MASTER'S DEGREE THESIS

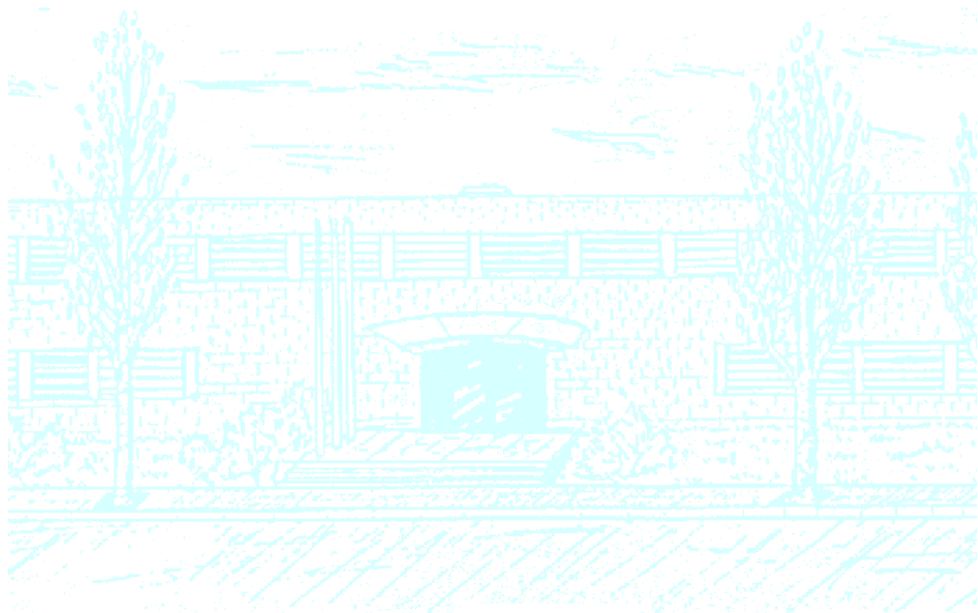# MSc in Statistics and Operations Research

Title: Spatial Distribution and Foraging Behaviour of Harbour seals (*Phoca vitulina*) in the Wadden Sea

Author: Laia Rojano Doñate

Advisor: Xavier Puig Oriol and Jakob Tougaard

Department: Estadística i Investigació Operativa

Academic year: 2013-2014

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

UPC

Facultat de Matemàtiques i Estadística

Spatial Distribution and Foraging Behaviour of Harbour seals (*Phoca vitulina*) in the Wadden Sea

Master's Thesis

Interuniversity Master in Statistics and Operations Research

Universitat Politècnica de Catalunya

Universitat de Barcelona

Laia Rojano Doñate

Supervised by Dr. Xavier Puig Oriol

Department of Statistics and Operations Research
Universitat Politècnica de Catalunya

and Senior Researcher Jakob Tougaard

Department of Bioscience - Arctic Research Centre
Aarhus University, Denmark

7th January 2014

Acknowledgments

I would like to thank my supervisors in this master thesis, Xavier Puig and Jakob Tougaard, for giving me the opportunity to develop this project and helping me during the process, despite the distances.

I would also like to mention in these acknowledgments the teachers of the different courses I have taken during the MEIO, specially Professor Lupe Gómez, not only for encouraging me to challenge myself and start this master in Statistics after my degree in Biology, but also for being always available for any help or support I would need.

I am also grateful to Professor Peter Madsen from Aarhus University, for giving me the opportunity of belonging to his lab and research group and being always ready to help and teach.

Furthermore, I would like to thank my classmates in this master for all the moments we spent together and my officemates in Aarhus, especially Mafalda and Line, for making the time spent in this master's project funnier and being always ready to help and proof-read everything.

I would also like to mention my family and friends, for understanding my absence, especially these last days and being always supporting me in my decisions and being always understanding.

Last, but not least, I would like to especially thank Oleguer for his constant support, reading again and again this memory and helping me with constructive comments.

Summary

Movement and habitat use are essential population processes fundamental for the management and conservation of animal species. Hence, the ability to track and infer specific animal behaviours are problems that have concerned scientists for many years. Technological advances have contributed to the improvement of tracking methods over the last century. Among them, ARGOS system has become an important tool for tracking animal movement globally, however it is limited to only providing geolocation. Therefore, it is still necessary to develop a methodology capable of inferring the behaviour of an animal using the available information.

The inherent complexity of animal movements and limitations in tracking systems have made necessary to develop statistical methods able to cope with such constraints. Recently, State-Space models (SSMs) have been presented as an approach capable of integrating the treatment of both limitations. In this project, it is presented a two-stages method able to determine, from movement metrics, the distribution of the animals and, more specifically, the foraging behaviour of Danish Wadden Sea population of harbour seals. Firstly, this methodology handles ARGOS inaccuracy positioning and, secondly, assigns behavioural states to each location.

From the methodological perspective, results from the filtering localization analysis showed that extreme or inaccurate locations were corrected, resulting in smoother and more realistic movement paths. Subsequently, results from the behavioural assignment showed that SSMs allowed us to distinguish two different states: one slower and more variable in terms of turning angles, and another, faster and less variable. From the biological perspective, results showed that animals traveled long distances in order to arrive to foraging spots, especially pups or young animals, probably due to competition with older, larger, dominant animals for the best foraging locations. Furthermore, results suggested that the study population tended to share hauling out banks with different individuals. On the contrary, results suggested that there were no defined or shared foraging spots for all the animals.

Finally, it was concluded that the application of this two-stage methodology was successful allowing for an improved localization of this seals' population and behavioural assignment to each location. It is important to remark that an improved localization of the population and its habitat use will allow for a better identification of critical habitats, such as overlap with fisheries and wind farms, and to test for effects of anthropogenic activities, such as seismic surveys and pile driving.

Contents

Chapter 1

Introduction

From an ecological point of view, movement is a fundamental population process essential for the management and conservation of many animal species that is still relatively poorly understood. Since the twentieth century, when the first birds were ringed, tracking animals has helped scientists to understand how individuals and populations move within local areas and migrate across oceans and continents.

Over the last century, technological advances have contributed to the improvement of animal tracking methods. The development of tracking tools, such as the Advanced Research and Global Observation Satellite (ARGOS) system and the Global Positioning System (GPS), have helped achieve a better understanding of animal movement and distribution and, consequently, improved in population management and conservation. In spite of these technological progresses, it is always challenging to estimate the movement behaviour of wild animals, particularly when referring to marine life, living most or all of their life submerged. Thus, advanced tracking methods are usually required to study marine species. In these cases, by equipping a number of animals with suitable sensors and recording electronics, as well as assuring a way of retrieving the data, important and detailed information can be obtained on migration, movement and even animal specific behaviour. However, these techniques normally present inherent limitations, such as low accuracy in localize (Tougaard et al., 2008), creating the need to apply complex smoothing and/or filtering methods. For instance, ARGOS locations have variable accuracy, the distance from the calculated to the true position ranges from a few hundred meters to more than 50 km (Vincent et al., 2002; White and Sjöberg, 2002). Therefore, conclusions made from those estimations can be widely affected, in particular, very specific questions and, especially, those depending on the behaviour of the tagged species.

Different methods have been used in order to deal with these limitations (e.g. Freeman et al., 1997 and Douglas et al., 2012), however, many have not yet been able to cope with the significant complexity of movement data. Lately, the use of State Space Models (SSM) has been suggested as a powerful and flexible approach, allowing for simultaneous estimation of both the inherent measurement errors and process noise of animal movement data (Jonsen et al., 2003).

The use of such time-series methods has risen due to belief that, for more efficient population

management, spacio-temporal information of a population needs to be combined with inform-
ation about its habitat use. Traditionally, population ecology has focused on understanding
temporal fluctuations in abundance, however, the focus has recently shifted to understanding
interactions between life history, physiology, behaviour and habitat, which make individual
movements an exceptionally complex phenomena. In order to deal with this increment in
complexity, SSMs have become a statistical tool that provides a powerful analytical foundation
for animal movement ecology by simultaneously capturing the essential ecological, physiological
and environmental factors driving movements (Patterson et al., 2008). The SSMs combine
a stochastic model of a biological, or physiological process, with a model of the observation
process. The process model predicts the future state of a system considering its previous states,
and the observation model describes how observations of the state are generated.

In this study, I use this recent approach to study the spatial distribution and foraging behaviour
of Danish Wadden Sea Harbour Seal population. This population of harbour seals is poorly
studied and little is known about its habitat use and, therefore, little can be done in order
to secure its management and conservation. The Wadden Sea area is an important zone for
industries, where many wind farms have been, and are planned to be, constructed. Thus, the
study of the animal population living in this environment is essential for population management
and successful coexistence of industry and biodiversity.

1.1   Objectives

In this Master's thesis, the main aim is to satisfactory apply a SSM to deal with the biological
and statistical complexities of movement data of harbour seals (*Phoca vitulina*) and use this to
determine the geographical distribution of the foraging areas as well as the site fidelity in terms
of foraging positions.

In order to study the spatial distribution and habitat use of this population, 21 harbour seals
were tagged in the Wadden Sea area using ARGOS system during 2002-2004. With this
information, I intend to use a combination of two Bayesian SSM frameworks that allow for the
estimation of movement pathways of 17 of those harbour seals, dealing with both the inherent
errors in ARGOS data and the complexity of spatio-temporal data, and the probabilistically
assignment of behavioural states to each location.

In order to accomplish these objectives, the following goals need to be achieved:

1. To learn about the inherent characteristics of movement data and animal tracking methods.
   In Chapter 2, I explain about the evolution of animal tracking methods and the analysis
   of movement data along with the limitations of the different approaches.

2. To learn and understand the theory behind State Space Models.
   In Section 2.2.2, I summarize the definition and properties of the SSMs and explain how
   they are used in a movement data context.

3. To moderate the error distribution among ARGOS locations.
   Excluding all low-quality ARGOS locations can be rarely done without eliminating potentially important biological information (Douglas et al., 2012). In section 3.3.1, I apply the SSM framework proposed by Jonsen et al. (2005) in order to deal with the inherent errors of ARGOS data and obtain a smoother movement track.

4. To select the best fitting model for assigning behavioural states.
   In Section 3.3.2, I generate several variants of the SSM frameworks proposed by Morales et al. (2004) and Eckert et al. (2008) in order to posteriorly compare them and assess which one fits better with the data, not only accounting for standard statistical criterion, but also accounting for their capability of capture the intrinsic data characteristics and biological meaning.

5. To determine specific foraging location for this seal population, assessing individual fidelity.
   Other telemetry studies of harbour seals have indicated that this species forages close to haul out banks and moves only short distances between banks (e.g. Dietz et al., 2003). This Satellite telemetry data helps in answer whether this is also true for harbour seals in the Danish Wadden Sea. Moreover, I use this data to evaluate whether a specific animal keeps specific spots and whether these spots are shared with other individuals.

6. To assess variation in foraging location.
   I assess whether the distribution of the foraging locations vary depending on physical characteristics of the animals, such as gender and age, as well as environmental characteristics, such as season, in order to evaluate the existence of specific foraging spots. It can be expected younger, smaller and weaker individuals will compete with stronger dominant ones for the best foraging locations.

Furthermore, there are some academic objectives to be accomplished during this master's thesis:

1. To improve my skills using the statistical software R and WinBUGS.
   During several of my master courses, I have gained skills for using both software R and WinBGUS. In this project, I aim to improve these skills by applying the SSM and doing posterior analyses using both softwares.

2. To learn how to use Adobe Illustrator and LaTeX in scientific writing
   On one hand, Adobe Illustrator is a very useful graphical software in terms of creating figures and graphs that can be used in many kinds of scientific texts. On the other hand, LaTeX is a very important tool for research when math formulas and symbols are needed. Thus, for this master's thesis, I decided to use both tools in the process of writing this report in order to learn the characteristics of these system.

3. To gain research skills and improve my writing in English
   This master's thesis is one of my first experiences doing research by myself, thus this project has given me a good opportunity to gain research skills for my future. In addition, knowing that English is the most used language in research, I have decided that learning to be able to write a scientific report is essential for me even though it has been a challenge.

Chapter 2

State of the Art

2.1 Animal Tracking

The ability to track animal movement is essential for understanding animal foraging ecology, migratory behaviour, habitat use and general life history parameters (Costa et al., 2009). In particular, understanding the movement of animals is a fundamental problem in ecology that has concerned scientists for many years. It was during the fall of 1803, when the French-American Naturalist John James Audubon conducted the first known bird-banding. He tied a string around the leg of a bird wondering whether birds returned to the same nesting place every year. The following spring, Audubon saw that the bird had indeed come back. Over the last century, the technological advances have contributed to improvements of the animal tracking methods. It was in the late 1950s when researchers began using radio transmitters to track wildlife and in the late 1970s when the ARGOS system was created, providing a method for tracking animals globally. Later, in the early 1990s, the GPS began to be used, providing the potential to collect high-resolution tracking data.

Prediction of species' distribution is central to diverse applications in ecology, evolution and conservation science. Historically, scientists have been concerned about the distribution and migrations of animals, but for many years the only way to track wildlife was to simply follow and observe the movement and habits of an animal or to capture, tag and wait to recapture the individual again in order to retrieve the data. Particularly, it becomes even more complicated when trying to track marine animals. Marine animals move over extensive areas and spend all or at least a significant proportion of time submerged, which makes them notoriously difficult to observe and track. Consequently, when it comes to marine animal behavioural studies, advanced tracking methods have become extremely important tools.

Nowadays, tags, such as metal bands, are still being used to track animal movements, but there are also new tools and a variety of electronic tagging methods available, which can link reliable location data with information about the animal's physiology (e.g. individual's temperature or hear-beat) and the environment (e.g. air or water temperature). The combination of these parameters can be used to determine a more accurate life history of the animals, such as migratory patterns, seasonal feeding movements, daily habits, etc.

Among the different tools available today, the most commonly used are electronic tags that can be classified into: archival tags and transmitting tags: 1) archival tags (or data-loggers) archive all of the information in an internal memory and, therefore, need to be recovered to retrieve the data; and 2) transmitting tags give off repeating signals that are picked up by, for example, radio or satellites, and therefore, do not need to be recovered since all of the information is sent to remote processing centers. Both systems have different advantages and disadvantages. Archival tags are usually cheaper and smaller than the others, but one of the biggest advantages of radio or satellite transmitters is that these tagging methods allow for tracking movements of tagged animals without following or recapturing them. This is especially useful for tracking the migration and territorial movements of marine species where recapturing tags can be very difficult. Nevertheless, radio systems are limited in terms of range detection, since radio signals can only be detected within close ranges and individuals need to be therefore actively followed in order to obtain a complete movement path.

In particular, satellite transmitter tags can localize animals using either ARGOS or GPS system. In order to localize the tagged animal, the ARGOS transmitter on the animal sends a signal that is picked up by an ARGOS satellite. On the contrary, the GPS satellite is in this case the one that sends a signal that is picked up by the GPS receiver on the animal. Despite this difference, both systems have the limitation that the signal path needs to be relatively clear for the signals to be transmitted or received, which can be difficult to achieve when a tagged animal is in a dense forest or underwater. Focusing on the underwater scenario, when studying aquatic wildlife, these tags can only be deployed on animals that spend sufficient time at the water surface, since the tag antenna must be above the water to transmit a signal. Marine mammals fulfill this requirement, as they have to surface to breathe, but it is important to point out that locations can only be obtained during animal surfaces.

When comparing the two most used transmitting systems, GPS and ARGOS, it is clear that GPS positions are usually more accurate than most of ARGOS positions (calibration studies indicate that 95% of locations are accurate to ±55 m). Nonetheless, it is important to note that the GPS system presents some limitations in terms of satellite covering, especially some years ago, in 2002, when the beginning of the data collection for the present study took place (Tougaard et al., 2003). On the contrary, ARGOS transmitters are normally better transmitting in difficult radio environments, in which satellites can receive signals even when sent in relatively extreme conditions.

In conclusion, as discussed above, several systems are presently available for animal tracking, but particularly, when this database was collected, there were three main options: Radio-transmitters, Satellite transmitters and Data Storage tags. Each system has its strengths and weaknesses and whether one or the other is a suitable solution depends on both the questions asked and the behaviour of the animal. For this project, due to the marine nature of seals and the available technology when the data was collected, ARGOS transmitters were selected as the most suitable transmitting system.

2.1.1 ARGOS System

The Advanced Research and Global Observation Satellite (ARGOS) service was launched in 1978 as a joint US-French initiative and, since then, it has been used as a tracking system mostly towards environmental applications, including oceanography, fishing vessel monitoring, maritime safety and wildlife tracking. One of the best advantages of this system is that it allows for worldwide, near real-time data collection and localization of any mobile object equipped with compatible transmitters (known as Platform Transmitter Terminals, PTTs). At the same time, ARGOS system offers the possibility of collecting biometric and environmental data, as some of the transmitters can be connected to additional measurement sensors.

In general, PTTs automatically transmit periodic signals messages which are received by networks or groups of polar-orbiting satellites each time a satellite passes over the platform. Satellites accurately measure the frequency of the received signals and relay messages and frequencies to ARGOS processing centers to compute the specific location. For this computation, the general principle used for ARGOS system is based on the Doppler Effect of the frequency measurements.

Location Calculation

Service ARGOS uses coordinates expressed in latitude and longitude and the WGS 84 (World Geodetic System 1984) as a reference system and calculates locations by measuring the Doppler Effect on transmission frequency. The Doppler Effect is the change in frequency of a sound or electromagnetic wave that occurs when the source of vibration and the observer are moving relative to each other. A classic example for this effect is when a stationary observer notices a change in the sound, as a noisy object approaches or moves away. Similarly, when a satellite approaches a transmitter, the frequency of the transmitted signal measured by the on-board receiver is higher than the actual transmitted frequency and lower when it moves away (Figure 2.1A). Therefore, by using this principle, each time the satellite instrument receives a message from a transmitter, it measures the frequency and time of arrival and estimates the PTT position by using a geometrical calculation. Messages need to be well distributed along the satellite pass in order to enable proper determination of the Doppler-curve and, in particular, of the inflection point which corresponds to the actual transmit frequency.

A major feature of the Doppler location is the fact that two positions of the PTT give exactly the same frequency measurements on-board the satellite. These locations, the nominal ("true") and the mirror ("virtual") location, are symmetrical about the satellite track and not distinguishable a priori. The reason for obtaining a true and a virtual position is because the location calculation, based on the first and last messages of a satellite pass, finds the intersection between two cones and the platform altitude ellipsoid and obtains two possible locations, one in each side of the satellite track (Figure 2.1B). In order to detect the nominal location, a Least-Squares

method is used in both locations to reduce the distance between recorded frequencies and theoretical (expected) frequencies as a function of longitude, latitude and actual frequency of the transmitter. During this process, the calculation can either converge or not converge. If the calculation converges, then latitude, longitude, transmit frequency and residual error may be estimated (the latter is used to estimate the location accuracy). From these results, using information of the previous location and assuming that the nominal location should have the minimum residual error, one of the locations is chosen and posteriorly validated. According to Service ARGOS, the risk of picking the mirror location using this method is estimated to be less than 1% (Argos, 2013).
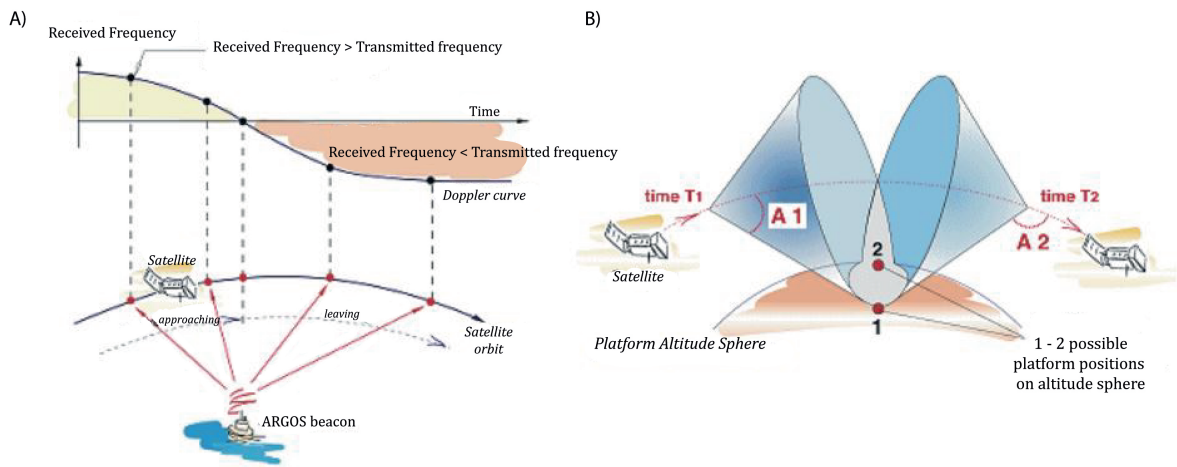
Figure 2.1: A) Schema representing the Doppler Effect procedure used for the location calculation of ARGOS transmitters. Location calculation is based on the difference between recorded frequency and real frequency of transmission. B) Schema representing the Geometric Location Principle used for ARGOS processing center for the calculation of the ARGOS transmitter's location. Figure modified from (Argos, 2013)

Location Accuracy

The satellites' capability to pick up signals from the PTT can be impaired by harsh weather conditions, specifically rough sea states can cause low transmitting power and results in low uplink rate and noisy tracks. In order to account for these inaccuracies, Service ARGOS provides estimates of location accuracy based on the number of messages received during a satellite pass. ARGOS positions are classified into seven Location Categories (LC): 3, 2, 1, 0, A, B and Z, which can be grouped into broader categories: *standard*, where estimated errors are provided, *auxiliary*, where no estimated errors are provided, and *invalid*, where all location calculations that do not converge are gathered. This latter category includes LCZ locations. On the contrary, all location calculations that do converge are classified into either standard (LC3, 2, 1 and 0) or auxiliary (LCA and B) locations, being part of one group or the other depends on the number of successive uplinks during a satellite pass (Table 2.1).

Table 2.1: Classification of ARGOS location into Location Categories is related to the number of message picked up during a satellite pass, locations accuracy is higher as the number of messages increase. Each Location Category has a estimated error assigned. Nominal values provided by Service ARGOS and Empirical determined errors from Vincent et al. (2002) are shown.

Location Category	Number of message per satellite	Service ARGOS	Vincent et al. (2002)	
			Latitude	Longitude
3	4 messages or more	<250 m	157 m	295 m
2	4 messages or more	250 m - 500 m	259 m	485 m
1	4 messages or more	500 m -1500 m	494 m	1021 m
0	4 messages or more	>1500 m	2271 m	3308 m
A	3 messages	No accuracy estimation	762 m	1244 m
B	2 messages	No accuracy estimation	4596 m	7214 m
Z	Location process failed	Invalid location	Invalid location	Invalid location

The number of parameters able to be estimated directly depends on the number of messages collected during a satellite pass. In order to calculate an accurate position, four parameters need to be estimated: latitude, longitude, transmit frequency and residual error, and for this, at least four messages need to be collected. Consequently, when at least four messages are picked up by the satellite, this location is classified as a standard location and location estimated errors can be provided. For classification purposes, Service ARGOS assumes an error to be isotropic and hence characterized by a single number called radius of error, that corresponds to one standard deviation of the estimated location error. According to Service ARGOS LC3, 2 and 1 location errors follow a normal bivariate distribution with standard deviations of 150 m, 350 m and 1000 m, respectively, and the 68th-percentile radius of error of 250 m, 500 m and 1500m. On the contrary, no bounded error is assigned to LC0 locations, it is just assumed to be >1500 m, being able to reach errors above 50, 100 or even 500 km. LC0 locations are a special type of standard positions in which, even having enough messages, the accuracy of the location is greatly variable due to random frequency instabilities observed between messages. Besides, when less than four messages are collected, the accuracy of the location cannot be estimated. Particularly, when three uplinks are collected, actual residual errors cannot be computed, and when two, only latitude and longitude can be estimated. In those cases, locations are classified as LCA and LCB, respectively.

In general, locations classified into LCs 0, A and B are predominant when there are low uplink rates, given that at least four homogeneous successive uplinks during a satellite pass are needed for a location to be assigned to an accurate category (Table 2.1). In particular, positions with LCs 0, A and B predominate in marine mammal tracking due to the difficulty of transmitting signals in an aquatic environment. A good example for this is the dataset used in this project, where 85.56% of location fixes are LC0, A or B. In consequence, due to the noisy tracks obtained from these locations, following what was said by White and Sjöberg (2002), LC0 A and B should only be used for studies involving long-range movements of 10 km or more, and would not be suitable for fine-scale analysis.

Studies on Accuracy of ARGOS Error Estimation

Understanding the limitations of a technique is crucial for the interpretation of the results and it is therefore highly relevant to perform studies in order to clarify these limitations. Even though Service ARGOS provides users with location accuracy, various experimental designs have been employed to estimate ARGOS location errors under field conditions (e.g. Vincent et al., 2002 and Patterson et al., 2010). It needs to be highlighted that the error estimation process is not entirely independent of either the transmitter frequency stability or the platform motion, which introduces an additional Doppler shift in the frequency signal measured by the satellite. As a consequence, frequency instability or a fast moving platform may lead to underestimation of the error.

One of the more relevant studies about ARGOS location accuracy was the one performed by Vincent et al. (2002). Concerned about the high proportion of location without guaranteed accuracy (LC0, A and B) in underwater studies, Vincent et al. estimated the accuracy of these locations by tagging gray seals with ARGOS system and keeping the animals in captivity in an outdoor tank. In the study, it was found that the error was generally greater in longitude than in latitude, resulting elliptical, and that LCA has errors similar to LC1 and far smaller than LC0. They also found that the average of the latitude and longitude 68th-percentile errors are bigger than those predicted by Service ARGOS in all LCs except in LCA (Table 2.1). Later, Royer and Lutcavage (2008) used data from freely swimming leatherback turtles and confirmed that locations LCA are more accurate on average than LC0 and that LC0 locations have a complex non-Gaussian structure, which makes even more complicated to estimate their error.

More studies have been done in order to determine some of the accuracy limitations of the use of ARGOS system, showing, for instance, that variations in sea temperature can degrade the quality of the locations and increase the occurrence of outliers (as the higher the temperature, the lower the accuracy of the locations) (Royer and Lutcavage, 2008).

In conclusion, even though Service ARGOS suggests that 68th-percentile accuracies of ARGOS telemetry range from 250 to >1500 m, in practice, it is quite common for errors to be an order of magnitude greater than this.

2.2   Analysing Spatial Data

Animals and environment interact in complex ways, which can be reflected in movement patterns. Understanding how these patterns arise and what their implications are for home-range, territorial dynamics or habitat use are important matters in ecology (Jonsen et al., 2003).

Traditionally, movement ecology has focused on studying the distribution and habitat use of populations but, due to either technological or analytical limitations, only home ranges were able to be estimated. Using classic tracking methods, home ranges were principally expressed using the adaptive Kernel method (Worton 1989), technique that places a normal distribution over each observed location, sums all the distributions together and normalizes them into a single kernel distribution (Breed et al., 2006).

Recently, this situation has changed. The present technological revolution has provided an improvement in terms of collecting movement and other kinds of data from roaming animals. The obsolete tagging and telemetry methods have evolved to small electronic components with much more autonomy and storage capacity. For instance, technologies such as ARGOS tags enable researchers to follow animal movements over long periods of time and large distances (e.g. Boeuf et al., 2000 and Eckert et al., 2008). As a result, this improved efficiency in data collection has created a data avalanche that, until recently, has not been matched by developments in analytical methods (Patterson et al., 2010). Nowadays, the combination of technological advances and present statistical methods has allowed us to move forward, undertake new challenges and attempt to understand interactions between movement, physiology and behaviour of individuals and populations.

2.2.1   Movement Data

Empirical movement data is formally represented as time-indexed positions of individual animals. When analysing movement data, it is important to take into account that temporal dependence is almost guaranteed, i.e., an individual is more likely to be doing the same thing at some time close to the present and, therefore, the current location of an individual will be near to its recent previous location. From a statistical point of view this autocorrelation can have important effects on the conclusions and, therefore, needs to be considered (Jonsen et al., 2013).

In general, despite the difficulties in predicting and modelling individual movement, it is easy to quantify movement data in terms of its metrics, i.e., quantities that might be calculated directly from raw movement data: speed, heading, turning angles between subsequent locations and rates of movement between regions (Patterson et al., 2008). Occasionally, values of these metrics can be taken as proxies of animal behaviour. For instance, when an animal encounters an area of sufficiently abundant food resource, it will often engage in area-restricted search by decreasing its travel rate and/or increasing its turning frequency and angle. On the contrary, an animal in an unsuitable habitat will often show faster travel rates and infrequent and small turning angles (Turchin, 1991). Consequently, given these values, it is tempting to derive

behavioural inferences solely from movement metrics. However, in most of the cases, it needs to be considered that such inferences can be confounded by the effects of observation error, such as the ones related to the observation method. Each observation method is subject to its own types of error: mark release recapture data suffers from variation in the probability of recapturing different individuals, electronic tagging and telemetry data often suffer from substantial positional imprecision and satellite telemetry is degraded by factors such as satellite coverage and interference. As a result, each source of error has the potential to bias ecological interpretation of movements, thus, handling location error becomes central when analysing movement data (Patterson et al., 2008).

Focusing on the method used in this study, data collected with ARGOS system, as previously stated, come with an inherent source of error which requires to be dealt with before making any inference. A great number of different methods have been used in order to handle these errors and provide realistic movement paths, among them:

1. Smoothing algorithms that weight ARGOS locations based on estimated error and time between locations, both in a generalized (Freeman et al., 1997) and in a species-specific framework (Thompson et al., 2003).

2. Bootstrapping random walks with forward-biased particle motion, while incorporating estimates of location error and physical boundaries (Tremblay et al., 2009).

3. Douglas Argos-filter algorithm (DAF), which filters locations using several user-prescribed parameters that allow its performance to be tuned to accommodate species' movement characteristics and study objectives (Douglas et al., 2012).

The main limitation of these methods is that they do not deal with the inherent movement data complexity. Nonetheless, statistical methods are improving all the time and more sophisticated statistical approaches have applied State-Space Models to deal with both ARGOS limitations and movement data complexity. One of the major advantages of these models is that they integrate the error correction, the calculation of metrics and the statistical analysis (Patterson et al., 2010).

2.2.2 State-Space Methods

State-Space Models (SSMs) include a range of time-series methods that estimate the state of an unobservable process from an observed dataset (Jonsen et al., 2013). SSMs combine two stochastic models:

1. **Process Model** that predicts the future state of an individual given its current state.

2. **Observation Model** that relates the unobserved state predicted by the process model to the observed data.

The main aim of an SSM is to draw inference from actual observations about how the state evolves through time, and about associated parameters that represent biological mechanisms underlying the state evolution and/or parameters that represent important features of the data collection process (Jonsen et al., 2013).

The **process model** is a very important part, mainly composed by a *transition equation* (see Box 1), that describes how the state evolves randomly through time. There are many alternative equation versions for representing different complexities of a process model. One of the simplest ones is the so-called Random Walk (RW), which assumes time steps at regular intervals, next state chosen at random and thus, zero correlation in the process variance. This can be verbally formulated as: *if the state of an animal at time t is known to be x_t, then the state at time $t+1$ is Gaussian with mean x_t and variance η*. Here η is a parameter representing the variance of the state between equal period of time, $\eta_t \sim \mathcal{N}(0, V)$. At the same time, it can be mathematically formulated as:

$$x_{t+1}|x_t \sim \mathcal{N}(x_t, V) \longrightarrow x_{t+1} = x_t + \eta_t \tag{2.1}$$

As it can be seen, the process model is written entirely without reference to available data and in terms of conditional probability distributions. These conditional distributions, which are known as *transition probabilities*, are sufficient to describe all state dynamics.

Box 1: *Transition Equation*

A transition equation describes the dynamics of a process, i.e. how unobserved states evolve over regular time. Many versions of different complexity models can define the transition equation, one of the simplest is the Random Walk (RW) model. Focusing on movement data, RWs analyses typically assume that movements are made in homogeneous environments and that animals do not change their behaviour drastically (Turchin, 1998). However, this is rarely the case in a natural environment. The behaviour of an individual is highly related to the habitat where the animal lives and its physiology, thus, individuals' behaviour may change as new stimuli are encountered and/or as an animal's internal state changes. Furthermore, dependence between movement variable and time is expected to be found. In consequence, movement pathways that span long time periods are likely to contain complex structure that is difficult to compare to a RW model (Jonsen et al., 2013). In order to solve this, different versions of RW models have been developed, among them the Correlated Random Walk (CRW).

A CRW model is a more natural way to think about animal movement, it describes many of the processes that we know to occur and provides the basis for more complex behaviour (e.g. Jonsen et al., 2005). In a simple RW, choice of direction at one point (step t) is unrelated to direction of the previous step, however, in animal motion it is known that there will be some persistence in the direction to take. An animal may change its direction gradually from step to step, which can be accommodated using CRW. In a CRW, the direction chosen at step $t+1$ is related to the direction at step t. Instead, what is randomly chosen is the change in direction (*turning angle*) from step t to step $t+1$.

The **observation model** describes what happens at the time of observation through the *measurement equation*, thus it does not make any reference to the dynamics in the underlying process model. This equation also specifies conditional probability distributions: *if the state at time t were known to be x_t, it is expected to find the distribution of the measurement y_t.* It is

important to appreciate the generality of this framework. The measurement may be a state variable subject to measurement error, for instance, a position estimated with the ARGOS system, but the measurement may also be any other quantity which holds some information about the state, for example, an animal's travel speed, turning angles or sea surface temperature. There are many potential options for distributions, and the choice would be determined by the data characteristics. For instance, continuous or discrete data, or the range of the expected values. A simple formulation of this equation would be:

$$y_t = x_t + \epsilon_t \tag{2.2}$$

where y_t is the observed state at time t, x_t is the true unobserved state and ϵ_t is the estimate observation error.

As it will be seen later in this project, for both the transition and the measurement equations, more complex models can be fitted. For instance, more sophisticated process models might include defined switches between states, in which state at time $t + 1$ depends on the state at time t via a matrix of switching probabilities, and irregular time lag can be corrected by adding data regulation parameters.

State-Space Models in a bio-logging context

Although the SSM concept has existed for some time, computational limitations have prevented its application in ecology until recently (Patterson et al., 2008). SSMs seem to represent a natural framework to model animal movement (Jonsen et al., 2003) where the process model formally describes a biological process of interest and predicts future states (e.g. the location and/or behavioural modes) of this process given its current state, and the observation model then weights the predictions by likelihood of the empirical data (e.g. locations obtained from ARGOS system) and may also provide interpolation of locations along the track.

The notion of *state* is crucial in a SSM. For example, in a bio-logging context, the most common state variables used are individuals' position, behavioural mode or even physiological response. In the case of behavioural mode, it is usually represented as a discrete variable with two or more nominal categories, such as "foraging" and "migration" or "encamped" and "exploratory" (e.g. Jonsen et al., 2005 and Morales et al., 2004). In terms of choosing the number of state categories, it is particularly important to find an appropriate balance between realism and feasibility, regarding both the biology of the animal and the quality of the data. At the same time, it is necessary to consider the capability of defining each state differently from the rest, as in some cases it can be very complicated to model states separately.

When analysing movement data, it is important to point out that in order to model the movement of an animal, it is required to model the trajectory of it through the space. However, although animal movement is a continuous-time phenomenon, it is often convenient to treat space as a discrete quantity and approximate a continuous situation with a finite set of locations,

for instance, in terms of daily positions (Jonsen et al., 2013).

In general, the nature of the SSM frameworks makes relatively straightforward to specify a biologically meaningful model of animal movement and to accommodate the properties of different data-collection methods. However, fitting these models to data is not straightforward (Patterson et al., 2008). A SSM and their variables can be fitted by using either maximum likelihood or Bayesian methods, but a common difficulty of both approaches is the evaluation of the likelihood, which is extremely demanding from a computational point of view. One of the firsts examples of a SSM was the *Kalman Filter* (KF) (Kalman, 1960) (see Box 2) which is still used in a wide range of applications. This two-pass recursive algorithm efficiently calculates the marginal likelihood when states are continuous and observations have Gaussian errors. This method is highly attractive because it involves rapid analytical updating of state predictions with data without recourse to intensive simulation. However, the KF is sub-optimal when errors are not Gaussian, thus, other approaches must be considered.

Box 2: *Kalman Filter*

The Kalman Filter provides an unbiased estimator for a stochastic linear system, written in the following state-space form:

$$x_t = Ax_{t-1} + Bu_{t-1} + v_{t-1} \tag{2.3}$$

$$y_t = Hx_t + e_t \tag{2.4}$$

where (2.3) describes the evolution of the animal position and state x_t and (2.4) defines how x is partially observed with measurements y_t through H. A and B are transition and control-input models, respectively, vector u_{t-1} is an optional drift component related to where next location is expected to be located, v_{t-1} and e_t are Gaussian distributed errors, $v \sim \mathcal{N}(0, Q)$ and $e \sim \mathcal{N}(0, R)$.

For non-linear models and non-Gaussian data (such as the dataset we used), non-linear SSM based on Bayesian simulation techniques are commonly used (Tanizaki, 2001). In particular, when studying movement trajectories, non-linear methods are often required since animal movements, or changes between behavioural modes, are inherently non-linear. In the simplest case, behaviour changes in a simple switch from one type of behaviour to another. In consequence, accounting for the nature of the dataset available for this project, from now on we will focus on non-linear SSM frameworks.

Under a non-linear State-Space representation, using ARGOS data as an example, the true unobserved location of the PTT, x, follows a process model described by the transition equation:

$$x_t = f_t(x_{t-1}) + \eta_t \tag{2.5}$$

where t is the time index of the satellite pass, f_t is the state-transition function and $\eta_t \sim \mathcal{N}(\bar{\eta}_t, Q_t)$ is the process noise that describes the inherent randomness in movement. The Doppler observation model yields the observed location (e.g. ARGOS location), y, from the true unknown location, x:

$$y_t = h_t\left(x_t\right) + \epsilon_t \tag{2.6}$$

where h_t is the Doppler observation function and $\epsilon_t \sim \mathcal{N}(\bar{\epsilon}_t, R_t)$ is the measurement noise that includes the observation error. For (2.5) and (2.6), the sequences $\{f_., \bar{\eta}_., Q_.\}$ and $\{h_., \bar{\epsilon}_., R_.\}$ are known, and the noises η_t and ϵ_t are mutually independently distributed and independent of x_{t-1}.

The combination of these two equations makes possible to estimate the unobserved location, x_t in this example, as well as behavioural modes or physiological response among others process.

2.3 Harbour seal (*Phoca vitulina*)

Harbour seals (*Phoca vitulina*, Linnaeus, 1758) are one of the most widely distributed pinnipeds. They can be found in coastal habitats of both the North Pacific and the North Atlantic, including the Baltic and North Seas (Bonner, 1989) and inhabiting different types of ecosystems, from high Arctic water and oceanic shores in temperate zone, to inshore waters and estuaries (Härkönen and Heide-Jorgensen, 1990). This broad distribution contributes to the creation of several subspecies. Currently, 5 subspecies are recognized: 2 in the Atlantic, the Eastern and the Western Atlantic subspecies, 2 in the Pacific, the Eastern and the Western Atlantic subspecies, and the last subspecies in freshwater lakes and rivers on the northern Quebec (Figure 2.2).
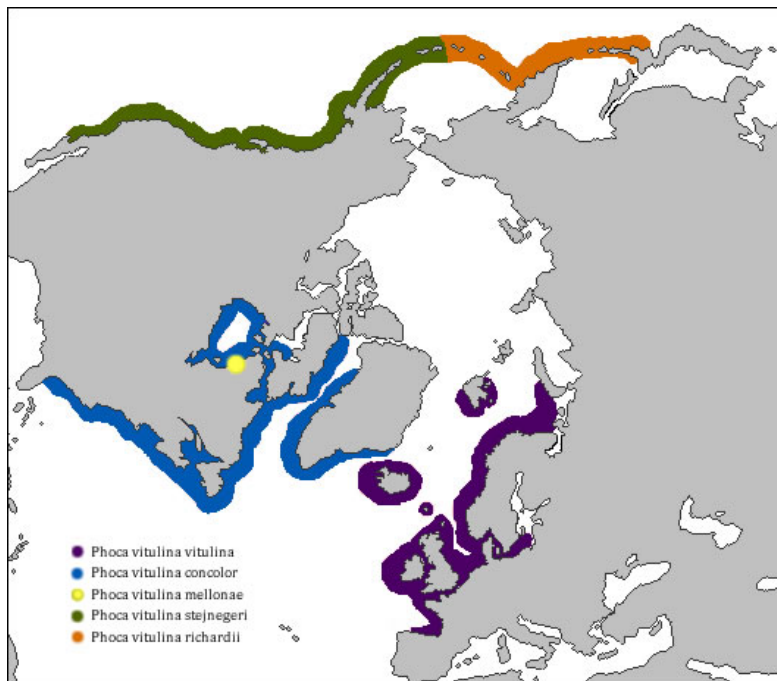

Figure 2.2: Distribution of 5 harbour seal subspecies around the world. Each colour represents a specific subspecies distribution. Purple, blue, yellow, green and orange areas mark the region of the Eastern Atlantic (P. v. vitulina), Western Atlantic (P. v. concolor), Northern Quebec (P. v. mellonae), Western Pacific (P. v. stejnegeri) and Eastern Pacific (P. v. richardii) subspecies, respectively. Figure modified from Bonner (1989).

This study is focused on the Eastern Atlantic subspecies, *Phoca vitulina vitulina*. This subspecies inhabits coastal habitats of Iceland, Ireland and Scotland, the east coast of England, the Wadden Sea, the western Scandinavia and the east Baltic (Härkönen et al. 2005), and it is therefore divided into 6 populations that can be clearly distinguished according to satellite tracking and genetic analysis results (Härkönen and Hårding, 2001; Olsen, 2006). In particular, this study is performed on the Danish Wadden Sea population.

The Wadden Sea is part of the North Sea. Extending over 500 km along the Northwestern coasts

of The Netherlands, Germany and Denmark (Figure 2.3), it has a surface of approximately 9000 km^2 of which 7350 km^2 are tidal areas. This shallow, highly productive habitat is characterised by high-sands and large inter-tidal sand banks, interspersed with tidal channels. It is in these high-sands and inter-tidal banks where this population of harbour seals is usually found.
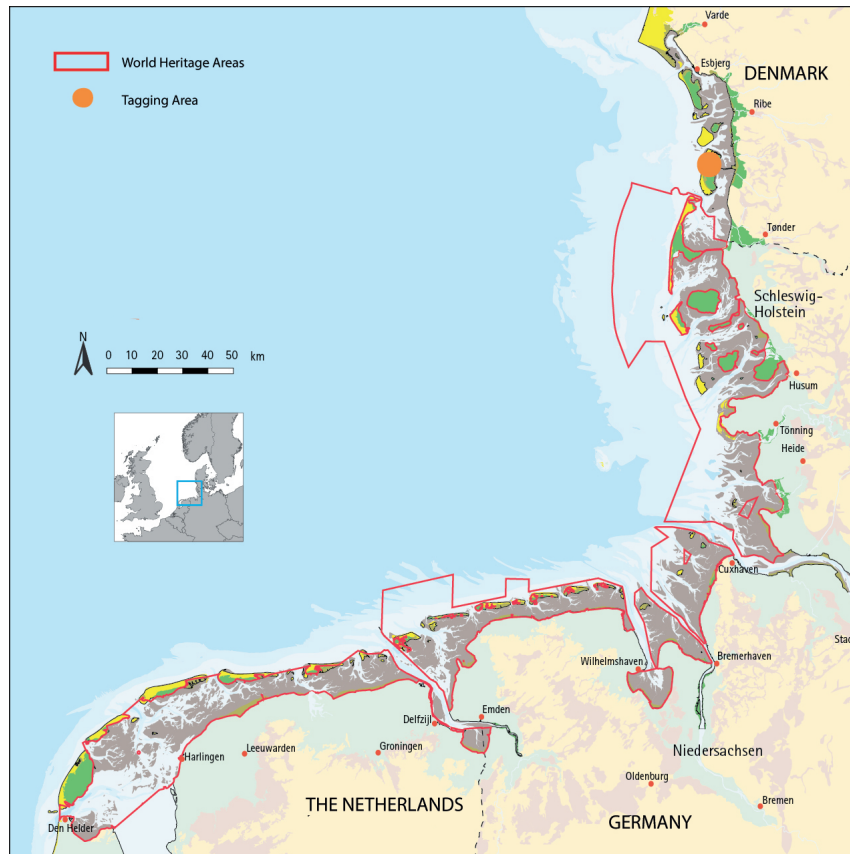
Figure 2.3: The Wadden Sea area in the North Sea stretching from Denmark in the north to the Netherlands in the south. The orange dot marks Rømø Island, where animals were caught, tagged and released during data collection. The red lines surround areas cataloged as World Heritage Areas by the UNESCO.

Harbour seals move little between haul out banks, and in consequence, different subpopulation can be distinguished within the International Wadden Sea harbour seal population, among them the Danish Wadden Sea harbour seal population (Tougaard et al., 2003). This population of harbour seals is poorly studied in terms of movement behaviour. Prior to this study, only a few studies using radio-transmitting systems have been performed.

Biology

Harbour seals are one of the smallest pinnipeds and, unlike other phocids species, has a barely visible sexual dimorphism in which males are slightly larger than females. A male or female adult can respectively reach a length of to 1.90 or 1.70 meters and weigh between 70 and 150

or 60 and 110 kilograms, which highly depends on the season (Härkönen and Heide-Jorgensen, 1990). Males reach their sexual maturity after 4-5 years and physical maturity after 7-9 years and females after 3-4 years and 6-7 years, respectively (Härkönen and Heide-Jorgensen, 1990). This lack of sexual dimorphism and the time until they reach the physical maturity is important in terms of individuals distribution, as travelling from haul out site to forage site may change depending on the size of the animal and storage energy.

In general, harbour seals spend most of their time in water, they optimise their foraging strategy by spending a high proportion of time submerged (Ries et al., 1997). Individuals usually forage in shallow bays and coastal zones, where they normally dive for 2 - 3 minutes between breaths. Nevertheless, this species is capable of diving down several hundred meters and remain submerged for more than 30 minutes (Ries et al., 1997). In terms of movement, the usual cruise speed of these animals does not reach more than few kilometers per hour, but speeds up to 20 km/h during short bursts have been reported (Williams and Kooyman, 1985).

Harbour seals breed in groups of two or more than several hundred of individuals. Females give birth to one pup at a time during the breeding season, which in our study population specifically occurs in late June. Pups suckle for a period of 3-4 weeks, during which females do no fast (Tougaard et al., 2006). After this period, around the middle of August, pups are totally independent and need to forage by themselves. The temporal and spatial behaviour pattern of males seem to relate to variations in females' distribution and density (Van Parijs et al., 1997). Mating occurs immediately after weaning of the pup, followed by gradual movement to the high sands along the western edge of the Wadden Sea, where moulting occurs. The Danish Wadden Sea is the only breeding area for harbour seals on the Danish west coast and also the only area where harbour seals haul out regularly in large numbers (Tougaard et al., 2006).

Haul out Behaviour

Hauling out is a behaviour associated with pinnipeds in which, for a period of time, individuals leave the water and stay on land. Hauling out is normally associated with mating and giving birth, but other benefits may include predator avoidance, thermal regulation, social activity and rest. When hauling out, seals normally are found in small scattered groups, nonetheless big aggregations of individuals can be seen in specific locations.

Haul out behaviour is typically observed during the pupping, mating and molting season, from May through August in the Eastern Atlantic (Härkönen and Hårding, 2001). A variety of habitats are used for hauling out, including rocky shores, reefs, ice floes or sand banks and intertidal mud, the most important ones for this study population. As this population mainly hauls out in tidal areas, they need the tide to be low and, therefore, the timing of haul outs often depends on tidal cycles. In the absence of tides, the time of day is the driving factor for harbour seal haul out behaviour.

Haul out sites are usually selected in connection with protection from land predators, access

to deep water, which means proximity to food sources, and protection from wind and waves. Harbour seals are known for moving little between haul out banks. They show high site fidelity and feed close to their haul out sites, individuals tend to stay in the same area all year round and make short trips within 50 km from haul out points (Härkönen, 1987). However, it has been seen that juveniles can travel long distances up to 500 km offshore (Bjorge et al., 2002). Particularly, a previous study in the Danish Wadden Sea area has shown that a majority of the studied seals remained connected to the same (56%) or within a radius of 25 km from the haul out bank (34%) where they were caught and tagged. Only 10% moved to banks about 75 km away and a single animal was found in the German Wadden Sea, about 150 km away (Nørgaard, 1996). In consequence, the exchange between population of the three Wadden Sea countries is considered to be limited (Tougaard et al., 2003). Besides, Nørgaard (1996) found that the animals showed a seasonal difference in location between winter and summer. The results showed that seals tend to spend more time out in winter than in summer and seals also spent significantly more time on land in summer compared to winter. No correlation between amount of time far away from the haul out point and sex or age of animals was found.

Foraging Ecology and Study Methods

Harbour seals mostly feed on fish and marine invertebrates but are generally opportunistic animals in which feeding depends on prey availability (Härkönen, 1987). It is known that seals perform foraging trips in which they leave the hauling out areas and make extensive several-day long trips (Härkönen, Hårding, and Lunneryd, 1999). Few studies are available on feeding biology of harbour seals in the Wadden Sea, but from them it can be seen that this area offers good foraging spots to this animal species (Tougaard et al., 2006).

Historically, the easiest way to study the foraging ecology of an animal has been to directly observe its behaviour (Naito et al., 2010). However, to directly observe the foraging behaviour of an animal in a marine environment can be extremely challenging if not impossible, as such behaviour often takes place below the surface and in remote areas. In particular, in order to deal with this limitation, a number of techniques and devices have been used for studying the foraging ecology of free ranging pinnipeds (Cooke et al., 2004; Naito et al., 2010). In some cases, diving behaviour, associated behavioural metric and physical and/or biological characteristics of the marine environment have been used as a proxy for the animals' behaviour when submerged. Diving behaviour of seals include several functions, such as foraging, reproduction, communication, travelling, resting or predator avoidance (Madden et al., 2008) and these different behaviours have mostly been related to specific dive shapes, such as the U-shaped, the V-shaped or W-shaped dives. Some studies suggest that a certain dive shape could be related to foraging behaviour of diving animals (Lesage et al., 1999). However, it is important to note that some of these behaviours can be performed simultaneously in the same dive, which can consequently be a limitation of this approach.

Recently, new methods have arisen in order to determine the underwater behaviour of individuals. Among them, temperature transmitters placed inside the animal, that can detect

changes of temperature as a result of incoming water due to the ingestion of prey, presumably, and mandibular sensors and accelerometers, that register jaw opening and closing, as well as changes in acceleration of the individual linked to prey hunting. Other methods are being developed for those cases in which this kind of physiological data is not available. The use of statistical models, in which the behaviour is estimated using parameters associated to the movement of the individuals, have recently become an important approach for determining the foraging behaviour of some population species (e.g. Boeuf et al., 2000).

Chapter 3

Case Study:
21 Harbour Seals from the Danish Wadden Sea

Little is known about the distribution and foraging ecology of the Danish Wadden Sea population of harbour seals. Thus, with the intention of better understanding both distribution and foraging ecology of this seal population, during 2002 and 2004, 21 individuals were caught, tagged and released on Rømø Island, at the west coast of Southern Denmark, on 7 separate occasions.

Animals were equipped with ARGOS transmitters (PTT), glued to seals' fur on the top of the head with fast hardening epoxy glue, which were specifically programed for surveys to be as efficient as possible. PTT were programmed to transmit daily, with a maximum of 500 uplinks allowed per day, and prevented from transmitting during night hours (22:00 to 03:00 h local time), due to poor satellite coverage during during night hours. Moreover, by means of a conductivity sensor carried for the PTT able to detect when it is dry or wet, PTT were programed to maximize its life by shutting down the transmission after two hours of dryness.

In addition to spatio-temporal information collected with the Satellite transmitters, biological characteristics of the tagged individuals were collected during the different surveys. When caught, animals were sexed, measured and weighted, and from this information the age of the individual was estimated. This kind of information is very important if differences between pattern distribution like to be related to biological characteristics of the individuals.

It has been a long time since this database was collected. Three previous studies have resulted from it, Tougaard et al. (2003), Tougaard et al. (2006) and Tougaard et al. (2008), but none of them present a clear method for determining where seals move and forage. With the present study, I aim to finally find a decisive method for solving questions about distribution and behaviour of the seal population.

3.1   Data Pre-Processing

In order to proceed with the study, data were pre-processed whereby movement metric variables were created from spatio-temporal information. Information about distance from tagging place, distance, time and speed between positions was calculated, as well as bearing of the animal, understanding bearing as the angular direction measured from one position to another using the North as a geographical reference line. Table 3.1 shows a brief example of the database.

Table 3.1: First 10 registers of the individual with ID = 1. For all the animals, the departure position was located at the same geographical point.

Date Time (dd/mm/yy hh:mm:ss)	Longitude (degrees)	Latitude (degrees)	LC	Distance (km)	Speed (km/h)	Bearing (degrees)
(01/04/02 07:00:00)	8.514	55.212	DP	0.000	0.000	0.000
(01/05/02 07:27:16)	8.471	55.209	1	2.747	0.112	263.046
(01/05/02 07:45:29)	8.482	55.192	A	2.014	6.633	159.727
(01/06/02 07:05:51)	8.540	55.220	B	4.817	0.206	49.745
(01/06/02 08:48:19)	8.445	55.276	B	8.656	5.069	316.000
(01/06/02 10:54:35)	8.412	55.221	0	6.459	3.069	198.895
(01/07/02 06:47:10)	8.511	55.213	A	6.338	0.319	98.022
(01/07/02 08:22:57)	8.431	55.205	1	5.150	3.226	260.092
(01/07/02 16:35:50)	8.424	55.204	0	0.458	0.056	255.948
(01/08/02 10:34:51)	8.464	55.207	A	2.558	0.142	82.496

LC = Location Category   DP = Departure Position

Due to the spatial location error inherent to ARGOS location data, it is advisable to exclude outliers locations before performing a more complex analysis (Royer and Lutcavage, 2008; Patterson et al., 2010). Consequently, a search of extreme observations was done. Firstly, a speed threshold was applied. According to biological reports, swimming speeds of harbour seals rarely reach speeds faster than 20 km/h (Williams and Kooyman, 1985), consequently, in order to avoid extreme positions and based on the idea that filters based on speed give good results, the database was filtered using this threshold. Secondly, we looked for extreme locations, understanding them as unrealistic positions far from the main cluster of dots. In this case, no extreme positions were found at sea, however, there were some on land. It is known that Wadden Sea harbour seals do not go far in land, but rather, remain in sandbanks close to shore, thus, positions further in to the last sea point were removed from the analysis.

3.2   Descriptive and Exploratory Analysis

After the pre-processing of the data, a descriptive analysis of our database was performed in order to inspect its nature. To start, a descriptive analysis of the biological characteristics of the individuals studied in this project was done. The frequency and percentage for each gender,

gender-age category and minimum and maximum values of length and weight are given. This was then followed by a description of the spatio-temporal information. Firstly, an accurate description of the frequency and percentage of positions belonging to each LC was done and, secondly, median follow-up days and uplink per day along with the frequency and percentage positions during each season were given. During this descriptive analysis, we also looked at how many days in a row the individuals seemed to remain offshore, presumably in foraging trips. This information was crucial in determining how many positions per day were needed in order to capture the actual behaviour of the seals.

In addition to the descriptive analysis, an exploratory analysis using graphical representations of the distribution of the animals in the geographical study area was performed. This allowed for visualization of different individuals' tracks and assess the existence of patterns. Moreover, visual inspection and boxplots were used in order to observe whether the distribution and distance travelled from tagging place were related to any of the main covariables of the study, both the ones related to biological characteristics, age, gender and length, and the environmental and temporal one, season.

3.3 State-Space Model Analysis

The database analysed in this project consists of seal ARGOS locations observed irregularly through time. Animal locations from the same individual are rarely independent, but rather, produce a complex, autocorrelated time-series that may comprise multiple types of movement behaviour. Moreover, one of the biggest problems of using ARGOS system is the low and variable accuracy of some of the locations and, therefore, the large number of extreme locations. In consequence, due to the nonlinear character of the movement trajectories data and the inherent source of error of ARGOS locations, a non-linear SSM was implemented using a Bayesian approach that affords efficient computation and allows prior information to be incorporated into the analysis.

In order to estimate the movement pathways of the tagged seals, and to probabilistically assign locations to a behavioural mode, we used a two-steps model result of the combination of two Bayesian SSM frameworks:

1. The Bayesian SSM described in Jonsen et al. (2005) was applied to analyse and filter movement data acquired through ARGOS system; and

2. The Bayesian SSM described in Morales et al. (2004) and in Eckert et al. (2008) was applied to assign behavioural states to each position; both explained below.

3.3.1 Filtering Stage

Different satellite uplink rates, duty cycles, battery life and location error rates caused large differences in both track duration and number of locations collected in both individual animals and per unit of time (Breed et al., 2006). Thus, to handle erroneous satellite locations and

normalize the number of locations per day, the framework of the SSM described in Jonsen et al. (2005) was applied.

In this stage, the *state variable* was a two-dimensional vector, representing the true location of the seal, $x_t = [longitude_t, latitude_t]'$, at regularly spaced intervals, t, while the *observed data*, $y_{t,i}$, consisted of locations from satellite tags recorded at irregular intervals (the subscript i indexes multiple locations recorded during interval t).

Based on the *mean turning angle* (θ) and the *movement correlation parameter* (γ), Jonsen's framework proposes a **process model** that uses a Correlated Random Walk (CRW) model as a transition equation (see Box 1 in Chapter 2). This CRW for location data is based on the differences in successive locations, i.e., the movements, and not on the locations themselves, as it is the case in equation (2.1). Intuitively, this approach seems more realistic as the behaviour in which we are interested is change in speed and direction, and not change in location per se. In addition, Jonsen et al. add the parameter γ, which allows for variability in the autocorrelation of direction and speed, with $\gamma = 0$ resulting in a simple RW and $0 < \gamma < 1$ yielding a RW with correlation in both direction and move speed. Finally, the CRW is written as:

$$d_t \sim \gamma T(\theta)d_{t-1} + \mathcal{N}_2\left(0, \Sigma\right) \tag{3.1}$$

where d_t is the difference between the locations x_t and x_{t-1} and T is the transition matrix that describes the rotational component of the CRW, i.e., provides the rotation required to move from d_{t-1} to d_t,

$$T\left(\theta\right) = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}, \tag{3.2}$$

where θ is the mean turning angle and $\mathcal{N}_2$ is a bivariate Gaussian distribution with mean 0 and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{lon}^2 & \rho\sigma_{lon}\sigma_{lat} \\ \rho\sigma_{lon}\sigma_{lat} & \sigma_{lat}^2 \end{pmatrix} \tag{3.3}$$

where σ_{lon}^2 and σ_{lat}^2 are the process variance in longitude and latitude, respectively, and ρ is the correlation coefficient.

The second component of Jonsen's framework, the **observation model** that relates the unobserved states predicted by the process model to the observed data, deals with the irregular sampling of the positions and the variable quality and non-Gaussian errors of ARGOS data. It is important to note that the observations recorded differ in both number and time lag between them, i.e., the number of positions and moment of the day they were collected were variable. Consequently, as the CRW assumes that the time between observations is constant, the observation model used in Jonsen's framework introduces intrinsic data regularization to obtain equal time interval, which is mathematically formulated as:

$$y_{t,i} = (1 - j_{t,i})\, x_{t-1} + j_{t,i}x_t + \epsilon_{t,i} \tag{3.4}$$

where i is the index for locations (if any are observed) between time t and $t + 1$ (i.e., $i = (0, 1, 2, ..., n_t)$), $j_{t,i}$ is the proportion of the regular time interval between x_{t-1} and x_t at which the ith observation was made ($0 < j_{t,i} < 1$). For this project, it was decided that the interval between consecutive positions was 1 day. This was chosen due to the number of positions per day, as it is advisable to select a number inferior to your average number of locations per day (Jonsen et al., 2013). Therefore, for regular intervals where no observation was made, these parameters are set to $i = 1$ and $j_t = 0.5$. Nevertheless, it is important to point out that, in doing this, we were presuming that animals travel in a straight line between x_t and x_{t+1}. Finally, $\epsilon_{t,i}$ is a random variable representing the estimation error. Due to the fact that some ARGOS position errors can be non-Gaussian and present large variability, the observation model of Jonsen's framework models location errors using a t-distribution. The t-distribution is robust for this kind of data as it has the effect of making extreme values less likely, thereby minimizing their influence on parameter estimation. In addition, given that not all ARGOS tags function equally, they include a scale parameter, ψ, estimated directly within the model that serves to inflate and deflate uniformly the standard error of the t-distribution as determined for the data. Particularly, they use ARGOS LCs error estimation from Vincent el al. (2002) (Table 3.2) to determine the appropriate t-distribution parameters to use for each position in the observation model. For estimation errors in latitude and longitude of LC, $\epsilon_{t,lat,LC} \sim t\left(0, \psi\tau_{lat,LC}, \nu_{lat,LC}\right)$ and $\epsilon_{t,lon,LC} \sim t\left(0, \psi\tau_{lon,LC}, \nu_{lon,LC}\right)$, where $\tau_{lat,LC}$ and $\tau_{lon,LC}$ are the scale parameters representing the precision, parametrized as the inverse of the variance, and $\nu_{lat,LC}$ and $\nu_{lon,LC}$ are the degrees of freedom. Therefore, if we define

$$z_{t,i} = (1 - j_{t,i})\, x_{t-1} + j_{t,i} x_t \qquad (3.5)$$

the formulation of $y_{t,i}$ will be:

$$y_{t,i} \sim t\left(z_{t,i}, \psi\tau_{t,i}, \nu_{t,i}\right) \qquad (3.6)$$

Table 3.2: Parameters for the t-distribution used by Jonsen et al. (2005) estimated from Vincent et al. (2002). For both longitude and latitude, the parameter σ represents the standard deviation used for the calculation of the scale parameter τ, $\tau = 1/\sigma^2$, and ν represents the degrees of freedom.

Location Category	σ_{lon}	σ_{lat}	ν_{lon}	ν_{lat}
3	0.290	0.122	3.071	2.076
2	0.312	0.261	1.221	6.314
1	0.902	0.460	2.299	3.897
0	2.163	1.607	0.914	1.011
A	0.507	0.511	0.787	1.058
B	4.205	3.041	1.079	1.331

For the rest of the model parameters, as it can be seen in Table 3.3, vague priors were chosen.

Table 3.3: Prior distributions and interpretation of the main parameter used in the 1st Stage of the methodology.

Parameter	Prior Distribution	Interpretation
Σ^{-1}	Wishart $(Id_2, 2)$	Process error
θ	Uniform(-π, π)	Mean direction for the turning angle of the movement path
γ	Uniform(0, 1)	Degree of autocorrelation between direction and speed of successive locations
ψ	Uniform(0, 10)	Scaling factor for the estimation error

The code used in order to adapt ARGOS data into WinBUGS software and the model applied in this stage can be found in the Appendix A and B, respectively.

From the resulting location estimations, only the first daily location estimation for each seal was selected for representing seals' trajectories. Subsequently, new movement metrics needed to be calculated: distance (km), speed (km/day) and bearing (radians) between consecutive locations, in order to adapt the data to fit into the second stage.

3.3.2   Behaviour discrimination Stage

Once the entire movement paths were modeled using Jonsen's SSM framework, SSM frameworks described in Morales et al. (2004) and in Eckert et al. (2008) were applied in order to assign a behavioural state to each position. In this context, behavioural states need to be understood as categorical actions that a certain animal can perform. Certainly, one can expect seals to perform several behaviours, but in a simplistic case, behaviours can be mainly classified into hauling out, travelling and foraging.

Jonsen et al. (2005) also present a method able to simultaneously model location errors and switches in movement behaviour but, as what happens with the filtering stage explained before, the characterization of the behavioural states is based on non-biologically intuitive parameters: mean turn angle and movement correlation parameters. Therefore, this approach was discarded and a more biologically meaningful one was used.

In this part, the method is purely based on movement parameters. Eckert's study is based on Morales' study, thus they both are based on the same framework. They assume that the movement path of an individual is composed by one or more RW (one RW for each behaviour) and expect to classify each observation as belonging to one of these RWs by taking into account the distribution of two specific movement parameters: turning angles and step length between positions and, in some of the cases, by relating the probability of changing from one RW to another to previous behavioural states or environmental conditions.

As a general model structure, each observation within an individual, $y_t(t = 1, ..., T)$, where T is the number of days, is associated with an unobserved state-indicator variable $I_t = i, i \in \{1, ..., M\}$, where M is the number of behavioural modes considered. Particularly, they define observations accounting for daily average movement rates and turning angles, $y_t = [r_t, \phi_t]$.

Moreover, conditioned on the ith behavioural state, $i \in \{1, ..., M\}$, each observation is assumed to be independently drawn from: 1) for turning angles, a Wrapped Cauchy distribution (see Box 3) with θ_i as a peak position of the "unwrapped" distribution and ρ_i as a scale parameter; and 2) for step length, a Weibull or Truncated Normal distribution (see Box 3), depending of the model variation, with a_i as a scale parameter and b_i as a shape parameter or μ_i as a location parameter and σ_i as a scale parameter, respectively.

Box 3: *Probability Distributions*

Wrapped Cauchy Distribution

The Wrapped Cauchy distribution is a wrapped probability distribution, i.e., a continuous probability distribution that describes data points that lie on a unit n-sphere. In 1 dimension, a wrapped distribution consists of points on the unit circle. Wrapped Cauchy distributions are governed by two parameters: the mean direction θ and the mean cosine of the angular distribution ρ, and takes the following density function:

$$C(\phi; \theta, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\phi - \theta)} \qquad (3.7)$$

where $0 \leq \phi \leq 2\pi$, $-\pi \leq \theta \leq \pi$ and $0 \leq \rho \leq 1$. As ρ goes to 0, the distribution converges to a uniform distribution over the circle. As ρ goes to 1, the distribution tends to one point, concentrated in the direction of θ.

Weibull Distribution

The Weibull distribution is a continuous probability distribution governed by two parameters: the shape parameter b and the scale parameter a, and takes the following density function:

$$W(r; a, b) = abr^{b-1} exp(-ar^b) \qquad (3.8)$$

where $r \geq 0$, $a > 0$ and $b > 0$. When $b = 1$, the distribution is reduced to an exponential distribution, when $b \geq 1$, it has an exponential tail and when $b < 1$ it has a fat tail.

Truncated Normal Distribution

The Truncated Normal distribution is a conditional distribution that results from restricting the domain of a Normal distribution either below, above or both. As in a Normal distribution, the Truncated Normal distribution is governed by two parameters: the location parameter μ and the scale parameter σ, but in this case, the variable r lies within a defined interval, $r \in (a, b)$ and takes the following density function:

$$N(r; \mu, \sigma)_{tr\{a,b\}} = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(r-\mu)^2}}{\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(r-\mu)^2} \, dr} \qquad (3.9)$$

We have used two options for modelling the step length of an individual: a Normal and a Weibull distribution, following Eckert et al. and Morales et al., respectively. Morales et al. propose the use of a Weibull distribution to model the distance moved between successive locations claiming that it is flexible enough to accommodate different forms of movement, for instance, departures from the RW or CRW models. On the contrary, Eckert et al. prefer to assume that movement rates of the behavioural states follow a Normal distribution, due to the ease of presenting one state as faster than another, $\mu_i < \mu_j$, especially, when more complex models are intended to be fitted. For this case, in order to cope with the possibility of having negative values, the distribution was truncated for those values. Therefore, with the aim of testing both approaches, both formulations were applied.

Finally, accounting for the two possible distributions for modelling the step length and for a given vector of behavioural states or modes I, the likelihood functions for our models were:

$$P(y|a, b, \theta, \rho) = \prod_{t=1}^{T} W(r_t|a_{I_t}, b_{I_t}) C(\phi_t|\theta_{I_t}, \rho_{I_t}) \qquad (3.10)$$

$$P(y|\mu, \sigma, \theta, \rho) = \prod_{t=1}^{T} N(r_t|\mu_{I_t}, \sigma_{I_t}) C(\phi_t|\theta_{I_t}, \rho_{I_t}) \qquad (3.11)$$

where $y = (y_1, ..., y_T)$ and W, C and N denote Weibull, Wrapped Cauchy and Normal distributions, respectively.

In addition, as previously stated, the model contains a state variable, I_t, created in order to have an indicator of the behavioural state of the position.

Different movement models can be constructed by fitting different numbers of RW models to the data and by making the switching rate between these different RWs fixed or dependent on one or more different features (Morales et al., 2004). For this study, 10 different model variations of the same general structure were applied (an example of the WinBUGS code is provided in Appendix C). On one hand, we present the possibility of having 1, 2 or 3 different behavioural states generating the entire path. In this case, it was assumed that seals perform either a single behaviour, mostly we use this assumption as a null hypothesis, or multiple behaviours during these trips: haul out, forage and/or travel. Accounting for the movement metrics used in this study, when two behaviours were assumed, it was expected to find one faster and less variable, in terms of turning angles, state and another one slower and more variable. In addition, when three behaviours were assumed, it was expected to find again a faster less variables state and, in this case, two slower and more variable states, where hauling out was assumed to be slower than foraging. On the other hand, we also present the possibility that consecutive behavioural states are dependent or independent, i.e., the fact of being in a determined behavioural state at time $t-1$ can affect or not the behavioural state at time t.

Model Variations

Here, the different model variations are explained according to the number of expected behaviours and relation between consecutive locations:

1. *"Single W"*. A single RW. The entire movement path was assumed to be generated within a single behavioural state, where step length was considered to follow a Weibull distribution.

2. *"Single N"*. The same as model (1), but step length was assumed to follow a Normal distribution. As explained before, in this case, the Normal distribution was truncated in order to not draw values < 0.

3. *"Double W"*. Mixture of two RW. The movement path was assumed to be generated by two behavioural states. In this case, the step length was considered to follow a Weibull distribution and it was constrained to have larger scale parameter in behavioural state 2 than 1. Each observation was assigned to one behavioural state independently of previous states and I_t was a binary indicator with probability distribution:

$$P(I_t) = \begin{cases} p & if \quad I_t = 1 \\ \\ 1-p & if \quad I_t = 2 \end{cases} \tag{3.12}$$

   where p is the probability of being in behavioural state 1.

4. *"Double N"*. The same as model (3), but step length was assumed to follow a Truncated Normal distribution and constrained to have smaller location parameter and larger scale parameter for behavioural state 1 and larger location parameter and smaller scale parameter for behavioural state 2.

5. *"Double-Switch W"*. The same as model (3), but the probability of being in a behavioural state was explicitly modelled and related to previous state. At each time step, an individual can decide to change from the current behavioural state to a different one with a fixed probability. When two possible behavioural states, a 2×2 matrix that contains probabilities q_{ij} was defined. These q_{ij} were the probability of being in the behavioural state j at time t given that the individual was in state i at time $t-1$. Now the conditional probability distribution for I_t will be:

$$P(I_t|I_{t-1}=1) = \begin{cases} q_{11} & if \quad I_t = 1 \\ \\ q_{12} = 1-q_{11} & if \quad I_t = 2 \end{cases} \tag{3.13}$$

   and

$$P(I_t|I_{t-1}=2) = \begin{cases} q_{21} & if \quad I_t = 1 \\ \\ q_{22} = 1-q_{21} & if \quad I_t = 2 \end{cases} \tag{3.14}$$

6. *"Double-Switch N"*. The same as model (4), but the probability of being in a behavioural state was explicitly modelled and related to previous states. The expression of I_t followed the same formulated as model (5).

7. *"Triple W"*. Mixture of three RW. The movement path was assumed to be generated by three behavioural states. In this case, the step length was considered to follow a Weibull distribution and was constrained to have larger scale parameter in behavioural state 3 than 2, and behavioural state 2 larger than 1. Each observation was assigned to one behavioural state independently of previous states and the conditional probability distribution for I_t was:

$$P(I_t) = \begin{cases} p_1 & if \quad I_t = 1 \\ p_2 & if \quad I_t = 2 \\ 1 - p_1 - p_2 & if \quad I_t = 3 \end{cases} \qquad (3.15)$$

8. *"Triple N"*. The same as model (7), but step length was assumed to follow a Truncated Normal distribution and constrained to have smaller location parameter and larger scale parameter for behavioural state related to shorter step length, and larger location parameter and smaller scale parameter for behavioural state related to longer step length.

9. *"Triple-Switch W"*. The same as model (7), but the probability of being in a behavioural state was explicitly modelled and related to previous state. In this case I_t was an indication with three categories and was conditionally formulated as:

$$P(I_t|I_{t-1} = 1) = \begin{cases} q_{11} & if \quad I_t = 1 \\ q_{12} & if \quad I_t = 2 \\ q_{13} = 1 - q_{11} - q_{12} & if \quad I_t = 3, \end{cases} \qquad (3.16)$$

$$P(I_t|I_{t-1} = 2) = \begin{cases} q_{21} & if \quad I_t = 1 \\ q_{22} & if \quad I_t = 2 \\ q_{23} = 1 - q_{21} - q_{22} & if \quad I_t = 3 \end{cases} \qquad (3.17)$$

and

$$P(I_t|I_{t-1} = 3) = \begin{cases} q_{31} & if \quad I_t = 1 \\ q_{32} & if \quad I_t = 2 \\ q_{33} = 1 - q_{31} - q_{32} & if \quad I_t = 3 \end{cases} \qquad (3.18)$$

10. *"Triple-Switch N"*. The same as model (8), but the probability of being in a behavioural state was explicitly modelled and related to previous state. The expression of I_t followed the same formulated as model (9).

For the rest of the model parameters, as it can be seen in Table 3.4, vague priors were chosen.

3.3.3   Parameter estimation

In both stages, parameter estimations were conducted using a Bayesian MCMC (see Box 4) approach in WinBUGS 1.4.3. For each model, MCMC runs consisted of three chains where a total of 100 000 samples were generated in each chain. A total of 50 000 samples were discarded as a burn-in and every fifth sample thereafter was retained to reduce autocorrelation.

Table 3.4: Prior distributions and interpretation of the main parameter used in 2nd stages of the methodology.

Parameter	Prior Distribution	Interpretation
a_1	Gamma(0.01, 0.01)	Scale parameter for Weibull distribution describing step length for the 1st behavioural state.
a_i	Gamma(0.01, 0.01)$_{tr\{a_{i-1},\}}$ $i \geq 2$	Scale parameter for Weibull distribution describing step length for the ith behavioural state.
b_i	Gamma(0.01, 0.01)	Shape parameter for Weibull distribution describing step length for the ith behavioural state.
μ_1	Normal(0, 0.01)$_{tr\{0,\}}$	Location parameter for Truncated Normal distribution describing step length for the 1st behavioural state.
μ_i	Normal(0, 0.01)$_{tr\{\mu_{i-1},\}}$ $i \geq 2$	Location parameter for Truncated Normal distribution describing step length for the ith behavioural state.
$1/\sigma_1^2$	Normal(0, 0.01)$_{tr\{0,\}}$	Scale parameter for Truncated Normal distribution describing step length for the $1^{s}t$ behavioural state.
$1/\sigma_i^2$	Normal(0, 0.01)$_{tr\{\mu_{i-1},\}}$ $i \geq 2$	Scale parameter for Truncated Normal distribution describing step length for the ith behavioural state.
θ_i	Uniform($-\pi$, π)	Mean direction for turning angles for the ith behavioural state.
ρ_i	Uniform(0, 1)	Mean cosine for turning angles for the ith behavioural state.
p_i	Uniform(0, 1)	Probability that the tth observation was in the behavioural state i.
q_{ij}	Uniform(0, 1)	Transition probability from the ith to the jth behavioural state.

Consequently, the posterior distribution of each parameter was estimated from a sample of $3 \times 10\,000$ independent MCMC observation. Model convergence and autocorrelation of the chains were assessed by visual inspection of the trace and autocorrelation plots and using the Gelman and Rubin diagnostic available in R package *coda* (Plummer et al., 2006).

Box 4: *Markov Chain Monte Carlo*

The Markov Chain Monte Carlo (MCMC) method is a general simulation method for iterative sampling from the posterior distribution and computing posterior quantities of interest. MCMC methods sample successively from a target distribution and each sample depends on the previous one. For this model, MCMC simulates values of the states and parameters, conditional on previously generated values, until the chain of samples converges to the posterior distribution. Any property of the posterior is approximated by the corresponding property of the samples in the same way that a sample mean approximates a population mean. Theory guarantees that an infinite MCMC sample will be representative of the posterior, but in practice only finite samples can be drawn and considerable care must be taken to ensure that these are representative of the posterior.

For all parameters, the expectation (i.e. mean) of the posterior distribution was selected as an appropriate estimator. However, as an exception, the mode of the posterior distribution of I_t was used as an estimator due to the discrete nature of the variable.

3.3.4 Model Selection - Validation

Once all the models were generated, the fit of the SSMs to the data needed to be checked, followed by the selection of most suitable model in representing the hidden process shaping the data. However, model checking and selection in a SSM context, and even, Bayesian models, remain areas of active statistical research. In general, it needs to be taken into account that there is no precise criterion for assessing whether a model is good enough, it will involve individual judgment and consideration of the model purpose. Therefore, standard tools for model checking and selection, such as Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and Deviance Information Criteria (DIC), need to be interpreted with caution, as it is not clear how the addition of one parameter to the model would affect the likelihood (Jonsen et al., 2013).

Presently, the main guidances is to check whether parameters estimates make intuitive and biological sense. Unrealistic parameters can result in over-parametrised models, often due to too many states. Therefore, it is advisable to create informative plots estimating functional relationships between parameters and covariables (Jonsen et al., 2013). Moreover, the examination of the posterior predictive is also recommended to assess the fit of a model to data. It consists of drawing simulated values from the posterior predictive distribution of replicated data and comparing these samples to the observed data. For a model to be good, replicated data should be indistinguishable, potential failings of the model would be indicated by systematic differences between the simulations and the data.

In this project, it was decided to do a preliminary filtration of the models, in which a range of possible models were selected according to visual inspection of the results. Posteriorly, from the resulting models, the joint posterior distribution of parameters generated by the MCMC simulation can be used to check the ability of models to reproduce observed properties of the data.

Biological Selection-Validation

Little is known about the foraging behaviour of this seals' population, thus it is complicated to proof whether these results are valid, as no valid record of foraging locations exists. From a visual point of view, the biological and ecological feasibility of the behavioural states distribution was evaluated. As stated before, harbour seals usually leave haul out areas to carry out foraging trips that can last few days. Consequently, taking into account this knowledge about seals' behaviour, the distribution of the behavioural states during a determined seal track was expected to show a realistic pattern in which the animals travel until they reached a certain point at which they remained for a while and foraged.

Posterior Predictive Assessment

From the selected models, following the same criterion as Morales et al. (2004), it was checked whether movement paths simulated with model parameters could produce autocorrelation functions (acf's) for mean daily movement rates similar to those observed in the data. In these cases, autocorrelation in movement rate reflected the temporal structure of changes in movement behaviour. During the analysis, we sampled from the joint posterior distribution of model parameters for the 30 000 replicates. Posteriorly, a movement path was simulated with each set of sampled parameters, and the acf of daily distance moved was calculated. In this way, a "posterior predictive distribution" was produced (Brooks and Gelman, 1998) for the acf that could be compared to the observed one. It is important to note that by checking on the posterior predictive distribution of the autocorrelation function, we are assessing the ability of models to fit a property of whole movement paths that are not explicitly included in the model (Morales et al., 2004).

3.4   Biological inference

Once the selection of the best model was done, the estimation of the different parameters was assessed with regard to their biological interpretation. Furthermore, the behavioural results were used in order to evaluate the spatial relationship of the different behaviour locations. The relationship between different behaviour locations and biological and environmental variables was analysed using graphical and geographical representations.

Chapter 4

Results

4.1 Descriptive and Exploratory Analysis

The first descriptive analysis showed that, among the 21 tagged animals, 13 (61.90%) were males and 8 (38.10%) females. Specifically, within males there were 7 adults (33.30%), 4 subadults (19.00%) and 2 pups (9.50%), and within females there was 1 adult (4.80%), 3 subadults (14.30%) and 4 pups (19.00%). The length of these individuals varied from 92 to 180 cm, and the weight ranged from 19 to more than 110 kg. It is important to remark that the weight of the animals is highly related to the season at which measurement was done, thus it is not a reliable information for posterior extrapolations (Table 4.1).

Table 4.1: Details of the 21 tagged harbour seals using ARGOS system in seven different occasions. Individuals of both sexes and three age categories were tagged. It can be observed in parenthesis the number of positions remaining after the filtering.

ID	Tagging Date	Follow-up (days)	N. of Positions	Sex	Age	Weight (kg)	Length (cm)
1	2002/01/04	83	181 (180)	Male	Subadult	33	141
2	2002/01/04	73	237 (234)	Female	Pup	24	107
3	2002/01/04	80	220 (218)	Female	Pup	26	111
4	2002/01/04	68	171 (170)	Male	Pup	25	116
5	2002/02/18	88	221 (220)	Male	Adult	>110	172
6	2002/02/18	76	288 (275)	Female	Subadult	43	143
7	2002/02/18	98	112 (108)	Male	Adult	>110	164
8	2002/05/06	49	109 (106)	Female	Subadult	44	134
9*	2002/05/06	60	225 (0)	Male	Subadult	47	139
10	2002/05/06	48	187 (178)	Female	Subadult	52	150
11	2003/12/02	180	266 (257)	Female	Adult	80	165
12	2004/09/18	92	320 (317)	Male	Adult	81	180
13*	2004/09/18	85	199 (0)	Male	Subadult	60	148
14*	2004/09/18	5	22 (0)	Male	Adult	82	175
15	2004/09/18	213	756 (741)	Male	Subadult	64	149
16*	2004/12/07	88	42 (0)	Male	Adult	85	154
17	2004/12/07	173	174 (171)	Male	Pup	30	101
18	2004/12/07	102	152 (149)	Female	Pup	19	92
19	2004/12/07	108	80 (77)	Female	Pup	24	94
20	2004/12/07	198	123 (122)	Male	Adult	92	150
21	2005/03/19	133	169 (169)	Male	Adult	87	172

ID = Animal identification

* = Animals removed from the final analysis

From these 21 individuals, a total of 4254 locations were collected. Out of these, 593 (12.46%) positions belonged to Location Classes 3 to 1 (most precise), while the rest of the positions were of quality 0, A and B (least precise). Individuals were followed-up a median of 100 days, ranging between 5 and 213 days. The average number of positions transmitted per day was 2.33, ranging between 0.48 and 4.40 (Table 4.1). As it can be observed in Figure 4.1, even though animals were tagged in seven different occasions, the study was mainly performed during two time periods, 2002 and 2004, in which observations were distributed all year around, with the exception of summer, where few positions were collected due to animals lost their tags during moulting.
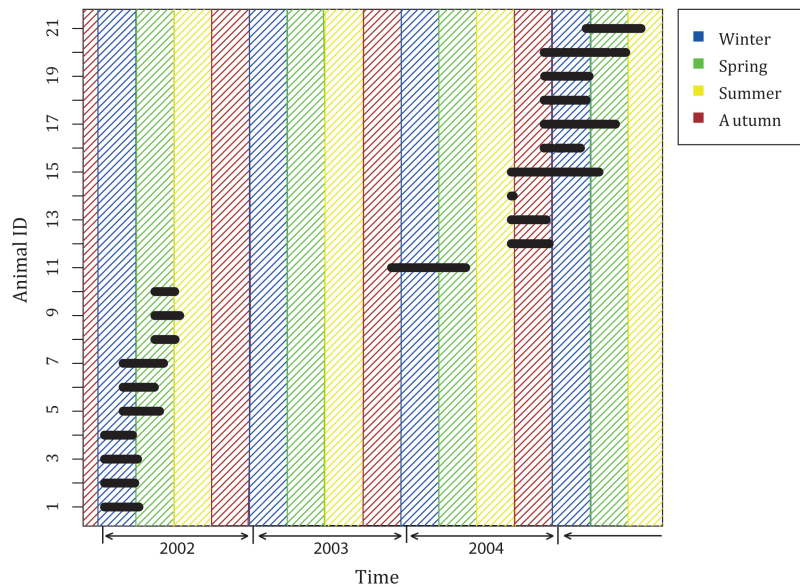

Figure 4.1: Individuals' time of follow-up according to seasons. In the graph, each black line represents the length of the period in which a specific individual was followed and the colour bars represent the different seasons of the year.

Posteriorly, 74 location were discarded: 57 of them due to extreme speed values and 17 of them due to locational filtering. As expected, most of the location (68%) belonged to LCB and A and only 1 (1.33%) single location belonged to LC3.

Following the descriptive analysis, in order to graphically see the tracks of the individuals, a visual inspection of animals' distribution was performed. During this inspection, it was observed that two animals, 9 and 13, went further south, close to the dutch coast (Figure 4.2). The fact that it is known that the Danish Wadden Sea population of harbour seals does not travel that much south (J. Tougaard, *personal communication*, June 2013), led to believe that these two animals belonged to the Dutch Wadden Sea population of seals. Consequently, these two animals were removed from the final database in order to maintain the analysis to a single population. Besides, two more individuals were also removed from the final analysis, animal 14 and 16. These two animals presented too few positions (22 and 42, respectively), which increased the probability of having inaccurate and unreliable tracks. After removing these 4

animals, a total of 488 positions were discarded.
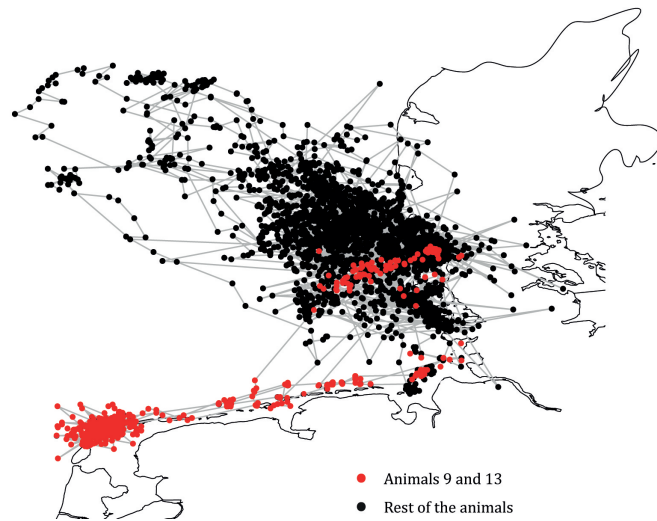
Figure 4.2: General distribution of the 21 individuals tagged in the Danish Wadden Sea. Red dots represent the locations corresponding to animals 9 and 13, which seem to belong to the seal population living in the Dutch Wadden Sea and, thus, removed from the final analysis.

After this filtering, the database included 9 (52.90%) males, with 5 adults (55.60%), 2 subadults (22.20%) and 2 pups (22.20%); and 8 (47.10%) females, with 1 adult (12.50%), 3 subadults (37.50%) and 4 pups (50.00%). In this case, the range of neither length nor weight of the individuals varied after filtering, thus, length ranged from 92 to 180 cm, and weight went between 19 to more than 110 kg (Table 4.1).

With regard to the spatio-temporal information, a total of 3692 positions were studied. As seen in Table 4.2, all the animals presented most of their locations in LC0, A and B, the least accurate categories, while only 506 (13.77%) of the positions belonged to LC3, 2 and 1, the most accurate ones. In this case, the spatio-temporal information showed that the individuals were followed-up a median of 92 days, where the minimum and maximum follow-up times were 48 and 213 days, respectively. The median number of positions transmitted per day was 2.17, ranging between 0.62 and 3.71 (Table 4.1). There were only 3 animals (17, 19 and 20) with daily update rates lower than 1: 0.99, 0.71 and 0.62, respectively. Finally, the number of positions per season was 1759 (47.60%) in Winter, 1070 (29.00%) in Spring, 112 (3.00%) in Summer and 751 (20.3%) in Autumn.

In Table 4.3, it can be observed a descriptive analysis of the main covariables related to positional information. At first glance, two values might be point out: time and speed between two consecutive positions. Particularly, when a closer look is taken at the position related to the maximum time lag between locations and the ones before and after, it seemed probable that this animal was hauling out, since the distance between locations and tagging position (which is a haul out spot) was not long. It is also important to note that, after applying a filter

Table 4.2: Number of Location Categories for each individual. Between parenthesis, percentage of the total within each animal can be observed. For all the animals the percentage of locations with LCA and LCB is significantly larger in compare to the rest of LC.

ID	LC3	LC2	LC1	LC0	LCA	LCB
1	4 (2.22%)	7 (3.89%)	17 (9.44%)	25 (13.89%)	39 (21.67%)	87 (48.33%)
2	6 (2.56%)	8 (3.42%)	22 (9.40%)	17 (7.26%)	66 (28.21%)	114 (48.72%)
3	2 (0.92%)	5 (2.29%)	31 (14.22%)	16 (7.34%)	53 (24.31%)	110 (50.46%)
4	4 (2.35%)	5 (2.94%)	17 (10.00%)	11 (6.47%)	39 (22.94%)	93 (54.71%)
5	3 (1.36%)	6 (2.73%)	8 (3.64%)	21 (9.55%)	46 (20.91%)	135 (61.36%)
6	8 (2.91%)	17 (6.18%)	16 (5.82%)	17 (6.18%)	68 (24.73%)	148 (53.82%)
7	3 (2.78%)	5 (4.63%)	10 (9.26%)	9 (8.33%)	25 (23.15%)	55 (50.93%)
8	1 (0.94%)	1 (0.94%)	7 (6.60%)	12 (11.32%)	29 (27.36%)	55 (51.89%)
10	6 (3.37%)	16 (8.99%)	12 (6.74%)	5 (2.81%)	46 (25.84%)	92 (51.69%)
11	7 (2.72%)	14 (5.45%)	40 (15.56%)	30 (11.67%)	56 (21.79%)	109 (42.41%)
12	12 (3.79%)	21 (6.62%)	30 (9.46%)	22 (6.94%)	91 (28.71%)	140 (44.16%)
15	6 (0.81%)	27 (3.64%)	41 (5.53%)	33 (4.45%)	186 (25.10%)	447 (60.32%)
17	1 (0.58%)	2 (1.17%)	6 (3.51%)	3 (1.75%)	45 (26.32%)	113 (60.08%)
18	3 (2.01%)	7 (4.70%)	7 (4.70%)	10 (6.71%)	35 (23.49%)	86 (57.72%)
19	0 (0.00%)	3 (3.90%)	5 (6.49%)	1 (1.30%)	18 (23.38%)	49 (63.64%)
20	5 (4.10%)	2 (1.64%)	4 (3.28%)	5 (4.10%)	30 (24.59%)	75 (61.48%)
21	3 (1.78%)	6 (3.55%)	7 (4.14%)	4 (2.37%)	40 (23.67%)	108 (63.91%)
TOTAL	74 (2.01%)	152 (4.13%)	280 (7.62%)	241 (6.56%)	912 (24.82%)	2016 (54.86%)

ID = Animal identification   LC= Location Category

of 20 km/h in our initial database, the maximum speed between locations in the final database was 25.50 km/h. This was due to the fact that the movement metrics were recalculated after some of the points were removed. Nonetheless, this maximum speed seemed reasonable enough for not applying any more filters. It can be relevant to point out as well that this descriptive analysis showed that faster speeds were usually related to positions belonging to low accuracy LC, LCA and B.

Table 4.3: Descriptive analysis of the main covariables of the database.

Variable	Median	Minimum	Maximum
Longitude	8.09°	4.12°	9.00°
Latitude	55.28°	53.72°	57.52°
Distance from tagging location	60.56 km	63.64 m	321.15 km
Distance between successive locations	6.63 km	63.16 m	153.45 km
Time since last location	4.28 hours	0.05 hours	269.29 hours
Speed between successive locations	1.80 km/h	0.01 km/h	25.50 km/h

During this descriptive analysis, we also studied how many days in a row the individuals seemed to remain offshore, presumably in foraging trips. Results suggested that trips ranged between a few days and more than 20. An example of these presumed foraging trips can be visualized in Figure 4.3, where the complete follow-up of animal 6 was plotted accounting for distance from tagging place and speed. In this figure, it can be observed approximately 4 trips in which,

taking the first one as example, the animal left the tagging place, went for a trip of more than 7 days and then came back to the initial haul out bank. In the graph, it can be seen as well the estimated speed of the individual. It shows faster speeds when the animal was leaving and going to hauling out banks and slower when the individual was there and at further distance from it. These results could suggest that animal was foraging in those areas. The same graphic for all the individuals, only including the distance from tagging place, are shown in the Figure D.1 of the Appendix D.
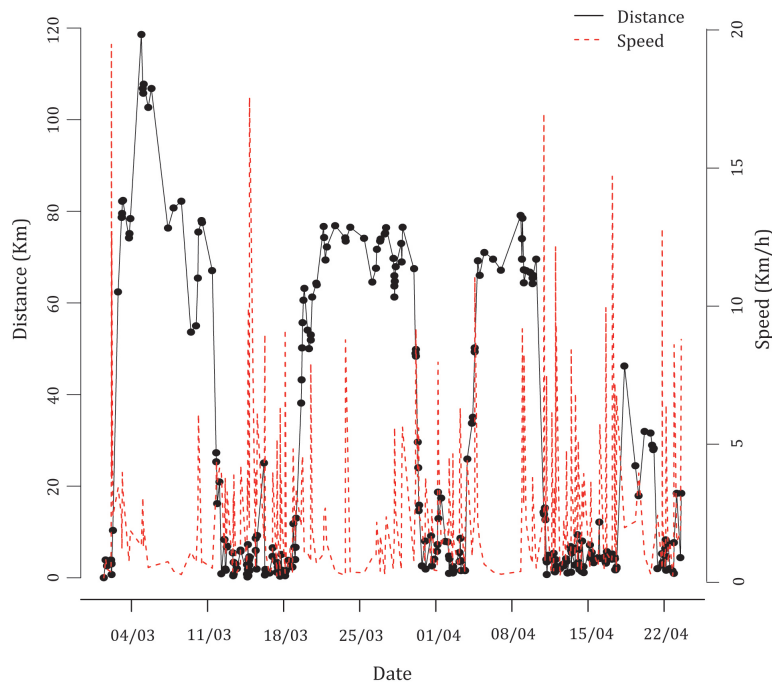
Figure 4.3: Plot of 76 days of follow-up for animal 6 showing distance from tagging place (in km, black dots and lines) and Speed (in km/h, red dashed lines) at which the individual was travelling at each point, accounting for time. It can be seen approximately 4 trips in which the animal left the initial haul out bank and went for trips that last a few days.

After the descriptive analysis, before performing the main analysis of this project, a preliminary visual and boxplot analysis of the database was performed in order to better understand the characteristics of the data and try to find differential patterns in this population's distribution.

First of all, the graphical inspection of the spatio-temporal data showed a considerable variation in the spatial distribution of the 17 tagged animals, specifically over the winter period. In particular, when accounting for the age of the individuals, it was observed that the animals that moved further away during the winter period were only some pups of the study (Figure 4.4). However, no differences seemed to be seen when comparing the distribution of the different age categories during the rest of season periods, which may indicate that the observed variation in the present data is due to the combination of the age and the season. Besides, no differences were observed when accounting for the gender of individuals.
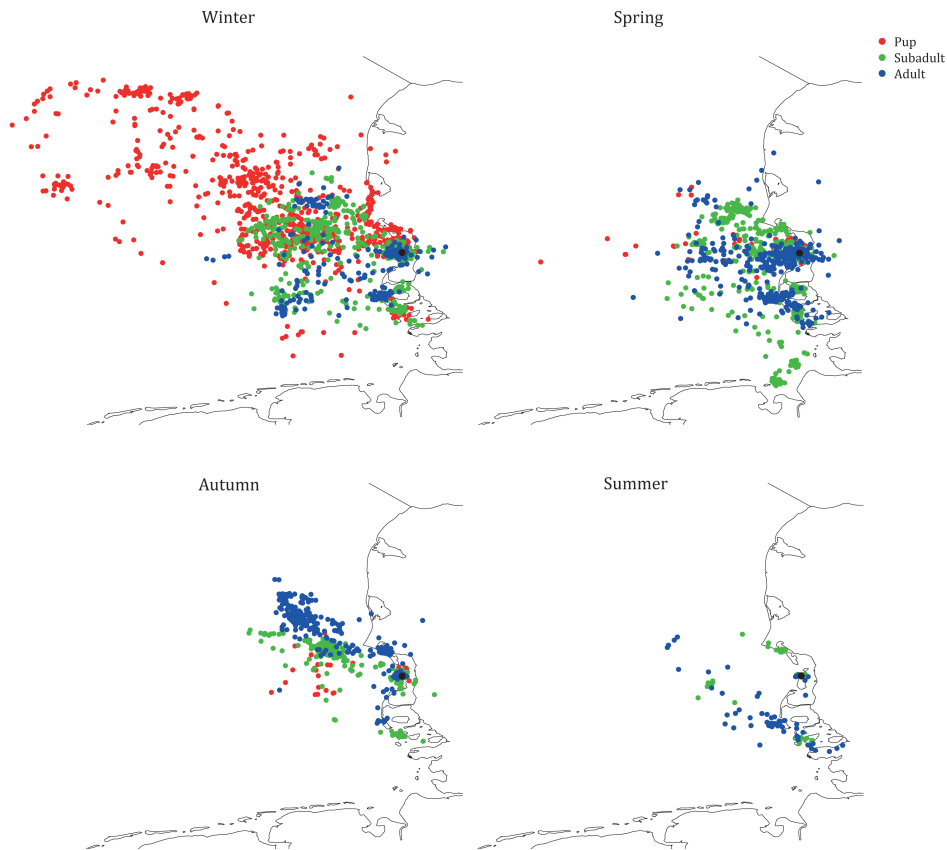
Figure 4.4: Spatial distribution in the Danish Wadden Sea of the final 17 seals studied in this project depending on age (red dots for pups, green dots for subadults and blue dots for adults) and season (each map corresponds to one season). In all seasons the three age classes were represented, except for Summer when no pups were followed-up.

Similar results were shown when using boxplot representation. In this case, latitude and longitude along with distance from tagging place were related to the main individuals' characteristics. Firstly, as it can be seen in Figure 4.5, when analysing the relation between latitude and longitude to the age of the individual, results showed that pups were present at higher latitudes and western longitudes and subadults at lower latitudes when comparing to the rest of the animals. At the same time, when comparing both latitude and longitude between genders, no differences seemed to be observed either. Secondly, when analysing the influence of gender, age or length on the distance travel from tagging place, the results showed some differences between age and length, but there were not differences between genders. As it can be seen in Figure 4.6, on one hand, pups were again the individuals going further away and, on the other hand, longer animals do not seem to travel longer distance, on the contrary, smaller/median animals were the individuals going further away in terms of distance, particularly, animals with a length between 100 and 130 cm.
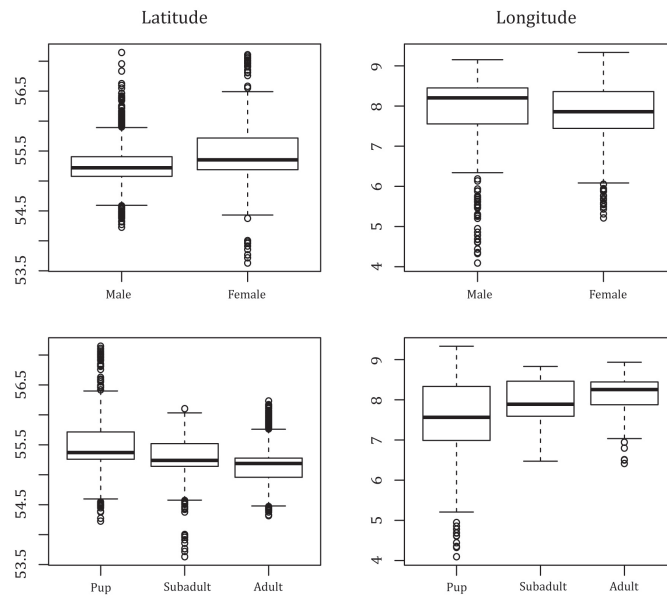
Figure 4.5: Latitude and Longitude represented accounting for age and gender of the individuals.
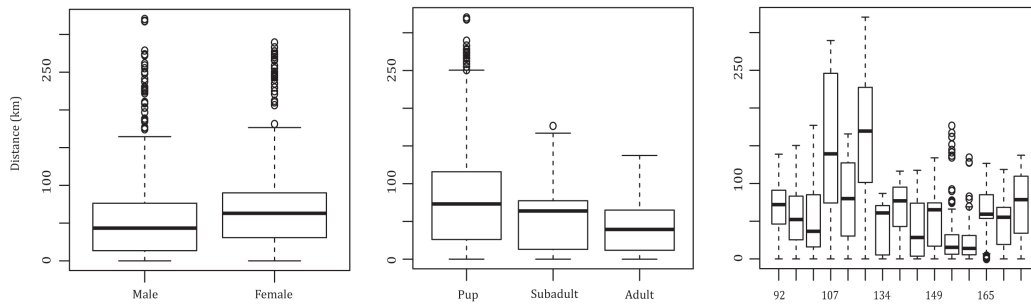
Figure 4.6: Distance from tagging place regarding gender, age, and length of the individuals.

4.2 State-Space Models Analysis

Following the descriptive and exploratory analysis, we proceeded with the main analysis of this project.

4.2.1 Filtering Stage

Finally, after removing 4 animals, a total of 17 seal tracks were analysed in this stage. The total number of evaluated positions was 3692, where the median number of locations per seal was 178 ranging from 77 to 741. Based on these observed locations collected with ARGOS system, the application of the SSM of this stage allowed us to obtain estimated "true" locations of these 17 seals.

As a global result, the model applied in this stage seemed to deal with extreme locations of all the animals, correcting most of the erroneous locations and smoothing the final trajectories of the seals. As an example, it can be seen in Figure 4.7 the observed and estimated trajectories of animal 11 using daily locations. The figure shows how the application of the SSM proposed by Jonsen et al. (2005) made the routes of the seals become smoother and more realistic and presented an improvement on the accuracy of the telemetry data. We ended up having a total of 1822 estimated positions, corresponding to one positions per day for animals. In Figure D.2 of the Appendix D, graphical results of the first stage for all the animals can be found.
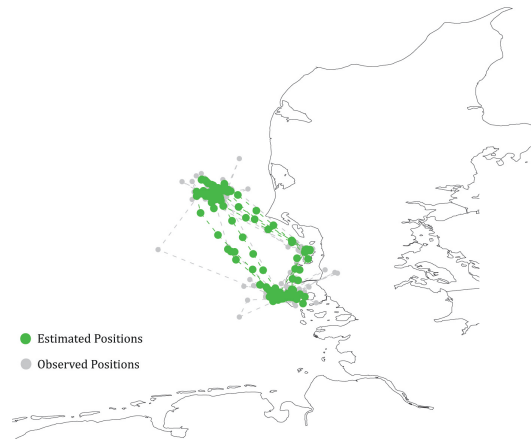
Figure 4.7: Observed (gray dots) and estimated (green dots) positions of animal 11.

4.2.2 Behaviour discrimination Stage

The number of evaluated positions was 1822, where the median number of locations per seal was 90 ranging from 48 to 213. The behavioural states of these positions were estimated using multiple variations of the frameworks proposed by Morales et al. (2004) and Eckert et al. (2008), as described in Section 3.3.2.

Previously to the deeper analysis of the results, a preliminary visual inspection was done. In this case, it was seen that most of the models made a good distinction between the different behavioural states, except for some animals in which all or most of the locations are represented as the same behavioural state, or the distribution of the behaviours did not reflect a defined pattern. Below, it is provided a closer description of all the models according to its number of behavioural states.

Models with one behavioural state

As refereed previously, it was not much meaningful to assume a single RW modelling the entire movement path, but it was useful as a null model to compare with the more complex ones.

When fitting the **"Single W"** or **"Single N"** model into the data, the results were very similar. The distribution of all step lengths were zero-modal and fat-tailed, with median values ranging from 3.68 to 39.07 km/day for "Single W", and from 7.45 to 28.77 km/day for "Single N". Besides, the distribution of the turning angles for all the animals was highly variable, as indicated for the low value of mean cosine of the turning angles (parameter ρ) raging from 0.04 to 0.21 in both models. As a remark, the distribution tends to one point concentrated in the direction of the parameter μ as ρ goes to 1.

Models with two behavioural states

When fitting models with two behavioural states, it was assumed that two RW modeled the entire movement path. In this case, it was expected to distinguish between two behaviours: a faster, less variable behaviour and a slower, more variable one. A priori, it would be expected that the faster behaviour would be related to travelling. This is not that clear for the slower behaviour, as at least both foraging and hauling out behaviour could be assumed. However, as a convenience, when referring to two behavioural states we classified them into travelling and foraging.

Results of "Double" models were quite similar in general. The percentage of time spent foraging and travelling was approximately constant, the travel rates in both behaviours were higher when travelling than foraging and mean cosine values of the turning angles were normally higher during foraging state, showing more variability in the turning angles of the animals.

When fitting the **"Double W"** model into the data, in which the step length was modeled by a Weibull distribution and there were not constraints on changing from one movement state to another, results placed seals in the foraging state a median percentage of approximately 54% of the time, ranging from 38% to 59% and in the travelling state in the rest of them. Expected daily movement rates (step length) ranged from 0.86 to 20.27 km/day during the foraging state and from 6.53 to 49.86 km/day during the travelling state. In this case, the mean cosine of the turning angles ranged from 0.04 to 0.59 when animals were foraging and from 0.12 to 0.69 when travelling.

When fitting the **"Double N"** model into the data, which was the same as "Double W" but using a Truncated Normal distribution for modelling step length, results were broadly similar to those of the "Double W" model. In these models, seals were placed in the foraging state a median percentage of approximately 56% of the time, ranging from 47% to 67%. Expected daily movement rates (step length) ranged from 2.28 to 13.98 km/day in the foraging state and from 12.18 to 42.69 km/day in the travelling state. The mean cosines of the turning angles were very similar in both states, ranging from 0.04 to 0.63 when foraging and from 0.08 to 0.63 when travelling.

When fitting the **"Double-Switch W"** model into the data, which was the same as "Double W" but defining a fixed probability of changing from one movement state to another, results

placed seals in the foraging state a median percentage of approximately 62% of the time, ranging from 35% to 89%. Expected daily movement rates (step length) ranged from 1.17 to 17.54 km/day in the foraging state and from 6.22 to 101.51 km/day in the travelling state. The mean cosines of the turning angles were very alike ranging from 0.04 to 0.35 in foraging state and from 0.12 to 0.84 in travelling state. The analysis of the daily switching probabilities showed broad variance between individuals, the median daily switching probabilities was 0.22, ranging from 0.14 to 0.44 from foraging to travelling states; and 0.31, ranging from 0.08 to 0.40 from travelling to foraging states.

When fitting the **"Double-Switch N"** model into the data, the same as "Double-Switch W" but using a Truncated Normal distribution for modelling step length, results were broadly similar to previous models. Results placed seals in the foraging state a median percentage of approximately 52% of the time, ranging from 8% to 76%. Expected daily movement rates (step length) ranged from 1.98 to 11.41 km/day in the foraging state and from 12.86 to 40.49 km/day in the travelling state. In this case, the mean cosines of the turning angles were similar between behaviours, ranging from 0.06 to 0.38 during foraging state and from 0.07 to 0.39 during travelling state. The median daily switching probabilities in this case was 0.25 ranging from 0.11 to 0.42 from foraging to travelling and 0.21 ranging from 0.11 to 0.47 from travelling to foraging.

Models with three behavioural states

When fitting models with three behavioural states, it was assumed that three RWs generated the entire movement path. In this case, it was expected to distinguish between a faster, less variable behaviour and two slower, more variable ones. Essentially, these behaviours are assumed to be travelling, and foraging and hauling out, respectively.

Results of "Triple" models were similar among them, except for model "Triple W" that in a few animals presented some convergence limitations and unreliable results. In general, results tended to leave the parameters for the travelling state quite similar compared to those of "Double" models. However, results tended to divide the previous foraging state into two further states: an almost stationary state in which movement rates were very low and a low movement state.

When fitting the **"Triple W"** model into the data, in which the step length was modeled by a Weibull distribution and there were not constraints on changing from one movement state to another, results tended to show the presence of a faster state in comparison to two slower ones. However, as stated before, this model presented some unreliable results, which provided unrealistic travel rates for some of the animals. The percentage of time spent in the different states was highly variable and meaningless.

When fitting the **"Triple N"** model into the data, which was the same as "Triple W" but using a Truncated Normal distribution for modelling step length, results showed clearly the

tendency to divide the foraging state into two further states: an almost stationary state in which movement rates were very low (1.89 - 9.63 km/day) and a low movement state (4.37 - 27.29 km/day). However, it left the parameters for the travelling state almost unchanged compared to those of "Double" models (16.16 - 44.75 km/day). The proportion of time spent in the travelling state was close to "Double N" and "Double-Switch N" (39%), but the proportions of time spent in the almost stationary and low movement states were variable between individuals, ranging between 15.56% and 70.79%, and 0% and 49.23%, respectively.

When fitting the **"Triple-Switch W"** model into the data, in which three RWs are fitted with switching parameters, results showed similar ranges for the travelling state compared to "Double-Switch W" (10.18 - 117.70 km/day) and two slower states ranging from 0.79 to 3.25 km/day and from 3.47 to 38.18 km/day. The proportion of time spent in the travelling state was similar to previous models, 38%, but the proportions of time spent in the almost stationary and low movement states were variable between individuals, ranging between 4% and 53%, and 15% and 74%, respectively.

When fitting the **"Triple-Switch N"** model into the data, the same as the "Triple-Switch W" but using a Normal distribution for modelling step length, results looked very similar to those observed in the "Triple N". Results for the travelling state remained in similar values, ranging from 15.72 to 48.29 km/day and the previous foraging state was divided into two further states, each of them ranging from 1.68 to 9.78 km/day and from 4.71 to 28.63 km/day. The proportion of time spent in the travelling state was almost identical to the previous model, 39%, but the proportions of time spent in the almost stationary and low movement states were variable between individuals, ranging between 0% and 63%, and 17% and 61%, respectively.

4.2.3 Model Selection - Validation

As previously stated in Section 3.3.4, there is not a precise criterion for checking and selecting SSMs. Therefore, following some of the main guidance, it was decided to use a visual inspection and parameters estimates in order to make a first selection of the most feasible models and, posteriorly, check-validate the selected group through autocorrelation function (acf).

Biological Selection-Validation

The visual inspection showed that models that made more biological and ecological sense were models that introduce two behaviours states for describing the entire path (Figure 4.8).When analysing the results of models that included a third behavioural state, it can be observed that behavioural states of successive locations did not seem to show a realistic pattern. It seemed that the third behaviour was complicated to model and distinguish from the others, especially from the other state with shorter and more variable step length. This was reflected in Figure 4.9, where animal 1 was used as example and behavioral state assignation results for models with two and three behavioural states were compared. In this figure, it is observed that

time periods at which the fastest behaviour ("travelling") was assigned usually remain in the same positions when comparing all the models. However, when analysing the dispositions of the other two states, it can be noticed that colours representing the slower behaviours are not disposed following any defined pattern.
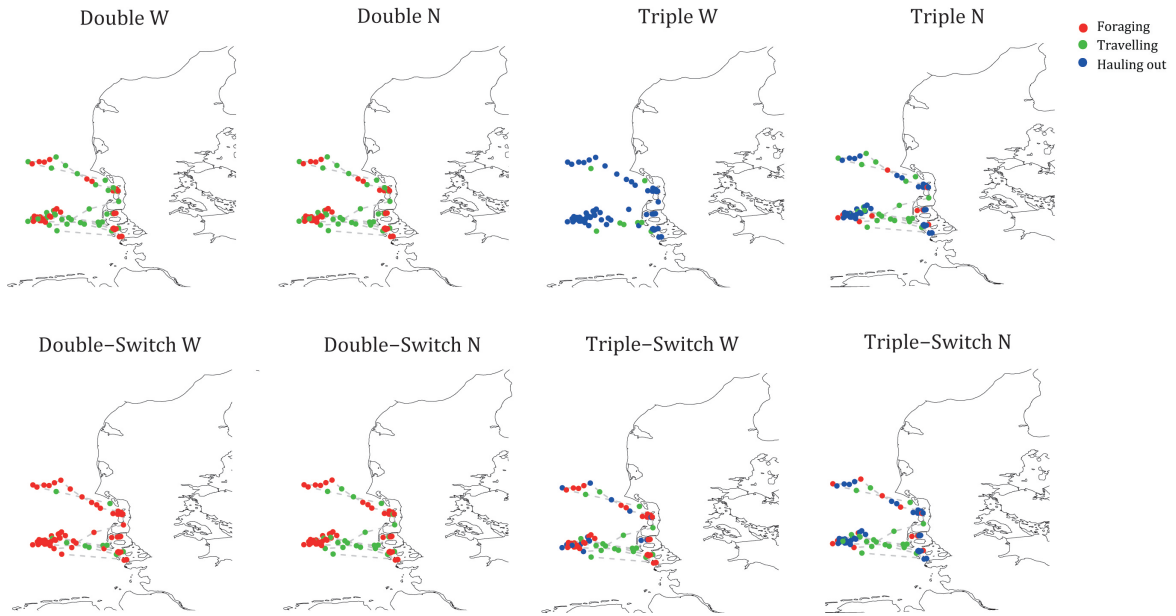
Figure 4.8: Distribution of animal 1 with colours representing behavioural states using results from models defining the movement path by two and three behavioural states. Red, green and blue dots represent foraging, travelling and hauling out behaviours, respectively.
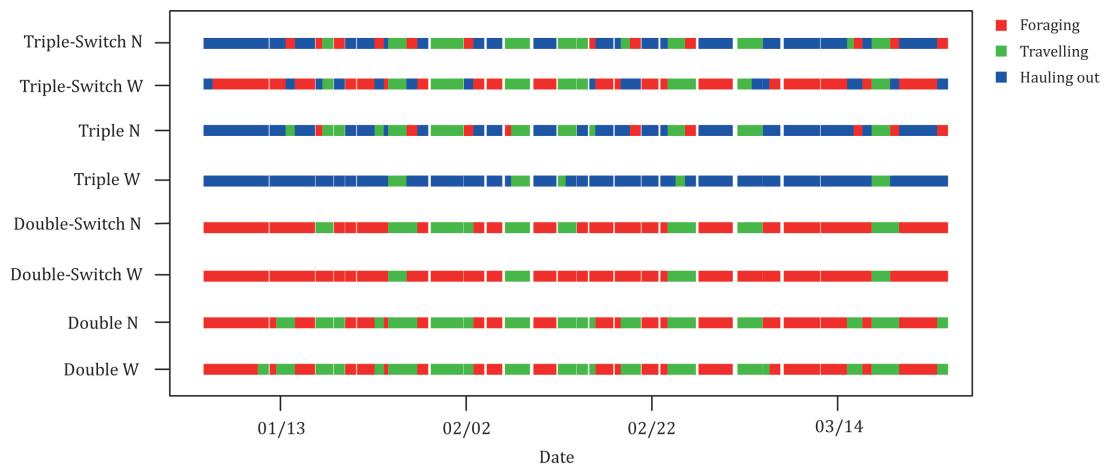
Figure 4.9: Activity bar showing the assignment of behavioural states through time for all the models fitted to animal 1. Red, green and blue bars correspond to foraging, travelling and hauling out state, respectively.

Besides, when visually comparing models that fitted two behavioural states, taking into account the feasibility of the results, it seemed that the "Double-Switch W" gave the least realistic results. For some animals, this model did not adjust well and some of the locations were likely to be wrong classified. For instance, using the example shown in Figure 4.8, it can be seen that "Double-Switch W" model is classifying some locations as "foraging", when it is likely that these positions were "travelling" positions. Among the other three "Double" models, it was quite complex to determine which one was the best one, as the visual results were indeed similar.

When distinguishing between behavioural states and looking at the parameters estimates results of the different models in terms of daily movement rates, turning angles and proportion of time spent in each behaviour, conclusions were similar to visual analysis results. Again, models fitting three behaviours showed difficulties distinguishing between the two slower behavioural states. Both the proportion of time spent and the movement rate in the faster ("travelling") state remained similar to the models with two behavioural states, showing constancy in the results. However, assignment results did not appear feasible when referring to the slower states. Results were not uniform when comparing the different "Triple" models and, in some animals, models were not able to distinguish between the two slower states. As previously, models that explain the movement path using two RWs were the ones showing a better fit. When looking at the results of the different "Double" model, expected daily movement results seemed realistic and robust. However, it was complicated to assess which one of the "Double" models showed a better fit just referring to the parameters estimates.

Consequently, the four models that included two behavioural states for representing the complexity of the movement path were kept for being checked-validated using the posterior predictive distribution. Additionally, as models that introduce a single behavioural state could not be assess by visual inspection, they were also kept for being checked-validated.

Posterior Predictive Assessment

Posteriorly, from the selected range of models, the selection of the best fitting model was done using the posterior predictive distribution. After the simulation and calculation of acf's of daily distance moved, comparing the autocorrelation structure in the model output and the data provided a further means by which model fit to the temporal structure of observed data could be judged. In Figure 4.10, acf's from observed data are compared with the 95% credibility intervals predicted using the posterior predictive distribution for the six models selected applied to 3 sample individuals: 2, 6 and 11.

On one hand, results showed that models with one behavioural state were not good representing the structure of the observed data, as the 95% credibility intervals for the acf's of the modeled paths were too wide and they did not adjust to the data. Similarly, "Double W" model that modeled the step length using a Weibull distribution showed similar results. Again, the 95% credibility interval of acf's did not adjust to the data (Figure 4.10) and, in some cases, such as animal 11, the model did not fit the data, leaving the observed data outside the 95% credibility

interval. On the other hand, results from models using a Truncated Normal distribution for representing step length of the model were better than the rest. As it can be seen in Figure 4.10, the 95% credibility interval adjust better the data, especially when the simulation were done using the "Double-Switch N" model. In the case of "Double N", the 95% credibility interval left outside a large number of observed data, meaning that in this case, the model did not represent well the complexity of the data. When comparing all the animals, results showed the same, particularly, "Double-Switch N" acf graphics for all the animals can be found in Figure D.3 in the Appendix D. Consequently, the "Double-Switch N" was the model chosen as the most reliable for representing the complexity of the movements and, therefore, for evaluating the characteristics of the foraging areas.
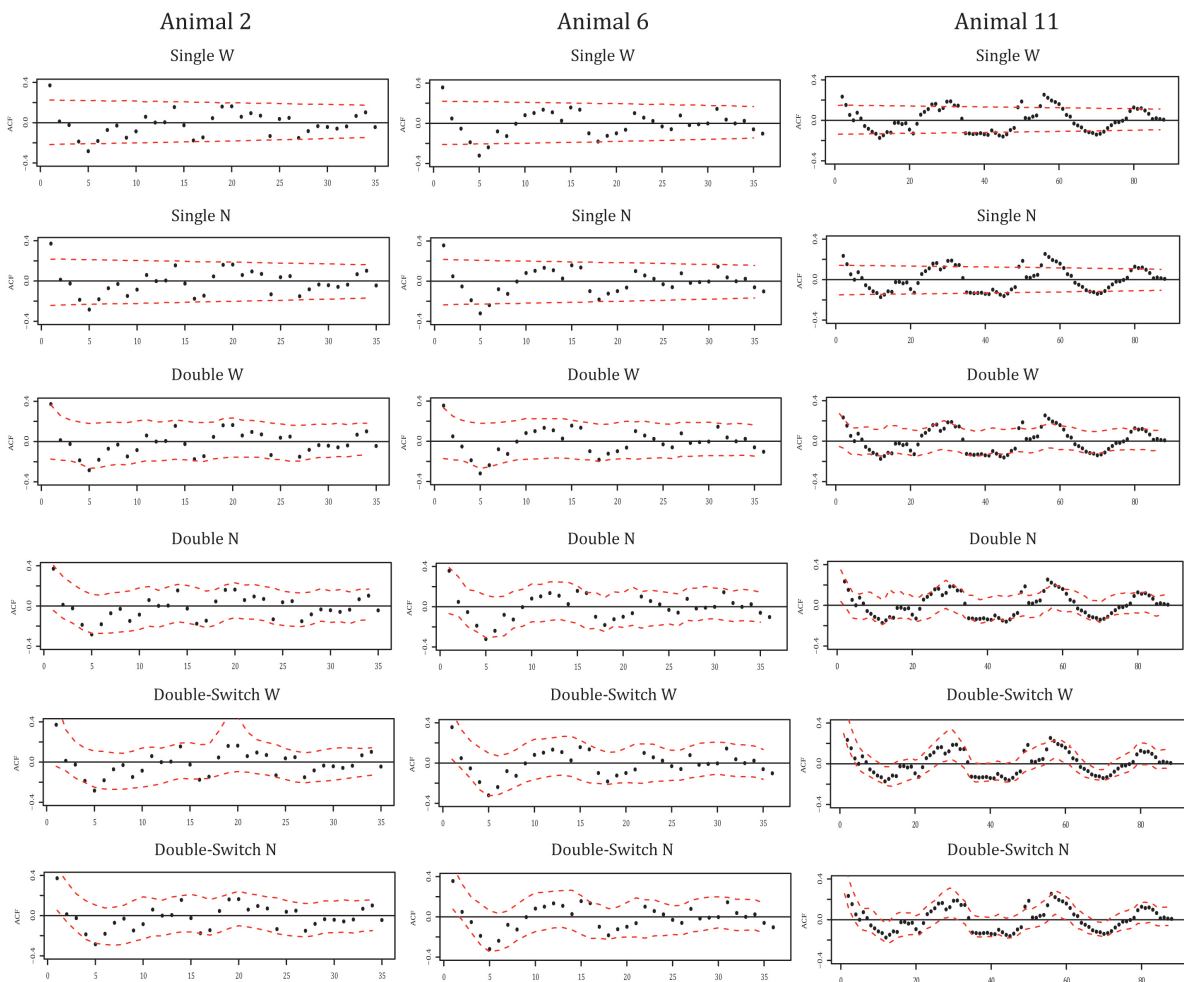
Figure 4.10: Autocorrelation functions (acf's) of daily movement rate for observed and modeled paths for lags 1-n/2, n = number of locations, for animals 2, 6 and 11. The graphics correspond to models: "Single W", "Single N", "Double W", "Double N", "Double-Switch W", "Double-Switch N". Black dots correspond to observed acf's and red dashed lines correspond to the 95% credibility interval for the acf's modeled with the different models for 30000 replicates.

4.3 Biological Inference

Once the "Double-Switch N" model was selected as the most suitable one for representing the behavioural states of the animals, all positions were classified into two behaviours: the faster and less variable, "travelling"; and the slower and more variable, "foraging". At first, a characterization of the behavioural states was performed in terms of the parameters estimates, step length and turning angles, of all the animals (Table 4.4).

Table 4.4: Expectation of the Posterior Distribution for Step length and the scale parameter of the Wrapped Cauchy distribution ρ of each individual using the resulting behavioural classification of the "Double-Switch N" model.

	Behavioural State 1		Behavioural State 2	
Animal ID	$E[r(y)]$	$E[\rho(y)]$	$E[r(y)]$	$E[\rho(y)]$
1	6.42	0.14	25.00	0.15
2	6.84	0.12	39.43	0.32
3	8.45	0.18	32.65	0.12
4	9.89	0.17	40.49	0.25
5	4.62	0.14	21.14	0.17
6	3.08	0.12	18.74	0.12
7	7.33	0.15	23.21	0.39
8	3.75	0.18	28.21	0.14
10	11.41	0.21	28.24	0.17
11	2.79	0.26	17.90	0.18
12	1.98	0.14	19.52	0.14
15	4.69	0.06	24.96	0.07
17	8.22	0.09	31.31	0.35
18	6.88	0.29	25.59	0.16
19	8.88	0.35	20.02	0.31
20	3.63	0.37	12.86	0.12
21	6.50	0.38	16.97	0.25

Moreover, Figure 4.11 illustrates fitted distributions for step length and turning angles for animal 11 in the two behavioural states. The figure showed a clear difference in terms of movement rate, as the "foraging" behaviour showed a slower and more variable movement rate than the "travelling" behaviour. The same happened to the turning angles results, the foraging behaviour appeared more variable, due to animals turning randomly during this state, than the travelling behaviour, where animals likely kept a defined direction when they were travelling to certain area.

Afterwards, in order to observe the spatial disposition of both behaviours, the distribution of all the animals was plotted together accounting for the two different behavioural states (Figure 4.12).
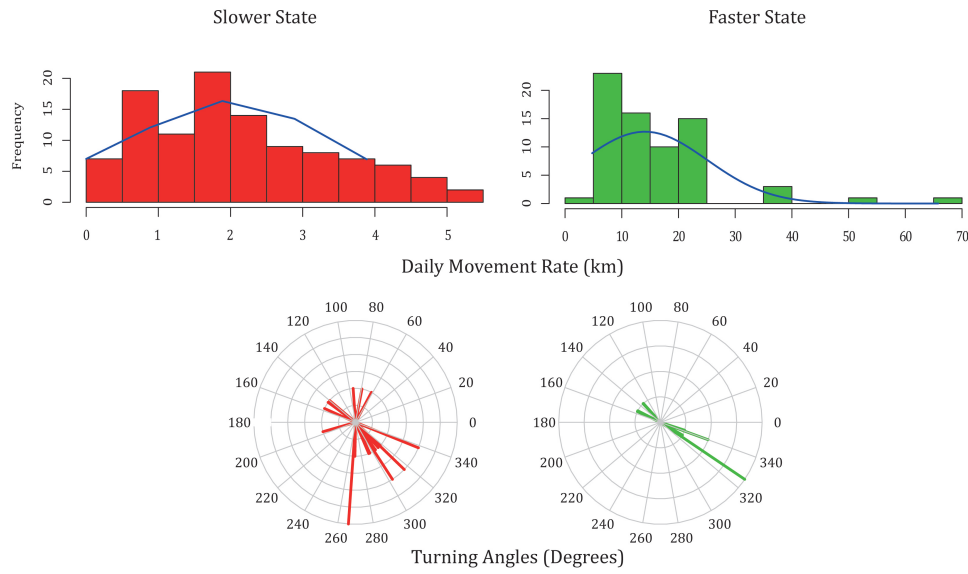
Figure 4.11: Step length and turning angles distribution for animal 11 in the two behavioural states as inferred using the "Double-Switch N" model. Step length have Truncated Normal distributions with parameters μ_I and σ_I corresponding to the mean of the posterior distributions. Turning angles have Wrapped Cauchy distributions with parameters θ_I and ρ_I, corresponding to the mean of the posterior distributions. Bars are observed frequencies of both movements (in km/day) for step length and angles (in degrees) for turning angles.


Figure 4.12: Behavioural distribution for all the animals. Red and green dots represent slower and faster state, respectively.

As it can be seen in the Figure 4.12, it seems that positions related to both behaviours are placed allover the area. However, when the individual tracks are analysed independently, a tendency or association among locations seems to be perceived. In Figure 4.13, the behavioural state classification using the "Double-Switch N" model can be observed for the same animals as the validation (2, 6 and 11). Results for the rest of the animals can be found in Figure D.4 of the Appendix D.
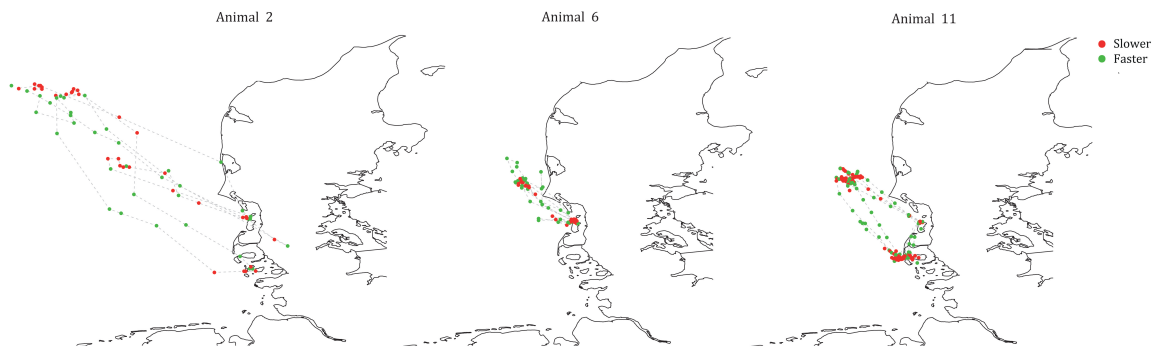
Figure 4.13: Distribution of animal 2, 6 and 11 with colours representing behavioural states using results from "Double-Switch N" model. Red and green dots represent slower and faster behaviours, respectively.

Particularly, the association between positions can be even more visible when relating the behaviour of seals with the time and the distances to the tagging place. An example of this can be observed in Figure 4.14, where animals' 1 and 2 position distance from tagging place were plotted referring to time, and behavioural state of each position was represented by different colours. In these graphs, it can be clearly seen that the chosen model distinguished and classified appropriately each position taking into account the characteristics of the model. In this case, the two RW and selected movement metrics, step length and turning angles, seem to capture the main essence of the movement path. It seems likely that the faster state was capturing the travelling behaviour but, looking at the distribution of the slower behavioural state, as previously stated, it seems very likely that the slower mode was fitting more than one behaviour, probably, foraging and hauling out.
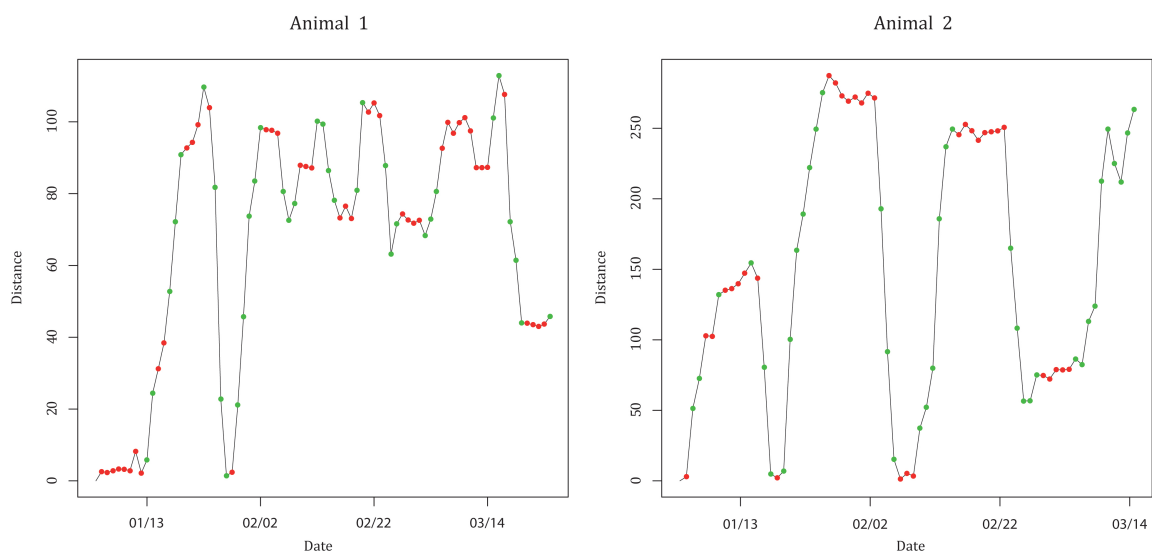

Figure 4.14: Distance (in km) from tagging point respecting to time for animal 1 and 2. Each position was coloured using results from the "Double-Switch N" model.

Posteriorly, as during the preliminary analysis results showed that both season and age of the individuals were variables related to the location of the animals, after assigning the behavioural state to each position, it was evaluated whether the spatial disposition of both behaviours was related to any of these covariables. Season and age results can be examined in Figure 4.15 and Figure 4.16, respectively, where it can be seen that, as in the preliminary results, pups during winter time were the individuals traveled further distances, in this case known to be foraging areas. Additionally, it could be suggested that subadults' foraging areas were also localized mostly at further distances from land than adults, although in both cases, pups and subadults can be also found close to land. Furthermore, when looking at these figures, it can also be visualized that positions related to the slower behaviour were more grouped than the ones related to the faster behaviour. These grouped positions were present allover the area, showing that when an animal gets to a specific area remains at the surroundings for a while, probably, foraging or hauling out, depending whether the position is at sea or on land, respectively.

If we distinguish between hauling out and foraging, by whether the individuals were on land or at sea, results showed that individuals had more than one haul out area and share these different banks with other individuals. At the same time, results did not suggest that there were defined foraging locations or that seals share foraging areas. In the graphs, substantial variations between animals were observed, each seal seemed to adopt its own foraging strategy, in terms of number of foraging areas and recurrence. Furthermore, it was not possible to localize delimited foraging areas and did not appear that more than one animal traveled to the same foraging spot.
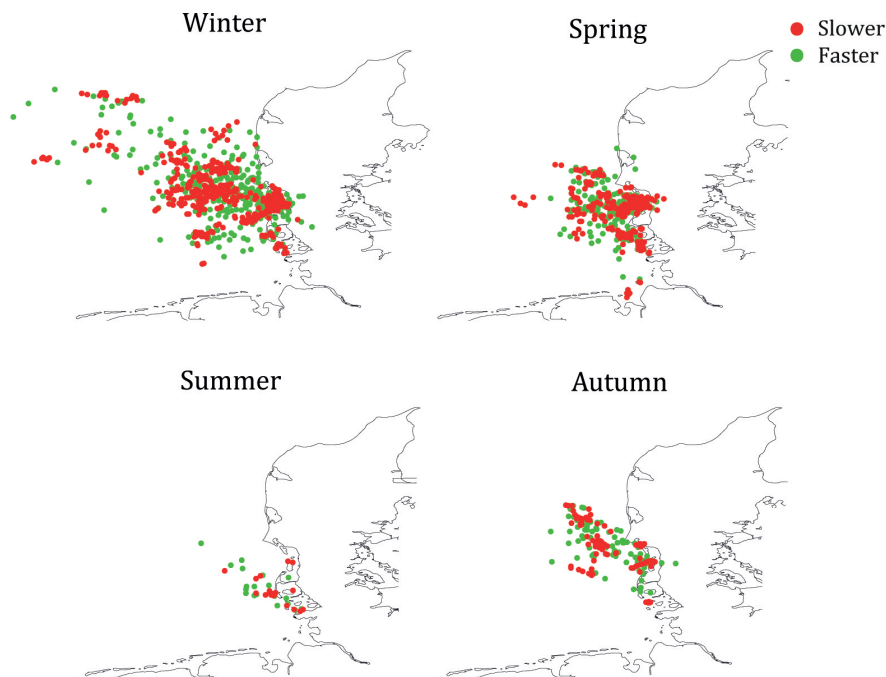
Figure 4.15: Distribution of positions related to slower (red dots) and faster (green dots) behaviour separated by season.
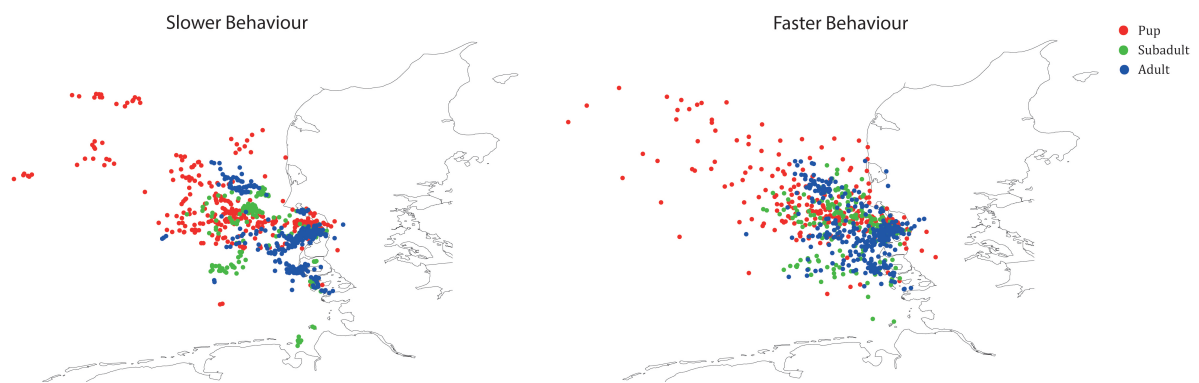
Figure 4.16: Distribution of positions accounting for the inferred behavioural state and age of the individuals. Red, green and blue dots represent pups, subadults and adults foraging positions, respectively.

Chapter 5

Discussion and Conclusion

In this project, I have presented and applied a methodology for determining the distribution and foraging behaviour of a harbour seal population, from which movement information was collected using ARGOS system between 2002 and 2004. This analysis has led to interesting hypothesis regarding the distribution and foraging behaviour of this population, that could be used for subsequent conservation applications. An improved understanding of the population's localization and habitat use will allow for a more informed conservation effort, better identification of critical habitats, such as overlap with fisheries and wind farms, and to test for effects of anthropogenic activities, such as seismic surveys and pile driving.

During this master's thesis I have highly increased my knowledge about movement data and animal tracking systems. I have learned about its inherent complexity and how it needs to be carefully treated. Moreover, I have presented a selected range of the wide variability of tracking methods and explained some of their limitations, always focusing on the ARGOS system, the tracking system used for this project. As explained during the State of the Art, GPS and ARGOS system are the most used tracking methods. Even though much higher accuracy can now be obtained using GPS technology, the number of existing satellite telemetry data sets collected via ARGOS system is large, and continuously increasing. Finding a methodology to obtain accurate positions from ARGOS data is therefore an important research area. In addition, ARGOS data only provide with information about geolocation, thus, it is still necessary to develop a methodology capable of inferring the behaviour of the animal using the available information.

Until now, different methods have been used in order to cope with limitations of both animal movement and inherent errors in ARGOS data. Recently, State-Space models (SSMs) have been presented as a method that provides a framework capable of integrating the treatment of both limitations in a same approach. In this project, we have developed a two-stages method that is able to determine from movement metrics the distribution of the animals and, more specifically, the foraging behaviour of Danish Wadden Sea population of harbour seals. To do so, we have used two SSMs in which, firstly, we handle ARGOS inaccuracy positioning and, secondly, we assign behavioural states to each location based on movement metrics.

For the development and application of this methodology, it has been necessary to learn the concept of SSMs from the beginning and, hence, I have carried out a significant bibliographic research about these models and studied them deeply in order to understand their meaning and implications.

After the application of the two-stages SSM, results showed that models seemed to properly fit the data. During the first step, extreme or inaccurate locations were corrected, resulting smoother and more realistic movement paths. During the second step, the SSM allowed us to assign a specific behaviour to each position, based on the step length and turning angle of each location. From all the different model variations applied in this project, the selected model was the one fitting two different Random Walk models for generating the entire movement path, modelling step length with a Truncated Normal distribution, and explicitly modelling the probability of switching from one behavioural state to another. Results of this stage were very satisfactory and the final model provided with very reliable results that allowed us to distinguish 2 different states: one slower and more variable in terms of turning angles and another one faster and less variable.

From a biological perspective, results showed that seals were likely to perform more than two behaviours during the entire movement path, probably a minimum of three. However, the potential for the model to determine the number of behavioural states was limited by the capability of the chosen movement metrics for capturing the complexity of the states and differently parameterising three behavioural states. In the results, the faster and less variable, in terms of turning angles, behaviour was clearly detected and classified as "travelling" behaviour. However, the slower and more variable behaviour seemed to contain at least two states, "foraging" and "hauling out" behaviours. Consequently, even though our results showed that the best fitting model was the "Double-Switch N", we concluded that it was probable that at least a third behavioural state was present in this seal population. The addition of more movement metrics or external movement variables would potentially facilitate the detection and proper assignment of a third behavioural state. For instance, in this case, it would be advisable in order to distinguish "foraging" from "hauling out" to add a variable related to the characteristic of being on land, such as information from a dryness sensor (ARGOS PTTs have this sensor, but the information was not available), information about distance to land or, even better, an incorporated Time-Depth Recorder (TDR) capable of taking pressure measurements and thus providing the depth at each location.

Focusing on the biological and behavioural results of the study, it was observed from the beginning of the study that pups were the individuals travelling the longest distances, reaching distances further than 250 km from the haul out bank. Posteriorly, it was proven that pups travelled these long distances for foraging and that it was likely that a similar behaviour was taking place when referring to subadults. A possible explanation for these results could be that young and smaller animals compete with large dominant ones for best resources and foraging locations close to hauls out sites, having to travel longer distances in order to forage. This competition for the "best" foraging location was not found between individuals of dif-

ferent gender, which could be also related to the slight sexual dimorphism present in this species.

Furthermore, results suggested an association between positions classified as slower state, which could be interpreted, in the case of foraging positions, as that animals do not forage randomly. It can be hypothesized that seals look for suitable foraging areas and, once there, remain in the area a few days after which they travel to either another foraging area or hauling out bank. From the final visual analysis, it was also observed that seals in the Danish Wadden Sea have more than one haul out area and share these different banks between individuals. On the contrary, results suggested that there were no defined or shared foraging spots for all the animals. It did not appear that more than a single individual travelled to the same spot to forage, but rather, that individuals left the haul out bank, started a search and stopped whenever they found a good area for foraging.

Lastly, further academic objectives were accomplished during this project. During the completion of this project, I have learned how to use LaTeX as a writing editor and I am able now to use it in any scientific work. I have also improved my knowledge on programming skills using software R and WinBUGS. For fitting the different SSM frameworks, it was necessary to understand and modify the different models. SSMs provided by Jonsen et al., Morales et al. and Eckert et al. were not directly applicable to our data and objectives and it was therefore necessary to develop suitable modification of these frameworks in order to be able to perform the desirable analysis. In addition, I gained skills in developing plots and figures using the software R jointly with Adobe Illustrator, one of the most used softwares for image creation and edition. I could therefore conclude that this master's thesis has not only helped me to better understand how to analyse and interpret the distribution and foraging behaviour of the Wadden Sea harbour seals but has also provided me with useful and important skills for my future as a scientist.

5.1 Future research

The completion of this master thesis does not signify the termination of the study. As referred before, an important improvement for this kind of models that could be done in a near future would be to add other movement metrics or external covariables in order to find a way for distinguishing between hauling out and foraging behaviours. Among the possible covariables with relevant information, the model could be supplemented with information from a pressure-dryness sensor that could provide data on dive depths and surface times, additionally, the distance from coast of each position or the bathymetry of the area of study would also be important covariables. Moreover, biological and environmental characteristics could also be added to the models, for instance, the age of the individual or the season, as it has been proven that they are related to foraging locations.

For future research, it would also be relevant to consider switching probabilities depending on landscape features, such as bathymetry, presence of food or anthropogenic presence. For

instance, the presence or distance to wind farms, an anthropogenic factor that is increasing every day in Danish waters, can be introduced into the model in order to check the effect of this anthropogenic object on the behaviour of the animals. Models can show a preference for foraging far or close from turbines, helping directly to conservation efforts. Preliminary results of a further study that I am carrying out applying these SSMs in another Danish harbour seals' population seem to denote that seals do forage closer to turbines, but the reasons of this apparent preference for the wind farm area are not clear. Two possible causes are discussed: an increased food availability inside the wind farm (reef effect) and/or the absence of vessels in an otherwise heavily trafficked part of the North Sea (sheltering effect).

Furthermore, in this project, we have only fitted models to single paths, but the models could be extended to include a population level by adding hyper-prior distributions. A hierarchical approach would permit the assessment of the degree of an individual variability in movement behaviour (Morales et al., 2004) and improve the estimation of the parameters (Silva et al., 2013). For instance, it would be assumed in a population study that animals move according to a RW but that each individual would be characterized by different diffusivity, and that these diffusivities would be random variables drawn from unknown distributions (Jonsen et al., 2013). Hierarchical SSMs (HSSMs) have the advantage of combining information from all tracking data to estimate parameters at both the individual and population levels. Consequently, these models enable parameter estimation even for shorter or incomplete tracks, by using information from other tracks. However, their use can lead to both convergence difficulties and longer computational time.

In terms of posterior biological inference, one of the biggest limitations of animal studies is the common scarcity of individuals sampled. In this project, we finally analysed 17 individuals, but the number was even more restricting when trying to determining age and gender differences. For future projects, it would be important to increase the number of seal tracks collected in order to be able to perform more complex biological inference. A larger sample would give robustness to the results and allow for more tests. In following studies, it would be interesting to analyse differences in travel rates between individuals of different age. For instance, it would be relevant to assess whether the speed of pups is higher when travelling to foraging areas or whether they spend more days offshore in order to get to further foraging spots. In the same direction, the existence of difference in travel rate or days offshore between different season it could be analysed.

Bibliography

Argos (**2013**). *User's Manual.* CLS/Service Argos, Toulouse, France.

Bjorge, A., Bekkby, T., and Bryant, E. B. (**2002**). "Summer home range and habitat selection of harbor seal (*Phoca vitulina*) pups." *Marine Mammal Science*, 18: 438–454.

Boeuf, B. J., Crocker, D. E., Costa, D. P., Blackwell, S. B., Webb, P. M., and Houser, D. S. (**2000**). "Foraging ecology of northern elephant seals." *Ecological Monographs*, 70(3): 353–382.

Bonner, W. N. (**1989**). *Natural History of Seals.* London.

Breed, G. A., Bowen, W. D., McMillan, J. I., and Leonard, M. L. (**2006**). "Sexual segregation of seasonal foraging habitats in a non-migratory marine mammal." *Proceedings of the Royal Society B*, 273(1599): 2319–2326.

Brooks, S. P. and Gelman, A. (**1998**). "General Methods for Monitoring Convergence of Iterative Simulations." *Journal of Computational and Graphical Statistics*, 7: 434–455.

Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G., and Butler, P. J. (**2004**). "Biotelemetry: a mechanistic approach to ecology." *Trends in Ecology & Evolution*, 19: 334–343.

Costa, D. P., Robinson, P. W., Arnould, J. P., Harrison, A. L., Simmons, S. E., Hassrick, J. L., Hoskins, A. J., Kirkman, S. P., Oosthuizen, H., Villegas-Amtmann, S., and Crocker, D. E. (**2009**). "Accuracy of ARGOS locations of Pinnipeds at-sea estimated using Fastloc GPS." *PLoS One*, 5(1): e8677.

Dietz, R., Teilmann, J., Henriksen, O. D., and Lai, K. (**2003**). "Movements of seals from Rødsand seal sanctuary monitored by satellite telemetry. Relative importance of the Nysted Offshore Wind Farm area to the seals." Technical report, National Environmental Research Institute Technical Report.

Douglas, D. C., Weinzierl, R., C. Davidson, S., Kays, R., Wikelski, M., Bohrer, G., and Giuggioli, L. (**2012**). "Moderating Argos location errors in animal tracking data." *Methods in Ecology and Evolution*, 3(6): 999–1007.

Eckert, S. A., Moore, J. E., Dunn, D. C., Sagaminaga Van Buiten, R., Eckert, K. L., and Halpin, P. N. (**2008**). "Modelling loggerhead turtle movement in the Mediterranean: Importance of body size and oceanography." *Ecological Applications*, 18(2): 290–308.

Freeman, A. N. D., Nicholls, D. G., Wilson, K. J., and Bart, J. A. (**1997**). "Radio- and Satellite-Tracking Westland Petrels *Procellaria Westlandica*." *Marine Ornithology*, 25: 31–36.

Härkönen, T. and Heide-Jorgensen, M. P. (**1990**). "Comparative Life Histories of East Atlantic and Other Harbor Seal Populations." *Ophelia*, 32: 211–235.

Härkönen, T. and Hårding, K. (**2001**). "Spatial structure of harbour seal populations and the implications thereof." *Canadian Journal of Zoology*, 79: 2115–2127.

Härkönen, T., Hårding, K., and Lunneryd, S. G. (**1999**). "Age- and sex-specific behaviour in harbour seals *Phoca vitulina* leads to biased estimates of vital population parameters." *Journal of Applied Ecology*, 36: 825–841.

Härkönen, T. J. (**1987**). "Influence of Feeding on Haul-Out Patterns and Sizes of Sub-Populations in Harbour Seals." *Netherlands Journal of Sea Research*, 21(4): 331–339.

Jonsen, I. D., Basson, M., Bestley, S., Bravington, M. V., Patterson, T. A., Pedersen, M. W., Thomson, R., Thygesen, U. H., and Wotherspoon, S. J. (**2013**). "State-space models for bio-loggers: A methodological road map." *Deep Sea Research Part II: Topical Studies in Oceanography*, 88-89: 34–46.

Jonsen, I. D., Mills Flemming, J., and Myers, R. A. (**2005**). "Robust state-space modeling of animal movement data." *Ecology*, 86(11): 2874–2880.

Jonsen, I. D., Myers, R. A., and Mills Flemming, J. (**2003**). "Meta-Analysis of animal movement using state-space models." *Ecology*, 84(11): 3055–3063.

Kalman, R. (**1960**). "A new approach to the lineal filtering and prediction problems." *Journal of Basic Engineering*, 82: 35–45.

Lesage, V., Hammill, M. O., and Kovacs, K. M. (**1999**). "Functional classification of harbor seal (*Phoca vitulina*) dives using depth profiles, swimming velocity, and an index of foraging success." *Canadian Journal of Zoology*, 77: 74–87.

Madden, K. M., Fuiman, L. A., William, T. M., and Davis, R. W. (**2008**). "Identification of foraging dives in free-ranging Weddell seals *Leptonychotes weddellii*: confirmation using video records." *Marine Ecology-Progress Series*, 365: 263–275.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (**2004**). "Extracting more out of relocation data: Building movement models as mixtures of random walks." *Ecology*, 85(9): 2436–2445.

Naito, Y., Bornemann, H., Takahashi, A., McIntyre, T., and Plötz, J. (**2010**). "Fine-scale feeding behavior of Weddell seals revealed by a mandible accelerometer." *Polar Science*, 4: 309–316.

Nørgaard, N. (**1996**). *Haul-out behaviour, movments, foraging strategies and population estimates of harbour seals* Phoca vitulina *in the Dansih Wadden Sea*. Ph.D. thesis, University of Aarhus.

Olsen, M. T. (**2006**). *Genetic Analysis of Harbour Seal (*Phoca Vitulina*) Population Structure and Dispersal Patterns in Danish and Western Swedish Waters.* Ph.D. thesis, University of Copenhagen, National Environmental Research Institute.

Patterson, T. A., McConnel, B. J., Fedak, M. A., Bravington, M. V., and Hindell, M. (**2010**). "Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error." *Ecology*, 91(1): 273–285.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (**2008**). "State-space models of individual animal movement." *Trends Ecology Evolution*, 23(2): 87–94.

Plummer, M., Best, N., Cowles, K., and Vines, K. (**2006**). "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News*, 6(1): 7–11. URL http://CRAN.R-project.org/doc/Rnews/.

Ries, E. H., Traut, I. M., Paffen, P., and Goedhart, P. W. (**1997**). "Diving patterns of harbour seals (*Phoca vitulina*) in the Wadden Sea, the Netherlands and Germany, as indicated by VHF telemetry." *Canadian Journal of Zoology*, 75: 2063–2068.

Royer, F. and Lutcavage, M. (**2008**). "Filtering and interpreting location errors in satellite telemetry of marine animals." *Journal of Experimental Marine Biology and Ecology*, 359(1): 1–10.

Silva, M. A., Prieto, R., Jonsen, I. D., Baumgartner, M. F., and Santos, R. S. (**2013**). "North Atlantic Blue and Fin Whales Suspend Their Spring Migration to Forage in Middle Latitudes: Building up Energy Reserves for the Journey?" *PLoS One*, 8(10): e76 507.

Tanizaki, H. (**2001**). "Estimation of unknown parameters in non linear and non gaussian state space models." *Journal of Statistical Planning and Inference*, 96: 301–323.

Thompson, D., Moss, S. E. W., and Lovell, P. (**2003**). "Foraging behaviour of South American fur seals *Arctocephalus australis*: extracting fine scale foraging behaviour from satellite tracks." *Marine Ecology Progress Series*, 260: 285–296.

Tougaard, J., Ebbesen, I., Tougaard, S., Jensen, T., and Teilmann, J. (**2003**). "Satellite tracking of Harbour Seals on Horns Reef: Use of the Horns Reef wind farm area and the North Sea." Technical report, Fisheries and Maritime Museum, Esbjerg.

Tougaard, J., Teilmann, J., and Tougaard, S. (**2008**). "Harbour seal spatial distribution estimated from Argos satellite telemetry: overcoming positioning errors." *Endangered Species Research*, 4: 113–122.

Tougaard, J., Tougaard, S., Jensen, R. C., Jensen, T., Teilmann, J., Adelung, D., Liebsch, N., and Müller, G. (**2006**). "Harbour seals at Horns Reef before, during and after construction of Horns Rev Offshore Wind Farm." Technical report, Final report to Vattenfall A/S.

Tremblay, Y., Robinson, P. W., and Costa, D. P. (**2009**). "A parsimonious approach to modeling animal movement data." *PLoSONE*, 4: e4711.

Turchin, P. (**1991**). "Translating foraging movements in heterogeneous environments into the spatial distribution of foragers." *Ecology*, 72: 1253–1266.

Turchin, P. (**1998**). *Quantitative Analysis of Movement: measuring and modeling population redistribution in plants and animals*. Sinauer Associates, Sunderland, MA.

Van Parijs, S. M., Thompson, P. M., Tollit, D. J., and Mackay, A. (**1997**). "Distribution and activity of male harbour seals during the mating season." *Animal Behaviour*, 54: 35–43.

Vincent, C., McConnel, B. J., Ridoux, V., and Fedak, M. A. (**2002**). "Assessment of ARGOS Location accurancy from satellite tags deployed on captive gray seals." *Marine Mammal Science*, 18(1): 156–166.

White, N. A. and Sjöberg, M. (**2002**). "Accuracy of satellite positions from free-ranging grey seals using ARGOS." *Polar Biology*, 25: 629–631.

Williams, T. M. and Kooyman, G. L. (**1985**). "Swimming Performance and Hydrodynamic Characteristics of Harbor Seals Phoca vitulina." *Physiological Zoology*, 58(5): 576–589.

Appendix A

R code for adapting data to WinBUGS: *dat4bugs*

As it has been explained in the methodology of this project, there is a R code to prepare ARGOS data for being fit into SSM frameworks. This code can be seen in the following text.

```
1  dat4bugs <- function(data, timestep = 1)
2  {
3  #
4  # Code to convert typical Argos data into proper format for WinBUGS
5  #
6  # Created by Ian Jonsen
7  #
8  #   created on: 01/04/2005
9  #   last modified on: 05/26/2005
10 #
11 #
12 #   Note data file must include the following named columns in the first 5 positions:
       date, time, lc, lat, lon
13 #
14 #     date must be a numeric object of class "dates times", this can be accomplished as
         follows:
15 #        data$date <- chron(as.character(data$date), format='d/m/y', out.format='m/d/y')
16 #        where format (eg, ='d/m/y') must be set to whatever format the date is in, and
       out.format is set to 'm/d/y'
17 #
18 #     time must be converted to "minutes after midnight" format, this can be
       accomplished as follows:
19 #        data$time <- as.numeric(times(as.character(data$time))) * 1440
20 #
21 #   timestep units are in days
22 #
23 #
24 #   try to convert date and time to correct format if not already so
25 #
26 if(is.factor(data$date) || is.character(data$date)) data$date <- chron(as.character(
       data$date), out.format='m/d/y')
27 else if(class(data$date)[1] != "dates") {stop('Error - date is not in correct format,
       see comments in dat4bugs')}
28
29 if(!is.numeric(data$time)) data$time <- as.numeric(times(as.character(data$time))) *
       1440
30 else if(max(data$time) > 1440 || min(data$time) < 0){stop('Error - time is not in
```

75

```
        correct format, see comments in dat4bugs')}
31
32 newdata <- data
33
34 time <- newdata$time
35 lon <- newdata$lon
36 lat <- newdata$lat
37 date <- newdata$date
38
39 # group observations into 'windows' of duration = timestep and create index (idx) for
       WinBUGS
40 start.date <- date[1]
41 end.date <- date[nrow(newdata)]
42 numday <- end.date - start.date
43 numsteps <- ceiling(numday / timestep)
44 steps <- seq(start.date, dates(as.numeric(start.date) + numsteps * timestep), by =
       timestep)
45 row.out <- sapply(2:numsteps, function(i){row.names(newdata)[date < steps[i] & date >=
       steps[i-1]]})
46 tmp <- sapply(1:length(row.out), function(i){if(length(row.out[[i]]) == 0) row.out[[i]]
        <- NA else row.out[[i]]})
47 idx <- cumsum(c(1, sapply(tmp, length)))
48 newdata <- newdata[unlist(tmp), ]
49 newdata <- data.frame(newdata, row.names = NULL)
50
51 # determine fraction of day at which each observation is made, j
52 j <- c()
53 for(i in 1:(length(idx) - 1)){
54   t <- newdata$time[idx[i]:(idx[i+1]-1)]
55   d <- as.numeric(newdata$date[idx[i]:(idx[i+1]-1)] - newdata$date[idx[i]])
56   j1 <- (t + d * 1440) / (timestep * 1440)
57   j <- c(j, j1)
58   }
59
60 # set up t-distribution parameters and create look-up values for the different location
        classess
61 sigma.lon <- c()
62 sigma.lat <- c()
63 nu.lon <- c()
64 nu.lat <- c()
65
66 # parameters for t-distributions (obtained by MLE); estimated from Vincent et al. 2002
       Argos data
67 #    these values are in Appendix A, Table A1
68 sigma.lon[1] <- 0.2898660
69 sigma.lon[2] <- 0.3119293
70 sigma.lon[3] <- 0.9020423
71 sigma.lon[4] <- 2.1625936
72 sigma.lon[5] <- 0.5072920
73 sigma.lon[6] <- 4.2050261
74
75 sigma.lat[1] <- 0.1220553
76 sigma.lat[2] <- 0.2605126
77 sigma.lat[3] <- 0.4603374
78 sigma.lat[4] <- 1.607056
79 sigma.lat[5] <- 0.5105468
80 sigma.lat[6] <- 3.041276
81
82 nu.lon[1] <- 3.070609
83 nu.lon[2] <- 1.220822
84 nu.lon[3] <- 2.298819
```

```
 85 nu.lon[4] <- 0.9136517
 86 nu.lon[5] <- 0.786954
 87 nu.lon[6] <- 1.079216
 88
 89 nu.lat[1] <- 2.075642
 90 nu.lat[2] <- 6.314726
 91 nu.lat[3] <- 3.896554
 92 nu.lat[4] <- 1.010729
 93 nu.lat[5] <- 1.057779
 94 nu.lat[6] <- 1.331283
 95
 96 # add following code to restrict nu to be >= 2; required for WinBUGS
 97 nu.lon[nu.lon < 2] <- 2
 98 nu.lat[nu.lat < 2] <- 2
 99
100 # convert sigma's from km back to degrees
101 sigma.lon <- (sigma.lon/6366.71 * 180)/pi
102 sigma.lat <- (sigma.lat/6366.71 * 180)/pi
103
104 # convert standard errors to precisions (se^-2); required for WinBUGS
105 itau2.lon <- sigma.lon[newdata$lc] ^ -2
106 itau2.lat <- sigma.lat[newdata$lc] ^ -2
107
108 # Note following code deals with timesteps that have no observations by interpolating a
        single observation at mid-day
109 # this could also be addressed by putting priors on the missing data for itau2, nu, and
        j.
110 #
111 # convert NA's in itau to very high standard errors
112 itau2.lon[is.na(itau2.lat)] <- max(itau2.lon, na.rm=T)
113 itau2.lat[is.na(itau2.lat)] <- max(itau2.lat, na.rm=T)
114
115 # converts NA's in nu to smallest nu allowed by Winbugs
116 nu.lon <- nu.lon[newdata$lc]
117 nu.lat <- nu.lat[newdata$lc]
118
119 nu.lon[is.na(nu.lon)] <- nu.lon[1]
120 nu.lat[is.na(nu.lat)] <- nu.lat[1]
121
122 # set missing observations to be at mid-day
123 j[is.na(j)] <- 0.5
124
125 return(list(y = cbind(newdata$lon,newdata$lat), itau2 = cbind(itau2.lon,itau2.lat), nu
      = cbind(nu.lon,nu.lat), idx = idx, j = j, RegN = length(idx) - 1))
126
127 }
```

dat4bugs.r

Appendix B

1st Stage Model: Jonsen's Framework

In order to perform the first part of the methodology, the following model was applied.

```
1   #   "DCRW" model from Ian D Jonsen, Joanna Mills Flemming and Ransom A Myers
2   #     Robust state-space modeling of animal movement data
3   #     jonsen@mathstat.dal.ca
4   #     Created by Ian Jonsen, 10/29/2004, last modified, 06/23/2005
5
6   model{
7
8     pi <- 3.141592653589
9     pi2 <- 2*pi
10    npi <- pi*-1
11
12    Omega[1,1] <- 1
13    Omega[1,2] <- 0
14    Omega[2,1] <- 0
15    Omega[2,2] <- 1
16
17    ## priors on process uncertainty
18    iSigma[1:2,1:2] ~ dwish(Omega[,],2)
19    Sigma[1:2,1:2] <- inverse(iSigma[,])
20
21    ## Priors for first location
22    for(k in 1:2){
23      x[1,k] ~ dt(y[1,k], itau2[1,k], nu[1,k])
24    }
25
26    ## Assume simple random walk to estimate 2nd regular position
27    x[2,1:2] ~ dmnorm(x[1,], iSigma[,])
28
29    theta ~ dunif(npi,pi) ## prior for theta (mean turn angle)
30    gamma ~ dbeta(1,1)  ## prior for gamma (persistence)
31    psi ~ dunif(0, 10)  ## scaling factor for estimation error
32
33    ## Transition equation
34    for(t in 2:(RegN-1)){
35      ## Build transition matrix for rotational component
36      T[t,1,1] <- cos(theta)
37      T[t,1,2] <- -sin(theta)
38      T[t,2,1] <- sin(theta)
```

```
39    T[t,2,2] <- cos(theta)
40
41    for(k in 1:2){
42      Tdx[t,k] <- T[t,k,1] * (x[t,1] - x[t-1,1]) + T[t,k,2] * (x[t,2] - x[t-1,2]) ##
            matrix multiplication
43      x.mn[t,k] <- x[t,k] + gamma * Tdx[t,k]  ## predict next location (no process
            error)
44    }
45
46    x[t+1,1:2] ~ dmnorm(x.mn[t,], iSigma[,])  ## predict next location (with process
          error)
47  }
48
49  ##  Measurement equation
50  for(t in 2:RegN){        ## loops over regular time intervals (t)
51    for(i in idx[t-1]:(idx[t]-1)){    ## loops over observed locations within interval
            t
52      for(k in 1:2){
53        itau2.psi[i,k] <- itau2[i,k] * psi
54        zhat[i,k] <- (1 - j[i]) * x[t-1,k] + j[i]*x[t,k]  ## interpolate irregularly
              observed locations
55        y[i,k] ~ dt(zhat[i,k], itau2.psi[i,k], nu[i,k]) ## robust measurement equation
56      }
57    }
58  }
59
60 }   #end model
61
62
63
64 # When running in WinBUGS gui mode:
65
66
67
68 #Paste Data here (eg. hsdata.txt)
69
70
71
72
73
74
75
76 #Paste Inits here (eg. hsDCRWinits.txt)
```

DCRW.r

Appendix C

2nd Stage Model: Morales' and Eckert's Frameworks

The following SSM is the "Double-Switch N", the model selected as the one that fits better the data.

```
## model "Double switch" used in:

model{

  ## priors

  b[1] ~ dnorm(0.0,0.01)I(0.0,)  ## mean parameter for slow movement
  b.diff ~ dnorm(0.0,0.01)I(0.0,)
  b[2] <- b[1]+b.diff       ## mean parameter for fast movement

  a[2] ~ dnorm(0.0,0.01) I(0.0,) ## variance parameter for fast movement
  a.diff ~ dnorm(0.0,0.01)I(0.0,)  ## a nonnegative variate
  a[1]  <- a[2] + a.diff   ## variance parameter for slow movement

  sigma[2] <- pow(1/a[2],0.5)
  sigma[1] <- pow(1/a[1],0.5)

  rho[1] ~ dunif(0,1)     ## mean cosine of turns for slow movement
  rho[2] ~ dunif(0,1)     ## mean cosine of turns for fast movement

  mu[1] ~ dunif(-3.14159265359, 3.14159265359) ## mean direction of turns for slow
       movement
  mu[2] ~ dunif(-3.14159265359, 3.14159265359) ## mean direction of turns for fast
       movement

  q[1] ~ dunif(0,1)     ## probability of being in state 1 at t given that individual
       was in state 1 at time t-1
  q[2] ~ dunif(0,1)     ## probability of being in state 1 at t given that individual
       was in state 2 at time t-1

  phi[1] ~ dunif(0,1)
  phi[2] <- 1 - phi[1]
  idx[1] ~ dcat(phi[])    ## asign state for first observation

  Pi <- 3.14159265359    ## define pi

```

```
34    for (t in 2:npts) {
35
36        nu[t,1] <- q[idx[t-1]]
37        nu[t,2] <- 1 - q[idx[t-1]]
38        idx[t] ~ dcat(nu[t,])   ##  idx is the latent variable and the parameter index
39
40        ## likelihood for steps
41        l[t] ~ dnorm(b[idx[t]], a[idx[t]])I(0.0,) # Normal distriution for step length
42
43        ## likelihood for turns.
44        ## use the "ones" trick (see WinBUGS manual) to sample from the Wrapped Cauchy
               distribution
45
46        ones[t] <- 1
47        ones[t] ~ dbern(wC[t])
48        ## below is the pdf for Wrapped Cauchy distribution, divided by 500 (arbitrary)
               to ensure that wC[t] will be less than one
49        wC[t] <- ( 1/(2*Pi)*(1-rho[idx[t]]*rho[idx[t]])/(1+rho[idx[t]]*rho[idx[t]]-2*rho[
               idx[t]]*cos(theta[t]-mu[idx[t]])) )/500
50
51    }
52 }
```
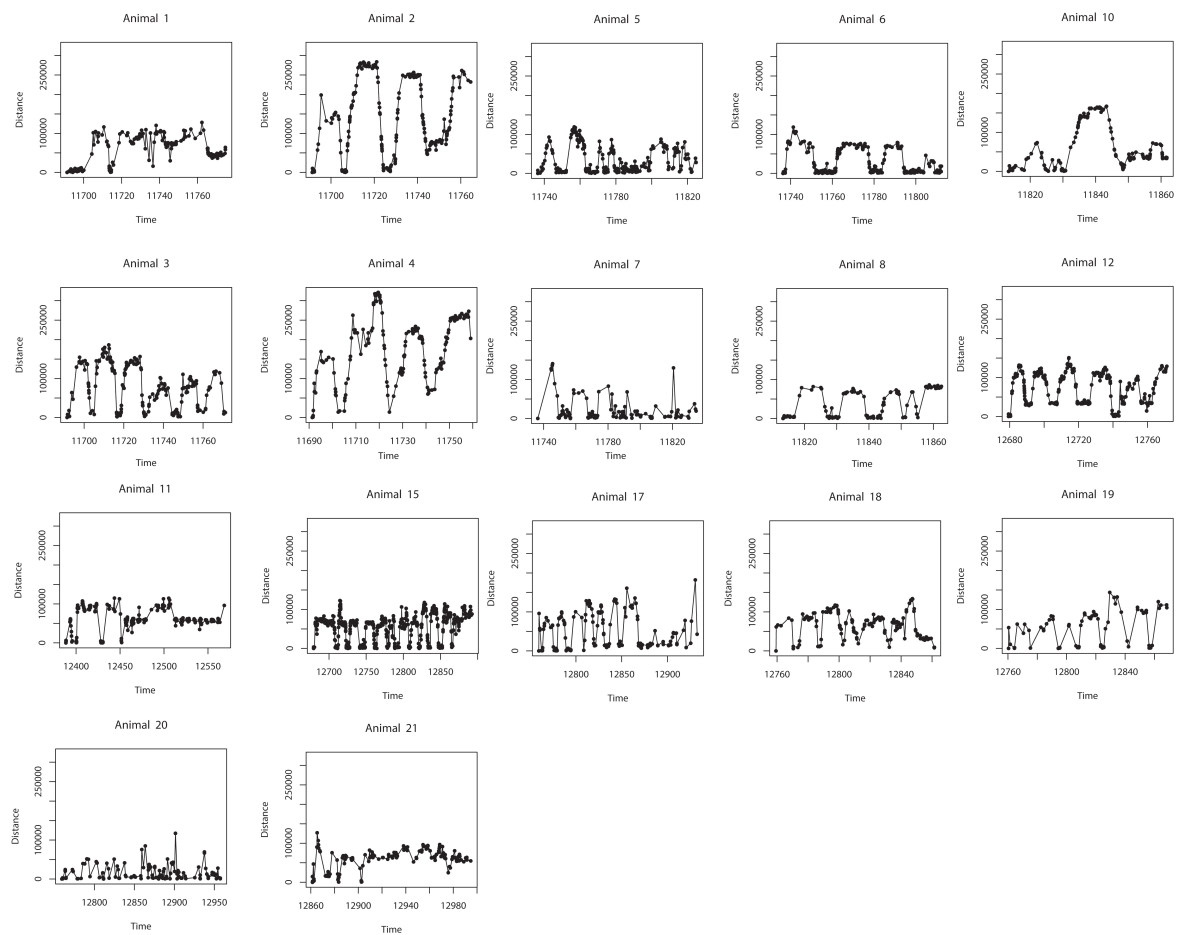
<div align="center">DoubleSNormal.r</div>

# Appendix D

Tables and Figures

Figure D.1: Distance from tagging place (in km, black dots and lines) regarding follow-up time for the final 17 animals. For all the animals, it can be seen the realization of trips in which the animals left the initial haul out bank and went for trips that last a few days.

Figure D.2: Results from applying the first stage of the project. Gray, red and green dots represent observed, removed during the filtering and estimated locations, respectively.
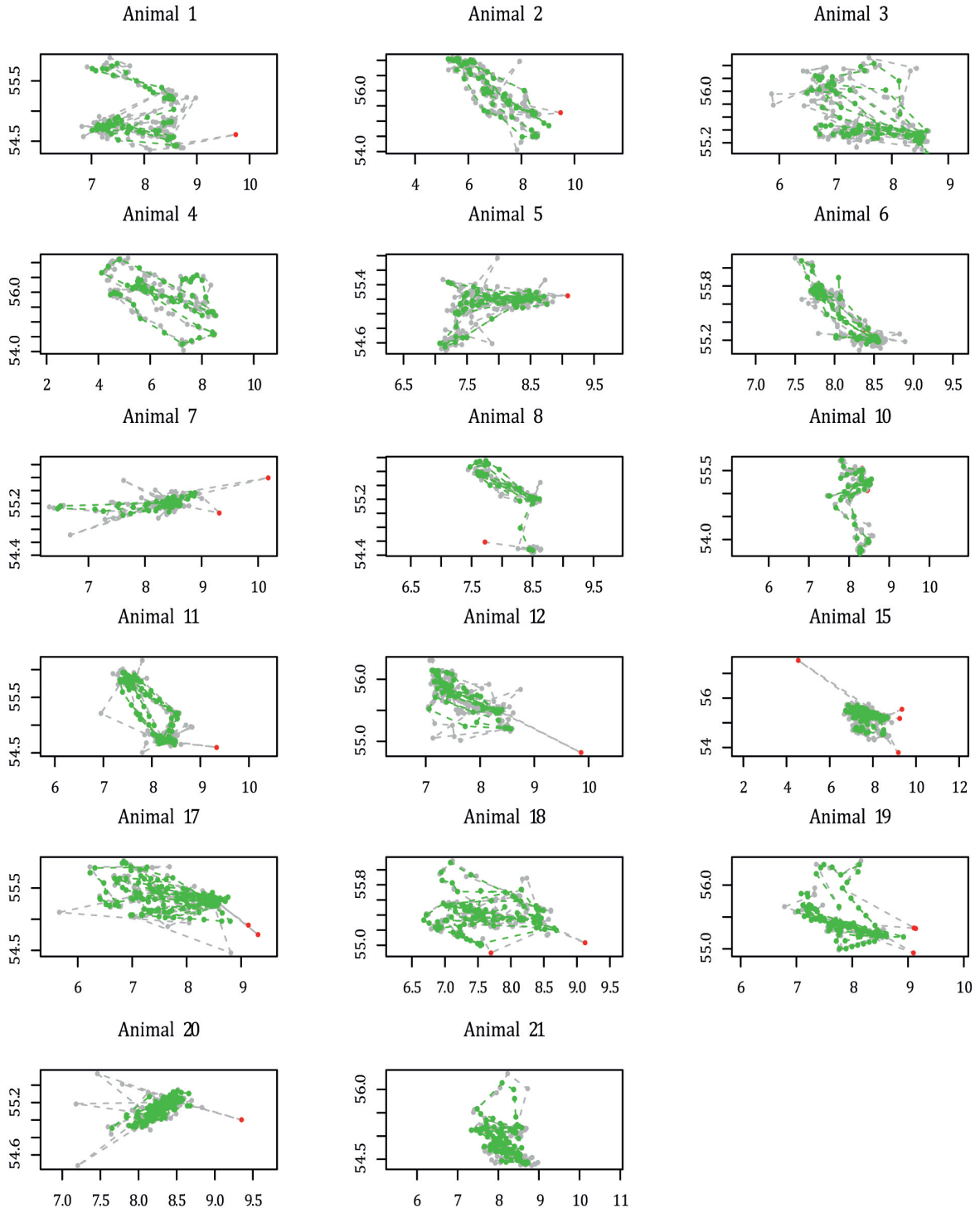
Figure D.3: Autocorrelation functions (acf's) of daily movement rate for observed and modeled paths for lags 1-n/2, n = number of locations, for all animals. The graphics correspond to the final model, "Double-Switch N". Dots correspond to observed acf's and dashed lines correspond to the 95% credibility interval for the acf's modeled with the different models for 30000 replicates.
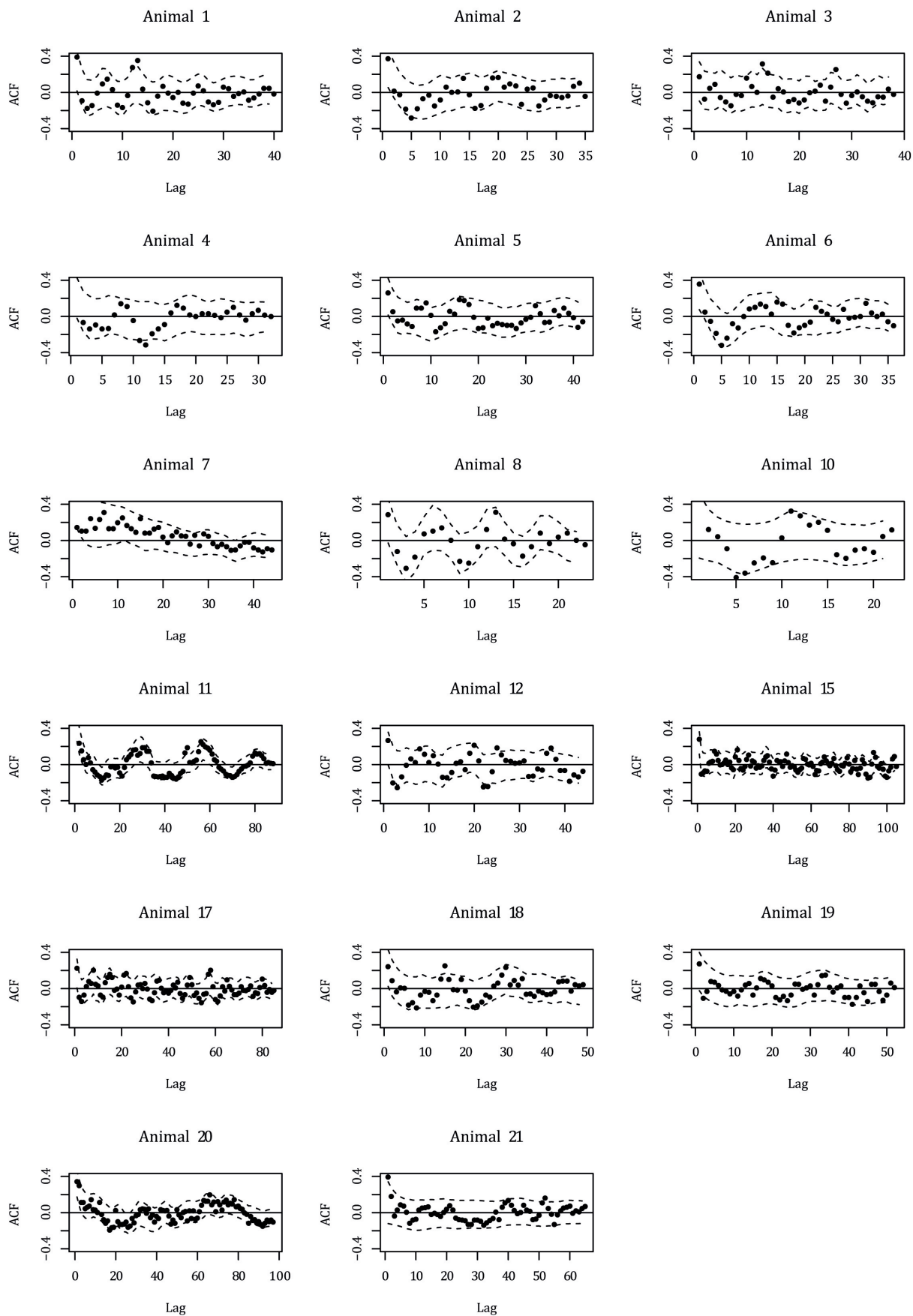
Figure D.4: Distribution of all animals with colours representing behavioural states using results from "Double-Switch N" model. Red and green dots represent slower and faster behaviours, respectively.