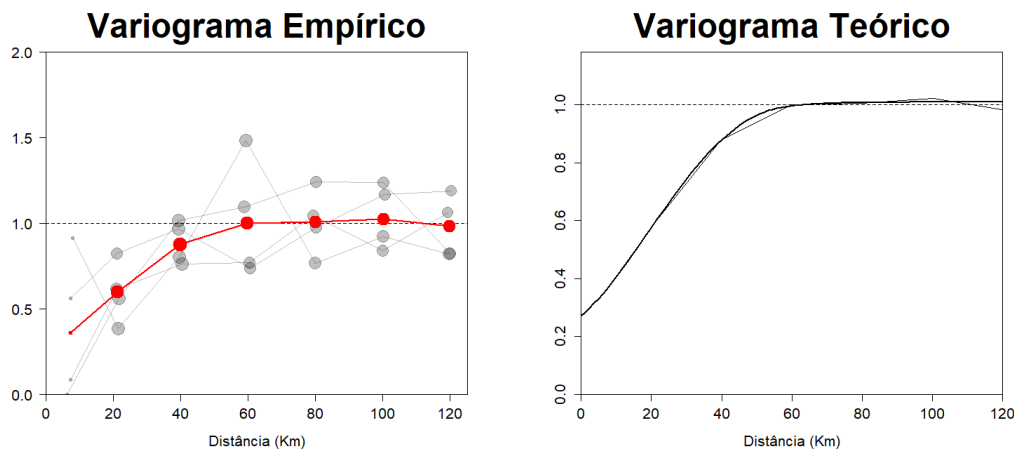


Figura 4.5: Abundância de raia-lenga capturada por hora. No painel da esquerda, variogramas empíricos individuais (linhas a cinzento) e médio (linha a vermelho); no painel da direita, variogramas empírico e teórico médios.

Nos variogramas empíricos (painéis da esquerda das Figuras 4.5 e 4.6), as linhas a cinzento representam os variogramas individuais de cada campanha e a vermelho destaca-se o variograma médio ponderado pelo número de pares. A dimensão dos pontos estão de acordo com o número de pares utilizado a determinada distância para o cálculo do variograma. Assim, quanto maior for a dimensão dos pontos, maior o número de pares considerados. A análise dos variogramas empíricos permite averiguar que os variogramas individuais são próximos do variograma médio, quer para a variável de interesse abundância como para a variável resposta peso dos elementos capturados. Dado o número elevado de observações nulas face ao número de realizações em cada campanha, a agregação das quatro campanhas em estudo é uma hipótese a ser considerada. Esta hipótese acaba por ser validada pelo cálculo do variograma médio, uma vez que a estrutura de associação espacial parece equivalente ao longo das campanhas.

Pela análise dos variogramas teóricos estimados para os indicadores de abundância e

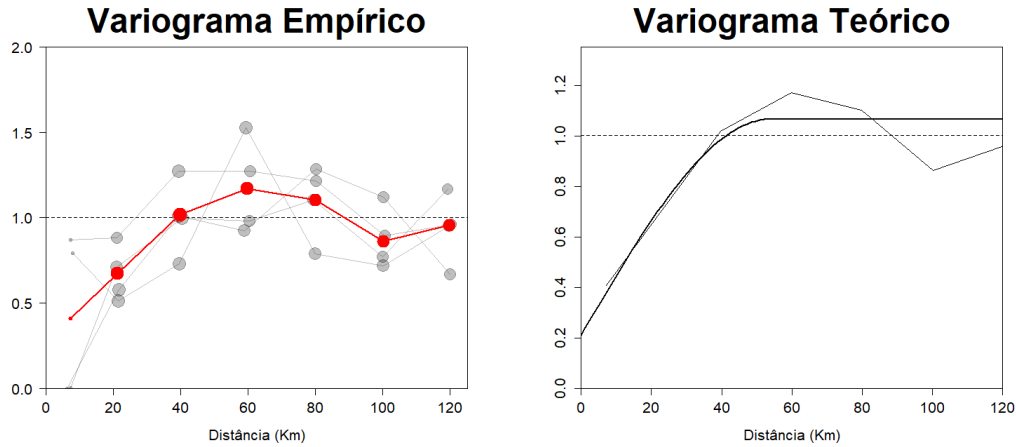


Figura 4.6: Biomassa da raia-lenga capturada por hora. No painel da esquerda, variogramas empíricos individuais (linhas a cinzento) e médio (linha a vermelho); no painel da direita, variogramas empírico e teórico médios.

biomassa (painéis da direita das Figuras 4.5 e 4.6), pode-se concluir que a correlação espacial se mantém para distâncias ligeiramente superiores no caso da abundância. Por outro lado, este caso também apresenta uma variância não-explicada (efeito pepita) superior ao caso da biomassa.

4.3 Interação espaço-tempo

Outra forma de perceber se a hipótese de agregação dos dados das quatro campanhas é plausível é testar a interação espaço-tempo. Os testes de hipóteses utilizados, nomeadamente o Teste de *Knox* e o teste de *Mantel* conforme discutido no Capítulo 2, verificam se os eventos próximos no tempo também tendem a ser próximos no espaço. A aplicação destes testes requer a escolha de distâncias “críticas”, quer para o espaço, quer para o tempo. A distância crítica estabelecida no espaço foi de 50 Km, tendo em conta os variogramas médios determinados anteriormente. No caso do tempo, considerou-se como distância crítica 1 ano, de forma a comparar as diversas campanhas. Em ambos os testes os valores de prova obtidos são altos, sendo pelo Teste de *Knox* 0,75 e pelo Teste de *Mantel* 1. Desta forma, não existe evidência estatística para rejeitar a hipótese nula, o que permite afirmar que a existência de interação espaço-tempo não é significativa e que é razoável a agregação das quatro campanhas.

4.3.1 Conclusões

A título de conclusão, as abordagens utilizadas no Capítulo 4 permitiram inferir sobre a distribuição espacial da raia-lenga e sobre a estrutura de correlação espacial inerente. Relativamente à estrutura de associação espacial, concluiu-se que não é significativa a

diferença entre as diversas campanhas em estudo. Deste modo, nos próximos capítulos será considerada a agregação dos dados referentes a todas as campanhas.

Capítulo 5

Estimação da abundância da *Raja clavata* e da *Merluccius merluccius*

No presente capítulo, ir-se-á prosseguir com a estimação da distribuição espacial da abundância da *R. clavata* na costa portuguesa, tirando partido dos dados ao longo das campanhas demersais realizados de 2013 a 2016. De modo semelhante, ir-se-á estimar a abundância da pescada, tendo em consideração os dados agregados de 2015 e 2016. A estimação da distribuição espacial de abundância e de biomassa da raia-lenga e da pescada deve-se ao facto de se pretender atualizar o desenho experimental de estações de pesca, com base nestas distribuições. Inicialmente, recorre-se a uma abordagem não-paramétrica, livre de pressupostos distribucionais sobre a população de onde os dados são provenientes. Trata-se de uma abordagem exploratória muito útil, que proporciona a construção simples de um mapa de distribuição que apenas considera os valores observados e a distância entre as observações. As abordagens não paramétricas apresentam vantagens, tais como não requererem modelos probabilísticos ou quaisquer hipóteses probabilísticas, como por exemplo a isotropia. Uma grande desvantagem prende-se com o facto de não permitirem uma quantificação fácil do erro associado à estimação [46].

Além da aplicação da suavização não-paramétrica, aplicar-se-ão outros métodos de estimação espacial nomeadamente o *kriging* com tendência externa e um modelo bayesiano geostatístico para zeros inflacionados. Em ambos os casos a variável de interesse $Z(\mathbf{s})$ é definida sobre uma região espacial $\mathbf{s} \in A \subset \mathbb{R}^2$, as observações são consideradas como realizações do processo nas localizações $\mathbf{s}_1, \dots, \mathbf{s}_n$.

Tal como no Capítulo 4, visando a aplicação das metodologias geoestatísticas referidas, as coordenadas geográficas foram convertidas em Km pelo sistema UTM.

5.1 Abordagem Não Paramétrica

O método não paramétrico utilizado na estimação da distribuição de abundância denomina-se por suavização *kernel* espacial e tem sido cada vez mais aplicado em diversas áreas científicas, nomeadamente na área da epidemiologia [47].

A suavização *kernel* espacial permite obter um mapa suavizado dos valores de densidade, de tal modo que cada localização reflete a concentração de pontos na área circundante. Os eventos dentro da área circundante são pesados de acordo com a sua distância ao centro da área. Eventos localizados perto do centro apresentam pesos mais elevados do que os mais distantes [48]. Este processo de suavização, também conhecido pelo método Nadaraya-Watson, pode ser representada por:

$$\hat{z}(s_0) = \frac{\sum_i K(s_0 - s_i) z(s_i)}{\sum_i K(s_0 - s_i) h}$$

onde \hat{z} identifica o valor estimado em \mathbf{s}_0 e h representa a *bandwidth* que define a área circundante. Note-se que $z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)$ são os valores observados nas localizações $\mathbf{s}_1, \dots, \mathbf{s}_n$ e K uma densidade de probabilidade (por exemplo, uma função *kernel* Gaussiana).

As superfícies estimadas da distribuição espacial dos indicadores de abundância e biomassa de raia-lenga estão apresentadas nas Figuras 5.1a e 5.1b, respetivamente. Na Figura 5.1c observa-se o mapa da distribuição espacial do indicador de abundância de pescada.

Em conformidade com os resultados apresentados na Secção 4.1 e pela análise das Figuras 5.1b e 5.1a, verifica-se que os indicadores de abundância de raia-lenga são mais elevados na zona de Lisboa e áreas adjacentes, em particular Sesimbra e Sines, e Peniche.

Da análise da Figura 5.1c, observa-se maior abundância de pescada na zona norte da costa continental portuguesa.

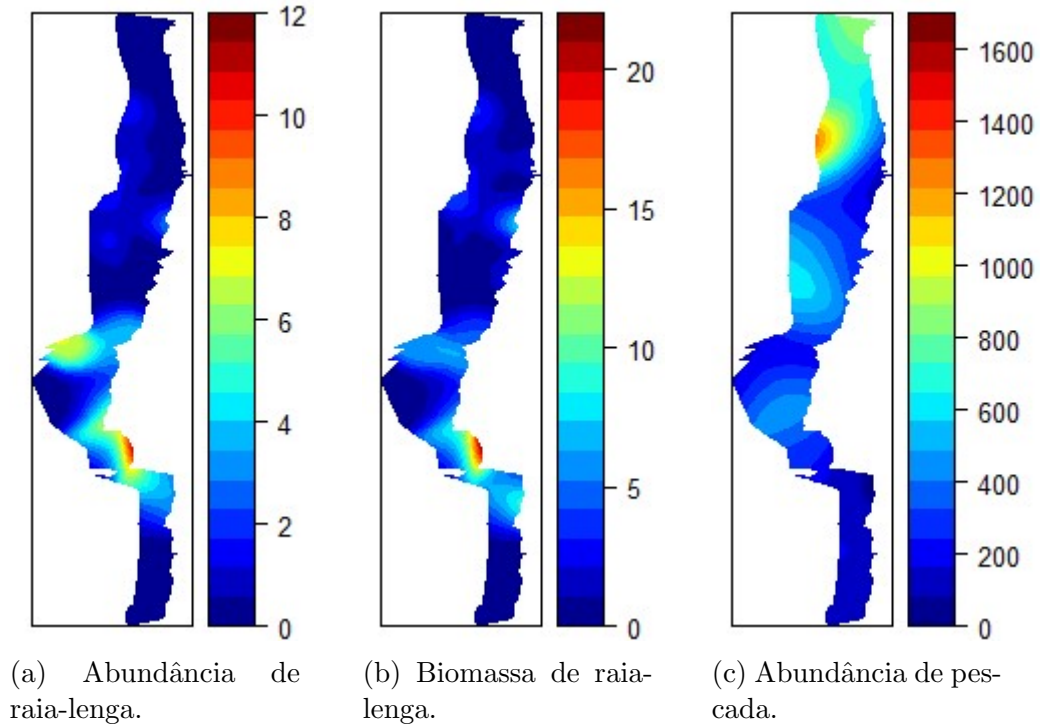


Figura 5.1: Superfícies previstas por abordagem NP das distribuições dos indicadores de **abundância** (painel da esquerda) e **biomassa** (painel central) de **raia-lenga** e do indicador de **abundância** de **pescada**.

5.2 *Kriging* com Tendência Externa

Esta metodologia será utilizada para estimar a distribuição espacial dos indicadores de abundância das duas espécies em estudo. Como referido na Secção 2.3, no método de *kriging* com tendência externa, a média da variável resposta é estimada recorrendo a variáveis explicativas.

As três variáveis resposta em estudo incluem dois indicadores de abundância e 1 indicador de biomassa. O *kriging* com tendência externa consiste na estimação dos coeficientes de regressão de forma a modelar a tendência, seguido do cálculo do variograma empírico à custa dos resíduos, recorrentes da modelação da tendência, e o ajuste do respetivo variograma teórico. Tal como na Secção 4.1, o variograma teórico é definido pela combinação linear de componentes referentes à estrutura de correlação espacial, que minimiza a distância entre os variogramas empírico e teórico através do método dos mínimos quadrados. Por último, é efetuada a predição de cada variável de interesse sobre uma grelha de estimação, dada a estrutura de correlação espacial estimada.

5.2.1 A *Raja clavata*

As covariáveis consideradas são a profundidade P (em metros), o tipo de fundo e o indicador de abundância de pescada pequena I_{P_P} .

A dependência da média da variável de interesse face às variáveis auxiliares é descrita por

$$\begin{aligned}
\mu(\mathbf{s}) = & \beta_0 + \beta_1 \times P(\mathbf{s}) + \beta_2 \times I_{PP}(\mathbf{s}) \\
& + \beta_3 \times Se(Tipo\ de\ Fundo(\mathbf{s}) = \text{“Mistura de Sedimentos”}) \\
& + \beta_4 \times Se(Tipo\ de\ Fundo(\mathbf{s}) = \text{“Lama”}) \\
& + \beta_5 \times Se(Tipo\ de\ Fundo(\mathbf{s}) = \text{“Lodo com Areia”}) \\
& + \beta_6 \times Se(Tipo\ de\ Fundo(\mathbf{s}) = \text{“Rocha”}) \\
& + \beta_7 \times Se(Tipo\ de\ Fundo(\mathbf{s}) = \text{“Areia”}) \\
& + \beta_8 \times Se(Tipo\ de\ Fundo(\mathbf{s}) = \text{“Areia Barrenta”}) \\
& + \beta_9 \times Longitude(\mathbf{s}) + \beta_{10} \times Latitude(\mathbf{s}) + \delta(\mathbf{s})
\end{aligned} \tag{5.1}$$

onde $\mu(\mathbf{s})$ representa o valor médio dos indicadores de abundância e biomassa de raia-lenga e $\delta(\mathbf{s})$ é o resíduo (especialmente correlacionado) associado à estimação na localização \mathbf{s} .

Indicador de abundância de *Raja clavata*

As estimativas dos coeficientes de regressão do modelo ajustado para a tendência do número de indivíduos capturados por hora de raia-lenga capturada por hora estão apresentadas na Tabela 5.1.

Tabela 5.1: Estimativas dos coeficientes de regressão da tendência para o indicador de abundância de raia-lenga.

Coefficiente	Estimativa
$\hat{\beta}_0$	60,34
$\hat{\beta}_1$	-0,02
$\hat{\beta}_2$	-1×10^{-4}
$\hat{\beta}_3$	-26,00
$\hat{\beta}_4$	-25,10
$\hat{\beta}_5$	-24,73
$\hat{\beta}_6$	-25,17
$\hat{\beta}_7$	-25,39
$\hat{\beta}_8$	-25,94
$\hat{\beta}_9$	-0,04
$\hat{\beta}_{10}$	-3×10^{-3}

Verifica-se que todas as variáveis em média apresentam uma influência negativa no indicador de abundância de raia-lenga. Em face dos resultados, espera-se que um aumento de profundidade em 100 metros diminua em cerca de duas unidades o valor do indicador de abundância de raia-lenga. Relativamente ao tipo de fundo, a espécie, em média, tem valor de indicador de abundância mais elevado sobre “lodo com areia”. A estimativa do coeficiente relativo à pescada pequena indica uma influência reduzida desta espécie na

abundância de raia-lenga, uma vez que o aumento de 1000 pescadas pequenas capturadas por hora apenas faz diminuir, em média, 0,1 unidades de raia-lenga capturada por hora. É de realçar que a interpretação destes números deve ser feita com cuidado, uma vez que existem diferenças apreciáveis nos valores dos indicadores de abundância de raia-lenga e de pescada (ver os valores máximos observados dos indicadores de abundância de raia-lenga e de pescada nas Tabelas 3.1 e 3.5).

O cálculo do variograma empírico foi baseado nos resíduos $\delta(\mathbf{s})$ tendo-se ajustado um variograma teórico (Figura 5.2). A vermelho destaca-se a representação gráfica do variograma teórico. Na Tabela 5.2 estão apresentadas as estimativas dos parâmetros do variograma teórico ajustado.

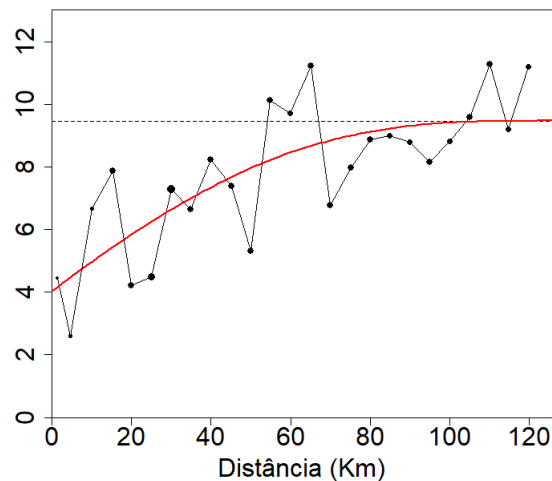


Figura 5.2: Variogramas empírico e teórico dos resíduos associados à estimação da tendência dos valores do indicador de abundância de raia-lenga.

Tabela 5.2: Estimativas dos parâmetros do variograma teórico dos resíduos relativos à estimação da tendência do indicador de abundância de raia-lenga.

Componente	Parâmetros		
	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$
Gaussiana	1,63	85,16	
Exponencial	1,91	77,10	4,02
Esférica	1,93	106,82	

A estimação da tendência espacial e da estrutura de associação espacial permite não só a predição da abundância de raia-lenga na região de estudo, mas também a estimação dos erros associados ao ajuste do modelo. Nas Figuras 5.3 e 5.4 estão representadas a superfície da distribuição espacial de predição do número de raia-lenga capturada por hora e a superfície dos erros associados à predição, respetivamente.

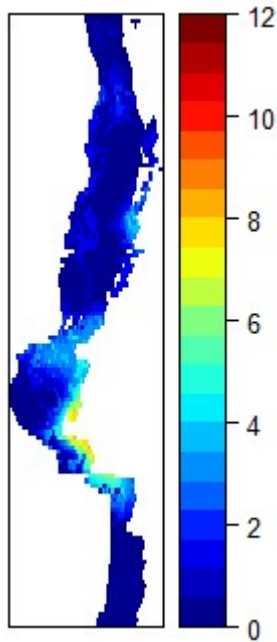


Figura 5.3: Superfície prevista por **KED** da distribuição espacial do indicador de abundância de raia-lenga.

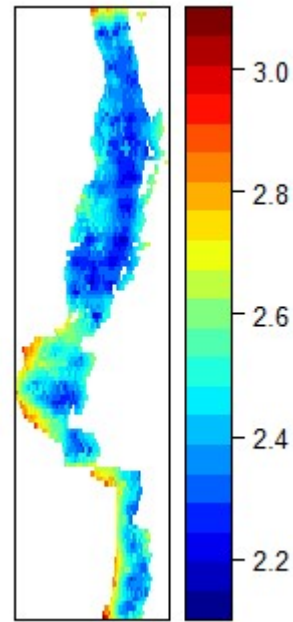


Figura 5.4: Superfície de desvios padrão das estimativas por **KED** do indicador de abundância de raia-lenga.

À semelhança da estimação não-paramétrica (Figura 5.1a), a abundância de raia-lenga parece superior na zona de Lisboa e arredores, em particular na zona de Sesimbra, Sines e Aveiro, e também na zona de Peniche. Os desvios padrão associados à estimação, como seria expectável, são mais elevados longe dos pontos observados e reduzidos nas localizações próximas donde se dispõe de observações.

Indicador de biomassa de *Raja clavata*

O processo de estimação do indicador de biomassa da raia-lenga por *kriging* com tendência externa é análogo ao realizado para a abundância desta espécie. As estimativas dos coeficientes apresentados na Expressão (5.1) encontram-se na Tabela 5.3.

O incremento de unidades nas covariáveis faça diminuir o indicador de biomassa de raia-lenga, com exceção da variável auxiliar tipo de fundo. O indicador de biomassa tem, em média, valor superior nas estações com tipo de fundo de “lama”. No entanto, o impacto médio deste tipo de fundo no indicador de biomassa de raia-lenga não parece ser muito diferente do estimado para o tipo de fundo de “lodo com areia”. De uma forma geral e como expectável, os resultados são semelhantes para o número e peso de elementos capturados de raia-lenga.

Estimados os parâmetros da tendência, em seguida, à custa dos erros são determinados os variogramas empírico e teórico. Na Figura 5.5 estão representados a preto o variograma empírico e a vermelho o variograma teórico.

Tabela 5.3: Estimativas dos coeficientes de regressão da tendência do indicador de biomassa de raia-lenga.

Coefficiente	Estimativa
$\hat{\beta}_0$	2,44
$\hat{\beta}_1$	-0,03
$\hat{\beta}_2$	-1×10^{-4}
$\hat{\beta}_3$	46,06
$\hat{\beta}_4$	47,14
$\hat{\beta}_5$	47,04
$\hat{\beta}_6$	45,93
$\hat{\beta}_7$	46,35
$\hat{\beta}_8$	45,17
$\hat{\beta}_9$	-0,04
$\hat{\beta}_{10}$	-0,01

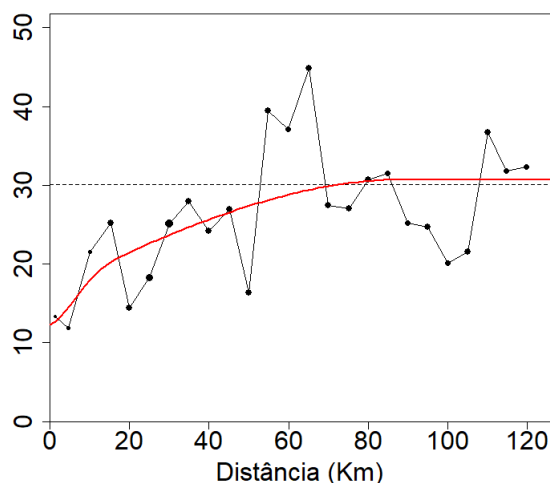


Figura 5.5: Variogramas empírico e teórico dos erros associados à estimação da tendência do indicador de biomassa de raia-lenga.

As estimativas dos parâmetros do variograma teórico encontram-se na Tabela 5.4. Verifica-se a existência de efeito pepita, que expressa a variância associada aos erros de medição do fenómeno.

Tabela 5.4: Estimativas dos parâmetros do variograma teórico dos erros relativos à estimação da tendência do indicador de biomassa de raia-lenga.

Componente	Parâmetros		
	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$
Gaussiana	4,64	15,49	12,30
Esférica	13,77	88,43	

Com base nos resultados do cálculo do variograma, foi possível realizar a predição espacial, representada na Figura 5.6 e obter o respetivo mapa de desvio padrão, apresentado na Figura 5.7.

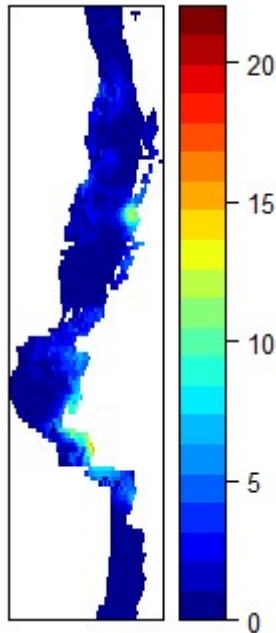


Figura 5.6: Superfície predita por **KED** da distribuição espacial do indicador de **biomassa de raia-lenga**.

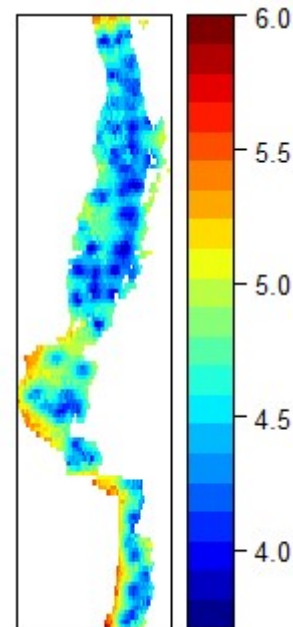


Figura 5.7: Superfície de desvios padrão das estimativas por **KED** do indicador de **biomassa de raia-lenga**.

Pela análise à superfície predita da distribuição espacial do indicador de biomassa de raia-lenga capturada por hora, conclui-se que nas zonas mais a este dos setores de Lisboa, Sines e Aveiro o valor deste indicador é mais elevado. Analogamente à interpretação efetuada da Figura 5.4, pela Figura 5.7 o desvio padrão das estimativas é menor perto dos pontos observados.

Ao comparar os resultados associados à estrutura de correlação espacial dos indicadores de abundância e biomassa da raia-lenga, verifica-se que o raio de influência estimado $\hat{\phi}$ para o indicador da biomassa é inferior ao da abundância. O que indica que no caso do número por hora a correlação espacial se mantém para distâncias mais longas.

Os resultados da estimação de ambos os indicadores de abundância e biomassa da raia-lenga são idênticos. Esta hipótese é corroborada pelo facto dos padrões espaciais encontrados nos mapas para a abundância e biomassa apresentarem semelhanças.

5.2.2 A *Merluccius merluccius*

A estimação do indicador de abundância da pescada é efetuada à custa das variáveis profundidade P , longitude e latitude. O tipo de fundo não é considerado pelo facto de não

se possuir esta característica nas localizações referentes às campanhas demersais IPMA para a pescada.

Assim, a média da abundância de pescada pode ser descrita à custa das variáveis auxiliares por:

$$\mu(\mathbf{s}) = \beta_0 + \beta_1 \times P(\mathbf{s}) + \beta_2 \times Longitude(\mathbf{s}) + \beta_3 \times Latitude(\mathbf{s}) + \delta(\mathbf{s}) \quad (5.2)$$

onde $\mu(\mathbf{s})$ indica o valor médio de abundância de pescada e $\delta(\mathbf{s})$ é o resíduo associado à estimação na localização \mathbf{s} . A abundância de pescada é maior quanto maior for a profundidade (ver Tabela 5.5). Em média, o indicador de abundância de pescada é mais elevado longe da costa e na região norte. Mais precisamente, em dois pontos com a mesma longitude distanciados em 100 Km verifica-se que, em média, na localização mais a norte o número de pescada capturada por hora em 144 elementos, considerando que a profundidade é a mesma.

Tabela 5.5: Estimativas dos coeficientes de regressão da tendência do indicador de abundância de pescada.

Coefficiente	Estimativa
$\hat{\beta}_0$	-6173,36
$\hat{\beta}_1$	3,56
$\hat{\beta}_2$	-0,31
$\hat{\beta}_3$	1,44

Com base nos valores dos resíduos obtidos determinou-se o variograma amostral e ajustou-se o variograma teórico. O primeiro está apresentado a preto e o segundo a vermelho na Figura 5.8. O variograma teórico é representado de acordo com a estimação dos parâmetros envolvidos no mesmo, que estão apresentados 5.6. Através desta tabela, verifica-se que a variância dos erros de medição é de 382723,12.

Tabela 5.6: Estimativas dos parâmetros do variograma teórico dos erros relativos à estimação da tendência do indicador de abundância de pescada.

Componente	Parâmetros		
	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$
Gaussiana	85580,22	40,06	382723,12

Determinada a estrutura de correlação espacial, o próximo passo na estimação espacial da abundância de pescada por *kriging* com tendência externa é a predição espacial, onde o seu mapa está representado na Figura 5.9. Consequentemente, foi também obtido o respetivo mapa de desvio padrão das estimativas da abundância, apresentada na Figura 5.10.

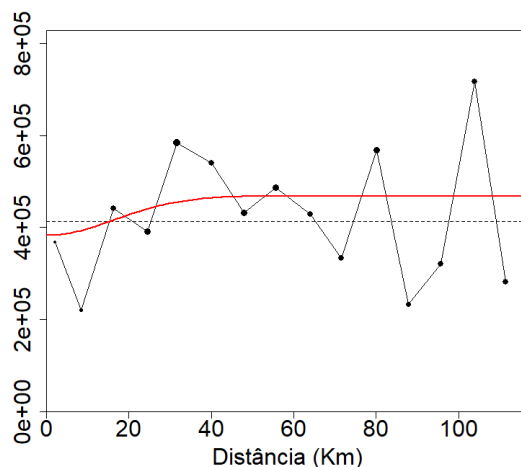


Figura 5.8: Variogramas empírico e teórico dos erros associados à estimação da tendência do indicador de abundância de pescada.

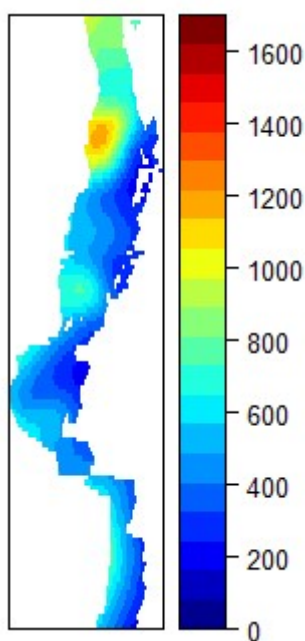


Figura 5.9: Superfície predita por **KED** da distribuição espacial do indicador de **abundância de pescada**.

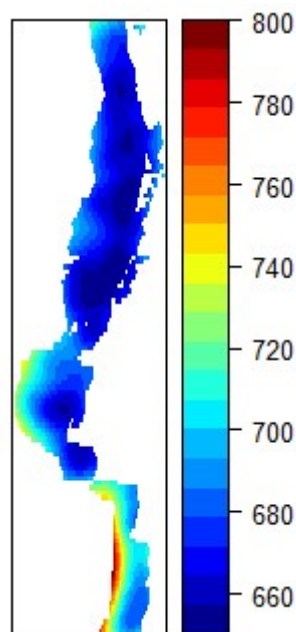


Figura 5.10: Superfície de desvios padrão das estimativas por **KED** do indicador de **abundância de pescada**.

É de realçar que a norte da costa continental portuguesa, a abundância de pescada estimada é maior quando comparada com outras zonas. Relativamente ao mapa da Figura 5.10, como expectável, verifica-se que os desvios padrões estimados são superiores onde não ocorreram observações. Isto deve-se ao facto de não se possuir informação real em tais localizações.

As análises realizadas nas secções 5.2.1 e 5.2.2 realçaram o facto de a distribuição espacial dos indicadores de abundância da raia-lenga e da pescada terem comportamentos diferentes. Primeiramente, este facto é revelado dado que as variáveis profundidade e latitude têm, em média, uma influência negativa nos indicadores de abundância de raia-lenga e positiva na abundância de pescada. Esta diferença é também visível pela comparação das superfícies preditas da distribuição espacial do número e peso dos indivíduos capturados por hora de raia-lenga capturada por hora (Figuras 5.3 e 5.6) com o mapa de predição espacial da abundância de pescada (Figura 5.9). Isto é, enquanto que a ocorrência de raia-lenga, em média, é superior na região de Lisboa e Sines, espera-se que a abundância de pescada seja maior na zona norte da costa.

5.3 Modelos de zeros inflacionados

Como referido na Secção 2.4, os ZIM visam a estimação da variável de interesse tendo em conta o número elevado de observações nulas. Em particular, propõe-se um modelo bivariado para modelar a probabilidade de presença em conjunto com as observações positivas.

Assim, denota-se por $T(\mathbf{s})$ a variável que indica a presença da espécie na localização \mathbf{s} , e por $Y(\mathbf{s})$ o valor do indicador de abundância ou de biomassa da espécie num ponto \mathbf{s} sob a condição de presença da mesma. Os parâmetros envolvidos nas Expressões (2.41)-(2.43) seguem as seguintes distribuições *à priori*:

$$\begin{aligned}\alpha_i &\sim Normal(0, 1000), i = 1, 2 \\ \beta_{i,j} &\sim Normal(0, 1000), i = 1, 2; j = 1, \dots, p \\ k &\sim Normal(0, 1000)\end{aligned}\tag{5.3}$$

de forma a que estas sejam o menos informativas possível, uma vez que não se possui qualquer conhecimento sobre estes parâmetros.

A aplicação dos ZIM inicia-se pela estimação dos parâmetros apresentados na Expressão (5.3) e dos relacionados com a estrutura de associação espacial, seguida da predição espacial da mediana *à posteriori* das variáveis $T(\mathbf{s})$ e $Y(\mathbf{s})$. De acordo com a Expressão (2.44), o processo $Z(\mathbf{s})$ para o indicador de abundância ou biomassa da espécie de interesse pode ser obtido como função de $T(\mathbf{s})$ e $Y(\mathbf{s})$. Por último também é possível obter a projeção do GMRF $W(\mathbf{s})$, que traduz a estrutura de correlação espacial inerente aos dados.

5.3.1 A *Raja clavata*

As variáveis profundidade P e abundância de juvenis de pescada I_{Pp} são utilizadas como covariáveis. A covariável tipo de fundo não pode ser considerada, porque não se possui essa informação para os pontos da *mesh*¹. A Expressão (2.43) pode ser reescrita da seguinte forma

$$\begin{aligned} \text{logit}(p(\mathbf{s})) &= \alpha_1 + \beta_{1,1} \times P(\mathbf{s}) \\ &+ \beta_{1,2} \times I_{Pp}(\mathbf{s}) + W(\mathbf{s}) \end{aligned} \tag{5.4}$$

$$\begin{aligned} \log(a(\mathbf{s})/b(\mathbf{s})) &= \alpha_2 + \beta_{2,1} \times P(\mathbf{s}) \\ &+ \beta_{2,2} \times I_{Pp}(\mathbf{s}) + kW(\mathbf{s}) \end{aligned}$$

É de realçar que apenas serão consideradas as covariáveis em que o respetivo coeficiente seja estatisticamente significativo, isto é, caso o valor nulo não pertença ao seu intervalo de credibilidade a 95%.

Indicador de abundância de *Raja clavata* capturada por hora

Recordando a análise efetuada na Secção 3.1, em 161 das 212 localizações observadas ao longo das quatro campanhas não se verificou a ocorrência de *R. clavata*. Deste modo, é importante analisar a distribuição do indicador de abundância de raia-lenga, apresentada na Figura 5.11, e a distribuição do indicador de abundância não nulo de raia-lenga, representada pelo histograma da Figura 5.12.

O histograma da Figura 5.11 realça o elevado número de observações nulas. A Figura 5.12 valida o uso da distribuição gama para modelar a abundância de raia-lenga capturada por hora sob a hipótese de presença, dado que parece ter um comportamento semelhante a esta.

Determinada uma grelha (denominada *mesh*) que permite aproximar o campo contínuo GF pelo campo discreto GMRF $W(\mathbf{s})$, segue-se a estimação das densidades marginais *à posteriori* de cada parâmetro do modelo, incluindo os parâmetros de regressão e os parâmetros associados ao efeito espacial $W(\mathbf{s})$. Na Tabela 5.7 são apresentados a média, o desvio padrão, os quantis 2,5% e 97,5%, a mediana, a moda e a divergência de *Kullback-Leibler*² (KLD). Valores pequenos do KLD indicam que a distribuição *à posteriori* é bem aproximada a uma distribuição normal.

¹Grelha que permite aproximar o campo contínuo GF pelo campo discreto GMRF $W(\mathbf{s})$.

²Descreve a diferença entre a Aproximação Gaussiana padrão e a Laplace Simplificada às densidades marginais posteriores.

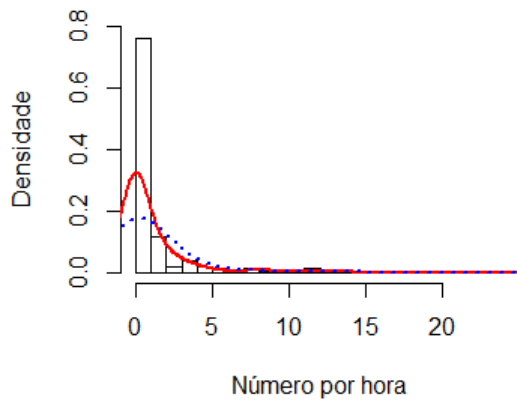


Figura 5.11: Histograma dos valores observados do número de elementos capturados por hora de raia-lenga.

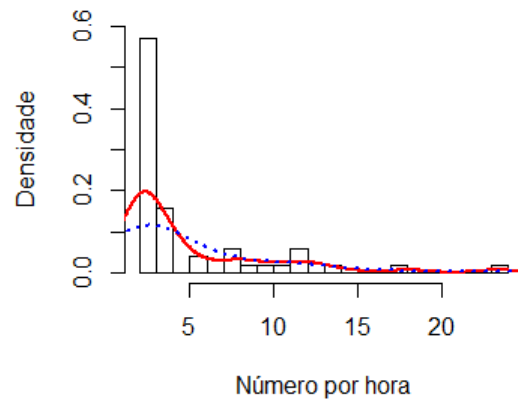


Figura 5.12: Histograma dos valores observados não nulos do número de indivíduos capturados por hora de raia-lenga.

Tabela 5.7: Medidas das distribuições dos parâmetros de regressão do modelo de zero inflacionados para abundância de raia-lenga capturada por hora.

	Média	Desvio Padrão	Quantil 2,5%	Mediana	Quantil 97,5%	Moda	KLD
α_1	-1,42	1,22	-4,06	-1,37	0,94	-1,30	0
α_2	1,80	0,60	0,52	1,82	2,93	1,8	0
$\beta_{2,1}$	-0,007	0,003	-0,013	-0,007	-0,001	-0,007	0
$\beta_{2,2}$	-0,001	3×10^{-4}	-0,001	-0,001	-1×10^{-4}	-7×10^{-4}	0

Através da consulta da Tabela 5.7 e tendo em conta a função de ligação *logit*, pode-se concluir que a probabilidade mediana de ocorrência de raia-lenga é de 0,20³. Caso se verifique a presença de raia-lenga, por cada aumento de 100 metros de profundidade, a abundância mediana de raia-lenga diminui cerca de 50%, considerando que a abundância de juvenis pescada é constante. Sob hipótese de presença de raia-lenga, o incremento de 100 elementos na abundância de pescada pequena reflete uma diminuição em cerca de 10% do número mediano de elementos capturados por hora de raia-lenga à mesma profundidade.

Relativamente às medidas de distribuição dos parâmetros associados ao efeito espacial $W(\mathbf{s})$, apresentadas na Tabela 5.8, verifica-se que à distância mediana de cerca de 84 Km a correlação entre variáveis tende a ser negligenciável. A mediana da distribuição à posteriori da variância σ^2 da abundância da raia-lenga capturada por hora é de cerca de 6,55.

$${}^3\hat{p} = \frac{\epsilon^{-1,37}}{1 + \epsilon^{-1,37}}$$

Tabela 5.8: Medidas das distribuições dos parâmetros associados ao efeito espacial $W(\mathbf{s})$ para o modelo de zero inflacionados para abundância de raia-lenga capturada por hora.

	Média	Desvio Padrão	Quantil 2,5%	Mediana	Quantil 97,5%	Moda
$1/\tau^2$	3,24	0,74	2,01	3,17	4,88	3,03
ϕ	91,78	37,32	43,21	83,64	186,16	70,32
σ	2,69	0,84	1,43	2,56	4,68	2,33
k	0,38	0,10	0,19	0,37	0,59	0,36

Um dos objetivos quando se está a modelar processos espacialmente contínuos é a predição do valor esperado numa localização alvo não observada. Desta forma, na Figura 5.13 estão apresentados, respetivamente, os mapas da mediana *à posteriori* da probabilidade de ocorrência de *R. clavata*, o número de elementos capturados de raia-lenga sob a condição de presença da mesma, e da abundância de raia-lenga capturada por hora na costa continental portuguesa.

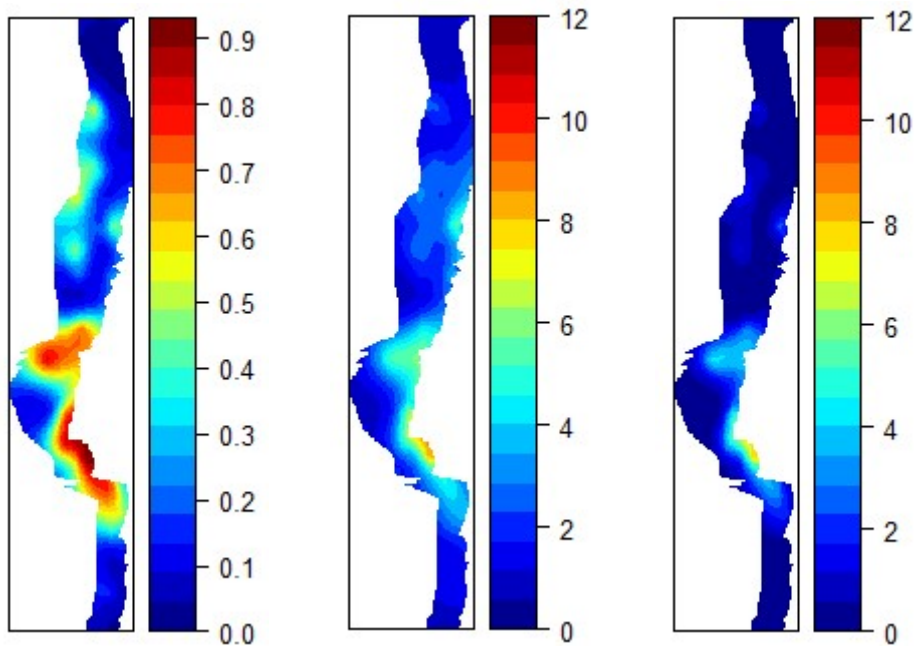


Figura 5.13: Superfícies previstas por **ZIM** das distribuições espaciais de: probabilidade de presença (painel da esquerda), abundância sob hipótese de presença da espécie (painel central) e da **abundância** de **raia-lenga** (painel da direita).

Através do painel da esquerda verifica-se que junto da costa nas zonas de Sines, Lisboa e Berlengas (BER) a probabilidade de ocorrência da espécie é mais elevada. O painel da direita, que se traduz pelo produto das estimativas apresentadas nos painéis da esquerda e central, indica que nas zonas de Aveiro, Berlengas, Lisboa e Sines estima-se que a abundância da espécie seja maior. É importante realçar, pela análise ao painel da direita da Figura 5.13, que em praticamente toda a costa o número mediano de raia-lenga

capturada é quase nulo.

A distribuição espacial estimada do indicador de abundância da raia-lenga determinada por *kriging* com tendência externa é semelhante à determinada com base no modelo de zeros inflacionados (ver Figuras 5.3 e 5.13).

Os mapas da média e do desvio padrão dos efeitos espaciais estão representados nos painéis da esquerda e da direita da Figura 5.14, respectivamente.

A Figura 5.14 mostra que existe uma variação de -2,4 a 4,1 no efeito espacial e os desvios padrão variam entre 1,3 e 2,8. O efeito espacial, painel da esquerda, apresenta uma estrutura de correlação, que se reflete na identificação de padrões espaciais.

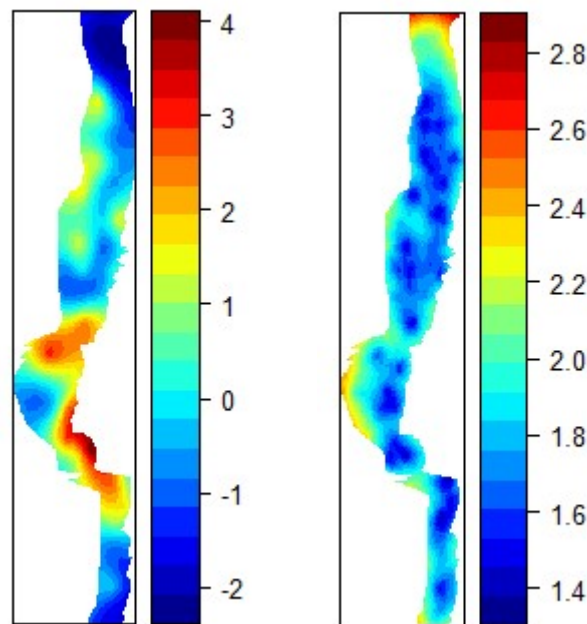


Figura 5.14: Superfícies por **ZIM** da média *à posteriori* (painel da esquerda) e do desvio padrão (painel da direita) dos efeitos espaciais do indicador de **abundância** de **raia-lenga**.

Indicador de biomassa de *Raja clavata* capturada por hora

Uma vez que os indicadores de abundância e de biomassa da *R. clavata* estão relacionados, também é visível o elevado número de zeros na variável peso de indivíduos (em Kg) capturados por hora.

Este facto, que é confirmado pela consulta do histograma da Figura 5.15, justifica o uso dos ZIM. Na Figura 5.16 está representada a distribuição do peso capturado por hora e verifica-se que os valores do peso capturado por hora estão mais concentrados à esquerda.

A aplicação de ZIM permitiu a estimação das densidades marginais *à posteriori* dos parâmetros de regressão, apresentados na Tabela 5.9, e dos parâmetros associados ao

efeito espacial $W(\mathbf{s})$, em que as suas estimativas se encontram na Tabela 5.10. Em cada uma das tabelas estão apresentados o valor médio, o desvio padrão, os quantis 2,5% e 97,5%, o valor mediano e a moda das densidades marginais. Na Tabela 5.9 ainda se encontra o valor da KLD associado a cada parâmetro.

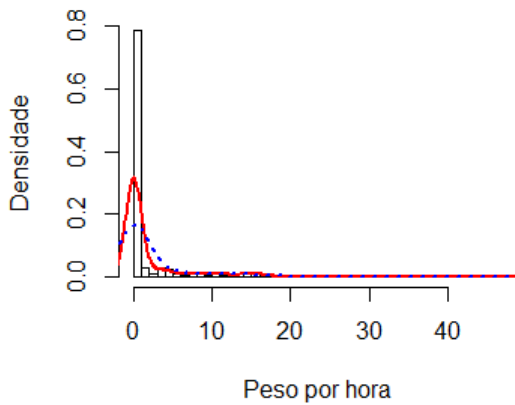


Figura 5.15: Histograma dos valores observados do peso dos indivíduos capturados por hora de raia-lenga.

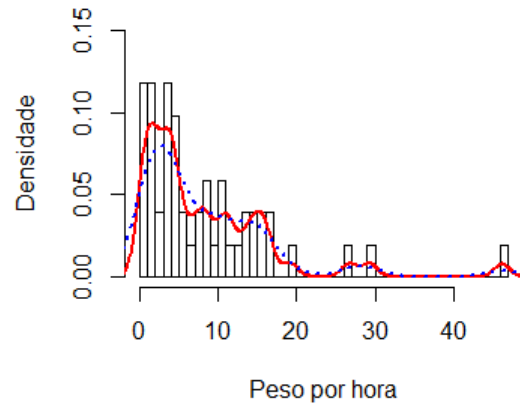


Figura 5.16: Histograma dos valores observados não nulos do peso dos elementos capturados por hora de raia-lenga.

Tabela 5.9: Medidas das distribuições dos parâmetros de regressão do modelo de zero inflacionados para o indicador de biomassa de raia-lenga capturada por hora.

	Média	Desvio Padrão	Quantil 2,5%	Mediana	Quantil 97,5%	Moda	KLD
α_1	-1,42	0,87	-3,33	-1,37	0,25	-1,30	0
α_2	2,61	0,62	1,26	2,64	3,73	2,71	0
$\beta_{2,1}$	-0,011	0,004	-0,018	-0,011	-0,004	-0,011	0

O valor da mediana da distribuição do parâmetro α_1 , indica que a probabilidade mediana de ocorrência da *R. clavata* é de cerca de 0,20. Sob a condição de que a espécie está presente, o incremento de 100 na profundidade da localização considerada resulta na diminuição em cerca de 67% do indicador de biomassa nessa mesma localização. Ao modelar o indicador de biomassa, a variável abundância de pescada pequena não influencia significativamente nem a presença da raia-lenga nem o peso da mesma capturado sob a hipótese de presença.

Ao analisar a Tabela 5.10 verifica-se, pelas medidas determinadas da distribuição do raio de influência, que à distância mediana de 61 Km deixa de existir correlação espacial. O valor da mediana da distribuição da variância do indicador de biomassa é 4,45 e o da variância dos erros de medição é de cerca de 0,78.

Tabela 5.10: Medidas das distribuições dos parâmetros associados ao efeito espacial $W(\mathbf{s})$ para o modelo de zero inflacionados para o indicador de biomassa da raia-lenga capturada por hora.

	Média	Desvio Padrão	Quantil 2,5%	Mediana	Quantil 97,5%	Moda
$1/\tau^2$	1,32	0,27	0,87	1,29	1,92	1,24
ϕ	70,00	35,33	27,71	61,41	161,59	48,64
σ	2,20	0,64	1,19	2,11	3,68	1,96
k	0,41	0,19	0,04	0,40	0,78	0,39

A predição espacial da biomassa em pontos não observados é obtida pelo valor mediano da distribuição das variáveis T , Y e Z , respetivamente representadas pelos painéis da esquerda, central e da direita da Figura 5.17. Os valores apresentados no painel da direita foram determinados com base no valor mediano das distribuições marginais $T(\mathbf{s})$ e $Y(\mathbf{s})$ de acordo com a Expressão (2.44).

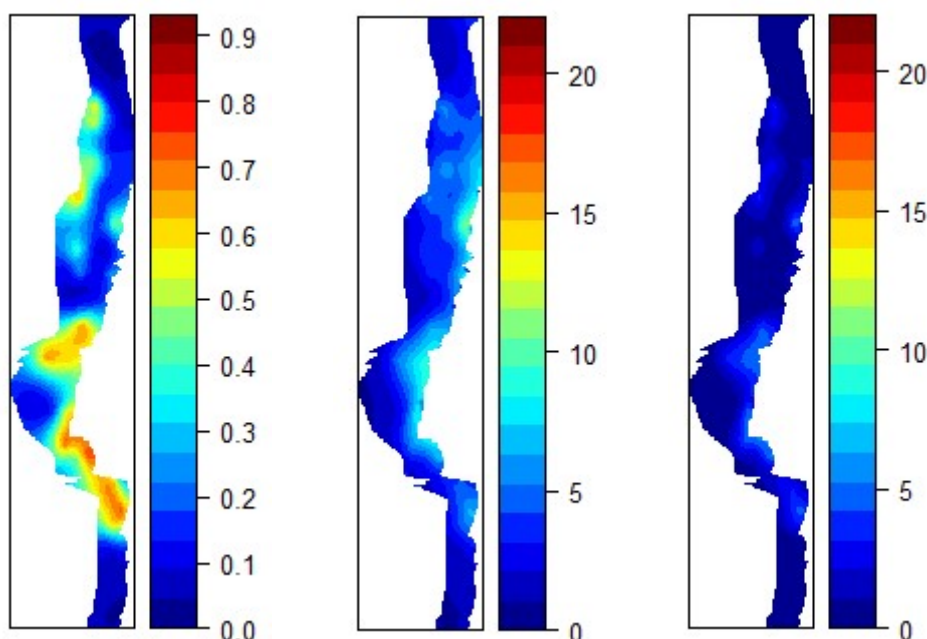


Figura 5.17: Superfícies previstas por **ZIM** das distribuições espaciais de: probabilidade de presença (painel da esquerda), biomassa sob hipótese de presença da espécie (painel central) e do indicador de **biomassa** de **raia-lenga** (painel da esquerda).

Através da superfície predita da distribuição espacial da probabilidade de presença de raia-lenga, verifica-se que o valor mediano desta é superior nas zonas das Berlengas, Lisboa e Sines junto da costa. Pelo painel da direita da Figura 5.17, é perceptível que nestas zonas o indicador de biomassa é mais elevado, em valor mediano. Nas restantes zonas da costa, a mediana do peso capturado por hora de arrasto é muito próximo de zero.

Os efeitos espaciais envolvidos na modelação da variável de interesse estão representados, na Figura 5.18, pelos mapa do seu valor médio (painel da esquerda) e do seu desvio padrão (painel de direita). Pela análise da Figura 5.18 verifica-se que existe uma variação da média dos efeitos espaciais de -2 a 2,6 e o desvio padrão varia entre 1,1 e 2,4.

Avaliando os resultados da Secção 5.3.1, faz-se notar que a variância dos erros de medição τ^2 , em termos medianos, é superior na modelação da abundância. Através da comparação dos painéis da esquerda das Figuras 5.13 e 5.17 verifica-se que a ocorrência de raia-lenga, tanto em peso como em número, é maior nas regiões de Lisboa, Sines e Berlengas.

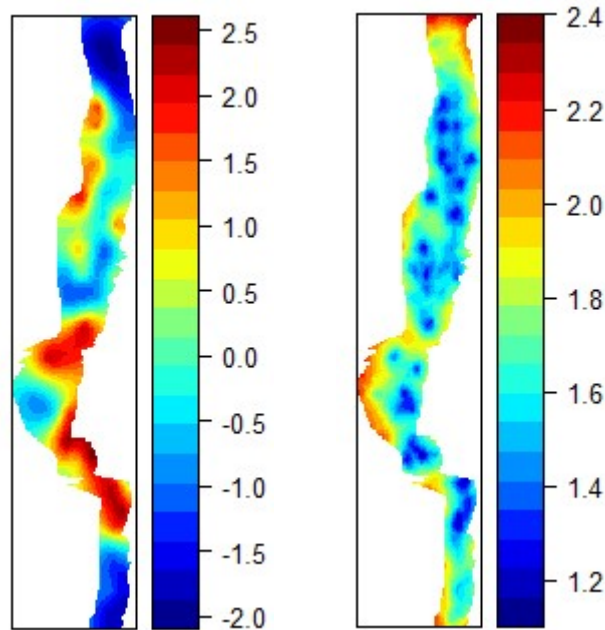


Figura 5.18: Superfícies por **ZIM** da média *à posteriori* (painel da esquerda) e do desvio padrão (painel da direita) dos efeitos espaciais do indicador de **biomassa** de **raia-lenga**.

5.3.2 A *Merluccius merluccius*

Analogamente ao efetuado para os indicadores de abundância da *R. clavata*, ir-se-á de seguida considerar um modelo Bayesiano de zeros inflacionados para a pescada, considerando a profundidade P como possível covariável. Por conseguinte, a Expressão (2.43) pode ser apresentada como

$$\text{logit}(p(\mathbf{s})) = \alpha_1 + \beta_{1,1} \times P(\mathbf{s}) + W(\mathbf{s}) \quad (5.5)$$

$$\log(a(\mathbf{s})/b(\mathbf{s})) = \alpha_2 + \beta_{2,1} \times P(\mathbf{s}) + kW(\mathbf{s})$$

Como referido na Secção 3.2, as espécies em estudo são capturadas em conjunto, em campanhas demersais governados pelo IPMA. Apesar da existência de localizações onde

não se observou pescada, a ocorrência da pescada é maior do que a da raia-lenga. Este facto pode ser confirmado pela análise das Figuras 5.19 e 5.20, uma vez que a diferença entre as distribuições das variáveis não é tão evidente. É no entanto plausível o uso da distribuição gama para modelar a abundância observada de pescada, dado que os valores desta variável possuem características como serem contínuos e simétricos à direita.

Um sumário dos parâmetros de regressão estimados para as densidades marginais à *posteriori*, é apresentado na Tabela 5.11. Os resultados para os respetivos parâmetros do efeito espacial $W(\mathbf{s})$ são apresentados na Tabela 5.12. Tal como acontece nas Tabelas 5.7 e 5.9, na Tabela 5.11 são apresentados os valores da média e da mediana, o desvio padrão, os quantis 2,5% e 97,5%, a moda e o valor da KLD.

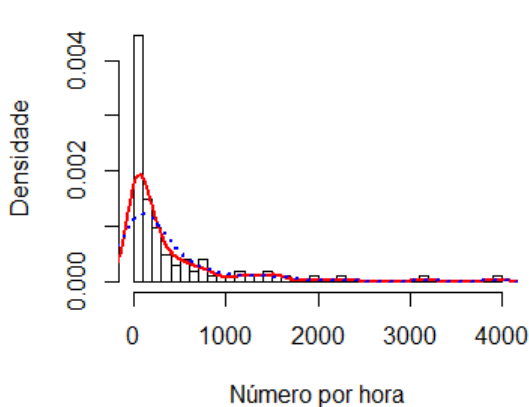


Figura 5.19: Histograma dos valores observados do indicador de abundância de pescada.

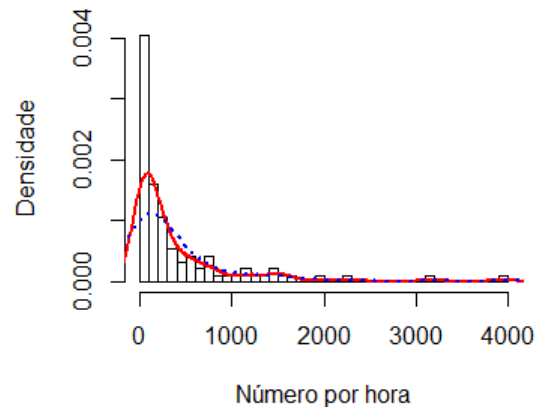


Figura 5.20: Histograma dos valores observados não nulos do indicador de abundância de pescada.

Pela consulta da Tabela 5.11 verifica-se que a profundidade influencia significativamente a presença e a abundância de pescada sob condição da presença da mesma. Mais concretamente, o aumento em 100 metros na profundidade aumenta, em termos medianos, mais de sete o valor do indicador de abundância. Os valores do KLD para todos os parâmetros são muito próximos de zero, pelo que a distribuição à *posteriori* destes parâmetros é bem aproximada a uma distribuição normal. Relativamente à análise da Tabela 5.12, é possível verificar que à distância mediana de 90 Km deixa de existir correlação espacial. É também visível que a variância associada aos erros de medição é de 1,28, em termos medianos.

Após a estimação da distribuição dos parâmetros de regressão e dos parâmetros associados ao efeito espacial $W(\mathbf{s})$, segue-se o objetivo da estimação, a predição espacial da abundância da espécie. Assim, no painel da esquerda da Figura 5.21 encontra-se a superfície predita da distribuição da probabilidade de presença da espécie, no painel central

Tabela 5.11: Medidas das distribuições dos parâmetros de regressão do modelo de zero inflacionados para o indicador de abundância da pescada capturada por hora.

	Média	Desvio Padrão	Quantil 2,5%	Mediana	Quantil 97,5%	Moda	KLD
α_1	-0,85	1,94	-4,70	-0,85	2,98	-0,85	0
α_2	3,28	0,91	1,47	3,29	5,08	3,34	1×10^{-4}
$\beta_{1,1}$	0,054	0,027	0,002	0,054	0,108	0,054	0
$\beta_{2,2}$	0,02	0,01	0,01	0,02	0,03	0,02	1×10^{-4}

Tabela 5.12: Medidas das distribuições dos parâmetros associados ao efeito espacial $W(\mathbf{s})$ para o modelo de zero inflacionados para o indicador de abundância de pescada.

	Média	Desvio Padrão	Quantil 2,5%	Mediana	Quantil 97,5%	Moda
$1/\tau^2$	0,77	0,10	0,56	0,78	0,97	0,80
ϕ	131,87	134,05	39,31	90,48	474,44	51,75
σ	1,58	0,47	0,78	1,55	2,59	1,49
k	0,57	0,19	0,26	0,54	1,01	0,44

o mapa de predição da distribuição do indicador de abundância de pescada considerando a ocorrência da espécie, e no painel da direita a superfície predita da distribuição do número de elementos capturados de pescada por hora.

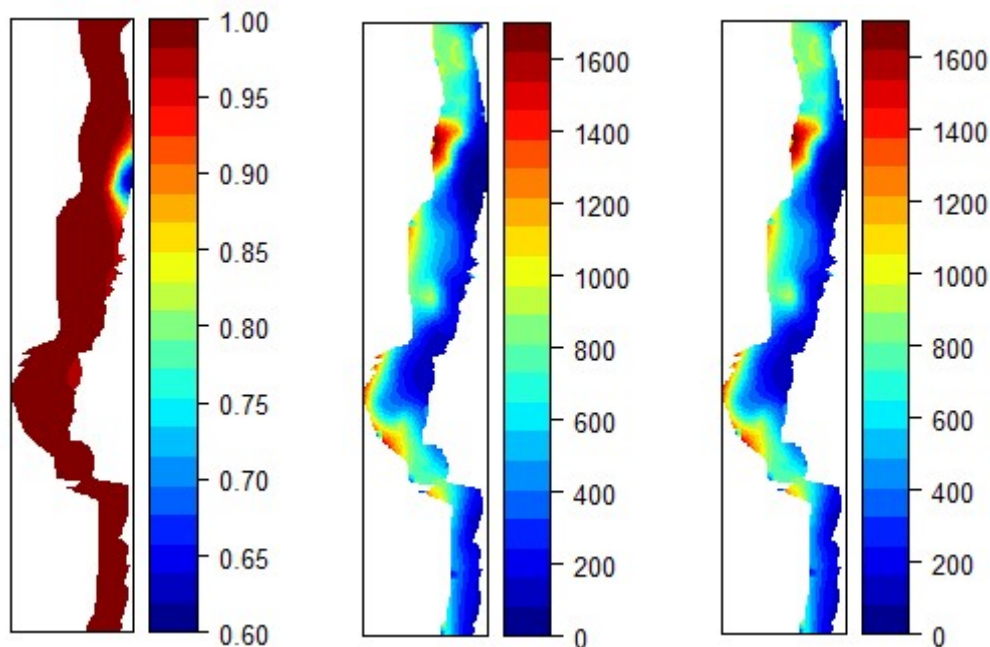


Figura 5.21: Superfícies preditas por **ZIM** das distribuições espaciais de: probabilidade de presença (painel da esquerda), abundância sob hipótese de presença da espécie (painel central) e **abundância** de **pescada** (painel da esquerda).

É fator relevante, pela análise da Figura 5.21, que a probabilidade mediana de ocorrência

desta espécie seja superior a 0,6 em toda a costa. Pelos painéis central e da direita, verifica-se que, no norte da costa e nas zonas mais afastadas da costa, a abundância mediana da pescada é superior.

Ao comparar as Figuras 5.9 e 5.21, verifica-se que as distribuições espaciais da abundância da pescada pelo método *kriging* com tendência externa e pelo modelo de zeros inflacionados são idênticas. Porém na zona norte da costa, os valores do indicador de abundância da pescada são superiores no mapa do painel da direita da Figura 5.21.

Na Figura 5.22 verifica-se que o o valor médio dos efeitos espaciais varia entre -2 e 3,1 e que a variabilidade destes varia entre 1,2 e 2,2, tomando valores mais baixos junto dos pontos observados.

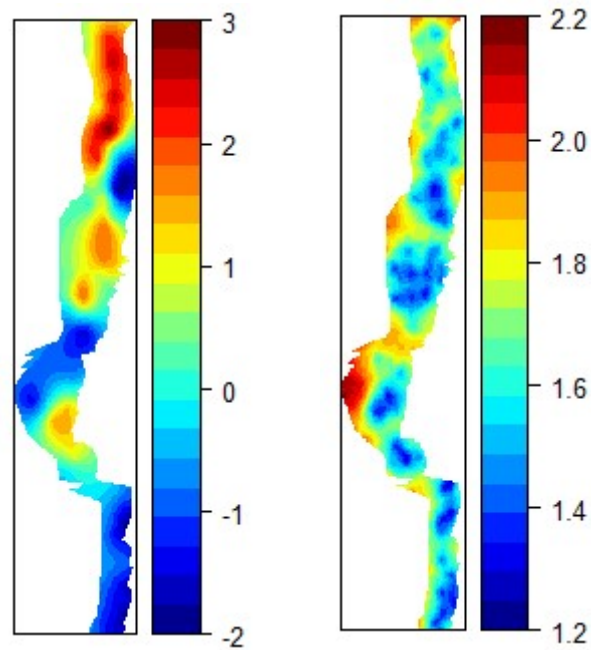


Figura 5.22: Superfícies por **ZIM** da média *à posteriori* (painel da esquerda) e do desvio padrão (painel da direita) dos efeitos espaciais do indicador de **abundância** de **pescada**.

Capítulo 6

Proposta e análise de plano de amostragem

6.1 Introdução

Desde 1979, o IPMA realiza campanhas de investigação dirigidos às espécies demersais que cobrem a costa continental de Portugal desde Caminha até Vila Real de Santo António. Até ao presente, foram realizadas 49 campanhas.

De junho de 1979 a julho de 1989, foi aplicado o plano de amostragem estratificado para a seleção de estações de pesca. Os estratos considerados nas campanhas do IPMA tinham por base a profundidade (até à batimétrica dos 500 metros) e as áreas geográficas da costa e as estações de pesca eram escolhidas aleatoriamente. De outubro de 1989 a 2005, passou-se a integrar estações fixas, visto que reduz a variância espacial da abundância das espécies capturadas. A par da mudança do programa de amostragem, a área de amostragem foi estendida até à batimétrica dos 750 metros. As localizações das estações de pesca fixas até à batimétrica dos 500 metros são locais arrastáveis, identificados nas estações de pesca realizados nas campanhas da série histórica 1979-1989. As estações de pesca relativas a batimetria compreendida entre 500 e 750 metros basearam-se em informações, sobre fundos arrastáveis, resultantes de estudos realizados no IPMA. O programa incluía 97 estações fixas, às quais são adicionadas 30 estações de pesca complementares para cada época, caso haja disponibilidade ou para substituir localizações impossíveis de arrastar. Atualmente, as campanhas demersais consideram estações de pesca até à batimétrica dos 500 metros, das quais 66 são fixas e 30 estações selecionadas aleatoriamente.

Tal como referido na Secção 1.2, um dos objetivos desta dissertação é a definição de um plano de amostragem sob a condição de representatividade na costa continental portuguesa. Isto é, o IPMA pretende que, nos vários estratos, representados na Figura A.1 do Apêndice A definidos pela combinação do setor e do grupo de profundidade,

sejam consideradas pelo menos duas unidades amostrais. O grupo de profundidade é uma variável binária que toma valor 1 caso a profundidade seja inferior a 100 metros e 2 quando a profundidade está compreendida entre 100 e 200 metros. Assim, o estrato de uma localização pertencente ao setor Caminha e de profundidade inferior a 100 metros é denominado por $Caminha_1$. Deste modo, uma vez que a região em estudo é composta por nove setores, no plano de amostragem serão considerados 18 estratos. Na Figura A.1 do Apêndice A, as linhas horizontais representam as divisões das áreas geográficas e as linhas a vermelho e a azul retratam as batimétricas dos 100 e dos 200 m, respetivamente. A Tabela 6.1 apresenta o número mínimo de localizações requeridas por estrato, sendo necessário considerar 36 estações para o cumprimento do critério da representatividade.

Tabela 6.1: Requisito do número de unidades amostrais por estrato.

Profundidade	Setor								
	CAM	MAT	AVE	FIG	BER	LIS	SIN	MIL	ARR
0m - 100m	2	2	2	2	2	2	2	2	2
100m - 200m	2	2	2	2	2	2	2	2	2

Tabela 6.2: Número de estações fixas por estrato (a preto) e número de estações em falta para cumprir o critério de representatividade dos estratos (a vermelho).

Profundidade	Setor									Total
	CAM	MAT	AVE	FIG	BER	LIS	SIN	MIL	ARR	
0m - 100m	4	4	4	4	1 (1)	1 (1)	2	0 (2)	0 (2)	20 (6)
100m - 200m	3	2	2	3	3	4	4	2	1 (1)	24 (1)
>200m	0	2	1	1	0	2	1	2	1	10
Total	7	8	7	8	4 (1)	7 (1)	7	4 (2)	2 (3)	54 (7)

A amostra selecionada será constituída por 65 estações conforme especificado pelo IPMA, das quais 54 serão fixas e as restantes escolhidas de acordo com o **plano de amostragem** proposto neste trabalho. É de realçar que esta proposta tem por objetivo permitir obter *melhores* estimativas para a abundância de raia-lenga, minimizando o impacto nas estimativas de abundância de pescada. Note-se que as estações de pesca consideradas fixas são aquelas que o IPMA pretende manter ao longo das campanhas, dado o histórico das mesmas. No mapa da Figura 6.1 estão representadas as localizações das estações fixas, confirmando os valores da última linha da Tabela 6.2.

No entanto, o conjunto das estações fixas não garante o critério da representatividade, faltando duas unidades amostrais aos estratos $Arrifana_1$ e $Milfontes_1$, e uma unidade amostral aos estratos $Arrifana_2$, $Berlengas_1$ e $Lisboa_1$ (ver Tabela 6.2 e Figura 6.2). Deste modo, das 11 novas localizações selecionadas debaixo de um plano de amostragem aleatória, sete terão em conta os estratos sub-representados, para cumprir o critério mencionado, e quatro serão selecionadas aleatoriamente considerando toda a costa. Pela

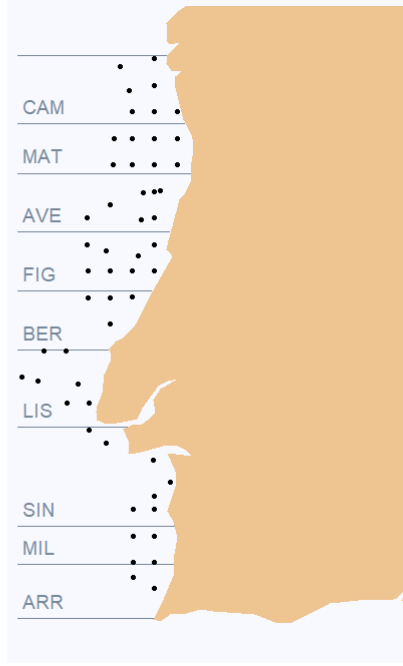


Figura 6.1: Mapa das estações fixas.

Tabela 6.2 verifica-se também que 10 das estações fixas não estão associadas nem a grupo de profundidade 1 nem 2, que são o objeto do nosso estudo. Pelo que as estações do tipo 3 serão desconsideradas do estudo.

6.2 Proposta de plano de amostragem

O plano de amostragem proposto utiliza as estimativas de abundância de ambas as espécies, pela abordagem bayesiana apresentadas na Secção 5.3, para determinar os pesos de probabilidade incluídos na seleção das unidades amostrais. Os pesos de probabilidades são encarados como a probabilidade de a localização \mathbf{s} pertencer à amostra selecionada, e podem ser descritos por:

$$w_{\mathbf{s}} = \frac{z_{\mathbf{s}}^R}{\max_{z^R}} \times \frac{z_{\mathbf{s}}^P}{\max_{z^P}}, \quad \mathbf{s} = \{\mathbf{s}_1, \dots, \mathbf{s}_m\} \quad (6.1)$$

onde $z_{\mathbf{s}}^R$ e $z_{\mathbf{s}}^P$ representam, respetivamente, as estimativas da abundância de *R. clavata* e de pescada na localização \mathbf{s} , e \max_{z^R} e \max_{z^P} são os respetivos máximos. m é o número de localizações que estão disponíveis para a seleção das unidades amostrais. Uma vez que as abundâncias das duas espécies em questão na costa continental portuguesa são de escalas bastante diferentes, cada estimativa é dividida pelo máximo estimado na área de estudo. Enquanto que o produto é justificado pelo facto de privilegiar as localizações onde se estimou a ocorrência de ambas as espécies. Ou seja, caso o valor estimado em determinada localização \mathbf{s} da abundância de uma das espécies seja nulo, a probabilidade

desta ser selecionada é zero independentemente do valor estimado da abundância da outra espécie.

A proposta de futura campanha científica é obtida através do seguinte algoritmo:

1. Selecionar as estações que servem de base para a estimação das abundâncias. Nomeadamente, os 212 pontos referentes às campanhas científicas 2013-2016 para a raia-lenga, e os 101 de 2015-2016 para a pescada.
2. Definir a *mesh* necessária para a abordagem bayesiana, tendo em consideração as localizações selecionadas.
3. Estimar a abundância das espécies dada a *mesh* do ponto 2 e utilizando a profundidade como covariável, ou outras covariáveis que venham a ser consideradas relevantes.
4. Selecionar, de acordo com os pesos de probabilidade dados pela Expressão (6.1), 11 localizações, que em conjunto com as 54 estações de pesca fixas formam a campanha científica proposta.

A campanha científica proposta será composta por 65 coordenadas (longitude e latitude), donde também se conhece o estrato associado a cada estação k , tal como apresentado na Tabela 6.3.

Tabela 6.3: Campanha proposta pelo método de amostragem utilizado.

Estação de Pesca	k	Longitude	Latitude	Estrato
Fixas	1	x_1	y_1	E_1
	\vdots	\vdots	\vdots	\vdots
	54	x_{54}	y_{54}	E_{54}
Aleatórias	55	x_{55}	y_{55}	E_{55}
	\vdots	\vdots	\vdots	\vdots
	65	x_{65}	y_{65}	E_{65}

6.2.1 Limitações do método proposto

Atualmente, a abordagem bayesiana não está a ter em conta o tipo de fundo para modelar a distribuição espacial da abundância de *R. clavata* e de pescada. Tal pode ser considerado uma limitação do método utilizado. Isto porque o tipo de fundo é uma variável importante na identificação dos locais pela qual a raia-lenga tem preferência. De qualquer modo, é de notar que os pesos de probabilidades apresentados na Expressão (6.1) deverão ter em conta implicitamente a informação do habitat preferido pela espécie.

Outra possível limitação inerente ao atual método é não considerar informação sobre se é possível realizar a arte do arrasto nas localizações aleatórias a serem incluídas

na campanha. No entanto, também aqui, as probabilidades em (6.1) deverão de modo implícito ajudar a mitigar este problema. Pois, se a ocorrência foi estimada como elevada numa determinada localização, então tal deverá traduzir-se numa probabilidade também elevada para a arte de arrasto.

6.3 Estudo de simulação

O objetivo do estudo de simulação é avaliar o método proposto para escolha de nova campanha científica.

Para a concretização deste objetivo, ir-se-á introduzir alguma aleatoriedade nas superfícies estimadas de abundância, gerando réplicas do conjunto de dados observados.

6.3.1 Geração de réplicas

A geração de réplicas será feita alterando o ponto 1 do algoritmo apresentado na Secção 6.1 para:

1. Selecionar aleatoriamente cerca de 50% das localizações observadas para a raia-lenga e 50% das localizações observadas para a pescada. Por exemplo, ter-se-á 100 e 50 localizações para a raia-lenga e a pescada, respetivamente.

Ir-se-á considerar um total de 100 réplicas, repetindo 100 vezes o algoritmo da Secção 6.1 com o novo ponto 1. Serão então geradas 100 amostras, de dimensão 150, selecionadas a partir dos dados observados

Para cada uma das 100 réplicas, será feita a proposta de uma campanha científica. Em concreto para avaliar o método sugerido ir-se-á usar a campanha de 2015 como referência, porque apresenta uma percentagem de valores não-nulos mais elevada face às restantes, para o caso da raia-lenga (ver Tabela 3.1).

Consequentemente, ire-se-á comparar os dados observados na campanha de 2015, e as respetivas estimativas obtidas à custa de 100 campanhas simuladas. É de notar que cada campanha simulada é composta pelas 11 localizações escolhidas aleatoriamente e pelas 44 estações de pesca fixas (grupo de profundidade 1 e 2 da Tabela 6.2).

Uma vez que as coordenadas das estações fixas da Tabela 6.2 nem sempre coincidem exatamente com as coordenadas observadas, então os valores dos indicadores de abundância nessas estações fixas serão “aproximados” pelos valores observados nas localizações mais próximas da campanha de 2015. A análise da Figura 6.2 indica que as estações fixas, de grupo de profundidade 1 e 2, estão relativamente próximas das observações de 2015, pelo que parece ser razoável a “aproximação” adotada.

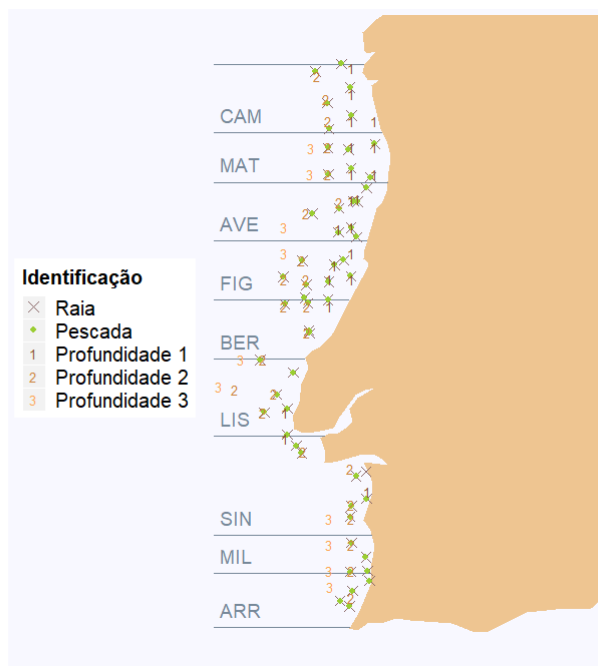


Figura 6.2: Mapa das estações fixas (identificadas por números) e das localizações observadas em 2015 (representadas por símbolos).

6.3.2 Critérios de avaliação

Com base em cada uma das 100 campanhas simuladas, cada composta por 55 estações, ir-se-á estimar os indicadores de abundância de raia-lenga e de pescada sobre as localizações observadas em 2015.

Em seguida, essas estimativas serão comparadas com os respetivos indicadores de abundância observados através de:

- correlação de *spearman* dado pela Expressão (6.2);
- estimador da média dos indicadores de abundância por SRS e da respetiva variância, apresentados em (6.3).

A correlação de *spearman* mede a associação entre duas variáveis e é denotada por:

$$r = 1 - \frac{6 \sum_{i=1}^v D_i^2}{v(v^2 - 1)} \quad (6.2)$$

sendo v o número de lances efetuados na campanha demersal IPMA referente ao ano de 2015 (isto é, $v = 54$ no caso da raia-lenga e $v = 53$ para a pescada) e D_i a diferença entre a ordem dos valores estimados e a ordem dos valores observados.

Por sua vez, o estimador para a média por SRS e a respetiva variância são descritos

por:

$$\bar{z}_{str} = \frac{1}{N} \sum_{h=1}^L N_h \bar{z}_h$$

$$s_{\bar{z}_{str}}^2 = \sum_{h=1}^L \left[\left(\frac{N_h}{N} \right)^2 \left(\frac{N_h - n_h}{N_h} \right) \left(\frac{s_h^2}{n_h} \right) \right] \quad (6.3)$$

onde L representa o número de estratos, N_h indica o tamanho do estrato h tal que $\sum_{h=1}^L N_h = N$ e n_h o número de observações no estrato h . A média amostral do estrato h é denotada por \bar{z}_h e s_h representa o desvio padrão amostral referente ao estrato h .

Neste estudo, serão considerados 18 estratos definidos pelo IPMA, resultantes da combinação do setor e do grupo de profundidade. O tamanho dos estratos representa a área em Km^2 de cada.

6.3.3 Análise dos resultados

Nas Tabelas 6.4 e 6.5 estão apresentados sumários das correlações entre os valores observados z e os valores estimados, \hat{z}_r $r = 1, \dots, 100$, a partir das 100 réplicas dos indicadores de abundância de raia-lenga e de pescada, respetivamente, para a campanha de 2015.

A análise destes resultados deverá ter em conta que os modelos abordados no Capítulo 5 tendem a subestimar o número de zeros, suavizando a superfície estimada. O problema é atenuado com a introdução dos ZIM. A análise das correlações deverá considerar estas dificuldades.

Tabela 6.4: Sumário das correlações entre os valores observados e os valores estimados do indicador de abundância da raia-lenga em 2015.

Medida	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Estimativa	0,38	0,50	0,55	0,57	0,63	0,78

Avaliando os resultados apresentados na Tabela 6.4 verifica-se que em 75 réplicas a correlação entre os valores observados e estimados da abundância da *R. clavata* foi no mínimo de 0,50.

Tabela 6.5: Sumário das correlações entre os valores observados e os valores estimados do indicador de abundância da pescada em 2015.

Medida	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Estimativa	0,49	0,49	0,60	0,59	0,62	0,84

A análise da Tabela 6.5 permite averiguar que a correlação entre os valores observados e estimados, na campanha de 2015, da abundância da pescada varia entre 0,49 e 0,84.

Dado que o plano de amostragem proposto tem em consideração os estratos, é relevante estudar o valor médio dos índices de abundância observados e estimados de ambas as espécies, ponderado pela dimensão dos estratos, e a sua variância.

Deste modo, nas Tabelas 6.6 e 6.8 encontram-se o valor médio \bar{z}_{STR} e a respetiva variância $Var(\bar{z}_{STR})$ para os dados observados na campanha de 2015. Nas Tabelas 6.7 e 6.9 estão apresentados sumários do valor esperado \bar{z}_{STR}^* e a respetiva variância $Var(\bar{z}_{STR}^*)$ para as réplicas simuladas de raia-lenga e de pescada, respetivamente.

Tabela 6.6: Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de raia na campanha de 2015.

Medida	Estimativa
\bar{z}_{STR}	1,87
$Var(\bar{z}_{STR})$	0,73

Tabela 6.7: Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de raia-lenga estimada nas 100 campanhas simuladas.

Medida	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
\bar{z}_{STR}^*	0,94	1,51	1,66	1,68	1,95	2,20
$Var(\bar{z}_{STR}^*)$	0,03	0,16	0,23	0,231	0,30	0,44

Ao comparar as Tabelas 6.6 e 6.7 verifica-se que o valor médio observado da abundância da raia-lenga está dentro da amplitude do valor médio estimado nas 100 campanhas simuladas. Relativamente à variância do estimador do valor médio, esta é superior nos dados observados quando comparada com a resultante de cada uma das réplicas.

Tabela 6.8: Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de pescada na campanha de 2015.

Medida	Estimativa
\bar{z}_{STR}	660,87
$Var(\bar{z}_{STR})$	7736,69

Tabela 6.9: Estimativas das medidas de SRS (Amostragem Aleatória Estratificada) para a abundância de pescada estimada nas 100 campanhas simuladas.

Medida	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
\bar{z}_{STR}^*	614,37	658,51	696,24	690,28	719,59	754,81
$Var(\bar{z}_{STR}^*)$	75,72	119,59	935,40	856,11	1284,53	3450,66

Os resultados para a pescada são idênticos aos da *R. clavata*, estando o valor médio da abundância de pescada observada compreendido entre os valores médio mínimo e máximo

estimados nas 100 réplicas obtidas. Da mesma forma, a variância máxima do estimador do valor médio, de entre as 100 campanhas simuladas, é inferior à variância resultante dos dados observados.

A análise dos resultados dos dois critérios de validação adotados demonstra que o método proposto para a seleção da campanha é plausível. Em particular, os valores das correlações entre os valores observados e estimados, nas 100 campanhas simuladas, indicam que as estimativas dos indicadores de abundância das espécies em estudo debaixo do plano de amostragem proposto são razoáveis. Por outro lado, a comparação dos estimadores da média para os valores observados e estimados nas 100 réplicas mostra que o método proposto vai ao encontro da realidade observada.

Capítulo 7

Conclusões e Trabalho futuro

A consciência ambiental e económica associada às variações temporais dos recursos marinhos são aspetos muito importantes que exigem um conhecimento robusto. A modelação espacial é uma ferramenta importante para melhorar o conhecimento sobre a dinâmica espacial das espécies. Em particular, é fundamental para responder a questões sobre a boa saúde dos ecossistemas e sobre a sustentabilidade da pesca.

A estimação da distribuição espacial dos indicadores de abundância e biomassa da raia-lenga e da pescada, a avaliação do grau de sobreposição espacial das distribuições espaciais dos indicadores de abundância das espécies e a definição de um desenho experimental para estações de pesca foram os objetivos desta dissertação. Apesar de serem estudadas as distribuições espaciais dos índices de abundância de ambas as espécies, o presente estudo focou-se essencialmente na raia-lenga. Dado que a abundância das espécies pode resultar de fatores ambientais, como a profundidade e o tipo de fundo, foi imperativo incluir tais fatores na estimação da distribuição espacial dos índices de abundância e biomassa das espécies em estudo.

Para a concretização dos objetivos estabelecidos foram utilizadas várias abordagens estatísticas, tais como: a caracterização da continuidade espacial através do cálculo de variogramas, a análise da interação espaço-tempo no período de estudo, a estimação da abundância da *R. clavata* e da pescada e a definição de um desenho amostral baseado nas estimativas de abundância das espécies.

A análise da interação espaço-tempo, realizada com base no cálculo do variograma médio e nos resultados provenientes dos testes de hipóteses de *Knox* e *Mantel*, permitiu averiguar que a interação espaço-tempo pode ser considerada desprezável e, por sua vez, procedeu-se à agregação das campanhas de investigação.

A estimação da abundância da raia-lenga foi realizada por três métodos, a suavização de *kernel* espacial, o *kriging* com tendência externa e os ZIM. Para os dois últimos, foram utilizadas as covariáveis profundidade e abundância de juvenis de pescada. No entanto, em *kriging* com tendência externa também foi considerada como variável explicativa o tipo de fundo oceânico. O método dos ZIM é um método robusto em que estimação da

abundância da espécie é feita sob condição da presença da mesma. Na abordagem *kriging* com tendência externa, verificou-se que os tipos de fundo de “lama” e de “lodo com areia” estão associados, em média, a indicadores de abundância e biomassa da raia-lenga mais elevados. Em ambas as metodologias, o aumento da profundidade parece impactar negativamente nos índices de abundância e biomassa da *R. clavata*. Relativamente à distribuição espacial da abundância da raia-lenga, verifica-se que as zonas de Lisboa e Sines se destacam pelo facto de os valores das estimativas dos indicadores de abundância e biomassa serem mais elevados, quando comparadas com as restantes.

Por último, foi apresentado um método de amostragem preferencial com base nas estimativas de abundância da *R. clavata* e da *M. merluccius*, onde através de um estudo de simulação se provou a adequabilidade do mesmo.

Em termos de trabalho futuro realça-se:

- a utilização de modelos aditivos generalizados de forma a que os indicadores de abundância e biomassa possam depender de funções não-lineares de covariáveis;
- a incorporação do tipo de fundo como covariável nos ZIM;
- a integração de informação sobre a costa continental portuguesa, nomeadamente a identificação de zonas não arrastáveis;
- a adoção de outros critérios para validar o método de amostragem proposto.

Bibliografia

- [1] A. Jorgenson and R. Dunlap, “Environmental problems,” *The Wiley-Blackwell Encyclopedia of Globalization*, 02 2012.
- [2] ONU, “The 2030 agenda for sustainable development.” <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>, 2015. 26-06-2019.
- [3] J. B. Jones, “Environmental impact of trawling on the seabed: A review,” *New Zealand Journal of Marine and Freshwater Research*, vol. 26, no. 1, pp. 59–67, 1992.
- [4] M. Stehmann and D. L. Bürkel, “Rajidae,” in *In Fishes of the North-eastern Atlantic and the Mediterranean* (P. J. P. Whitehead, M. BL. Bauchot, J. BC. Hureau, J. Nielse, and E. Tortonese, eds.) (S. c. UNESCO (United Nations Educational and C. Organization), eds.), vol. 1, (Paris), pp. 163–196, 1984.
- [5] E. Comission, “Hake *merluccius merluccius*.” https://ec.europa.eu/fisheries/marine_species/wild_species/hake. 26-06-2019.
- [6] P. A. Walker, *Fleeting images dynamics of North Sea ray populations*. PhD thesis, Faculty of Science, University of Amsterdam, 1999.
- [7] P. Bordalo-Machado, L. S. Gordo, and I. Figueiredo, “Skate and ray species composition in mainland portugal from the commercial landings,” in *Aquatic and Living Resources*, vol. 17, pp. 231–234, 2004.
- [8] C. S. Martin, S. Vaz, B. Ernande, J. R. Ellis, P. Eastwood, F. Coppin, S. Harrop, G. Meaden, and A. Carpentier, “Spatial distributions (1989-2004) and preferential habitats of thornback ray and lesser-spotted dogfish in the eastern english channel,” in *ICES CM 2005/N:23*, 2005.
- [9] IPMA, “Estado dos stocks em 2018 e aconselhamento científico para a sua gestão em 2019.” https://ec.europa.eu/fisheries/sites/fisheries/files/docs/body/op-portugal-fact-sheet_en.pdf, 2018. 29-06-2019.

- [10] A. N. Svetovidov, *Merluccidae*. . In: Whitehead, P. J. P., Bauchot, M. L., Hureau, J. C., Nielsen, J. & Tortonese, E (eds) 1986. In *Fishes of the North-eastern Atlantic and the Mediterranean*, vol. I. Paris: UNESCO, 1986.
- [11] S. J. M. Blaber and C. D. of Marine Research, *Fish and fisheries of tropical estuaries*. London ; New York : Chapman & Hall, 1st ed ed., 1997. Includes bibliographical references (p. [297]-335 and indexes.
- [12] M. Carvalho and I. Natário, *Análise de Dados Espaciais*. Congresso Anual, Sociedade Portuguesa de Estatística, Vila Real, 2008.
- [13] G. P. Robertson, “Geostatistics in ecology: Interpolating with known variance,” *Ecology*, vol. 68, p. 744, 06 1987.
- [14] H. Kondoh, T. Koizumi, and K. Ikeda, “A geostatistical approach to spatial density distributions of sika deer (*cervus nippon*),” *Journal of Forest Research*, vol. 18, pp. 93–100, Feb 2013.
- [15] P. Garcia-Soidan and R. Menezes, “Estimation of the spatial distribution through the kernel indicator variogram,” *Environmetrics*, vol. 23, p. 535–548, 09 2012.
- [16] P. G. Fernandes and J. Rivoirard, “A geostatistical analysis of the spatial distribution and abundance of cod, haddock and whiting in north scotland,” in *geoENV II — Geostatistics for Environmental Applications* (J. Gómez-Hernández, A. Soares, and R. Froidevaux, eds.), (Dordrecht), pp. 201–212, Springer Netherlands, 1999.
- [17] R. Webster and M. A. Oliver, *Geostatistics for Environmental Scientists*. Statistics in Practice, second ed., 2001.
- [18] M. Ridout, C. Demétrio, and J. Hinle, “Models for count data with many zeros. international biometric conference,” *Cape Town*, vol. 13, pp. 1–13, 01 1998.
- [19] P. J. Royle and D. Nychka, “An algorithm for the construction of spatial coverage designs with implementation in s-plus,” *Comput. Geosci.*, pp. 479–488, 1998.
- [20] J. Ver Hoef, “Sampling and geostatistics for spatial data,” *ECOSCIENCE*, vol. 9, pp. 152–161, 01 2002.
- [21] J. Wang, Y. Ge, G. Heuvelink, and C. Zhou, “Spatial sampling design for estimating regional gpp with spatial heterogeneities,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, pp. 539–543, 02 2014.
- [22] J. W. Hardin and J. M. Hilbe, *Generalized Linear Models and Extension*, vol. 2. Stata Press, 2007.

- [23] A. P. Pousinho, “Modelos lineares generalizados tweedie aplicados ao cálculo de provisões para sinistros,” Master’s thesis, Instituto Superior de Economia e Gestão, 2013.
- [24] F. Taroni, A. Biedermann, and S. Bozza, “Statistical hypothesis testing and common misinterpretations: Should we abandon p-value in forensic science applications?,” *Forensic Science International*, vol. 259, pp. e32–e36, 02 2016.
- [25] A. Gorbunova and B. Lemeshko, “Application of parametric homogeneity of variances tests under violation of classical assumption,” 06 2012.
- [26] S. Meyer, I. Warnke, W. Rössler, and H. Leonhard, “Model-based testing for space-time interaction using point processes: An application to psychiatric hospital admissions in an urban area,” 2016.
- [27] P. Petitgas, M. Woillez, J. Rivoirard, D. Renard, and N. Bez, *Handbook of Geostatistics in R for Fisheries and Marine Ecology*, vol. 338. ICES Cooperative Research Report, 2017.
- [28] G. Matheron, *Traité de Géostatistique Appliquée*. Technip, 1962.
- [29] P. Diggle and P. Ribeiro, *Model-based Geostatistics*. Springer, USA, 2007.
- [30] C. V. Deutsch and J. L. Deutsch, “Introduction to choosing a kriging plan,” 2015. In J. L. Deutsch (Ed.), *Geostatistic Lesson*.
- [31] J. Rivoirard, “On the structural link between variables in kriging with external drift,” *Mathematical Geology*, vol. 34, pp. 797–808, Oct 2002.
- [32] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical Modeling and Analysis of Spatial Data*, vol. 101. 01 2004.
- [33] S. Wenger and M. Freeman, “Estimating species occurrence, abundance, and detection probability using zero-inflated distributions,” *Ecology*, vol. 89, pp. 2953–9, 11 2008.
- [34] E. W. Frees, R. A. Derrig, and G. Meyers, *Predictive Modeling Applications in Actuarial Science*, vol. 1 of *International Series on Actuarial Science*. Cambridge University Press, 2014.
- [35] F. Lindgren, H. Rue, and J. Lindst, “An explicit link between gaussian fields and gaussian markov random fields: The spde approach,” 01 2011.
- [36] G. Soutinho, “Estimação do uso e ocupação do solo ao nível das nuts 2 e nuts 3,” Master’s thesis, Universidade do Minho, 2014.

- [37] J. Wang, A. Stein, B. Gao, and Y. Ge, “A review of spatial sampling,” *Spatial Statistics*, vol. 2, pp. 1–14, 12 2012.
- [38] W. G. Cochran, *Sampling Techniques, 3rd Edition*. John Wiley, 1977.
- [39] B. Matern, *Spatial Variation*. New York: Springer-Verlag Berlin Heidelberg, 2 ed., 1986.
- [40] D. J. Brus, “Design-based and model-based sampling strategies for soil monitoring,” in *19th World Congress of Soil Science, Soil Solutions for a Changing World*, pp. 32–34, 2010.
- [41] P. Diggle, R. A. Moyeed, and J. A. Tawn, “Model-based geostatistics (with discussion),” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 47 (3), pp. 299–550, 1998.
- [42] D. J. Stevens, “Spatial properties of design-based versus model-based approaches to environmental sampling,” pp. 119–125, 2006.
- [43] P. Diggle, R. Menezes, and T. Su, “Geostatistical inference under preferential sampling,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 59, pp. 191–232, 03 2010.
- [44] L. Guedes, *Otimização de amostragem espacial*. PhD thesis, Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2008.
- [45] A. Soares, *Geoestatística para as ciências sa terra e do ambiente*. IST Press, 2000.
- [46] A. Journel, “Nonparametric estimation of spatial distributions,” *Journal of the International Association for Mathematical Geology*, vol. 15, pp. 445–468, 06 1983.
- [47] T. Gatrell, T. C. Bailey, P. Diggle, and B. S. Rowlingson, “Spatial point pattern analysis and its application in geographical epidemiology,” *Transactions of the Institute of British Geographers*, vol. 21, pp. 256–274, 01 1996.
- [48] V. Goldsmith, P. G. McGuire, J. H. Mollenkopf, and T. A. Ross, *Analyzing Crime Patterns*. Sage Publications, Inc, 2000.

Apêndice A

Limite das zonas e setores

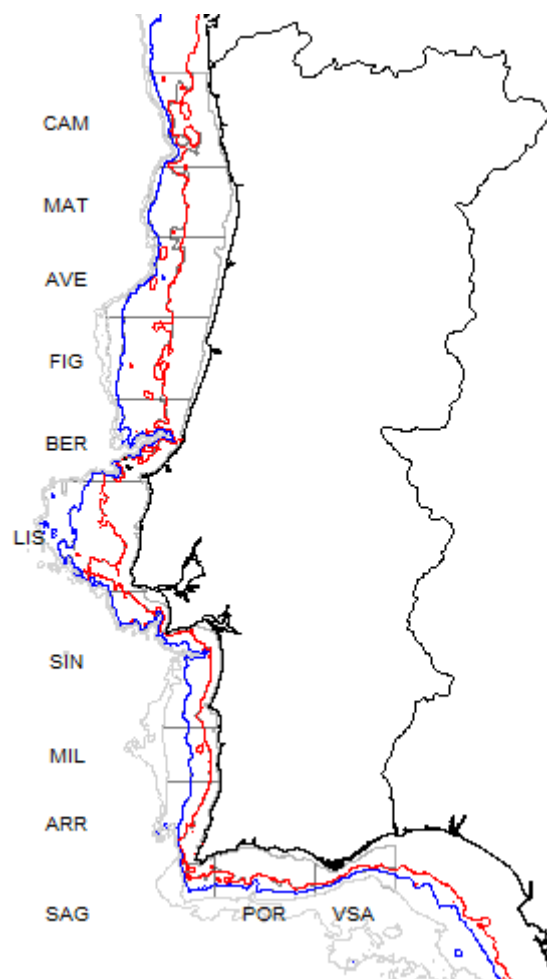


Figura A.1: Mapa da costa continental portuguesa com os estratos (combinação da área geográfica e do grupo de profundidade: 0m-100m e 100m-200m). As áreas geográficas são representados pelas linhas horizontais e as linhas vermelha e azul são as linhas batimétricas dos 100 e 200 m, respetivamente.

Apêndice B

Análise exploratória

Tabela B.1: Resultados do teste de homogeneidade da variância do número de elementos capturados por hora de raia-lenga em cada setor.

Setor 1	Setor 2	Variância (Setor 1)	Variância (Setor 2)	Estatística de Teste	<i>P</i> -valor ($\alpha = 0,05$)
AVE	MAT	6,385	1,228	-0,636	0,525
AVE	CAM	6,385	2,159	-1,648	0,099
AVE	FIG	6,385	0,939	1,070	0,285
AVE	BER	6,385	6,687	-0,017	0,986
AVE	SIN	6,385	27,336	-1,320	0,187
AVE	LIS	6,385	23,538	-0,929	0,353
AVE	MIL	6,385	0,308	1,348	0,178
AVE	ARR	6,385	0,000	1,731	0,083
MAT	CAM	1,228	2,159	-1,583	0,113
MAT	FIG	1,228	0,939	0,714	0,475
MAT	BER	1,228	6,687	-0,591	0,554
MAT	SIN	1,228	27,336	-2,307	0,021
MAT	LIS	1,228	23,538	-1,524	0,127
MAT	MIL	1,228	0,308	1,319	0,187
MAT	ARR	1,228	0,000	1,820	0,069
CAM	FIG	2,159	0,939	-1,170	0,242
CAM	BER	2,159	6,687	-1,901	0,057
CAM	SIN	2,159	27,336	-2,803	0,005
CAM	LIS	2,159	23,538	-2,305	0,021
CAM	MIL	2,159	0,308	-0,051	0,959
CAM	ARR	2,159	0,000	0,924	0,355
FIG	BER	0,939	6,687	-1,170	0,242
FIG	SIN	0,939	27,336	-3,164	0,002

FIG	LIS	0,939	23,538	-1,958	0,050
FIG	MIL	0,939	0,308	0,938	0,348
FIG	ARR	0,939	0,000	1,605	0,108
BER	SIN	6,687	27,336	-0,946	0,344
BER	LIS	6,687	23,538	-0,924	0,356
BER	MIL	6,687	0,308	1,498	0,134
BER	ARR	6,687	0,000	1,941	0,052
SIN	LIS	27,336	23,538	-0,166	0,868
SIN	MIL	27,336	0,308	1,219	0,223
SIN	ARR	27,336	0,000	1,111	0,266
LIS	MIL	23,538	0,308	1,762	0,078
LIS	ARR	23,538	0,000	2,015	0,044
MIL	ARR	0,308	0,000	1,000	0,317

Tabela B.2: Resultados do teste de homogeneidade da variância do peso total de indivíduos capturados por hora de raia-lenga em cada setor.

Setor 1	Setor 2	Variância (Setor 1)	Variância (Setor 2)	Estatística de Teste	P -valor ($\alpha = 0,05$)
AVE	MAT	38,604	6,682	-0,608	0,543
AVE	CAM	38,604	7,632	-1,644	0,100
AVE	FIG	38,604	5,050	1,093	0,275
AVE	BER	38,604	12,852	0,516	0,606
AVE	SIN	38,604	97,177	-0,844	0,399
AVE	LIS	38,604	31,049	-0,410	0,682
AVE	MIL	38,604	0,001	1,470	0,142
AVE	ARR	38,604	0,000	1,725	0,085
MAT	CAM	6,682	7,632	-1,478	0,142
MAT	FIG	6,682	5,050	0,930	0,353
MAT	BER	6,682	12,852	0,350	0,726
MAT	SIN	6,682	97,177	-2,035	0,042
MAT	LIS	6,682	31,049	-0,868	0,385
MAT	MIL	6,682	0,001	1,565	0,118
MAT	ARR	6,682	0,000	1,786	0,074
CAM	FIG	7,632	5,050	-0,957	0,339
CAM	BER	7,632	12,852	-1,596	0,110
CAM	SIN	7,632	97,177	-2,539	0,011
CAM	LIS	7,632	31,049	-1,975	0,048
CAM	MIL	7,632	0,001	0,042	0,967
CAM	ARR	7,632	0,000	0,924	0,355
FIG	BER	5,050	12,852	-0,595	0,552
FIG	SIN	5,050	97,177	-3,001	0,003
FIG	LIS	5,050	31,049	-1,356	0,175
FIG	MIL	5,050	0,001	1,223	0,221
FIG	ARR	5,050	0,000	1,570	0,117
BER	SIN	12,852	97,177	-1,570	0,116
BER	LIS	12,852	31,049	-0,648	0,517
BER	MIL	12,852	0,001	1,728	0,084
BER	ARR	12,852	0,000	1,914	0,056
SIN	LIS	97,177	31,049	-0,567	0,571
SIN	MIL	97,177	0,001	1,266	0,205
SIN	ARR	97,177	0,000	1,054	0,292
LIS	MIL	31,049	0,001	1,898	0,058
LIS	ARR	31,049	0,000	2,008	0,045
MIL	ARR	0,001	0,000	1,000	0,317

Tabela B.3: Resultados do teste de homogeneidade da variância do número de elementos capturados por hora de pescada adulta em cada setor.

Setor 1	Setor 2	Variância (Setor 1)	Variância (Setor 2)	Estatística de Teste	P -valor ($\alpha = 0,05$)
CAM	MAT	436692,600	2237352,710	-2,999	0,003
CAM	AVE	436692,600	86148,420	-0,573	0,567
CAM	FIG	436692,600	233770,070	0,240	0,811
CAM	BER	436692,600	262062,560	-1,102	0,271
CAM	LIS	436692,600	146139,990	0,510	0,610
CAM	SIN	436692,600	22753,990	0,949	0,342
CAM	MIL	436692,600	25626,870	0,542	0,588
CAM	ARR	436692,600	7235,810	0,108	0,914
MAT	AVE	2237352,710	86148,420	-0,090	0,928
MAT	FIG	2237352,710	233770,070	2,879	0,004
MAT	BER	2237352,710	262062,560	2,055	0,04
MAT	LIS	2237352,710	146139,990	2,435	0,015
MAT	SIN	2237352,710	22753,990	1,604	0,109
MAT	MIL	2237352,710	25626,870	0,921	0,357
MAT	ARR	2237352,710	7235,810	0,748	0,455
AVE	FIG	86148,420	233770,070	0,895	0,371
AVE	BER	86148,420	262062,560	0,240	0,810
AVE	LIS	86148,420	146139,990	0,989	0,323
AVE	SIN	86148,420	22753,990	1,741	0,082
AVE	MIL	86148,420	25626,870	0,805	0,421
AVE	ARR	86148,420	7235,810	1,256	0,209
FIG	BER	233770,070	262062,560	-0,777	0,437
FIG	LIS	233770,070	146139,990	0,177	0,860
FIG	SIN	233770,070	22753,990	0,938	0,348
FIG	MIL	233770,070	25626,870	0,797	0,425
FIG	ARR	233770,070	7235,810	0,002	0,998
BER	LIS	262062,560	146139,990	1,035	0,301
BER	SIN	262062,560	22753,990	1,562	0,118
BER	MIL	262062,560	25626,870	1,511	0,131
BER	ARR	262062,560	7235,810	0,749	0,454
LIS	SIN	146139,990	22753,990	0,375	0,708
LIS	MIL	146139,990	25626,870	0,291	0,771
LIS	ARR	146139,990	7235,810	0,157	0,875
SIN	MIL	22753,990	25626,870	-0,096	0,923
SIN	ARR	22753,990	7235,810	0,483	0,629
MIL	ARR	25626,870	7235,810	0,579	0,563

Tabela B.4: Resultados do teste de homogeneidade da variância do número de indivíduos capturados por hora de pescada pequena em cada setor.

Setor 1	Setor 2	Variância (Setor 1)	Variância (Setor 2)	Estatística de Teste	P -valor ($\alpha = 0,05$)
CAM	MAT	428624,456	2207069,933	-2,407	0,016
CAM	AVE	428624,456	78167,815	-0,032	0,975
CAM	FIG	428624,456	170841,426	0,847	0,397
CAM	BER	428624,456	161769,084	0,000	1,000
CAM	LIS	428624,456	130648,486	1,006	0,314
CAM	SIN	428624,456	20811,441	0,989	0,323
CAM	MIL	428624,456	23845,905	0,992	0,321
CAM	ARR	428624,456	5510,667	0,496	0,620
MAT	AVE	2207069,933	78167,815	0,683	0,495
MAT	FIG	2207069,933	170841,426	2,576	0,010
MAT	BER	2207069,933	161769,084	2,240	0,025
MAT	LIS	2207069,933	130648,486	2,670	0,008
MAT	SIN	2207069,933	20811,441	1,914	0,056
MAT	MIL	2207069,933	23845,905	1,374	0,169
MAT	ARR	2207069,933	5510,667	1,374	0,169
AVE	FIG	78167,815	170841,426	0,945	0,345
AVE	BER	78167,815	161769,084	0,802	0,422
AVE	LIS	78167,815	130648,486	1,270	0,204
AVE	SIN	78167,815	20811,441	1,627	0,104
AVE	MIL	78167,815	23845,905	1,074	0,283
AVE	ARR	78167,815	5510,667	1,702	0,089
FIG	BER	170841,426	161769,084	-0,112	0,911
FIG	LIS	170841,426	130648,486	0,461	0,644
FIG	SIN	170841,426	20811,441	0,917	0,359
FIG	MIL	170841,426	23845,905	0,783	0,433
FIG	ARR	170841,426	5510,667	0,465	0,642
BER	LIS	161769,084	130648,486	1,054	0,292
BER	SIN	161769,084	20811,441	1,346	0,178
BER	MIL	161769,084	23845,905	1,317	0,188
BER	ARR	161769,084	5510,667	0,749	0,454
LIS	SIN	130648,486	20811,441	0,459	0,646
LIS	MIL	130648,486	23845,905	0,537	0,591
LIS	ARR	130648,486	5510,667	0,090	0,929
SIN	MIL	20811,441	23845,905	0,031	0,975
SIN	ARR	20811,441	5510,667	0,815	0,415
MIL	ARR	23845,905	5510,667	0,579	0,563

Apêndice C

Resultados da Análise da distribuição espacial de *Raja clavata*

Tabela C.1: Estimativas dos parâmetros envolvidos nas estruturas do valor médio do indicador de biomassa de raia-lenga.

Campanha	Tendência					
	$\hat{\beta}_0$	$\hat{s}(\beta_0)$	$P(\hat{\beta}_0)$	$\hat{\beta}_1$	$\hat{s}(\beta_1)$	$P(\hat{\beta}_1)$
2013	16,45	10,64	0,13	-0,003	0,002	0,13
2014	16,73	12,09	0,17	-0,004	0,003	0,170
2015	0,33	9,30	0,97	0,000	0,949	
2016	31,13	10,59	0,01	-0,007	0,002	0,006

Tabela C.2: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para o indicador de biomassa de raia-lenga em cada campanha.

Campanha	Componente	Parâmetros		
		$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
2013	Exponencial	0	10,73	44,45
2014	Exponencial	0	8,34	59,68
2015	Exponencial	0	37,80	31,87
2016	Exponencial	0	3,14	30,30
	Esférica	0	4,63	55,38

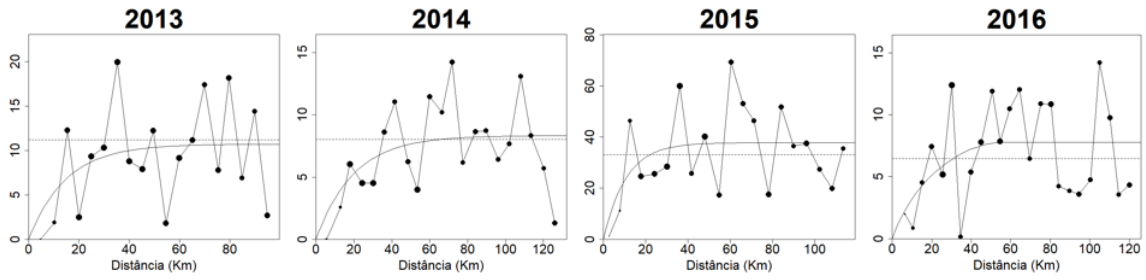


Figura C.1: Variogramas amostrais do indicador de biomassa de raia-lenga capturada por hora em cada campanha.

Tabela C.3: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para a ocorrência de raia-lenga em cada campanha.

Campanha	Componente	Parâmetros		
		$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
2013	Exponencial	0	0,17	35,45
2014	Gaussiana	0	0,18	27,44
2015	Exponencial	0,11	0,11	18,27
2016	Gaussiana	0	0,18	45,44

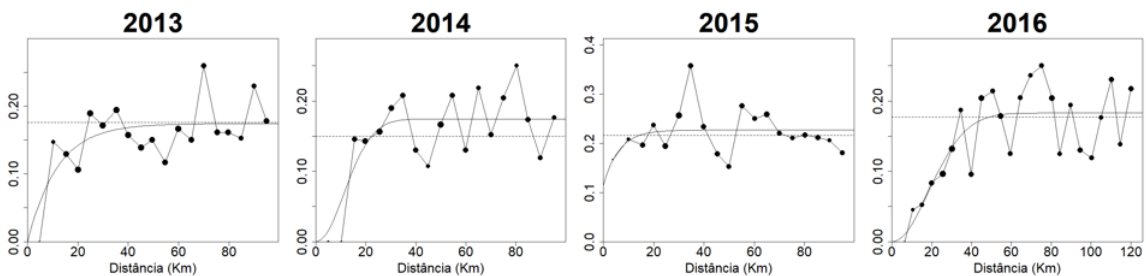


Figura C.2: Variogramas amostrais da ocorrência de raia-lenga em cada campanha.