Near real-time distribution modelling of cetacean distribution off the Azores

Tese de doutoramento

Marta Tobeña Morcillo

Doutoramento em

CIÊNCIAS DO MAR

SICUT AURORA SCIENTIA LUCET
UNIVERSIDADE DOS AÇORES

Horta
2022

Near real-time distribution modelling of cetacean distribution off the Azores

Tese de doutoramento

Marta Tobeña Morcillo

Orientadores

Doutora Mónica Cordeiro de Almeida e Silva, Okeanos- Instituto de Investigação em Ciências do Mar, Universidade dos Açores

Professora Doutora Maria Manuela Fraga Juliano, Universidade dos Açores

Doutor Patrick Lehodey, Pacific Community and Mercator Ocean

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências do Mar

This thesis was supported by the doctoral grant M3.1.a/F/072/2015 from the Fundo Regional da Ciência e Tecnologia (FRCT).

And co-funded by the following research projects and institutions:
TRACE (PTDC/ MAR/74071/2006)
MAPCET (M2.1.2/F/012/2011) and 2020 (M2.1.2/I/026/2011)
WATCH IT (ACORES-01-0145-FEDER-000057)
CMEMS Service Evolution GREENUP project funded by Mercator Ocean

GOVERNO
DOS AÇORES

FRCT
FUNDO REGIONAL PARA A CIÊNCIA E TECNOLOGIA

FCT
FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DOS AÇORES

MERCATOR
OCEAN
INTERNATIONAL

Pacific
Community
Communauté
du Pacifique

SRCTE
SECRETARIA REGIONAL DA CIÊNCIA, TECNOLOGIA E EQUIPAMENTOS

CLS
COLLECTE LOCALISATION SATELLITES

GOVERNO DE
PORTUGAL
MINISTÉRIO DA EDUCAÇÃO
E CIÊNCIA

Host Institutions:
Institute of Marine Research, Azores (IMAR)
Okeanos – Instituto de Investigação em Ciências do Mar, Universidade dos Açores

UAc
UNIVERSIDADE
DOS AÇORES

OKEANOS

imar

ACKNOWLEDGMENTS

I am thankful to my three supervisors, Mónica Silva, Manuela Juliano and Patrick Lehodey, for agreeing to supervise this thesis. Thank you very much for your support and patience.

I couldn't have done Chapter 2 without the shared data from the whale-watching operators, thanks for your collaboration.

Thank you so much Rosa Crujeiras for taking the time you I need to understand things and giving me help with statistical doubts about bimodal distribution and Chiara Mandl for help with the whale-watching dataset, which are decisive in Chapter 2. Thank you so much Rui Prieto, your concepts and suggestions on AIS data have helped me developing Chapter 2 and 3.

Sergi Pérez-Jorge introduced me to the world of GAMs, which was crucial for developing the cetacean models.

Thanks for your hard work to all members of POPA and the program director; without it, there would be no cetacean data to analyse, and I would not have made the models for this thesis.

Ricardo Medeiros, thank you for giving me honest feedback to help me improve my skills with ArcGis.

I thank the Collecte Localisation Satellites for access to the SEAPODYM reanalyses variables, used in the Chapter 4 and 5. Special thanks are owed to Patrick Lehodey and Beatriz Calmettes for provide simulated prey data and help with the database and the concepts.

Therefore, this research has been based on the work and collaboration of many people.

I also want to thank the Azores Whale Lab team, Mónica, Rui, Cláudia, Maria João, Irma, Sergi and Miriam and all the people who work now or worked there for built a nice environment to research. Without this, I would not have decided to write a thesis; thank you all very much.

I was supported by a doctoral grant M3.1.a/F/072/2015 from the Fundo Regional da Ciência e Tecnologia (FRCT). Part of the work presented here was supported by projects TRACE (PTDC/ MAR/74071/2006), MAPCET (M2.1.2/F/012/2011), 2020 (M2.1.2/I/026/2011), WATCH IT (ACORES-01-0145-FEDER-000057), and CMEMS Service Evolution GREENUP project funded by Mercator Ocean

I was already warned that doing a thesis was a challenging and absorbing process. I always thought the thesis complexity was related to intense work, with many hours of work and marathon days. Besides, there are unforeseen in the projects, the animals are missing or the sea condition is not good enough, and your own life. A researcher is also a person; it is already known that living beings depend on their health. Therefore, in addition to thanking all the people who have helped me academically to carry out this work. I also have to thank all those who have helped me improve my life quality in record time, to restructure all my life habits, food, hobbies, exercise routines, and friendliness. Without them, I would not have been able to adapt quickly to so many changes or do this thesis.

I want to thank family and friends who have strengthened the bond between us during a ton of changes. Os quiero mucho a todos.

Estoy muy agradecida a mis padres por cómo me han criado, con amor y cariño, y por cómo han luchado para que yo pudiera seguir mis propios pasos. Muchas gracias por todo os quiero mucho . Tengo que agradecerle a mi tío y padrino muchas cosas, pero últimamente ha conseguido transformar idas y esperas en hospitales en momentos divertidos e inolvidables, no hay mucha gente que consiga hacer eso .

Quien me iba a decir que mis ojos cambiarían de color, de verde a azul y marrón, muchas gracias Mariona, Núria y María por ser mis ojos, incluso con olas de calor cuando uno lo único que quiere es perderse en la playa. No me voy a olvidar de esto en la vida.

Moltes gràcies, muchas gracias, muito obrigado and thanks to my friends for having my back, especially Ana, Anita, Alves, Carla, Cristina, Isaac and Silver.

I could not have finish this thesis without the love and dedication of my partner.

ABSTRACT

The cetaceans have inhabited the mid-Atlantic Ocean for a long time, and the Azorean community has depended on them for centuries. At the beginning of the last century, the Azorean community were whalers, but after the whaling moratoria, the tourism business growth around the cetaceans. In addition, cetaceans have an essential role in the environment, supporting the integrity of the marine ecosystem.

Nevertheless, the processes that drive their presence in the Azores are complex, and identifying them is a challenge. Therefore, there are multiple options: ecological processes, anthropogenic activities, or blended.

Clear up the relationships between cetacean and their drivers is essential to predict how the changes in the environment and anthropogenic activities can influence the group's resilience and the service ecosystem support.

The species distribution maps will help to improve the efficiency of cetaceans sightings search without increasing the impact on the animals in two ways, to select lookout places and identify with precision the most relevant areas for cetaceans.

The cetacean distribution is also crucial for managing other marine activities like shipping traffic, which could use the information to implement measures to avoid collisions with cetaceans. Therefore, the maps will help improve the marine economy's sustainability.

Most of the cetacean distribution models in the literature focus on environmental variables because they are relevant and available. Here to create more realistic models, the candidate variables include, besides the traditional variables, prey and anthropogenic variables. SEAPODYM provides the simulated prey variables. However, the anthropogenic variables are built in this thesis, including a method to detect when the cetacean watching boats are in a cetacean sighting.

Besides the model evaluation, the animal exposure to the cetacean watching and the risk of lethal collision of cetaceans with shipping traffic are analysed.

The models, maps and method produced in the thesis are helpful tools for managing the marine spatial in the Azores, which will help implement the blue growth agenda. Besides, the results are essential to cetaceans in other areas and for other taxa.

RESUMO

Os cetáceos têm habitado o Oceano Atlântico desde há muito tempo, e a comunidade açoreana tem dependido deles por séculos. No início do último século, a comunidade açoreana eram balieiros, mas depois da moratória de caça de baleias, o negócio do turismo cresceu a volta dos cetáceos. Além disso, os cetáceos têm um papel essencial no meio ambiente, sustentando à integridade do ecosistema marinho. Os processos que guiam a sua presença nos Açores são complexos e identificá-los é um reto. Há multiplas opções: processos ecológicos, atividades antropogénicas, o a mistura deles.

Encontrar as relações entre cetáceos y os seus geradores de presença é essencial para predizer como as alterações no ambiente e das atividades humanas podem influir na resiliência de grupo e no serviço deles ao ecosistema.

Os mapas de distribuição de espécies vai ajudar a melhorar a eficiência dos avistamentos de cetáceos sem incrementar o impacto nos animais de duas maneiras, a escolher os lugares de vigias e a identificar com mais precisão as áreas relevantes para os cetáceos.

A distribuição dos cetáceos é também importante para a gestão de outras atividades marinhas como o tráfego marítimo, que poderia utilizar esta informação para implementar medidas para evitar as colisões com cetáceos. Então, os mapas vão ajudar a melhorar a sustenibilidade da economia marinha.

Muitos dos mapas de distribuição de cetáceos na literatura estão desenvolvidos focados nas variáveis ambientais porque são relevantes e disponivéis. Aquí para desenvolver modelos mais reais, para além das tradicionais foram utilizadas variáveis de presa y de atividades humanas. O modelos SEAPODYM proporcionou presas simuladas. Mas as variavéis de atividades humanas foram desenvolvidas nesta tese, além dum método para detetar quando os barcos de observação de cetáceos estão num avistamento de cetáceos.

Além da evaluacão dos modelos, também se analisou a exposição dos cetáceos à atividade de observação de cetãceos e o risco de colisão letal com o tráfego maritimo.

Os modelos, mapas e o método producidos nesta tese son ferramentas úteis para a gestão do espaço marinho nos Açores, podem ajudar a implementar a “Blue Growth agenda”. Os resultados são essencias para os cetáceos na região mas também em outras áreas e para outras taxa.

Table of contents

Acknowledgments ii

Abstract iv

Resumo v

CHAPTER 1

1.1. Background 1

1.2. The Azores: a hotspot for large whales 2

1.3. Modelling whale distribution in the Azores 3

1.4. Anthropogenic threats to large whales in the Azores 5

1.5. Objectives and outline 6

1.6. References 7

CHAPTER 2

2.1. Introduction 14

2.2. Methods 16

2.2.1. Data pre-processing 17

2.2.2. Trip Identification Algorithm 21

2.2.3. Data exploration 22

2.2.3.1. Time interval standardization 22

2.2.3.2. Boat activity by home port 22

2.2.4. Identifying cetacean watching behaviour 23

2.2.4.1. Method development 23

2.2.4.2. Validation method 23

2.2.4.3. Cetacean watching effort maps 24

2.3. Results 24

2.3.1. Data exploration 24

2.3.1.1. Time interval standardization 24

2.3.1.2. Boat activity by home port 26

2.3.2. Identification of cetacean watching behaviour 27

2.3.2.1. Method development 27

2.3.2.2. Validation method 29
2.3.2.3. Cetacean watching effort maps 29
2.4. Discussion 32
2.4.1. Cetacean watching effort 34
2.5. Conclusions 34
2.6. References 35

CHAPTER 3

3.1. Introduction 40
3.2. Methods 41
3.2.1. AIS data 41
3.2.2. Data processing 42
3.2.3. Analysis of Maritime traffic 44
3.2.3.1. Shipping density 44
3.2.3.2. Shipping density with higher probability of lethal collisions 45
3.2.3.3. Transformed shipping density 45
3.3. Results 46
3.3.1. Number of vessels and trips 46
3.3.1.1. Total shipping density vs lethal probability (LP) shipping density 50
3.3.2. Spatial and temporal patterns in shipping density 52
3.3.3. Spatial and temporal patterns in LP shipping density 56
3.4. Discussion 60
3.4.1. Caveats and limitations 60
3.4.1.1. Shipping density 61
3.4.1.2. Shipping density with high probability of lethal collisions with cetaceans 62
3.5. Conclusions 63
3.6. References 64

CHAPTER 4

4.1. Introduction 69
4.2. Methods 71
4.2.1. The study area 71
4.2.2. Cetacean survey data 71
4.2.3. Habitat and prey data 73

4.2.4. Modelling ... 74
4.2.4.1. Model evaluation ... 75
4.2.4.2. Model predictions ... 75
4.3. Results ... 79
4.3.1. Species habitat preferences ... 79
4.3.1.1. Sperm whales ... 79
4.3.1.2. Fin whale ... 79
4.3.1.3. Blue whale ... 79
4.3.1.4. Sei whale ... 79
4.3.2. Predicted relative density ... 81
4.3.2.1. Sperm whale ... 81
4.3.2.2. Fin whale ... 81
4.3.2.3. Blue whale ... 81
4.3.2.4. Sei whale ... 82
4.4. Discussion ... 94
4.4.1. Model limitations and Predictive performance ... 94
4.4.2. Habitat preferences ... 95
4.4.3. Predicted relative densities of large whales ... 98
4.5. Conclusion ... 99
4.6. References ... 101

CHAPTER 5

5.1. Introduction ... 108
5.2. Methods ... 111
5.2.1. The study area ... 111
5.2.2. Cetacean survey data ... 111
5.2.3. Environmental data ... 112
5.2.4. Anthropogenic data ... 112
5.2.5. Influence of anthropogenic activities on species distribution ... 113
5.2.5.1. Model development ... 113
5.2.5.2. Model selection and evaluation ... 113
5.2.5.3. Model predictions ... 114
5.2.6. Assessing the risk of exposure to anthropogenic activities ... 114

5.2.6.1. Risk of lethal ship collisions 114
5.2.6.2. Exposure to cetacean watching 115
5.2.6.3. Assessing the amount of preferred habitat exposed to cetacean watching activity 115
5.3. Results 115
5.3.1. Influence of shipping density on the distribution of sperm and fin whales 115
5.3.2. Risk of lethal ship collisions 121
5.3.3. Exposure to cetacean watching 124
5.3.4. Assessing the amount of preferred habitat exposed to cetacean watching activity 124
5.4. Discussion 124
5.4.1. Caveat of the approach 124
5.4.2. Influence of anthropogenic activities on species distribution 128
5.4.2.1. Risk of lethal ship collisions 129
5.4.3. Exposure to cetacean watching 130
5.4.4. Cumulative exposure to shipping collisions and cetacean watching 131
5.4.5. Management and conservation implications 131
5.5. Conclusion 132
5.6. References 133

CHAPTER 6

6.1. Cetacean watching and marine traffic patterns 142
6.2. Modelling whale distribution 143
6.2.1. Incorporating prey covariates in whale models 143
6.2.2. Whale distribution and habitat preferences 145
6.2.3. Effect of marine traffic on whale distribution 146
6.2.4. Whale exposure to cetacean watching and risk of lethal ship collisions 147
6.3. Final considerations 148
6.4. References 149

LIST OF FIGURES AND TABLES

List of figures 154
List of tables 157

APPENDICES

Supplementary data of Chapter 2 160
Appendix A. Time interval 160
Appendix B. Bimodality results... 160
Supplementary data of Chapter 3 161
Appendix C. Land stations 161
Appendix D. NA in raw data 161
Appendix E. Process to detect duplicate vessel names: 162
Appendix F. Process to detect systematic errors in vessel coordinates 164
Appendix G. Harbour boxes.. 164
Appendix H. Inter-annual patterns of mean daily shipping density on cruise from 2013 to 2016... 165
Supplementary data of Chapter 4 165
Appendix I. Train and test data summary 165
Appendix J. Supplementary maps 167
Supplementary data of Chapter 5 169
Appendix K. Results of the discarded models 169
References 172

CHAPTER 1

1. INTRODUCTION

1.1. BACKGROUND

Cetaceans have moved the curiosity of humans since ancient times [1] but for many centuries they were seen exclusively as a resource to be exploited [2]. Today, the role of cetaceans in maintaining the delicate balance of marine ecosystems is increasingly appreciated [3].

Due to their huge body sizes and high consumption rates, large whales (baleen whales (Mysticete) and the sperm whale, *Physeter macrocephalus*) cetaceans can exert strong effects on the energy flow in marine food webs through direct predation or trophic cascades [4]. Therefore, the extraction of large whales from the trophic chain, as of other large predators, can significantly alter the structure and dynamics of marine food webs and the functioning of marine ecosystems [5]. For example, the reduction of whale populations during whaling, in combination with possible environmental alterations, is believed to have contributed to a 400% increment in biomass of pollock (*Pollachius virens*) in the Bearing Sea [6]. Large whales are important vectors of organic material and nutrients [7, 8]. Through their high mobility and ocean-wide migrations, they can contribute to the horizontal transport of biomass and nutrients. They also play an important role in nutrient recycling, by feeding at depth and releasing nutrients-rich faecal plumes in photic waters, which can stimulate phytoplankton growth and influence ecosystem dynamics [9]. In addition, the carcasses of whales that sink to and rest on the bottom of the oceans provide an abundant food source to various organisms and become colonized by rich macrofauna communities [7]. Therefore, large whales play a crucial role in fundamental ecological processes, such as the provisioning, transfer and cycling of biomass and nutrients, shaping biodiversity and structure of pelagic and deep-sea communities, and climate regulation [10].

Most large cetacean species were severely overexploited by commercial whaling over the last two centuries [11]. Since the international moratorium on commercial whaling in 1986, some populations have recovered but most have not yet reached their pre-whaling numbers [12-15]. These whales now experience numerous other pressures including ship collisions, entanglement in fishing gear, exposure to noise (e.g., shipping traffic, seismic exploration, military sonars), chemical, nutrient, and plastic pollution, changes in food webs from overfishing and introduction of invasive species, disturbance from maritime activities, possible impacts from global climate change, among others [16]. These pressures, acting singly or in combination, threaten the recovery of large whales, and can affect the services provided by marine ecosystems [17, 18].

Large whales range widely but information on their distribution, as well as on potential impacts from human activities, are only available for limited parts of a species' distributional range. This lack of knowledge prevents an understanding of the consequences of anthropogenic impacts at the population level, and is a major impediment to the development and implementation of conservation actions at relevant geographic scales [19].

1.2. THE AZORES: A HOTSPOT FOR LARGE WHALES

The Azores is the most remote archipelago in the North Atlantic, distancing about 1,000 nm from continental Europe and 3,000 nm from North America (Figure 1). The archipelago consists of nine volcanic islands divided into three groups: Central (Pico, Faial, S. Jorge and Terceira islands), Eastern (S. Miguel and Sta. Maria islands) and Western (Corvo and Flores islands). The archipelago is located on the Mid-Atlantic Ridge (MAR), where the African, Eurasian and North American tectonic plates converge. The Exclusive Economic Zone (EEZ) of the Azores, of ~1 million km^{2,} encompasses diverse topographic features, including the narrow island shelves, steep slopes, a fraction of the Mid-Atlantic Ridge, and vast expansions of deep ocean (>3000 m) areas with numerous seamounts (~480) [20]. The region is characterized by complex ocean circulation patterns resulting from the interaction of the North Atlantic Current (in the North) and the Azores Current (in the south) with many unstable eddies and meanders [21-25]. The Azores Current depends on the subtropical Azores Front has branches with seasonal variations [26]. This front defines the limit with the South Subtropical Mode Water, warmer than in the north [27]. Waters are oligotrophic and most of the nutrients are expected to emerge by dynamic processes like oceanic filaments or eddies [28]. This unique blend of a dynamic oceanography interacting with high seafloor complexity in the middle of the North Atlantic basin is thought to provide the particular conditions which attract pelagic megafauna [29], namely large whales. The richness of the area motivated several Marine Protected Areas (MPAs) under regional, national or international regulations (Rede Natura 2000, Ospar, IBA, NEAFC, Gift to the Earth) [30] (Governo dos Açores ,2021).

Sperm whales are distributed worldwide but males and females have distinct ranges [31]. Females stay in tropical and subtropical waters year-round, where they live in long-term social groups with immature and calves of both sexes, while males disperse from their natal group and move to higher latitudes, reaching polar waters [31]. Adult males migrate periodically to the warm waters inhabited by females to mate [32].

The waters around the Azores archipelago are an important feeding, calving and possibly mating ground for sperm whales in the North Atlantic [33, 34]. Whales of both sexes and all age classes use the area year-round, but the majority of the observations consist of social groups, formed by adult females, and their immatures and calves, and occur in late spring and summer [34]. Sperm whale social groups are nomadic, and the Azores encompasses only a part of their range. Although the movements of the population are not well known, the information currently available suggests that this population may have its core habitat within Macaronesian waters (Azores, Madeira and Canary Islands; Steiner et al. 2015 [35]).

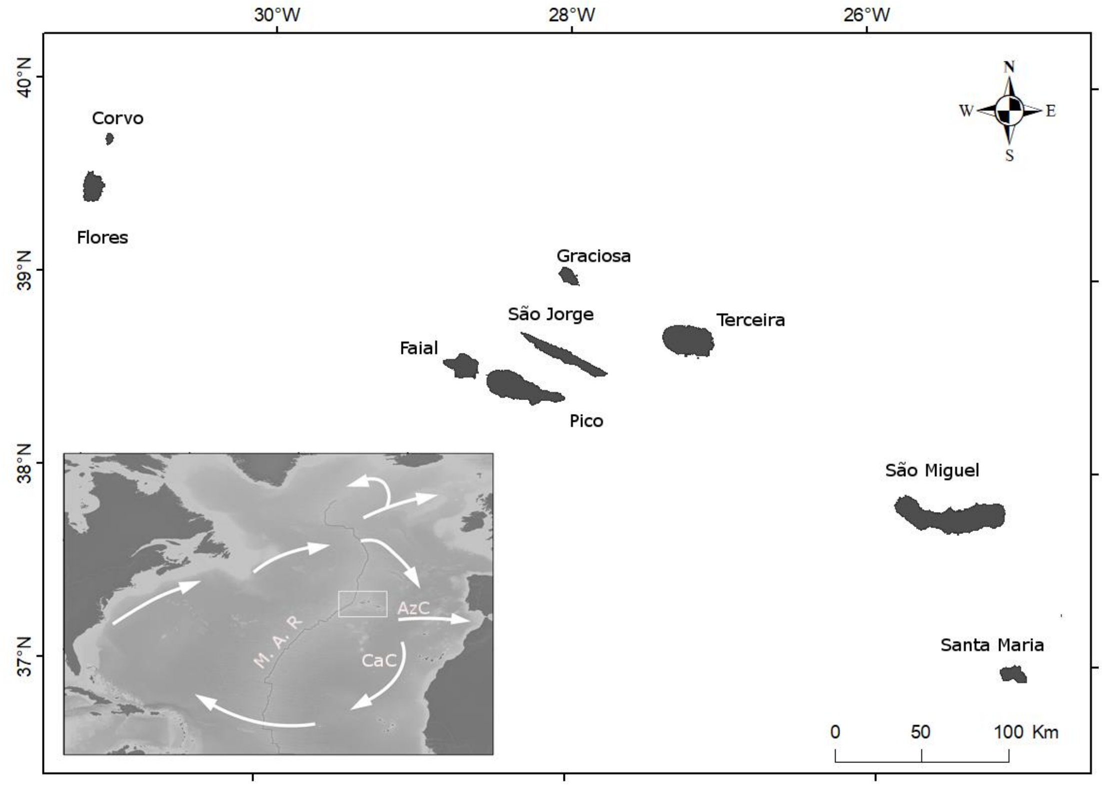


Figure 1. Study area map. Az: Azores Current, CaC: Canary Current. M.A.R.: Mid-Atlantic Ridge. 9 islands: Corvo, Flores, Faial, Pico, São Jorge, Graciosa, Terceira, São Miguel and Santa Maria island.

Blue (*Balaenoptera musculus*), fin (*B. physalus*) and sei (*B. borealis*) whales are also cosmopolitan, found in all ocean basins, especially in offshore waters [36]. The Azores are also an important migratory and mid-latitude foraging habitat for North Atlantic blue, fin and sei whales [37-39]. Earlier studies using satellite tags showed that blue and fin whales remained foraging around the Azores for periods of up to a few months, before resuming their migration towards the central and northeastern Atlantic waters (e.g., Greenland, Iceland British Islands) [37], whereas sei whales left the Azores shortly after tagging and continued their northbound migration towards the Labrador Sea [38].

Notwithstanding, the patterns and underlying mechanisms of sperm whale and baleen whale presence and distribution in the region are still not understood.

1.3. MODELLING WHALE DISTRIBUTION IN THE AZORES

Shipboard or aerial line-transect surveys have been proved the most adequate tool to obtain information on cetacean distribution and density. So far, a single line-transect survey was conducted in the Azores, covering only a small part of the waters around the Azores [40]. Use of opportunistic data can provide reliable estimates of cetacean distribution and absolute or relative density, as long as appropriate statistical methods are employed to address bias associated with the opportunistic sampling [41].

The Azorean Fisheries Observer Program (POPA) is a fishery-based data collection program that collects cetacean sighting and effort information in the Azores since 1999. The potential of POPA dataset to monitor cetacean populations is well established [42, 43] but use of these data requires advanced processing and modelling to minimize bias from the extreme heterogeneous distribution of sampling effort. Species distribution models (SDMs) relate species' occurrence data with relevant environmental variables and enable projecting their potential distribution over unsampled areas [44, 45]. The models' reliability depends partly on the sample size of species' data [46], on the availability in space and time of environmental variables from the sampling area [47], and on the influence of these variables on the target species [45].

Traditionally, the primary sources of environmental data for whale SDMs have been physiographic data (bathymetry and derived variables), and in-situ or remotely sensed oceanographic data (e.g., sea surface temperature or chlorophyll a concentration), used as proxies of prey availability and distribution [45, 48]. However, such distal predictors have limited explanatory power due to potential large lags between oceanographic processes and biological response at trophic levels relevant to whale ecology [49]. Prey abundance and quality directly influence whale distribution [50, 51], and as such should ideally be included in most cetacean SDMs as proximal predictors. However, examples of whale SDMs incorporating prey are still rare since those data are extremely challenging to obtain and are rarely unavailable at the spatiotemporal scales used to build the models [49].

Recent advances in ecosystem models enable overcoming these limitations, by providing spatially-explicit predictions of low to mid-trophic organisms to incorporate into predator SDMs. The Spatial Ecosystem And Population Dynamics Model (SEAPODYM) [52, 53] is one of such models; it is driven by biophysical variables (including ocean currents, temperature and primary productivity) and uses a system of advection–diffusion–reaction equations to generate spatiotemporally resolved predictions of production and biomass of mesozooplankton and six functional groups of mid-trophic level (micronekton). SEAPODYM has been applied to successfully model the distribution of blue, fin and sei whales during their migration from the Azores to their northern foraging grounds [54], to explain fin whale acoustic detections and vocal behaviour in the Azores [55], and to model delphinid usage of two Azorean seamounts [56]. In addition, prey simulations have been used to develop SDMs for sperm whales, baleen whales, and other cetaceans, in various other regions [57, 58].

Incorporating prey covariates into SDMs of large whales can therefore improve spatial and temporal predictions of whale distribution, contributing to a better understanding of their ecology and enabling to predict and manage potential impacts from human activities in the area.

1.4. ANTHROPOGENIC THREATS TO LARGE WHALES IN THE AZORES

Entanglements in fishing gear and collisions with ships are the two major sources of injury and direct mortality for large whales worldwide [16]. While records of whales entangled in the Azores are extremely rare [59], the number of ship collisions with sperm and baleen whales, as well as of observations of whales with recent injuries likely caused by interaction with ships, has increased in recent years [60].

The demand for maritime transport is growing worldwide. In the last decade, the number of large merchant vessels went from 83,000 to 99,000 [61]. The Azores archipelago is located nearby major shipping routes connecting both sides of the Atlantic Ocean and the main Azorean harbours are used by some of these vessels. In addition, local maritime traffic links the islands of the Azores, the Azores islands to the mainland, and other countries [62]. Marine traffic is expected to continue increasing, increasing collisions with megafauna [63], chemical and underwater noise pollution [64, 65], gas emissions [66], and transport of non-indigenous species [67]. Large whales are particularly vulnerable to ship collisions, because they are large and slow-moving [68, 69], spend extended time near the surface [70, 71], and transverse heavy traffic areas during their seasonal migrations and movements across ocean basins [72]. Ship collisions have become the main cause of large whale mortality in several regions around the world, leading the International Whaling Commission to identify ship strikes as a priority conservation concern back in 2005 [73].

A major challenge to estimating large whale mortalities or injuries from ship collisions in the Azores (and elsewhere) is the fact that many interactions are not reported and, because they occur far from the islands, carcasses remain floating and never wash ashore, or they sink. Hence, the known collision cases most likely underestimate the true number of whale strikes in the Azores. Given the need to identify where hidden ship collisions may be occurring, estimating collision risk from ship density patterns and whale distributions are required.

Most previous studies estimating ship-whale collision risk for various locations around the world overlayed a map of shipping traffic (i.e., using data from a single year) with predictions of whale distribution averaged over several years and months [74-76]. This approach cannot explicitly account for the spatiotemporal variability in shipping patterns, whale distributions, or both. Thus, near real-time modelling tools are necessary to provide more accurate predictions of collision risk, as well as to evaluate the effectiveness of static area-based management approaches, commonly used to mitigate collision risk [76, 77].

Until recently, human activities causing non-lethal impacts on cetaceans, such as cetacean watching activities, were assumed to have insignificant effects [78-80]. However, in the last three decades, it has been discovered that long-term effects from non-lethal activities can adversely impact population viability [81-83]. Presence and noise from cetacean watching boats can induce avoidance behaviours, disrupt critical activities, and increase stress levels [84]. Noise from boats can mask important sounds and reduce the functional acoustic ranges of cetaceans, interfering with their ability to receive and transmit acoustic information [65, 85]. Repeated exposure of whales to cetacean watching disturbance may lower reproductive and foraging success, ultimately impacting population dynamics [78].

The growth of the cetacean watching activity in the Azores has increased concerns for potential impacts from repeated disturbance to some cetacean species. Sperm whales may be particularly vulnerable, not only because they are the main target of the activity, but because some social groups use the area for foraging, nursing and mating for several months [86]. Baleen whales could also be adversely affected if repeated exposure to cetacean watching interfered with their ability to forage successfully and regain energy during their spring migration [37]. If the proportion of habitat or individuals exposed to whale-watching is relatively small, negative effects of the activity at the population level could be minor. Investigating the overlap of cetacean watching activity with predicted distributions of whales and quantifying the amount of non-exposed habitat, may provide important insights into the potential impacts of the activity on whale populations using the Azores

1.5. OBJECTIVES AND OUTLINE

The main goals of this thesis are to produce accurate predictions of the distribution of sperm, blue, fin and sei whales in the waters around the Azores, and to assess their exposure to cetacean watching and the risk of fatal collisions with large vessels. To achieve these objectives, a series of modelling tools will be developed to characterize the spatiotemporal patterns of cetacean watching, shipping traffic, and the relative density of the four whale species. Then, these outputs will be combined to identify the areas of greatest concern for the conservation of these whales species in the region. The purpose of this thesis is to contribute to a better comprehension of the ecology and conservation needs of sperm, blue, fin and sei whales in the Azores, and to provide a modelling approach and tools to support cetacean conservation and management of marine activities in the region.

Beyond the first chapter, which aims to provide a general background and some insights into the methodologies used throughout the work, this dissertation is organized into four research chapters.

Chapter 2 reports the development and validation of a method to identify the behaviour of cetacean watching boats during cetacean observation, using only vessel positioning data. This chapter presents the outcomes of the application of the method to available data to produce effort maps of the activity for cetacean watching zone A. The tools created in this chapter are transferable to other regions with cetacean watching activities. The results of this chapter will be the basis for studying cetacean exposure to cetacean watching activity in Chapter 5.

Chapter 3 uses AIS data to characterize and map shipping traffic for all ships and for ships with high risk of lethal collisions with large whales. The outputs are used in Chapter 5 to examine the influence of shipping density on the distribution of large whales and to estimate their risk to lethal collisions.

Chapter 4 investigates the spatial and temporal distribution of sperm, blue, fin and sei whales in the Azores waters using sighting and effort data collected by the POPA (2001-2016), environmental variables, and simulated prey by the SEAPODYM model.

Chapter 5 incorporates shipping traffic calculated in Chapter 3 into SDMs of large whales to investigate the influence of this stressor on whale relative density. In addition, results from Chapter 2 are overlapped with predictions of whale distribution (Chapter 4) to identify areas of higher cetacean watching exposure and quantify the amount of habitat available free of this activity for the four species. Finally, predictions of shipping traffic (Chapter 3) and whale distribution (Chapter 4) are combined to produce estimates and identify areas of potential fatal collision risk by species. These results are discussed in an integrated way to present the areas of highest concern for these large whales in the study area.

Chapter 6 discusses the main problems and limitations, and summarizes and discusses key findings of the work in light of available knowledge on the ecology and conservation of sperm, blue, fin and sei whales.

1.6. REFERENCES

1. Castro A, Zubimendi M, Ambrústolo P, Mazzitelli L, Beretta M, Ciampagna L, et al., editors. Sitio Cueva del Negro: Un caso de aprovechamiento intensivo de los recursos marinos en la costa norte de Santa Cruz (Patagonia Argentina). XVII Congreso Nacional de Arqueología Argentina Arqueología Argentina en el Bicentenario de la Revolución de Mayo; 2010.
2. Flower WH. Whales, and British and colonial whale fisheries1895.
3. Wallach AD, Izhaki I, Toms JD, Ripple WJ, Shanas U. What is an apex predator? Oikos. 2015;124(11):1453-61. doi: DOI: 10.1111/oik.01977.
4. Roman J, Estes JA, Morissette L, Smith C, Costa D, McCarthy J, et al. Whales as marine ecosystem engineers. Frontiers in Ecology and the Environment. 2014;12(7):377-85. doi: DOI: 10.1890/130220.
5. Jones CG, Gutiérrez JL. On the purpose, meaning, and usage of the physical ecosystem engineering concept. First ed2007. 3-20 p.

6. Trites AW, Livingston PA, Mackinson S, Vasconcellos M, Springer AM, Pauly D. Ecosystem change and the decline of marine mammals in the Eastern Bering Sea: testing the ecosystem shift and commercial whaling hypotheses. Fisheries Centre Research Reports 1999 ISSN 1198-6727.

7. Baco AR, Smith CR. High species richness in deep-sea chemoautotrophic whale skeleton communities. Marine Ecology Progress Series. 2003;260:109-14.

8. Butman CA, Carlton JT, Palumbi SR. Whaling effects on deep-sea biodiversity. JSTOR; 1995. p. 462-4.

9. Roman J, McCarthy JJ. The whale pump: marine mammals enhance primary productivity in a coastal basin. PLoS ONE 16(8): e0255667. 2010;5(10):e13255. doi: https://doi.org/10.1371/journal.pone.0013255.

10. Cook D, Malinauskaite L, Davíðsdóttir B, Ögmundardóttir H, Roman J. Reflections on the ecosystem services of whales and valuing their contribution to human well-being. Ocean & Coastal Management. 2020;186:105100. doi: https://doi.org/10.1016/j.ocecoaman.2020.105100.

11. Clapham P, Van Waerebeek K. Bushmeat and bycatch: the sum of the parts. Wiley Online Library; 2007.

12. Clapham PJ. Managing leviathan: conservation challenges for the great whales in a post-whaling world. Oceanography. 2016;29(3):214-25.

13. Noad MJ, Kniest E, Dunlop RA. Boom to bust? Implications for the continued rapid growth of the eastern Australian humpback whale population despite recovery. Population Ecology. 2019;61(2):198-209. doi: https://doi.org/10.1002/1438-390X.1014.

14. Baines M, Kelly N, Reichelt M, Lacey C, Pinder S, Fielding S, et al. Population abundance of recovering humpback whales Megaptera novaeangliae and other baleen whales in the Scotia Arc, South Atlantic. Marine Ecology Progress Series. 2021;676:77-94.

15. Jackson JA, Kennedy A, Moore M, Andriolo A, Bamford CC, Calderan S, et al. Have whales returned to a historical hotspot of industrial whaling? The pattern of southern right whale Eubalaena australis recovery at South Georgia. Endangered Species Research. 2020;43:323-39. doi: https://doi.org/10.3354/esr01072

16. Nicol C, Bejder L, Green L, Johnson C, Keeling L, Noren D, et al. Anthropogenic threats to wild cetacean welfare and a tool to inform policy in this area. Frontiers in veterinary science. 2020;7:57. doi: doi: 10.3389/fvets.2020.00057.

17. Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, et al. Impacts of biodiversity loss on ocean ecosystem services. Science. 2006;314(5800):787-90.

18. Estes JA, Terborgh J, Brashares JS, Power ME, Berger J, Bond WJ, et al. Trophic downgrading of planet Earth. Science. 2011;333(6040):301-6. doi: DOI: 10.1126/science.1205106.

19. Edwards EF, Hall C, Moore TJ, Sheredy C, Redfern JV. Global distribution of fin whales B alaenoptera physalus in the post-whaling era (1980–2012). Mammal Review. 2015;45(4):197-214. doi: https://doi.org/10.1111/mam.12048.

20. Tempera F, Giacomello E, Mitchell NC, Campos AS, Henriques AB, Bashmachnikov I, et al. Mapping Condor seamount seafloor environment and associated biological assemblages (Azores, NE Atlantic). In: Peter Harris EB, editor. Seafloor Geomorphology as Benthic Habitat. Second ed: Elsevier; 2012. p. 807-18.

21. Alves ML, Colin de Verdière A. Instability dynamics of a subtropical jet and applications to the Azores Front Current System: eddy-driven mean flow. Journal of Physical Oceanography. 1999;29(5):837-64.

22. Santos RS, Gonçalves J, Isidro E. Marine Research: The role of the Department of Oceanography and Fisheries of the University of the Azores. Higher Education Policy. 1995;8(2):25-8.

23. Johnson J, Stevens I. A fine resolution model of the eastern North Atlantic between the Azores, the Canary Islands and the Gibraltar Strait. Deep Sea Research Part I: Oceanographic Research Papers. 2000;47(5):875-99.

24. Bashmachnikov I, Machín F, Mendonça A, Martins A. In situ and remote sensing signature of meddies east of the mid-Atlantic ridge. Journal of Geophysical Research: Oceans. 2009;114(C5).

25. Santos M, Pierce G, Boyle P, Reid R, Ross H, Patterson I, et al. Stomach contents of sperm whales Physeter macrocephalus stranded in the North Sea 1990-1996. Marine Ecology Progress Series. 1999;183:281-94. doi: 10.3354/meps183281.

26. Juliano M. Determination and analysis of the threedimensional Atlantic Ocean circulation and climate: An integrated vision: Ph. D. dissertation, University of the Azores; 2003.

27. Pollard RT, Griffiths MJ, Cunningham SA, Read JF, Perez FF, Rios AF. Vivaldi 1991—A study of the formation, circulation and ventilation of eastern North Atlantic Central Water. 1996. p. 167-92.

28. Caldeira R, Reis JC. The Azores confluence zone. Frontiers in Marine Science. 2017;4:37. doi: https://doi.org/10.3389/fmars.2017.00037.

29. Afonso P, Fontes J, Giacomello E, Magalhães MC, Martins HR, Morato T, et al. The Azores: a mid-Atlantic hotspot for marine megafauna research and conservation. Frontiers in Marine Science. 2020;6:826. doi: https://doi.org/10.3389/fmars.2019.00826.

30. Abecasis RC, Afonso P, Colaço A, Longnecker N, Clifton J, Schmidt L, et al. Marine conservation in the Azores: evaluating marine protected area development in a remote island context. Frontiers in Marine Science. 2015;2:104. doi: https://doi.org/10.3389/fmars.2015.00104.

31. Whitehead H. Sperm whales: social evolution in the ocean. first ed: University of Chicago press; 2003.

32. Rice DW. Sperm whale: Physeter macrocephalus Linnaeus, 1758. In: Ridgway SH HR, editor. Handbook of marine mammals. 4: river dolphins and the larger1989.

33. Clarke R. Sperm Whales of the Azores: University Press; 1956.

34. Silva MA, Prieto R, Cascão I, Seabra MI, Machete M, Baumgartner MF, et al. Spatial and temporal distribution of cetaceans in the mid-Atlantic waters around the Azores. Marine Biology Research. 2014;10(2):123-37. doi: https://doi.org/10.1080/17451000.2013.793814.

35. Steiner L, Pérez M, der Van Linde M, Freitas L, Santos R, Martins V, et al., editors. Long distance movements of female/immature sperm whales in the North Atlantic. Poster presented at the Biennial Society for Marine Mammalogy Conference, San Francisco, USA; 2015.

36. Cooke JG. Balaenoptera borealis. The IUCN Red List of Threatened Species 2018: 2018.

37. Silva MA, Prieto R, Jonsen I, Baumgartner MF, Santos RS. North Atlantic blue and fin whales suspend their spring migration to forage in middle latitudes: building up energy reserves for the journey? PloS one. 2013;8(10):e76507. doi: 10.1371/journal.pone.0076507.

38. Prieto R, Silva MA, Waring GT, Gonçalves JM. Sei whale movements and behaviour in the North Atlantic inferred from satellite telemetry. Endangered Species Research. 2014;26(2):103-13. doi: 10.3354/esr00630.

39. Visser F, Hartman KL, Pierce GJ, Valavanis VD, Huisman J. Timing of migratory baleen whales at the Azores in relation to the North Atlantic spring bloom. Marine Ecology Progress Series. 2011;440:267-79. doi: doi: 10.3354/meps09349.

40. Freitas R, Romeiras M, Silva L, Cordeiro R, Madeira P, González JA, et al. Restructuring of the 'Macaronesia'biogeographic unit: A marine multi-taxon biogeographical approach. Scientific reports. 2019;9(1):1-18.

41. Bouveroux T, Waggitt JJ, Belhadjer A, Cazenave PW, Evans PG, Kiszka JJ. Modelling fine-scale distribution and relative abundance of harbour porpoises in the Southern Bight of the North Sea using platform-of-opportunity data. Journal of the Marine Biological Association of the United Kingdom. 2020;100(3):481-9. doi: 10.1017/S0025315420000326.

42. Tobeña M, Prieto R, Machete M, Silva MA. Modeling the potential distribution and richness of cetaceans in the Azores from fisheries observer program data. Frontiers in Marine Science. 2016;3:202. doi: doi: 10.3389/fmars.2016.00202.

43. Prieto R, Tobeña M, Silva MA. Habitat preferences of baleen whales in a mid-latitude habitat. Deep Sea Research Part II: Topical Studies in Oceanography. 2017;141:155-67. doi: https://doi.org/10.1016/j.dsr2.2016.07.015.

44. Elith J, Leathwick JR. Species distribution models: ecological explanation and prediction across space and time. Annual review of ecology, evolution, systematics. 2009;40:677-97. doi: 10.1146/annurev.ecolsys.110308.120159.

45. Zurell D, Franklin J, König C, Bouchet PJ, Dormann CF, Elith J, et al. A standard protocol for reporting species distribution models. Ecography. 2020;43(9):1261-77. doi: doi: 10.1111/ecog.04960.

46. Liu C, Newell G, White M. The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. Ecography. 2019;42(3):535-48. doi: https://doi.org/10.1111/ecog.03188.

47. Smith AB, Santos MJ. Testing the ability of species distribution models to infer variable importance. Ecography. 2020;43(12):1801-13. doi: doi: 10.1111/ecog.05317.

48. Lambert C, Mannocci L, Lehodey P, Ridoux V. Predicting cetacean habitats from their energetic needs and the distribution of their prey in two contrasted tropical regions. PloS one. 2014;9(8):e105958. doi: doi:10.1371/journal.pone.0105958.

49. Guisan A, Zimmermann NE. Predictive habitat distribution models in ecology. Ecological modelling. 2000;135(2-3):147-86. doi: https://doi.org/10.1016/S0304-3800(00)00354-9.

50. Croll DA, Marinovic B, Benson S, Chavez FP, Black N, Ternullo R, et al. From wind to whales: trophic links in a coastal upwelling system. Marine Ecology Progress Series. 2005;289:117-30.

51. Friedlaender AS, Halpin PN, Qian SS, Lawson GL, Wiebe PH, Thiele D, et al. Whale distribution in relation to prey abundance and oceanographic processes in shelf waters

of the Western Antarctic Peninsula. Marine Ecology Progress Series. 2006;317:297-310. doi: 10.3354/meps317297.

52. Lehodey P, Conchon A, Senina I, Domokos R, Calmettes B, Jouanno J, et al. Optimization of a micronekton model with acoustic data. ICES Journal of Marine Science. 2015;72(5):1399-412.

53. Lehodey P, Murtugudde R, Senina I. Bridging the gap from ocean models to population dynamics of large marine predators: a model of mid-trophic functional groups. Progress in Oceanography. 2010;84(1-2):69-84.

54. Pérez-Jorge S, Tobeña M, Prieto R, Vandeperre F, Calmettes B, Lehodey P, et al. Environmental drivers of large-scale movements of baleen whales in the mid-North Atlantic Ocean. Diversity Distributions. 2020;26(6):683-98.

55. Romagosa M, Pérez-Jorge S, Cascão I, Mouriño H, Lehodey P, Pereira A, et al. Food talk: 40-Hz fin whale calls are associated with prey biomass. Proceedings of the Royal Society B. 2021;288(1954):20211156.

56. Romagosa M, Lucas C, Pérez-Jorge S, Tobeña M, Lehodey P, Reis J, et al. Differences in regional oceanography and prey biomass influence the presence of foraging odontocetes at two Atlantic seamounts. Marine Mammal Science. 2020;36(1):158-79. doi: 10.1111/mms.12626.

57. Roberts JJ, Best BD, Mannocci L, Fujioka E, Halpin PN, Palka DL, et al. Habitat-based cetacean density models for the US Atlantic and Gulf of Mexico. Scientific reports. 2016;6(1):1-12.

58. Mannocci L, Roberts JJ, Pedersen EJ, Halpin PN. Geographical differences in habitat relationships of cetaceans across an ocean basin. Ecography. 2020;43(8):1250-9.

59. Silva MA, Machete M, Reis D, Santos M, Prieto R, Dâmaso C, et al. A review of interactions between cetaceans and fisheries in the Azores. Aquatic Conservation: Marine Freshwater Ecosystems. 2011;21(1):17-27.

60. DRPM. Rede de Arrojamento de Cetáceos dos Açores: Direção Regional de Políticas Marítimas; 2021 [cited 2021]. Available from: https://servicos-sraa.azores.gov.pt/doit/servicos.asp?id_dep=10&id_form=84.

61. UNCTAD. Review of Maritime Transport. 2021 eISSN: 2225-3459.

62. Açores EdAd. Transporte marítimo. Direção Regional do Ambiente e Alterações Climáticas: Governo dos Açores, 2017.

63. Conn PB, Silber GK. Vessel speed restrictions reduce risk of collision-related mortality for North Atlantic right whales. Ecosphere. 2013;4(4):16. doi: 10.1890/es13-00004.1.

64. MacGillivray AO, Li Z, Hannay DE, Trounce KB, Robinson OM. Slowing deep-sea commercial vessels reduces underwater radiated noise. Acoust Soc Am. 2019;146(1):340. Epub 2019/08/03. doi: 10.1121/1.5116140. PubMed PMID: 31370655.

65. Croll DA, Clark CW, Calambokidis J, Ellison WT, Tershy BR, editors. Effect of anthropogenic low-frequency noise on the foraging ecology of Balaenoptera whales. Animal Conservation forum; 2001: Cambridge University Press.

66. Fagerholt K, Gausel NT, Rakke JG, Psaraftis HN. Maritime routing and speed optimization with emission control areas. Transportation Research 2015;52:57-73. doi: 10.1016/j.trc.2014.12.010.

67. Kotta J, Nurkse K, Puntila R, Ojaveer H. Shipping and natural environmental conditions determine the distribution of the invasive non-indigenous round goby

Neogobius melanostomus in a regional sea. Estuarine, Coastal and Shelf Science. 2016;169:15-24. doi: 10.1016/j.ecss.2015.11.029.

68. Nowacek DP, Johnson MP, Tyack PL. North Atlantic right whales (Eubalaena glacialis) ignore ships but respond to alerting stimuli. Proceedings Biological sciences. 2004;271(1536):227-31. Epub 2004/04/03. doi: 10.1098/rspb.2003.2570. PubMed PMID: 15058431; PubMed Central PMCID: PMC1691586.

69. Harris K, Gende SM, Logsdon MG, Klinger T. Spatial pattern analysis of cruise ship–humpback whale interactions in and near Glacier Bay National Park, Alaska. Environmental management. 2012;49(1):44-54. doi: 10.1007/s00267-011-9754-9.

70. Goldbogen JA, Calambokidis J, Shadwick RE, Oleson EM, McDonald MA, Hildebrand JA. Kinematics of foraging dives and lunge-feeding in fin whales. Journal of Experimental Biology. 2006;209(7):1231-44. doi: 10.1242/jeb.02135.

71. Acevedo-Gutiérrez A, Croll D, Tershy B. High feeding costs limit dive time in the largest whales. Journal of Experimental Biology. 2002;205(12):1747-53. doi: https://doi.org/10.1242/jeb.205.12.1747.

72. Kraus SD, Brown MW, Caswell H, Clark CW, Fujiwara M, Hamilton PK, et al. North Atlantic right whales in crisis. Science. 2005;309(5734):561-2.

73. IWC. Annual Report of the International Whaling Commission 2005. 2006.

74. Frantzis A, Leaper R, Alexiadou P, Prospathopoulos A, Lekkas D. Shipping routes through core habitat of endangered sperm whales along the Hellenic Trench, Greece: Can we reduce collision risks? PloS one. 2019;14(2):e0212016. Epub 2019/02/28. doi: 10.1371/journal.pone.0212016. PubMed PMID: 30811429; PubMed Central PMCID: PMC6392247 not alter our adherence to PLOS ONE policies on sharing data and materials.

75. Williams R, OHara P. Modelling ship strike risk to fin, humpback and killer whales in British Columbia, Canada. Journal of Cetacean Research and Management. 2009;11(1):1-8.

76. Redfern J, McKenna M, Moore T, Calambokidis J, Deangelis M, Becker E, et al. Assessing the risk of ships striking large whales in marine spatial planning. Conservation Biology. 2013;27(2):292-302. doi: 10.1111/cobi.12029

77. Pirotta E, New L, Marcoux M. Modelling beluga habitat use and baseline exposure to shipping traffic to design effective protection against prospective industrialization in the Canadian Arctic. Aquatic Conservation: Marine Freshwater Ecosystems. 2018;28(3):713-22. doi: https://doi.org/10.1002/aqc.2892.

78. Christiansen F, Bertulli CG, Rasmussen MH, Lusseau D. Estimating cumulative exposure of wildlife to non-lethal disturbance using spatially explicit capture–recapture models. The Journal of Wildlife Management. 2015;79(2):311-24. doi: http://dx.doi.org/10.1002/jwmg.836.

79. Christiansen F, Lusseau D. Linking behavior to vital rates to measure the effects of non-lethal disturbance on wildlife. Conservation Letters. 2015;8(6):424-31. doi: 10.1111/conl.12166.

80. Palazzo Jr JT. Whose Whales? Developing countries and the right to use whales by non-lethal means. 1999.

81. Creel S, Christianson D, Liley S, Winnie JA. Predation risk affects reproductive physiology and demography of elk. Science. 2007;315(5814):960-.

82. Preisser EL, Bolnick DI, Benard MF. Scared to death? The effects of intimidation and consumption in predator–prey interactions. Ecology. 2005;86(2):501-9.

83. Parsons E, Brown DM. Recent advances in whale-watching research: 2015–2016. Tourism in Marine Environments. 2017;12(2):125-37.

84. Senigaglia V, Christiansen F, Bejder L, Gendron D, Lundquist D, Noren D, et al. Meta-analyses of whale-watching impact studies: comparisons of cetacean responses to disturbance. Marine Ecology Progress Series. 2016;542:251-63. doi: 10.3354/meps11497.

85. Castellote M, Clark CW, Lammers MO. Acoustic and behavioural changes by fin whales (Balaenoptera physalus) in response to shipping and airgun noise. Biological Conservation. 2012;147(1):115-22. doi: 10.1121/2.0000311.

86. Boys RM, Oliveira C, Pérez-Jorge S, Prieto R, Steiner L, Silva MA. Multi-state open robust design applied to opportunistic data reveals dynamics of wide-ranging taxa: the sperm whale case. Ecosphere. 2019;10(3):e02610.

CHAPTER 2

2. CETACEAN WATCHING

2.1. INTRODUCTION

Commercial whale-watching (or more appropriately, cetacean watching) has become a fast-growing, multi-million euros industry throughout the world. At a global level, in 2009 the activity was estimated to attract 13 million tourists and be worth at least US $2.1 billion annually [1-3], but these figures are outdated and are likely considerably higher at present. In fact, a more recent estimate indicate that the number of tourists going cetacean watching in Europe incremented at least by 30% from 2008 to 2018 [4].

For small and peripheral regions, marine ecotourism can have a big the socioeconomic impact, and constitute one of the pillars of a sustainable blue economy strategy [5]. Case in point, Ressurreiçao *et al* (2022)(5) estimated that, in 2014, marine ecotourism generated nearly 80 million euros for the economy of the Azores archipelago (Portugal), equivalent to 2.2% of the Gross Domestic Product of the region. Cetacean watching direct and indirect revenues accounted for nearly 71% of that value [5].

The rapid worldwide expansion and growth of this activity has motivated dedicated research as well as implementation of legal regulations and/or best-practice guidelines in several places [6-8]. Most of the research related to this activity has focused on economic aspects, and on the short- and (less frequently) long-term effects of boat presence and noise on the targeted animals and populations [2, 9]. Due to great uncertainties about the effects of the activity on animal populations, and conflicting concerns in animal conservation and local economic interests, cetacean watching regulations and best-practice guidelines often combine and have to balance science-based principles, a precautionary approach, and socio-economic growth goals [10-12]. For those reasons, and while many share the same basic points, regulations and guidelines can vary widely at national and regional levels [2, 13-15].

The concept of carrying capacity has been hailed as a way of guaranteeing the activity’s sustainability over time, by estimating the amount of interference from the activity which the animals can sustain without incurring in population-level adverse effects [16-19]. As such, many regulations include measures to limit the number of boats (using licensing schemes), activity-specific regulations (e.g. swimming, professional audio-visual recordings, scientific work), maximum observation time, seasonal closures and spatial zoning [2, 7, 20, 21]. However, these limitations are mostly based on a precautionary approach rather than on science-based evaluation [22]. In fact, assessing the carrying capacity is a complex endeavour that requires knowledge on cetaceans’ distribution and abundance, behavioural, energetic, and physiological effects of stressors, as well as the effective distribution and effort of the cetacean watching activity.

Cetacean abundance and distribution can be estimated with models fitted to sighting survey or mark-recapture data [23, 24]. There are also efforts to understand the long-term effects of cetacean watching, giving rise to a growing body of literature on the matter [9, 25]. Strikingly, much less effort has been put into developing methodologies to characterize and quantify cetacean watching distribution and density over spatial and temporal scales [26, 27]. This information is essential to guide management efforts towards the long-term sustainability of the cetacean watching activity.

Calculating the effort of an activity, using some appropriate metric, is crucial to estimate its effects on a given resource. Assessing cetacean watching effort helps to identify the areas where exposure of cetaceans is higher, and scale the potential effects of the activity to the animals [26, 27]. Cetacean watching effort can be calculated by reconstructing boat tracks based on geographic positions recorded at regular intervals [26, 27]. However, disentangling actual cetacean watching periods from transiting and other boat activities is a complex task that has not been fully resolved. Furthermore, those data are not always available, and currently there is no standard to record and process cetacean watching vessel activity.

Currently, vessels can be outfitted with a variety of positioning equipment, allowing for remote tracking and storage of information on their cruising behaviour, along with relevant metadata. The Automatic Identification System (AIS), a standard radio communication and maritime safety system, is now a widespread ship tracking system. As of 2004, merchant ships over 300 gross tonnage and all passenger vessels are required to be equipped with AIS, but the system is also widely used by other ship types [28]. The system has increasingly been used in smaller ships either as a result of evolving regulations or voluntarily, as a safety measure, and the monitoring network is expanding around the world [28, 29]. AIS equipment broadcasts encoded messages in real-time that include information on the ship's geographic position derived from a global navigation satellite system (GPS or other), navigation status, identity, and other metadata such as dimensions, cargo, origin and destination, and security related messages. AIS messages are transmitted via a radio VHF-transceiver, which can be then received by other ships or land stations, and more recently by satellite, and can be stored efficiently on servers in near real-time [30].

There are two AIS classes (termed class A and class B), with slightly different transceivers. Class A transponders are the standard in the shipping industry and transmit a large set of data about the ship identity and configuration, status, cargo, and safety-related information. Class B is a simplified version that transmits less information and at a lower transmission frequency than class A. Other differences among the two classes are that class A has priority over class B when there is a bandwidth overload, and that class A transceivers transmit at a higher power output (usually 12.5 Watt) compared to class B (usually 2.0 Watt). Despite its reduced capabilities, those differences make class B more appealing for smaller ships due to smaller size, lower power consumption, and reduced cost [30].

Recently, the Regional Government of the Azores started a program to instrument several cetacean watching boats with AIS class B units, on a voluntary basis, in order to enable the characterization of the activity. The cetacean watching operators usually prepare their boats with GPS equipment to monitor their surveys, some of the operators of the Azores have shared this data.

The time-series of high-precision geolocation data based on GPS and relayed by systems such as AIS, enable accurate reconstruction of ship tracks, calculation of vessel speeds, and can be used to infer boat activities and behaviours. Several methods have been used to infer fishing ship behaviour (e.g., transiting, searching, fishing and handling time) from AIS or Vessel Monitoring Systems (VMS) data [31-35]. Fish behaviour can be classified into multiple behavioural states or commonly simplified into two behaviours, when the speed distribution follows a clear bimodal pattern [36], such as transiting (higher speeds) or fishing (lower speeds) [31]. Machine learning, data mining and filtering approaches have been applied to discriminate among multiple fishing ship behaviours [36]. Among these, the most widely used procedure includes an algorithm to define ship states and a model to estimate the component parameters to define the speed interval [32]. Some examples are a density-based clustering algorithm using a Gaussian curve, an expectation maximization (EM), or the Gaussian Mixture Model (GMM) and Bayesian Hidden Markov Models (BHMM) [31-35].

The most appropriate approach to infer boat behaviour ultimately depends on the type of activity and the way boats manoeuvre during different behaviours [36], as well as on the available data. Some algorithms require extra information on onboard operations not provided from the AIS or VMS systems, and that require observers or ship logbooks [31, 33, 36, 37]. Other limitations are the processing power necessary to handle large datasets, the application of time-consuming algorithms or models, and how to validate or test the methods [30]. Vespe et al. [33] described a method for mapping fishing activities using the bimodal distribution of speed when they are travelling and during fishing events. The method is elegant for its simplicity and presents a good performance with a relatively small amount of ship information. While the method has good potential for characterising the behaviour of other shipping activities, it has never been used to characterize the behaviour of cetacean watching boats.

This chapter presents a methodology for classifying cetacean watching boat activities from archived positional data from GPS loggers or via AIS, inspired on Vespe et al. [33] with changes to conform with cetacean watching specificities. The algorithm presented by Vespe et al. [33] was modified to split the cetacean watching trips into different behaviours and identifying periods when boats were transiting or watching cetaceans. The method was validated by applying it to an independent dataset, for which boat behaviour was known. Finally, method was applied to the cetacean watching dataset collected in the Azores archipelago to assess the distribution of cetacean watching in the area.

2.2. METHODS

Data analysis was structured in five sections: data pre-processing, development of the algorithm for trip identification, data exploration, development of the algorithm to identify cetacean observation periods, and the creation of effort maps.

Data pre-processing aimed at unifying in a single database all cetacean watching tracking data from different sources (GPS and AIS). Subsequently, an algorithm was developed to split the boat data into unique trips. Data exploration included two steps: 1) standardizing the time interval between boat positions; and 2) characterizing boat activity by home port. The methodology to identify periods when cetacean watching boats were engaged in observation of cetaceans (hereafter called cetacean watching boat behaviour) also included two steps: 1) developing a method to identify the behaviour of cetacean watching boats during cetacean observation that could be adapted to the large databases available; 2) validating this method using the boat's logbook and a sequence of maps as an example of all phases of the validation process. Finally, the methodology was applied to create effort maps using only trips when boats were engaged in cetacean observation.

2.2.1. DATA PRE-PROCESSING

Cetacean watching geolocation data were collected for boats from Faial and Pico Islands, in the Central Group of the Azores archipelago (Figure 2.1). These islands are inside a cetacean watching spatial management area (Zone A around the triangle of Faial, Pico and S. Jorge islands) defined by specific cetacean watching regulations issued by the Azores Autonomous Regional Government (DLR 10/2003/A, 22 Mar.; Portaria n.º 5/2004, 29 Jan.). The study area includes coastal waters with complex bathymetry that can present depths in excess of 1,000 meters, between and around these islands.

Geolocation data from cetacean watching boats were collected in one of two ways. Several cetacean watching operators provided data from GPS loggers placed on board their boats from 2013 to 2018 (Table 2.1). Additionally, in 2018 several cetacean watching boats were equipped with AIS class B transceivers (Table 2.1). A dedicated antenna and a server for data storage were installed in Faial to record AIS messages relayed by boats in the study area. Hereafter, data originating from GPS loggers will be referred as 'GPS' data and from AIS transceivers as 'AIS' data.

A total of 675,405 unique records with geolocation and metadata were collected from 14 (of a total of 24) cetacean watching boats licensed to operate in zone A (Table 2.1). Boat identity was given either by the boat operator or by the unique boat identifier (Maritime Mobile Service Identity – MMSI) present in all AIS messages. Boat identities were only used to derive the physical characteristics of the boats. All operators gave their consent for data storage and processing. Data included two types of variables: 1) static, related to the characteristics of the boat (MMSI, name, size, type, flag), and 2) dynamic (date, time, latitude, longitude, and speed).

Messages from the AIS transceivers were decoded using the software AISDecoder (https://arundaleais.github.io), and stored as daily CSV tables. To decode data from GPS loggers, either Garmin Basecamp® or DNRGPS (https://www.dnr.state.mn.us) software were used to decode and convert raw data into CSV tables with a similar structure to that of the AIS CSV tables. Custom R scripts (R Core Team 2020) were written to clean, normalize and merge the data. For AIS data, a filter for MMSI was applied to retain only records from cetacean watching boats. Static information from boats was added to the GPS tables using a custom script, following the same format used in the AIS data tables.

The home port, corresponding to the operation port (which may or may not coincide with the registration port), was also added for all boats (AIS and GPS) to aid in some of the subsequent analyses.

AIS transponders send different messages with redundant information for safety and technical reasons, producing a large amount of duplicate information. Another issue with the system is that some of the static information must be written manually, and as a consequence is prone to errors. Therefore, CSV tables originating from both data sources were cleaned by removing incomplete and redundant records and uninformative characters and words. Superfluous fields containing information not used for the analysis were also removed to reduce the database size. Variables were renamed and consolidated (for example, columns containing dates and times in different formats were transformed into a single format). Finally, a point-in-polygon approach was used to remove all data points falling on land. Once GPS and AIS data were cleaned and harmonized, tables were merged into a single database for subsequent processing. The workflow and R packages used to build the database are represented in Figure 2.2.

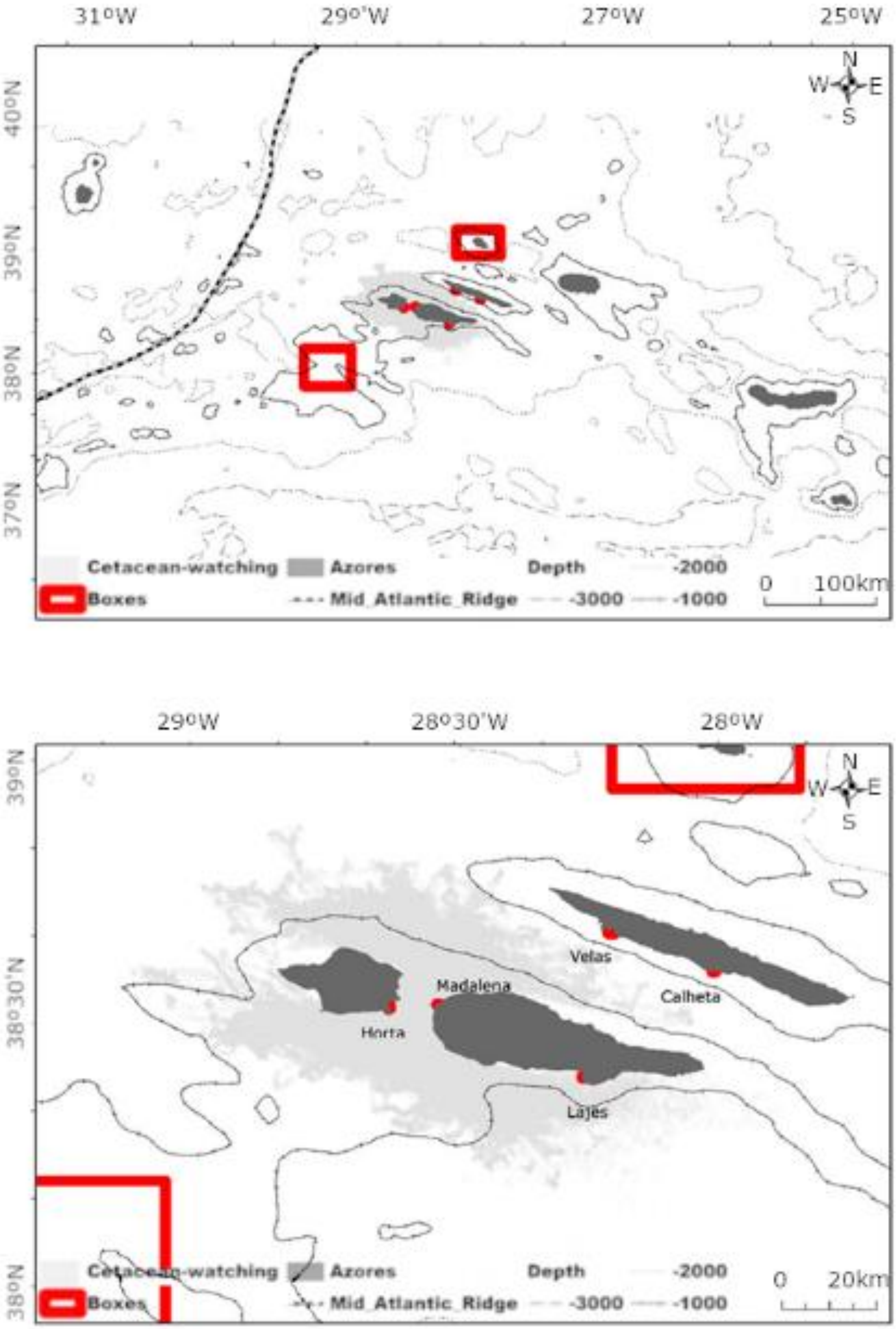


Figure 2. 1. Study area. Top: Azores archipelago. Red boxes: exclusion boxes for the ports and other marine activities used to identify and classify different trips (refer to text for further details). Bottom: Detail of the study area. In light grey: Cetacean watching recorded positions.

Table 2. 1. Dataset used to calculate cetacean watching effort. Boat: each boat was identify by a unique number (except boat 3 for which the same GPS was shared by two different boats); Boat length (m); Type: RIB (rigid inflatable boat) or cabin; Year: year data recorded; Month_i: initial month of monitoring; Month_f: final month of monitoring; Home port: port from which boat operates; System: AIS or GPS data; Process: data used to calibrate the method or to validate it; CW. Dataset: Azorean cetacean watching fleet (yes) or survey (no) datasets; Count: total number of locations.

Boat	Boat length (m)	Type	Year	Month_i	month_f	Home port	System	Process	CW. dataset	Count
1	9	RIB	2018	3	10	Horta,Faial	AIS	calibration	yes	53987
2	7,5	RIB	2018	3	10	Madalena,Pico	GPS	calibration	yes	48725
3	7,5-10	RIB	2016	9	10	Madalena,Pico	GPS	calibration	yes	9319
3	7,5-10	RIB	2017	3	9	Madalena,Pico	GPS	calibration	yes	224938
4	8	RIB	2013	4	9	Lajes,Pico	GPS	calibration	yes	17479
5	9	RIB	2018	3	10	Horta,Faial	AIS	calibration	yes	38971
6	11	cabin	2018	5	10	Horta,Faial	AIS	calibration	yes	27863
7	10	RIB	2018	3	10	Madalena,Pico	AIS	calibration	yes	12293
8	8,5	RIB	2017	6	10	Lajes,Pico	GPS	calibration	yes	35202
9	8	RIB	2018	6	10	Madalena,Pico	AIS	calibration	yes	24420
10	13	cabin	2017	6	9	Lajes,Pico	GPS	calibration	yes	4396
11	12	RIB	2017	6	9	Lajes,Pico	GPS	calibration	yes	8484
12	13	cabin	2018	3	10	Horta,Faial	AIS	calibration	yes	32649
13	12	RIB	2018	5	9	Madalena,Pico	AIS	calibration	yes	31638
14	9	RIB	2018	3	10	Horta,Faial	AIS	calibration	yes	56444
14	9	RIB	2017	4	6	Horta,Faial	GPS	validation	no	40427
15	8,6	cabin	2015	4	8	Horta,Faial	GPS	calibration	no	3000
15	8,6	cabin	2014	4	8	Horta,Faial	GPS	validation	no	5170

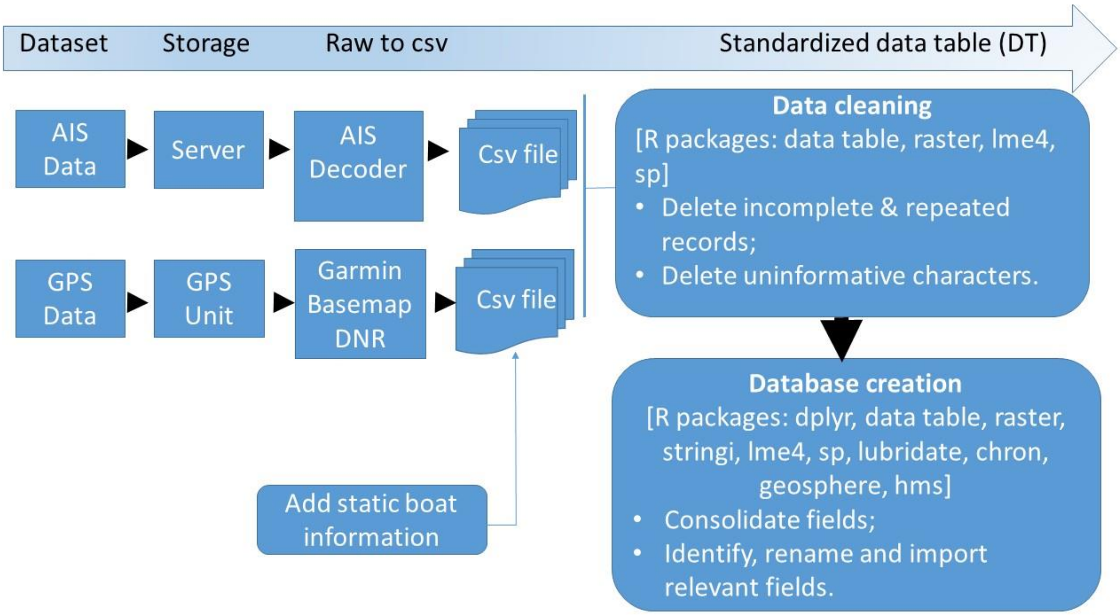


Figure 2. 2. Data pre-processing and standardization. Prior to analysis, data from the AIS and GPS were cleaned and standardized using custom R scripts. After pre-processing, data were merged into a single database. R packages used to run the scripts are shown. Information on software used to process raw GPS and AIS data is given in the main text.

2.2.2. TRIP IDENTIFICATION ALGORITHM

The data from each boat included a series of date and time-stamped geolocations with associated static and dynamic variables. By connecting geolocations in chronological order it was possible to create pseudo-tracks that closely reproduce the actual boat trajectory [34]. However, a single boat often made multiple trips on a single day. Additionally, some boats may be utilized interchangeably for cetacean watching, other touristic activities (such as scuba diving or bird watching), and passenger ferrying within the study area. Thus, prior to analysis, boat tracks had to be split into single trips, and individual trips were classified into cetacean watching or non-cetacean watching.

In studies evaluating fishing activity, the separation of tracks into inactivity/transiting/fishing segments is often based on long periods of inactivity, periods with no transmissions, or with speeds thresholds <5 kn to infer stops at ports. These periods are then used to split boat tracks into different trips [33-35]. However, that approach was not suitable in the present case, since cetacean watching boats can make multiple trips with very short intervals between them. Instead, the *Trip Identification Algorithm* developed here used a spatial approach to identify the start and end of a trip. Polygons were defined around each harbour with cetacean watching boats within the study area (cetacean watching ports), and a point-in-polygon approach was used to identify any points falling inside the ports. The algorithm used that information to split cetacean watching data into separate trips.

Subsequently, the *Trip Identification Algorithm* used a series of criteria to identify cetacean watching trips. First, polygons were defined around the main diving spots, bird-watching areas, and principal harbours (aside from those already defined as cetacean watching ports). Second, the start and finish times of each trip were used to calculate trip duration and to identify the diel period (night time was defined as the period from 22:00 and 07:00 to allow for variation in length of daylight across the cetacean watching season and for trips ending later than sunset time). Finally, cetacean watching trips were defined as those meeting all of the following criteria: 1) occurred during daytime, 2) occurred within waters delimited by the cetacean watching management zone A (around Faial, Pico and S. Jorge islands (Figure 2.1), 3) did not stay extended periods of time in diving areas or ports, 4) had a duration of one to five hours, and 5) were longer than 20 km in length [2, 39, 41-43].

2.2.3. DATA EXPLORATION

2.2.3.1. TIME INTERVAL STANDARDIZATION

The time interval between consecutive boat locations varies depending on the monitoring system, the type of equipment, reception, transmission and recording [44, 45]. Also, messages or signals transmitted via satellite or terrestrial antennas sometimes are lost, which can be due to 1) the low quality of the transmitting or reception equipment; 2) the low quality of messages/propagation signal due to atmospheric/ionosphere conditions, physical obstacles, or the low number of satellites available; or to 3) disconnection of GPS devices or AIS transponders [34, 46, 47]. Furthermore, in AIS the propagation is related to the distance to the nearest AIS station, and its transmission quality can be modified by the saturation of the system in high-density areas [48]. In AIS data, the time interval between position records can also be influenced by the vessel speed [49, 50]. Therefore, the number of cetacean watching trips calculated in this work will likely underestimate the true number of trips.

The final database included (non-duplicated) AIS and GPS positions recorded every 10 min between 2013 and 2018. This database was subsequently divided into two datasets: a subset of the data from 15 boats (14 cetacean watching and one research cetacean boats) were used to develop and calibrate the method (train data), and a second subset containing the records from two boats in two different years was used to validate the method (test data) (Table 2.1).

2.2.3.2. BOAT ACTIVITY BY HOME PORT

Differences in cetacean watching trips (length, duration and spatial distribution) by home port (Figure 2.3), were tested using only the training data. The distribution of trip length and duration was analysed using the Anderson-Darling Normality test and the Shapiro test. A Kruskal-Wallis rank sum test was performed to investigate whether trip length and duration varied significantly among home ports. When the Kruskal-Wallis was significant, a post-hoc analysis was performed using the correction method from Benjamini and Hochberg 1995 [51] to identify which ports differed from the others. Finally, Kernel density maps were created including 90% of the total positions to analyse the spatial range of the activity based on the homeport.

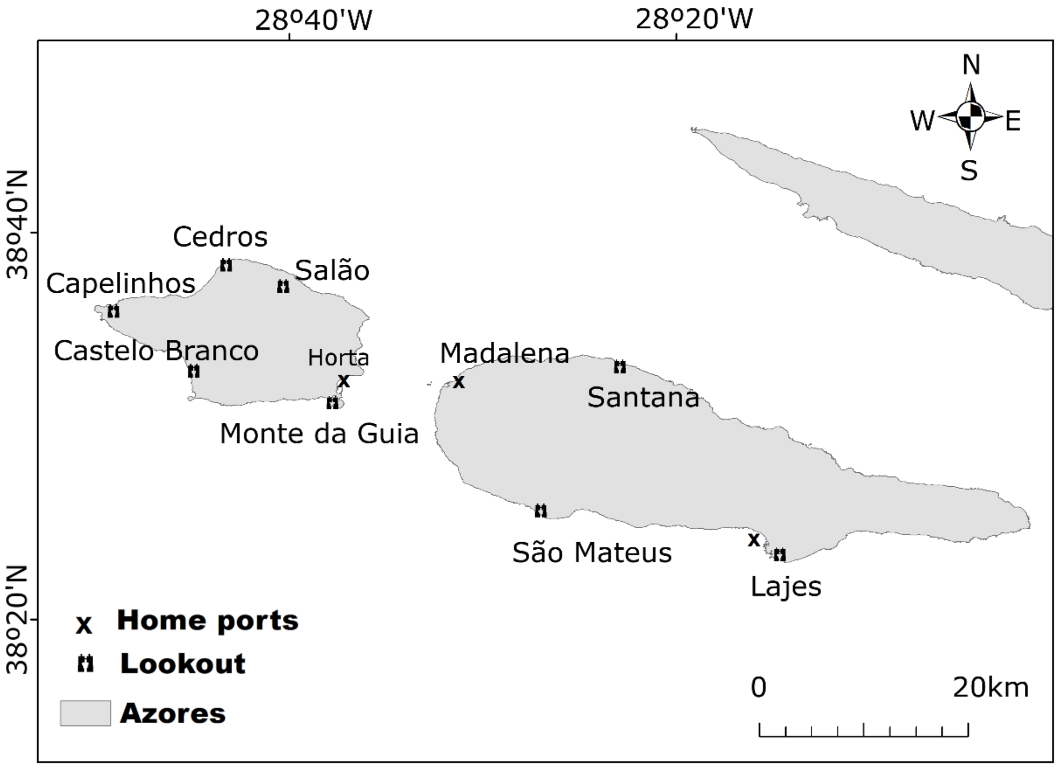


Figure 2. 3. Homeports of cetacean watching in the study area and the positions of the lookout hotspots used to detect cetaceans for which data is available

2.2.4. IDENTIFYING CETACEAN WATCHING BEHAVIOUR

2.2.4.1. METHOD DEVELOPMENT

The first step was to define the behaviour of the boats when engaged in cetacean observation or when approaching an observation area. Boats are expected to move slowly than when transiting, not only because of cetacean travel speed, but also because the current legislation sets a maximum speed in the vicinity of the animals [16, 39, 52]. Hence, I performed a test and plotted a density graph to test this hypothesis using the train data. The distribution of the speed variable was studied based on excess mass.

Graphs for feature significance were created to estimate the kernel density of the speed of each boat and to attempt to identify sections within trips where speed differed. Then, the Kolmogorov-Smirnov test for two samples, the Hartigan’s test, and the bimodality coefficient were used to further investigate the distribution of speed values among different sections and detect significant differences. Finally, a Gaussian Mixture Model (GMM), with the unsupervised classification method, was used to classify trip segments in two distinct activity classes, based on the confidence interval of speeds during cetacean watching and transiting periods.

2.2.4.2. VALIDATION METHOD

The method validation was carried out using the test dataset with 39,059 locations from 65 trips conducted by two boats (boat 14 in 2017, and boat 15 in 2014; Table 2.1). One of these boats was dedicated to cetacean watching activity and the other to cetacean research. These boats were selected to validate the method because they were the only ones with a logbook containing the required information for every cetacean sighting: trip identification, the initial time and location of a sighting, the end time and location of that sighting.

The trips were split into 5-minute segments and each segment was classified according to two activity classes: transiting or cetacean watching. Predictions of a GMM model (described above) fitted to the same data were then used to build a confusion matrix (Table 2.2), to assess performance of the model according to the true activities recorded for the boats in the logbooks.

Table 2. 2. *Confusion matrix for the validation of the method identifying cetacean watching behaviour. GMM: Gaussian Mixture Model. Logbook: boat register of the cetacean sighting.*

Results		**Logbook**	
		Absence	Presence
GMM confidence interval (Prediction)	Transiting -	**True negative** (negative prediction CO absence)	**False negative** (negative prediction CO occurrence)
	Cetacean Observation (CO) +	**False positive** (positive prediction CO absence)	**True positive** (positive prediction CO occurrence)

2.2.4.3. CETACEAN WATCHING EFFORT MAPS

The final step of the process was to build effort maps for cetacean watching activity. The effort maps were built from the train dataset using only data from 2013 and 2018 because these years had a larger sample size the season. Kernel density Silverman (1986) of cetacean watching effort was calculated within a 10 km^2 grid, from the activity class attributed to each of the segments produced by the GMM model. Yearly (2013 and 2018), and monthly (May to September) cetacean watching density maps were produced using the kernel density tool in ARCGIS 10.3.1, in order to identify temporal variations in the activity distribution.

2.3. RESULTS

2.3.1. DATA EXPLORATION

2.3.1.1. TIME INTERVAL STANDARDIZATION

After processing the data, a total of 1,335 cetacean watching trips were identified in the training dataset (Figure 2.4).

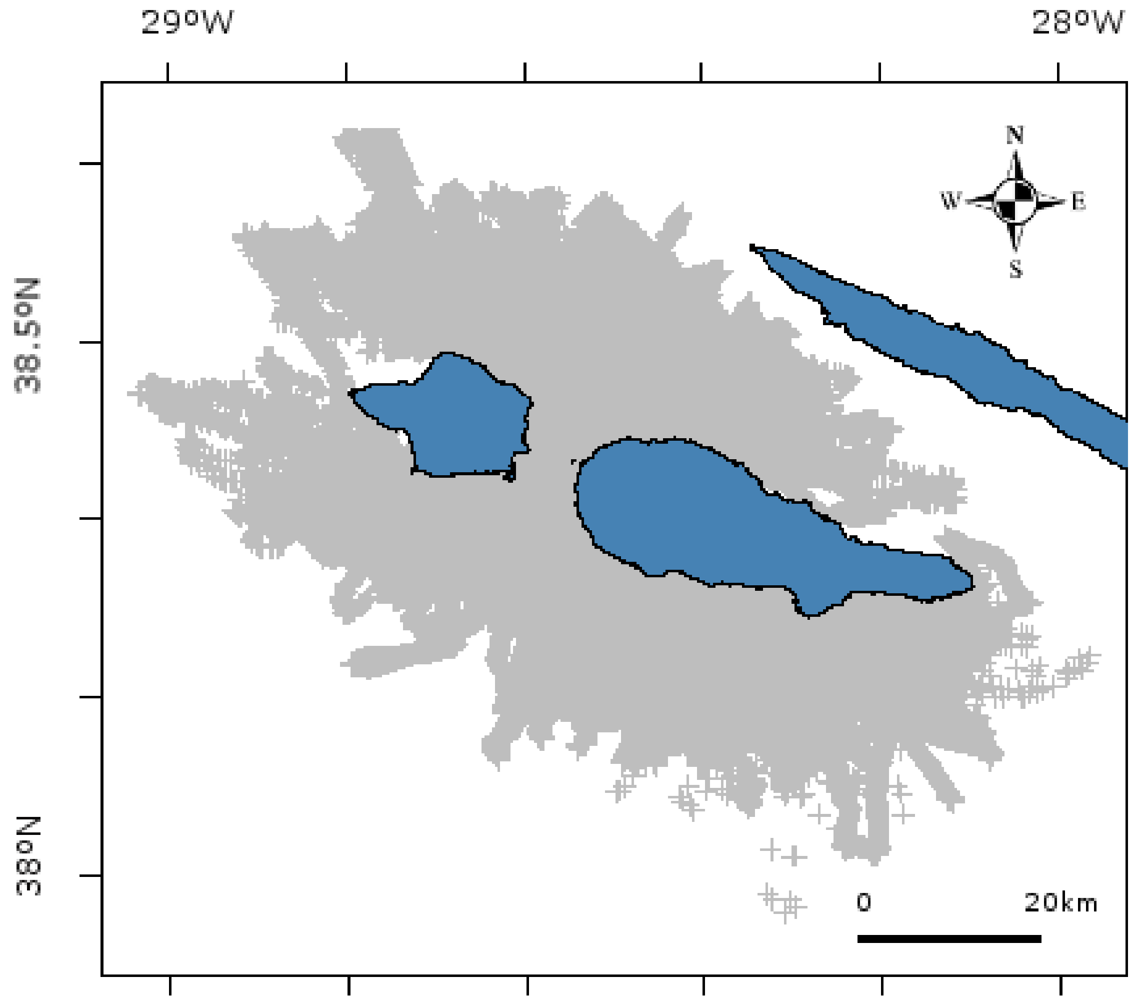


Figure 2. 4. Geographic positions recorded for 1,335 cetacean watching trips used to fit the GMM predictive model.

Once the cetacean watching trips were selected, the time intervals between consecutive geographic positions for each cetacean watching boat were calculated (Figure 2.5). The average intervals ranged from 10 s to 10 min (Table App. 2.1). Most boats with AIS devices had intervals of approximately 40 s. All the GPS equipment had time steps of about 10 s except for one boat, where the interval was set to 10 m for technical reasons.

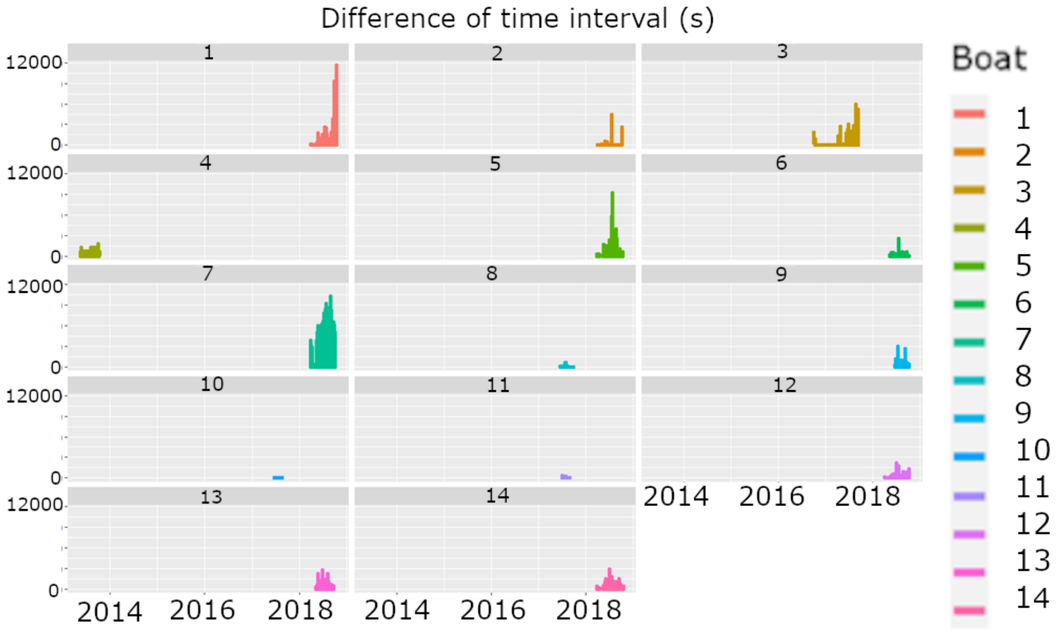


Figure 2. 5. Daily variation in the time interval (s) between consecutive geographic positions per boat. The x-axis shows the julian day, from the first day of the cetacean watching season.

2.3.1.2. BOAT ACTIVITY BY HOME PORT

Trip length and duration varied significantly between home ports (Kruskal-Wallis test, trip length: H=221,5, p<0,05; trip duration: H=60,9, p<0,05). The post-hoc analysis showed that boat trips from Horta were significantly longer in duration and distance travelled, followed by those from Madalena and lastly those from Lajes (Figure 2.6). The kernel maps showed that the boats from Horta used an area of 1,5x10^3 km^2, those from Madalena 1,4x10^3 km^2, and those from Lajes 6,2x10^2 km^2 (Figure 2.7). While the boats from Lajes used a much smaller area south of Pico island, they shared almost half of that area with the boats from the other ports (Figure 2.7).

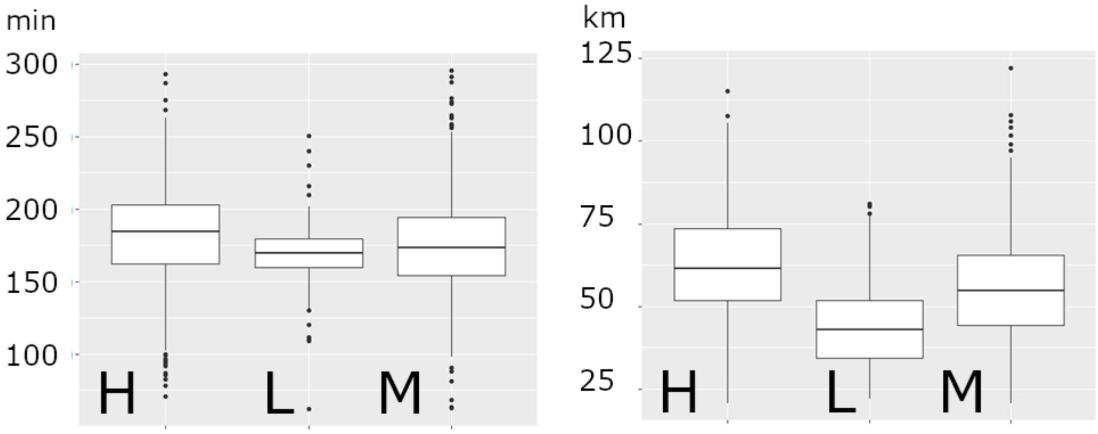


Figure 2. 6. Duration (left) and length (right) of the cetacean watching trips (min and km respectively) based on the home port (H: Horta, L: Lajes, and M: Madalena). The boxplot shows the median (horizontal bar), the interquartile range (IQR) (box), 1.5 x IQR (whiskers) and outliers (dots).

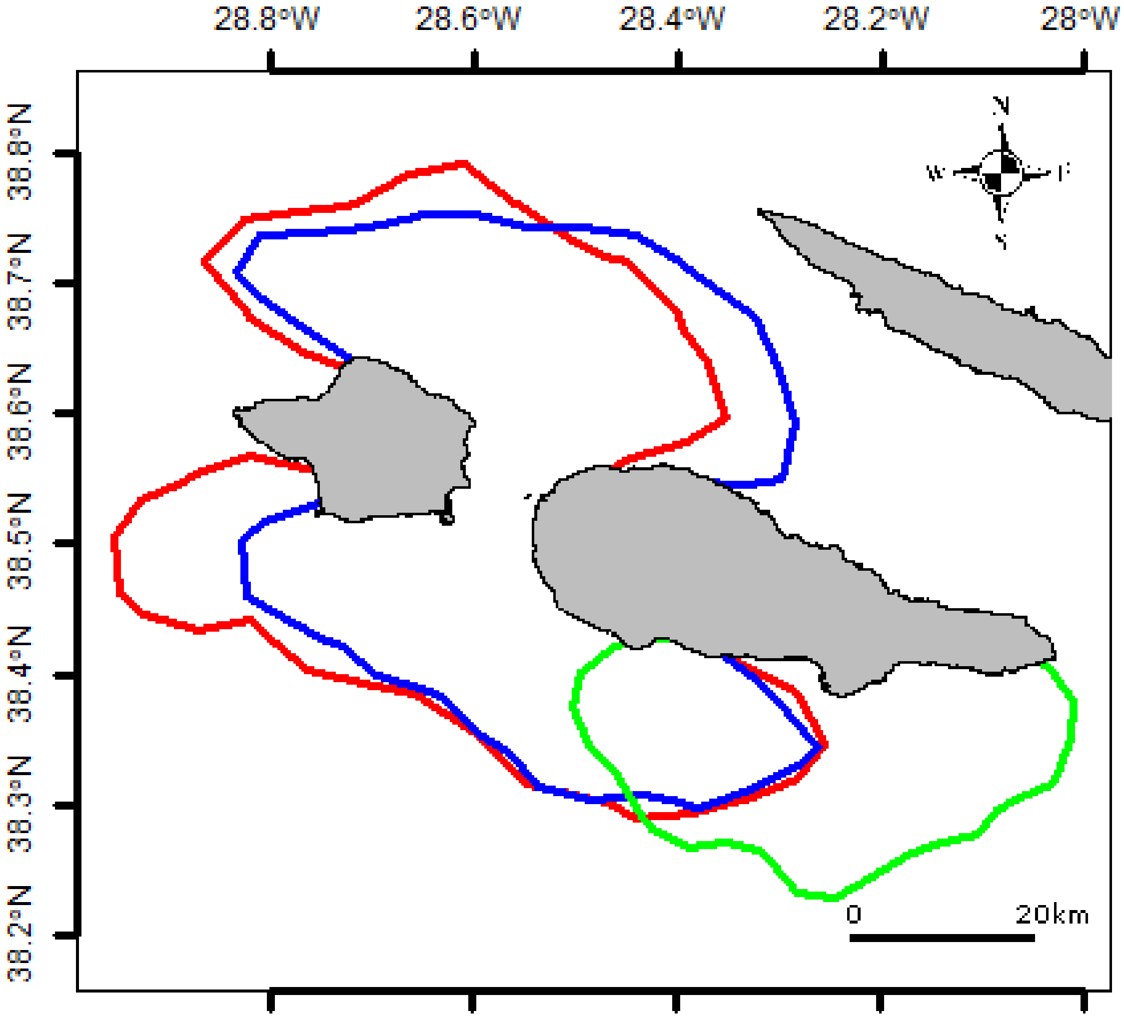


Figure 2. 7. Area used by the cetacean watching boats monitored in 2013 and 2018. The contour lines represent the 90% kernel utilization distribution of boats from Horta (red), Madalena (blue) and Lajes (green).

2.3.2. IDENTIFICATION OF CETACEAN WATCHING BEHAVIOUR

2.3.2.1. METHOD DEVELOPMENT

The analyses of the speed densities showed the same pattern across all trips and therefore only an example is presented below. The kernel density estimator (KDE) indicated that boats attained higher speeds at the beginning and end of the trips, than during the remainder of the trip. In 11 of 14 boats, the initial section of the trip had significantly higher speeds than the final section (Figure 2.8).

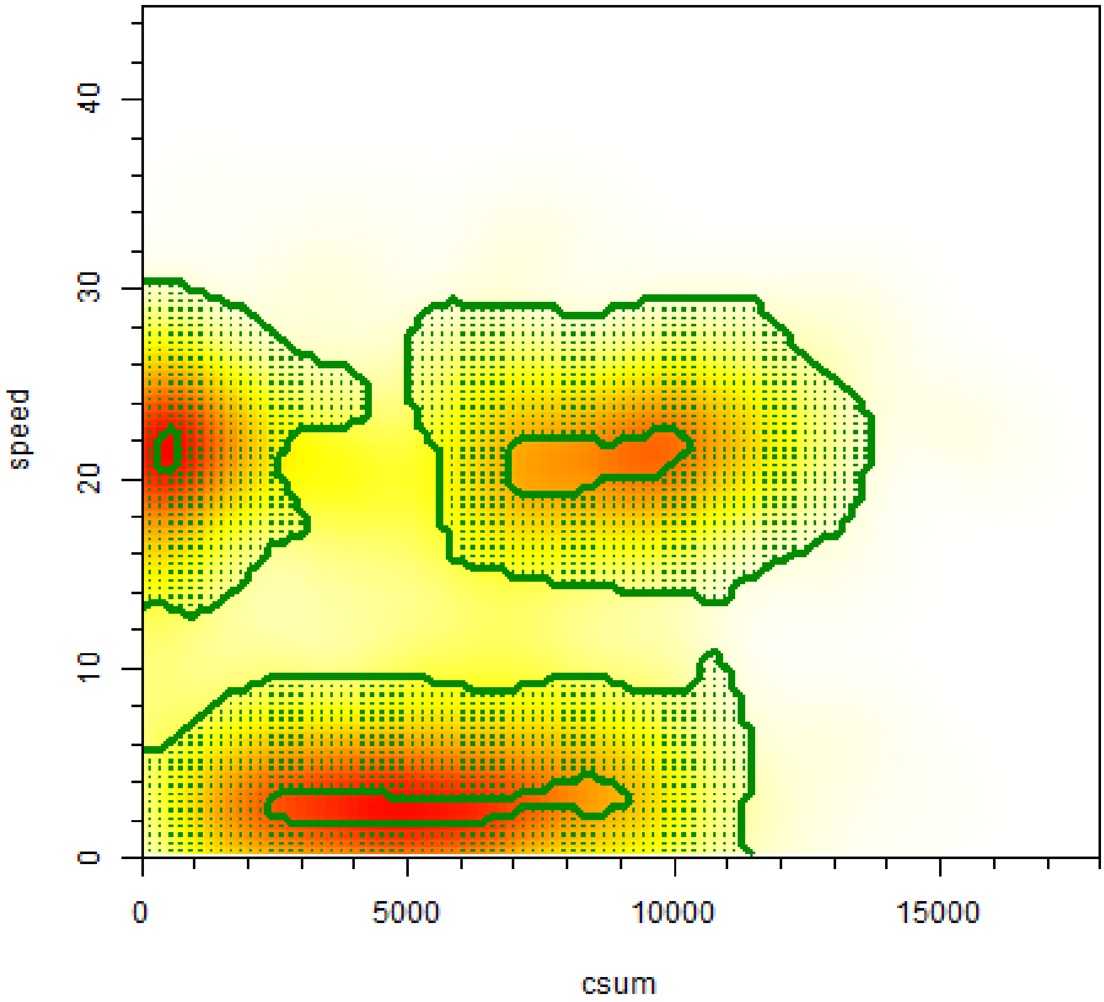


Figure 2. 8. Example of Feature significance for kernel density estimation for one boat. The graph shows how speed (y-axis) varied with the duration of trip (cumulative sum of time, csum). Red and yellow indicate higher and lower density estimates, while the green contours flag significant KDE gradient regions.

Based on the results of the kernel density estimator, the trips were split into three sections: *leaving the port, core* and *entering the port*. The Two-sample Kolmogorov-Smirnov test showed that the speed distributions when leaving or entering the port were significantly different from that of the core section of the trip, both for each boat and when all boats were pooled together ($p<0,05$ for all comparisons; results not shown here). Hence, the variable speed was used to indicate the two main activities of cetacean watching boats: cetacean watching behaviour or transiting.

The two modality tests, Dip and bimodality coefficients, yielded slightly different results (Tables in appendix B) but showed a bimodal distribution of speeds for the majority of boats (the most conservative test rejected the hypothesis of bimodality in one out of 14 boats) (Table App. 2.2). The bimodal distribution of speeds was associated with their two main activities, the mode with the lowest speeds corresponding to cetacean watching, and the mode of higher speeds, reflecting transiting.

Subsequently, the behaviour of boats was classified based on the confidence interval of speeds calculated with the GMM for each boat (Table 2.3). Using the bimodal distribution of speeds, the mean and variance of speed for boats engaged in cetacean observation was estimated as ranging from 1-5,8 kn and 1,9-15,1, respectively, for all the boats studied.

Table 2. 3. Confidence interval by boat. Boat: id number of the boat, n: number of modes, μ: mean of the lowest speed mode, s2: variance of the lowest speed mode.

Boat	n	μ	s^2
1	2	3,58	5,44
2	2	2,74	2,11
3	2	3,06	3,36
4	2	2,20	3,34
5	2	3,48	5,32
6	2	4,84	14,61
7	2	3,49	5,18
8	2	2,79	2,34
9	2	3,34	5,16
10	2	2,58	1,97
11	2	5,11	9,94
12	2	5,81	15,18
13	2	3,80	6,72
14	2	3,40	5,92
15	2	2,54	3,60

2.3.2.2. VALIDATION METHOD

The validation was carried out using data from boat 15 in 2014 (235 sightings) and from boat 14 in 2017 (321 sightings). The method correctly identified 81% of cetacean observation periods in trip segments, and 81-88% of transiting periods (Table 2.4).

Table 2. 4. Results of confusion matrix. Boat: id number of the boat. True_P: true positive, False_N: false negative, True_N: true negative and False_P: false positive. COB: percentage of segments (5min) where boats were engaged in cetacean observation, S: percentage of segments (5min) where boats were transiting.

Boat	True_P	False_N	True_N	False_P	COB	S
14	81	19	88	12	55	45
15	81	19	81	19	66	34

2.3.2.3. CETACEAN WATCHING EFFORT MAPS

The effort maps included 11,753 data points from 10 cetacean watching boats, from 2018 and 2013. These two years were used as case studies of the Faial-Pico channel and the Lajes do Pico fleets, respectively.

Figure 2.9 (top) shows the effort of cetacean watching in 2013 and Figure 2.9 (bottom) in 2018 calculated from the combination of AIS and GPS data. For boats from Lajes, cetacean watching was entirely concentrated south of Pico, and mostly around the home port. For boats from Madalena and Horta, most cetacean watching occurred N-NE of Faial Island and along the southern coasts of Faial and Pico Islands. Cetacean watching effort clearly increased from May to July and decreased again in September for boats from Lajes home ports (Figure 2.10). Boats from Madalena and Horta also exhibited monthly variations in the location of the activity, with the effort mostly concentrated in the southern sector on most months, with the exception of August, where the majority of cetacean watching occurred north of Faial (Figure 2.11).

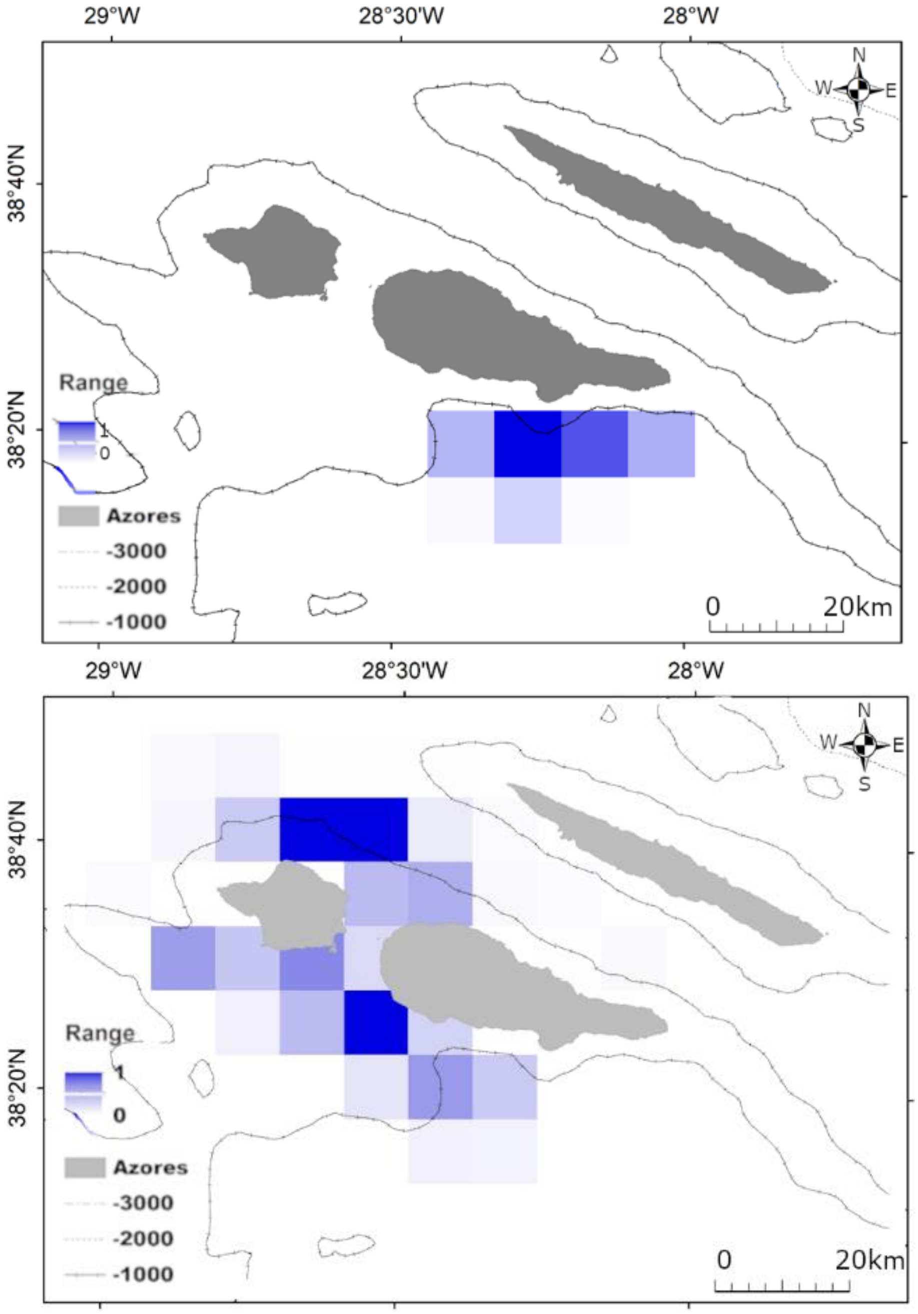


Figure 2. 9. Spatial pattern of cetacean watching effort from boats from Lajes in 2013 (top) and Madalena and Horta (bottom). The kernel density of the number of trips was calculated by 10 km x 10 km cells, and scaled to 0-1. Effort was calculated from GPS data.

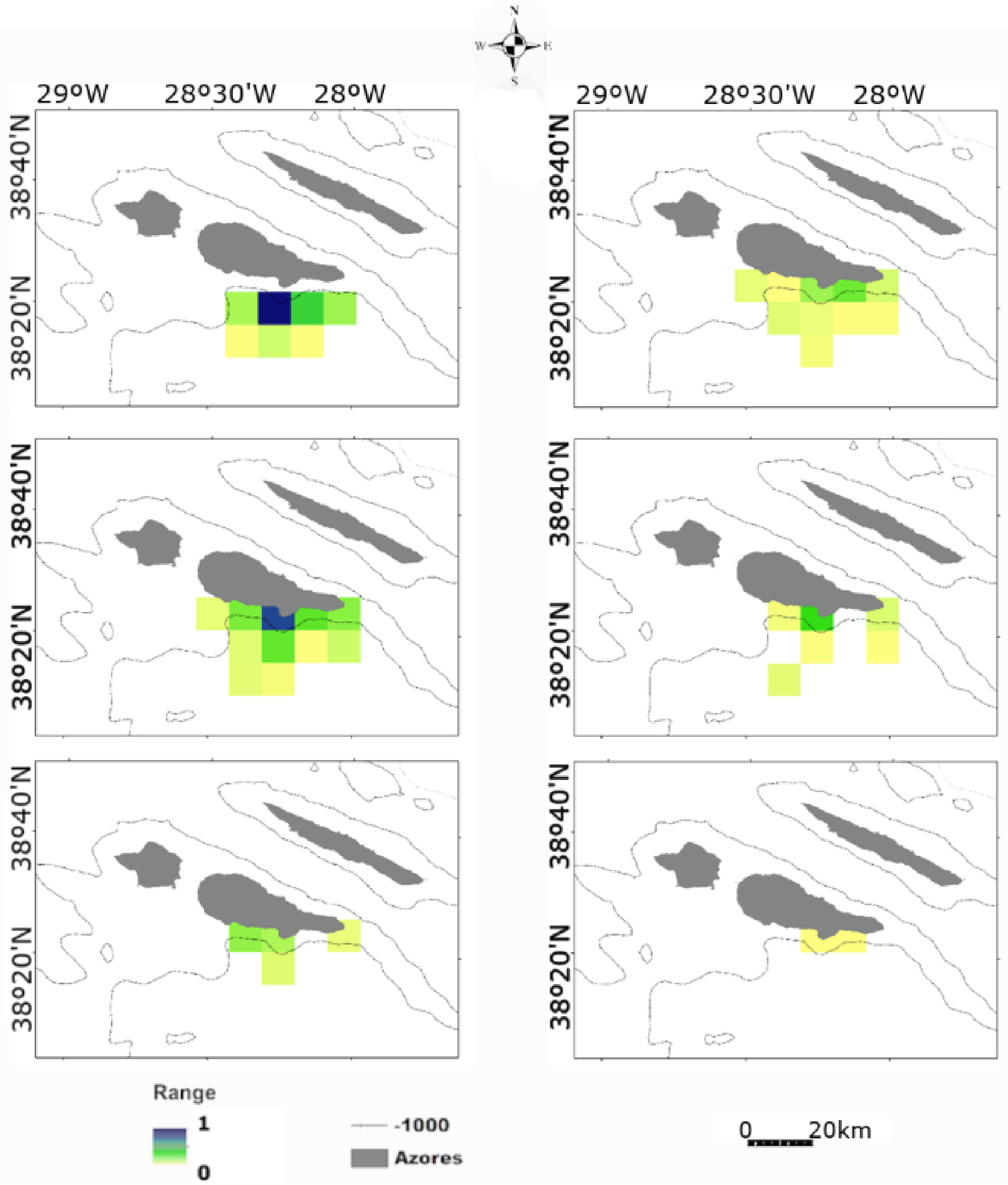


Figure 2. 10. Monthly variation in the spatial pattern of cetacean watching effort from boats from Lajes in 2013. May: (top left), June (middle left), July (bottom left), August (top right), September (middle right), October (bottom right).The kernel density of the number of trips was calculated by 10 km x 10 km cells, and scaled to 0-1. Effort was calculated from GPS data.

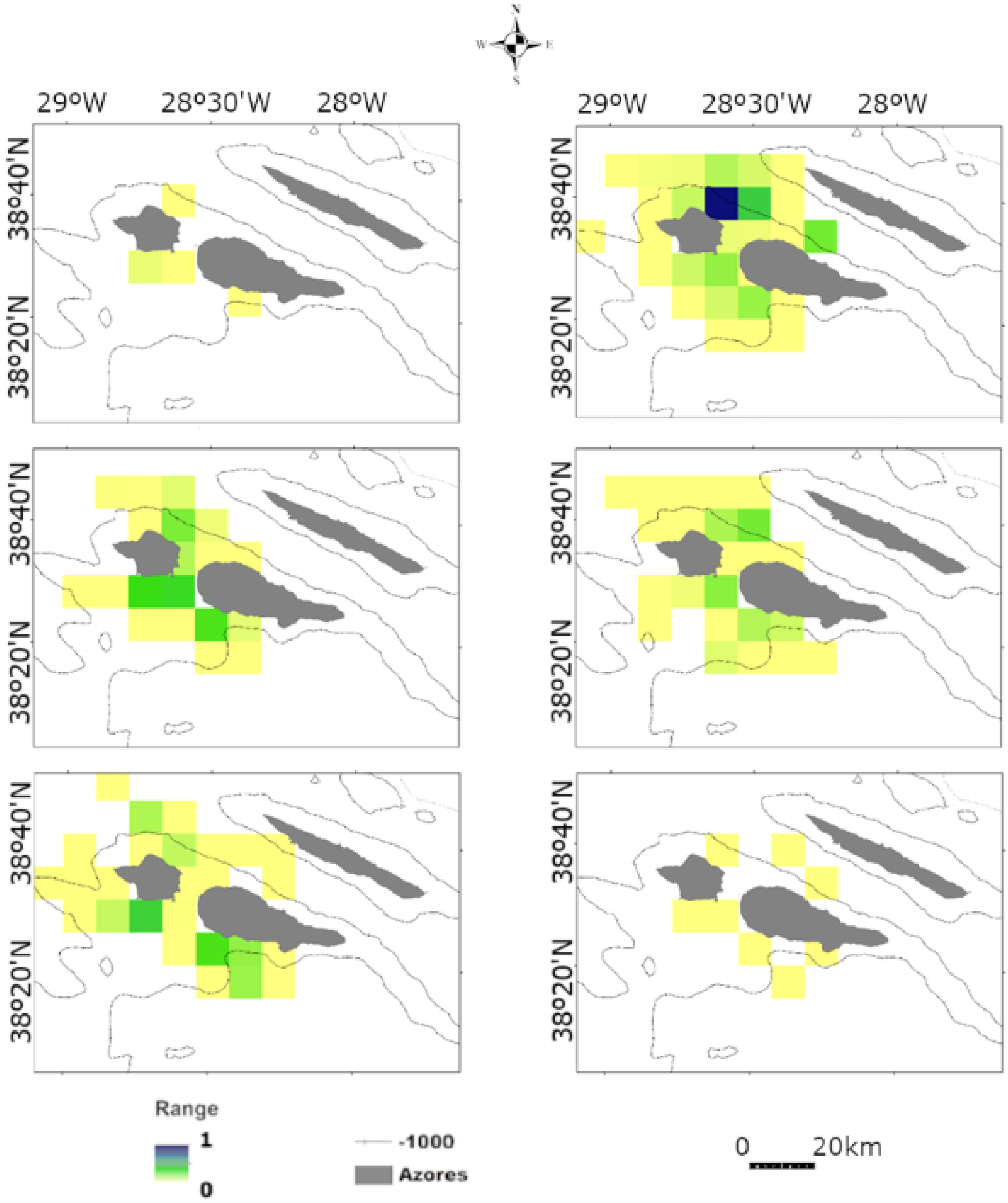


Figure 2. 11. Monthly variation in the spatial pattern of cetacean watching effort from boats from Horta and Madalena in 2018. May: (top left), June (middle left), July (bottom left), August (top right), September (middle right), October (bottom right). The kernel density of the number of trips was calculated by 10 km x 10 km cells, and scaled to 0-1. Effort was calculated from AIS and GPS data.

2.4. DISCUSSION

In this chapter I developed a multiple-step method to automatically identify the activity of cetacean watching boats using AIS and GPS tracking data, and to map cetacean watching effort, without inflating effort by including time when boats are engaged in other activities, including transiting. Similar approaches have been applied to classify activity of fishing boats [32-35], but to the best of my knowledge, this is the first time the method has been used to clearly distinguish activities of cetacean watching vessels.

The data exploration process revealed that cetacean watching trips from different home ports have different lengths, durations and spatial distributions. These differences have not been previously described but have significant implications for the estimation of cetacean watching effort and to the characterization of the activity. The boats with home ports in Horta and Madalena, both located in the channel between Pico and Faial islands, have a greater home range than those from Lajes do Pico. In addition, boats from Lajes make shorter trips (in length and time). Despite these differences, all the boats analysed showed the same distribution of speeds. Thus, the approach to identify cetacean watching activity was applied to all boats, regardless of their home port.

One possible caveat of using AIS derived positions to estimate boat activity is that messages may be missed depending on boat distance to ground stations, atmospheric conditions, and equipment failure [34, 54]. Notwithstanding, in this work, missing messages did not have an effect on the distribution of speed frequencies. That may be in part because cetacean watching vessels stayed within reception range from ground stations and the contact is seldom lost. This result supports the use of AIS derived data for this purpose.

As expected, boat speed was a reliable indicator of the activity of cetacean watching boats. All the methods used demonstrated that, in general, boat speed followed a bimodal distribution, with the highest speeds attained in the beginning and end of the trips, when the boats were transiting from and to the home port, and the slowest speeds in the middle (core) of trips, when boats spent more time engaged in cetacean watching.

Validation of the method using an independent dataset from two boats showed that the method was capable of identifying cetacean watching behaviour with very good accuracy. Validation showed that 81% of the trip segments were correctly identified by the algorithm as periods of cetacean observation, and 81-88% of the segments were correctly classified as not being in cetacean observation. Interestingly, although I could expect different behaviours from cetacean watching and research boats, given they have different goals, they showed similar speed distribution patterns while searching and observing cetaceans.

The method developed proved reliable and enables calculating and mapping cetacean watching effort much more easily and quickly than by crossing annotated sighting locations to the trips of each boat. Furthermore, it enables calculating the effort for the boats that do not record sighting information during their trips, which is the large majority of cetacean watching boats in the study area. However, the method does not enable detecting individual cetacean encounters, nor does it enable identifying the cetacean species encountered. To study in more detail the location and frequency of encounters per species, it would be necessary to collect that information from the boat logbook. This was beyond the focus of this chapter, however, which aimed to find a robust way to characterize the activity effort in general, and not to know which species were targeted.

2.4.1. CETACEAN WATCHING EFFORT

Lost AIS messages can underestimate the calculation and mapping of cetacean watching effort. Therefore, the results presented in this chapter represent the minimum effort made in the area by the boats analysed. These maps should not be interpreted as cetacean distribution because other factors influence the distribution effort of the activity, including proximity to the home port, sea state and weather conditions, information on the presence of cetaceans from lookouts and from other boats, among others [41, 42, 55, 56].

Annual maps revealed a heterogeneously distributed cetacean watching effort within the study area. The 2013 map shows that boats from Lajes do Pico concentrated their effort around the port. On the 2018 map, Horta and Madalena boats concentrated their activity north of Faial and south of Faial and Pico islands. Maps show an overlap in use from boats of all home ports southwest of Lajes do Pico. The areas of greatest effort reflect the most widely used observation hotspots by the lookouts: Lajes do Pico, São Mateus, Monte da Guia, Cedros, and Santana. This reinforces the idea that the distribution of cetacean watching effort is strongly influenced by the location of the lookouts that guide the boats to the cetacean sightings. The monthly maps of cetacean watching effort agree with published information on the cetacean watching activity in the area, which begins in spring, increases in summer and decreases in autumn [42, 56, 57]. These maps also show a slight displacement of the effort of the boats operating from Horta and Madalena towards N of Faial in August.

Mapping the effort of the cetacean watching activity is an important step for the characterization of this activity in the Azores archipelago and fundamental for managing this activity. In chapter 5, these maps were used to study the overlap with distribution of blue, fin, sei and sperm whales, and to evaluate whether these whales have suitable habitats outside the whale-watching area.

2.5. CONCLUSIONS

Cetacean watching boats of at least eight countries have AIS devices, voluntarily or mandatory [58, 59]. The number of cetacean watching boats equipped with AIS devices is expected to increase in the coming years, providing a unique opportunity to monitor cetacean watching boats remotely. In addition, several cetacean watching boats routinely collect their GPS tracks. However, neither AIS nor GPS raw data provide detailed information on the activity of the boats that enables separating navigation or transiting periods from times when the boat is actively engaged in cetacean searching or watching. If cetacean watching effort is calculated without differentiating the behavioural state of the boat (transiting versus cetacean watching), it could lead to the overestimation of the effort of the activity and to the incorrect identification of observation areas.

The method applied here provides a robust way of detecting periods of cetacean watching during trips, enabling a more accurate quantification of the effort and areas used by the activity. The method follows the approach applied to identify fishing events. It relies on the analysis of the distribution of boat speeds and on using robust methods (such as the GMM used here) to find the confidence interval of speed values associated with different boat activities. The method is easily reproducible and provides an efficient and fast way for processing large AIS or GPS datasets, in an automated manner.

Since AIS is a near-real time data system, after model parametrization this method enables classifying individual trips also in near-real time, offering a powerful tool for management and enforcement. For example, it can be used to monitor breaches of restricted areas or periods. It also enables adaptive and dynamic management, by providing up to date information on the activity dynamics. The method can also be tuned to other marine tourism activities (such as scuba diving, bird-, pinniped-, and shark-watching) and support policy making towards better management of marine ecotourism.

2.6. REFERENCES

1. Hoyt E. The worldwide value and extent of whale watching 1995. First ed. Bath, UK: Whale and Dolphin Conservation Society; 1995. 1-36 p.
2. Hoyt E. Whale watching 2001: worldwide tourism numbers, expenditures, and expanding socioeconomic benefits. Yarmouth Port, MA, USA: International Fund for Animal Welfare; 2001. 1-158 p.
3. O’Connor S, Campbell R, Cortez H, Knowles T. Whale Watching Worldwide: tourism numbers, expenditures and expanding economic benefits, a special report from the International Fund for Animal Welfare. International Fund for Animal Welfare, 2009.
4. Hoyt E. Whale and Dolphin Watching in Europe. In: Laetitia Nunny WAW, La Garriga, Spain, editor. Under Pressure. First ed: OceanCare; 2021. p. 86-96.
5. Ressurreiçao A, Cardigos F, Giacomello E, Leite N, Oliveira F, Kaiser MJ, et al. The value of marine ecotourism for an European outermost region. Ocean and Coastal Management. 2022;222:106129. doi: https://doi.org/10.1016/j.ocecoaman.2022.106129.
6. Hoyt E, Hvenegaard GT. A review of whale-watching and whaling with applications for the Caribbean. Coastal Management. 2002;30(4):381-99. doi: 10.1080/08920750290027 3.
7. Carlson C. A review of whale watch guidelines and regulations around the world: version 2012. International Whaling Commission, 2012.

8. Lambert E, Hunter C, Pierce GJ, MacLeod CD. Sustainable whale-watching tourism and climate change: towards a framework of resilience. Journal of Sustainable Tourism. 2010;18(3):409-27. doi: 10.1080/09669581003655497.

9. Lynch H, Crosbie K, Fagan W, Naveen R. Spatial patterns of tour ship traffic in the Antarctic Peninsula region. Antarctic Science. 2010;22(2):123-30. doi: 10.1017/S0954102009990654.

10. Schaffar A, Garrigue C. Review of commercial humpback whale watching activities in the South Pacific. IWC, 2007.

11. Carson-Jackson J. Satellite AIS–developing technology or existing capability? The Journal of Navigation. 2012;65(2):303-21. doi: https://doi.org/10.1017/S037346331100066X.

12. IWC. General Principles for Whalewatching International Whaling Commission; 1996 [cited 2020 6 july 2020]. Available from: https://iwc.int/wwguidelines#manage.

13. Prakash TSL, Gangodawila N, Jayakody S, Makandura G, Amarasinghe N, Leisure. Current perceptions and the need for a strategic plan for the whale watching industry in Mirissa, Sri Lanka. African Journal of Hospitality, Tourism. 2019;8(3):1-16.

14. Buultjens J, Ratnayke I, Gnanapala A. Whale watching in Sri Lanka: Perceptions of sustainability. Tourism Management Perspectives. 2016;18:125-33. doi: https://doi.org/10.1016/j.tmp.2016.02.003.

15. Kessler M, Harcourt R. Whale watching regulation compliance trends and the implications for management off Sydney, Australia. Marine Policy. 2013;42:14-9. doi: http://dx.doi.org/10.1016/j.marpol.2013.01.016.

16. Hoyt E. Sustainable ecotourism on Atlantic islands, with special reference to whale watching, marine protected areas and sanctuaries for cetaceans. Biology and environment: proceedings of the Royal Irish Academy. 2005;105B(3):141-54.

17. WCED. Our common future. The report of the World Commission on Environment and Development. Oxford: Oxford University Press., 1987.

18. Saarinen J. Traditions of sustainability in tourism studies. Annals of tourism research. 2006;33(4):1121-40. doi: doi:10.1016/j.annals.2006.06.007.

19. Butler RW. Sustainable tourism: A state-of-the-art review. Tourism Geographies. 1999;1:1: 7-25. doi: 10.1080/14616689908721291.

20. Bentz J, Rodrigues A, Dearden P, Calado H, Lopes F. Crowding in marine environments: Divers and whale watchers in the Azores. Ocean & Coastal Management. 2015;109:77-85. doi: https://doi.org/10.1016/j.ocecoaman.2015.03.001.

21. Tezanos-Pinto G, Constantine R, Mourão F, Berghan J, Scott Baker. High calf mortality in bottlenose dolphins in the Bay of Islands, New Zealand–a local unit in decline. Marine Mammal Science. 2015;31(2):540-59. doi: 10.1111/mms.12174.

22. Higham JE, Bejder L, Allen SJ, Corkeron PJ, Lusseau D. Managing whale-watching as a non-lethal consumptive activity. Journal of sustainable tourism. 2016;24(1):73-90. doi: http://www.tandfonline.com/doi/full/10.1080/09669582.2015.1062020#.VcQtO_mqpBc.

23. Redfern J, Ferguson MC, Becker E, Hyrenbach K, Good C, Barlow J, et al. Techniques for cetacean–habitat modeling. Marine Ecology Progress Series. 2006;310:271-95. doi: doi:10.3354/meps310271.

24. Durban J, Elston D, Ellifrit D, Dickson E, Hammond P, Thompson PM. Multisite mark-recapture for cetaceans: Population estimates with Bayesian model averaging.

Marine Mammal Science. 2005;21(1):80-92. doi: https://doi.org/10.1111/j.1748-7692.2005.tb01209.x.

25. Christiansen F, Lusseau D. Linking behavior to vital rates to measure the effects of non-lethal disturbance on wildlife. Conservation Letters. 2015;8(6):424-31. doi: 10.1111/conl.12166.

26. Berrow SD, Holmes B. Tour boats and dolphins: A note on quantifying the activities of whalewatching boats in the Shannon estuary, Ireland. Journal of Cetacean Research Management. 1999;1(2):199-204.

27. Almunia J, Delponti P, Rosa F. Using big data to estimate whale watching effort. bioRxiv. 2020:1-19. doi: http://doi.org/10.1101/2020.11.30.403923.

28. Lee E, Mokashi AJ, Moon SY, Kim G. The maturity of automatic identification systems (AIS) and its implications for innovation. Journal of Marine Science. 2019;7(9):287. doi: 10.3390/jmse7090287.

29. McCauley DJ, Woods P, Sullivan B, Bergman B, Jablonicky C, Roan A, et al. Ending hide and seek at sea. Science. 2016;351(6278):1148-50. doi: 10.1126/science.aad5686.

30. Robards M, Silber G, Adams J, Arroyo J, Lorenzini D, Schwehr K, et al. Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)—a review. Bulletin of Marine Science. 2016;92(1):75-103. doi: http://dx.doi.org/10.5343/bms.2015.1034.

31. Vermard Y, Rivot E, Mahévas S, Marchal P, Gascuel D. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. Ecological Modelling. 2010;221(15):1757-69. doi: 10.1016/j.ecolmodel.2010.04.005.

32. Mazzarella F, Vespe M, Damalas D, Osio G, editors. Discovering vessel activities at sea using AIS data: Mapping of fishing footprints. 17th International conference on information fusion (FUSION); 2014: IEEE.

33. Vespe M, Gibin M, Alessandrini A, Natale F, Mazzarella F, Osio GC. Mapping EU fishing activities using ship tracking data. Journal of Maps. 2016;12(sup1):520-5. doi: 10.1080/17445647.2016.1195299.

34. Mazzarella F, Vespe M, Alessandrini A, Tarchi D, Aulicino G, Vollero A. A novel anomaly detection approach to identify intentional AIS on-off switching. Expert Systems with Applications. 2017;78:110-23. doi: 10.1016/j.eswa.2017.02.011.

35. Natale F, Gibin M, Alessandrini A, Vespe M, Paulrud A. Mapping Fishing Effort through AIS Data. PloS one. 2015;10(6):e0130746. Epub 2015/06/23. doi: 10.1371/journal.pone.0130746. PubMed PMID: 26098430; PubMed Central PMCID: PMC4476776.

36. de Souza EN, Boerder K, Matwin S, Worm B. Improving fishing pattern detection from satellite AIS using data mining and machine learning. PloS one. 2016;11(7):e0158248. doi: 10.1371/ journal.pone.0158248.

37. Bez N, Walker E, Gaertner D, Rivoirard J, Gaspar P. Fishing activity of tuna purse seiners estimated from vessel monitoring system (VMS) data. Canadian Journal of Fisheries Aquatic Sciences. 2011;68(11):1998-2010. doi: 10.1139/f2011-114.

38. Decreto Legislativo Regional n.º 10/2003/A, de 22 de março - Altera o Decreto Legislativo Regional n.º 9/99/A, de 22 de março (2003).

39. Decreto Legislativo Regional n.º 13/2004/A, de 23 de março - Altera o Decreto Legislativo Regional n.º 10/2003/A, de 22 de março (2004).

40. Portaria n.º 5/2004, de 29 de janeiro (2005).

41. Bentz J, Dearden P, Calado H. Strategies for marine wildlife tourism in small islands–the case of the Azores. Journal of Coastal Research. 2013;65(sp1):874-9. doi: https://doi.org/10.2112/SI65-148.1.

42. Sequeira M, Elejabeitia C, Silva M, Dinis A, de Stephanis R, Urquiola E, et al. Review of whalewatching activities in mainland Portugal, the Azores, Madeira and Canary archipelagos and the Strait of Gibraltar. Cetacean Research and Management. 2009; SC61/WW11:1-41.

43. Oliveira CI. A actividade de observação turística de cetáceos no arquipélago dos Açores: contribuição para o seu desenvolvimento sustentável [Master's thesis]: Universidade dos Açores; 2005.

44. Wing MG, Eklund A, Kellogg LD. Consumer-grade global positioning system (GPS) accuracy and reliability. Journal of Forestry. 2005;103(4):169-73. doi: https://doi.org/10.1093/jof/103.4.169.

45. Eriksen T, Olsen Ø. Vessel tracking using automatic identification system data in the Arctic. Sustainable Shipping in a Changing Arctic. 2018;7(WMU Studies in Maritime Affairs):115-36. doi: https://doi.org/10.1016/j.marpol.2018.03.028.

46. Iphar C, Napoli A, Ray C, editors. Detection of false AIS messages for the improvement of maritime situational awareness. Proceedings of the 30th Annual Computer Security Applications Conference; 2015.

47. Spaans J. GPS: The Holy Grail? Journal of Navigation. 2000;53(2):293-7. doi: 10.1017/S0373463300008742.

48. Ford JH, Peel D, Kroodsma D, Hardesty BD, Rosebrock U, Wilcox C. Detecting suspicious activities at sea based on anomalies in Automatic Identification Systems transmissions. PloS one. 2018;13(8):e0201640. Epub 2018/08/10. doi: 10.1371/journal.pone.0201640. PubMed PMID: 30091985; PubMed Central PMCID: PMC6084947.

49. Pallotta G, Vespe M, Bryan K. Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction. Entropy. 2013;15(12):2218-45. doi: 10.3390/e15062218.

50. Sector I-RR. Technical characteristics for a universal shipborne Automatic Identification System using time division multiple access in the VHF maritime mobile band, ITU-R. Rec. ITU-R M.1371-5 2014.

51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B. 1995;57(1):289-300. doi: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

52. Williams TM. Swimming. In: Würsig B, Thewissen J, Kovacs KM, editors. Encyclopedia of Marine Mammals. Third ed: Elsevier; 2018. p. 970-9.

53. Silverman BW. Density estimation for statistics and data analysis. First ed. New York: Routledge; 1986.

54. Nguyen D, Vadaine R, Hajduch G, Garello R, Fablet R, editors. A multi-task deep learning architecture for maritime surveillance using AIS data streams. 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA); 2018: IEEE.

55. Gomes-Pereira JND. Daily species checklist from whale-watching—studying the research potential with an Azorean case study. Journal of the Marine Biological Association of the United Kingdom. 2008;88(6):1283-8.

56. Magalhães S, Prieto R, Silva MA, Gonçalves J, Afonso-Dias M, Santos RS. Short-term reactions of sperm whales (Physeter macrocephalus) to whale-watching vessels in the Azores. Aquatic Mammals. 2002;28(3):267-74.

57. Bentz J, Lopes F, Calado H, Dearden P. Managing marine wildlife tourism activities: Analysis of motivations and specialization levels of divers and whale watchers. Tourism Management Perspectives. 2016;18:74-83. doi: https://doi.org/10.1016/j.tmp.2016.01.004.

58. TFSR. National GMDSS Task Force. Newsletter and Summary Record of 16 May 2019 Meeting. 2019. p. 1-8.

59. Nazha N. Expanding AIS requirements In: Bulletin SS, editor. Ship Safety Bulletin ed: Government of Canada; 2019.

CHAPTER 3

3. SHIPPING TRAFFIC

3.1. INTRODUCTION

Maritime traffic was the most significant contributor to international trade in Europe in the last decade (2010-2019) [1]. In 1980, it was estimated that the world fleet had a carrying capacity of 672,142 dead-weight tons (dwt, an indicator of the cargo carrying capacity of a ship); in 2021, this capacity exceeded 2,134,640 dwt [2]. In the last decade (2011-2021), the number of vessels went from 83,000 to 99,000. International trade encourages the global circulation of goods and people, favouring the growth of the economy. However, marine traffic can pose a significant threat to biodiversity [3], by increasing collisions with wild animals [4], chemical and underwater noise pollution [5, 6], gas emissions [7], and transport non-indigenous species [8].

The first scientific studies of vessel collisions with cetaceans began in the early 2000s and were based on recordings in the vessels' logbook and analysis of stranded cetaceans [9, 10]. These earlier studies provided the first information about the severity of the problem, identified some high risk areas and species [9]. In 2007 a global database of vessel collisions with cetaceans was established [11, 12]. Nevertheless, quantifying vessel collisions with cetaceans is difficult because many incidents go unnoticed or are not recorded, and carcasses of cetaceans struck can sink to the bottom of the sea [13].

Maritime traffic safety was improved by monitoring the vessels with the Automatic Identification System (AIS) to prevent collisions between ships and other accidents [14]. The monitoring data includes enough information to estimate the shipping density, which provides relevant information for maritime spatial planning and impact assessment [15]. Shipping density can also be used to estimate the risk of oil spills or of bioinvasions, predict shipping noise levels and noise exposure of different marine taxa [16, 17]. AIS data also provides an opportunity to estimate the probability of vessel collisions with cetaceans, identify high-risk areas, or evaluate effectiveness of measures to mitigate collisions [12, 18-21].

The Azores archipelago is located nearby major shipping routes connecting countries on both sides of the Atlantic Ocean and the Azorean harbours are also used by some of these vessels. In addition, local maritime traffic links the islands of the Azores, the Azores islands to the mainland, and other countries [22]. Marine traffic can pose a significant threat to cetaceans using the Azores region, particularly large baleen whales and sperm whales (*Physeter macrocephalus*) that are very susceptible to collisions with vessels. In addition, marine traffic may also directly impact other taxa (e.g., sea turtles, seabirds, large sharks), alter important habitats through chemical or acoustic pollution (many of which are protected under regional, national or international regulations [23-27]), thereby affecting the structure and functioning of marine ecosystems in the Azores. It is therefore fundamental to characterize the shipping traffic in the region and to identify high-shipping density areas to investigate potential impacts to marine life.

Freely available data on shipping traffic usually consists of images or layers hosted in an online map viewer with aggregated information for whole months or years [28]. Although the data provide a general vision of the maritime activity, they contain insufficient detail to characterize the shipping density or to support predictions of vessel collisions with cetaceans. To overcome this limitation, in this chapter I analysed a dataset of AIS class A collected by ten land stations in the Azores archipelago obtained in the scope of project MACAIS and REDAIS [29]. The aim was to develop a framework for characterizing shipping traffic, taking into account differences in vessel characteristics and operations known to influence the potential severity of collisions with cetaceans. This was achieved by calculating shipping density for different vessel types, speeds and sizes, as these attributes determine not only the probability of collisions with cetaceans but also the probability that those interactions are fatal.

3.2. METHODS

3.2.1. AIS DATA

Vessels equipped with AIS class A transmitted information to each other and land stations. The transmission of messages occurred via VHF, every few seconds or minutes, with dynamic or static information of each vessel [30]. These messages were received by ten land stations in the Azores archipelago (Table 3. 1, appendix C) and were individually stored in the AIS database.

Table 3. 1. Land stations in the Azores archipelago.

Station	Island
Pico Alto	Sta Maria
Pico Bartolomeu	S. Miguel
Pico das Camarinhas	S. Miguel
Serra do Cume	Terceira
Serra de Sta Bárbara	Terceira
Pico Timão	Graciosa
Macelinha	S. Jorge
Pico Geraldo	Pico
Cabeço Gordo	Faial
Morro Alto	Flores

The spatial coverage of these VHF antennas depends on their range to the horizon from the coordinates they receive, in addition to atmospheric conditions [15, 31]. The study area was defined as the area where the estimated detection probability of each station was equal to one. The detection range of each station D was calculated by the simplified formula of distance between transmitter antenna and horizon

$$D = \sqrt{2 * Re * h1}$$

where h1 was the height of the transmitter antenna. This was then multiplied by the radius of Earth (6370 km) in wgs84 datum [32] and by 2 [33]. The sum of the VHF coverage of each antenna defined the extent of the study area (Figure 3.1).

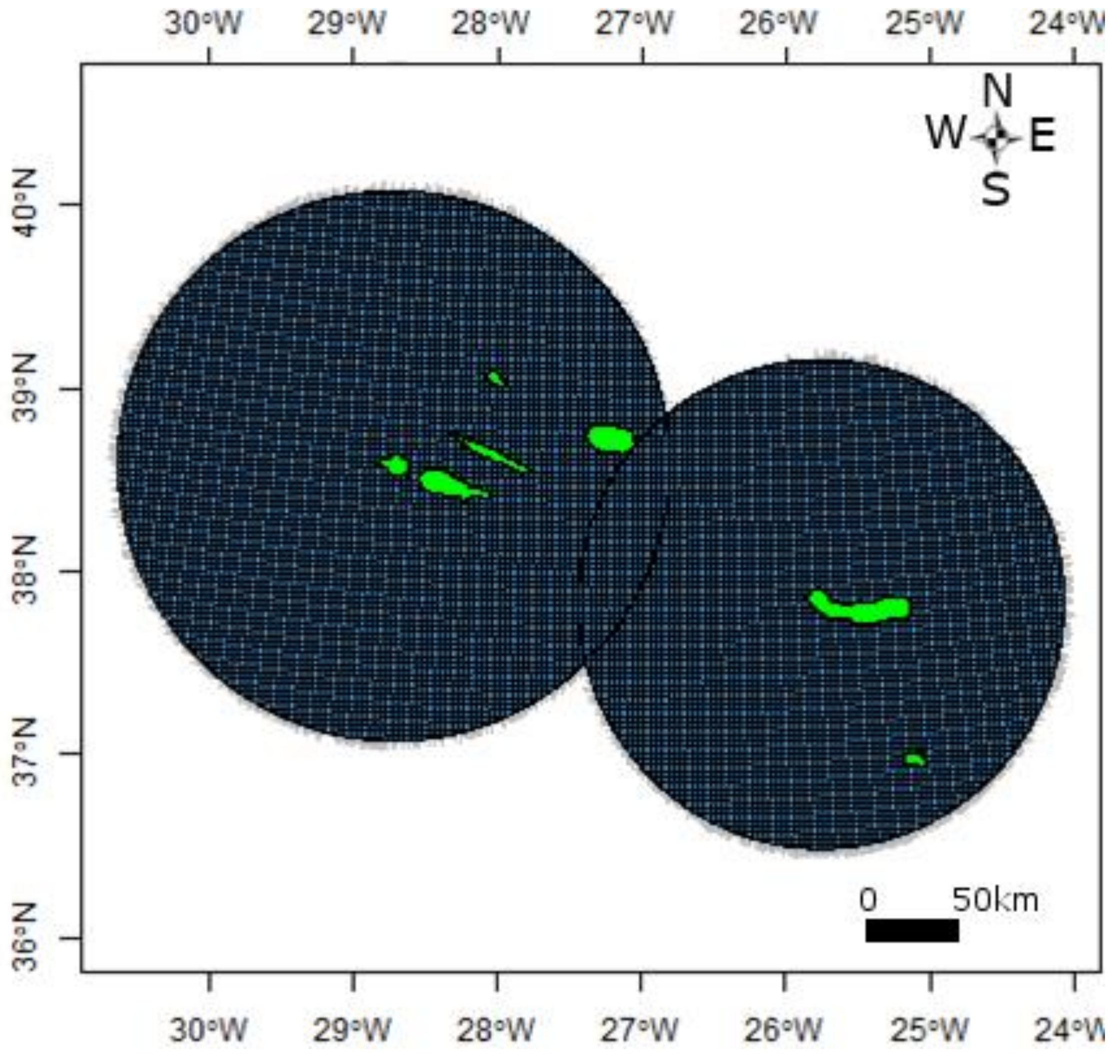


Figure 3. 1. Study area defined by the vhf coverage.

3.2.2. DATA PROCESSING

The process of decoding AIS messages and exporting to CSV format was explained in Chapter 2. Maritime traffic data were collected from 22 of October 2012 until 31 of December 2016, with some missing days (25 days in 2013, 32 in 2014, 31 in 2015) (appendix D, Table App. 3. 2). Data from December 2014 and January 2015 were removed from the analysis due to the detection of a large number of mistakes.

The huge amount of AIS data analysed in this chapter required the development of scripts to automatically clean and correct errors and inconsistencies in the database (as detailed in appendices E and F) in an iterative process. The AIS records contained several variables with information on the identity of the vessels (Maritime Mobile Service Identity – MMSI, International Maritime Organization number - IMO, Name, Callsign) (Table 3. 2). Although the IMO code is vessel-specific, only 75% of the AIS records included this code. The MMSI code has the disadvantage that is not vessel-specific but was available for 100% of the records. Therefore, MMSI was the primary code used to identify the vessels, and IMO was only used as a secondary identifier.

Table 3. 2. Summary of dynamic and static AIS variables.

Name	Description	Units	Priority
MMSI	Nine-digit numerical code		Essential
Datetimechanged	Real time coordinates	UTC sec	Essential
IMO	Seven-digit numerical code		Optional
Stationid	Identifier for this Message		Optional
Name	Vessel registered name		Optional
Callsign	7 bit ASCII characters		Optional
Country	Country of vessel registration		Optional
Cargotype	Type of vessel/cargo		Optional*
Length	Length of the boat	meters	Optional*
Beam	Beam of the boat	meters	Optional
Draught	Draught of the boat	meters	Optional
Latitude	Coordinate of the boat at that time		Essential
Longitude	Coordinate of the boat at that time		Essential
Speed	Speed over the ground (SOG)	knots	Optional*
True heading	Correcting for magnetic variation yields	degrees	Optional
Status	Vessel activity (15 categories related with common activities; ex. anchored)		Optional
Destination	Expected destination		Optional
Arrivaldatetime	Expected date and time of arrival	UTC sec	Optional

*Optional variable to identify vessels by AIS, but essential to the analysis of this chapter.

In the first step, AIS records with invalid MMSI code, or without latitude or longitude were removed. Then, the data on the type and the size of the vessel were analysed and completed. The vessels suspected of having erroneous size information, with a length outside the range 10-400 m, or with a beam larger than the length, were corrected.

Vessel types were classified into ten categories [34] (Table 3. 3), with some modifications, such as the inclusion of the High-Speed Craft (HSC) category that presents a high probability of collisions with cetaceans [35-37].

Table 3. 3. Vessel classification1.

Standard AIS classification	Simplified AIS classification
Anti-pollution equipment, law enforcement, medical transport, rescue vessels and Search and rescue vessel (SAR)	Rescue vessels
Bulk carrier, cargo (Hazmat, A,B,C or D), cointainer ship	Cargo
Cable layer, diving Ops, dredging, motorhopper, offshore supply vessel, offshore support vessel, other (Hazmat, A,B,C or D), Reserved, ship according to RR resolution no.18, spare-local vessel, suction dredger, towing, WIG(Hazmat, A, B, C or D)	Other
Fishing	Fishing
HSC(Hazmat, A,B,C or D)	HSC
Military ops, military vessels	Military vessels
Passenger(Hazmat A,B,C or D)	Passenger
Pilot, port tender, tug, vessels used for port services	Vessels used for port services
Pleasure boats, pleasure craft, sailing, scientific ship, recreational vessel	Pleasure boats
Tanker(Hazmat A,B,C or D)	Tanker

Messages with mistakes in vessel coordinates giving unreliable positions were detected and subsequently removed by: i) finding duplicate coordinates (with the same date and time); ii) calculating the vessel speed between consecutive coordinates of the same trip; iii) identifying records with speed over ground (SOG) > 80 kn. Finally, true heading values outside the 0-360º range were eliminated.

The cleaned AIS data were processed to split the vessels´ trips based on the MMSI, IMO and date. Trips were also split every time a vessel arrived at a harbour. A rectangular buffer was created around all harbours in the Azores archipelago ([38], in appendix G Figure .App. 3. 2), and the messages falling within these areas were used to split trips.

3.2.3. ANALYSIS OF MARITIME TRAFFIC

3.2.3.1. SHIPPING DENSITY

The shipping density was calculated at different temporal scales and by type of vessel, using only AIS records with speed over the ground (SOG)>1 kn [39]. The AIS records were used to calculate the time intervals between messages. This interval was influenced by several conditions, as explained in Chapter 2. Therefore, AIS records were pre-filtered to a consistent time interval of 20 min to standardise the samples to calculate shipping density. The daily shipping density was calculated by counting daily messages per unique trip within a 25x25 km grid cell (the resolution of environmental and prey variables used in Chapter 4). The daily shipping density was stored in individual rasters to use as a covariate in species distribution models in Chapter 5. The daily shipping density was

---

Note 1: adapted from Le Tixerant *et al* (2018) [34].

averaged for 2013-2016 to investigate overall spatial patterns, and the standard deviation was used to assess variability in these patterns. All analyses were done in Arcgis10.3 and R version 4.0.3 (R Core Team 2020).

3.2.3.2. SHIPPING DENSITY WITH HIGHER PROBABILITY OF LETHAL COLLISIONS

The probability of lethal collisions between vessels and cetaceans increases at vessel speeds over 14 kn [35]. At these speeds, the highest probability of lethal collision was found for vessels over 80 m in length [9]. Therefore, to identify the spatial patterns of shipping with a higher likelihood of lethal interactions with cetaceans (hereafter called lethal probability (LP) shipping), two different categories were created: one using only data from vessels travelling at speeds >14 kn, and the other using data from vessels >80 m length at speeds >14 kn (Table 3. 4). Each category was then used to compute the daily shipping density in 25x25 km grid cells. As above, daily shipping density was then averaged over years and the standard deviation was calculated to examine variability over the years.

Table 3. 4. Categories of lethal probability (LP) shipping. SOG: speed over ground, Length: length of the vessel.

Category LP shipping	Speed (SOG) (kn)	Length (m)
Speed	14 - 80	All
Speed & Size	14 - 80	>80

3.2.3.3. TRANSFORMED SHIPPING DENSITY

The total shipping density and the shipping with higher probability of lethal collisions were transformed to improve the visualization of the spatial and temporal patterns (Figure 3.2). The daily shipping density was averaged per vessel type by month and year to investigate intra- and inter-annual patterns, respectively. Then, values were scaled (by dividing by the maximum value) to range from 0 to 1 (1 being the maximum density) to account for differences in transmission and reception conditions, enabling comparison of shipping density across temporal periods and vessel types.

Figure 3. 2. Workflow to map shipping density per month and year.

3.3. RESULTS

3.3.1. NUMBER OF VESSELS AND TRIPS

A total of 8509 vessels were detected in the study area between 2012 to 2016. The number of vessels rose in the three first years of the study (from 852 to 3294), remaining above 3060 vessels in the following years (Table 3. 5). The number of vessel nationalities followed a similar trend, reaching 70 different nationalities in the latter years.

Table 3. 5. Summary of AIS data used in this study (October 2012- December 2016).

Year	Vessels	Countries	Days	Messages	Trips
2012	852	51	65	17487	3646
2013	2818	63	339	94542	22921
2014	3221	68	333	109464	24794
2015	3062	74	334	118306	28483
2016	3294	71	366	132533	31990

Cargo vessels accounted for the highest number of vessels in all years (64%), followed by tankers (24%), pleasure boats (4%), other (3.8%), fishing (1.5%) and passenger (1.5%), with the remaining vessel types comprising only 1.2% (Figure 3.3). This same pattern was observed in all years of the study (data from 2012 was excluded because it was only from the end of October onwards).

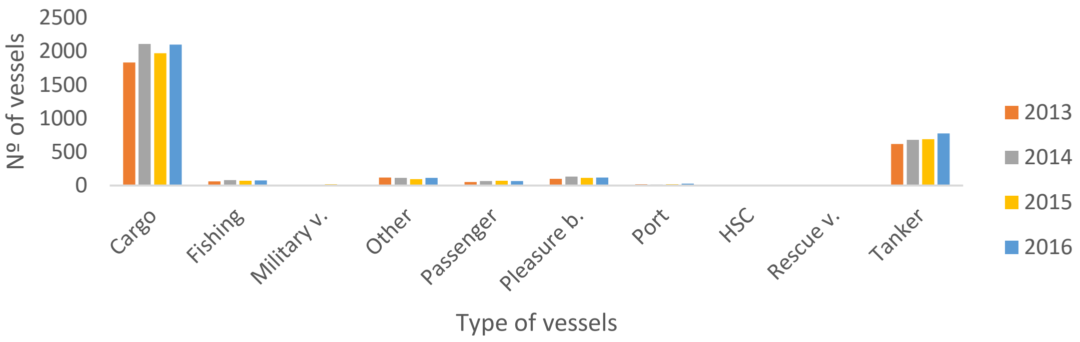


Figure 3. 3. Annual number of vessels in the study area per vessel type (October 2012 -December 2016). Military vessel (Military v.), other vessels (Other), pleasure boat (Pleasure b.), vessels used for port services (Port), High-speed craft (HSC), and rescue vessel (Rescue v.)

The number of trips by vessel type also followed similar trends throughout the years (Figure 3.4). The higher number of trips was made by cargos (~9,000 trips/year), followed closely by passenger vessels (7,000-10,000 trips/year). Fishing vessels reached almost 4,000 trips in 2015 and 2016, while tankers made around 2,000 trips in all years. The remaining vessels seldomly reached 1,000 trips/year (Figure 3.4).

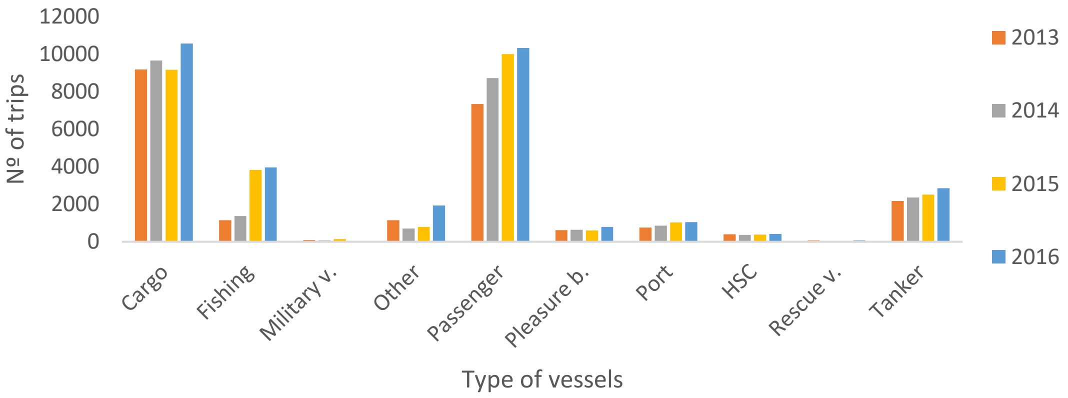


Figure 3. 4. Annual number of trips in the study area per vessel type (October 2012 -December 2016). Military vessel (Military v.), other vessels (Other), pleasure boat (Pleasure b.), vessels used for port services (Port), High-speed craft (HSC), and rescue vessel (Rescue v.)

The number of vessels per month usually ranged from 800 to 1400, with three peaks in February, June-July, and November (Figure 3.5). The number of trips ranged from 2000 to 5000 in January-March and September-December, and between 3000 and 7200 from April to August (Figure 3.5).

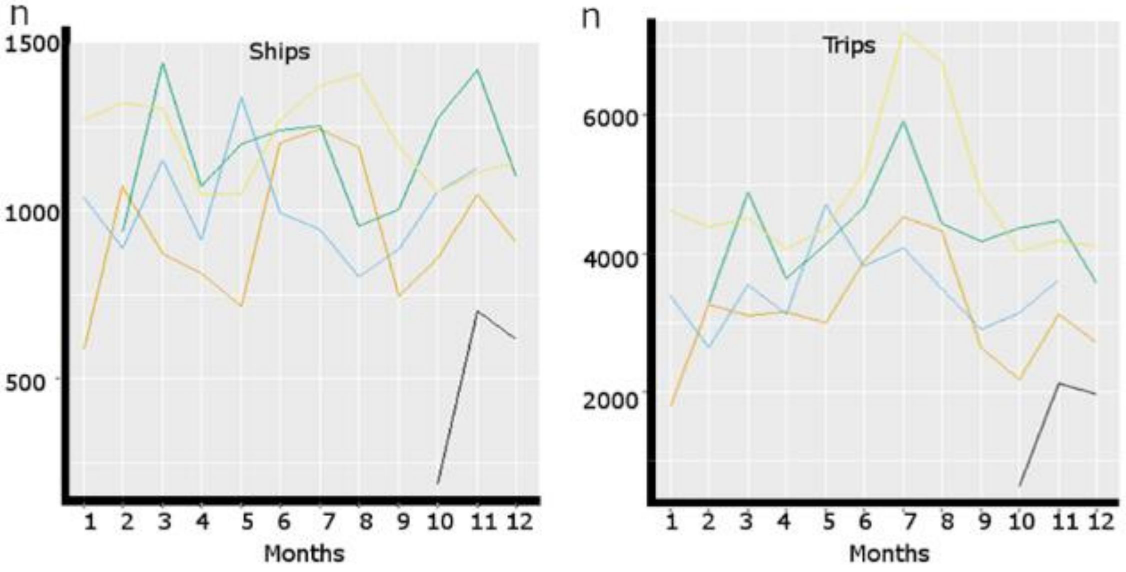


Figure 3. 5. Monthly variation in the number (n) of vessels (left) and trips (right) (October 2012 - December 2016). Black 2012; orange 2013; blue 2014; green 2015 and yellow 2016. Left to right, January to December.

The temporal patterns in the number of trips per vessel type showed two different situations (Figure 3.6). On the one hand, a group of vessel types had consistent patterns across the years. One of the most consistent was the passenger vessels, which increased in March, reached a peak in July-August, and decreased after the summer. The HSC vessels regularly transited through the area between May and September, whereas the pleasure boats had more trips between March and July. The tanker, rescue and military vessels had relatively constant values over time. The other group of vessel types showed a great variation in the number of trips across the years. Cargo vessels had some peaks in March, May, July and November in two study years. In 2015 and 2016 there were clear summer peaks in the number of trips from fishing vessels.

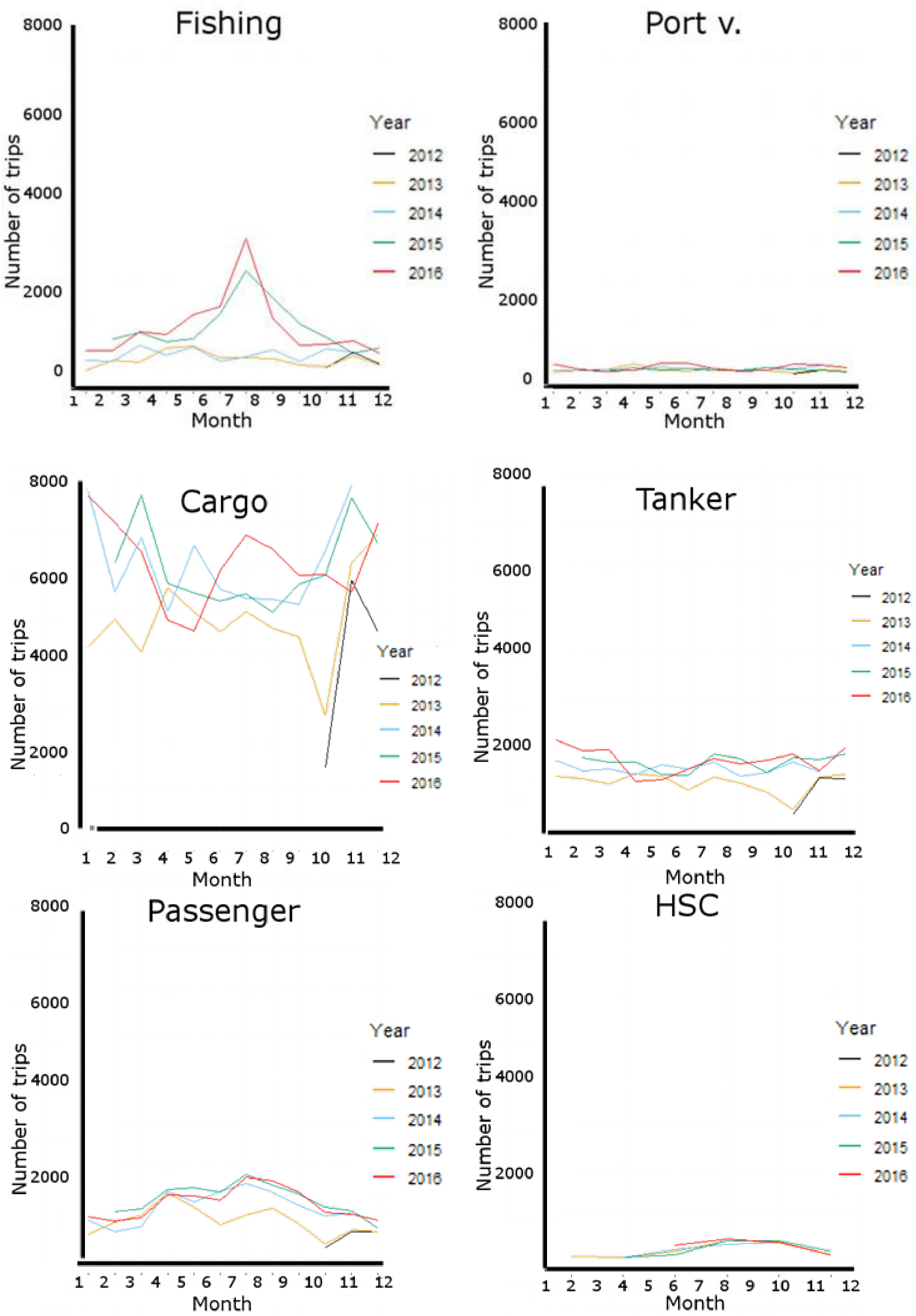


Figure 3. 6. Monthly variation in the number of trips per year for each vessel (October 2012 - December 2016).

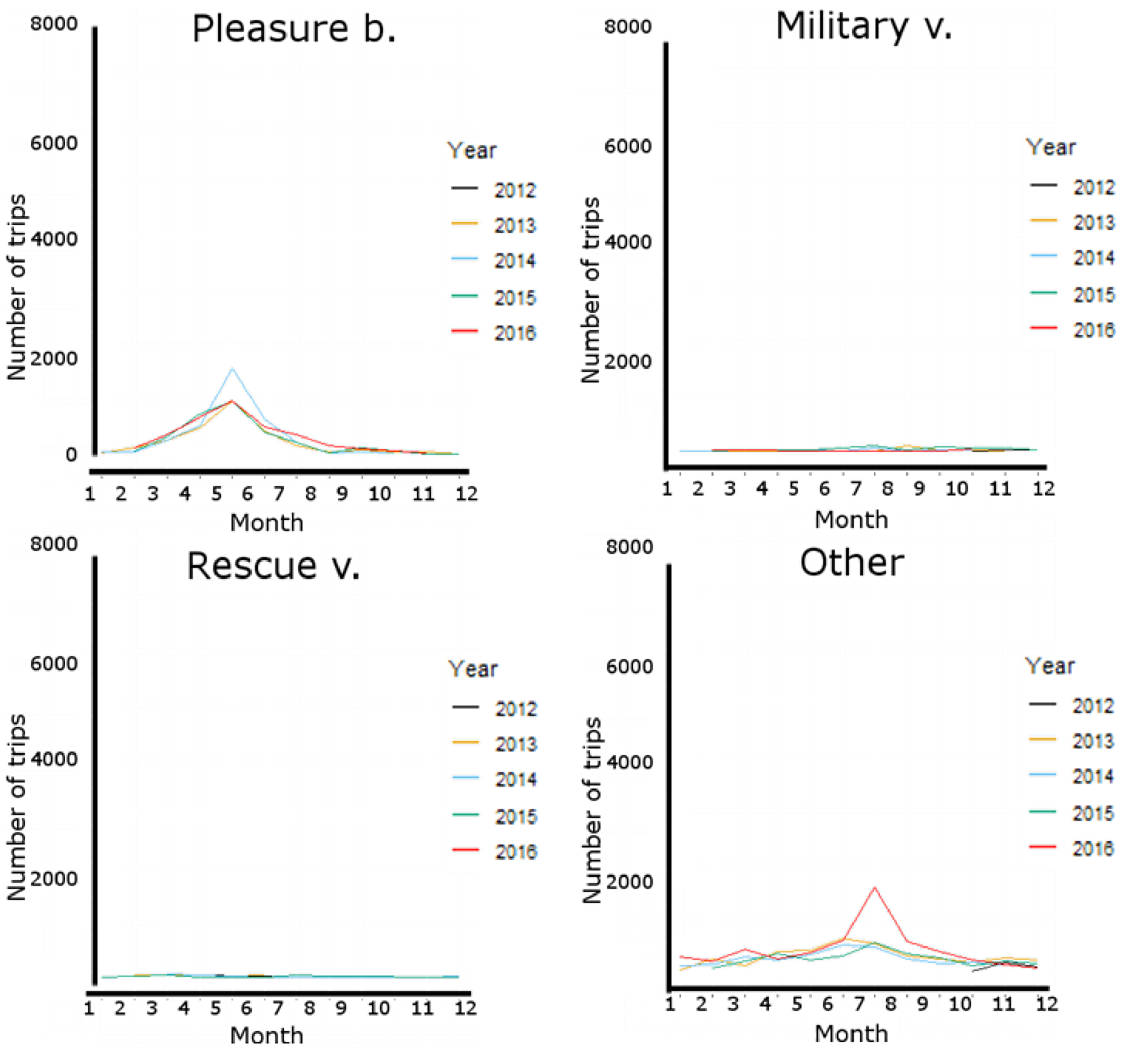


Figure 3. 6 (CONT). Monthly variation in the number of trips per year for each vessel (October 2012 -December 2016).

3.3.1.1. TOTAL SHIPPING DENSITY VS LETHAL PROBABILITY (LP) SHIPPING DENSITY

As expected, density was higher for total shipping compared to shipping of vessels >80 m length at speeds >14 kn, associated with higher probability of lethal collisions (Figure 3.7). Total shipping density reached the maximum values of 18 contacts/day per 25km^2 in areas connecting the islands 3-18 contacts/day per 25km^2 and intermediate values of 2-3 contacts/day per 25km^2 on the N and S routes in the archipelago. In comparison, the LP shipping density did not surpass three contacts/day per 25km^2, with the highest densities in between the islands. The standard deviation (SD) of shipping density was small suggesting little variability over time (Figure 3.7). The highest SD values occurred close to the ports (Horta, Ponta Delgada, Velas), and the intermediate values lied on the main traffic routes. Interestingly, there was higher temporal variability in LP shipping density than in the density of total traffic, with the highest variability E of Graciosa and of São Jorge, and intermediate SD values distributed throughout the study area.

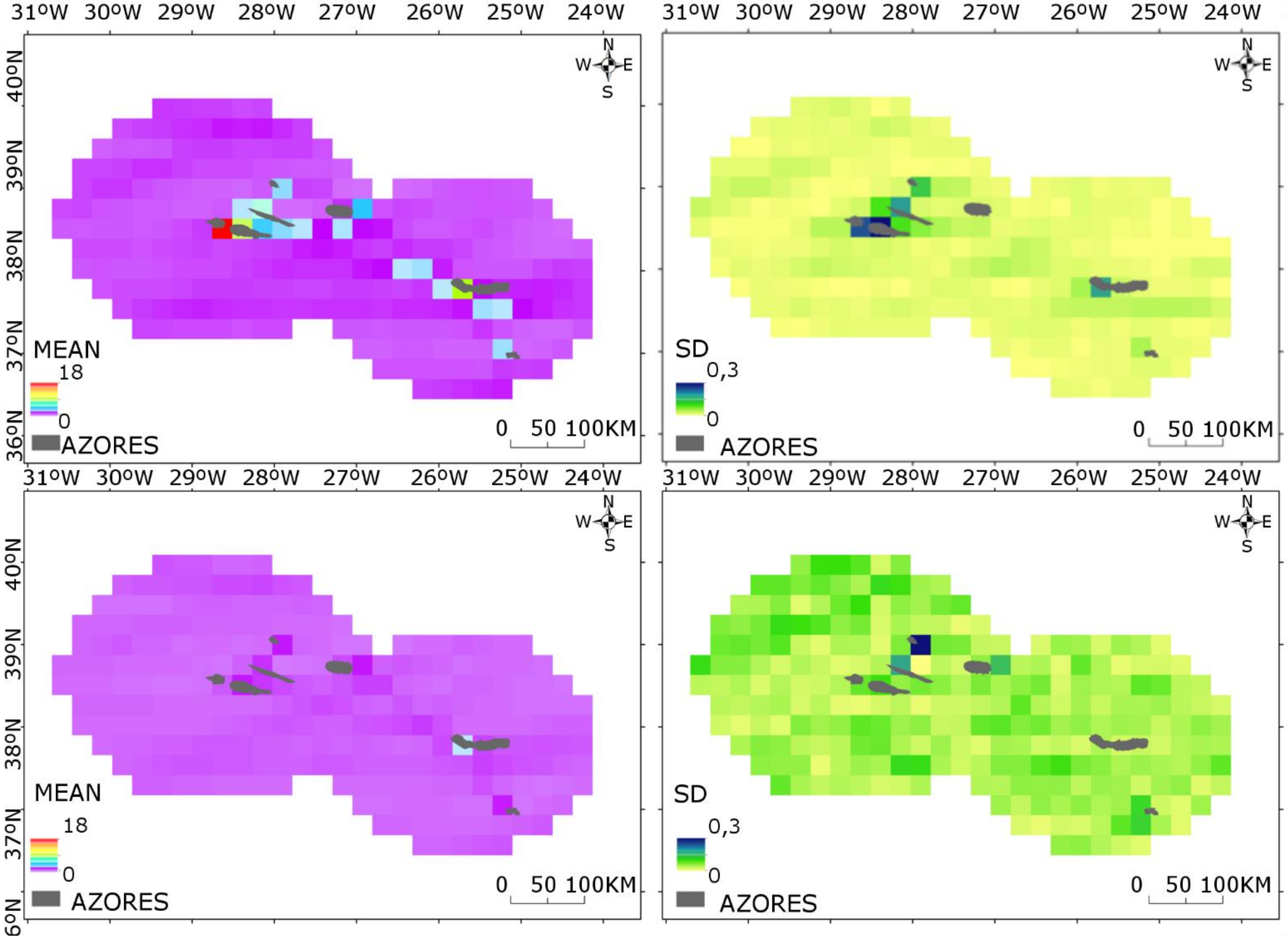


Figure 3. 7. Mean and standard deviation (SD) of daily shipping density for total (TOP) and LP shipping (bottom).

3.3.2. SPATIAL AND TEMPORAL PATTERNS IN SHIPPING DENSITY

Shipping density was scaled (0-1) to assess spatial and temporal patterns in shipping density, without the influence of other confounding factors. Mapping the mean daily shipping density for the entire study period (January 2013-December 2016; data from 2012 was excluded because it only covers autumn and winter) revealed four main shipping routes: one to the north of the archipelago, one on the south, a route linking S. Miguel to the central group of islands, and another one connecting the islands within the central group (Figure 3.8). The highest shipping density was around the main harbours and in the Faial-Pico channel. Distribution of shipping density was very similar among years (Appendix H), especially from 2014 to 2016, when the number of AIS records was similar.

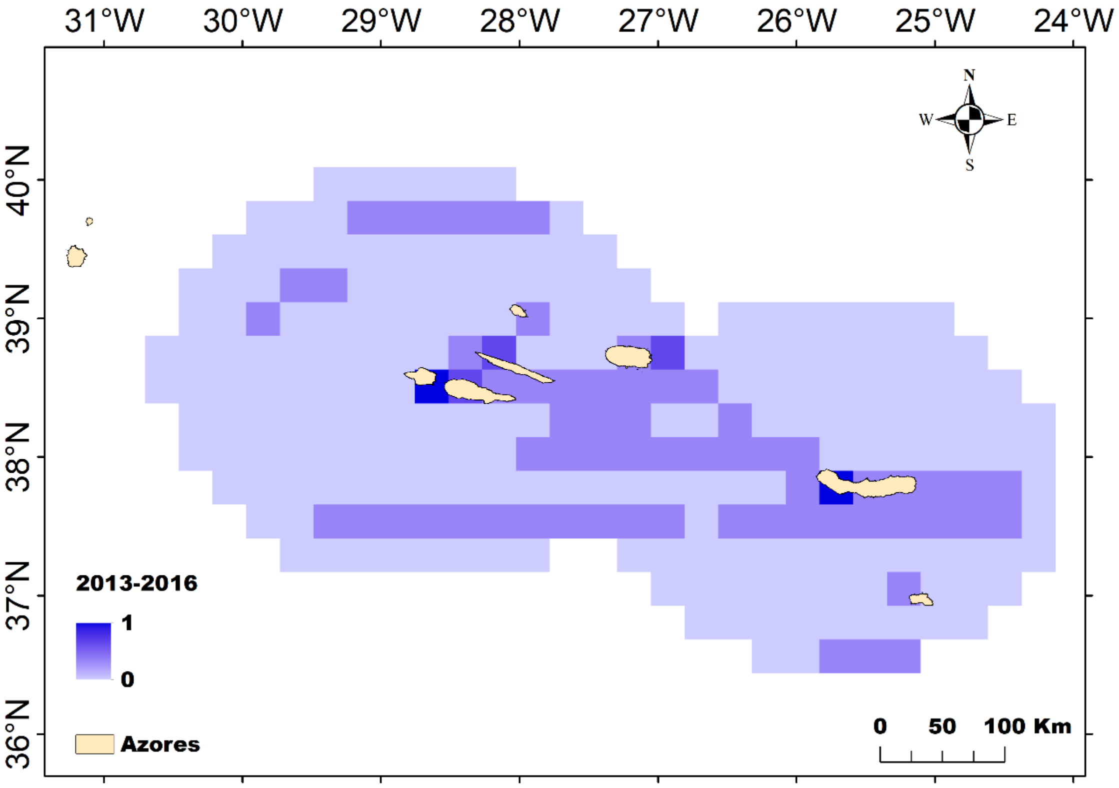


Figure 3. 8. Mean daily shipping density for the study period (January 2013 - December 2016).

The four main shipping routes were also evident in the monthly density maps (Figure 3.9). The main routes tended to be more prominent from November through March. The north route had the maximum values in June, while the connection between S. Miguel and the central group of islands reached maximum values in July and August, and the southern routes had maximum values in December and January. Despite the intra-annual spatial and temporal differences, peaks in density were consistently detected around the Faial Pico channel and the Ponta Delgada harbour in S. Miguel.

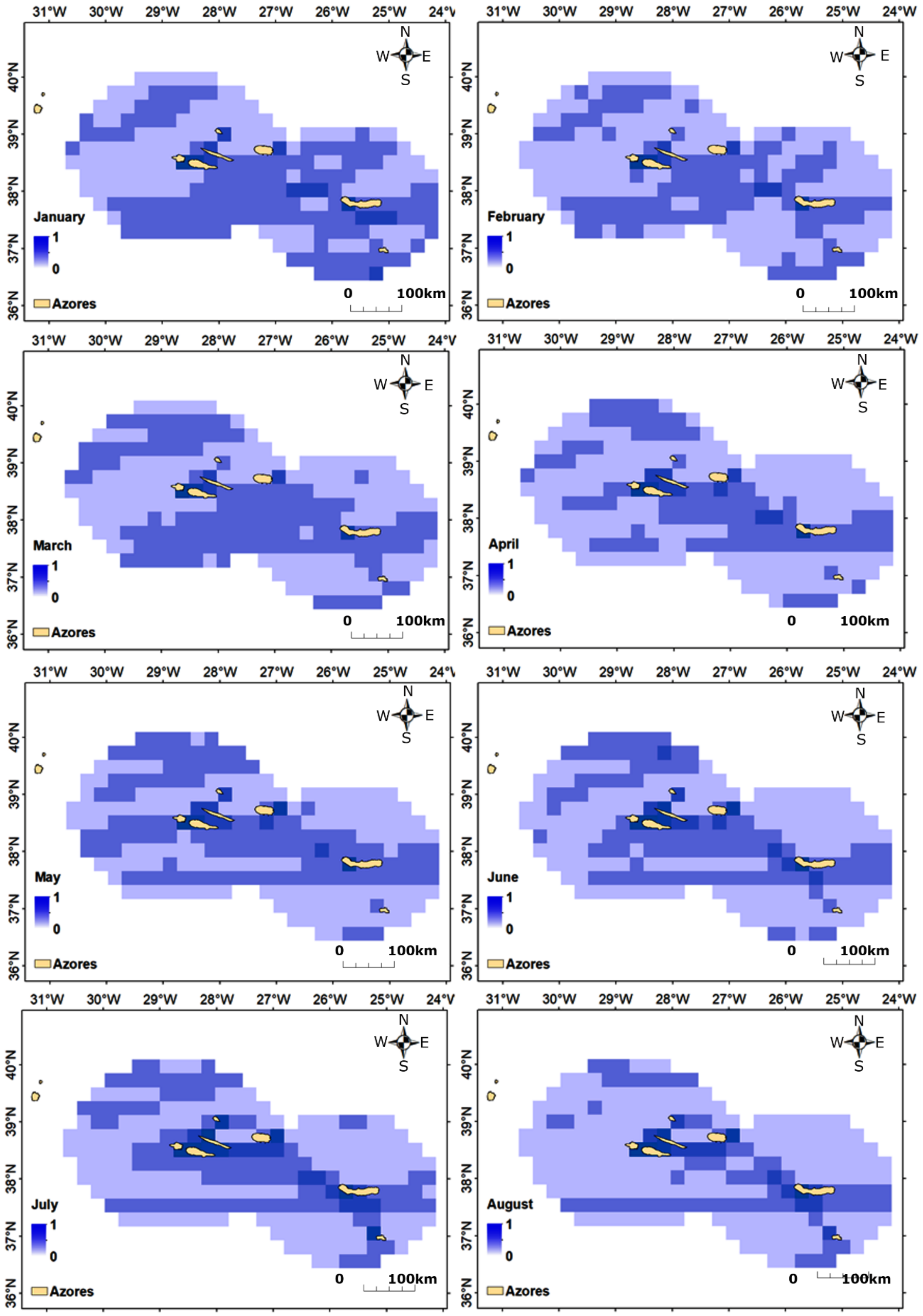


Figure 3. 9. Monthly patterns in mean daily shipping density (October 2012 – December 2016).

.

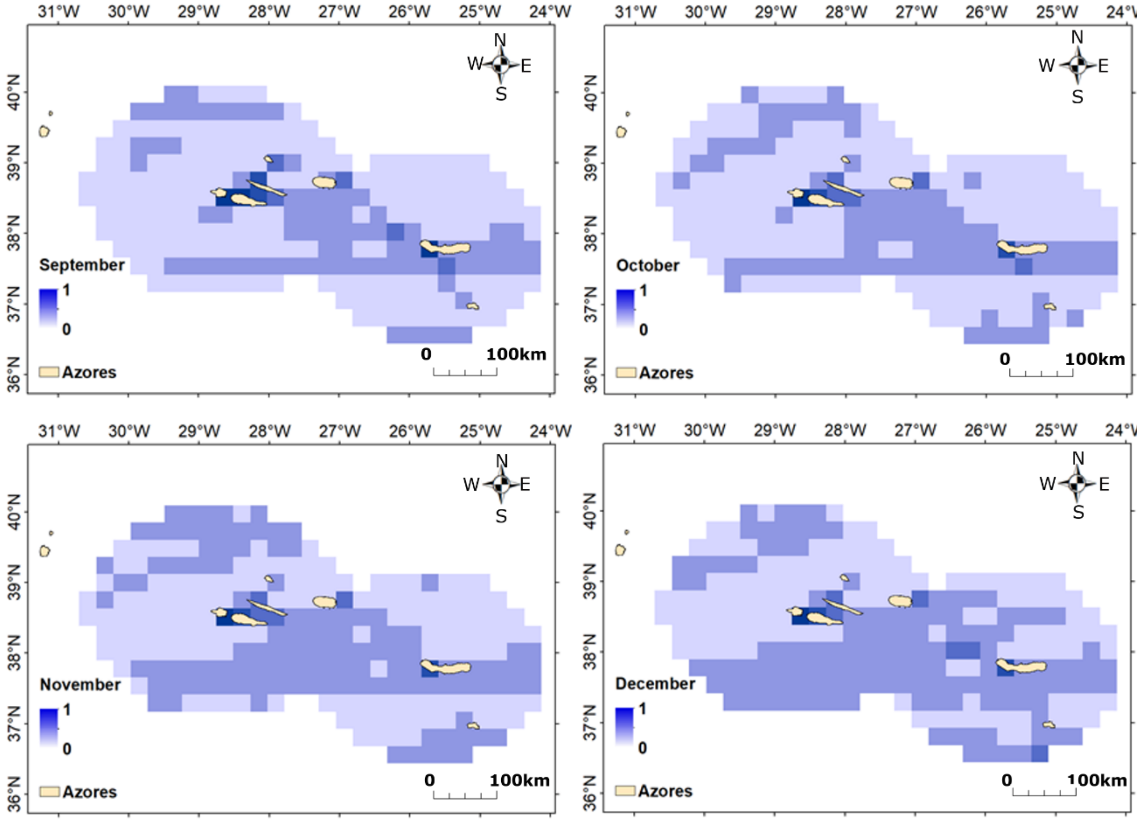


Figure 3. 9. (Continued). Monthly patterns in mean daily shipping density (October 2012 – December 2016).

Figure 3.10 presents maps of mean daily shipping density by vessel type. These maps use the same density scale, so the results are comparable. Most vessel types (HSC, port services ships, military, rescue, and other vessels) had lower densities in the whole area and traffic was concentrated mainly around the islands. Traffic of cargo and tankers was more spread out within the study area, although the highest densities occurred around the islands and along some of the main shipping routes. Interestingly, these routes were not clear in the maps of passenger vessels, which showed highest density around the islands. The pleasure boats had the highest values within the inter-island routes and SW of Faial. The fishing vessels were detected around the islands, seamounts and in a new route to the south of Terceira. Traffic from military vessels was mainly along a straight line from S. Miguel to the west. Although the vessel types displayed different spatial patterns, most had maximum density values located in some of the main ports of the Azores archipelago (Ponta Delgada in S. Miguel, Angra do Heroísmo in Terceira, Horta in Faial) except for military vessels.

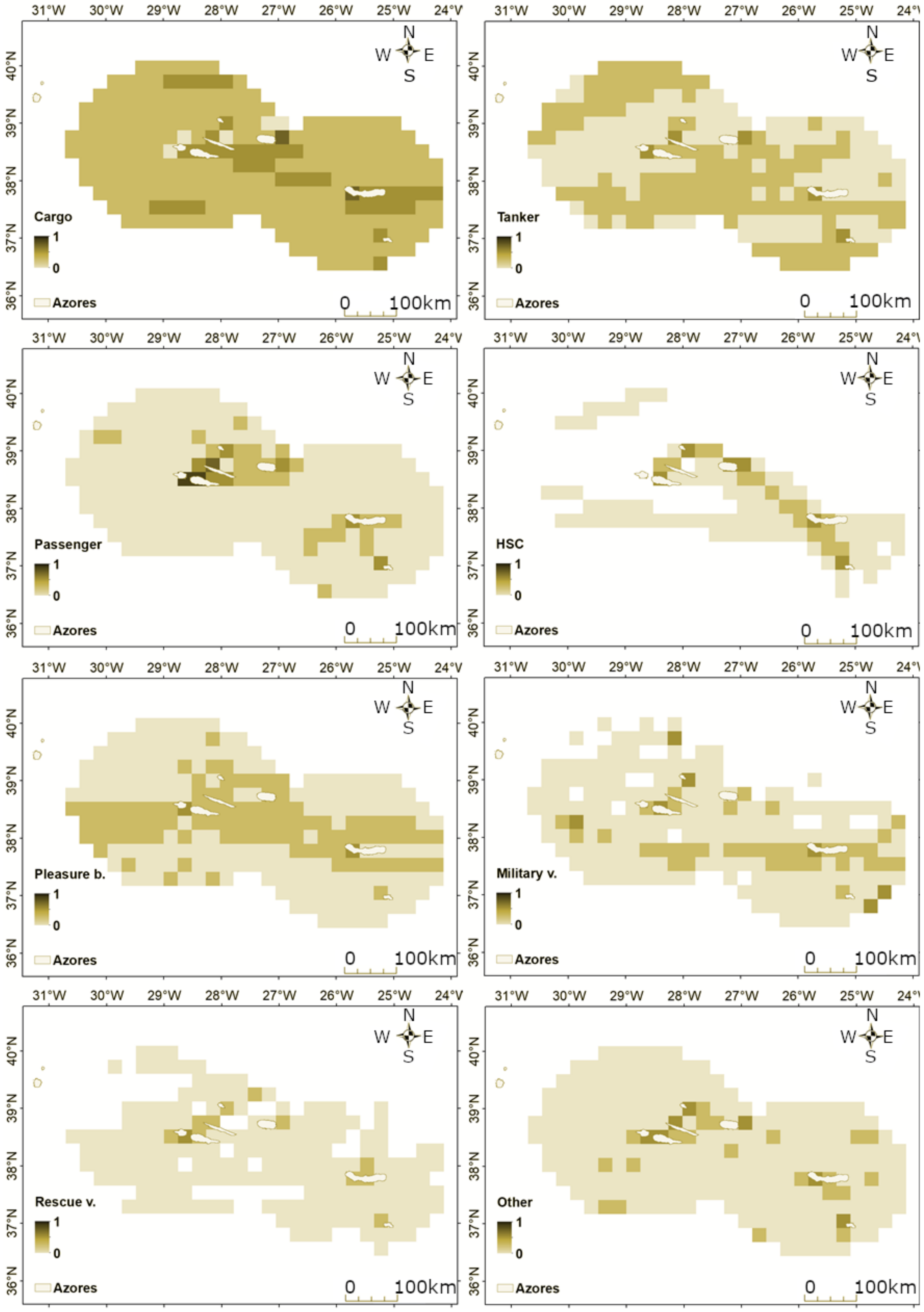


Figure 3. 10. Mean daily shipping density per vessel type (October 2012 – December 2016). The density scale is the same for all maps.

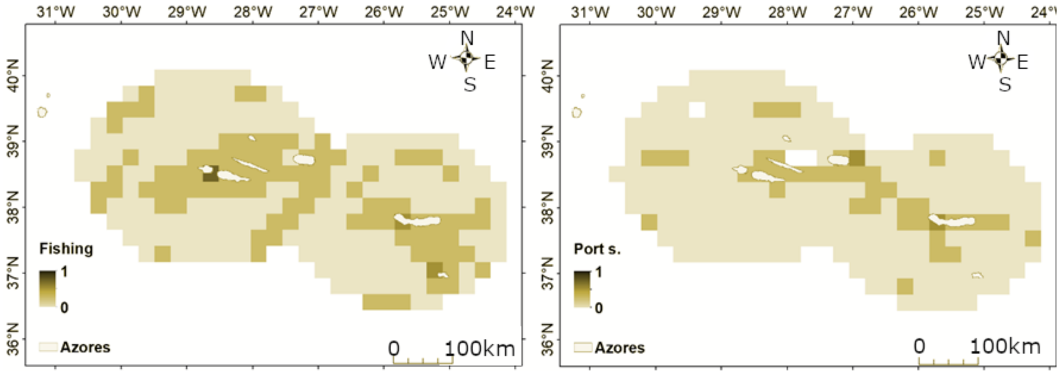


Figure 3. 10. (Continued). Mean daily shipping density per vessel type (October 2012 – December 2016). The density scale is the same for all maps.

Figure 3.11 shows the shipping density on the 2nd January 2013 and on the 1st July 2013, to illustrate differences in density between winter and summer days. Both maps showed the highest density in the Faial-Pico channel and around S. Miguel but density was lower, more scattered and with large gaps in the winter than in summer .

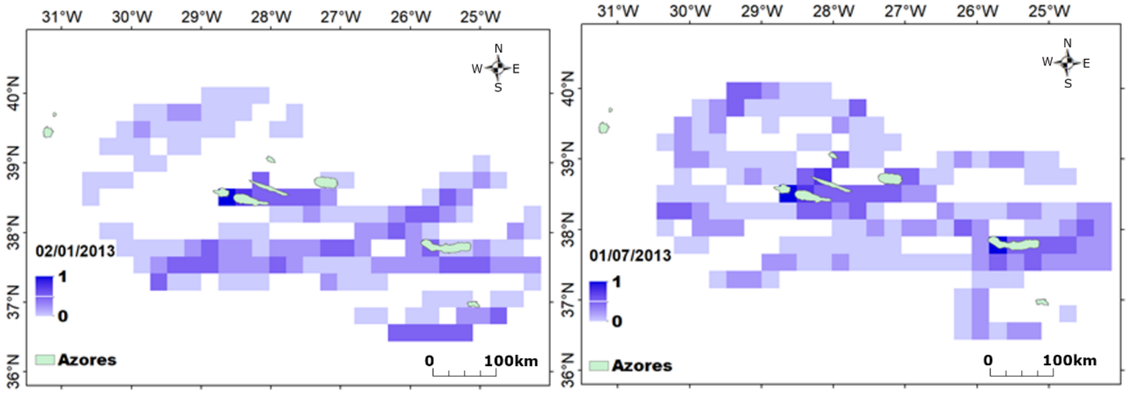


Figure 3. 11. Shipping density on 2nd January (LEFT) and 1st July (RIGHT of 2013)

3.3.3. SPATIAL AND TEMPORAL PATTERNS IN LP SHIPPING DENSITY

The two subsets of LP shipping categories (vessels >14 kn and vessels >14 kn and > 80 m length) shared almost identical records (94%) and the same vessels (85%), and showed indistinguishable shipping density patterns. Therefore, only maps for vessels >14 kn and >80 m in length are shown in this work.

The mean daily density map of LP shipping (Figure 3.12) was very similar to that of total shipping (Figure 3.8), showing the same traffic routes and increased density around the islands, particularly near the ports and on inter-islands routes. Similarly, intra-annual differences were minor and monthly patterns were essentially the same as those for total shipping (Figure 3.9). Thus, only annual maps are presented below (Figure 3.12). The year 2013 was the most similar to the mean daily shipping density. In 2014, the density between Faial and Flores was slightly higher, increasing in 2015 and 2016.

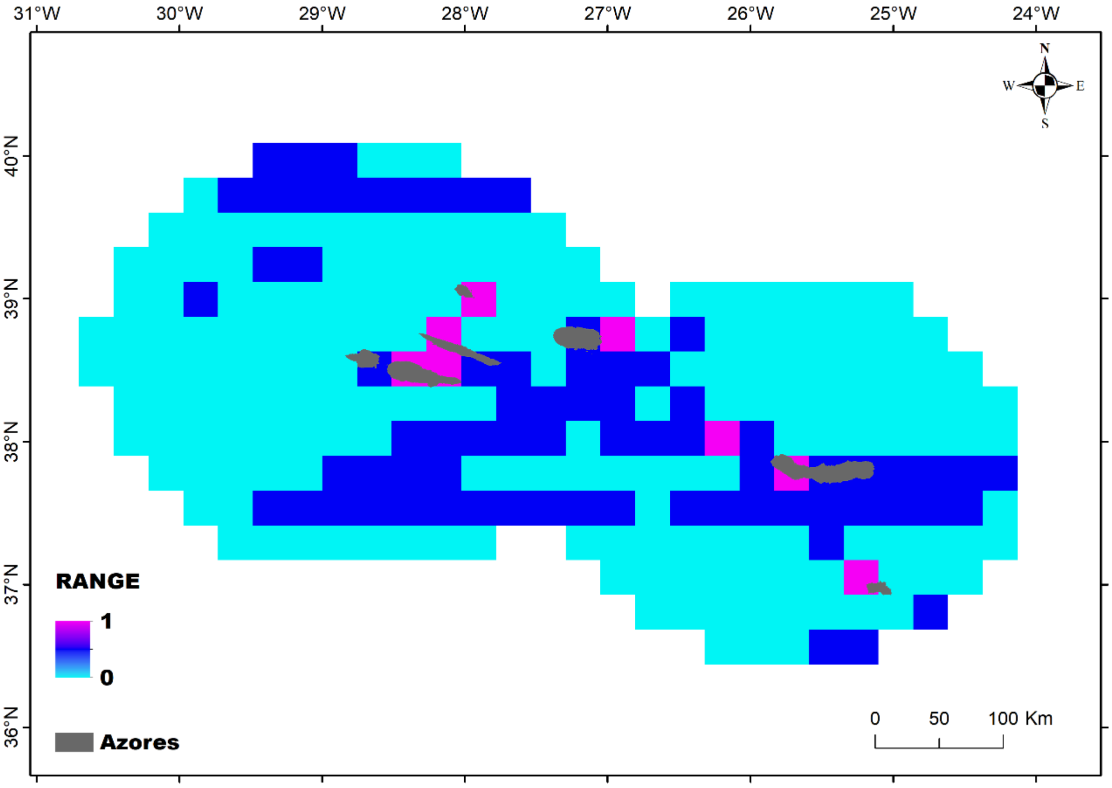


Figure 3. 12. Mean daily LP shipping density (speed >14 kn, length >80 m; January 2013 – December 2016).

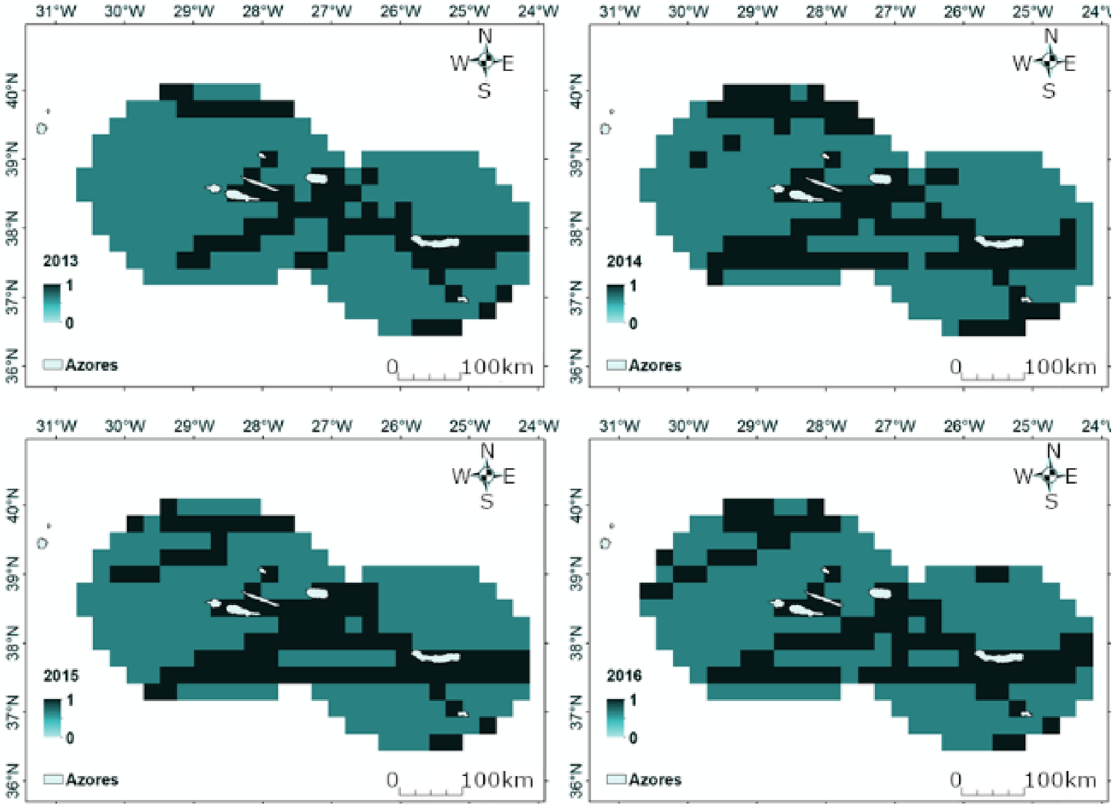


Figure 3. 13. Annual patterns in mean daily LP shipping density (speed >14 kn, length >80 m; January 2013 – December 2016).

The fishing and port services vessels showed speeds consistently lower than 14 kn, and therefore did not fit the criteria of vessels with high probability of lethal collisions with cetaceans (Figure 3.14). For most vessel types (e.g., tanker, pleasure boat, HSC, military and rescue vessels), there were huge areas where LP shipping density was null. LP shipping was highest for cargos, followed by tankers and passenger vessels and occurred throughout the study area, albeit more concentrated around the islands and along the main shipping routes. For HSC ships, LP shipping was restricted to a well defined corridor linking the eastern and western groups and the islands within these groups.

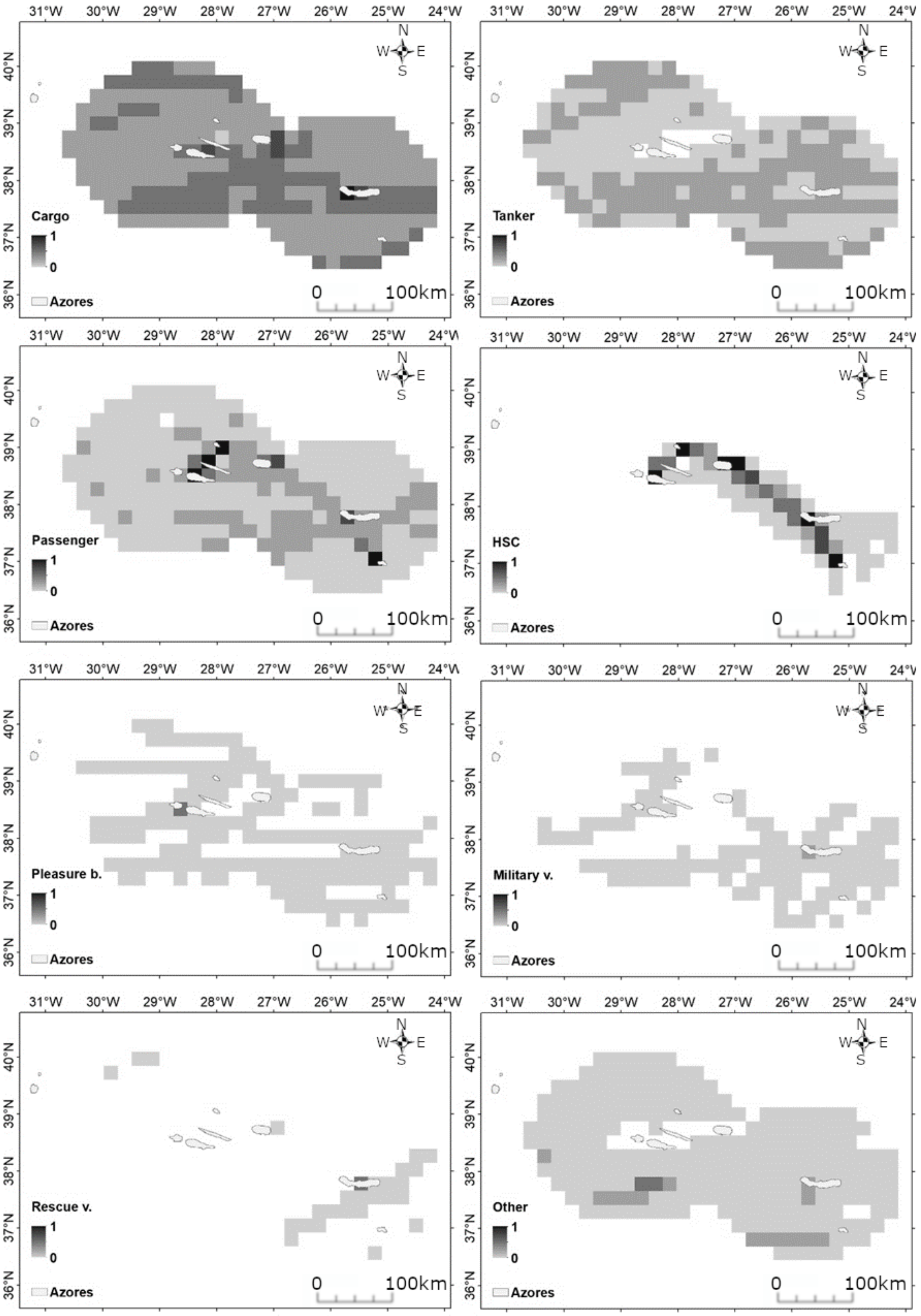


Figure 3. 14. Mean daily LP shipping density per vessel type (speed >14 kn, length >80 m; October 2012 – December 2016).

3.4. DISCUSSION

3.4.1. CAVEATS AND LIMITATIONS

Different approaches have been used to calculate shipping density from AIS data: a) construct vessel tracks from the consecutive AIS positions for each vessel, and sum total tracks (number or length in kilometres) [40, 41] within a grid square [42]; b) calculate how long a vessel takes to travel between two consecutive AIS locations, and calculate the time spent in each grid cell by all vessels [43-45]; and c) count the number of AIS messages within a grid cell [46]. In this study, the median interval between AIS records was 20 min. It could not be assumed that vessels navigated in a straight line between AIS records amidst the islands of the Azores archipelago. The chosen method adopted here was to calculate the shipping density using the number of AIS records by grid cell. This method avoids vessel tracking inferences and reduces bias from assigning vessel presences to incorrect grid cells. Nevertheless, this method is also more prone to errors resulting from varying time intervals between messages. To reduce this source of bias, AIS records with time intervals <20 min were removed. In addition, the shipping density was calculated per day to avoid differences in the number of days with AIS records per year.

It is also important to emphasise that the AIS dataset used in this study came only from tracking through AIS land stations, and only includes messages broadcasted within the radio range of the antennas. To minimise bias from variation in the probability of detecting AIS messages across the study area, the detection range of each receiving antenna was used to define the study area. I adopted a conservative approach by setting the study area as the area where the detection probability of each antenna equalled 1, and disregarded all messages transmitted from outside this area. As a consequence, the study area did not cover the western islands of the Azores (Corvo and Flores) and only covered a limited area to the north of Terceira and south of Sta. Maria islands. In spite of the conservative approach, given than detection range is also influenced by environmental conditions [47, 48] , some bias arising from differences in detectability is unavoidable, although at present I am unable to assess its severity.

Finally, calculations of shipping density presented here underestimate maritime traffic [49] in the area because not all vessels are equipped with AIS (see also Chapter 2). As a result of IMO requirements, all large vessels are fitted with AIS, but smaller vessels are outside the scope of these regulations. In addition, EU requires that all fishing vessels larger than 15 m long carry AIS but this has not been fully implemented in several places, including the Azores [31, 50]. In 2016, only 28 of the 600 Azorean fishing vessels were detected in this work. Although there is no obligation to use this equipment for pleasure boats, some ships have started using it for safety reasons. In addition, in 2016, this study detected 799 trips from class A AIS transponders whereas the official records indicated 4048 trips for the same year [51]. This shows that AIS equipment covered less than 5% of fishing boats and 25% of pleasure boats. Nonetheless, smaller vessels, such as pleasure boats, can reach high speeds on transit and potentially pose a high risk of collision with cetaceans [9, 37, 52, 53]. The characterisation of the vessel types is essential to interpret shipping patterns. The discrepancies between the number of fishing vessels detected and present in the fishing fleet show that this segment is highly underestimated and shipping patterns of fishing vessels may be biased. Future studies should attempt to complement AIS data with other data sources (e.g., VMS).

As expected, LP shipping density was lower than total shipping density. The characterization of the maritime traffic indicate that the majority of vessels that currently emit with AIS (class A) in the region are longer than 80 m. Therefore, as anticipated, AIS data better represents traffic of large vessels in the area. This also implies that vessel speed was the main driver of differences in total and LP shipping density in Azorean waters.

3.4.1.1. SHIPPING DENSITY

Despite the shortcomings mentioned above, to the best of my knowledge, this study provides the first characterization of shipping density for the Azores archipelago, allowing the identification of important shipping routes and traffic hotspots. Not surprisingly, two of these routes and some of the hotspots are related to inter-island connections (from commercial ships like cargo and tankers; from passenger vessels and pleasure boats) and to port approaches (e.g., the Faial-Pico channel, the area around Ponta Delgada, in S. Miguel, and NE of Terceira). Although to a lesser extent, the rescue and others vessel types also contribute to the inter-island routes and traffic around the ports. Another route was located to the north of the archipelago and the other crossed the archipelago just south of S. Miguel. These routes reflected the traffic of cargo ships and tankers, which together accounted for the large majority of vessels, trips and, consequently, messages. However, the activity of passenger ships also contributed to explain these spatial patterns, especially in late spring and summer (see below). Although fishing vessels showed slightly different patterns, they matched partially with the main traffic routes.

The main routes and density hotspots persisted over all months and years of the study. In general, shipping patterns varied more within months than between the years. Although the main routes were apparent in all monthly maps, density along those routes was higher from April to September, than in the remainder of the year. Overall, the highest shipping density was found in the Faial-Pico channel, followed by the area offshore Ponta Delgada and NE Terceira. The intense traffic between Faial, Pico and S. Jorge can primarily be attributed to the daily ferry connections, which became more frequent in spring and summer (April to August), but also to the movement of fishing vessels in and out of the harbours in Horta (Faial) and Madalena (Pico). The shipping density between S. Miguel and the central group of islands (Terceira, Graciosa, Faial, Pico and S. Jorge) was also higher in July and August than in the rest of the year. Unexpectedly, this increase in density during the summer appears to be more related to higher traffic of the HSC, cargo and pleasure boats, than to the weekly inter-island ferries.

3.4.1.2. SHIPPING DENSITY WITH HIGH PROBABILITY OF LETHAL COLLISIONS WITH CETACEANS

The physical encounter between a moving vessel and a cetacean may have very different impacts on the animal, depending on the vessel characteristics and operations [9, 54]. In particular, vessel size (i.e., length) and speed are positively associated with the increased probability of a collision resulting in physical injuries on the animals, resulting in mortality [9, 35]. These factors were used to filter the AIS records in this study to describe shipping patterns associated with a higher probability of lethal collisions and to predict the areas where risk of these lethal encounters is higher (Chapters 4 and 5).

In this study, both categories of risk factors (speed>14 kn, or speed>14 kn and length>80 m) returned nearly the same patterns. Vessels that are mostly cruising through the area or between ports (merchant and passengers ships) are associated with higher risk of lethal collisions, because they are large and maintain constant high speeds. Cargo and tankers generally exceeded the speed and length thresholds and were responsible for much of the traffic associated with increased likelihood of lethal collisions. Even though there were fewer passenger and HSC vessels in the area, they were also relevant for LP density because they make more trips. Traffic of pleasure boats, military, rescue and other vessels longer than 80 m and faster than 14kn was low and areas where potential collisions are likely to be fatal were scattered within the study area. Fishing boats and service port boats were usually smaller than 80m and thus did not fit with the criteria of LP traffic. However, as explained above, only approximately 5% of the fishing vessels registered in the Azores were detected in the AIS records. Some fishing vessels attain speeds >14 kn, implying that LP shipping density may be underestimated and the spatio-temporal patterns slightly biased. Moreover, fishing vessels can threaten cetaceans in other ways, namely through entanglement in fishing gear [55], and their distribution should be carefully monitored.

Distribution of shipping traffic associated with high probability of lethal collisions with cetaceans followed similar spatial and temporal patterns to that of overall traffic. The areas in the vicinity of the ports, two inter-island routes, and the northern and southern routes, were the areas where collisions with cargos and tankers have higher probability of being lethal. Cargos, HSC and passenger vessels were also responsible for the high LP shipping traffic on the inter-island routes. In general, vessels slow down when approaching ports; it is evident that the spatial resolution of this analysis did not capture this variation in manoeuvres. The shipping density patterns shows that many vessels with the highest probability of a lethal collision approach the islands. These hotspots are magnified by the pleasure boats. Even though tankers are highly relevant in the area, they only had a minor contribution to LP traffic, and restricted to Ponta Delgada port in

S. Miguel. The activity of tankers at high-speed was mainly observed when these vessels crossed the archipelago along the main routes. The passenger vessels, pleasure boats, and HSC depend on tourism demand, sailing events and weather conditions, and their traffic increased in spring and summer. The two main LP traffic routes crossing the archipelago were located north and south of S. Miguel. Both routes reflected the traffic of cargo ships and tankers, and to a lower extent pleasure boats. However, passenger, military, and other vessel types also contributed to LP density of the south route, especially passenger vessels in late spring and summer (see below).

The spatial pattern of traffic associated with a high probability of lethal collisions with cetaceans is maintained inter-annually, although with some variations. The main routes had higher shipping density from April to September, associated with increasing traffic of goods and passengers in the region, namely around S. Miguel. The heavy traffic between Faial, Pico and S. Jorge is related to the daily ferry connections, increasing their regular travels in spring and summer. The shipping density between S. Miguel and the central group of islands was also higher in July and August. Unexpectedly, this increase in density during the summer appears to be more related to higher traffic of the HSC, cargo and pleasure boats than the weekly inter-island ferries.

The influence of some variables on maritime traffic is evident in this work. The most touristic season and the good weather implied a greater influx of passenger vessels, HSC and pleasure boats. However, even something so easily predictable and constant can be altered by political decisions, such as the one made in 2021 to replace most inter-island ferry transport with cheap flights in the Azores archipelago. Therefore, care should be taken when attempting to extrapolate maritime traffic patterns beyond the studied period.

3.5. CONCLUSIONS

This chapter presents shipping patterns within the Azores archipelago, based on AIS class A records received at land stations from October of 2012 to December of 2016. In addition, using criteria known to affect severity of potential vessel-cetacean collisions, I characterized the shipping activity with a greater probability of causing cetacean mortality. Our results show that cargo ships were responsible for the majority of shipping traffic during the study period, followed by tankers. Fishing vessels and passenger vessels, albeit less, made frequent trips, thereby contributing to spatial and temporal patterns of shipping density. I have identified four main shipping routes within the study area: one to the north of the archipelago, one on the south, a route linking S. Miguel to the central group of islands, and another one connecting the islands within the central group. Not surprisingly, shipping density was always higher around the harbours, especially in Faial, S. Miguel and Terceira. These were also the areas where potential collisions with cetaceans have a high probability of resulting in the death of the animals.

Global marine traffic is expected to continue increasing as a result of the growing human population and increasing demand for goods and services. Additionally, patterns of global shipping might change in the future as a consequence of environmental and economical shifts, including the melting of the ice [56, 57] in the Arctic and Antarctica, the emergence of HSC [58] and trends in the structure of the international maritime market [57]. Shipping traffic is also expected to increase locally, with the implementation of the Azores Oceanic Hub project (https://www.azoresoceanichub.com/) that foresees the construction of new container and cruise terminals, and a Liquefied Natural Gas bunker station. Therefore, shipping patterns presented here might change profoundly in the near future and should be monitored closely.

3.6. REFERENCES

1. Eurostat. International trade in services (since 2010) Eurostat: © European Union, 1995-2021; 2009 [cited 2021 6 April 2021]. Available from: https://ec.europa.eu/eurostat/databrowser/view/bop_its6_det/default/table?lang=en.
2. UNCTAD. Review of Maritime Transport. 2021 eISSN: 2225-3459.
3. Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, et al. Impacts of biodiversity loss on ocean ecosystem services. Science. 2006;314(5800):787-90.
4. Conn PB, Silber GK. Vessel speed restrictions reduce risk of collision-related mortality for North Atlantic right whales. Ecosphere. 2013;4(4):16. doi: 10.1890/es13-00004.1.
5. MacGillivray AO, Li Z, Hannay DE, Trounce KB, Robinson OM. Slowing deep-sea commercial vessels reduces underwater radiated noise. Acoust Soc Am. 2019;146(1):340. Epub 2019/08/03. doi: 10.1121/1.5116140. PubMed PMID: 31370655.
6. Croll DA, Clark CW, Calambokidis J, Ellison WT, Tershy BR, editors. Effect of anthropogenic low-frequency noise on the foraging ecology of Balaenoptera whales. Animal Conservation forum; 2001: Cambridge University Press.
7. Fagerholt K, Gausel NT, Rakke JG, Psaraftis HN. Maritime routing and speed optimization with emission control areas. Transportation Research 2015;52:57-73. doi: 10.1016/j.trc.2014.12.010.
8. Kotta J, Nurkse K, Puntila R, Ojaveer H. Shipping and natural environmental conditions determine the distribution of the invasive non-indigenous round goby Neogobius melanostomus in a regional sea. Estuarine, Coastal and Shelf Science. 2016;169:15-24. doi: 10.1016/j.ecss.2015.11.029.
9. Laist DW, Knowlton AR, Mead JG, Collet AS, Podesta M. Collisions between ships and whales. Marine Mammal Science. 2001;17(1):35-75. doi: https://doi.org/10.1111/j.1748-7692.2001.tb00980.x.
10. Knowlton A, Kraus S. Mortality and serious injury of northern right whales (Eubalaena glacialis) in the western North Atlantic Ocean. Journal of Cetacean Resource Management. 2001;2:193-208.
11. The IWC Global Ship Strikes Database [Internet]. International Whaling Commission. 2007. Available from: https://iwc.int/ship-strikes.

12. Peel D, Smith JN, Childerhouse S. Vessel strike of whales in Australia: the challenges of analysis of historical incident data. Frontiers in Marine Science. 2018;5:69. doi: 10.3389/fmars.2018.00069.

13. Panigada S, Pesante G, Zanardelli M, Capoulade F, Gannier A, Weinrich MT. Mediterranean fin whales at risk from fatal ship strikes. Marine pollution bulletin. 2006;52(10):1287-98. Epub 2006/05/23. doi: 10.1016/j.marpolbul.2006.03.014. PubMed PMID: 16712877.

14. IMO. Recommendation on performance standards for an universal shipborne automatic identification system (AIS). International Maritime Organization, 1998.

15. Eriksen T, Høye G, Narheim B, Jensløkken Meland B. Maritime traffic monitoring using a space-based AIS receiver. Acta Astronautica. 2006;58:537–49.

16. Fournier M, Hilliard RC, Rezaee S, Pelot R. Past, present, and future of the satellite-based automatic identification system: areas of applications (2004–2016). WMU journal of maritime affairs. 2018;17(3):311-45. doi: https://doi.org/10.1007/s13437-018-0151-6.

17. Renner M, Kuletz KJ. A spatial–seasonal analysis of the oiling risk from shipping traffic to seabirds in the Aleutian Archipelago. Marine pollution bulletin. 2015;101(1):127-36. doi: http://dx.doi.org/10.1016/j.marpolbul.2015.11.007.

18. Greig NC, Hines EM, Cope S, Liu X. Using Satellite AIS to Analyze Vessel Speeds Off the Coast of Washington State, U.S., as a Risk Analysis for Cetacean-Vessel Collisions. Frontiers in Marine Science. 2020;7(109). doi: 10.3389/fmars.2020.00109.

19. Nichol LM, Wright BM, Hara PO, Ford JK. Risk of lethal vessel strikes to humpback and fin whales off the west coast of Vancouver Island, Canada. Endangered Species Research. 2017;32:373-90. doi: https://doi.org/10.3354/esr00813.

20. Cates K, DeMaster D, Brownell R, Silber G, Gende S, Leaper R. Strategic Plan to Mitigate the Impacts of Ship Strikes on Cetacean Populations: 2017-2020. IWC 2017;IWC Strategic Plan to Mitigate Ship Strikes.

21. Constantine R, Johnson M, Riekkola L, Jervis S, Kozmian-Ledward L, Dennis T, et al. Mitigation of vessel-strike mortality of endangered Bryde’s whales in the Hauraki Gulf, New Zealand. Biological Conservation. 2015;186:149-57. doi: http://dx.doi.org/10.1016/j.biocon.2015.03.008.

22. Açores EdAd. Transporte marítimo. Direção Regional do Ambiente e Alterações Climáticas: Governo dos Açores, 2017.

23. Silva MA, Prieto R, Cascão I, Seabra MI, Machete M, Baumgartner MF, et al. Spatial and temporal distribution of cetaceans in the mid-Atlantic waters around the Azores. Marine Biology Research. 2014;10(2):123-37. doi: https://doi.org/10.1080/17451000.2013.793814.

24. PORTUGAL National Report UNESCO Man & Biosphere Programme, 2018.

25. SIARAM. Zonas húmidas Portal do Governo dos Açores: Governo dos Açores; 2021 [cited 2021 3 March 2021]. Available from: http://siaram.azores.gov.pt/vegetacao/zonas-humidas/_intro.html.

26. EIONET. Macaronesian Region: European Enviroment Agency; 2012 [cited 2021 20/feb/2021]. Available from: https://www.eea.europa.eu/data-and-maps/figures/natura-2000-sites-biogeographical-regions-4/macaronesian_3-maps_mid2009.eps.

27. DRAM. Conveção Ospar: Governo dos Açores; 2004 [12/feb/2021]. Available from: http://www.azores.gov.pt/Gra/Lixo+Marinho+-+DRAM/conteudos/livres/Convencao_OSPAR.htm.
28. Commission E. European Atlas of the Seas. In: Affairs M, editor.: EMODnet; 2021.
29. Açores Pd. Projetos: Porto dos Açores; 2021. AcoresPRO:[Available from: https://portosdosacores.pt/projetos/.
30. ITU-R RS. Recomendation M.2092. 2015.
31. Eriksen T, Olsen Ø. Vessel tracking using automatic identification system data in the Arctic. Sustainable Shipping in a Changing Arctic. 2018;7(WMU Studies in Maritime Affairs):115-36. doi: https://doi.org/10.1016/j.marpol.2018.03.028.
32. Ghasemi A, Abedi A and Ghasemi F. Propagation Engineering in Wireless Communication. Second ed. Springer, editor2012.
33. Gonçalo V. Macaronésia Automatic Identification System Porto dos Açores: Programa de Iniciativa Comunitária INTERREG III B 2000-2006 AÇORES-MADEIRA-CANÁRIAS; 2006 [cited 2013 22 July 2013]. Available from: http://www.macais.org/omacais_es.asp.
34. Le Tixerant M, Le Guyader D, Gourmelon F, Queffelec B. How can Automatic Identification System (AIS) data be used for maritime spatial planning? Ocean Coastal Management. 2018;166:18-30.
35. Jensen AS, Silber GK. Large whale ship strike database. U.S. Department of Commerce, NOAA 2003. p. 37.
36. Fais A, Lewis TP, Zitterbart DP, Álvarez O, Tejedor A, Aguilar Soto N. Abundance and Distribution of Sperm Whales in the Canary Islands: Can Sperm Whales in the Archipelago Sustain the Current Level of Ship-Strike Mortalities? PloS one. 2016;11(3):e0150660. Epub 2016/03/22. doi: 10.1371/journal.pone.0150660. PubMed PMID: 26999791; PubMed Central PMCID: PMC4801403.
37. Carrillo M, Ritter F. Increasing numbers of ship strikes in the Canary Islands: proposals for immediate action to reduce risk of vessel-whale collisions. Journal of Cetacean Research and Management. 2010;11(2):131-8.
38. Eurostat, cartographer Transport networks: © European Union, 1995-2021; 2009.
39. Rockwood RC, Calambokidis J, Jahncke J. High mortality of blue, humpback and fin whales from modeling of vessel collisions on the US West Coast suggests population impacts and insufficient protection. PloS one. 2017;12(8):e0183052. doi: https://doi.org/10.1371/journal.pone.0183052.
40. Ferrà C, Tassetti AN, Armelloni EN, Galdelli A, Scarcella G, Fabi G. Using AIS to attempt a quantitative evaluation of unobserved trawling activity in the Mediterranean Sea. Frontiers in Marine Science. 2020;7:1036. doi: https://doi.org/10.1016/j.marpol.2017.12.013.
41. Pizzolato L, Howell SE, Dawson J, Laliberté F, Copland L. The influence of declining sea ice on shipping activity in the Canadian Arctic. Geophysical Research Letters. 2016;43(23):12,146-12,54. doi: 10.1002/2016GL071489.
42. David L, Alleaume S, Guinet C. Evaluation of the potential of collision between fin whales and maritime traffic in the north-western Mediterranean Sea in summer, and mitigation solutions. Journal of Marine Animals Their Ecology Vol. 2011;4(1).
43. Shepperson JL, Hintzen NT, Szostek CL, Bell E, Murray LG, Kaiser MJ. A comparison of VMS and AIS data: The effect of data coverage and vessel position

recording frequency on estimates of fishing footprints. ICES Journal of Marine Science. 2018;75(3):988-98. doi: https://doi.org/10.1093/icesjms/fsx230.

44. Wu L, Xu Y, Wang Q, Wang F, Xu Z. Mapping global shipping density from AIS data. The Journal of Navigation. 2017;70(1):67-81. doi: https://doi.org/10.1017/S0373463316000345.

45. Coello J, Williams I, Hudson DA, Kemp S. An AIS-based approach to calculate atmospheric emissions from the UK fishing fleet. Atmospheric Environment. 2015;114:1-7. doi: https://doi.org/10.1016/j.atmosenv.2015.05.011.

46. Cazzanti L, Pallotta G, editors. Mining maritime vessel traffic: Promises, challenges, techniques. OCEANS 2015-Genova; 2015: IEEE.

47. Wright D, Janzen C, Bochenek R, Austin J, Page EJFiMS. Marine Observing Applications Using AIS: Automatic Identification System. 2019;6:537.

48. Metcalfe K, Bréheret N, Chauvet E, Collins T, Curran BK, Parnell RJ, et al. Using satellite AIS to improve our understanding of shipping and fill gaps in ocean observation data to support marine spatial planning. Journal of Applied Ecology. 2018;55(4):1834-45. doi: 10.1111/1365-2664.13139.

49. Smith JN, Kelly N, Childerhouse S, Redfern JV, Moore TJ, Peel D. Quantifying ship strike risk to breeding whales in a multiple-use marine park: the Great Barrier Reef. Frontiers in Marine Science. 2020;7:67. doi: 10.3389/fmars.2020.00067.

50. Hermannsen L, Mikkelsen L, Tougaard J, Beedholm K, Johnson M, Madsen PT. Recreational vessels without Automatic Identification System (AIS) dominate anthropogenic noise contributions to a shallow water soundscape. Scientific reports. 2019;9(1):1-10. doi: https://doi.org/10.1038/s41598-019-51222-9.

51. Embarcações de Recreio [Internet]. Serviço Regional de Estatística dos Açores. 2021. Available from: https://srea.azores.gov.pt/ReportServer/Pages/ReportViewer.aspx?%2fRelatoriosVarios%2f15.TransporteEmbarcacoesRecreio&rs:Command=Render.

52. Di Sciara GN, Birkun A. Conserving whales, dolphins and porpoises in the Mediterranean and Black Seas. 2010.

53. Palomo LE, Hernández-Flores A. Integrating a spatial model and decision theory towards optimal boating density and carrying capacity in a recreational fishery. Marine Policy. 2020;112:103740. doi: https://doi.org/10.1016/j.marpol.2019.103740.

54. Ritter F. Collisions of sailing vessels with cetaceans worldwide: First insights into a seemingly growing problem. Journal of Cetacean Research Management. 2012;12(1):119-27.

55. Tulloch V, Pirotta V, Grech A, Crocetti S, Double M, How J, et al. Long-term trends and a risk analysis of cetacean entanglements and bycatch in fisheries gear in Australian waters. 2020;29(1):251-82. doi: https://doi.org/10.1007/s10531-019-01881-x.

56. Lasserre F. Arctic shipping: A contrasted expansion of a largely destinational market. In: Finger M, Heininen L, editors. The GlobalArctic Handbook. First ed: Springer; 2019. p. 83-100.

57. Chen J-L, Kang S-C, Guo J-M, Xu M, Zhang Z-M. Variation of sea ice and perspectives of the Northwest Passage in the Arctic Ocean. Advances in Climate Change Research. 2021;12(4):447-55. doi: https://doi.org/10.1016/j.accre.2021.02.002.

58. Kartoğlu C, Kum S. The Place of High Speed Crafts (HSCs) in Maritime Transportation. In: Akyüz E, editor. Handbook of Research on the Applications of

International Transportation and Logistics for World Trade. Turkey: IGI Global; 2020. p. 258-87.

CHAPTER 4

4. ENVIRONMENTAL AND PREY INFLUENCES IN THE CETACEAN DISTRIBUTION

4.1. INTRODUCTION

Species distribution models (SDM) are numerical tools that combine observations of species occurrence or abundance with environmental estimates, enabling to predict species distribution/abundance across landscapes [1]. SDMs have become a fundamental tool in both ecological and evolutionary studies, enabling to document spatio-temporal patterns of single species or species assemblages, species interactions, habitat use, community structure, and assess or predict consequences of natural or anthropogenic environmental changes, among others [2, 3]. The performance, and therefore usefulness, of SDMs are strongly influenced by methodological factors, especially the sample size, modelling method [4], as well as the selection of explanatory variables [5].

Until very recently, most efforts to predict cetacean distribution have only considered cetacean occurrence data and distal environmental predictors, such as bathymetry or oceanographic variables [6-8]. As such, previous studies modelling the presence and distribution of cetaceans in the Azores archipelago relied on traditional environmental variables [9-13]. However, prey variables, such as distribution, density or biomass, should be more important at explaining cetacean distribution [14-16]. Therefore, these variables are instrumental in species models (e.g. Pendleton et al. 2012 [17]).

Synchronising the sampling of cetaceans with their potential food is very challenging, explaining why prey data is not usually integrated into cetacean models. Some attempts to obtain concurrent data on cetaceans and their potential prey have used cetacean sightings and visual or acoustic prey sampling [18-20]. However, cetacean prey can hardly be detected from visual observations made at the surface. On the other hand, active acoustic surveys for prey may influence the detectability and behaviour of cetaceans that can be disturbed by the ping frequencies emitted by echosounders [19, 20]. Prey models can provide a suitable alternative.

The Spatial Ecosystem and Population Dynamics Model (SEAPODYM) is a numerical modelling framework for the management of marine resources and ecosystems. It includes representations of low and intermediate trophic levels (zooplankton and micronekton), age-structured fish populations, and fisheries. SEAPODYM MTL simulates biomass distributions and production of one zooplankton and six micronekton functional groups, according to their diel vertical migration behaviour and enables obtaining hindcast and forecast estimates [21].

The use of prey variables requires prior knowledge about the diet of the species to be modelled. The diet of the target species of this chapter - blue, fin, sei and sperm whales - has been studied in several world regions and in the Azores archipelago. The sperm whale is a generalist feeder that preys on deep-water cephalopods, primarily meso- and bathypelagic squids, and occasionally on fish [22, 23]. In the Azores, their diet is dominated by the families Octopoteuthidae (39.8% by weight), Histioteuthidae (32.7%), and Architeuthidae (12.1%), with *Taningia danae*, *Histioteuthis bonnellii* and Architeuthis dux representing the most important species [24]. Sperm whales' prey show a wide range of sizes but most cephalopods consumed in the Azores had a mantle length lower than 50 cm and weighted 400-450 g [24]. However, prey size can vary with gender and age [25]. In the Azores, sperm whales forage mostly between 600 and 900 m depth, both during the day and at night [26].

Blue whales are dietary specialists, feeding almost exclusively on krill. While there is little information about the prey species taken in the Azores, in the northwest and northeast Atlantic Ocean they feed mainly on *Thysanoessa raschii, Meganyctiphanes norvegica* and *T. inermis* [27, 28]. In the western Atlantic ocean, the distribution of blue whales is strongly related to *Nyctiphanes simplex* [29, 30]. North Atlantic fin whales also consume mostly euphausiids (especially *M. norvegica* and *T. raschii)* but they also prey on a variety of small epipelagic and mesopelagic schooling fish, including capelin (*Mallotus villosus*) and herring (*Clupea harengus*) [27]. Sei whales feed on copepods, euphausiids, amphipods, and small fishes, but their diet composition seems to differ markedly by region, season and year. In the northeast and northwest Atlantic Ocean, they appear to rely mainly on calanoid copepods (especially *Calanus finmarchicus*) and less on euphausiids

Recently, prey data derived from the SEAPODYM-MTL (Mid Trophic Level) model [31] were used to successfully predict the movements and habitat preferences of fin, blue and sei whales during their northward migration across the mid-North Atlantic Ocean, and the acoustic presence of foraging odontocetes, including sperm whales, at seamounts [32, 33].

In this chapter, environmental variables and simulated micronekton distribution from SEAPODYM-MTL were used to develop SDMs of sperm whales, blue whales, fin whales, and sei whales for the Azores archipelago. The goal was to improve previous SDMs for the region that were based only on physiographic and dynamic ocean variable, taken as proxies of prey availability [9, 10]) to create more realistic models. I utilised a Generalised Additive Model (GAM) modelling approach to create monthly predictive distribution maps for the four species in spring and summer. The models produced in the scope of this chapter are an essential tool to inform marine ecosystem conservation and management in the Azores region.

4.2. METHODS

4.2.1. THE STUDY AREA

The environmental setting and the biological importance of the study area for the studied species were described in the general Introduction of the thesis. Data analysed in this chapter were collected in an area of approximately 260.000 km^2 around the Azores archipelago, between 35ºN 23ºW and 41ºN 33ºW (Figure 4.1). The study area includes diverse bathymetry (1-4,000 m depth) and various underwater topographic features, including the shelves and slopes of the nine Azorean islands, shallow and deep-water seamounts, canyons and ridges, and the Mid-Atlantic Ridge.

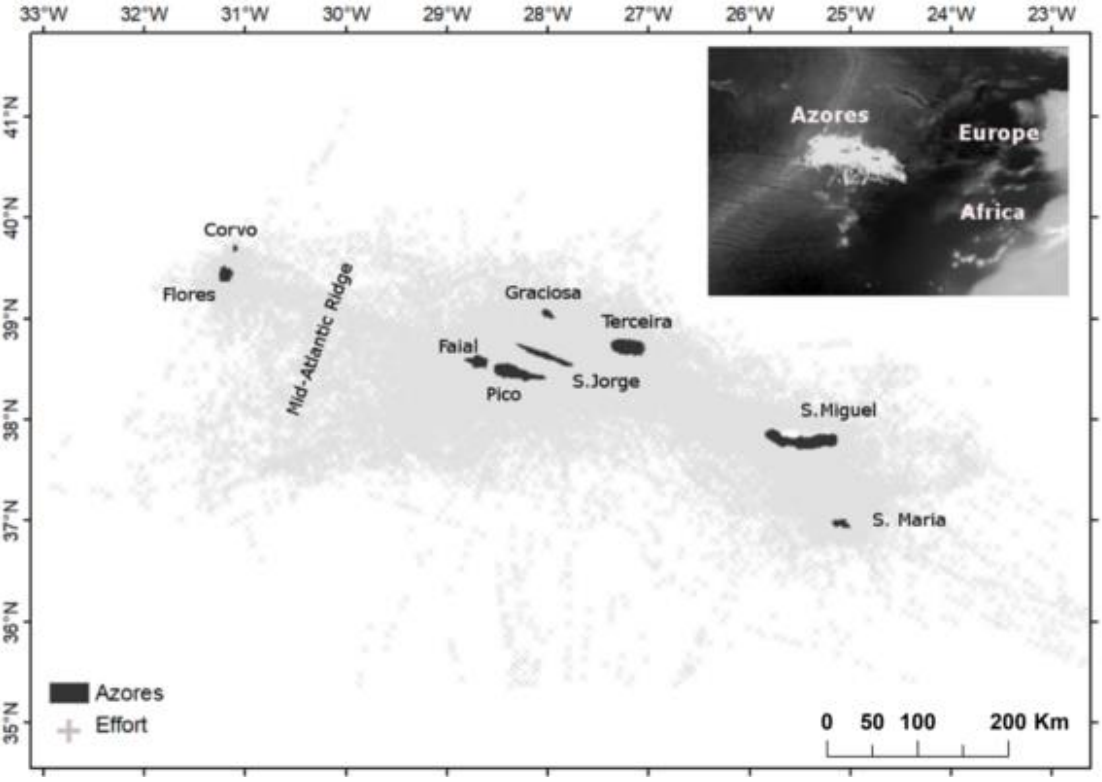


Figure 4. 1. Cetacean on-effort positions of POPA transects between 2001 and 2016 around the azores.

4.2.2. CETACEAN SURVEY DATA

Cetacean and effort data were collected during 2436 fishing trips (corresponding to 2481 days at sea) conducted from May 2001 to October 2016 by tuna-fishing vessels monitored by the Azorean Fisheries Observer Program (POPA). Further details on the POPA and on the methodology used to collect the data are provided by Tobeña et al. 2016 [10]. POPA places trained observers on board the Azorean tuna fishing fleet (some Madeiran vessels) operating within the Azores EEZ from April/May to October. While the vessels are travelling to or from the harbours, or searching for tuna, observers located on a high platform (~8m height) searched for cetaceans by naked eye or binoculars [34, 35]. During these dedicated cetacean-surveying periods (hereafter cetacean on-effort), the observers recorded information about the vessel (speed, heading, etc), weather and environmental conditions (Beaufort sea state, glare, cloud cover), and cetacean sightings (location, species, group size, behaviour). For each sighting, observers also recorded the reliability of species identification ranging from 0 (low confidence) to 3 (high confidence).

Fishing vessels occasionally operated together, duplicating survey tracks within the same area and day. In order to minimise spatial and temporal oversampling, vessel track lines were split into continuous on-effort segments of ~10 km. This involved computing the length of all on-effort vessel tracks: tracks shorter than 5 km were removed; tracks 5 to 10 km long were retained as segments; and tracks >10 km were split into 'n' segments of 10 km, until only one segment <10 km remained. If the remainder was <5 km, it was distributed uniformly among the other n segments (resulting in segments slightly larger than 10 km). Subsequently, to avoid duplication of sightings made by different fishing vessels, segments of different vessels <5 km from each other within the same day were identified. From this selection, a single segment was retained based on the following decreasing order of priorities: segment that had i) more blue whale sightings, ii) sightings of larger number of target species, iii) larger number of individuals. This was done to remove duplicate on-effort segments, while retaining the segments with the most information for modelling. Blue whale sightings were prioritised because there were fewer sightings compared to the other species. The current protocol resulted in 30,120 segments with a mean length of 8 km (SD= 2.1 km, range of 5-15 km), totalling 242,239,625 km of survey effort (Table 1).

The total number of individuals from each whale species was calculated for each segment. Only sightings made in good sea state conditions (i.e., Beaufort scale ≤3) and with high confidence in species identification (score 3) were used in the analysis. After data quality control (sea state condition, reliability of species identifications, oversampling correction and variable availability), the final dataset used for modelling included 1580 sperm whale individuals, 199 fin whales, 91 sei whales, and 44 blue whales. Most of the individuals were sighted singly, 44% for sperm whales, 50% for fin whales, 60% for sei and blue whales. Most of the sightings corresponded to adult individuals but calves or small juveniles made up 9% of the data records of blue whales, 4% of sperm whales, 2% of fin whales and 1% of sei whales.

Finally, the centroid of each effort segment was calculated and used to extract values of candidate explanatory variables, as described below. Besides the information on explanatory variables, each centroid includes data on the segment length (in kms) and the number of individuals by species (Table 4.1).

Table 4. 1. Cetacean data used for modelling.

	Sperm whale	Fin whale	Blue whale	Sei whale
Nº individuals	1580	199	44	91
Effort (Km)	242,232,281	242,239,625	242,239,625	242,239,625

4.2.3. HABITAT AND PREY DATA

Knowledge on the ecology of the four species was used to select the candidate physiographical, biogeochemical, topographical and prey candidate variables (Table 4.2) [7, 36, 37]. All the variables were first processed in UTM zone 26N with horizontal datum WGS84. Then all variables were projected to an Equidistant Cylindrical projection with horizontal datum WGS84 and resampled to the exact extent, with 2.5 arc-minute resolution. Weekly numerical data of all variables used for the model development were extracted from various sources (Table 4.2). Several variables were obtained from GLORYS.2v3 ocean reanalysis carried out by the European Copernicus Marine Environment Monitoring Service (hereafter, CMEMS ocean reanalysis), while others were output from the SEAPODYM-MTL model. The variable set was ordered in four groups: general variables, Derived environmental variables, Outputs from SEAPODYM-MTL, and Time variables. Depth was acquired from the grid-centered bedrock version of the ETOPO-1 digital elevation model [38]. Seamount coordinates and type (large or small) were defined by Morato et al. (2008). The Euclidean distance to seamount and by type (large or small) variables were derived from Seamount variable using a Geographic Information System (ArcGIS 10.3.1; ESRI, Inc.; hereby referred to as ArcGIS). The physical variables (Sea surface temperature and Sea water potential temperature) were obtained from the *CMEMS ocean reanalysis*. Sea surface temperature (sst) was based on analyses combining in-situ and satellite sources from both infrared and microwave radiometers. Sea water potential temperature (temperature) was acquired for the surface down to 5500 m of depth. The biogeochemical variables (Mass concentration of chlorophyll a in sea water (CHLA), Net primary production of carbon (NPP) and Euphotic depth (ZEU)) were obtained either from the sum of small and large phytoplankton production simulated using the PISCES or the ESSIC biogeochemical model (CMEMS ocean reanalyses).

Derived environmental variables were transformed (logarithm, square root). Logarithm of sea surface temperature (sstL) was derived from sst provided by CMEMS ocean reanalyses. Logarithm of depth (DepthL) and Slope within a 3x3 pixel kernel (Slope) were derived from Depth variable. Derived variables with the square root transformation from Depth and Seamount variables were: Euclidean distance to shoreline (Distance to shore), Euclidean distance to n meters isobaths (Dist(n), with “n” representing isoline depth), Euclidean distance to seamount (sd-Seamounts), Euclidean distance to large seamount (sdL-Seamounts), Euclidean distance to small seamount (sdS-Seamounts). Time-lagged Chlorophyll-a concentration and NPP for one and two months prior to the sighting month were logarithmized.

The Euclidean distance to SST fronts (Ed_Fronts) and Euclidean distance to eddies (Ed_Eddies) were run in the MGET-tool on ArcGis, with Canny algorithm, and Find eddies Okubo-Weiss, respectively. Ed_Fronts was calculated in WGS84_UTM_26N [39]. The Ed_Eddies downloads the SSH images from AVISO DUACS 2014 gridded sea surface height using the OPeNDAP protocol Chelton et al 2011 [40].

The SEAPODYM-MTL estimates the biomass and production of macrozooplankton and six functional groups of micronekton (consisting mainly of various fish, cephalopods, crustaceans of 2-20 cm length) between the surface and 1,000 m depth [41]. The model is driven by a series of advection–diffusion-reaction equations linked to environmental and physical variables, to describe the spatial dynamics and recruitment of the micronekton population [36]. The model defines three micronekton layers (epi, meso and bathypelagic) based on the euphotic depth, which are further divided into vertical migrant and non-migrant. The SEAPODYM-MTL candidate variables were: biomass and production of epipelagic, mesopelagic, migrant mesopelagic, highly migrant bathypelagic, and lower trophic level plankton. Baleen whales forage predominantly on prey available within the epipelagic layer [42]. Therefore, biomass and production outputs from the SEAPODYM-MTL pertaining to the bathypelagic and migrant bathypelagic prey, which are not available within the upper layer of the water column, were not included as candidate variables in the baleen whale models. Mesopelagic, bathypelagic and epipelagic prey layers were used to develop the sperm whale model [22, 24, 43, 44].

SEAPODYM-MTL outputs were provided at 0.25x 0.25 degrees and at a weekly resolution. Thus, all other variables were aggregated to match this spatial and temporal resolution on the exact grid using ArcGis.

Month and Year were included to investigate potential variations in the functional relationship of four whale species with the habitat and prey variables within and between years.

4.2.4. MODELLING

Generalised Additive Models (GAM) were used to model the relative density of each whale species, calculated as the total number of individuals of each species per survey segment (km) [45]. GAM were used because they have better responses to changes in the number of individuals than other types such BRT [45]. The natural logarithm of the survey effort (i.e, the segment length) was included as an offset, and habitat and prey data were used as covariates. I used a Tweedie distribution because of the large number of zeros in the sighting data [46, 47]. The smoothing splines function of the predictors were limited to 5 degrees of freedom to avoid overfitting, and preserve functional relationships and ecological interpretability. A process was used to select a subset of predictors based on their p-value (<0.05) [39] and on the performance of GAM models built for each variable of the candidate set. Covariates with a pair-wise correlation higher than 0.7 were removed. These correlations were calculated using the Pearson correlation test and the GAM concurvity measure, both implemented in R, using the *Hmisc* and *mgcv* package [48, 49]. The restricted maximum likelihood (REML) optimization method was used to fit the models [50]. The final model for each species contained the selected significant variables ($p<0.05$) that were not correlated with each other.

4.2.4.1. MODEL EVALUATION

A cross-validation procedure was used to assess the accuracy of the final model for each species. The data were randomly split into training (75% of records) and testing (remaining 25% of records) dataset (Table App. 4.1). Models built with the training dataset were fit to the test data and their predictive performance was assessed using the concordance index (C-index), which can be applied to continuous and categorical data. This index is equivalent to the area under the Receiver Operating Characteristic curve (AUC). It is generally assumed that a C-index between 0-0.5 indicates "no prediction", 0.7-0.8 "moderate discrimination", 0.8–0.9 "good discrimination", and 0.9–1 "excellent discrimination" [51]. Models and Relative explained deviance (R.ExpD) were performed using several R packages (R Core Team (2021)) [52]. The effect of each covariate included in the best GAM models was plotted to visually inspect the functional form to determine whether species densities exhibit comparable peaks or changes in relation to other covariates during the whole period or by month.

4.2.4.2. MODEL PREDICTIONS

Finally, I produced prediction maps based on the best model for each whale species using outputs averaged for 2001-2016. I predicted only within the range of sampled covariate values to avoid extrapolation to non-modelled conditions. In addition, monthly predictions were only produced for May-August, because number of observations outside this period was considerably lower. In order to facilitate comparison of prediction maps, predicted densities (individuals per km^2) were transformed using the natural logarithm of cell values in a raster. Model predictions were performed in R using the *mgcv* package [53].

Table 4. 2. Description of the candidate variables used in the sdms. pb: biomass distribution, pp: production distribution.

Environmental variable	Acronym	Transformation	Resolution Spatial/temporal	Units	Source
Depth	Depth	none	1 arcmin/static	m	NationalGeophysicalDataCenter(NGDC), National OceanicandAtmospheric Administration(NOAA) http://www.ngdc.noaa.gov/mgg/global/global.html. (Amante & Eakins 2008)
Seamounts	None	none	10 m	unitless	www.int-res.com/articles/suppl/m357p017_app.pdf. (Morato et al. 2008).
Euclidean distance to seamount	None	none	1 arcmin /static	m	Seamounts
Euclidean distance to large seamount	None	none	1 arcmin /static	m	Seamounts
Euclidean distance to small seamount	None	none	1 arcmin /static	m	Seamounts
Sea surface temperature	sst		0.25°	°C	CMEMS ocean reanalysis (1998-2015)*
Sea water potential temperature	temperature		0.25°	°C	CMEMS ocean reanalysis (1998-2015)*
Euphotic Depth	ZEU		0.25°	m	CMEMS ocean reanalysis (1998-2015)*
Mass concentration of chlorophyll a in sea water (glo glorys2 vgpm 025x7d)	CHLA	Log10	0.25°	mg/m^3	CMEMS ocean reanalysis (1998-2015)*
Net primary production of carbon (glo glorys2 vgpm 025x7d)	NPP	Log10	0.25°	mmolC/m^2/d	CMEMS ocean reanalysis (1998-2015)*
Derived environmental variables					Original variable

Logarithm of Sea surface temperature	sstL	Log10	0.25°	°C	CMEMS ocean reanalysis (1998-2015)*
Logarithm of depth	DepthL	Log10	1 arcmin /static		Depth
Slope within a 3x3 pixel kernel	Slope	log10	1 arcmin /static	degrees from the horizontal	Depth
Euclidean distance to shoreline	Distance to shore	square root	1 arcmin /static	m	Depth
Euclidean distance to 200 meters isobath	Dist(200)	square root	1 arcmin /static	m	Depth
Euclidean distance to 500 meters isobath	Dist(500)	square root	1 arcmin /static	m	Depth
Euclidean distance to seamount	sd-Seamounts	square root	1 arcmin /static	m	Seamounts
Euclidean distance to large seamount	sdL-Seamounts	square root	1 arcmin /static	m	Seamounts
Euclidean distance to small seamount	sdS-Seamounts	square root	1 arcmin /static	m	Seamounts
Time-lagged Chlorophyll-a concentration (_1 month)	CHLA (_1 m)	Log10	0.25°	mg/m^3	CHLA
Time-lagged Chlorophyll-a concentration (_2 month)	CHLA (_2 m)	Log10	0.25°	mg/m3	CHLA
Time-lagged NPP (_1 month)	NPP (_1 m)	Log10	0.25°	mmolC /m^2/d	NPP

Time-lagged NPP (_2 month)	NPP (_2 m)	Log10	0.25°	mmolC/m^2/d	NPP
Square root of Euclidean distance to SST fronts	Ed_Fronts	Canny algorithm, MGET	0.25°	m	Sea surface temperature
Euclidean distance to Eddies	Ed_Eddies	Okubo-Weiss, MGET	0.25°	m	Sea surface height, AVISO, MADT
Outputs from SEAPODYM-MTL					
Epipelagic micronekton(pb)	epi_pb		0.25°	g/m^2	SEAPODYM-MTL
Epipelagic micronekton(pp)	epi_pp		0.25°	g/m^2/d	SEAPODYM-MTL
Upper-mesopelagic(pb)	meso_pb		0.25°	g/m^2	SEAPODYM-MTL
Upper-mesopelagic(pp)	meso_pp		0.25°	g/m^2/d	SEAPODYM-MTL
Migrant upper-mesopelagic(pb)	mmeso_pb		0.25°	g/m^2	SEAPODYM-MTL
Migrant upper-mesopelagic(pp)	mmeso_pp		0.25°	g/m^2/d	SEAPODYM-MTL
High migrant lower-mesopelagic (pb)	hmbathy_pb		0.25°	g/m^2	SEAPODYM-MTL
High migrant lower-mesopelagic (pp)	hmbathy_pp		0.25°	g/m^2/d	SEAPODYM-MTL
Lower trophic level plankton(pb)	pk_pb		0.25°	g/m^2	SEAPODYM-MTL
Lower trophic level plankton(pp)	pk_pp		0.25°	g/m^2/d	SEAPODYM-MTL
Time variables					Original variable
Year	Year			year	Date
Month	Month			month	Date

4.3. RESULTS

4.3.1. SPECIES HABITAT PREFERENCES

4.3.1.1. SPERM WHALES

The best model for the sperm whale explained only 4.62% of the deviance and had a mean C-index of 0.60 (Table 4.3). The model included four covariates (Table 4.4), in decreasing order of contribution to deviance explained: Depth, Euclidean distance to small seamount (sdS-Seamounts), Month, and production of upper-mesopelagic (meso_pp). Sperm whales preferred areas of 900-2,200 m depth, at increasing distances from small seamounts, and with increasing mesopelagic production (Figure 4.2). In addition, there was a slight decrease in sperm whale density from august to november.

4.3.1.2. FIN WHALE

The best model for the fin whale explained 17.2% of the deviance and had a mean C-index of 0.76 (Table 4.3). Fin whale density was best predicted by 1-month lagged chlorophyll concentration (CHLA (_1 m)), 2-month lagged net primary production (NPP (_2 m)), and biomass of epipelagic micronekton (epi_pb) (Table 4.4). Density of fin whales increased linearly with NPP(_2m) and epi_pb, and peaked at at lagged chlorophyll values ≥19 mg/m^3 (Figure 4.3)

4.3.1.3. BLUE WHALE

The best model for blue whale had the highest deviance explained (20.6%) and a mean C-index of 0.86 (Table 4.3). The blue whale density was predicted by the logarithms of Sea surface temperature (sstL) and Depth (DepthL), and by 1-month lagged chlorophyll (CHLA (_1 m) (Table 4.5). Blue whales preferred colder, deeper waters which showed high (≥0.16 mg/m^3) chlorophyll concentration on the previous month (Figure 4.4).

4.3.1.4. SEI WHALE

The best fitting GAM for sei whale explained 13.4% of the deviance with a mean C-index of 0.76 (Table 4.3) and included only two covariates (Table 4.6): 1-month lagged NPP (_1 month) (NPP (_1 m)), and biomass of upper-mesopelagic (meso_pb). Sei whale density correlated with increasing NPP (_1 m) and decreased with biomass of upper-mesopelagic (Figure 4.5).

Table 4. 3. Summary of the output and evaluation of the best model for each whale species.

Species	Covariates in best model	Deviance explained (%)	AIC	TRAIN DATA		TEST DATA	
				Mean C-index	SD C-index	Mean C-index	SD C-index
Sperm whale	Environmental +SEAPODYM	4.62	20080	0.60	0.02	0.55	0.04
Fin whale	Environmental +SEAPODYM	17.2	18950	0.76	0.05	0.71	0.09

Blue whale	Environmental	20.6	17636	0.86	0.04	0.76	0.1
Sei whale	Environmental +SEAPODYM	13.4	19094	0.76	0.06	0.67	0.1

Table 4. 4. Covariates retained in the best sperm whale model. Depth, Depth; Euclidean distance to small seamount, sdS-Seamounts; upper-mesopelagic (pp), meso_pp.

Covariates	edf	Ref.df	F	p-value	R.ExpD(%)
Depth	3.111	4.0	9.320	2x10^{-16}	48
sdS-Seamounts	1.097	4.0	5.510	2.01x10^{-6}	27
month	2.980	3.64	5.197	8.8x10^{-4}	24
meso_pp	0.826	4.0	1.049	0.024	1

Table 4. 5. Covariates retained in the best fin whale model. Time-lagged chlorophyll-a concentration (_1 month), chla(_1m); time-lagged npp (_1 month), npp (_1m), and epipelagic micronekton (pb), epi_pb.

Covariates	edf	Ref.df	F	p-value	R.ExpD(%)
NPP(_2m)	1.023	4	3.845	4.71x10^{-5}	45
CHLA(_1m)	2.336	4	2.686	0.0035	41
epi_pb	0.816	4	1.020	0.0240	14

Table 4. 6. Covariates retained in the best blue whale model. Logarithm of sea surface temperature, sstl; logarithm of depth, depthl; time-lagged chlorophyll-a concentration (_1 month), chla(_1m)

Covariates	edf	Ref.df	F	p-value	R.ExpD(%)
sstL	1.086	4	2.906	0.0004	49
DepthL	0.857	4	1.341	0.0120	34
CHLA(_1m)	0.773	4	0.794	0.0401	17

Table 4. 7. Covariates retained in the best sei whale model. Time-lagged npp (_1 month), npp (_1m), and; upper-mesopelagic (pb), meso_pb.

Covariates	edf	Ref.df	F	p-value	R.ExpD(%)
NPP (_1 m)	1.051	4	6.172	4.22x10^{-7}	71

meso_pb	0.946	4	2.196	0.0019	29

4.3.2. Predicted relative density

Predicted relative density of sperm whales was considerably higher than densities of baleen whale species, overall and for all the months of the study period. Among the baleen whales, fin whales showed the highest relative density, followed by sei, and then blue whales. Relative density of baleen whales tended to decrease from spring to summer, generally reaching very low values in July and August. A comparison of model predictions with the pattern of sightings showed a good match for all the four species in the study area (Figure 4.App. 1).

4.3.2.1. Sperm whale

The model predicted relative density ranged from a maximum of 3,81 x 10^{-3} and a minimum of 5,9 x 10^{-5} individuals per 27,5 km^2 (1/4 degree 2) (logarithm of density ranged between -5 and -9) (Figure 4.6, Table App.2.). The highest relative densities occurred around NE Terceira, W Corvo and Flores, N of S. Miguel, NE of Sta. Maria, and S of Faial. The lowest predicted relative densities were around Princesa Alice seamount (within the 1000 m depth contour south of Pico and Faial islands), and SE of Sta. Maria. A low standard deviation (SD) was associated with low densities, but SD increased in the areas where predicted densities were higher (Figure 4.App. 2).

There were some spatial variations in predicted density across months (Figure 4.7). Density of sperm whales NE of Terceira decreased from spring to summer. The areas around Corvo and Flores had the highest densities in May and June. For the area in S. Miguel and Sta. Maria, the highest densities were in June and July, whereas S of Faial and N of Graciosa the highest densities occurred in July. The density of sperm whales in August was lower than in previous months.

4.3.2.2. Fin whale

Mean predicted fin whale density across the study area varied from 1,22 x 10^{-3} to 1,48 x 10^{-6} individuals per 27,5 km^2 (1/4 degree 2) (logarithm of density ranged from -6 to -13) (Figure 4.8). The highest relative density was predicted off the Mid-Atlantic Ridge, N of Faial, and between Terceira and S. Miguel. The lowest densities were predicted in the southern part of the study area (35-37.7º N), especially around Sta. Maria. Standard deviation of mean densities was generally low, increasing with increased predicted densities (Figure 4.App.2). The model predicted higher fin whale relative densities in May off the Mid-Atlantic Ridge, N of Faial, and NE of S. Miguel (Figure 4.9). There was a peak in density off the Mid-Atlantic Ridge and N of the Faial in June. Fin whale relative density decreased in July, with higher values NW of S. Miguel, and by August, density was low across most of the area.

4.3.2.3. Blue whale

The mean predicted density ranged from 5,95 x 10^{-4} to 3,31 × 10^{-8} individuals per 27,5 km^2 (1/4 degree 2) (logarithm of density -7 to -17) (Figure 4.10). The highest relative densities were SW of Sta. Maria, off the Mid-Atlantic Ridge and N of Corvo, and between S. Miguel and the central group of islands. The lowest density was predicted E of the Mid-Atlantic Ridge. As with other models, SD of mean densities augmented with increasing density values (Figure 4.App.2). Blue whale density decreased markedly from May to August (Figure 4.11). The highest values in May occurred SW of Sta. Maria, NE of Corvo, in the Mid-Atlantic Ridge, and between Terceira and S. Miguel, while in June were N of S. Miguel and off the Mid-Atlantic Ridge. Density of blue whales was low in July and August in the whole study area.

4.3.2.4. SEI WHALE

The mean predicted density from the best model ranged from 1,02 x 10^{-3} to 5,05 × 10^{-6} individuals per 27,5 km^2 (1/4 degree 2) (-6 to -12 after logarithmized) (Figure 4.12). The highest relative densities were predicted N of Flores and Corvo, followed by the Mid-Atlantic Ridge, W and S of Faial, between Terceira and S. Miguel, N of Terceira and NW of S. Miguel. The lowest densities were predicted E of the Mid-Atlantic Ridge and around S. Maria. As with the other baleen whales, sei whale predicted density decreased from May to August (Figure 4.13). The highest relative densities were predicted in May NE of Corvo, in the Mid-Atlantic Ridge, between S. Miguel and Terceira, and around the Pico and Faial islands. In June, sei whales showed a similar distribution but density was slightly lower. Density continued to decrease throughout July and August.

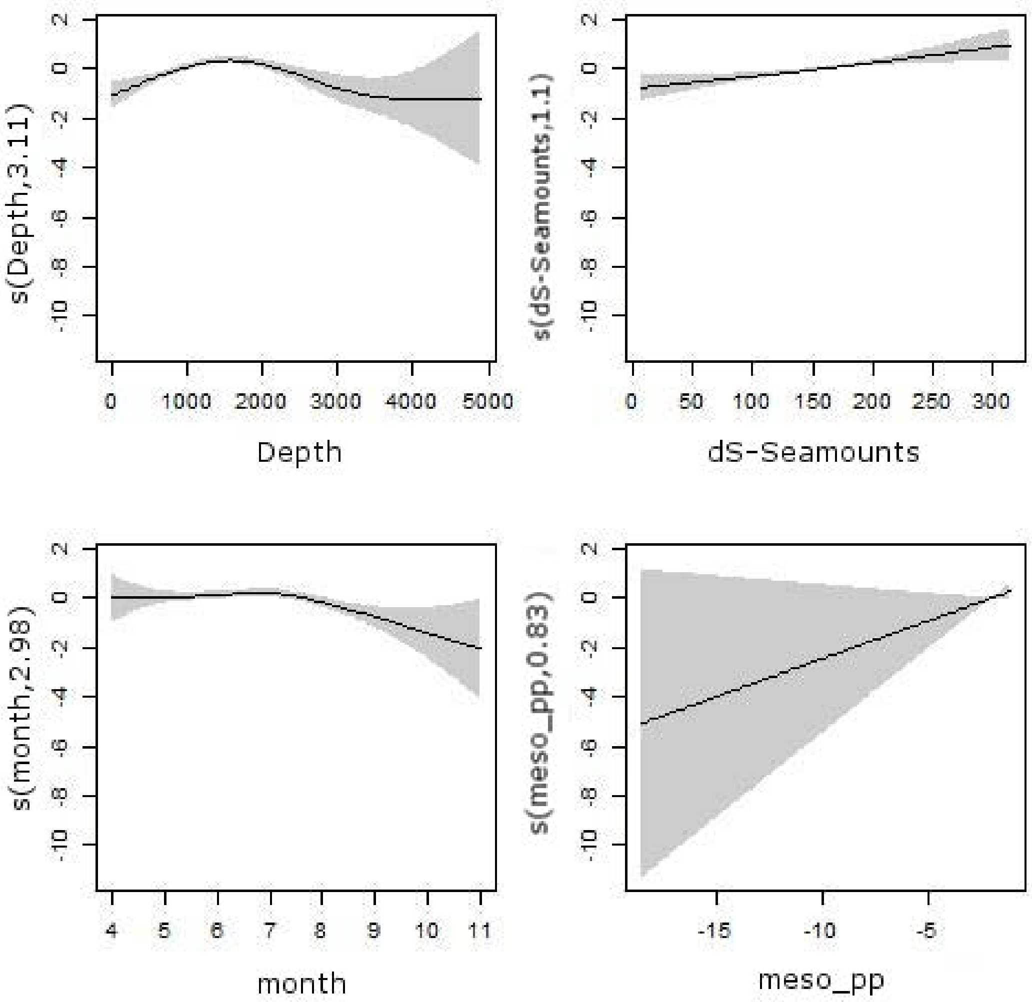


Figure 4. 2. Gam smoothers for sperm whales: on top left depth and on right distance to small seamounts (ds-seamounts), on down left month and right upper-mesopelagic(pp), meso_pp. Smooth function are shown as a solid line, and shaded regions represent 95% confidence intervals.

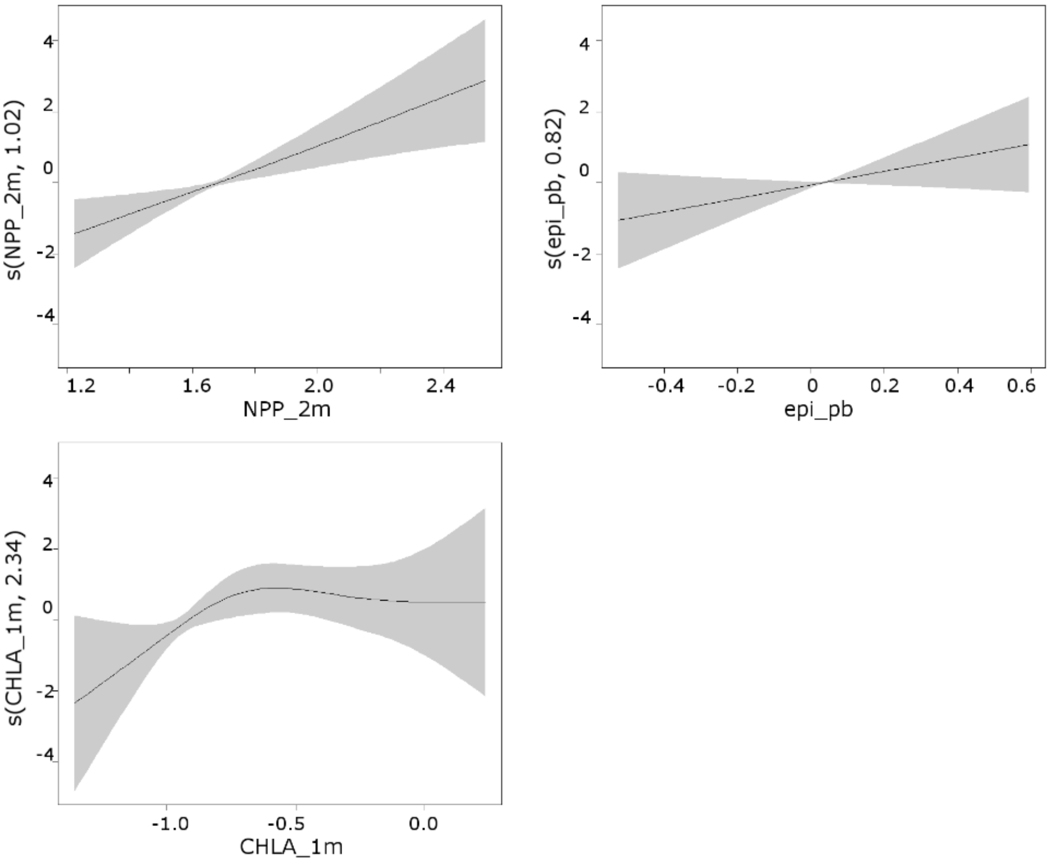


Figure 4. 3. Smoothed functions for the selected covariates for fin whale: on top left time-lagged npp (_2 month), npp (_2m) and right epipelagic micronekton (pb), on down left time-lagged chlorophyll-a concentration (_1 month), chla(_1m). Smooth function (solid line), shaded regions (95% confidence intervals).

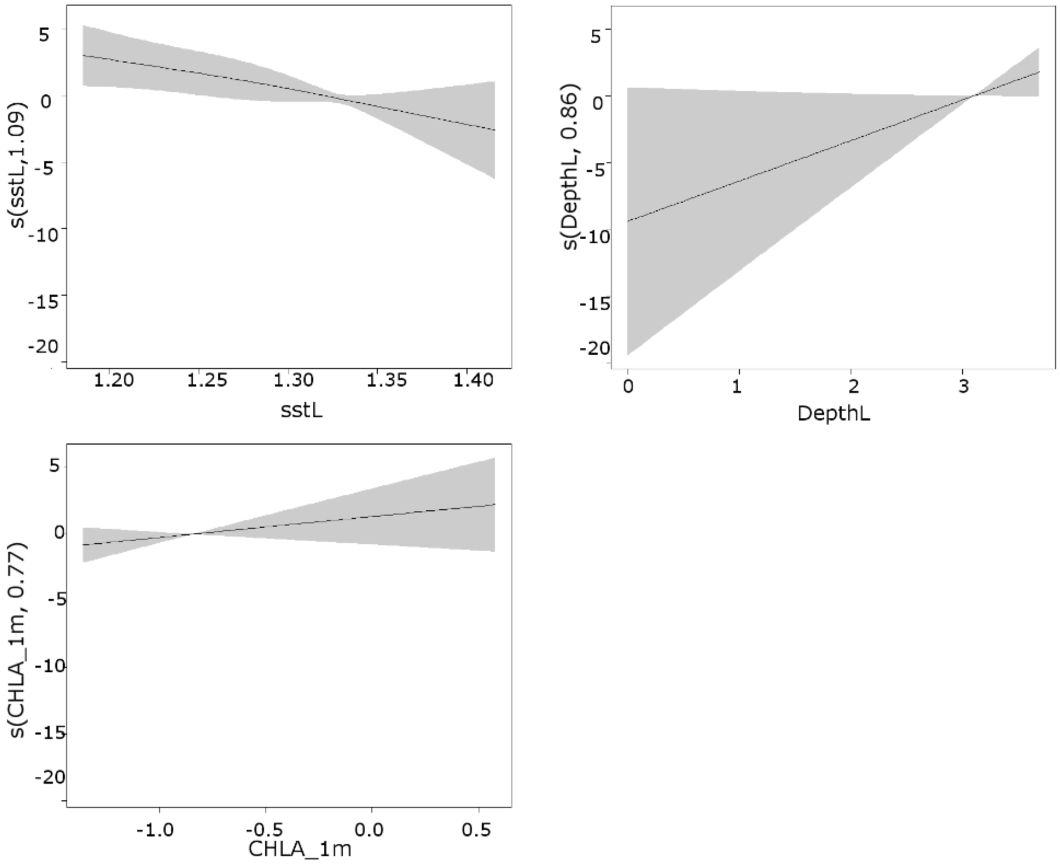


Figure 4. 4. Smoothed functions for the selected covariates for blue whale: on top left logarithm of sea surface temperature, sstl; and on right logarithm of depth, depthl; on down left time-lagged chlorophyll-a concentration (_1 month), chla(_1m). Smooth function (solid line), shaded regions (95% confidence intervals).

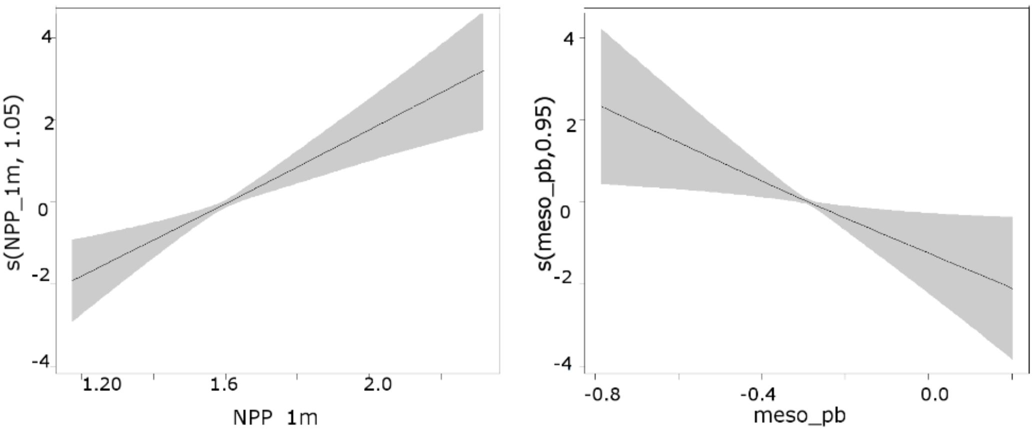


Figure 4. 5. Smoothed functions for the selected covariates for sei whale: on left time-lagged npp (_1 month), npp (_1m), and on right; upper-mesopelagic (pb), meso_pb. Smooth function (solid line), shaded regions (95% confidence intervals).

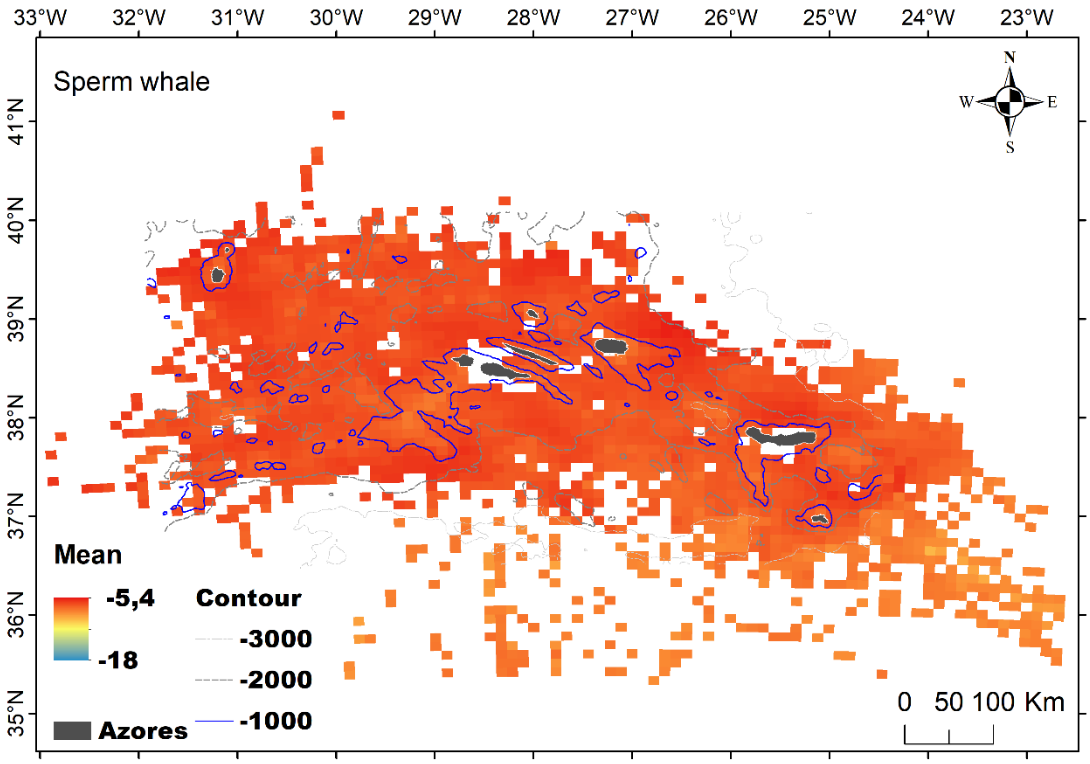


Figure 4. 6. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sperm whales based on the train dataset

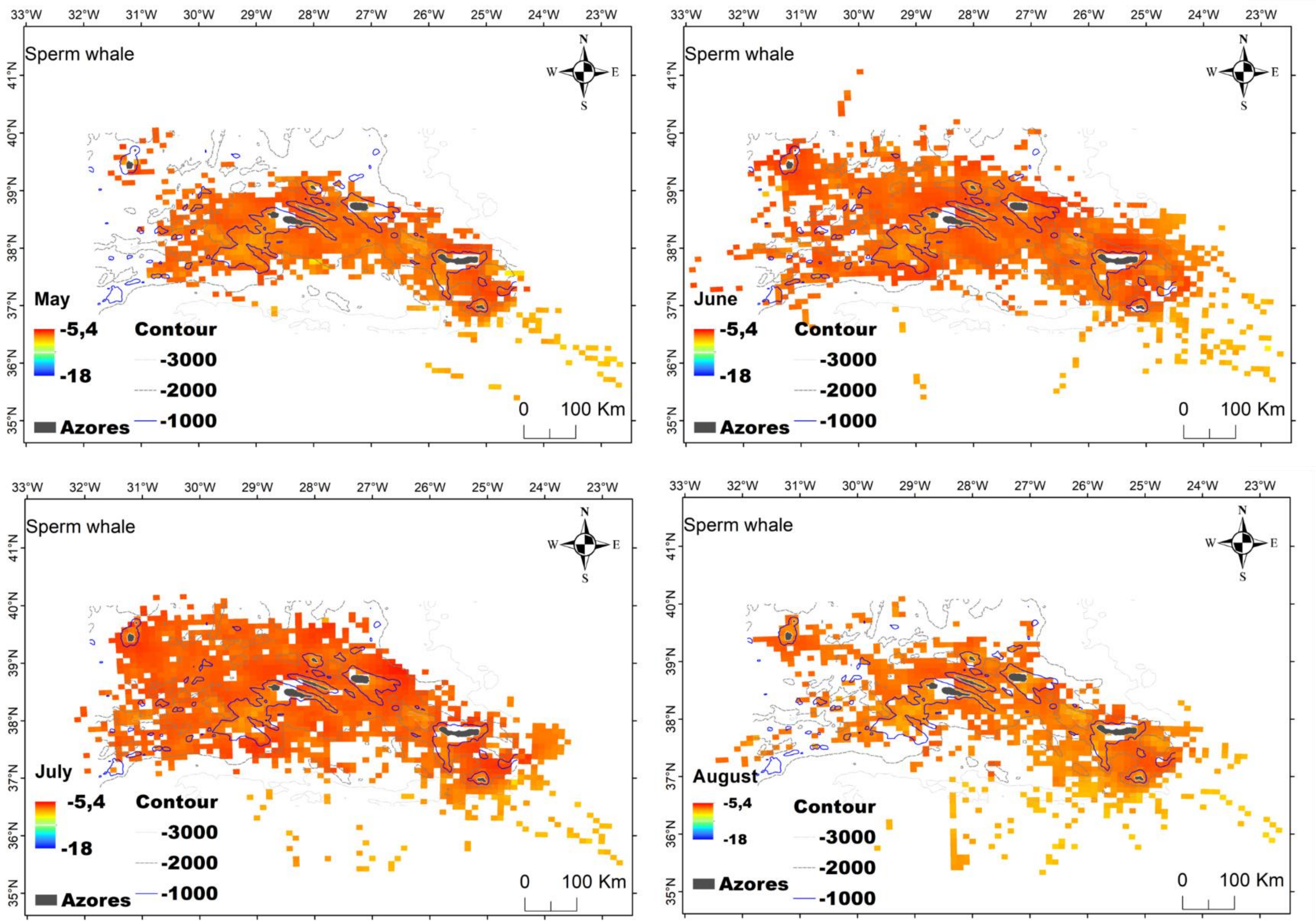


Figure 4. 7. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sperm whales from May to August based on the train dataset

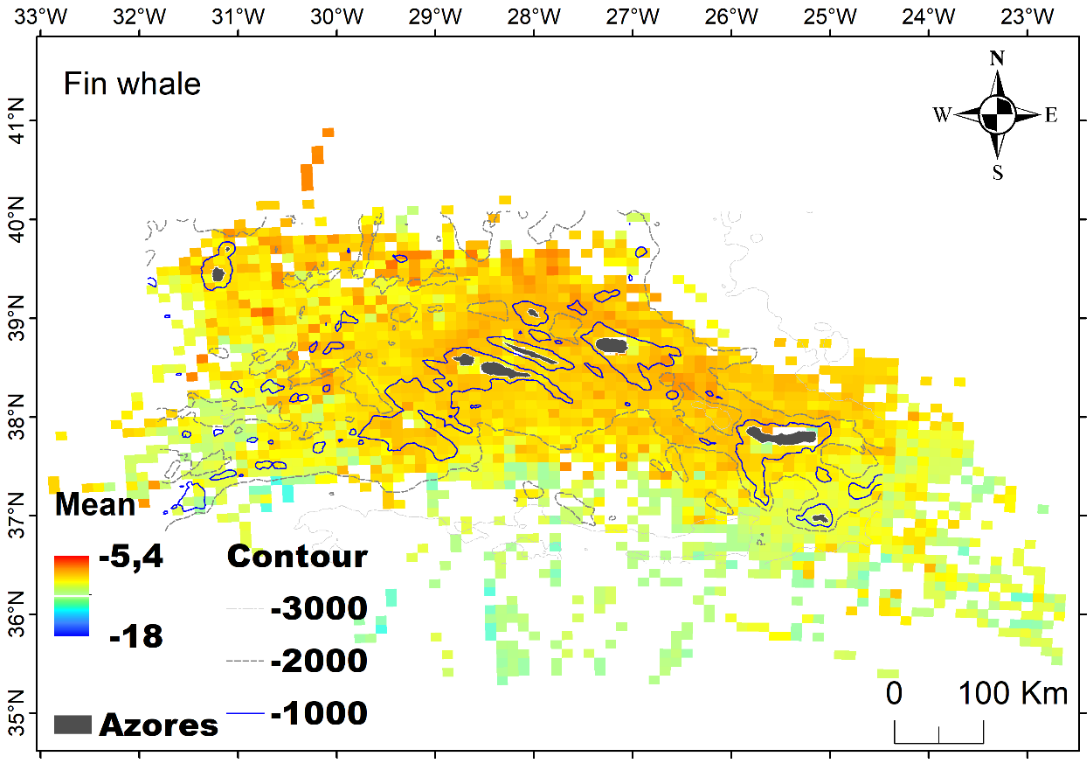


Figure 4. 8. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of fin whales based on the train dataset.

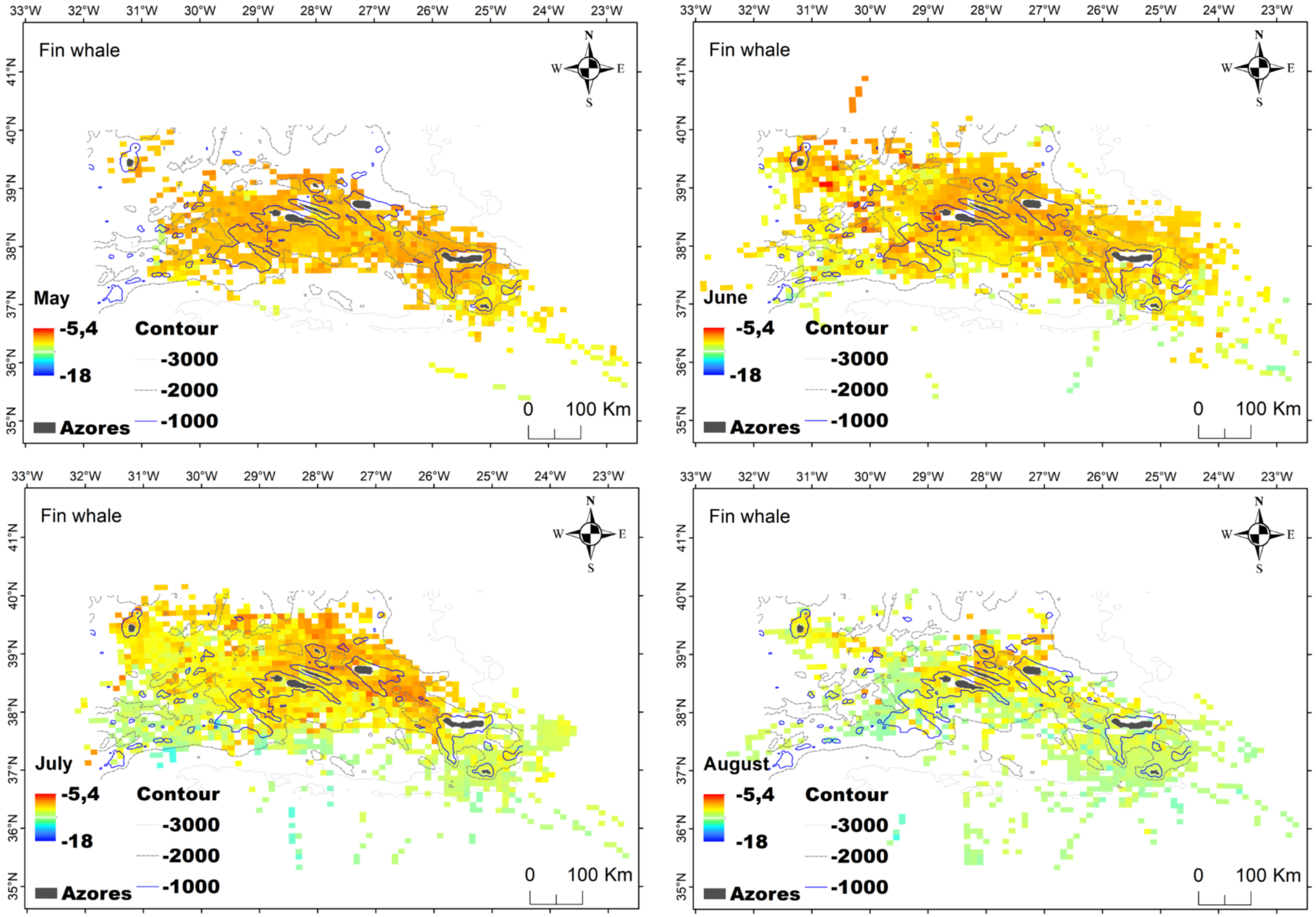


Figure 4. 9. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of fin whales from May to August based on the train dataset.

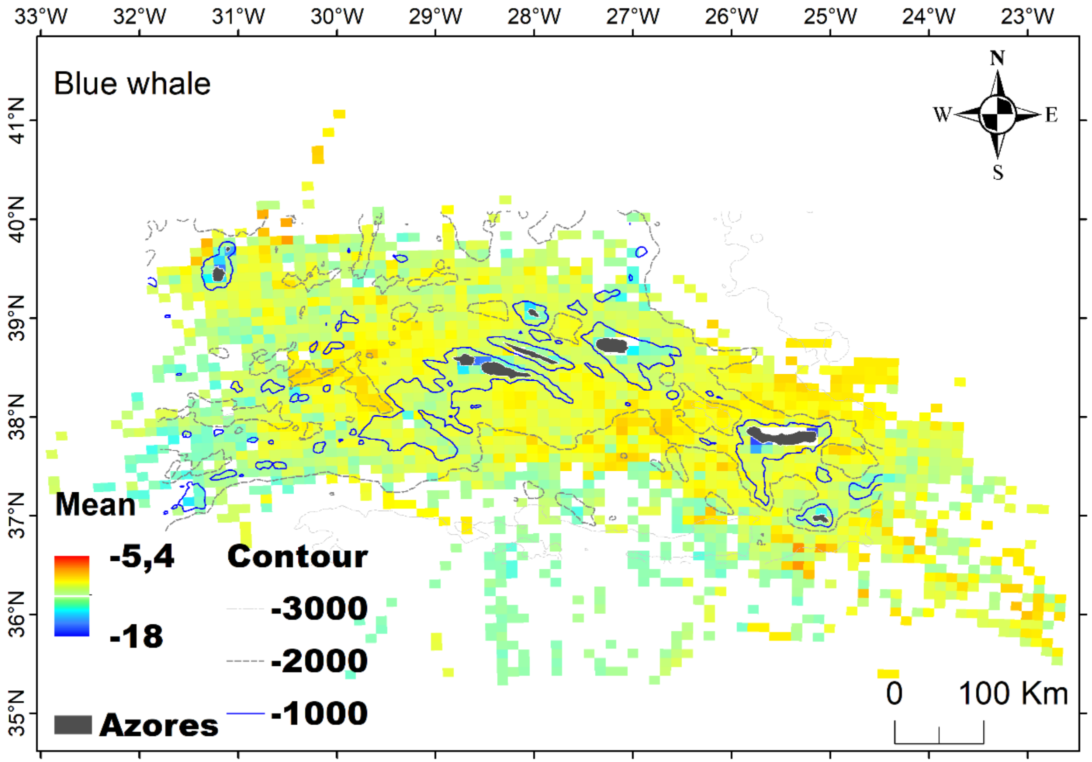


Figure 4. 10. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of blue whales based on the train dataset.

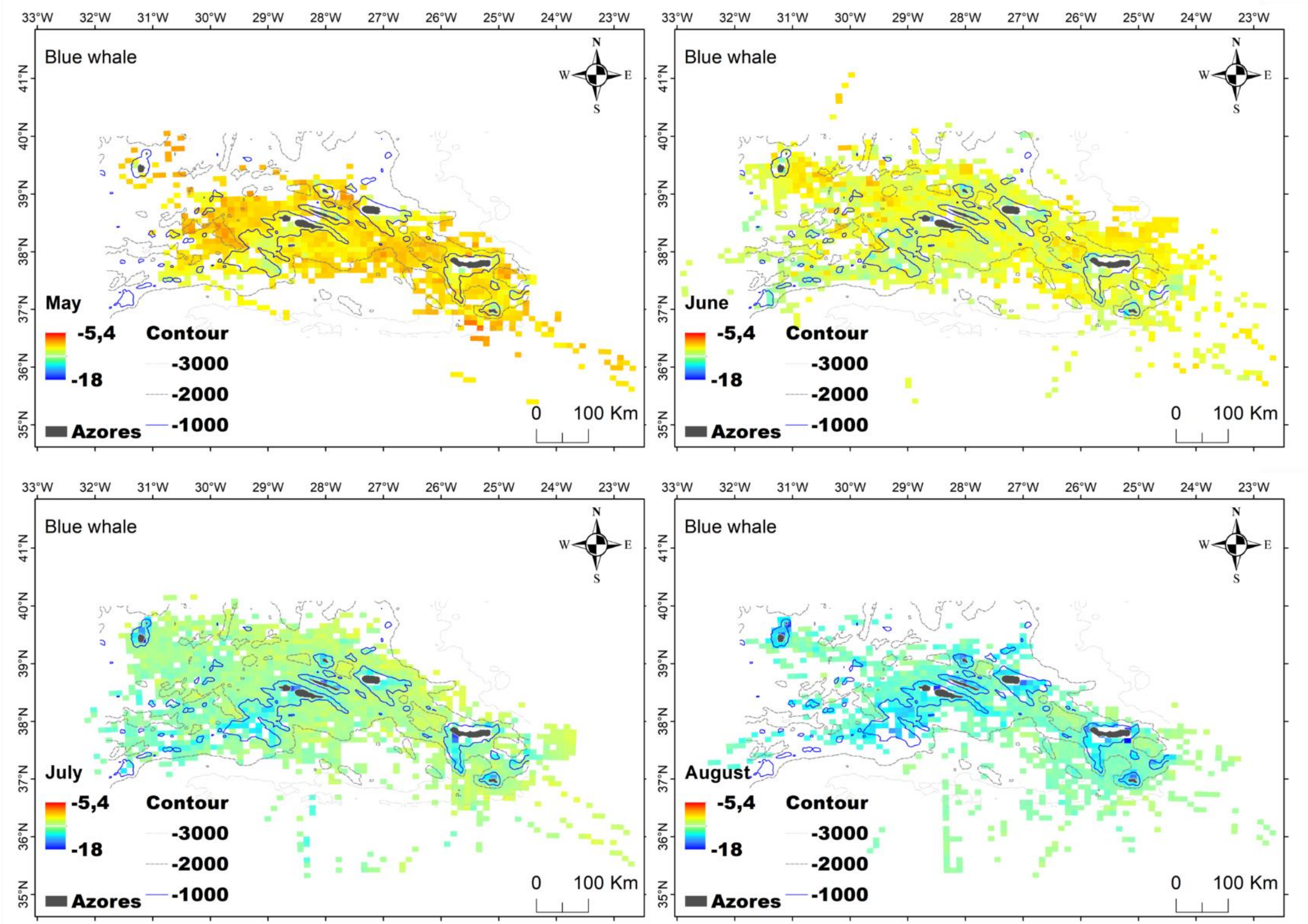


Figure 4. 11. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of blue whales from May to August based on the train dataset.

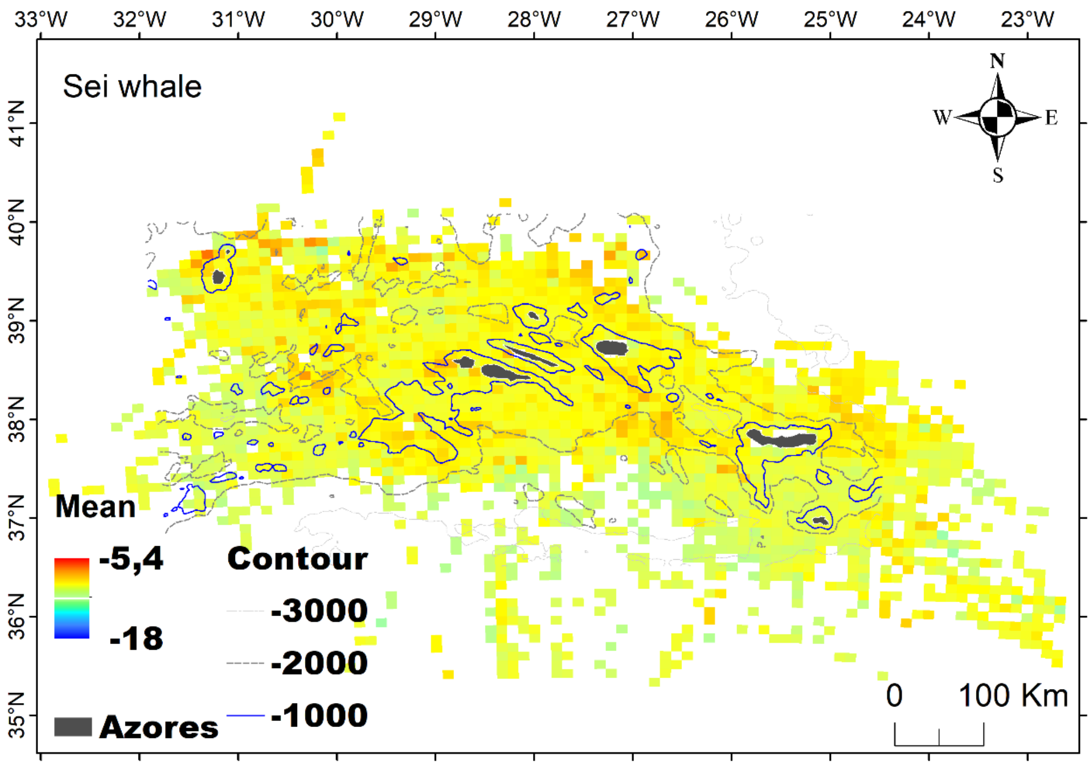


Figure 4. 12. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sei whales based on the train dataset.

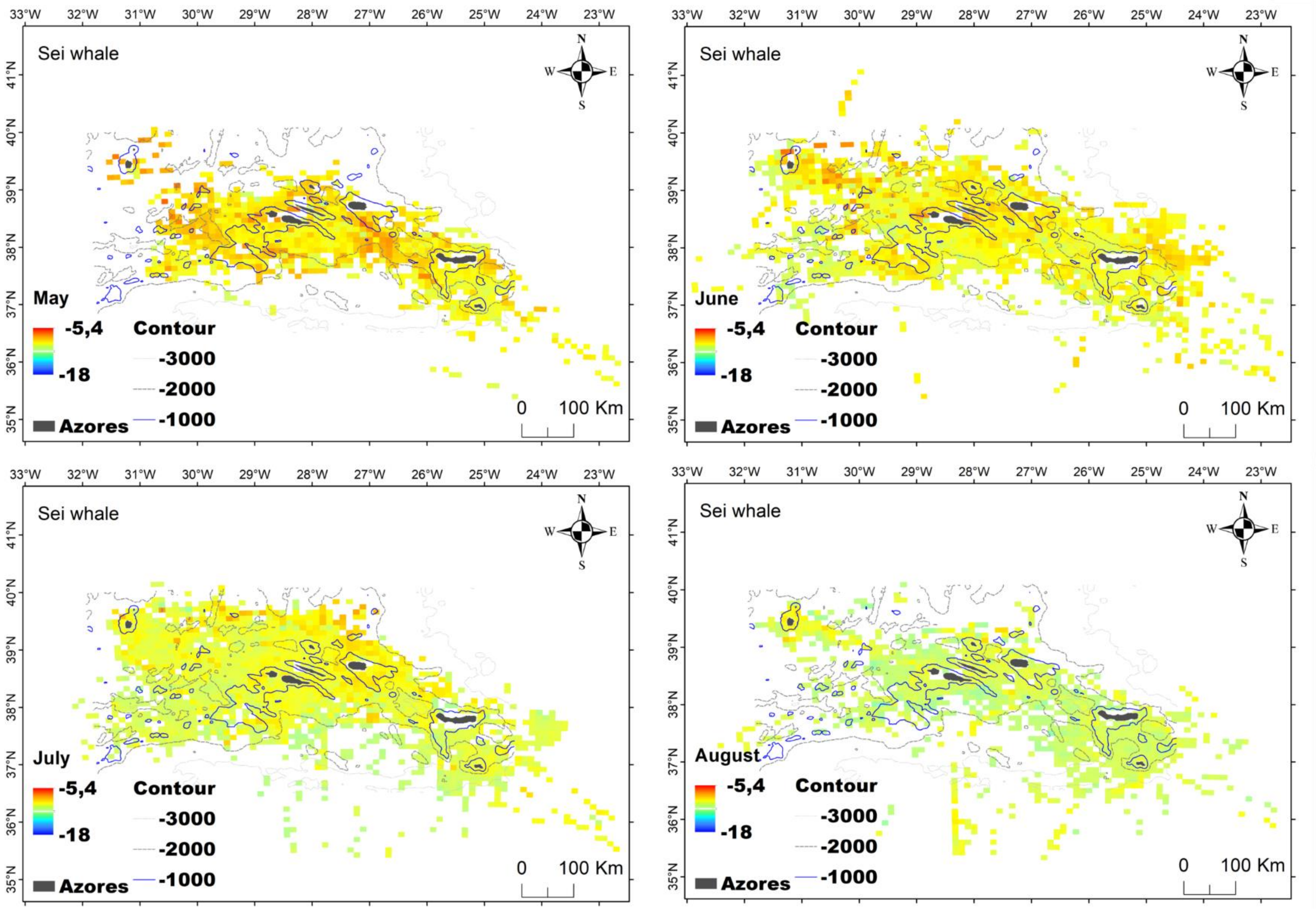


Figure 4. 13. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sei whales from May to August based on the train dataset.

4.4. DISCUSSION

This chapter applied GAM models to predict the relative density of sperm, blue, fin and sei whales around the Azores, in spring and summer. The models were based on a long-term dataset of effort and cetacean sightings, and included simulated prey variables that have rarely been used in SDMs of large whales.

4.4.1. MODEL LIMITATIONS AND PREDICTIVE PERFORMANCE

The dataset used in this study covered 16 years of cetacean sightings collected over a large area around the Azores. However, the dataset does not contain data from all seasons, precluding examination of year-round habitat preferences and density of large whales. In addition, inter-annual differences in habitat preferences could not be investigated, because sightings were insufficient to run models by year. Moreover, at the beginning and end of the fishing season, the dataset seems to cover a smaller geographic area, so results must be compared with caution. Another disadvantage of POPA data is that it was not collected along random or systematic transects, and may be spatially and temporally biased. POPA's sighting effort depend on the distribution of tuna, the tuna species being targeted, the proximity of the vessel's ports of departure/arrival, among other factors. Many of these factors contribute for the concentration of fishing activity and the aggregation of several vessels within the same area and period. To address the spatial and temporal bias in the survey effort, transects conducted in the same area and day were removed from the dataset.

Modelling requires selecting variables that are ecologically relevant for each species. However, many covariates, including habitat and prey variables from SEAPODYM, were often correlated. An example was the strong collinearity (>0.7) between chlorophyll concentration, net primary production of carbon and biomass of lower trophic level plankton. There were also strong correlations between some SEAPODYM prey variables. When this happened, I kept the prey variable that made most sense, based on the known feeding behaviour of each species. Nevertheless, by removing some of the correlated variables, some ecological or oceanographic processes driving whale density may go unnoticed or be difficult to interpret. Other correlated variables were more difficult to select because they may indicate diverse processes or events. As expected, month and temperature variables were highly correlated, showing the known seasonal patterns. However, other environmental dynamics have also clear seasonal patterns [54]. Hence, temperature can indicate diverse environmental processes or serve as a proxy for prey (e.g., associated with the spring bloom) [55].

Another limitation in this study was the spatial resolution of the SEAPODYM data, at 0.25°. The highly dynamic environment of the Azores may not be adequately captured by variables with a resolution of more than 500 km^2. In fact, recent studies revealed that using coarse resolution climatological data in habitat-based models for migratory marine predators can decrease models' predictive performance, particularly in highly dynamic regions [56]. While in this case, contemporary and not climatological data were used, the low spatial resolution of the data may have contributed to the poor performance of some of the models, and to the wide confidence intervals of the relationship of whale density with several covariates.

Despite these shortcomings, with the exception of the sperm whale model, the models for the other whale species performed relatively well. The deviance explained for the baleen whale models ranged from 13.4% (sei whale) to 20.6% (blue whale). In agreement with the deviance results, the blue whale model showed the highest predictive performance, as indicated by the C-index >0.85, and the fin and sei whale models were close to a good accuracy, with a C-index around 0.8. In contrast, the sperm whale model explained 4.6% of the deviance and showed a poor ability to predict the species density (C-index 0.6). These results based on the training dataset (75% random data from 2001 to 2016) were almost identical to the C-index calculated from the test data.

In general, SDM models built based on GAM perform better with more sighting data (of species or individuals) [57]; accordingly, I would expect the sperm whale model to have the best results, which was not the case. The accuracy of the sighting locations has also been shown to improve the model results [57]. However, all the species were recorded in the same way and by the same observers and accuracy should be the same. On the other hand, there is evidence that it is more challenging to estimate large groups than small groups accurately [58]. This may have affected the sperm whale model because sperm whales occasionally occur in numerous groups [34] with a wider separation between individuals [59-61]. Finally, the lower performance of the sperm whale model suggests that available covariates were unable to capture the drivers of the specie's density, namely, their prey distribution. This hypothesis is discussed in more detail below.

Nevertheless, the values of deviance explained obtained in this study generally agreed with those reported in other studies using GAMs. Deviance explained by models of sperm whale distribution and density commonly ranged from ~5%, when using only oceanographic covariates, to 8-12%, when including prey variables [6, 7, 16, 62, 63]. Usually, deviance reported for baleen whale models is slightly higher, varying between 6 and 34% [6, 13, 15, 62, 64].

4.4.2. HABITAT PREFERENCES

The density of large whales was strongly associated with prey availability and dynamic variables linked to biological productivity, as well as depth. In addition, none of the developed models retained more than four variables. Therefore, they avoid complexity and favour their interpretation. Prey variables were retained in the best models of three out of four whale species, reflecting the tight relationship between whale occurrence in the Azores and the availability of food in the area, and illustrating the importance of including prey data when attempting to predict whale distribution. Net primary production (NPP) and chlorophyll-a concentration (CHLA) were also important predictors of baleen whale density, but the correlation between these predictors and whale density was lagged by one or two months, which likely results from the time mediating the peak in phytoplankton and the subsequent development of the zooplankton, which are prey to baleen whales [65, 66]. Among the static covariates, depth was important at predicting the density of sperm and blue whales, with sperm whales showing a preference for a narrow range of depths, while density of blue whales increased linearly with depth. Nevertheless, the strength and shape of the relationship with these predictors varied between species, indicating distinct habitat preferences arising from differences in diet and foraging behaviour.

Sperm whales in the Azores consistently forage below 600 m depth, often reaching 900-1100 m [67], so the preference for deeper waters seems reasonable. Perhaps more surprising was that depth was still the most important variable in the sperm whale model, despite the prey variables being available to fit the model. According to sperm whale depth profiles in the region, most of their foraging activity occurs within the mesopelagic layer (<1000 m depth). Most of their prey reaches more than 20 cm in size, and are not part of the micronekton, so they would not be included in any of the prey variables output from SEAPODYM. Nevertheless, sperm whale prey are known to feed on small micronekton organisms [24]. Thus, I expected SEAPODYM variables to be a good proxy of the distribution of sperm whale prey. As anticipated, the prey layer retained in the best sperm whale model was the upper-mesopelagic. Still, upper-mesopelagic production had a small contribution to the variability in the data, and other static and time variables had a much stronger relationship with sperm whale density. In addition, variations in prey preferences (species or size) by year, whale sex or age class [25], could confound the modelling results. Future studies may consider developing annual models to capture the possible inter-annual prey variations or include lagged prey variables.

Another critical variable in the sperm whale best model was month, indicating a seasonality in the density of the species, with lower values in late summer and autumn. Care should be taken, however, as these results may be influenced by the contraction of the surveyed area in the end of the fishing season previously mentioned.

Despite the evidence of a relationship between sperm whales and seamounts [68] and the abundance of these topographic structures in the study area [69], the distribution of sperm whales does not seem to be categorically linked to the seamounts in the Azores [10, 69]. Besides, there is evidence that not all seamounts have a regular presence of sperm whales [32]. The discrepancy in sperm whale usage of seamounts could be due to limitations in sampling or differences in the physiographical or oceanographic setting, affecting prey availability [32]. In this study, sperm whale density increased with increasing distance from small seamounts and generally preferred areas about 100 km away from these structures. It is possible that these structures create physical processes (e.g., turbulence, increased mixing) that propagate horizontally or vertically across space. It is also possible that this relationship is an artefact and reflects the association of sperm whales with a variable not included in the models. The simulated SEAPODYM-MTL has a spatial resolution on 25º and the latest SEAPODYM-LMTL on 0,08º, with this finer resolution it would expect to capture better the accumulation of biomass and their specific location. Because the new simulated SEAPODYM prey (zooplankton and micronekton) with higher resolution will predict accurately mesoscale features favourable to retention prey around seamounts and clarify whether there is an influence of these topographical elements on sperm whales in the area. Future studies should attempt to classify seamounts according to their oceanographic conditions and size to attempt to resolve this potential association.

While all three baleen whale species are known to feed on epipelagic organisms, especially euphausiids and copepods, epipelagic micronekton was only retained in the fin whale model, and even so, it was not the most influential variable. One-month lagged chlorophyll and 2-month lagged net primary production were more important at explaining fin whale density than biomass of epipelagic micronekton. The former variables were also retained in the blue whale and in the sei whale model, respectively. Both these variables are indicative of the primary productivity potentially available for the zooplankton on which baleen whales feed. The retention of these variables in the models indicates that the simulated SEAPODYM prey do not accurately reflect the prey available for baleen whales in the area, or that the spatial and/or temporal scales of the outputs are unable to capture the predator-prey relationships, as mentioned previously. It is also possible that SEAPODYM is not capturing well the temporal responses of lower trophic levels to changes in productivity. In future models, epipelagic micronekton from the previous months could be included to assess a possible mismatch between simulated prey and these whales.

Blue whales preferred slightly higher chlorophyll values than the fin whale, which might be related to their seasonal patterns. Predicted density of blue and fin whales was higher in the months around the spring bloom; however, blue whale density dropped sharply after the spring bloom, while fin whales were still present at moderate densities. These results suggest a closer relationship with this phenomenon for the blue whale than the fin whale. In addition to lagged chlorophyll concentration, blue whale density was inversely associated with sea surface temperature and positively related with depth. The strong association with temperature likely results from the seasonality of blue whales in the area, and the lower temperatures that occur in the development of the spring phytoplankton bloom. The relationship with depth could reflect a genuine preference of blue whales for deeper or more offshore areas, although the wide confidence intervals raise some suspicion and the relationship needs to be further investigated.

The diet of the sei whales in the Atlantic ocean is mainly based on calanoid copepods and less frequently on euphausiids and small fishes [70, 71]. Thus, epipelagic micronekton appeared as the strongest prey candidate in the set. Surprisingly, the best model for the species included 1-month lagged NPP and upper-mesopelagic biomass. Although previous studies indicated that sei whales rarely forage in the Azores archipelago [70], the best model was strongly related to a prey proxy (NPP). One hypothesis, is that NPP in the area reflects processes happening on large geographic scales and could indicate the future conditions on sei whale feeding grounds to which they were heading. Alternatively, sei whales may forage more frequently than assumed, and the relationship with lagged NPP reflects the temporal delay between primary productivity and the development of zooplankton on which these whales forage. The negative relationship with biomass of upper-mesopelagic was intriguing. It could be an indirect consequence of sei whales avoiding oceanographic structures or environmental conditions preferred by upper-mesopelagic, although this seems unlikely, as many of these organisms have the same dietary preferences as sei whales. Sei whales could also avoid areas of high mesopelagic abundance to avoid other large megafauna species, that could be competitors or potential predators. Despite GAM allows investigate the ecological relationships and is a sensible technique that give insight into the ecology, even fitting complex non-linear responses (asymmetrical or skewed), better than conventional parametric techniques [72]. At present, I cannot explain the relationship of sei whales with mesopelagic biomass, and future studies should explore this in greater detail.

4.4.3. PREDICTED RELATIVE DENSITIES OF LARGE WHALES

Overall predicted relative densities of whale species obtained in this chapter are consistent with known relative abundance of the species in the area, the sperm whale being the most frequently sighted species, followed by the fin, sei and blue whales [34] In addition, the decreasing trend in the relative density of fin, blue and sei whales from spring to summer is also in agreement with the known seasonality of these species in the region. Relative density of sperm whales also showed some variability with season, albeit less than baleen whales, which also agrees with previous information [34].

No single area stood out as a high-density area for the sperm whale or any of the baleen whale species. Hotspots of higher whale density identified by the models were scattered across most of the study area. Moreover, most of these hotspots were not static but changed over time, reflecting the dynamics in local and regional oceanography. Nevertheless, some general patterns emerged from predicted density maps of the four species. Sperm whales tended to avoid waters shallower than ~1000 m depth, and therefore the areas close to the islands, banks (Princesa Alice, D. João de Castro, Mar da Prata) and around the Mid-Atlantic Ridge, showed lower than average densities, while density was generally high in most of the remainder waters. Interestingly, several high-density hotspots of baleen whales occurred along or in the vicinity of the Mid-Atlantic

Ridge, in addition to the area around Faial and Pico, and between Terceira and S. Miguel. With the exception of blue whales, predicted density of whales around Sta. Maria, or at least south of this area, was low. While this result could partly reflect variation in survey effort at the limits of the study area, the same was not observed at the northern limit, suggesting that the low overall whale density around Sta. Maria was true.

The habitat preferences of the four species were strongly linked to dynamic variables, which naturally vary throughout the year, creating temporal fluctuations in their density. For baleen whales, seasonality in density is evident, and creates a latitudinal gradient, that reflects the movement of blue, fin and sei whales through the area as they track the northward progression of biological productivity across the North Atlantic from early spring to the end of summer [9, 33, 70, 73, 74]. Conversely, no obvious spatial pattern emerged from the maps of sperm whale density. Sperm whales occur year-round in the Azores, although there are no resident groups in the area [75, 76]. They are primarily females with juveniles, although adult males are seen occasionally [76]. While sperm whales can cover extensive ranges [77] and adult males can travel long distances between their polar/subpolar feeding grounds and the temperate/tropical mating grounds [78, 79], there is no evidence of seasonally directional migration in this species.

Blue whale density has a temporal pattern decreasing from May to August, which match with the migrations known in the mid-Atlantic [33, 34]. This whale is the only baleen whale with a static variable in the best model. The spatial pattern is very marked, around -1000 m, especially in May. This relationship may be because this month has the highest density values and stands out more. Alternatively, the importance of depth varies by season.

The sei whale's density is strongly linked to the mid-Atlantic ridge (NW Azores archipelago) crossing the Azores archipelago. However, the initially important density of whales between islands is rapidly lost over the months. The differences between these two physical structures suggest that the mid-Atlantic ridge is more relevant for the sei whale migration. Accordingly, as seen in previous studies [80], physical structures are relevant for sei whales, but the best model did not capture them. In future studies, new topographic variables should be candidates to build sei whales models.

4.5. CONCLUSION

This chapter developed predictive models of density of four large whale species in the Azores archipelago. These models have expanded the knowledge about the seasonal changes in spatial density and enabled identifying important habitat areas for these species within the archipelago [81]. Such knowledge is fundamental to inform future research programmes focusing on these species, as well as marine spatial planning initiatives [1, 7, 82, 83]. This is especially relevant for species that are challenging to monitor continuously, like large whales, due to their high mobility.

This study was based on data collected by the fisheries observer programme that has been running in the Azores since 1998. Data collection by POPA follows the same standardized protocol and observers receive intensive training on data collection. In the absence of a dedicated line-transect monitoring programme, platforms of opportunity like POPA are an excellent way for collecting scientific data on cetaceans (and other taxa), that can then be used for multiple purposes, such as developing SDMs. Nevertheless, data collected by platforms of opportunity suffer from several limitations. In the case of POPA, one of the main limitations is that it only covers the spring and summer seasons. Combining data collected by POPA sighting surveys with data from other sources (e.g, satellite tagging, acoustic recordings, and photo-identification) could increase the knowledge on these endangered and vulnerable whale species.

The best models for three of the four species included SEAPODYM prey variables. Still, prey variables always explained less variance than the other variables retained in the models. Therefore, the density of these whale species was only weakly related to SEAPODYM prey variables. There may be several explanations for this. First, these variables may not represent well the availability of the prey taxa being targeted by these whales. I believe this could be the case for the sperm whale which prey on deep water squid that are usually larger than the micronekton simulated by SEAPODYM. In addition, cephalopods have only a minor contribution to the acoustic backscatter [84] at the frequencies (38 kHz) used to collect data to feed and calibrate the SEAPODYM model [31]. Thus, data output from the model only partially reflects the density of cephalopods preyed by sperm whales. Second, prey metrics output from SEAPODYM may not be the most relevant for these whales. Several studies have shown that the density and depth of krill patches is more important at driving distribution and foraging behaviour of baleen whales than total abundance or biomass [18, 20, 28]. Third, the spatial and temporal resolution of SEAPODYM data may not reflect the dynamics of prey at scales relevant to these predators. SEAPODYM data is now available including zooplankton and micronekton (upper and lower mesopelagic groups) at finer spatial and temporal resolutions [31]. Future SDMs incorporating these data could enable examining the habitat preferences and estimating the relative density of large whales at much finer scales. However, it is unclear how these models would perform, as recent studies suggest that models with large-scale prey information performed better than those with finer scale [14, 16]. Future studies should strive to investigate the effects of changing the spatial and temporal resolution of input data and the scale of analysis on SDM performance

The models of all four whale species highlighted variations in spatial density across months, and this must be taken into account when developing spatial conservation measures, like Marine Protected Areas (MPAs). This can be especially challenging because the dynamic changes imply considerable adaptive management to achieve successful implementation [85, 86]. Regular updates of dynamic MPAs need to be transferred into regulatory frameworks and be quickly disseminated to all stakeholders, from competent authorities to commercial and recreational users [87]. The recent Dynamic Ocean Management (DOM) tool integrates near real-time covariates (new biological, oceanographic, economic and/ or social data) to refine spatial and temporal scales of management areas [88, 89]. DOM could be a valuable tool to improve the protection of large whales and ensure the sustainability of marine activities in the Azores archipelago.

4.6. REFERENCES

1. Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, et al. Novel methods improve prediction of species' distributions from occurrence data. Ecography. 2006;29(2):129-51. doi: doi/epdf/10.1111/j.2006.0906-7590.04596.x.
2. Mannocci L, Roberts JJ, Pedersen EJ, Halpin PN. Geographical differences in habitat relationships of cetaceans across an ocean basin. Ecography. 2020;43(8):1250-9.
3. Araújo MB, Pearson RG, Thuiller W, Erhard M. Validation of species–climate impact models under climate change. Global change biology. 2005;11(9):1504-13. doi: 10.1111/j.1365-2486.2005.001000.x.
4. Gaul W, Sadykova D, White HJ, Leon-Sanchez L, Caplat P, Emmerson MC, et al. Data quantity is more important than its spatial bias for predictive species distribution modelling. PeerJ. 2020;8:e10411. doi: 10.7717/peerj.10411.
5. Brodie SJ, Thorson JT, Carroll G, Hazen EL, Bograd S, Haltuch MA, et al. Trade-offs in covariate selection for species distribution models: a methodological comparison. Ecography. 2020;43(1):11-24. doi: https://doi.org/10.1111/ecog.04707.
6. Becker E, Forney K, Ferguson M, Foley D, Smith R, Barlow J, et al. Comparing California Current cetacean–habitat models developed using in situ and remotely sensed sea surface temperature data. Marine Ecology Progress Series. 2010;413:163-83. doi: 10.3354/meps08696.
7. Mannocci L, Laran S, Monestiez P, Dorémus G, Van Canneyt O, Watremez P, et al. Predicting top predator habitats in the Southwest Indian Ocean. Ecography. 2014;37(3):261-78. doi: doi: 10.1111/j.1600-0587.2013.00317.x.
8. Praca E, Gannier A, Das K, Laran S. Modelling the habitat suitability of cetaceans: example of the sperm whale in the northwestern Mediterranean Sea. Deep Sea Research Part I: Oceanographic Research Papers. 2009;56(4):648-57. doi: doi:10.1016/j.dsr.2008.11.001.
9. Prieto R, Tobeña M, Silva MA. Habitat preferences of baleen whales in a mid-latitude habitat. Deep Sea Research Part II: Topical Studies in Oceanography. 2017;141:155-67. doi: https://doi.org/10.1016/j.dsr2.2016.07.015.
10. Tobeña M, Prieto R, Machete M, Silva MA. Modeling the potential distribution and richness of cetaceans in the Azores from fisheries observer program data. Frontiers in Marine Science. 2016;3:202. doi: doi: 10.3389/fmars.2016.00202.

11. Correia AM, Gil Á, Valente RF, Rosso M, Sousa-Pinto I, Pierce GJ. Distribution of cetacean species at a large scale-Connecting continents with the Macaronesian archipelagos in the eastern North Atlantic. Diversity Distributions. 2020;26(10):1234-47. doi: https://doi.org/10.1111/ddi.13127.

12. González García L, Pierce GJ, Autret E, Torres-Palenzuela JM. Multi-scale habitat preference analyses for Azorean blue whales. PloS one. 2018;13(9):e0201786. doi: 10.1371/journal.pone.0201786.

13. Valente R, Correia AM, Gil Á, González García L, Sousa-Pinto I. Baleen whales in Macaronesia: occurrence patterns revealed through a bibliographic review. Mammal Review. 2019;49(2):129-51. doi: https://doi.org/10.1111/mam.12148.

14. Torres LG, Read AJ, Halpin P. Fine-scale habitat modeling of a top marine predator: Do prey data improve predictive capacity. Ecological Applications. 2008;18(7):1702-17. doi: https://doi.org/10.1890/07-1455.1.

15. Fiedler PC, Redfern JV, Forney KA, Palacios DM, Sheredy C, Rasmussen K, et al. Prediction of large whale distributions: a comparison of presence–absence and presence-only modeling techniques. Frontiers in Marine Science. 2018;5:419. doi: 10.3389/fmars.2018.00419.

16. Lambert C, Mannocci L, Lehodey P, Ridoux V. Predicting cetacean habitats from their energetic needs and the distribution of their prey in two contrasted tropical regions. PloS one. 2014;9(8):e105958. doi: doi:10.1371/journal.pone.0105958.

17. Pendleton DE, Sullivan PJ, Brown MW, Cole TV, Good CP, Mayo CA, et al. Weekly predictions of North Atlantic right whale Eubalaena glacialis habitat reveal influence of prey abundance and seasonality of habitat preferences. Endangered Species Research. 2012;18(2):147-61. doi: https://doi.org/10.3354/esr00433

18. Barlow DR, Bernard KS, Escobar-Flores P, Palacios DM, Torres LG. Links in the trophic chain: modeling functional relationships between in situ oceanography, krill, and blue whale distribution under different oceanographic regimes. Marine Ecology Progress Series. 2020;642:207-25. doi: https://doi.org/10.3354/meps13339.

19. Friedlaender AS, Halpin PN, Qian SS, Lawson GL, Wiebe PH, Thiele D, et al. Whale distribution in relation to prey abundance and oceanographic processes in shelf waters of the Western Antarctic Peninsula. Marine Ecology Progress Series. 2006;317:297-310. doi: 10.3354/meps317297.

20. Rockwood RC, Elliott ML, Saenz B, Nur N, Jahncke J. Modeling predator and prey hotspots: Management implications of baleen whale co-occurrence with krill in Central California. PLoS ONE 16(8): e0255667. 2020;15(7):e0235603. doi: https://doi.org/10.1371/journal.pone.0235603.

21. Virgili A, Hedon L, Authier M, Calmettes B, Claridge D, Cole T, et al. Towards a better characterisation of deep-diving whales' distributions by using prey distribution model outputs? . PLoS ONE 16(8): e0255667 2021. doi: https://doi.org/10.1371/journal.pone.0255667.

22. Kawakami T. A review of sperm whale food. Sci Rep Whales Res Inst. 1980;32:199-218.

23. Santos M, Pierce G, Boyle P, Reid R, Ross H, Patterson I, et al. Stomach contents of sperm whales Physeter macrocephalus stranded in the North Sea 1990-1996. Marine Ecology Progress Series. 1999;183:281-94. doi: 10.3354/meps183281.

24. Clarke M, Martins H, Pascoe P. The diet of sperm whales (Physeter macrocephalus Linnaeus 1758) off the Azores. Philosophical Transactions of the Royal

Society of London Series B: Biological Sciences. 1993;339(1287):67-82. doi: https://doi.org/10.1098/rstb.1993.0005.

25. Flinn RD, Trites AW, Gregr EJ, Perry RI. Diets of fin, sei, and sperm whales in British Columbia: an analysis of commercial whaling records, 1963–1967. Marine Mammal Science. 2002;18(3):663-79. doi: https://doi.org/10.1111/j.1748-7692.2002.tb01065.x.

26. Oliveira CI. Behavioural ecology of the sperm whale (Physeter macrocephalus) in the North Atlantic Ocean 2014.

27. Gavrilchuk K, Lesage V, Ramp C, Sears R, Bérubé M, Bearhop S, et al. Trophic niche partitioning among sympatric baleen whale species following the collapse of groundfish stocks in the Northwest Atlantic. Marine Ecology Progress Series. 2014;497:285-301. doi: 10.3354/meps10578.

28. McQuinn IH, Gosselin J, Bourassa M, Mosnier A, St-Pierre J, Plourde S, et al. The spatial association of blue whales (Balaenoptera musculus) with krill patches (Thysanoessa spp. and Meganyctiphanes norvegica) in the estuary and northwestern Gulf of St. Lawrence. Canadian Science Advisory Secretariat (CSAS): Fisheries and Oceans Canada, Ecosystems and Oceans Science; 2016.

29. Acevedo-Gutiérrez A, Croll D, Tershy B. High feeding costs limit dive time in the largest whales. Journal of Experimental Biology. 2002;205(12):1747-53. doi: https://doi.org/10.1242/jeb.205.12.1747.

30. Gendron Laniel D. Relaciones entre la abundancia de eufáusidos y de ballenas azules Balaenoptera musculus en el Golfo de California: Instituto Politécnico Nacional. Centro Interdisciplinario de Ciencias Marinas; 1990.

31. Lehodey P, Conchon A, Senina I, Domokos R, Calmettes B, Jouanno J, et al. Optimization of a micronekton model with acoustic data. ICES Journal of Marine Science. 2014;72(5):1399-412.

32. Romagosa M, Lucas C, Pérez-Jorge S, Tobeña M, Lehodey P, Reis J, et al. Differences in regional oceanography and prey biomass influence the presence of foraging odontocetes at two Atlantic seamounts. Marine Mammal Science. 2020;36(1):158-79. doi: 10.1111/mms.12626.

33. Pérez-Jorge S, Tobeña M, Prieto R, Vandeperre F, Calmettes B, Lehodey P, et al. Environmental drivers of large-scale movements of baleen whales in the mid-North Atlantic Ocean. Diversity Distributions. 2020;26(6):683-98.

34. Silva MA, Prieto R, Cascão I, Seabra MI, Machete M, Baumgartner MF, et al. Spatial and temporal distribution of cetaceans in the mid-Atlantic waters around the Azores. Marine Biology Research. 2014;10(2):123-37. doi: https://doi.org/10.1080/17451000.2013.793814.

35. Silva MA, Feio R, Prieto R, Gonçalves J, Santos R. Interactions between cetaceans and the tuna fishery in the Azores. Marine Mammal Science. 2002;18(4):893-901.

36. Lehodey P, Murtugudde R, Senina I. Bridging the gap from ocean models to population dynamics of large marine predators: a model of mid-trophic functional groups. Progress in Oceanography. 2010;84(1-2):69-84.

37. Houghton L, Ramirez-Martinez N, Mikkelsen B, Víkingsson G, Gunnlaugsson T, Øien N, et al. Oceanic drivers of sei whale distribution in the North Atlantic. NAMMCO Scientific Publications. 2019;11.

38. Amante C, Eakins BW. ETOPO1 arc-minute global relief model: procedures, data sources and analysis. 2009.

39. Roberts JJ, Best BD, Mannocci L, Fujioka E, Halpin PN, Palka DL, et al. Habitat-based cetacean density models for the US Atlantic and Gulf of Mexico. Scientific reports. 2016;6(1):1-12.

40. Chelton DB, Schlax MG, Samelson RM. Global observations of nonlinear mesoscale eddies. Progress in oceanography. 2011;91(2):167-216. doi: https://doi.org/10.1016/j.pocean.2011.01.002.

41. Lehodey P, Senina I, Murtugudde RJPiO. A spatial ecosystem and populations dynamics model (SEAPODYM)–Modeling of tuna and tuna-like populations. 2008;78(4):304-18. doi: 10.1016/j.pocean.2008.06.004.

42. Croll DA, Marinovic B, Benson S, Chavez FP, Black N, Ternullo R, et al. From wind to whales: trophic links in a coastal upwelling system. Marine Ecology Progress Series. 2005;289:117-30.

43. Mendes S, Newton J, Reid RJ, Frantzis A, Pierce GJ. Stable isotope profiles in sperm whale teeth: variations between areas and sexes. Journal of the Marine Biological Association of the United Kingdom. 2007;87(2):621-7. doi: 10.1007/s00442-006-0612-z.

44. Jaquet N, Whitehead H. Scale-dependent correlation of sperm whale distribution with environmental features and productivity in the South Pacific. Marine ecology progress series. 1996;135:1-9.

45. Becker EA, Carretta JV, Forney KA, Barlow J, Brodie S, Hoopes R, et al. Performance evaluation of cetacean species distribution models developed using generalized additive models and boosted regression trees. Ecology evolution. 2020;10(12):5759-84. doi: 10.1002/ece3.6316.

46. Shono H. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research. 2008;93(1-2):154-62. doi: 10.1016/j.fishres.2008.03.006.

47. Peel D, Bravington M, Kelly N, Wood SN, Knuckey I. A model-based approach to designing a fishery-independent survey. Journal of agricultural, biological, environmental statistics. 2013;18(1):1-21. doi: 10.1007/s13253-012-0114-x.

48. Harrell Jr FE, Harrell Jr MFE. Package ‘hmisc’. CRAN. 2019;2019:235-6.

49. Wood SN. Introducing GAMs. In: Group TF, editor. Generalized additive models: an introduction with R. 4. Second ed. A Chapman and Hall Book: CRC press; 2017. p. 161-320.

50. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B. 2011;73(1):3-36. doi: https://doi.org/10.1111/j.1467-9868.2010.00749.x.

51. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996;15(4):361-87.

52. Team RC. R: A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Austria.; 2021.

53. Wood SN. mgcv. 1.8-34 ed. Journal of the Royal Statistical Society (B), 73(1), 3-36.2011.

54. Amorim P, Perán AD, Pham CK, Juliano M, Cardigos F, Tempera F, et al. Overview of the Ocean Climatology and Its Variability in the Azores Region of the North Atlantic Including Environmental Characteristics at the Seabed. Frontiers in Marine Science. 2017;4. doi: 10.3389/fmars.2017.00056.

55. Martins A, Amorim A, Figueiredo M, Souza R, Mendonça A, Bashmachnikov I, et al., editors. Sea surface temperature (AVHRR, MODIS) and ocean colour (MODIS) seasonal and interannual variability in the Macaronesian islands of Azores, Madeira, and Canaries. Remote Sensing of the Ocean, Sea Ice, and Large Water Regions 2007; 2007: International Society for Optics and Photonics.

56. Scales KL, Hazen EL, Jacox MG, Edwards CA, Boustany AM, Oliver MJ, et al. Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. Ecography. 2017;40(1):210-20. doi: 10.1111/ecog.02272.

57. Soultan A, Safi K. The interplay of various sources of noise on reliability of species distribution models hinges on ecological specialisation. PLoS ONE 16(8): e0255667. 2017;12(11):e0187906.

58. Gerrodette T, Perryman WL, Oedekoven CS. Accuracy and precision of dolphin group size estimates. Marine Mammal Science. 2019;35(1):22-39.

59. Whitehead H. Formations of foraging sperm whales, Physeter macrocephalus, off the Galapagos Islands. Canadian journal of zoology. 1989;67(9):2131-9.

60. Smultea MA. Behavioral ecology of cetaceans in the Southern California Bight 2016.

61. Weilgart L, Whitehead H. Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales. Behavioral Ecology Sociobiology. 1997;40(5):277-85.

62. Forney KA, Ferguson MC, Becker EA, Fiedler PC, Redfern JV, Barlow J, et al. Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. Endangered Species Research. 2012;16(2):113-33. doi: 10.3354/esr00393.

63. Forney KA, Becker EA, Foley DG, Barlow J, Oleson EM. Habitat-based models of cetacean density and distribution in the central North Pacific. Endangered Species Research. 2015;27(1):1-20. doi: 10.3354/esr00632.

64. Tepsich P, Schettino I, Atzori F, Azzolin M, Campana I, Carosso L, et al. Trends in summer presence of fin whales in the Western Mediterranean Sea Region: new insights from a long-term monitoring program. PeerJ. 2020;8:e10544. doi: 10.7717/peerj.10544.

65. Szesciorka AR, Ballance LT, Širović A, Rice A, Ohman MD, Hildebrand JA, et al. Timing is everything: Drivers of interannual variability in blue whale migration. Scientific reports. 2020;10(1):1-9. doi: 10.1121/1.5136714

66. Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, Bograd SJ, et al. Tracking apex marine predator movements in a dynamic ocean. Nature. 2011;475(7354):86-90. doi: 10.1038/nature10082.

67. Oliveira C, Wahlberg M, Silva MA, Johnson M, Prieto R, Wisniewska DM, et al., editors. Foraging behaviour of the sperm whale in the Azores. European Cetacean Society Conference; 2016 30 of March of 2016; Madeira2016.

68. Wong SN, Whitehead H. Seasonal occurrence of sperm whales (Physeter macrocephalus) around Kelvin Seamount in the Sargasso Sea in relation to oceanographic processes. Deep Sea Research Part I: Oceanographic Research Papers. 2014;91:10-6.

69. Morato T, Varkey DA, Damaso C, Machete M, Santos M, Prieto R, et al. Evidence of a seamount effect on aggregating visitors. Marine Ecology Progress Series. 2008;357:23-32. doi: 10.3354/meps07269.

70. Prieto R, Silva MA, Waring GT, Gonçalves JM. Sei whale movements and behaviour in the North Atlantic inferred from satellite telemetry. Endangered Species Research. 2014;26(2):103-13. doi: 10.3354/esr00630.

71. Silva MA, Borrell A, Prieto R, Gauffier P, Bérubé M, Palsbøl PJ, et al. Stable isotopes reveal winter feeding in different habitats in blue, fin and sei whales migrating through the Azores. Royal Society open science. 2019;6(8):181800. doi: https://doi.org/10.1098/rsos.181800.

72. Leathwick J, Elith J, Hastie T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological modelling. 2006;199(2):188-96. doi: https://doi.org/10.1016/j.ecolmodel.2006.05.022.

73. Visser F, Hartman KL, Pierce GJ, Valavanis VD, Huisman J. Timing of migratory baleen whales at the Azores in relation to the North Atlantic spring bloom. Marine Ecology Progress Series. 2011;440:267-79. doi: doi: 10.3354/meps09349.

74. Silva MA, Prieto R, Jonsen I, Baumgartner MF, Santos RS. North Atlantic blue and fin whales suspend their spring migration to forage in middle latitudes: building up energy reserves for the journey? PloS one. 2013;8(10):e76507. doi: 10.1371/journal.pone.0076507.

75. Silva M, Magalhães S, Prieto R, Quérouil S, Pinela A, Seabra M, et al. Ecologia e estrutura populacional dos roazes e cachalotes nos Açores: relação com as características do habitat. Arquivos DOP Á Série Estudos. 2006;4.

76. Boys RM, Oliveira C, Pérez-Jorge S, Prieto R, Steiner L, Silva MA. Multi-state open robust design applied to opportunistic data reveals dynamics of wide-ranging taxa: the sperm whale case. Ecosphere. 2019;10(3):e02610.

77. Whitehead H. Variation in the feeding success of sperm whales: temporal scale, spatial scale and relationship to migrations. Journal of Animal Ecology. 1996:429-38. doi: https://doi.org/10.2307/5778

78. Whitehead H. Sperm whales: social evolution in the ocean. first ed: University of Chicago press; 2003.

79. Engelhaupt D, Rus Hoelzel A, Nicholson C, Frantzis A, Mesnick S, Gero S, et al. Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (Physeter macrocephalus). Molecular Ecology. 2009;18(20):4193-205. doi: 10.1111/j.1365-294X.2009.04355.x.

80. Skov H, Gunnlaugsson T, Budgell W, Horne J, Nøttestad L, Olsen E, et al. Small-scale spatial variability of sperm and sei whales in relation to oceanographic and topographic features along the Mid-Atlantic Ridge. Deep Sea Research Part II: Topical Studies in Oceanography. 2008;55(1-2):254-68. doi: 10.1016/j.dsr2.2007.09.020.

81. Baylis AM, Tierney M, Orben RA, Warwick-Evans V, Wakefield E, Grecian WJ, et al. Important at-sea areas of colonial breeding marine predators on the southern Patagonian shelf. Scientific reports. 2019;9(1):1-13. doi: https://doi.org/10.1038/s41598-019-44695-1.

82. Hastie TJ, Tibshirani RJ. Generalized Additive Models. 1st ed. Routledge, editor1990.

83. Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AI, et al. Predicting species distributions for conservation decisions. Ecology letters. 2013;16(12):1424-35. doi: 10.1111/ele.12189.

84. Proud R, Handegard NO, Kloser RJ, Cox MJ, Brierley AS. From siphonophores to deep scattering layers: uncertainty ranges for the estimation of global mesopelagic fish biomass. ICES Journal of Marine Science. 2019;76(3):718-33. doi: 10.1093/icesjms/fsy037.

85. Hooker SK, Cañadas A, Hyrenbach KD, Corrigan C, Polovina JJ, Reeves RR. Making protected area networks effective for marine top predators. Endangered Species Research. 2011;13(3):203-18. doi: 10.3354/esr00322.

86. Agardy T, Bridgewater P, Crosby MP, Day J, Dayton PK, Kenchington R, et al. Dangerous targets? Unresolved issues and ideological clashes around marine protected areas. Aquatic conservation: marine freshwater ecosystems. 2003;13(4):353-67. doi: 10.1002/aqc.583.

87. O'Leary BC, Brown RL, Johnson DE, von Nordheim H, Ardron J, Packeiser T, et al. The first network of marine protected areas (MPAs) in the high seas: the process, the challenges and where next. Marine Policy. 2012;36(3):598-605. doi: 10.1016/j.marpol.2011.11.003.

88. Hazen EL, Scales KL, Maxwell SM, Briscoe DK, Welch H, Bograd SJ, et al. A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. Science advances. 2018;4(5):eaar3001. doi: 10.1126/sciadv.aar3001.

89. Maxwell SM, Hazen EL, Lewison RL, Dunn DC, Bailey H, Bograd SJ, et al. Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. Marine Policy. 2015;58:42-50. doi: http://dx.doi.org/10.1016/j.marpol.2015.03.014.

CHAPTER 5

5. ANTHROPOGENIC ACTIVITIES INFLUENCE IN THE DISTRIBUTION OF CETACEANS

5.1. INTRODUCTION

Anthropogenic activities can have adverse effects on marine wildlife [1-4]. Shipping traffic represents one of the lethal threats for marine wildlife and is regarded with increasing concern [5, 6]. There have been reports of ship collisions with 75 marine species [7]. Among these, large whales are particularly susceptible to mortality or suffering severe injuries from ship collisions [8-11], due to their large body size [12, 13], limited ability to manoeuvre [14, 15], and extended time near the surface [16, 17].

The fin whale (*Balaenoptera physalus*) is the species most commonly struck by shipping traffic worldwide, and the sperm whale (*Physeter macrocephalus*) is in the top list of cetaceans killed by collision, followed by blue *(B. musculus*) and sei whales (*Balaenoptera borealis*) [18, 19]. Despite the difficulties in monitoring cetacean strandings and identifying the causes of these events, there are well documented cases of large whale deaths and injuries from ship strikes in some areas. Of 35 sperm whales stranded in the Canary Islands from 1991 to 2017, 60% were confirmed fatal victims of collisions with ships [20]. About 15% of the 81 sperm whales stranded in the western Mediterranean Sea from 2005 to 2017 had been struck by ships [21]. More than 10% of fin whale strandings in the western Mediterranean Sea between 2005 and 2017 were caused by ship collisions [21]. In Washington, five of the seven fin whale strandings from 1980-2006 were confirmed as fatal collisions with ships [22]. Although there is less data on sei and blue whales, there are records of two sei whales killed by ships in the Canary Islands [23] and Washington [22], two blue whales in California [24], and another two in Washington [22]. In Sri Lanka, nine blue whales out of 20 stranded animals (from 2010-2014) were killed by collisions [25]. Ship collisions can hinder the recovery of populations that have been severely depleted by commercial whaling [26, 27] and are still considered as Endangered [28, 29] or Vulnerable [30-32]. Nevertheless, the actual number of whales killed by ship strikes each year is unknown because the probability of detecting these events is low [18, 19, 22, 26, 33].

Information on ship collisions with cetaceans comes either from observations of cetaceans (dead or alive) with injuries [34], logbook records of ship damages, or from governmental or non-governmental databases, such as the International Whaling Commission (IWC) Global Ship Strikes Database [35]. Ship collisions with large whales are occasionally reported to the authorities by the ships involved. Other than this, traces of ship collisions are difficult to detect because most interactions happen in remote places, far away from land and without witnesses. Marks, cuts and amputations caused by propellers or the ship hull on live cetaceans are sometimes recorded by researchers [36]. Less commonly, individuals bearing these typical injuries are found stranded, usually dead [34]. While there are protocols to identify traumas caused by ship collisions, these are often challenging to follow due to the advanced state of decomposition of carcasses, lack of conditions to perform a thorough examination of the carcass or lack of trained personnel. In the case of fatal collisions, the carcasses can rapidly sink and go undetected [7]. These circumstances make it difficult to accurately assess the number of collisions [37] and identify areas or time periods where these events are persistent.

Despite all these limitations, there is evidence of increased cetacean strandings related to the rise in shipping traffic worldwide [24, 38]. In the Azores Archipelago, the Regional Stranding Network (RACA) collects information on cetacean strandings since 1990 [39]. RACA database includes approximately 450 stranding records, including 73 sperm whales, eight fin whales, four sei whales, and no record of blue whales. Only seven individuals were registered as victims of collision with ships, four sperm whales, one fin whale, one minke whale (*Balaenoptera acutorostrata*) and one common dolphin (*Delphinus delphis*). The cause of death has only rarely been identified, but the number of reports of ship collisions with large whales has increased in recent years.

Human activities that have non-lethal impacts, like wildlife tourism [40-45], are usually seen as a minor threat to population viability. The cetacean watching industry has grown exponentially [46, 47] and continues to increase in the number of boats, places and number of days. The activity can contribute to the economy of small communities, help raise ocean literacy and awareness [48, 49], and provide opportunities for scientific research. Nevertheless, non-lethal detrimental effects are evident [42], and cetacean watching is now viewed as a sub-lethal consumptive industry.

Presence and noise from cetacean watching boats can induce avoidance behaviours [50, 51], such as changes in respiration, swimming speed and direction, diving behaviour, disrupt critical activities like foraging, resting or socialising, and increase stress levels [52]. Noise from boats can mask important sounds and reduce the functional acoustic ranges of animals, interfering with their ability to receive and transmit acoustic information that is essential for communication, navigation, prey and predator detection [53, 54]. When these short-term effects occur repeatedly over time, such as with the case of populations residing in cetacean watching areas that are routinely exposed to the activity, they may lower individual fitness, reproductive and foraging success, ultimately impacting population dynamics [41].

The cetacean watching in the Azores archipelago has increased exponentially since the early 90s [55]. The main target are the sperm whales, but the activity also focuses on baleen whales during their spring migration [56]. Little is known about the long-term effects of the activity on these species in the Azores but several short-term changes consistent with increased avoidance have been documented in sperm whales [57, 58]. In addition, sperm whales using the Azores also occur in Norway (males) [59], Gulf of Mexico or Bahamas [60], Madeira and Canary Islands (females) [61]. In all these regions there is a cetacean watching industry [62-64] and individuals may, therefore, be exposed in multiple areas within their range.

The growing recognition of the need to address both lethal and sub-lethal impacts on cetacean populations, has led to the implementation of various management initiatives in different areas. Measures to reduce ship collisions with cetaceans include the use of lookouts or acoustic recorders for whales [65], restriction of ship speed [66], fixed or seasonal modification of maritime routes, depending on the distribution and abundance of vulnerable species. Management of cetacean watching activities has often been based on site- or time-specific restrictions of the number of boats, as well as on regulations about the boats (e.g., changing from outboard to jet engines) or their behaviour near the animals (e.g., speed, movement direction) [45, 67].

In the Azores, there are no specific regulations to reduce risks from shipping traffic on cetaceans. In August 2021, the Megafauna Observation Program in Local Transportation (POMET) has been implemented to place observers aboard ferries but these observers do not have the authority to impose any changes in traffic. Conversely, cetacean watching regulations have been in place since 1999 with the goal of protecting cetaceans from disturbance [68-73]. Measures enforced include limitations in the number of boat licenses for different zones of the archipelago, limitations in the number of boats around each group of animals, duration of the observation period, minimum distance, speed and the manoeuvring around these animals. Nevertheless, a key impediment to developing effective management strategies in the Azores has been the lack of information on the distribution of the activities and cetaceans. This information is critical to quantify the frequency of potential interactions of cetaceans with shipping and cetacean watching activities, and identify areas and times of year where these interactions are more likely to occur.

This chapter addresses this gap by assessing the influence of shipping traffic and cetacean watching activities on the distribution of sperm, fin, blue and sei whales in the Azores archipelago, and quantifying the risk of exposure of these species to these activities. First, habitat models developed for these species in Chapter 4 using environmental and simulated prey variables were used as a basis to develop models incorporating also variables describing shipping traffic and cetacean watching activity. The relative density of each species was therefore modelled as a function of anthropogenic, environmental and prey variables, to evaluate the importance of anthropogenic activities on species distribution. Ecological relationships are complex, so the distribution of animals is often related to environmental factors and prey but can also be affected by anthropogenic factors. The anthropogenic impact is evaluated by incorporating their variables into the model and analysing whether they are relevant to the species. It can also be analysed if they have more explanatory power than the traditional variables. Then, the best fitting models were used to assess the risk of lethal collision and the exposure to cetacean watching. Model-base predictions of whale density were overlapped with maps of shipping density and cetacean watching effort to identify areas of risk of lethal collisions and quantify the proportion of the suitable habitat with a high risk of exposure to the cetacean watching activity. These models and analyses can inform marine ecosystem conservation and management of anthropogenic activities in the waters around the Azores archipelago.

5.2. METHODS

5.2.1. THE STUDY AREA

The description of the oceanographic and geophysical characteristics of the marine environment of the Azores is given in Chapter 1. The study area lies within 36ºN 24ºW and 40ºN 32ºW, where most of the data analysed in this chapter was available (Figure 5.1).

5.2.2. CETACEAN SURVEY DATA

Cetacean survey data from the POPA programme used in Chapter 4 was subset to the same spatial (~90,000 km^2) extent and temporal period (from October 2012 to October 2016) of the anthropogenic data (Table 5.1). A total of 628 surveys conducted within this period produced 2302 “on-effort” segments. After removing records with missing values and spatial autocorrelation only two cetacean species had enough sightings to develop SDMs: sperm whales (n=135) and fin whales (n=26). Blue and sei whales had less than 10 sightings each. Thus, for the latter species, spatial distribution models with anthropogenic variables could not be developed and overlap with anthropogenic activities was calculated using models with only habitat and prey variables developed in Chapter 4.

*Table 5. 1. Summary of cetacean and anthropogenic activities data. Dataset, type of data; source, source of the cetacean data; period, period of the data; platform type, type of boat; analyses, the anthropogenic activity analysed and cw, cetacean watching. * searching cetaceans ~meters above the water. **cetacean watching boats have at least one person looking for cetaceans.*

Dataset	Source	Period	Platform type	Analyses
Cetacean	POPA	2012-2016	~8m*; 20-m fishing Vessel	Shipping
Shipping density	AIS class A	2012-2016	>80-m	Shipping

Cetacean watching effort**	GPS & AIS class B	2013 & 2018	<15-m motor boat and RHIB	CW

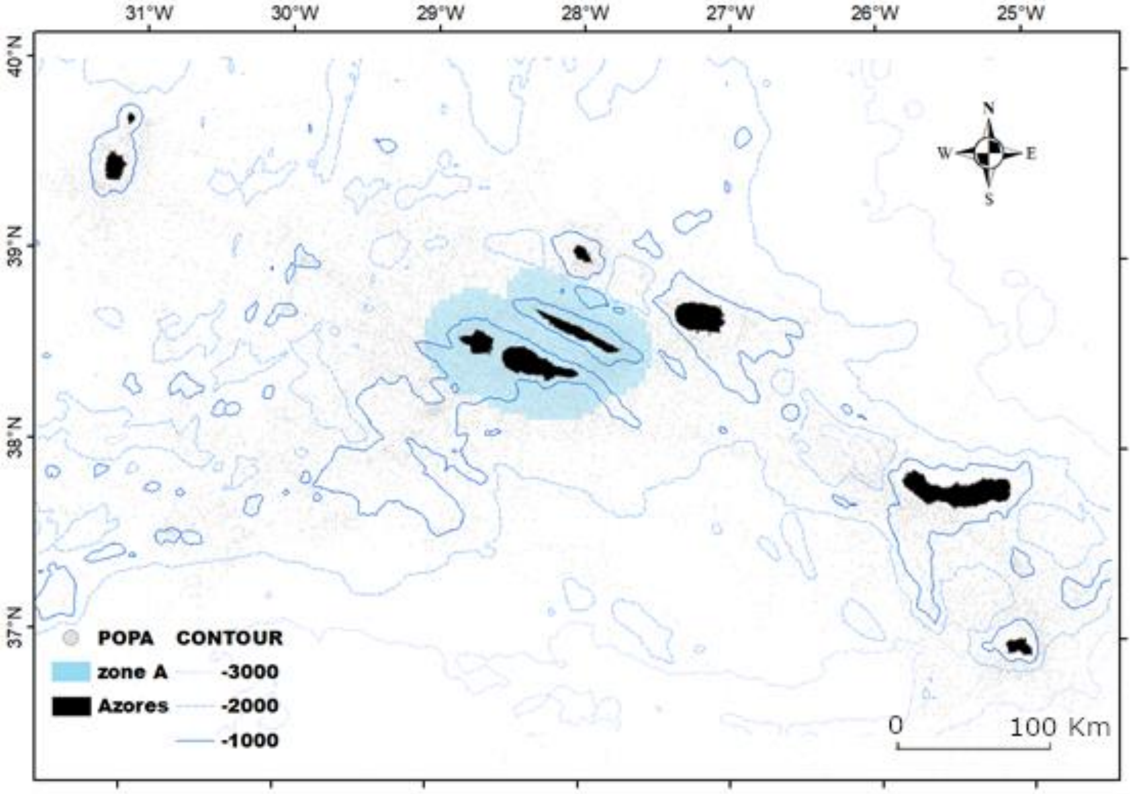


Figure 5. 1. Location of the surveys conducted by the POPA programme (grey circles) from 2012 to 2016 in the Azores archipelago. Zone a (blue) delimits one of the cetacean watching activity areas and was the focus of this study.

5.2.3. ENVIRONMENTAL DATA

Habitat and prey data used in Chapter 4 (Table 5.2) were also subset to have the same spatial and temporal coverage of the anthropogenic data. The non-migratory upper and lower-mesopelagic prey variables were discarded to avoid missing data in coastal and shallower waters where the cetacean watching activity concentrated.

5.2.4. ANTHROPOGENIC DATA

Cetacean watching effort and shipping density data were subset from those used in Chapters 2 and 3, respectively. An annual cetacean watching effort was calculated by 10x10km cells, using only data from boats engaged in cetacean observation as previously explained (Chapter 2). The resulting variable was scaled to the same spatial resolution as the simulated prey variables (Chapter 4, Table 5.3).

The risk of collisions of ships with cetaceans is influenced by the probability of spatial and temporal matches between the two, as well as by the avoidance response of cetaceans [12]. On the other hand, the severity of the collision is related to the ship's speed and the size, with a higher risk of lethal collisions when ships travel at >14 knots, and the highest risk for ships over 80 m cruising at these speeds [19]. Thus, the weekly shipping density was calculated separately for these two categories (Table 5.2). Each variable was then standardised to values ranging from 0 to 1, with 1 representing the maximum weekly shipping density.

Table 5. 2. Candidate anthropogenic variables used in SDMs of sperm whales and fin whales.

Anthropogenic variable	**Acronym**	**Transformation**	**Spatial**	**Description**
Shipping density, "Speed"	14kn	0 -1	0.25 degrees	Weekly shipping density with higher probability of lethal collisions
Shipping density, "Speed and size"	14kn 80m	0 -1	0.25 degrees	Weekly shipping density with highest probability of lethal collisions
Cetacean watching effort*	CWe	0 -1	0.25 degrees	Annual effort of cetacean watching (CWe)
Cetacean watching effort categorical*	CWec	0=absence, 1=presence	0.25 degrees	Transformation of the CWe variable.
Cetacean watching effort by 5 categories*	effwwc4	0-4	0.25	

*Discarded variables.

5.2.5. INFLUENCE OF ANTHROPOGENIC ACTIVITIES ON SPECIES DISTRIBUTION

5.2.5.1. MODEL DEVELOPMENT

Generalised Additive Models (GAMs) [74] were used to investigate the influence of shipping density and cetacean watching density on the relative density of each whale species, while accounting for their habitat preferences described by the environmental and prey variables [75]. Separate species distribution models (SDMs) were built for sperm whales and fin whales using number of individuals as a response variable, and using as predictors the habitat and prey variables described in Chapter 4, and anthropogenic variables described in Table 5.2. Models were developed using data from 2012-2016 .

The GAMs were fitted with a Tweedie distribution to account for overdispersion [76, 77], using the restricted maximum likelihood (REML) to optimise the parameter estimates [78] and adding the natural logarithm of the survey effort as an offset. GAM knots were restricted to 5 degrees of freedom to avoid overfitting and preserving ecological interpretability of functional relationships. The same modelling procedures described in Chapter 4 was used here. Only pairs of covariates with a Pearson's correlation coefficient <0.7 were used in the models. Pearson and concurvity were calculated in R, using the *Hmisc* package [79]. Models were developed using the *mgcv* package in R version 4.1.0 (R Core Team) [80, 81].

5.2.5.2. MODEL SELECTION AND EVALUATION

Model selection was based on the Akaike Information Criterion corrected for small sample sizes (AICc). The AICc assesses how much of the variation in the data can be explained by the predictor variables in the model. The model with the lowest AICc is considered the best-fitting model, explaining the greatest amount of variation with the lowest number of variables [82].
All attempts to include cetacean watching variables (continuous or categorical) resulted in models with a very poor fit to the data. Therefore, these variables had to be discarded from all models.

To evaluate the prediction performance of the models I used a cross-validation process following the methodology described in Chapter 4. For model training, 75% of the data from each species was randomly selected, and the rest (25%) was used for validation (Table 5.3). Models and Relative explained deviance (R.ExpD) were evaluated using the *mgcv* package in R version 4.1.0 [80, 81].

Table 5. 3. Presences and absences used to fit SDMs. SDM-p-es: model with environmental and prey variables. SDM-p-aa: model with environmental, prey, and anthropogenic variables.

Species	**Model**	**TRAIN DATA**		**TEST DATA**	
		Presences	**Absences**	**Presences**	**Absences**
Fin whale	SDM-p-es/aa	18	3477	8	1541
Sperm whale	SDM-p-es/aa	108	3483	27	851

5.2.5.3. MODEL PREDICTIONS

The best fitting models were used to estimate the mean relative density of sperm and fin whales in the study area for 2012-2016. I predicted only within the range of values of the sampled covariates to avoid extrapolating to unknown conditions. Maps of predicted relative density maps represented the number of individuals per 100 km^2. Model predictions were performed in R version 4.1.0 using the *mgcv* package [80, 81].

5.2.6. ASSESSING THE RISK OF EXPOSURE TO ANTHROPOGENIC ACTIVITIES

The risk of exposure of large whales to shipping traffic and cetacean watching was investigated using the relative density of each species, as predicted from the best fitting models, the shipping density of ships >80m at >14kn as calculated by the AIS data (Chapter 3), and the cetacean watching effort calculated in Chapter 2. For blue and sei whales, for which there was not enough data to fit models with anthropogenic variables, the relative density was predicted for the area and period of interest using the best models developed in Chapter 4.

5.2.6.1. RISK OF LETHAL SHIP COLLISIONS

The risk of lethal collisions (RLC_i) was calculated for each grid cell *i* as:

$$RLC_i = WD_i \; x \; SD_i$$

where WD_i is the relative density of each whale species predicted for grid cell *i* by the best SDMs, and SD_i is the weekly shipping density, averaged across the years of study, of ships >80m travelling at >14 kn in that grid cell. RLC_i was mapped for the entire study area to identify areas where the risk of lethal collision was higher for each whale species.

5.2.6.2. EXPOSURE TO CETACEAN WATCHING

To quantify the level of exposure of large whales to cetacean watching (CWE) I used the formula:

$$CWE_i \; = WDi \; x \; CWe_i$$

Where WD_i is the relative density of each whale species per grid cell, as above, and CWe_i is the average cetacean watching effort for that cell. CWE was only calculated for zone A (Faial, Pico and S. Jorge islands) because data on cetacean watching effort was not available for the other zones. Cetacean watching exposure was mapped for each species to identify areas where exposure is higher.

5.2.6.3. ASSESSING THE AMOUNT OF PREFERRED HABITAT EXPOSED TO CETACEAN WATCHING ACTIVITY

I quantified the amount of preferred habitat for each species exposed to cetacean watching in two ways. I identified grid cells with high whale density by determining the percentage of cells that fell in the third quartile of the distribution of density values for each species. Then, I calculated the proportion of high whale density cells within cetacean watching areas (including all zones). Second, I calculated the proportion of cells with high species density that overlapped cells with cetacean watching effort higher than average (only for zone A).

5.3. RESULTS

5.3.1. INFLUENCE OF SHIPPING DENSITY ON THE DISTRIBUTION OF SPERM AND FIN WHALES

The best fitting models for sperm whales incorporating only habitat and prey variables (SDM-p-es model), plus shipping density at high speeds (SDM-p –aa-1), and shipping density at high speeds for large vessels (SDM-p –aa-2) were (Table 5.4):

SDM-p-es model: Sperm whale relative density ~ s(distance to SST fronts) +s(epipelagic production) + s(highly migrant bathypelagic production)

SDM-p –aa-1 model: Sperm whale relative density ~ s(distance to SST fronts) + s(epipelagic production) + s(highly migrant bathypelagic production)+ s(shipping density >14kn)

SDM-p –aa-2 model: Sperm whale relative density ~ s(distance to SST fronts) + s(epipelagic production) + s(highly migrant bathypelagic production)+ s(shipping density >14kn >80m)

The percentage of deviance explained by each model was very similar but slightly higher for the SDM-p-aa-2 model (4.73%), including shipping density for vessels >80m long travelling at high speeds (>14kn) (Table 5.5). This model also showed the lowest AIC and highest C-Index (0.60), although values were also very close to those of the other two models (Table 5.5).

The relationship of sperm whale relative density and each covariate was very similar among the three models, and only outputs of the best model SDM-p-aa-2 are shown below (results for the other models are presented in Tables App.1 and App.2, Figures App.1 and App.2). Relative density of sperm whales increased linearly with the production of epipelagic and highly migrant lower-mesopelagic (although variance was very high for both covariates), with decreasing distance from fronts and with decreasing shipping density for large vessels travelling at high speeds (Figure 5.2 and 5 on top). Sperm whales preferred low traffic areas, relatively close to the ocean mesoscale structures (less than 25km), and with epipelagic production higher than that of highly migrant bathypelagic organisms.

The addition of anthropogenic variables to the habitat model of fin whales did not improve or modify the model performance (Table 5.4 and 6). The model with the lowest AIC included only environmental and prey variables (SDM-p-es), explaining 14.3% of the deviance, and had a C-index of 0.8 (Table 5.5):

SDM-p-es model: Fin whale density ~ s(epipelagic production) + s(Lower trophic level plankton production) + s(temperature)

The relative density of fin whales increased with increasing production of lower trophic level and epipelagic micronekton, as decreased with temperature (Figure 5.3). Fin whales preferred colder waters with higher productivity of phytoplankton and epipelagic organisms (Figure 5.5 on bottom).

*Table 5. 4 Summary of the best fitting models of shipping density for sperm whales. Akaike information criteria, aic or * small-sample equivalent (aicc); concordance index, C-index. Model type abbreviations are as follows: SDM-p-es: environmental and prey variables. SDM-p-aa: environmental, prey and anthropogenic variables. The SDM-p-aa1 used shipping density at >14kn, and sdm-p-aa2 used shipping density at >14kn for ships >80m long.*

Species	**Model name**	**Model type**	**Deviance (%)**	**AIC**	**TRAIN DATA**		**TEST DATA**	
					C-index	SD	C-index	SD
Sperm whale	SDM-p –es	Environmental +SEAPODYM	3.72	3870	0.58	0.005	0.49	0.006
	SDM-p –aa-1	SDM-p-es+Risk14_80kn	4.21	3871	0.59	0.005	0.49	0.006
	SDM-p –aa-2	SDM-p-es+risk14_80kn_80m	4.73	3870	0.60	0.005	0.49	0.006
Fin whale	SDM-p –es	Enviromental +SEAPODYM	14.3	3513	0.80	0.008	0.82	0.1
	SDM-p –aa-1	SDM-p-es+risk14_80kn	14.3	3514	0.80	0.008	0.82	0.1
	SDM-p –aa-2	SDM-p-es+risk14_80kn_80m	14.3	3514	0.80	0.008	0.82	0.1

Table 5. 5. Covariates retained in the best sperm whale model (SDM-p-aa-2). Square root of euclidean distance to sst fronts, Ed_fronts; epipelagic micronekton (pp), epi_pp and highly migrant lower-mesopelagic (pp), hmbathy_pp, and Shipping density, "speed and size", 14-80kn 80m.

Covariates	**edf**	**Ref.df**	**F**	**p-value**	**R.ExpD**
Ed_fronts	0.8223	4	1.071	0.0215	28
epi_pp	0.8165	4	1.140	0.0168	26
hmbathy_pp	0.7909	4	0.860	0.0351	25
14-80kn 80m	0.8336	4	1.142	0.0189	20
Parametric coefficients:	**Estimate**	**SE**	**T value**	**p-value**	
Intercept	-6.7817	0.1206	-56.23	<2e-16	

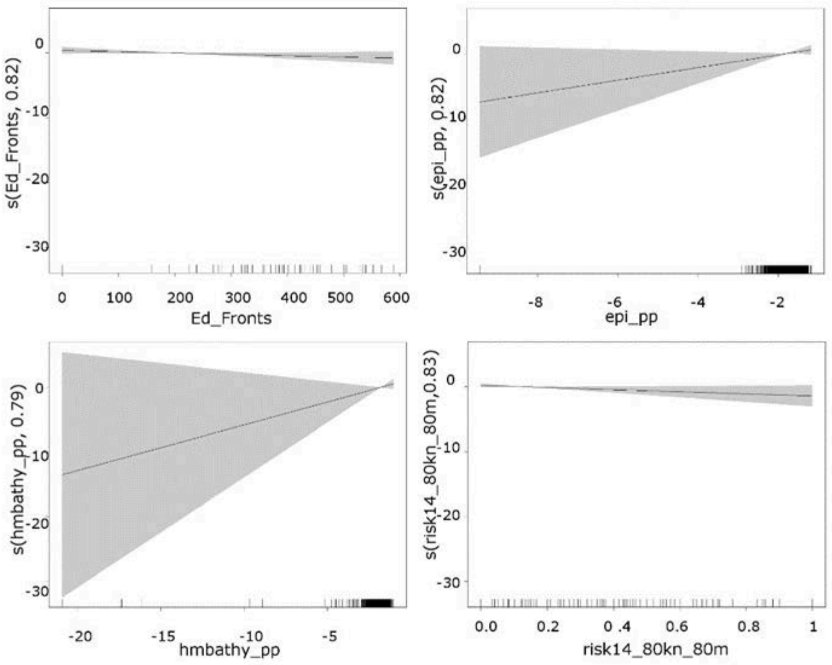


Figure 5. 2. Gam smoothers for the best fitting model (sdm-p-aa-2) for sperm whales: top left) shipping density at >14kn and >80m, top right) square of euclidean distance to fronts (ed_fronts), bottom left) epipelagic micronekton production (epi_pp), and bottom right) highly migrant Lower-mesopelagic production (hmbathy_pp). Smooth functions are shown as a solid line, and shaded regions represent 95% confidence intervals.

Table 5. 6. Covariates retained in the best fin whale model (SDM-p-es). Epipelagic micronekton (pp), epi_pp; Sea water potential temperature, temperature and lower trophic level plankton (pp), pk_pp.

Covariates	**edf**	**Ref.df**	**F**	**p-value**	**R.ExpD**
epi_pp	0.8148	4	1.015	0.0241	34
pk_pp	0.8677	4	1.274	0.0130	33
Temperature	0.8110	4	0.955	0.0285	33
Parametric coefficients:	**Estimate**	**SE**	**T value**	**p-value**	
Intercept	-9.440	0.3432	-26.93	2.16e-10	

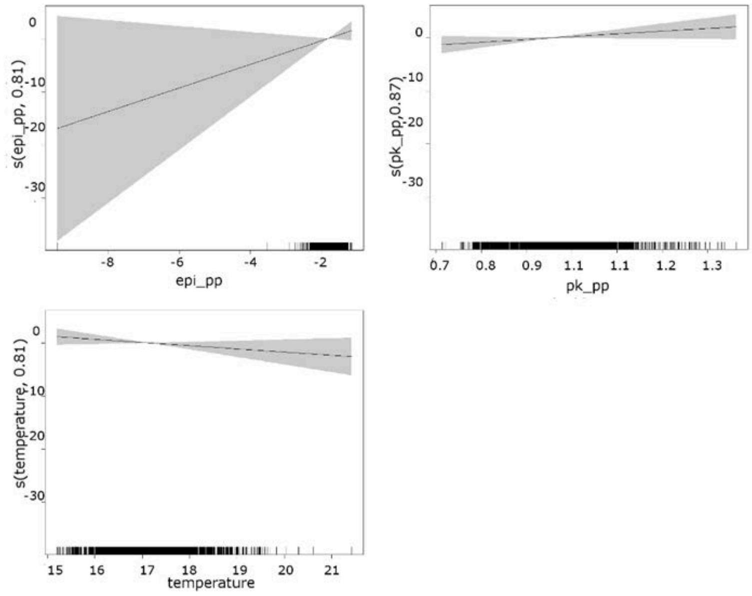


Figure 5. 3. Gam smoothers for fin whales (sdm-p-es) on shipping extent: top left) lower trophic level plankton (pp), pk_pp, top right) epipelagic micronekton (pp), epi_pp, and bottom left) sea water potential temperature, temperature. Smooth functions are shown as a solid line, and shaded regions represent 95% confidence.

Predictions of the relative density of sperm whales averaged for 2012-2016 from the best model (SDM-p-aa2) are shown in Figure 5.4 (on top). Hotspots of sperm whale distribution occurred SW of Faial, along 1000 m isolines, in the area between Graciosa and Terceira islands, and W of S. Miguel Island, along 1000 m isolines. Other important areas also occurred between Terceira and S. Miguel and around the Mid-Atlantic Ridge. Areas with low density were identified SE of S. Jorge and SW of S. Miguel.

The fin whale best model (SDM-p-es) predicted the highest mean relative density in the Mid-Atlantic Ridge, in the area between Terceira and S. Miguel, and SW of Faial, N of Terceira and SW of S. Miguel (Figure 5.4, on bottom). Overall, low densities of fin whales were predicted in the channel Pico-S. Jorge and around Sta. Maria.

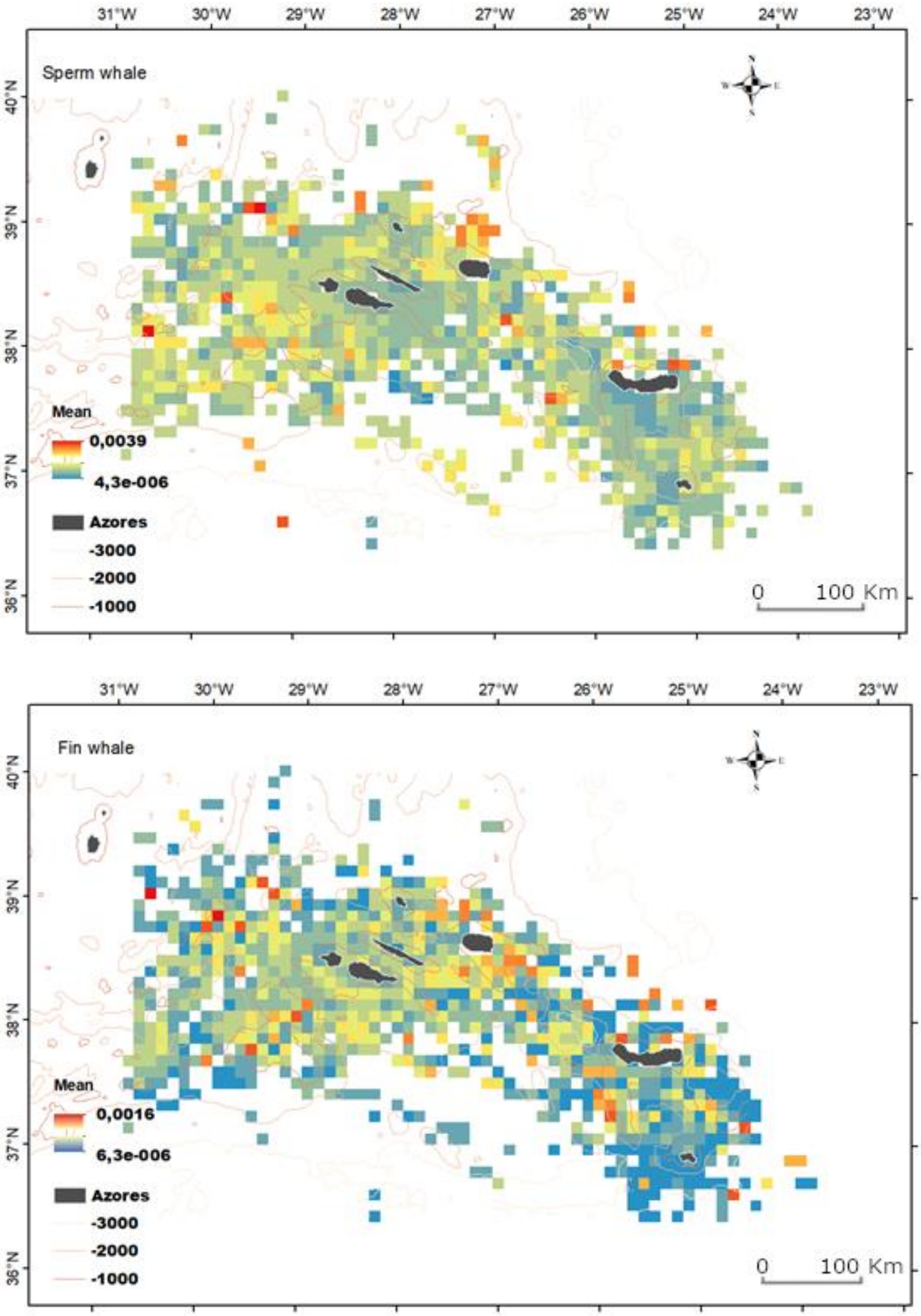
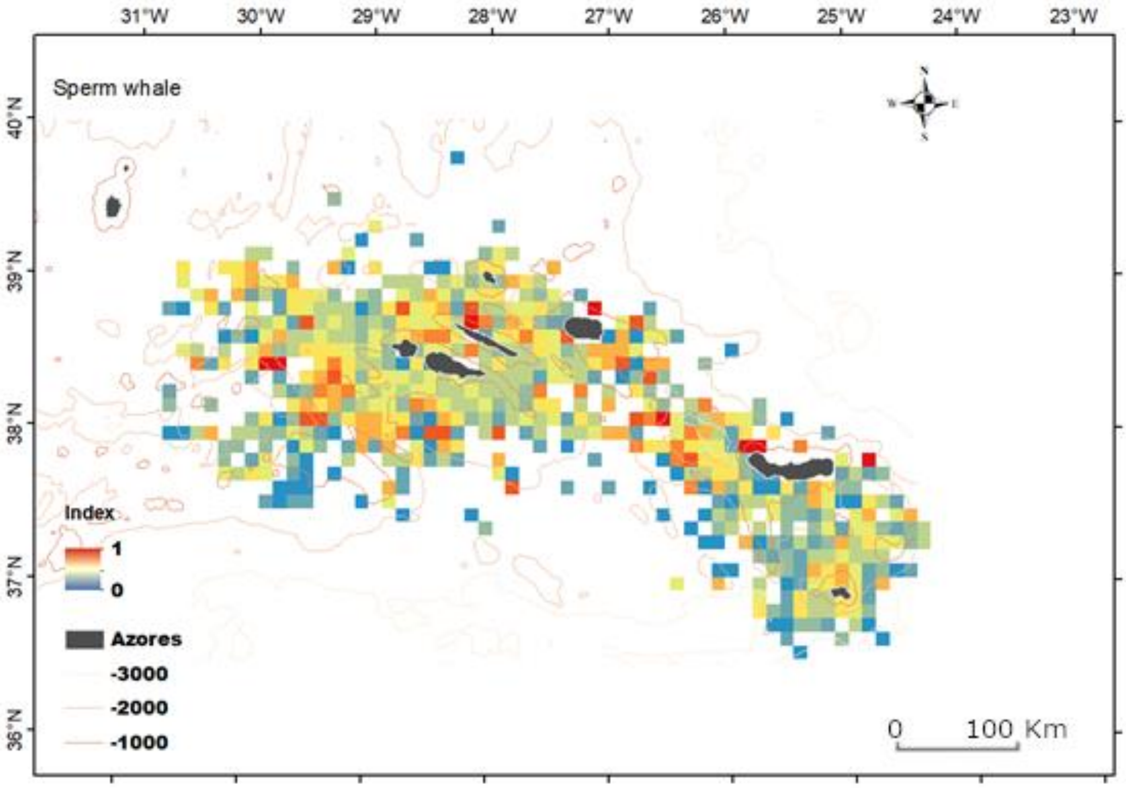


Figure 5. 4. Mean predicted relative density of sperm whales (on top) and fin whales (on bottom) (individuals per 100 km2) for 2012-2016 based on the training dataset for SDM-p-aa2 and SDM-p-es respectively..

5.3.2. RISK OF LETHAL SHIP COLLISIONS

Based on the overlap of the predicted relative density of each whale species and the density of ships >80m travelling at >14kn, sperm whales are at risk of ship collisions in 55% of their distribution area, sei whales in 42%, fin whales in 41%, and blue whales in 14%. These percentages of habitat at risk are not comparable among species because SDMs used to calculate risk for sperm and fin whales predict for a smaller area than those used for blue and sei whales.

Risk of lethal collisions for sperm whales was high and spread nearly homogeneously throughout the study area. The highest risk occurred along the corridor linking S. Miguel and Terceira, and the corridor that crosses the archipelago and passes N of S. Miguel and S of Pico and Faial, as well as between Faial, Pico, S. Jorge, Terceira and Graciosa islands (Figure 5.5 on top). Higher risk of lethal collisions with fin whales was SW of S. Miguel, S of Terceira, in an area over the Mid-Atlantic Ridge between Flores and Faial, and along the shipping route crossing the archipelago. The lowest risk occurred around Sta. Maria (Figure 5.5 on bottom). The blue whale was at higher risk near the shipping route that passes S of the archipelago (S of Sta. Maria), in a large area to the N and W of S. Miguel, and over the Mid-Atlantic Ridge, between Flores and Faial (Figure 5.6 on top). With the exception of the area extending from the S of S. Miguel to Sta Maria, risk of lethal collisions with sei whales was relatively high in most of the archipelago, with hotspots along the corridor linking the eastern and central groups, an area over the Mid-Atlantic Ridge, between Flores and Faial, along the shipping route that passes north of the archipelago, and close to Faial, Pico and S. Jorge (Figure 5.6 on bottom).



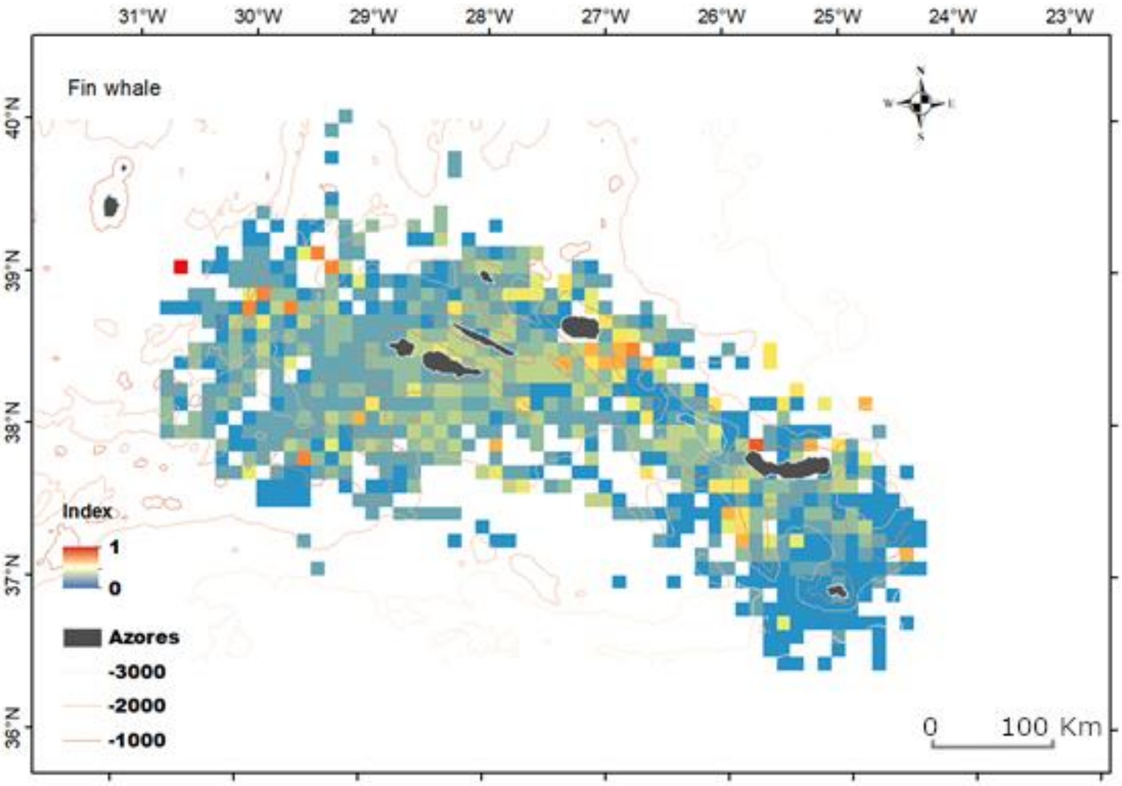
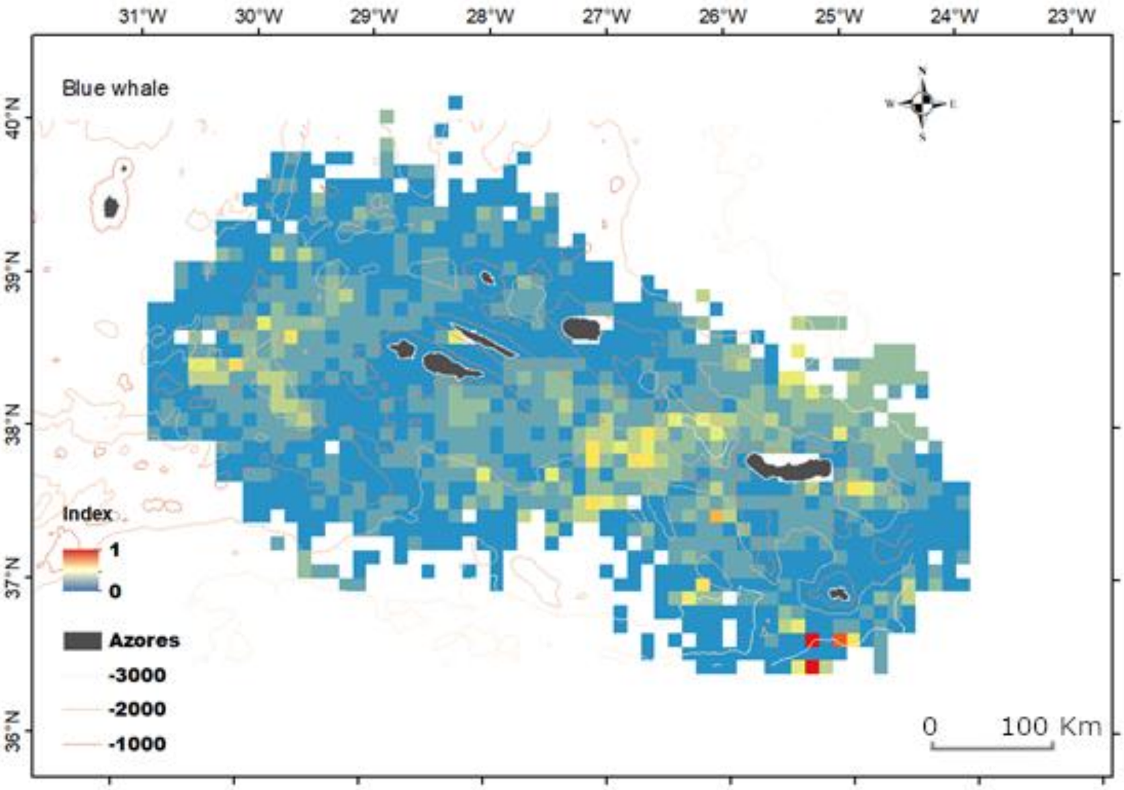


Figure 5. 5. Risk of lethal ship collisions for sperm (on top) and fin whales (on bottom). The risk was calculated using sperm whale predicted mean density and fin whale (individuals per 100 km^2) based on the SDM-p-aa2 model and the SDM-p-es model, respectively, the density of ships >80m travelling at >14kn.



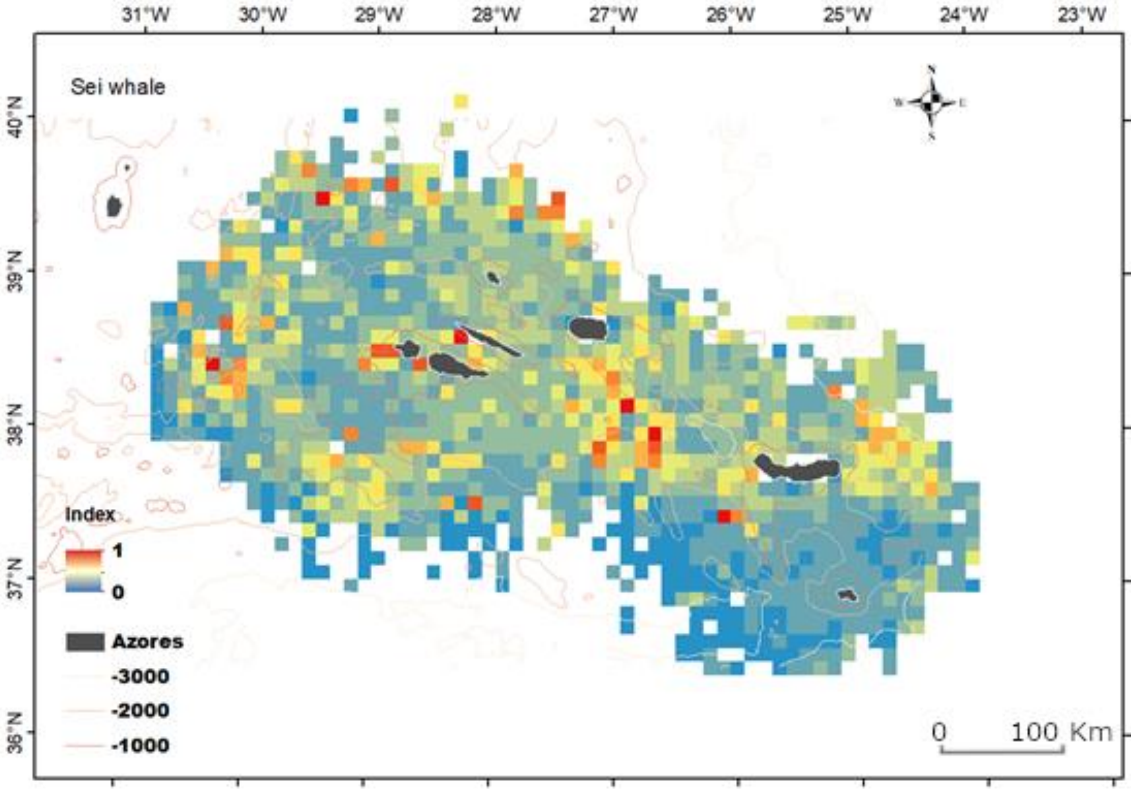


Figure 5. 6. Risk of lethal ship collisions for blue (on top) and sei whales (on bottom). The risk was calculated using each whale predicted mean density (individuals per 100 km^2) based on the each SDM-p model from chapter 4 and the density of ships >80m travelling at >14kn.

5.3.3. EXPOSURE TO CETACEAN WATCHING

The highest exposure of sperm whales to cetacean watching occurred N and NE of Faial, off Lajes do Pico, and between S. Mateus, S of Pico, and the channel Faial-Pico. The area SW of Capelinhos, in Faial, was only of minor importance (Figure 5.7 on top). The main areas of fin whale exposure were off Lajes and S. Mateus (Pico) and NE of Faial (Figure 5.7 on bottom), while blue whale exposure was mainly restricted to the S of Pico, especially off Lajes and Ribeiras, and S. Mateus (Figure 5.8 on top). The sei whale was the species for which exposure was more widespread including off S. Mateus, N and NE Faial Island, Lajes, S of the channel Faial-Pico, and off Capelinhos (Figure 5.8 on bottom).

5.3.4. ASSESSING THE AMOUNT OF PREFERRED HABITAT EXPOSED TO CETACEAN WATCHING ACTIVITY

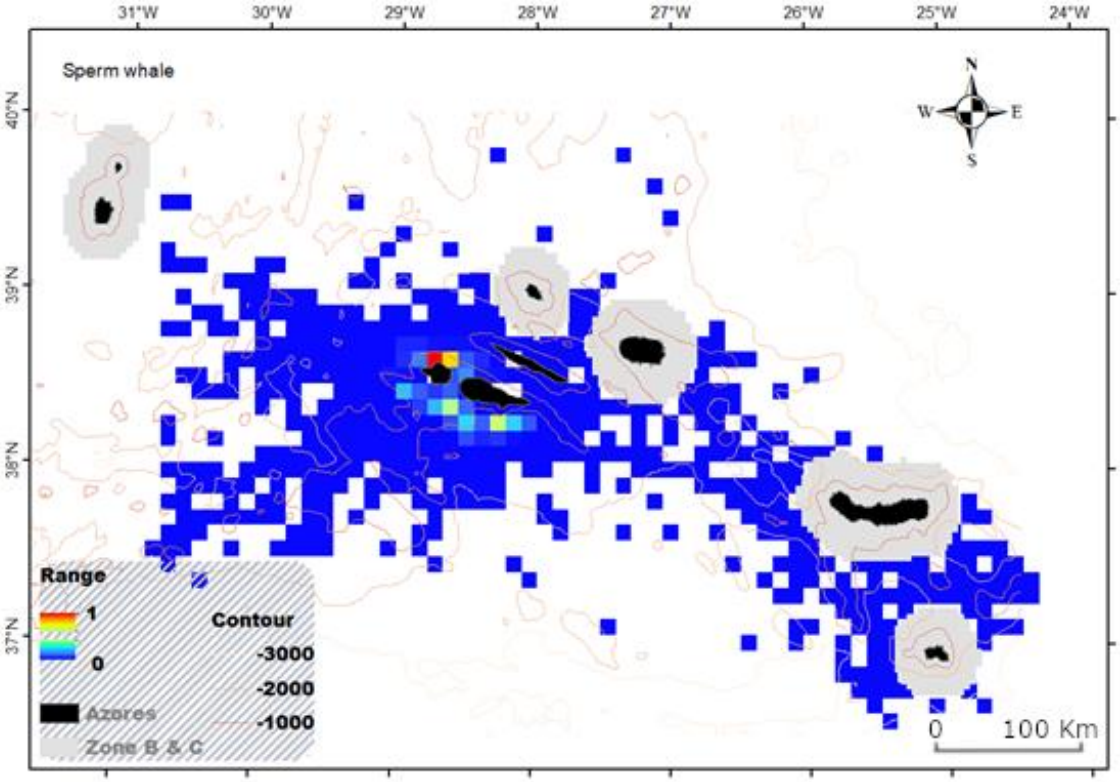
All studied whales had a highly widespread distribution in the waters around the Azores, which meant that neither species had >25% of their habitat within cetacean watching areas (zones A, B and C). The species with the highest value was the fin whale (25%), followed by the sperm whale (16%), the sei whale (12%) and the blue whale (8%). About 7% of fin whale preferred habitat within zone A was exposed to high cetacean watching effort, whereas sei and sperm whales had only 3% of their habitat highly exposed, and blue whale 2%.

5.4. DISCUSSION

This chapter describes a method to assess the exposure of large whales to shipping traffic and cetacean watching activity. Predicted relative density of sperm, fin, blue and sei whales, and density of shipping and cetacean watching were analysed to identify spatial patterns of exposure and collision risk.

5.4.1. CAVEAT OF THE APPROACH

It must be taken into account that the dataset used to develop models and maps in chapter 5 has been spatially and temporally limited due to data on anthropogenic activities. Therefore, data was subset from 2012 to 2016 and the area around Corvo and Flores are out of this chapter's study area. In addition, two prey variables have been eliminated from the set (The non-migratory upper and lower-mesopelagic) to predict also in swallow coastal areas, relevant to the anthropogenic activities. While GAMs perform reasonably well with small sample sizes, some caution is necessary when interpreting the results, because models are based on few occurrences [83]. The model predictions may be affected by differences in temporal resolution between the best model and the mapped predictions. The dynamic covariates included in the models were contemporaneous (same week) to the sightings. These data will present more variability than the climatologies used in the model's prediction. In addition, samples with consecutive zeros tend to underestimate the density [84].



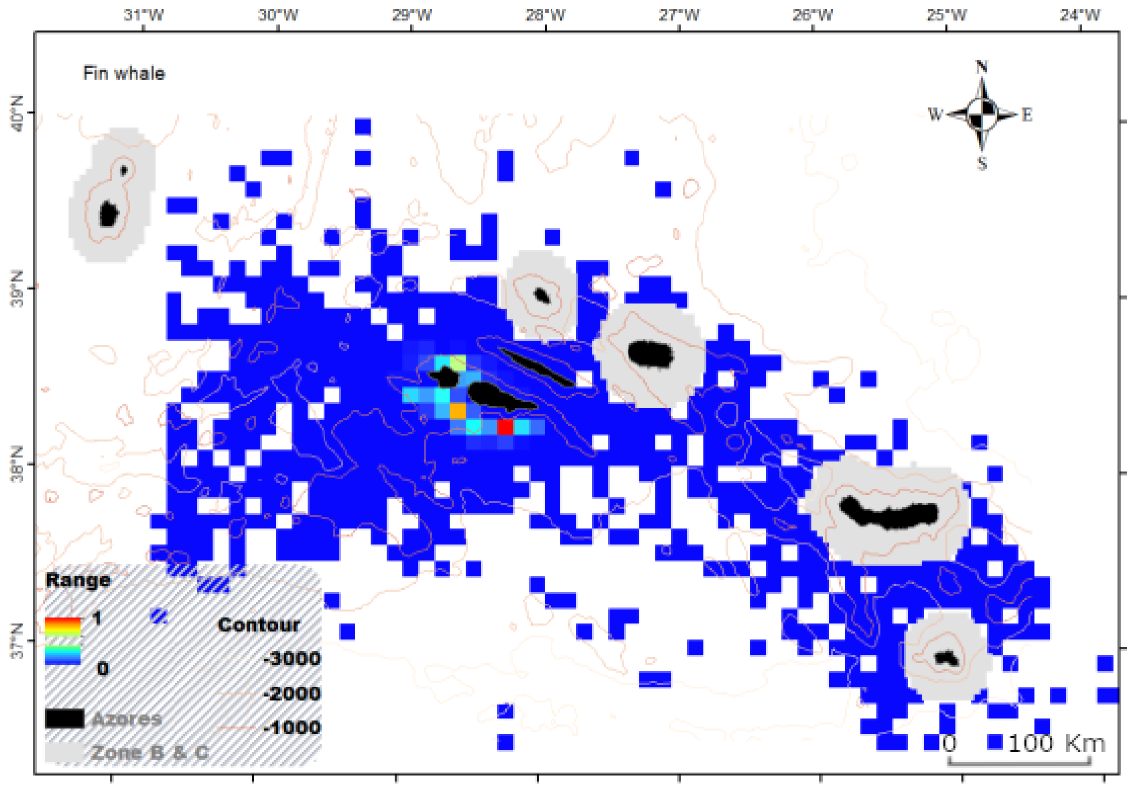
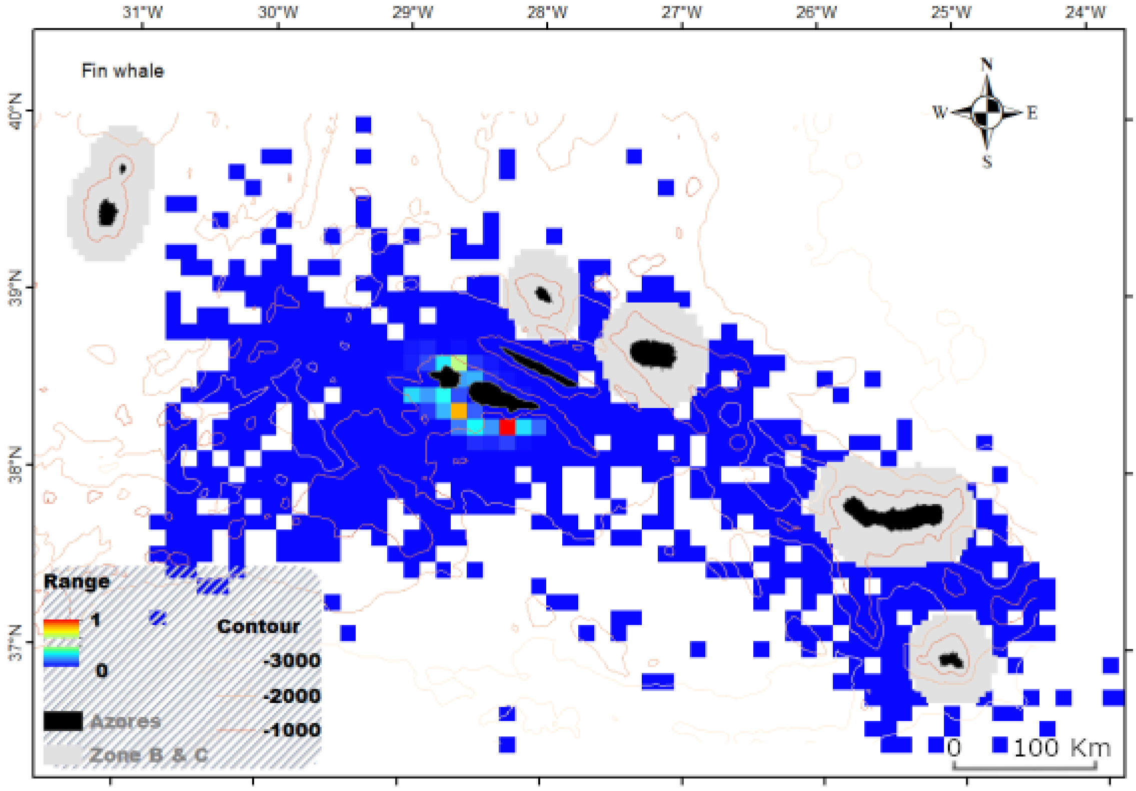


Figure 5. 7. Exposure to cetacean watching activity for sperm (on top) and fin whales (on bottom). The exposure was calculated using fin whale predicted mean density (individuals per 100 km^2) based on the SDM-p-aa and SDM-p-es model, respectively, and the cetacean watching effort estimated in chapter 2. Grey areas represent zones excluded from the predictions due to the lack of cetacean watching data (zones b and c).



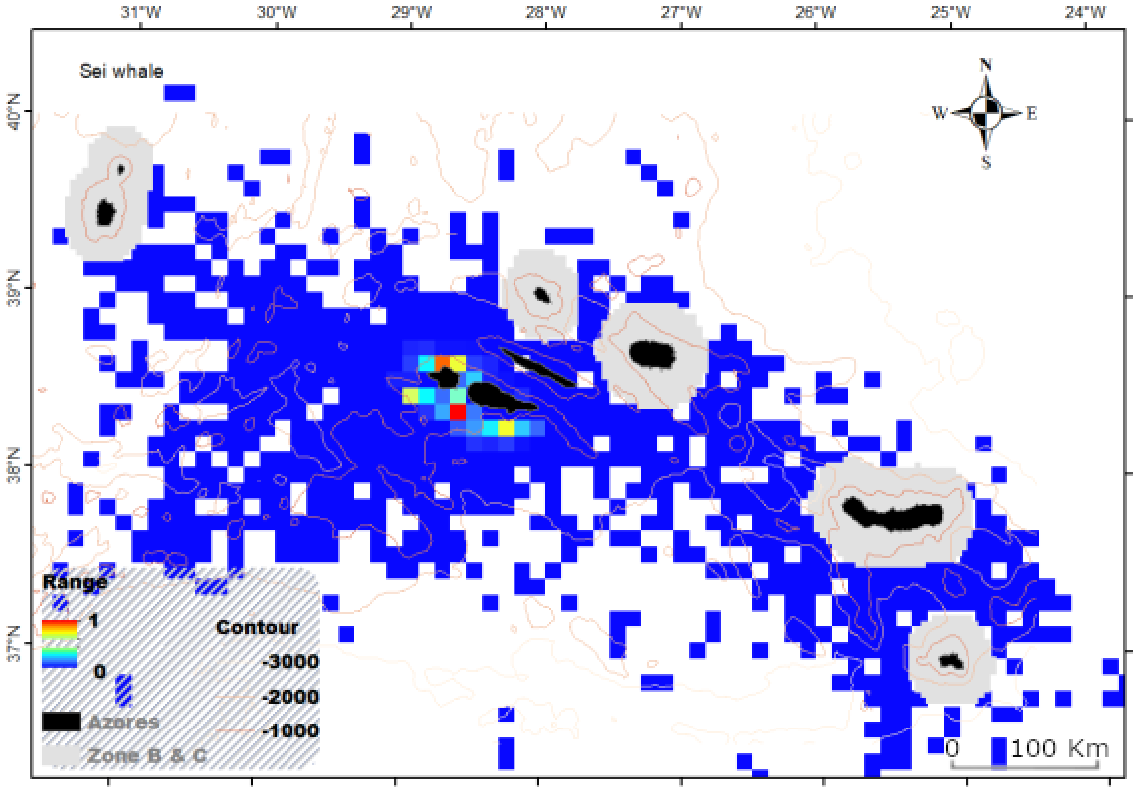


Figure 5. 8. Exposure to cetacean watching activity for blue (on top) and sei whales (on bottom). The exposure was calculated using each whale predicted mean density (individuals per 100 km^2) based on the each SDM-p from chapter 4 model and the cetacean watching effort estimated in chapter 2. Grey areas represent zones excluded from the predictions due to the lack of cetacean watching data (zones b and c).

Testing model accuracy with internal performance evaluation is an effective alternative to testing with an independent dataset, which is undoubtedly most robust [85-87] but also more costly for highly mobile animals.

As mentioned in the section "Model limitations and predictive performance" in Chapter 4, the models were seasonally restricted because POPA only collects data in spring and summer. In addition, by subsampling data to only four years to cover a similar time period of the anthropogenic activities studied, SDM performance may decrease. Shipping traffic and cetacean watching activities are driven by dynamic factors, such as economic changes, government restrictions, demand [88, 89], changes in fishing, environmental conditions, climate change, new routes [90-92], and cetacean distribution [47]. Therefore quantifying the activities at the lowest temporal resolution, as done for shipping traffic, is the best option. Unfortunately, similar data was not available for cetacean watching, and the analysis was done using a static representation for the closest period of the cetacean data. Besides, the cetacean watching data covers only part of the activity, precluding examination of this activity for the whole archipelago.

Despite the limitations mentioned, the fin whale model showed a reasonable predictive performance. Despite having more observations to fit the models, the sperm whale model fitted the data poorly and showed a relatively low predictive performance. As explained in the previous chapter, the reasons for this difference may lie in the difficulty of count in numerous groups, or in the inability of the candidate variables to capture with more accuracy the sperm whale density drivers. One example is that SEAPODYM model simulate smaller preys as included in sperm whale diet. Nevertheless, both models are helpful for ecological interpretation and assessing risk from human activities. The predictions of the new models for sperm and fin whales, as well as those of the models developed in chapter 4 for sei and blue whales, were used to estimate the risk of lethal collisions of these species with ships, and examine their exposure to cetacean watching. I expect the predictions of the new models to provide more accurate results relative to the old models, because they match the temporal scale and resolution of dynamic anthropogenic activities. If these activities vary greatly over time, the models fitted to the same period will be more realistic. Conversely, if differences are small, the long-term models may provide more accurate results. At present, I have no data to assess those differences, but this would be interesting to pursue in future studies.

5.4.2. INFLUENCE OF ANTHROPOGENIC ACTIVITIES ON SPECIES DISTRIBUTION

The density of ships >80m long travelling at >14kn was statistically significant at explaining the relative density of sperm whales in the study area. Inclusion of this variable in the SDM of sperm whales slightly improved model performance, indicating that in addition to oceanographic features and distribution of prey, shipping traffic also influences sperm whale distribution. Sperm whales prefer areas close to mesoscale structures, greater production of epipelagic and highly migrant bathypelagic organisms, and lower traffic density. Mesoscale structures can increase or aggregate productivity, which in turn attract micronekton organisms and their predators [93]. This is consistent also with the relationship of sperm whales with areas of increased epipelagic and migrant bathypelagic production which likely serve as prey to the large mesopelagic and bathypelagic cephalopods on which sperm whales feed [94, 95]. The confirmation of a contemporaneous and direct relationship between maritime traffic and sperm whales is highly relevant because there is hardly any information on how human activities shape the distribution of cetaceans [96, 97].

A critical determinant of landscape use by animals is the perception of risk, and individuals must balance the rewards (e.g., foraging opportunities, shelter) with the potential costs (e.g., increased predation risk)[98]. Human activity typically elicits fear responses in wildlife similar to those caused by non-human predators, such as altered habitat selection, activity, or vigilance [98]. Avoidance of areas with higher shipping density indicates that sperm whales perceive traffic as an important threat. Sperm whale behavioural response likely decrease the probability of encounters and the risk of collisions with ships, and is therefore critical for population persistence. However, avoidance of high traffic areas could also have negative consequences for sperm whales, if animals are displaced from higher quality habitats used for foraging. A recent study showed huge variations in squid density between neighbouring, bathymetrically similar areas, and over small horizontal distances [99]. Thus, use of alternative habitats with less traffic could mean poorer foraging options with adverse energetic consequences for sperm whales.

Conversely, the models did not show a significant relationship between fin whale relative density and shipping traffic. The lack of a relationship could be due to the small number of whale observations used to fit the model. In small datasets, the relationship of the response variable with the covariates can be disproportionately affected by observations at the extreme of the distribution, i.e, in this case, by observations with more than one individual whale [100], which often respond less to large vessels than single animals [101]. Previous studies detected a response in the behaviour of animals to the shipping, using minor spatial scales than that used in this thesis and acoustic methods [54]. More data and the analysis at minor scales would be necessary to confirm the absence of a relationship of shipping density with fin whale relative density, and to investigate the influence on blue and sei whales.

The best model revealed that fin whales preferred cold waters, consistent with knowledge of their seasonal migration in the area, with most sightings occurring in spring [102], when waters are colder. In addition, the fin whale response to higher productivity of phytoplankton is likely related to the timing of the spring bloom and the subsequent rise in zooplankton on which these whales feed [103].

5.4.2.1. RISK OF LETHAL SHIP COLLISIONS

The higher risk of lethal collisions for sperm, fin, blue and sei whales occurred along the main shipping routes that cross the archipelago. The primary areas of risk of lethal strikes were at the main routes connecting the American and European continents, some of which go through the archipelago, or cross just to the north or south of the islands, and the routes between islands groups (Eastern-Central, Central-Western).

Routes between islands of the central group posed a higher risk of lethal collisions for sperm and sei whales than for fin and blue whales. These later species whales are at higher risk in the main shipping routes linking continents and to a lower extent in the local routes linking the island groups.

Overall, risk of lethal collisions was higher and more widespread within the area for sperm whales and sei whales, than for fin and blue whales, reflecting the species' distribution patterns and relative densities in the area. However, this analysis does not take into account the diving behaviour and patterns of each species, which is also known to affect the probability of collisions with ships [104]. Specifically, the time spent at or near the surface, increases the chance of the whales being struck [105]. Foraging sperm whales spend prolonged periods of time underwater (average of 43 min of foraging dives), and 10 min at the surface in between foraging dives [106]. They spent the 70% of the time foraging and the rest of the time can remain at the surface resting or socializing. Conversely, fin, blue and sei whales exhibit shorter foraging dives (8-12 min) and shorter surfacing periods between dives [107]. Differences in diving behaviour between sperm and sei, fin and blue whales can have profound consequences for the risk of collision with ships, and should be accounted for in future studies. In addition, other factors likely influence the probability of a whale being struck by a ship. Cetaceans have a lower response to anthropogenic activities when feeding, resting, or there are calves in the group [100, 101, 108]. Therefore, future work should evaluate the influence of these factors on the risk of lethal collisions.

5.4.3. EXPOSURE TO CETACEAN WATCHING

Although cetacean watching effort has been increasing [89] and its geographic distribution expanding with the establishment of the activity in new islands and ports, the overlap with preferred habitat areas of each whale species was relatively small (8-25%). High-density areas of sperm, fin, blue and sei whales were mainly in deep, offshore waters, often tens of kms away from the home ports of cetacean watching boats and from the areas observed by lookouts. A clear example was the sei whale, for which there a peak in density W of Faial, 40km from Horta and Madalena ports, and 60km from Lajes, thus too far to be frequently targeted by boats. Whether the preference for more offshore areas is related to the higher cetacean watching effort in nearshore waters is currently unknown, as models incorporating cetacean watching effort could not be fit for any whale species.

Within zone A, the areas of highest overlap occurred off Lajes do Pico for fin and blue whales, and north of Faial and at the southern entrance of the channel Faial-Pico for sperm and sei whales. These findings are not surprising, as these areas correspond to major hotspots of cetacean watching activity, respectively, for boats operating from the port of Lajes and the ports of Horta and Madalena do Pico. While these areas make up a relatively small proportion (2-7%) of the suitable habitat available for each species in the study area, metrics used in this study to evaluate habitat quality are relatively crude and coarse-scaled and may not capture fine-scale oceanographic and prey dynamics that drive foraging success in these whales. Moreover, this analyses does not enable assessing the short and long-term consequences of whale exposure to cetacean watching activity. Hence, it is necessary to investigate if and to what extent exposure to whale-watching boats influences foraging behaviour and movements in these species.

5.4.4. CUMULATIVE EXPOSURE TO SHIPPING COLLISIONS AND CETACEAN WATCHING

This work identifies areas where sperm, fin, blue and sei whales have a higher probability of lethal collisions and of being exposed to cetacean watching boats. Within cetacean watching zone A, there were three areas where risk of exposure to both activities was high: N-NE of Faial, and S of Pico, off S. Mateus and Lajes. The co-occurrence of cetacean watching boats and commercial ships may exert accumulating and interacting impacts on large whales, both in terms of physical disturbance and underwater noise. For example, disturbance from both activities can increase the magnitude or frequency of avoidance behaviours, causing them to shift habitat selection, increase movements or vigilance, with important energetic consequences, or cause strong behavioral reactions associated with potentially disproportionate physiological responses, that may lead to severe injuries or death. Behavioural response to one of the activities may also condition response or make animals less vigilant to other approaching threats, and anthropogenic noise may distract whales [109]. Hence, it is critical that both activities are studied and managed in concert.

5.4.5. MANAGEMENT AND CONSERVATION IMPLICATIONS

While this study contributes to identifying risk areas of lethal collisions for large whales, it does not provide information on the actual number of lethal collisions. Quantifying the number of collisions requires that collisions are systematically reported in ship logbooks [66]and that observer programs be established to monitor higher risk routes and ships. Nonetheless, the present study provides important hints of potential measures to reduce the risk to large whales. Permanently re-routing shipping away from critical whale habitats and high-density areas [66], which has been successfully implemented in other regions, could be an option but must be carefully evaluated because these species are relatively widespread in the area, and alternative traffic corridors may increase collision risks elsewhere. Limiting vessel speed on the main shipping routes identified by this work and for faster vessels could be very effective but usually comes at high economic costs and is difficult to implement. However, this mitigation measure could be temporary if established along with a cetacean alert system, as done in other regions [110]. This could be done by placing visual observers aboard some vessels, such as high-speed ferries, and/or by deploying hydrophones in high-risk routes to enable whale detection in real-time [22, 111]. Both monitoring schemes would allow activating precautionary measures (e.g., reduce speed, change route) to reduce the risk of collisions [112, 113]. Acoustic monitoring could work well for sperm whales, that emit echolocation or social clicks very frequently but can be challenging for large baleen, which show lower calling rates that change seasonally [114].

This study also revealed that the cetacean watching effort within zone A is highly concentrated in very restricted areas that, together, total <5,000 km^2. Such concentration seems to be primarily driven by the location of the lookouts, followed by distance to ports, and only secondarily by whale distribution. As such, increasing and or a better spreading of the lookouts along the islands could contribute to a more even spatial distribution of cetacean watching effort and reduce exposure in specific areas. The potential benefits of this strategy in terms of behavioural disturbance to animals relative to concentrating most of the activity in fewer whale habitats should be assessed.

Finally, shipping traffic and cetacean watching are expected to continue increasing and changing according to the region's new commercial and tourism demands. These changes in both anthropogenic activities should be closely monitored and continually updated in future management assessments [115].

5.5. CONCLUSION

Analysis of exposure to shipping traffic and cetacean watching analysis based on spatial models of whale density provided an effective approach for assessing whale response to these activities and identifying areas of higher risk. These analyses are crucial in the mid-Atlantic Ocean where shipping density of commercial and passenger vessels is projected to continue increasing. These areas are also critical habitats for migratory baleen whales [102, 116] and for sperm whales [117].

I found evidence of shipping avoidance by sperm whales. The sperm whale was also the species more at risk of lethal collisions and with the higher proportion of suitable habitat overlapped with cetacean watching activity. Thus, both lethal and non-lethal impacts from these activities on sperm whales could be substantial and deserve to be further investigated. Cetacean visual and acoustic surveys throughout the year improve our understanding their abundance and seasonality in the area. These data could improve SDMs developed for the species and their predictive ability which, together with a longer AIS dataset, could be used to refine the risk analysis. These would also improve the models and risk analysis for the baleen whales. Analysis of the exposure to cetacean watching data was only carried out for a limited area of the archipelago and data from other areas is necessary to expand the analyses to cover the spatial distribution of the activity. Besides, more years of cetacean watching data could help create a near real-time covariate to analyse the species' response through SDMs. Finally, data on the behavioural responses of individual animals and on the fine-scale distribution and density of preferred prey of these species could help understand the energetic impacts of disturbance. This could then be linked with species distribution modelling and risk maps to obtain spatially-explicit estimates of disturbance, thereby improving spatial management of these activities.

5.6. REFERENCES

1. Maxwell SM, Hazen EL, Bograd SJ, Halpern BS, Breed GA, Nickel B, et al. Cumulative human impacts on marine predators. Nature communications. 2013;4:2688. Epub 2013/10/29. doi: 10.1038/ncomms3688. PubMed PMID: 24162104.
2. Fleishman E, Costa DP, Harwood J, Kraus S, Moretti D, New LF, et al. Monitoring population-level responses of marine mammals to human activities. Marine Mammal Science. 2016;32(3):1004-21. doi: 10.1111/mms.12310.
3. Williams R, Thomas L, Ashe E, Clark CW, Hammond PS. Gauging allowable harm limits to cumulative, sub-lethal effects of human activities on wildlife: A case-study approach using two whale populations. Marine Policy. 2016;70:58-64. doi: http://dx.doi.org/10.1016/j.marpol.2016.04.023.
4. Nabe-Nielsen J, van Beest FM, Grimm V, Sibly RM, Teilmann J, Thompson PM. Predicting the impacts of anthropogenic disturbances on marine populations. Conservation Letters. 2018;11(5):e12563. doi: 10.1111/conl.12563.
5. Rockwood RC, Calambokidis J, Jahncke J. High mortality of blue, humpback and fin whales from modeling of vessel collisions on the US West Coast suggests population impacts and insufficient protection. PloS one. 2017;12(8):e0183052. doi: https://doi.org/10.1371/journal.pone.0183052.
6. Nelms SE, Alfaro-Shigueto J, Arnould JP, Avila IC, Nash SB, Campbell E, et al. Marine mammal conservation: over the horizon. Endangered Species Research. 2021;44:291-325. doi: https://doi.org/10.3354/esr01115.
7. Schoeman RP, Patterson-Abrolat C, Plön S. A Global Review of Vessel Collisions With Marine Animals. Frontiers in Marine Science. 2020;7(292). doi: 10.3389/fmars.2020.00292.
8. García-Cegarra AM, Pacheco AS. Collision risk areas between fin and humpback whales with large cargo vessels in Mejillones Bay (23°S), northern Chile. Marine Policy. 2019;103:182-6. doi: 10.1016/j.marpol.2018.12.022.
9. Van Waerebeek K, Baker AN, Félix F, Gedamke J, Iñiguez M, Sanino GP, et al. Vessel collisions with small cetaceans worldwide and with large whales in the Southern Hemisphere, an initial assessment. Latin American Journal of Aquatic Mammals. 2007;6(1):43-69. doi: 10.5597/lajam00109.
10. Potvin J, Cade DE, Werth AJ, Shadwick RE, Goldbogen JA. A perfectly inelastic collision: Bulk prey engulfment by baleen whales and dynamical implications for the world's largest cetaceans. American Journal of Physics. 2020;88(10):851-63. doi: doi: 10.1119/10.0001771.
11. Williams SH. Quantifying the ecological processes underlying collisions between large baleen whales and large ships to evaluate risk. University of Montana. 2017;10(4):1-15. doi: 10.1002/ecs2.2713.

12. Silber GK, Slutsky J, Bettridge S. Hydrodynamics of a ship/whale collision. Journal of Experimental Marine Biology and Ecology. 2010;391(1-2):10-9. doi: 10.1016/j.jembe.2010.05.013.

13. McKenna MF, Calambokidis J, Oleson EM, Laist DW, Goldbogen JA. Simultaneous tracking of blue whales and large ships demonstrates limited behavioral responses for avoiding collision. Endangered Species Research. 2015;27(3):219-32. doi: 10.3354/esr00666.

14. Nowacek DP, Johnson MP, Tyack PL. North Atlantic right whales (Eubalaena glacialis) ignore ships but respond to alerting stimuli. Proceedings Biological sciences. 2004;271(1536):227-31. Epub 2004/04/03. doi: 10.1098/rspb.2003.2570. PubMed PMID: 15058431; PubMed Central PMCID: PMC1691586.

15. Harris K, Gende SM, Logsdon MG, Klinger T. Spatial pattern analysis of cruise ship–humpback whale interactions in and near Glacier Bay National Park, Alaska. Environmental management. 2012;49(1):44-54. doi: 10.1007/s00267-011-9754-9.

16. Goldbogen JA, Calambokidis J, Shadwick RE, Oleson EM, McDonald MA, Hildebrand JA. Kinematics of foraging dives and lunge-feeding in fin whales. Journal of Experimental Biology. 2006;209(7):1231-44. doi: 10.1242/jeb.02135.

17. Acevedo-Gutiérrez A, Croll D, Tershy B. High feeding costs limit dive time in the largest whales. Journal of Experimental Biology. 2002;205(12):1747-53. doi: https://doi.org/10.1242/jeb.205.12.1747.

18. Winkler C, Panigada S, Murphy S, Ritter F. Global Numbers of Ship Strikes: An Assessment of Collisions between Vessels and Cetaceans Using Available Data in the IWC Ship Strike Database. IWC B. 2020;68.

19. Laist DW, Knowlton AR, Mead JG, Collet AS, Podesta M. Collisions between ships and whales. Marine Mammal Science. 2001;17(1):35-75. doi: https://doi.org/10.1111/j.1748-7692.2001.tb00980.x.

20. Arregui M, Bernaldo de Quiros Y, Saavedra P, Sierra E, Suárez-Santana CM, Arbelo M, et al. Fat embolism and sperm whale ship strikes. Frontiers in Marine Science. 2019;6:379. doi: https://doi.org/10.3389/fmars.2019.00379.

21. Peltier H, Beaufils A, Cesarini C, Dabin W, Dars C, Demaret F, et al. Monitoring of Marine Mammal Strandings Along French Coasts Reveals the Importance of Ship Strikes on Large Cetaceans: A Challenge for the European Marine Strategy Framework Directive. 2019;6(486). doi: 10.3389/fmars.2019.00486.

22. Douglas AB, Calambokidis J, Raverty S, Jeffries SJ, Lambourn DM, Norman SA. Incidence of ship strikes of large whales in Washington State. Journal of the Marine Biological Association of the United Kingdom. 2008;88(6):1121-32. doi: 10.1017/S0025315408000295.

23. Carrillo M, Ritter F. Increasing numbers of ship strikes in the Canary Islands: proposals for immediate action to reduce risk of vessel-whale collisions. Journal of Cetacean Research and Management. 2010;11(2):131-8.

24. Berman-Kowalewski M, Gulland FM, Wilkin S, Calambokidis J, Mate B, Cordaro J, et al. Association between blue whale (Balaenoptera musculus) mortality and ship strikes along the California coast. Aquatic Mammals. 2010;36(1):59-66. doi: 10.1578/am.36.1.2010.59.

25. Nanayakkara RP, Herath H. Report on 14 large whales that died due to ship strikes off the coast of Sri Lanka, 2010–2014. Journal of Marine Biology. 2017;2017. doi: https://doi.org/10.1155/2017/6235398.

26. Monnahan CC, Branch TA, Punt AE. Do ship strikes threaten the recovery of endangered eastern North Pacific blue whales? Marine Mammal Science. 2015;31(1):279-97. doi: 10.1111/mms.12157.

27. White A, Perry R, Koops M, Randall R, Bundy A, Lawton P, et al. A National Sythesis of the Fisheries and Oceans Canada Ecosystem Research Initiative: Canadian Science Advisory Secretariat; 2013.

28. Cooke JG. Balaenoptera borealis. The IUCN Red List of Threatened Species 2018: 2018.

29. Cooke JG. Balaenoptera musculus (errata version published in 2019). The IUCN Red List of Threatened Species 2018: 2018.

30. Cooke JG. Balaenoptera physalus. The IUCN Red List of Threatened Species 2018: : 2018.

31. Taylor B, Baird R, Barlow J, Dawson S, Ford J, Mead J, et al. Physeter macrocephalus (amended version of 2008 assessment). The IUCN Red List of Threatened Species 2019: IUCN, 2019.

32. IUCN. The IUCN red list of threatened species. Version 2019-1. IUCN, Gland, Switzerland. 2019 [cited 2021 31-08-2021]. Available from: Available from http://www.iucnredlist.org. .

33. Peel D, Smith JN, Childerhouse S. Vessel strike of whales in Australia: the challenges of analysis of historical incident data. Frontiers in Marine Science. 2018;5:69. doi: 10.3389/fmars.2018.00069.

34. Panigada S, Pesante G, Zanardelli M, Capoulade F, Gannier A, Weinrich MT. Mediterranean fin whales at risk from fatal ship strikes. Marine pollution bulletin. 2006;52(10):1287-98. Epub 2006/05/23. doi: 10.1016/j.marpolbul.2006.03.014. PubMed PMID: 16712877.

35. Wright AJ, Simmonds MP, Galletti Vernazzani B. The international whaling commission—beyond whaling. Frontiers in Marine Science. 2016;3:158. doi: https://doi.org/10.3389/fmars.2016.00158.

36. Bradford AL, Weller DW, Ivashchenko YV, Burdin AM, Brownell J, Robert L. Anthropogenic scarring of western gray whales (Eschrichtius robustus). Marine Mammal Science. 2009;25(1):161-75. doi: 10.1111/j.1748-7692.2008.00253.x.

37. de Vere AJ, Lilley MK, Frick EE. Anthropogenic impacts on the welfare of wild marine mammals. Aquatic Mammals. 2018;44(2):150-80. doi: 10.1578/AM.44.2.2018.150.

38. Redfern J, McKenna M, Moore T, Calambokidis J, Deangelis M, Becker E, et al. Assessing the risk of ships striking large whales in marine spatial planning. Conservation Biology. 2013;27(2):292-302. doi: 10.1111/cobi.12029

39. DRPM. Rede de Arrojamento de Cetáceos dos Açores: Direção Regional de Políticas Marítimas; 2021 [cited 2021]. Available from: https://servicos-sraa.azores.gov.pt/doit/servicos.asp?id_dep=10&id_form=84.

40. Dyck MG, Baydack RK. Vigilance behaviour of polar bears (Ursus maritimus) in the context of wildlife-viewing activities at Churchill, Manitoba, Canada. Biological Conservation. 2004;116(3):343-50. doi: 10.1016/S0006-3207(03)00204-0.

41. Christiansen F, Bertulli CG, Rasmussen MH, Lusseau D. Estimating cumulative exposure of wildlife to non-lethal disturbance using spatially explicit capture–recapture models. The Journal of Wildlife Management. 2015;79(2):311-24. doi: http://dx.doi.org/10.1002/jwmg.836.

42. Christiansen F, Lusseau D. Linking behavior to vital rates to measure the effects of non-lethal disturbance on wildlife. Conservation Letters. 2015;8(6):424-31. doi: 10.1111/conl.12166.

43. Christiansen F, Rasmussen M, Lusseau D. Whale watching disrupts feeding activities of minke whales on a feeding ground. Marine Ecology Progress Series. 2013;478:239-51. doi: 10.3354/meps10163.

44. Crossman CA, Barrett-Lennard LG, Frasier TR. An example of DNA methylation as a means to quantify stress in wildlife using killer whales. Scientific reports. 2021;11(1):1-8. doi: https://doi.org/10.1038/s41598-021-96255-1.

45. Higham J, Bejder L, Williams R. Whale-watching: Sustainable tourism and ecological management: Cambridge University Press; 2014.

46. Hoyt E. Whale watching 2001: worldwide tourism numbers, expenditures, and expanding socioeconomic benefits. Yarmouth Port, MA, USA: International Fund for Animal Welfare; 2001. 1-158 p.

47. O'Connor S, Campbell R, Cortez H, Knowles T. Whale Watching Worldwide: tourism numbers, expenditures and expanding economic benefits, a special report from the International Fund for Animal Welfare. International Fund for Animal Welfare, 2009.

48. Filby NE, Stockin KA, Scarpaci C. Social science as a vehicle to improve dolphin-swim tour operation compliance? Marine Policy. 2015;51:40-7. doi: https://doi.org/10.1016/j.marpol.2014.07.010.

49. Pacheco AS, Silva S, Alcorta B, Gubbins S, Guidino C, Sanchez-Salazar F, et al. Short Note Cetaceans Diversity Revealed from Whale-Watching. Aquatic Mammals. 2019;45(1):116-22. doi: 10.1578/AM.45.1.2019.116.

50. Holt MM, Tennessen JB, Ward EJ, Hanson MB, Emmons CK, Giles DA, et al. Effects of vessel distance and sex on the behavior of endangered killer whales. Frontiers in Marine Science. 2021;7:1211. doi: 10.3389/fmars.2020.582182.

51. Arranz P, de Soto NA, Madsen PT, Sprogis KR. Whale-watch vessel noise levels with applications to whale-watching guidelines and conservation. Marine Policy. 2021;134:104776. doi: 10.1016/j.marpol.2021.104776.

52. Senigaglia V, Christiansen F, Bejder L, Gendron D, Lundquist D, Noren D, et al. Meta-analyses of whale-watching impact studies: comparisons of cetacean responses to disturbance. Marine Ecology Progress Series. 2016;542:251-63. doi: 10.3354/meps11497.

53. Croll DA, Clark CW, Calambokidis J, Ellison WT, Tershy BR, editors. Effect of anthropogenic low-frequency noise on the foraging ecology of Balaenoptera whales. Animal Conservation forum; 2001: Cambridge University Press.

54. Castellote M, Clark CW, Lammers MO. Acoustic and behavioural changes by fin whales (Balaenoptera physalus) in response to shipping and airgun noise. Biological Conservation. 2012;147(1):115-22. doi: 10.1121/2.0000311.

55. Oliveira C, Gonçalves J, Magalhães S, Prieto R, Silva M, Santos R. Whale watching management in the Azores: An updated review of the regulations. International Whaling Commission. Anchorage, Alaska. SC/59/WW7: 4pp, 2007.

56. Silva MA, Prieto R, Cascão I, Seabra MI, Machete M, Baumgartner MF, et al. Spatial and temporal distribution of cetaceans in the mid-Atlantic waters around the Azores. Marine Biology Research. 2014;10(2):123-37. doi: https://doi.org/10.1080/17451000.2013.793814.

57. Richter CF, Dawson S, Slooten E. Sperm whale watching off Kaikoura, New Zealand: effects of current activities on surfacing and vocalisation patterns: Department of Conservation Wellington, NZ; 2003.

58. Magalhães S, Prieto R, Silva MA, Gonçalves J, Afonso-Dias M, Santos RS. Short-term reactions of sperm whales (Physeter macrocephalus) to whale-watching vessels in the Azores. Aquatic Mammals. 2002;28(3):267-74.

59. Steiner L, Lamoni L, Plata MA, Jensen S-K, Lettevall E, Gordon J. A link between male sperm whales, Physeter macrocephalus, of the Azores and Norway. Marine Biological Association of the United Kingdom. 2012;92(8):1751. doi: 10.1017/S0025315412000793.

60. Mullin KD, Steiner L, Dunn C, Claridge D, García LG, Gordon J, et al. Long-Range Longitudinal Movements of Sperm Whales (Physeter macrocephalus) in the North Atlantic Ocean Revealed by Photo-Identification. Aquatic Mammals. 2022;48(1):3-8. doi: 10.1578/AM.48.1.2022.3.

61. Steiner L. Reseaclarch whalewacthazores2021 [cited 2021 26/08/2021]. Available from: http://www.whalewatchazores.com/index.php/research.

62. Ritter F, Ernert A, Smit V, editors. A long-term cetacean sighting data set from whale watching operations as a reflection of the environmental dynamics in a multi-species cetacean habitat. Poster presented at the Annual Conference of the European Cetacean Society (ECS) in Cadiz (Spain); 2011.

63. Rødland ES, Bjørge A. Residency and abundance of sperm whales (Physeter macrocephalus) in the Bleik Canyon, Norway. Marine Biology Research. 2015;11(9):974-82. doi: 10.1080/17451000.2015.1031800.

64. Sequeira M, Elejabeitia C, Silva M, Dinis A, de Stephanis R, Urquiola E, et al. Review of whalewatching activities in mainland Portugal, the Azores, Madeira and Canary archipelagos and the Strait of Gibraltar. Cetacean Research and Management. 2009; SC61/WW11:1-41.

65. Williams SH, Gende SM, Lukacs PM, Webb K. Factors affecting whale detection from large ships in Alaska with implications for whale avoidance. Endangered Species Research. 2016;30:209-23. doi: 10.3354/esr00736.

66. Mullen KA, Peterson ML, Todd SK. Has designating and protecting critical habitat had an impact on endangered North Atlantic right whale ship strike mortality? Marine Policy. 2013;42:293-304. doi: 10.1016/j.marpol.2013.03.021.

67. Hoyt E, Parsons E. The whale-watching industry. In: Laetitia Nunny WAW, La Garriga, Spain, editor. Whale-watching: Sustainable tourism ecological management. 57. First ed. OceanCare2014. p. 86-96.

68. Decreto Legislativo Regional n.º 9/99/A, de 22 de março - Observação de cetáceos (1999).

69. Decreto Legislativo Regional n.º 10/2003/A, de 22 de março - Altera o Decreto Legislativo Regional n.º 9/99/A, de 22 de março (2003).

70. Decreto Legislativo Regional n.º 13/2004/A, de 23 de março - Altera o Decreto Legislativo Regional n.º 10/2003/A, de 22 de março (2004).

71. Portaria n.º 5/2004, de 29 de janeiro (2005).

72. Portaria n.º 70/2005, de 8 de setembro - Altera a Portaria n.º 5/2004, de 29 de janeiro (2005).

73. Portaria n.º 17/2007, de 29 de março, (2007).

74. Hastie TJ, Tibshirani RJ. Generalized Additive Models. 1st ed. Routledge, editor1990.

75. Becker EA, Forney KA, Fiedler PC, Barlow J, Chivers SJ, Edwards CA, et al. Moving towards dynamic ocean management: how well do modeled ocean products predict species distributions? Remote Sensing. 2016;8(2):149. doi: https://doi.org/10.3390/rs8020149.

76. Shono H. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research. 2008;93(1-2):154-62. doi: 10.1016/j.fishres.2008.03.006.

77. Peel D, Bravington M, Kelly N, Wood SN, Knuckey I. A model-based approach to designing a fishery-independent survey. Journal of agricultural, biological, environmental statistics. 2013;18(1):1-21. doi: 10.1007/s13253-012-0114-x.

78. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B. 2011;73(1):3-36. doi: https://doi.org/10.1111/j.1467-9868.2010.00749.x.

79. Harrell Jr FE, Harrell Jr MFE. Package ‘hmisc’. CRAN. 2019;2019:235-6.

80. Wood SN. Thin plate regression splines. Journal of the Royal Statistical Society: Series B. 2003;65(1):95-114. doi: https://doi.org/10.1111/1467-9868.00374.

81. Wood SN. Introducing GAMs. In: Group TF, editor. Generalized additive models: an introduction with R. 4. Second ed. A Chapman and Hall Book: CRC press; 2017. p. 161-320.

82. Burnham KP, Anderson DR. Practical Use of the Information-Theoretic Approach. Model selection multimodel inference. 2002;2. doi: 10.1007/978-1-4757-2917-7_3.

83. Smith AB, Santos MJ. Testing the ability of species distribution models to infer variable importance. Ecography. 2020;43(12):1801-13. doi: doi: 10.1111/ecog.05317.

84. Jensen OP, Seppelt R, Miller TJ, Bauer LJ. Winter distribution of blue crab Callinectes sapidus in Chesapeake Bay: application and cross-validation of a two-stage generalized additive model. Marine Ecology Progress Series. 2005;299:239-55. doi: 10.3354/meps299239.

85. Araújo MB, Pearson RG, Thuiller W, Erhard M. Validation of species–climate impact models under climate change. Global change biology. 2005;11(9):1504-13. doi: 10.1111/j.1365-2486.2005.001000.x.

86. Olden JD, Jackson DA, Peres-Neto PR. Predictive models of fish species distributions: a note on proper validation and chance predictions. Transactions of the American fisheries society. 2002;131(2):329-36. doi: 10.1577/1548-8659(2002)131<0329:PMOFSD>2.0.CO;2.

87. Guisan A, Thuiller W, Zimmermann NE. Habitat suitability and distribution models: with applications in R: Cambridge University Press; 2017.

88. Hoyt E. Whale watching. In: WF Perrin BWaJT, editor. Encyclopedia of marine mammals. Third ed. Academic Press: Elsevier; 2009. p. 1223-7.

89. Ressurreiçao A, Cardigos F, Giacomello E, Leite N, Oliveira F, Kaiser MJ, et al. The value of marine ecotourism for an European outermost region. Ocean and Coastal Management. 2022;222:106129. doi: https://doi.org/10.1016/j.ocecoaman.2022.106129.

90. March D, Metcalfe K, Tintoré J, Godley BJ. Tracking the global reduction of marine traffic during the COVID-19 pandemic. Nature communications. 2021;12(1):1-12. doi: https://doi.org/10.1038/s41467-021-22423-6.

91. Breeze H, Li S, Marotte EC, Theriault JA, Wingfield J, Xu J. Changes in underwater noise and vessel traffic in the approaches to Halifax Harbor, Nova Scotia, Canada. Frontiers in Marine Science. 2021;8:796. doi: 10.3389/fmars.2021.674788.

92. Jensen CM, Hines E, Holzman BA, Moore TJ, Jahncke J, Redfern JV. Spatial and temporal variability in shipping traffic off San Francisco, California. Coastal Management. 2015;43(6):575-88. doi: http://dx.doi.org/10.1080/08920753.2015.1086947.

93. Abrahms B, Scales KL, Hazen EL, Bograd SJ, Schick RS, Robinson PW, et al. Mesoscale activity facilitates energy gain in a top predator. Proceedings of the Royal Society B. 2018;285(1885):20181101. doi: http://dx.doi.org/10.1098/rspb.2018.1101.

94. Biggs DC, Leben RR, Ortega-Ortiz JG. Ship and satellite studies of mesoscale circulation and sperm whale habitats in the northeast Gulf of Mexico during GulfCet II. Gulf of Mexico Science. 2000;18(1):2. doi: 10.18785/goms.1801.02.

95. Griffin RB. Sperm whale distributions and community ecology associated with a warm-core ring off Georges Bank. Marine Mammal Science. 1999;15(1):33-51. doi: https://doi.org/10.1111/j.1748-7692.1999.tb00780.x.

96. Campana I, Crosti R, Angeletti D, Carosso L, David L, Di-Méglio N, et al. Cetacean response to summer maritime traffic in the Western Mediterranean Sea. Marine Environmental Research. 2015;109:1-8. doi: https://doi.org/10.1016/j.marenvres.2015.05.009.

97. Chou E, Kershaw F, Maxwell SM, Collins T, Strindberg S, Rosenbaum HC. Distribution of breeding humpback whale habitats and overlap with cumulative anthropogenic impacts in the Eastern Tropical Atlantic. Diversity Distributions. 2020;26(5):549-64. doi: 10.1111/ddi.13033.

98. Frid A, Dill L. Human-caused disturbance stimuli as a form of predation risk. Conservation ecology. 2002;6(1). doi: 10.5751/ES-00404-060111.

99. Southall BL, Benoit-Bird KJ, Moline MA, Moretti D. Quantifying deep-sea predator–prey dynamics: Implications of biological heterogeneity for beaked whale conservation. Journal of Applied Ecology. 2019;56(5):1040-9. doi: 10.1111/1365-2664.13334.

100. Elgar MA. Predator vigilance and group size in mammals and birds: a critical review of the empirical evidence. Biological Reviews. 1989;64(1):13-33. doi: 10.1111/j.1469-185x.1989.tb00636.x

101. Aniceto AS, Carroll J, Tetley MJ, Van Oosterhout C. Position, swimming direction and group size of fin whales (Balaenoptera physalus) in the presence of a fast-ferry in the Bay of Biscay. Oceanologia. 2016;58(3):235-40. doi: http://dx.doi.org/10.1016/j.oceano.2016.02.002.

102. Silva MA, Prieto R, Jonsen I, Baumgartner MF, Santos RS. North Atlantic blue and fin whales suspend their spring migration to forage in middle latitudes: building up energy reserves for the journey? PloS one. 2013;8(10):e76507. doi: 10.1371/journal.pone.0076507.

103. Amorim P, Perán AD, Pham CK, Juliano M, Cardigos F, Tempera F, et al. Overview of the Ocean Climatology and Its Variability in the Azores Region of the North Atlantic Including Environmental Characteristics at the Seabed. Frontiers in Marine Science. 2017;4. doi: 10.3389/fmars.2017.00056.

104. Caruso F, Hickmott L, Warren JD, Segre P, Chiang G, Bahamonde P, et al. Diel differences in blue whale (Balaenoptera musculus) dive behavior increase nighttime risk

of ship strikes in northern Chilean Patagonia. Integrative Zoology. 2021;16(4):594-611. doi: 10.1111/1749-4877.12501.

105. Calambokidis J, Fahlbusch JA, Szesciorka AR, Southall BL, Cade DE, Friedlaender AS, et al. Differential vulnerability to ship strikes between day and night for blue, fin, and humpback whales based on dive and movement data from medium duration archival tags. Frontiers in Marine Science. 2019;6:543. doi: 10.3389/fmars.2019.00543.

106. Oliveira CI. Behavioural ecology of the sperm whale (Physeter macrocephalus) in the North Atlantic Ocean 2014.

107. Croll DA, Acevedo-Gutiérrez A, Tershy BR, Urbán-Ramírez J. The diving behavior of blue and fin whales: is dive duration shorter than expected based on oxygen stores? Comparative Biochemistry Physiology Part A: Molecular Integrative Physiology. 2001;129(4):797-809. doi: https://doi.org/10.1016/S1095-6433(01)00348-8.

108. Dolman S, Williams-Grey V, Asmutis-Silvia R, Isaac S. Vessel Collisions and Cetaceans: What Happens If They Don't Miss the Boat. Wiltshire: Whale and Dolphin Conservation Society., 2006.

109. Wright AJ, Kyhn LA. Practical management of cumulative anthropogenic impacts with working marine examples. Conservation Biology. 2015;29(2):333-40. doi: 10.1111/cobi.12425.

110. Best P, Marzetti S, Poupard M, Ferrari M, Paris S, Marxer R, et al., editors. Stereo to five-channels bombyx sonobuoys: from four years cetacean monitoring to real-time whale-ship anti-collision system. e-Forum Acusticum 2020; 2020.

111. Kowarski KA, Gaudet BJ, Cole AJ, Maxner EE, Turner SP, Martin SB, et al. Near real-time marine mammal monitoring from gliders: Practical challenges, system development, and management implications. The Journal of the Acoustical Society of America. 2020;148(3):1215-30. doi: 10.1121/10.0001811.

112. Soldevilla MS, Rice AN, Clark CW, Garrison LP. Passive acoustic monitoring on the North Atlantic right whale calving grounds. Endangered Species Research. 2014;25(2):115-40. doi: 10.3354/esr00603.

113. Vanderlaan AS, Taggart CT, Serdynska AR, Kenney RD, Brown MW. Reducing the risk of lethal encounters: vessels and right whales in the Bay of Fundy and on the Scotian Shelf. Endangered Species Research. 2008;4(3):283-97. doi: 10.3354/esr00083.

114. Romagosa M, Baumgartner M, Cascão I, Lammers MO, Marques TA, Santos RS, et al. Baleen whale acoustic presence and behaviour at a Mid-Atlantic migratory habitat, the Azores Archipelago. Scientific reports. 2020;10(1):1-11. doi: https://doi.org/10.1038/s41598-020-61849-8.

115. Trave C, Brunnschweiler J, Sheaves M, Diedrich A, Barnett A. Are we killing them with kindness? Evaluation of sustainable marine wildlife tourism. Biological conservation. 2017;209:211-22. doi: http://dx.doi.org/10.1016/j.biocon.2017.02.020.

116. Prieto R, Silva MA, Waring GT, Gonçalves JM. Sei whale movements and behaviour in the North Atlantic inferred from satellite telemetry. Endangered Species Research. 2014;26(2):103-13. doi: 10.3354/esr00630.

117. Clarke R. Sperm Whales of the Azores: University Press; 1956.

CHAPTER 6

6. DISCUSSION

Most large whales were subjected to severe overexploitation by commercial whaling and have not yet attained their historical numbers, remaining endangered or vulnerable [1, 2]. These whales now experience numerous other pressures from human activities [3, 4] that, acting singly or in combination, threaten the recovery of populations. Large whales are highly mobile and their range can extend throughout entire ocean basins. Yet, information on these pressures is only available for limited parts of whales' range, preventing an understanding of the consequences at the population level, and hindering implementation of effective management and regulation measures.

The waters around the Azores archipelago are an important feeding, calving and possibly mating ground for sperm whales in the North Atlantic [5]. The Azores are also an important migratory and mid-latitude foraging habitat for North Atlantic blue, fin and sei whales [6-8]. These whales support a thriving cetacean watching industry in the area that generates over 50M€ [9]. However, potential negative impacts of the local cetacean watching activity on these whales have long been a cause for concern. Cetacean watching elicit avoidance responses from targeted cetaceans that can disrupt individual's energy budgets. Repeated disturbance can lead to displacement from important habitats used for critical activities and reduced population fitness [10]. Recently, the number of ship collisions with sperm and baleen whales, as well as of observations of whales with recent injuries likely caused by interaction with ships, has increased in the Azores [11]. The known collision cases most likely underestimate the number of whale strikes, as many collisions are not reported and whale carcasses often sink [12]. This raised the concern that ship collisions may be an important source of mortality for whales in the region.

This thesis developed a robust modelling approach and tools to assess the exposure of large whales to cetacean watching and marine traffic. The application of these tools contributed new knowledge on the drivers and patterns of distribution of four whale species (sperm, blue, fin and sei whales) in the Azores (Chapter 4), and provided the first estimates for the region of the cetacean watching and marine traffic spatio-temporal patterns (Chapters 2 and 3), the risk of fatal collisions of these four whale species with large vessels (Chapter 5), and their exposure to cetacean watching activity (Chapter 5).

The main findings of the thesis are summarized and discussed below.

6.1. CETACEAN WATCHING AND MARINE TRAFFIC PATTERNS

The incorporation of Automatic Identification System (AIS) in vessels adopted by IMO in 2000 to increase the security of maritime operations has brought benefits in other aspects [13]. This system allows capturing a large volume of data, enabled identifying shipping traffic patterns by large vessels over the last two decades. However, in small coastal vessels, such as those that offer cetacean watching trips, the equipment has only started to be implemented recently.

Cetacean watching is a global billionaire business and is expected to continue to grow [14]. Yet, with few exceptions [15, 16], the effort of the activity and the exposure of cetaceans has not been determined, although this information is critical to monitor the activity, estimate the carrying capacity for a given population/area, or assess the effectiveness of management measures

AIS and other vessel positioning systems (e.g., Global Positioning System (GPS)), enable reconstructing the vessel trajectories but do not provide information on their activities. Using data from AIS and GPS systems installed on cetacean watching operators (that voluntarily agreed to it) and logbook data kindly shared by some operators, a method was developed to identify the behavioural state of cetacean watching vessels, enabling to distinguish periods of cetacean observation, from other activities (Chapter 2). The method inspired on Vespe et al. [17], estimated the speed density distribution of vessels engaged in different activities and applies these speed parameters to automatically identify vessel behaviour. Validation of the method using an independent dataset showed high accuracy, with 81% of cetacean observation periods, and 81-88% (depending on the boat) of transiting periods correctly identified.

The method enables a more accurate quantification of cetacean-effort and a more realistic representation of the areas where cetaceans are exposed to the activity. Importantly, the method is easily transferable to other cetacean watching areas, requiring only adjustment and validation of the speed parameters used to discern boat behaviour. Therefore, it can be a powerful tool for monitoring and managing cetacean watching activity.

Cetacean watching activity in zone A (Faial, Pico and S. Jorge islands) mainly concentrated north of Faial, in an area extending off the channel Faial-Pico to Cedros, and south of Faial and Pico, between Lajes do Pico and Castelo Branco (Chapter 2). Boats from Lajes do Pico only used the south of Pico, but boats from Horta and Madalena used all areas. The effort increased in spring, peaked in summer and decreased again in autumn [18-20]. Monthly maps showed that boats from Horta and Madalena used more the southern sector in April-June but increased the effort north of Faial in August. These results confirm empirical knowledge: i) cetacean watching activity tends to focus around the home ports of the operating vessels, especially off the areas where there are lookouts;) ii) vessels from the same home port exhibit similar spatial patterns; iii) boats from Horta, Madalena and Lajes do Pico overlap south of Pico.

AIS data was also used in Chapter 3 to characterize and map shipping traffic, with a focus on the traffic by large and fast vessels known to be responsible for most fatal collisions with large whales [21, 22]. A series of algorithms were specifically created to correct and standardize AIS data records, and to calculate metrics used to estimate shipping density patterns in the Azores. Results show that cargo ships and tankers were responsible for the majority of shipping traffic from 2012-2016, with fishing and passenger vessels, having also an important contribution. Four main shipping routes were identified within the study area: one to the north of the archipelago, one on the south, a route linking S. Miguel to the central group of islands, and another one connecting the islands within the central group. The highest shipping density was found close to the main ports (e.g., the Faial-Pico channel, the area around Ponta Delgada, in S. Miguel, and NE of Terceira), and along direct routes connecting all islands. Two trans-Atlantic routes of cargos and tankers crossed the Azores, one north of the archipelago and the other south of S. Miguel. These four main routes were observed throughout the study, with little variation.

Large vessels (>80 m long) travelling at high speeds (> 14 kn) have been recorded more often in fatal collisions with large whales [21, 22]. These two criteria were therefore used to filter the AIS data to describe shipping patterns associated with a higher risk of lethal collisions (Chapter 3). Most cargo and tankers detected in the AIS records were at least 80 m in length. As these vessels were responsible for most traffic in the area, the patterns of traffic with high risk of lethal collisions were similar to those of total traffic, although shipping density was naturally lower. Passenger and HSC vessels also contributed to shipping density with increased probability of lethal collisions, especially in the summer.

6.2. MODELLING WHALE DISTRIBUTION

6.2.1. INCORPORATING PREY COVARIATES IN WHALE MODELS

Modelling species' habitat preferences is critical to understand their spatio-temporal dynamics and to assess their exposure to human activities. A major impediment to assessing whale overlap with cetacean watching activities and marine traffic in the Azores has been the lack of accurate predictions of whale density at relevant spatial and temporal scales. Previous models of whale habitat preferences in the Azores relied on oceanographic (e.g., temperature, chlorophyll-*a* concentration) and static (e.g., depth, bottom slope) variables [23, 24]. These models lacked information on prey distribution, a critical driver of the distribution of top predators, including cetaceans [25, 26]. The absence of prey data is believed to have affected the ability of these models to describe the dynamics of whale populations in the region at finer scales.

Prey data is rarely included in marine top predator models because of the difficulty of simultaneously sampling prey and predator [25-27]. In this work, the whale distribution models included simulated prey variables from the SEAPODYM model [28], in addition to a selection of environmental variables (Chapter 4). The SEAPODYM predicts the biomass and production of mesozooplankton organisms (i.e. holo- and mero-zooplankton) and six functional groups of micronekton (small (2-20 cm) pelagic fish, cephalopods and crustaceans), in different depth layers from the surface down to 1,000m depth [29]. Selection of prey data to be included in the models was based on the diet preferences and diving behaviour of each whale species.

Surprisingly, none of the prey variables selected was retained in the best model for the blue whale, while sei whale relative density increased in waters with low biomass of upper-mesopelagic layer. Relative density of sperm and fin whales increased with increasing upper-mesopelagic production and epipelagic biomass, respectively. But even when significant, prey variables were never the most important covariates at explaining whale density, with the exception of fin whales model in Chapter 5.

SEAPODYM model outputs may not represent well the main prey of the whales studied here. This likely is the case of the deep water squid preyed on by sperm whales, most of which are not part of the micronekton, and have a small contribution to the acoustic backscatter at the frequencies used for parameter optimization in SEAPODYM (Lehodey et al. 2014). Even so, upper-mesopelagic production was retained in the sperm whale model, in contrast to what happened in other areas, where all prey variables from SEAPODYM were discarded [30]. Still, the wide confidence intervals of the smoother for upper-mesopelagic production reveal a very high uncertainty in the results.

Manocci et al [30] showed that production of epipelagic micronekton was the most important covariate in fin whale density models and the fourth most important covariate in sei whale density models developed for different locations in the northwest, central and northeast Atlantic [30]. Distribution models of blue, fin and sei whales instrumented with satellite telemetry tags in the Azores showed a link to zooplankton, rather than epipelagic micronekton [31]. Hence, in contrast to what happens with sperm whales, SEAPODYM may represent at least to some extent the planktonic and micronekton organisms consumed by baleen whales in the Azores and elsewhere. Nevertheless, the latter study showed that blue and sei whales were associated to intermediate values of zooplankton production or biomass from the previous month. These results suggest a temporal mismatch between SEAPODYM zooplankton variables and whale distribution, or that prey metrics output from SEAPODYM are not the most relevant for baleen whales. In fact, patch density, rather than total prey abundance or biomass, appear more important at driving foraging effort and energy intake of baleen whales [32].

Finally, the spatial scale and resolution of the analysis of Perez-Jorge [31] was totally different from the one used here. In addition to prey abundance, the prey spatial scale also influences the predator's response. Some studies suggest that predator-prey relationships are better detected on a fine-scale [26, 33], and others suggest that the scales should be larger than the prey aggregations [34].

Despite these shortcomings, prey covariates improved the predictive performance of the models in which they were retained, and contributed to unveil the dynamics of whale distribution in the area.

6.2.2. WHALE DISTRIBUTION AND HABITAT PREFERENCES

The local density of large whales was strongly associated with dynamic variables linked to biological productivity, prey availability, and as well as depth, reflecting the tight relationship between whale occurrence in the Azores and the availability of food in the area (Chapter 4 and 5). The models also were influenced by other power explanatory variables see more details in 6.2.3.Effect of marine traffic on whale distribution (Chapter 5).

The whales's preference in the Azores by dynamic variables reflect their migratory status in the area.

In this thesis the sperm whale's preference by the food availability was not that important as for the baleen whales. This can be influence by the more complex relationship between the predator and prey, by the level in the trophic chain or by the need of include bigger prey size for sperm whale [35].

Besides, the baleen whales preferred productivity lagged variables. The temporal delay between primary productivity and the development of the whales prey are linked to their life cycles [36, 37]. The main prey of the baleen whales and sperm whales have different temporal delays [36, 38]; the analysed periods here it could be more accurate for the baleen whales than for sperm whales. Therefore, the lagged relationship between prey and predator would be dismissed for sperm whales.

These whale species occasionally have several feed strategies in one region and sometimes linked to age or gender [39-41]. The sperm whale is the most extreme case, because the adult males use to feed on bigger size animals and include more fish than the rest of their congeners [35, 42].

The depth preference characterize the species distribution, for the sperm whale is an interval and for the blue whale a linear increment. The relationship between the blue whale and this variable should show their migratory patterns, it would be more individuals away from the shoreline and for the sperm whales would reflect their link with submarine irregularity, like seamounts [43].

6.2.3. EFFECT OF MARINE TRAFFIC ON WHALE DISTRIBUTION

Animal distribution can be influenced by anthropogenic activities, directly by disturbing or changing their behaviour, or indirectly through changes in their ecosystem (e.g., physical changes to the habitat, prey or predator distribution) [32, 33]. Although explicitly modelling the effects of human activities on animal distribution has been common practice in terrestrial systems for some time (e.g. [44, 45]), only recently it started to be attempted in marine organisms [46, 47].

Vessels of all sizes are known to elicit an array of behavioural changes in cetaceans [48], particularly large whales [49]. These responses range from changes in swimming speed and direction, and in diving patterns, increases in aerial behaviour, mother-calf separation, cessation of feeding and calf nursing, and displacement from preferred feeding sites [50].

In Chapter 5, shipping density was incorporated as a covariate in whale SDMs, to understand its influence on whale distribution at broad spatial and temporal scales. Because shipping density was only available for four years, the cetacean and effort data had to be subset to match the same period, resulting in a considerable reduction in sample size, and in the selection of different covariates. Thus, predictions from these models cannot be compared with those from Chapter 4. Shortening the time period covered also implied that the number of sightings of blue and sei whales were insufficient to fit the model with shipping density.

Results show that sperm whale relative density was significantly influenced by shipping density, with density decreasing as shipping density of large vessels (> 80 m) travelling at high speeds (>14 kn) increased. These findings suggest that sperm whales avoid areas of high marine traffic (Chapter 5). It is unknown whether this implies that sperm whales were displaced from preferred habitats. Future studies should try to investigate this and to estimate the energetic consequences for sperm whales of avoiding high traffic areas.

In contrast, there was no evidence that shipping density influenced fin whale relative density (Chapter 5), despite the co-occurrence of whales (Chapter 4) and ships (Chapter 3) throughout most of the area. Presently, it cannot be ascertained if the lack of a relationship was an artifact from the low sample size used to fit the model. Nevertheless, these results do not mean that fin whales do not respond to traffic. Responses may be temporary and subtle, and would therefore require fine-scale sampling to detect. In any case, these findings suggest that fin whales may be at higher relative risk of being disturbed, given the potential for multiple encounters with vessels. Evasive responses to vessels likely increase the energetic costs of whales and may force them to interrupt or cease their previous activity. Fin whales use the Azores for foraging during migration [6], a life stage during which individuals must carefully balance their energy budget, as they cross vast ocean areas where feeding opportunities are reduced [51]. Repeated interruption of feeding can thus have a cumulative adverse effect on the energy balance of migratory fin whales. Additionally, shipping noise can disrupt fin whale acoustic behaviour, mask their vocalisations thus reducing their communication space, and increase stress levels [52].

Incorporating shipping density into whale distribution models enabled a first insight into the responses of large whales to marine traffic at large spatial and temporal scales. This approach could be applied to sighting and AIS data at smaller resolutions (e.g., daily time scales), as well as to satellite-tracked movements and vessel activity.

6.2.4. WHALE EXPOSURE TO CETACEAN WATCHING AND RISK OF LETHAL SHIP COLLISIONS

A conservation strategy commonly applied to cetaceans is to protect areas where populations are known to be vulnerable to specific threats, or protect their core habitats [53, 54]. Risk-based approaches are increasingly used to assess potential impacts from anthropogenic activities on wildlife and to support area-based management efforts (e.g., Merchant *et al* (2018) [55]).

Borrowing from this concept, in Chapter 5, estimates of cetacean watching effort produced using the methods and data described in Chapter 2 were combined with predictions of whale density (Chapter 4), to produce an exposure map for each whale species and calculate the percentage of suitable habitat where exposure occurred. Habitat exposure to cetacean watching activity varied among species, ranging from a maximum of 25% in fin whales to 8% in blue whales. These results suggest that proportion of suitable habitat within the region exposed to whale-watching was relatively small for most species, and negative effects at the population level are likely to be minor, especially considering that none of the species is resident in the area year-round. In the case of fin whales, the analysis showed fewer high-quality habitats available to move away and avoid being exposed to the activity. Future analysis of the carrying capacity of the cetacean watching activity should [46] take special attention to this species.

Risk maps of lethal collisions revealed that collision hotspots were mainly on the trans-Atlantic traffic lanes, and on the routes connecting the Azorean islands groups. Overall risk was greater for sperm whales, the species with higher relative density in the area, followed by fin, blue and sei whales. In some areas, risk of lethal collisions and exposure to cetacean watching overlap. The two activities may exert accumulating and interacting impacts on large whales, both in terms of physical disturbance and underwater noise, and it is critical to consider these pressures collectively [56].

The risk of shipping traffic to whale populations can be reduced in various ways, for example, by re-routing existing traffic routes to less suitable habitats. However, this work showed that high-density areas of these whale species were not distributed in a single nucleus but were spread throughout the region. So protecting them is more complicated. A more feasible solution could be restricting the speed of the large vessels [57] in areas where whales were more abundant, to reduce the severity of potential collisions and disturbances [58]. This would also benefit other taxa as the relationship between speed and the severity of injuries is the same for a wide variety of marine species [59, 60]. Nevertheless, any potential measures to reduce collisions with marine taxa must be carefully evaluated before and after to assess its effectiveness [61-63]. Therefore, it is essential to implement a regional program to record and monitor collisions with marine wildlife.

6.3. FINAL CONSIDERATIONS

This work provides a set of modelling tools to characterize maritime activities, cetacean habitat preferences and distribution, and the interaction between the two. The algorithms developed in Chapter 2 enable identifying areas of cetacean watching and quantifying the effort of the activity using only boat positioning data. These algorithms can be applied to data from anywhere in the world. The tools for processing large volumes of AIS data and the filters created to select records fitting to vessel speed and length (Chapter 3) have made it possible to map shipping traffic and identify areas with the highest concentration of ships associated with a greater risk of lethal collisions. SDMs developed in Chapter 4 improved the understanding of the spatio-temporal dynamics of cetaceans and enabled estimating their exposure to cetacean watching and a marine traffic. Together, these tools can increase the effectiveness of conservation measures for these whale species, allowing managers and decision-makers to prioritize conservation actions and focus in areas identified as having the greater risk.

These tools can be refined and fed with additional data and new sources of data. Incorporation of AIS in smaller vessels (including cetacean watching boats) would permit a more robust analyses of the influence that ships have on cetacean distribution and on the risk of lethal and non-lethal collisions, as well evaluating the risk to other pressures (e.g., noise, oil spills, gas emissions). Fitting AIS equipment in cetacean watching boats would help characterize the activity at the scale of the Azores archipelago and even in other parts of cetacean's range enabling a better understanding on the exposure of the populations across their range, and using these data to develop cetacean SDMs. SEAPODYM is being constantly improved and recent developments output prey variables at much lower resolutions than the ones used, which can significantly increase the accuracy of predictions.

The Azores archipelago depends on the sea. In the future, it is expected that anthropogenic activities in the marine environment continue to expand and new ones emerge. The modelling tools developed in this dissertation can assist reducing the impact of these activities on cetacean species.

6.4. REFERENCES

1. Carroll G, Hedley S, Bannister J, Ensor P, Harcourt R. No evidence for recovery in the population of sperm whale bulls off Western Australia, 30 years post-whaling. Endangered Species Research. 2014;24(1):33-43.
2. Garrigue C, Derville S, Bonneville C, Baker CS, Cheeseman T, Millet L, et al. Searching for humpback whales in a historical whaling hotspot of the Coral Sea, South Pacific. Endangered Species Research. 2020;42:67-82. doi: https://doi.org/10.3354/esr01038.
3. Nicol C, Bejder L, Green L, Johnson C, Keeling L, Noren D, et al. Anthropogenic threats to wild cetacean welfare and a tool to inform policy in this area. Frontiers in veterinary science. 2020;7:57. doi: doi: 10.3389/fvets.2020.00057.
4. Dias LA, Litz J, Garrison L, Martinez A, Barry K, Speakman TJESR. Exposure of cetaceans to petroleum products following the Deepwater Horizon oil spill in the Gulf of Mexico. 2017;33:119-25.
5. Clarke R. Sperm Whales of the Azores: University Press; 1956.
6. Silva MA, Prieto R, Jonsen I, Baumgartner MF, Santos RS. North Atlantic blue and fin whales suspend their spring migration to forage in middle latitudes: building up energy reserves for the journey? PloS one. 2013;8(10):e76507. doi: 10.1371/journal.pone.0076507.
7. Prieto R, Silva MA, Waring GT, Gonçalves JM. Sei whale movements and behaviour in the North Atlantic inferred from satellite telemetry. Endangered Species Research. 2014;26(2):103-13. doi: 10.3354/esr00630.
8. Visser F, Hartman KL, Pierce GJ, Valavanis VD, Huisman J. Timing of migratory baleen whales at the Azores in relation to the North Atlantic spring bloom. Marine Ecology Progress Series. 2011;440:267-79. doi: doi: 10.3354/meps09349.
9. Ressurreiçao A, Cardigos F, Giacomello E, Leite N, Oliveira F, Kaiser MJ, et al. The value of marine ecotourism for an European outermost region. Ocean and Coastal Management. 2022;222:106129. doi: https://doi.org/10.1016/j.ocecoaman.2022.106129.
10. Christiansen F, Rasmussen M, Lusseau D. Whale watching disrupts feeding activities of minke whales on a feeding ground. Marine Ecology Progress Series. 2013;478:239-51. doi: 10.3354/meps10163.
11. DRPM. Rede de Arrojamento de Cetáceos dos Açores: Direção Regional de Políticas Marítimas; 2021 [cited 2021]. Available from: https://servicos-sraa.azores.gov.pt/doit/servicos.asp?id_dep=10&id_form=84.
12. Peltier H, Beaufils A, Cesarini C, Dabin W, Dars C, Demaret F, et al. Monitoring of Marine Mammal Strandings Along French Coasts Reveals the Importance of Ship Strikes on Large Cetaceans: A Challenge for the European Marine Strategy Framework Directive. 2019;6(486). doi: 10.3389/fmars.2019.00486.
13. Robards M, Silber G, Adams J, Arroyo J, Lorenzini D, Schwehr K, et al. Conservation science and policy applications of the marine vessel Automatic

Identification System (AIS)—a review. Bulletin of Marine Science. 2016;92(1):75-103. doi: http://dx.doi.org/10.5343/bms.2015.1034.

14. Hoyt E, Parsons E. The whale-watching industry. In: Laetitia Nunny WAW, La Garriga, Spain, editor. Whale-watching: Sustainable tourism ecological management. 57. First ed. OceanCare2014. p. 86-96.

15. Almunia J, Delponti P, Rosa F. Using Automatic Identification System (AIS) Data to Estimate Whale Watching Effort. Frontiers in Marine Science. 2021:827. doi: 10.3389/fmars.2021.635568.

16. Almunia J, Delponti P, Rosa F. Using big data to estimate whale watching effort. bioRxiv. 2020:1-19. doi: http://doi.org/10.1101/2020.11.30.403923.

17. Vespe M, Gibin M, Alessandrini A, Natale F, Mazzarella F, Osio GC. Mapping EU fishing activities using ship tracking data. Journal of Maps. 2016;12(sup1):520-5. doi: 10.1080/17445647.2016.1195299.

18. Magalhães S, Prieto R, Silva MA, Gonçalves J, Afonso-Dias M, Santos RS. Short-term reactions of sperm whales (Physeter macrocephalus) to whale-watching vessels in the Azores. Aquatic Mammals. 2002;28(3):267-74.

19. Bentz J, Lopes F, Calado H, Dearden P. Managing marine wildlife tourism activities: Analysis of motivations and specialization levels of divers and whale watchers. Tourism Management Perspectives. 2016;18:74-83. doi: https://doi.org/10.1016/j.tmp.2016.01.004.

20. Sequeira M, Elejabeitia C, Silva M, Dinis A, de Stephanis R, Urquiola E, et al. Review of whalewatching activities in mainland Portugal, the Azores, Madeira and Canary archipelagos and the Strait of Gibraltar. Cetacean Research and Management. 2009; SC61/WW11:1-41.

21. Laist DW, Knowlton AR, Mead JG, Collet AS, Podesta M. Collisions between ships and whales. Marine Mammal Science. 2001;17(1):35-75. doi: https://doi.org/10.1111/j.1748-7692.2001.tb00980.x.

22. Jensen AS, Silber GK. Large whale ship strike database. U.S. Department of Commerce, NOAA 2003. p. 37.

23. Tobeña M, Prieto R, Machete M, Silva MA. Modeling the potential distribution and richness of cetaceans in the Azores from fisheries observer program data. Frontiers in Marine Science. 2016;3:202. doi: doi: 10.3389/fmars.2016.00202.

24. Prieto R, Tobeña M, Silva MA. Habitat preferences of baleen whales in a mid-latitude habitat. Deep Sea Research Part II: Topical Studies in Oceanography. 2017;141:155-67. doi: https://doi.org/10.1016/j.dsr2.2016.07.015.

25. Redfern J, Ferguson MC, Becker E, Hyrenbach K, Good C, Barlow J, et al. Techniques for cetacean–habitat modeling. Marine Ecology Progress Series. 2006;310:271-95. doi: doi:10.3354/meps310271.

26. Torres LG, Read AJ, Halpin P. Fine-scale habitat modeling of a top marine predator: Do prey data improve predictive capacity. Ecological Applications. 2008;18(7):1702-17. doi: https://doi.org/10.1890/07-1455.1.

27. Watanabe YY, Takahashi A. Linking animal-borne video to accelerometers reveals prey capture variability. Proceedings of the National Academy of Sciences. 2013;110(6):2199-204. doi: https://doi.org/10.1073/pnas.1216244110.

28. Lehodey P, Senina I, Murtugudde RJPiO. A spatial ecosystem and populations dynamics model (SEAPODYM)–Modeling of tuna and tuna-like populations. 2008;78(4):304-18. doi: 10.1016/j.pocean.2008.06.004.

29. Lehodey P, Murtugudde R, Senina I. Bridging the gap from ocean models to population dynamics of large marine predators: a model of mid-trophic functional groups. Progress in Oceanography. 2010;84(1-2):69-84.

30. Mannocci L, Roberts JJ, Miller DL, Halpin PN. Extrapolating cetacean densities to quantitatively assess human impacts on populations in the high seas. Conservation Biology. 2017;31(3):601-14. doi: 10.1111/cobi.12856.

31. Pérez-Jorge S, Tobeña M, Prieto R, Vandeperre F, Calmettes B, Lehodey P, et al. Environmental drivers of large-scale movements of baleen whales in the mid-North Atlantic Ocean. Diversity Distributions. 2020;26(6):683-98.

32. Goldbogen JA, Calambokidis J, Shadwick RE, Oleson EM, McDonald MA, Hildebrand JA. Kinematics of foraging dives and lunge-feeding in fin whales. Journal of Experimental Biology. 2006;209(7):1231-44. doi: 10.1242/jeb.02135.

33. Ishii M, Murase H, Fukuda Y, Sawada K, Sasakura T, Tamura T, et al. Diving behavior of sei whales Balaenoptera borealis relative to the vertical distribution of their potential prey. Mammal study. 2017;42(4):1-9. doi: https://doi.org/10.3106/041.042.0403.

34. Rose GA, Leggett WC. The importance of scale to predator-prey spatial correlations: An example of Atlantic fishes. Ecology. 1990;71(1):33-43. doi: https://doi.org/10.2307/1940245

35. Clarke M, Martins H, Pascoe P. The diet of sperm whales (Physeter macrocephalus Linnaeus 1758) off the Azores. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 1993;339(1287):67-82. doi: https://doi.org/10.1098/rstb.1993.0005.

36. Szesciorka AR, Ballance LT, Širović A, Rice A, Ohman MD, Hildebrand JA, et al. Timing is everything: Drivers of interannual variability in blue whale migration. Scientific reports. 2020;10(1):1-9. doi: 10.1121/1.5136714

37. Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, Bograd SJ, et al. Tracking apex marine predator movements in a dynamic ocean. Nature. 2011;475(7354):86-90. doi: 10.1038/nature10082.

38. Bruggeman J, Jacobs ZL, Popova E, Sauer WH, Gornall JM, Brewin RJ, et al. The paralarval stage as key to predicting squid catch: Hints from a process-based model. Deep Sea Research Part II: Topical Studies in Oceanography. 2022:105123.

39. Víkingsson GA. Feeding of fin whales (Balaenoptera physalus) off Iceland-diurnal and seasonal variation and possible rates. Journal of Northwest Atlantic Fishery Science. 1997;22:77-89.

40. Christensen I, Haug T, Øien N. A review of feeding and reproduction in large baleen whales (Mysticeti) and sperm whales Physeter macrocephalus in Norwegian and adjacent waters. Fauna Norvegica Series A. 1992;13:39-48.

41. Jory C, Lesage V, Leclerc A, Giard J, Iverson S, Bérubé M, et al. Individual and population dietary specialization decline in fin whales during a period of ecosystem shift. Scientific reports. 2021;11(1):1-14. doi: https://doi.org/10.1038/s41598-021-96283-x.

42. Flinn RD, Trites AW, Gregr EJ, Perry RI. Diets of fin, sei, and sperm whales in British Columbia: an analysis of commercial whaling records, 1963–1967. Marine Mammal Science. 2002;18(3):663-79. doi: https://doi.org/10.1111/j.1748-7692.2002.tb01065.x.

43. Wong SN, Whitehead H. Seasonal occurrence of sperm whales (Physeter macrocephalus) around Kelvin Seamount in the Sargasso Sea in relation to

oceanographic processes. Deep Sea Research Part I: Oceanographic Research Papers. 2014;91:10-6.

44. Morán-Ordóñez A, Suárez-Seoane S, Elith J, Calvo L, de Luis E. Satellite surface reflectance improves habitat distribution mapping: a case study on heath and shrub formations in the Cantabrian Mountains (NW Spain). Diversity

Distributions. 2012;18(6):588-602.

45. Robinson L, Elith J, Hobday A, Pearson R, Kendall B, Possingham H, et al. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Global Ecology

Biogeography. 2011;20(6):789-802.

46. Lundquist D, Gemmell N, Würsig B, Markowitz T. Dusky dolphin movement patterns: short-term effects of tourism. New Zealand Journal of Marine

Freshwater Research. 2013;47(4):430-49.

47. Currie JJ, McCordic JA, Olson GL, Machernis AF, Stack SH. The impact of vessels on humpback whale behavior: the benefit of added whale watching guidelines. Frontiers in Marine Science. 2021;8:601433.

48. Nowacek DP, Thorne LH, Johnston DW, Tyack PL. Responses of cetaceans to anthropogenic noise. Mammal Review. 2007;37(2):81-115. doi: https://doi.org/10.1111/j.1365-2907.2007.00104.x.

49. Croll DA, Clark CW, Calambokidis J, Ellison WT, Tershy BR, editors. Effect of anthropogenic low-frequency noise on the foraging ecology of Balaenoptera whales. Animal Conservation forum; 2001: Cambridge University Press.

50. Smultea MA. Behavioral ecology of cetaceans in the Southern California Bight 2016.

51. Lockyer C, Brown S. The migration of whales. In: Aidley DJ, editor. Animal migration. 13. Cambridge: Cambridge University Press New York, NY; 1981. p. 105-37.

52. Erbe C, Dunlop R, Dolman S. Effects of noise on marine mammals. In: Hans Slabbekoorn RJD, Arthur N. Popper, Richard R. Fay, editor. Effects of anthropogenic noise on animals. Springer Handbook of Auditory Research 66. First ed: Springer; 2018. p. 277-309.

53. Frantzis A, Leaper R, Alexiadou P, Prospathopoulos A, Lekkas D. Shipping routes through core habitat of endangered sperm whales along the Hellenic Trench, Greece: Can we reduce collision risks? PloS one. 2019;14(2):e0212016. Epub 2019/02/28. doi: 10.1371/journal.pone.0212016. PubMed PMID: 30811429; PubMed Central PMCID: PMC6392247 not alter our adherence to PLOS ONE policies on sharing data and materials.

54. Panigada S, Notarbartolo Di Sciara G, Zanardelli Panigada M. Fin whales summering in the Pelagos Sanctuary (Mediterranean Sea): Overview of studies on habitat use and diving behaviour. Chemistry ecology. 2006;22(sup1):S255-S63. doi: 10.1080/02757540600720680.

55. Merchant ND, Faulkner RC, Martinez R. Marine noise budgets in practice. Conservation Letters. 2018;11(3):e12420. doi: 10.1111/conl.12420.

56. Murray CC, Hannah LC, Doniol-Valcroze T, Wright BM, Stredulinsky EH, Nelson JC, et al. A cumulative effects model for population trajectories of resident killer whales

in the Northeast Pacific. Biological Conservation. 2021;257:109124. doi: 10.1016/j.biocon.2021.109124.

57. Silber GK, Vanderlaan AS, Arceredillo AT, Johnson L, Taggart CT, Brown MW, et al. The role of the International Maritime Organization in reducing vessel threat to whales: process, options, action and effectiveness. Marine Policy. 2012;36(6):1221-33. doi: https://doi.org/10.1016/j.marpol.2012.03.008.

58. Vanderlaan AS, Taggart CT. Vessel Collisions with Whales: The Probability of Lethal Injury Based on Vessel Speed. Marine Mammal Science. 2007;23(1):144-56. doi: 10.1111/j.1748-7692.2006.00098.x.

59. Vanderlaan AS, Taggart CT, Serdynska AR, Kenney RD, Brown MW. Reducing the risk of lethal encounters: vessels and right whales in the Bay of Fundy and on the Scotian Shelf. Endangered Species Research. 2008;4(3):283-97. doi: 10.3354/esr00083.

60. Schoeman RP, Patterson-Abrolat C, Plön S. A Global Review of Vessel Collisions With Marine Animals. Frontiers in Marine Science. 2020;7(292). doi: 10.3389/fmars.2020.00292.

61. van der Hoop J, Barco SG, Costidis AM, Gulland FM, Jepson PD, Moore KT, et al. Criteria and case definitions for serious injury and death of pinnipeds and cetaceans caused by anthropogenic trauma. Diseases of aquatic organisms. 2013;103(3):229-64. doi: 10.3354/dao02566.

62. McKenna MF, Katz SL, Condit C, Walbridge S. Response of commercial ships to a voluntary speed reduction measure: are voluntary strategies adequate for mitigating ship-strike risk? Coastal Management. 2012;40(6):634-50. doi: https://doi.org/10.1080/08920753.2012.727749.

63. Vanderlaan AS, Taggart CT. Efficacy of a voluntary area to be avoided to reduce risk of lethal vessel strikes to endangered whales. Conservation biology : the journal of the Society for Conservation Biology. 2009;23(6):1467-74. Epub 2009/09/25. doi: 10.1111/j.1523-1739.2009.01329.x. PubMed PMID: 19775275.

List of figures

CHAPTER 1

Figure 1. 1. Study area map. Az: Azores Current, CaC: Canary Current. M.A.R.: Mid-Atlantic Ridge. 9 islands: Corvo, Flores, Faial, Pico, São Jorge, Graciosa, Terceira, São Miguel and Santa Maria island 3

CHAPTER 2

Figure 2. 1. Study area. Top: Azores archipelago. Red boxes: exclusion boxes for the ports and other marine activities used to identify and classify different trips (refer to text for further details). Bottom: Detail of the study area. In light grey: Cetacean watching recorded positions.. 19

Figure 2. 2. Data pre-processing and standardization. Prior to analysis, data from the AIS and GPS were cleaned and standardized using custom R scripts. After pre-processing, data were merged into a single database. R packages used to run the scripts are shown. Information on software used to process raw GPS and AIS data is given in the main text................................ 21

Figure 2. 3. Homeports of cetacean watching in the study area and the positions of the lookout hotspots used to detect cetaceans for which data is available 23

Figure 2. 4. Geographic positions recorded for 1,335 cetacean watching trips used to fit the GMM predictive model. 25

Figure 2. 5. Daily variation in the time interval (s) between consecutive geographic positions per boat. The x-axis shows the julian day, from the first day of the cetacean watching season...... 26

Figure 2. 6. Duration (left) and length (right) of the cetacean watching trips (min and km respectively) based on the home port (H: Horta, L: Lajes, and M: Madalena). The boxplot shows the median (horizontal bar), the interquartile range (IQR) (box), 1.5 x IQR (whiskers) and outliers (dots) 26

Figure 2. 7. Area used by the cetacean watching boats monitored in 2013 and 2018. The contour lines represent the 90% kernel utilization distribution of boats from Horta (red), Madalena (blue) and Lajes (green) 27

Figure 2. 8. Example of Feature significance for kernel density estimation for one boat. The graph shows how speed (y-axis) varied with the duration of trip (cumulative sum of time, csum). Red and yellow indicate higher and lower density estimates, while the green contours flag significant KDE gradient regions... 28

Figure 2. 9. Spatial pattern of cetacean watching effort from boats from Lajes in 2013 (top) and Madalena and Horta (bottom). The kernel density of the number of trips was calculated by 10 km x 10 km cells, and scaled to 0-1. Effort was calculated from GPS data. 30

Figure 2. 10. Monthly variation in the spatial pattern of cetacean watching effort from boats from Lajes in 2013. May: (top left), June (middle left), July (bottom left), August (top right), September (middle right), October (bottom right).The kernel density of the number of trips was calculated by 10 km x 10 km cells, and scaled to 0-1. Effort was calculated from GPS data...... 31

Figure 2. 11. Monthly variation in the spatial pattern of cetacean watching effort from boats from Horta and Madalena in 2018. May: (top left), June (middle left), July (bottom left), August (top right), September (middle right), October (bottom right). The kernel density of the number of trips was calculated by 10 km x 10 km cells, and scaled to 0-1. Effort was calculated from AIS and GPS data 32

CHAPTER 3

Figure 3. 1. Study area defined by the vhf coverage. 42

Figure 3. 2. Workflow to map shipping density per month and year. 46

Figure 3. 3. Annual number of vessels in the study area per vessel type (October 2012 -December 2016). Military vessel (Military v.), other vessels (Other), pleasure boat (Pleasure b.), vessels used for port services (Port), High-speed craft (HSC), and rescue vessel (Rescue v.) 47

Figure 3. 4. Annual number of trips in the study area per vessel type (October 2012 -December 2016). Military vessel (Military v.), other vessels (Other), pleasure boat (Pleasure b.), vessels used for port services (Port), High-speed craft (HSC), and rescue vessel (Rescue v.) 47

Figure 3. 5. Monthly variation in the number (n) of vessels (left) and trips (right) (October 2012 -December 2016). Black 2012; orange 2013; blue 2014; green 2015 and yellow 2016. Left to right, January to December. 48

Figure 3. 6. Monthly variation in the number of trips per year for each vessel (October 2012 - December 2016). 49

Figure 3. 7. Mean and standard deviation (SD) of daily shipping density for total (TOP) and LP shipping (bottom). 51

Figure 3. 8. Mean daily shipping density for the study period (January 2013 - December 2016). 52

Figure 3. 9. Monthly patterns in mean daily shipping density (October 2012 – December 2016). 53

Figure 3. 10. Mean daily shipping density per vessel type (October 2012 – December 2016). The density scale is the same for all maps. 55

Figure 3. 11. Shipping density on 2nd January (LEFT) and 1st July (RIGHT of 2013) 56

Figure 3. 12. Mean daily LP shipping density (speed >14 kn, length >80 m; January 2013 – December 2016). 57

Figure 3. 13. Annual patterns in mean daily LP shipping density (speed >14 kn, length >80 m; January 2013 – December 2016). 57

Figure 3. 14. Mean daily LP shipping density per vessel type (speed >14 kn, length >80 m; October 2012 – December 2016). 59

CHAPTER 4

Figure 4. 1. Cetacean on-effort positions of POPA transects between 2001 and 2016 around the azores. 71

Figure 4. 2. Gam smoothers for sperm whales: on top left depth and on right distance to small seamounts (ds-seamounts), on down left month and right upper-mesopelagic(pp), meso_pp. Smooth function are shown as a solid line, and shaded regions represent 95% confidence intervals. 83

Figure 4. 3. Smoothed functions for the selected covariates for fin whale: on top left time-lagged npp (_2 month), npp (_2m) and right epipelagic micronekton (pb), on down left time-lagged chlorophyll-a concentration (_1 month), chla(_1m). Smooth function (solid line), shaded regions (95% confidence intervals). 84

Figure 4. 4. Smoothed functions for the selected covariates for blue whale: on top left logarithm of sea surface temperature, sstl; and on right logarithm of depth, depthl; on down left time-lagged chlorophyll-a concentration (_1 month), chla(_1m). Smooth function (solid line), shaded regions (95% confidence intervals). 85

Figure 4. 5. Smoothed functions for the selected covariates for sei whale: on left time-lagged npp (_1 month), npp (_1m), and on right; upper-mesopelagic (pb), meso_pb. Smooth function (solid line), shaded regions (95% confidence intervals). 86

Figure 4. 6. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sperm whales based on the train dataset........ 86

Figure 4. 7. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sperm whales from May to August based on the train dataset........ 87

Figure 4. 8. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of fin whales based on the train dataset. 88

Figure 4. 9. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of fin whales from May to August based on the train dataset........ 89

Figure 4. 10. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of blue whales based on the train dataset. 90

Figure 4. 11. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of blue whales from May to August based on the train dataset. 91

Figure 4. 12. Predicted mean density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sei whales based on the train dataset. 92

Figure 4. 13. Predicted density (logarithm of individuals per 27,5 km2 (1/4 degree 2)) of sei whales from May to August based on the train dataset........ 93

CHAPTER 5

Figure 5. 1. Location of the surveys conducted by the POPA programme (grey circles) from 2012 to 2016 in the Azores archipelago. Zone a (blue) delimits one of the cetacean watching activity areas and was the focus of this study. 112

Figure 5. 2. Gam smoothers for the best fitting model (sdm-p-aa-2) for sperm whales: top left) shipping density at >14kn and >80m, top right) square of euclidean distance to fronts (ed_fronts), bottom left) epipelagic micronekton production (epi_pp), and bottom right) highly migrant Lower-mesopelagic production (hmbathy_pp). Smooth functions are shown as a solid line, and shaded regions represent 95% confidence intervals. 118

Figure 5. 3. Gam smoothers for fin whales (sdm-p-es) on shipping extent: top left) lower trophic level plankton (pp), pk_pp, top right) epipelagic micronekton (pp), epi_pp, and bottom left) sea water potential temperature, temperature. Smooth functions are shown as a solid line, and shaded regions represent 95% confidence. 119

Figure 5. 4. Mean predicted relative density of sperm whales (on top) and fin whales (on bottom) (individuals per 100 km2) for 2012-2016 based on the training dataset for SDM-p-aa2 and SDM-p-es respectively. 120

Figure 5. 5. Risk of lethal ship collisions for sperm (on top) and fin whales (on bottom). The risk was calculated using sperm whale predicted mean density and fin whale (individuals per 100 km2) based on the SDM-p-aa2 model and the SDM-p-es model, respectively, the density of ships >80m travelling at >14kn. 122

Figure 5. 6. Risk of lethal ship collisions for blue (on top) and sei whales (on bottom). The risk was calculated using each whale predicted mean density (individuals per 100 km2) based on the each SDM-p model from chapter 4 and the density of ships >80m travelling at >14kn. 123

Figure 5. 7. Exposure to cetacean watching activity for sperm (on top) and fin whales (on bottom). The exposure was calculated using fin whale predicted mean density (individuals per 100 km2) based on the SDM-p-aa and SDM-p-es model, respectively, and the cetacean watching effort estimated in chapter 2. Grey areas represent zones excluded from the predictions due to the lack of cetacean watching data (zones b and c). 126

Figure 5. 8. Exposure to cetacean watching activity for blue (on top) and sei whales (on bottom). The exposure was calculated using each whale predicted mean density (individuals per 100 km2) based on the each SDM-p from chapter 4 model and the cetacean watching effort estimated in chapter 2. Grey areas represent zones excluded from the predictions due to the lack of cetacean watching data (zones b and c). 127

List of tables

CHAPTER 2

Table 2. 1. Dataset used to calculate cetacean watching effort. Boat: each boat was identify by a unique number (except boat 3 for which the same GPS was shared by two different boats); Boat length (m); Type: RIB (rigid inflatable boat) or cabin; Year: year data recorded; Month_i: initial month of monitoring; Month_f: final month of monitoring; Home port: port from which boat operates; System: AIS or GPS data; Process: data used to calibrate the method or to validate it; CW. Dataset: Azorean cetacean watching fleet (yes) or survey (no) datasets; Count: total number of locations. 20

Table 2. 2. Confusion matrix for the validation of the method identifying cetacean watching behaviour. GMM: Gaussian Mixture Model. Logbook: boat register of the cetacean sighting. 24

Table 2. 3. Confidence interval by boat. Boat: id number of the boat, n: number of modes, µ: mean of the lowest speed mode, s2: variance of the lowest speed mode 29

Table 2. 4. Results of confusion matrix. Boat: id number of the boat. True_P: true positive, False_N: false negative, True_N: true negative and False_P: false positive. COB: percentage of segments (5min) where boats were engaged in cetacean observation, S: percentage of segments (5min) where boats were transiting.......... 29

CHAPTER 3

Table 3. 1. Land stations in the Azores archipelago. 41

Table 3. 2. Summary of dynamic and static AIS variables. 43

Table 3. 3. Vessel classification 44

Table 3. 4. Categories of lethal probability (LP) shipping. SOG: speed over ground, Length: length of the vessel.. 45

Table 3. 5. Summary of AIS data used in this study (October 2012- December 2016). 46

CHAPTER 4

Table 4. 1. Cetacean data used for modelling. 73

Table 4. 2. Description of the candidate variables used in the sdms. pb: biomass distribution, pp: production distribution. 76

Table 4. 3. Summary of the output and evaluation of the best model for each whale species. 79

Table 4. 4. Covariates retained in the best sperm whale model. Depth, Depth; Euclidean distance to small seamount, sdS-Seamounts; upper-mesopelagic (pp), meso_pp. 80

Table 4.5. Covariates retained in the best fin whale model. Time-lagged chlorophyll-a concentration (_1 month), chla(_1m); time-lagged npp (_1 month), npp (_1m), and epipelagic micronekton (pb), epi_pb. 80

Table 4. 6. Covariates retained in the best blue whale model. Logarithm of sea surface temperature, sstl; logarithm of depth, depthl; time-lagged chlorophyll-a concentration (_1 month), chla(_1m) 80

Table 4. 7. Covariates retained in the best sei whale model. Time-lagged npp (_1 month), npp (_1m), and; upper-mesopelagic (pb), meso_pb. 80

CHAPTER 5

Table 5. 1. Summary of cetacean and anthropogenic activities data. Dataset, type of data; source, source of the cetacean data; period, period of the data; platform type, type of boat; analyses, the anthropogenic activity analysed and cw, cetacean watching. * searching cetaceans ~meters above the water. **cetacean watching boats have at least one person looking for cetaceans. 111

Table 5. 2. Candidate anthropogenic variables used in SDMs of sperm whales and fin whales 113

Table 5. 3. Presences and absences used to fit SDMs. SDM-p-es: model with environmental and prey variables. SDM-p-aa: model with environmental, prey, and anthropogenic variables 114

Table 5. 4 Summary of the best fitting models of shipping density for sperm whales. Akaike information criteria, aic or * small-sample equivalent (aicc); concordance index, C-index. Model type abbreviations are as follows: SDM-p-es: environmental and prey variables. SDM-p-aa: environmental, prey and anthropogenic variables. The SDM-p-aa1 used shipping density at >14kn, and sdm-p-aa2 used shipping density at >14kn for ships >80m long..117

Table 5. 5. Covariates retained in the best sperm whale model (SDM-p-aa-2). Shipping density, “speed and size”, 14-80kn 80m; square root of euclidean distance to sst fronts, Ed_fronts; epipelagic micronekton (pp), epi_pp and highly migrant lower-mesopelagic (pp), hmbathy_pp. 118

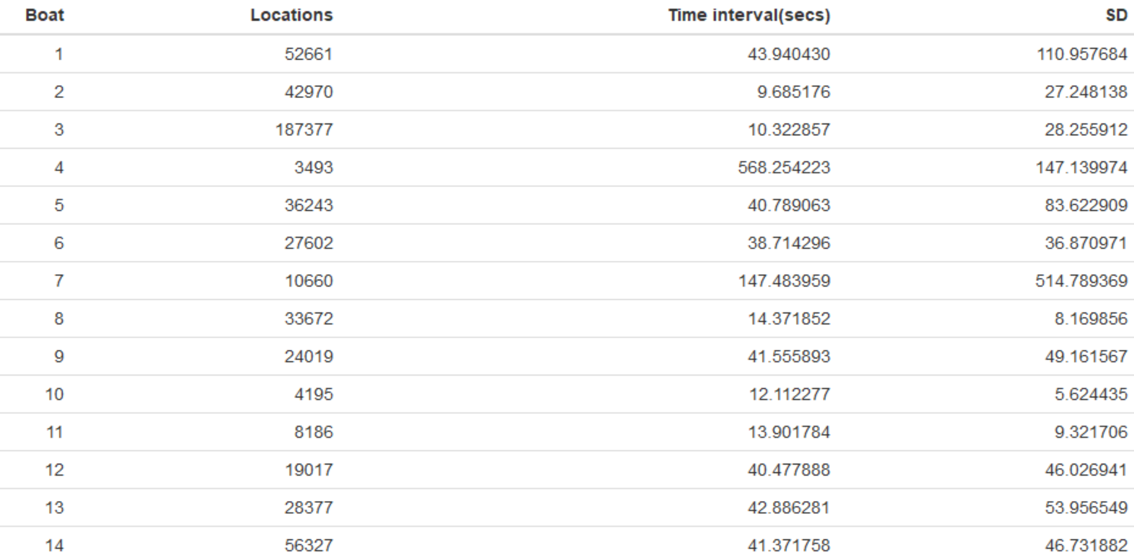
Table 5. 6. Covariates retained in the best fin whale model (SDM-p-es). Sea water potential temperature, temperature; lower trophic level plankton (pp), pk_pp; epipelagic micronekton (pp), epi_pp. 118

APPENDICES

SUPPLEMENTARY DATA OF CHAPTER 2

APPENDIX A. TIME INTERVAL

Table App. 2.1. Time interval between messages/signals by boat. Boat: id number of the boat; Locations: number of positions; Time interval: mean of time interval in seconds; SD: standard deviation of time interval in seconds.



Boat	Locations	Time interval(secs)	SD
1	52661	43.940430	110.957684
2	42970	9.685176	27.248138
3	187377	10.322857	28.255912
4	3493	568.254223	147.139974
5	36243	40.789063	83.622909
6	27602	38.714296	36.870971
7	10660	147.483959	514.789369
8	33672	14.371852	8.169856
9	24019	41.555893	49.161567
10	4195	12.112277	5.624435
11	8186	13.901784	9.321706
12	19017	40.477888	46.026941
13	28377	42.886281	53.956549
14	56327	41.371758	46.731882

APPENDIX B. BIMODALITY RESULTS

Table App. 2.2. Dip test and bimodality coefficient results. Boat: id number of the boat, Count: number of positions, Dip statistic: value of the statistic, Dip pvalue. BC: bimodality coefficient value.

Boat ID	Count	Dip Statistic	Dip pvalue	BC
1	39743	0,065702	<0,01	0,738995
2	42970	0,032494	<0,01	0,640054
3	187377	0,052109	<0,01	0,682073
4	3493	0,040724	<0,01	0,73093
5	27345	0,067955	<0,01	0,712708
6	21224	0,101855	<0,01	0,794464
7	7488	0,093499	<0,01	0,717595
8	33672	0,047111	<0,01	0,572008
9	22457	0,058403	<0,01	0,692441
10	4195	0,031259	<0,01	0,41369
11	8186	0,045804	<0,01	0,68244
12	12866	0,041158	<0,01	0,696065

Boat ID	Count	Dip Statistic	Dip pvalue	BC
13	18270	0,036696	<0,01	0,749689
14	44286	0,092753	<0,01	0,792203
15	2673	0,011501	<0,01	0,681311

SUPPLEMENTARY DATA OF CHAPTER 3

APPENDIX C. LAND STATIONS

Table App. 3.3. Estimated range covered (nm)

Station	Range (nm)
Sta Maria (Pico Alto)	66,70
S. Miguel (Pico Bartolomeu)	81,70
S. Miguel (Pico das Camarinhas)	66,80
Terceira (Serra do Cume)	65,70
Terceira (Serra de Sta Bárbara)	87,50
Graciosa (Pico Timão)	49,80
S. Jorge (Macelinha)	59,80
Pico (Pico Geraldo)	60,70
Faial (Cabeço Gordo)	88,90
Flores (Morro Alto)	83,90

APPENDIX D. NA IN RAW DATA

The AIS raw data had missing data; five variables had not available information for 2-34 % of the total messages (IMO, length, beam, speed and status). These variables were the same over the years.

Table App. 3.4. The total number values of the AIS raw records from the messages, vessels, trips, countries and dates for the period (2012-2016).

Year	Messages	Vessels	Trips	Countries	Dates
2012	82706	1206	4696	55	65
2013	608849	5648	37702	87	339
2014	663575	5845	38459	79	333
2015	793428	6146	47560	95	334
2016	1035874	6649	58276	95	366

Table App. 3.5. Number of vessels per year and vessel type (AIS raw records for the period (2012-2016)). Military v.: military vessels; pleasure b.: pleasure boats; port: vessels used for port services; rescue v.: rescue vessels

Year	Cargo	Fishing	Military v.	Other	Passenger	Pleasure b.	Port	HSC	Rescue v.	Tanker
2012	831	36	4	35	17	2	7	0	0	274
2013	3451	133	16	348	121	145	35	7	12	1380
2014	3727	113	16	306	140	200	35	5	12	1291
2015	3694	139	25	366	141	206	58	10	8	1499
2016	3958	120	15	553	138	197	59	9	12	1588

Table App. 3.6. Number of trips per year and vessel type (AIS raw records for the period (2012-2016)). Military v.: military vessels; pleasure b.: pleasure boats; port: vessels used for port services; rescue v.: rescue vessels

Year	Cargo	Fishing	Military v.	Other	Passenger	Pleasure b.	Port	HSC	Rescue v.	Tanker
2012	2412	340	15	125	1034	4	94	0	0	672
2013	17939	1679	117	1965	8215	833	920	646	217	5171
2014	17560	2325	98	2047	9495	855	1060	370	77	4572
2015	19363	4749	180	2288	11130	1016	1494	747	109	6484
2016	22858	5009	58	5305	12085	1406	2045	969	262	8279

APPENDIX E. PROCESS TO DETECT DUPLICATE VESSEL NAMES:

The algorithm described here allows evaluating the reliability of the data. The process of the algorithm started once the vessels are identified by MMSI and IMO. Some vessels showed more than one name. Therefore, the frequency of each vessel's names was calculated, names under 10% of frequency were labelled as suspicious. The compute Similarity Scores between the vessel names was run in R using the package *stringdist v0.96.3 [1]*. The function stringsim returns a value between zero and one, where zero is complete dissimilarity and one perfect similarity. The next step was to run the index to identify the same vessel with former names or some differences in the static provided information. This index was based on MMSI, IMO, name, callsign, type of vessel (cargotype) and country. If the index concluded that they were different, they were selected to be corrected consulting the cleaning AIS list, which was developed by analysing the raw data, and external online databases. The external data consulted were Fishing Vessels Finder (FVF) of the Food and Agriculture Organization of the United Nations (FAO) and the free AIS vessels tracking web site VesselFinder which displays vessel positions and static information [2, 3].

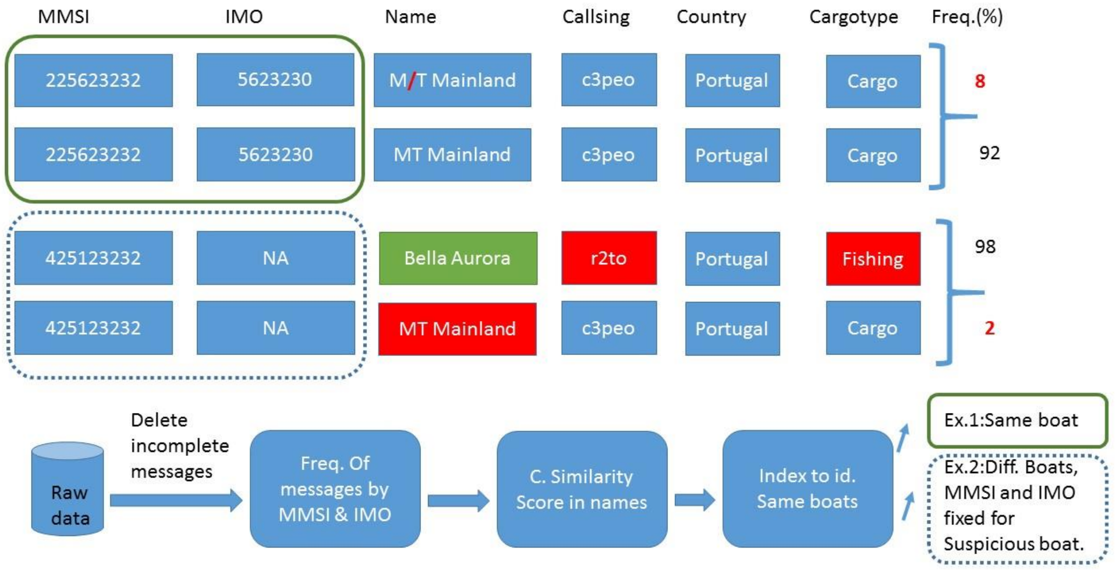


Figure App. 3. 1. *Scheme of the process to detect duplicate vessel names*

Table App. 3.7. Algorithm of the Process to detect duplicate vessel names

Algorithm	
Require	AIS shipping dataset
Require	Locations (latitude and longitude)
Require	AIS messages identified by MMSI and IMO, and include date, time. Also optional variables: name, callsign, type, country, speed etc
1	Frequency of names by MMSI_IMO
2	Suspicious Names; with a <10% freq of the messages
3	Compute similarity scores between the names by MMSI_IMO
4	Index to identify same vessels with one or several differences based on MMSI, IMO, name, callsign, type and country

Table App. 3.8. Classification of the same vessel index

Classification	Variable	Description
0	Prob. different name	Similarity scores, is not probably the same boat's name
0.5	Country	Static boat information
0.5	Cargotype	Static boat information
1	Prob. same name	Similarity scores, is probably the same boat's name
2	MMSI	The unique MMSI number of a vessel can changes (sold, owner, flag)
2	IMO	Once assigned, the Number remains unchanged for a company and/or registered owner
2	Callsign	Unique VHF call sign, is programmed when installing the equipment a

APPENDIX F. PROCESS TO DETECT SYSTEMATIC ERRORS IN VESSEL COORDINATES:

A systematic mistake was detected in the transmission from the vessels and from the land station. Messages with coordinates and all the data required but wrong. The algorithm was built here to detect duplicated messages based on the exact coordinates, date and hour (without including minutes and hours). For example a vessel acting as a sea station transmit five messages with the information of five different vessels (5 MMSI), and all of them had the exact coordinates that belong to the intermediary vessel between the vessel and the land station, not to the original vessels (MMSI). The validation of this method was using land station's data because all the messages with a MMSI for a vessel with the land station' coordinates were an apparent mistake. The subset of messages with land station's coordinates was tested with the algorithm, and was detected a 97-100% of the mistakes. After the validation of the process, a filter was created to remove all messages detected as duplicates and messages within the coordinates of the land stations.

APPENDIX G. HARBOUR BOXES

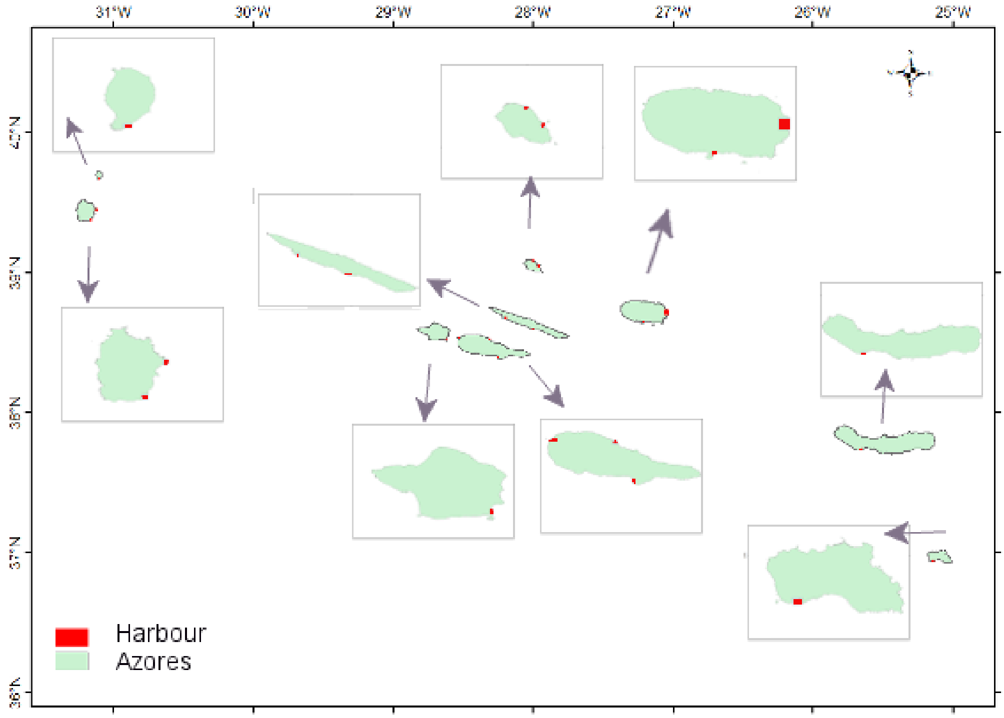


Figure App. 3. 2. Harbour boxes used in the split trips process[1]

---

[1] Adapted from 4. Eurostat, cartographer Transport networks: © European Union, 1995-2021; 2009.

Appendix H. Inter-annual patterns of mean daily shipping density on cruise from 2013 to 2016.

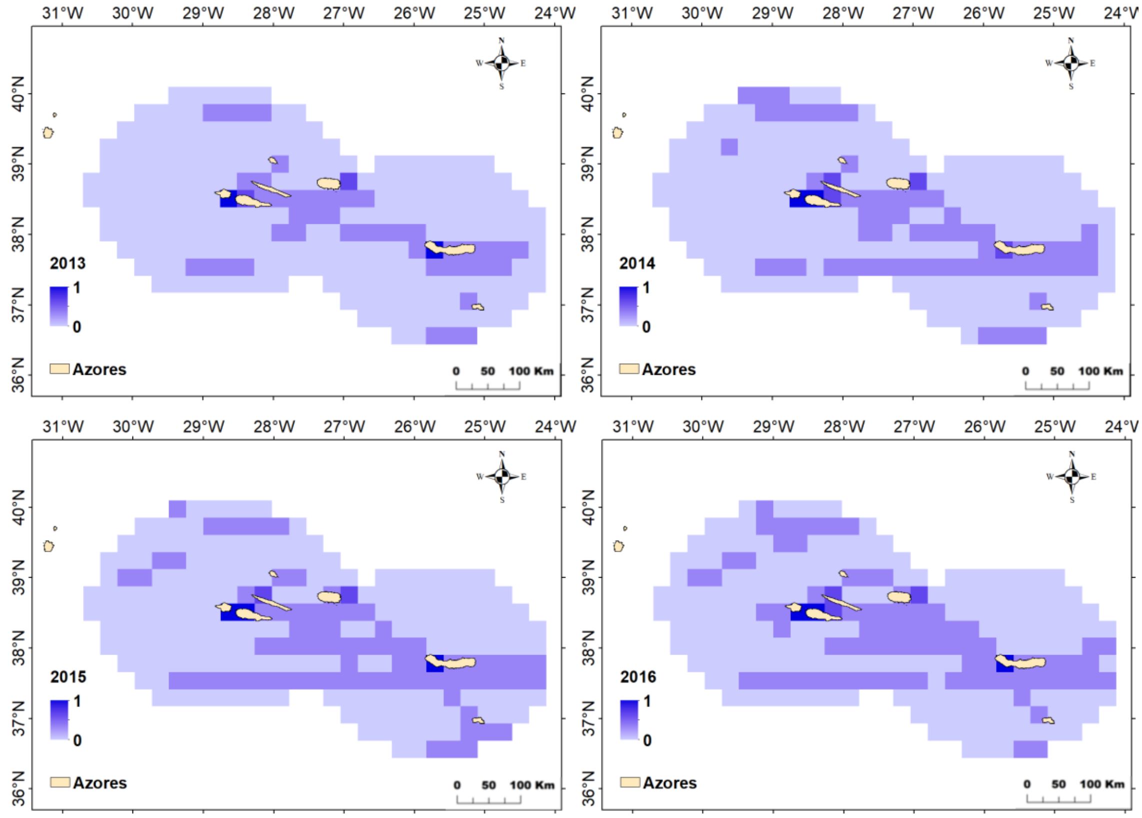


Figure App. 3. 3. *Inter-annual patterns of mean daily shipping density on cruise from 2013 to 2016.*

Supplementary data of Chapter 4

Appendix I. Train and test data summary

Table App. 4.9. Presences and absences used to fit SDMs. Number of individuals (presences).

Species	TRAIN DATA		TEST DATA	
	Presences	Absences	Presences	Absences
Fin whale	154	18798	45	6539
Blue whale	31	19714	11	10377
Sei whale	61	19051	30	6303
Sperm whale	1204	18371	376	6807

Table App. 4.10. The models predicted relative density ranged from a maximum and a minimum of individuals per 27,5 km2 (1/4 degree 2) and transformation to per 100km^2.

Species	1/4 degree^2		100km^2	
	Min	Max	Min	Max
Sperm whale	5,94E-05	3,81E-03	2,16E-04	1,39E-02
Fin whale	1,48E-06	1,22E-03	5,40E-06	4,42E-03
Blue whale	3,31E-08	5,95E-04	1,20E-07	2,17E-03
Sei whale	5,05E-06	1,02E-03	1,83E-05	3,70E-03

APPENDIX J. SUPPLEMENTARY MAPS

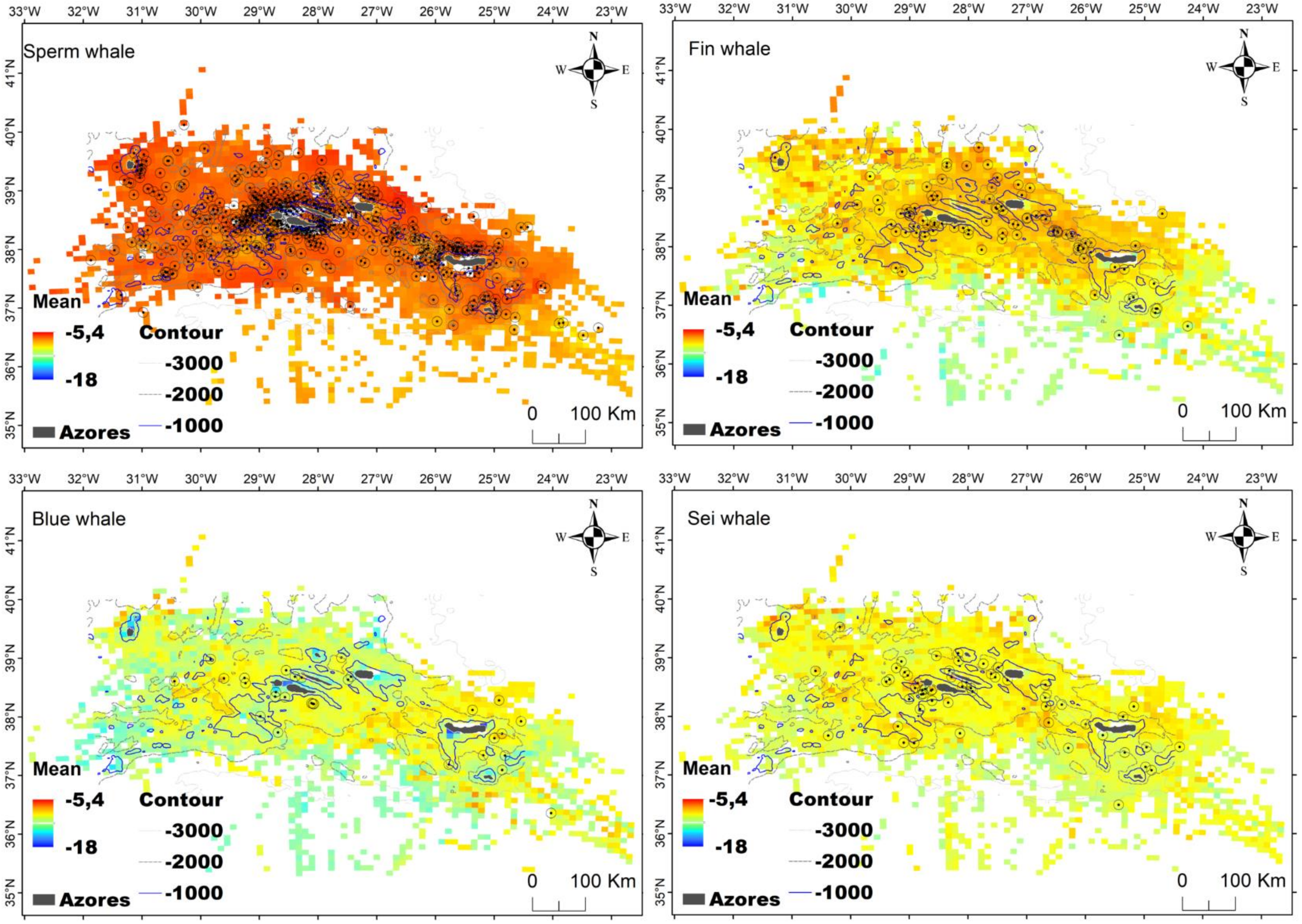


Figure App 4. 1. Mean overall predictions with sightings. Sperm whale on top left, fin whale on top right, blue whale on down left and sei whale on down right.

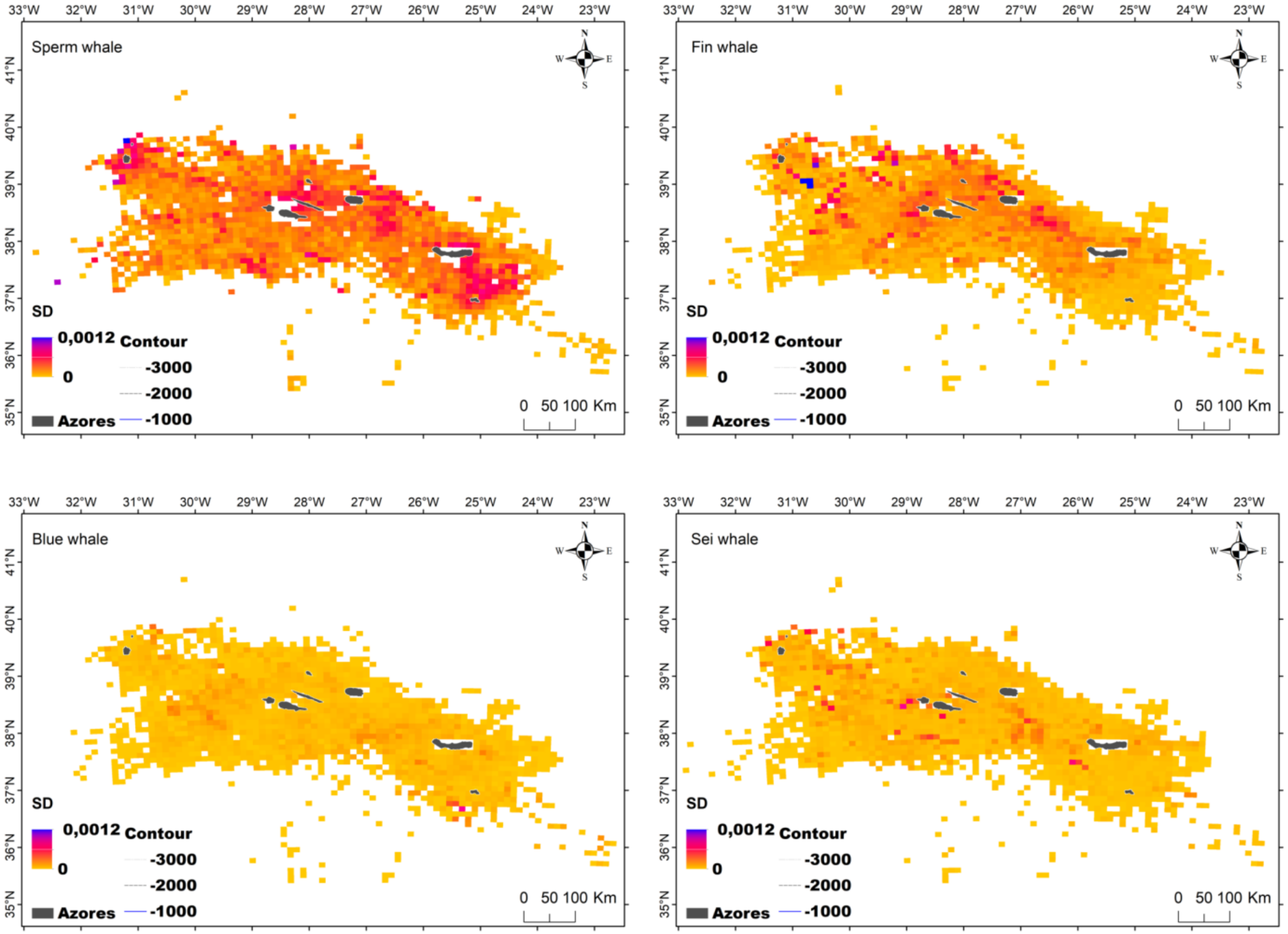


Figure App 4. 2. Standard deviation (SD), best model for sperm whale (top left), fin whale (top rigth), blue whale (down left) and sei whale (down rigth).

SUPPLEMENTARY DATA OF CHAPTER 5

APPENDIX K. RESULTS OF THE DISCARDED MODELS

Table App. 5.11. Covariates retained in the best sperm whale model (SDM-p-es). Square root of Euclidean distance to SST fronts, Ed_Fronts; Epipelagic micronekton (pp), epi_pp and highly migrant bathypelagic micronekton (pp), hmbathy_pp.

Covariates	**edf**	**Ref.df**	**F**	**p-value**
epi_pp	0.8207	4	1.173	0.0155
Ed_Fronts	0.8075	4	0.982	0.0264
hmbathy_pp	0.7973	4	0.896	0.0321
Parametric coefficients:	**Estimate**	**SE**	**T**	**p-value**
Intercept	-6.7552	0.1195	-56.52	<2e-16

Table App. 5.12. Covariates retained in the best sperm whale model (SDM-p-aa-1). Shipping density, "Speed", 14-80kn; pk_pp, square root of Euclidean distance to SST fronts, Ed_Fronts; Epipelagic micronekton (pp), epi_pp and highly migrant bathypelagic micronekton (pp),

Covariates	**edf**	**Ref.df**	**F**	**p-value**
Ed_Fronts	0.8161	4	1.031	0.0235
epi_pp	0.8028	4	1.046	0.0208
hmbathy_pp	0.7822	4	0.811	0.0392
14-80kn	0.7048	4	0.533	0.0811
Parametric coefficients:	**Estimate**	**Std. Error**	**T value**	**Pr(>|t|)**
Intercept	-6.7643	0.1197	-56.5	<2e-16

Table App. 5.13. Covariates retained in the best fin whale model (SDM-p-aa-1). Shipping density, "Speed", 14-80kn; Sea water potential temperature, temperature; Lower trophic level plankton (pp), pk_pp; Epipelagic micronekton (pp), epi_pp.

Covariates	**edf**	**Ref.df**	**F**	**p-value**
pk_pp	0.8677	4	1.274	0.0130
epi_pp	0.8148	4	1.015	0.0241
temperature	0.8110	4	0.955	0.0285
14-80kn	6e-5	4	0.000	0.9288
Parametric coefficients:	**Estimate**	**Std. Error**	**T value**	**Pr(>|t|)**
Intercept	-9.440	0.3432	-26.93	2.16e-10

Table App. 5.14. Covariates retained in the best fin whale model (SDM-p-aa-2). Shipping density, "Speed and size", 14-80kn 80m; Sea water potential temperature, temperature; Lower trophic level plankton (pp), pk_pp; Epipelagic micronekton (pp), epi_pp.

Covariates	**edf**	**Ref.df**	**F**	**p-value**
pk_pp	0.8677	4	1.274	0.0130
epi_pp	0.8148	4	1.015	0.0241
temperature	0.8110	4	0.955	0.0285
14-80kn 80m	8e-5	4	0.000	0.7694
Parametric coefficients:	**Estimate**	**Std. Error**	**T value**	**Pr(>|t|)**
Intercept	-9.440	0.3432	-26.93	2.16e-10

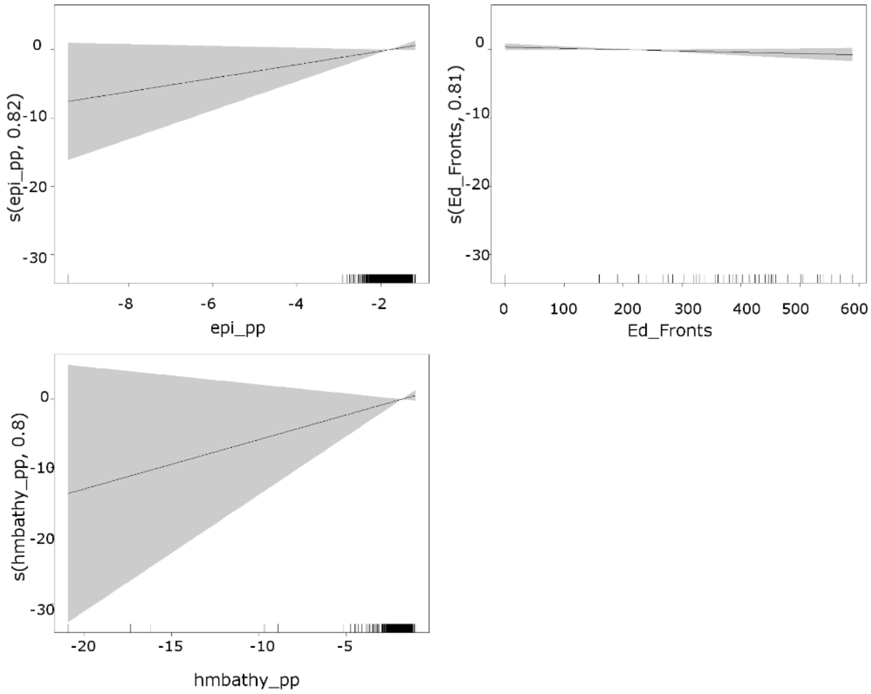


Figure App. 5. 1. GAM smoothers for sperm whales (SDM-p-es): top left) Epipelagic micronekton (pp), epi_pp, top right) Square of Euclidean distance to Fronts, Ed_Fronts, and bottom left) highly migrant bathypelagic micronekton (pp), hmbathy_pp. Smooth functions are shown as a solid line, and shaded regions represent 95% confidence intervals.

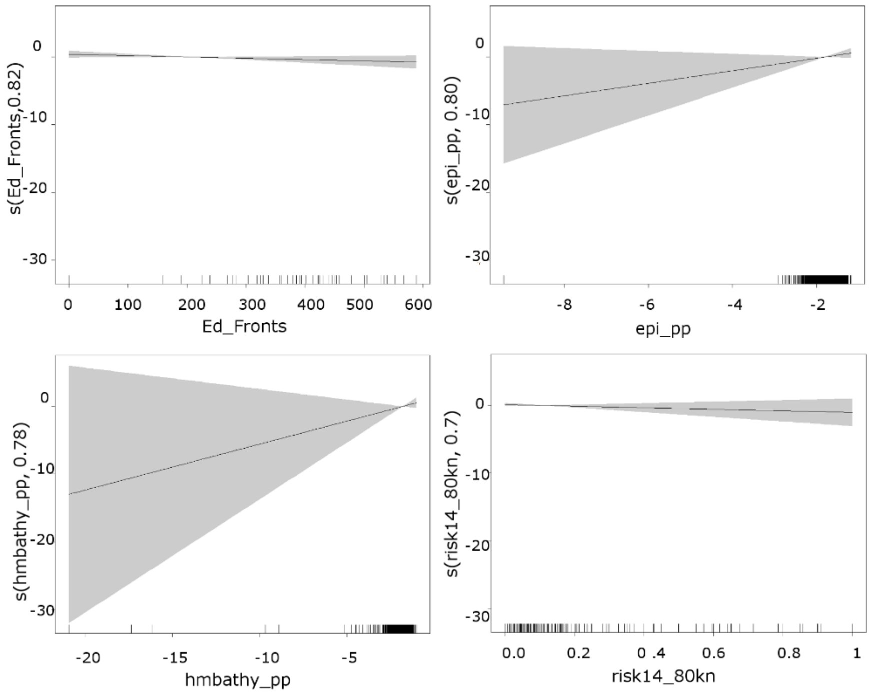


Figure App. 5. 2. GAM smoothers for sperm whales (SDM-p-aa-1): top left) Shipping density, "Speed", 14-80kn, top right) Square of Euclidean distance to Fronts, Ed_Fronts, bottom left) Epipelagic micronekton (pp), epi_pp, and bottom right) highly migrant bathypelagic micronekton (pp), hmbathy_pp. Smooth functions are shown as a solid line, and shaded regions represent 95% confidence intervals.

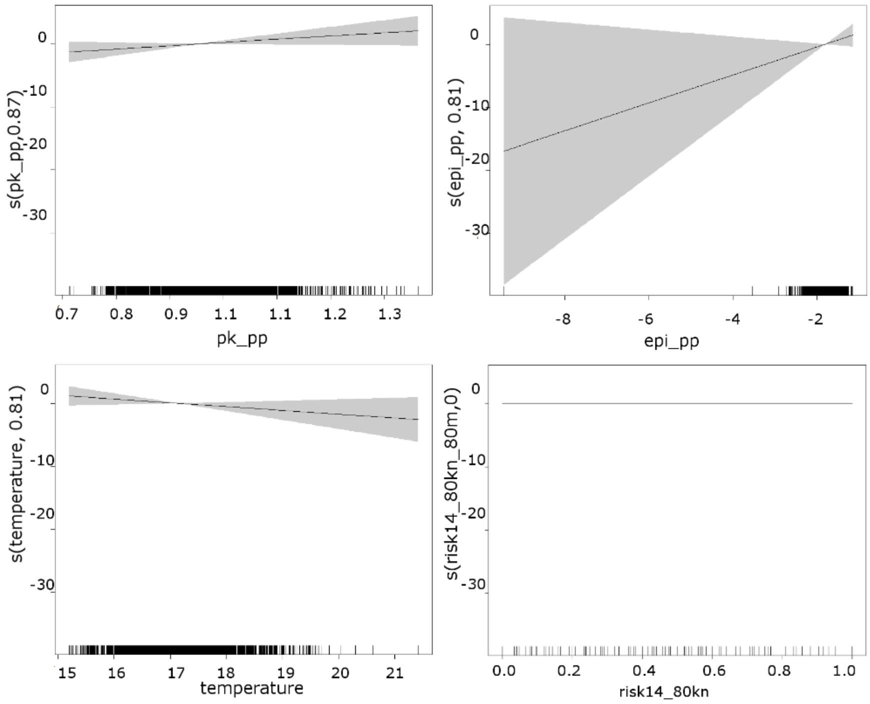


Figure App. 5. 3. GAM smoothers for fin whales of the SDM-p-aa1 and SDM-p-aa2 models on shipping extent. The SDM-p-aa1 and SDM-p-aa2 models were identical for fin whales, thus represented only once: top right) Lower trophic level plankton (pp), pk_pp, top left) Epipelagic *micronekton (pp), epi_pp, bottom left) Sea water potential temperature, temperature and bottom right) Shipping density, "Speed and size", 14-80kn 80m and Shipping density, "Speed", 14-80kn same flat line. Smooth functions are shown as a solid line, and shaded regions represent 95% confidence.*

REFERENCES

1. van der Loo M. The stringdist package for approximate string matching. 2014;6:111-23.
2. Fishing Vessels Finder for EU (European Union Fleet Registry On The Net (FRONT))vessels [Internet]. Fisheries and Aquaculture Department. 2011 [cited 25/feb/2021]. Available from: http://www.fao.org/figis/vrmf/finder/EU/search/#.YDfV-XnO-1s.
3. VesselFinder [Internet]. 2011. Available from: www.vesselfinder.com.
4. Eurostat, cartographer Transport networks: © European Union, 1995-2021; 2009.

UNIVERSIDADE DOS AÇORES
Faculdade de Ciências e Tecnologia

Rua Professor Doutor Frederico Machado, N.º 4
9901-862 Horta
Açores, Portugal

Near realtime distribution modelling of cetacean distribution off Azores

Marta Tobeña Morcillo

[TD]

SICUT AURORA SCIENTIA LUCET
UNIVERSIDADE DOS AÇORES

2022