Forecasting the trajectories of Southern Resident Killer Whales with stochastic continuous-time movement models

by

Teng-Wei Lin

M.Sc., National Taiwan University, 2020
B.Sc., National Taiwan University, 2018

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2023

Declaration of Committee

Name:	**Teng-Wei Lin**
Degree:	**Master of Science**
Thesis title:	**Forecasting the trajectories of Southern Resident Killer Whales with stochastic continuous-time movement models**
Committee:	**Chair:** Liangliang Wang Professor, Statistics and Actuarial Science
	Ruth Joy Supervisor Assistant Professor, Environmental Science Adjunct Professor, Statistics and Actuarial Science
	Derek Bingham Committee Member Professor, Statistics and Actuarial Science
	Michael Dowd Examiner Professor Department of Mathematics and Statistics Dalhousie University

Abstract

The Southern Resident Killer Whale (SRKW) is an endangered population of killer whales that is present in the Salish Sea. This fish-eating predator has been heavily impacted by human activities in the region, particularly by commercial vessels in shipping lanes that traverse federally-designated SRKW critical habitat. Forecasting the movement trajectories of these whales would help provide early warning alerts to slow down or reroute commercial vessels, and reduce the risks of ships overlapping with whale presence. In this study, we develop a stochastic animal movement model that is guided by a historical database of sighting records of SRKW. Specifically, we make use of a continuous-time Ornstein-Uhlenbeck (O-U) velocity process that provides the basis for a movement forecast system and simulates realizations of SRKW velocities and trajectories given initial conditions. However, if the forecast system were to simply rely on the O-U velocity process alone, it would steer simulated whale trajectories to areas where SRKWs are rarely found. To address this, we propose a direction blending scheme to project the simulated velocities in more realistic directions. It makes use of historical directional information along with the O-U velocity process to create more probable pathways consistent with observed SRKW movement patterns. By integrating the simulated trajectories generated from the simulated velocities, we establish a dynamic probability-based forecast scheme that demonstrates skill in forecasting SRKW trajectories on a time-scale that aligns with the time to slow and reroute commercial vessels.

Keywords: Southern Resident Killer Whales; Animal movement modelling; Continuous-time model; Directional persistence; Trajectory forecast

Acknowledgements

I thank Alexis Morrigan for her assistance in sharing and organizing the sighting data of the Southern Resident Killer Whale in the Orca Master dataset, managed by the Whale Museum in the United States. Additionally, I extend my thanks to all individuals involved in documenting daily Southern Resident Killer Whale sightings from the past until now, contributing to the success of this research.

I deeply appreciate all the guidance and support from my supervisors, Ruth Joy and Michael Dowd, during my time at SFU. Thank you for bringing me to the world of animal movement modelling and the conservation of Southern Resident Killer Whales. Coming from the other end of the Pacific Ocean, I am fortunate to build connections with the incredible marine life in the Salish Sea and the local community on the Gulf Islands.

Many thanks to all faculty and staff members in the Department of Statistics and Actuarial Science at SFU. Without your help, I could not have smoothly completed my studies at SFU in Canada.

Thank you, to my fellow graduate students, who helped me overcome all challenges in these two years. Special thanks to Zhi Yuh, Yirong, Junpu, Nikhil, Kalpani, David, Gurashish, and Hashan for your precious friendship and support.

Finally, a sincere thank you to my parents, brother, and grandparents for your unconditional love and support in helping me achieve my dreams.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Southern Resident Killer Whale (SRKW), *Orcinus Orca*, is a population of killer whales listed as 'endangered' in Canada and the United States. SRKWs inhabit the waters of the northeastern Pacific Ocean and are present year-round, but more prevalent in the Salish Sea (Fig 1.1) from spring to fall [13]. The SRKW is also a highly social animal in which closed family groups form one of three pods (J, K, and L pod), and whales belonging to the same pod tend to move together. Evidence suggests that the noise and physical disturbance from commercial vessels alters the movement of SRKWs[25, 26, 39, 33, 18]. Waters such as the Haro Strait, Boundary Pass, and Active Pass are busy shipping lanes that connect Canadian and American Pacific Ports with the Pacific Ocean. There are over 10,000 vessels annually operating in the Salish Sea [10], which pose an important threat to SRKWs through acoustic and physical disturbance [18, 38, 35]. However, if ship pilots can adjust their paths and/or reduce their speeds before approaching killer whales, the noise footprint and the risk of vessel collisions with SRKW are reduced. Forecasting the trajectories of moving SRKWs or probable regions where SRKW may be present can reduce the impacts of commercial shipping by alerting ship pilots and providing enough time to change the ships speed and direction. The core tool needed to do this is a statistical dynamical whale movement model that can forecast the trajectories of an SRKW pod on time-scales of an hour or more. Development of this SRKW forecasting model is the central objective of this thesis.

Animal movement modelling has been widely used in ecological research to understand the behaviours of animals and predict their trajectories. There have been two major approaches used in animal movement modelling: discrete-time and continuous-time models

[22, 14]. Discrete-time models are an intuitive choice and have been widely used in animal movement research [23, 21, 11]. They are popular because they are straightforward to apply with simulation procedures that are typically discrete [14], and there are abundant statistical techniques for their implementation such as the correlated random walk proposed by Clifford et al. [29] as well as various standard time series models that can be applied. However, true animal movement patterns are continuous in both time and space and thus, discrete-time models do not reflect the reality of animal movement [14]. In addition, discrete-time models are problematic for irregularly sampled data [17, 11] because observation times may not match the time steps of the models, and data may need to be thinned or interpolated to match the discrete time step [6, 17, 11]. In contrast, continuous-time models allow us to directly use raw data by defining animal movement as a stochastic process which is controlled by scale-invariant parameters that can be estimated under irregularly sampled data and various time steps [17, 21, 9, 11, 14]. Therefore, even though continuous-time models will generally be discretized for numerical solutions in computers, it is still worth constructing animal movement models from a continuous-time perspective [14].

The Ornstein-Uhlenbeck Process (O-U Process) proposed by Ornstein and Uhlenbeck [36] is one of the most commonly used continuous-time models that describes a Gaussian stationary process composed of a mean-reverting process and random fluctuations, or the Wiener process. The O-U process can be used to model either animal positions or velocities [30]. For the position-based O-U process, the random variable is animal positions, and the O-U process model represents the instantaneous change in position, which is velocity. The mean-reverting process implies that the movement of animals will be steered towards specific locations such as their core habitat or foraging areas [31]. The velocity-based O-U process uses a random variable corresponding to animal velocity and models the instantaneous change in velocity, which is acceleration. Since velocity is the prognostic variable, the animal position must be obtained by integrating the velocity process [30].

The O-U process has been applied to simulate a variety of animal movements including marine mammals [11]. For example, Brillinger et al. [5] integrated the O-U process on the surface of a sphere to model the migration of elephant seals (*Mirounga angustirostris*).

Johnson et al. [17] utilized the O-U velocity process within a state-space framework to analyze the movement of harbour seals (*Phoca vitulina*) and northern fur seals (*Callorhinus ursinus*). For our target species, Randon et al. [32] pioneered an SRKW movement forecast framework that used the velocity-based O-U process within a state-space framework with data assimilation to forecast the trajectories of SRKWs. This data assimilative SRKW movement forecast was demonstrated to have predictive skill out to 2.5 hours when applied to a single test case in the Salish Sea.

Although the approach of Randon et al. [32] worked well for a single trajectory, it was not expanded to forecast trajectories in other regions of the Salish Sea and did not provide a comprehensive movement modelling simulator. In addition, it did not statistically estimate each parameter in the O-U process, instead specifying them by assumptions. Furthermore, the framework used the gradient of potential functions defined by monthly sighting data to indicate the mean velocity in the mean-reverting process in the O-U process. Potential functions are a physical concept to illustrate the motion of particles in Newtonian dynamics [3, 31]. The mean-reverting process in the position-based O-U process can be characterized by the gradient of a potential function [4], which assumes animals tend to move along the gradient of a potential field (i.e., from low potential to high potential). Following this concept, potential functions can also be used to control the direction or speed of motion of animals [3]. For example, if preferred habitats are assumed to have high potential and current locations are found in an area of low potential, animals will move along the gradient toward the high-potential valued habitat. Note that the potential function in the position-based O-U process is a velocity potential, aligning with the unit of what is modelled: instantaneous change in position. However, in the velocity-based O-U process, the mean-reverting process steers velocity toward a specified velocity, instead of directly to preferable positions. Therefore, steering trajectories toward preferable positions by controlling the velocity-based O-U process is more complicated. The most key issue is that the potential function inserted in the velocity-based O-U process is an acceleration potential, aligning with the unit of what is modelled: the instantaneous change in velocity, which requires more high-quality and precise position datasets. Therefore, it is worth finding an alternative method to steer

the trajectories generated from the velocity-based O-U process to preferable positions. We will do this via a directional blending approach using historical pathway information.

Building on the research of Randon et al. [32], this thesis is focused on forecasting the trajectories of SRKWs given an initial location and velocity in the Salish Sea based on the historical sightings of SRKWs. We applied historical sighting data of SRKWs from the Orca Master dataset managed by the Whale Museum in Washington, USA to access the historical trajectories and velocities of SRKWs. Archived (historical) trajectories were compiled and spatially indexed to provide information about direction across SRKW habitats in the Salish Sea. Based on the historical velocities, we aimed to simulate realistic trajectories of SRKWs and propose a preliminary forecast scheme. As mentioned before, SRKWs usually move with the same pod members, and thus, it is more realistic to forecast the trajectories of a pod than an individual whale. However, forecasting a pod is challenging because the information about a pod such as range and pod leaders is dynamic and hard to collect. Therefore, we will assume that we are simulating the trajectories of a pod leader, and other SRKWs of the same pod would follow the movement of the pod leader. Based on the historical velocities, we constructed the velocity-based O-U process as a basis to simulate velocity. Simultaneously, we solved the issue of estimating the parameters of the velocity-based O-U process with the historical irregularly sampled velocity data. Then, we established a new direction-blending method as an alternative to potential functions to guide simulated velocities in preferred directions based on historical records. In addition, we proposed a coastal avoidance algorithm to ensure the trajectories integrated from the simulated velocities stay in the waters. Finally, to test our forecasting system, we predict three trajectories with different initial locations in the Salish Sea to illustrate our approach.

This thesis report is organized as follows. The Observations and Data Sources chapter will introduce the historical sighting data of SRKW in the Orca Master dataset curated and provided by The Whale Museum. The Methods section will dive deeper into the velocity-based O-U process and introduce the direction-blending method, the coastal avoidance algorithm, and the SRKW simulation scheme. The implementation and simulation of tra-

jectories will be presented in the Results chapter. Various aspects of our approach and findings will be explored in the Discussion chapter.

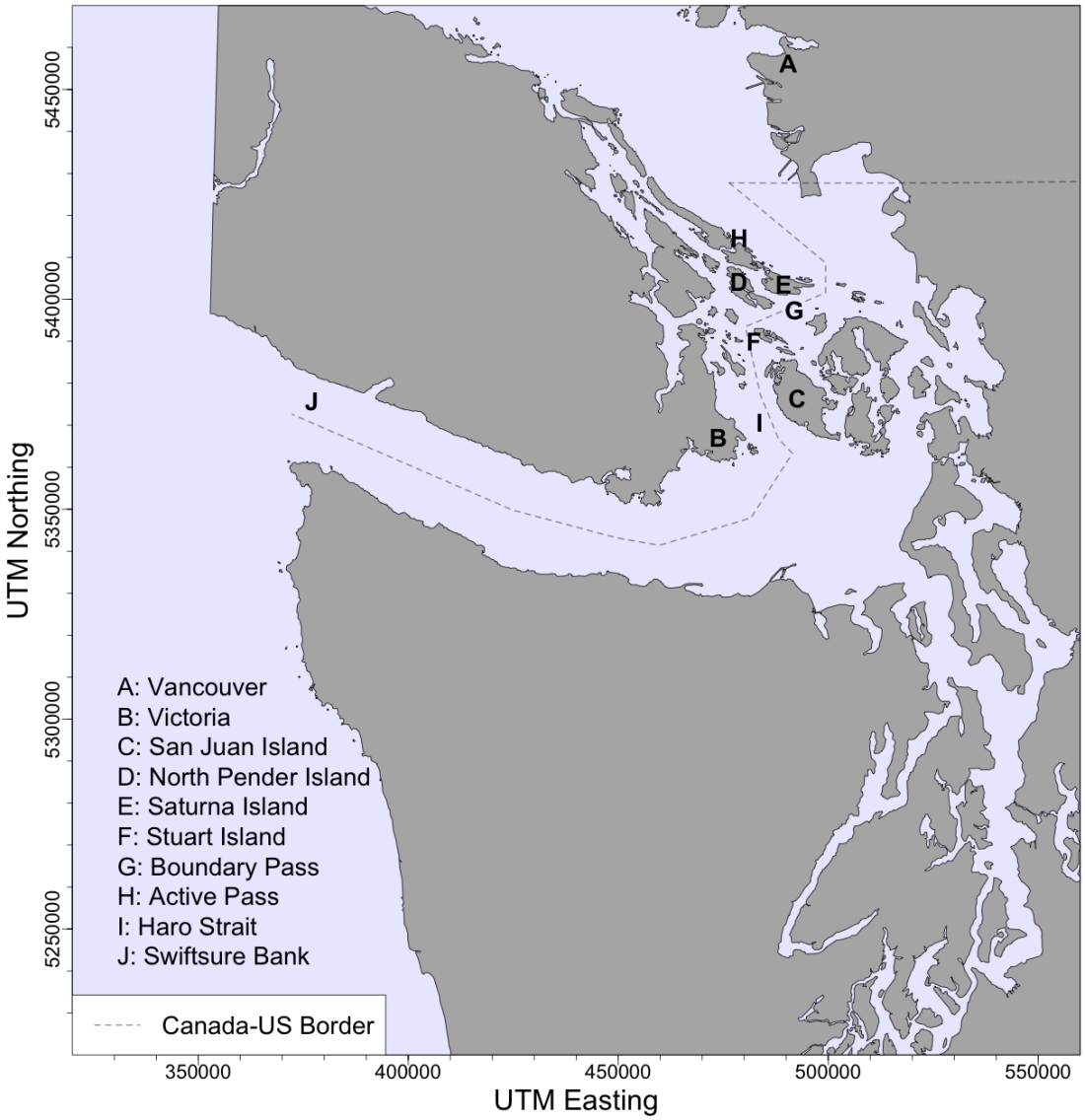


Figure 1.1: Area of Interest: Transboundary waters at the core of SRKW summer range including the Southern Gulf Islands in Canada (North Pender Island: D, Saturna Island: E), San Juan Islands (San Juan Island=C), and important commercial shipping zones (Boundary Pass: G, Active Pass: H, and Haro Strait: I).

Chapter 2

Observations and Data Sources

Our observations comprise trajectories of SRKW in the Salish Sea and provide the basis for both an empirical understanding of SRKW movement in the Salish Sea, as well as guidance for making a whale movement model capable of successfully forecasting SRKW trajectories. The dataset and processing steps are outlined below.

2.1 The Orca Master Dataset

We require trajectories and velocities of SRKWs to construct and validate our stochastic movement model and simulate trajectories. Trajectory information used in this study comes from the sighting records of SRKWs contained in the Orca Master dataset from 2012 to 2022. These data are curated by the Whale Museum in the US, which is a non-govermental organization whose mission is to promote stewardship of whales and the Salish Sea ecosystem through education and research. The sighting records of SRKWs contained in the Orca Master dataset were collected by many individuals and organizations, including researchers, citizen scientists, and commercial whale watch vessels. The Whale Museum is responsible for merging and filtering the sighting observations and maintaining the dataset. Further details can be found in [27].

There are a total of 7097 SRKW sighting records available between 2012 and 2022. As shown in Table 2.1, the sighting records in the Orca Master dataset include date, time, pod, direction of travel, and location [27]. A pod is a closely-related family group of SRKWs comprised of mothers and their offspring. Currently, SRKWs consist of three pods: designated J, K, and L [27]. These three pods can sometimes converge, hunt, socialise, mate and

travel together, and then separate back into their individual pods. Figure 2.1 shows the locations of all sightings in the Orca Master dataset between 2012 and 2022. The sightings cover most of the Salish Sea. The locations were recorded as either: (i) positions nearest the pre-specified Whale Museum quadrants (grids approximately 4.6 km $\times$ 4.6 km) or (ii) the actual positions reported as longitude and latitude and converted to the Universal Transverse Mercator (UTM) coordinate system. Although the coarse-resolution Whale Museum grids are designed to take measurement errors into consideration and make the dataset easier to manage, they cause geographical bias when using the data to build statistical models. We also emphasize that because of the hundreds of citizen scientists who contributed to the Orca Master dataset and their investment in continuous marine mammal observation efforts coupled with real-time connections through social media (Facebook, Discord, WhatsApp, etc), SRKWs are unlikely to be missed if they are indeed present in the Salish Sea. Therefore, SRKWs can be tracked continuously once they are reported, which provides fairly precise sighting locations which together allow us to infer the speed and direction of whale movement.

Date	Time	Pod	Direction	Latitude	Longitude
2/12/2012	17:00	J	W	48.5799	-122.8595
2/12/2012	17:07	J	W	48.5248	-122.6544
2/12/2012	17:20	J	SW	48.5386	-122.9544

Table 2.1: An example of three rows of the 7097 SRKW sightings in the Orca Master dataset.

2.2 Sighting Density

Based on whale locations plotted in Figure 2.1, we created the sighting density of SRKWs in the Salish Sea from 2012 to 2022 using kernel density estimation as shown in Figure 2.2. Higher sighting density represents a higher frequency of observing SRKWs, but it is important to point out that this is influenced by, and confounded with, the concentrated effort of whale sighters in specific geographical locations. The shoreline and waters around San Juan Island have the highest sighting density due to a high number of whales and a high number of human observers. Boundary Pass and Active Pass also have high sighting densities. Since SRKWs are mainly observed by personnel on vessels and citizen scientists

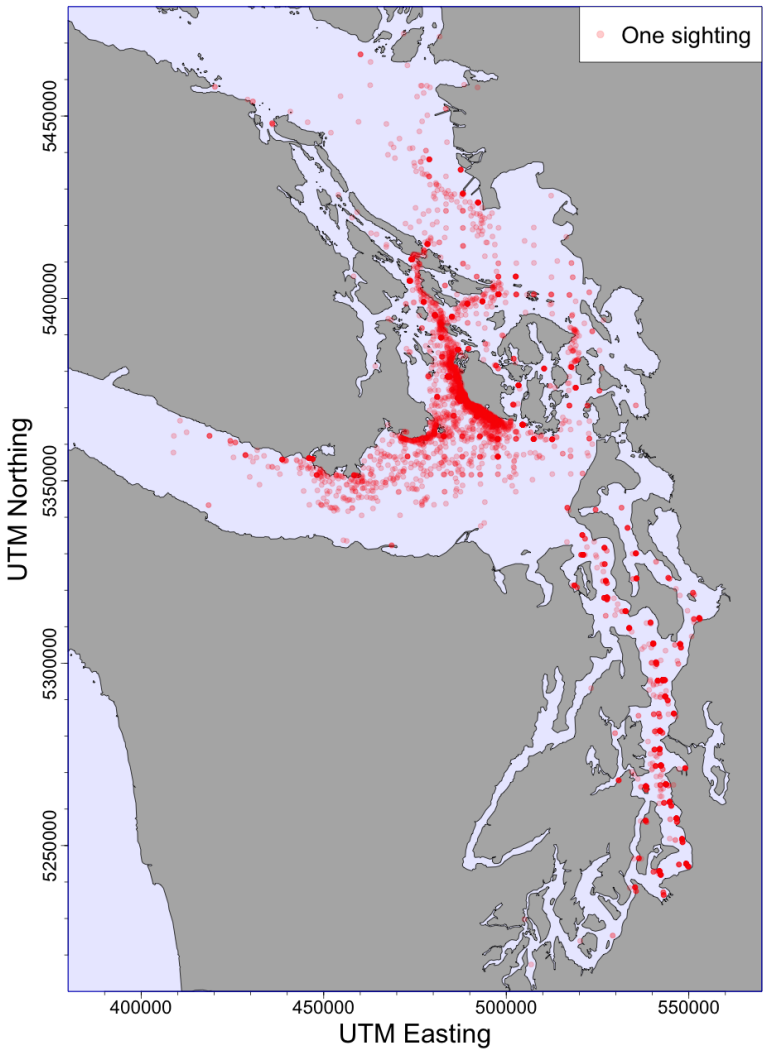


Figure 2.1: The sighting records of SRKWs in the Salish Sea from 2012 to 2022. A single Whale location data in the the Orca Master dataset is plotted with transparency as shown in the legend. This distinguishes between a single whale at a single location and multiple sightings reported at the same location where the sightings are represented as an opaque red point. Grid centroids of The Whale Museum quadrants are associated with multiple sightings and are particularly visible in Puget Sound in the southern region of the map.

on land, there is sampling bias in the sighting density [37], or a preferential sampling of whale sightings due to human observer concentration. That is, places with high sighting density may tell us where active vessels or scientists are, and may not actually indicate higher chances of observing SRKWs. Places with low sighting density may also have high chances of SRKW being present but remaining unobserved due to lack of effort. In fact, because there are so many vessels going through Boundary Pass and Active Pass, along

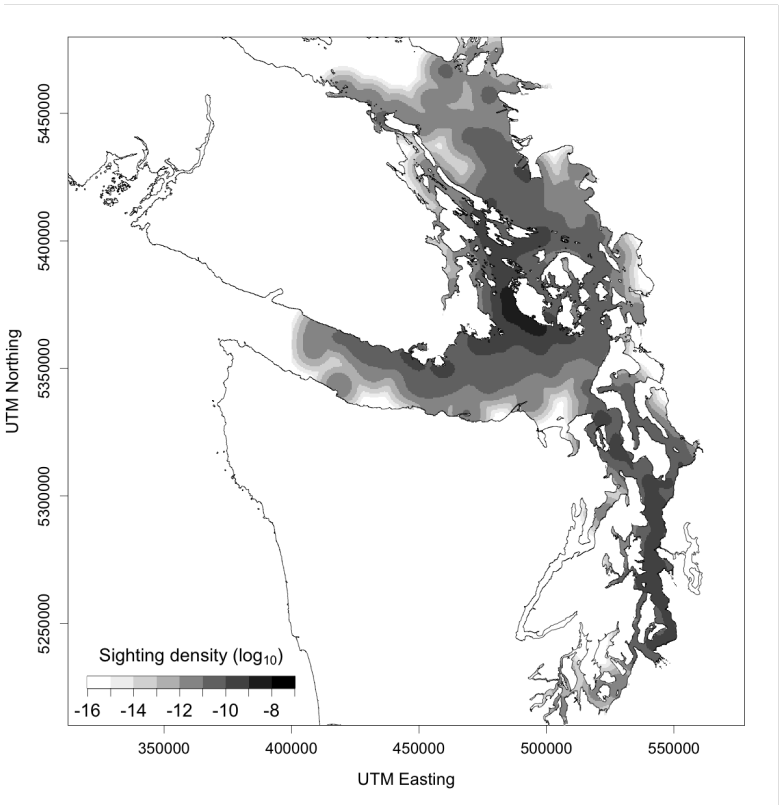


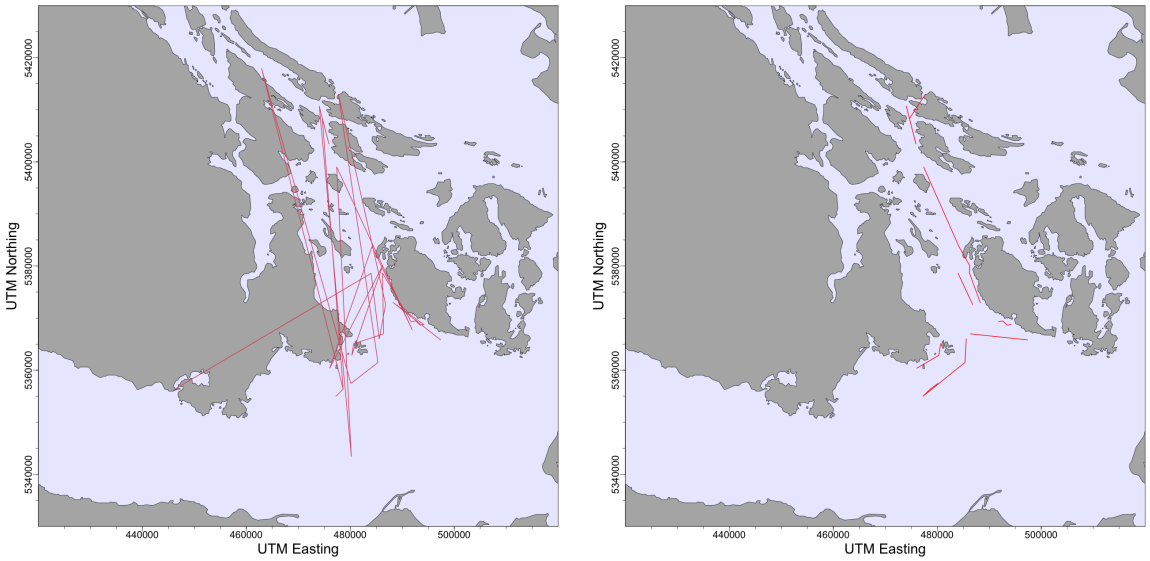
Figure 2.2: The sighting density of SRKWs in the Salish Sea from 2012 to 2022.

with co-located live-streaming hydrophones as an additional source of information on 'when' observers should look, the sighting density in these regions is high. However, even though sampling bias exists, the whale density is not directly used. Instead, the archived sighting data provides excellent and valuable information about the direction and velocity of the SRKW trajectories for use in our study.

2.3 Trajectory Extraction

In this study, we need the historical trajectories of SRKWs and velocities generated from the trajectories as the inputs to our model. Unfortunately, the most straightforward approach of simply connecting the sighting records in sequence does not give us reasonable trajectories. For example, Figure 2.3a shows the results of sequentially connecting all observations within one day. The pattern is messy and can be difficult to interpret. Moreover, the displacement between two widely separated locations does not represent a viable route for

SRKWs. SKRWs sometimes can widely spread out across dozens of kilometres even though they belong to the same pod [40]. Therefore, multiple observations can be recorded within a short time period and be located far from each other. In addition, observations may be just a single matriline (one mother and her offspring) within a pod may stray away from other pod members for a short period of time. Therefore some observations should not be joined together into a trajectory, as they would not represent a true movement trajectory. Thus, the sighting records must be systematically reordered and split into more realistic trajectories. A procedure for this is outlined below.



(a) Before reordering and splitting consecutive SRKW observations

(b) After reordering and splitting consecutive SRKW locations into realistic trajectories

Figure 2.3: Trajectories from sighting data. By reordering and splitting the sighting records on the left panel above that suggests an unrealistic trajectory is transformed into the 8 trajectories seen on the right panel constrained to 'realistic' speeds, time steps, and to avoid land.

Although obtaining complete and long trajectories is always desired, we compromise by isolating shorter trajectories that have reasonable values in terms of both speed and direction, and are constrained by certain observed limitations so as to be realistic. Assume there are n sighting records in a day with sighting location $\mathbf{x_i} = \left[x_1^{(i)}, x_2^{(i)}\right]^T$ at time t_i, $i = 1, ..., n$. Here, the first element of the vector represents the east-west coordinate, and

the second element represents the north-south coordinate. We used the following method to isolate realistic trajectories from consecutive sighting records.

1. Extract all pod-specific sighting records within the same day, and assign them to the same trajectory.

2. Assume a time interval between two sighting records greater than 1 hour is too long to be the same trajectory, and hence break the trajectory into two different trajectories.

3. Assume that a travel speed between two locations should be smaller than a reasonable value, e.g., <30 km/hr. That is, if $t_i - t_{i-1} > 1$ hour or $\frac{\|\mathbf{x_i} - \mathbf{x_{i-1}}\|}{t_i - t_{i-1}} > 30$ km/hr, $\mathbf{x_i}$ and $\mathbf{x_{i-1}}$ belong to two different trajectories.

4. If there is more than one trajectory within an hour and less than 30 km away, the location $\mathbf{x_i}$ is connected to the closest trajectory.

5. If there are no trajectories to connect, $\mathbf{x_i}$ will become the start of a new trajectory. If there are no following sightings connected behind $\mathbf{x_i}$, then $\mathbf{x_i}$ will be a single unconnected point and not included in any trajectories.

The results of using the trajectory splitting procedure are shown in Figure 2.4. We managed to separate 1266 possible trajectories from the sighting records database for the 11 years between 2012 and 2022. It is worth noting that this trajectory extraction method is not perfect but overall seems to work quite well. We may sacrifice some chances to connect more complete trajectories, and many of the separated trajectories were quite short. In addition, there are still a few trajectories crossing over land. However, this method is easy and quick to apply, and the separated trajectories generally maintain the information of moving direction at different locations, which is helpful in this study. These trajectory data will be used as the basis for both calibration and validation of the stochastic whale movement model which is outlined in the next chapter.

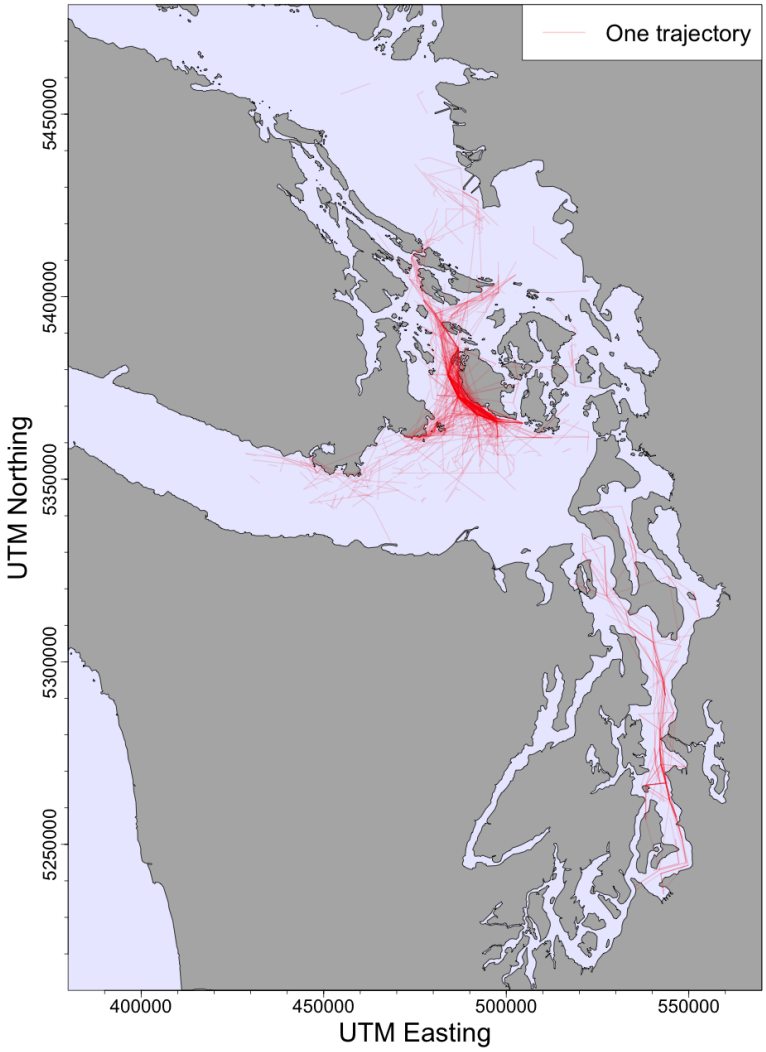


Figure 2.4: The result of creating split trajectories to provide information about the direction and velocity of SRKW movement in the Salish Sea study area.

Chapter 3

Methods

We aim to forecast SRKW movement using a stochastic simulation model of whale trajectories given initial locations and directional information. Using the velocities obtained from historical sightings data derived from the Orca Master database, our approach applies the Ornstein-Uhlenbeck (O-U) velocity model to simulate realizations of velocities for an SRKW pod. Note that an O-U process unconstrained by any observational information will steer simulated trajectories to areas where SRKWs rarely transit. Simply put, the raw O-U model is too flexible and hence unrealistic. To address this, we incorporate historical trajectory information to project the velocities in more realistic directions. In addition, it is necessary to modify the stochastic process model to ensure the trajectories drawn from the simulated velocities stay in water (i.e. SKRW cannot be permitted to go on land, or transit over land). In this chapter, we will provide an overview of the O-U velocity model, the direction-blending method to project velocities, and the coastal avoidance algorithm to maintain trajectories in water, and present examples of the whale simulation scheme in different regions of the Salish Sea.

3.1 Ornstein-Uhlenbeck Velocity Model

3.1.1 Overview

The Ornstein-Uhlenbeck (O-U) process is a stochastic process and the continuous-time analog of the discrete-time AR(1) stochastic process. The O-U process resolves the issue of irregularly sampled data due to its continuous time formulation which accommodates any observation times. It also has consistent parameters for different discretization scenarios

and time steps [11]. This is important since the velocities of SRKWs drawn from sighting records are not regularly observed in time. Therefore, we choose the O-U process to account for the gross features seen in the irregular time series and then discretize the process into the AR(1) model under controlled time steps to simulate the velocities of SRKWs.

Since we are modelling two-dimensional position (i.e. longitude or UTM easting, and latitude or UTM northing), we also consider the velocities of SRKW in 2 dimensions with the bivariate velocity $\mathbf{v_t} = [v_{tx}, v_{ty}]^T$. Anderson-Sprecher and Ledolter [2] and Johnson et al. [17] suggested that treating bivariate velocity components independently of one another is more realistic. For example, a positive correlation would mean SRKWs tend to have a rotational motion. However, in reality there is little evidence for this and it is possible for SRKWs to randomly switch directions between northeast and southwest, or indeed over all directions. Therefore, we follow the advice of [17] and consider the parameters of the bivariate velocity as independent in this study. This also means we can actually model our system as 2 independent univariate processes. As a consequence, for the convenience of explanation, we use the univariate velocity, $v_t = v_{tx}$ or v_{ty}, to illustrate the following theories and inferences.

General Form

The instantaneous velocity of a killer whale under the O-U process can be defined by the following equation [36]:

$$dv_t = \alpha(\mu - v_t)dt + \sigma dW_t\,, \tag{3.1}$$

where v_t is the velocity at time t, α is a persistence parameter or the speed of mean reversion associated with autocorrelation, μ is the mean velocity or a drift term, σ is the magnitude of the stochasticity [11], and W_t represents the Wiener process with mean zero and the property: $W_{t_2} - W_{t_1} \sim N(0, t_2 - t_1),\ 0 \leq t_1 < t_2$ [15, 17]. Note that α has units of time^{-1}, the dW_t term has units of time$^{1/2}$, and σ has units of distance $\times$ time$^{-3/2}$ [11]. The α parameter can be interpreted as the time-scale for reverting to the mean velocity, and the σ parameter has no obvious biological interpretation [11].

It is worth noting the analytic solution to the Equation (3.1), from [15], is

$$v_{t+\delta} = \mu - e^{-\alpha\delta}[\mu - v_t] + \sigma \int_t^{t+\delta} e^{\alpha(t-u)} dW(u)\,, \tag{3.2}$$

where δ is a small time interval. The analytic solution also implies $v_{t+\delta}$ and v_t are almost independent if $\delta \geq 3/\alpha$ [17].

Inference such as maximum likelihood estimation (MLE) based on (3.1) and (3.2) is applicable if observations are continuous and meet the asymptotic assumptions of MLE, which will be explained later. In our data, the sighting records of SRKWs are discrete, and time increments are too large, and the observations too sparse, to meet the asymptotic assumptions of MLE. Therefore, we apply numerical approximation to estimate parameters in the O-U process.

Euler Method

The Euler-Maruyama method (Euler Method) is the easiest and the most common numerical approximation scheme to discretize the continuous time formulation of the O-U process. Equation (3.1) can discretize (following [15]) as:

$$v_{t+\Delta t} - v_t = \alpha(\mu - v_t)\Delta t + \sigma[W_{t+\Delta_t} - W_t]$$

$$v_{t+\Delta t} - v_t = \alpha(\mu - v_t)\Delta t + \sigma\sqrt{\Delta t}\epsilon_t\,, \tag{3.3}$$

where $\epsilon_t \sim N(0,1)$, and Δt can be any arbitrary time step. Often in practice $\mu = 0$ is assumed for the velocity model [17]. If this is the case then, Equation (3.3) will become:

$$v_{t+\Delta t} = (1 - \alpha\Delta t)v_t + \sigma\sqrt{\Delta t}\epsilon_t. \tag{3.4}$$

Note that this corresponds to the form of auto-regressive order one, or AR(1) process

$$v_{i+1} = \phi_{\Delta t} v_i + \eta_i\,, \tag{3.5}$$

where $\phi_{\Delta t} = (1 - \alpha\Delta t)$ and $\eta_i \sim N(0, \sigma^2\Delta t)$, and α and σ are two unknown parameters that must be specified.

3.1.2 Parameter Estimaton

In general, we aim to estimate μ, α, and σ in the O-U process. If observations of velocity, v_t, are continuous and meet the asymptotic assumptions, the maximum likelihood estimator (MLE) is commonly used. If observations are discrete, the pseudo-likelihood estimator (PLE) is commonly used, and the results align with the least squares estimator (LSE). These cases are each explained below.

Continuous Case

We define the time increment as δ and sample size as n. The MLE is asymptotic normal and consistent under two asymptotic schemes. The first scheme is the large sample scheme where δ is a fixed constant, and $n \to \infty$. The second scheme is rapidly increasing design scheme wherein δ decreases with n but the total observation time goes to infinity, namely:

$$\begin{aligned} &\delta \to 0,\ n \to \infty,\ n\delta \to \infty \\ &\text{and for some } k \geq 2,\ n\delta^k \to 0. \end{aligned}$$

The likelihood function can be derived from the conditional density of Equation (3.1) [15, 34]:

$$\begin{aligned} &v_t \mid v_{t-1} \sim N(m, \Lambda) \\ &m = v_{t-1}e^{-\alpha\delta} + \mu(1 - e^{-\alpha\delta}) \\ &\Lambda = \frac{\sigma^2(1 - e^{-2\alpha\delta})}{2\alpha}. \end{aligned}$$

Let $\phi(x)$ be the probability density function of standard normal distribution. The likelihood function of (μ, α, σ), from [34], is

$$L(\mu, \alpha, \sigma) = \phi\left(\frac{\sqrt{2\alpha}(v_0 - \mu)}{\sigma}\right) \times \prod_{t=1}^{n} \phi\left(\frac{\sqrt{2\alpha(1 - e^{-\alpha\delta})}(v_t - v_{t-1}e^{-\alpha\delta} - \mu(1 - e^{-\alpha\delta}))}{\sigma}\right).$$

By maximizing the likelihood function above using observations on v_t, the MLE $\hat{\mu}$, $\hat{\alpha}$, and $\hat{\sigma}$ can be obtained.

Discrete Case

For the discrete case under the Euler method, the distribution of $v_{t+\Delta t} - v_t$ in Equation (3.3), from [15], is

$$\begin{aligned} &v_{t+\Delta t} - v_t \sim N(m, \Lambda) \\ &m = \alpha(\mu - v_t)\Delta_t,\ \Lambda = \sigma^2 \Delta_t \,. \end{aligned}$$

When the time series is irregular, the pseudo-log-likelihood function is [15]

$$\ell(\mu, \alpha, \sigma) = -\frac{1}{2} \sum_{i=1}^{n-1} \left[\frac{(v_{i+1} - v_i - \alpha(\mu - v_i \Delta_i))^2}{\sigma^2 \Delta_i} + \log(2\pi\sigma^2 \Delta_i) \right],$$

where $\Delta_i = t_{i+1} - t_i$, and Δ_i's can be different time intervals. Taking σ^2 as constant, maximizing the log-likelihood function is equal to minimizing the function

$$\sum_{i=1}^{n-1} (v_{i+1} - v_i - \alpha(\mu - v_i \Delta_i))^2 \,.$$

As mentioned in Subsection 3.1.1, we most often assume $\mu = 0$, and the function above simplifies to

$$\sum_{i=1}^{n-1} (v_{i+1} - v_i + \alpha v_i \Delta_i)^2 \,. \tag{3.6}$$

The consistent estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n-1} (v_{i+1} - v_i)^2}{\sum_{i=1}^{n-1} \Delta_i} \,.$$

From the perspective of linear models, Equation (3.4) can be rewritten as

$$\frac{v_{t+\Delta t}}{\sqrt{\Delta t}} = \frac{v_t}{\sqrt{\Delta t}} - \alpha \sqrt{\Delta t} v_t + \sigma \epsilon_t \,, \tag{3.7}$$

which aligns with the form of a linear model. The least-squares estimates (LSE) can be obtained by minimizing the function:

$$\sum_{i=1}^{n-1}\left[\frac{v_{i+1}}{\sqrt{\Delta_i}}-\frac{v_i}{\sqrt{\Delta_i}}+\alpha\sqrt{\Delta_i}v_i\right]^2, \tag{3.8}$$

which is equivalent to minimizing Equation (3.6). In addition, we can also show the variance of $\hat{\alpha}$ calculated from the perspectives of PLE and LSE are the same. For the PLE, the variance of α can be obtained by the Fisher information, the inverse of the negative expectations of the second derivative of the log-likelihood function.

$$\ell(\alpha)=-\frac{1}{2}\sum_{i=1}^{n-1}\left[\frac{(v_{i+1}-v_i+\alpha v_i\Delta_i)^2}{\sigma^2\Delta_i}+\log(2\pi\sigma^2\Delta_i)\right]$$

$$Var(\hat{\alpha})=\left(-E\left[\frac{d^2}{d\alpha^2}\ell(\alpha)\right]\right)^{-1}=\frac{\sigma^2}{\sum_{i=1}^{n-1}\Delta_i v_i^2}$$

For LSE, the variance of α is straightforward.

$$\begin{aligned}\frac{v_{i+1}-v_i}{\sqrt{\Delta_i}}&=-\alpha\sqrt{\Delta_i}v_i+\epsilon_i\\ \text{Let } X&=\left[\sqrt{\Delta_1}v_1,...,\sqrt{\Delta_{n-1}}v_{n-1}\right]^T\\ Y&=\left[\frac{v_2-v_1}{\sqrt{\Delta_1}},...,\frac{v_n-v_{n-1}}{\sqrt{\Delta_{n-1}}}\right]^T\\ Var(\hat{\alpha})&=Var[(X^TX)^{-1}XY]=\sigma^2(X^TX)^{-1}\\ &=\frac{\sigma^2}{\sum_{i=1}^{n-1}\Delta_i v_i^2}\end{aligned}$$

The similarity is unsurprising because the discrete O-U process with no drift term corresponds to the AR(1) process. Estimates from PLE and LSE are consistent for the AR(1) process.

If Δ_i's are small enough, the pseudo-likelihood estimator is consistent and asymptotically normal [15], the impact of the value of Δ_i does exist, and large Δ_i can make estimates biased [15]. In our case, the time interval between any two observations in the same trajectory is always less than 1 hour. Although there would be some bias existing for the Euler method,

compromising its easiness to apply, we still choose to apply the Euler method to estimate the parameters.

3.2 Direction-blending

3.2.1 Motivations

As noted earlier, the simulated trajectories generated by realizations of the velocities simulated purely from the standard O-U process are not very useful for SRKW forecasts. The reason is that predicted trajectories often pass through regions that SRKWs rarely visit in practice. To solve this issue, we propose a direction-blending method to alter the directions of simulated velocities by making use of historical directional information to alter simulated trajectories to be more like whale movement tracks.

3.2.2 Assumptions

Directional Memory

Figures 2.1 and 2.4 indicate that the movement of SRKWs has a certain level of persistence in direction and a tendency to move along oceanic highways, or pathways, in certain regions (e.g., in Haro Strait). We assume SRKWs do not significantly change their directions over a period of time. We also assume in our simulation that the directional memory persists for 3 steps, or about 9 minutes. That is, we store the movement direction in the previous 9 minutes as a directional memory, designated θ_ℓ. A long-directional memory would make simulated trajectories too persistent, whereas a short-directional memory makes simulated trajectories too random. The directional memory will play a role in altering the direction of the current velocity simulated from the O-U process, which will be explained later Section 3.2.4.

Historical Directional Samples

To appropriately alter the directions of SRKW velocities simulated from the O-U process, we need to make use of observations of directional tendency from historical data as references. We assume the movement directions of SRKWs are influenced by both the directional memory and the location of the whale. During the simulation, we collect directional samples

from the split trajectories of the Orca Master dataset given the current directional memory and location.

At any given time in the simulation, we define the current location as (x_i, y_i) and the directional memory as $\theta_{\ell i}$, where i is the current time step. There are N sighting locations (x_j, y_j) corresponding to moving directions $\theta_j,\ j = 1, ..., N$, that can be obtained from the Orca Master dataset, We include θ_j's in the directional sample Θ_i with two conditions:

1. $x_i - 2.5 \text{ km} \leq x_j \leq x_i + 2.5 \text{ km}$ and $y_i - 2.5 \text{ km} \leq y_j \leq y_i + 2.5 \text{ km}$
2. $\theta_{\ell i} - \pi/2 \leq \theta_j \leq \theta_{\ell i} + \pi/2$.

The first condition means that we use a 5 km $\times$ 5 km square with its centroid at the current location (x_i, y_i). This approach allows us to include other directional observations nearby in the sample because it is unlikely to have directional observations precisely at the current location. The second condition controls the permissible range of directions that can be used. We assume SRKWs do not make sudden large-angle turns inconsistent with their current direction of travel, and so only include historical directions whose difference with the directional memory is no greater than $\pi/2$, or 90°. It is worth noting that in the simulation, the directional memory is allowed to change as time progresses because it only traces back over the past 9 minutes.

After gathering the relevant historical directional samples, we can alter the directions of the velocities simulated from the O-U process as follows. We assume the movement direction given a location and directional memory will follow a Von Mises distribution which is known as circular normal distribution. Then, we randomly draw a direction from this distribution and use it as the next moving direction. However, historical directional samples are sometimes not available at more remote locations due to a lack of observations. Therefore, we modify the velocities by making use of the directional memory, which will be further explained in Section 3.2.4.

3.2.3 Von Mises Distribution

The ith bivariate velocity simulated from the O-U process can be written as

$$\mathbf{v_i} = v_i \begin{bmatrix} \cos\theta_i \\ \sin\theta_i \end{bmatrix}, 0 \leq \theta_i < 2\pi, \tag{3.9}$$

where v_i is the speed, and θ_i is the angle. By changing θ_i, the velocity can be projected in another direction under a Von Mises distribution, but maintain the same speed.

The probability density function of the Von Mises distribution given the current location (x_i, y_i) and directional memory $\theta_{\ell i}$ is

$$f(\theta_m | x_i, y_i, \theta_{\ell i}) = \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(\theta_m - \tau_i)}, \tag{3.10}$$

$$I_0(\kappa_i) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa_i \cos u} du,$$

where τ_i is the mean direction, κ_i is the concentration, as the variance in a normal distribution, at the current location, $I_0(\kappa_i)$ is the modified Bessel function of order 0, and $0 \leq \theta_m < 2\pi$. With the sample of historical directions Θ_i, we can estimate μ_i and κ_i by MLE. The MLE of those two parameters can be found in [16], and R provides the function, 'mle.vonmises' in the circular package [1]. By generating a random number from the von Mises distribution, we can then project the direction θ_i of the velocity simulated from the O-U process in θ_m, a moving direction more likely to happen based on historical information.

3.2.4 Blending Principles

The ideal situation is that we can project all velocities in realistic directions using the von Mises distribution conditioned on different locations and previous directional movements. However, the sighting data of SRKW are often too sparse to cover all locations within the Salish Sea due in part to lack of observer effort, and also in part due to low probability of SRKW occurrence. Therefore, there are no historical directional observations at some locations. Even though there are no directional observations, we still, however, want to maintain directional consistency. Therefore, we consider projecting the moving direction

as the average between the directional memory and the direction simulated from the O-U process. In this way, the difference between the current direction and the directional memory is still no greater than $\pi/2$. Specifically, we define the blended direction for these two cases as

$$\theta_i' = \begin{cases} \theta_m \sim von\ Mises(\tau_i, \kappa_i \mid x_i, y_i, \theta_{\ell i}) & \text{if historical directional observations are available} \\ (\theta_{\ell i} + \theta_i)/2 & \text{otherwise} \end{cases} \tag{3.11}$$

After projection, the velocity becomes:

$$\mathbf{v_i'} = v_i \begin{bmatrix} \cos\theta_i' \\ \sin\theta_i' \end{bmatrix},\ 0 \leq \theta_i' < 2\pi, \tag{3.12}$$

where v_i is the same speed in Equation (3.9). Equation (3.12) then replaces (3.9), and stochastic simulation can continue following the O-U process.

3.3 Coastal Avoidance Algorithm

When simulating animal trajectories, the simulated trajectories must stay within a valid area of interest. For birds or some terrestrial animals, the geometry of boundaries is usually simple such as a polygon, or a bounding box delineating a habitat type. For marine mammals, areas of interest must fall within the ocean, and the shapes of coastal boundaries may be highly irregular. In our case, the geometry of the Salish Sea is complex, with numerous large and small islands distributed throughout as depicted in Figure 1.1. During the simulation of SRKW trajectories, the algorithm must be adapted to prevent simulated whale trajectories from crossing land, or entering shallow water. We do this by further adjusting the movement direction in Equation (3.12) if the current simulated velocity from the O-U velocity process and direction-blending pushes the trajectory onto land or shallow water ($<$ 2 m deep).

We consider a simple algorithm to maintain trajectories to be in water and to avoid coastlines. The idea behind the approach is to find the direction aligning best with the current directional memory, but making the trajectory stay in water. In the simulation,

we make use of the following quantities: the current location $\mathbf{x_i} = (x_i, y_i)$, the velocity $\mathbf{v_i} = (v_{xi}, v_{yi})$ from the O-U process and the blending method, the directional memory $\theta_{\ell i}$, the time step Δt, and the next predicted destination $\mathbf{x_{i+1}} = (x_{i+1}, y_{i+1})$ derived from $\mathbf{x_i} + \mathbf{v_i}\Delta t$. When there is land or shallow water obstructing the route from $\mathbf{x_i}$ to $\mathbf{x_{i+1}}$, the algorithm helps to directly adjust the next destination and force the whale trajectory to remain in water while keeping the moving direction consistent with the directional memory as much as possible. Finally, the entire route is constrained to waters > 2 metres deep.

The procedures underlying the algorithm are shown in Figure 3.1. As shown in Figure 3.1a, there is a direction memory (black arrow) associated with the current location in the simulation. The current moving direction (red arrow) pushes the next destination onto land. We then adjust the next destination in three steps:

1. We use the distance between the current location and the original next destination as the radius r (the red arrow in Figure 3.1a) to generate 40 candidates $\mathbf{Z}$, covering different angles.

$$\mathbf{Z} = \mathbf{z_1}, ..., \mathbf{z_k}$$

$$\mathbf{z_k} = \mathbf{x_i} + r \begin{bmatrix} \cos\theta_k \\ \sin\theta_k \end{bmatrix}$$

$$\theta_k = \frac{\pi}{20}k, k = 1, ..., 40$$

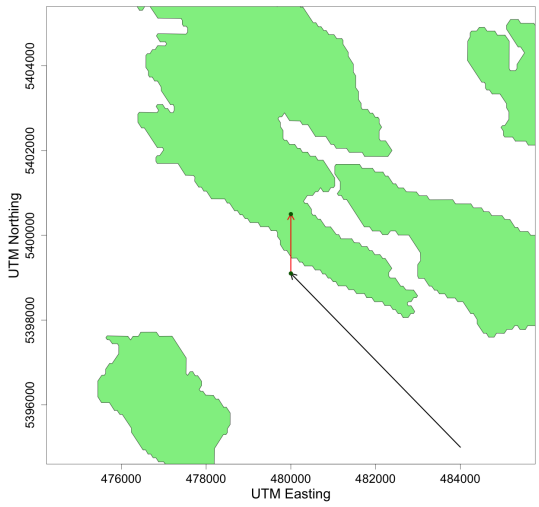
   There should be no land or shallow water (less than 2 m depth) between the current location and candidates. For some angles, the radius allows some candidates to occur on land, and thus, we put those candidates along with the same angles back in the water but adjacent to the coastline. The candidates are shown in Figure 3.1b.

2. Among the 40 candidates, we take the median of their distances between the current location $\mathbf{x_i}$ as a threshold. Then, we discard the candidates whose distances from the current location are under the threshold. The 20 filtered candidates $\{\mathbf{z_k}', \theta_k'\}$ are shown in Figure 3.1c.

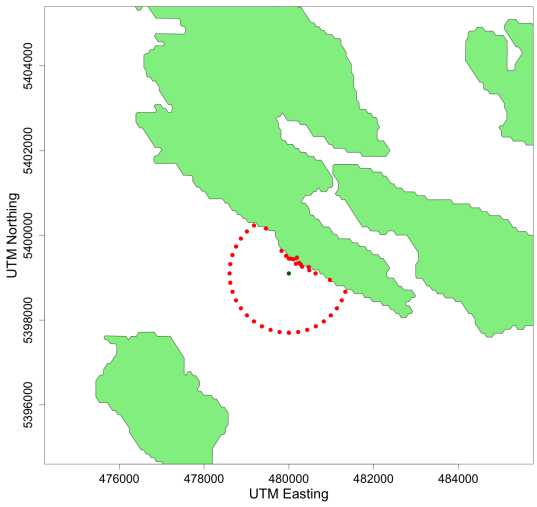
3. Of the remaining filtered candidates, we pick the one whose moving direction is closest to the directional memory as the next destination.

$$max\left\{cos\Big(|\theta'_k - \theta_{\ell i}|\Big)\right\}$$

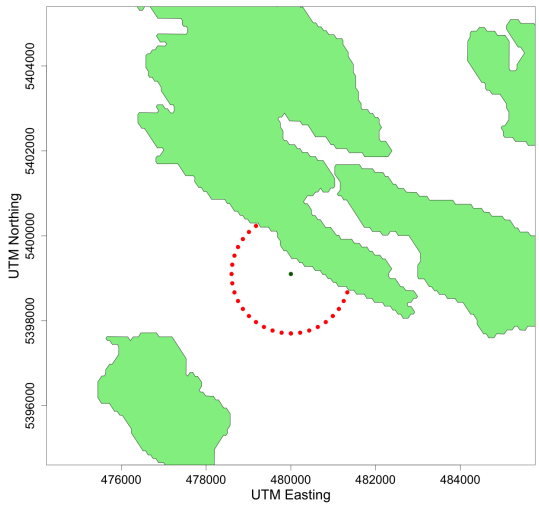
By following the steps of the algorithm above, we not only avoid pushing trajectories onto land or shallow water but also ensure trajectories are close to the distance to the original destination, thereby reducing the chance of trapping trajectories by the coastline.



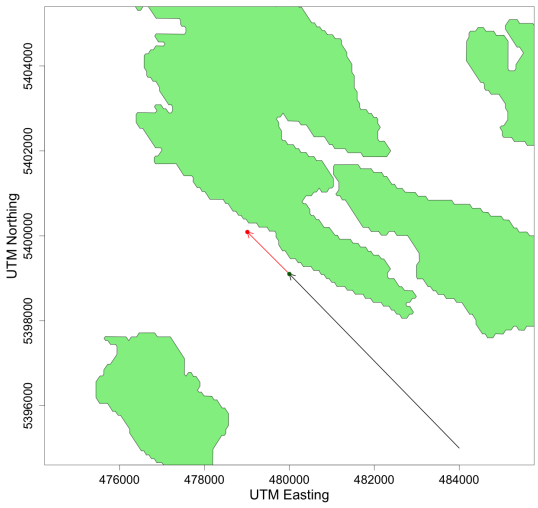
(a) The moving direction to the next destination (red arrow) and current directional memory (black arrow)



(b) Candidate destinations **Z**



(c) Candidate destinations with enough displacement



(d) Final destination

Figure 3.1: Steps in the coastal avoidance algorithm

3.4 Simulation Procedures

The goal of the simulation is to forecast possible travelling areas of SRKW pods given that a pod or pod leader is observed at a specific location. Given an initial condition (position and velocity), we simulate 1000 whale pods, make them travel by following the O-U velocity process with direction-blending and using the coastal avoidance algorithm, and finally draw simulated trajectories. Each simulated trajectory is a realization, representing a possible scenario under the initial condition and the proposed models. Using many simulated trajectories, we build all possible travelling areas under all scenarios, which is also termed an ensemble. Using these ensembles of realizations, we can build probability forecasts of possible travelling areas based on the simulated trajectories.

The procedures for simulating SRKW trajectories can be summarized as follows.

1. Simulate an SRKW pod trajectory starting from generating a whale pod with initial conditions, including initial location $\mathbf{x_0} = (x_0, y_0)$, initial velocity $\mathbf{v_0} = (v_{0x}, v_{0y})$.

2. The velocity of the simulated pod follows the O-U velocity process under Euler numerical approximation in Equation (3.4), which is an AR(1) process. We model the SRKW pod velocity forward in time by generating $\mathbf{v_i}$, the i$^{\text{th}}$ velocity of the whale pod, with a constant time step $\Delta t = 0.05$ hours (3 minutes) from the following 2-D AR(1) process.

$$v_{ix} = (1 - \hat{\alpha}\Delta t)v_{i-1x} + z_{ix}$$
$$v_{iy} = (1 - \hat{\alpha}\Delta t)v_{i-1y} + z_{iy}$$
$$z_{ix}, z_{iy} \sim N(0, \hat{\sigma}^2\Delta t)$$

3. The pod velocity simulated from the O-U process (the AR(1) process) would steer the trajectory to locations where SRKW rarely transit. Therefore, we modify $\mathbf{v_i}$ by the direction-blending method as Equation (3.11) and (3.12) to steer the simulated pod to locations aligning with historical observations.

4. The pod trajectory drawn from simulated trajectories must stay in water (and not intersect land). Therefore, when any piece of the simulated pod trajectory is steered toward coastlines, the simulated velocity from the O-U process with direction-blending should be adjusted by the coastal avoidance algorithm in Section 3.3. In the simulation, when the route from $\mathbf{x_i}$ to the destination $\mathbf{x_{i+1}}$ passes land or shallow water, we again modify $\mathbf{v_i}$ using the direction selected by the coastal avoidance algorithm.

5. Repeat Steps 2-4 until reaching the end of the desired simulation period. The simulated pod location $\mathbf{x_i}$ along with the trajectory can be drawn by

$$\mathbf{x_i} = \mathbf{x_0} + \sum_{j=1}^{i-1} \mathbf{v_j} \Delta t$$

This algorithm will be applied in the following chapter for SRKW trajectory forecasting.

Chapter 4

Results

In this chapter, we will present the outcomes of implementing the methods introduced in Chapter 3. Initially, we carry out parameter estimation for the O-U velocity process discretized via the Euler approximation. Subsequently, we will demonstrate how the direction-blending method is able to simulate realistic trajectories that conform to probable pathways. Finally, we will show the results of forecasting experiments simulating three trajectories using specified initial conditions and observed pathways.

4.1 Parameter Estimation

We used the split trajectories obtained from the Orca Master dataset using the approach described in Subsection 3.1.2 to estimate α and σ in Equation (3.4). Using the estimate $\hat{\alpha}$, obtained from Equation (3.5) the parameters ϕ and η of the O-U process can be obtained. The estimate of the persistence parameter ϕ was calculated as $\hat{\phi}_{\Delta t} = 1 - \hat{\alpha}\Delta t$, and the estimate of the standard deviation of η is $\hat{\sigma}\sqrt{\Delta t}$.

The numerical values for these parameters are shown in Table 4.1. The estimated speed of mean reversion, $\hat{\alpha}$, is 2.48 units/h, and the estimated magnitude of the stochasticity or variability, $\hat{\sigma}$, is 26.77 km/h$^{3/2}$. Based on these results, the estimates of persistence parameter, $\hat{\phi}$, and standard deviation, $\hat{\eta}$, obtained from $\hat{\alpha}$ and used in Equation 3.5 to simulate the trajectories of SRKWs with a time step 0.05 h, are 0.876 and 6 km/h. The high auto-correlation coefficient implies the velocities of SRKWs are persistent over a 3-min time step, and the persistence decreases as the size of time steps increases. If a chosen time step is larger than 24 min (0.4 h), ϕ will decay to nil, which means two velocities with an

interval greater than 24 min can be considered independent. Because the sample size, or the number of available split trajectories is large (1266), the standard error (SE) of $\hat{\alpha}$ is quite small, and thus, the 95% confidence interval (CI) is narrow. The standard deviation of the velocity for the time step of 0.05 h appears quite realistic and shows SRKWs do not accelerate or slow down too much within 3 minutes.

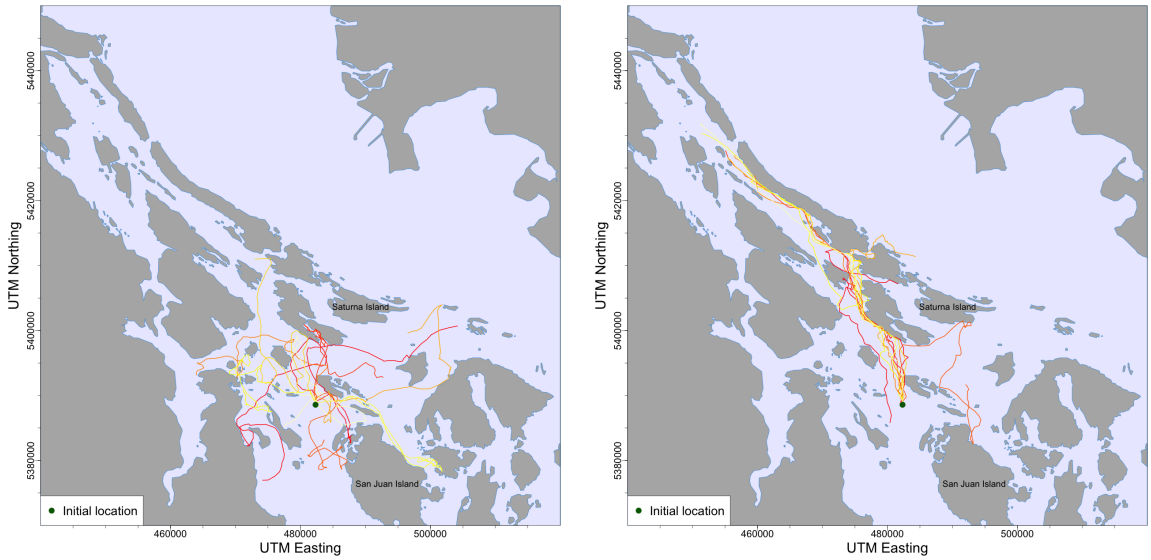
Table 4.1: Estimation of the parameters in Equation (3.3) and (3.5)

	α	$\phi_{0.05}$	σ	$\sigma_{0.05}$
Estimate	2.48	0.876	26.77	6.00
SE	0.072	0.004	-	-
Upper 95% CI	2.62	0.883	-	-
Lower 95% CI	2.34	0.869	-	-

4.2 Direction-blending

To illustrate the influence of direction-blending on the simulation of SRKW movement, we simulated ten trajectories (realizations) using the same initial location and with the same initial velocity (10 km/h north) for 4 hours with and without direction blending. Figure 4.1 shows the simulation results of the ten trajectories for each case. The dark green dots and coloured lines in both plots indicate the initial location and simulated trajectories, respectively. The trajectories simulated purely from the O-U velocity model without direction-blending turned radiated the initial location, reversed direction relative to the initial velocity, and hence did not explore the relevant spatial region in a realistic manner. In contrast, the trajectories simulated from the O-U velocity model and adjusted by the direction-blending were smooth, expanded to cover relevant regions in the Salish Sea, and maintained a degree of consistency; all in all they are more like the movement of SRKW in Figure 2.4.

As mentioned in Section 3.2, velocities generated from the raw O-U velocity model cannot guide trajectories to specific locations and along specific pathways. By direction-blending, we maintained consistency and adjusted the angles of the velocities in more re-



(a) Trajectories simulated without direction-blending

(b) Trajectories simulated with direction-blending

Figure 4.1: Comparison of 10 trajectories with the same initial conditions. The left panel makes use of a correlated random walk movement model without direction blending, while the right panel includes direction-blending in the movement model. The right panel demonstrates how the direction-blending method is able to simulate realistic trajectories that conform to historic trajectories 2.4.

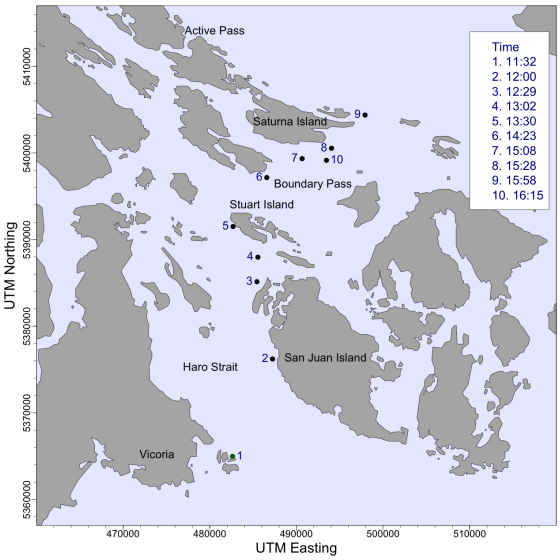
alistic directions based on historical records. The SRKW forecasts in the simulation then can explore the regions where SRKW were present, aligning with Figure 2.2. In addition, even though we maintained a degree of consistency in direction-blending, it still allowed simulated whale pods to travel back or make a large-angle turn, as SRKWs are observed to do in the Salish Sea [40]. Following Section 3.2, we only used the moving direction within 9 min as a directional memory to collect historical directional information to form a sample of possible directions, and thus SRKW trajectories did not maintain the same directional memory from the initial location to the end of the simulation. This allows simulated SRKW to travel in directions opposite to the directions at the beginning of the trajectory, but the change comes about in a slow and smooth way. Overall, adjusting moving directions by historical records provides a bridge between velocities simulated by stochastic processes and realistic trajectories.

4.3 Trajectory Forecasting

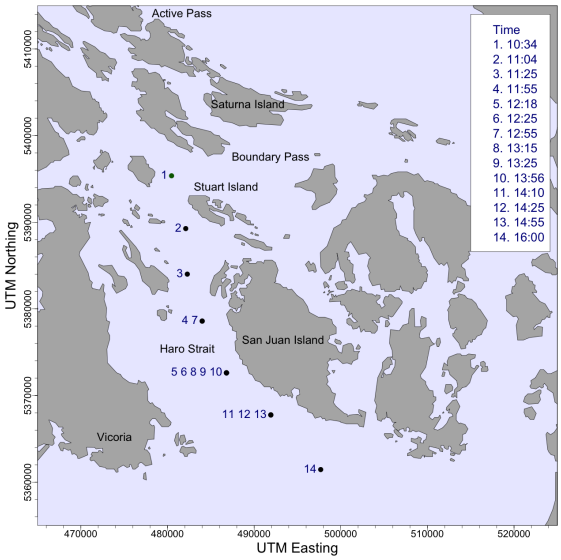
Following the approach described in Section 3.4, we aimed to forecast the movement of SRKWs by simulating the trajectories of SRKWs with only the information on initial conditions. We selected three historical observed trajectories to validate our simulation and forecast methods. These trajectories are independent of the trajectories we used to build our model. As shown in Figure 4.2, the dark green dots in all plots are the initial locations. We also calculated the initial velocity from the data as the distance between the initial and the second observed locations divided by the time interval. In Figure 4.2a, the first trajectory was observed in July 2012, starting around Vancouver Island, moving north to Stuart Island, and entering Boundary Pass. In Figure 4.2b, the second trajectory was observed in August 2016, starting around Stuart Island and moving south around San Juan Island. The last trajectory in Figure 4.2c was a reverse trajectory observed in July 2012, starting around the north of Pender Island, moving south first, and finally turning north to Active Pass.

Figures 4.3-4.5 show the forecasted locations for the SRKW for various time horizons for the three test cases. These are expressed as the envelope of 90% probabilities, generated from the kernel-density estimated forecast probability density functions using the ensemble of forecast trajectories.

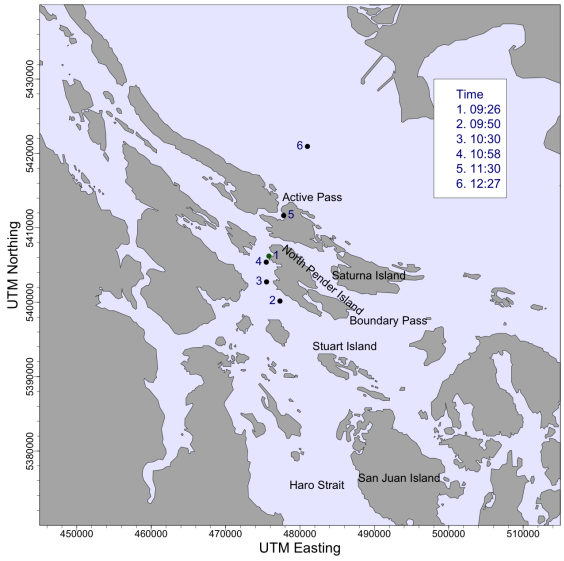
In Figure 4.3 our forecasts missed the observation at 0.47 h, captured observations from 0.95 h to 2.85 h, and missed again at 3.6 h. The plausible forecast region increased as the forecast time increased. Beyond 2.85 h, the forecast regions already covered a large portion of the Salish Sea. In the first 1.97 h, the entire forecast regions moved along with the observations. However, at 1.97 h, the forecast showed a trend of the second peak that SRKWs might transit to the south instead of sticking around Stuart Island. At 2.85 h and 3.6 h, there were two obvious peaks in the probability density, one was around Active Pass, and the other was around San Juan Island. These forecast results implied direction-blending worked in the model to allow SRKWs to reverse in the opposite direction based on historical observations. However, the forecasts showed a higher chance that SRKWs would transit through Active Pass, and SRKWs finally transited through Boundary Pass. In reality,



(a) Trajectory from Haro Strait to Boundary Pass in July 2012. Note the trajectory does a back step in Boundary Pass.



(b) Trajectory through Haro Strait in August 2016. Note the trajectory does a back step mid-sequence.



(c) Trajectory through Active Pass in July 2012.

Figure 4.2: Three historical data trajectories used to evaluate forecasting performance.

Stuart Island is an important turning point where SRKWs reaching from the south choose to transit to either Active Pass or Boundary Pass. That difference between the forecasts and observations may imply there were more observations transiting from Stuart Island to Active Pass than Boundary Pass. It also implies forecasts starting from the south can only

follow long-term directional probabilities around the turning point to select Active Pass or Boundary Pass as the final destination.

The forecast results of the second trajectory in 2016 around San Juan Island are shown in Figure 4.4. Compared to the forecasts of the trajectory in 2012, the forecast regions covered the observations well from the beginning to 4.35 h. The forecast regions were also growing as forecast time increased. After 2.85 h, the forecast regions were stable. There were also two peaks in the probability density showing up after 2.85 h. One was around Active Pass, and the other was around San Juan Island. It is worth noting that the forecast regions for the two target trajectories both had similar ranges and two peaks around Active Pass and San Juan Island after 2.85 h. However, they displayed varying probability densities for the peaks. Similar forecast ranges imply that our forecasts consider most possible directions that SRKWs can transit through, ending up covering a large portion of the Salish Sea.

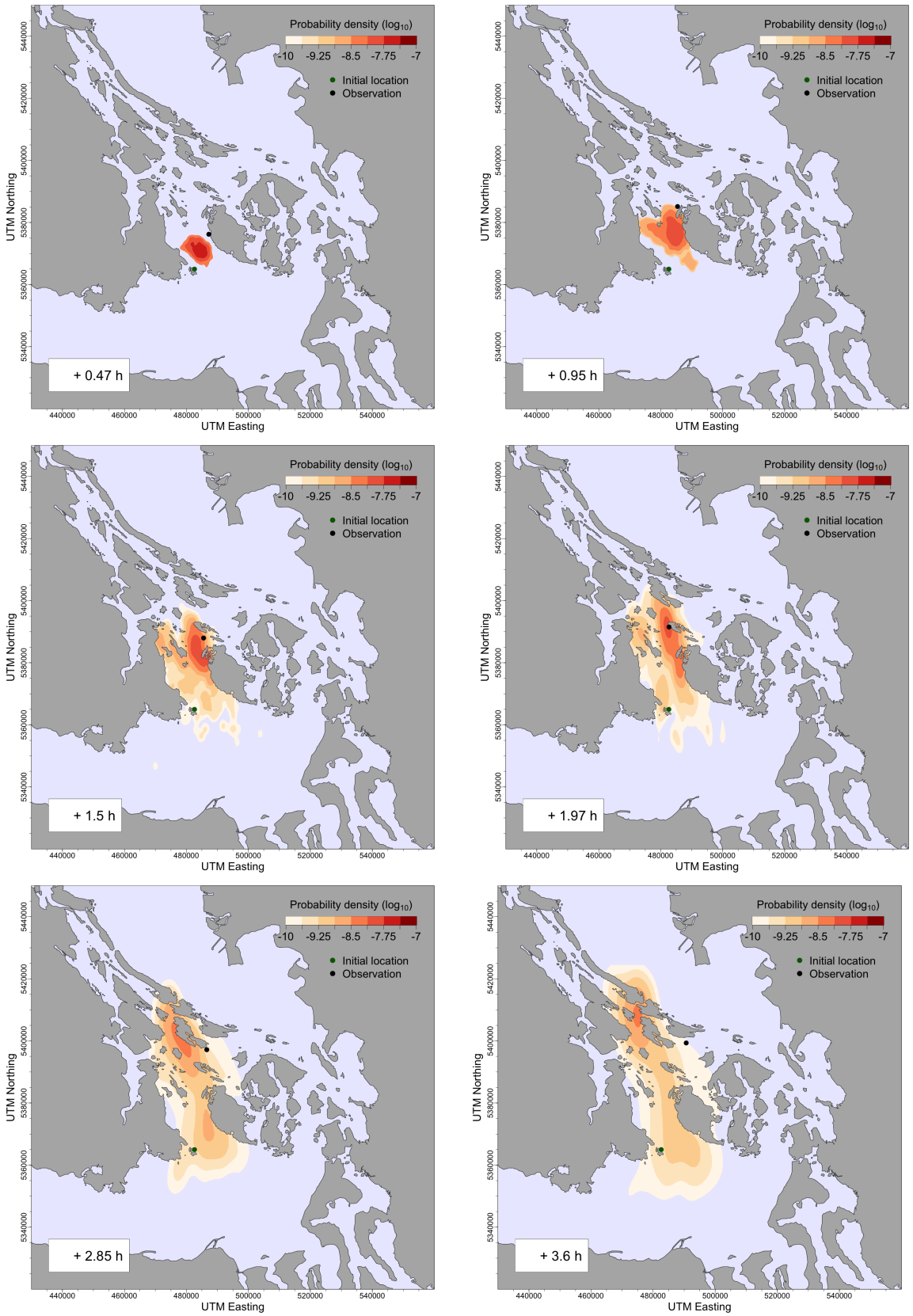


Figure 4.3: The results of a forecasting experiment showing probability forecasts in Haro Strait and Boundary Pass. Units of time are discretized to align with Orca Master whale observations.

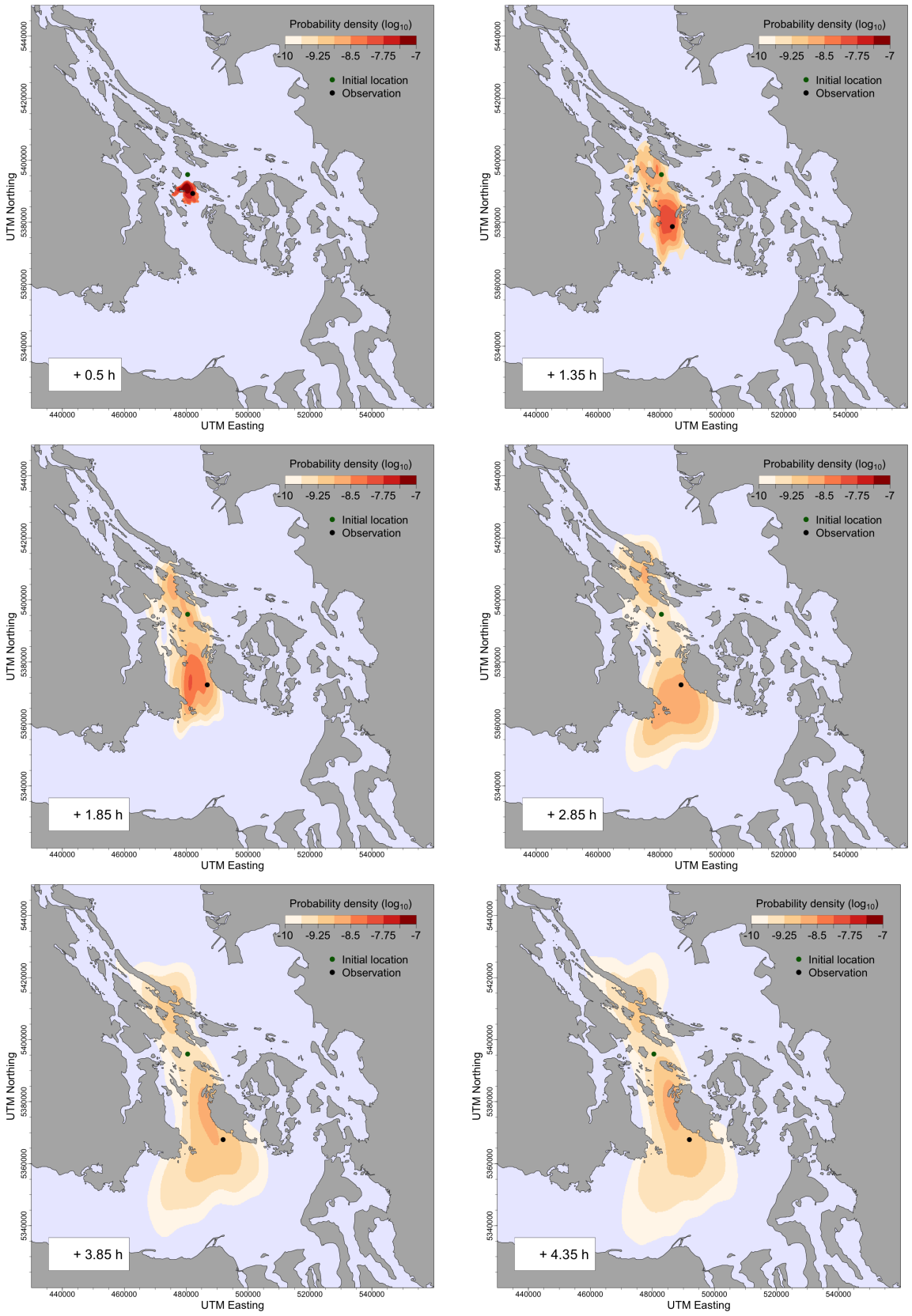


Figure 4.4: The results of a forecasting experiment showing probability forecasts in Haro Strait and Active Pass. Units of time are discretized to align with Orca Master whale observations.

Figure 4.5 shows the forecast regions of the reverse trajectory around Active Pass in 2012. The initial velocity at the initial location was south, and the forecast region covered the observation well and was small at 0.4 h. At 1.07 h, the forecast started spreading out, considering a strong tendency to continue to transit south and possibility to reverse northward to Active Pass. The observation at 1.07 h indeed reversed in the opposite direction to the north and kept heading to Active Pass in the next few hours. The forecast regions still covered the observations even though the moving directions were totally different from the initial velocity. Therefore, although our forecast regions were large after the forecast time reached 3 h, those simulated whale pods in the simulation explored most possible scenarios of the movement of SRKWs, including the possibility of reversing.

Boxplots in Figure 4.6 show the distribution of direct position errors (in km) for the three historical trajectories in Figure 4.3, 4.4 and 4.5. Direct position error is defined as the distances between simulated whale locations and observations matched to the observation time. The direct position errors generally grew as forecast time increased, and trajectories diverged from those observed. The median direct position errors generally remained under 20 km for forecasts out to 3 hours. The ranges of direct position errors were also generally stable after the forecast time reached 3 hours. This aligns with the results we mentioned before, wherein forecast envelopes were generally stable (and large) after 3 hours. Combining the median direct position errors with forecast ranges, our forecasts can be said to work up to 3 hours for different initial conditions in the Salish Sea.

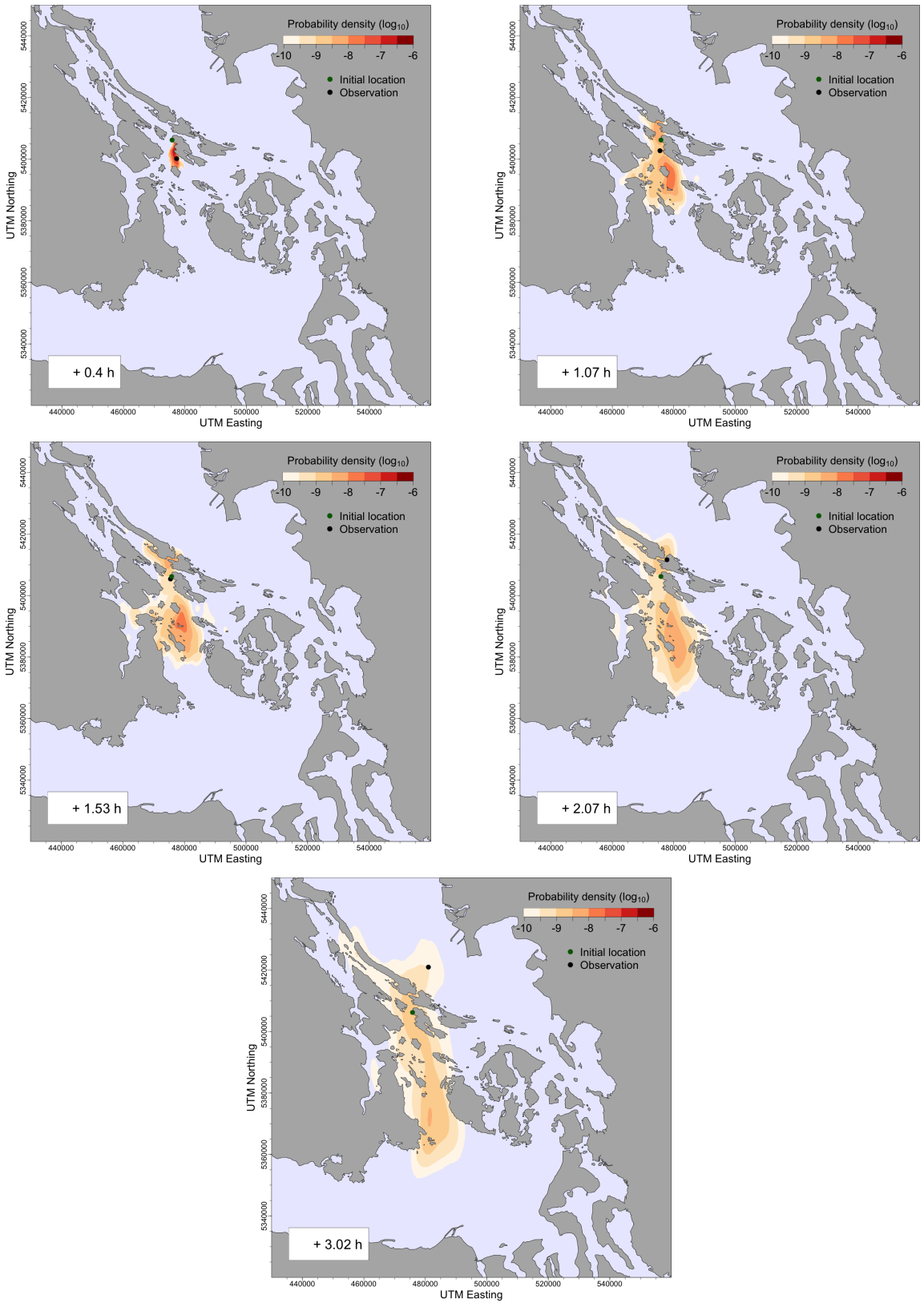
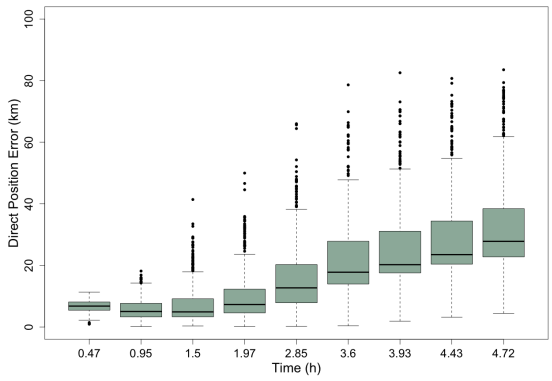
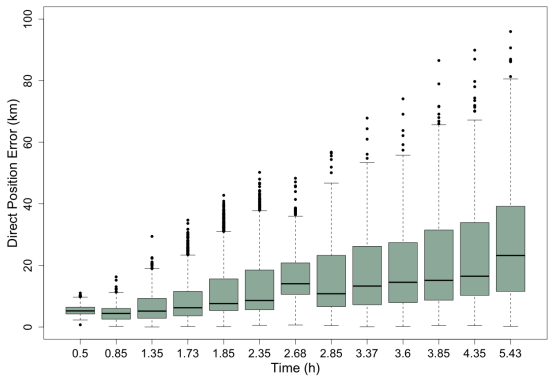


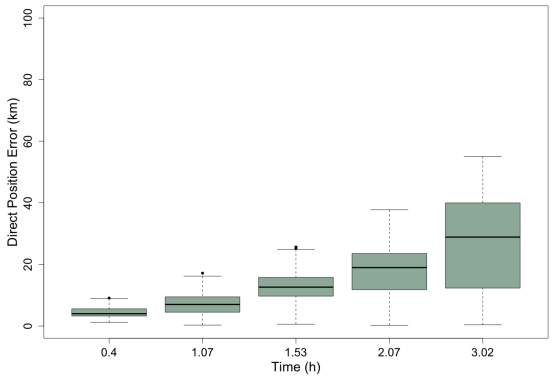
Figure 4.5: The results of a forecasting experiment showing probability forecasts through Active Pass and Haro Strait. Units of time are discretized based on Orca Master whale observations. Probability forecast around Active Pass



(a) Trajectory from Haro Strait to Boundary Pass in July 2012



(b) Trajectory through Haro Strait in August 2016.



(c) Trajectory through Active Pass in July 2012

Figure 4.6: Boxplots showing the distribution of direct position error of the forecasts of the three chosen trajectories

Chapter 5

Discussion

In this study, we developed a forecasting framework for the movement of SRKWs for the purpose of providing early warning alerts to mitigate collisions with commercial vessels and reduce noise exposure. To do this we modified a basic animal movement model in the form of a continuous-time velocity-based Orstein-Uhlenbeck (O-U) process. Our central modification was to incorporate historical pathway information into the stochastic simulation, thereby making SRKW trajectories more realistic and improving forecast skills. The dataset we used to do this was comprised of archived sighting data from the Orca Master dataset that was analyzed and processed to yield reasonable and realistic historical whale trajectories. A direction blending algorithm provided the means to incorporate these historical movement directions into the O-U velocity process. A coastline avoidance algorithm was also included. The result is an SRKW movement model that, given initial position and velocity, is able to successfully simulate SRKW trajectories.

Both Johnson et al. [17] and Gurarie et al. [11] suggested that using continuous-time models to model animal movement is wiser than using discrete-time models because it better matches irregularly sampled data, and most importantly aligns with actual animal movement which is a continuous process. We support this viewpoint, especially when dealing with irregularly sampled data. Given that the sighting records of SRKWs are discrete and opportunistically sampled, the continuous time model allowed us to directly fit our raw velocity observations to the model without further processing. The time-invariant parameters also allowed us to model the velocities and trajectories of SRKWs at any chosen time scale.

Therefore, continuous-time models are able to utilize the full information of data and have more flexibility compared to discrete-time models.

In continuous-time models, We chose the independent two-dimensional O-U velocity process to explore the statistical regularity hidden behind the irregular sighting data. By transforming the O-U velocity process into the AR(1) process through discretization, the results showed the velocity of SRKWs is highly persistent ($\hat{\phi} = 0.876$) within a single time step of 3 minutes, also implying that system memory is erased in about 24 minutes. The standard deviation of the velocity within a single time step of 3 minutes is 6 km/h, which is much larger than the 1 km/h that Randon et al. [32] assumed in their study for the SRKW. Therefore, properly estimating the parameters by the data across the area of interest is important.

We then simulated the velocities of a pod of SRKWs, or a pod leader, from the O-U velocity process. However, our results showed trajectories determined from the simulated velocities from the O-U velocity process were not close to realistic SRKW movement patterns. The realizations had far too much directional spread. Therefore, we proposed the direction-blending method to project the velocities in more probable directions based on the assumed directional memory of 9 minutes (3 time steps) and historical directional information. This approach chose the direction by generating a random number from a von Mises distribution as dictated by the current directional memory and historical directional information, or the average of directional memory and the direction of the velocity simulated from the O-U process if the historical directional sample is unavailable.

We suggest that considering directional persistence and preference in trajectory simulation is important for animal movement modelling, especially for the velocity-based O-U process. When modelling the two-dimensional O-U velocity process, assuming independence between two coordinates is natural so that unwanted rotational movement can be prevented. In addition, potential functions that govern preferable destinations are not easy to be used with the O-U velocity process. These settings and limitations do not take into account directional persistence and preference of animal movement and thus, we recommend animal movement models should be adjusted by using this directional information. There have

been a few studies focusing on applying directional persistence and circular distributions to animal movement such as Duchesne et al. [7], Nicosia et al. [24], and Mastrantonio [20]. Nevertheless, most animal movement studies applying the O-U velocity process did not consider directional preference (Gurarie et al. [12], Johnson et al. [17], Mcclintock et al. [21]). Velocities simulated from pure continuous-time correlated random-walk models may steer trajectories to places that animals rarely visit. Therefore, adjusting the velocities toward preferable positions such as our approach, provides a variable alternative to potential functions.

Three historical trajectories were chosen to illustrate our simulation and forecast framework. Clearly, this is not enough to definitely validate our movement model, but serves to hint at its efficacy. From the simulated trajectories, we generated probability density functions, covering 90% probability of being within a region, as estimated by kernel density to indicate regions where SRKW may be found in the next few hours. For all three cases, our forecast regions covered the observations well. Although the forecast regions enlarged as the forecast time increased and ended up reaching stable long-term average patterns, the possibility of SRKWs reversing was considered, and the median direct position errors were generally under 20 km in the first 3 hours. Overall, our forecasts are valid to provide the ranges of SRKWs transiting up to 3 hours and thus, able to improve upon currently implemented conservation measures for vessel-whale interaction.

Joy et al. (2019) indicated that commercial vessels transiting with a speed under 11 knots could significantly reduce lost foraging time for SRKWs [18]. Leaper (2019) also indicated slower vessel speeds are beneficial to reduce collision risk with whales [19]. Transport Canada has implemented two speed restricted zones (SRZs) in the Salish Sea near Swiftsure Bank and 2 interim sanctuary zones (ISZs) between Pender Island and Saturna Island in effect from June to November to protect killer whales. All vessels are required to stay under a speed of 10 knots in the speed restricted zones, and no vessels are allowed to enter the interim sanctuary zones except local groups with an exemption. In addition, all vessels must keep a distance of at least 400 meters and 300 yards within the border of Canada and the USA, respectively, from all killer whales in the Salish Sea. Our movement model can also

incorporate recorded vessel tracks to simulate and assess the cumulative sound exposure level (cSEL) for SRKWs. In fact, we are now assisting Transport Canada in assessing the necessity of extending the ISZ around Saturna Island to Tumbo Channel by calculating the cSEL that SRKWs would receive when they transit through Tumbo Channel, which relies on the simulated trajectories of SRKWs from our model.

The SRZs and the ISZs are only a small portion of the Salish Sea. However, almost the whole Salish Sea is federally designated as 'critical habitat' of SRKWs in both Canada [8] and the USA [28]. If speed-restricted measures could be implemented across the Salish Sea, the impact of noise for SRKWs would be substantially reduced. However, the speed-restricted measures in place across the Salish Sea also significantly influence economic activities. To reach a balance between conservation objectives and economic cost, a dynamic structure can be developed to implement speed-restricted measures based on our forecast scheme. Given an initial condition, we can keep forecasting and updating probable areas where SRKWs would transit and advise commercial vessels in the forecast areas to slow their speeds. The procedures could be adjusted dynamically. For example, once a citizen scientist or a hydrophone detects SRKWs in the Salish Sea, we can send a slow-down warning to commercial vessels that overlap with the whale movement forecast region of the first few hours. Then, we can update the forecast region and the slow-down warning to the vessels in the following hours using the initial conditions of the most recent sighing, or if new observations become available, the forecast region can be updated in real-time. There should be a variety of strategies to utilize the forecast information. In addition, because our forecast regions include most areas that SRKWs may transit, our forecast regions may be seen as the largest region where commercial vessels should decrease their speeds. If speed-restricted orders are assigned temporarily, instead of all the time, any potential economic loss from slowing the vessel will also be mitigated.

Currently, our forecasts solely rely on initial whale location and velocity. In fact, SRKWs are highly likely to be tracked continuously because of the large amount of observation effort invested by citizen scientists and research groups in the Salish Sea. Therefore, it is worth introducing data assimilation, such as in Randon et al. [32] as the forecasting framework.

Our improved SRKW movement model can be directly incorporated within that framework and will improve predictive skills. We are also interested in exploring how to adjust the estimates in the O-U process and update direction-blending information once new observations emerge. Our direction-blending scheme would provide useful directional information to improve the process of data assimilation for more accurate forecasts. In addition, exploring the relationship between the movement of SRKWs and the behavioural state of SRKWs and how it links to environmental or physical factors such as tidal state, currents, sea temperature, and salinity, may also help improve the forecasting framework and contribute information that would narrow down the forecast region. In addition, our approach is transferable to different animals and different data sources such as satellites, hydrophones, and other telemetry data. As long as we have initial location and velocity and historical velocity data, we can forecast the trajectories of a specific animal. We anticipate making our forecasting framework more comprehensive under different scenarios and applications in the real world.

Bibliography

[1] C. Agostinelli and U. Lund. *R package* `circular`*: Circular Statistics (version 0.4-95)*. CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA, 2022.

[2] Richard Anderson-Sprecher and Johannes Ledolter. State-space analysis of wildlife telemetry data. *Journal of the American Statistical Association*, 86(415):596–602, 1991.

[3] David Brillinger, Haiganoush Preisler, Alan Ager, John Kie, and Usda Service. The use of potential functions in modelling animal movement. 05 2001.

[4] David R Brillinger, Haiganoush K Preisler, Alan A Ager, and JG Kie. The use of potential functions in modelling animal movement. *Selected Works of David Brillinger*, pages 385–409, 2012.

[5] David R. Brillinger and Brent S. Stewart. Elephant-seal movements: Modelling migration. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 26(3):431–443, 1998.

[6] Stéphane Dray, Manuela Royer-Carenzi, and Clément Calenge. The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*, 25(3):673–681, 2010.

[7] Thierry Duchesne, Daniel Fortin, and Louis-Paul Rivest. Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. *PLOS ONE*, 10(4):1–12, 04 2015.

[8] Fisheries and Oceans Canada. Recovery strategy for the northern and southern resident killer whales (orcinus orca) in canada [proposed]. species at risk act recovery strategy series, fisheries and oceans canada, 2018.

[9] Christen H. Fleming, Justin M. Calabrese, Thomas Mueller, Kirk A. Olson, Peter Leimgruber, and William F. Fagan. Non-markovian maximum likelihood estimation of autocorrelated movement processes. *Methods in Ecology and Evolution*, 5(5):462–472, 2014.

[10] Georgia Strait Alliance. Vessel traffic. `https://georgiastrait.org/issues/vessel-traffic/`.

[11] Eliezer Gurarie, Christen Fleming, William Fagan, Kristin Laidre, Jesús Hernández-Pliego, and Otso Ovaskainen. Correlated velocity models as a fundamental unit of animal movement: Synthesis and applications. *Movement Ecology*, 5, 05 2017.

[12] Eliezer Gurarie, Daniel Gruenbaum, and Michael T. Nishizaki. Estimating 3d movements from 2d observations using a continuous model of helical swimming. *Bulletin of Mathematical Biology*, 73(6):1358–1377, 2011.

[13] Marla Holt. *Sound Exposure and Southern Resident Killer Whales (Orcinus orca): A Review of Current Knowledge and Data Gaps*. 02 2008.

[14] Mevin B. Hooten, Devin S. Johnson, Brett T. McClintock, and Juan M. Morales. *Animal Movement: Statistical Models for Telemetry Data*. Taylor and Francis Group, LLC, 1 edition, 2017.

[15] Stefano M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples (Springer Series in Statistics)*. Springer Publishing Company, Incorporated, 1 edition, 2008.

[16] S.R. Jammalamadaka, A. Sengupta, and A. Sengupta. *Topics in Circular Statistics*. Series on multivariate analysis. World Scientific, 2001.

[17] Devin Johnson, Josh London, Mary-Anne Lea, and John Durban. Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89:1208–15, 06 2008.

[18] Ruth Joy, Dominic Tollit, Jason Wood, Alexander MacGillivray, Zizheng Li, Krista Trounce, and Orla Robinson. Potential benefits of vessel slowdowns on endangered southern resident killer whales. *Frontiers in Marine Science*, 6:344, 2019.

[19] Russell Leaper. The role of slower vessel speeds in reducing greenhouse gas emissions, underwater noise and collision risk to whales. *Frontiers in Marine Science*, 6, 2019.

[20] Gianluca Mastrantonio. Modeling animal movement with directional persistence and attractive points. *ANNALS OF APPLIED STATISTICS*, 16(3):2030–2053, SEP 2022.

[21] Brett Mcclintock, Devin Johnson, Mevin Hooten, Jay Ver Hoef, and Juan Morales. When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*, 2:21, 10 2014.

[22] Brett T. McClintock, Ruth King, Len Thomas, Jason Matthiopoulos, Bernie J. McConnell, and Juan M. Morales. A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, 82(3):335–349, 2012.

[23] Vilis Nams. Sampling animal movement paths causes turn autocorrelation. *Acta biotheoretica*, 61:269–284, 06 2013.

[24] Aurélien Nicosia, Thierry Duchesne, Louis-Paul Rivest, and Daniel Fortin. A general hidden state random walk model for animal movement. *Computational Statistics & Data Analysis*, 105:76–95, jan 2017.

[25] NOAA Fisheries. Southern resident killer whale (orcinus orca). `https://www.fisheries.noaa.gov/west-coast/endangered-species-conservation/southern-resident-killer-whale-orcinus-orca`.

[26] Dawn Noren, AH Johnson, D Rehder, and A Larson. Close approaches by vessels elicit surface active behaviors by southern resident killer whales. *Endangered Species Research*, 9, 07 2009.

[27] Jennifer Olson, J Wood, RW Osborne, L Barrett-Lennard, and Shawn Larson. Sightings of southern resident killer whales in the salish sea 1976-2014: The importance of a long-term opportunistic dataset. *Endangered Species Research*, 37, 10 2018.

[28] Jennifer K Olson, Jason Wood, Richard W Osborne, Lance Barrett-Lennard, and Shawn Larson. Sightings of southern resident killer whales in the salish sea 1976–2014: the importance of a long-term opportunistic dataset. *Endangered Species Research*, 37:105–118, 2018.

[29] Clifford S. Patlak. A mathematical contribution to the study of orientation of organisms. *Bulletin of Mathematical Biology*, 15:431–476, 1953.

[30] Toby Patterson, Alison Parton, Roland Langrock, Paul Blackwell, Len Thomas, and Ruth King. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101, 07 2017.

[31] Haiganoush K. Preisler, Alan A. Ager, Bruce K. Johnson, and John G. Kie. Modeling animal movements using stochastic differential equations. *Environmetrics*, 15(7):643–657, 2004.

[32] Marine Randon, Michael Dowd, and Ruth Joy. A real-time data assimilative forecasting system for animal tracking. *Ecology*, 103(8):e3718, 2022.

[33] Rosalind M. Rolland, Susan E. Parks, Kathleen E. Hunt, Manuel Castellote, Peter J. Corkeron, Douglas P. Nowacek, Samuel K. Wasser, and Scott D. Kraus. Evidence that ship noise increases stress in right whales. *Proceedings of the Royal Society B: Biological Sciences*, 279(1737):2363–2368, 2012.

[34] Cheng Yong Tang and Song Xi Chen. Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149(1):65–81, 2009.

[35] M Scott Taylor and Fruzsina Mayer. International trade, noise pollution, and killer whales. Technical report, National Bureau of Economic Research, 2023.

[36] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36:823–841, Sep 1930.

[37] Joe Watson, Ruth Joy, Dominic Tollit, Sheila J Thornton, and Marie Auger-Méthé. Estimating animal utilization distributions from multiple data types: a joint spatiotemporal point process framework. *The Annals of Applied Statistics*, 15(4):1872–1896, 2021.

[38] Rob Williams, Erin Ashe, Laurel Yruretagoyena, Natalie Mastick, Margaret Siple, Jason Wood, Ruth Joy, Roland Langrock, Sina Mews, and Emily Finne. Reducing vessel noise increases foraging in endangered killer whales. *Marine Pollution Bulletin*, 173:112976, 2021.

[39] Rob Williams, David Bain, J.C. Smith, and David Lusseau. Effects of vessels on behaviour patterns of individual southern resident killer whales orcinus orca. *Endangered Species Research*, 6:199–209, 01 2009.

[40] Jeannette E Zamon, Troy J Guy, Kenneth Balcomb, and David Ellifrit. Winter observations of southern resident killer whales (orcinus orca) near the columbia river plume during the 2005 spring chinook salmon (oncorhynchus tshawytscha) spawning migration. *Northwestern Naturalist*, pages 193–198, 2007.