Title
Whither the whale shark wanders: Tools and methods for modelling whale shark movement

Author
Wickham, Charlotte Victoria

Publication Date
2011

Peer reviewed|Thesis/dissertation

Whither the whale shark wanders: Tools and methods for modelling whale shark movement

by

Charlotte Victoria Wickham

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor David Brillinger, Chair
Professor John Rice
Professor Dennis Levi

Spring 2011

Abstract

Whither the whale shark wanders:
Tools and methods for modelling whale shark movement
by
Charlotte Victoria Wickham
Doctor of Philosophy in Statistics
University of California, Berkeley
Professor David Brillinger, Chair

The whale shark is the largest fish living in our oceans but little is known about its ecology and natural history. Tagging studies allow us to learn more about whale shark movements and habitat use. This knowledge is essential for planning the management and conservation of whale sharks to ensure their continued coexistence with humans.

This thesis is concerned with modelling the movements of whale sharks using stochastic models estimated from tracks obtained by tagging studies. In particular, it focusses on using stochastic differential equations being driven by potential functions. Approximations are presented that reduce the task of estimating the potential function to regression problems. I present a method for obtaining smooth potential functions and explicitly modelling measurement error.

The primary scientific questions addressed by the thesis are: Where do whale sharks go, both in terms of geographical location and habitat? Do whale sharks interact with each other? Remote sensing data on sea surface temperature, chlorophyll concentration and sea surface currents are readily available. These time varying covariates are incorporated into the potential function model allowing investigation of their influence on the whale shark's movements.

Acknowledgements

Thank you to my advisor, David Brillinger, for providing direction and much appreciated advice. I've learnt so much from you about being a statistician and how to value the important things in life.

Many thanks to Brent Stewart for providing the whale shark data and many useful comments on the manuscript.

Thank you Shuang Song for sharing your eye movement data and helping me understand your field.

Thank you to my other committee members, Dennis and John, for your comments and encouragement.

Special thanks to all my fellow grad students but especially Margaret, Irma, Frances, Cathy and Harry, for providing their friendship and support through the highs and lows of grad school.

Thank you to the many other friends that took me snowboarding, sailing, hiking, swimming, backpacking, biking and out for a drink. You reminded me of the world outside Evans Hall and kept me sane.

Matt, thanks for being the first friend I made in Berkeley and providing many enjoyable coffee breaks in the ensuing six years.

My big brother, Hadley, deserves a huge thank you for being a great role model of a young statistician, building amazing R tools that saved me many hours and giving firm motivational advice when I needed it.

Thanks Mum and Dad for giving me such a happy, loving and intellectually stimulating start in life. Thanks for always being there.

Thank you Josh for being patient and providing many pleasant distractions from work.

To Joshua

Contents

Introduction

> "Movement rules are critical to realistic simulation of how individuals and therefore populations respond to changes in habitat and population density"
>
> Railsback et al. (1999)

This thesis is primarily concerned with modelling the movement of whale sharks off the coast of East Africa. The measurements of their movement come from tags attached for a year that estimate the sharks location by processing light levels. The primary goals are to increase understanding of the movements of whale sharks in relation to environmental variables and each other.

Chapter 1 introduces the whale shark and the reasons for tracking it. The primary data set is also introduced. Chapter 2 gives an overview of some of previous work and approaches to trajectory modelling with a focus on animal movements. The approach built upon in this thesis is described in Chapter 3, which reviews the relevant theory and presents the least squares approach to estimating the potential function parameters. Different forms for the potential are discussed in Chapter 4 along with an approach to estimate a smooth potential function. The tags used to track the whale sharks have particularly large errors associated with them. The process of obtaining locations from the tags, prior work on the magnitude of the errors in the locations and an approach to directly model the errors are discussed in Chapter 5. In Chapter 6 the methods are applied to one shark using a potential that only depends on the shark's position. Covariates from remote sensing are explored and added to the model of the sharks movement in Chapter 7. Chapter 8 investigates interactions between two of the sharks.

While the primary focus of the thesis is on modelling animal trajectories the methods are generally applicable to any kind of trajectory. An example of previous work includes modelling the movement of a soccer ball (Brillinger, 2007b). In Chapter 9 the methods developed in the thesis will be applied to eye movement.

A comment on software

All of the analyses in this thesis are performed in `R` (R Development Core Team, 2010). Heavy use is made of the packages `ggplot2` (Wickham, 2009) and `plyr` (Wickham, 2011) for visualization and data manipulation. Use of other packages is noted throughout the text.

Chapter 1

Meet the Whale Shark (*Rhincodon typus*)

Figure 1.1: Dr Brent Stewart swimming with a whale shark.

The whale shark (*Rhincodon typus*) is the largest fish living in our oceans but little is known about its ecology and natural history. They are known to grow up to 14m long and weigh up to 15 tons (Colman (1997), Stewart and Wilson (2005)). Its slow growth, long life and low fecundity make it susceptible to overfishing whether from intentional or unintentional catches. Its gentle and slow moving nature, aggregations in shallow warm water and striking appearance make it an attractive animal for tourism. Learning more about feeding, breeding, movements, habitat use and ecology in general is essential for planning the conservation and management of the whale shark to ensure its continued coexistence with humans. Tracking studies provide an important window into the behaviour of these huge fish.

Whale sharks feed through a filter mechanism. They are known to suction feed, gulping

mouthfuls of prey laden water, and ram feed, swim with an open mouth passively filtering out prey. They feed on small nekton and perhaps actively prey on larger schooling fishes and squids.

Whale sharks are found around the globe in tropical and warm temperate seas (Compagno (2001) see Figure 1.2). However, recent tracking studies demonstrated diving to depths where water temperature was as cold as 4–5° C (Wilson et al., 2006) which suggests they are not constrained by cold water. Whale sharks appear in large numbers seasonally in shallow waters near Western Australia, the Seychelles, the Sea of Cortez, Belize, Kenya, South Africa, Galapagos, Caribbean, India and even occasionally in New Zealand (Duffy, 2002). This thesis concentrates on whale sharks tagged of the coast of Kenya where they aggregate from November to March. Below a depth of 10m, whale sharks are practically invisible from the surface so it is not known where they are when not visible from the surface, hence the usefulness of tracking studies.

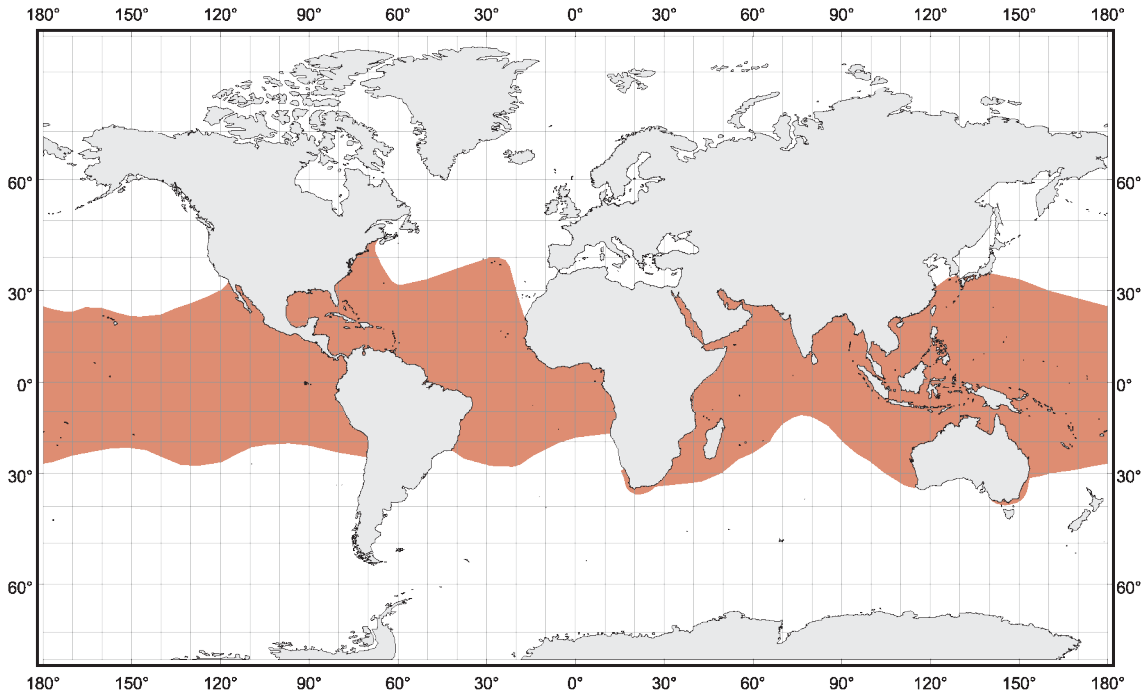


Figure 1.2: Distribution of whale sharks. Picture adapted from of Compagno (2001)

Tagging studies of whale sharks seek to answer many questions including: when are the sharks resident in the areas of congregation? Where do they go when they aren't visible from the surface? What is the spatial and temporal scope of the seasonal aggregations? Studies have demonstrated some genetic similarities among whale shark populations in different ocean basins (Castro et al. (2007)). Is this a result of regular migration and mixing? What habitats they are attracted to? These are especially important questions for the appropriate

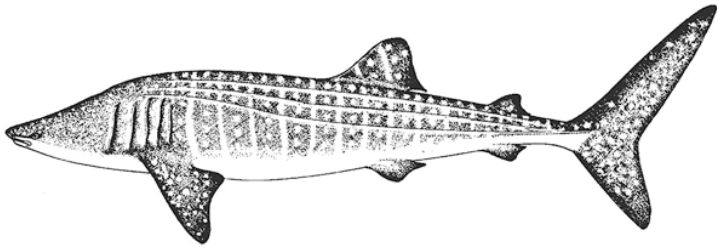


Figure 1.3: Diagram of a whale shark. Taken from Compagno (2001).

design and scope of protected areas.

This thesis extends beyond simple plotting of tracks. Formal statistical models of movement allow the exploration of hypotheses on the animals use of habitat and interactions, as well as generating simulations of movement and uncertainty evaluation.

This thesis will focus on the following two scientific questions:

- Is there a relationship between the movement of the whale sharks and environmental variables such as chlorophyll, temperature and currents?
- Is there interaction between whale sharks?

addressing them via stochastic modelling.

1.1 Kenya tagging study

The data set used in this thesis comes from a tagging study conducted off the coast of Kenya led by Dr Brent Stewart (Stewart et al. (2007), Stewart (2009)). I will give an overview of the study here. Some specifics of the study, particularly the type of tag used, and the available covariate data, will guide the development of the models used.

The tagging study took place in 2007 and 2008. Twenty sharks were tagged near Diani Beach and Watamu, Kenya. Figure 1.4 displays a map of the area. The tags used were archival pop-up tags (PAT) which periodically measure pressure, temperature and ambient light levels. The tag remains tethered to the shark until scheduled or accidental release. It then floats to surface and transmits its archived data to the ARGOS satellite system. The time available to transmit the recorded data depends on the battery life of the tag but is typically no more than two weeks. It is possible to retrieve all of the data if the tag is physically retrieved. However, it is not generally practical to search for the released tag so this only occurs if the tag is washed up on a beach. Tags process light level data to estimate the times of sunrise and sunset. These are then used to make daily estimates of the whale sharks latitude and longitude. More details on this method of location are discussed in Section 5.2.

Figure 1.4: Map of the tagging study area. Sharks were tagged near Watamu and Diani Beach, Kenya. Map data from `http://www.naturalearthdata.com/`

Tags can release prior to their scheduled release date. If the tag measures constant pressure for more than 4 days, signaling a dead shark for example, or reaches a pressure of 2000 psi, the crushing limit of the buoyancy float, it is programmed to release from the shark and float to the surface to begin transmitting. Often the tags detach for unknown reasons.

Of the 20 tags that were attached to whale sharks during 2007 and 2008, four tags never made satellite contact after their attachment. One tag only transmitted very briefly and no location data were obtained. The number of locations and time of observation for the remaining 15 tags are shown in Table 1.1. Of these tags, two reached their release date, nine released due to constant pressure, two due to high pressure and three for unknown reasons. Only one tag reached its release date and managed to transmit all of the recorded data.

The reported locations for the 15 whale sharks are shown in Figure 1.5. The majority of them stayed near the coast. The exception being 80180 which made a journey to the East ending up near the Seychelles. There isn't any evidence here for systematic migrations (B. Stewart, pers., comm.). The dates of the location estimates for all 15 sharks are shown in Figure 1.6. While the procedure for obtaining locations should result in one location per day, we can see that the data received is far from this regular. Daily locations are missing either because the tag did not have a chance to transmit all of the recorded data or that the light level data used to calculate the location was not reliable enough to calculate a location (often near an equinox). Tags that only managed to transmit a fraction of the data have gaps as long as a month between observations. This highlights the need for methods that can deal with unequispaced times.

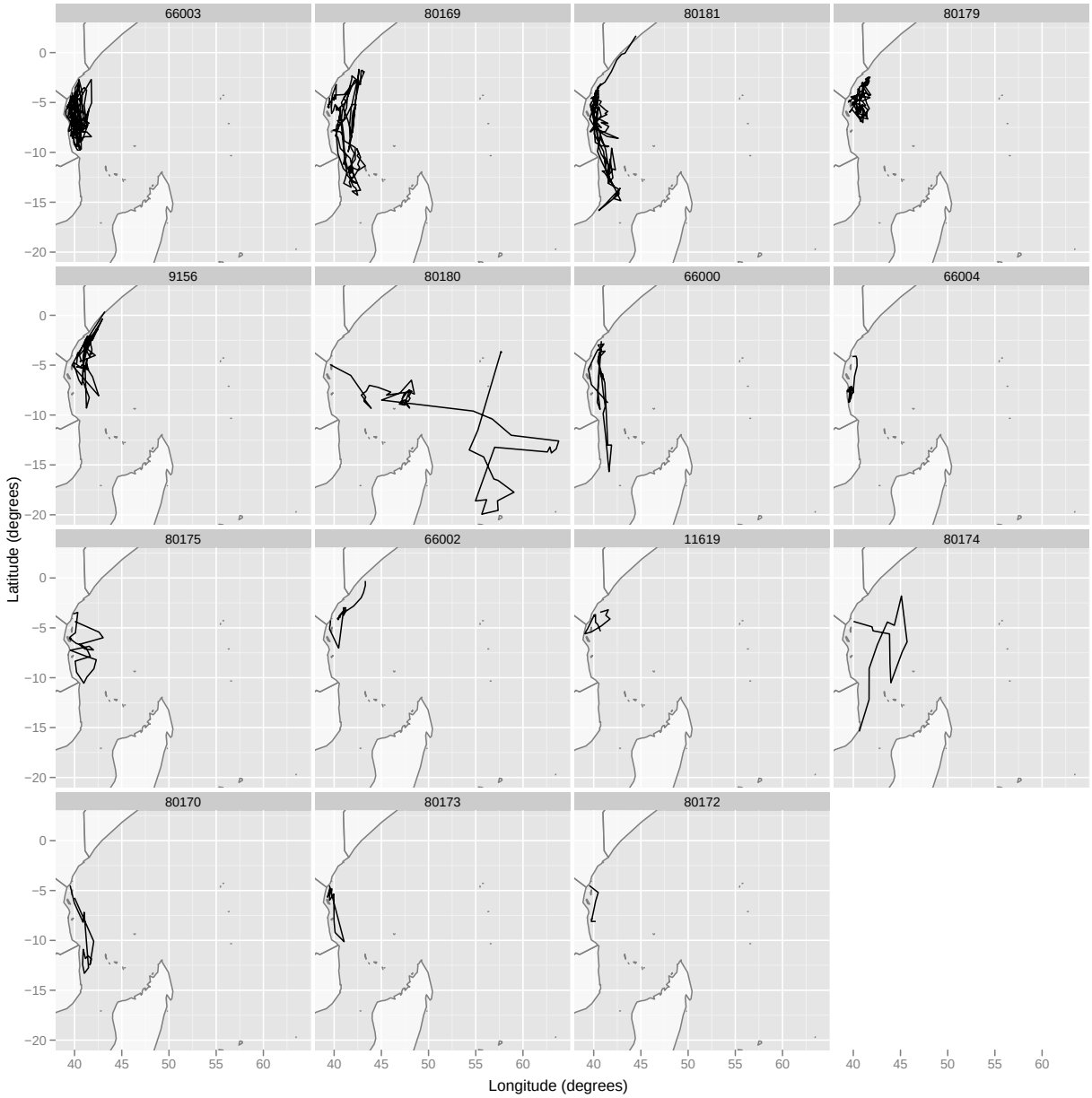


Figure 1.5: The recorded tracks of the 15 sharks. The numbers at the top of each panel indicate the individual tag id numbers.

	Tag ID	Tagged	Time between first and last location (days)	Number of locations	Total distance covered (km)	Average speed (km/day)
1	66003	Feb 2007	381	201	17773	77
2	80169	Feb 2008	242	162	14634	78
3	80181	Mar 2008	135	128	13587	106
4	80179	Mar 2008	342	77	6318	38
5	9156	Feb 2008	175	61	7760	74
6	80180	Mar 2008	122	56	10154	158
7	66000	Apr 2008	215	38	5512	52
8	66004	Feb 2007	142	37	1592	24
9	80175	Feb 2008	157	28	2997	44
10	66002	Feb 2007	74	21	1705	55
11	11619	Feb 2008	16	17	1070	67
12	80174	Feb 2008	190	15	3724	33
13	80170	Feb 2008	277	14	2580	11
14	80173	Feb 2008	276	10	1673	8
15	80172	Feb 2008	7	7	512	75

Table 1.1: (From Stewart (2009)) Summary of location data for sharks tagged as part of the study. Distance covered is approximated by the sum of euclidean distances between successive locations. In practice the total distance covered by the sharks will be greater than this because of deviations from a straight line path and the curvature of the Earth.

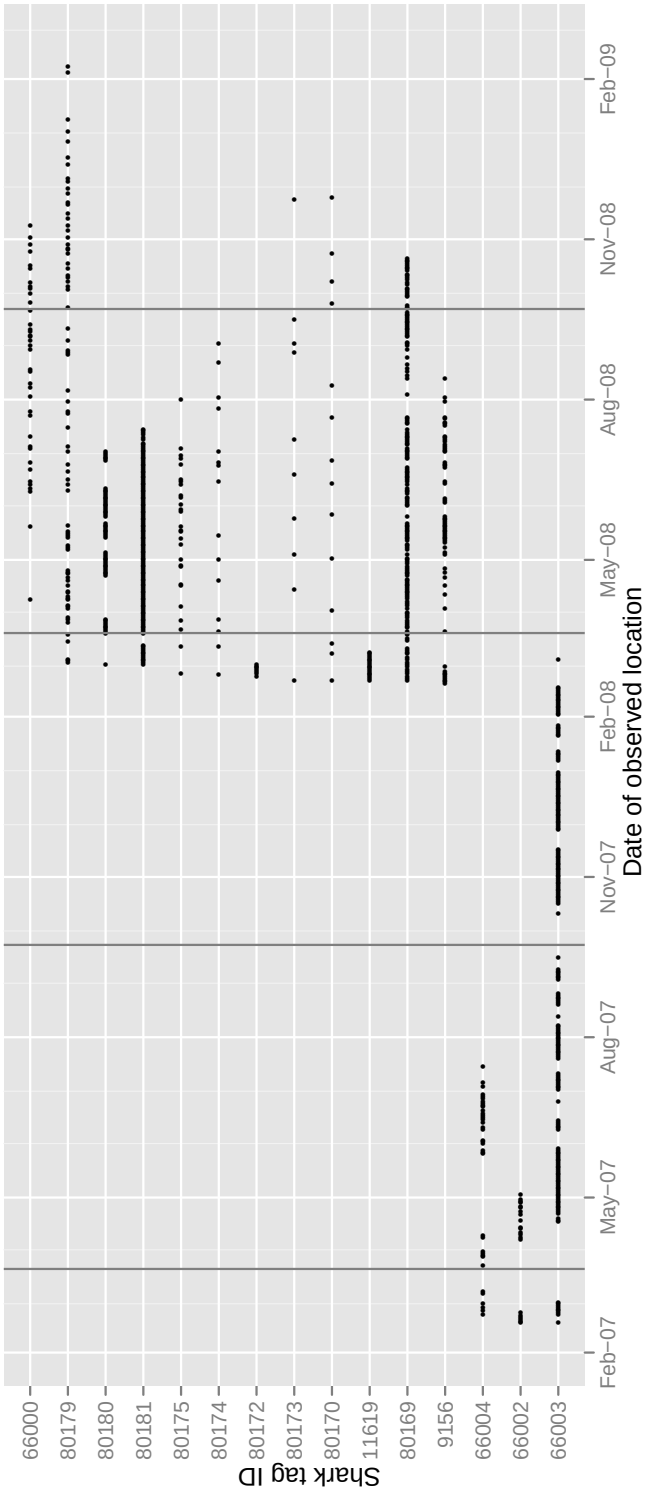


Figure 1.6: The times of recorded locations for the 15 sharks. Equinoxes are marked with grey lines.

Examining the sharks' latitude and longitude against time is another way to take a preliminary look at possible patterns amongst the sharks (Figure 1.7). There is no indication that the sharks are grouping together at any particular time. An alternative approach is to use bagplots, a 2-dimensional generalization of boxplots (Rousseeuw et al., 1999), of the sharks' locations by season (Figure 1.8). Although the Dec–Feb time period shows all the sharks in a similar location, this is a result of the bulk of the observations during these months corresponding to three freshly tagged sharks in 2007 who have not had time to disperse from the tagging location. None of the other seasons show evidence of sharks sharing common ranges.

Figure 1.7: Latitude and longitude against time for all sharks except shark 80180.

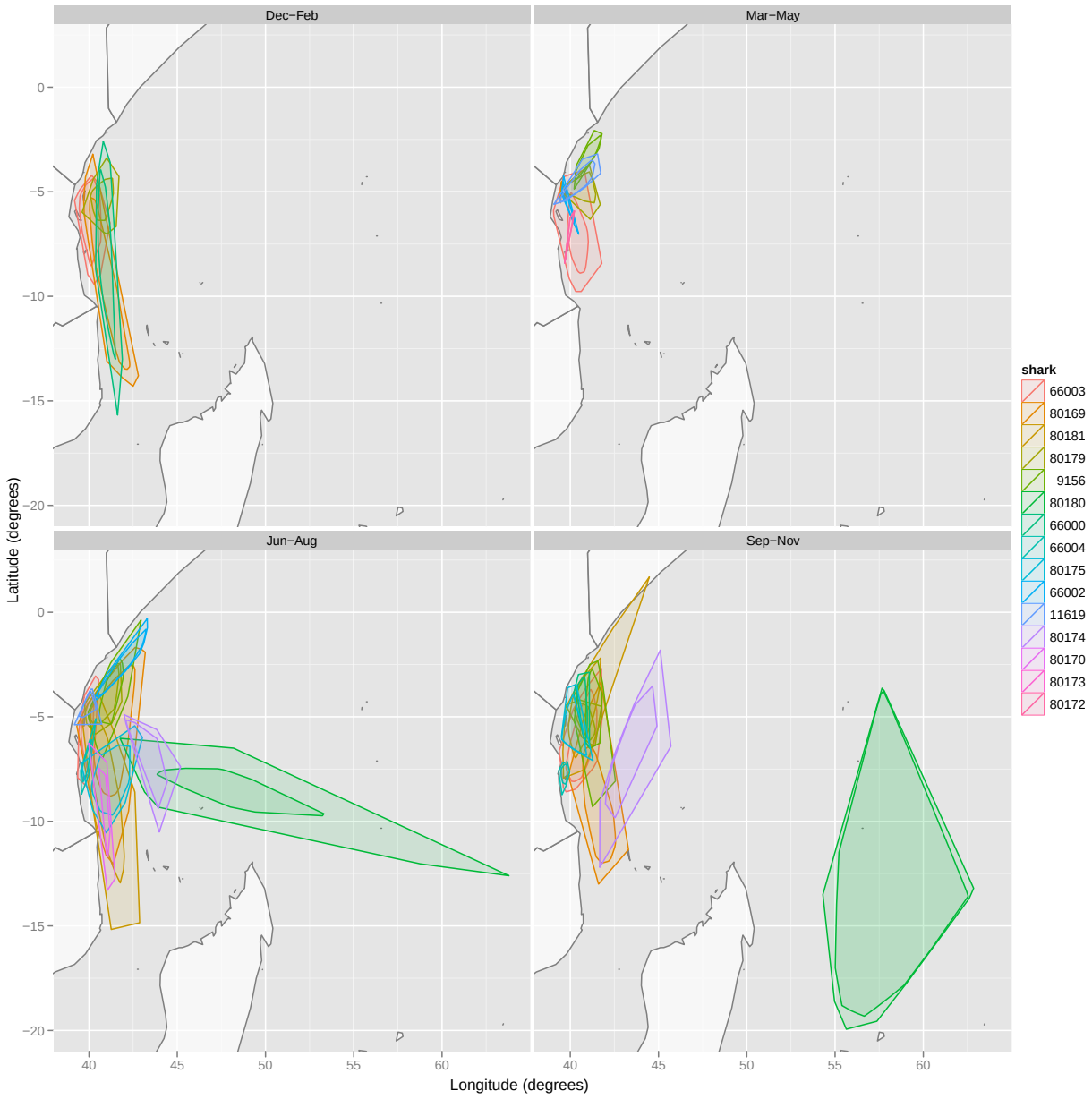


Figure 1.8: Bagplots for all sharks by season.

1.2 Coordinate systems

It's important to take some time to describe the various coordinates systems used in this thesis. The most familiar coordinate system for locations on the Earth is longitude and latitude. This also happens to be the output of most geolocation tags. Successive locations of an animal are reported as (longitude, latitude) pairs.

Longitude measures position in the East-West direction. It is the angle between the projection of the position of interest and a reference position onto the equatorial plane. By convention the reference position is the Royal Observatory in Greenwich, England. The longitude is zero on the great circle path from the poles passing through the Royal Observatory, Greenwich. Locations to the East of zero are positive and to the West are negative and are reported in degrees. Places of equal longitude form half circles of the same size passing through the poles, and are also known as meridians.

Latitude measures position in the North-South direction. It is the angle between the equatorial plane and a line that is normal to the Earth's reference ellipsoid at the position of interest. The reference ellipsoid is a simplified model of the earth accounting for its flattened shape compared to a sphere. By convention latitude is 0° at the equator. The convention used in this thesis is that the North pole is 90° and the South pole -90°. Lines of equal latitude form circles parallel to the equator. Latitude and longitude are not uniform measures. A movement of one degree of latitude is always about 111km, but a movement of one degree of longitude varies from about 111km at the equator to 0km at the poles (it is about 78km at a latitude of 45°). Assuming the Earth is spherical, 1° of longitude at ϕ° latitude is approximately,

$$\cos\left(\phi\frac{2\pi}{360}\right)M_R\frac{2\pi}{360} \quad \text{km}$$

where M_R is the average meridional radius of the Earth (≈ 6367 km). This is an important fact when considering motion on the surface since moving one degree of longitude can take significantly less effort than moving one degree of latitude. The distance between two points on the Earth is generally calculated along the great circle path between them. The great circle path between two points is the shortest path between them on the surface of the earth. For two points at (longitude, latitude)$= (\lambda_1, \delta_1)$ and (λ_2, δ_2) the great circle distance can be calculated approximately by

$$\text{great circle distance (km)} = R\cos^{-1}\left(\cos\delta_1\cos\delta_2\cos\left(\lambda_1-\lambda_2\right)+\sin\delta_1\sin\delta_2\right)$$

where R is the equatorial radius of the earth (≈ 6371 km). Both of the above equations are approximate in the sense that they assume the Earth is spherical. The great circle distance formulae above may also be inaccurate for small distances and the Haversine method is preferred (Sinnott, 1984).

It is convenient for modelling animal movement in the oceans that all directions are

treated equally. There are two approaches. One is to work in latitude and longitude but base the underlying process on spherical Brownian motion with drift as in Brillinger (1997). The other is to work with a projection of latitude and longitude into cartesian space so that distances calculated using the Pythagorean theorem in this new space closely match those using great circle distances in the latitude-longitude space. The choice between the two approaches depends on the entities of interest. If the area covered by the animal is small approximating it by a flat surface will not introduce significant distortion. If the animal is covering huge distances then a spherical Brownian motion will be more appropriate. Here the whale sharks stay near the coast of Africa and we take the approach of working in cartesian coordinates.

Universal Transverse Mercator (UTM) coordinates are one approach and the transformation utilised here. It is not single map projection but a collection of sixty zones each employing a different projection (Kemp, 2008). The zones are each centered over a meridian of longitude and 6 are degrees in width. Within a zone the degrees latitude and longitude are transformed to metres Easting and Northing using a transverse mercator projection. To visualise the transverse mercator projection, imagine taking a rectangular sheet of paper with a line drawn down the middle lengthways. Wrap the sheet of paper around a globe to form a cylinder by aligning the center line with a line of constant longitude. Locations are then projected onto the cylinder and it is unrolled. The projection has very little distortion near the chosen longitude meridian. By using multiple meridians distortion within a UTM zone is kept below 1 part in 1,000. In UTM, a position on the earth is specified by a zone, and an Easting and a Northing in kilometres. The majority of the whale sharks tagged in Kenya stay within zone 37, and this will be the zone used for all UTM coordinates unless specified.

Chapter 2

An overview of some prior work on animal trajectory modelling

I define a trajectory as a curve providing an animal's location, $\mathbf{r}(t)$, through time, t. The location is often two dimensional ($\mathbf{r}(t) \in \mathbb{R}^2$), the position of the animal on the globe in some coordinate system, for example latitude and longitude or in UTM as Easting and Northing, but could be higher dimensional, with the addition of depth below sea level for example. In this thesis I focus on the two dimensional case. The trajectory is assumed to be a continuous path but only observed at discrete times, often unequally spaced. The observation times are denoted, t_i, so the data become $\mathbf{r}(t_i) = (x(t_i), y(t_i)), i = 1, \ldots, n$.

For animals, these types of data occur most often when the animals are fitted with a remote telemetry device (more details can be found in Section 5.1). They can also result from direct observation or the capture of images by camera. The advent of GPS allows locations to be captured with very short time intervals and very high accuracy (at least for terrestrial animals). It is expected there will be more and more of these types of data.

Two concepts that occur frequently in the animal tracking literature are home range and habitat use. Home range was defined informally by Burt (1943) as "the area, usually around a home site, over which the animal normally travels for food". This has been replaced by the more statistical definition "the range is some fixed percentage (usually 95%) confidence region obtained from the animal's utilization distribution function" (Worton, 1987). Where the use distribution function is the frequency distribution of an animal's location over time.

Habitat is defined by Hall et al. (1997) as "the resources and conditions present in an area that produce occupancy—including survival and reproduction—by a given organism." Habitat use attempts to infer the desired habitat of the organism by comparing the habitat it spends time in to the available habitat.

Often the goal in an appropriate model of an animal's movement is to estimate home range or habitat use. In this chapter I give a summary of some previous work on the modelling of animal trajectories.

2.1 Descriptive methods

Perhaps the most basic analysis is to simply plot the trajectory, joining successive locations with straight lines and examine the result. Stewart and DeLong (1995) used this successfully to describe the twice yearly migration of northern elephant seals. The addition of important features such as coastlines, roads or fences is often crucial in obtaining insight into patterns of movement.

Calenge et al. (2009) reviewed some of the most common derived quantities from trajectories. These included the straight line distances between successive locations, the apparent average speed between locations, the increments in the X and Y directions, the absolute angles between steps and the X axis, the mean squared displacements from the initial position, and relative turning angles between steps. Interpolation to equispaced time is generally necessary for distances or angles to be comparable. Summaries (statistical or graphical) of these variables often provide useful information to compare across different individuals in a single study.

Models for the complete trajectory can be split into two categories, those that result from the consideration of continuous time stochastic processes and those that result from discrete equispaced time stochastic models. In practice they share a lot of common properties. A common assumption is that the successive positions are Markov processes.

2.2 Continuous time stochastic approaches

The continuous time approaches share the property that they assume the location of the animal behaves like a diffusion process. Kendall (1974) introduced diffusions as a movement model for birds. In particular he used Brownian motion with attractive polar drift to model their navigation towards a pole. As an approach to home range estimation, Dunn and Gipson (1977) took a very similar approach and used the Multivariate Ornstein-Uhlenbeck (MOU) process as a model for an animal's position.

Worton (1995) discussing Dunn and Brisbin Jr (1985) and Dunn and Gipson (1977) suggested even if the MOU assumptions were violated, the parameters of the model provided useful summaries of trajectories but that an effort should be made to find robust standard errors.

In an attempt to make the movement more realistic Blackwell (1997) extended the simple Ornstein-Uhlenbeck process by assuming an underlying unobserved discrete state variable. This variable allowed the animal to switch between moving according to different Ornstein Uhlenbeck models (corresponding to "foraging" or "migrating" for exmaple). Parameters and states were estimated via MCMC and applied to the movements of a wood mouse.

In a similar vein to Kendall (1974), but employing stochastic differential equations, Brillinger and Stewart (1998) used spherical Brownian motion with attraction to a point to model the migrations of northern elephant seals. The parameters were fit by maximizing

the likelihood of the discretized process. A specific form for the measurement error was set down and its parameters estimated by Monte Carlo.

Brillinger et al. (2001) extended the Brownian motion idea to modelling the movement as stochastic differential equations with the drift determined by a potential function. That approach was extended to include terms to model attraction, repulsion from boundaries, repulsion from disturbances and time varying covariates (Brillinger et al., 2004a, 2006a,b, 2004b; Brillinger, 2007a; Brillinger and Stewart, 2010). Measurement error can be added as in Brillinger and Stewart (1998) and Brillinger (2000). This work will motivate the approach taken in this thesis and is discussed in detail in Chapter 3.

2.3 Discrete random walk approaches

The discrete time random walk approaches share a lot of their development with the continuous time ones and in many cases are specific cases of discretized continuous time models. The early methods used a simple autoregressive structure either on positions or their differences, for example, a simple random walk

$$\mathbf{r}(t) = \mathbf{r}(t-1) + \boldsymbol{\epsilon}_t$$

or a correlated random walk

$$\mathbf{r}(t) - \mathbf{r}(t-1) = T\left(\mathbf{r}(t-1) - \mathbf{r}(t-2)\right) + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{\epsilon}_t$ is a noise process and T is a 2×2 matrix describing the average turning angle.

Extensions to include unobserved behavioral states and measurement error lead to the application of state space models (Morales et al., 2004; Royer et al., 2005; Jonsen et al., 2003, 2005; Flemming et al., 2006). Estimates were made robust by assuming t-distributions for the measurement error (Jonsen et al., 2003, 2005; Flemming et al., 2006). In general the estimation of the state space model parameters occured via a Kalman filter or a Bayesian MCMC procedure. The discrete models employed have the disadvantage that equally spaced times are needed, typically constructed by interpolation.

Another model for discrete time trajectories that has been proposed is the Lévy flight. Originally applied to the times between an albatross' touch down on water (Viswanathan et al., 1996) a Lévy flight has step lengths drawn from a long-tailed power law distribution. Subsequently Edwards et al. (2007) found no evidence for this type of process after reanalysis of the data.

2.4 Estimating home range and adding covariates

A simple method for calculating a home range is to use the minimum convex polygon of the observed locations (Mohr, 1947). A more statistical approach is advocated in Worton (1989) and uses kernel density estimation. The home range is then the 95% confidence region of the estimated density of the animals position.

The method of Dunn and Gipson (1977) based on multivariate Orstein-Uhlenbeck processes could only produce elliptical home ranges. Later, Dunn and Brisbin Jr (1985) suggested that the MOU approach could be extended by suitable transformations to allow other shapes, for example to account for a boundary. They also suggested exploring interactions between two animals by setting up a four dimensional problem and looking for terms on the off diagonal indicating dependence between animals.

Calenge (2007) discussed using trajectories to investigate habitat choice. Habitat at the observed locations of the animal is compared to the habitat available (generally within the animal's home range). The comparison was made using marginality, defined to be the squared Euclidean distance between the conditions encountered by the animal and the average conditions in the area. Martin et al. (2008) emphasized that the correlation between positions should be taken into account and advocated using habitat along a null track (created by rotating the observed trajectory) as a comparison to the habitat along the observed track.

Brillinger et al. (2004b) assessed the effect of human disturbance (a time varying covariate) on the behaviour of tagged elk by adding a dependence on the distance to an ATV in the model of the elk movement. For another experiment on elk, Preisler et al. (2006) developed a model for the probability of flight, that the velocity was above a time varying threshold, as a function of the distance to a disturbance.

Brillinger and Stewart (2010) used the stochastic differential equation with potential function model to incorporate the effect of covariates on the trajectory of a detached drifting whale shark tag. The most relevant covariate was the surface currents obtained from remote sensing.

2.5 Summary

The advantage of the stochastic differential equation approach is that it deals naturally with unequispaced time points, maintains flexibility in the deterministic motions of the animal and directly extends to account for covariates and boundaries. The details and theory of the approach are discussed in the following chapter.

Essential to the modelling of the whale sharks off Kenya will be investigating the effect of the environmental covariates and the sharks interactions with each other.

Chapter 3

An SDE approach to animal movement

The methods in this thesis build on the approach to modelling a trajectory, $\mathbf{r}(t)$, as a stochastic differential equation (SDE) where the drift is the gradient of a potential function $H(\mathbf{r})$, that is,

$$d\mathbf{r}(t) = -\nabla H(\mathbf{r}(t), t)dt + \sigma d\mathbf{B}(t). \tag{3.1}$$

Here $\mathbf{B}(t)$ is a Brownian motion, $H(\mathbf{r})$ a potential function while ∇ denotes the gradient operator. Most generally $\mathbf{r}(t) \in \mathbb{R}^d$, but I will restrict discussion to $d \leq 2$. The potential function, H, models the deterministic movement of the object, animal for example, in response to its environment or describes preferential movement. The random part makes the position, $\mathbf{r}(t)$, stochastic, allows for variation in the deterministic movement and accounts for measurement error, excluded explanatories and model error. The potential function, H, is the basic object of interest.

This chapter describes some historical use of potential functions for modelling movement of objects in classical physics and reviews the mathematical formalism of stochastic differential equations. A series of results are presented that allow estimation of the potential function given discrete trajectory data, $\mathbf{r}(t_i), i = 1, \ldots, n$, with the t_i being the times of observation, under certain assumptions.

3.1 Potential function

3.1.1 Classical mechanics

Potential functions arise naturally in classical mechanics when dealing with motion under a force. Newton's second law states a body experiencing a force F experiences an acceleration a related to F by $F = ma$, where m is the mass of the body (Taylor, 2005). In differential form, an object at location, x, with mass m under the influence of a force, F, evolves according to the equation,

$$m\frac{d^2x}{dt^2} = F\left(x, \frac{dx}{dt}\right),$$

where the derivatives are with respect to time.

A conservative force is one such that the force depends only on position, x, and the work done between two positions (the line integral of force) is independent of the path taken. For a conservative force a **potential function** V can be defined such that

$$-\nabla V(x) = F(x)$$

The negative sign is a historical convention. If the potential function, $V(x)$, describes a surface, then the force $F(x)$ pushes the object downhill. Physical systems that can be described in this way include objects under gravitational, electrostatic and interatomic forces. Another example is a simple harmonic oscillator, a mass on a spring type problem, where $F(x) = -kx$, $V(x) = 1/2kx^2$ where k is a spring constant. The quadratic form of the potential here means the mass is attracted to the point zero.

If friction is also present the system, then the equations satisfied by the object become

$$\frac{d^2x}{dt^2} = F(x) - \beta\frac{dx}{dt}.$$

with β the coefficient of friction. For large values of β this can be approximated by

$$\frac{dx}{dt} = F(x)$$

also known as Aristotle's Law (Andrianov et al., 2002). Figure 3.1 provides examples of the simple harmonic oscillator with $k = 0.1$ and under different levels of friction.

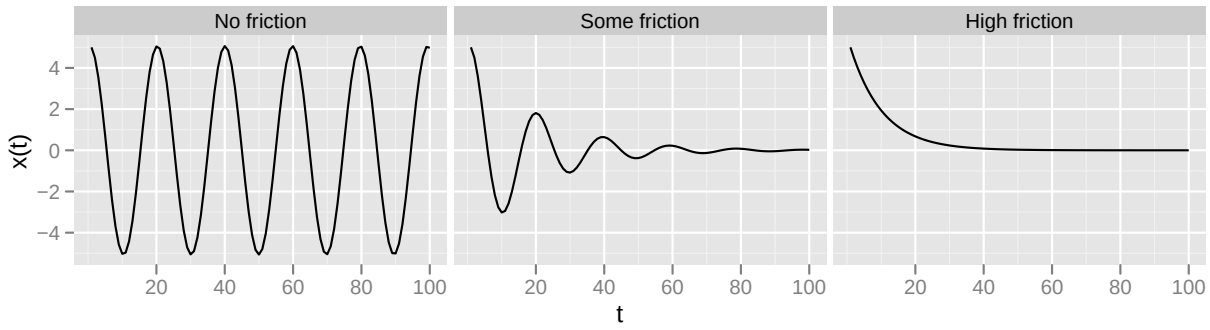


Figure 3.1: Examples of the simple harmonic oscillator ($k = 0.1$) with no friction, some friction and using the high friction approximation. Solutions are numerically simulated using time steps of size 1, $x(0) = 5$ and a unit mass.

Moving from deterministic to stochastic systems, the random movement of particles sus-

pended in a fluid has been described by a Langevin equation,

$$m\frac{d^2x}{dt^2} = F(x) - \beta\frac{dx}{dt} + b(t)$$

where b, termed Brownian motion, was used to describe a stochastic force due to interaction with the many atoms in the fluid in which the particle is suspended (Phillies, 2000). Later this process was formalised mathematically as as Brownian or Weiner process and will be described in the next section.

Definition 1. Brownian motion [Shreve (2004) pg 94] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$ there is a process $B(t,\omega), t \geq 0$ that is a measurable function of ω (the dependence on ω is usually supressed). $B(t)$ is Brownian motion if for all n and $0 < t_1 < t_2 < \ldots < t_n$ the increments,

$$B(t_1) - B(0), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$$

are independent and normally distributed with

$$\begin{aligned} \mathbb{E}\left(B(t_{i+1}) - B(t_i)\right) &= 0 \\ \text{Var}\left(B(t_{i+1}) - B(t_i)\right) &= t_{i+1} - t_i. \end{aligned}$$

Conventionally it is often assumed $B(0) = 0$. One of the surprising properties of Brownian motion is that it is continuous but nowhere differentiable. A d-dimensional Brownian motion is simply a collection of d independent Brownian motions $\mathbf{B}(t) = (B_1(t), \ldots, B_d(t))$ taking values in $\mathbb{R}^d$. It is also know as a Weiner process after the mathematician that initially formalized many of its properties.

We now introduce stochastic differential equations.

3.2 Stochastic Differential Equations

A stochastic differential equation (SDE) is a differential equation where some of the terms are stochastic processes. Hence, under appropriate assumptions solutions to stochastic differential equations, when they exist, are also stochastic processes. Typically a stochastic differential equation is written in the form,

$$d\mathbf{x}(t) = \mu(\mathbf{x}(t), t)dt + \sigma(\mathbf{x}(t), t)d\mathbf{B}(t), \tag{3.2}$$

where $\mathbf{B}(t)$ is a d-dimension Brownian motion. The functions μ and σ are called the drift and diffusion coefficients respectively. Equation 3.2 is known as an Ito stochastic differential equation.

When a solution, $\mathbf{x}(t)$, to the above SDE exists it is called a diffusion process. Under conditions it is a Markov process. The equation 3.2 is formally written

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t \mu(\mathbf{x}(t), t)dt + \int_0^t \sigma(\mathbf{x}(t), t)d\mathbf{B}(t),$$

Conditions on the functions μ and σ to guarantee a unique solution and will be set down later. First, some sense needs to be made of the second integral since $\mathbf{B}(t)$ is not of bounded variation.

3.2.1 Ito Integral

Equation 3.2 contains a stochastic integral. The basic approach to the definition of the stochastic, or Ito, integral,

$$\int_0^T G(t)dB(t),$$

is to first define the integral for simple functions, then to show that a general function can be successively approximated by simple functions and define the integral of a class of functions to be the limit of the integral of the sequence of simple functions. Here, I will outline the definition of the Ito integral. A more thorough treatment can be found in Oksendal (2003) (pages 26-29).

Let G be a simple function, that is given by a partition $0 = t_0 < t_1 < \ldots < t_n = T$ such that,

$$G(t) = G(t_j)1_{t\in[t_j,t_{j+1})} \quad 0 \leq t \leq T$$

$G(t)$ can depend on the same ω as $B(t)$ but for a general definition must only depend on information up to and including time t. In this case the stochastic integral up to time t is defined as,

$$\int_0^t G(u)dB(u) = \sum_{j<k} G(t_j)\left(B(t_{j+1} - B(t_j)\right) + G(t_k)\left(B(t) - B(t_k)\right) \tag{3.3}$$

where k is such that $t_k < t < t_{k+1}$.

It can be shown for a general $L_2(R)$ function G there exists a sequence of simple functions G_n that converge to G in the sense that

$$\mathbb{E}\left(\int_0^T |G_n(t) - G(t)|^2 dt\right) \to 0 \text{ as } n \to \infty.$$

Then integral of a general function is defined to be the limit in mean square of the

integrals of the sequence of simple functions

$$\int_0^t G(u)dB(u) = \underset{n\to\infty}{\text{l. i. m.}} \int_0^t G_n(u)dB(u)$$

The limit exists because we are working in L_2. There are other ways of defining a stochastic integral. For example the Stratonovich definition takes the function value in equation 3.3 at the midpoint (as opposed to the left hand point) in the limit (Shreve (2004) pg 191). Other definitions have proved useful in other situations but won't be pursued here.

3.2.2 Diffusion processes

Having defined the Ito integral the properties of the solution of 3.2 are now discussed. If a solution exists it is a diffusion process which is a special case of a Markov process.

Definition 2. Markov process Bhattacharya and Waymire (1990) (Def 1.2 pg 110) A stochastic process $\mathbf{x}(t)$ has the *Markov property* if for each $s < t$ the conditional distribution of $x(t)$ given $x(u), u \leq s$ is the same as the conditional distribution of $x(t)$ given $x(s)$.

For a Markov process to be a diffusion the following conditions are required.

Definition 3 (Bhattacharya and Waymire (1990)(Def 1.1 pg 368))**.** A Markov process $\{\mathbf{x}(t) = (x_1(t), \ldots, x_d(t))\}$ on the state space S, where S is a "regular" open subset of $\mathbb{R}^d$, is said to be a *diffusion* with drift coefficient $\boldsymbol{\mu}(\mathbf{x}) = (\mu_1(\mathbf{x}), \ldots, \mu_d(\mathbf{x}))$ and diffusion coefficient $\boldsymbol{\sigma}(\mathbf{x})$, a $d \times d$ matrix with elements $\sigma_{ij}(\mathbf{x})$, if it has continuous sample paths and satisfies,

$$\begin{aligned}
\mathbb{E}\left(x_i(s+t) - x_i(s)\middle|\mathbf{x}(s) = \mathbf{x}\right) &= t\mu_i(\mathbf{x}) + o(t) \\
\mathbb{E}\left((x_i(s+t) - x_i(s))(x_j(s+t) - x_j(s))\middle|\mathbf{x}(s) = \mathbf{x}\right) &= t\sigma_{ij}(\mathbf{x}) + o(t) \\
\mathbb{E}\left((x_i(s+t) - x_i(s))(x_j(s+t) - x_j(s))(x_k(s+t) - x_k(s))\middle|\mathbf{x}(s) = \mathbf{x}\right) &= o(t)
\end{aligned}$$

for all $i, j, k \in \{1, \ldots, d\}$ as $t \downarrow 0$, for every $\mathbf{x} \in S$.

The following provides conditions on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ for the solution of an SDE to be a diffusion.

Theorem 1 (Bhattacharya and Waymire (1990), Thm 2.5 pg 580)**.** *If,*

$$\begin{aligned}
&|\boldsymbol{\mu}(\mathbf{x}, t)| + |\boldsymbol{\sigma}(\mathbf{x}, t)| \leq C(1 + |\mathbf{x}|) \quad \mathbf{x} \in \mathbb{R}^d, t \in [0, T] \\
&|\boldsymbol{\mu}(\mathbf{x}, t) - \boldsymbol{\mu}(\mathbf{y}, t)| \leq M|\mathbf{x} - \mathbf{y}| \quad \textit{and} \quad \|\boldsymbol{\sigma}(\mathbf{x}, t) - \boldsymbol{\sigma}(\mathbf{y}, t)\| \leq M|\mathbf{x} - \mathbf{y}| \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, t \in [0, T]
\end{aligned}$$

for some constants C and M, where $\|\cdot\|$ denotes the matrix norm $\|\mathbf{D}\| := \sup\{\|\mathbf{Dz}\| : \mathbf{z} \in \mathbb{R}^k\}$. Then, the process satisfying

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t \mu(\mathbf{x}(t), t)dt + \int_0^t \boldsymbol{\sigma}(\mathbf{x}(t), t)d\mathbf{B}(t),$$

is a diffusion with drift $\boldsymbol{\mu}(\cdot)$ *and diffusion matrix* $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$, ^{T} *denoting transpose here.*

3.3 Movement Model

With a concrete definition of what the solution to an SDE is and some conditions on its existence, the SDE becomes the basis for the model of movement. The drift function, μ, will be assumed to have the special property that it is the gradient of a potential function. Also in the empirical applications later in the thesis the diffusion coefficient will generally be assumed to be independent of time and location so that $\boldsymbol{\sigma}(\mathbf{x}(t), t) = \boldsymbol{\sigma}$, a $d \times d$ matrix. Relaxations of this constraint are possibilities for future work. The model for movement becomes,

$$d\mathbf{x}(t) = -\nabla H(\mathbf{x}(t), t)dt + \boldsymbol{\sigma}\mathbf{B}(t). \tag{3.4}$$

One might ask if the conditions for the solution to be an diffusion are satisfied. With constant σ this condenses to asking whether $-\nabla H(\cdot)$ is Lipschitz continuous. Some functions of interest, for example, $\nabla H(x) = x^3$ are not globally Lipschitz continuous but are locally so. Since the animals' positions are bounded (they cannot leave the surface of the Earth), the local property will be enough.

3.3.1 Ornstein-Uhlenbeck

The Ornstein-Ulhenbeck (O-U) process is a classical example of a diffusion process. It is a solution to the SDE,

$$dx(t) = k(\mu - x)dt + \sigma dB(t)$$

with k and μ constants. A one dimensional process of the type described in equation 3.2 with $\mu(x(t), t) = k(\mu - x)$. It is an SDE with drift defined as the gradient of the potential function,

$$H(x) = \frac{1}{2}k(x - \mu)^2$$

When $\mu = 0$ this is the same potential as that of the simple harmonic oscillator in the high friction setting. Sample paths of an O-U process are shown in Figure 3.2 and the similarity can be seen by comparing these to the right most plot of Figure 3.1. The potential function of the O-U process can be described as providing attraction to the point μ. For this reason it is often referred to as a mean reverting process.

3.4 Estimation of potential

A primary objective of this section is to estimate and make inferences about the potential function. The remainder of this chapter will focus on the approach taken in this thesis to estimate the potential and some large sample results.

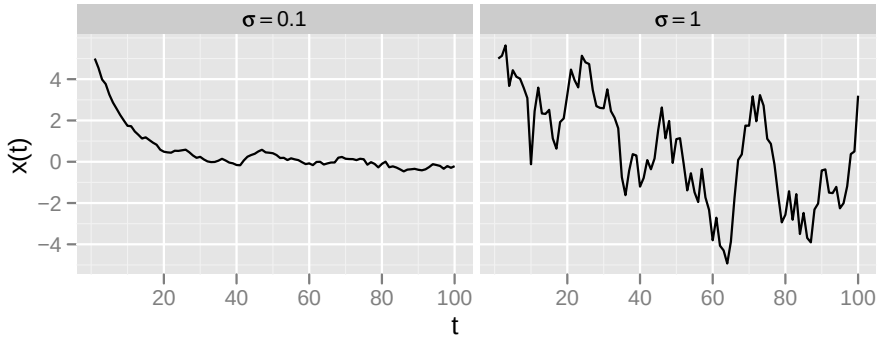


Figure 3.2: Examples paths of an Ornstein-Uhlenbeck process with varying noise variance. Both have $\mu = 0, x(0) = 5$.

Since the trajectory is measured at discrete times it is first modelled by a discrete time process. Under some conditions on the potential the resulting model can be written in the form of a linear or nonlinear regression. A theorem is presented that supports using the usual least squares estimator for the unknown parameters in the potential despite the presence of stochastic regressors. The application of the theorem relies on further assumptions on the potential to ensure that the discrete Markov process representing the movement has a stationary distribution.

I start by discussing the Euler approximation which will be applied to the continuous time movement model resulting in a discrete time approximation.

3.4.1 Discrete approximation

The Euler approximation will provide a way of simulating a SDE and an approximate model for the movement. This approximate model will be used in the development of the approach to estimating the potential function.

Given a stochastic differential equation,

$$d\mathbf{x}(t) = \boldsymbol{\mu}(\mathbf{x}(t), t)dt + \boldsymbol{\sigma}(\mathbf{x}(t), t)d\mathbf{B}(t),$$

the **Euler approximation** is a stochastic process $y(t)$ satisfying the iterative scheme,

$$\mathbf{y}(t_{n+1}) = \mathbf{y}(t_n) + \boldsymbol{\mu}(\mathbf{y}(t_n), t_n)(t_{n+1} - t_n) + \boldsymbol{\sigma}(\mathbf{y}(t_n), t_n)(\mathbf{B}(t_{n+1}) - \mathbf{B}(t_n)) \quad n = 1, 2, \ldots$$

given a starting point, y_0, where $\mathbf{B}(t)$ is a standard multivariate Brownian motion (Kloeden and Platen, 1992). For Brownian motion the increments $\mathbf{B}(t_{n+1}) - \mathbf{B}(t_n)$ are simply independent Gaussian random variables with mean zero and variance $t_{n+1} - t_n$. These random increments can be simulated by standard computer programs and the recursion provides one way to simulate a given SDE.

As an example, the O-U processes in Figure 3.2 were simulated using the Euler approximation with equispaced time points, $t_{i+1} - t_i = 1$, initial point $x(0) = 5$ and $k = 0.1$ resulting in the recursion,

$$x(t+1) = x(t) + kx(t) + \sigma\epsilon_t, t = 1, \ldots, 100$$

where the ϵ_t are randomly generated variates with iid pseudo Normal(0,1) distribution obtained from the `rnorm` command in R.

Higher order approximations providing quicker convergence are possible. However, I prefer the Euler approximation because of its interpretability and because it leads to estimation methods that are computationally simple.

3.4.2 Using the Euler approximation

The Euler approximation provides a route for approximating the continuous time movement diffusion by a discrete time Markov process. Applying the Euler approximation to the 1 dimensional case of Equation 3.4, $dx(t) = -H'(x)dt + \sigma dB(t)$, at the observation times $t_i, i = 1, \ldots, n$, gives

$$x(t_{i+1}) - x(t_i) = -H'\left(x(t_i)\right)\left(t_{i+1} - t_i\right) + \sigma\sqrt{t_{i+1} - t_i}Z_{i+1}, \quad i = 1, \ldots, n-1,$$

where Z_i are bivariate standard Normal random variates. By dividing through by $\sqrt{t_{i+1} - t_i}$ and assuming the potential is linear in its parameters this equation can be written in the form of a linear regression (see Section 3.5 for details). One might consider estimating the potential parameters by least squares. The primary difference between this and classical linear regression is that the covariates depend on past values of the response. A theorem that deals with this special case is now introduced and the large sample properties of the least squares estimates may be obtained.

3.5 Lai-Wei Theorem

Theorem 2 (Lai and Wei (1982))**.** *Consider the regression equation,*

$$Y_n = X_n\beta + \boldsymbol{\epsilon}_n$$

where $Y_n = (y_1, y_2, \ldots, y_n)^T$ are responses observed with respect to the design variable X_n an $n \times p$ matrix where each row i contains $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$, β is a vector of p parameters and $\boldsymbol{\epsilon}_n = (\epsilon_1, \ldots, \epsilon_n)$ unobservable errors.

If the following conditions hold:

1. *The sequence $\{\boldsymbol{\epsilon}_n\}$ is a martingale difference sequence with respect to an increasing set of σ-fields $\{\mathcal{F}_n\}$.*

2. $\sup_n \mathbb{E}\left(|\epsilon_n|^\alpha \middle| \mathcal{F}_{n-1}\right) < \infty$ *a.s. for some* $\alpha > 2$.

3. $\lim_{n\to\infty} \mathbb{E}\left(\epsilon_n^2 \middle| \mathcal{F}_{n-1}\right) = \sigma^2$ *a.s. for some constant* σ.

4. $\mathbf{x}_n$ *is* $\mathcal{F}_{n-1}$ *measurable.*

5. *There exists a non random matrix symmetric positive definite matrix* B_n *for which*

$$B_n^{-1}(X_n^T X_n)^{1/2} \to_P I_p \quad \textit{and} \quad \max_{1\le i\le n} \|B_n^{-1}\mathbf{x}_i\| \to_P 0,$$

where $\to_P$ *denotes convergence in probability as* $n \to \infty$.

Then, the usual least squares estimate $\mathbf{b}_n = (X_n^T X_n)^{-1} X_n^T Y_n$ *has an asymptotically normal distribution in the sense that,*

$$(X_n^T X_n)^{1/2}(\mathbf{b}_n - \beta) \to_D \mathbf{N}(0, \sigma^2 I_p).$$

Applying this to the movement model is just a matter of checking the conditions. The following theorem will be useful in showing that certain processes satisfy condition 5.

3.5.1 Law of large numbers for discrete parameter Markov process

Theorem 3 (Billingsley (1961) (Thm 1.1 pg 6))**.** *If for each* $\theta \in \Theta$ *the stationary distribution, which by assumption exists and is unique, has the property that for each* $s \in S$ *(state in state space)* $p_\theta(s,\cdot)$ *is absolutely continuous with respect to* $p_\theta(\cdot)$*, then for any* $\theta \in \Theta$ *the process* $\{x_n\}$ *is ergodic if the initial distribution is the stationary one. No matter what the initial distribution is, if* ϕ *is measurable* $\mathcal{F}_x \times \mathcal{F}_x$ *and if* $\mathbb{E}_\theta\big(|\phi(x_1,x_2)|\big) < \infty$*, then*

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \phi(x_k, x_{k+1}) = \mathbb{E}_\theta\big(|\phi(x_1,x_2)|\big)$$

with probability 1.

A special case of this theorem will be useful for satisfying condition 5 by letting $\phi(x_k, x_{k+1}) = x_k^2$. The above theorem requires the existence of a stationary distribution for the movement model. The following theorem provides some conditions on the potential that will help in this regard.

3.5.2 Relationship between invariant distribution and potential function

For the special case of SDE's in equation 3.4 with the additional assumptions that $\boldsymbol{\sigma} = \sigma I$ and the potential is not time dependent there is a useful result that relates the potential function to the invariant distribution.

Theorem 4 (Aldous (1989) (I3 pg 169)). *Let, $H : \mathbb{R}^d \to \mathbb{R}$. If,*

$$d\mathbf{x}(t) = -\nabla H(\mathbf{x}(t))dt + \sigma I d\mathbf{B}(t)$$

and

$$\int \exp\left(-2H(x)/\sigma^2\right)dx < \infty$$

then the invariant distribution of x is,

$$\pi_\sigma(x) = c \exp\left(\frac{-2H(\mathbf{x})}{\sigma^2}\right)$$

with c a constant.

As an example, in the case of the O-U process, the invariant distribution is

$$\pi_\sigma(x) = c \exp\left(\frac{-k(x-\mu)^2}{\sigma^2}\right)$$

which is recognized as a normal distribution with mean μ and variance $\sigma^2/2k$.

3.6 Lai-Wei applied to the movement model

3.6.1 The 1D case

Consider the model laid out in Equation 3.4. The simplest case is for a univariate process, $dx = -H'(x)dt + \sigma dB$. Consider the Euler approximation of this process at the fixed observation times $t_i, i = 1, \ldots, n$,

$$x(t_{i+1}) - x(t_i) = -H'\left(x(t_i)\right)\left(t_{i+1} - t_i\right) + \sigma\sqrt{t_{i+1} - t_i}Z_{i+1}, \quad i = 1, \ldots, n-1,$$

where Z_{i+1} is distributed as $N(0,1)$. Let $\mathcal{F}(t)$ be a filtration for $B(t)$ representing the information available up to and including time t and write $\mathcal{F}(t_i)$ as $\mathcal{F}_i$, then Z_{i+1} is $\mathcal{F}_{i+1}$ measurable (since it depends only on $B(t_i)$ and $B(t_{i+1})$). Divide through by $\sqrt{t_{i+1} - t_i}$ to get,

$$\frac{x(t_{i+1}) - x(t_i)}{\sqrt{t_{i+1} - t_i}} = -H'\left(x(t_i)\right)\sqrt{t_{i+1} - t_i} + \sigma Z_{i+1}.$$

Consider a potential function linear in p-dimensional vector β, for example

$$H(x) = f(x)^T\beta.$$

Then the model becomes,

$$\frac{x(t_{i+1}) - x(t_i)}{\sqrt{t_{i+1} - t_i}} = -\sqrt{t_{i+1} - t_i} f'\left(x(t_i)\right)^T \beta + \sigma Z_{i+1}.$$

It is direct to see that by letting,

$$y_i = \frac{x(t_{i+1}) - x(t_i)}{\sqrt{t_{i+1} - t_i}}, \quad x_i = -\sqrt{t_{i+1} - t_i} f'\left(x(t_i)\right)^T \quad \text{and } \epsilon_i = \sigma Z_{i+1}$$

the model is in the form required for the Lei-Wei theorem. We need only check that the conditions are satisfied. Relabel the filtration so that $\mathcal{F}_i = \mathcal{F}_{n-1}$. In this case ϵ_i are $\mathcal{F}_n$ measurable, i.i.d Normal$(0, \sigma^2)$ so they have expectation zero, finite moments and constant variance, σ^2, satisfying conditions 1, 2 and 3. Condition 4 is satisfied since x_i is $\mathcal{F}_{n-1}$ measurable, that is we only need information up until time t_i to calculate the gradient term.

For condition 5, assume the potential is such that the discrete process above has a stationary distribution. Then using the LLN for discrete parameter Markov processes (section 3.5.1) the required sum will converge in probability to a constant and this can be used to obtain an appropriate B_n.

With the conditions satisfied the least squares estimate of β has the asymptotic approximation for large n,

$$\mathbf{b}_n \dot{\sim} N(\beta, \sigma^2 (X_n^T X_n)^{-1}),$$

where $\dot{\sim}$ denotes "is approximately distributed as" and

$$X_n = -\begin{pmatrix} \sqrt{t_2 - t_1} f_1'(x(t_1)) & \dots & \sqrt{t_2 - t_1} f_p'(x(t_1)) \\ \vdots & \ddots & \vdots \\ \sqrt{t_n - t_{n-1}} f_1'(x(t_{n-1})) & \dots & \sqrt{t_n - t_{n-1}} f_p'(x(t_1 n - 1)) \end{pmatrix}$$

3.6.2 The bivariate case

I now consider the bivariate case, $\mathbf{r} \in \mathbb{R}^2$,

$$d\mathbf{r}(t) = -\nabla H(\mathbf{r}) dt + \boldsymbol{\sigma} d\mathbf{B}(t),$$

where $\mathbf{r}(t) = (x(t), y(t))^T$. The argument largely follows that of the 1D case. One problem arising is how to deal with the structure of the diffusion coefficient, $\boldsymbol{\sigma}$, which is now a matrix. The potential is again assumed to be linear, there is a known function $f : \mathbb{R}^2 \to \mathbb{R}^p$, such that,

$$H(\mathbf{r}) = f(\mathbf{r})^T \beta.$$

The scaled Euler approximation is,

$$\frac{\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)}{\sqrt{t_{i+1} - t_i}} = -\nabla H\left(\mathbf{r}(t_i)\right)\sqrt{t_{i+1} - t_i} + \boldsymbol{\sigma}\mathbf{Z}_{i+1}.$$

where $\boldsymbol{\sigma}$ is a 2×2 matrix, and $\mathbf{Z}$ is a 2 dimensional Normal random variable with mean zero and covariance matrix I.

In the special case that,

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{pmatrix},$$

the Euler approximation reduces to,

$$\begin{aligned}
\frac{x(t_{i+1}) - x(t_i)}{\sqrt{t_{i+1} - t_i}} &= -f_x\left(\mathbf{r}(t_i)\right)^T\sqrt{t_{i+1} - t_i}\beta + \sigma_x Z_{i+1} \\
\frac{y(t_{i+1}) - y(t_i)}{\sqrt{t_{i+1} - t_i}} &= -f_y\left(\mathbf{r}(t_i)\right)^T\sqrt{t_{i+1} - t_i}\beta + \sigma_y W_{i+1} \quad i = 1, \ldots, n-1,
\end{aligned}$$

where $f_x(\mathbf{r}) = -\frac{\partial}{\partial x}f(\mathbf{r})$, $f_y(\mathbf{r}) = -\frac{\partial}{\partial y}f(\mathbf{r})$, and Z and W are independent standard Normals. This case occurs when the random motions in the vertical and horizontal directions are independent. If $\sigma_x = \sigma_y = \sigma$ then the required form for the Lai-Wei theorem maybe obtained directly by stacking the equations in x and y on top of each other. So that for example,

$$Y = \begin{pmatrix} \frac{x(t_2)-x(t_1)}{\sqrt{t_2-t_1}} \\ \vdots \\ \frac{x(t_n)-x(t_{n-1})}{\sqrt{t_n-t_{n-1}}} \\ \frac{y(t_2)-y(t_1)}{\sqrt{t_2-t_1}} \\ \vdots \\ \frac{y(t_n)-y(t_{n-1})}{\sqrt{t_n-t_{n-1}}} \end{pmatrix}, \quad X = -\begin{pmatrix} \sqrt{t_2-t_1}f_{x1}(\mathbf{r}(t_1)) & \ldots & \sqrt{t_2-t_1}f_{xp}(\mathbf{r}(t_1)) \\ \vdots & \ddots & \vdots \\ \sqrt{t_n-t_{n-1}}f_{x1}(\mathbf{r}(t_{n-1})) & \ldots & \sqrt{t_n-t_{n-1}}f_{xp}(\mathbf{r}(t_{n-1})) \\ \sqrt{t_2-t_1}f_{y1}(\mathbf{r}(t_1)) & \ldots & \sqrt{t_2-t_1}f_{yp}(\mathbf{r}(t_1)) \\ \vdots & \ddots & \vdots \\ \sqrt{t_n-t_{n-1}}f_{y1}(\mathbf{r}(t_{n-1})) & \ldots & \sqrt{t_n-t_{n-1}}f_{yp}(\mathbf{r}(t_{n-1})) \end{pmatrix}.$$

where $f_{\cdot i}(\mathbf{r}) = -\frac{\partial}{\partial \cdot}f_i(\mathbf{r})$. If $\boldsymbol{\sigma}$ isn't diagonal multiplying through by it's square root inverse allows the same decomposition.

3.6.3 Example: Quadratic

As an example of the calculation of the design matrix, consider a quadratic potential,

$$\begin{aligned} H(\mathbf{r}) &= b_{10}x + b_{01}y + b_{11}xy + b_{20}x^2 + b_{02}y^2 \\ &= \begin{pmatrix} x & y & xy & x^2 & y^2 \end{pmatrix} \begin{pmatrix} b_{10} \\ b_{01} \\ b_{11} \\ b_{20} \\ b_{02} \end{pmatrix} \end{aligned}$$

Then the gradient is,

$$-\frac{\partial}{\partial x}H(\mathbf{r}) = -\begin{pmatrix} 1 & 0 & y & 2x & 0 \end{pmatrix} \begin{pmatrix} b_{10} \\ b_{01} \\ b_{11} \\ b_{20} \\ b_{02} \end{pmatrix} \quad -\frac{\partial}{\partial y}H(\mathbf{r}) = -\begin{pmatrix} 0 & 1 & x & 0 & 2y \end{pmatrix} \begin{pmatrix} b_{10} \\ b_{01} \\ b_{11} \\ b_{20} \\ b_{02} \end{pmatrix}$$

So the design matrix is,

$$X = -\begin{pmatrix} \sqrt{\Delta_2} & 0 & \sqrt{\Delta_2}y(t_1) & \sqrt{\Delta_2}2x(t_1) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sqrt{\Delta_n} & 0 & \sqrt{\Delta_n}y(t_{n-1}) & \sqrt{\Delta_n}2x(t_{n-1}) & 0 \\ 0 & \sqrt{\Delta_2} & \sqrt{\Delta_2}x(t_1) & 0 & \sqrt{\Delta_2}2y(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \sqrt{\Delta_n} & \sqrt{\Delta_n}x(t_{n-1}) & 0 & \sqrt{\Delta_n}2y(t_{n-1}) \end{pmatrix}.$$

where $\Delta_i = t_i - t_{i-1}$

3.7 Discussion of the model and extensions

I have shown a potential linear in its parameters can be estimated by least squares with the usual large sample distributional properties under the following assumptions

- The potential function is known and linear up to set of parameters and of a form such that the solution to the SDE is a diffusion (i.e. satisfies the conditions of Theorem 1).
- The potential function is of a form that the resulting diffusion has a stationary distribution (satisfies the conditions of Theorem 4).
- The time steps are close enough that the Euler approximation is reasonable.

- The true form of σ is chosen.

In practice I will take the Euler form of the model,

$$\frac{\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)}{t_{i+1} - t_i} = -\nabla H\left(\mathbf{r}(t_i)\right) + \boldsymbol{\sigma}\sqrt{t_{i+1} - t_i}\mathbf{Z}_{i+1}.$$

and treat it as a regression problem. It has the nice interpretation that the empirical velocity in the next time step is approximately the gradient of the potential evaluated at the current location. The above work allows the use standard least squares results as a first approach but careful checking of the residuals will be necessary to determine the appropriateness of the assumptions. The residuals are naturally defined as

$$\mathbf{e}_i = \frac{\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)}{t_{i+1} - t_i} - -\nabla \widehat{H}\left(\mathbf{r}(t_i)\right), \quad i = 1, \ldots, n-1$$

where $\widehat{H}$ is the estimated potential function. Examining standardized residuals will be appropriate in many cases and they can be calculated by taking the above and dividing by $\sqrt{t_{i+1} - t_i}$. Bootstrapping standard errors will also be important to guard against assumption violations. As always care needs to be taken if estimates of the potential function are to be used in locations far from from observed trajectories.

A similar theorem to the Lai-Wei result is proved in Lai (1994) for nonlinear regression. This allows estimation of potentials with non-linear terms.

Chapter 4

Potential functions

In this chapter I describe various possibilities for the form of the potential function.

4.1 Parametric forms for emulating desired behaviour

One way to estimate the potential function is to choose a form that emulates the desired behaviour of the animal and estimate the missing parameters. A simple linear parametric form that can account for a point of attraction or repulsion is a **quadratic** in location,

$$H(x, y) = \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy$$

This allows for potentials that are planes, hollows or mounds. If there is no xy term ($\beta_5 = 0$) then it has a simple interpretation from the parameters. If both β_3 and β_4 are positive then $\left(-\frac{\beta_1}{2\beta_3}, -\frac{\beta_2}{2\beta_4}\right)$ is the point of attraction (if β_3 and β_4 are both negative it is a point of repulsion). The speed in a unit time step is proportional to the distance from the point of attraction. In reality an animal may have multiple points it is attracted to. One could consider the potential being time varying and allow a different quadratic attractor at different time points.

A direct extension of the quadratic model is a polynomial in location. A general polynomial of degree d can be written

$$\sum_{i,j:i+j\leq d} \beta_{i,j} x^i y^j$$

Choosing d becomes a balance between capturing the relevant amount of spatial variation and the amount of data while not introducing too wiggly a term. This is an example of using a basis for the potential. The generalized additive model discussed later is another example where the coefficients are penalized to avoid excessive wiggliness.

A problem with the quadratic attractor is that the attraction is greater (in terms of speed) to the point of attraction the farther you are from it. In reality one might expect an animal to have a constant speed toward an attractor or some limit of perception: that

an animal isn't aware of an attractor if it is greater than a certain distance from it. This suggests forms like

$$H(r) = \begin{cases} \beta\frac{1}{2}d_A(\mathbf{r})^2 & d_A(\mathbf{r}) \leq c \\ \beta c\left(|d_A(\mathbf{r})| - \frac{c}{2}\right) & d_A(\mathbf{r}) > c \end{cases}$$

where $d_A(\mathbf{r})$ is the distance to the attractor (or distance to the closest point of a region of attraction). The speed toward the attractor is constant when the animal is greater than a constant, c, from the point of attraction and proportional to distance when it is closer than c. Something like the integral of the Tukey Biweight would allow a flat surface outside a radius c from the attractor,

$$H(r) = \begin{cases} \beta\frac{1}{2}d_A(\mathbf{r})^2\left(1 - \frac{d_A(\mathbf{r})^2}{c^2} + \frac{1}{3}\frac{d_A(\mathbf{r})^4}{c^4}\right) & d_A(\mathbf{r}) \leq c \\ 0 & d_A(\mathbf{r}) > c \end{cases}$$

A point of attraction could be time varying. For example it might be the location of another individual: this would allow attraction between animals (or equivalently repulsion). This will be explored further in Chapter 8 where two sharks have a common position based potential in addition to a potential on the distance between the two animals. The point of attraction could also be an area of attraction, for example a marine reserve, or equivalently an area of repulsion, for example land for a ocean going animal.

Many parametric forms are possible for the potential. The main requirement is the availability of its partial derivatives in x and y. When the partial derivatives are not readily available analytically they can be approximated by using differences. For example, for a potential function linear in its parameters,

$$\frac{\partial}{\partial x}\beta H(x,y) \approx \beta\frac{H(x+\delta,y) - H(x,y)}{\delta}$$

for small δ.

4.2 Generalized additive models as a non parametric approach

Another approach to estimating the potential function is to choose a flexible form for the potential and let the data speak for themselves. Generalized additive models will be used as a non parametric model for potentials in this thesis. Their attractiveness lies in providing a simple approach to estimating smooth functions. The approach advocated by Simon Wood in Wood (2006) and implemented in the R package `mgcv` is particularly nice as it provides a way to objectively choose the amount of smoothing based on the data. Following Wood (2006), the approach used in this thesis to estimate smooth potentials is to represent them by a thin plate regression spline basis. The smoothness is regulated by penalizing the basis coefficients based on the "wiggliness" of their corresponding basis functions. The smoothness

is chosen by generalized cross validation. I will summarise the details of the implementation here but stress that any set of basis functions could be used in a similar way to estimate a smooth non-parametric potential.

Generalized additive models assume the response of interest is an additive combination of smooth functions of relevant covariates. For example,

$$y_i = f(x_i, z_i) + g(v_i) + \epsilon_i$$

where y_i is the response of interest, x_i, z_i and v_i are covariates, f and g are some unknown smooth functions and ϵ_i a mean zero error term.

There are three components to a generalized additive model,

- the method used to represent the smooth functions,
- the way to adjust the smoothness of the functions,
- and the way to choose the appropriates smoothness based on the data at hand.

Here I will represent the functions using thin plate regression splines, adjust the smoothness by imposing a penalty on the second derivations of the splines and choose the smoothness by generalized cross validation.

Thin plate regression splines

Thin plate splines result from considering the model

$$y_i = f(\mathbf{x}_i) + \epsilon_i$$

where f is a smooth function and ϵ_i a random error term with the intent to estimate f by minimizing

$$\hat{f} = \text{argmin}_f \|y - f\|^2 + \lambda J(f)$$

where J is a penalty on the wiggliness of f and for $\mathbf{x} \in \mathbb{R}^2$ is defined as,

$$J(f) = \int\int \left\{ \left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 \right\} dx_1 dx_2 \tag{4.1}$$

Given this setup the general functional form for f can be found up to a set of parameters. The resulting functional form defines **thin plate splines**. The problem of finding f then reduces to estimation of the unknown parameters by penalized least squares. The downside of thin plate splines is that there are as many parameters to be estimated as data points. This can be computationally expensive. A compromise is to reduce the number of parameters to be estimated by taking an eigen truncation of the relevant design matrix. These low rank

approximations to thin plate splines are known as **thin plate regression splines**. As a basis for smooth regression they have the advantages that no knots need to be chosen, they are computationally cheap, they extend easily to functions of multiple variables and have some nice optimality properties.

Estimation of smoothing parameter

The parameter λ controls the amount of smoothing. It can be estimated by minimizing the generalized cross validation score. The computational details are complicated and won't be discussed here. More details of the procedure can be found in Wood (2006) (Section 4.5).

4.2.1 Setting up the potential model

Allow the potential function to be represented by a basis, $\{f_j(x, y)\}$. Then

$$H(x, y) = \sum_{j=1}^{k} \beta_j f_j(x, y)$$

The terms for each i stacked for the design matrix in the least squares regression are

$$\frac{\partial}{\partial x} f_j(x_i, y_i) \text{ and } \frac{\partial}{\partial y} f_j(x_i, y_i)$$

If the partial derivatives are readily available then they can be directly evaluated at the observed locations. If the partial derivatives are complicated, as in the thin plate regression spline case, then simple differences can be used as an approximation, and the terms stacked in the design matrix become,

$$\frac{f_j(x_i + \delta, y_i) - f_j(x_i, y_i)}{\delta} \text{ and } \frac{f_j(x_i, y_i + \delta) - f_j(x_i, y_i)}{\delta}$$

for small δ. For the generalized additive model case penalized least squares is used to estimate the β_j where the penalization is of the form defined in 4.1.

4.3 Boundaries

Boundaries on the animals movement deserve some attention. For ocean dwelling whale sharks land provides a barrier. Here I discuss a few ways to incorporate boundaries. Also see Brillinger (2003). One approach is to include a term that acts as repulsion from the coast, for example

$$H(\mathbf{r}) = \frac{C}{d_c(\mathbf{r})^\alpha}$$

as used in Brillinger et al. (2006a). Any function that increases rapidly enough as the animal approaches the boundary will work. This can be used in simulation to avoid the whale sharks moving onto land, although due to the discreet nature of simulation a step may still result in the whale shark being on land. One possibility in this case is to move that position to the nearest on the coast or halve the step size until the position is back in the ocean.

Soap film smoothing (Wood et al., 2008) provides an extension to the penalization of the thin plate splines in the gam methodology described above for the situation when the surface is bounded. Instead of the form in 4.1 the penalization takes the form

$$J_\Omega(f) = \int\int_\Omega \left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2}\right)^2 dx_1 dx_2$$

where Ω defines the area of interest. For the whale sharks this would be the ocean.

4.4 Adding covariates

Covariates can be integrated directly into the potential function. For example, the surface of the ocean floor could define a potential surface. Including it as a term in the potential function allows for preference to move toward deeper (if the coefficient is positive) or shallower (if the coefficient is negative) water. An example of this approach is used in Brillinger and Stewart (2010) where sea surface currents and winds were used in the movement model to model the movement of a drifting tag on the surface of the ocean. This approach will be used in Chapter 7 to explore the influence of ocean covariates on a shark's movement.

Chapter 5

Measurement Error

The tags used in the Kenyan tracking study produce locations for the sharks that can be subject to particularly large errors. In this chapter I summarize the methods used to locate the sharks and previous work on the measurement error. I present a method for modelling the measurement explicitly and a way to estimate its parameters.

5.1 Methods of telemetry

Wildlife telemetry involves "methods of obtaining information on living free-ranging animals by remote means" (Pride and Swift, 1992). The study of tracks concentrates on obtaining an animal's location but many other measurements are of course possible. There are a number of available telemetry methods. The whale sharks in the Kenya study are tagged with pop-up archival tags which use light levels to estimate the location of the fish. This method and the errors associated with it are discussed in depth in the next section but I give a brief overview of some of the other technologies here.

VHF radio and acoustic tags work on the same principle. A tag is attached to an animal that transits a regular pulsing signal. An operator then uses a handheld antenna and receiver to pinpoint the direction to the animal from the operators current location. A series of bearings can be estimated to pinpoint the location by triangulation and is often confirmed by a visual sighting. Acoustic tags are used in the marine environment as sound travels through salt water better than the higher frequency radio waves. This tends to be a very labour intensive method and is impractical for long studies or animals that travel great distances.

Satellite tags (SAT) offer a completely remote system using transmitting tags in conjunction with the ARGOS satellite platform. Tags transmit in the UHF band at a set frequency. With at least two transmissions received by an ARGOS satellite in a short time period the location can be inferred from the Doppler shift of the frequency of the transmission. Locations obtained are often accompanied by a categorization of the amount of error in the estimate. In general the estimates with this method are quite accurate (within 150 m) however the tag

needs to have a clear view of the satellites which makes them unsuitable for animals that spend considerable amounts of time beneath the surface of the water.

GPS tags offer the best in location accuracy and frequency however they suffer from the same disadvantage as SAT tags in that they will not able to communicate with satellites when under water.

For marine animals that spend very little time near the surface the most successful technology has been pop-up satellite archival tags (PSAT or PATs). These tags are attached to the fish and programmed to record light levels and often temperature and pressure until a specified release date. Once released the tags float to the surface and transmit their recorded data to the ARGOS satellites which relay the data to an establishment which computes estimates of the locations. The daily locations of the tagged fish are estimated using the recorded light levels, as explained in the next section, and the results are sent to the investigator.

5.2 Calculating position from light levels

While the sun has been used for navigation for centuries (Bowditch (1821) pages 120 & 168), Hunter (1986) was the first to suggest logging light levels as a method for tracking marine animals, and Delong et al. (1992) and Stewart and DeLong (1995) first applied the method to a mobile marine animal.

Hill (1994) provides a good overview of using light levels to calculate the position of tagged animals and the pertinent sources of error. The general idea is well illustrated by considering the locations experiencing sunrise or sunset at the same time, A. The places are illustrated by the circle on the leftmost diagram in Figure 5.1. At a time B, about 12 hours later, a new set of locations are experiencing sunrise or sunset (the middle diagram). The locations experiencing sunrise and sunset at both times A and B lie at the intersection of the two circles. At the equinox, when the length of night and day are the same for all locations, the intersection of the two circles will be a circle of points and it is impossible to determine latitude. Otherwise, there are only two points on the Earth experiencing sunrise and sunset at times A and B, one of which experienced sunrise at time A and the other experienced sunset at time A.

The PSATs/PATs record light levels at a specified interval and are equipped with an on-board clock (standardized to some time zone, generally Universal Time (UT)). To determine location, first sunrise and sunset must be estimated from the time series of recorded light levels. There are multiple definitions for sunrise and sunset. The most useful here are civil dawn and civil dusk. These occur when the center of the sun is 6 degrees below the horizon. This is the time at which the sun can first be seen and at which it first disappears (the sun's radius being about 6 degrees when viewed from the Earth). They also correspond to the times at which light levels are changing most rapidly making them easier to estimate from the light level records.

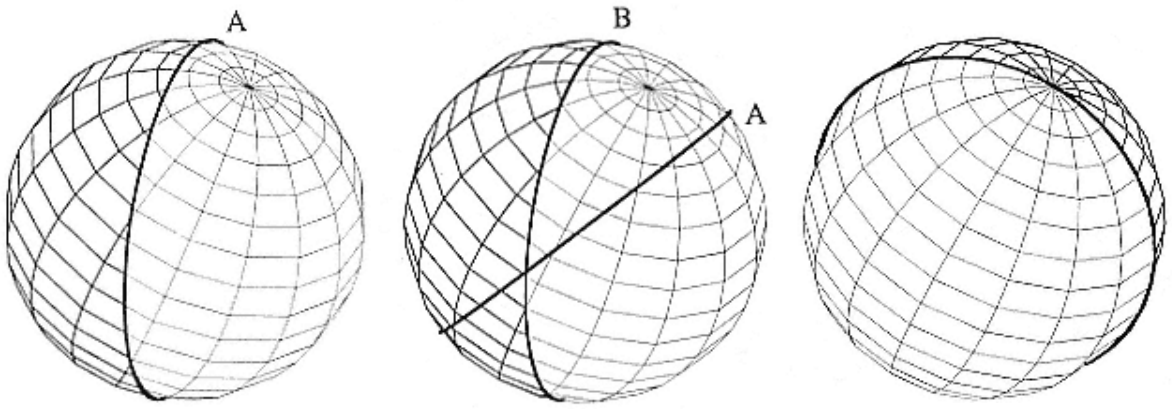


Figure 5.1: Illustration of geolocation using sunrise and sunset from Hill (1994). In the two left figures the sun is to the right, the shading representing day and night. The locations experiencing sunrise or sunset are on the boundary between the light and dark shading. Left: time A, center: time B, right: at the equinox.

Standard astronomical equations can be used to predict the time of sunrise and sunset at a given location (Meeus (1991), Chapter 14; Nautical Almanac). To estimate a location from a given sunrise and sunset an iterative process is generally used. An initial guess of longitude is local midnight in Universal Time (UT) $\times$ 15° and an initial guess for latitude is taken as a recent determined location of the tag. Sunrise and sunset are calculated for this postulated location and then a new position is generated by comparing these to the observed sunrise and sunset. The iteration is continued until the change between the postulated and observed sunrise and sunset is below a specified accuracy. Under the best conditions this results in one estimate of the animal's position per day.

5.2.1 Sources of error

The light from the sun is bent by refraction through the atmosphere. This means the time the sun appears to sink below the horizon may be some time after it is physically 6° below the horizon. The amount of refraction depends of atmospheric conditions and can never be fully accounted for. This results in errors in the determination of sunrise and sunset, leading to variability in the estimates of latitude and longitude of at least 0.7° and 0.32° respectively (Hill and Braun (2001)).

Around the time of the equinox, latitude is very poorly determined. Often another measurement can be used to help determine the latitude. A common choice is sea surface temperature, where the tag measurements can be compared to those generated by remote sensing and oceanic buoys.

If the clock on the tag is not accurate the times of sunrise and sunset will be incorrect. Hill (1994) comments that "one minute of inaccuracy in the estimate of dawn and dusk times

will generate an error of 0.25 degrees of longitude".

Another source of error comes from the tag being attached to a fish which may be at great depth or changing depth around dawn and dusk.

There have been a number of papers attempting to quantify the errors in position from this light level method of geolocation. Some of the relevant ones are reviewed in the next section.

5.3 Accuracy of light level location estimates

It is important to note that there are a variety PSAT/PAT tags and tag manufacturers. Often the manufacturer or service provider will not reveal all the details of their location estimation algorithm. This is the case for the Microwave Telemetry PTT-100 tags used in the Kenya study. While studies using any archival light level tag will give some idea of the magnitude of errors in the light level method, studies using the same tag, similar animals and a similar geographic location will give the best guide to the errors we may expect in the Kenya tagging study.

There are two common approaches in the literature to estimating the amount of error in the reported locations. The first is to track a fish using a light level archival tag and an additional method known to provide accurate locations. The archival tags reported positions are then compared to these "true" positions. The second approach is to model the PSAT reported positions and directly incorporate terms for the measurement error which are then estimated from the tracks.

Examples of the first kind include Musyl et al. (2001) and Teo et al. (2004). Musyl et al. (2001) attached tags to stationary mooring lines which ensured a constant known position (166°42' W, 24°00' N). They used Northwest Marine Technologies and Wildlife computers MK 7 tags. They reported standard deviations of the errors in latitude and longitude to be 3.08° and 0.24° respectively. They suggested the errors were non-normal but did not give indication of in what respect. Teo et al. (2004) provided a more relevant study as they attached tags to free ranging sharks. They compared locations from archival pop-up tags (PAT 2.0 Wildlife computers) and SPOT 2.0 satellite telemetry tags on Salmon and Blue sharks. They reported root mean square errors (RMSE) of 0.89° and 0.55° for longitude and 1.47° and 1.16° for latitude. The study area was a quite high latitude 40° N– 60° N.

Early examples of the second kind of study are Brillinger and Stewart (1998) and Brillinger (2000) where a measurement model was set down in addition to the movement model for a migrating elephant seal. Errors were estimated by Monte Carlo as about 1°.

The later examples of the second kind of studies mostly use a model derived from that presented in Sibert and Fournier (2001). The movement of the animal was assumed to be a random walk with drift. The evolution of the animals actual position was modelled as,

$$\mathbf{r}_t = \mathbf{r}_{t-1} + \mathbf{c}_t + \boldsymbol{\eta}_t \quad t = 1, \ldots, T$$

with $\boldsymbol{\eta}_t$ independent Normal$(0, \sigma^2 I)$ and $\mathbf{c}_t = (u, v)$. This is seen to be a simplification of the potential model with a planar potential and equi-spaced unit time steps. The observed locations, $\mathbf{r}'_t$, are the actual locations plus measurement error modelled as,

$$\mathbf{r}'_t = \mathbf{r}_t + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{\epsilon}_t$ represents the error in the measurement process and is assumed Normal with zero mean and covariance matrix,

$$\begin{pmatrix} \delta_x^2 & 0 \\ 0 & \delta_y^2 \end{pmatrix}$$

Sibert and Fournier (2001) suggested estimating the unknown parameters and obtaining the "most probable" track (the $\mathbf{r}_t$) using maximum likelihood in combination with the Kalman filter. The estimates of δ_x and δ_y provided estimates of the standard deviation of the error in the light level locations in longitude and latitude respectively.

Sibert et al. (2003) improved upon the model by allowing unequispaced times, and modelling the measurement error in latitude as a function of time to account for the increase in error around the equinoxes. They tested this method on light level archival tags (Northwest Marine Technologies and Wildlife Computers) in moored positions (the same position as Musyl et al. (2001)) as well as on free swimming bigeye tuna (tagged around 20°N). They estimated the latitude measurement error was maximal about 10 days before the equinox. Without adjusting for latitude errors changing through time they estimated δ_x as 0.16° and δ_y as 3.08° for a moored tag and δ_x ranged from 0.22° to 0.77° and δ_y ranged from 0.95° to 7.56° (although all but one value was less than 2.49°) for free ranging fish.

In a later study using the same tag as in the Kenya whale shark study (Microwave Telemetry PTT-100) Sibert et al. (2006) fit a similar state space model to a collection of 43 Atlantic bluefin tuna tracks over 3 tagging seasons. They allowed the model to have a changepoint estimated from the data to account for tags that continued reporting after detachment from the fish. Using the equinox singularity model (Sibert et al., 2003) for latitude error they reported median errors for the three seasons to be 0.41°, 0.35° and 0.93° for longitude and 1.03°, 2.09° and 2.11° for latitude.

Perhaps the most directly relevant study specifically designed to test the accuracy of the PTT-100 Microwave Telemetry tags with whale sharks is Wilson et al. (2007). They tagged a single fish at Ningaloo Reef, Australia (about 20°S) with two PTT-100 tags and a SAT tag. The SAT tags provided estimates of location when the shark was at surface using the ARGOS doppler shift method. These locations were reported with an error estimate. Those with an error estimate less than 1 km were taken as the true location of the shark and compared with the locations given by the light level method of the archival tags. Using the raw light levels the RMS error was 5.1° latitude and 2.0° longitude. The Kalman filter method of Sibert et al. (2003) and Sibert and Fournier (2001) was also applied to the light level data improving the RMSE to 2.97° latitude and 0.78° longitude, however they did not

report the parameter estimates of the model so it is hard to compare these calculated errors to those estimated by the state space model. Further improvements are reported by also including sea surface temperature via the methods in Nielsen et al. (2006).

Some general conclusions can be drawn from this collection of studies. The errors in latitude can be much larger than those in longitude. It seems at higher latitudes the errors in latitude are smaller. It seems reasonable to expect in the Kenya tagging studies that the standard deviation of the errors could be expected to be between 1–3° in latitude and 0.2–1° in longitude.

One would like a general approach that allows for a flexible specification of the potential function and allows for measurement error. The following presents a model that includes a potential that is linear in its parameters and additive Gaussian measurement errors (with the possibility of different variances and correlation). The model is recognized as a non linear state space model and the methods of estimation for these models can be applied. In particular I apply the Gibbs sampling method of Carlin et al. (1992) to produce samples from the posterior distributions of the parameters. The method is illustrated using simulated data from a quadratic potential model.

5.4 Derivation of Gibbs sampler for potential model

Let the potential be represented by a linear combination of known differentiable functions,

$$H(\mathbf{r}) = \sum_{j=1}^{k} \beta_j G_j(\mathbf{r})$$

where G_j and H are real valued. This allows a lot of flexibility in the form of the potential and includes the possibility of using a set of basis functions, such as splines of wavelets, for the $\{G_j\}$. Denote the gradients of each function by $g_j : \mathbb{R}^2 \to \mathbb{R}^2$,

$$\begin{aligned} h(\mathbf{r}) := -\nabla H(\mathbf{r}) &= -\sum_{i=1}^{n} \beta_j g_j(\mathbf{r}) = -\sum_{j=1}^{k} \beta_j \left(\frac{\delta}{\delta x} G_j(\mathbf{r}), \frac{\delta}{\delta y} G_j(\mathbf{r}) \right) \\ &= -g(\mathbf{r})\boldsymbol{\beta} \end{aligned}$$

where $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_k)$ and $g(\mathbf{r})$ is a $2 \times k$ matrix with columns containing $g_j(\mathbf{r})$. For equispaced times the models for the animal movement and location measurement are,

$$\mathbf{r}_t = \mathbf{r}_{t-1} + h(\mathbf{r}_{t-1}) + \boldsymbol{\epsilon}_t \quad \text{Movement model} \tag{5.1}$$

$$\mathbf{r}'_t = \mathbf{r}_t + \boldsymbol{\eta}_t \quad \text{Measurement model} \tag{5.2}$$

where $\boldsymbol{\epsilon}_t \sim N_2(0, \Sigma)$ and $\boldsymbol{\eta}_t \sim N_2(0, \mathcal{T})$ independently. I redefine h to contain the previous location by letting $\beta_0 = 1$, $g_0(\mathbf{r}_{t-1}) = \mathbf{r}_{t-1}$ and letting the other β's absorb the negative sign. Then I can write the movement model as,

$$\mathbf{r}_t = h(\mathbf{r}_{t-1}) + \boldsymbol{\epsilon}_t$$

The model is known up to the parameters $\boldsymbol{\beta}, \Sigma$, and $\mathcal{T}$. The combination of the movement model and measurement model are an example of a state space model with a linear measurement equation and a non-linear transition equation both with additive Gaussian errors.

The most general state space model takes the form,

$$\begin{aligned} \mathbf{x}_t &= F_t(\mathbf{x}_{t-1}, \mathbf{w}_t) \\ \mathbf{y}_t &= H_t(\mathbf{x}_t, \mathbf{v}_t) \qquad \text{(Shumway and Stoffer (2000), page 377)} \end{aligned}$$

where F_t and H_t are known functions (up to a finite dimensional set of parameters) and $\mathbf{w}$ and $\mathbf{v}$ are white noise processes. If $H_t = \Psi_t \mathbf{x}_{t-1} + \mathbf{w}_t$ and $F_t = A_t \mathbf{x}_{t-1} + \mathbf{v}_t$ with $\mathbf{w}_t$ and $\mathbf{v}_t$ Gaussian then the process is known as a dynamic linear model. If all the parameters are known the Kalman filter provides estimates for the underlying states, $\{\mathbf{x}_t\}$. If unknown, the parameters can be estimated by using the Kalman filter in conjunction with maximum likelihood (for example either by a Newton-Raphson or E-M algorithm). The model I wish to fit isn't linear, and I want the flexibility to allow the errors to be non-Gaussian, so these methods aren't applicable. There are extensions to the Kalman filter for non-linear non-normal models such as Kitagawa (1987) who used linear approximations to the density functions. Here I will take the approach of Carlin et al. (1992) who used a Bayesian approach and Gibbs sampling to obtain samples from the joint posterior of the states and parameters.

Let $R_n = (\mathbf{r_1}, \ldots, \mathbf{r}_n)$ and $R'_n = (\mathbf{r}'_1, \ldots, \mathbf{r}'_n)$. The goal is estimate the parameters of the potential function, measurement and movement variances as well as the true locations of the observed animal given the data observed. Taking a Bayesian approach I use the marginal posterior distributions to calculate a relevant estimate. That is using the distributions

$$p(\mathbf{r}_t | R'_n) \quad t = 1, \ldots, n; \quad p(\boldsymbol{\beta} | R'_n), \quad p(\Sigma | R'_n) \quad \text{and} \quad p(\mathcal{T} | R'_n)$$

A Markov Chain Monte Carlo (MCMC) approach uses an iterative method to draw samples from these distributions to approximate them. The Gibbs sampler is one MCMC approach that samples from their joint distribution utilizing only information on their complete conditional distributions.

The Gibbs sampler was first developed by Geman and Geman (1984). The algorithm explained here comes from Gelfand and Smith (1990).

Gibbs Sampler

Assume we have a collection of k random variables $U_1, \ldots, U_k$ and that we know their conditionals, $f(U_k \big| U_{j \neq k})$, and they can be sampled from readily. The Gibbs sampler works as follows, start with a set of initial values $U_1^{(0)}, \ldots, U_k^{(0)}$. Then sample $U_1^{(1)}$ from $f(U_1 \big| U_2^{(0)}, \ldots, U_k^{(0)})$. Next sample $U_2^{(1)}$ from $f(U_1 \big| U_1^{(1)}, U_3^{(0)} \ldots, U_k^{(0)})$. Continue with the remaining variables up to $U_k^{(1)}$ from $f(U_k \big| U_1^{(1)}, \ldots, U_{k-1}^{(1)})$. Repeat this process of drawing k samples i times to arrive at $U_1^{(i)}, \ldots, U_k^{(i)}$. Geman and Geman (1984) show that $(U_1^{(i)}, \ldots, U_k^{(i)}) \rightarrow^D (U_1, \ldots, U_k)$ as $i \rightarrow \infty$ and hence for any S, $U_S^{(i)} \rightarrow^D U_S$.

Applying this to the animal movement model (as in Carlin et al. (1992)), to sample from the joint distributions and hence the marginals given the data, the following conditional distributions are required,

$$\begin{aligned}
&\mathbf{r}_t \big| \mathbf{r}_{s \neq t}, R_n', \boldsymbol{\beta}, \Sigma, \mathcal{T} \\
&\boldsymbol{\beta} \big| R_n, R_n', \Sigma, \mathcal{T} \\
&\Sigma \big| R_n, R_n', \boldsymbol{\beta}, \mathcal{T} \\
&\mathcal{T} \big| R_n, R_n', \boldsymbol{\beta}, \Sigma
\end{aligned}$$

I now proceed to calculate these conditional distributions in this specific case. First prior distributions for the parameters need to be set down,

$$\begin{aligned}
(\beta_1, \ldots, \beta_k) &\sim \text{Normal}_k(\mu, W) \\
\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \quad \text{with} \quad \sigma^2 &\sim \text{InvGamma}(a, b) \\
\mathcal{T} &\sim \text{InvWishart}(m, \Psi)
\end{aligned}$$

The hyperparameters (μ, W, a, b, m, Ψ) could be chosen using prior information or in a way to be weakly informative. These priors are chosen for their convenience in calculating the required conditional distributions. A discussion of alternative priors is given later in the example.

The full likelihood for the specific model (say $x_0 = y_0$ fixed, as would be the case if a GPS fix is taken at the place an animal is first tagged) is,

$$p(R_n, R_n'; \Theta) = \prod_{t=1}^{n} \phi\left(\mathbf{r}_t | h(\mathbf{r}_{t-1}, \boldsymbol{\beta}), \Sigma\right) \phi\left(\mathbf{r}_t' | \mathbf{r}_t, \mathcal{T}\right)$$

where $\phi(\mathbf{x}|\mu, \Sigma)$ is the bivariate Normal density function (of $\mathbf{x}$) with mean μ and variance Σ. By Bayes rule the full conditionals will be proportional to the full Bayesian model

specification

$$p(R_n, R_n'; \Theta)p(\boldsymbol{\beta})p(\Sigma)p(\mathcal{T})$$

where $p(\boldsymbol{\beta})$, (Σ) and $p(\mathcal{T})$ are the prior distributions on the unknown parameters. To find the full conditionals it's just a matter of multiplying likelihood by priors, discarding irrelevant terms and normalizing. This is now done for the required conditionals.

Conditional distribution of states

$$\begin{aligned}
p(\mathbf{r}_t \big| \mathbf{r}_{s\neq t}, R_n', \Theta) &\propto \phi\left(\mathbf{r}_t | h(\mathbf{r}_{t-1}), \Sigma\right) \phi\left(\mathbf{r}_{t+1} | h(\mathbf{r}_t), \Sigma\right) \phi\left(\mathbf{r}_t' | \mathbf{r}_t, \mathcal{T}\right) \quad t = 1, \ldots, n-1 \\
&\propto \exp\Big(-\frac{1}{2}(\mathbf{r}_t - h(\mathbf{r}_{t-1}))^T \Sigma^{-1} (\mathbf{r}_t - h(\mathbf{r}_{t-1})) \\
&\quad -\frac{1}{2}(\mathbf{r}_{t+1} - h(\mathbf{r}_t))^T \Sigma^{-1} (\mathbf{r}_{t+1} - h(\mathbf{r}_t)) \\
&\quad -\frac{1}{2}(\mathbf{r}_t' - \mathbf{r}_t)^T \mathcal{T}^{-1} (\mathbf{r}_t' - \mathbf{r}_t) \Big) \\
&\propto \exp\Big(-\frac{1}{2}(\mathbf{r}_{t+1} - h(\mathbf{r}_t))^T \Sigma^{-1} (\mathbf{r}_{t+1} - h(\mathbf{r}_t)) \Big) \\
&\quad \times \exp\Bigg(-\frac{1}{2}\Big(\mathbf{r_t}^T (\Sigma^{-1} + \mathcal{T}^{-1}) \mathbf{r}_t - \mathbf{r}_t^T \left(\Sigma^{-1} h(\mathbf{r}_{t-1}) + \mathcal{T}^{-1} \mathbf{r}_t'\right) \\
&\quad\quad - \left(h(\mathbf{r_{t-1}})^T \Sigma^{-1} + \mathbf{r}_t'^T \mathcal{T}^{-1}\right) \mathbf{r}_t \Big) \Bigg) \\
&\propto w(\mathbf{r}_t) \times \phi(\mathbf{r}_t | B_t b_t, B_t) \quad \text{by completing the square}
\end{aligned}$$

where

$$\begin{aligned}
w(\mathbf{r}_t) &= \exp\Big(-\frac{1}{2\sigma^2} \|\mathbf{r}_{t+1} - h(\mathbf{r}_t)\|^2 \Big) \\
B_t^{-1} &= \Sigma^{-1} + \mathcal{T}^{-1} \quad \text{and} \\
b_t^T &= h(\mathbf{r}_{t-1})^T \Sigma^{-1} + \mathbf{r}_t'^T \mathcal{T}^{-1}
\end{aligned}$$

The $w(\mathbf{r}_t)$ term disappears for $t = n$. Since $0 \leq w(\mathbf{r}_t) \leq 1$ we can sample from this distribution using rejection sampling. First sampling from Normal$(B_t b_t, B_t)$ and accepting the number with probability $w(\mathbf{r}_t)$. This works well as long as $w(\mathbf{r}_t)$ is not too close to zero. In this case adaptive rejection sampling (Gilks et al., 1995) may be used. In `R` one can use `arms` in the package `dlm`, (Petris, 2010; Petris et al., 2009).

Conditional distribution of the movement variance

$$p(\sigma^2 \big| R_n, R'_n, \boldsymbol{\beta}, \mathcal{T}) \propto \left(\frac{1}{\sigma^2}\right)^{a+1} \exp \frac{-b}{\sigma^2} \times \prod_{t=1}^{n} \frac{1}{\sigma} \exp\left(-\frac{1}{2}(\mathbf{r}_t - h(\mathbf{r}_{t-1}))^T \frac{1}{\sigma^2} I(\mathbf{r}_t - h(\mathbf{r}_{t-1}))\right)$$
$$= \left(\frac{1}{\sigma^2}\right)^{a+n/2+1} \exp\left(\frac{-b}{\sigma^2} - \frac{1}{2}\sum_{t=1}^{n} \frac{\|\mathbf{r}_t - h(\mathbf{r}_{t-1})\|^2}{\sigma^2}\right)$$

With the result,

$$\sigma^2 \big| R_n, R'_n, \boldsymbol{\beta}, \mathcal{T} \sim \text{InvGamma}(a + n/2, b + 1/2 \sum_{t=1}^{n} \|\mathbf{r}_t - h(\mathbf{r}_{t-1})\|^2)$$

The effect of the prior parameters a and b, is to act like we have $2a$ observations of squared deviation b between the true position and expected position under the potential. The factor of 2 is curious but a consequence of having an two observations of σ^2 one from the x direction and one from the y direction. By choosing a small compared to n, and b not wildly far from the truth the prior will have little effect on the posterior.

Conditional distribution of the measurement variances

$$p(\mathcal{T} \big| R_n, R'_n, \boldsymbol{\beta}, \Sigma) \propto |\mathcal{T}|^{-(m+3)/2} \exp\big(-\text{trace}(\Psi\mathcal{T}^{-1})/2\big) \times$$
$$\prod_{t=1}^{n} |\mathcal{T}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{r}'_t - \mathbf{r}_t)^T \mathcal{T}^{-1}(\mathbf{r}'_t - \mathbf{r}_t)\right)$$
$$= |\mathcal{T}|^{-(m+n+3)/2} \times$$
$$\exp\left(-\frac{1}{2}\text{trace}(\Psi\mathcal{T}^{-1}) - \frac{1}{2}\sum_{t=1}^{n}(\mathbf{r}'_t - \mathbf{r}_t)^T \mathcal{T}^{-1}(\mathbf{r}'_t - \mathbf{r}_t)\right)$$
$$= |\mathcal{T}|^{-(m+n+3)/2} \times$$
$$\exp\left(-\frac{1}{2}\text{trace}(\Psi\mathcal{T}^{-1}) - \frac{1}{2}\sum_{t=1}^{n}\text{trace}\big((\mathbf{r}'_t - \mathbf{r}_t)^T \mathcal{T}^{-1}(\mathbf{r}'_t - \mathbf{r}_t)\big)\right)$$
since $\text{trace}(c) = c$
$$= |\mathcal{T}|^{-(m+n+3)/2} \times$$
$$\exp\left(-\frac{1}{2}\text{trace}(\Psi\mathcal{T}^{-1}) - \frac{1}{2}\sum_{t=1}^{n}\text{trace}\big((\mathbf{r}'_t - \mathbf{r}_t)(\mathbf{r}'_t - \mathbf{r}_t)^T \mathcal{T}^{-1}\big)\right)$$

So, that

$$\mathcal{T} \sim \text{InvWishart}\left(m+n, \Psi + \sum_{t=1}^{n}(\mathbf{r}'_t - \mathbf{r}_t)(\mathbf{r}'_t - \mathbf{r}_t)^T\right)$$

The effect of the prior parameters m and Ψ is to act like we have m observations of empirical covariance Ψ between the true position and the observed position. For the prior to be weakly informative we want m to be small compared to n.

Conditional distribution of the potential parameters

Since β_0 is fixed, here we work with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ and $g(\mathbf{r}) = (g_1(\mathbf{r}), \ldots, g_k(\mathbf{r}))$.

$$\begin{aligned}
p(\boldsymbol{\beta}\big|R_n, R'_n, \Sigma, \mathcal{T}) & \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\mu)^T W^{-1}(\boldsymbol{\beta}-\mu)\right) \\
&\quad \times \exp\left(-\frac{1}{2}\sum_{t=1}^{n}(\mathbf{r}_t - \mathbf{r}_{t-1} - g(\mathbf{r}_{t-1})\boldsymbol{\beta})^T\Sigma^{-1}(\mathbf{r}_t - \mathbf{r}_{t-1} - g(\mathbf{r}_{t-1})\boldsymbol{\beta})\right) \\
&\propto \exp\Bigg(-\frac{1}{2}\Big(\boldsymbol{\beta}^T W^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}^T W^{-1}\mu - \mu^T W^{-1}\boldsymbol{\beta} \\
&\quad + \sum_{t=1}^{n}\boldsymbol{\beta}^T g(\mathbf{r}_{t-1})^T\Sigma^{-1}g(\mathbf{r}_{t-1})\boldsymbol{\beta} - \boldsymbol{\beta}^T g(\mathbf{r}_{t-1})^T\Sigma^{-1}(\mathbf{r}_t - \mathbf{r}_{t-1}) \\
&\quad - (\mathbf{r}_t - \mathbf{r}_{t-1})^T\Sigma^{-1}g(\mathbf{r}_{t-1})\boldsymbol{\beta})\Big)\Bigg)
\end{aligned}$$

So, that the conditional posterior distribution of the parameter $\boldsymbol{\beta}$ is $\text{Normal}_k(Bb, B)$ where,

$$B^{-1} = W^{-1} + \frac{1}{\sigma^2}\sum_{t=1}^{n}\|g(\mathbf{r_{t-1}})\|^2 \text{ and } b^T = \mu^T W^{-1} + \frac{1}{\sigma^2}\sum_{t=1}^{n}(\mathbf{r}_t - \mathbf{r}_{t-1})^T g(\mathbf{r}_{t-1})$$

5.4.1 Non-Guassian errors

In their paper Carlin et al. (1992) allowed the error distribution to be non Gaussian by introducing two nuisance parameters λ_t and ω_t so that the transition density $x_t\big|x_{t-1}$ and the measurement density $y_t\big|x_t$ become scaled mixtures of normals. This allows the errors to be double exponential, logistic, exponential power or t densities. The added complexity is that the conditional distributions of λ_t and ω_t need to be computed and sampled from. This isn't implemented in this thesis but is an area for future work.

5.4.2 Using WinBUGS or JAGS

WinBUGS (Lunn et al., 2000) and JAGS (Plummer, 2003) are programs designed for Gibbs sampling and can be run standalone or through R interfaces. They allow easy adjustment of the error structures and prior distributions without the user having to re-derive the conditional distributions of the parameters.

The primary disadvantage is that they are limited in the complexity of functions that can be written. This places a limit on the types of potential functions that can easily be used. For example, they can be used for a quadratic potential function very easily but not for a potential function based on a complicated basis. Writing a custom Gibbs sampler in `R` has the advantage that there are many tools available for generating and evaluating useful bases and their derivatives, such as splines and wavelets.

In the work in this thesis I will use JAGS to estimate the parameters of the state space formulation, through the `rjags` (Plummer, 2011) package in `R`. Setting the model up in JAGS is simply a matter of setting down equations 5.2 and 5.3. An example model file for a quadratic potential can be found in Appendix A.

5.5 Illustration of parameter estimation in a quadratic potential model

The Gibbs sampler method of estimation will be illustrated using simulated data. Data were generated according to equations 5.1 and 5.2 with a quadratic potential model,

$$H(\mathbf{r}) = \beta_{10}x + \beta_{01}y + \beta_{20}x^2 + \beta_{02}y^2 + \beta_{11}xy$$

with the following parameters,

$$g(\mathbf{r}) = g(x,y) = \begin{pmatrix} x & -1 & 0 & -2x & 0 & -y \\ y & 0 & -1 & 0 & -2y & -x \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_{10} \\ \beta_{01} \\ \beta_{20} \\ \beta_{02} \\ \beta_{11} \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 3 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix},$$

$$\sigma^2 = 1, \quad \mathcal{T} = \begin{pmatrix} \tau_x^2 & \tau_x\tau_y\rho \\ \tau_x\tau_y\rho & \tau_y^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

The parameters were chosen somewhat arbitrarily but note that the potential has an attractor at $(-10, -10)$. A track of length $n = 100$ was generated with unit time steps starting from $\mathbf{r}_0 = (10, 10)$. The resulting simulated track is shown in Figure 5.2. The prior hyper-

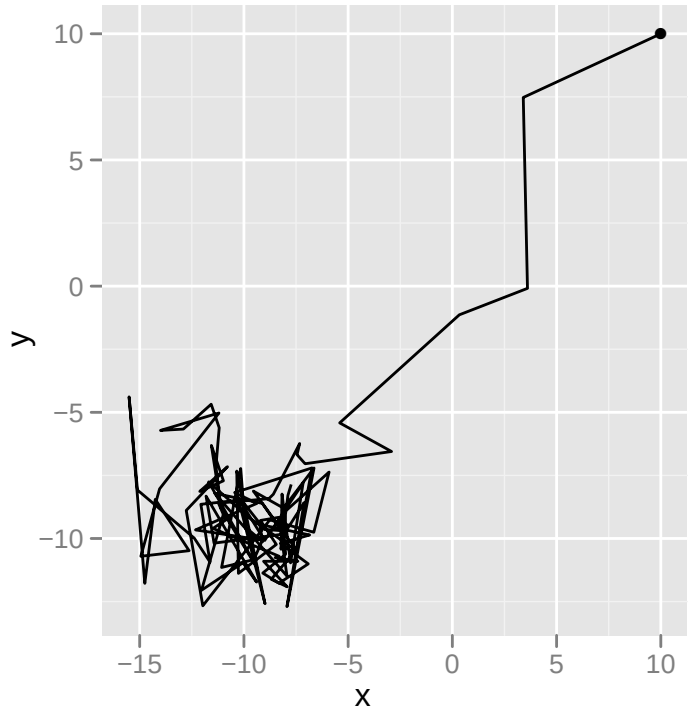


Figure 5.2: An example of the simulated track, under a quadratic potential. The initial position is marked with a point.

Table 5.1: Estimates and approximate 95% credible intervals estimated using Gibbs sampling for the simulated track

Parameter	Estimate	95% CI
β_{10}	2.26	(1.61, 2.96)
β_{01}	2.87	(2.14, 3.63)
β_{20}	0.052	(−0.029,0.116)
β_{02}	0.089	(0.053,0.136)
β_{11}	0.127	(0.079,0.185)
ρ	−0.12	(−0.46, 0.22)
σ	1.01	(0.73, 1.36)
τ_x	0.91	(0.56, 1.20)
τ_y	1.25	(0.93, 1.55)

parameters were set at

$$a = 0.001, b = 0.001, m = 3, \Psi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mu = \mathbf{0}, W = 100I$$

Three chains with random starting points were generated and allowed to run through 1000 iterations. Traces of 100 following iterations are shown in Figure 5.3. The chains appeared to be mixing well: the traces do not exhibit any trends and rapidly explore the extent of their range. An additional 10,000 samples were generated for each of the three chains, every 10th value was retained, resulting in a total of 3000 samples from the posterior distributions of the parameters of interest.

Estimate and confidence intervals for each parameter were calculated by taking the median and 2.5% and 97.5% percentiles from these posterior samples (Table 5.1).

The estimates and intervals for the quadratic potential looked acceptable. The intervals, except for β_{10}, contained the true parameter values. To evaluate the behavior of the estimates and confidence intervals the experiment was repeated by simulating multiple tracks with the same true parameters and comparing the intervals. One hundred tracks were simulated with the same parameters, the parameters were then estimated using the Gibbs sampler. To save computational time only one chain was used per simulation (examining the quantiles based on the three chains separately revealed differences only in the third significant digit). Credible intervals for each parameter over the 100 simulations are displayed in Figure 5.4. If intervals are "good" one would expect to see about 5 ± 3 intervals missing the true parameters (we note that the intervals are independent between simulations but not between parameters within a simulation). The intervals for the linear terms in the quadratic potential, β_{10} and β_{01}, seem appropriate, however there seem to be more intervals missing the true parameter than expected for the other parameters. This can be explained by the approximate nature of

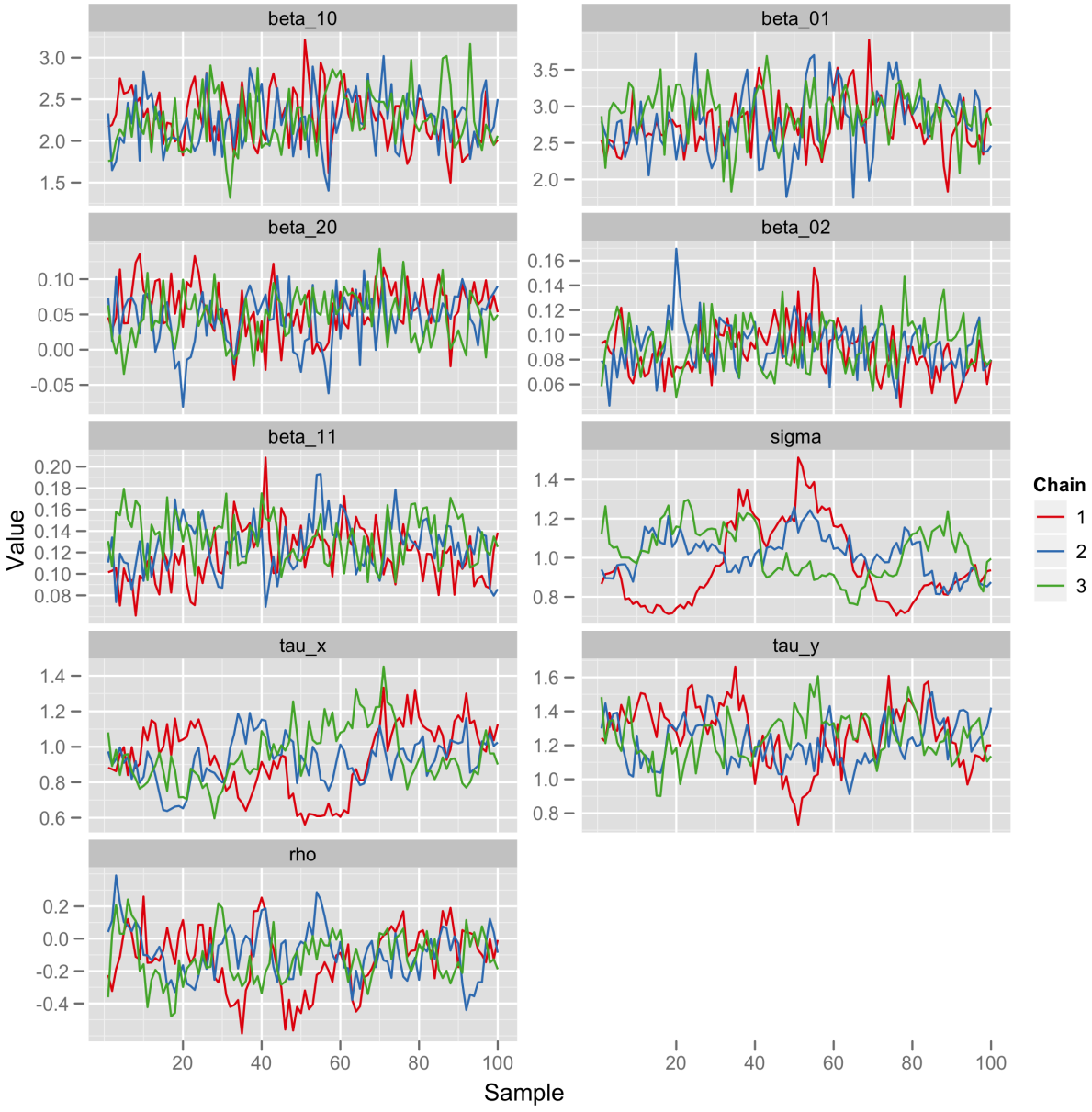


Figure 5.3: Trace of the progress of the Gibbs sampler for the three independent chains after discarding 1000 iterations.

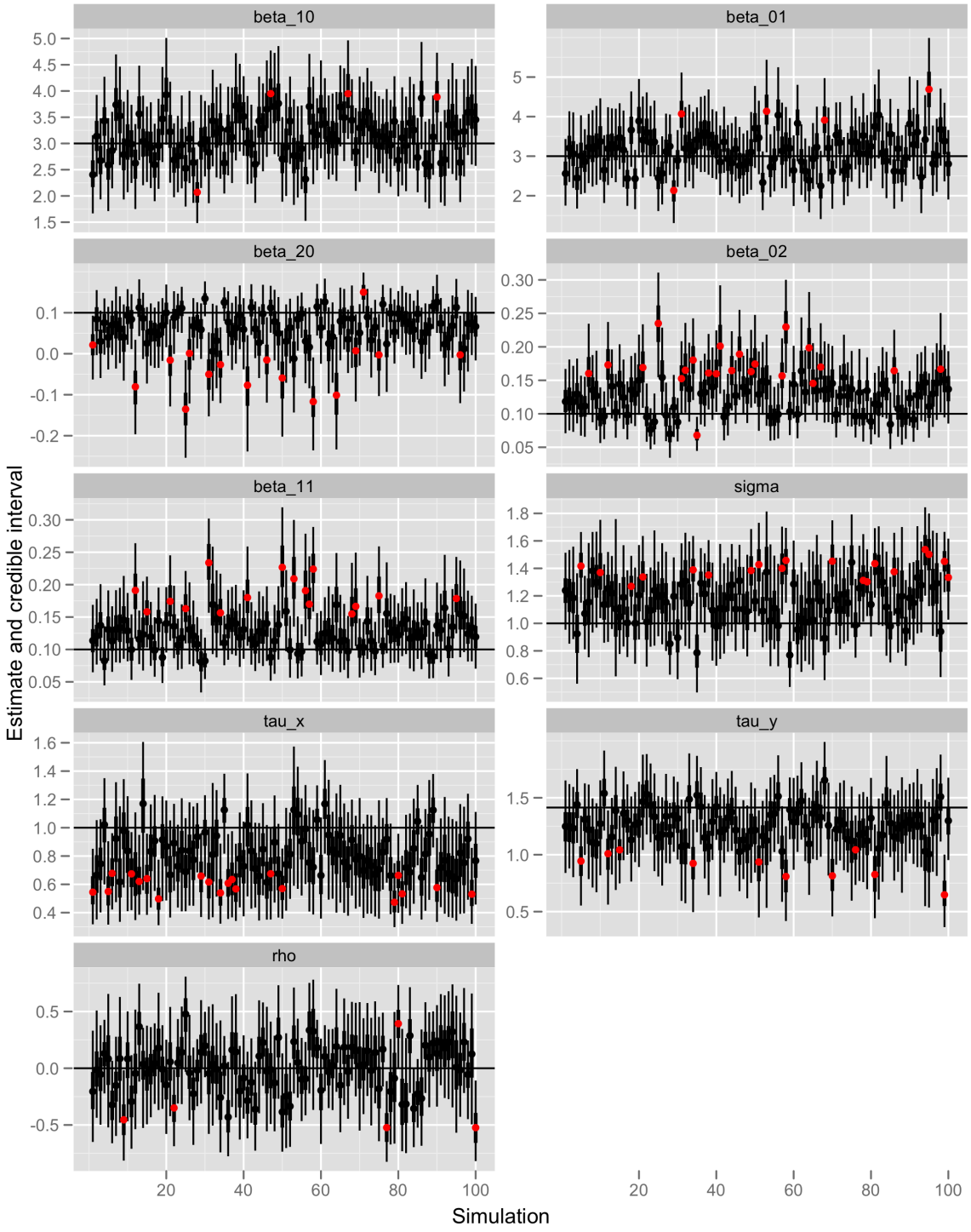


Figure 5.4: Credible intervals (50% heavy line, 95% in light line) estimated using the Gibbs sampler for 100 simulated tracks with 100 observations each ($n = 100$). Intervals that do not contain the true parameters are red.

the intervals. Repeating the experiment with tracks of length 200 showed not only that the intervals get narrower with larger samples, but that the intervals are more likely to contain the true value (see Figure 5.5).

5.5.1 Some alternative priors

It's important to check the choice of prior isn't overly influential in the analysis. Alternative parameters for the priors used above are,

$$a = 0.01, b = 0.01, m = 10, \Psi = 10 \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mu = \mathbf{0}, W = 10I \quad \textbf{More informative}$$

$$a = 1, b = 1, m = 10, \Psi = 10 \times \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \mu = (3, 3, 0.1, 0.1, 0.1), W = I \quad \textbf{Extra informative}$$

The "more informative" prior representing slightly more weight given to the prior than in the original analysis and the "extra informative" prior being representative of the case where we have some good prior knowledge of the parameters.

Gelman (2006) suggested using a Uniform prior distribution for error standard deviations. So, an alternative set of priors might be,

$$(\beta_1, \ldots, \beta_k) \sim \text{Uniform}_k(-B, B)$$

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \quad \text{with} \quad \sigma \sim \text{Uniform}(0, S)$$

$$\mathcal{T} = \begin{pmatrix} \tau_x^2 & \rho\tau_x\tau_y \\ \rho\tau_x\tau_y & \tau_y^2 \end{pmatrix} \quad \text{with} \quad \tau_x, \tau_y \sim \text{Uniform}(0, T) \quad \text{and} \quad \rho \sim \text{Uniform}(-1, 1)$$

The choice of B, S and T depends on the situation. Poor choices can lead to very poor mixing of the MCMC chains. Here, I illustrate this prior with $B = 5, S = 5$ and $T = 5$ which could be chosen quite readily just by examining the range of the trajectory.

Samples of the variance parameters (σ^2, τ_x^2, τ_y^2 and ρ) drawn from these three alternate priors along with samples from the original prior are shown in Figure 5.6. The analysis was run using these three priors and the posterior densities of the parameters of interest are shown in Figure 5.7. Despite the clear differences in priors the posteriors are very similar and inferences are not significantly changed by using one of these alternate priors.

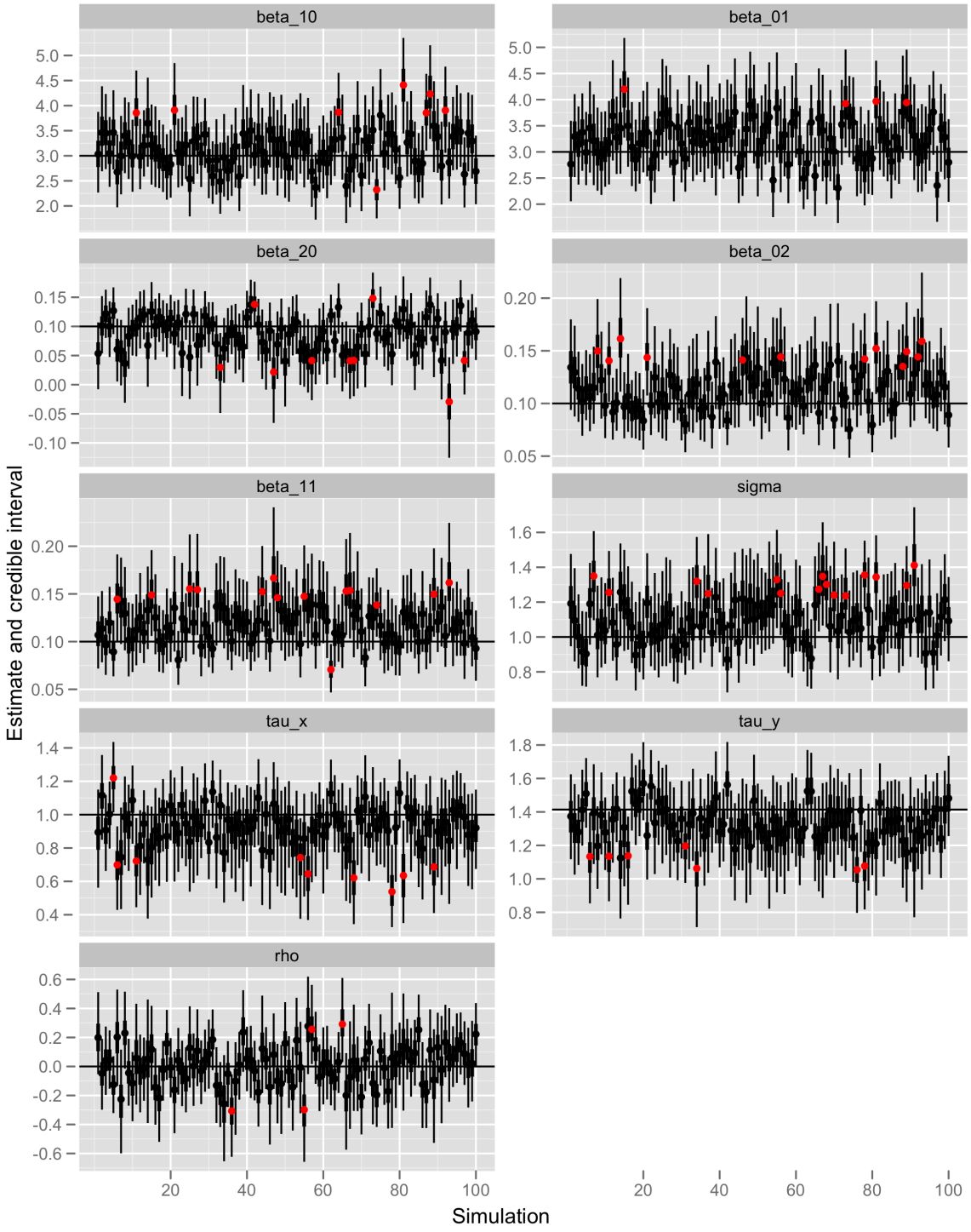


Figure 5.5: Credible intervals (50% heavy line, 95% in light line) estimated using the Gibbs sampler for 100 simulated tracks with 200 observations each ($n = 200$). Intervals that do not contain the true parameters are red.

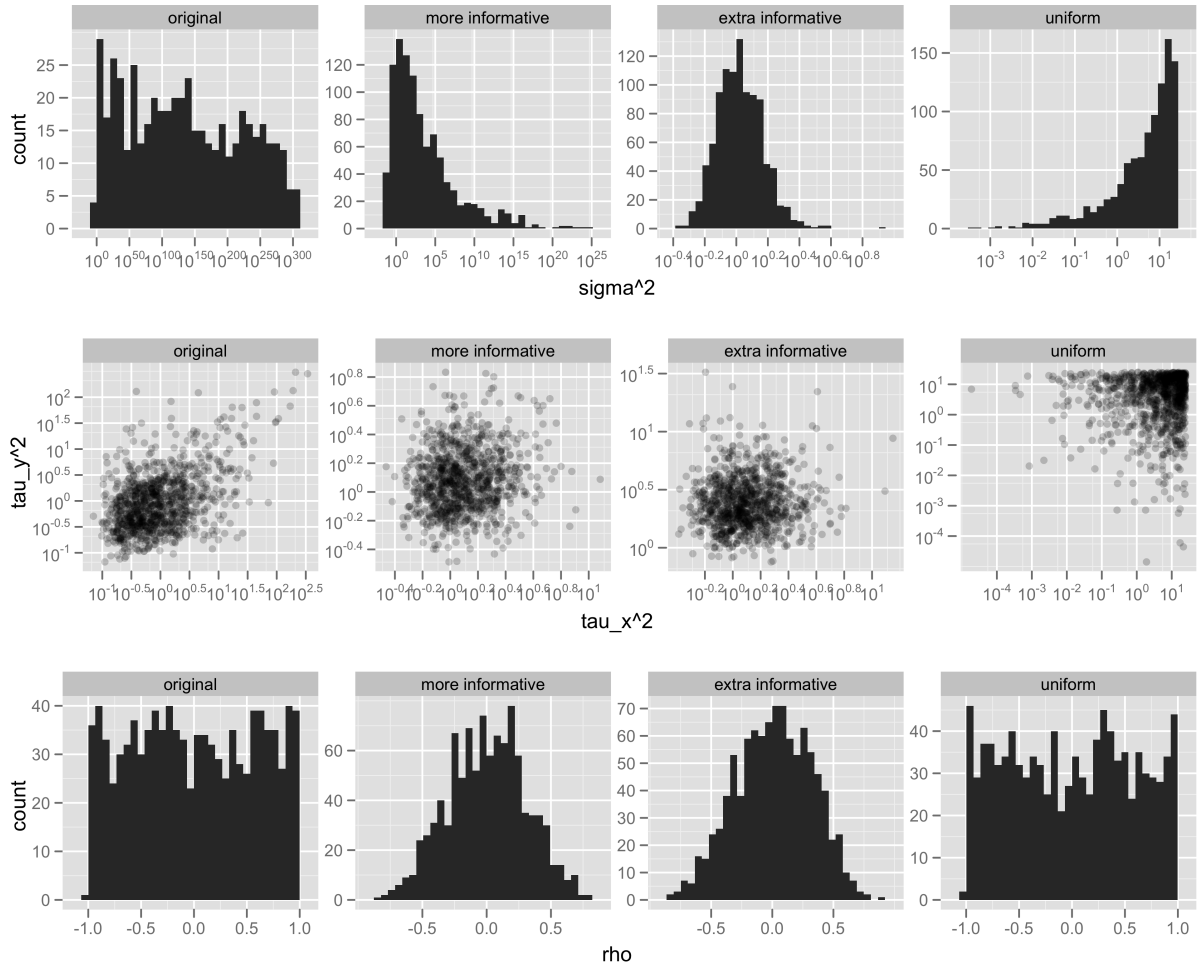


Figure 5.6: Samples from the four different prior distributions.

5.6 Extension to unequal times

When the positions occur at unequally spaced time points $t_i, i = 1, \ldots, n$, some adjustments can be made. The measurement model remains the same. The movement model becomes,

$$\mathbf{r}(t_i) = \mathbf{r}(t_{i-1}) + h(\mathbf{r}_{t-1})\Delta_i + \boldsymbol{\epsilon}_i \text{ where } \Delta_i = t_i - t_{i-1} \tag{5.3}$$

where ϵ_i now has variance $\Delta_i\Sigma$. The Gibbs sampler procedure follows as above resulting in the following conditional distributions. We separate the $\mathbf{r}(t_{i-1})$ term from $h(\mathbf{r}(t_{i-1}))$ in the notation that follows since it is not scaled by the time step.

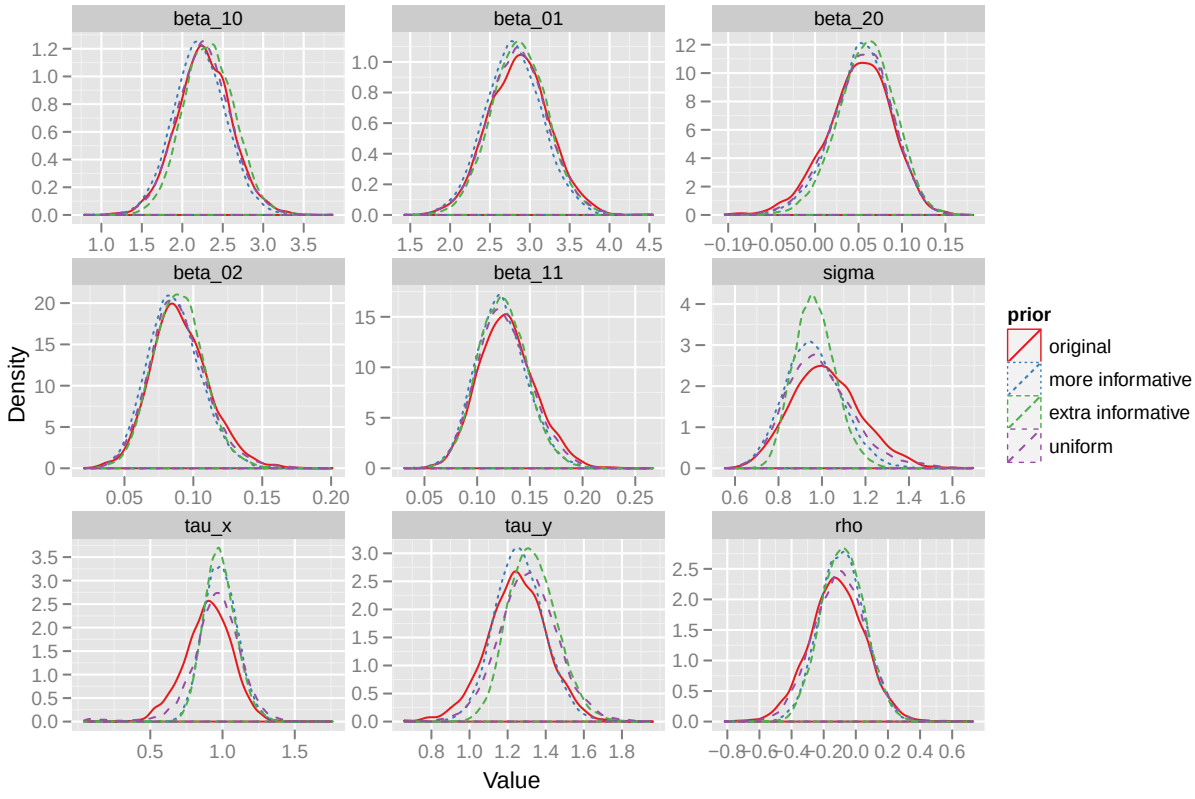


Figure 5.7: Posterior densities for the parameters of interest under different priors.

States

The conditional distribution of the states becomes,

$$p(\mathbf{r}(t_i)\big|\mathbf{r}(t_{j\neq i}), R_n', \Theta) \propto w(\mathbf{r}(t_i)) \times \phi(\mathbf{r}(t_i)|B_t b_t, B_t)$$

where

$$\begin{aligned} w(\mathbf{r}(t_i)) &= \exp\Big(-\frac{1}{2\sigma^2\Delta_{i+1}} \|\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i) - h(\mathbf{r}(t_i))\Delta_{i+1}\|^2 \Big) \\ B_t^{-1} &= (\Delta_i \Sigma)^{-1} + \mathcal{T}^{-1} \quad \text{and} \\ b_t^T &= \left(\mathbf{r}(t_{i-1}) + h(\mathbf{r}(t_{i-1}))\right)^T (\Delta_i\Sigma)^{-1} + \mathbf{r}'(t_i)^T \mathcal{T}^{-1} \end{aligned}$$

Movement Variance

The conditional distribution of the movement variance becomes,

$$\sigma^2 \big| R_n, R'_n, \boldsymbol{\beta}, \mathcal{T} \sim \text{InvGamma}(a + n/2, b + 1/2 \sum_{i=1}^{n} \frac{\|\mathbf{r}(t_i) - \mathbf{r}(t_{i-1}) - h(\mathbf{r}(t_{i-1}))\Delta_i\|^2)}{\Delta_i}$$

Measurement Variance

The measurement variance is the same as in the equispaced time case,

$$\mathcal{T} \sim \text{InvWishart}\left(m + n, \Psi + \sum_{t=1}^{n} (\mathbf{r}'(t_i) - \mathbf{r}(t_i)(\mathbf{r}'(t_i) - \mathbf{r}_(t_i))^T\right)$$

Potential Parameters

The conditional posterior distribution of the potential parameters $\boldsymbol{\beta}$ is $\text{Normal}_k(Bb, B)$ where,

$$B^{-1} = W^{-1} + \frac{1}{\sigma^2} \sum_{t=1}^{n} \Delta_i \|g(\mathbf{r}(t_{i-1}))\|^2 \text{ and } b^T = \mu^T W^{-1} + \frac{1}{\sigma^2} \sum_{t=1}^{n} (\mathbf{r}(t_i) - \mathbf{r}(t_{i-1}))^T g(\mathbf{r}(t_{i-1}))$$

5.7 Extension to several animals

When multiple animals are tracked the question arises: do they have the same potential? A simple approach is to assume they have the same parametric form for their potential and test whether they have the same coefficients. The model described can be extended to multiple animals. Let $j = 1, \ldots, J$ index the different individuals, with each individual having observed locations, $\mathbf{r}_j$, at times $t_{ij}, i = 1, \ldots, n_j$. The likelihood for all of the animals assuming different potential function parameters, different movement variances and different measurement variances is

$$\prod_{j=1}^{J} \prod_{i=1}^{n_j} \phi\left(\mathbf{r}_j(t_{i,j}) | \mathbf{r}_j(t_{i-1,j}) + h(\mathbf{r}_j(t_{i-1,j})\Delta_{ij}, \boldsymbol{\beta_j}), \Sigma_j\right) \phi\left(\mathbf{r}'_j(t_{i,j}) | \mathbf{r}_j(t_{i,j}), \mathcal{T}_j\right)$$

where $\Delta_{i,j} = t_{i,j} - t_{i-1,j}$. The conditional distributions follow directly. With no common parameters the conditional distributions can be factored by individual and calculations are the same as for a single shark. When parameters are shared between sharks the conditional distributions for the shared parameters involve a sum over the individuals. The models can be compared using the Deviance Information Criterion for example.

5.7.1 Deviance Information Criterion

The deviance information criterion (DIC) (Spiegelhalter et al., 2002) provides a way to compare Bayesian models in similar way to AIC and BIC. Models are compared by their likelihood at the fitted parameters with some penalty for how many parameters were needed to fit the model.

Deviance is defined as

$$\text{Deviance} = D(\theta) = -2l(y\big|\theta) + C$$

where $l(y\big|\theta)$ is the log likelihood of the data given the parameter vector θ. C is a constant depending solely on the data and is irrelevant for model selection. The Deviance Information Criterion summarizes the fit of the model by the sum of the posterior expectation of the deviance and the effective number of parameters, p_D,

$$\text{DIC} = E_{\theta|y}D + p_D$$

Small values of DIC are to be preferred. Hence, a small expected deviance is to be preferred but not at the expense of a large number of estimated parameters. The effective number of parameters can be defined as,

$$p_D = E_{\theta|y}D - D(E_{\theta|y}(\theta))$$

which can be rewritten as

$$p_D = \bar{D} - D(\bar{\theta})$$

where $\bar{D}$ can be estimated by the average deviance over MCMC samples and $D(\bar{\theta})$ can be estimated by the deviance at the average posterior parameter values over MCMC samples.

Chapter 6

Shark 66003

Shark 66003 represents the best case scenario for an archival tag: the tag reached its release date and all the data were recovered from the tag. It will be used to demonstrate methods for estimating a potential. The potential functions explored will only depend on position. Interactions with the environment and other sharks will be explored in later chapters.

The track of the shark is shown in Figure 6.1 with the colour of the track representing the progression of time. We can see this shark stays near the coast. To get a clearer picture of the progression in time, easting and northing are plotted separately in Figure 6.2. Shark 66003 appears to navigate its whole range with no seasonal preference: it isn't favouring the top or bottom of its range in a particular season.

Speeds, estimated by simple differences at the observed locations (Figure 6.3), show some evidence of an attractor near (650E, -700N). Speeds in the north-south direction are generally positive (blue) in the south and negative (green) in the north, speeds in the east-west direction are generally positive in the west (or near the coast) and negative in the east. A smooth nonparametric estimate for the potential can help clear up the picture.

6.1 Smooth potential estimates

Using the methods outlined in Section 4.2 a potential function was fit using `mgcv::gam` with the amount of smoothing chosen by generalized cross validation. The fit is shown in Figure 6.4. There is a region of attraction estimated near (650E, -825N).

Instead of choosing the amount of smoothing by generalized cross validation it can be set directly. Figure 6.5 shows the fitted surface when the smoothing parameter is set to 0.01, 0.1, 5 and 10 times that found using generalized cross validation.

Although the smooth was fit under the assumption that the movement errors are independent gaussian coordinates with the same variance in both directions, σ be estimated from the residuals separately for each direction. In this case the standard deviation, σ, is estimated to be 38.3km in the easting direction and 94.3km in the northing direction. This is a combination of both measurement and movement model errors. To allow for different

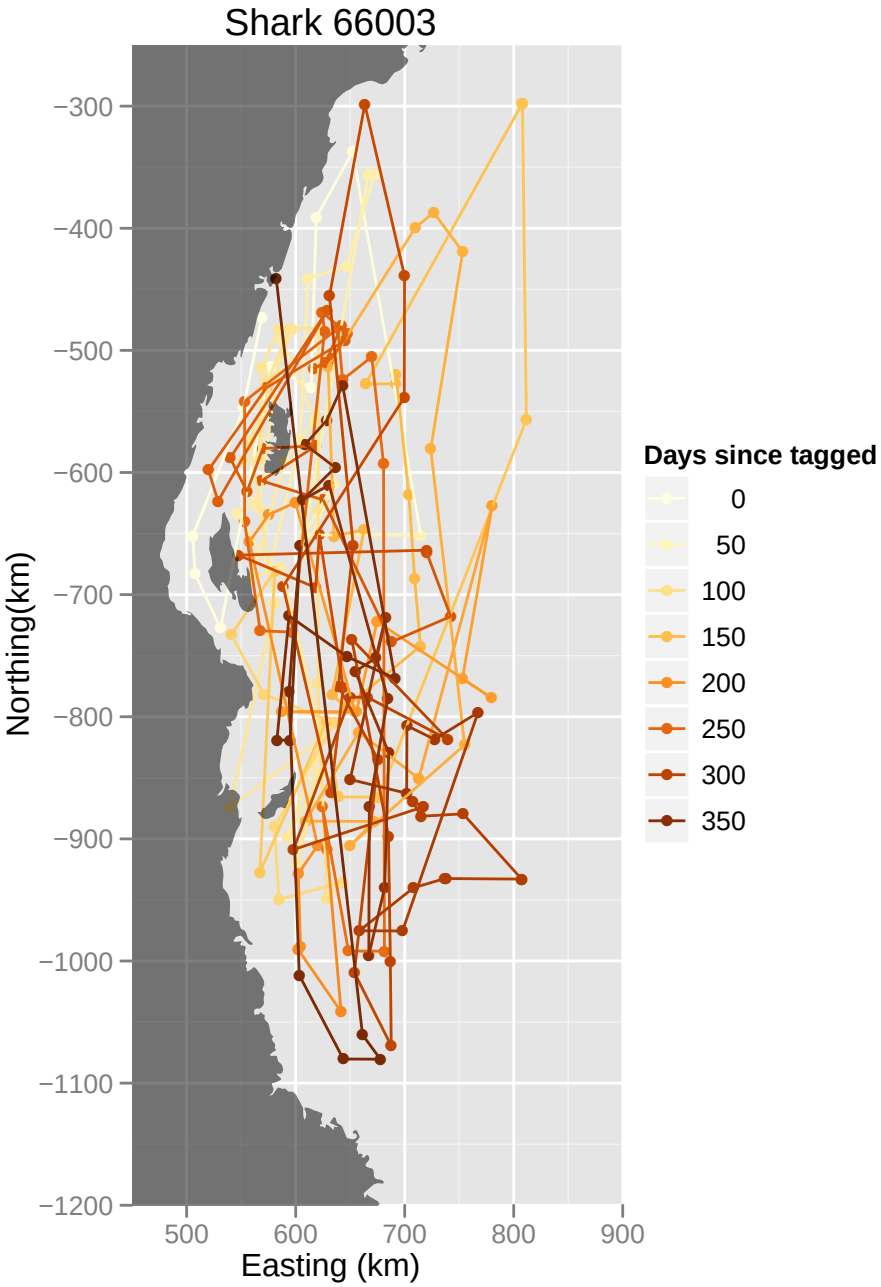


Figure 6.1: Recorded track for shark 66003. The colour of the line indicates the progression of time since the initial position. The sharks final position is seen to be quite close to its initial position near (600E, -450N).

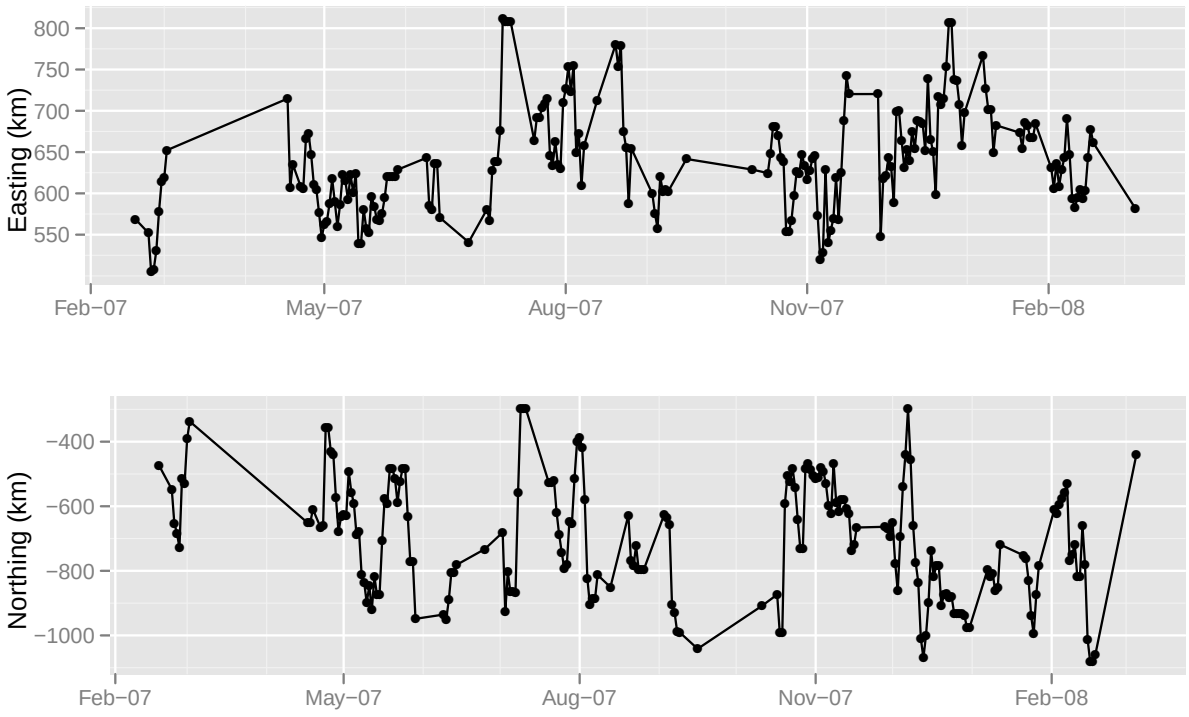


Figure 6.2: Reported easting and northing against time for shark 66003.

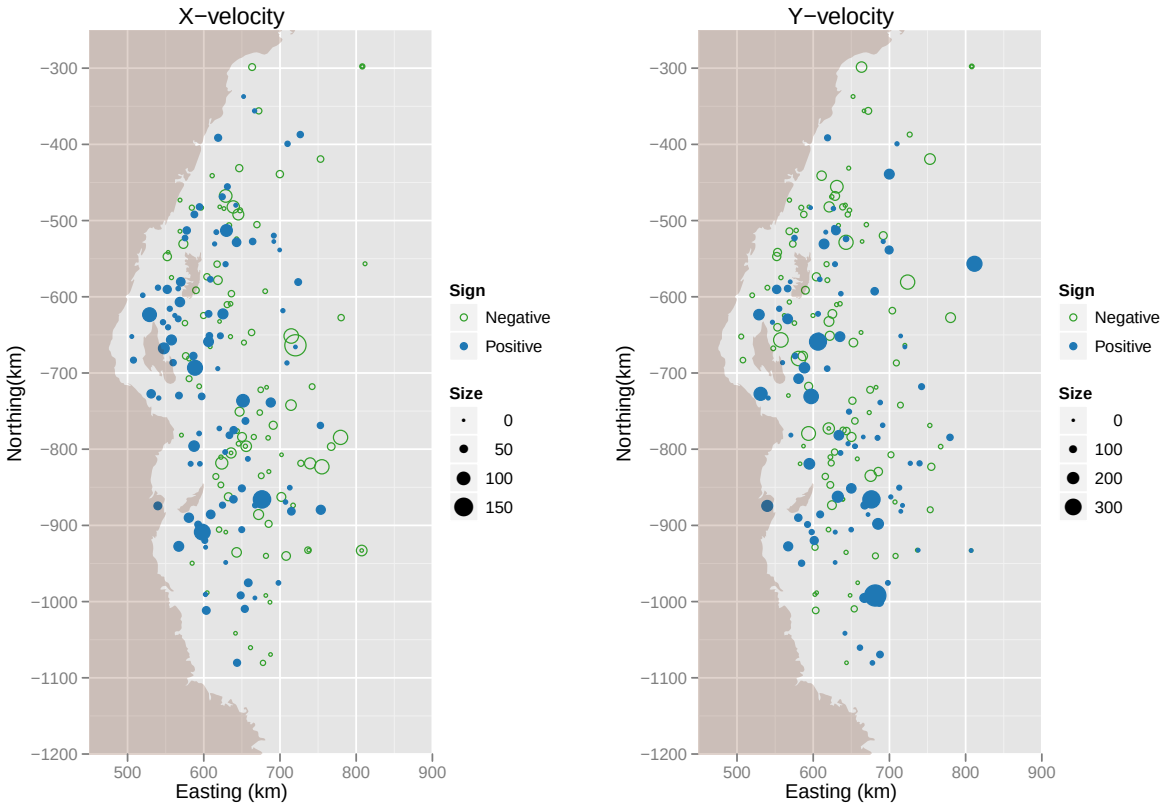


Figure 6.3: Speeds at the observed locations as approximated by differencing. There is weak evidence for an attractor near (650E, -700N).

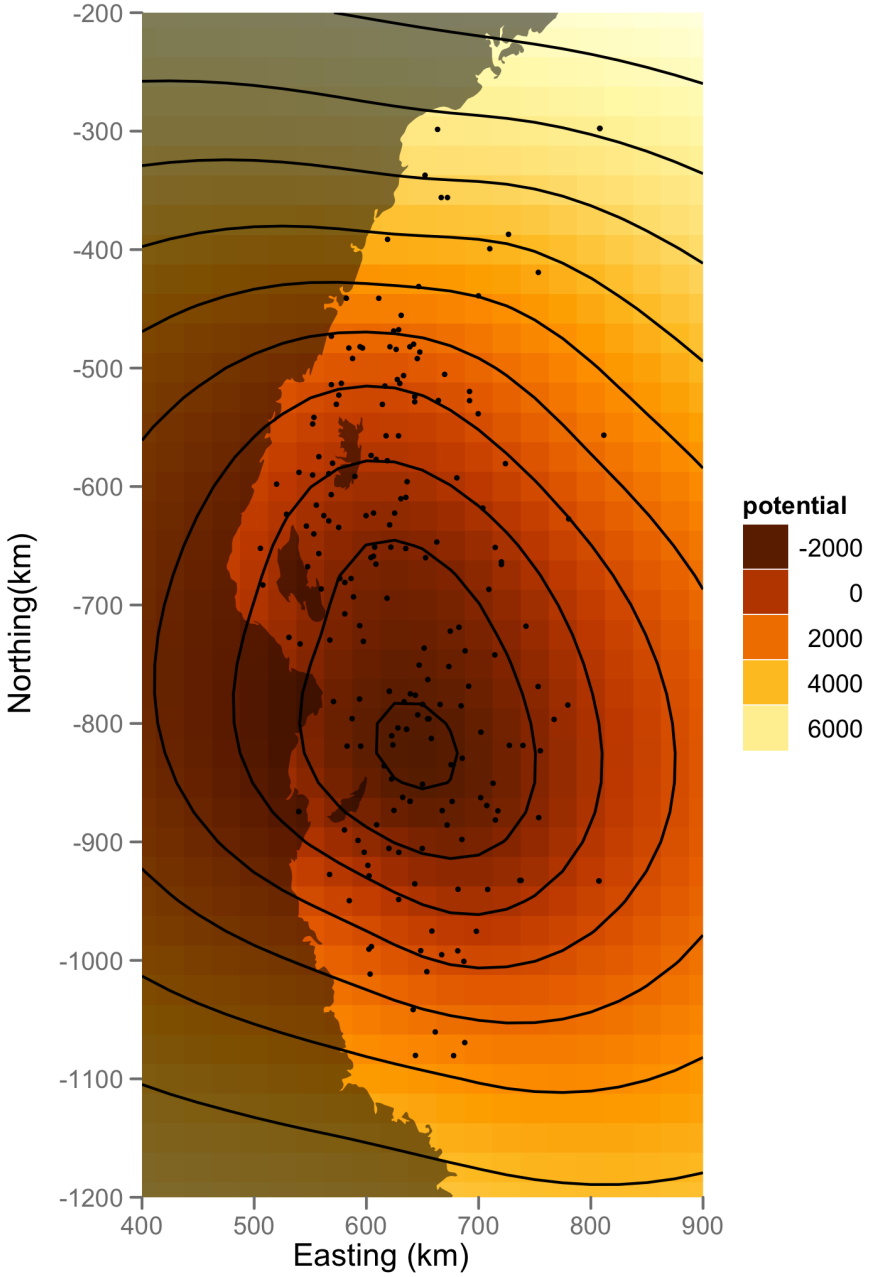


Figure 6.4: Fitted potential function using penalized thin plate regression splines. For convenience the contours are allowed to run over the land. Black dots are actual observed locations for shark 66003.

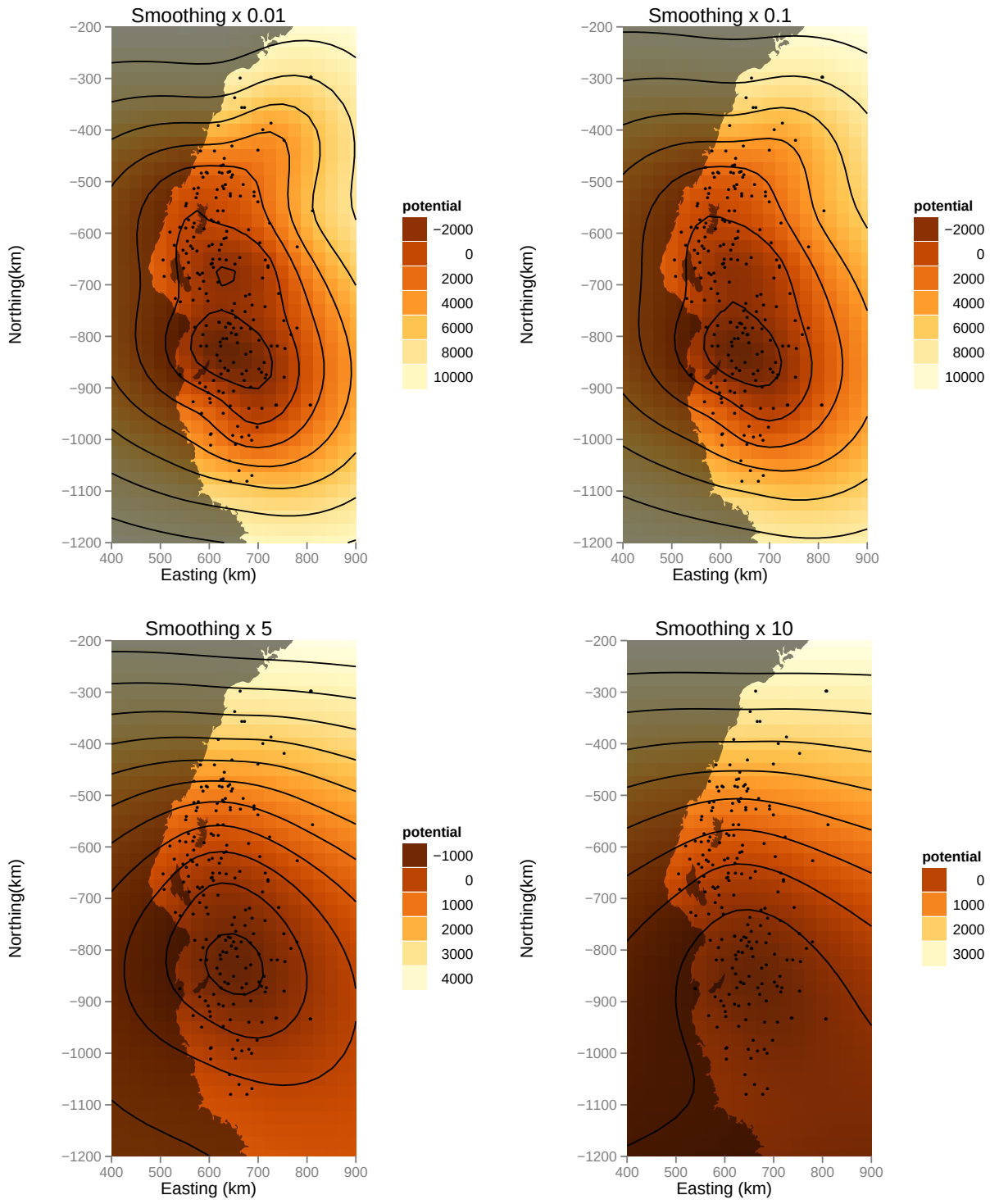


Figure 6.5: Fitted potential function using penalized thin plate regression splines varying the smoothing parameter as a multiple of that found using generalized cross validation.

Table 6.1: Estimate and 95% CI.

Parameter	Lower	Estimate	Upper
b_1	-137.8	-91.5	-45.3
b_2	8.6	41.3	73.9
b_3	0.04	0.07	0.11
b_4	0.01	0.03	0.05
b_5	-0.01	0.00	0.02
S_y	2.1	2.4	2.8
σ	35.1	38.8	42.7

error variances in the northing and easting directions, a simpler potential model that allows for a more complicated error structure was fit.

6.2 A quadratic potential estimate

A simple parametric model suggested by the smooth estimate is a quadratic in easting and northing. That is,

$$H(x, y) = \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy$$

One can allow for different variances in the easting and northing directions, that is the errors have mean zero with covariance matrix,

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & S_y^2 \sigma^2 \end{pmatrix}$$

This model was fit using generalized least squares (using `gls` in the R package `nlme` (Pinheiro et al., 2011)). The coefficients for this fit, along with standard errors (under the assumptions for the Lai-Wei theorem, see Section 3.5) are shown in Table 6.1.

The resulting potential is plotted in Figure 6.6. An attractor was estimated at 653E -688N. The standard deviation of the errors in the northing direction were estimated to be 2.4 times that in the easting direction. This is consistent with the tag technology if most of the error is coming from measurement error.

A linear, quadratic and cubic potential were fit and tested for significance using a likelihood ratio test. The quadratic was shown to be significantly better than a linear potential, and a cubic potential not significantly better than the quadratic (at the 5% level under the assumptions of Section 3.5).

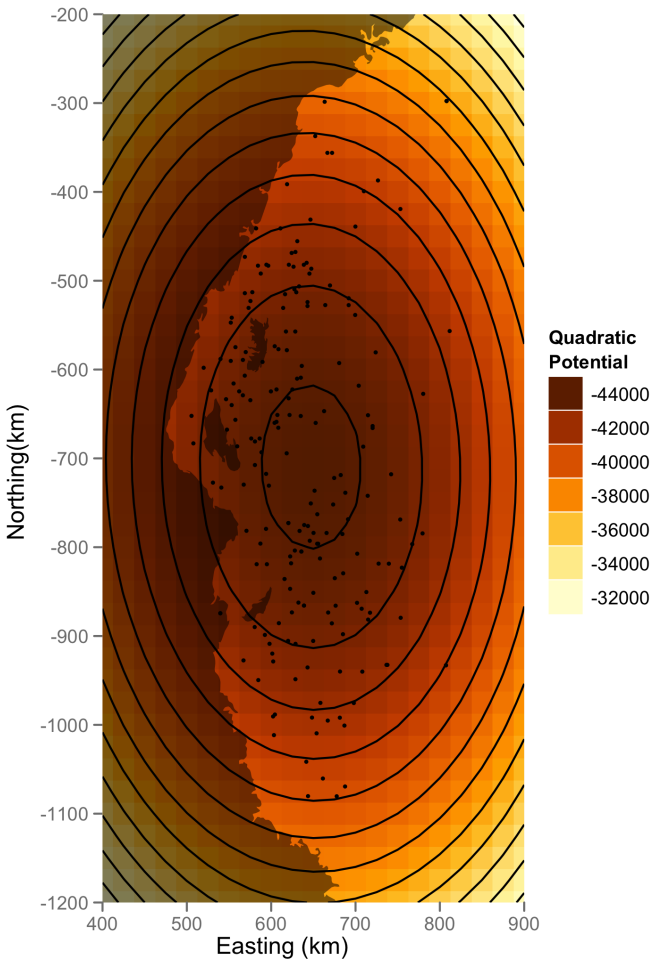


Figure 6.6: Estimated quadratic potential for shark 66003.

6.3 Checking the adequacy of the quadratic model

The model fit can be examined by looking at the residuals. Here the standardized velocity residuals are used. Figure 6.7 examines the suitability of the quadratic form of the potential. Comparing this to Figure 6.3, there are no obvious spatial patterns left in the residuals suggesting that the quadratic surface was adequate.

The potential was assumed to not change through time. This assumption is examined in Figure 6.8 by plotting the velocity residuals against time. It shows no obvious trends and the residuals are consistent with no autocorrelation. This was confirmed by examining plots of the estimated autocorrelation function of the residuals (not shown), only the complete cases at each lag were used.

In setting down the the model, the errors were assumed to be normally distributed. This can be examined marginally, by looking at quantile-quantile plots of the residuals. These

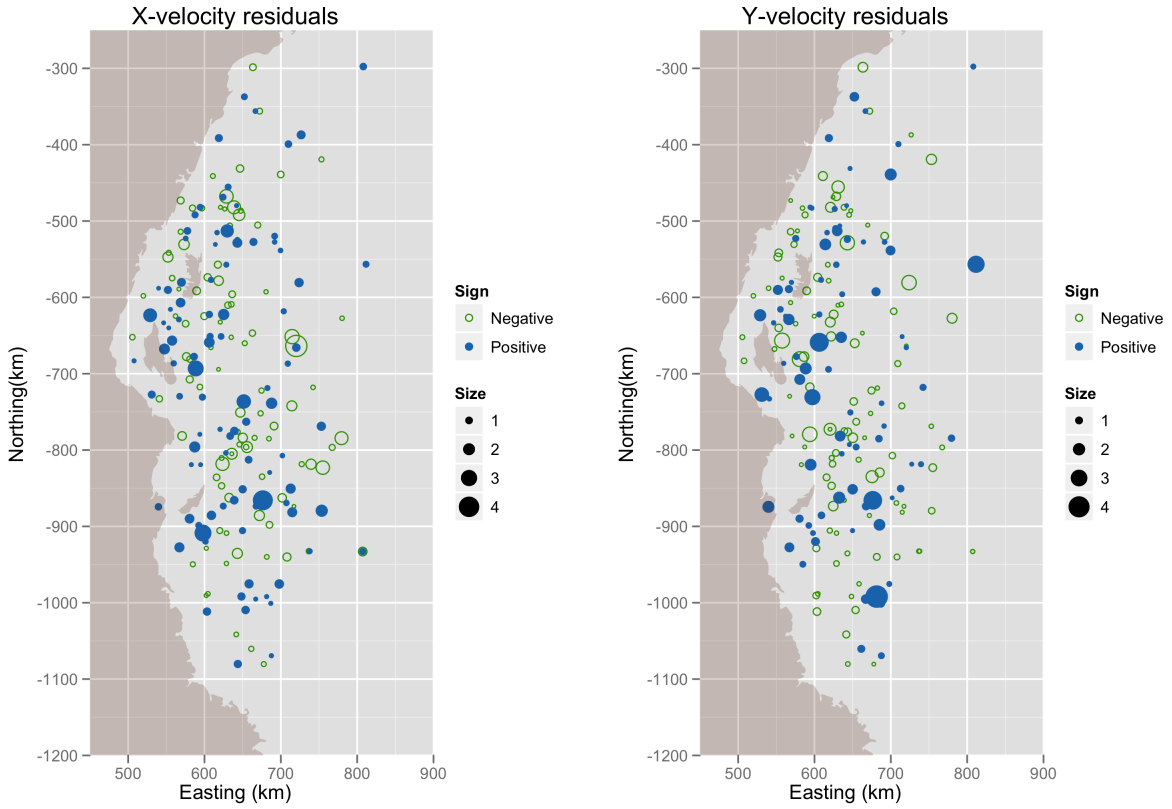


Figure 6.7: Standardised velocity residuals in the easting (X) and northing (Y) directions from the quadratic potential fit.

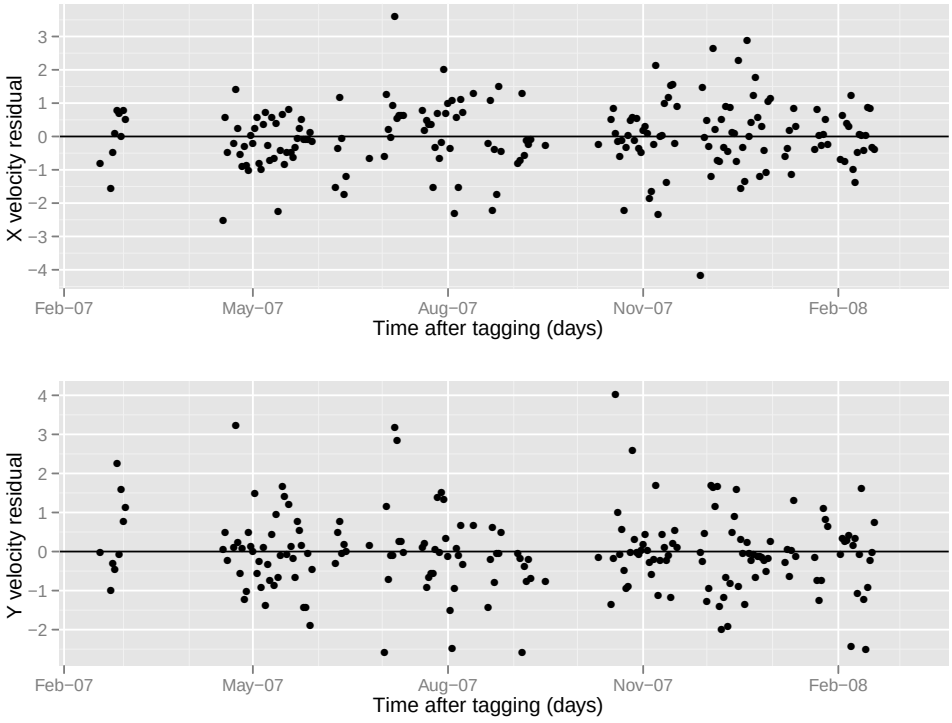


Figure 6.8: Standardised velocity residuals in the easting and northing directions from the quadratic potential fit against time.

are shown in Figure 6.9. The residuals appear approximately symmetric but perhaps have slightly longer tails than would be expected with a normal distribution. This places some doubt on the appropriateness of the normal based confidence intervals reported. To address this bootstrapped confidence intervals were calculated.

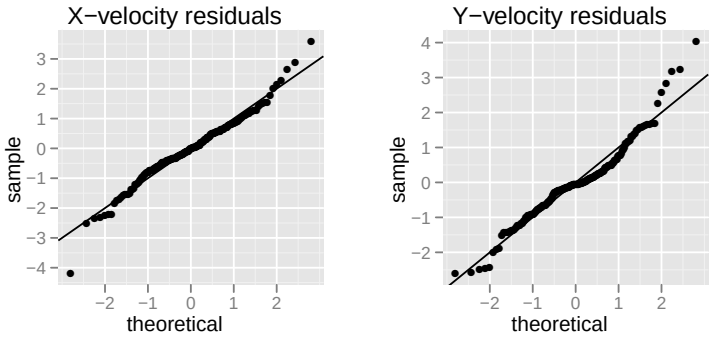


Figure 6.9: Quantile-quantile plots of residuals

6.4 Bootstrapping the quadratic estimates

A bootstrap track was produced by taking the starting point from the track of shark 66003 and generating daily time steps ($\Delta_i = 1, \forall i$) using the potential estimated by the quadratic fit. The residuals (scaled appropriately) from the original fit were resampled to produce the errors. The track was then reconstructed at each successive time point and then subsampled to the days a location for shark 66003 was actually observed. One thousand such tracks were produced and the quadratic model fitted to them. Examples of the bootstrapped tracks can be seen in Figure 6.10. The tracks look reasonable except for the fact they go on land. This will be addressed later. Bootstrapped confidence intervals were produced by taking the empirical 2.5% and 97.5% quantiles of the fitted coefficients and are displayed in Table 6.2. The bootstrapped estimates suggest there may be some bias in our estimates. All of the bootstrap samples resulted in quadratic functions with unique attraction points. These points are plotted in Figure 6.11. The figure gives an indication of the uncertainty in the estimated point of attraction.

6.4.1 What about the coast?

The quadratic model, as set down, places no boundaries on the shark's range. This may be appropriate for open ocean but since this shark is near the coast this allows the shark to move onto land. The simulated tracks can be kept off the land by adding an analytic term. If the current position is on the land or within 50km of the coast, a term to repel the particle

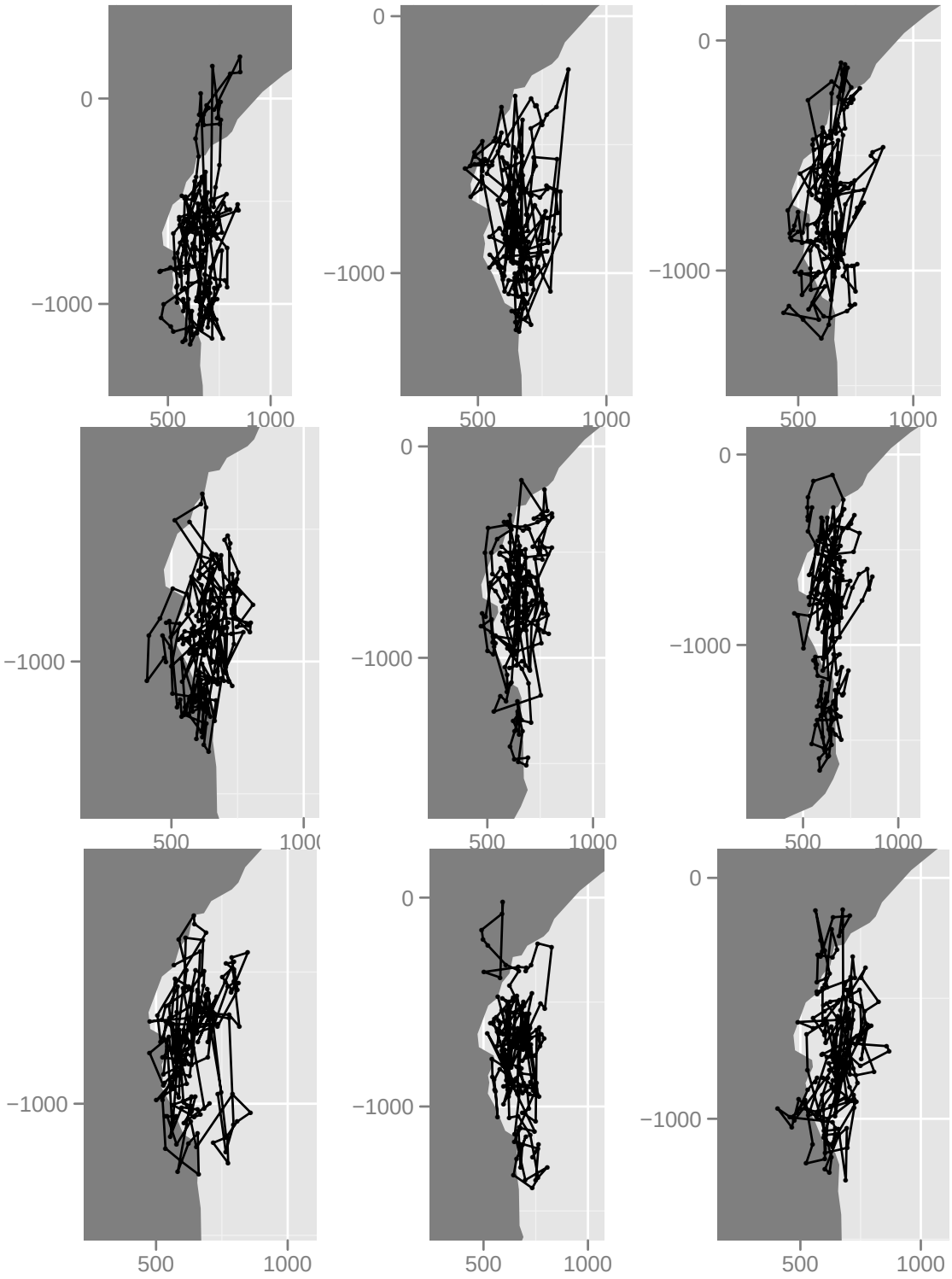


Figure 6.10: Examples of bootstrapped tracks using the estimated quadratic potential and resampled residuals.

Table 6.2: Estimates based on bootstrapped trajectories with 95% confidence intervals. Truth indicates the potential parameters used to create the bootstrapped realizations.

Parameter	Truth	Estimate	95% CI
β_1	-91.5	-75.7	$(-124.2, -40.4)$
β_2	41.3	40.7	(21.4, 76.7)
β_3	0.07	0.06	(0.04, 0.09)
β_4	0.03	0.03	(0.02, 0.05)
β_5	0.00	0.002	$(-0.015, 0.019)$

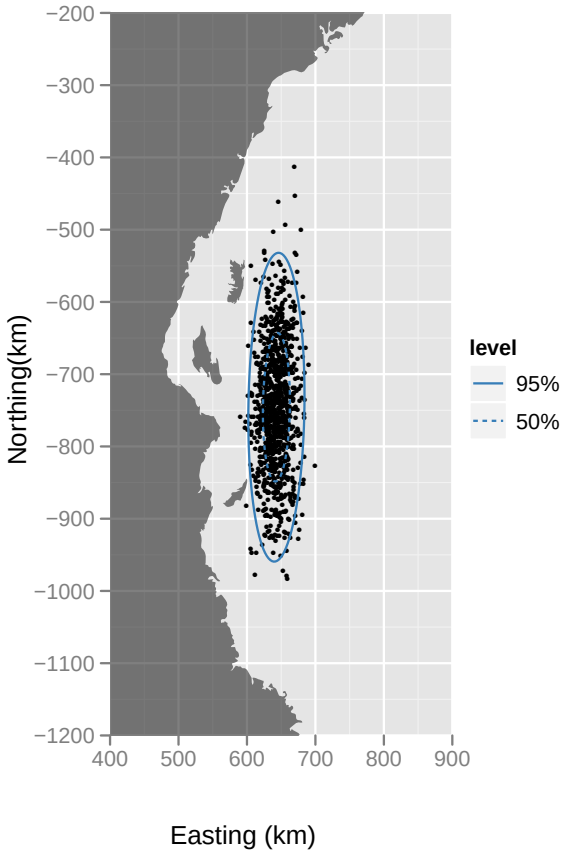


Figure 6.11: Estimated points of attraction under the quadratic model of 6.2 for the bootstrap samples. All samples resulted in a quadratic with a unique point of attraction. The ellipses show 50% and 95% quantiles based on a bivariate normal assumption.

off the coast is added,

$$H(x,y) = \text{quadratic}(x,y) + \left((50 - d_C(x,y))_+\right)^2$$

where $d_C(x,y)$ is the distance from (x,y) to the closest point on the coast. This keeps the trajectory off the land (ignoring the effects of the quadratic and random noise). The resulting potential function is shown in Figure 6.12. If a simulated location ended up on

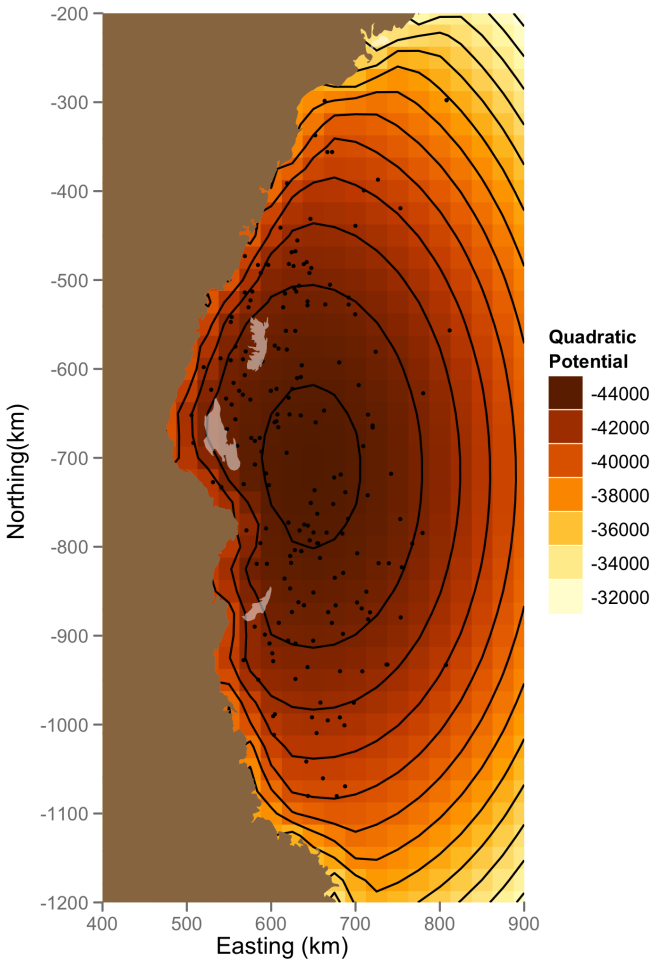


Figure 6.12: The potential for the bootstrapped tracks combining a quadratic potential with repulsion from coast.

land by chance, the errors were resampled. This affects the distribution of the errors but I leave dealing with this until later work. Some examples of bootstrapped tracks generated using this potential are shown in Figure 6.13. Confidence intervals based on these tracks are shown in Table 6.3

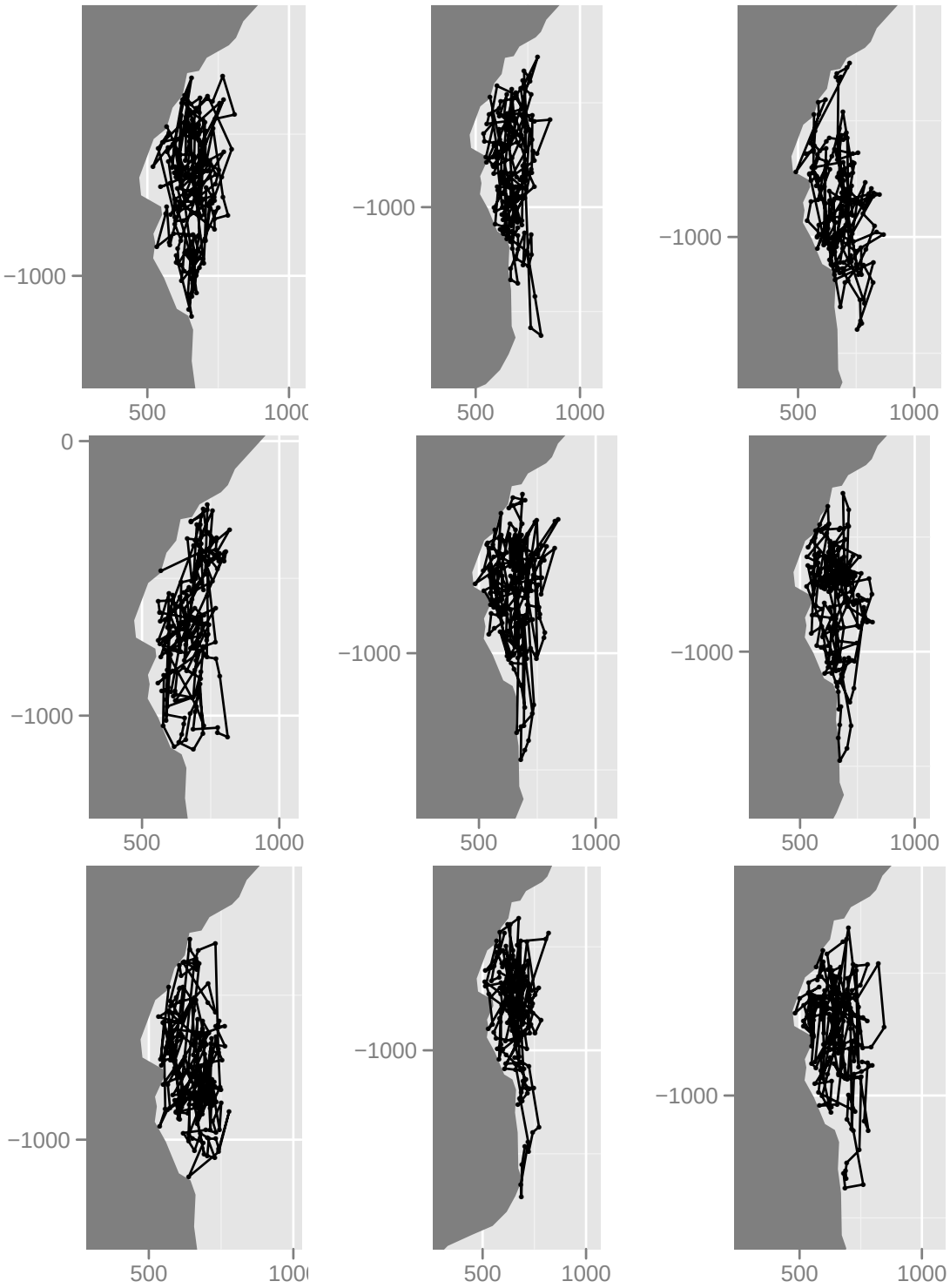


Figure 6.13: Examples of bootstrap tracks using a quadratic potential along with a term to repel the sharks off the coast.

Table 6.3: Bootstrapped estimate and 95% confidence interval for trajectories kept off the land.

Parameter	Truth	Estimate	95% CI
β_1	-91.5	-79.0	$(-127.8, -46.3)$
β_2	41.3	60.2	(33.1, 98.4)
β_3	0.07	0.06	(0.04, 0.10)
β_4	0.03	0.04	(0.02, 0.07)
β_5	0.00	0.006	$(-0.012, 0.025)$

6.4.2 Boundaries

There are ways to incorporate boundaries in fitting smooth functions. Soap film smoothing (Wood et al., 2008) is an extension of the gam methodology that only calculates the penalty within the boundary of interest. This was used to estimate a non parametric potential function for shark 66003 calculating the penalty only over the ocean. The resulting potential function is shown in Figure 6.14. Comparing this to Figure 6.4 a difference appears to be that the contours follow the coast. In particular, near the outcropping of land at (600E, -800N), following a contour will avoid the outcropping. In the non-soap-film smooth, moving along a contour would move through the land. The soap film smooth provides a more realistic model from the shark's perspective. Adding structure to the error variance in the soap film smooth is an area for further research.

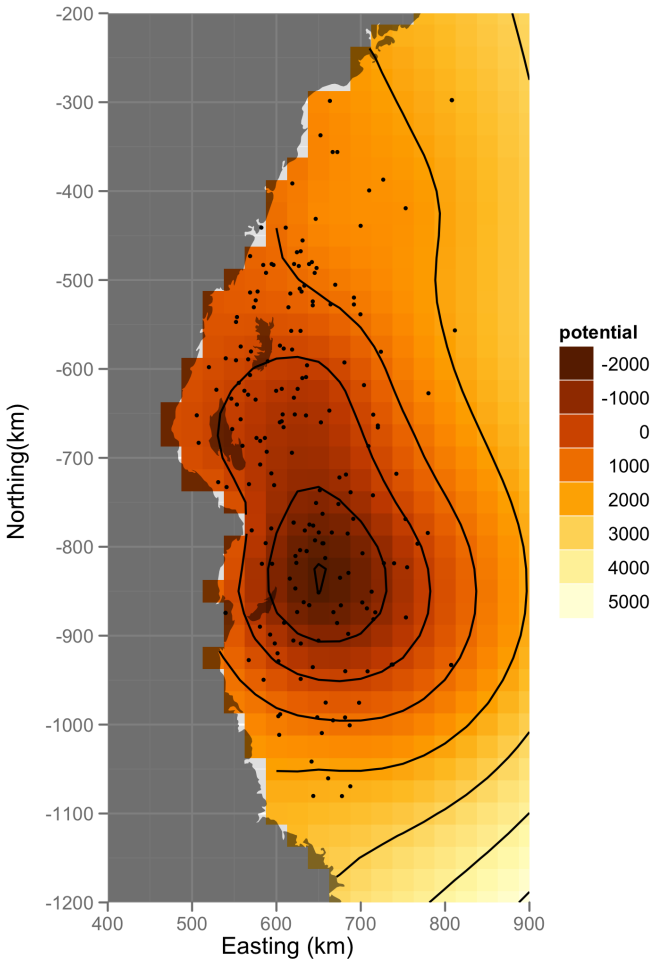


Figure 6.14: Fitted potential using soap film smoothing: a penalty on wiggliness only over the area of interest (not mainland Africa). The dots represent observed locations.

6.5 Measurement error

The preceding models for shark 66003 combined the measurement and movement error together. In this section the Gibbs sampling methodology developed in Section 5.4 to model measurement and movement error explicitly is applied to the track of shark 66003.

The potential was assumed to be a quadratic, so that the movement model is

$$\begin{pmatrix} x(t_{i+1}) \\ y(t_{i+1}) \end{pmatrix} = \begin{pmatrix} x(t_i) \\ y(t_i) \end{pmatrix} - (t_{i+1} - t_i) \begin{pmatrix} \beta_1 + \beta_3 x(t_i) + \beta_5 y(t_i) \\ \beta_2 + \beta_4 y(t_i) + \beta_5 x(t_i) \end{pmatrix} + \boldsymbol{\sigma} \sqrt{t_{i+1} - t_i} \mathbf{Z}_{i+1}$$

where the $\mathbf{Z}_i$ are bivariate standard normals. The errors in location due to the PSAT tagging methods are modelled by the measurement equation,

$$\mathbf{r}'(t_{i+1}) \quad = \quad \mathbf{r}(t_{i+1}) + \boldsymbol{\epsilon}_{i+1}$$

where the $\boldsymbol{\epsilon}_i$ are independent

$$\text{Normal}_2 \left(\mathbf{0}, \mathcal{T} = \begin{pmatrix} \tau_x^2 & \rho\tau_x\tau_y \\ \rho\tau_x\tau_y & \tau_y^2 \end{pmatrix} \right)$$

The approach outlined in Section 5.4 was used to estimate the parameters. Vague Normal priors were used for the parameters of the potential function, $\boldsymbol{\beta} \sim Normal(0, 100^2)$. Two different sets of priors were used for the error variance parameters.

The information from prior studies on the measurement error was incorporated into priors on the measurement error parameters. From previous studies the error is expected to be between 1–3° in latitude and 0.2–1.0° in longitude. Since the tracks are in UTM coordinates, these ranges were converted from degrees to kilometres. 1–3° in latitude is about 111–333km. The area the shark moved in was roughly centered around 7°S 40°E. So, 0.2–1.0° in longitude is about 22–110 km. Based on these ranges, two Inverse Wishart priors on the measurement error covariance matrix were chosen with the following parameters,

$$\mathcal{T} \sim \text{Inverse Wishart} \left(m_1 = 10, \Psi_1 = 7 \times \begin{pmatrix} 4900 & 0 \\ 0 & 49000 \end{pmatrix} \right) \qquad \text{Prior \#1}$$

$$\mathcal{T} \sim \text{Inverse Wishart} \left(m_2 = 5, \Psi_2 = 2 \times \begin{pmatrix} 5200 & 0 \\ 0 & 60000 \end{pmatrix} \right) \qquad \text{Prior \#2}$$

The Ψ are scaled by $m - 3$ because the mean of an Inverse Wishart is $\Psi/(m-3)$. The 2.5% and 97.5% quantiles for draws from these distributions are roughly (based on 10000 draws) (43, 114) for τ_x and (135, 352) for τ_y from the first prior (30, 146) and (103, 507) respectively for the second prior.

The Gibbs sampler applied to shark 66003 was very unstable unless a somewhat infor-

mative prior for σ^2 was used. The following two alternatives were used,

$$\sigma^2 \sim \text{Inverse Gamma}(0.25, 100) \quad \text{Prior \#1}$$
$$\sigma^2 \sim \text{Inverse Gamma}(0.2, 80) \quad \text{Prior \#2}$$

The 2.5% and 97.5% quantiles for σ from 10,000 draws from these distributions are (2.8, 135) and (2.8, 348) for prior #1 and #2 respectively. Prior #1 places more weight on the prior distributions than prior #2.

For each prior three chains were run from random starting points, with a 1000 iteration burn in period. Ten thousand more iterations were run per chain and every tenth retained. The posterior estimates of the parameters for each prior are shown in Figure 6.15 along with the estimates from the methods used earlier in the chapter. The first row of the figure shows the intervals for the parameters of the quadratic potential. All of the methods agree on the sign of the terms: β_1 is negative, β_2 is positive, β_3 and β_4 are small and positive and β_5 is not inconsistent with zero. These parameter values are consistent with a point of attraction. The two state space models shrink the coefficients of the potential function towards zero compared to the intervals produced without an explicit measurement error term.

The second row of Figure 6.15 displays the estimates of the variance parameters. The σ are not directly comparable as the first three methods include measurement error in its definition, however we see they are all roughly the same value, around 37km. The state space models give slightly different estimates of σ and $\mathcal{T}$. Prior # 1 placed more weight on the prior information and resulted in larger estimates of the measurement error but smaller estimates of the movement error, both estimated the measurement error in the north-south direction to be about four times as large as in the east-west direction.

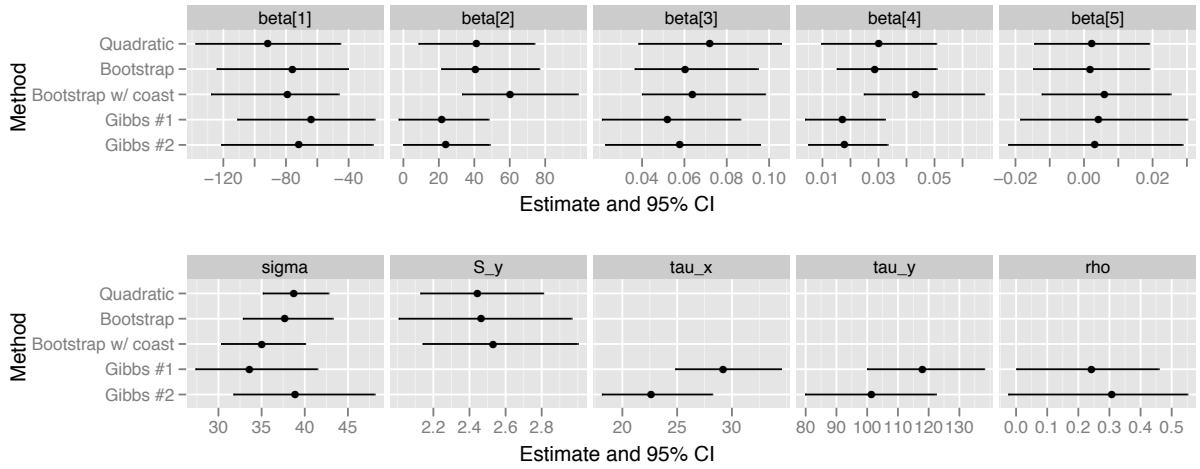


Figure 6.15: Estimates of the parameters of the movement model for the methods applied to shark 66003.

6.5.1 Residuals in state space model

The measurement residuals,

$$\mathbf{d}(t_i) = \mathbf{r}'(t_i) - \mathbf{r}(\hat{t}_i).$$

for the state space model estimated with prior #1 are shown in 6.16. These correspond to the observed locations minus the estimated true locations. The residuals in the x and y direction are plotted separately. There are strong spatial patterns. The measurement residuals in the east-west direction are negative in the west and positive in the east, the north-south residuals are positive in the north and negative in the south. The true positions have been estimated closer to the centre of the shark's range. This is seen more directly in

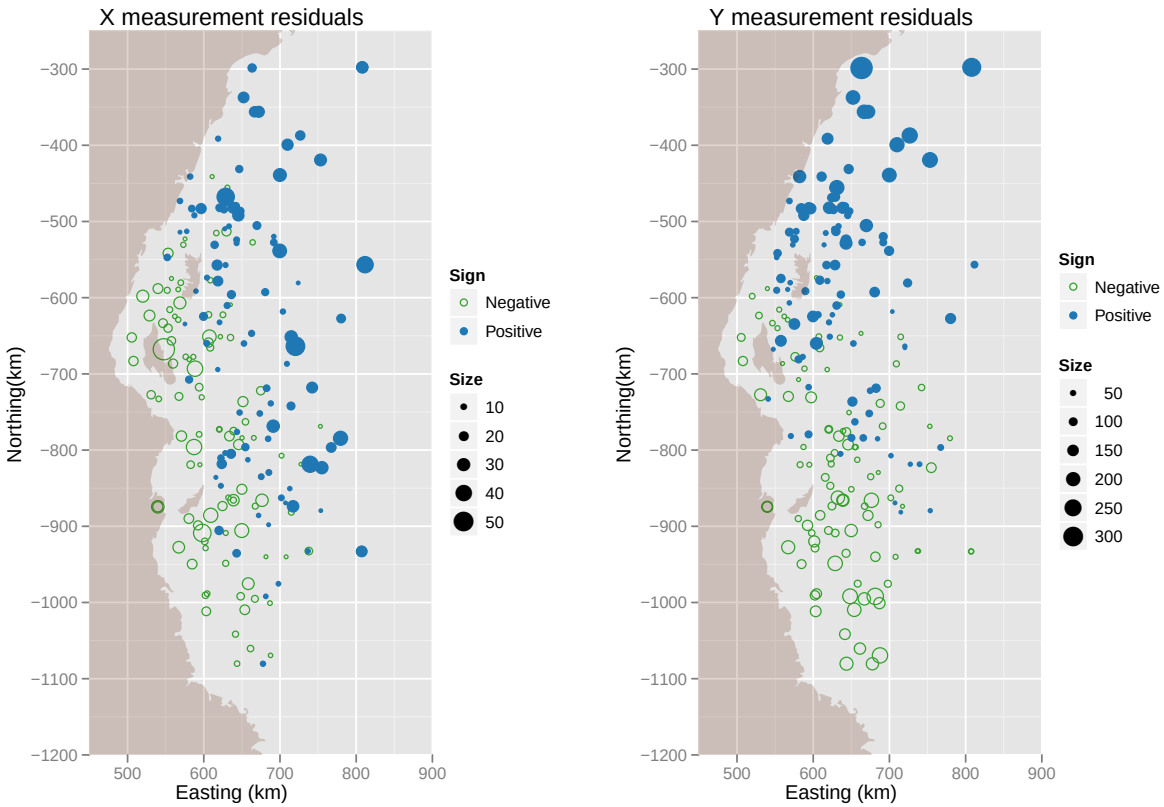


Figure 6.16: Measurement residuals (observed position - estimated position) for the estimates using the state space model and prior #1.

Figure 6.17 where the estimated true positions of shark 66003 are plotted along with the observed locations against time. The estimated locations have been smoothed compared to the observed ones, the locations in the north-south receiving much more smoothing as a result of τ_y being estimated as much larger than τ_x. Although not shown here the residuals and locations for the state space model estimated with prior #2 show the same patterns. The true measurement errors should not show such a strong spatial pattern. The pattern

is most likely a result of the strong effects of a quadratic attractor at distances far from the point of attraction. A potential with an attractor but flatter at larger distances than a quadratic may be more appropriate and is an avenue for future work.

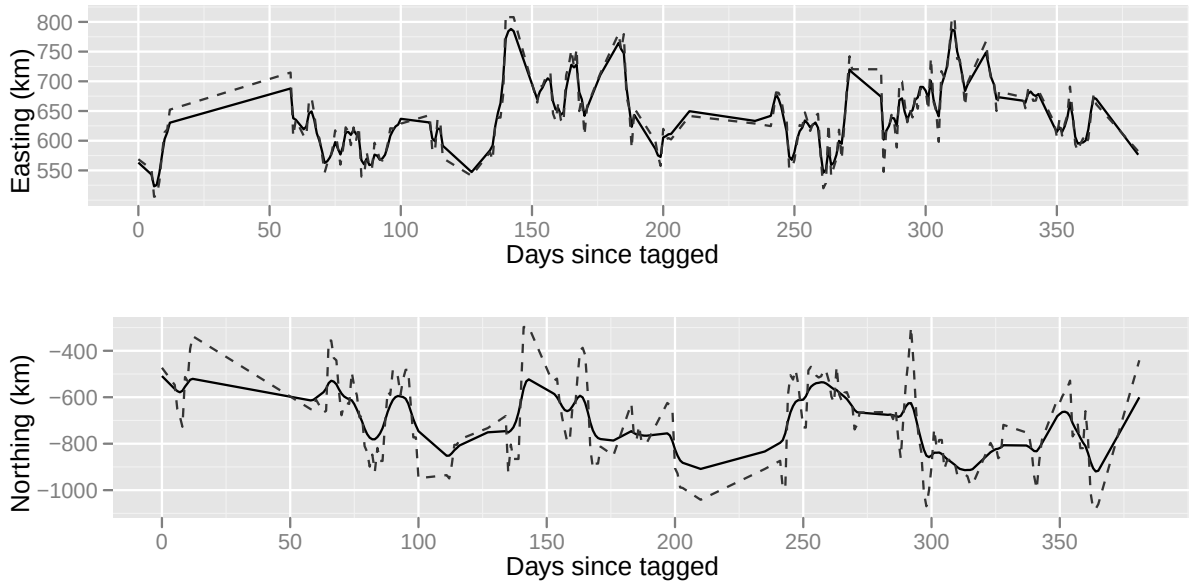


Figure 6.17: Estimated true positions (solid black line) along with observed positions (dashed grey line).

6.6 Discussion

One concludes the observed track of Shark 66003 is consistent with an attractor around 650km easting and -700km northing (UTM zone 37). For a year this shark stays in the same area. The following chapter will investigate relationships with the environmental covariates. What about the other sharks? Chapter 8 explores the question of whether the sharks share a common quadratic potential. It's clear that there is greater variance in the north-south direction than east-west which is consistent with the method of geolocation used.

The methods applied in this chapter have varied strengths. The generalized additive model allows the data to speak with minimal assumptions on the functional form of the potential, however details in implementation place more restriction on the form of the errors. The quadratic least squares fit is the easiest to implement and quickest to run, and in this situation models the data quite well. The state space quadratic model explicitly models the measurement and movement variances separately but it isn't clear that it is an improvement in the present situation as the resulting measurement residuals appear to have more spatial structure than would be expected.

Chapter 7

Exploring covariates

The previous chapter investigated modelling the movement of shark 66003 as a function of only geographical position. Here, I investigate the influence of environmental covariates on his or her movement. Two approaches are taken in this chapter. First, the levels of the covariates the shark spends time in are explored. Second, the influence of the covariates on the movement of the shark is tested by including them in the potential function movement model. The potential functions used now have time varying components.

I will start with a description and exploration of the covariates available for the area: bathymetry, chlorophyl concentration, sea surface temperature, sea surface current and surface winds.

7.1 The covariates

The covariates used in this analysis are available from remote sensing satellites and are all obtained from `http://coastwatch.pfeg.noaa.gov/erddap/griddap/index.html`. There are some general caveats to make. The covariates available are only proxies for the things of interest. In an ideal world, complete three dimensional data would be available on the environment in the ocean: temperature, productivity and currents. The only data available are estimates of surface values and whale sharks are not expected to spend most of their time at the surface. At best one has an interpolated model produced value of the covariate at the shark's observed location. However, the covariates may still prove useful as explanatory variables.

7.1.1 Bathymetry

Bathymetry is the mapping of the depth of the ocean floor. The National Geophysical Data Center at NOAA provides 2-minute gridded global topography data (ETOPO2v2 (2006)). These are plotted for the West Indian ocean in Figure 7.1. The convex hull for shark 66003 and all of the tracked sharks (excluding 80180) are shown as black polygons.

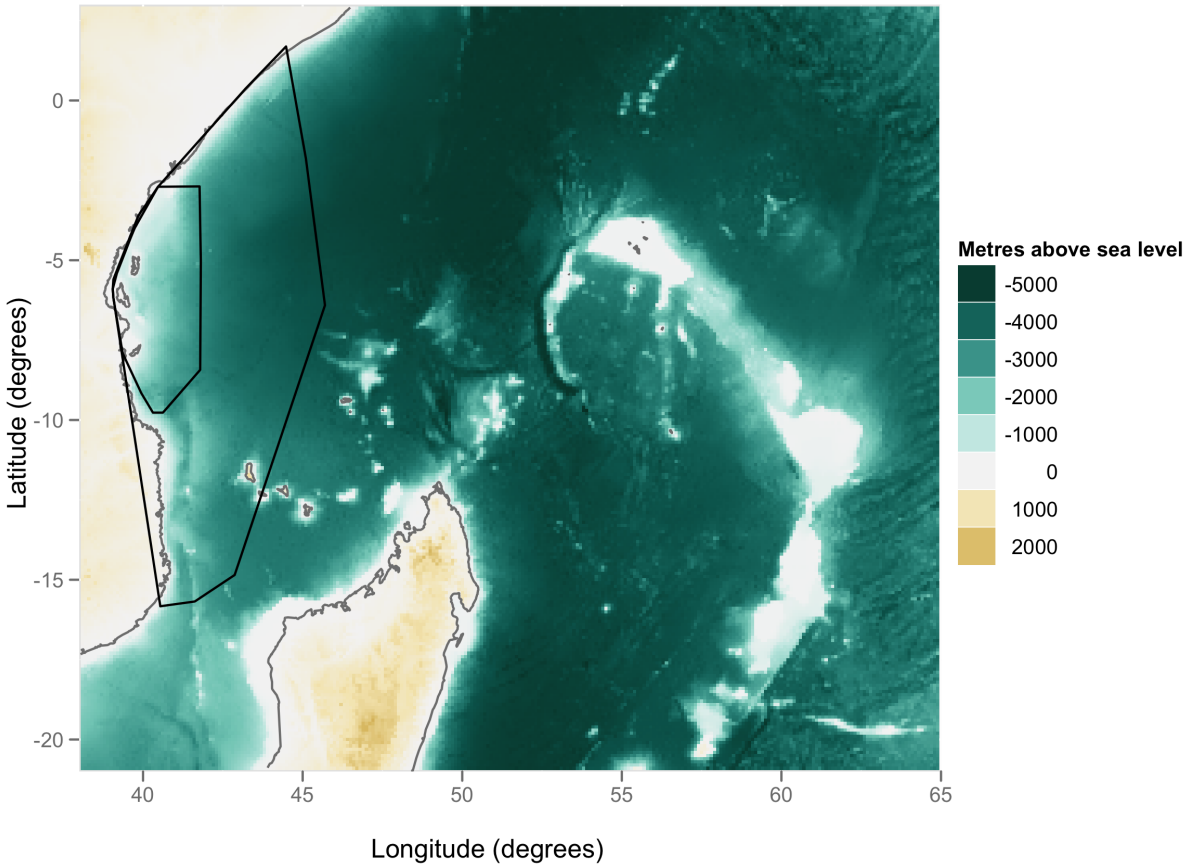


Figure 7.1: Ocean depth obtained from ETOPO2v2 (2006). The smaller black polygon is the convex hull of the positions of shark 66003, that large black polygon is the convex hull for all of the sharks excluding shark 80180.

7.1.2 Chlorophyll concentration

Chlorophyll concentration is of interest as it can be a proxy for near surface ocean productivity. It measures the amount of phytoplankton in the water. Phytoplankton are the primary plant life in the ocean and the direct or indirect food source for most marine animals.

One might expect the whale sharks prey to be more plentiful in areas of high chlorophyll concentration. The NOAA CoastWatch Program provides chlorophyll-a concentration data captured by the NASA's Aqua Spacecraft MODIS Chlorophyll. Measurements are made using the Moderate Resolution Imaging Spectroradiometer (MODIS) device and processed according to the algorithms in O'Reilly et al. (2000) to estimate the chlorophyll concentration. I use the 8 day product at a resolution of 0.1 degrees in latitude and longitude. The 8 day product refers to the fact that 8 days of data are used in the estimation. The chlorophyll is measured in milligrams per metre cubed. It has a very skewed distribution so I will generally use it after a log transform.

An example of these data for February 22nd 2007 is shown in Figure 7.2. There are a large amount of missing data, primarily due to cloud cover. Areas with high chlorophyll can be found by averaging over the two years of tracking, as shown in Figure 7.3. Coastal areas of the African continent and Madagascar have on average higher levels than the open ocean. In general the depth of the water seems to be an indicator of the average chlorophyll. To help elucidate the different annual patterns, the time series for chlorophyll at each location were clustered using hierarchical clustering with Ward linkage (Ward, 1963) on the centered time series. The distance between two locations is calculated using the Euclidean distance. This is complicated by missing values. The times one or both series have missing values are excluded from the sum and the sum is scaled by the number of times they are both observed. Clustering was not used because I believe there are distinct clusters but as a way to help understand the geographical variation in chlorophyll levels throughout the year. The results of cutting with 4 clusters are shown in Figures 7.4 and 7.5. There is no restriction in the clustering algorithm for locations nearby in geographical space to be in the same cluster, but they were generally assigned to the same clusters. The time series of the clusters (Figure 7.5) show slightly higher chlorophyll levels in May through November. The local variation is high. For example, clusters 2 and 3 exhibit a quicker rise in levels in July 2007. Shark 66003 primarily occupies cluster 1 during the time it was tracked. This cluster exhibits the most sinusoidal like variation with a period of one year.

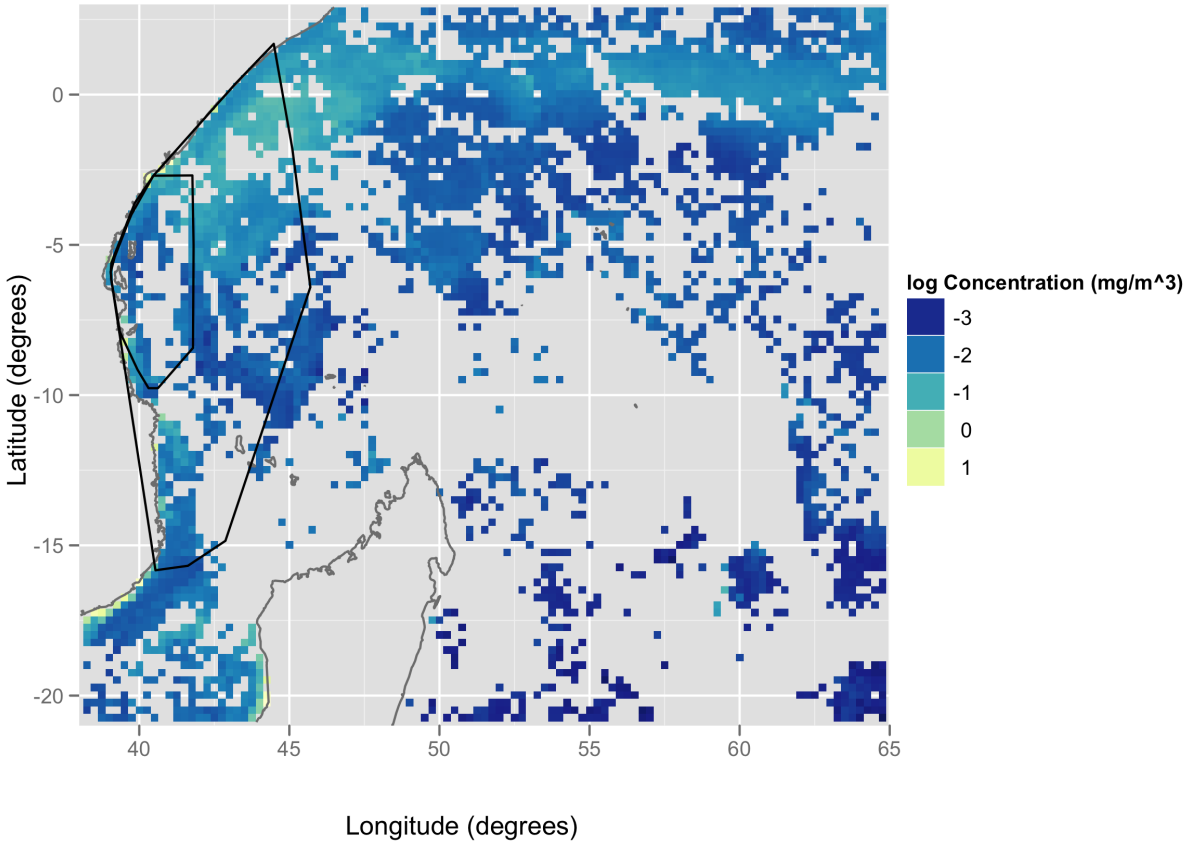


Figure 7.2: Reported chlorophyll concentration for Feb 22nd 2007.

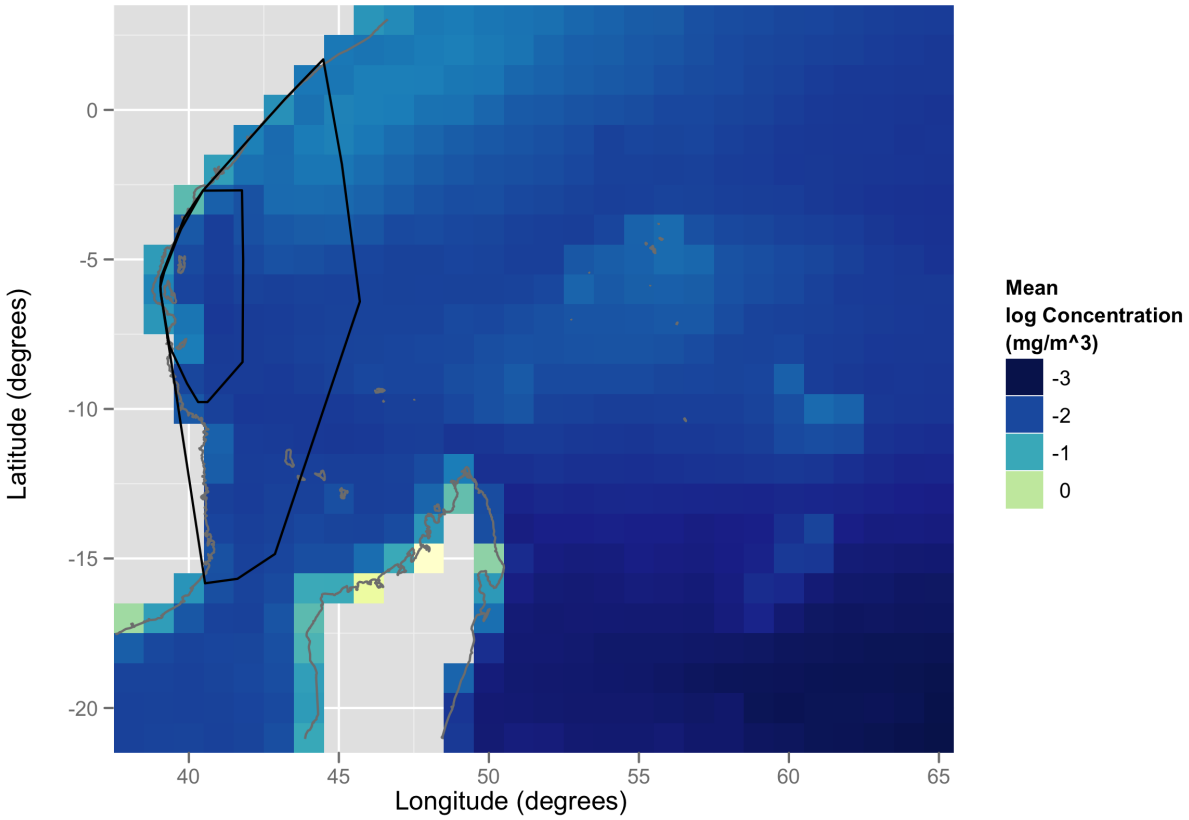


Figure 7.3: Average chlorophyll concentration over two years in degrees C. The colour represents the average log concentration.

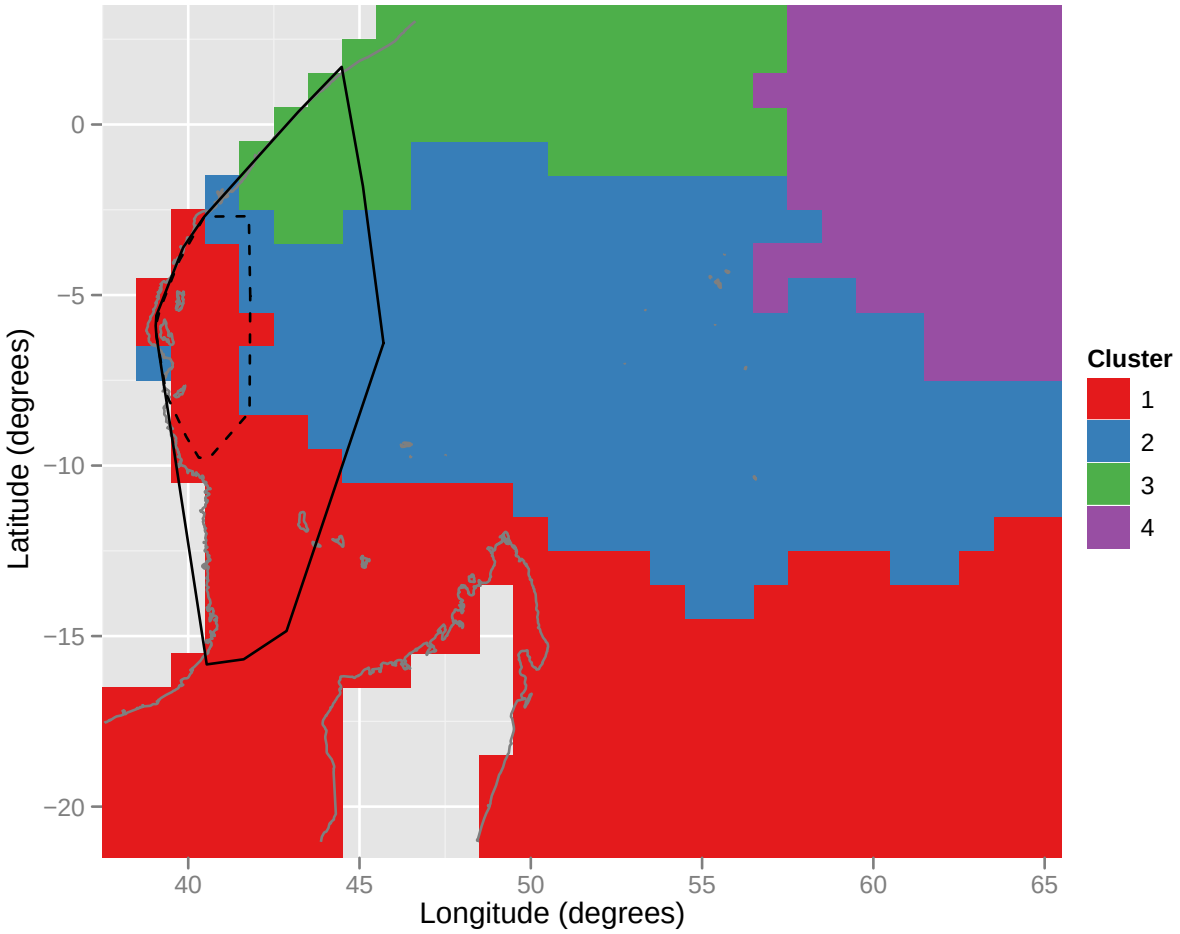


Figure 7.4: Clusters of chlorophyll concentration based on hierarchical clustering with 4 clusters.

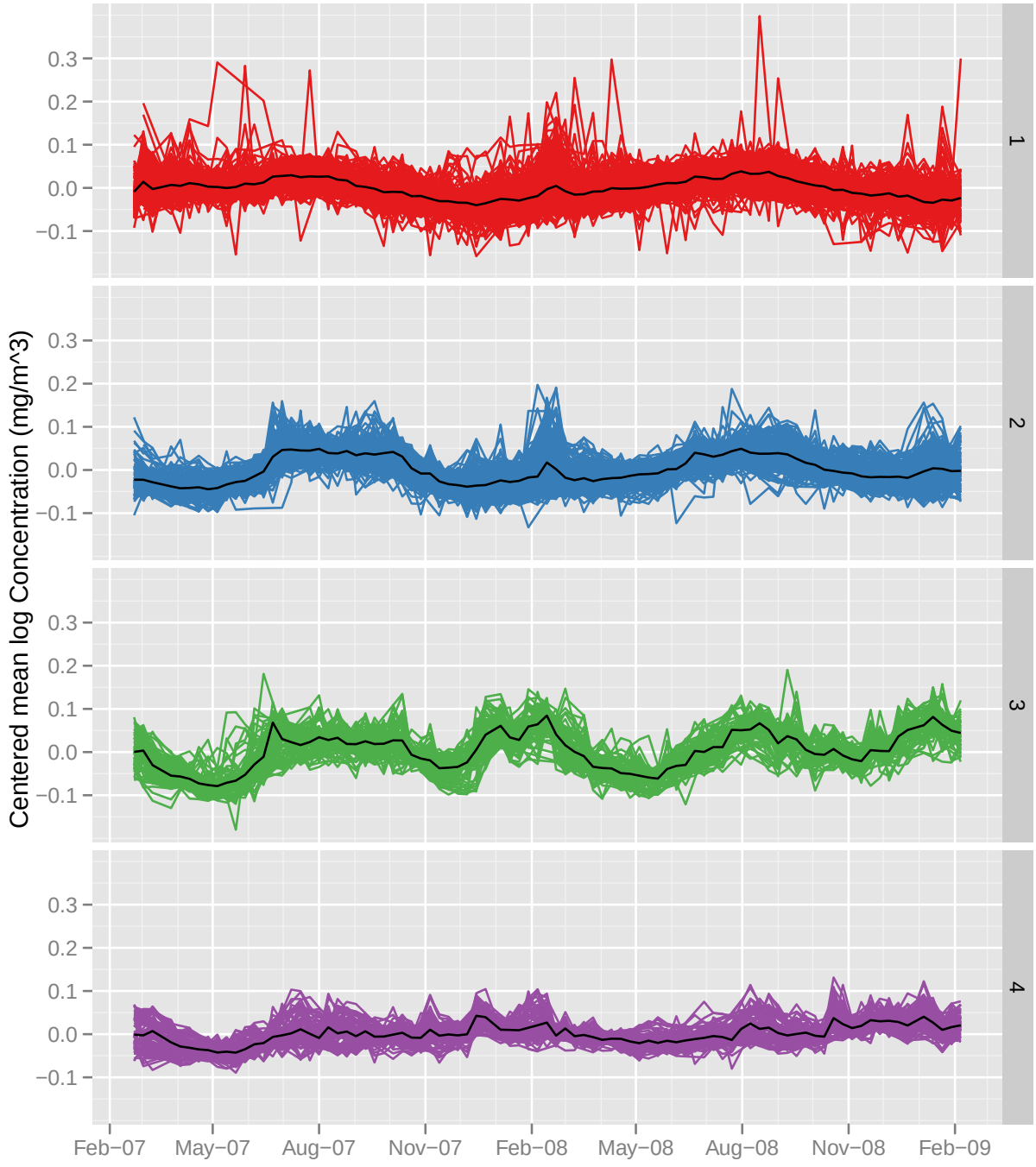


Figure 7.5: Time series of chlorophyll concentration by cluster based on hierarchical clustering with 4 clusters.

7.1.3 Sea surface temperature

Sea surface temperature is also available via the MODIS device on Aqua (MODIS SST). It is reported daily in degrees Celsius and downloadable at a resolution of 0.1 degrees latitude and longitude. A days worth of data are shown in Figure 7.6. There are again a lot of missing values. Figure 7.7 shows the average temperature over the two years the sharks are tracked. There is a warm area from the eastern equator that flows down towards the coast where the whale sharks were often observed. The time variation of sea surface temperature was also

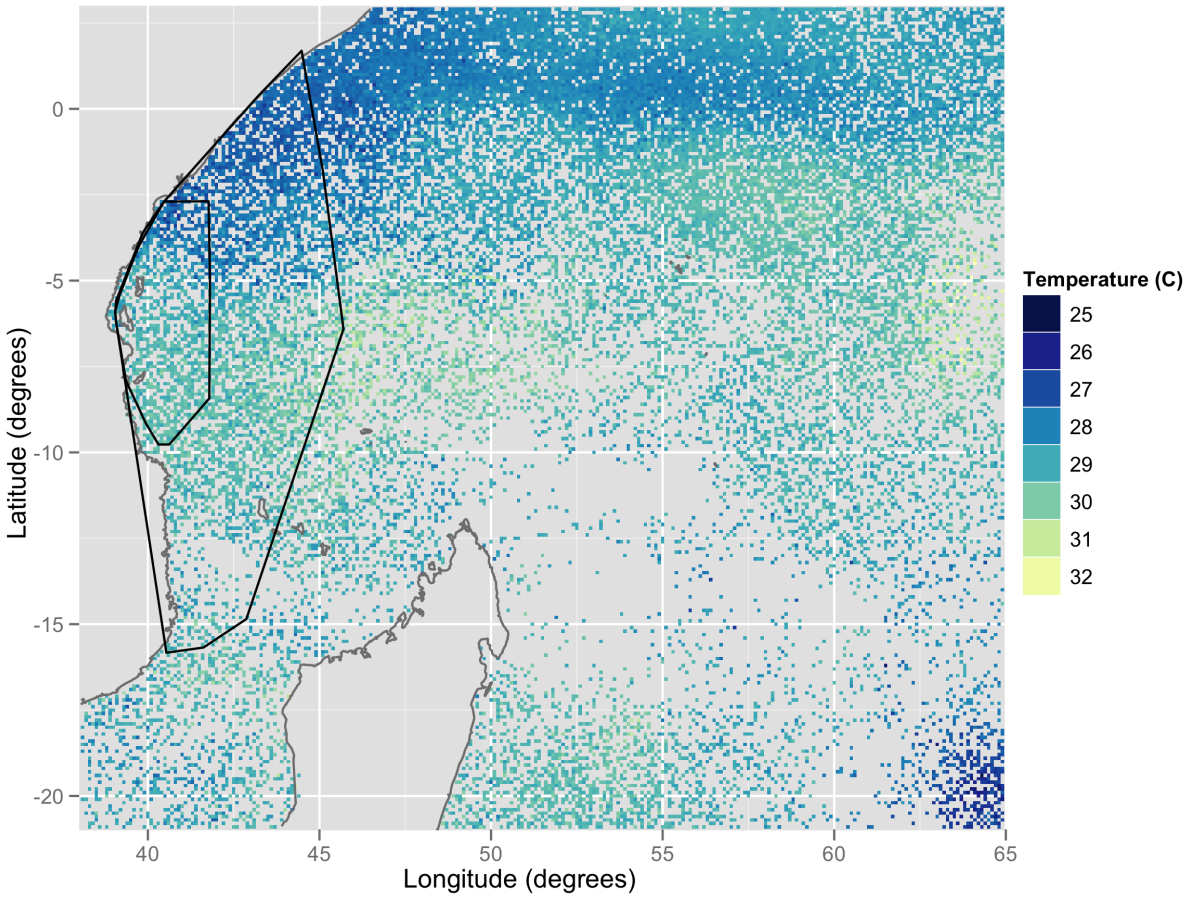


Figure 7.6: Sea surface temperature for Feb 18th 2007.

explored by clustering the centered time series (see Figures 7.8 and 7.9). Six clusters were chosen. Time series for the clusters (Figure 7.9) show in general higher temperatures in November through May. The exceptions are clusters 2 and 6, both corresponding to areas near the equator, where temperatures seem to be more stable throughout the year. Shark 66003 along with the other whale sharks occupies clusters 2, 3 and 4. These clusters are less sinusoidal than the clusters with lower latitudes but still exhibit some seasonality.

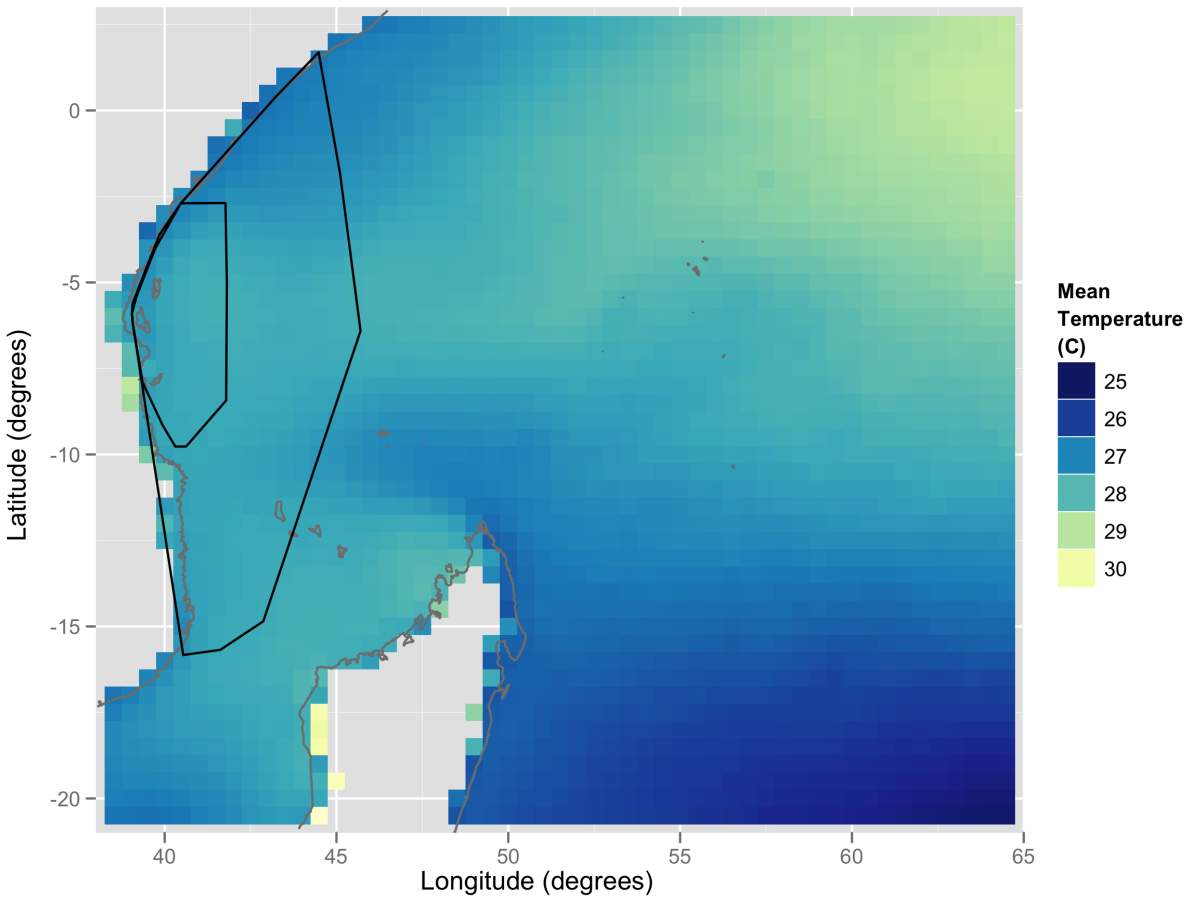


Figure 7.7: Average sea surface temperature over two years starting February 2007 (°C).

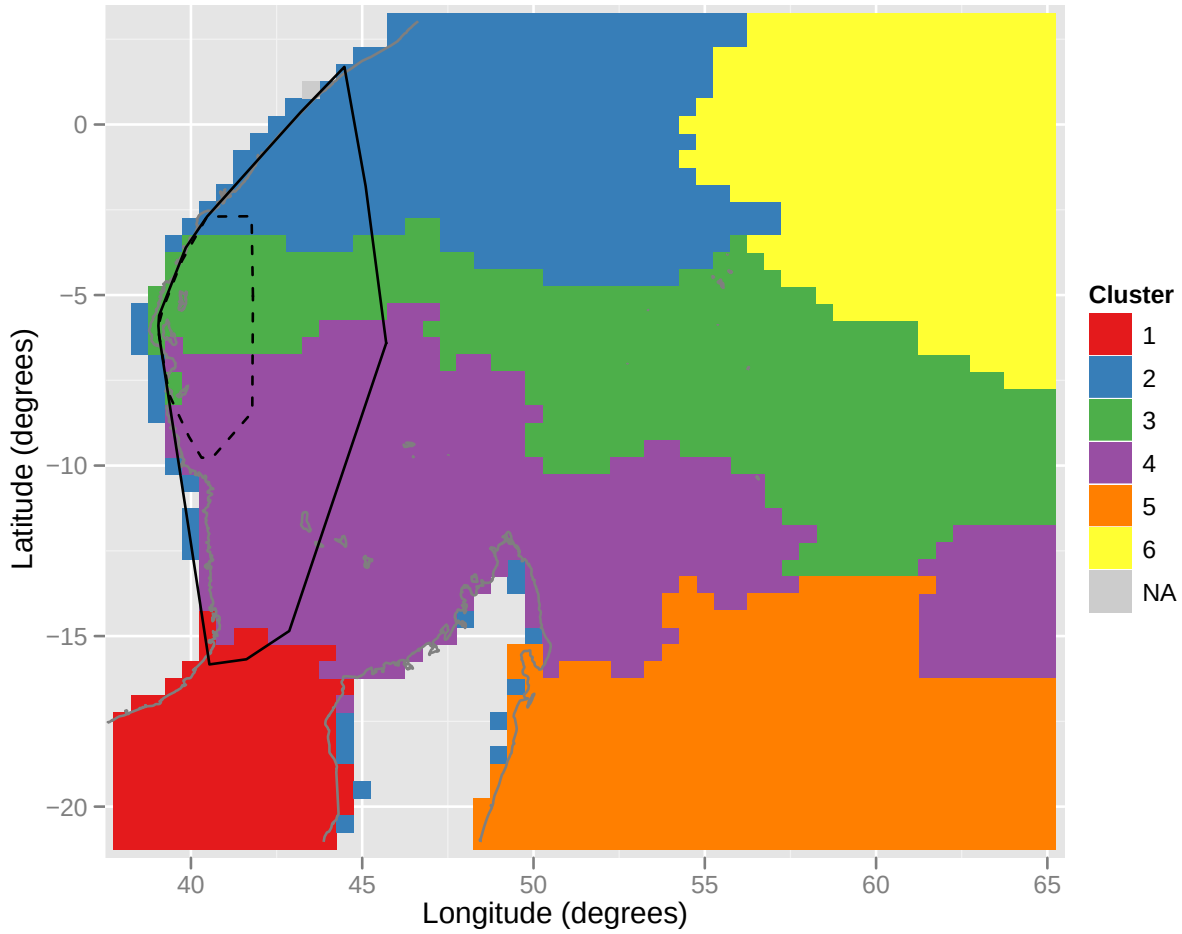


Figure 7.8: Clusters of sea surface temperature based on hierarchical clustering with 6 clusters.

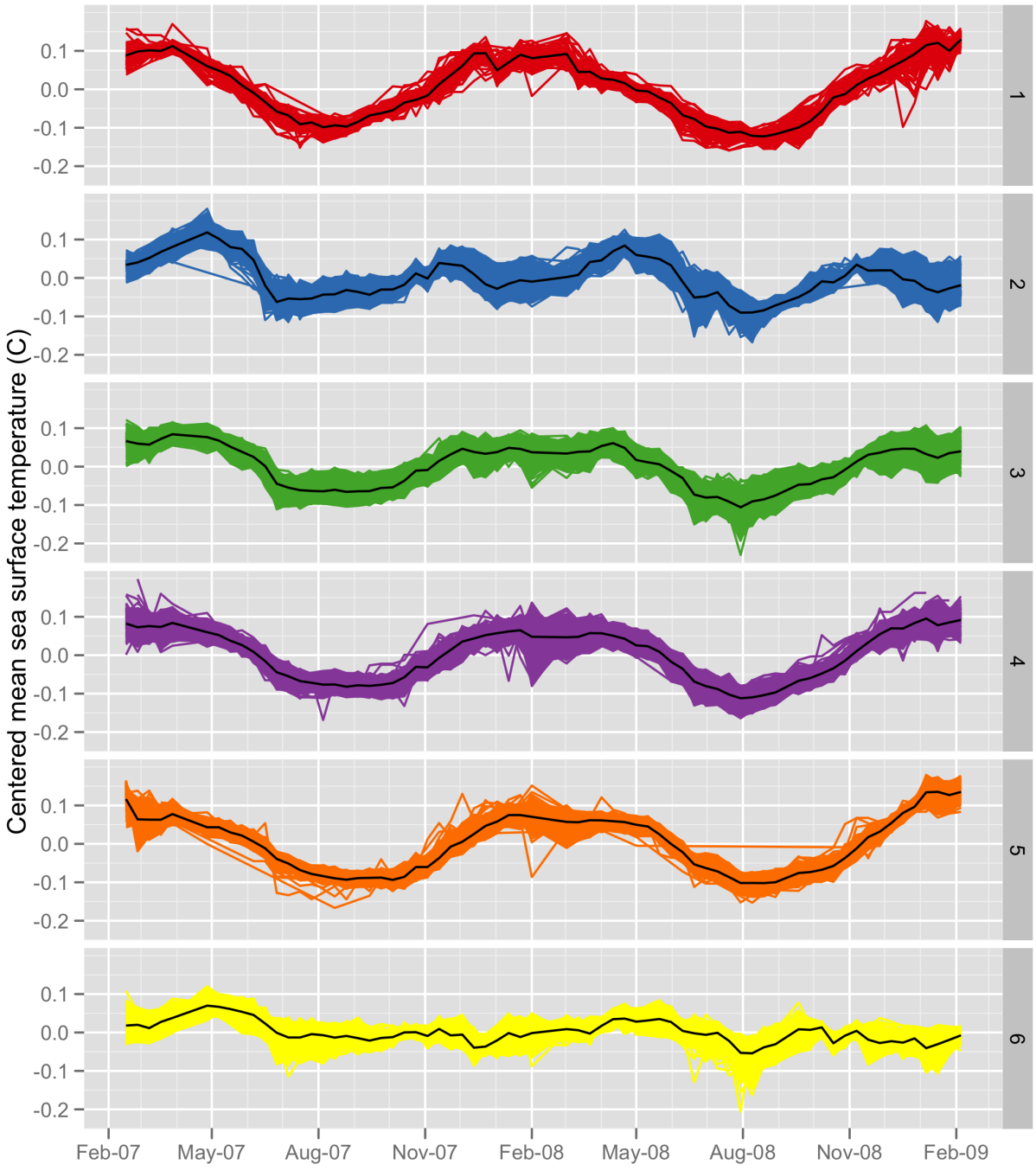


Figure 7.9: Time series of sea surface temperature by cluster based on hierarchical clustering with 6 clusters.

7.1.4 Sea surface currents

The sea surface currents are available from Coastwatch based on measurements from multiple satellites (Aviso Currents). The currents are inferred using the sea surface height (as a deviation from a reference level) and basic fluid mechanics: the two main drivers being gravity and the Coriolis force. The current is reported as a vector (eastward sea water speed, and northward sea water speed) in metres per second. Daily data with 0.25 degrees resolution is available.

Reported currents for one day, Feb 21st 2007, are shown in Figure 7.10, where the colour represents the magnitude of the current. Simply averaging the currents over the two year

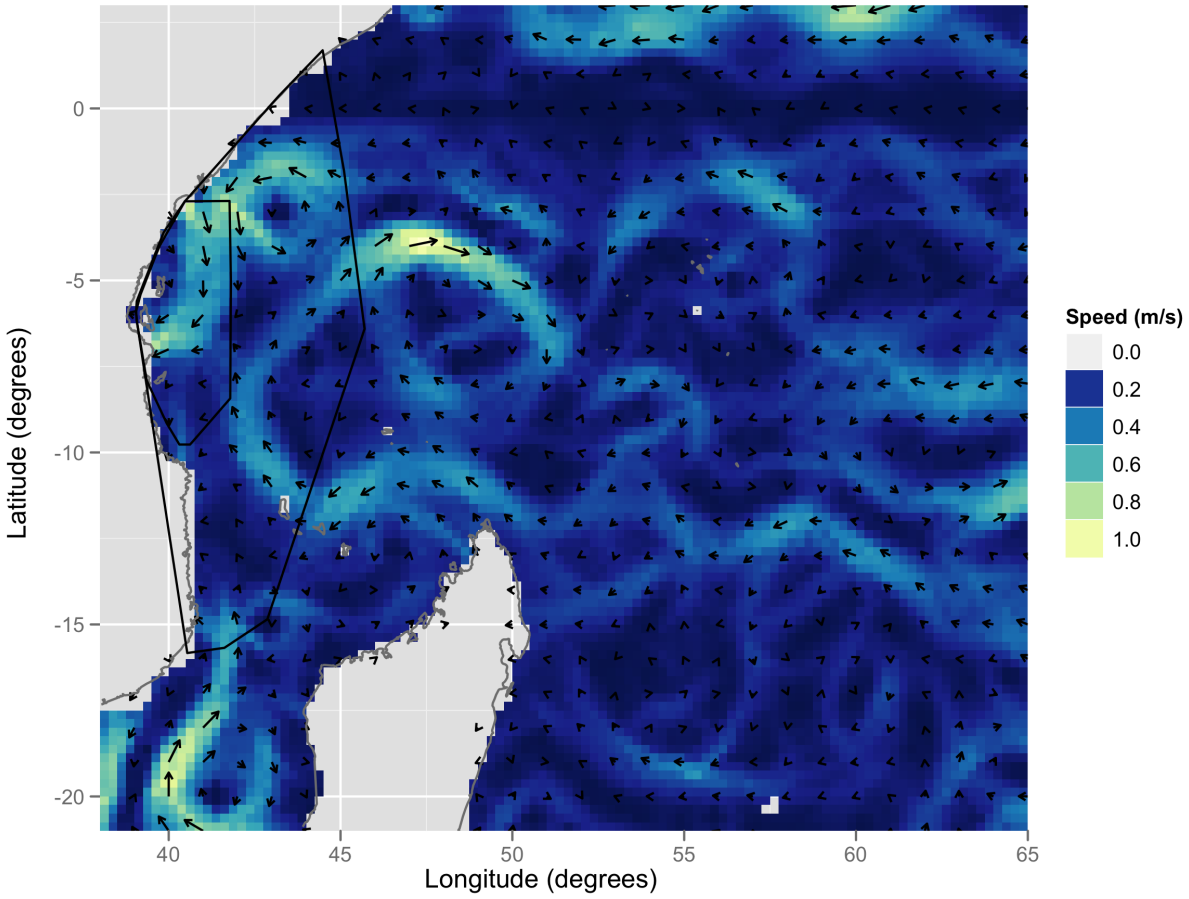


Figure 7.10: Sea surface currents denoted by arrows for Feb 21st 2007. The colour of the tiles corresponds to the magnitude of the current.

period is inappropriate as a location with opposite currents in Summer and Winter will on average have little current. Instead to understand the variation over time, all current vectors are plotted as points around their location in Figure 7.11. The red dots indicate the case where the current $= (0, 0)$, and the currents are plotted relative to this for each

location. For example, the currents at -8°– -9° N 40°– 45° E longitude were consistently onshore. In contrast, in the bottom left the currents were in highly variable directions but had much larger magnitude. Some current data appear on land. This is a result of the gridded resolution of the data. If a grid square contains some ocean, then currents are reported, and I am plotting them at the center of the grid square even though this could be land. The currents are zero at the equator due to the model used to move from ocean height to currents breaking down for latitude zero.

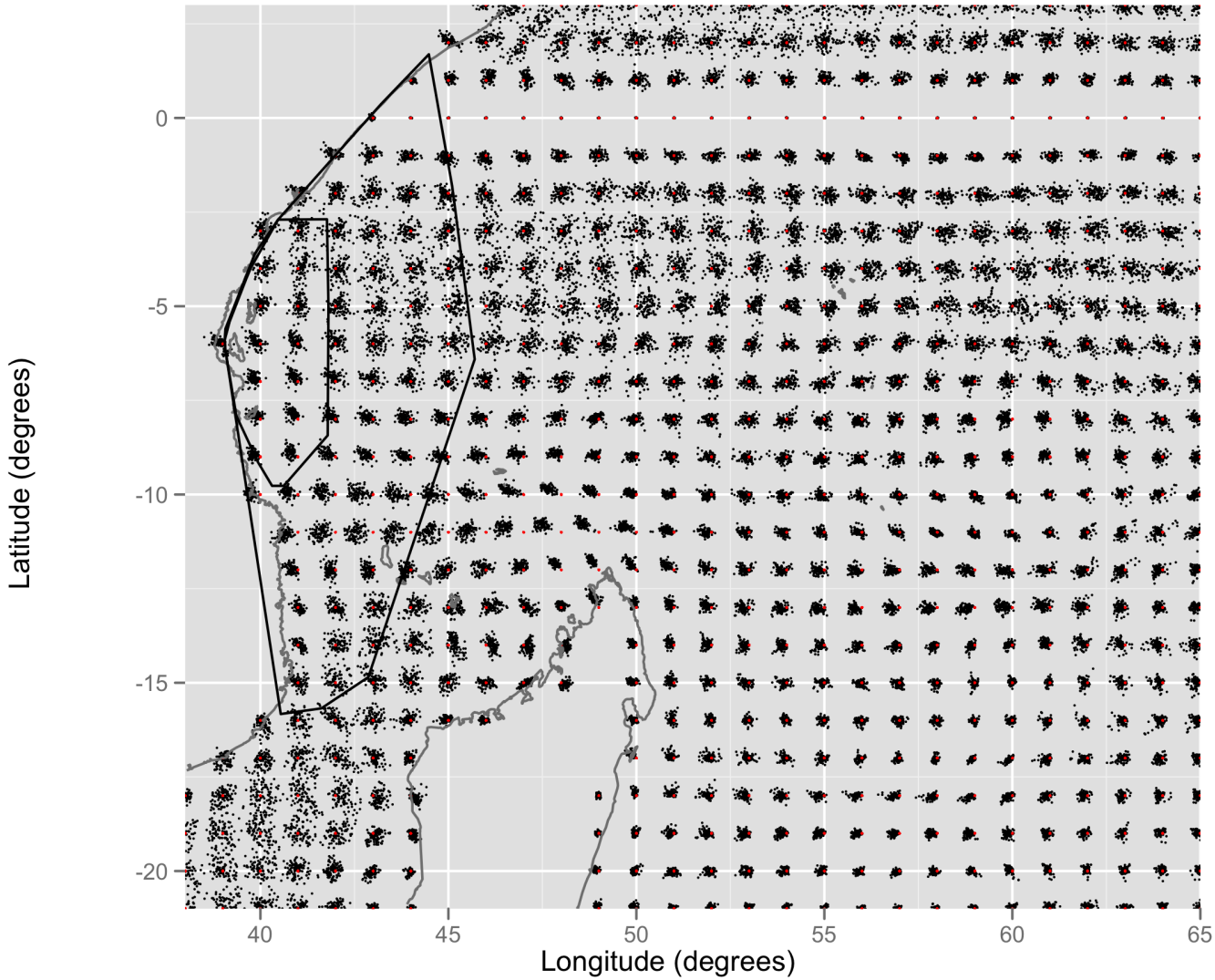


Figure 7.11: Sea surface currents over the whole two years at one degree resolution. Red dots indicate the case where the current $= (0, 0)$, and the currents are plotted relative to this for each location.

7.1.5 Surface winds

Surface wind data are available from Remote Sensing Inc. measured using the SeaWinds instrument on NASA's QuikSCAT satellite: "The SeaWinds scatterometer bounces radar signals off the ocean surface and measures the strength of the returning backscatter signal. Wind roughens the ocean surface, which affects the backscatter signal and enables measurement of wind velocity"(NASA QuikSCAT Winds). Wind is reported as a vector (meridional wind, zonal wind) both in metres per second and available as a 3-day composite, with 0.125° resolution in longitude and latitude.

One day, Feb 26th 2007, is shown in Figure 7.12 where the resolution has been downsampled to 1 degree in latitude and longitude for readability. On this day the largest feature is the cyclonic system near (53E, -18N): a low pressure area combined with the Coriolis force produces strong clockwise winds. Two years worth of wind data are shown in Figure 7.13.

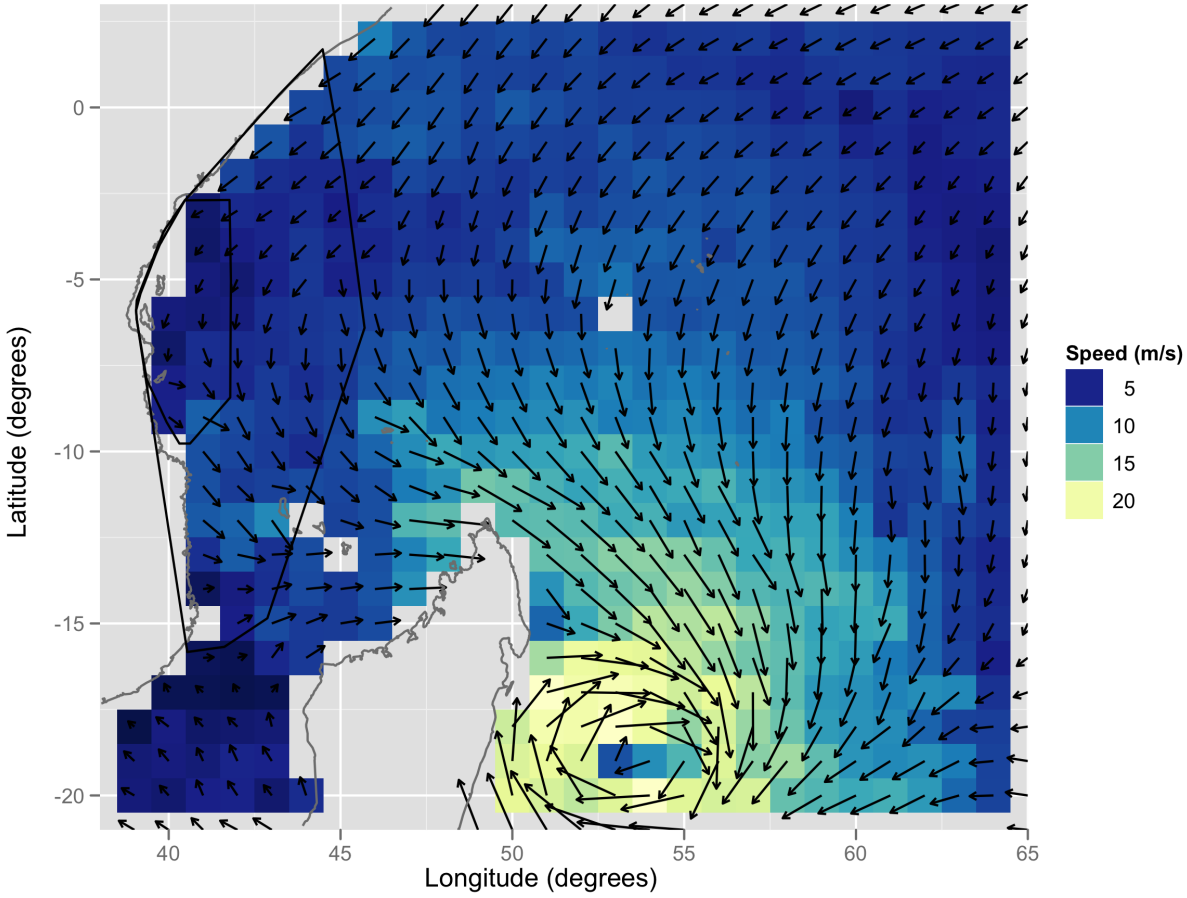


Figure 7.12: Surface winds denoted by arrows for Feb 26th 2007. The colour of the tiles corresponds to the magnitude of the wind.

There is a much more systematic pattern than that in the currents. In the area shark 66003

is observed the winds are usually from the north or south. Further investigation found the winds were generally northerly (from the north) November through March and southerly during the rest of the year.

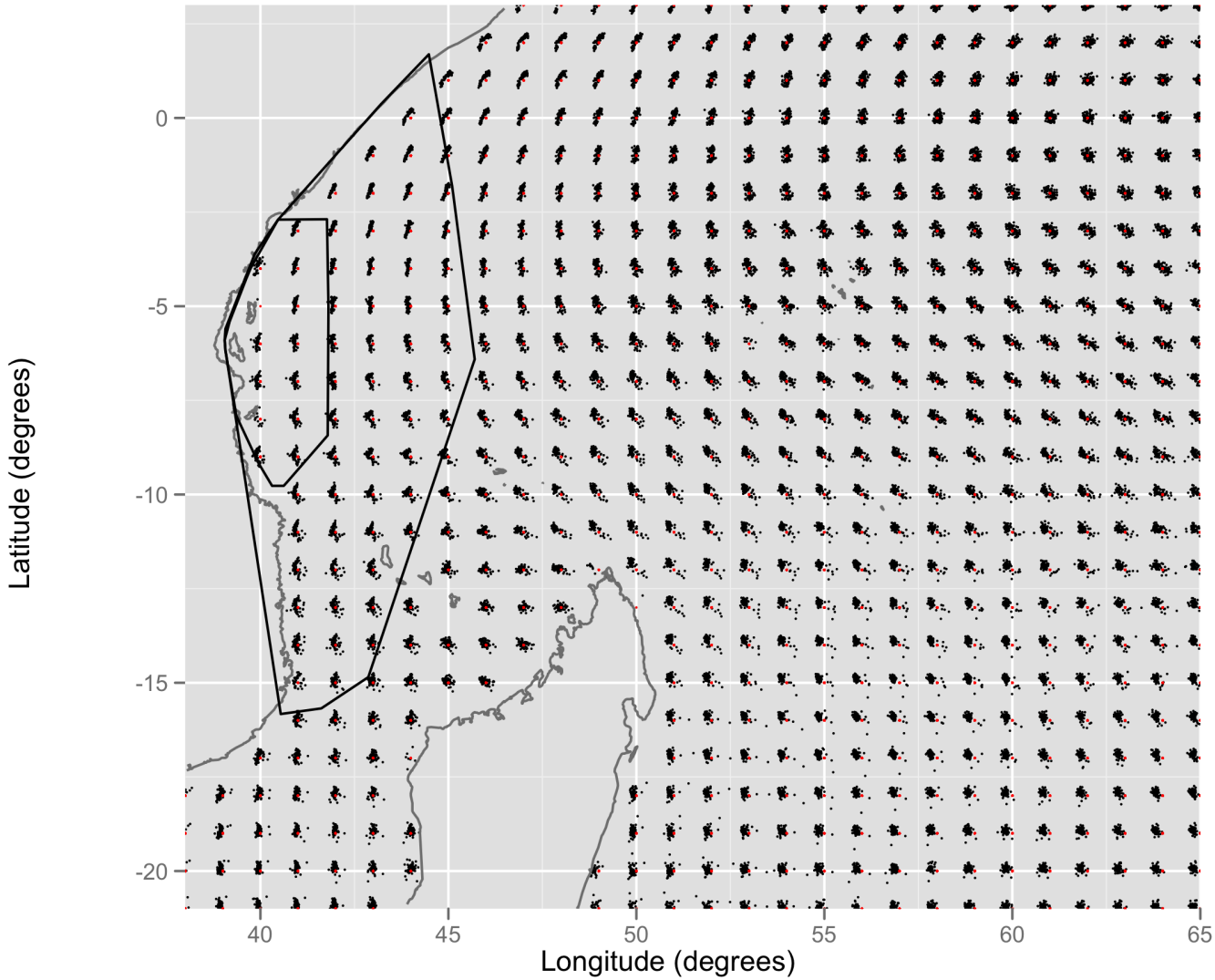


Figure 7.13: Winds over the the two years starting in February 2007. Red dots indicate the case where the wind $= (0, 0)$, and the winds are plotted relative to this for each location.

7.2 Habitat association

While the above exploration gives an overview of the covariates in the area of interest it is hard to determine if shark 66003 is showing any associations with their levels. One way to explore the habitat shark 66003 travels through is to simply look at a density plot of the covariates at the observed locations of the shark. To have some frame of reference one needs something to compare this to. Can the area the shark spends time in be distinguished by the level of the covariates? Framing the question in a more statistical way might be: do the covariates sampled by the shark differ from a random sample of the available covariates? The distribution of available covariates form a null distribution and one wants to ask if the covariates seen by the shark could be a random sample from this distribution.

A pertinent question regarding methods is: how do we define the available covariates? One way is to define it as the distribution of the covariates within our area of interest (the West Indian ocean). The distribution can be expected to be highly dependent on how this area is chosen.

Perhaps a better way is to restrict it to the area that shark 66003 travels in. The definition I will use here is the convex hull of the track (the smallest convex polygon that includes all of the observed locations for a given time period). Martin et al. (2008) suggested rotating the track to provide a null sample. This is inappropriate here as it will most likely result in a track on land. A better approach is to use the bootstrap tracks from the previous chapter based on a quadratic potential function with repulsion off the land.

There are three "null" distributions that I will use to sample 201 locations (the same number of observed locations for shark 66003). The densities of their covariate values are compared to those observed along the actual track. To evaluate sampling variability each null case has 10 tracks sampled. Comparing the densities from samples from the entire range to the convex hull of shark 66003 indicates how the area where shark 66003 was observed varies from the rest of the ocean. Comparing the densities from the convex hull of shark 66003 to the bootstrapped tracks highlights the importance of location on the covariate by indicating how much of the covariate range is accounted for by the quadratic model. Comparing the bootstrapped tracks to the observed track tells us if the quadratic model is enough to explain the range of covariate encountered. A significant deviation from the bootstrapped tracks can indicate that the quadratic potential isn't capturing association between the shark and ocean habitat.

The densities for the scaler covariates are shown in Figure 7.14. For bathymetry, the distributions defined by the range of shark 66003 and the bootstrap tracks, have much shallower water than the entire West Indian ocean. The shark spent more time in areas where the water was 1000m deep than expected under the bootstrapped track. The observed sea surface temperature and chlorophyll concentration were not inconsistent with either the bootstrapped track or random sample from the sharks range. They both exhibit differences from the whole area consistent with being near the coast.

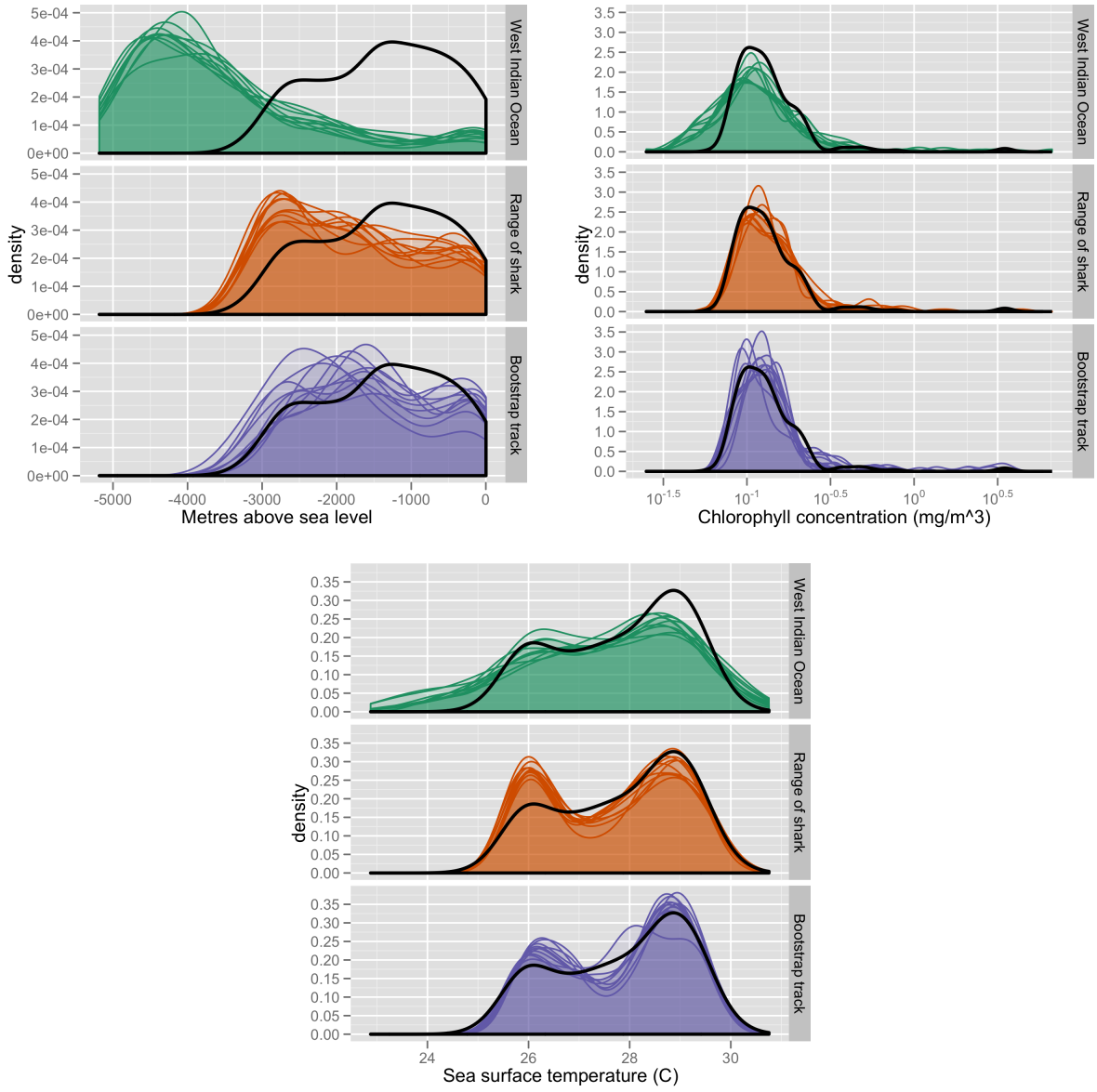


Figure 7.14: Estimated densities for the scalar valued covariates encountered by shark 66003 (black) compared to 201 randomly sampled points in the West Indian ocean (green), 201 randomly sampled from the range of shark 66003 (blue) and a bootstrapped track using the quadratic potential (red).

The densities for the vector-valued covariates are shown in Figure 7.15. For the surface current first, the shark spends a slightly longer time in positive currents in both directions than expected. The observed surface winds appear consistent with random samples within the shark's home range and the bootstrapped tracks.

7.3 Covariates as components of potential functions

The real-valued covariates describe a surface over the area of interest and one can ask whether they influence the sharks movement via gradients of potential functions. The chlorophyll, sea surface temperature and bathymetry are reported as surfaces, (i.e. height(x, y, t)), the surface current and surface winds are reported as gradients of potential functions for water and air movement. A model incorporating all the covariates is,

$$\begin{aligned} d\mathbf{r}(t) = \big(&- \gamma_1 \nabla \text{Temp}(\mathbf{r}(t), t) - \gamma_2 \nabla \text{Chlorophyll}(\mathbf{r}(t), t) - \gamma_3 \nabla \text{Bathymetry}(\mathbf{r}(t), t) \\ &+ \gamma_4 \text{ Water}(\mathbf{r}(t), t) + \gamma_5 \text{Air}(\mathbf{r}(t), t) - \nabla H(x, y) \big) dt + \sigma d\mathbf{B}(t) \end{aligned}$$

Here $H(x, y)$ allows for a time homogeneous potential term on position. Positive values for γ_1, γ_2, γ_3 would indicate a tendency to move towards lower levels of chlorophyll, cooler surface temperatures and deeper waters. Positive values on γ_4 and γ_5 would indicate a tendency to move with the surface currents and with the surface winds.

The gradients of the chlorophyll, temperature and bathymetry were approximated using Sobel operators (Acharya and Ray (2005) pg 135). Let $C(i, j)$ denote the value of the covariate at gridpoint i, j and let $\mathbf{C}_3(i, j)$ be the 3×3 matrix of the covariates values consisting of the point (i, j) and its 8 direct neighbours. That is,

$$\mathbf{C}_3(i, j) = \begin{pmatrix} C(i-1, j+1) & C(i, j+1) & C(i+1, j+1) \\ C(i-1, j) & C(i, j) & C(i+1, j) \\ C(i-1, j-1) & C(i, j-1) & C(i+1, j-1) \end{pmatrix}$$

The gradient at grid point (i, j) is calculated as a weighted sum of the surrounding points,

$$\nabla C(i, j) \approx \Big(S_x * \mathbf{C}_3(i, j), S_y * \mathbf{C}_3(i, j) \Big) \quad \text{where} \quad S_x = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \quad \text{and} \quad S_y \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

where $*$ indicates the two dimensional convolution operator. In fact this gives the derivative multiplied by 8 and in units of the resolution of the observation grid. The empirical derivatives were transformed to "change in covariate per km".

The gradient value of the covariate for a given location in UTM coordinates was found by first finding the closest measured grid point and then interpolating the value at this grid point linearly between the two closest times with measurements.

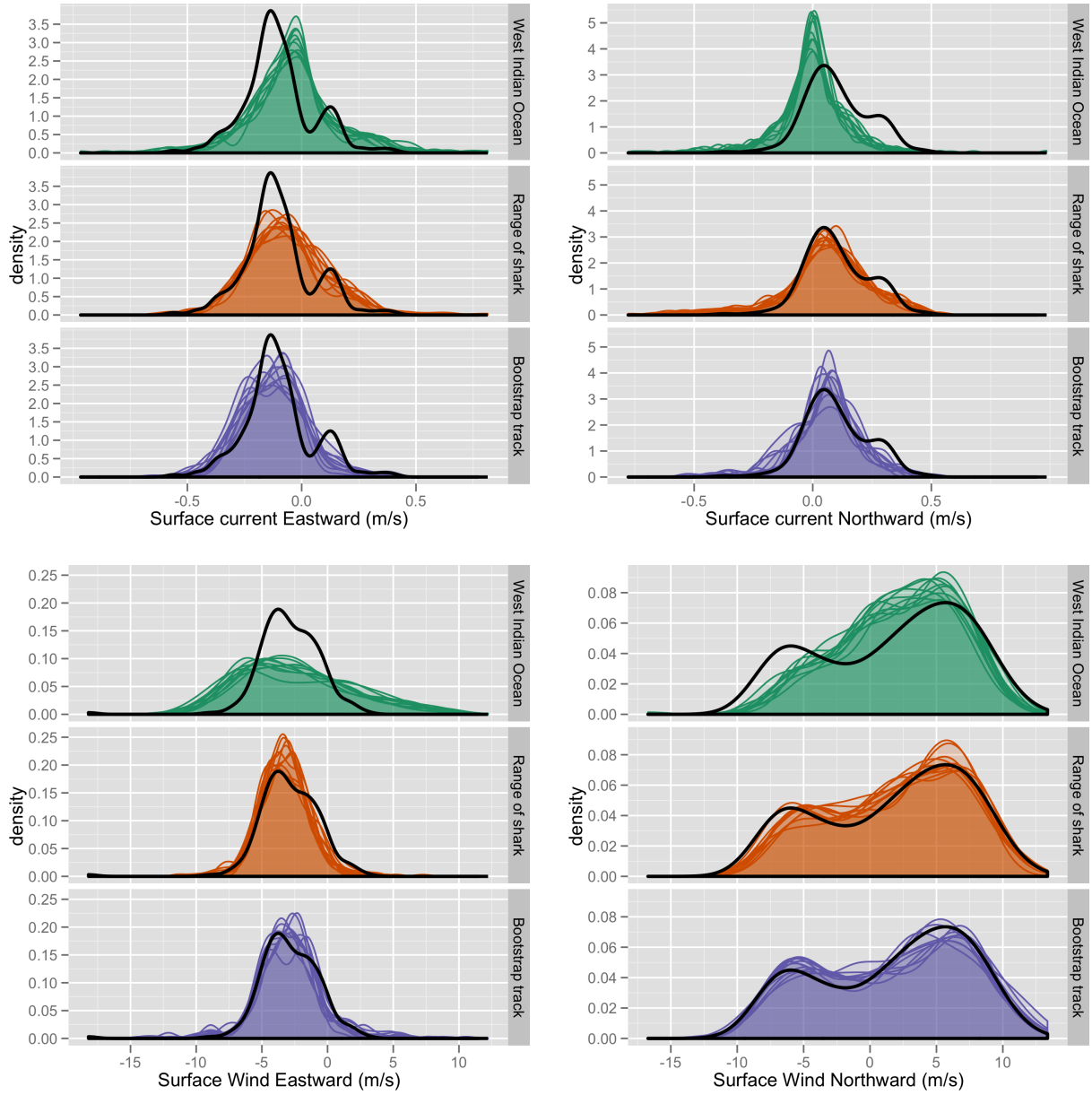


Figure 7.15: Estimated densities for the vector valued covariates encountered by shark 66003 (black) compared to 201 randomly sampled points in the West Indian ocean (green), 201 randomly sampled from the range of shark 66003 (blue) and a bootstrapped track using the quadratic potential (red).

Table 7.1: Estimates and approximate 95% confidence intervals for the potential model including covariates

Parameter	Estimate	95% CI
β_1	−95.9	(−144.9,−46.8)
β_2	53.0	(15.6, 90.5)
β_3	0.07	(0.03, 0.10)
β_4	0.03	(0.01, 0.05)
β_5	−0.01	(−0.03, 0.01)
log(Chlorophyll)	43.9	(−110.3, 198.1)
Bathymetry	143.0	(−182.1, 468.0)
Sea Surface Temperature	−0.15	(−0.56, 0.26)
Current	−28.1	(−56.6, 0.5)
Wind	−0.22	(−1.54, 1.11)
S_y	2.4	(2.1, 2.8)
σ	38.5	(34.9, 42.5)

The above model with a quadratic for $H(x, y)$, as in the previous chapter, was fit using generalised least squares allowing different variances on the errors in the east-west and north-south directions.

Coefficient estimates for the fitted model are shown in 7.1. A likelihood ratio test gave no evidence against the hypothesis that all of the covariate parameters were zero (p-value = 0.38). None of the covariates reached significance at the 5% level (under the assumptions of the Lai-Wei theorem, see 3.5) but the coefficient on sea surface current was close (p-value = 0.053). The negative sign on this coefficient could imply an association with moving in the opposite direction to the current.

In an attempt to capture the time spent in 1000m deep water (Bathymetry$(\mathbf{r}(t), t)-1000)^2$ was substituted into the potential for the Bathymetry term. This term acts as an attractor to areas where the sea floor is 1000m below the surface. This new term was not significant (p-value = 0.86) and did not significantly alter the quadratic potential parameters. This attraction term was also added to the quadratic potential (without including the other covariates) and did not significantly improve the fit (p-value = 0.95).

7.4 Discussion

Shark 66003 spent more time in 1000 m deep water than would be expected from just a quadratic potential function model. Adding an attraction to 1000m deep water via a quadratic potential term did not significantly improve the fit of the quadratic potential model.

The form of the potential used to incorporate the covariates only allowed their influ-

ence through movement towards lower (or higher) values. The addition of the quadratic bathymetry term illustrated one way to add attraction to a specific value of a covariate. It may be more realistic to model the influence of covariates as an attraction to a range of values, or even that the shark follows a contour (a level curve) of a covariate. Describing and fitting appropriate potentials for these cases is an area for future work.

One of the reasons for failing to find many covariate effects is the poor resolution of the data. The shark is probably making movement decisions based on the environment within metres of its location but we only have information on its location and its environment at a resolution of tens of kilometers. Whale sharks are being found to spend very little time at the surface (B. Stewart, pers. comm.) so even finding an association with the covariates available does not give much information on the sharks preferences as the true covariates of interest are below the surface.

There was a very weak suggestion that this whale shark's movement tended against the surface current. More evidence might be built by looking at other sharks and this is work in progress.

Chapter 8

One shark two shark red shark blue shark

This chapter investigates methods for exploring interactions among several animals. Two sharks that are observed at the same time are used to illustrate the methods. There is no conclusive evidence for these sharks interacting but perhaps that is expected when they are frequently greater than 500km apart. It is hoped that these methods will be more fruitful on animals tracked at better temporal and spatial resolution, particularly when multiple animals from the same group (family or herd, say) are tracked.

Sharks 80181 and 80169 were tracked for a common period of about four and a half months from the 2rd of March to 15th of July 2008, a total of 136 days. Both sharks stayed near the coast of West Africa. The times measurements were recorded on each shark during this period are shown in Figure 8.1. Shark 80181 (blue) is observed almost every day throughout the time period, for a total of 128 days. Shark 80169 (red) has a few more missing days with a total of 101 observations. Since there is roughly one measurement per day I take a common equispaced time scale, $t(i) = t_i = \text{day } i$, where days in which a shark is not measured are considered missing. In order to proceed the missing observations are linearly interpolated from the surrounding non-missing positions. The resulting positions are shown in Figure 8.2.

A plot of the distance between the two sharks from this interpolated data is shown in Figure 8.3. The sharks are at closest about 100km apart which is within the errors associated with the light level geolocation. There is some oscillation, but the sharks are not more than about 1200km apart except for near the end of the period when shark 80181 (blue) headed northwards along the coast. One might wonder if this restriction in distance apart is a result of similar potentials based on position (for example, the sharks have attractors near the same point) or an interaction between the sharks. Over such large distances the sharks are not expected to be directly communicating but exploring their observed interaction might uncover an common underlying driver, such as temperature or bathymetry, in their movement.

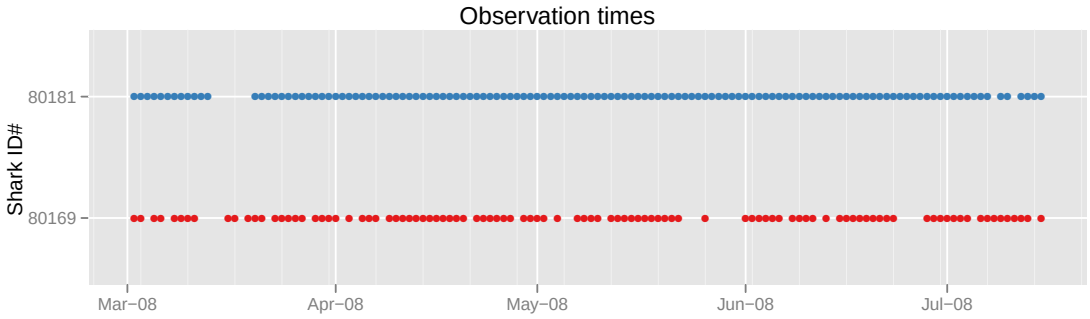


Figure 8.1: Position measurement times (marked by a point) for the two sharks during a common period. The period consists of a total of 136 days. Shark 80181 (blue) had measurements for 128 of these days and shark 80169 (red) for 101 of these days.

8.1 Modelling interactions

To explore the interaction between the two sharks I introduce a potential model that includes a dependence on the difference in their positions. Let $\mathbf{r}_i(t) = (x_i(t), y_i(t))$, for the two sharks, $i = 1, 2$. To simplify, I will assume the interaction between the animals only depends on the distance between them. Let the distance from shark 1 to shark 2 be denoted, $d_{1,2}(t) = \sqrt{(x_2(t) - x_1(t))^2 + (y_2(t) - y_1(t))^2}$. The model of the movement of the two sharks is now assumed to be,

$$d\mathbf{r}_1(t) = -\nabla H(\mathbf{r}_1(t))dt - \nabla V(d_{1,2}(t))dt + \mathbf{\Sigma}_1 d\mathbf{B}_1(t)$$
$$d\mathbf{r}_2(t) = -\nabla G(\mathbf{r}_2(t))dt - \nabla U(d_{2,1}(t))dt + \mathbf{\Sigma}_2 d\mathbf{B}_2(t)$$

The H and G are potential functions on the position of the sharks and U and V are the potential functions on the distance to the other shark. $\mathbf{B}_1$ and $\mathbf{B}_2$ are independent, two-dimensional Brownian motion processes. As seen in earlier chapters, a quadratic in x and y may be appropriate for the position terms. That is let,

$$H(\mathbf{r}) = \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy$$
$$G(\mathbf{r}) = \gamma_1 x + \gamma_2 y + \gamma_3 x^2 + \gamma_4 y^2 + \gamma_5 xy$$

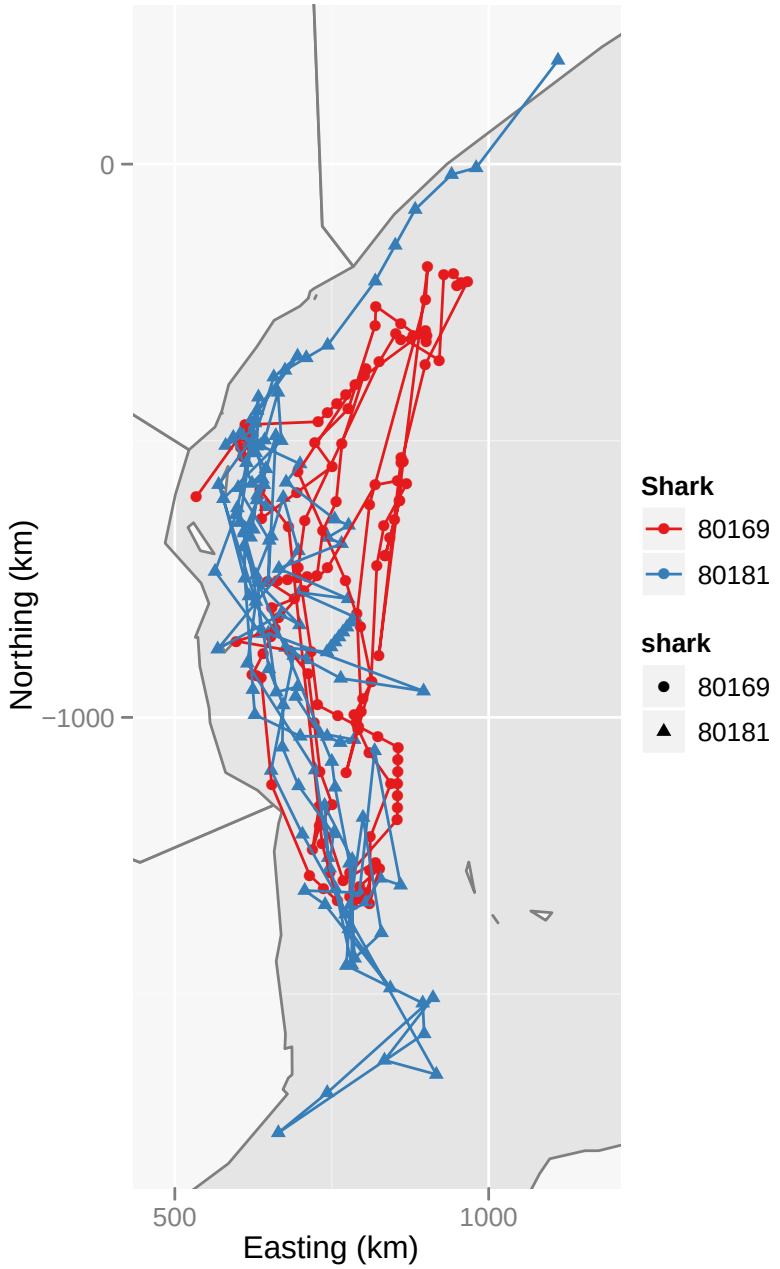


Figure 8.2: Interpolated positions of the two sharks for the union of observed times (136 time points).

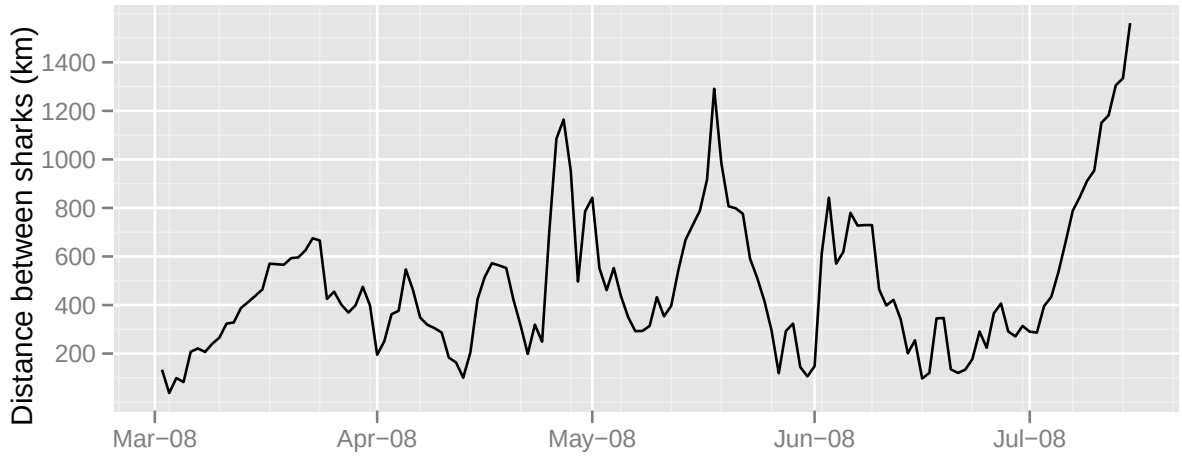


Figure 8.3: Distance between the two sharks over the commonly observed period.

For generality, let the potential on the distance to the other shark be any smooth differentiable function. Consider the first shark,

$$\begin{aligned}
&-\nabla V(d_{1,2}(t)) \\
&\qquad = -\left(\frac{\partial}{\partial x_1} V(d_{1,2}(t)), \frac{\partial}{\partial y_1} V(d_{1,2}(t))\right) \\
&\qquad = -\left(\frac{\partial V(d_{1,2}(t))}{\partial d}\frac{\partial d}{\partial x_1}, \frac{\partial V(d_{1,2}(t))}{\partial d}\frac{\partial d}{\partial y_1}\right) \\
&\qquad = -\left(\frac{\partial V(d_{1,2}(t))}{\partial d}\frac{x_1(t)-x_2(t)}{d_{1,2}}, \frac{\partial V(d_{1,2}(t))}{\partial d}\frac{y_1(t)-y_2(t)}{d_{1,2}}\right).
\end{aligned}$$

The x and y directions contain the same term for the dependence on d with different multipliers. Let $\frac{\partial V(d)}{\partial d}$ be denoted $v(d)$. Then by stacking the equations for the x and y directions a common $v(d)$ can be estimated. There are a number of options for estimating v. One would be to use a generalized additive model, another would be to represent v with a b-spline basis. I will use the later here as it allows more complete modelling of the variance structure within the current implementations of `mgcv::gam` and `nlme::gls` in R. There is also a choice between placing the bspline basis on v or V. Here, I choose to put it on V approximating its derivative by first differences. This allows easy calculation of the fitted function and point-wise standard errors.

To be specific I will model V using a cubic b-spline basis with five degrees of freedom. That is

$$V(d) = \sum_{i=1}^{k} B_i^m(d)\theta_i$$

with $m = 3$ and $k = 5$. The basis functions, B_i^m are defined recursively by,

$$B_i^m(d) = \frac{d - d_i}{d_{i+m+1} - d_i} B_i^{m-1}(d) + \frac{d_{i+m+2} - x}{x_{i+m+2} - x_{i+1}} B_{i+1}^{m-1} \quad i = 1, \ldots, k$$

where d_i are knot locations, chosen here as quantiles of the data. The derivative of this function will be needed to fit to the data. It is approximated by

$$\frac{\partial V(d)}{\partial d} \approx V(d+1) - V(d)$$

The same calculation can be done for the second shark resulting in,

$$-\nabla U(d_{2,1}(t)) = -\left(\frac{\partial U(d_{2,1}(t))}{\partial d} \frac{x_2(t) - x_1(t)}{d_{2,1}}, \frac{\partial U(d_{2,1}(t))}{\partial d} \frac{y_2(t) - y_1(t)}{d_{2,1}} \right)$$

If both sharks have the same relationship to the distance of the other shark, that is $V(d) = U(d)$, then since $d_{1,2} = d_{2,1} = d$, both sharks will have a common $\frac{\partial V(d(t))}{\partial d} \frac{1}{d}$ term and the multipliers of shark 2 will have the opposite sign of shark 1.

Let sharks 1 and 2 be shark 80169 and shark 80181 respectively.

8.1.1 Is there a distance based potential?

To test for the distance based potential two models were fitted. Model 1 has a position and distance to the other shark potential for each shark,

$$\begin{aligned} d\mathbf{r}_1(t) &= -\nabla H(\mathbf{r}_1(t))dt - \nabla V(d_{1,2}(t))dt + \mathbf{\Sigma}_1 d\mathbf{B}_1(t) \\ d\mathbf{r}_2(t) &= -\nabla G(\mathbf{r}_2(t))dt - \nabla U(d_{2,1}(t))dt + \mathbf{\Sigma}_2 d\mathbf{B}_2(t) \end{aligned}$$

Model 2 has no distance to the other shark potential,

$$\begin{aligned} d\mathbf{r}_1(t) &= -\nabla H(\mathbf{r}_1(t))dt + \mathbf{\Sigma}_1 d\mathbf{B}_1(t) \\ d\mathbf{r}_2(t) &= -\nabla G(\mathbf{r}_2(t))dt + \mathbf{\Sigma}_2 d\mathbf{B}_2(t) \end{aligned}$$

In both models the Σ's are assumed diagonal but are allowed different values in the x and y directions. They are parametrized in terms of the standard deviation of shark 1 in the x direction, so that

$$\mathbf{\Sigma}_1 = \begin{pmatrix} \sigma & 0 \\ 0 & S_y\sigma \end{pmatrix} \quad \mathbf{\Sigma}_2 = \begin{pmatrix} T_x\sigma & 0 \\ 0 & T_y\sigma \end{pmatrix}$$

In order to fit the models the Euler approximation is used to reduce things to a linear regression problem. For example the observed velocity in the x direction for shark 1 can be

written as,

$$\mathbf{x}_1(t+1) - \mathbf{x}_1(t) \approx \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ -2x \\ 0 \\ -y \\ -\frac{x_1(t)-x_2(t)}{d(t)} B_1^3(d+1) - B_1^3(d) \\ -\frac{x_1(t)-x_2(t)}{d(t)} B_2^3(d+1) - B_2^3(d) \\ -\frac{x_1(t)-x_2(t)}{d(t)} B_3^3(d+1) - B_3^3(d) \\ -\frac{x_1(t)-x_2(t)}{d(t)} B_4^3(d+1) - B_4^3(d) \\ -\frac{x_1(t)-x_2(t)}{d(t)} B_5^3(d+1) - B_5^3(d) \end{pmatrix}^T + \sigma\epsilon_{x,t}$$

The directions and sharks can be stacked (and parameter vector and design matrix augmented) so that all the unknown parameters can be fit in a single generalized least squares fit. For simplicity the models are written in their differential form, but are all fit using the Euler approximation.

The models were compared using a likelihood ratio test. The results are shown in Table 8.1. The model with only a position potential was rejected at the 1% level.

	model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
Position only	2	14.00	6152.86	6212.94	-3062.43			
Position & distance	1	24.00	6145.17	6248.17	-3048.59	1 vs 2	27.68	0.002

Table 8.1: Likelihood ratio test comparing the model with position potentials for shark 80181 and 80169 with the model with position potentials and a distance to the other shark potential.

8.1.2 Do the sharks have the same potential?

Consider some simplified models where the sharks have the same potential for either position or distance to the other shark.

Model 3 $H = G$ Same position potential

$$d\mathbf{r}_1(t) = -\nabla H(\mathbf{r}_1(t))dt - \nabla V(d_{1,2}(t))dt + \mathbf{\Sigma}_1 d\mathbf{B}_1(t)$$
$$d\mathbf{r}_2(t) = -\nabla H(\mathbf{r}_2(t))dt - \nabla U(d_{2,1}(t))dt + \mathbf{\Sigma}_2 d\mathbf{B}_2(t)$$

Model 4 $V = U$ Same distance potential

$$d\mathbf{r}_1(t) = -\nabla H(\mathbf{r}_1(t))dt - \nabla V(d_{1,2}(t))dt + \mathbf{\Sigma}_1 d\mathbf{B}_1(t)$$
$$d\mathbf{r}_2(t) = -\nabla G(\mathbf{r}_2(t))dt - \nabla V(d_{2,1}(t))dt + \mathbf{\Sigma}_2 d\mathbf{B}_2(t)$$

Both of these models are nested within Model 1 and were compared to Model 1 using the likelihood ratio test(Table 8.2). The results suggested different distance potentials are needed for each shark ($p < 0.001$) but there was not enough evidence to reject the hypothesis they have the same position potential ($p = 0.15$). I concluded that model 3 is the most appropriate

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
	1	24.00	6145.17	6248.17	-3048.59			
$H = G$	3	19.00	6143.30	6224.84	-3052.65	1 vs 3	8.13	0.15
$V = U$	4	19.00	6159.27	6240.81	-3060.64	1 vs 4	24.10	< 0.001

Table 8.2: Likelihood ratio test for simplified models with either the sharks with the same position potential (Model 3) or the same distance to other shark potential (Model 4).

and explored it further. The estimated model parameters for the common quadratic potential and variance parameters are displayed in Figure 8.4.

The B-spline fits for the distance potentials are most easily interpreted by examining their fitted forms as shown in Figure 8.5. The plot suggests that beyond about 800km, shark 80169 (red) is attracted to shark 80181 and shark 80181 is repelled from shark 80169. Extreme care needs to be taken in interpreting this. The bulk of the observed distances are less than 800km and these patterns seem to be driven primarily by the outliers beyond 800km. The zoomed in portion of the figure shows both sharks have relatively flat potentials below 800km (and could indeed be flat within the 95% pointwise error bars shown). The common quadratic potential is also shown in Figure 8.4. The quite planar nature of the estimate also suggests the fitted distance potential might be modelling some inadequacies in the position potential.

Parameter	Estimate	95% CI
b_1	102.4	(52.7, 152.1)
b_2	−56.0	(−101.1,−10.9)
b_3	0.06	(0.03, 0.10)
b_4	0.03	(0.01, 0.06)
b_5	0.00	(−0.02, 0.01)
S_y	3.19	(2.69, 3.79)
T_x	1.76	(1.48, 2.09)
T_y	3.11	(2.62, 3.69)
σ	33.8	(29.9, 38.1)

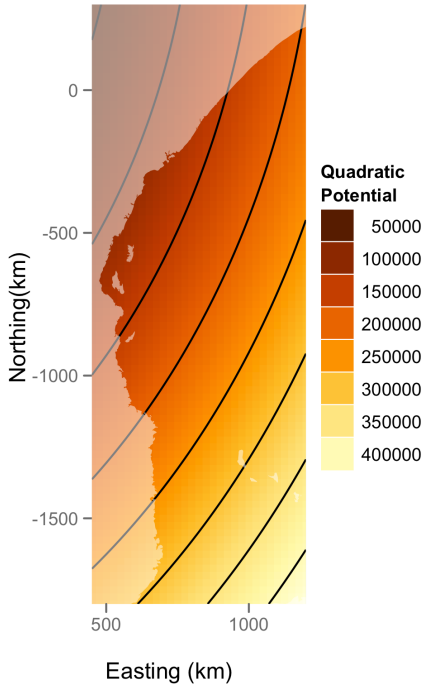


Figure 8.4: Estimates and 95% confidence intervals for the parameters of the common quadratic potential for the two sharks (Model 3).

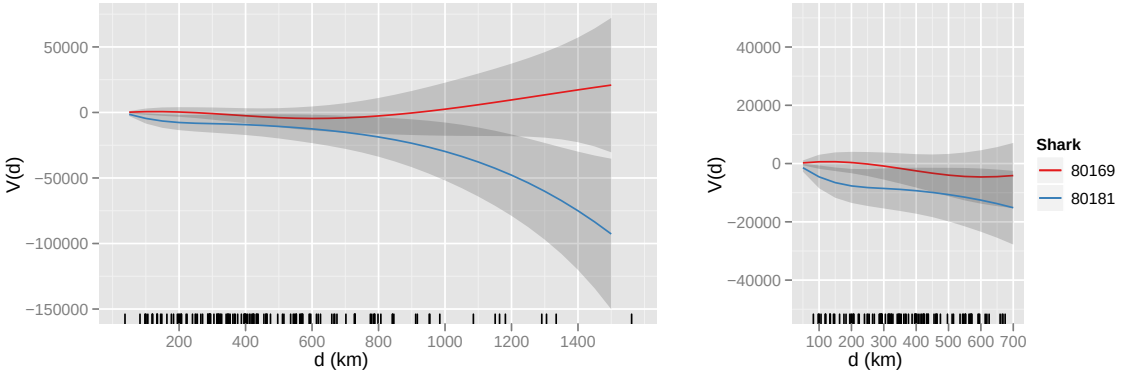


Figure 8.5: Estimated distance to other shark (d) potentials for the two sharks. Approximate 95% pointwise confidence bands are shown in grey. A zoomed in section is shown on the right. Actual observed distances between the sharks are shown in the rug along the bottom.

8.2 Potential of all sharks

Smooth potential estimates for all sharks in the Kenya tagging study were calculated using the methodology in Section 4.2 and are displayed in Figures 8.6 and 8.7. The first figure contains estimates from the sharks with less than 40 observations of their location. With so few points it is hard to draw any solid conclusions. The second figure contains the remaining sharks. Sharks 9156, 66003, 80169, 80179 and 80181 all exhibit an area of attraction off the coast of Kenya. The question arises: do these sharks share a common potential? This question is answered is this section using quadratic potentials.

Consider two models: one where each shark has its own parameters for the a quadratic potential and one where they all have the same quadratic potential. Let $j = 1, \ldots, 5$ index the sharks of interest. For shark j, the movement model is,

$$\begin{pmatrix} x_j(t_{j,i+1}) \\ y_j(t_{j,i+1}) \end{pmatrix} = \begin{pmatrix} x_j(t_{j,i}) \\ y_j(t_{j,i}) \end{pmatrix} - (t_{j,i+1} - t_{j,i}) \begin{pmatrix} \beta_{10,j} + \beta_{11,j} y_j(t_{j,i}) + \beta_{20,j} x_j(t_{j,i}) \\ \beta_{01,j} + \beta_{11,j} x_j(t_{j,i}) + \beta_{02,j} y_j(t_{j,i}) \end{pmatrix} + \Sigma_j \sqrt{t_{j,i+1} - t_{j,i}} \mathbf{Z}_{j,i+1}$$

where

$$\Sigma_j = \begin{pmatrix} \sigma_j^2 & 0 \\ 0 & S_{jy}^2 \sigma_j^2 \end{pmatrix}$$

The extra subscript, j, on the time points is required as different sharks may not be observed at the same times. The hypothesis of interest is whether the sharks share a common potential, that is,

$$H_0 : \quad (\beta_{10,j}, \beta_{01,j}, \beta_{11j}, \beta_{20j}, \beta_{02j}) = (\beta_{10}, \beta_{01}, \beta_{11}, \beta_{20}, \beta_{02}) \quad \text{for all } j.$$

Error structure

Three different models for the error structure were considered: one where the sharks share the same error variance parameter, one where each shark has its own variance parameter but they share the scale factor for the north-south direction, and one where they sharks all have different diagonal variance matrices. The models were parameterised as multiples of the variance in the east-west direction of the first shark (9156). That is,

$$\Sigma_j = \Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & T_y^2 \sigma^2 \end{pmatrix}, \forall j \qquad \text{Model 1}$$

$$\Sigma_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & T_y^2 \sigma^2 \end{pmatrix}, \quad \Sigma_j = \begin{pmatrix} S_j^2 \sigma^2 & 0 \\ 0 & S_j^2 T_y^2 \sigma^2 \end{pmatrix} \quad \text{for } j > 1 \qquad \text{Model 2}$$

$$\Sigma_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & T_{1y}^2 \sigma^2 \end{pmatrix}, \quad \Sigma_j = \begin{pmatrix} T_{jx}^2 \sigma^2 & 0 \\ 0 & T_{jy}^2 \sigma^2 \end{pmatrix} \quad \text{for } j > 1 \qquad \text{Model 3}$$

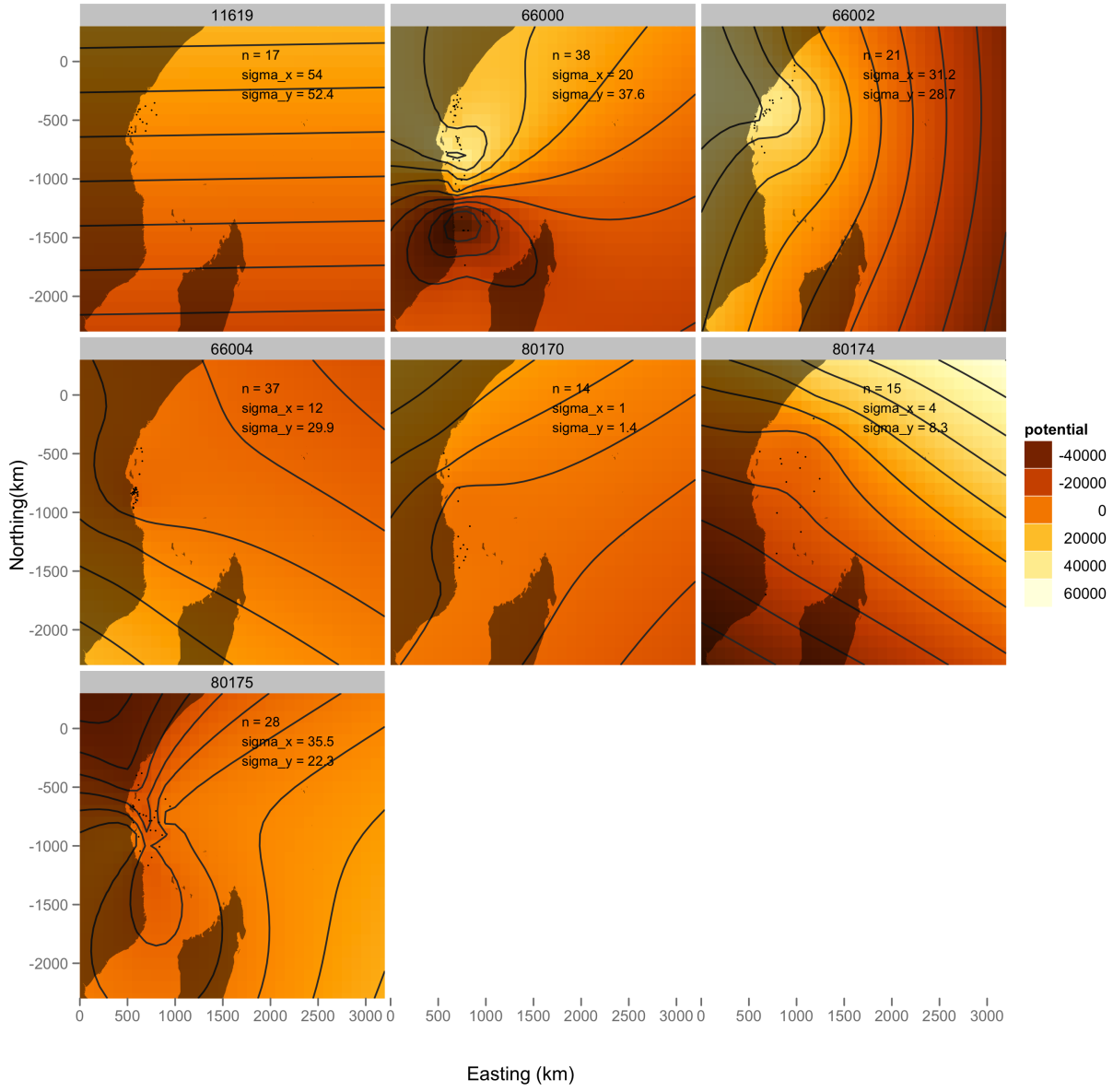


Figure 8.6: Smooth potential function estimates for the sharks in the Kenya study with less than 40 locations. The number of locations, n, and the estimated standard deviation in each direction, σ_x and σ_y, are noted on each shark's figure.

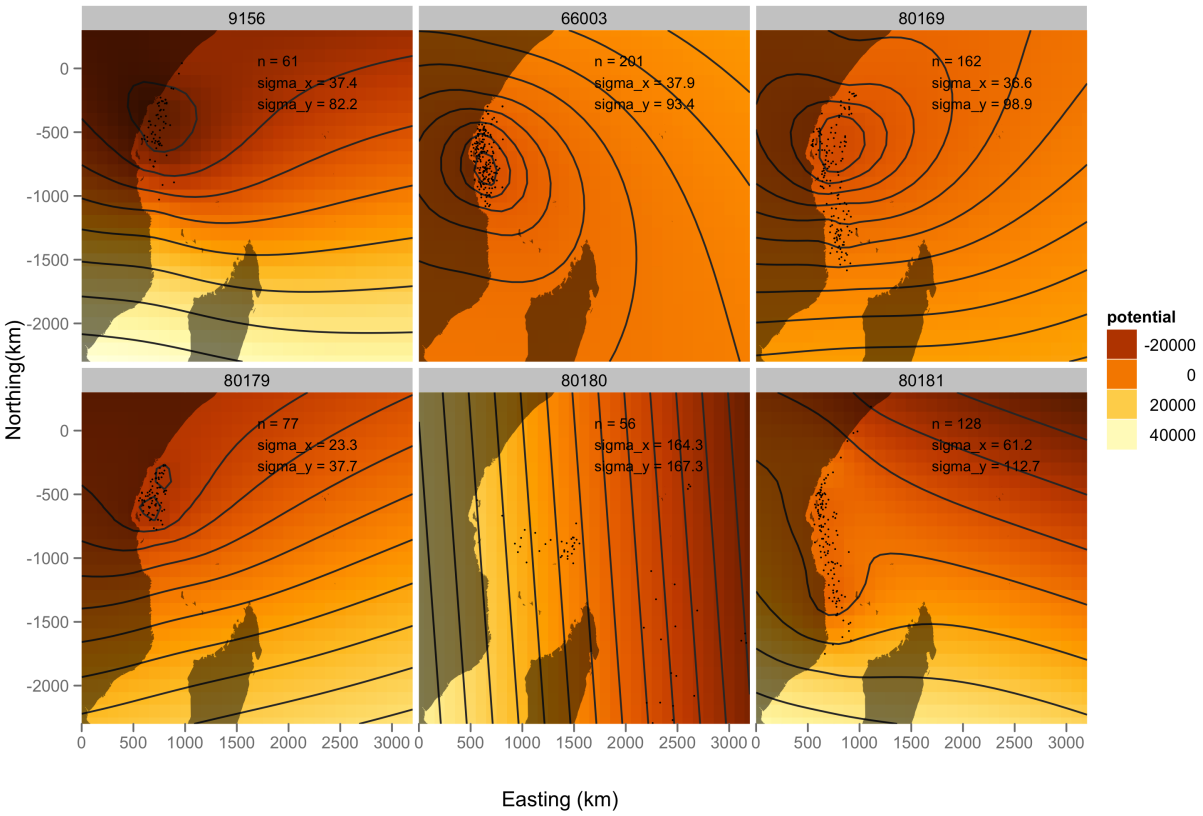


Figure 8.7: Smooth potential function estimates for the sharks in the Kenya study with more than 40 locations. The number of locations, n, and the estimated standard deviation in each direction, σ_x and σ_y, are noted on each shark's figure.

All three error models were fit allowing each shark to have its own set of potential function parameters using restricted maximum likelihood. Since the models are nested they can be compared using the likelihood ratio test, under the assumptions set out in Chapter 3. The results of this test are shown in Table 8.3. The simpler structures are rejected and Model 3 is retained.

Table 8.3: Likelihood ratio test of the three nested error structures for the five sharks.

Model	df	AIC	BIC	log(Likelihood)	Test	Likelihood Ratio	p-value
1	27	13694.21	13832.15	-6820.11			
2	31	13622.73	13781.11	-6780.37	1 vs 2	79.48	< 0.0001
3	35	13603.56	13782.37	-6766.78	2 vs 3	27.17	< 0.0001

With the error structure chosen the hypothesis of the sharks sharing a common quadratic potential could be answered. Two models with the error structure defined above were fit: one with a separate set of potential function parameters for each shark, and one with a common set of potential function parameters for all five sharks. Again the models can be compared using the likelihood ratio test, Table 8.4. The null model with common potential parameters was rejected. Figure 8.8 explores the differences in the potential function parameters.

Table 8.4: Likelihood ratio test of a model with common potential parameters against a model with separate potential parameters for each shark.

Model	df	AIC	BIC	log(Likelihood)	Test	Likelihood Ratio	p-value
Common β	15	13551.94	13628.88	-6760.97			
Different β	35	13548.88	13728.40	-6739.44	1 vs 2	43.06	< 0.0001

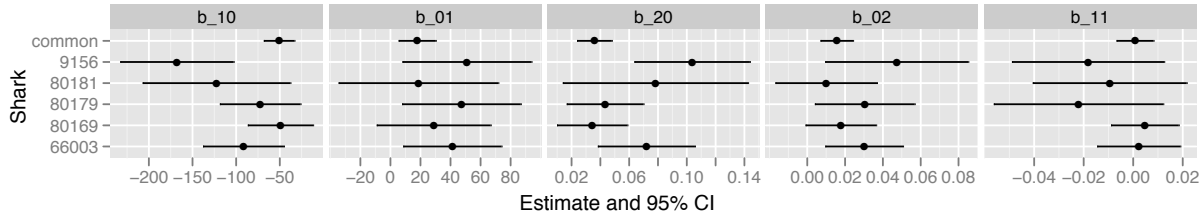


Figure 8.8: Estimates of the parameters for different quadratic potential functions for five sharks, along with a pooled estimate.

8.3 Discussion

This chapter presented two extensions to the potential function approach when several animals are involved. The interaction between two animals was modelled with a potential depending on the distance between the animals. The methods presented are quite general and should prove useful with other data. The position potentials used here were quite structured but with more data one could imagine using a smooth non-parametric potential instead. The spline used for the distance potential allows unstructured examination of the relationship between the sharks movements and their distance apart.

The method presented is directly extended to more animals, the only consideration being how to relate the movement of one animal to its distance from many. Obvious choices include only depending on the distance to the closest (or furthest) animal, or a average distance from all animals.

The potential models can be fit to multiple animals allowing for the sharks to share both potential and error structure parameters. For five sharks in the Kenya study the hypothesis of a common quadratic potential was rejected.

The models were compared here using the likelihood ratio test which depends on large sample approximations and the result from Lai & Wei (see Section 3.5). Ideally more robust tests would be used. This could be done by using simulations of tracks from the null hypotheses to obtain a more accurate null distribution for the likelihood ratio statistic.

Chapter 9

Another application: Eye movements

This chapter presents an application of the methods in this thesis to a different type of trajectory: eye movements. The trajectory is the path traced out by an image on the retina. This path corresponds directly to the movements made by the eye if the head remains in a fixed position relative to the target image.

The movements of interest here are fixational eye movements, the tiny ones made by the eye when the subject fixates on a small target. The primary goal is to characterise these movements and compare them between different eyes. In particular, comparing normal eyes to amblyopic eyes (commonly known as "lazy" eyes). For vision scientists a characterisation of the movements is an initial step in exploring possibilities for improving the vision of amblyopes.

The chapter begins with some background on fixational eye movements and amblyopia. The trajectories are decomposed empirically into two components based on speed. The two components are characterised using the potential function method to gain insight into their functions.

The data come from an experiment conducted by Shuang Song and is also discussed and analyzed in her thesis (Song, 2010) and reported in Song et al. (2010).

9.1 Background

A schematic of the eye is shown in Figure 9.1. The retina and fovea will be of most interest in describing eye movements. The retina (the yellow surface in the diagram) is a complex layered surface containing photoreceptor cells that respond to light. The optics of the eye focus an image upon the retina and the signals from the photoreceptor cells undergo complex processing by other cells in the retina before being transmitted to the brain. The fovea (the depression in the retina in the diagram) is the part of the retina with the greatest density of photoreceptor cells. In a normal eye the most detailed vision occurs for images falling upon the fovea. For this reason an image is generally kept on, or near, the fovea during fixation.

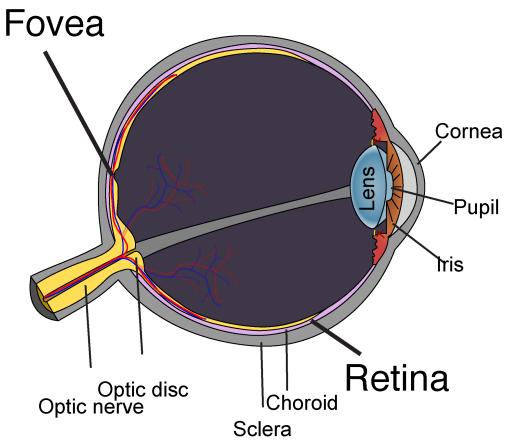


Figure 9.1: A schematic of the eye. Adapted from http://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_en.svg

9.1.1 Fixational eye movements

Fixational eye movements describe the movements of the eye when attempting to keep an image stable with respect to the retina. For normal eyes this generally means keeping the image stable and on the fovea. Even though the image may appear stable to the subject the eye is making small involuntary movements. These movements are imperceptible to the subject in most circumstances but can be illuminated via certain optical illusions (for example, see Figure 1 in Rolfs (2009)). If an image is completely still on the retina our vision fades away as a result of neural adaptation. It is generally agreed (Rolfs (2009), Martinez-Conde et al. (2004)) that fixational eye movements are designed to prevent this from happening when fixating on a point. The movement is generally decomposed into three components: microsaccades, drift and tremor. **Microsaccades** are large jerk like movements that occur on the order of a second apart. They last about 25 milliseconds and can travel over dozens to hundreds of photoreceptors in distance. **Drift** occurs between microsaccades and is characterized by a slow wander. **Tremor** is the smallest component overlaid upon drift with amplitude around the diameter of a foveal photoreceptor cell. It is a wave like motion with frequency around 90Hz (Martinez-Conde et al., 2004). The system used to measure the movements in the experiment analysed here is not sensitive enough to distinguish the tremor from measurement noise.

There has been much research into the respective roles of the three components in maintaining fixation. Rolfs (2009) provides a thorough review and concludes,

> "...a contribution of microsaccades is not a necessity for accurate fixation when observers try to hold their eyes as still as possible, however, they are clearly a significant source of position correction during visual fixation. Based on several lines of evidence I argue that microsaccades and drift serve the same behavioral goal, potentially implemented by a common neurophysiological machinery."

The focus in this chapter will not be to enter in to the debate on function but to demonstrate a potential function model provides a useful description for the fixational movements.

9.1.2 Amblyopia

Amblyopia is a condition resulting is impaired vision. Commonly known as "lazy" eye it may be better termed "lazy" brain. Generally a physical abnormality causes the brain to prefer one eye over the other. The normal neural development is interrupted and vision remains poor even if the physical abnormality is corrected.

Here the amblyopic subject has strabismus which results in eccentric fixation: the center of fixation is not at the fovea.

9.2 The experiment

In the experiment performed by Shuang Song the fixational eye movements are measured using a Adaptive Optics Laser Scanning Opthalmoscope (AOLSO) (Roorda et al., 2002), a method that can obtain an image of the retina detailed enough to distinguish individual photoreceptor cells. This results in a movie of the moving retina (relative to the target). An example of one frame of a movie is shown in Figure 9.2. A single frame can be thought of as pixel intensity, $I(\mathbf{r})$, (for example $I(\mathbf{r}) = 0$ would be a black pixel, $I(\mathbf{r}) = 1$ a white pixel) as a function of the pixel position, $\mathbf{r} = (x, y)$. The movie provides a time series of these functions, $I(\mathbf{r}, t)$, where t indexes the movie frame. The frame rate of the movie is 30 frames per second. So, the observed positions are equispaced in time, $t = i/30$, $i = 1, \ldots, n$. The movie is transformed into a series of translations by aligning each frame with the previous one using their correlation. For example, the movement from the frame at time t and the frame at time $t + 1$ is

$$\arg\max_{\delta} \mathrm{Cor}(I(\mathbf{r}, t), I(\mathbf{r} + \boldsymbol{\delta}, t + 1).)$$

Translations resulting in low correlations are also checked manually and flagged if the estimate of translation isn't considered reasonable.

In this experiment there are two subjects and a total of three eyes. One eye from a subject with normal vision, and both eyes from a subject with amblyopia (one amblyopic eye and one nonamblyopic eye). For the amblyopic patient, the left eye is the amblyopic eye and the right eye is the nonamblyopic eye. For the normal subject the data are from his left eye. In a single experimental run the subject continuously fixates on a target for a set amount of time. The target and time vary throughout the experiment. In particular, the amblyope needs larger targets. The total number of runs is 76 for the normal eye, 28 for the nonamblyopic eye and 199 for the amblyopic eye.

For the normal and nonamblyopic eye a reference frame is created for each experiment. The fovea provides a common reference point for the separate experiments on the same eye

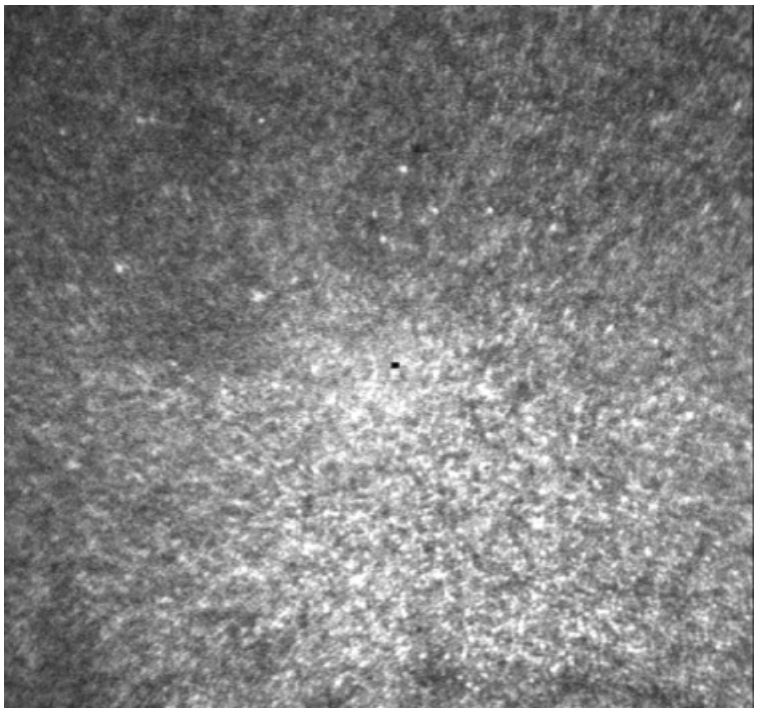


Figure 9.2: One frame from the movie captured by AOLSO. The target is the black square in the center of the frame.

at $\mathbf{r} = (0, 0)$. For the amblyopic eye one reference frame is created using all the experiments and they are directly comparable. For ease of comparison between the eyes an empirical center of fixation in defined as the median fixation location. This is subtracted from the position for the set of experiments for each eye. For the normal eye this is (16, 36) pixels, for the nonamblyopic eye (9, -14) and the amblyopic eye (1372, 932). The amblyopic center of fixation is very far from the fovea as a result of their strabismus.

The locations in pixels and converted to locations in degrees using a linear transformation.

9.3 Drifts

9.3.1 Defining drifts empirically

The empirical speeds of the target at each time step are calculated by simple differencing of position,

$$\text{Speed}(t) = \sqrt{\left(\frac{x(t+1) - x(t)}{1/30}\right)^2 + \left(\frac{y(t+1) - y(t)}{1/30}\right)^2}$$

They have the units degrees per second (deg/sec).

The histogram for the empirical speeds is shown in Figure 9.3 for the normal eye. There are two groups seen in the histogram. An empirical cutoff between drifts and microsaccades

is chosen at 3.16 deg/sec corresponding to about a 0.1 degree jump in one time step. A cut off of 2 deg/sec (a 0.06 degree jump) is chosen for the amblyopic eye but due to the larger nature of the movements in the amblyopic eye a larger cutoff of 5.6 deg/sec (a 0.18 degree jump) is chosen.

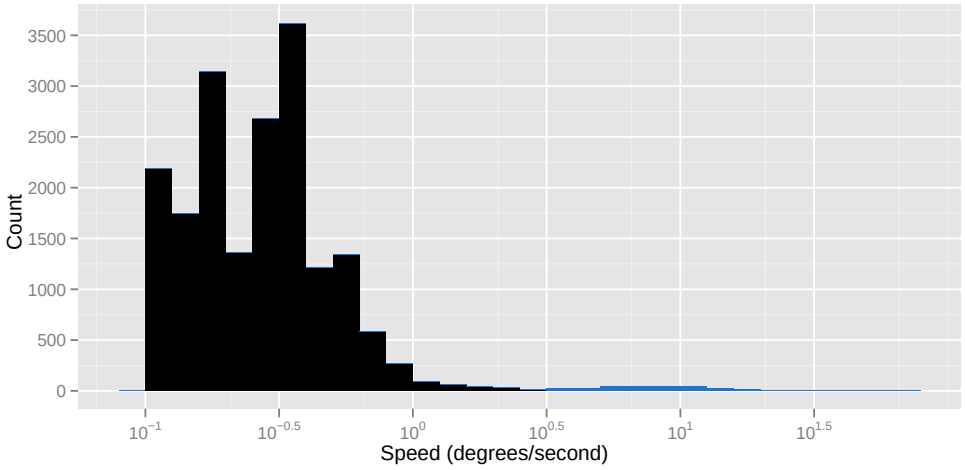


Figure 9.3: Histogram of the empirical speeds (on a log scale) for the normal eye over all experimental runs. The cutoff is picked to be $10^{0.5} = 3.16$ and speeds above this are classified as microsaccades and are coloured blue.

An example of the trajectory generated by one experimental run is shown in Figure 9.4. This trajectory comes from a normal eye and is over a 10 second period. The empirically classified microsaccades are coloured blue. In general we see the target stays near the center of fixation but moves up to 0.3 degrees away from the center. To get an idea of the range of the target on the retina we can use all the positions from the trajectories from each eye to create an empirical density of the target location. This is shown in Figure 9.5 using hexagonal binning for each eye. The target moves about 0.5 degrees away from the center of fixation for the normal eye and about 0.3 degrees for the nonamblyopic eye. The amblyopic eye ranges over a much larger scale: 3 degrees in the horizontal direction and 1.5 degrees in the vertical direction. The density also appears much flatter for the amblyopic eye.

9.3.2 Exploring trends in the drifts

A subdrift is defined as the period of movement started by a microsaccade and ended by either another microsaccade or a missing value. For the normal eye this results in a total of 887 subdrifts, with a median length of 0.9 seconds. All of the subdrifts for the normal eye are examined in Figure 9.6 where the position of the target is plotted against time for the horizontal and vertical directions separately. There is some suggestion that the horizontal position tends to the left (negative direction) and the vertical position tends downwards over

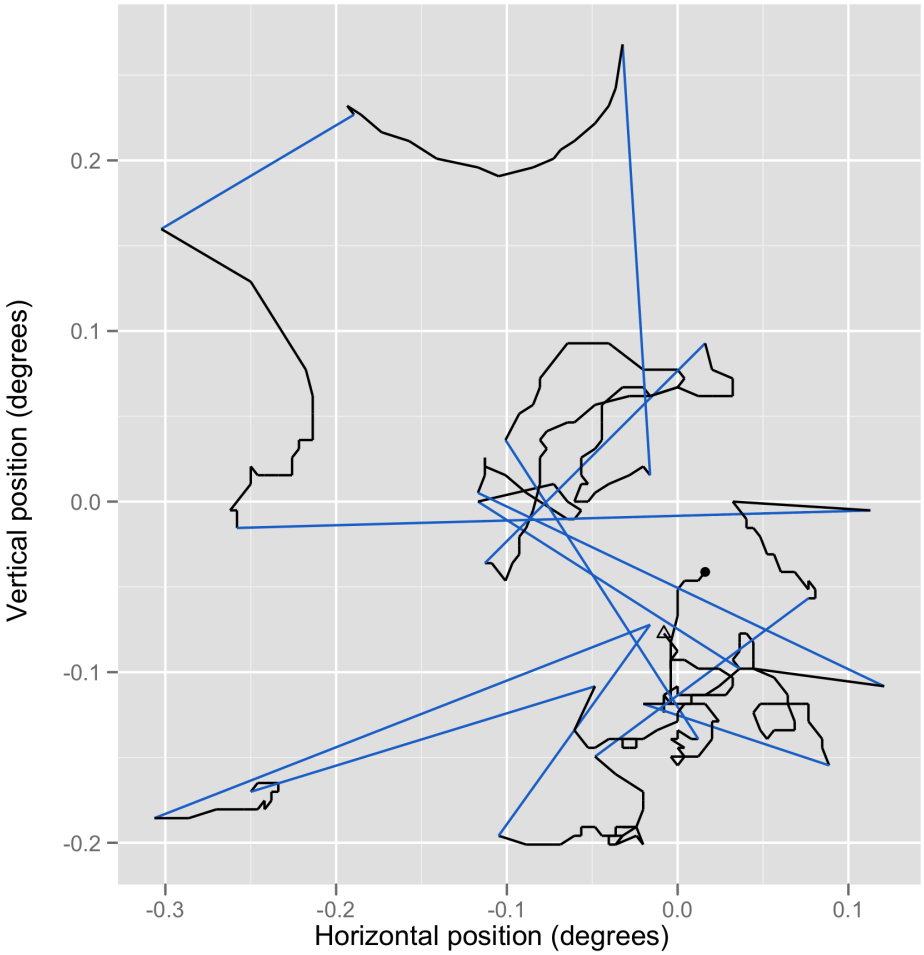


Figure 9.4: An example eye movement trajectory. Each successive position of the traget is joined with a straight line. The empirically classified microsaccades are highlighted in blue

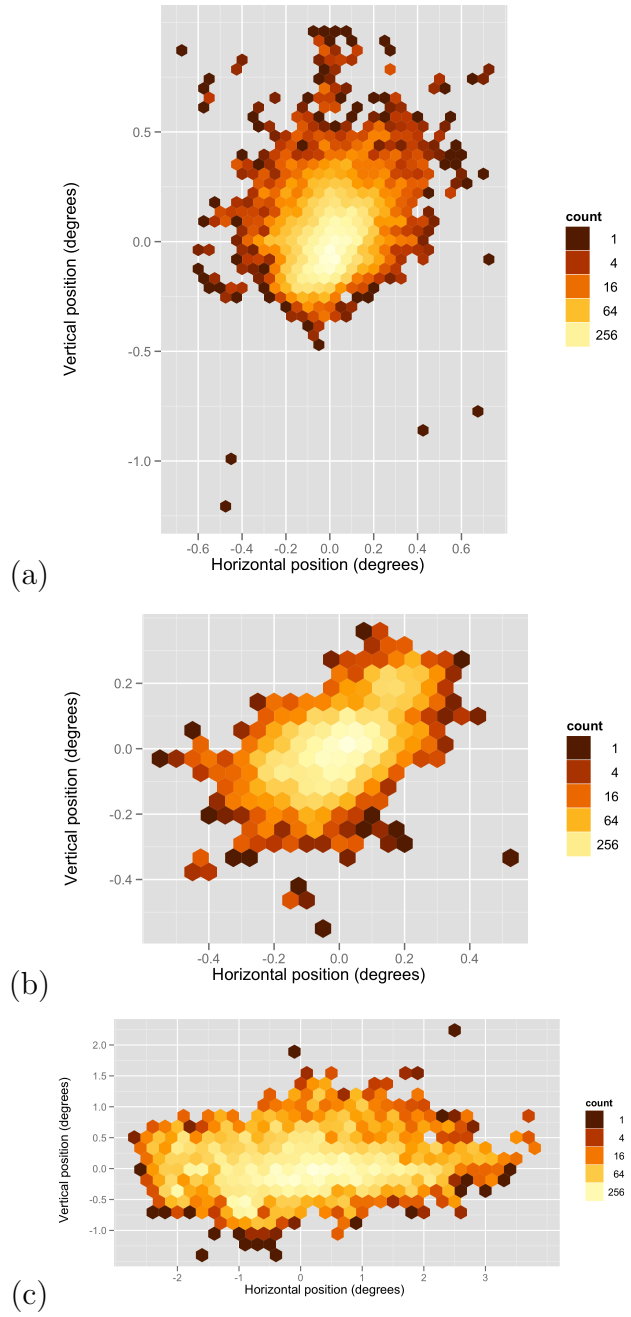


Figure 9.5: Density of locations over all experiments for the normal eye (a), nonamblyopic eye (b) and amblyopic eye (c).

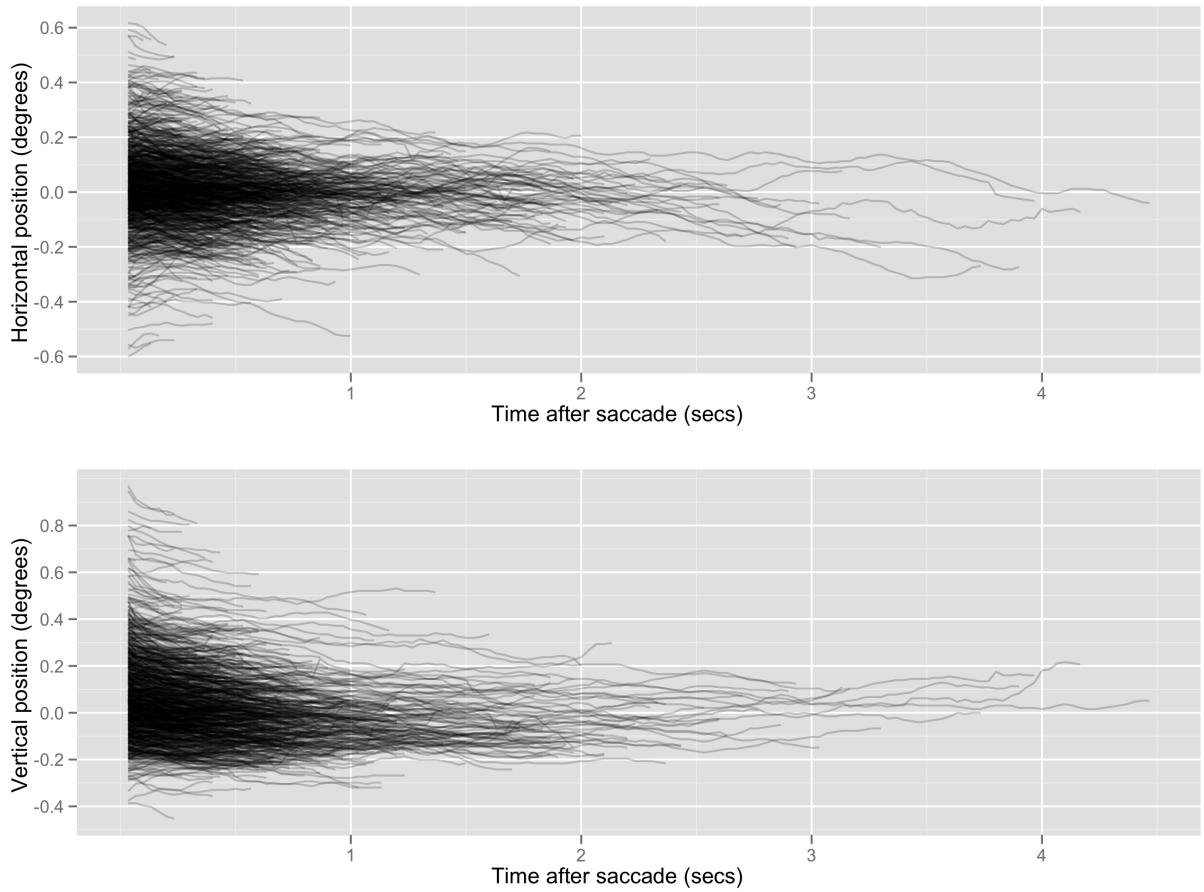


Figure 9.6: Horizontal and vertical position of target as a function of the time since a microsaccade for the normal eye.

time. This pattern can be emphasized by subtracting off the first position of each subdrift from the successive positions as shown in Figure 9.7. The tendency of the subdrifts can

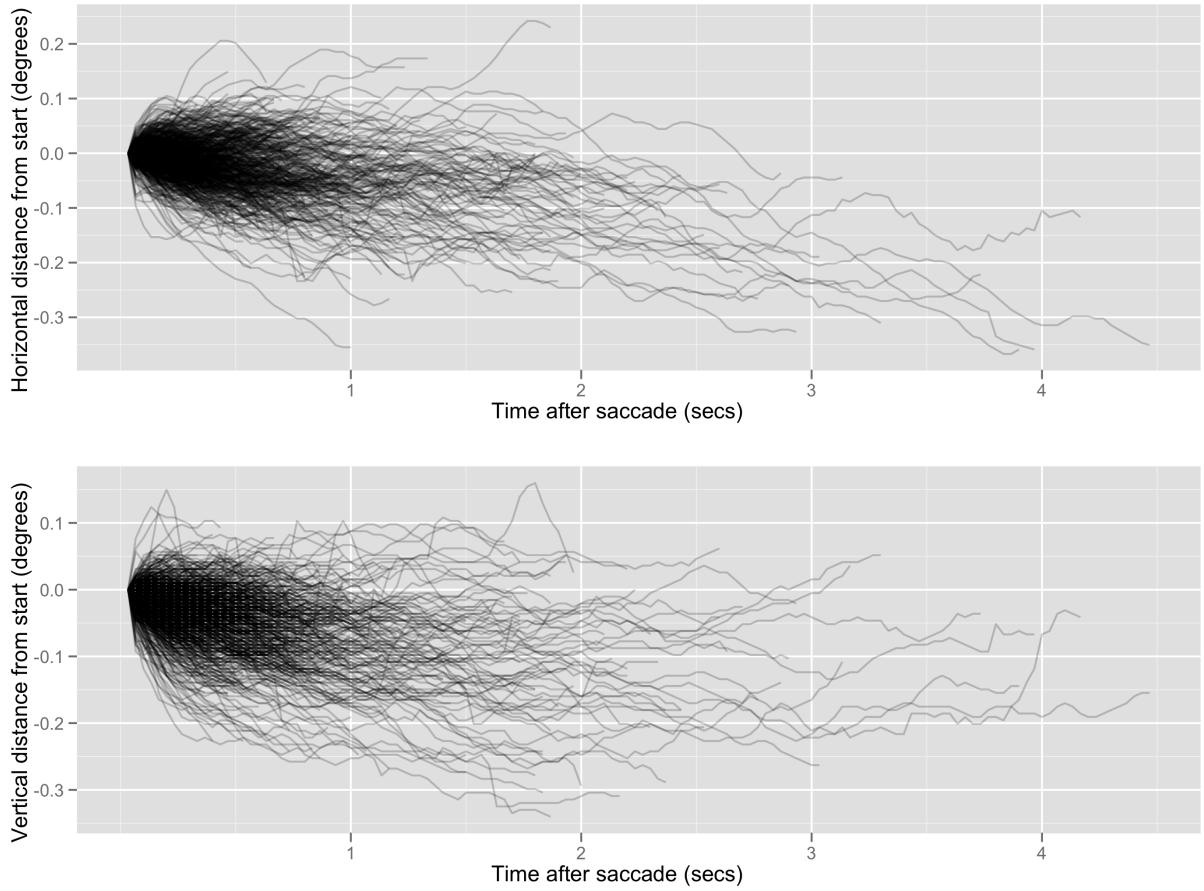


Figure 9.7: Horizontal and vertical positions of target with initial position subtracted for the normal eye.

be summarised in both directions by calculating the average velocity by position. This is shown in Figure 9.8 where the average velocity is calculated within 0.05 degree square bins. We can see a systematic trend to move down and to the left. Since this is the left eye, this corresponds to a temporal drift. It is interesting that there is no visual evidence that the subdrifts tend towards the center of fixation.

This can be repeated for the other two eyes as shown in Figure 9.9. We see similar evidence for a systematic trend although not in the same direction as the normal eyes. The nonamblyopic eye shows an up and to the right trend and the amblyopic an up and to the left trend. Both exhibit temporal (as opposed to nasal) drift.

A smooth non-parametric potential function can be fitted to the subdrift positions using the generalized additive model methodology outlined in Section 4.2. Some care needs to be taken here to avoid overfitting. The 887 subdrifts result in a total of 18838 velocities. With

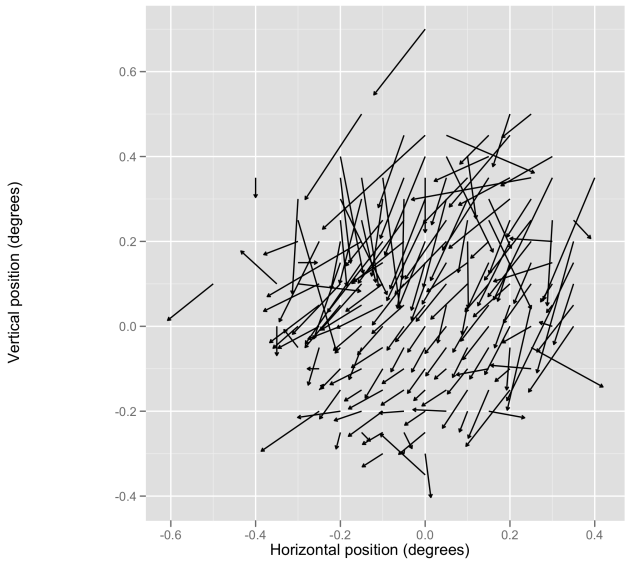


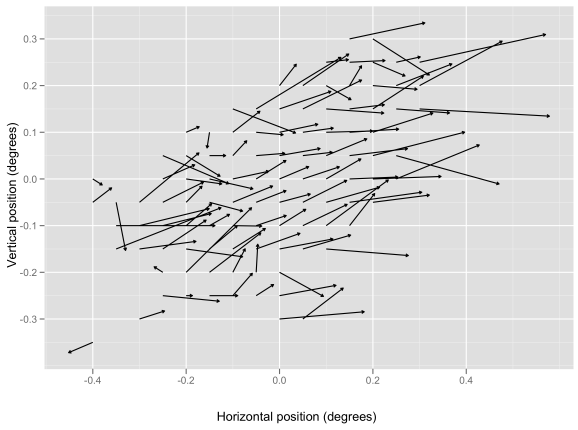
Figure 9.8: Empirical average velocity of subdrifts as a function of position on the retina for the normal eye (left eye). The length of the arrow directly corresponds to the velocity in deg/sec.

this many observations, very complicated surfaces can be shown to be significantly better than simple ones when there is no practical difference. A simple approach to avoid this problem is to fit the surface to 1000 randomly chosen positions. An example fit for the normal eye is shown in Figure 9.10. The subdrifts will flow downhill, from white towards black, confirming the down and to the left drift. The assumptions for this nonparametric smooth are independent constant variance errors. This is found to be unsuitable and remedied in the next section.

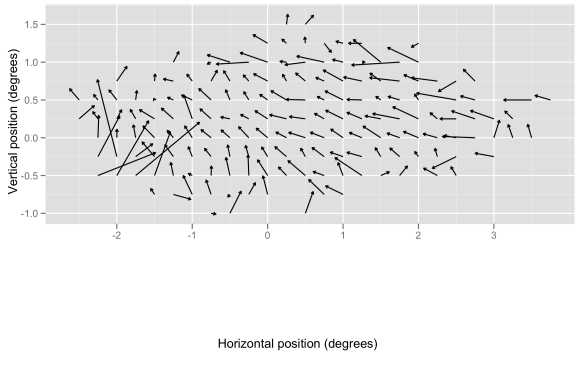
The smooth potentials for the other eyes is shown in Figure 9.11. Both smooths are close to linear, show a slight upward trend and opposite left to right trends.

9.3.3 Modelling the drifts

The subdrifts show some systematic trends. In this section we will use a quadratic model of the potential to describe and compare the trends between the eyes. The quadratic model is chosen as a parsimonious model that will allow for the trends seen in the previous section and has the natural property that if the target is far from the center of fixation it will be rapidly returned (assuming the quadratic is concave with a center somewhere near the center of fixation). It also allows for more complex modelling of the error structure. The errors are modelled as an autoregressive process.



(a)



(b)

Figure 9.9: Empirical average subdrifts as a function of position on the retina for the (a) nonamblyopic (right) and (b) amblyopic (left) eyes. The length of the arrow directly corresponds to the velocity in deg/sec.

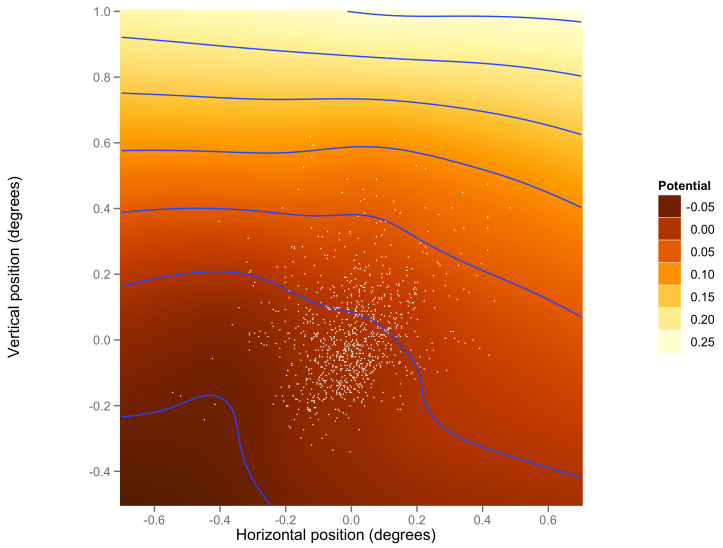


Figure 9.10: A non parametric smooth potential fit to 1000 randomly chosen velocites from the subdrifts of the normal eye. The white points are the positions used for the fit.

9.3.4 Model

The structure in the movement of the target on the retina is modelled by a quadratic potential. That is,

$$\frac{\mathbf{r}(t+1) - \mathbf{r}(t)}{\Delta t} = -\nabla H(\mathbf{r}(t)) + \boldsymbol{\epsilon}_t$$

where

$$H(\mathbf{r}) = H(x, y) = \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy$$

The errors $\boldsymbol{\epsilon}_i$ are modelled as an autoregressive process of order 2 (chosen by examination of the residuals of fit without correlated errors),

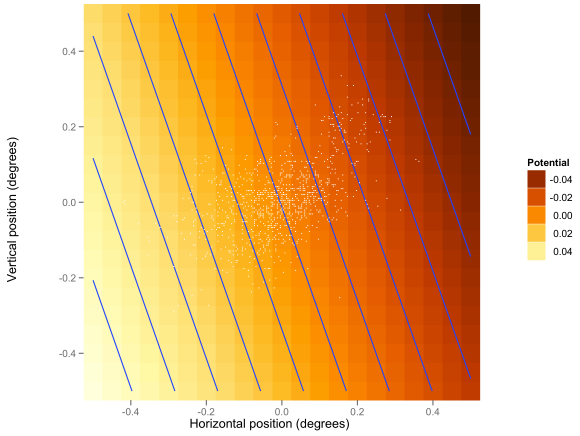
$$\boldsymbol{\epsilon}_t = \phi_1 \boldsymbol{\epsilon}_{t-1} + \phi_2 \boldsymbol{\epsilon}_{t-2} + \boldsymbol{\eta}_t$$

The variance of $\boldsymbol{\eta}$ is allowed to be different in the horizontal and vertical directions (although this is only found to be necessary in the amblyopic eye),

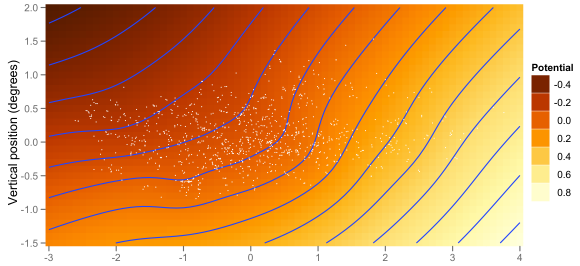
$$\boldsymbol{\eta}_t \sim \text{Normal}\left(\mathbf{0}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & S_y^2 \sigma^2 \end{pmatrix}\right).$$

9.3.5 Model results

The model is fit to each eye separately. The parameter estimates are shown in Table 9.1. The potential parameters provide a useful summary of the trends in the subdrifts for each eye but are most usefully interpreted by a image plot of the estimated potential. These



(a)



(b)

Horizontal position (degrees)

Figure 9.11: Fitted smooth potential models for the a) nonamblyopic and b) amblyopic eyes.

Potential parameters	x	y	x^2	y^2	xy
Normal	-0.052 (0.004)	-0.085 (0.004)	-0.12 (0.01)	-0.27 (0.01)	0.04 (0.02)
Nonamblyopic	0.087 (0.003)	0.029 (0.003)	-0.03 (0.01)	-0.15 (0.02)	0.08 (0.02)
Amblyopic	-0.128 (0.014)	0.122 (0.012)	-0.06 (0.01)	-0.20 (0.01)	-0.07 (0.01)
Error parameters	ϕ_1	ϕ_2	S_y	σ	
Normal	0.37	0.21	-	0.26	
Nonamblyopic	0.27	0.23	-	0.22	
Amblyopic	0.30	0.47	0.80	0.81	

Table 9.1: The parameter estimates for the model fitted to each eye. Only the model for the amblyopic eye uses separate error variances in each direction.

are shown in Figure 9.12. All three eyes have quite large parameters for the autoregressive errors. This breaks the assumptions required for the subdrifts to be Brownian motions with drift. However, the model as written above is still a useful representation of the movement. The standard deviations of the errors are comparable for the normal and nonamblyopic eye. The standard deviation of the errors in the amblyopic eye is about 3.5 times larger than the other eyes and there is evidence for different standard deviations in the horizontal and vertical directions. The standard deviation in the vertical direction is estimated to be 80% of that in the horizontal direction.

9.3.6 Checking the model

Some subdrift trajectories are simulated from the model using 100 real subdrift lengths and starting positions. For normal eyes the horizontal positions for the simulated and real data are shown in Figure 9.13. The simulated data emulate a number of the properties of the real subdrifts. The downward trend is well modelled, the quadratic terms ($\beta_3 x^2$ and $\beta_4 y^2$) prevent subdrifts near the bottom from drifting further from the center of fixation and the smoothness over time is well modelled. It seems the real subdrifts are a little more variable. The residuals can also be examined to check for inadequacies in the model. The residuals being used here are normalized velocity residuals, where the residuals are transformed to be approximately uncorrelated and have common variance.

Figure 9.14 displays binned velocity residuals for the x and y directions by position of the target. There is no evidence of a spatial pattern which suggests we have captured the major spatial trends with the quadratic potential.

The residuals are plotted against the time after a saccade in Figure 9.15. At the first position following a microsaccade there is some suggestion that the residuals are higher than expected in the horizontal direction and lower than expected in the vertical direction. Capturing this using time varying potentials is an avenue for future work.

Figure 9.16 shows normal quantile-quantile plots for the residuals in both directions. The shape suggests the residuals have much heavier tails compared to a normal distribution. This implies caution needs to be exercised in interpreting the estimates of the standard errors on the coefficients. Quantile-quantile plots of the residuals compared to a t-distribution with four degrees of freedom are shown in Figure 9.17. Apart from a few outliers the residuals form a relatively straight line suggesting the t-distribution is a more suitable description of the residuals. Future work will involve using this observation to perform robust methods of estimation.

A comparison of the horizontal to vertical residuals provides no evidence of correlation.

9.3.7 Summary of subdrift analysis

The quadratic potential has proved a useful model to compare the systematic trends in the subdrifts. All of the eyes demonstrated an attraction to a point away from the center of

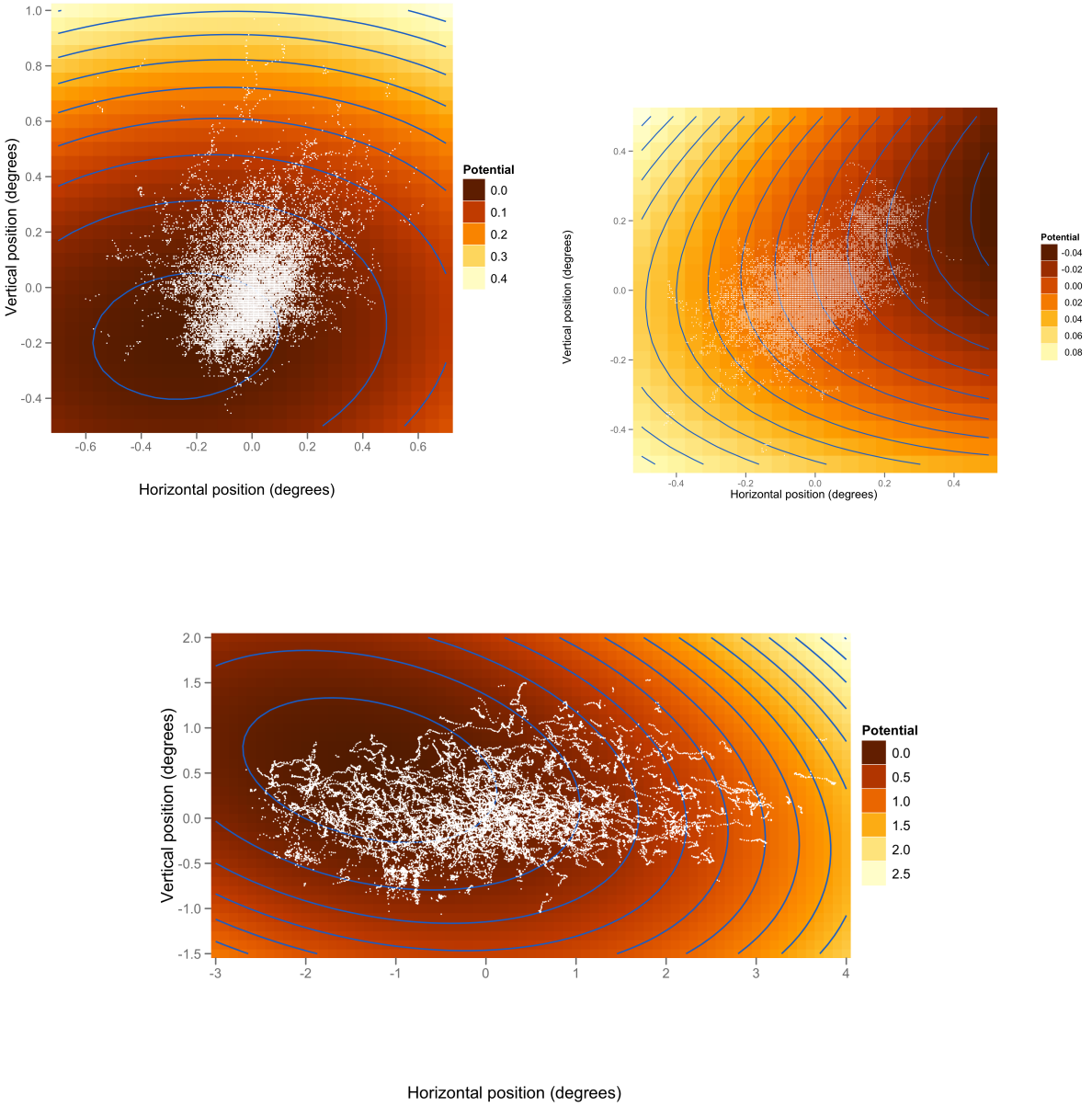


Figure 9.12: Estimated quadratic potentials for the subdrifts in the normal eye (top left), the nonamblyopic eye (top right) and the amblyopic eye (bottom), with actual target locations added in white.

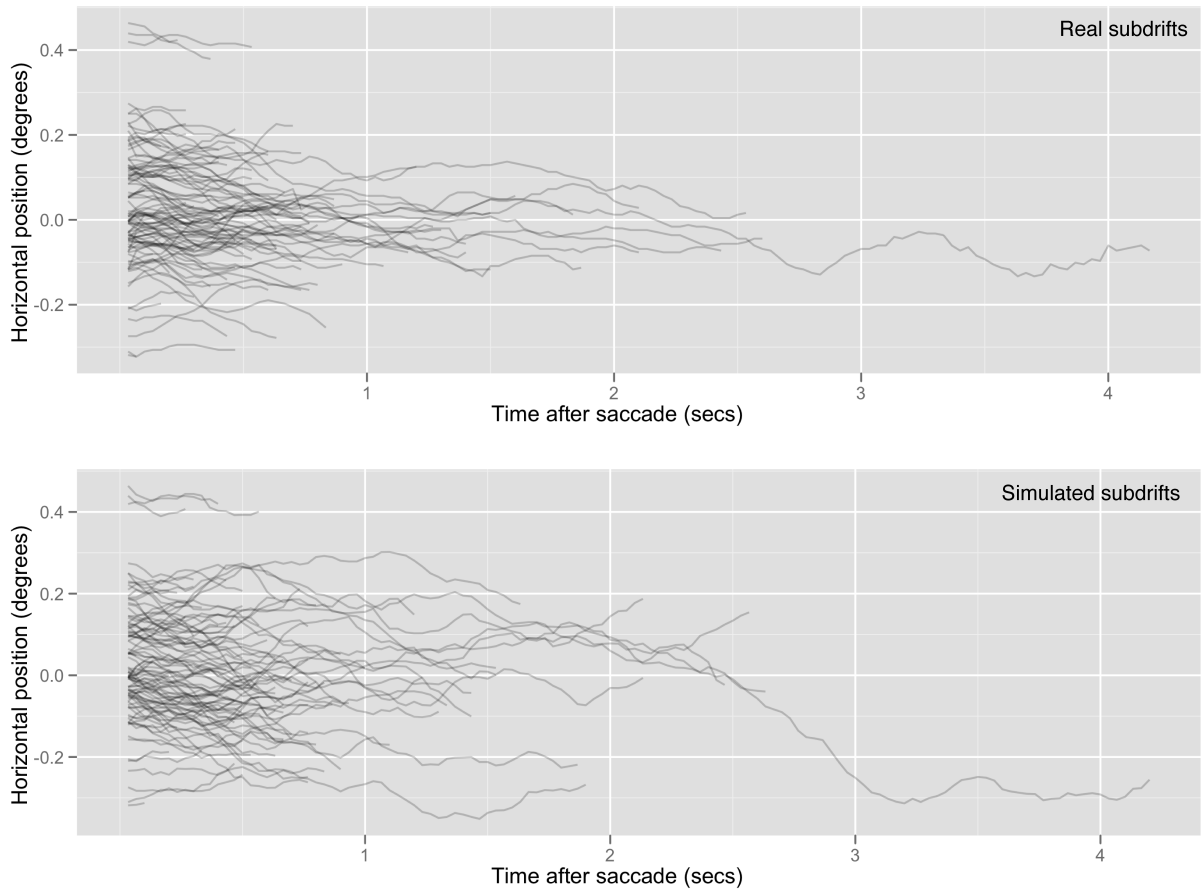


Figure 9.13: Horizontal positions of target of 100 real subdrifts (top) and 100 simulated subdrifts (bottom) using the quadratic potential with correlated errors.

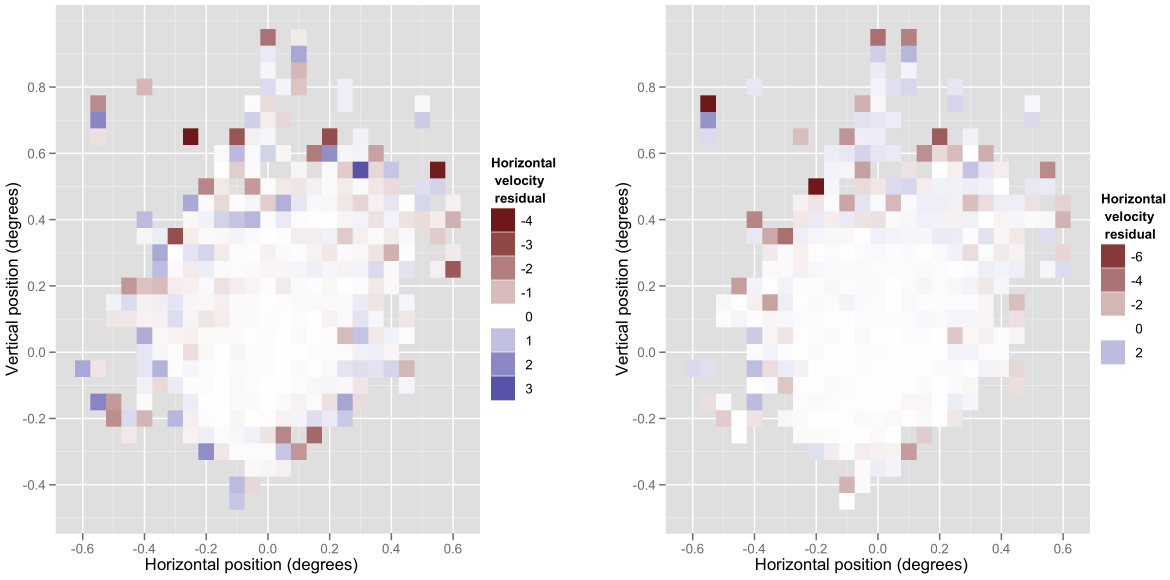


Figure 9.14: Binned horizontal and vertical velocity residuals. No obvious spatial pattern.

fixation which suggests drifts alone would not maintain fixation. The errors were found to be autocorrelated which may be due to the physical nature of this system; the muscles in the eye cannot instantaneously change velocity. The amblyopic eye demonstrated a similar pattern to the normal and nonamblyopic eyes but the drift and error are on much larger scales.

The eyes all demonstrated a tendency for temporal drift, combined with different vertical drifts. It is not clear whether this is something that changes randomly from person to person, eye to eye or deterministically left eye to right eye. Further experimentation would be required to determine how the trends vary between eyes and subjects.

9.4 Microsaccades

Microsaccades are treated here as a marked point process defined by the non-drift locations. They consist of a time, place of jump and velocity of jump. The time and place are the point, (t, x, y), and the empirical velocity is the mark (V_x, V_y).

There are two questions to be explored: Where do the microsaccades go? When do the microsaccades occur?

9.4.1 Where do they go?

The mircosaccades can also be treated as acting under a potential. Their velocities can be fit to the derivative of a smooth potential. Since there are many fewer microsaccades than

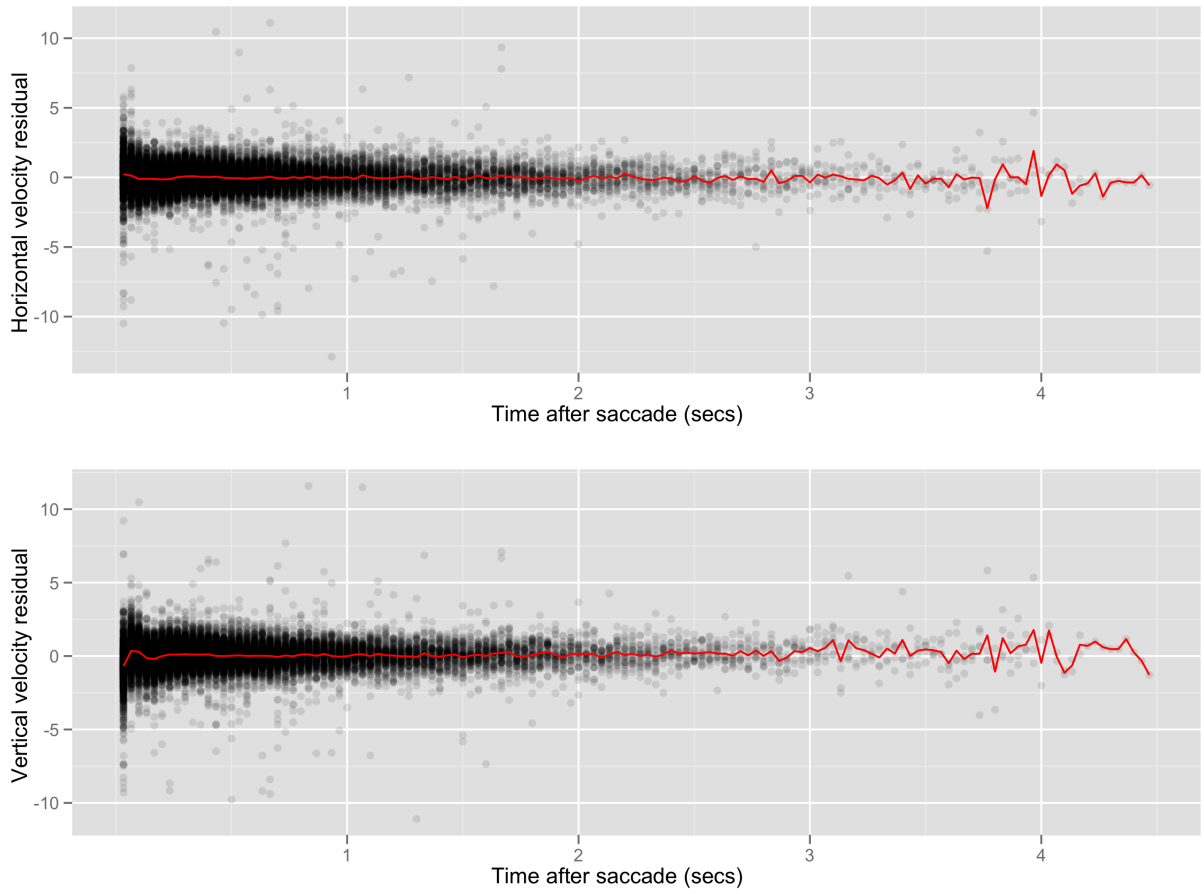


Figure 9.15: Horizontal and vertical velocity residuals against time. Average residual at each time point overlaid in red.

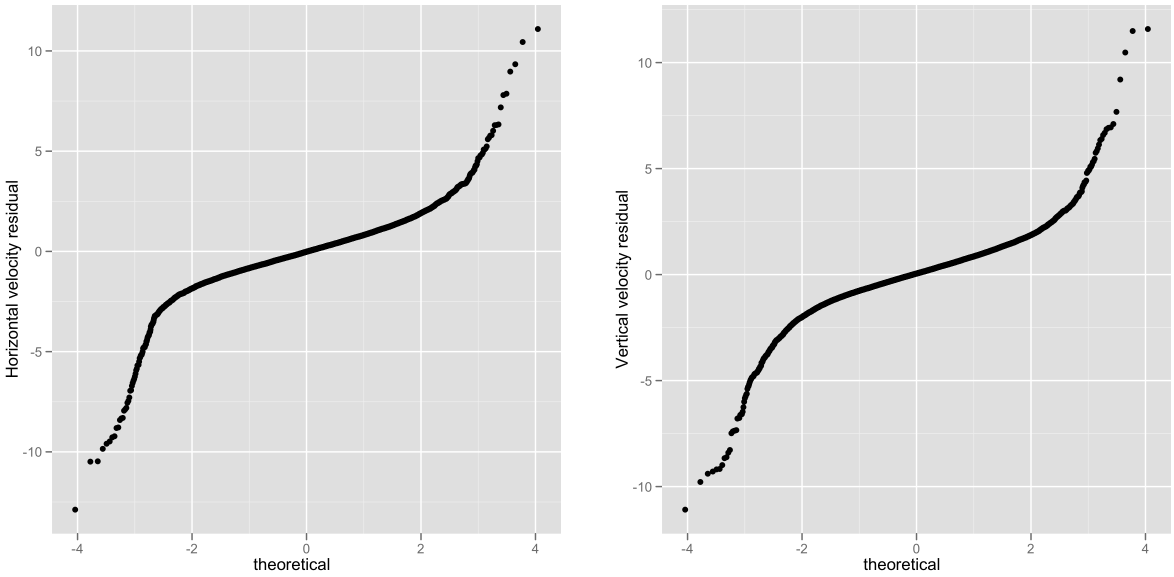


Figure 9.16: Normal quantile-quantile plots for horizontal and vertical velocity residuals.

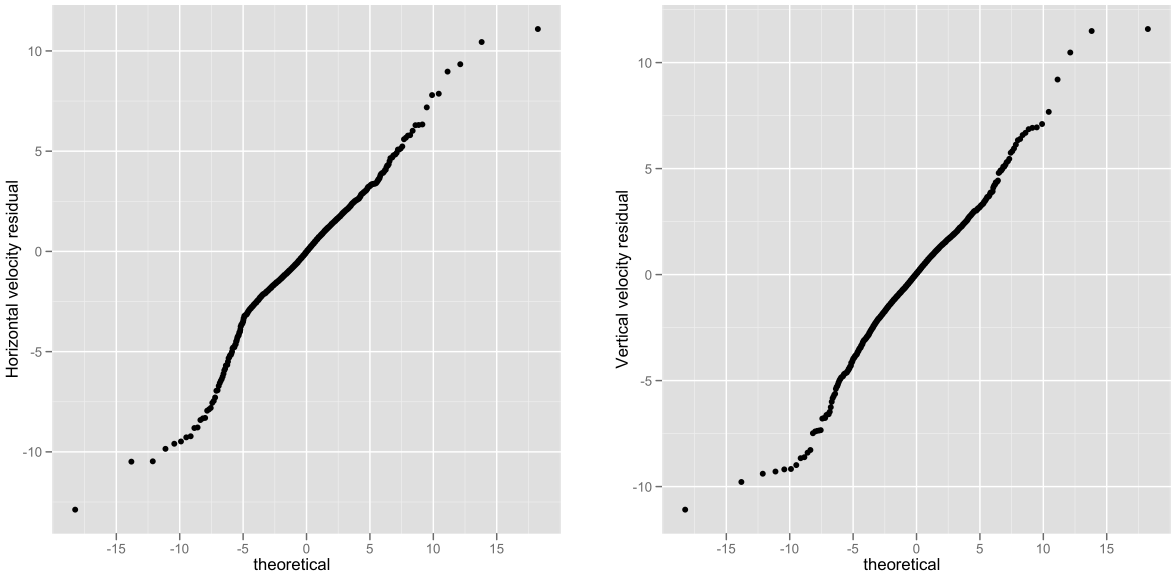


Figure 9.17: Quantile-quantile plots based on a t-distribution with four degrees of freedom for horizontal and vertical velocity residuals.

drifts the over fitting problem is reduced here. Smooth non-parametric fits of the potential are shown in Figure 9.18 for all eyes.

Interestingly all eyes show a point of attraction at the center of fixation. This suggests the microsaccades are correcting the target towards the center of fixation (although they may overshoot).

9.4.2 When do they occur?

First we look at how long after one microsaccade another one will follow. The auto intensity, $A(s)$, at lag s, can be defined as,

$$A(s) = P(\text{microsaccade at time } t+s \big| \text{microsaccade at time } t)$$

For a point process, $\{\tau_i, i = 1, \ldots, n\}$, this can be estimated empirically by,

$$\hat{A}(s) = \frac{1}{n} \sum_{i>j} 1\left(s - b < \tau_i - \tau_j < s + b\right)$$

where b is a binwidth parameter.

Since each experiment is short compared to the length of the time between microsaccades we need to adjust this value. For example, if the experiment is 10 seconds long and we are interested in lags of 5 seconds, only microsaccades occurring in the first half of the time period are available for estimation. To correct for this we can simply scale the value at each lag by $(10 - \text{lag})/(10)$. The auto-intensity is calculated within each experiment and then averaged across experiments for each lag. The results are shown in Figure 9.19. The normal and nonamblyopic eyes demonstrate an inhibition of microsaccades for the about the first second before a roughly constant probability. The amblyopic eye shows an opposite effect where microsaccades are more likely for the first half second following another.

Figure 9.20 explores the relationship between the probability of a microsaccade and location of the target. For the normal and nonamblyopic eyes there is evidence that the farther from the center of fixation the higher the probability of a microsaccade. For the amblyopic eye the area of highest probability of saccade seems to be in the lower left of the fixational area (this is the point closest to the fovea due to their eccentric fixation).

In order to model both the dependence on time and space we assume microsaccades are a Bernoulli process with the probability of occurring in the next time step, $\pi(x, y, t)$, when it has been t seconds since the last microsaccade and the target is at position (x, y). We model this probability as,

$$\text{logit } \pi(x, y, t) = f(t) + g(x, y)$$

where f and g are assumed to be smooth functions. This determines a generalized additive model and can be fit using the usual techniques. Fitted functions are shown in Figure 9.21.

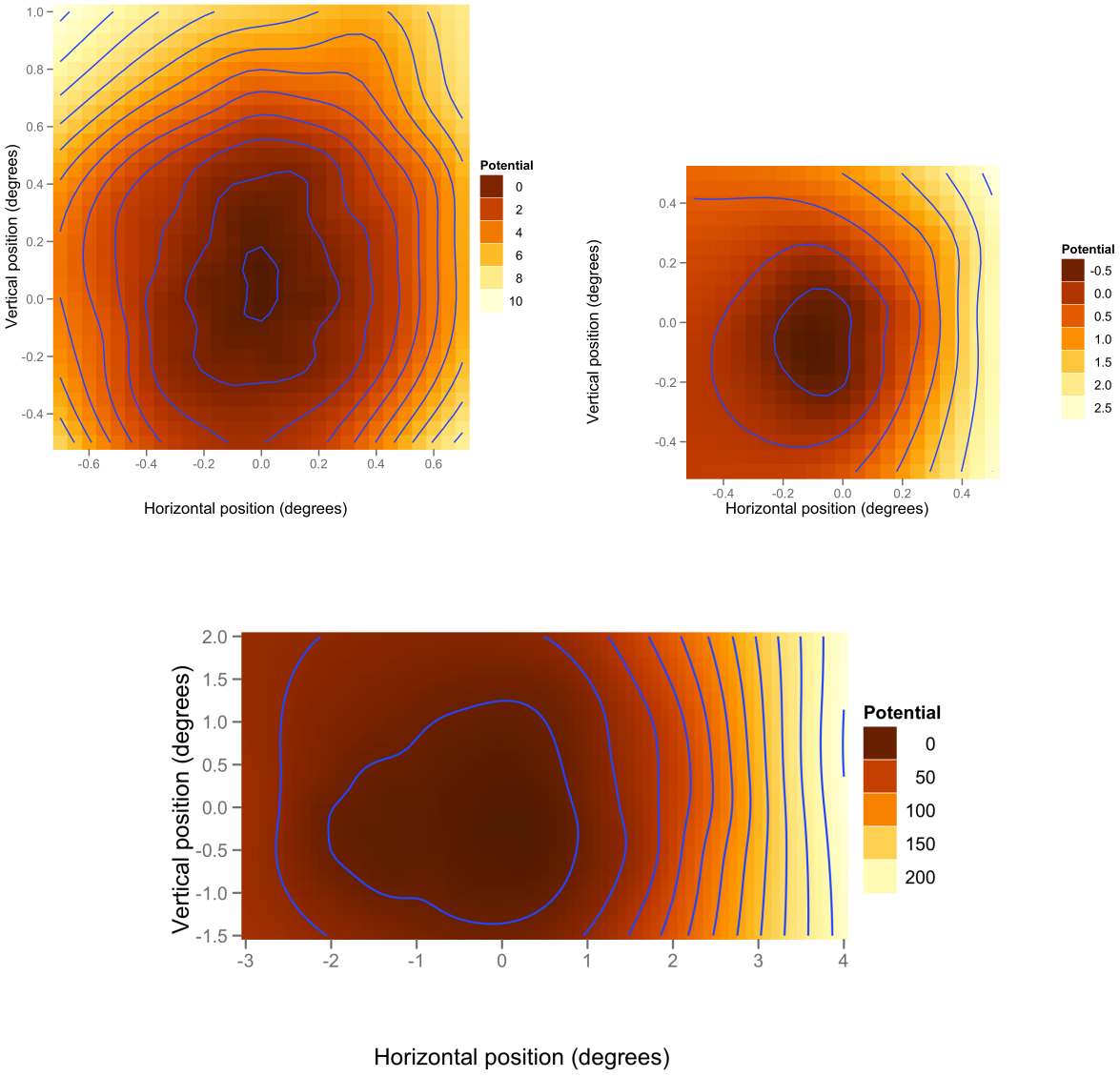


Figure 9.18: The fitted non parametric smooth potential for the microsaccades in the normal eye (top left), the nonamblyopic eye (top right) and the amblyopic eye (bottom).

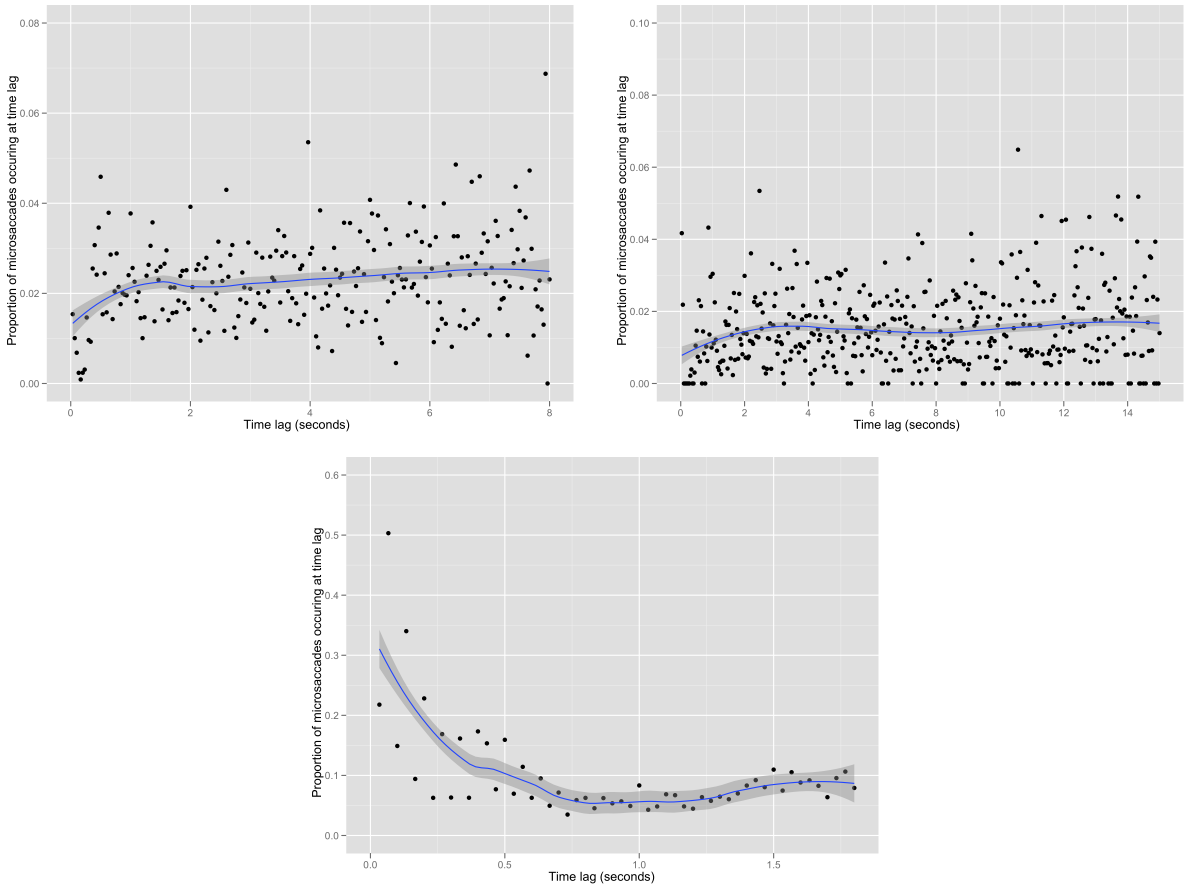


Figure 9.19: Scaled average empirical auto intensities for the three eyes.

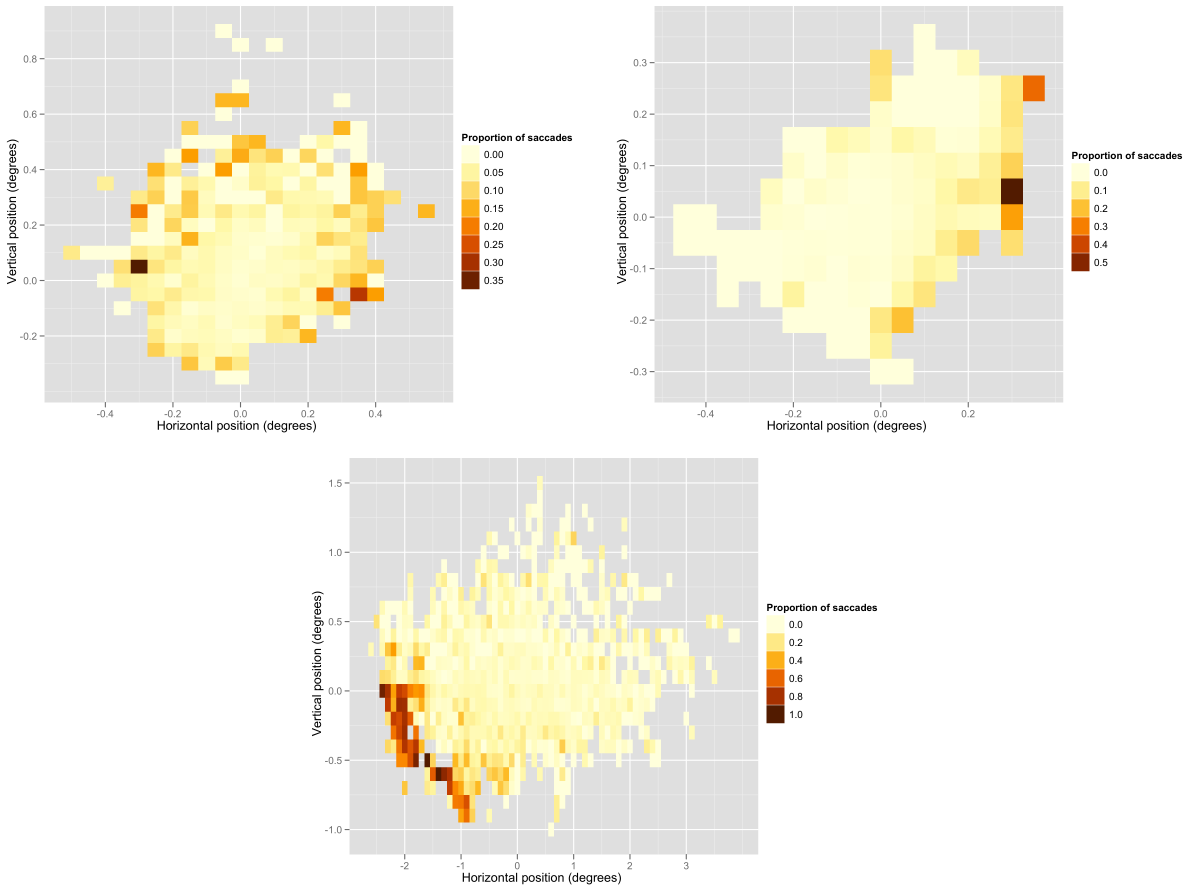


Figure 9.20: Binned proportion of microsaccades for the three eyes. Only bins with more than 5 observations are plotted.

The fit for the normal eye shows an inhibition of microsaccades directly after another and an increase in the probability of a microsaccade the further the target is from the center of fixation.

The probability of microsaccade by position for the nonamblyopic eye is interesting as there is evidence of two areas where the probability is low — one near the center of fixation and the other near (0.2, 0.2).

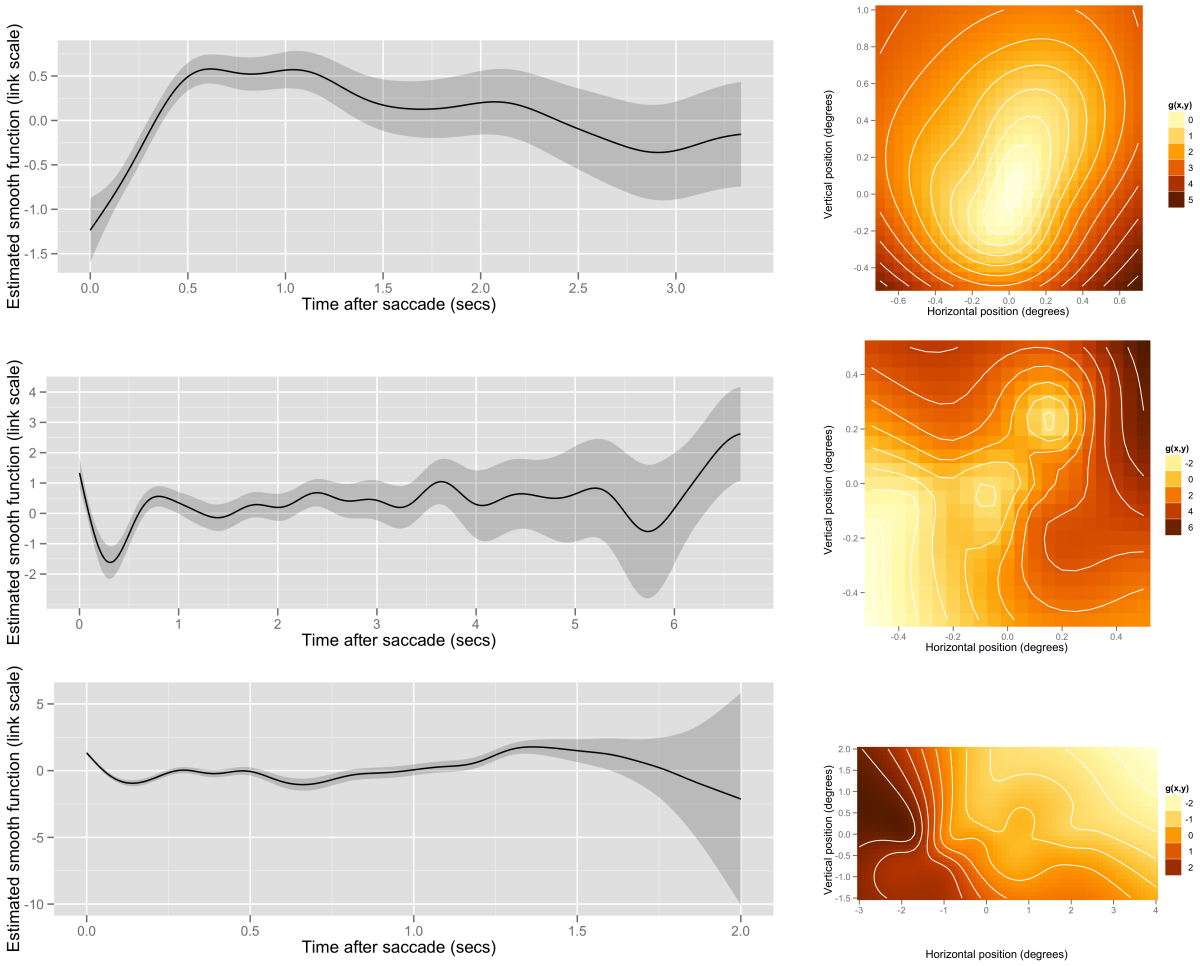


Figure 9.21: Fitted smooth functions of time and distance from center on the logit scale of the probability of microsaccade for the three eyes.

9.4.3 Summary of microsaccades

We find evidence in the eyes studied here that the microsaccades move in the direction of the center of fixation. In the normal and amblyopic eyes there is inhibition of microsaccades following another but the opposite is observed in the amblyopic eye. In the normal eye microsaccades are more likely the further from the center of fixation. The nonamblyopic eye

shows some evidence for microsaccades being less likely to occur if the target is near the center of fixation or an area up and to the right of the center of fixation. In the amblyopic eye microsaccades appear to be most likely in the area closest to the fovea.

9.5 Discussion

The potential function model proves useful both to model the subdrifts and microsaccades. They provide an intuitive way to visualize the tendencies in the movements and show the microsaccades tend to move toward the center of fixation but the drifts do not.

There are some questions left unanswered. How do the tendencies in subdrifts vary from eye to eye and person to person? Further experiments would be required to answer this question and the approach explored here would be useful tool.

9.5.1 Future Work

The approach here treated the behavior of microsaccades and drifts independently. Examining the robustness of the results to variation in the choice of empirical cutpoint for determining drifts and microsaccades is one area of future research. One could imagine modeling both at once with a state space model where the states are "drift" or "microsaccade" and switch according to a probability model like that explored in Section 9.4.2. A choice of cutpoint would then be unnecessary as the state at each time point would be estimated as part of the model fitting.

The residuals in the subdrift model exhibited much heavier tails than would be expected under a Normal assumption. This suggests the need for a robust method and is a focus for future work.

There was also evidence that the potentials modelling the subdrifts should be time dependent. This could be accounted for by allowing the quadratic potential coefficients to be smooth time varying functions and could be fit using `mgcv::gamm`.

It seems natural to model the subdrift lengths as survival times with occurrence of microsaccades signifying a death and a missing value censoring. Applying the methods of survival analysis to the subdrifts is another avenue for future research.

Chapter 10

Conclusions and Future Work

This thesis illustrated the usefulness of the potential function approach to modelling trajectories. Non-parametric and quadratic functions were used to demonstrate a point of attraction for shark 66003 off the coast of Kenya during the year 2007. The use of the model to generate bootstrap tracks allowed the discovery that the shark spent more time in 1000m deep water than could be explained by the quadratic potential alone. However, no covariates were found to be significant predictors of the sharks movement.

Chapter 8 demonstrated how appropriate choice of potential functions could be used to model the interaction between several animals. Although no evidence of interaction was found, the method will be useful for modelling other animals. The chapter also demonstrated how animals may share movement parameters. For five sharks that had similar ranges a hypothesis that they share common quadratic potential parameters was rejected.

The high error in location estimates of the PSAT tag technology illustrates the need for measurement error models. There is still work to be done on this approach. I only used quadratic potentials in the state space models in the analysis in this thesis. The equations laid out in Section 5.4 equations apply in the general case of a linear potential. This means non-parametric (based on a linear basis) potential functions with measurement error are possible and the implementation of the methods is an area of future work.

Accounting for boundaries is often important for accurate models of animal movement. Since the whale sharks in the Kenya study stayed near the coast it was an important consideration in creating bootstrap tracks. The true position of shark should only be in water (unless perhaps at the very end of a track if it has been caught or washed up on land), but the observed position could be on land. This is complicated slightly by the tag providers doing their own cleaning first, they do not report a position as on land even if it is the most likely position given the light level observations. Future work will concentrate on setting down relevant models for how animals move near boundaries. These could easily be incorporated into the state space model approach.

The eye movement data demonstrated the general applicability of the methods in this thesis. Since it was an experimental set up it had the advantage of replicates. This makes

estimating non parametric time-varying potentials quite feasible and is an avenue for future research. The eye movement data also illustrated a need for more research on for long tailed error distributions and robust potential function estimates. A promising avenue for research is using t-distributions for the measurement error in the state space approach. There is still work to do on working with large samples and avoiding overfitting.

References

Nautical Almanac. U.S. Naval Observatory and H.M. Nautical Almanac Office. URL `http://www.usno.navy.mil/USNO/astronomical-applications/publications/naut-almanac`.

T. Acharya and A. Ray. *Image processing: principles and applications*. John Wiley, 2005. ISBN 9780471719984. URL `http://books.google.com/books?id=smBw4-xvfrIC`.

D. Aldous. *Probability Approximations via the Poisson Clumping Heuristic*. Springer, 1989.

I. Andrianov, L. Manevich, and M. Hazewinkel. *Asymptotology: ideas, methods, and applications*. Kluwer Academic Publishers, 2002. ISBN 9781402009600. URL `http://books.google.com/books?id=IuFleWtEvt0C`.

Aviso Currents. Currents, geostrophic, aviso, global (1 day composite). URL `http://coastwatch.pfeg.noaa.gov/erddap/info/erdTAgeo1day/index.html`.

R. Bhattacharya and E. Waymire. *Stochastic processes with applications*. John Wiley Sons, 1990.

P. Billingsley. *Statistical inference for Markov processes*. University of Chicago Press, 1961.

P. Blackwell. Random diffusion models for animal movement. *Ecological Modelling*, 100(1-3): 87–102, 1997.

N. Bowditch. *American practical navigator*. Edmund M. Blunt, New York, 5th edition, 1821.

D. R. Brillinger. A particle migrating randomly on a sphere. *Journal of Theoretical Probability*, 10(2):429–443, 1997.

D. R. Brillinger. Some examples of random process environmental data analysis. *Handbook of Statistics*, 18:33–56, 2000.

D. R. Brillinger. Simulating constrained animal motion using stochastic differential equations. *Probability, Statistics and their Applications*, 2003.

D. R. Brillinger. Learning a potential function from a trajectory. *IEEE Signal Processing Letters*, 14(11):867–870, November 2007a.

D. R. Brillinger. A potential function approach to the flow of play in soccer. *J. Quant. Anal. Sports*, 3(1), 2007b.

D. R. Brillinger and B. S. Stewart. Elephant-seal movements: Modelling migration. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 26(3):431–443, sep 1998. ISSN 0319-5724. URL `http://links.jstor.org/sici?sici=0319-5724%28199809%2926%3A3%3C431%3AEMMM%3E2.0.CO%3B2-8`.

D. R. Brillinger and B. S. Stewart. Stochastic modeling of particle movement with application to marine biology and oceanography. *Journal of Statistical Planning and Inference*, 2010.

D. R. Brillinger, H. K. Preisler, A. A. Ager, J. G. Kie, and B. S. Stewart. Modelling movements of free-ranging animals. Technical Report 610, University of California, Berkeley, 2001. URL `citeseer.ist.psu.edu/487010.html`.

D. R. Brillinger, H. K. Preisler, A. A. Ager, and J. G. Kie. An exploratory data analysis (EDA) of the paths of moving animals. *Journal of Statistical Planning and Inference*, 122 (1-2):43–63, 2004a.

D. R. Brillinger, H. K. Preisler, A. A. Ager, and M. J. Wisdom. Stochastic differential equations in the analysis of wildlife motion. *2004 Proceedings of the American Statistical Association, Statistics and the Environment Section*, 2004b.

D. R. Brillinger, B. S. Stewart, and C. Littnan. A meandering hylje. *Festschrift for Tarmo Pukkila on His 60th Birthday*, 2006a.

D. R. Brillinger, B. S. Stewart, and C. L. Littnan. Three months journeying of a Hawaiian Monk seal. *IMS Lecture Notes - Monograph*, 2006b.

W. H. Burt. Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy*, 24(3):346–352, Aug 1943. URL `http://www.jstor.org/stable/1374834`.

C. Calenge. Exploring habitat selection by wildlife with adehabitat. *Journal of statistical software*, 22(6):2–19, 2007.

C. Calenge, S. Dray, and M. Royer-Carenzi. The concept of animals' trajectories from a data analysis perspective. *Ecological Informatics*, 4(1):34–41, 2009.

B. Carlin, N. Polson, and D. Stoffer. A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87(418):493–500, 1992.

A. L. F. Castro, B. S. Stewart, S. G. Wilson, R. E. Hueter, M. G. Meekan, P. J. Motta, B. W. Bowen, and S. A. Karl. Population genetic structure of earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology*, 16(24):5183–5192, 2007. URL `http://dx.doi.org/10.1111/j.1365-294X.2007.03597.x`.

J. Colman. A review of the biology and ecology of the whale shark. *Journal of Fish Biology*, 51(6):1219–1234, 1997.

L. Compagno. *Sharks of the world: an annotated and illustrated catalogue of shark species known to date.*, volume 2 of *FAO Species Catalogue for Fishery Purposes*. FAO, 2001.

R. L. Delong, B. S. Stewart, and R. D. Hill. Documenting migrations of northern elephant seals using day length. *Marine Mammal Science*, 8(2):155–159, 1992.

C. A. J. Duffy. Distribution, seasonality, lengths, and feeding behaviour of whale sharks (*Rhincodon typus*) observed in New Zealand waters. *New Zealand Journal of Marine and Freshwater Research*, 36:565–570, 2002.

J. Dunn and I. Brisbin Jr. Characterizations of the multivariate ornstein-uhlenbeck diffusion process in the context of home range analysis. North-Holland, 1985.

J. Dunn and P. Gipson. Analysis of radio telemetry data in studies of home range. *Biometrics*, pages 85–101, 1977.

A. M. Edwards, R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S. V. Buldyrev, M. G. E. da Luz, E. P. Raposo, H. E. Stanley, and G. M. Viswanathan. Revisiting levy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165):1044–1048, 2007. URL `http://dx.doi.org/10.1038/nature06199`.

ETOPO2v2. 2-minute gridded global relief data (etopo2v2), 2006. URL `http://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html`.

J. Flemming, C. Field, M. James, I. Jonsen, and R. Myers. How well can animals navigate? estimating the circle of confusion from tracking data. *Environmetrics*, 17(4):351–362, 2006.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, Jun. 1990. URL `http://www.jstor.org/stable/2289776`.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–533, 2006.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6): 721–741, 1984.

W. Gilks, N. Best, and K. Tan. Adaptive rejection metropolis sampling within gibbs sampling. *Applied Statistics*, 44(4):455–472, 1995.

L. Hall, P. Krausman, and M. Morrison. The habitat concept and a plea for standard terminology. *Wildlife Society Bulletin*, 25(1):173–182, 1997.

R. Hill. Theory of geolocation by light levels. *Elephant Seals. Population Ecology, Behavior, and Physiology*, pages 227–236, 1994.

R. Hill and M. Braun. Geolocation by light level. Kluwer Academic Pub, 2001.

J. Hunter. *The dynamics of tuna movements: an evaluation of past and future research*. Food & Agriculture Org, 1986.

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

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

K. Kemp. *Encyclopedia of geographic information science*. SAGE Publications, 2008. ISBN 9781412913133. URL `http://books.google.com/books?id=FrUQHIzXK6EC`.

D. G. Kendall. Pole-seeking brownian motion and bird navigation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(3):365–417, 1974. ISSN 0035-9246. URL `http://links.jstor.org/sici?sici=0035-9246%281974%2936%3A3%3C365%3APBMABN%3E2.0.CO%3B2-G`.

G. Kitagawa. Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041, Dec. 1987. URL `http://www.jstor.org/stable/2289375`.

P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, 1992.

T. Lai. Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *The Annals of Statistics*, pages 1917–1930, 1994.

T. L. Lai and C. Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics*, 10(1): 154–166, Mar 1982. URL `http://www.jstor.org/stable/2240506`.

D. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337, 2000.

J. Martin, C. Calenge, P. Quenette, and D. Allaine. Importance of movement constraints in habitat selection studies. *Ecological Modelling*, 213(2):257–262, 5 2008. URL http://www.sciencedirect.com/science/article/B6VBS-4RR86V8-1/2/ebd0b766a3b6c1601486c5a6d8e560b0.

S. Martinez-Conde, S. Macknik, and D. Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3):229–240, 2004.

J. Meeus. *Astronomical algorithms*. Willmann-Bell, Incorporated, 1991.

MODIS Chlorophyll. Science quality chlorophyll-a concentration from modis/aqua. URL http://coastwatch.pfeg.noaa.gov/infog/MH_chla_las.html.

MODIS SST. Science quality sea surface temperature from modis/aqua. URL http://coastwatch.pfeg.noaa.gov/infog/MH_sstd_las.html.

C. O. Mohr. Table of equivalent populations of north american small mammals. *American Midland Naturalist*, 37(1):223–249, Jan 1947. URL http://www.jstor.org/stable/2421652.

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

M. Musyl, R. Brill, D. Curran, J. Gunn, C. Boggs, and R. Brainard. Ability of archival tags to provide estimates of geographical position based on light intensity. Kluwer Academic Pub, 2001.

NASA QuikSCAT Winds. QuikSCAT global ocean surface wind velocity and divergence, science quality. URL http://coastwatch.pfeg.noaa.gov/infog/QS_ux10_las.html.

A. Nielsen, K. Bigelow, M. Musyl, and J. Sibert. Improving light-based geolocation by including sea surface temperature. *Fisheries Oceanography*, 15(4):314–325, 2006.

B. Oksendal. *Stochastic differential equations*. Springer Berlin, sixth edition, 2003.

J. O'Reilly, S. Maritorena, D. Siegel, M. O'Brien, D. Toole, B. Mitchell, M. Kahru, F. Chavez, P. Strutton, and G. Cota. Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. *NASA Tech. Memo*, 206892(11):9–27, 2000.

Petris, Petrone, and Campagnoli. *Dynamic Linear Models with R*. Springer, 2009.

G. Petris. An r package for dynamic linear models. *Journal of Statistical Software*, 36(12): 1–16, 2010. URL http://www.jstatsoft.org/v36/i12/.

G. Phillies. *Elementary lectures in statistical mechanics*. Springer, 2000. ISBN 9780387989181. URL http://books.google.com/books?id=tSJGiq4xUP8C.

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Development Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2011. R package version 3.1-98.

M. Plummer. Jags: A program for analysis of bayesian graphical models using gibbs sampling, 2003. URL http://sourceforge.net/projects/mcmc-jags/. Version 2.1.

M. Plummer. *rjags: Bayesian graphical models using MCMC*, 2011. URL http://CRAN.R-project.org/package=rjags. R package version 2.2.0-4.

H. Preisler, A. Ager, and M. Wisdom. Statistical methods for analysing responses of wildlife to human disturbance. *Ecology*, 43:164–172, 2006.

I. G. Pride and S. M. Swift. *Wildlife telemetry: remote monitoring and tracking animals*. Ellis Horwood Series in Environmental Management, Science and Technology. Ellis Horwood, 1992.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL http://www.R-project.org/. ISBN 3-900051-07-0.

S. Railsback, R. Lamberson, B. Harvey, and W. Duffy. Movement rules for individual-based models of stream fish. *Ecological Modelling*, 123(2-3):73–89, 1999.

M. Rolfs. Microsaccades: Small steps on a long way. *Vision research*, 49(20):2415–2441, 2009.

A. Roorda, F. Romero-Borja, W. Donnelly III, H. Queener, T. Hebert, and M. Campbell. Adaptive optics scanning laser ophthalmoscopy. *Optics Express*, 10(9):405–412, 2002.

P. J. Rousseeuw, I. Ruts, and J. W. Tukey. The bagplot: A bivariate boxplot. *The American Statistician*, 53(4), 1999.

F. Royer, J. Fromentin, and P. Gaspar. A state-space model to derive bluefin tuna movement and habitat from archival tags. *Oikos*, 109(3):473–484, 2005.

S. Shreve. *Stochastic calculus for finance: Continuous-time models*. Birkhuser, 2004.

R. Shumway and D. Stoffer. *Time series analysis and its applications*. Springer Verlag, 2000.

J. Sibert and D. Fournier. Possible models for combining tracking data with conventional tagging data. Kluwer Academic Pub, 2001.

J. Sibert, M. Musyl, and R. Brill. Horizontal movements of bigeye tuna (thunnus obesus) near hawaii determined by kalman filter analysis of archival tagging data. *Fisheries Oceanography*, 12(3):141–151, 2003.

J. Sibert, M. Lutcavage, A. Nielsen, R. Brill, and S. Wilson. Interannual variation in large-scale movement of atlantic bluefin tuna (thunnus thynnus) determined from pop-up satellite archival tags. *Canadian Journal of Fisheries and Aquatic Sciences*, 63(10):2154–2166, 2006.

R. W. Sinnott. Virtues of the haversine. *Sky and telescope*, 68, 1984.

S. Song. *Acuity, Crowding, Feature Detection, and Fixation in Normal and Amblyopic Vision*. PhD thesis, University of California Berkeley, 2010.

S. Song, E. A. Rossi, C. Wickham, A. Roorda, D. R. Brillinger, and D. M. Levi. Fixational eye movements for normal and strabismic amblyopic observers. *Journal of Vision*, 10(7): 456–456, 08 2010. URL `http://www.journalofvision.org/content/10/7/456`.

D. Spiegelhalter, N. Best, B. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

B. Stewart and S. Wilson. Threatened fishes of the world: Rhincodon typus (Smith 1828) (*Rhincodontidae*). *Environmental Biology of Fishes*, 74(2):184–185, 2005.

B. S. Stewart. Movement and habitats of whale sharks of East Africa: Results of tagging studies in 2007 & 2008. Technical Report 368, Hubbs-SeaWorld Research Institute, San Diego, California, 2009.

B. S. Stewart and R. L. DeLong. Double migrations of the Northern Elephant seal, *Mirounga angustirostris*. *Journal of Mammalogy*, 76(1):196–205, feb 1995. ISSN 0022-2372. URL `http://links.jstor.org/sici?sici=0022-2372%28199502%2976%3A1%3C196%3ADMOTNE%3E2.0.CO%3B2-L`.

B. S. Stewart, S. G. Wilson, N. Njonjo, and V. Bassen. Ecology of the whale sharks of East Africa. Technical Report 361, Hubbs-SeaWorld Research Institute, San Diego, California, 2007.

J. Taylor. *Classical mechanics*. University Science Books, 2005. ISBN 9781891389221. URL `http://books.google.com/books?id=P1kCtNr-pJsC`.

S. Teo, A. Boustany, S. Blackwell, A. Walli, K. Weng, and B. Block. Validation of geolocation estimates based on light level and sea surface temperature from electronic tags. *Marine Ecology Progress Series*, 283:81–98, 2004.

G. Viswanathan, V. Afanasyev, S. Buldyrev, E. Murphy, P. Prince, and H. Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415, 1996.

J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL `http://had.co.nz/ggplot2/book`.

H. Wickham. *plyr: Tools for splitting, applying and combining data*, 2011. URL `http://CRAN.R-project.org/package=plyr`. R package version 1.4.

S. Wilson, B. Stewart, J. Polovina, M. Meekan, J. Stevens, and B. Galuardi. Accuracy and precision of archival tag data: a multiple-tagging study conducted on a whale shark (rhincodon typus) in the indian ocean. *Fisheries Oceanography*, 16(6):547–554, 2007.

S. G. Wilson, J. J. Polovina, B. S. Stewart, and M. G. Meekan. Movements of whale sharks (rhincodon typus) tagged at ningaloo reef, western australia. *Marine Biology*, 148(5): 1157–1166, 2006. URL `http://dx.doi.org/10.1007/s00227-005-0153-8`.

S. Wood. *Generalized additive models: an introduction with R*. CRC Press, 2006.

S. Wood, M. Bravington, and S. Hedley. Soap film smoothing. 2008.

B. Worton. A review of models of home range for animal movement. *Ecological Modelling*, 38(3-4):277–298, 1987.

B. Worton. Modelling radio-tracking data. *Environmental and Ecological Statistics*, 2(1): 15–23, 1995.

B. J. Worton. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, 70(1):164–168, 02 1989. URL `http://dx.doi.org/10.2307/1938423`.

Appendices

Appendix A

Example Jags model file

Listing A.1: Jags model for quadratic potential and measurement error

```
var beta[5], pos[n,2], obs[n,2], pot[n, 2], timediff[n];
# pos[i, j] = true position at time i in direction j
# obs[i, j] = observed location at time i in direction j
# pot[i, j] =  gradient of potential at true position
# timediff[i] = time between observations at time i and i+1
# n = number of observations
model {
  pos[1, ] ~ dmnorm(obs[1, 1:2], InvTau)

  # movement model
  for (i in 2:n){
    pot[i-1, 1] <- -1 * timediff[i-1] * (beta[1] + beta[3] * pos[i
       -1, 2] +
      2 * beta[4] * pos[i-1, 1])
    pot[i-1, 2] <- -1 * timediff[i-1] * (beta[2] + beta[3] * pos[i
       -1, 1] +
      2 * beta[5] * pos[i-1, 2])
    pos[i, ] ~ dmnorm(pos[i-1,] + pot[i-1, ],
      invSigma2 * (1/timediff[i-1]) * Ident[1:2, 1:2])
  }

  # measurement model
  for (i in 2:n){
    obs[i,] ~ dmnorm(pos[i,], InvTau)
  }

  # priors for potential parameters
  beta[1:5] ~ dmnorm(mu, W)
```

```
  invSigma2 ~ dgamma(a, b)
  sigma <- 1/sqrt(invSigma2)

  InvTau ~ dwish(Phi[,], m)
  Tau <- inverse(InvTau[,])
  tau_x <- sqrt(Tau[1,1])
  tau_y <- sqrt(Tau[2,2])
  rho <- Tau[1,2] / (tau_x * tau_y)
}
```