Models and analysis of animal movements: From individual tracks to mass dispersal

Eliezer Gurarie

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

Doctor of Philosophy

University of Washington

2008

Program Authorized to Offer Degree:
Quantitative Ecology and Resource Management

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Eliezer Gurarie

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

_____

James Anderson

Reading Committee:

_____

James Anderson

_____

Daniel Grünbaum

_____

Richard Zabel

Date: _____

University of Washington

Abstract

Models and analysis of animal movements:
From individual tracks to mass dispersal

Eliezer Gurarie

Chair of the Supervisory Committee:
Professor James Anderson
School of Aquatic and Fisheries Sciences

Almost every process related to animal ecology, including foraging, predator avoidance, mate encounter, invasion, dispersal and migration, is intimately related to animal movements. In recent years, improvements in tracking and observation technologies have led to an explosion of movement data on all manner of organisms. Movement processes are, however, difficult to model mathematically. They are the result of extremely complex interactions between an organism's internal state, behavioral tendencies and environmental cues. The data are multi-dimensional and almost always non-independent, and there is no consensus on the appropriate statistical summaries or underlying models. This dissertation aims to address several issues in quantitative movement ecology beginning with rigorous mathematical descriptions of individual movement and culminating in models of mass movements and dispersal.

Part I, containing Chapters 2 and 3, is devoted to theoretical and mathematical considerations related to parameterizations of movement process. In Chapter 2, the correlated random walk model (CRW) is discussed in detail. Generalizations of the CRW for arbitrarily sampled data are presented, as well as methods to obtain the essential summaries of homogeneous movements: characteristic length scales and time-scales of independence. In Chapter 3, a continuous, autocorrelated stochastic

model of movement and its parameterization is discussed, along with algorithms for estimating the associated parameters. The second part covers several applications and extensions of the basic movement models on an ecological scale. In Chapter 4, theoretical relationships between the fundamental time and length scales of movement and encounter rates are derived and applied to survival of migrating salmon. Chapter 5 presents a statistically robust and informative method for identifying multiple behavioral modes in irregularly sampled individual animal track data. In Chapter 6, migratory and dispersing mass movements are considered in terms of population level heterogeneity in the movement parameters. Methods of separating intrinsic randomness of movement from differences within and between populations from dispersal and travel-time distributions are derived.

The models and methods throughout this work are applied to a wide range of aquatic organisms: microscopic algae in a laboratory, migrating salmonids and dispersing cyprinids in freshwater environments, dugongs in subtropical Australian waters and northern fur seals in the northwestern Pacific Ocean.

TABLE OF CONTENTS

iv

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

In the highly tortuous and meandering path that culminated in the writing of this dissertation, I have been blessed with inspiration, support and encouragement of a great many individuals. My advisor, Jim Anderson, beyond bottomless support, patience and encouragement, has amazed me time and time again as a fount of original ideas. Seeing these ideas coming to a boil in the crucible of our many extended, sometimes chaotic "whiteboard sessions" is a real vision of how great science can be done. My other committee members, Danny Grünbaum, Julia Parrish and Rich Zabel, provided unique and complementary combinations of knowledge about the natural world, insights into both the means and meanings of mathematical modeling, perspective in issues of natural resource management and conservation, and experience and wisdom regarding the scientific process itself.

Financial support over the years has come from the College of Arts and Sciences via the QERM program, the Bonneville Power Administration, Northwest Fisheries Science Center (NOAA-NMFS) and the National Marine Mammal Lab (NOAA-NMFS).

I am deeply grateful to everybody involved in the process of collecting, processing and sharing the datasets analyzed in this dissertation. Heterosigmna movement data was provided by D. Grünbaum's lab, with thanks to Mike Nishizaki. Russian fur seal movement data was provided by Russell Andrews in the context of the Russian Far East Marine Mammal Research Program. The dugong movement data was collected by the Department of Environment and Conservation in Western Australia and generously shared by David Holley. Salmon migration data was provided by the BPA, the Army Corps of Engineers and the Chelan County Public Utility District No. 1,

with special thanks to Tracy Steig and Chris van Holmes.

On questions of mathematical and statistical modeling, I have benefited greatly from interactions and discussions with Tilmann Gneiting, Mark Kot, Vladimir Minin and Sidney Redner. The past and present students of our relatively small QERM program, despite diverse backgrounds and interests, have been an exceptional source of both intellectual and moral companionship. Other peers - too many to list off - at the SAFS, CFR and the Biology department, as well as students at the Pfister-Wootton Lab at the University of Chicago have provided everything from intellectual stimulation to data to play with to opportunities to muck around in the field. I would like to thank Willy Eldridge, in particular, for sharing a base of knowledge and experience surpassed only by good humor and an childlike glee at exploring the world. Who else could make memories of flinging salmon carcasses across the rain drenched tributaries of the Nooksack River among the more cherished of graduate school? Also, Elizabeth Skewgar, whose interests overlapped closely with mine, revealed to me most directly the joys of collaboration.

I would like to thank Vladimir Burkanov for having given me the opportunity to spend do field work in Eastern Russia. Though little of the work done there is directly present in this dissertation, much of the wonder, awe and humility I have before the dynamic complexity, mystery and beauty of the natural world comes from those many months on uninhabited islands. Opportunities to work on marine mammals with people from Kamchatka, Vladivostok, Magadan and Moscow, as well as with scientists in Alaska from the USGS, the FWS and the ADFW have broadened my horizons considerably.

Finally, my closest friends and family, mama, papa, babushka, Mark, and to Kristin, who have, with their unconditional support, made this journey not only possible but often a joy. And their love gives this, and any other endeavor, meaning.

Chapter 1

INTRODUCTION

1.1   Overview of movement modeling in ecology

Movement is one of the defining features of many forms of life. Most representatives of the protist and animal kingdoms move actively. Even those relatively few animals that are sessile, including sponges, bryozoans, some mollusks, cnidarians, and annelid worms, generally display active motility in the larval stages. Furthermore, many alga move actively while passive movements, most notably of genetic material by physical or animal related dispersal, is an essential process for most other plants.

It can consequently be argued that most processes in ecology are directly related to movement. Survival is in large part the result of successful acquisition of food and avoidance of predators; reproductive success depends on encounters with mates; habitats and home ranges are defined as the geographical areas in which physical displacement of individuals occurs; invasions are mass movements toward previously unoccupied habitats; and so on. Conservationists and resource managers need to identify home ranges and migration pathways as well as the interactions between an organism and its habitat. Evolutionary ecologists are interested in teasing out the selective forces related to successful food encounter, mate encounter and predator avoidance. Ethologists investigate the strategies and tactics by which animals navigate their environment to fulfill their survival needs.

It might therefore seem surprising that for the bulk of the history of ecology as a scientific discipline, the study of movement has played a somewhat peripheral role. This is illustrated by a casual overview of "classic" papers in *Foundations of Ecology: Classic Papers with Commentaries* (Real and Brown, 1991). Of the 40 papers in

the compilation, roughly 15 deal with population structure, predation and feeding strategies, community structure and habitat of animal species. Of these only one article, C. S. Skellam's (1951) seminal "Random dispersal in theoretical populations", addresses the analysis of movement explicitly.

This disparity between the self-evident centrality of movement processes and its historical absence in ecological theory can be attributed in part to the lack of appropriate data to motivate theoretical development. In the aforementioned study, Skellam (1951) develops many elaborate mathematical formulations of expected rates of dispersal and migration based on a random-walked derived diffusion approximation combined with simple models of reproduction and population growth. But the only data Skellam considered was the multi-generational expansion of the range of introduced North American muskrat (*Ondatra zibethicus*) in Central Europe documented by Ulbrich (1930), and the comparisons to the data are essentially qualitative. Skellam also had no access to data on individual movements, nor did he make a rigorous attempt to relate individual movements to the parameters of his equations, except to briefly discuss the discrete, unbiased, one-dimensional random walk.

From the perspective of individual movements, a very rigorous analysis of individual movements was provided in a pair of excellent early papers by C. Patlak (1953a,b). In these largely forgotten papers the effect of various forms of correlation and anisotropy in individual movement steps are related to diffusion approximations, with explicit derivations of expected displacements and rigorous tests for external orientation. Patlak considers a discretization of movement that consists of step lengths l, the tuning angles θ and time intervals t drawn from some arbitrary distributions and predicted the displacements and theoretical rates of diffusion for such movers in terms of the mean and variance or step lengths, the mean cosine of turning angles, and the mean time interval. Both of his papers were purely theoretical, but the author concludes somewhat hopefully that these models

"... will provide a systematic method for experimental investigation of the problem of the orientation of organisms ... [and] that by use of the knowledge of the various parameters not only may the behavior of the organism in its natural environment be predicted but also the search for the underlying physiological mechanisms for this behavior may be facilitated" (Patlak, 1953a).

Patlak's mathematical results were far ahead of the data, and consequently mostly ignored. Some of his results were rederived in an extremely influential study by Kareiva and Shigesada (1983), who were apparently unaware of Patlak's work. These authors essentially recreated Patlak's movement, ignoring the variable time intervals, and dubbed it a *correlated random walk* (CRW). In contrast to any earlier studies, however, the authors could apply their model on actual data on the movement of cabbage white butterflies (*Pieris rapae*) through an experimental field. Based on the expected squared displacement according to the theory, they conclude, for example, that while ovipositing butterflies move in a "correlated random" manner, nectar-feeding butterflies displace in an manner that violates the CRW assumptions. The authors summarize the utility of quantifying individual movements:

It may be especially fruitful to investigate the foraging or searching movements of animals as correlated random walks whose parameters (turning angle or move length) depend on local ecological conditions. Thus, for example, an organism might always move randomly, but with sharper turning angles in the presence of food resources. (Kareiva and Shigesada, 1983)

This is a concise summary of the simplest sort of hypothesis relating measurable changes in animal movements to an environmental covariate.

4

1.1.1 Tracking data

The data analyzed by Kareiva and Shigesada (1983) was meticulously collected by manually pursuing over 200 butterflies at a respectful distance of four to five feet as they navigated through a planted field of potatoes and greens and meticulously attaching color-coded clothespins to the landing sites (see Root and Kareiva, 1984, for details on the methods). Since then, the development and application of animal tracking technology have undergone extraordinary advances leading to an explosion of data. Radio tags, ARGOS and GPS satellite tags have been developed that follow everything from thousand mile migrations of birds such as storks and albatross (Jouventin and Weimerskirch, 1990; Blouin et al., 1999; Weimerskirch et al., 2000; Martell et al., 2001; Kanai et al., 2002; Fritz et al., 2003; Grünbaum and Veit, 2003; Chernetsov et al., 2004; Shimazaki and Tamura, 2004; Berthold et al., 2006, others), free ranging terrestrial mammals such as elk, reindeer, wolves and bears (Keating et al., 1991; Bascompte and Vilà, 1997; Mårell et al., 2002; Morales et al., 2004; Fortin et al., 2005b; Forester et al., 2007; Johnson et al., 2007, others), foraging pinnipeds (Andrews, 1998; Fedak et al., 2002; Raum-Suryan et al., 2004; Speckman et al., 2006), large pelagic creatures including cetaceans (Laidre et al., 2004; Johnson et al., 2008, Zerbini, unpublished data) and bluefin tuna (Block et al., 1998; Royer and Gaspar, 2005), to the daily peregrinations of tree-dwelling monkeys (Ramos-Fernández et al., 2004), marten (Nams and Bourgeois, 2004) and dugong (Holley, 2006; Sheppard et al., 2006). Hydroacoustic technologies are used to obtain three-dimensional tracks of salmon approaching dams (Ransom et al., 2007; Goodwin et al., 2006; Steig et al., 2006) and foraging sturgeon (M. Parsley, USGS Cook Biological Station, *pers. comm.*). Digital video setups tracking movements of individual plankters, various larvae and fish in laboratory settings are increasingly common (Parrish et al., 2002; Bearon et al., 2004; Seuront et al., 2004; Uttieri et al., 2005). If Root and Kareiva were to repeat their collection of the butterfly movement data now, they might consider attaching 22

mg radar transponders to the butterflies' thoraxes and using a large harmonic radar to obtain the butterfly data, as was successfully performed by Cant et al. (2005). The preceding list is necessarily far from comprehensive, but it gives an idea of the range and diversity of the kinds of individual movement data, that researchers only 20 years ago could scarcely have imagined.

1.1.2 Analysis methods

In the wake of this ever-growing body of data, there has been a balkanization of approaches to the quantification of movement data. Many of these methods have moved away from the explicit organism-centric modeling implicit in the correlated random movement models, focusing more on descriptive statistics that quantify some characteristic of the path itself. To this family of descriptive analysis methods one can ascribe fractal analysis and the Lévy walk models.

Fractal analysis

The fractal dimension D of an object measures its ability to fill Euclidian space (Mandelbrot, 1967). A straight line is a 1 dimensional object, whereas a pure Brownian motion - i.e. a random walk composed of independent random steps - has fractal dimension 2. A fractal dimension of a curve is estimated using the simple regression method applied by Mandelbrot for the English coast.[1]

The ease of application and intuitive nature of fractal analysis have led to an abundance of studies applying the technique. A short list of organisms whose paths have been analyzed in terms of fractal dimensions includes copepods, spider mites,

---

[1] In a brief note published in *Science*, Benoît Mandelbrot (1967) noted that coastline's measured length $L(r)$ increases as the ruler of length r becomes smaller and smaller. The relationship is given by the simple empirical formula: $L \propto r^{1-D}$, where D is interpreted as a fractional dimension of the curve. The primary purpose of Mandelbrot's note was to demonstrate that a fractional dimension, far from being an esoteric "invention of mathematics", did, in fact, have "simple and concrete applications and great usefulness".

grasshoppers, clownfish, albatrosses, sheep, martens, wolves, polar bears and narwhals. Most commonly, authors compare estimated fractal dimension D against some environmental or internal state factor, like season (Laidre et al., 2004), life cycle (Bascompte and Vilà, 1997), abundance of food, predators or competition (Coughlin et al., 1992; Dicke and Burroughs, 1988; Doerr and Doerr, 2004), landscape structure (With, 1994). For example, Laidre et al. (2004) comprehensively analyzed the movement patterns of narwhals in Greenland, comparing these patterns between different bays, seasons, and different age and sex classes. The lower fractal dimension of migratory behavior was well constrasted with the more tortuous movements during the summer and winter periods. The relationship between the nature of the more tortuous movements could be realted to the fractal structure of the ice fields or the fjord-lined coast. Dicke and Burroughs (1988) performed a behavioral manipulation experiment, tracking spider mites in an attractive field of Lima bean odor and in a repellent field of Lima bean odor plus other spider-mite pheromone. The spider mites in the repellent field left with more directed, less tortuous, paths than those that made more wiggly foraging movements in the attractive field. In both of these cases, the fractal dimension provided a robust and relatively simple indirect index of behavior.

Other researchers have explored how the estimate for D varies across scale rather than calculating a single value for a curve (Fritz et al., 2003; Nams, 2005). The purpose of these studies is to elucidate and quantify the hierarchical nature of an organism's spatial interaction. For example, Fritz et al. (2003), analyze albatross tracks up to 1100 km long with a resolution on the order of a single meter. The authors note that there is a consistent and distinct dip in tortuosity between the 100 m and 10 km scales. The authors conclude that albatrosses are responding to their environment variously on different scales. The small scale reflects adjustments to small changes in wind direction and intensity, the medium (low tortuosity) zone reflects the effectiveness of the albatross' specific ability to soar, and the higher tortuosity at a longer scale

reflects responses to larger scale weather and wind patterns.

There have been debates as to whether a fractal curve is an appropriate model for animal movement data from a technical point of view since a truly fractal object is formally non-differentiable, whereas the motion of an organism space is necessary smooth. In a fairly influential critique of fractal analysis, Turchin (1996; 1998) argues that at the scale of the organism itself, all motion is approximately linear, whereas at a large temporal and spatial scales (over an entire range, or an entire lifetime), most organisms movementd will appear uncorrelated and therefore more truly random. The measured estimated fractal dimension can therefore be an artifact of the temporal resolution of the data collection.

Another potential shortcoming of fractal analysis is the treatment of an organism's track as a purely spatial object. Tracks, in fact, are objects in space and time. The spatial component can be thought of is a projection of a three-dimensional $\{X, Y, T\}$ object onto a two-dimensional $\{X, Y\}$ plane. A purely spatial characterization of a trajectory risks losing some information, especially when velocities are variable. However, if velocities are fairly consistent and the sampling is more or less regular, the fractal dimension seems to be a generally robust measure.

In general fractal analysis of paths is a straightforward and relatively simple to implement characterization of movement paths. When applied judiciously, it can be an excellent tool for identifying observed patterns of relatively coarsely measured movements on large scales (as with the narwhals), for characterizing movements in a controlled behavioral experiments (as with the spider mites), and for parsing the cross-scale nature of movements for high resolution data (as with the albatross). Methods presented in this dissertation hope to provide a similar application while accounting both for the ultimately smooth and time-dependent nature of movement.

Lévy walks

The analysis of animal movements in terms of Lévy walks (or Lévy flights) has also led to some fruitful analyses. A Lévy walk is essentially an uncorrelated random walk distinguished by a step length distribution following some power law: $f(l) \sim t^{-\alpha}$. Higher values of α are marked by many small steps, and an occasional very long movement, and therefore have a fairly distinctive "look". For values of $\alpha < 3$, the movements are super-diffusive and the dispersals become leptokurtic (Zhang et al., 2007). Lévy walks have been applied to fit long time-scale animal movement data where periods of tight movement are interrupted with occasional long voyages, including albatross (Viswanathan et al., 1996, 1999), plankton (Levandowsky et al., 1988; Klafter et al., 1989), amoebas (Levandowsky et al., 1997), jackals (Atkinson et al., 2002), reindeer (Mårell et al., 2002), spider monkeys (Ramos-Fernández et al., 2004) and humans (Brockmann et al., 2006).

Beyond the qualitative observation that organisms display rare long distance movements, some researchers have made an effort to explain the adaptive significance of Lévy movements. Thus, motivated by albatross movements, Bartumeus et al. (2002, 2005) found via simulation that for sparcely distributed, slow-moving prey and large interaction distances, encounter rates can be significantly higher for a Lévy random search than for a Brownian random search. This is explained by the ability of the Lévy walker to explore a greater area as well as minimizing the probability of reaching areas already visited.

Boyer et al. (2006) demonstrate that a Lévy walk can emerge from a foraging process that is not itself random. In their simulation of monkey foraging, resources are allocated patchily in trees and each tree has a power-law distributed random quality (i.e. a few high quality spots, many poorer quality spots). Simulated foraging monkeys with full knowledge of their environment aim to minimize their energy expenditure, spending an amount of time in a tree proportional to its quality, and often

preferring long trips to other high-quality trees over shorter trips to lower quality trees. The results of their model conform broadly with the behavior of spider monkeys (*Ateles geoffroyi*) and tree quality distributions measured in the field (Ramos-Fernández et al., 2004). Unlike the encounter rate simulations by Bartumeus and colleagues, the Lévy walk here is not a strategy *per se*, but a consequence of resource distributions. Similarly, the Lévy-type scaling associated with human movements is related to the spatial distribution of urban centers (Brockmann et al., 2006).

Several recent studies, including work done by some of the early proponents of the Lévy walk model, suggest that patterns which fit the Lévy walk model are more informatively fit by multi-modal random walks or other more behaviorally sophisticated models (Edwards et al., 2007; Benhamou, 2007). Lévy-walk like statistical properties can therefore be considered an "emergent property" of more classical movement models that undergo behavioral shifts.

Correlated random walk and extensions

Many studies apply or explore some version of the Patlak-Kareiva-Shigesada-type correlated random walk model alluded to earlier (Alt, 1990; Firle et al., 1998; Bergman et al., 2000; Mårell et al., 2002; Morales et al., 2004; Fortin et al., 2005b,a; Forester et al., 2007, others). In an example of a relatively sophisticated application of the CRW to study the interaction of animals and the environment, Fortin et al. (2005b) assess movement of elk (*Cervus canadensis*) in Yellowstone National Park in the context of a trophic-cascade hypothesis that the reintroduction of wolves encourages aspen by controlling overgrazing by the elk. In this study, the researchers radio tagged 14 female elk and described the environment using several relevant factors: aspen stands, distance to roads, slopes, use by wolves. They then compared the actual paths of the elks with simulated paths generated using turning angles and step-lengths estimated from the actual elk and explored whether the paths the elks chose differed from the random paths, and how. They concluded convincingly that

the elk do respond to wolves on the winter range by shifting their habitat selection from aspen, a preferred habitat of wolves, to conifer forests.

Nonetheless, there are some recurrent fallacies that emerge in the CRW literature. In particular, there is a tendency in many studies to "test" whether a movement track is a CRW or a simple RW. In fact, at a small enough temporal resolution *all* movements are necessarily correlated and a long enough temporal resolution, *all* movements are uncorrelated. For example, in another study of foraging elk by the same group, Fortin et al. (2005a) repeatedly test the movement data against a simple random walk (RW) hypothesis, concluding that: "This directional persistence indicated that foraging elk do not travel as simple random walkers." In fact, a more accurate statement would be: "The temporal resolution of our data was within the auto-correlation lag of the foraging elk's movement." An explicit awareness of the meaning of temporal scale of the correlation in movement is generally lacking in CRW literature, which contents itself with reporting clustering coefficients.

Extensions of correlated random walk models

Recently, several more mathematically and statistically sophisticated approaches to interpreting movement data have emerged.

In order to address the fact that much movement data is error-ridden, several excellent studies on the use of state-space model (SSM) approaches to analyzing animal movement data have appeared (Jonsen et al., 2003; Morales et al., 2004; Jonsen et al., 2005; Royer and Gaspar, 2005; Forester et al., 2007; Ovaskainen et al., 2008., see review by Patterson et al. 2008). A SSM is a stochastic time series model in which the observed data is interpreted as the result of a coupling between a *process* model with an *observation* model. A process model might be a correlated random walk with parameters that change in response to some environmental variable, while the observation model might be a standard Gaussian error. Parameters of a model are estimable using maximum-likelihood or Bayesian approaches. The important

advantage of the state-space framework is the ability to estimate the extent and effect of measurement errors. The limitation of the state-space framework is that it is ultimately only as good as the underlying movement model and the behavioral hypotheses associated with it. There is little direct need for an state-space framework for data for which the errors are essentially negligible, e.g. GPS, acoustic and video-tracking data.

The use of state-space models necessarily begs the question of what explicit model of movement is, in fact, most realistic or informative. Several kinds of models have been proffered in the literature. In an analysis and simulation of data on foraging elk (*Cervus elaphus*), Morales et al. (2004) suggests a CRW with Weibull distributed steps and wrapped Cauchy turning angles (see chapter 2 for definitions of these distributions) where the parameters have some probability of switching from a traveling-like mode to a foraging-like mode. The authors compare several different switching models (for example: no switching, constant switching probability, switching probability which depends somehow on habitat type), estimate parameters and compare models. They conclude that a model with two types of behaviors dependent on covariates is the most parsimonious and informative model. Similarly, Skewgar et al. (*in prep.*) parse foraging trips of nesting Magellanic penguings (*Spheniscus magellanicus*) in South America into directed travel modes and foraging/searching modes using the identical Weibull-Wrapped Cauchy model, and estimate Markovian transition probabilities between the states. Forester et al. (2007) test models, again working with elk movement data, in which switching between states depends on a complex of landscape factors. Ovaskainen (2004) presents a rigorous method for using diffusion models for modeling butterfly and wolf movements,[2] in which the diffusion parameters depend on certain landscape features. It should be noted that the butterfly data the authors analyze is mark-recapture data which, though fundamentally similar to individual

---

[2]The latter example comes from personal communications and unpublished data.

track data, is far coarser with, typically, a complete loss in auto-correlations.

Several older studies have developed continuous stochastic models of movement and applies them to movements of microscopic organisms (Dunn and Brown, 1987; Alt, 1988, 1990). These models are discussed in detail in chapter 3. A recent study by Johnson et al. (2008) presents a continuous-time correlated random walk in terms of an Ornstein-Uhlenbeck process and applies the models to interpreting movements of several species of marine mammal. Perhaps the greatest advantage of a rigorously parameterized continuous-time model is that data that is irregularly sampled can be modeled (chapter 5).

Individual-based simulation models

All of the analysis methods presented thus far describe movement as stochastic processes with some estimable statistical properties. Behavioral changes, when included, are modeled as basic responses to coarse environmental covariates. This reduction is almost unavoidable, since movement is a complex outcome of behavioral responses to individual's internal states, biophysical constraints, external stimuli and simple whimsy, and these interactions are daunting to model. Nonetheless some attempts at modeling explicit behavioral rules have been made, with occasionally striking results.

For example, Goodwin et al. (2006) developed a highly involved simulation model for the movements of juvenile salmonids near dams in the Columbia river. The model combines hydrodynamic models of water flow velocities, shears and strains with a behavioral model of fish response to the flow structure based on a event-based memory model of behavior proposed by Anderson (2002). While computationally very intensive, the model does reproduce observed fish behaviors in dams forebays with good qualitative agreement. Notably, the duration and nature of milling before passing through the dam and the specific routes in the flow environment that fish select are well captured. Furthermore, the model allows for experimental passage setups

to be simulated virtually, for example to test the effects of shifting the position of a fish passage entrance or changing the orientation of a boom. A drawback of the method is one common to many very detailed individual based models: It is difficult to compare models or test how realistic a model might be since the structure of the model itself is based largely on guesswork. Nonetheless, in view of the primarily pragmatic, engineering-oriented purpose of the model, accurate and useful predictions are a satisfactory objective.

Considerable research has been devoted to investigating clearly non-independent aggregated dynamics, such as insect swarm, fish schools, bird flocks, herds of ungulates, even human aggregations. Group dynamics are extremely complex, and understanding their formation, ecological function and evolutionary genesis calls for high quality empirical datasets together with novel theoretical and modeling frameworks (Parrish and Edelstein-Keshet, 1999). A direct approach to exploring these dynamics is to measure the movements of multiple individuals in a controlled environment and attempt to obtain empirically derived behavioral rules of interaction. Studies along these lines have been performed by taking video measurements on groups of tropical, freshwater cyprinids (giant danio - *Danio aequipinnatus*) in a tank (Parrish et al., 2002; Viscido et al., 2004, 2005; Grünbaum et al., 2005). From the data, individual parameters such as preferred nearest neighbor distance and orientation, turning curvature, and velocities were measured and used to calibrate individual based models. The group dynamics of the real and simulated fish, based on several metrics of aggregated movements, were compared, shedding some insights into the factors that control the size, coherence and behavior of aggregations.

These last two examples illustrate ways in which quantification of individual movements can be integrated into individual based simulation modeling frameworks to explore "emergent" properties of nearly intractably complicated systems, whether due to complex environments such as the flow regime in the forebay of a dam or to interactions between individuals in an aggregation.

1.1.3 Comment on randomness

Since almost all the sections of this work deal to some extent with versions of so-called "random walk" models, it is worth pausing to discuss the extent to which movement can legitimately be considered "random".

Organism movement is due in some proportion to internal choice and direct external control or constraints. Seeds or spores in the wind or non-motile larvae in water are transported entirely passively, and their dispersal is well-described by a statistical description of their medium. A large land-based vertebrate moves through the thin soup of air more willfully, but is often constrained by topographical or landscape-defined features that might be considered randomly distributed. The flight of a bird in strong winds or the movement of a fish in strong flows is some mixture of willful and medium-constrained motion. In short, most willful movement processes are very complicated deterministic processes with multiple unknown factors, some of which might be stochastic in nature. All of these processes combining into a single movement output are most conveniently summarized statistically. The observed movement which is the output of these processes is then often considered "random", though there can be very little that is truly random about the movement.

As information is collected on possible stimuli or explanatory factors to describe a movement, researchers ideally separate deterministic factors, explicitly model-able stochastic processes, and are left with a remaining error, whether related to measurement that most parsimoniously contains the 'intrinsic' randomness and those deterministic factors which we cannot explain. An analogy can be made to standard regression models, which generally attribute a certain amount of variation to "random error". In fact, such errors are most commonly not truly random, but merely unexplained and occasionally unexplainable.

Researchers would be well-advised to follow the recommendation of Turchin (1998)[3]

---

[3]made in a section of his seminal book appropriately entitled: "The bugbear of 'randomness' "

that the term "random movement" be replaced as much as possible with "stochastically modeled", if only for the sake of precision of meaning. However, historical convention and strong roots in the idealized mathematical model known as the "random walk" make use of the term quite common, and this dissertation is no exception.

1.2   Fundamental assumptions and structure of dissertation

Despite the great diversity and rapidly increasing number of movement studies in ecology, I would tend to agree with Jonsen et al. (2003) that "[o]ur ability to analyze movement patterns ... has been far outstripped by our ability to collect individual movement data". The overarching purpose of this dissertation is to move the mathematical and statistical analysis of movement data forward with the goal of developing practical methods to obtain biologically and ecologically relevant information from various kinds of movement data. The subjects of the various chapters of the dissertation range from rigorous mathematical parameterizations and predictions of ecological processes based on movement parameters, to identifying behavioral changes in individual tracks and characterizing mass movements. The fundamental feature of all the models considered here is that they are *stochastic* and *organism-centric*. Because these are such fundamental aspects, they merit discussion in some depth.

1.2.1   Heuristic of movement analysis

Throughout this dissertation, movement is considered from the reference frame of the moving organism as opposed to a fixed reference frame. A useful term for this distinction can be borrowed from fluid mechanics, in which models that consider changes that occur to a moving particle are termed *Lagrangian*, while models that consider the changes that occur at given points in space are termed *Eulerian*. Random walks, correlated random walks and Lévy flight models are essentially Lagrangian analyses, while home range analysis (Millspaugh and Marzluff, 2001; Moorcroft and Lewis, 2006) might be considered Eulerian since its primary concern is the characteristics and

the geographical area used by an organism. Though there is, in fact, much overlap between these two approaches, the basic difference between these types of analysis are the sorts of questions that are typically asked: Lagrangian models probe the detailed behavioral responses and mechanics of movement, while Eulerian analyses generally ask larger scale questions about spatial distributions and habitat usage of animals.

Environmental variables and behavioral responses can be related to a Lagrangian model of movement using by fromulating a very general heuristic model:

$$\frac{\partial \mathbf{S}(\mathbf{X}, t)}{\partial t} = f\left(\mathbf{S}(t), \mathbf{E}(\mathbf{X}(t))\right) \tag{1.1}$$

Here, $\mathbf{X}(t)$ represents the spatial location of the individual (whether in one, two or three dimensions) at time t, $\mathbf{E}$ is some vector of local environmental variables at location $\mathbf{X}(t)$ (e.g. temperature, food density, light availability, presence of predators), and the innocuous looking $f(\cdot)$ represents the behavioral response of the organism to environmental conditions and its state and $\mathbf{S}$ is a description of an organism's state. This variable is a function of extrinsic properties, notably (for our purposes) spatial location, but also any number of physiological properties, size, strength, hunger/satiation, energetic reserves, and also intangible ones like knowledge, memory and desire. Using mathematical jargon, *Behavior* can be viewed as an operator that transforms internal states and environmental parameters into new states. *Movement behavior* would be the component of the behavior operator that acts only on the change of physical location of the organism.

In the most simple model of movement, an organism is unaware of its absolute spatial position but makes its movement choices based purely on immediate environmental information. For example, a foraging organism that encounters a food item might change its movement pattern to intensify a food search by slowing down and increasing its turning rate, a fish might move more slowly in warmer waters, any diurnal organism spends some part of the night inactive. The step length is a function of the

environmental information, which is itself a function of the absolute location of the organism, while the intrinsic state is independent of absolute position. Displacement can be expressed as

$$\mathbf{V}(t) = \frac{\partial \mathbf{X}}{\partial t} = f(S(t), \mathbf{E}(X(t))), \qquad (1.2)$$

or, in the discrete case, the movement step can be expressed as:

$$\Delta \mathbf{X}_t = f(\mathbf{S}_{t-1}, \mathbf{E}(X_{t-1})) \qquad (1.3)$$

This general model suggests a framework within which information about movement from the Lagrangian perspective of the organism can be interpreted in terms of Eulerian knowledge of the environment. Once positional data, $\mathbf{X}_t$, is decomposed into actual parameters of movement, such as step lengths R_t and turning angles θ_t, and the environmental history $\mathbf{E}_t$ is obtained, the specific analysis needs to be tailored to the study organism, the relevant question at hand and to the quality of the available data.

1.2.2 Objectives and structure

This dissertation does not, in fact, contribute concretely to inferences on the behavioral operator presented above. However, its broad goal is to contribute to the development of an understanding of the movement models themselves, whether complex and behaviorally heterogeneous or diffusive simplifications. All the movements in the dissertation are ultimately parameterized by a Lagrangian, stochastic model of movement. Basically, much concern is given to modeling and parameterizing $\mathbf{\Delta X}(t)$, whatever the application.

The dissertation is divided into two parts. Part 1 (Chapters 2 and 3) discusses and develops mathematical parameterizations of movement at the most fundamental level with emphasis on estimating relevant parameters from movement data. Some

of the ambiguities and shortcomings of existing methods are addressed, refinements and alternatives are proposed, and practical examples are provided.

Chapter 2 discusses, dissects and decomposes the correlated random walk model. In most applications, the CRW assumes *independent* distributions of move lengths and move angles that describe an organism's path. The "correlated" nature of the CRW is captured by turning angles that are clustered around zero degrees, but are not, in fact, autocorrelated themselves. The consequences of this omission for analyzing movement tracks are discussed and presented. A further drawback of the CRW is that the parameters are only meaningful when the data is "regular", i.e. the time interval of sampling is constant. However, real data is often very irregular, particularly in marine environments. I explore what happens to the parameters of a CRW under irregular samplings and suggest a parameterization in terms of a time-scale of loss of correlation, illustrating the method with data on dugong (*Dugong dugon*) movements in Western Australia. Finally, I suggest that the two fundamental parameters of homogeneous random movement are the length and time scales at which correlation is lost, and present a method to determine these scales. In later chapters, these scaling parameters are shown to be largely sufficient for predicting larger-scaled ecological processes such as dispersal, migration and encounter rates.

In Chapter 3, I present in detail a continuous, stochastic, auto-correlated model of movement originally developed by Dunn and Brown (1987) and Alt (1990). The model is versatile, capable of capturing advective movements, helical trajectories, and turning persistence, all features which traditional CRW models are incapable of capturing. The parameters of the model are readily interpretable in terms of temporal and spatial scales of autocorrelation and turning radii. Efficient methods of estimating the parameters are developed and applied to movements of unicellular algae.

Part II (Chapters 4, 5 and 6) presents diverse ecological applications of various ideas developed in the Part I. Chapter 4 is devoted to mathematical predictions of encounter rates, whether between predator and prey or foragers and search targets.

Under certain assumptions, encounter rates can be derived entirely from characteristic time and length scales of movement presented in Chapter 2. The one-dimensional encounter rate model is applied to data on survival of migrating juvenile salmon in the Columbia River basin. Finally, the two-dimensional case is discussed with some general applications about the movement scales which optimize foraging in two dimensional environments.

In Chapter 5, I develop a comprehensive method of interpreting animal movement data in terms of discrete behavioral switches. Movement is parameterized in terms of means, variance and autocorrelations of persistence and turning components. Robust statistical methods to identify discrete shifts in the values of these parameters are developed and tested. The method is notably geared to appropriate interpretation of gappy movement data, that is data that has been sampled at irregular intervals, as is particularly common in marine environments where the receipt and transmission of a signal relies on the organism being on the surface of the water. I apply the method to data of a GPS-tracked foraging female northern fur seal (*Callorhinus ursinus*) acquired in the Kuril Islands in Russia.

Chapter 6 departs from the analysis of individual track data to consider two other fundamental kinds of movement data: mark-recapture of a dispersing population and first-passage times of a migrating population. Beginning with the fundamental length and time scales of movement defined in chapter 2, the effects of a heterogeneous population on the observed distributions and travel times of moving organisms are derived in terms of a general population-heterogeneity framework. Parameters of the resulting distributions are estimated for dispersing blue head chub, migrating salmonids in the Columbia river and human runners in a marathon. I present a straightforward quantification of population level heterogeneity and provide a mechanistic explanation for the long-tails observed both in dispersal and first-passage time distributions.

I conclude in chapter 7 with some general thoughts about the study of animal movements, possible future directions in fundamental developments and relevance to

applied problems in ecology.

This work is essentially methodological. An attempt has been made to include all the necessary estimation and analysis routines, including annotated `R`-code in the appendices which should allow the reader to readily implement any of the analyses. Because of the wide scope of this dissertation, with objects of study ranging from microscopic alga to top predator marine mammals, from tropical waters in Australia to North American rivers, it is necessarily difficult to fully explore the practical consequences and nuances of any of the ecological systems considered. Nonetheless, it is hoped that the mixture of theoretical rigor, practical algorithms and concrete examples will provide some inspiration for attacking a variety of problems in movement ecology.

Part I

MATHEMATICAL MODELS OF MOVEMENT AND THEIR PARAMETERIZATIONS

[T]he type of motion presented by living organisms ... can be regarded as containing both a systematically directed and also a random component... The mathematical treatment of the statistical mechanics of the kind of systems here taken in view may appear to threaten formidable difficulties. It is hoped that this will not altogether prevent its attack.

Alfred J. Lotka (1924)

Chapter 2

THE CORRELATED RANDOM WALK: OVERVIEW AND EXTENSIONS

[But] it is on such false correlations that men found half their inferences...

<div align="right">

GEORGE ELIOT (1865)

</div>

2.0 Introduction

Actual animal movements are the result of a complicated behavioral and biophysical responses to internal states, environmental cues, and constraints. Because it is impossible to model all of these interactions, especially for mass movements, movements of organisms are often modeled stochastically, i.e. with intrinsically random velocities and orientations that can be summarized by well-defined probability densities and associated parameters. While actual movement contains many different behavioral modes, the discussion here is constrained to a single behavioral idealized "unit" of movement that is *homogeneous* and *stationary*, i.e. described by the same set of parameters independent of absolute spatial location and time respectively. Despite these simplifications, it is not immediately obvious how a continuous movement process should be modeled and parameterized, since movement is a multi-dimensional, combining both spatial and a temporal dimension, and auto-correlated process.

While a true consensus in appropriate models still eludes movement ecologists, an increasingly widespread general description of homogeneous unoriented random movements can be broadly classified as the "correlated random walk" (CRW). The most common CRW is modeled as one where step-lengths r have some non-negative,

right-skewed distribution and turning angles θ are symmetrically clustered around zero degrees (Patlak, 1953b; Shigesada, 1980; Kareiva and Shigesada, 1983; Marsh and Jones, 1988; Bergman et al., 2000; Morales et al., 2004; Bartumeus et al., 2008, others). Widely used example of a circular distribution for turning angles are the wrapped normal (WND) von-Mises (VMD) and wrapped Cauchy distributions (WCD) (see Fisher (1993) for a review of circular distributions). Most recently, attention has been paid to a circular distributions known as the Jones-Pewsey distributions which generalize wrapped normal and Cauchy and other distributions into a single family (Jones and Pewsey, 2005; Bartumeus et al., 2008).

Here and in subsequent chapters I will be primarily applying the Weibull distribution for step-lengths r and the wrapped Cauchy distribution for turning angles θ (see figure 2.1 for examples of trajectories and distributions). The Weibull distribution is a positive unimodal distribution whose density function is given by

$$R \sim f(r) = \frac{a}{b} \left(\frac{r}{b}\right)^{a-1} \exp\left[\left(-\frac{r}{b}\right)^a\right] \tag{2.1}$$

with shape parameter a and scale parameter b. Several special cases for the shape parameter merit discussion. The case when $a = 2$ is the Cauchy distribution

$$f(r|t) = \frac{2r}{b\sqrt{t}} \exp\left(\frac{-r^2}{bt}\right) \tag{2.2}$$

which is the solution for the distribution of an unconstrained two-dimensional diffusion process where the diffusion constant $D = (b/2)^2$. As a approaches infinity, the Weibull distribution becomes a delta function at value b. If $a = 1$, (2.2) is the exponential distribution (the spatial Poisson process), and the resulting movement corresponds to jumps between neighboring points randomly distributed in space.
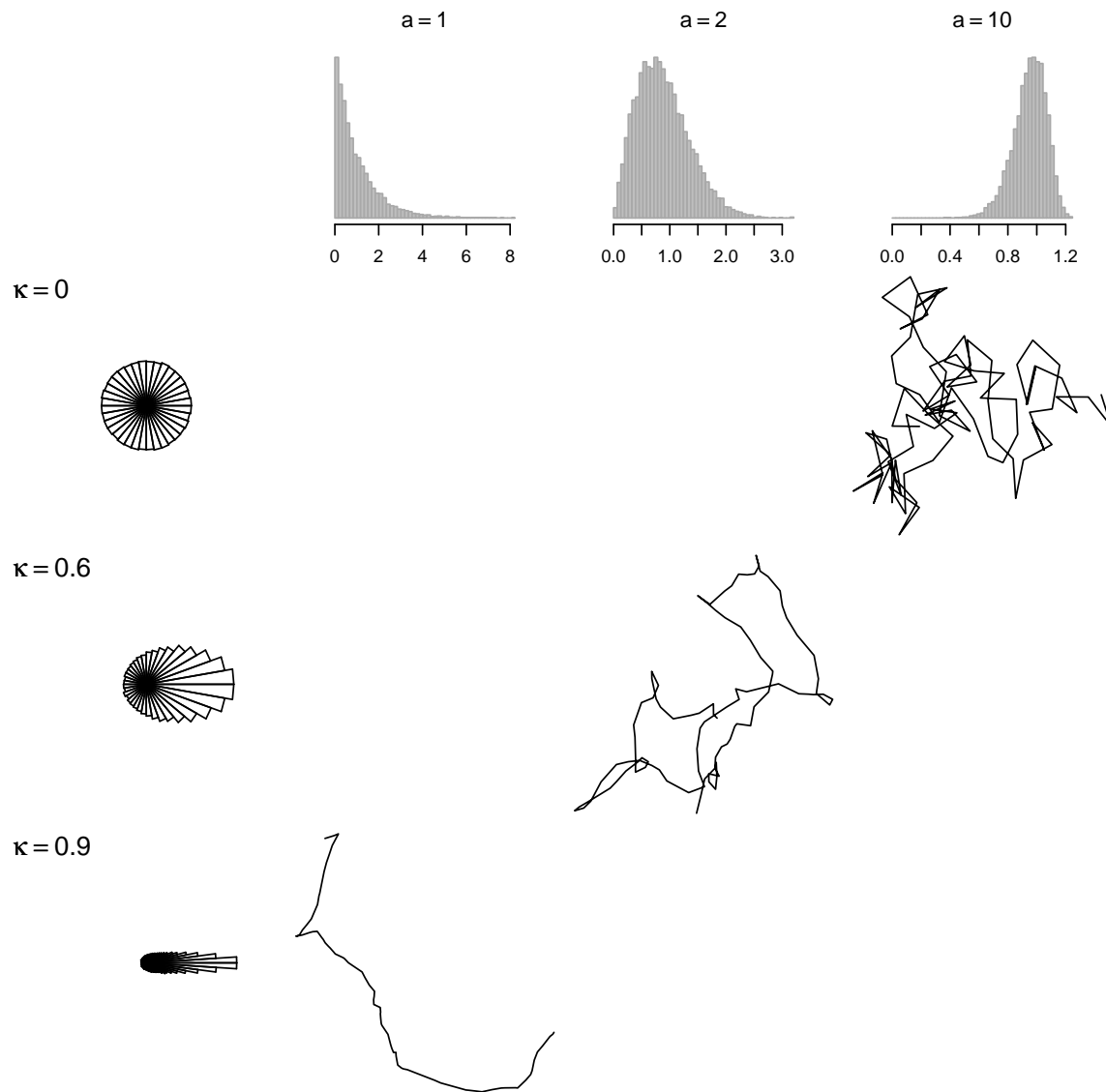
Figure 2.1: Examples of correlated walks along with rose diagrams of wrapped Cauchy distributions for turning angles (along left) and Weibull distributions for the step lengths (along top).

The mean and variance of the Weibull distribution are:

$$\mathrm{E}\,[R] \;\; = \;\; b\Gamma(1+1/a) \tag{2.3}$$

$$\mathrm{Var}\,[R] \;\; = \;\; b^2[\Gamma(1+2/a) - \Gamma(1+1/a)^2] \tag{2.4}$$

where Γ is the gamma function. These moments can be used to obtain moments of methods estimators (MME's) from the sample mean and variance. Maximum likelihood estimates can also be obtained, though they require iterative procedures (Gove, 2003).

The wrapped Cauchy distribution is defined over the interval $(-\pi, \pi)$ as follows

$$\Theta \sim g(\theta) = \frac{1}{2\pi} \left(\frac{1 - \kappa^2}{1 + \kappa^2 - 2\kappa\cos(\theta - \mu)} \right) \tag{2.5}$$

where μ is the mean direction and $\kappa \in \{0, 1\}$ is a clustering parameter: When $\kappa = 0$, the angles are distributed uniformly between $-\pi$ and π and when $\kappa = 1$, the distribution is a delta function at $\theta = 0$. Uniformly distributed turning angles correspond to a true random walk (RW), or drunkard's walk, or Pearson's walk[1]. Clustering greater than zero between successive steps give a movement some directional persistence (figure 2.2B).

An important feature of the WCD is that the clustering parameter κ is the ex-

---

[1] The study of random walks can be said to have been fathered by Karl Pearson (1857-1936), who posed the following problem in a brief letter in *Nature* (Pearson, 1905):

> A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r + dr$ from his starting point, O.

An approximate solution provided by Lord Rayleigh of, essentially, a two-dimensional Gaussian distribution around the origin prompted the following quip by Pearson:

> The lesson of Lord Rayleigh's solution is that in open country the most probable place to find a drunken man who is at all capable of keeping on his feet is somewhere near his starting point!

pectation of the cosine of the angles: $\kappa = \langle \cos(\theta) \rangle$; making κ an intuitive and readily estimable parameter. This property also dovetails nicely with a fundamental result in the CRW literature by Kareiva and Shigesada (1983): The expected square of the displacement (MSD) after n steps of a CRW can be written out in terms of κ and the moments of the step length distribution as

$$\left\langle R^2(n) \right\rangle = n \left\langle r^2 \right\rangle + 2 \left\langle r \right\rangle^2 \frac{\kappa}{1 - \kappa} \left(n - \frac{1 - \kappa^n}{1 - \kappa} \right) \tag{2.6}$$

For $\kappa = 0$, the second term drops off and the MSD becomes simply $\langle R^2(n) \rangle = n \langle r^2 \rangle$, which is the familiar relationship of a true random walk[2] that the variance of a cumulative independent random process increases linearly with time. For non-zero values of κ, at the long-time limit of large n the MSD becomes

$$\lim_{n \gg 1} \left\langle R^2(n) \right\rangle = n \left(\left\langle r^2 \right\rangle + \frac{2\kappa - 1}{1 - \kappa} \left\langle r \right\rangle^2 \right) \tag{2.7}$$

which is also proportional to the length of the walk. Thus, a CRW will eventually approximate standard diffusion-like dispersal at a long enough time-scale. This is an expected result of the essential Markovian property of all homogeneous random walks. Deviations from the Gaussian approximation are only relevant at time-scales where the term $\frac{1 - \kappa^n}{1 - \kappa}$ is on the order of n.

The MSD is analytically solvable for CRW's, and therefore a commonly used property in analyses. However, the actual expected displacement, which is often a parameter of greater interest, is often difficult to solve for analytically. Bovet and Benhamou (1988) suggest that for a reasonable amount of steps ($N > 10$) a good estimate of the actual expected net displacement is

$$\langle R(n) \rangle \sim \sqrt{\frac{\pi}{4} \langle R^2(n) \rangle} \tag{2.8}$$

---

[2]or of Brownian motion, or of diffusion approximations, or even of the central limit theorem

This approximation gives good predictions under assumptions of homogeneous, "independent" CRW's, and will be referred to later.

2.1 Is the correlated random walk really correlated?

A fundamental assumption of the correlated random walk as it is most commonly applied is that the step-lengths and turning angles are independent. The "smoothness" of the walk is accounted for by angles that are clustered around zero, but the fluctuations around a directed orientation, while small, oscillate independently between positive values and negative values. The radius of the curvature is limited by the step-length. Real animal movements, however, often display large radia of curvature (see, for example, fur seal tracks in figures 5.5), which cannot be accounted for by a classic CRW model. Furthermore, at a high temporal resolution of sampling of a true smooth trajectory, turning angles and velocities will necessarily be auto-correlated due to rotational and linear inertia. Thus, a sampling of a true movement that may be well-modeled by a CRW at a particular time interval will not be a sub-sampling of another CRW, but of a truly correlated random walk (what we might call a TCRW).

There are any number of ways of modeling a truly correlated random walk; the only requirement is that the autocorrelation structure between subsequent parameters be well defined. Because turning angles are circular and velocities are positive and skewed, standard Gaussian autocorrelated models are, however, not appropriate. A simple model which explicitly autocorrelates the turning angles can be written out as follows:

$$\theta_i = \rho\,\theta_{i-1} + \epsilon_i \tag{2.9}$$

where ϵ are independent and identically distributed wrapped Cauchy random variables with clustering coefficient κ and $0 < \rho < 1$. In this parameterization, $\rho = \kappa = 1$ is

linear movement, $\rho = 0$ collapses to the CRW and $\rho = \kappa = 0$ is the uncorrelated RW.

Realizations of the TCRW are presented in figure 2.2. Note that the CRW (figure 2.2B) is somewhat more zig-zaggy, even at a high value for $\kappa = 0.9$, whereas at $\rho = 0.5$ (figure 2.2C) the path is somewhat smoother. At high values of ρ (fig. 2.2D), we observe several periods of tight looping, corresponding to periods where larger turning angles are auto-correlated, which the standard CRW is incapable of capturing.

One immediate impact of a TCRW is that the expected distances no longer conform with Patlak's expectation. I performed 1000 simulations apiece of TCRW with turning angles generated from equation 2.9 for values of ρ ranging from 0 to 1. The results (figure 2.3) indicate that as ρ approaches 1, the expected displacement drops from 25.83 (as predicted by Patlak's formula 2.7) by more than half to 10.97. These results suggest that auto-correlation in the turning angles can have a significant impact on the statistical nature of correlated random walk models.

A more mathematically sophisticated discussion of auto-correlation structures in continuous stochastic movement models is presented in chapter 3.

2.2 Dealing with irregular intervals: Estimating $\Theta(t)$ for gappy time-series

A further limitation to applying correlated random walk models to continuous movements is the fact that they are predicated on regular observations between intervals. It is, however, not meaningful to fit a CRW to gappy data, since neither the $R(t_i)$ nor $\Theta(t_i)$ are identically distributed when the interval between measurements changes.

Here, I explore approximations of the expected distribution of Θ as a function of irregular intervals. Again, the basic CRW model I consider is one in which Θ and R are independently drawn from the wrapped Cauchy and Weibull model respectively, such that:
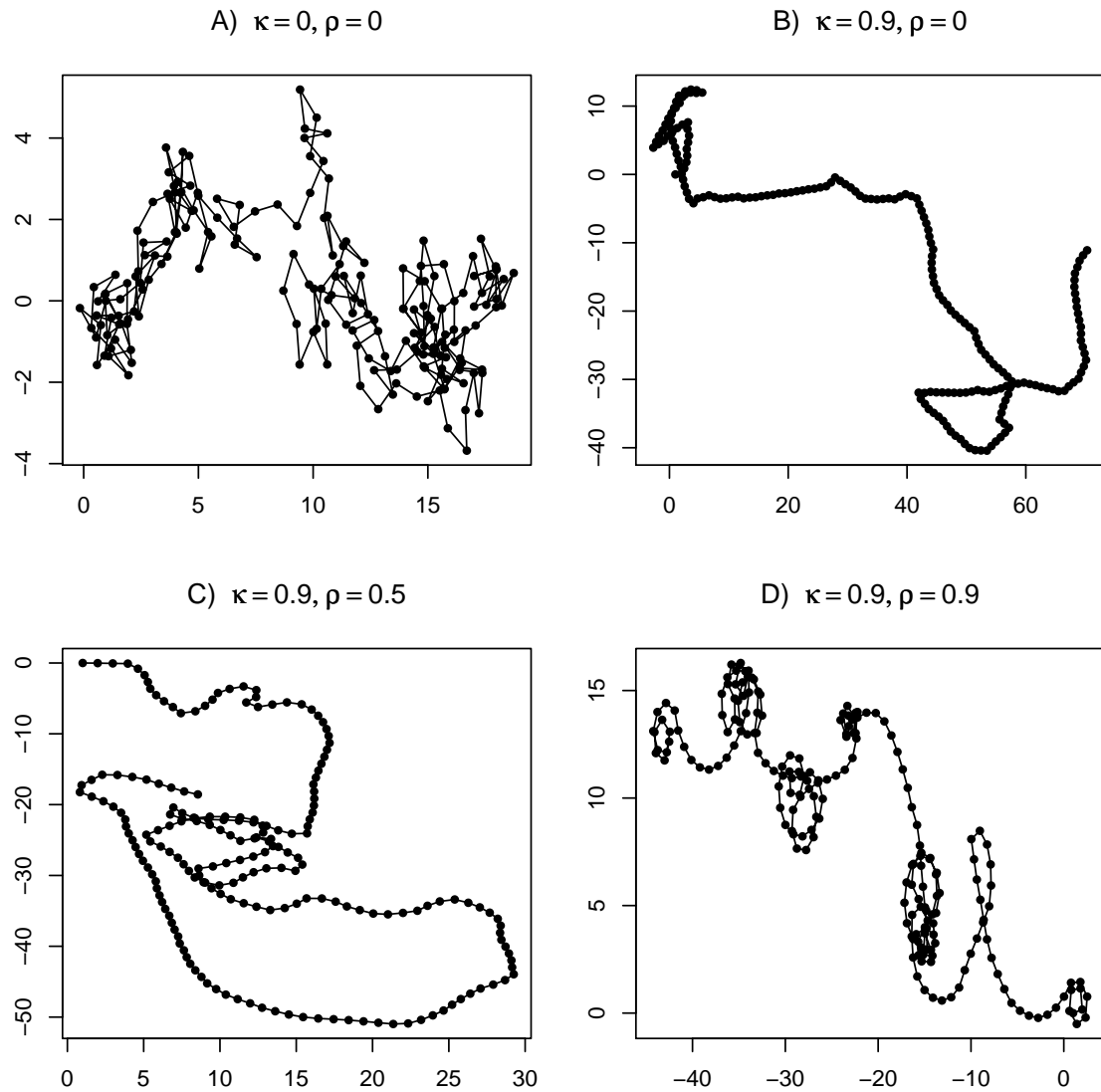
Figure 2.2: Examples of several kinds of stochastic models of movement. A) Uncorrelated random walk. B) Weibull-wrapped Cauchy random walk. C) True CRW (2.9) with $\kappa = 0.9$, $\rho = 0.5$. D) True CRW with $\kappa = 0.9$ and $\rho = 0.9$.
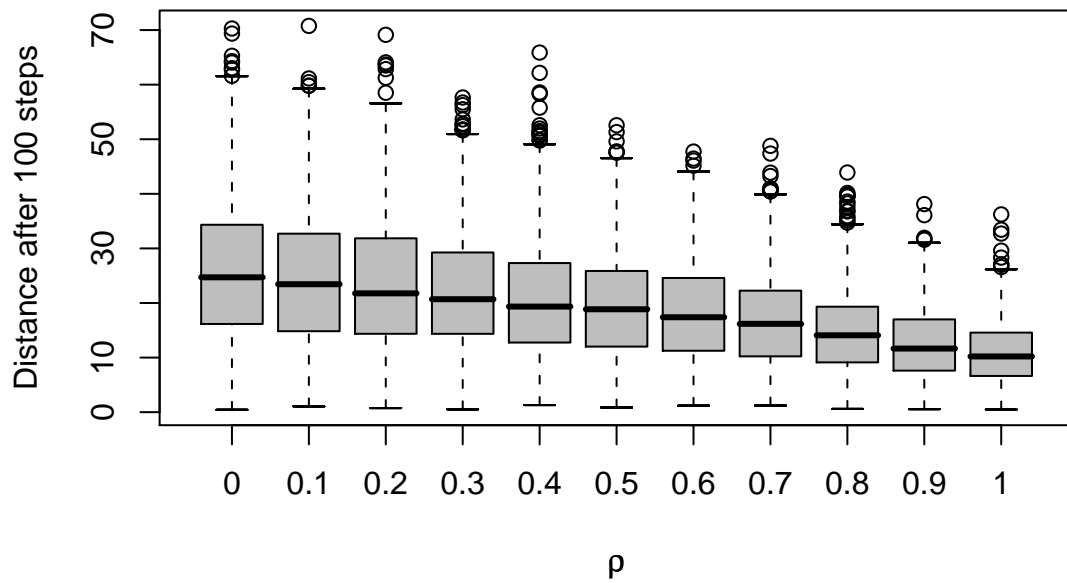
Figure 2.3: Expected displacements after 100 steps of a TCRW process (2.9) with $\kappa = 0.8$ and ρ values ranging from 0 to 1.

$$\Theta \sim f(\theta) = \frac{1}{2\pi} \left(\frac{1 - \kappa^2}{1 + \kappa^2 - 2\kappa \cos(\theta - \mu)} \right) \tag{2.10}$$

$$R \sim g(r) = \frac{a}{b} \left(\frac{r}{b} \right)^{a-1} \exp \left[\left(-\frac{r}{b} \right)^a \right] \tag{2.11}$$

2.2.1 Modified wrapped-Cauchy distribution

When we consider turning angles a problem with irregular measurements immediately arises: at very small time intervals, the probability that the organism will continue to move in the previous direction is very high and κ is close to one. At long time intervals, the tendency to continue in the same direction is closer to zero (e.g. an organism is typically very little affected at any given moment by its orientation months before a given moment). The κ parameter is therefore a continuous function depending on the time between measurements (τ). In order to find the relationship between the effective κ and the interval of measurement, I performed ten thousand simulations of CRW curves with high values of κ_1 (so denoted to reflect the fact that it is the true clustering coefficient at a time interval of 1), subsampled these at every possible interval $\tau \in \{1, 30\}$, obtained values of $\widehat{\kappa}_\tau$ and performed linear regressions of $1/\widehat{\kappa}_\tau$ against τ. The regression yielded the following semi-empirical relationship

$$\kappa(\tau) = \frac{1}{\left(\frac{1}{2}(1 - \kappa_1)\tau + 1 \right)^2} \tag{2.12}$$

This relationship provides excellent fits to the simulation results (figure 2.4). According to equation 2.12 the measured κ drops as expected from a value of 1 at $\tau = 0$ to zero, suggesting that a "characteristic" time-scale of the decrease might be a parameter of interest, especially since actual irregularly collected animal movement data does not lend itself to an immediately discretizable, "natural" value of τ_1. Thus, if we define a parameter $\tau_{1/2}$ such that $\kappa(\tau_{1/2}) = 0.5$, equation 2.12 can be reexpressed

as

$$\kappa(\tau) = \frac{\tau_{1/2}^2}{\left((\sqrt{2}-1)\tau + \tau_{1/2}\right)^2} \tag{2.13}$$

and the complete distribution for turning angles is

$$f(\theta, \tau | \tau_{1/2}) = \frac{1}{2\pi} \left(\frac{1 - \kappa(\tau)^2}{1 + \kappa(\tau)^2 - 2\kappa(\tau)\cos(\theta - \mu)} \right). \tag{2.14}$$

where $\kappa(\tau)$ is a function of $\tau_{1/s}$ as in (2.14). I will refer to this distribution as the *adjusted wrapped Cauchy* distribution. The $\tau_{1/2}$ parameter is the first of several possible indices of characteristic time-scales of movement that I introduce in this dissertation. R-code for generating the distribution (2.14) and for simulating random draws from it are provided in appendix A.1 and A.2.

2.2.2 *Estimating parameters*

The existence of a well-defined form for the distribution of turning angles as a function of the gap interval allows for ready application to maximum likelihood methods for estimating a value for $\tau_{1/2}$ from data in the form $\{X_i, Y_i, T_i\}$, where $i \in \{1, 2, ..., n\}$. We reduce this dataset to turning angles Θ_i (see section 3.3 for the most efficient method to obtain turning angles from spatial data) and gaps $\tau_i = T_i - T_{i-1}$. Then the likelihood of parameter $\widehat{\tau}_{1/2}$ given the data is

$$L(\tau_{1/2} | \Theta, \tau) = \prod_{i=1}^{n} f(\theta, \tau | \tau_{1/2}), \tag{2.15}$$

where $f(\cdot)$ is the adjusted wrapped-Cauchy distribution (2.14). This is smooth, well-behaved function of $\tau_{1/2}$ and can readily be maximized numerically (see R-code in appendix A.3).
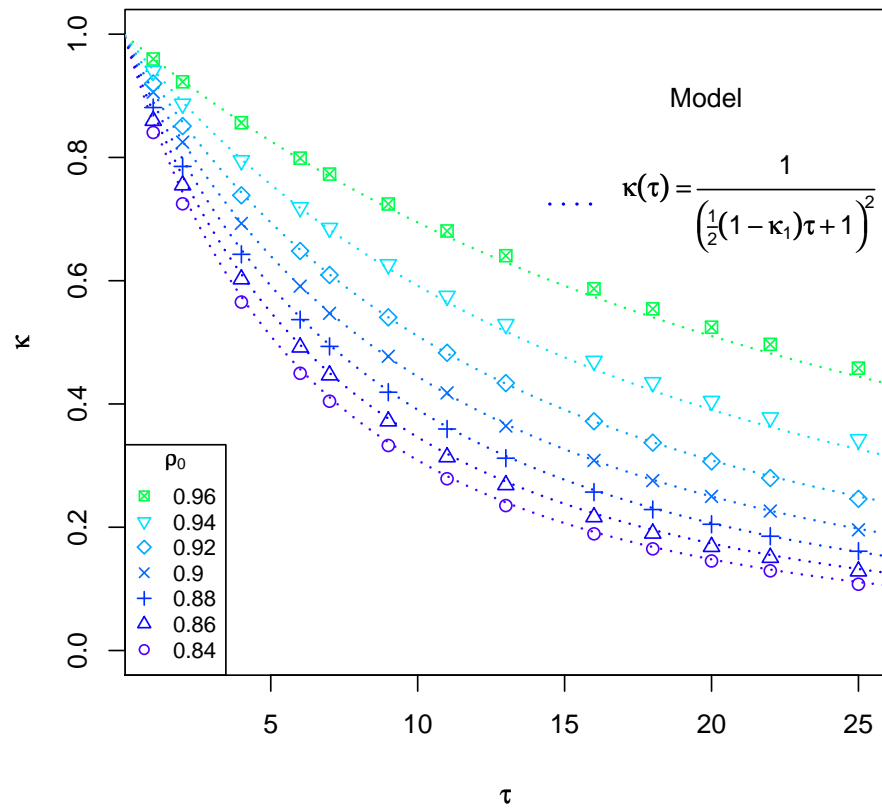
Figure 2.4: Comparison of simulation results and semi-empirically fit model for estimated values of κ at different values of κ_1 ranging from 0.84 to .98
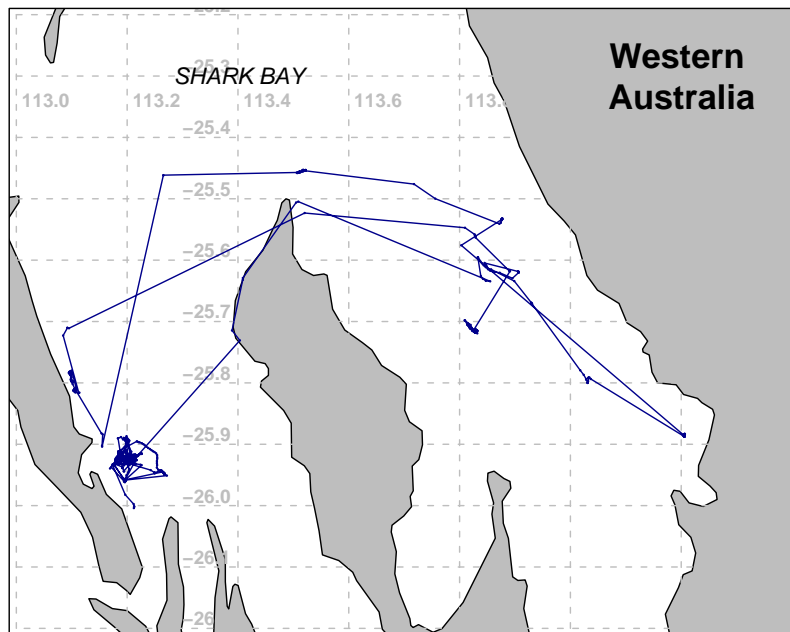
Figure 2.5: Movement track of dugong D0606 in Shark Bay, Western Australia.
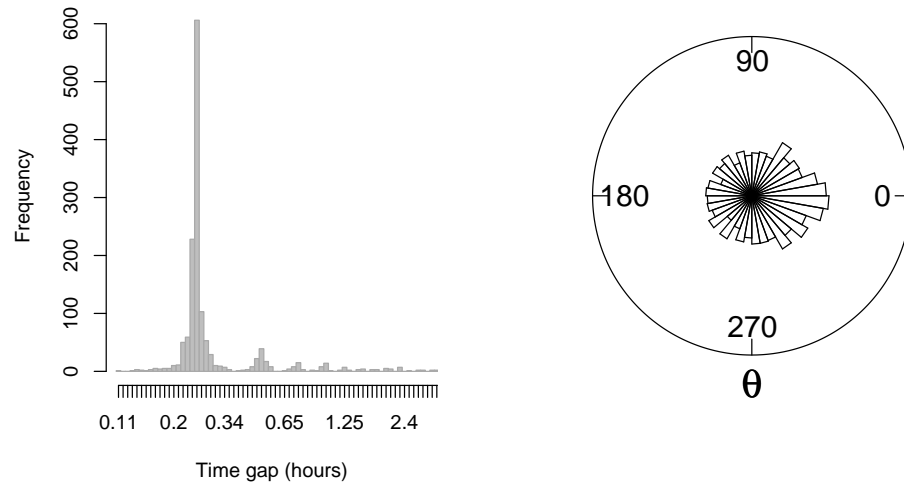
Figure 2.6: Histogram of time intervals between measurements and rose diagram of turning angles for D0606 track.

2.2.3 Example with dugong (Dugong dugon) movements

The dugong (*Dugong dugon* Müller, 1776) is a species of marine mammal in the order *Sirenia* which also includes sea cows and manatees (*Trichechus spp.*) It is the smallest of the sirenians, measuring usually less than 3 m. and is distinguished by a fluked rather than paddle-shaped tail. Once found throughout the Indian Ocean and tropical South Pacific, the dugong's populations have been greatly reduced and fragmented, primarily due to hunting and habitat degradation. Dugongs are associated with shallow, protected near-shore areas and spends most of their time foraging on various species of seagrass (de Iongh et al., 1995). One of the healthier dugong populations resides in Shark Bay, a shallow (6-15 m) semi-enclosed bay on Australia's west-central coast (25^oS, 113^oE) with the most extensive seagrass shoals in the world (Marsh et al., 2005).

From 2000 to 2002, dugongs have been captured, instrumented with GPS tags

and released. Their locations (latitude and longitude to millisecond resolution) have been subsequently monitored remotely, and at a later date some of the instruments were retrieved for additional data. The tags floated on the surface of the water, towed by the animal, and emitted a signal every 15 minutes identifying the latitude and longitude. For more details on the tagging methods and data see Holley (2006). Here, I discuss a track drawn by a single dugong, referred to in this report as D0606, that was tracked from June 18 to August 9, 2002 (see Figure 2.5).

Due to variable satellite exposure, linkage delays and deeper diving behavior, the resulting data is very gappy. D0606 was covered for 52.5 days or 5040 possible 15 min. periods, but only 1471 positions were registered. Time intervals between measurements therefore ranged from a few minutes to over a day, with roughly 90% under 45 minutes and 75% under 16 minutes (see figure 2.6A). The turnings angles are weakly clustered around zero degrees (figure 2.6A).

Since a straightforward estimation of a wrapped Cauchy parameter ($\widehat{\kappa} = \langle \cos(\theta_i) \rangle$ is relatively meaningless, the identification of the half-clustering time is a useful parameter for characterizing the time-scale at which the movement loses its autocorrelation. For dugong D0606, the resulting value for the estimate is 9.55 minutes, suggesting that if positions were samples *exactly* every ten minutes, the estimate of the clustering coefficient would be around 0.5. From (2.13), we calculate that the expected clustering at 1 minute intervals would be a high 0.92, whereas as a 1 hour sampling rate, $\widehat{\kappa}$ drops to 0.077, i.e. undetectable clustering in the turning angles. Thus, we can generalize the movement of the dugong to say that it's general persistence or "memory" or movement is on the order of tens of minutes.

In reality a dugong's movements are highly heterogeneous. There are periods when it transitions from one feeding ground to another, in which case the persistence scale is much higher, as opposed to the periods when it is feeding and making much slower, more random movements. The separation of behavioral modes within a single movement track is the topic of chapter 5. Within a relatively homogeneous move-

ment period, however, the half-time of persistence $\tau_{1/2}$ is an excellent indicator of the temporal scale of the animals directional changes, whereas the traditional κ has essentially no meaning for irregularly sampled data.

2.3   The mu-sigma-tau (MST) walk

There is much discussion in the literature as to whether actual movement data constitutes a true random walk (RW) or a correlated random walk (CRW), how to characterize the "sinuosity" or "tortuosity" of a walk, and what impact different models have on expected dispersal distances (Skellam, 1951; Patlak, 1953b; Shigesada, 1980; Kareiva and Shigesada, 1983; Byers, 2001; Benhamou, 2004). However, much of this debate seems somewhat spurious, since almost all movements are necessarily correlated at a small enough time scale and uncorrelated at a large enough time-scale. Exceptions to this rule are processes that are inherently discrete, such as in the original application of the CRW in the analysis of moth movements between flowers by Kareiva and Shigesada (1983) or of elk moving between snow craters (Fortin et al., 2005b).

It is generally not meaningful to discuss whether a movement itself is "correlated", rather, *data* acquired from an organism may or may not be correlated depending on the temporal resolution of the data acquisition compared to a time scale of independence (τ^*). A given time scale of independence is associated with a characteristic distance scale (σ^*) which is a measure of the typical random component of distance that a moving organism covers in time τ^*. A third parameter, μ^*, characterizes the advective component of a movement. I make the assertion that for any homogeneous random walk, the identification of these time and length scales is sufficient to determine such larger scaled ecological processes as encounter rates (chapter 4) and dispersal rates (chapter 6), regardless of the specific "nature" of the appropriate correlated random walk model.

2.3.1 Formal definition of an MST-walk

Most generally, a generalized homogeneous[3] Markovian movement process is defined as a continuous time process, $\mathbf{Z}(t)$, that can be discretized such that at at some interval τ

$$\mathbf{Z}(t + \tau) = \mathbf{Z}(t) + \mathbf{X}(\tau) \tag{2.16}$$

where $\mathbf{Z}$ is the *step distribution*, a random variable with some constant mean vector μ and variance-covariance matrix $\mathbf{\Sigma}^2$. The time interval τ at which subsequent $\mathbf{X}$'s are fully independent, such that $\text{Cov}[X(t), X(t + \tau)] = 0$, is defined as the time scale of independence.

In two dimensions, for example, the process defined as $\{W_x(t), W_y(t)\}$ would be fully characterized by six parameters of step length: μ_x, μ_y, σ_x^2, σ_y^2, σ_{xy} and the temporal independence scaling parameter τ. The case where $\mu_x = \mu_y = 0$ can capture the statistical behavior of any correlated random walk, since there is no absolute trend in orientation to typical CRW movements. In that case, τ is merely sub-sampling interval at which turning angles become uniformly distributed. If the mean parameter is different from zero, then the movement has an advective trend the velocity of which is $\langle V \rangle = \mu/\tau$.

The movement defined in equation 2.16 will be referred to as an MST-walk, since it is fully defined by the parameters μ, σ and τ. Note that both μ and σ have units of distance.

One immediate application of this parameterization is in the identification of diffusion parameters. The "biodiffusion" model, which has been widely used to describe either distribution of large numbers of independently moving organisms of probability distributions of displacement for a single randomly moving individual (Skellam, 1951; Zabel, 1994; Turchin, 1998; Okubo and Levin, 2001; Ovaskainen, 2004, others). The

---

[3]It would be more technically correct to use the time-series jargon: "stationary" rather than "homogeneous", but I sacrifice the obvious benefits of precision in order to avoid a distressingly oxymoronic discussion of "stationary movements".

simplest form of one-dimensional, homogeneous, unconstrained diffusion of a group of organisms all leaving the origin at a time 0 with mean velocity v suggests that their spacial distribution at time t is given by

$$f(x|t) = \phi(v\,t, Dt/2) \tag{2.17}$$

where $\phi(\mu, \sigma^2)$ is the Gaussian distribution and D is a diffusion parameter with units of distance^2 over time. The diffusion parameter describes the rate at which the spatial distribution of the population increases with time. Equation (2.17) is often derived from the solution of Fick's equation which integrates the flux of infinitesimally moving Brownian particles described in partial differential equation form. Diffusion parameters are rarely defined explicitly in terms of the parameters of individual movement. The MST parameterization and the central limit theorem, however, explicitly suggest that a cumulative number of independent MST-steps yields

$$f(x|t) = \phi\left(\frac{\mu}{\tau}t, \frac{\sigma^2}{\tau}t\right). \tag{2.18}$$

This derivation and parameterization is biologically highly intuitive, and contributes to the encounter rate theory presented in chapter 4 and derivations of migration and dispersal presented in chapter 6.

2.3.2 Estimating parameters

There are presumably as many ways to identify the MST parameters as there are ways to model stochastic movements. The half-time of clustering under the correlated walk model discussed in the previous section is one example. In chapter 3, I present a mathematically more sophisticated method by exploring the continuous autocorrelation structure of a complex track with oscillatory components. The persistence and turning autocorrelation calculated for foraging northern fur seals in 5 is related

to another such index. Here, I present a method that is based on a semi-empirical analysis of subsamplings of a CRW.

Consider data separated into turning angles $\mathbf{\Theta}$ and step lengths $\mathbf{R}$. Subsampling this walk at various intervals τ yields sets of step lengths $\mathbf{R}_\tau$ and turning angles $\mathbf{\Theta}_\tau$. (It should be noted that a set of these gapped values can be obtained beginning at any of the first τ points, such that the length of these subsample vectors is $N - \tau$.) Maximum likelihood estimates of the clustering coefficients $\widehat{\kappa_\tau}$ and step-lengths $\widehat{R_\tau}$ at different intervals can be obtained using methods described in Fisher (1993). At a high gap interval, the turning angles will be impossibly to distinguish from uniformity. The size of that interval, τ^* and the corresponding expected step-length σ^* are the time and length scales of independence.

An example of the application of this methodology is presented in figure 2.7. A smooth curve is generated with constant step length $l = 1$ and clustering parameter $\kappa = 0.9$ (fig. 2.7A). The curve is sampled at various intervals τ as described above. In order to determine the point at which any correlation is lost, I generated an empirical null-distribution of κ-estimates by simulating 1000 datasets from a uniform angular distribution ($\kappa_0 = 0$) of length N/τ and obtaining 5% and 95% quantile envelopes around estimates of $\hat{\kappa}_0$ (fig. 2.7B). In the example, the data-derived $\bar{\kappa}$ crosses the envelope at $\tau^* = 21$, indicating that at that interval length the turning angles are no longer meaningfully clustered. The corresponding empirical step length is $\sigma^* = 14.4$, or, using the Bovet-Benhamou approximation, $\sigma^* = 13.7$ (fig. 2.7C).

I repeated this process for a range of clustering coefficient values $0 < \kappa < 0.98$, simulating 20 walks at each value and obtaining estimates and scatters for τ^*. The resulting relationship is plotted in figure 2.8. At $\kappa = 0$, i.e. the uncorrelated walk, the expected value for τ^* is 1, while at $\kappa = 1$, i.e. the linear, directed movement, $\tau^* \to \infty$. An empirical equation that fits the data well is

$$\tau^* = A \log \left(\frac{1}{1 - \kappa^2} \right) + 1 \tag{2.19}$$
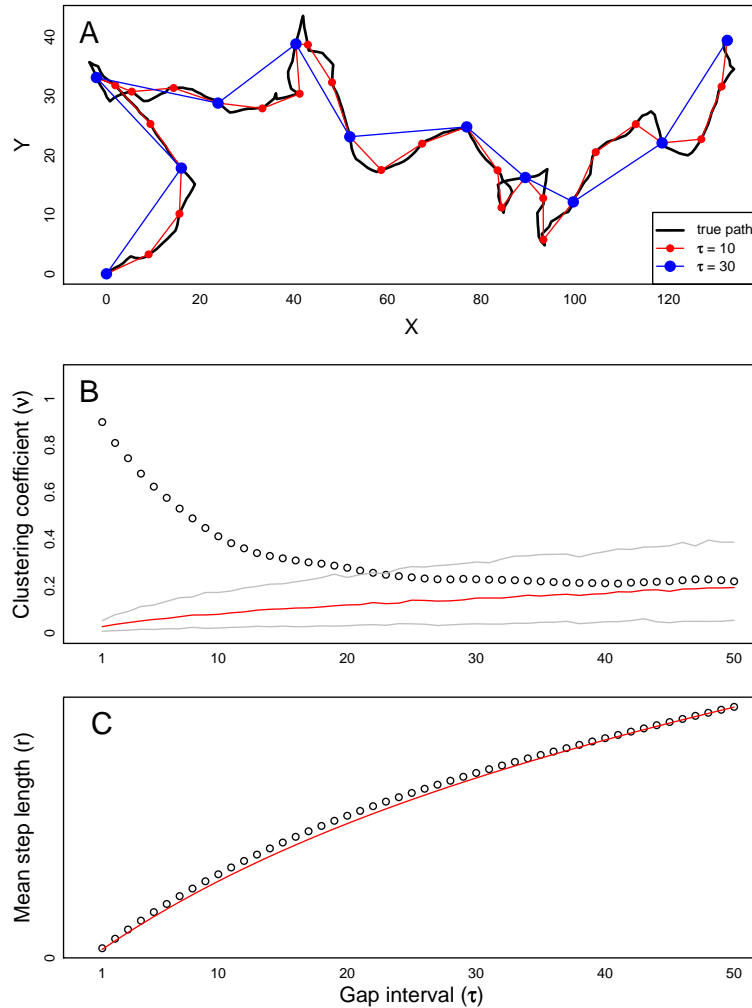
Figure 2.7:   Example of decomposition of smooth track. (A) represents a smooth CRW trajectory, with wrapped Cauchy turning angles with clustering coefficient $\kappa = 0.9$ and step lengths equal to one. Subsamplings of the data at intervals $\tau = 10$ and 30 are illustrated in red and blue respectively. While the first subsample still shows correlation, at $\tau = 30$ the walk can be considered truly random. Maximum likelihood estimates of κ against τ are shown in plot (B). The red and grey lines at the bottom are the median and 5% and 95% quantile envelopes of estimates of κ for data generated with no clustering ($\kappa = 0$). At the point where the κ estimates enters the envelope (approximately $\tau = 21$), the estimate cannot be considered significantly different than zero. Plot (C) shows the mean step length at each τ along with the Bovet-Benhamou prediction. At $\tau^* = 21$, the step length of independence $\sigma^* \sim 14.4$.

where A is found via regression on a linearization of the formula to be equal to about 13.71. This formula gives a rough idea of the relationship we might expect between the strength of the autocorrelation in a random walk and the expected time lag of independence.

2.3.3   Total distance of subsampled movement

When applying this subsampling to an true correlated movement, there is clearly some additional movement that is not accounted for in the non-correlated approximation. When calculating encounter rates, it is of some importance to know how much of the true path length is "lost" by the subsampling procedure. Denoting the total path-length at subsampling τ as L_τ, the ratio between expected subsampled pathlength and the total true path-length after N steps is

$$\frac{L_\tau}{L_1} = \frac{\frac{N}{\tau} \langle R(n) \rangle}{N \langle r \rangle}. \tag{2.20}$$

For high values of κ, this quantity drops relatively slowly against τ, while for lower values of κ there is a rapid drop (see figure 2.8); however, this loss is compensated for by the correspondigly much shorter time-scales of independence for less correlated walks. A superposition of the expected values for τ^* onto the distance ratio curves suggests that the ratio of distance traveled to total distance does not drop below 0.6.

While this is a semi-rigorous quantification of the impact of rescaling on the total distance swept out by the true path, the main message is that we can rescale with relative impunity. As long as the time and distance scale of independence is chosen as close as possible to the autocorrelation threshold, we can be confident that the bulk of the properties of the movement will be retained despite the simplification. Only a very coarse sub-sampling will lead to total distances that are less than half the true distance.
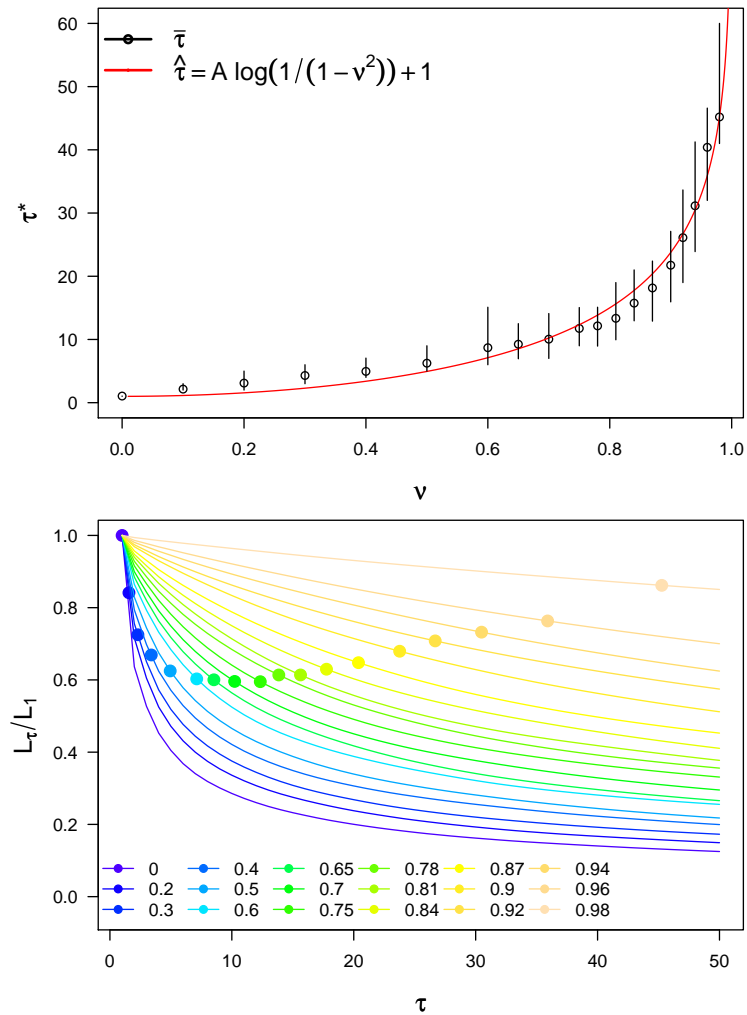
Figure 2.8: Simulation results exploring the relationship between clustering coefficient κ, estimated independence time-scale τ^* and the length ratio L_τ/L_1. In the plot, 20 trajectories were generated for each value of κ and values of τ^* were estimated using the method described in the text. The median values and 5 and 95% quantile values are plotted, as well as an empirically fit curve (in red). The lower plot shows the ratio of the subsampled path lengths (L_τ) compared to total path length for different values of κ, based on simulated paths of length 10000. The bold points represent the expected values of τ^* for each of the curves based on the results of the simulation above.

2.4 Conclusions

Lack of independence is one of the primary difficulties in analyzing movement data since it precludes application of many standard statistical analysis techniques. However, there is much potential information in the temporal and spatial scales at which movement data can be considered independent. These scales lead directly to a parameterization of diffusion approximations as well as predictions of encounter rates, which have consequences both for predation avoidance and successful foraging. Furthermore, these time and length scale are also more fundamentally *intrinsic* parameters of movement in the sense that they are independent of the frequency of measurement. The half-time parameterization of clustering angles and step lengths are useful not only because they deal with the issue of irregular sampling intervals, but because the parameters themselves have some biological meaning.

In this chapter, I presented a few examples of methods to estimate these scales from various kinds of data. Others are implicit in the parameterizations of movement presented further. While comparison of the various methods awaits further research, the theme of scales of independence and autocorrelation is a recurrent one throughout this dissertation. The MST-walk in particular will be referred to on several occasions. A general recommendation can be made to researchers of animal movement to embrace rather than avoid the issue of lack of independence, since identifying the intrinsic scales of independence can provide considerable insight into movement processes.

Chapter 3

CONTINUOUS STOCHASTIC AUTO-CORRELATED MODELS OF MOVEMENT

[T]he complexity of movement data has often led researchers to seek out methods that avoid its complexities rather than embrace them.

PATTERSON ET AL. (2008)

3.1 Introduction

A fundamental feature of physical movement is that it occurs in two or more spatial dimensions and in time. The CRW model considers the two polar dimensions of discretized steps (step length and turning angle), but does not account for any relationship between the two or, as we have seen, for any kind of temporal structure beyond independent samplings from appropriate distributions. In this section, I present some mathematically convenient methods of modeling movements as continuous stochastic processes, highlighting the natural way in which persistence and turning velocities can be auto- and cross-correlated. I also present ways to discretize, simulate, and estimate the necessary parameters for such processes.

I use *complex vector notation* to model continuous movement. Complex numbers are ordered pairs of numbers, separated into *real* and *imaginary* components. Complex numbers map onto two-dimensional space and can therefore represent two-dimensional vectors. A variety of straightforward arithmetic and algebraic operations (addition, multiplication, integration, etc.) have ready geometric interpretations. In short, they are an efficient and natural mathematical framework for describing two-dimensional processes like movement.
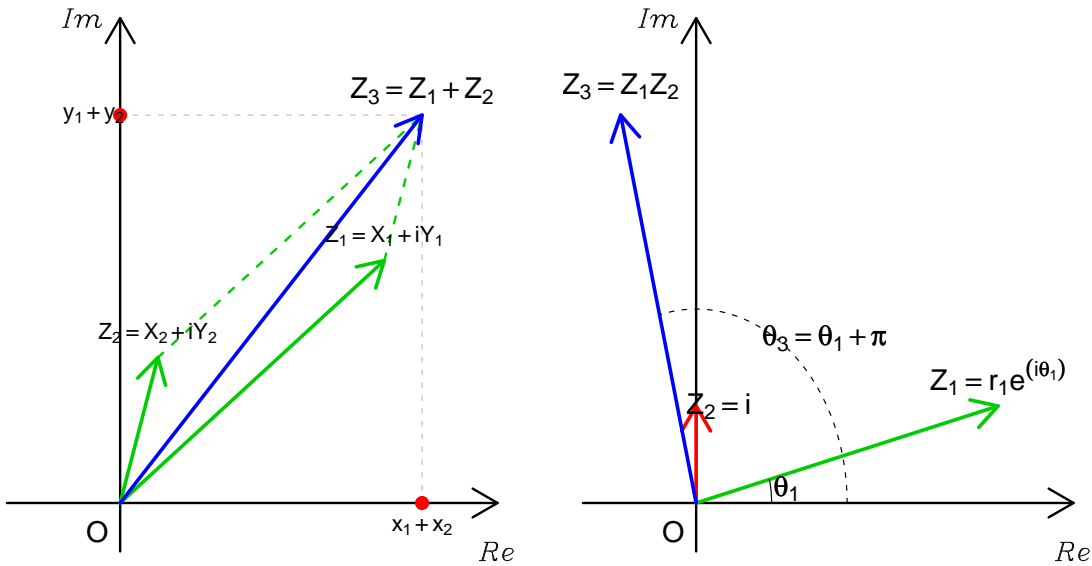
Figure 3.1: Geometric representation of complex numbers, their addition and multiplication.

Complex variables were most notably used to describe biological movement by Dunn and Brown (1987) and Alt (1988, 1990), who developed a complex autocorrelation analysis for modeling continuous stochastic movements with applications to various single-celled organisms. However, these methods do not appear to have caught on widely, especially among investigators of larger organisms. The interested reader is encouraged to study the original papers.

3.2   Brief review of complex notation

(N.B. The following is a brief review of complex numbers and their manipulation. Readers comfortable with complex analysis can proceed directly to section 3.3.)

Complex numbers are written in the form $z = x + iy$ where x is the *real part* and y is the *imaginary part* (since it is multiplied by the *imaginary unit*: $i = \sqrt{-1}$). These components are often notated as $x = \Re(z)$ and $y = \Im(x)$. Complex numbers can be

visualized conveniently as vectors in the *complex plane*, where the x and y axes are the real and imaginary axes respectively (figure 3.1). A complex number z can also be summarized by the polar quantities r, also denoted $|z|$ and referred to as the *modulus*, and θ, known as the *argument*. Geometrically, these correspond to the distance from the origin and the angular deviation from the positive x-axis respectively, and are related to x and y via the standard Cartesian-polar relationships:

$$r = \sqrt{x^2 + y^2} \tag{3.1}$$

$$x = r\cos(\theta) \tag{3.2}$$

$$y = r\sin(\theta). \tag{3.3}$$

A complex number can be expressed in terms of the modulus and argument as: $z = r\exp(-i\,\theta)$. Thus, the real complex number $Z = r$ can be expressed as $Z = r\exp(0)$ and the imaginary complex number $Z = ir = r\exp(i\pi)$.

The sum of two complex numbers $z_1 + z_2$ is defined as $(x_1 + x_2) + i(y_1 + y_2)$ and is readily visualized as the sum of two vectors. The product is given by

$$z_1 z_2 \;=\; (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) \tag{3.4}$$

$$\;=\; r_1 r_2 \exp{(\theta_1 + \theta_2)}i. \tag{3.5}$$

It is perhaps most clear in the polar notation that the product of the complex numbers z_1 and z_2 with arguments θ_1 and θ_2 is equivalent to a rotation of z_1 by θ_2 degrees, accompanied by some stretching (figure 3.1).[1]

A final definition that will be useful later is the *complex conjugate*, defined for vector $z = x + iy$ as $z^\star = x - iy$, or, in polar notation: $z^\star = r\exp(-i\,\theta)$. The product of a complex number z_1 and the complex conjugate of another number z_2

---

[1] This interpretation helps clarify the otherwise cryptic definition of $i = \sqrt{-1}$: Multiplying a number by i (equivalent to $r = 1$, $\theta = \pi/2$) is just a 90 degree rotation. Thus the value i^2 corresponds to rotating the unit vector twice, from $+1$ to $+i$ to -1.

can be thought of as a measure of their similarity. Consider the following set of relationships:

$$
\begin{aligned}
\text{Identical vectors:} && z_1 \cdot z_1^\star &= |z_1|^2 \\
\text{Diametrically opposite vectors:} && z_1 \cdot (z_1 \, e^{i\pi})^\star &= -|z_1|^2 \\
\text{Perpendicular vectors:} && z_1 \cdot (z_1 \, e^{i\frac{\pi}{2}})^\star &= i\,|z_1|^2
\end{aligned}
$$

Thus, the product $z_1 z_2^\star$ has the greatest positive value for vectors that are oriented in the same direction, the greatest negative value for opposite orientations, and a zero real component for pairs of perpendicular vectors. These relationships will be exploited to construct autocorrelation structures.

3.3 Movement in complex notation

A continuous two-dimentional trajectory $Z(t) = \{X(t), Y(t)\}$ is fully determined by the velocity vector $V(t)$, such that

$$
Z(t) = \int_{t'=0}^{t} V(t') \, dt'. \tag{3.6}
$$

The velocity vector can be denoted in complex notation as

$$
V(t) = S(t) \, e^{i\Phi(t)}, \tag{3.7}
$$

where $S = |V|$ is the instantaneous speed (modulus of V) and Φ is the absolute orientation.

Actual movement data is almost always discrete. Consider data consisting of ordered pairs of x and y locations X_t and Y_t where $t \in \{0, 1, 2, ..., n\}$ (for simplicity we assume that the time interval is equal to 1). These locations can be recast as a complex vector of positions $Z_t = X_t + i\,Y_t$ and the velocity vector is given by

$$
V_t = Z_t - Z_{t-1} \text{ for } t \in \{1, 2, ..., n\}. \tag{3.8}
$$
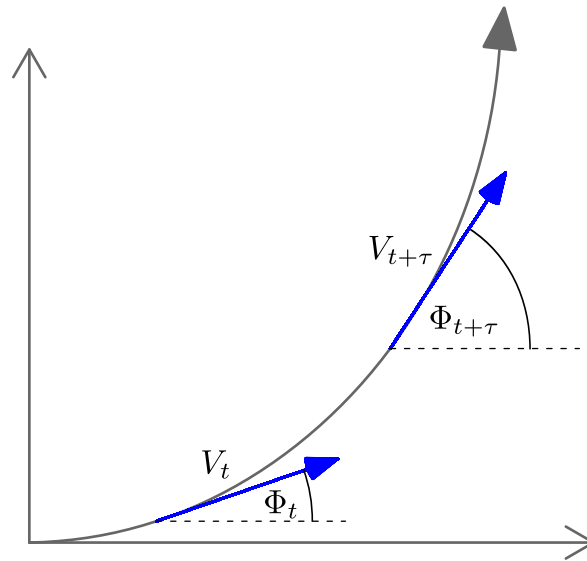
Figure 3.2: Velocity vectors displaced by interval τ.

Obtaining turning angles and step lengths from complex velocity vectors V_t is notationally and computationally very compact. This is illustrated in the following example (given in R-code but analogous in any standard mathematical programming language):

```
# Complex position vector from positions "X" and "Y"
  Z <- complex(re=X,im=Y)
# Velocity vector
  V <- diff(Z)
# Absolute orientations and step lengths
  S <- Mod(V)
  Phi <- Arg(V)
# Turning angles and differential velocity changes
  dV <- diff(V)
  Theta <- Arg(dV)
```

It is arguably worth the effort of familiarizing oneself with complex notation and manipulations for the sake of this efficiency alone.

3.3.1 Complex velocity autocorrelation function

Consider $V(t+\tau)$, the velocity vector at some small interval τ later (figure 3.2). The complex product between two velocity vectors :

$$G(t, t+\tau) = V(t) \cdot V^\star(t+\tau) = S(t)\, S(t+\tau)\, e^{-i\,(\Phi(t+\tau)-\Phi(t))}. \qquad (3.9)$$

This complex product has intuitive meaning in the context of the properties of the complex product listed above. Consider a decomposition into real and imaginary components

$$\Re(G(t, t+\tau)) = S(t)\, S(t+\tau)\, \cos(\Phi(t+\tau) - \Phi) \qquad (3.10)$$
$$\Im(G(t, t+\tau)) = S(t)\, S(t+\tau)\, \sin(\Phi(t+\tau) - \Phi) \qquad (3.11)$$

The real component reflects the persistence of the movement in a given direction, while the imaginary component represents the tendency to turn to the right and left. A positive, purely real value for G indicates that the movement is continuing in exactly the same direction as before, whereas an imaginary value for G indicates that the movement has made a 90 degree turn to the right or left.[2].

We consider stationary, stochastic processes, i.e. movements that are homogeneous with time but include a random component that can be only be described statistically. This constraint is referred to the *Markov condition*, which suggests that at any given moment, the subsequent evolution of the process depends only on the current state. Under this assumption, the expected value of $G(t, t+\tau)$ should depend only on τ.

---

[2]These components of movement are related to those which are analyzed for Northern fur seals in chapter 5 (see also Gurarie and Andrews *in prep*)

We can then define a *complex velocity autocorrelation function* (CVAF)

$$\mathcal{G}_{\mathcal{V}}(\tau) \;=\; \langle V(t+\tau) \cdot V^{\star}(t) \rangle \tag{3.12}$$

At $\tau = 0$, the value of this autocorrelation function is the mean speed squared $\mathcal{G}_{\mathcal{V}}(\prime) = \langle |V(t)| \rangle^2$. As τ approaches infinity and all information between sampled velocity vectors is lost, the function approaches the modulus of the expected square velocities: $\mathcal{G}_{\mathcal{V}}(\tau \to \infty) = \langle |V(t)|^2 \rangle$. For a movement that has no persistence in any given direction, $\mu_v = 0$ and $\mathcal{G}_{\mathcal{V}}(\tau \to \infty) = 0$. While both of these limits are confined to the real line, intermediate values of the complex velocity autocorrelation function can well be imaginary. The imaginary component reflects tendencies to turn in a consistent way at certain intervals τ.

`R` code for obtaining the CVAF from discrete data is provided in Appendix B.1.

3.4   The Dunn-Brown-Alt movement model

A versatile continuous model for homogeneous autocorrelated movement with a random component can be expressed as a stochastic differential equation

$$dV = \alpha \cdot (\mu_v - V)\, dt + \beta \cdot dW_t \tag{3.13}$$

where μ_v is a mean drift, $\alpha = \alpha_1 + i\,\alpha_2$ is a complex parameter which determines the turning radius and strength of autocorrelation, $W_t = W_1 + i\,W_2$ is a white noise Wiener process composed of real and imaginary independent standard Gaussian variables and β is a real parameter that represents the strength of the randomness. This model was proposed by Dunn and Brown (1987) to model the motility of unicellular organisms and was further developed by Alt (1990). In deference to these researchers I will refer to it as the Dunn-Brown-Alt (DBA) model.[3] A movement is determined by the two

---

[3]In the original references, a, b and $\overline{V}$ are used in place of α, β and μ_v. My choice of notation is an attempt to be consistent with the statistical convention of using Greek letters for parameters

complex parameters: μ_v and α and the real parameter β.

The simplest discretization of the DBA model (which we can refer to as the *DBA walk*) is to set $dt = 1$ as follows:

$$V_{t+1} = V_t + \alpha \cdot (\mu_v - V_t) + \beta \cdot W_t. \tag{3.14}$$

However, care is required when discretizing the DBA model in this way, since coarse time samplings will lead to unstable solutions. Specifically, if the value for the α parameter lies outside of the circle defined by $(\alpha_1 - 1)^2 + \alpha_2^2 = 1$, the mode of the velocity will either explode or collapse exponentially. This problem is mitigated by introducing a scaling coefficient $K > 1$ to make the temporal increments smaller, such that

$$V_{t+1} = V_t + \frac{\alpha}{K} \cdot (\mu_v - V_t) + \frac{\beta}{\sqrt{K}} \cdot W_t. \tag{3.15}$$

As an example, to convert this equation from units of minutes to units of seconds, K would be set equal to 60. The square root before stochastic part of the equation is there to balance out the increase in the variance of the white noise process as the square root of the number in increments.

`R` code for simulating DBA walks is provided in appendix B.2 and several realizations under various values for the parameters are presented in figure 3.3.

A feel for the behavior of this model is obtained by considering several special cases. If there is no drift ($\mu_v = 0$) and $\alpha = 1$ (figure 3.3a), the model reduces to $V_{t+1} = \beta \cdot W_t$, i.e. a pure random walk where each step is independently drawn from a bivariate normal. Real quantities for α between 0 and 1 yield classic correlated random walks (figure 3.3b). The addition of an imaginary component to α introduces a tendency to turn in circles (figure 3.3c), and $\mu_v > 0$ causes advective movement (figure 3.3d).

---

to be estimated and capital letters for random variables.
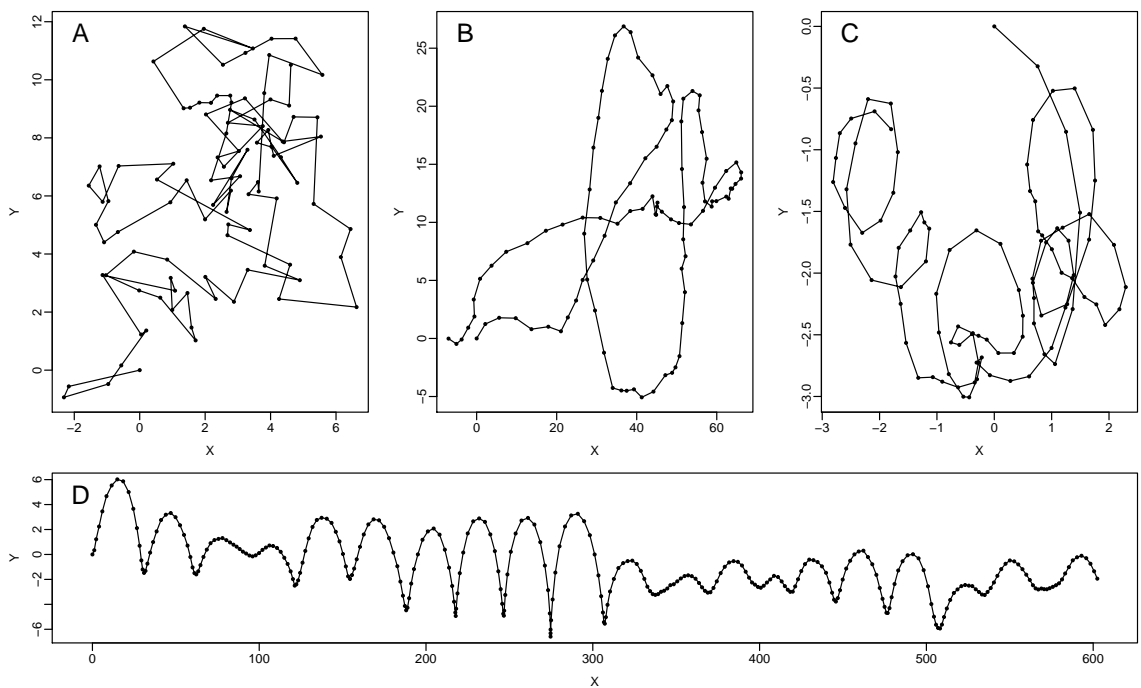
Figure 3.3: Realizations of discrete DBA walks (3.15) for different parameter values. (A) $\alpha = 1$, $\beta = 1$, $\mu_v = 0$; (B) same as (A), except $\alpha = 0.1$, (C) $\alpha = 0.1 + 0.4i$, $\beta = .1$, $\mu_v = 0$; (D) same as (C), except $\mu_v = 2$.

3.4.1 Complex velocity autocorrelation function

The velocity autocorrelation function (3.9) of a DBA model can be expressed as

$$\mathcal{G}_{\mathcal{V}}(\tau) = \mu_v^2 + \frac{\beta^2}{2\Re(\alpha)} e^{-\alpha\tau} \tag{3.16}$$

which can be decomposed as

$$\Re(\mathcal{G}_{\mathcal{V}}(\tau)) = \mu_v^2 + \frac{\beta^2}{2\alpha_1} e^{-\alpha_1\tau} \cos(\alpha_2\tau) \tag{3.17}$$

$$\Im(\mathcal{G}_{\mathcal{V}}(\tau)) = \frac{\beta^2}{2\alpha_1} e^{-\alpha_1\tau} \sin(\alpha_2\tau) \tag{3.18}$$

Empirical autocorrelation functions for the four parameterizations illustrated in figure 3.3 are presented in figure 3.4. Recalling from (3.12) that $\mathcal{G}_{\mathcal{V}}(\prime) = \langle |V(t)| \rangle^2$, 3.16 can be rewritten at lag zero as

$$\langle |V(t)| \rangle^2 - \mu_v^2 = \text{Var}\,[V(t)] = \frac{\beta^2}{2\alpha_1}. \tag{3.19}$$

The autocorrelation function suggests that the velocity of an organism fluctuates as a stationary Gaussian process around the mean drift μ_v with variance $\sigma_v^2 = \beta^2/2\alpha_1$, a mean angular velocity given by the frequency of the periodic component of the autocorrelation function $\bar{\omega} = -\alpha_2$, and a *relaxation time* (i.e. a characteristic time scale at which the autocorrelation is lost) given by the exponent on the decay term: $T_v = 1/a_1$.
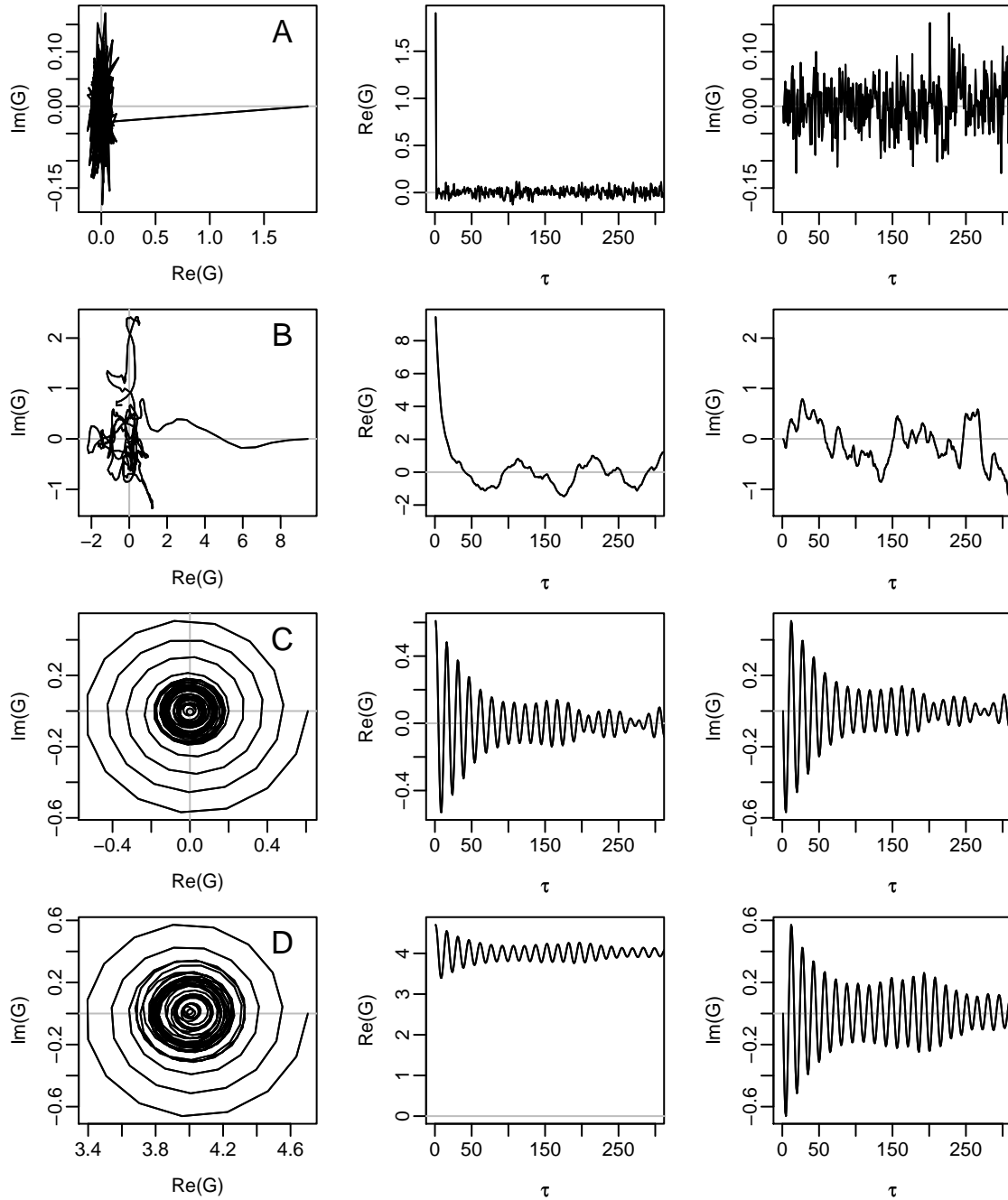
Figure 3.4: Velocity auto-correlation functions for the four models in figure 3.3. The trace of the function in the complex plane, the real and imaginary components are presented in the left, middle and right columns, respectively.

3.4.2 Estimating parameters

The estimate for the mean drift $\hat{\mu}_v$ is obtained directly by taking the average velocity,

$$\widehat{\mu_v} = \overline{V_t} \tag{3.20}$$

Note that the mean of a vector of complex values has a standard definition, such that for vector $Z_k = Z_k + i\,Y_k$, where $k \in (1, 2, ..., n)$, the mean is given by

$$\bar{Z} = \sum_{k=1}^{N} X_k + i\,Y_k, = \overline{X} + i\,\overline{Y} \tag{3.21}$$

The estimate for stochasticity strength parameter β is constrained by its relationship to the variance of the velocity and to α_1 (3.19) such that

$$\widetilde{\beta} = \sqrt{\frac{\widehat{\sigma_v}^2}{2\widehat{\alpha_1}}} = \sqrt{\frac{\overline{|V_t|^2} - \overline{|V_t|}^2}{2\widehat{\alpha_1}}} \tag{3.22}$$

Since α_2 is related to the period of rotations, a good initial estimate can be obtained identifying the lag at which the standard autocorrelation function of the real or imaginary component of V attains a minimum. Formally:

$$\widetilde{\alpha_2} = \frac{\pi}{\underset{\tau}{\operatorname{argmax}}\left(-\gamma(\Re(V_t))\right)} \tag{3.23}$$

where $\gamma_\tau(X)$ is the autocorrelation function $\gamma_\tau = \sum_{j=1}^{N-\tau} X_j \cdot X_{j+\tau}$.

The final estimate $\widehat{\alpha_1}$ can be obtained by fitting the real part of the empirical complex autocorrelation function to equation (3.17). In the examples that follow, I performed a minimization of the least-squared difference between the theoretical equation (3.17) and the empirical CVAF over values of α_1, α_2 and β using $\widetilde{\alpha_2}$ and $\widetilde{\beta}$ as seed parameters in the optimization routine (see figure 3.5 for a comparison between the null parameters and the fitted parameters). R-code for estimating all

parameters is provided in appendix B.3.

Obtaining errors or confidence intervals for these parameters is a subject for future work.

3.5   *Fitting* Heterosigma *movement data*

I fitted this model to several tracks of *Heterosigma akashiwo*, a motile biflagellate unicellular alga. *Heterosigma* episodically form toxic aggregations known as harmful algal blooms (HAB) or red tide. Since these can have ecological and public health impacts, there is considerable interest in understanding and explaining *Heterosigma* movement patterns and behaviors. It has been suggested, for example, that statistically significant differences in the movement parameters between different strains can explain differences in the resistance to dispersion of toxic slicks(Bearon et al., 2004).

The alga were videotaped in a cylindrical chamber and their tracks were automatically digitized as a two-dimensional track (see Bearon et al. (2004, 2006) for details on the experimental setup). *Heterosigma* typically follow helical trajectories along vertical axes, but there is high variability between individuals as well as distinct behavioral changes within single trajectories.

Figure 3.5 represents three sample tracks for individual *Heterosigma* labeled H49, H45 and H52, along with fitted complex velocity autocorrelation curves. The data consist of 600 to 1200 two dimensional locations, accounting for 20 to 40 seconds of movement at 30 frames per second. The total distance traveled range is on the order of 0.5 to 2.5 millimeters.

The estimates for the parameter DBA parameter values are given in table 3.1. Examples of tracks simulated using these estimated parameters are plotted in figure 3.5 for visual comparison.
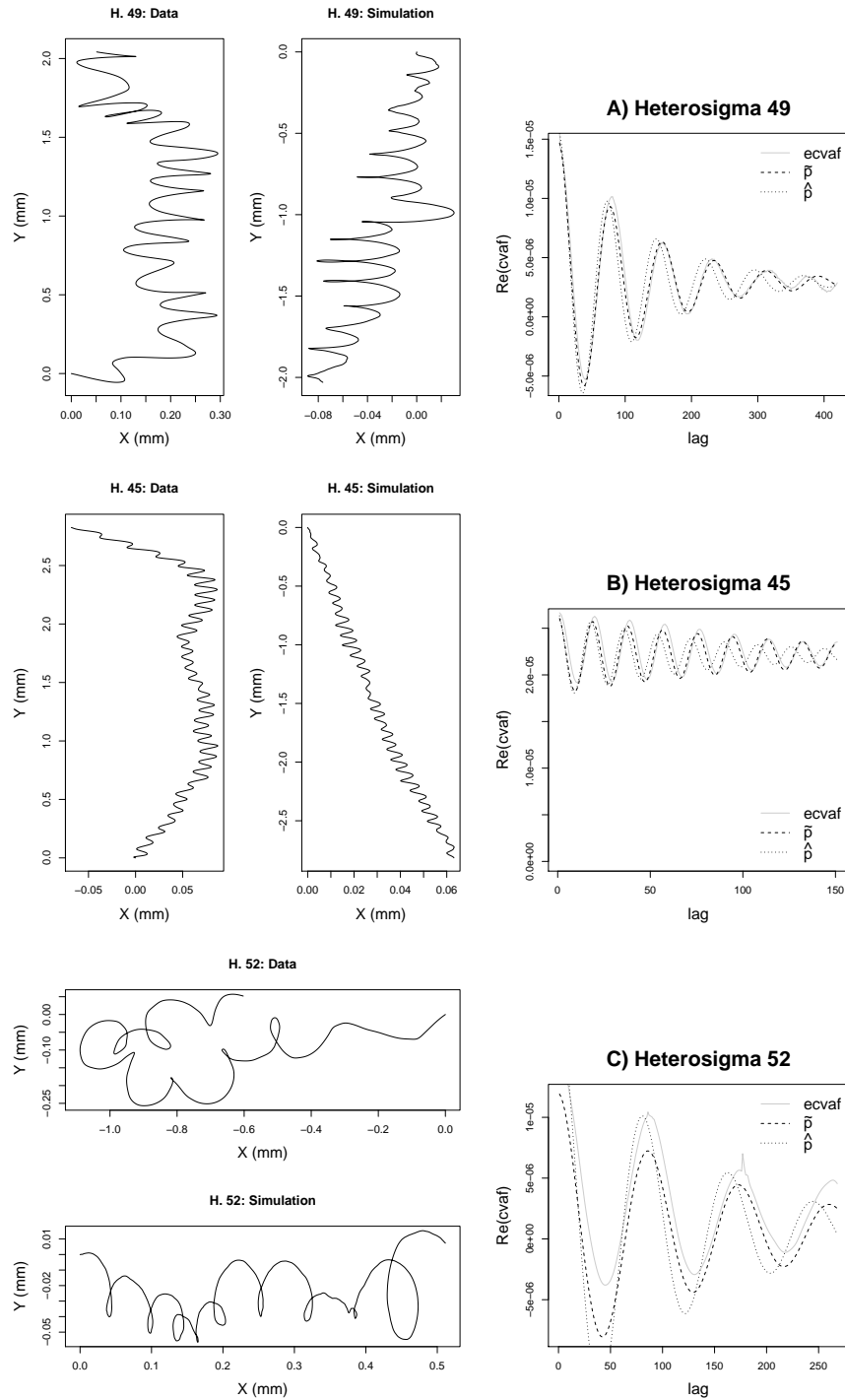
Figure 3.5: Data on 3 sample *Heterosigma* tracks and simulated tracks based on parameters estimated from the movement data. The plots on the right represent the corresponding real part of the empirical complex velocity autocorrelation functions (ecvaf, in grey) with fitted curves (dotted and dashed lines) representing the null parameters ($\tilde{p}$) and the optimized parameters ($\hat{p}$). See text for details.

Table 3.1: Estimates of parameters for the DBA model estimated for example *Heterosigma* tracks. T_t and D_t are the total time elapsed and net displacement of the tracks respectively. $T_v = 1/\Re(\alpha)$ is the characteristic time-scale of auto-correlation.

	Frames (1/30 sec)	D_t (mm)	$\widehat{\mu_v}$ (mm/sec)	$\widehat{\alpha}$ (1/sec)	$\widehat{\beta}$ (mm/sec$^{1/2}$)	T_v (sec)
H49	1197	2.042	$-0.00128 - 0.051i$	$0.239 + 2.40i$	2.38×10^{-3}	4.18
H45	602	2.826	$0.00340 - 0.141i$	$0.213 + 10.0i$	1.31×10^{-3}	4.69
H52	765	0.605	$0.02366 - 0.002i$	$0.189 + 2.16i$	2.07×10^{-3}	5.30

3.6 Discussion

A visual comparison of the true tracks and the simulated tracks (figure 3.5) suggest that the estimation routine captures many of the essential features of the data. Specifically, the net displacements, the number of oscillations, the nature of the oscillations (loops versus swoops) and the variability in the magnitude of the oscillations are rendered well. The parameter values are readily interpretable in terms of these visual characteristics. Thus: H49 and H45 have predominantly downward net movement (large negative imaginary components of $\hat{\mu}$), with the latter moving almost three times faster ($\Im(\mu_v) = -.141$ compared to -0.051), while H52 has the slowest aggregate displacement. The tightest oscillations are displayed by H45, as reflected in the high imaginary component of $\widehat{\alpha}$ (10.0 compared to 2.4 and 2.16). The greater "loopiness" of H52 is explained by the relatively slow advection term compared to a comparable angular velocity with the other tracks. The variability coefficient β is lowest for H45, but quite similar for the other two tracks. The time-scale of autocorrelation is fairly consistent between all three tracks, between 4.2 and 5.5 seconds. In summary, H45 can be classified as a fast, tightly coiled, less variable helix, while H49 is somewhat intermediate in speed, less tightly coiled and variable, and H52 is a slow, loopy, variable track.

The movement model is most inadequate in its inability to capture the larger scale direction changes evident in the tracks even within the relatively short durations. While H49 maintains the most consistent mean direction, H45 displays several gradual direction shifts and H52 doubles back on itself, making the use of net displacement in calculating μ_v obviously inappropriate. One possible way of dealing with this problem is to identify the trend in the data by filtering away the periodic component via an averaging of points over a scale larger than the period of oscillation (reminiscent of standard "de-trending" techniques in time-series analysis. Subtracting away the trend from the raw data should give a linearized, oscillating remainder which will conform with the DBA model. This is a subject for future research.

3.7   *Conclusions*

The DBA model is a more versatile model for movement than the classic correlated random walk. It contains relatively few parameters, all of which are readily interpretable in terms of random fluctuations, scales of auto-correlations, mean drift and turning radii. The parameters are straightforward to estimate via an analysis of the complex velocity auto-correlation function. For trajectories like those of the *Heterosigma*, which combine drift, oscillations and randomness, the DBA model does an excellent job of quantifying the main features of these movements in a straightforward and easy to implement way. Informative quantification of complex movement, especially when collected in large quantities as in the video collection of microscopic plankter movements, is in itself an important goal as it is a first step in being able to compare differences between strains, between individuals, and between several behaviors performed by a single individual. Furthermore, a consideration of the values of these parameters in terms of the actual biomechanics of *Heterosigma* is a potentially informative exercise. For example, the interaction between flagellar length, power and body mass must be related to the speed (μ), angular velocity ($\Im(\alpha)$) of the organism's movement, while ambient variations or fluctuations in the water medium might

control the time-scales of autocorrelation and the strength of the random component $(\beta, \Re(\alpha))$.

In further chapters, I am interested in the movement of much larger organisms for which the techniques outlined here do not at first appear immediately applicable. The movement of a top predator is presumably controlled by much more complicated behavioral and biophysical mechanisms than two flagella. The tendency or necessity to oscillate in a helical manner is largely absent. Satellite tracking data is generally two-dimensional and often very irregularly sampled, especially in the marine environment. However, there are several important lessons to be learned from the DBA model. The first is that the decomposition of a movement into turning and persistence components and the modeling of each of these as essentially stationary Gaussian processes is conceptually and mathematically justifiable. This decomposition is the fundamental premise of the technique employed in Chapter 3 for the analysis of remotely sensed large animal movements at sea. The DBA model, however, adds to these analyses a rigorous understanding of the relationship between these orthogonal components of movement. The second contribution of the DBA model is the straightforward way in which characteristic time and length scales of randomness can be obtained directly from the parameters. I discuss the importance of these scales for parameterizing rates of diffusion and dispersal and for predicting encounter rates in subsequent sections and chapters of this dissertation.

Part II

ECOLOGICAL APPLICATIONS

[The study of animal movement] is fundamentally important to future advances in ecological, ethological and evolutionary theory, as well as to applied disciplines such as natural resource management, pest control and public health ...

W. Lidicker and R. Caldwell 1982

Chapter 4

ENCOUNTER RATE MODELS

[T]rusting their lives ... they descended the Mississippi, running the gantlet between hostile tribes, who fiercely attacked them.

FRANCIS PARKMAN (1865)

4.1 Introduction

Encounters between organisms are universal prerequisites for many fundamental ecological processes. Feeding depends on the ability to encounter food items or prey, survival can be strongly dependent on avoiding encounters with predators, and successful reproduction depends on mates encountering each other. The most widely applied null-model of encounter rates is essentially a variation of one developed by Clausius (1859) and Maxwell (1860) to describe the statistical-mechanical behavior of gases. Their result relates collision rates of "ideal free gases", i.e. perfectly elastic, deterministically moving spheres of fixed radius, via a straightforward function of mean velocities, interaction area and density. A fundamental assumption behind the derivation is that each particle moves linearly with a constant velocity. Clausius assumed all particles move with a single homogeneous velocity while Maxwell solved the collision rate problem for particles whose velocities are distributed according to the Maxwell-Boltzmann distribution - i.e. where the x, y and z components of velocity were each normally distributed.

These equations were generalized to a predator-prey type scenario in a widely cited and influential model of Gerritsen and Strickler (1977) (referred to as the GS model), which predicted encounter rates between prey (in their case, zooplankton) moving

at velocity u and predators moving at velocity v. The GS model accounts for any population-wide distribution of velocities $g(u)$ and $f(v)$ for the prey and predators respectively, leading to the following result

$$E_r = \frac{\pi\, r^2 \rho}{6} \frac{(\bar{u} + \bar{v})^3 - |\bar{u} + \bar{v}|^3}{\bar{u} \cdot \bar{v}} \tag{4.1}$$

where E_r is the stable encounter rate of a system of predators and prey moving with mean velocities $\bar{v}$ and $\bar{v}$ are the mean speeds of the predator and prey populations, ρ is the density of prey and r is the encounter radius. Evans (1989) suggests that a simplification of the GS-model to a simple root squared sum of the velocities, such that

$$E_r = \pi\, r^2 \rho \sqrt{\bar{u}^2 + \bar{v}^2)} \tag{4.2}$$

While the GS and Evans models account for the possibility of heterogeneous populations, the fundamental movement of the individuals remains linear and deterministic as in the classic ideal gas.

While encounter rate modeling is dominated by heterogeneous populations of individuals moving in a deterministic way, most historical models of population dispersal are governed by the inverse case, namely a homogeneous population of randomly moving individuals. Among many studies along these lines, one might cite Skellam (1951); Turchin (1998); Okubo and Levin (2001). In another example of biology taking its cues from results in physics, these movements are described as approximately Brownian, i.e. consisting of very short, independent steps in random directions at small time intervals. This model has the advantage of being mathematically related to processes of diffusion and an extensive family of differential equation-based approaches to modeling dispersal (Okubo and Grünbaum, 2001). Rates of 'diffusion' have been related to individual parameters of movement and thereby to actual measurements of animal movements via random walk (RW) and correlated random walk models (CRW), notably in an several early mathematical studies by Patlak (1953b,a) and a highly

influential model of butterfly movements between flowers by Kareiva and Shigesada (1983). Much of the theoretical work in diffusive dispersion has considered homogeneous populations of randomly moving individuals, though several recent studies have explicitly modeled heterogeneity in the populations (Skalski and Gilliam, 2000; Yamamura, 2002; Skalski and Gilliam, 2003; Gurarie et al., 2008).

Lotka (1924), who used an ideal gas model to justify the density-dependence of his famous predation model, was fully aware that his models were great simplifications of reality and called on researchers to attack the mathematical complexities of movement models that consist of mixtures of directed and random components (see epigraph of Part I). Several studies have since addressed the issue of random movement in ideal gas models. In particular, Hutchinson and Waser (2007), in their excellent review of ideal gas models, performed simulations of encounter rates between organisms moving as correlated random walks, stating that "turning angle distributions, detection distance, step length and step number all have interacting non-linear effects, so quantitative predictions are only possible using simulation".

Anderson et al. (2005) present an analytical model of encounter rates referred to as the XT-model. This model predicts encounter rates between prey migrating through a field of predators with mean velocity w such that

$$E_r = \pi\, r^2 \rho \sqrt{(w^2 + \omega^2)} \tag{4.3}$$

where w is the relative advective speed of prey moving through a field of predators and ω is the root mean squared velocity of prey and predator movement, i.e. a statistical measure of the net randomness of the process. This result is reminiscent of Evans' model (4.2), with the relative speeds of predator and prey replaced by a decomposition in to advection and random movement. The empirically supported conclusion of this model is that the strength of dependence of survival on travel distance versus travel time is controlled by the relative net randomness of the movement. Specifically,

linearly traveling advective prey survival is independent of the speed of travel since the same gauntlet of predators is encountered, whereas more random movements increase encounter rates. "Randomness of movement" is, however, somewhat ambiguously defined in their model, with no suggestion as to how to estimate ω from individual movements, or what it's relationship to a rate of dispersion or diffusion might be.

In this chapter, I present mathematical models of encounter rates which incorporate a mixture of random and directed movements for heterogeneous populations. The resulting expressions are relatively simple and tractable. While broadly corroborating the XT-model, it is more rigorously parametrized in terms of individual movements.

In developing the argument, I rely on the MST parameterization of model from chapter 2. I derive one-dimensional encounter rates for unbiased and biased walkers moving through a field of moving and stationary predators based on the parameterization and discuss applications to survival probabilities of migrating salmon. Finally, I extend the encounter rate model to two dimensions.

A note on jargon: In developing the following arguments, I begin with the simplest of model of a moving organism, referred to as a "searcher" encountering stationary organisms referred to as "targets". In applications where the "targets" represent, for example, food items and "searchers" represent foragers, the jargon is intuitive. In other applications, the "searchers" represent prey items who are presumably attempting to avoid collisions or interactions with "targets" that represent predators, though it is a distinctly masochistic prey that "seeks" a predator. In applications where both "searchers" and "targets" are moving, there is an inevitable symmetry between the two. It should therefore be made clear that the two terms refer merely to two classes of organisms. The objective of the derivations is to predict rates of encounter given certain parameters of movement, regardless of intent.

4.2   One-dimensional encounter rate models

4.2.1   Unbiased walker

Assume that an individual is moving one-dimensionally as an unbiased MST-walk, i.e. $\mu = 0$, $\sigma > 0$ and $\tau > 0$ through a field of immobile targets (either food items or ambushing predators) with fixed density ρ. In the one-dimensional case, we assume that the characteristic distance between targets $(\frac{d=1}{\rho})$ is greater than the size of the random step σ. The MST-walk is defined as

$$W(t) = \sum_{i=0}^{t/\tau} X_i \tag{4.4}$$

where X_i is the step random variable: X_i are iid random variables with mean $\mu = 0$ and variance σ^2. Several important quantities related to $W(t)$ are the *total length of the trajectory*

$$L_X(t) = \sum_{i=1}^{t} |X_i| \tag{4.5}$$

and the *total range of the walk*

$$W_R(t) = \max(W(t)) - \min(W(t)). \tag{4.6}$$

Non-destructive encounters

If the encounters are not "destructive" - i.e. the searcher can reencounter any given target an unlimited amount of times - and the system is isotropic, encounters must occur randomly with an asymptotically constant rate.

The expected number of encounters as a function of time is

$$E_p(t) = N_p(t)R_p \tag{4.7}$$

where N_p is the total number of *unique* targets encountered and R_p is the mean number of times each target is encountered. The total number of unique targets encountered is just the density ρ times the total range of a given random walk. The mean number of revisits per target is the total expected length of the trajectory, denoted L_W, divided by the total range. It should be clear that total length is the sum of the absolute lengths of each time step, distinct from the net displacement. Thus, we can rewrite 4.7 as

$$E_p(t) = \rho\, W_R(t)\frac{L_X(t)}{W_R(t)} = \rho\, L_X(t) \tag{4.8}$$

Recalling that time is discretized in units of τ, the total length of the walk is:

$$L_X(t) = \sum_{i=1}^{\frac{t}{\tau}} |X_i| \tag{4.9}$$

and the expected number of encounters is

$$E_p(t) = \rho\, \langle L_X(t) \rangle = \frac{\rho t}{\tau}\, \langle |X| \rangle \tag{4.10}$$

The encounter rate is just the number of encounters per unit time,

$$E_r = \frac{\rho\, \langle |X| \rangle}{\tau}. \tag{4.11}$$

The value of the expected absolute step length ($\langle |X| \rangle$), and consequently the encounter rate, *depends on the nature of the step distribution.* This is a somewhat surprising result, since most aggregate statistical properties of random walks and diffusion processes, including dispersal rates and arrival times, are identical as long as the first two moments are equivalent. Indeed, this principle has been formalized and is referred to as *Wald's identity* (Wald, 1947).

We now consider three different random walk models, each with step mean equal
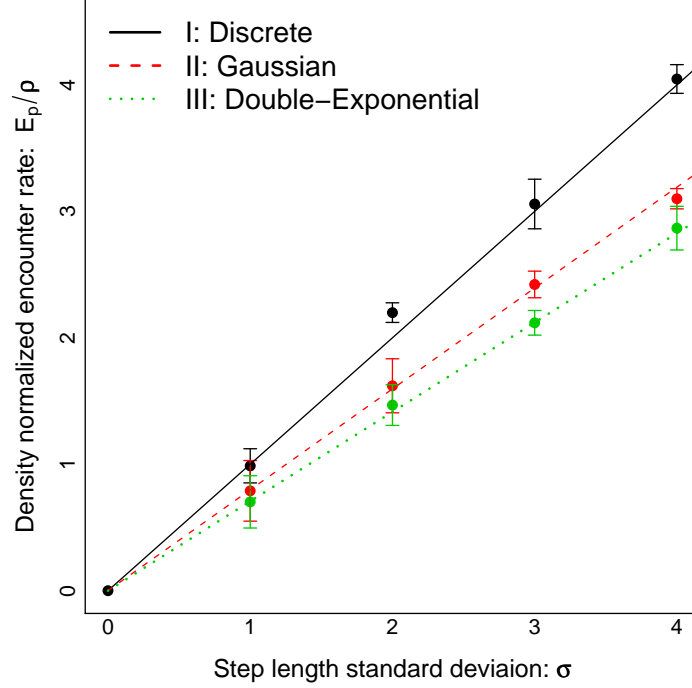
Figure 4.1: Results of encounter rate simulations. The number line was randomly seeded with targets in a range of densities ρ from 0.01 to 0.1. Unbiased random walks on the number line were generated and the resulting encounter rates (E_p) obtained. The slope of E_p against ρ is plotted here with the associated estimated standard deviations (error bars) for different values of σ and three different step distributions: discrete unbiased random walk (black), the Gaussian random walk (red), and the double exponential random walk (green). The lines represent the theoretically predicted slopes: 1, $\sqrt{2/\pi}$ and $1/\sqrt{2}$ respectively.

to zero and variance σ^2:

Case I: A discrete unbiased random walk with variance σ^2 has fixed step length σ. The encounter rate then becomes:

$$E_r = \frac{\rho\sigma}{\tau} \tag{4.12}$$

Case II: A Gaussian random walk, or discrete Gaussian jump process, has a normal step distribution: $X \sim \text{Normal}\{0, \sigma^2\}$. The expected absolute step length $\langle |X| \rangle =$

$\sqrt{2/\pi}\sigma$, and the encounter rate is

$$E_r = \sqrt{\frac{2}{\pi}}\frac{\rho\sigma}{\tau} \qquad (4.13)$$

Case III: Another reasonable model for step length distributions is a double exponential model where the step length distribution is

$$X \sim f(x) = \frac{\sqrt{2}}{\sigma}e^{-\frac{\sqrt{2}}{\sigma}|x|} \qquad (4.14)$$

The expected absolute value of a step is $\langle|X|\rangle = \sigma/\sqrt{2}$ such that the encounter rate becomes

$$E_r = \frac{1}{\sqrt{2}}\frac{\rho\sigma}{\tau} \qquad (4.15)$$

In general, the encounter rate for this process is:

$$E_r = \kappa\frac{\rho\sigma}{\tau} \qquad (4.16)$$

where κ is a constant less than or equal to 1. The case where $\kappa = 1$, the discrete unbiased random walk, is in some ways an extreme case, as the distribution is a split Delta dirac function. For "reasonable" step distributions, κ is inversely related to the kurtosis of X. The kurtoses of the discrete, Gaussian and double-exponential cases are {-2, 0 and 3} respectively, corresponding to values for κ of {1, 0.798 and 0.707}. The double-exponential case is particularly fat-tailed and it is safe to assume that for any real movements the coefficient κ will range between 0.5 and 1, It should be noted that ore extreme tails which fall into the class of walks known as Lévy walks (Bartumeus et al., 2002)) and may have somewhat different properties.

Simulations of the movement process confirm these results with great precision (figure 4.1).

Destructive encounters

A destructive encounter is one in which the target is annihilated upon encounter. In this case, there is no constant encounter rate, as the density of targets is consistently depleted. Nonetheless, the expected number of encounters as a function of time is simply the range of the walk times the density. Thus

$$E_p(t) = \rho W_R(t) \tag{4.17}$$

The expected range of a random walk $\langle W_R \rangle = \frac{2}{\pi}\sigma\sqrt{t}$ (see Weiss (1994) for a derivation). The encounter rate of a destructive, unbiased one-dimensional walk consequently decreases as the square root of time as

$$E_r(t) = \frac{2}{\pi}\frac{\sigma\rho}{\sqrt{t}} \tag{4.18}$$

Note that this result holds regardless of the nature of the step length distribution.

4.2.2 Advective random walkers

If an MST-walker moves deterministically through a field of targets with velocity μ/τ and step variance $\sigma^2 = 0$, the expected number of targets encountered at time t is just $E_p(t) = \rho\mu\frac{t}{\tau}$. The addition of a random component in the movement will have an impact on encounter rates only if there is some probability that the walker returns to a previously visited site. Clearly, this probability will be positive only if the random component has some probability of exceeding the magnitude of advective step μ. Thus, in the advective case, the nature of the step-process has particular bearing on the encounter rate. Since it is straightforward to scale any of these processes in time, we will assume $\tau = 1$ for the subsequent dicussion.
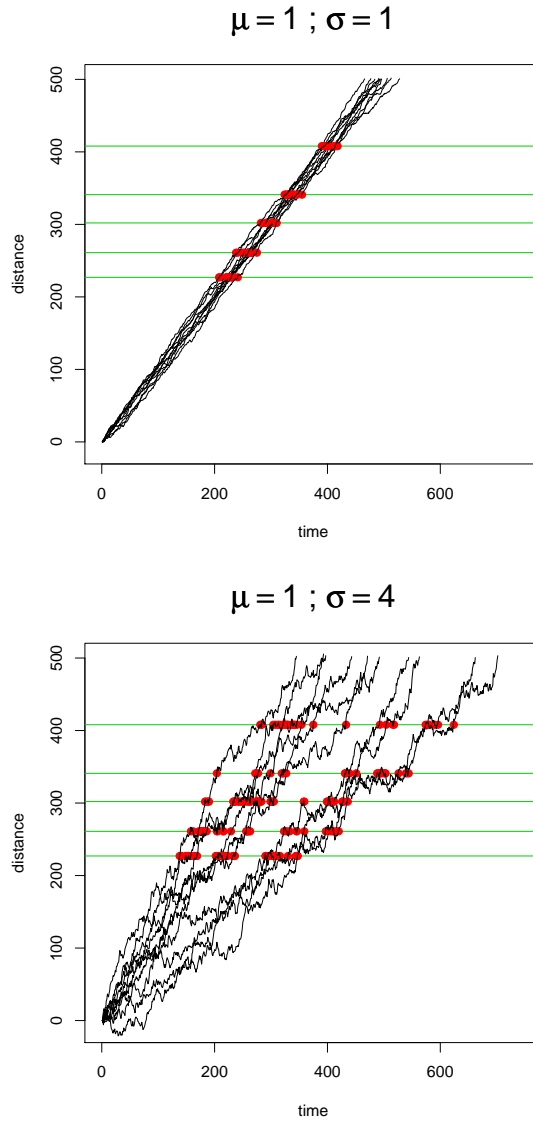
Figure 4.2: Trajectories of 10 simulated 1-D Gaussian MST walkers moving with $\mu = 1$ and (A) $\sigma = 1$ and (B) $\sigma = 4$. Green lines represent the location of stationary targets; red dots represent encounters. In these realizations, the number of encounters per walker per target are (A) 1.24, compared to a predicted value (4.31) of $\sqrt{\frac{2}{\pi}\left(\frac{\sigma}{\mu}\right)^2 + 1} = 1.28$ and (B) 2.72, compared to a predicted value of 2.47.

The biased discrete random walk

A form of random movement for which an analytic solution of advective encounter rates can be obtained is the biased discrete random walk, denoted W_B, where a random walker takes a step of length b to the right with probability p or to the left with probability $q = 1 - p$ such that

$$W_B(t) = b \sum_{i=0}^{t} (2B_i - 1) \tag{4.19}$$

where B are independent Bernoulli trials with probability p. The mean and variance of this walk are

$$\langle W_B(t) \rangle = b(p - q)t \tag{4.20}$$

$$\mathrm{Var}\left[W_B(t)\right] = 4b^2 pqt. \tag{4.21}$$

The analogous statistical properties of the MST walk are

$$\langle W(t) \rangle = \mu t \tag{4.22}$$

$$\mathrm{Var}\left[W(t)\right] = \sigma^2 t. \tag{4.23}$$

Equating the expressions for the means (4.20 and 4.22) and the variances (4.21 and 4.23), the following equivalences are obtained:

$$b = \sqrt{\sigma^2 + \mu^2} \tag{4.24}$$

$$p = \frac{1}{2}\left(1 + \frac{\mu}{\sqrt{\sigma^2 + \mu^2}}\right) \tag{4.25}$$

$$q = \frac{1}{2}\left(1 - \frac{\mu}{\sqrt{\sigma^2 + \mu^2}}\right). \tag{4.26}$$

Note that if $\sigma = 0$, then $p = 1$ and $b = \mu$, i.e. the motion is deterministic with fixed step-length μ. If $\mu = 0$, the motion is asymptotically equivalent to an unbiased random walk with $p = q = 1/2$ and step-length $b = \sigma$.

While every target must eventually be encountered by a discrete biased random walker, some targets will be encountered multiple times as there always exists some finite probability of returning to a previously visited site. The expected number of times a target is encountered is denoted R_p. In order to derive this quantity, we first consider the *probability of recurrence* p_{00}, defined as the probability that a random walker revisits *any* site at least twice. For a discrete biased random walk,

$$p_{00} = 1 - \sqrt{1 - 4pq}. \tag{4.27}$$

[See Hughes (1995) for a derivation of this result. The expected number of returns is given by

$$R_p = \sum_{i=1}^{\infty} i\, P_i \tag{4.28}$$

where P_i is the probability distribution of i returns. The probability of arriving the first time (to any point) is unity. The probability of returning at least once after that is p_{00}, the probability of returning at least twice is p_{00}^2, and so on. Thus, the probability of returning *exactly* twice is $P_2 = p_{00} - p_{00}^2$. In general,

$$P_i = p_{00}^{i-1} - p_{00}^i \tag{4.29}$$

The sum (4.28) can be solved:

$$R_p = \frac{1}{1 - p_{00}} \tag{4.30}$$

Plugging in our expression for p_{00} from equation (4.27) and converting to the MST parameters via equations 4.25 and 4.26 we obtain the following expression for the

number of encounters per target:

$$R_p = \sqrt{\left(\frac{\sigma}{\mu}\right)^2 + 1} \tag{4.31}$$

We note that when $\sigma = 0$, there is exactly one encounter per target over the entire trajectory, whereas as if $\mu = 0$ there is an infinite expected number of encounters, a restatement of the recurrence property of unbiased random walks.

As time goes to infinity, the expected number of unique targets encountered is the expected distance travelled times the density:

$$N_p(t) = \langle W(t) \rangle = \rho \mu t \tag{4.32}$$

The product of the number of individual targets encountered per unit time and the number of encounters per target yields the encounter rate

$$E_r = N_p R_r = \rho \sqrt{\sigma^2 + \mu^2} \tag{4.33}$$

Scaling to discrete units of τ gives

$$E_r = \frac{\rho}{\tau} \sqrt{\sigma^2 + \mu^2} \tag{4.34}$$

Equating equation (4.34) to the non-advective encounter rate (4.16), we can generalize the advective encounter rate to accomodate different distribution shapes as

$$E_r = \frac{\rho}{\tau} \sqrt{(\kappa \sigma)^2 + \mu^2} \tag{4.35}$$

where κ is defined by $\langle |X| \rangle = \kappa \sigma$, with the very important constraint that if $\max(|X|) < \mu$, then revisits are impossible, $\kappa = 0$, and the expected encounter rate will always be equal to $E_r = \rho \mu$. Simulations of the process indicate that this model does an
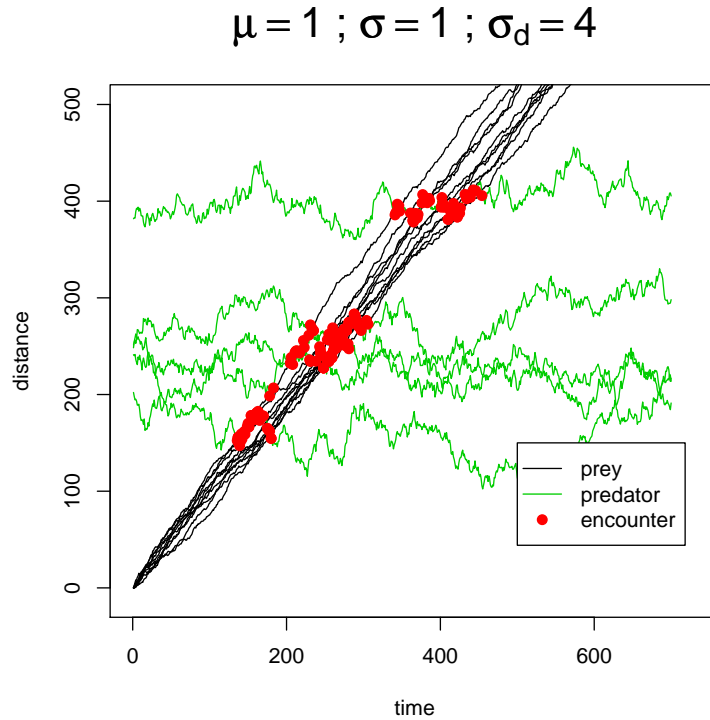
Figure 4.3: Motions of 10 simulated one-dimensional walkers moving as low-randomness a Gaussian MST-walkers with $\mu = 1$ and $\sigma = 1$ through a field of randomly moving targets with $\sigma_d = 4$. The number of encounters per walker per target in this simulation is 3.4 compared to a theoretically predicted value of $\sqrt{\frac{2}{\pi}\left(\frac{\sigma^2+\sigma_d^2}{\mu^2}\right)} + 1 = 3.44$ from equation 4.37.

excellent job of predicting encounter rates for the biased discrete random walk (figure 4.4A). For the advective Gaussian walk (figure 4.4B), the model has the appropriate limiting behavior ($E_r = \rho\kappa\sigma$ where $\mu \ll \sigma$ and $E_r = \rho\kappa\mu$ for $\mu \gg \sigma$), but slightly overestimates actual encounter rates for intermediate values (figure 4.4B). While a mathematical explanation for this divergence is elusive, for all practical purposes it can be considered negligible.
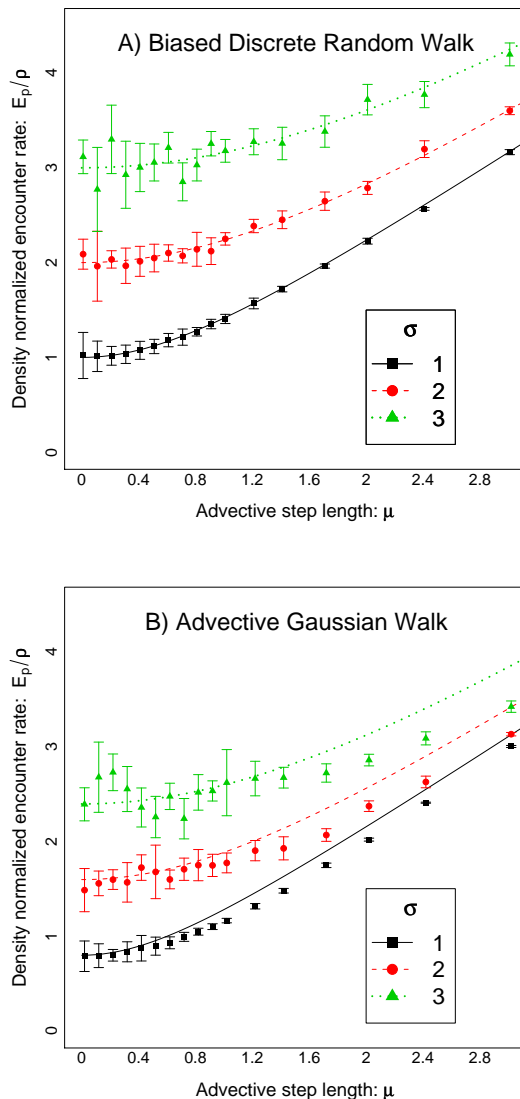
Figure 4.4: Results of encounter rate simulations. The numberline was randomly seeded with targets in a range of densities ρ from 0.01 to 0.1. Random walks on the number line were generated and the resulting encounter rates (E_p) obtained. The slope of E_p against ρ with the associated estimated standard deviation is plotted here against advective component mu for three values of σ. The solid lines represent the predicted encounter rates according to equation 4.35: $E_r = \rho\sqrt{(\kappa\sigma)^2 + \mu^2}$, where $\kappa = 1$ for (A) biased discrete random walk and $\sqrt{\frac{2}{\pi}}$ for (B) the Gaussian random walk.

Randomly moving targets

Within the framework of this derivation, it is relatively straightforward to predict the effect of randomly moving targets. We consider an advective random walker performing movement $W_a(t)$ with parameters $\mu_a > 0$, $\sigma_a > 0$ and $\tau > 0$ through a field of unbiased randomly moving targets $W_b(t)$ with step distribution X_b having $\mu_b > 0$ and some $\sigma_b > 0$ at a similar time-scale τ. The distance $\Delta W(t) = W_a(t) - W_b(t)$ between the walkers and any moving target is simply

$$
\begin{aligned}
\Delta W(t) &= W_a(t) - W_b(t) \\
&= \sum_{i=1}^{t/\tau} X_{ai} - \sum_{i=1}^{t/\tau} X_{bi} \\
&= \sum_{i=1}^{t} \Delta X_i
\end{aligned}
\tag{4.36}
$$

where $\Delta X = X_a - X_b$. For independent processes with well-behaved distributions, ΔX will have mean $\mu_a - \mu_b$ and variance $\sigma_a^2 + \sigma_b^2$. Thus, from the reference point of any given moving target, the process is identical as if the target were stationary and the random walker were moving with transformed MST parameters $\mu' = \mu_a + \mu_b$ and $\sigma' = \sqrt{\sigma_a^2 + \sigma_b^2}$. We can substitute these results directly into (4.35) and obtain the following expression for the encounter rate

$$
E_r = \frac{\rho}{\tau} \sqrt{\kappa^2(\sigma_a^2 + \sigma_b^2) + (\mu_a - \mu_b)^2}
\tag{4.37}
$$

In most practical applications, where we are concerned about encounter rates between a migrating population and a stationary population, μ_b can be set equal to zero. The κ coefficient is generally close to 1 (see equation 4.16 and subsequent discussion). Thus, a widely applicable simple expression for the encounter rate of an advective, randomly moving population moving through a field of randomly moving non-migratory targets
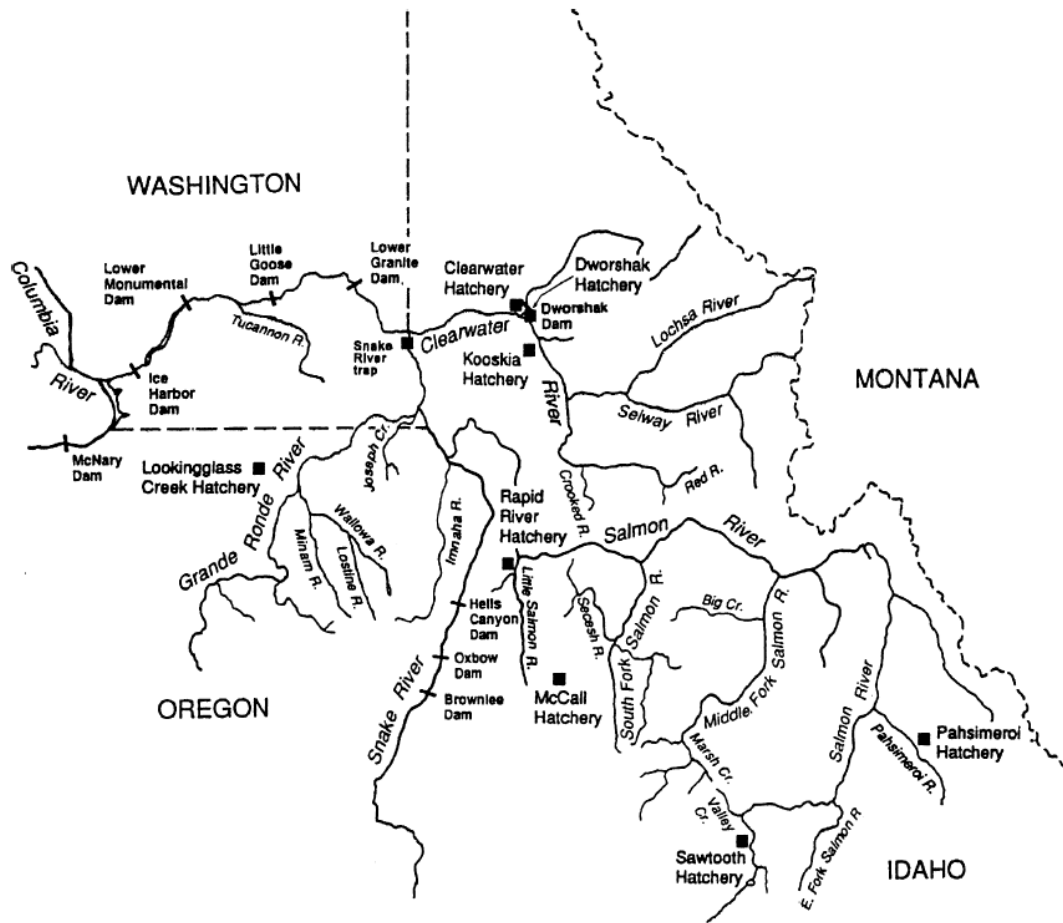
Figure 4.5: Map showing Snake River tributaries and the hatcheries from which juvenile salmonds were released for survival studies.

is given by

$$E_r = \frac{\rho}{\tau}\sqrt{\sigma_a^2 + \sigma_b^2 + \mu_a^2} \tag{4.38}$$

Simulations (figure 4.3) confirm that this model gives accurate predictions of encounter rates.

4.3   Application to juvenile salmon survival

The one-dimensional encounter rate model can be applied in a very natural way to survival analysis of migrating juvenile salmonids (*Oncorhynchus spp.*). Millions of juvenile salmonids in the heavily impounded Snake and Columbia rivers are consumed by predators (Rieman et al., 1991). The main piscivorous fish are the native northern pikeminnow (*Ptychocheilus oregonensis*), and three non-native species: smallmouth bass (*Micropterus dolomieu*), walleye (*Stizostedion vitreum*) and channel catfish (*Ictalurus punctatus*) (Poe et al., 1991). Significant avian predators include the Caspian tern (*Sterna caspia*), the double-crested cormorant (*Phalacrocorax auritas*) and several gull species (*Laridae spp.*) (Collis et al., 2001, 2002; Roby et al., 2005).

There is considerable concern about the impact of predation on salmon populations, especially since predation has likely increased since the construction of dams. Several reasons can be cited for this effect. Deeper, wider reservoirs provide more net habitat for predators, especially non-native ones, while warmer net temperatures increase their bioenergetic demands (Petersen and Ward, 1999; Petersen, 2001). Passage through hydroelectric systems can cause disorientation of the juveniles as well as predictable, spatial concentrations of prey at the tailraces (Rieman et al., 1991). Finally, the slower net flows in the reservoirs increase the residency time of the juveniles, increasing the probability of predation. This latter assumption, that the probability of survival is dependent exponentially on time of residency in a reservoir, has been an essential component in most models of fish passage survival models, such as the Columbia River Survival Passage (CRiSP, Anderson et al., 1996) and Fish Leaving Under Several Hypotheses (FLUSH, Wilson, 1994) models.

One accessory benefit of major anthropogenic transformations of the Columbia River has been the opportunity to obtain very detailed data on its denizens. Since the 1990s, hundreds of thousands of juvenile salmon have been implanted with individually identifiable PIT (passive integrated transponder) tags, providing an exten-

sive dataset of survival estimates and environmental covariates at many reservoirs. Anderson, Gurarie, and Zabel (2005) analyzed 287 groups of tagged juvenile spring chinook (*O. tschawytscha*) released at 17 different locations in tributary streams on the Snake River upstream from Lower Granite Dam in eastern Washington state between 1993 and 2003 (figure 4.5). The distances of the release locations ranged from 31 to 772 km. Survival estimates were obtained using the multiple-recapture model for single-release groups (Muir et al., 2001) and straightforward regressions of log-survival ($\log(S)$) against travel time and travel distance were performed (figures 4.6 and 4.7).

The relationship between log survival and migration travel time was generally very weak, with only five significant years (1996-1998, 2001, 2002 with $p < 0.1$) all of which displayed low correlations $r^2 < 0.5$. The relationship with distance was significant for all years except 1997 and the correlations were greater than 0.7 for six of the remaining 10 years.

Anderson et al. (2005) developed a model referred to as the *XT model*, arguing that the strength of the dependence on distance or time is controlled by the extent of the relative randomness in the movement of the predator and prey according to

$$S(x,t) = \exp\left[-\rho a \sqrt{x^2 + (\omega t)^2}\right] \tag{4.39}$$

where S is survival, a is a measure of the interaction area between predator and prey, ρ is predator density, and ω is defined as the "relative random velocity" of the predator-prey system. Small values of ω suggest a survival that is entirely dependent on distance traveled, regardless of the mean speed of travel. This model is referred to as the "gauntlet" model, since lack of random movement suggests that the prey must pass a fixed gauntlet of predators, regardless of the amount of time spent in the reservoir. The case where there is no net advection allows for multiple possible encounters between predator and prey and corresponds to something more similar
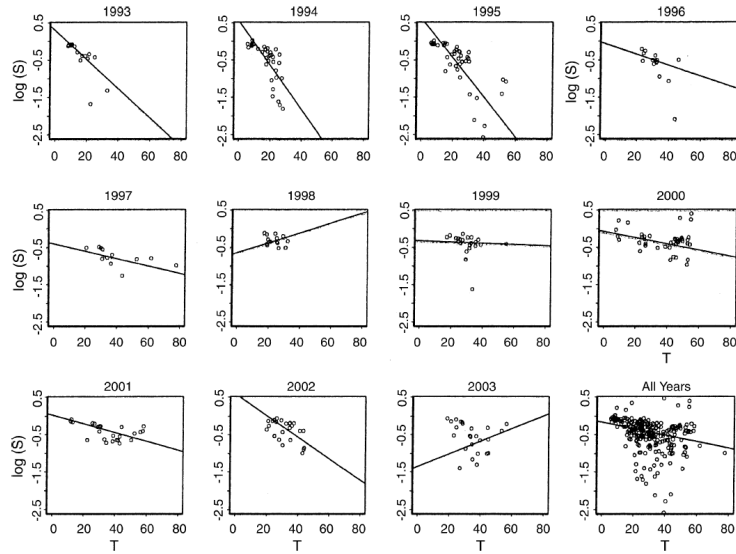
Figure 4.6: Regression of log survival, $\log(S)$, vs. migration travel time, T (days). Regressions are weighted by $1/\text{S.E.}^2$ of S (from Anderson, Gurarie, and Zabel 2005)
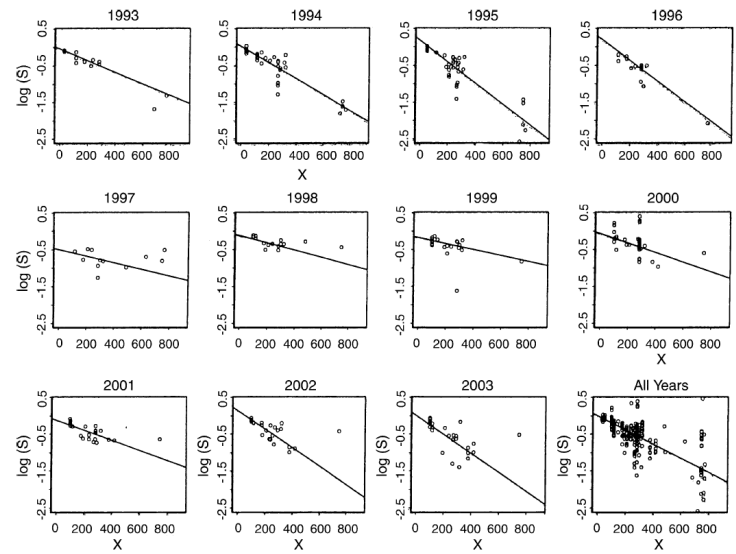
Figure 4.7: Regression of log survival, $\log(S)$, vs. migration travel distance, X (km). Regressions are weighted by $1/\text{S.E.}^2$ of S (from Anderson, Gurarie, and Zabel 2005)

to the ideal-free gas model implicitly assumed by the time-dependent Lotka-Volterra type predation models.

The estimated "random velocity" (on the order of 5 cm/s) was generally less than the mean advective velocity (mean 15 cm/s) for the data in the analysis. The broad implication of this result is that increasing flows is not theoretically expected to significantly mitigate survival of juvenile salmon, or at least not for the reasons worked into the passage survival models, namely reduction of residency time. The empirical evidence seems to largely bear this conclusion out.

To the extent that the XT model is tractable, intuitive, parameterizable with existing data and ecologically insightful, it can be considered a success. However, the meaning of the ω parameter, defined somewhat heuristically as "mean squared *random* encounter speed" in terms of actual parameters of individual movement remains somewhat unclear. Furthermore, a statistical analysis which separates time and distance is somewhat unsatisfactory since the two are clearly related via the mean velocity.

The relationship between velocity, distance and time is, in fact, less simple than might be expected. It can be expressed in terms of the mean distance traveled in a fixed time: $\bar{x} = \bar{v}\,t$ or as a mean time for arrival over a fixed distance: $\bar{t} = \bar{v}/x$. Migrating salmon disperse as they travel, and the rate of that dispersal is presumably somehow related to the ω parameter.[1] Interpreting survival in terms of the individually explicit one dimensional encounter rate models clarifies the relationships between travel distances, travel times and encounter rates.

4.3.1 Survival analysis

While the migrating and potential encounters salmonids through a reservoir occurs in three dimensions, certain simplifying assumptions allow us to essentially project

---

[1] For more detail on the interpretation, inference and meanings of travel times and spatial distributions of migrating organisms, see chapter 6.
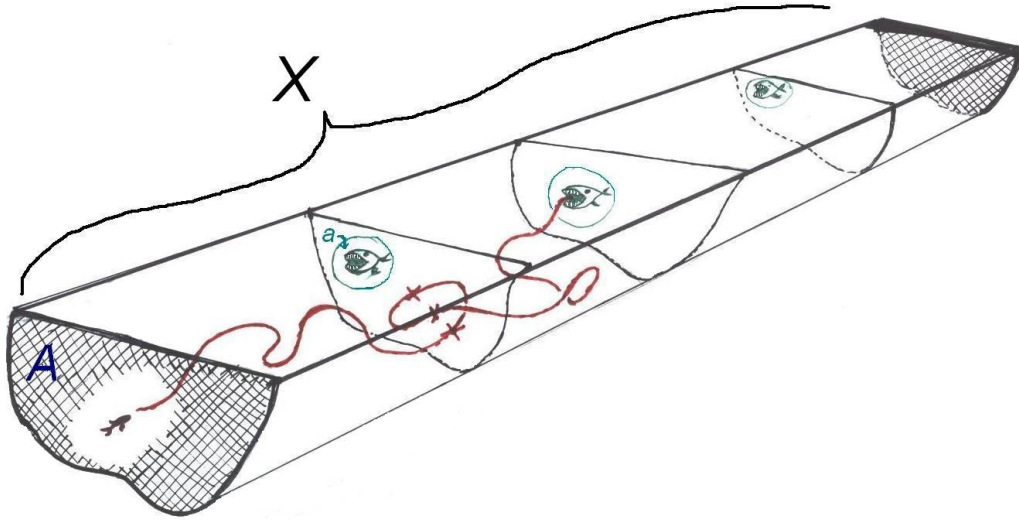
Figure 4.8: Schematic of salmonid passage survival through a field of stationary predators. The randomly moving fish survives three possible encounters with the first predator, only one unfortunate encounter with the second, and no opportunities for an encounter with the third.

the geometry of the problem onto the single natural dimension of the river's axis.

Consider a potential prey item moving through a reservoir of length X and cross-sectional area A containing N predators (figure 4.8). A basic mechanistic assumption is that the probability of a prey being consumed by any given predator is the ratio of the interaction area between predator and prey a and cross-sectional area of the reservoir A. The total probability of survival in the reservoir is given by:

$$
\begin{aligned}
S &= \prod_{i=1}^{N^\star}(1 - P\{\text{predation event}\}_i) \\
&= \left(1 - \frac{a}{A}\right)^{N^\star}
\end{aligned}
\tag{4.40}
$$

where N^* is the number of total encounters during its journey. If we consider that the predators are stationary and the prey is moving via an MST walk, then the number

of possible encounters per predator is given by equation (4.31)

$$N^* = N\sqrt{\left(\frac{\sigma}{\mu}\right)^2 + 1} \tag{4.41}$$

Substituting in $N = \rho A X$ and rearranging we obtain

$$S = \left[\left(1 - \frac{a}{A}\right)^A\right]^{\rho X \sqrt{\left(\frac{\sigma}{\mu}\right)^2 + 1}} \tag{4.42}$$

If we assume $a/A \ll 1$, this expression reduces to

$$S(X) \approx \exp\left(-a\,\rho\,X\,\sqrt{\left(\frac{\sigma}{\mu}\right)^2 + 1}\right) \tag{4.43}$$

Expressed in this way, survival probability can be viewed as being dependent only on the distance traveled and a ratio of random and directed length scales. Since the average velocity of a cohort of fish is given by $V = \mu/\tau$ and the mean time spent in the reservoir is given by $T = X/V$ equation (4.43) can be recast as

$$S(X) = \exp\left(-a\,\rho\,\sqrt{\left(\frac{\sigma}{\tau}T\right)^2 + X^2}\right) \tag{4.44}$$

This form corresponds exactly with equation (4.39), and the "random velocity" ω is revealed to be the ratio σ/τ, i.e. the ratio of the length and time scale of independence. For both randomly moving predators and prey, the equivalence becomes

$$\omega^2 = \sqrt{\frac{\sigma_a^2 + \sigma_b^2}{\tau^2}} \tag{4.45}$$

where σ_a and σ_b are the length scales of independence of prey and predator respectively and τ is a common time scale of independence. The relationship between the scales

of randomness of predator and prey are illustrated in figure 4.9.

Interpretation of survival model

This formula has some ecological and behavioral interpretations. Since it is not in the interest of the migrating smolt to be captured, there is little incentive for migrating fish to increase the randomness of their migration, as it is in the prey's interest to drive σ_a to 0 and drive the system as close as possible to a gauntlet-like situation. The caveat to this strategy is that prey items themselves might be foraging themselves on smaller prey and attempting to increase their encounter rates in turn. The actual extent of foraging among outmigrating salmon varies considerably between species (Quinn, 2005), leading to substantially different migratory behavior (see chapter 6).

The onus on increasing encounters lies with the predators, either by increasing its σ_b or encounter radius. The first strategy is constrained by the energetic costs associated with active foraging. If passive encounter rates with no random movements are sufficient to fulfill the needs of a predator, then a stationary "sit-and-wait" strategy is sufficient.

Conceptually, this model gives a framework for classifying in a quantifiable manner a broad range of potential predator strategies. For example, a gliding bird low has a relatively low energetic to searching, and its σ_b can be quite high. On the other hand, fish predators like the northern pikeminnow, are more likely to be satiated by sit-and-wait strategies. When the flux of passing prey is too low, many have alternative foraging options to switch to, such as feeding on benthic crustaceans and resident fish Naughton and Bennett (2003). Certain predators such as orb-weaving spiders don't need to move at all, concentrating on increasing their encounter radius and relying on the sufficiency of passive prey flux.
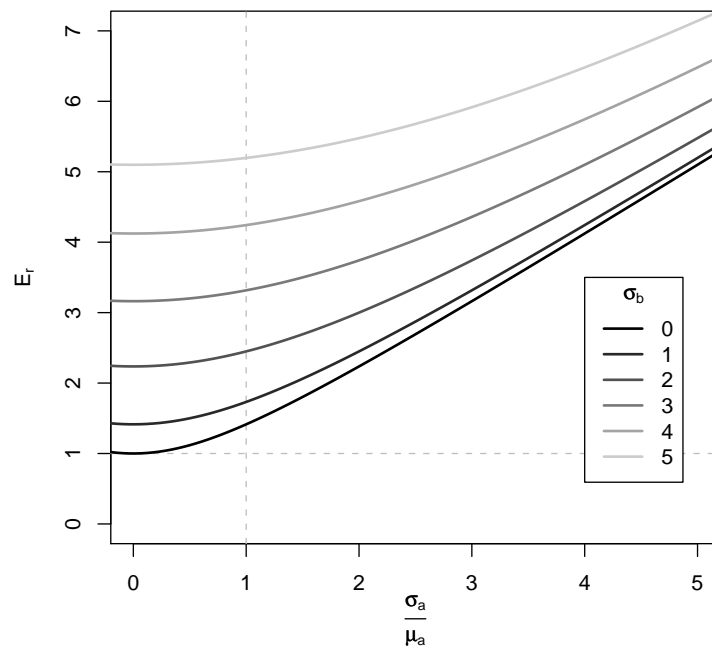
Figure 4.9: Theoretically expected number of encounters per predator at different prey and predator distance scales of indepedence (σ_a and σ_b respectively). It is generally in the prey's interest to keep the encounter rates low by driving σ_a to zero, but the predator can increase encounter rates itself by increasing σ_b.

4.3.2 Relating encounter velocity to spread of migrating populations

Intuitively, the random velocity that governs encounter rates (ω) should be related to the rate of spread of a migrating cohort of random walkers (see sections 2.3 and 6.3.2), defined by

$$f(x|t, v, D) = \frac{1}{\sqrt{\pi Dt}} \exp\left(-\frac{(x - vt)^2}{Dt}\right) \tag{4.46}$$

where $f(x)$ is either the spatial distribution of a population of identical random walkers or the probability distribution of a single walker at time t, v is the advective velocity and D is a standard diffusion rate of spread in units of distance squared per time. This diffusion rate D is given from equations (2.17) and (2.18) , the rate of diffusion D of a population is simply

$$D = \frac{\sigma_a^2}{\tau} \tag{4.47}$$

Arrival time data at a detection site can be modeled using used the inverse Gaussian distribution, from which estimates for D are straightforward to obtain (see Zabel (1994); Zabel and Anderson (1997) and chapter 6 for details). For spring chinook in the Lower Goose reservoir, for example, estimates for D were on the order of 5 to 4.6 to 12.5 km $\cdot$ day$^{-0.5}$ (see Zabel and Williams, 2002, Table 1).

From equation (4.45) the net random velocity of encounter (ω) is related to the individual parameters of movement by

$$\omega^2 = \frac{\sigma_a^2 + \sigma_b^2}{\tau^2}. \tag{4.48}$$

According to the XT-model paper, estimated values for ω were on the order of 4 to 8 km/day. Equating equations (4.48) and (4.48) and rearranging yields the following

relationship for the time-scale of independence τ

$$\tau = \frac{D}{\omega^2} \left(\frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2} \right) \tag{4.49}$$

The *minimum* value for τ is obtained when there is no random movement of the predator, and the ratio D/ω^2 yields a lower limit on τ. Using the reported values for D and ω into (4.49) results in time scales of independence with a lower bound ranging from 0.5 to 2 days. If the random movement of the predator is roughly on the order of the random velocity of the prey then the estimate of τ increases correspondingly.

At first glance, this estimate for τ as a time scale of independence for the movement of an individual fish seems quite large. Several factors may give rise to this curious result. First, recall that τ is the unit of time at which subsequent displacements for the migrating organism are independent. Juvenile salmon display strong diurnal patterns during their migration: moving primarily at night and presumably holding station during the day (Zabel and Williams, 2002; Quinn, 2005). Furthermore, due to daily fluctuations in power demand, there is a strong one day cycle in the flow regimes of the reservoirs. Thus, in this context of the diffusion process, a one day scale of independence might be reasonable.

Furthermore, estimates of D are inflated, sometimes significantly, by the assumption in the inverse Gaussian model of homogeneous populations. For certain populations, the bulk of the spread in arrival times can be explained by a model in which fish move more or less deterministically, but with a population-level variance in mean travel velocity. The effect of population-level heterogeneity on models of migration are discussed in detail in chapter 6.

4.3.3 Discussion

The major theoretical contribution of the one-dimensional models is a rigorous unifi-cation of advection-diffusion and the XT survival model in terms of individual, mea-

surable movement parameters. Practically, the results make clear why survival of migrating smolt is, indeed, more gauntlet-like and distance dependent than a classic reaction-type predation model might predict. The exercise of thinking about survival in terms of individual movements is a useful one, for it clarifies the essentially simple fact that time dependence only plays a role in survival if there is a possibility of multiple encounters between predators and prey. It is highly unlikely and rarely, if ever, observed that juvenile salmon spend any significant effort swimming upstream (figure 4.8 notwithstanding), and the theory provides a vivid explanation why. In the Columbia river system, the onus on having multiple encounters and increasing the probability of encounter success lies with the predators. In those situations in which the flux of juvenile salmon prey passing is sufficient to energetically satisfy the predators, we can predict that a wholly distance-dependent gauntlet system is in effect. In this case, increasing flows (a common management scheme to mitigate survival) will probably have little effect on encounter rates. However, as soon as the predators begin significant random foraging movements themselves, the travel speed of the prey suddenly does have an impact on survival. Thus, according to this theory and in support of conclusions in Anderson et al. (2005), *active foraging of predators is a requisite condition for flow increases to have a beneficial effect on juvenile salmon survival.*

While PIT-tag data is voluminous and excellent, it necessarily aggregates everything that happens in a reservoir into a single travel time and survival estimate. However, actual reservoirs are heterogeneous environments, swimming and predation behavior are complex and little understood, and the distribution of predators and migration of prey are both rather patchy. In Appendix C, I provide a brief analysis of migration and survival of juvenile salmon in the mid-Columbia river using hydroacoustic data which provides a crude idea of the differences between mid-reservoir, tailrace and forebay predation. The differences between survival rates in different regions are striking and significant, not only within a reservoir but between species

(figure C.5 and C.2) as are the travel times in different sections of the river (figure C.2). The variability in survival between sections can be quite striking. For example, nearly 10 times more sockeye died in the tailraces or passing through two of the dams than in any of the mid-reservoirs (figure C.5).

In light all of these environmental heterogeneities, it is unclear how to interpret parameters estimated over an entire aggregated reservoir. I would suggest, however, that the insight provided by the models can inform recommendations for future exploration of the system. Currently, a massive effort is underway in finding correlations and explanatory factors for salmon survival in terms of all sorts of reservoir-wide covariates (flows, spills, temperatures, turbidity, dissolved gas content). Large projects to individually track acoustically tagged salmon in three dimensions as they approach and pass through the dam are in effect. Juveniles are dutifully inspected by hand for blemishes as they pass through multi-million dollar fish passage systems. The fish have been genetically sequenced, transported in barges, tagged and clipped every which way. But most models of survival still rely on murky time-dependent functions of survival. There is an increasingly growing body knowledge about the distribution and movement behavior of redator and prey in the river itself (R. W. Zabel, NOAA Fisheries, *pers. comm.*), but these data have yet to be actively implemented in understanding proceses of survival.

If I were to make concrete recommendations to managers on the Columbia River concerned about mitigating the survival of juvenile salmon, I would suggest that considerably more effort be devoted to understanding sub-reservoir dynamics, predator distributions and movement behavior of all the agents. This sort of information is difficult to obtain and would necessarily involve some elaborate and clever monitoring efforts. But it would greatly facilitate the realism and informativeness of survival models from a process- and biology-based perspective in the Columbia River. In this context, the rigorous development of the encounter models presented here can serve an important role.
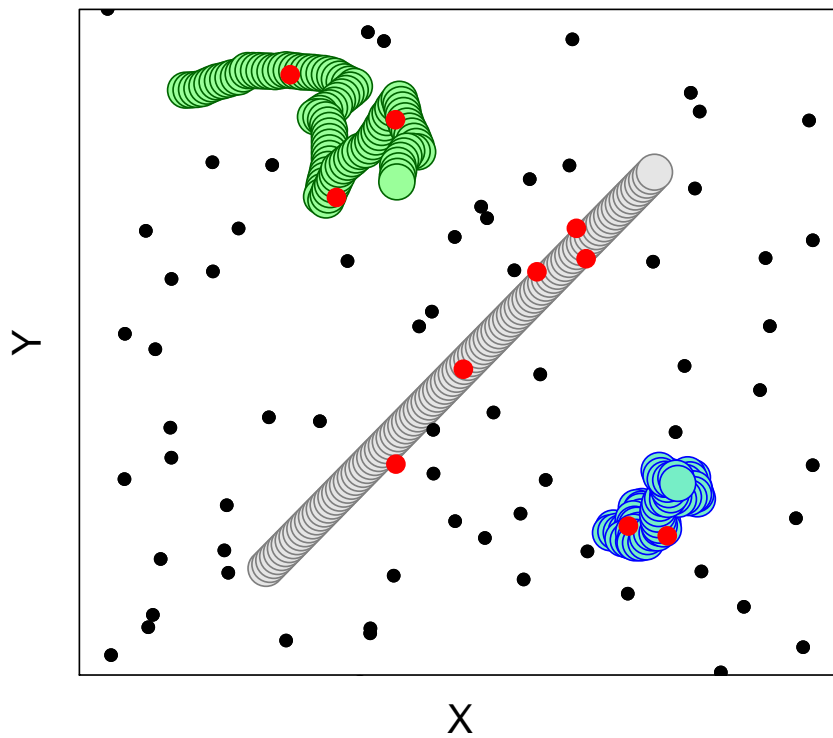
Figure 4.10: Three trajectories showing the effect of randomness on encounter rates. All three movements consist of 100 steps of length 1. The grey walker is moving linearly, and has the highest encounter rate. The green walker is walking with relatively high correlation between turning angles, and the blue walker displays very low auto-correlation.

4.4 Two-dimensional encounter rate models

Consider a walker moving in a straight line through a two-dimensional field of stationary targets of density ρ at constant velocity v with an encounter radius a. An encounter occurs whenever the detection area of the walker includes a target. The number of encounters is given by $N(t) = \rho A(t)$ where $A(t)$ is the area swept out by the movement: $A = 2avt$. Thus the encounter rate is

$$E_r = 2\rho av \tag{4.50}$$

For random walkers, this expression ceases to hold. Intuitively, the straighter a trajectory, the closer an encounter rate estimate will be to the deterministic model above. A more tortuous path, with slower dispersal and higher probability of reencounters, will have lower encounter rates (see schematic in figure 4.10). The only attempts I am aware of to consider the impact of random movements on encounter rates are simulation studies performed by Bartumeus et al. (2002) comparing the encounter rate efficiency of "Brownian walks" to "Lévy walks" and by Hutchinson and Waser (2007) where the effect of a correlated random walk on expected ideal-free gas type encounter rates was numerically assessed. The latter authors conclude, somewhat pessimistically, that: "It is apparent that turning angle distribution, detection distance, step length and step number all have interacting non-linear effects, so quantitative predictions are possible only using simulation."

Indeed, mathematical predictions of encounter rates depend to some extent on the way in which random movement is modeled. A greatly simplifying assertion can be made that the identification of the time and length scales of independence (i.e. the MST parameters) largely determine encounter rates. This is supported by the observation made in section 2.3.2 that the actual total distance swept out by a homogeneous random movement is unlikely to be more than 1.5 times greater than

the MST decomposition of the movement (see figure 2.8). Thus, many of the factors cited by Hutchinson and Waser collapse into the essential parameters of an MST walk, and quantitative predictions are possible after all.

4.4.1 Derivation

The following derivation assumes a searcher moving with a non-correlated random walk with characteristic time and step-length scales τ and σ in a uniform random field of predators. Targets are distributed randomly according to a Poisson process with density ρ such that the expected distance between targets is much smaller than the detection range a of the searcher. We are interested in comparing the nature of the movement rather than the effect of absolute velocity. Thus, we fix the absolute speed of the movement $\sigma/\tau = 1$. It is important to note that small values of σ and large values of σ correspond to movements whose speed, and, presumably, energetics, are identical.

Analogously to the one-dimensional case, encounter rates in two dimensions are related to the ratio of the area swept out by a movement to the potential area explored. This ratio is controlled by the relationship between the length scale of movements and the distance between targets. For uniformly distributed targets with density ρ in two dimensions, the mean distance between neighboring targets is $\lambda = \frac{1}{2\sqrt{\rho}}$. If the step length σ is on the order of λ or greater, the encounter rate asymptotically approximates

$$E_r(\sigma|\sigma \geq \lambda) = 2a\rho. \tag{4.51}$$

This expression is equivalent to equation (4.50) with the "random speed" $\sigma/\tau = 1$ replacing the velocity term.

The limiting case where $\sigma \to 0$ is an infinitesimal Brownian motion, for which expected displacement scales as the square of time rather than linearly with time.

When the step lengths are much smaller than characteristic distance λ, the path

to any given target becomes highly tortuous, such that the time needed to traverse a distance λ scales as t^2 rather than t and instead of a linear relationship with velocity σ/τ, the encounter rate is proportional to σ^2/τ, such that

$$E_r(\sigma|\sigma \ll \lambda) = 2\,k\,a\,\rho\,\sigma, \tag{4.52}$$

where k is an empirical slope parameter that describes the encounter rate limit for highly tortuous paths. It should be noted that the encounters considered here are destructive. Otherwise, a highly tortuous movement that disperses slowly continuously reencounters a given target thereby inflating the apparent encounter rate.

A function that fits both of these asymptotic behaviors is the von Bertalanffy curve

$$E_r(\sigma) = 2a\rho(1 - e^{-k\sigma}). \tag{4.53}$$

It is straightforward to estimate the k parameter from simulated data. I used a non-linear least squares fitting routine in R.

The model can readily be scaled into real time by dividing by τ, such that a final encounter rate model for a non-directed randomly moving searcher in a uniform field of targets is given by:

$$E_r(\sigma) = 2a\rho\frac{\sigma}{\tau}(1 - e^{-k\sigma}) \tag{4.54}$$

For the case where the targets are also moving with step length σ_b at time scale τ, the encounter rate is adjusted in a manner analogous to the one-dimensional case (4.38), such that

$$E_r(\sigma) = 2a\rho\frac{\sqrt{\sigma_a^2 + \sigma_b^2}}{\tau}(1 - e^{-k\sqrt{\sigma_a^2 + \sigma_b^2}}). \tag{4.55}$$

This result was obtained by analogy to the one-dimensional case: from the perspective

of the random searcher, all of the targets are moving with a random velocity that adds their movement with the searcher's own movement. A rigorous derivation of these results is left as a future project.

4.4.2 Simulations

I performed many thousands of simulations to test these results, seeding a virtual environment with targets at density $\rho = 0.001$, interaction distance $a = 1$ and measured encounter rates for random walkers with step-lengths σ varying from 0.4 to 20 (figure 4.11). I fitted equation (4.53) to the resulting encounter rates, and obtained an estimated value for k of 0.316 (95% C.I.: 0.302, 0.331).[2] The fit is excellent and yields good predictions for diverse values of a and ρ.

4.4.3 Discussion

One contribution of the two-dimensional encounter rate model is to illustrate the use of MST parameters in reducing an otherwise complex problem of encounter rates to tractable and relatively simple expressions. Practically, encounter rate models based on individual movements are difficult to compare to data since data that combines both individual movements and observations of encounters is rare. However, several fundamental theoretical principles can be made regarding foraging strategies.

Perhaps the most important principle is that *when searching randomly for sparcely distributed prey, the fundamental length scale of movement should be on the order of or greater than the distance between targets.* An upper limit on the scale of the movement is the size of the greater patch. Thus, the intrinsic scales of a movement behavior of a foraging organism provides some indirect information on the distribution of the prey. Often, the movement behavior of forager can be measured while much less is known about the prey items. The best example of this is marine predators, such as pinnipeds

---

[2]This value is tantalizingly close to $\frac{1}{\pi}$. A more satisfactory derivation of the encounter rate awaits further work.
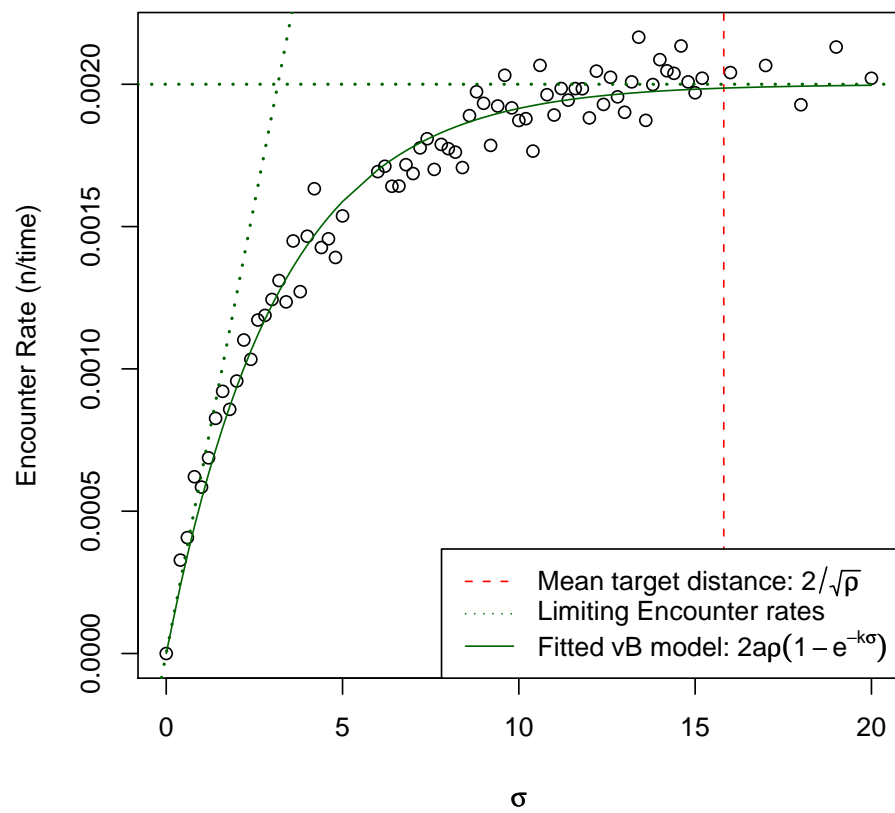
Figure 4.11: Simulation results of 2D encounter with a Gaussian walker moving through a field of targets with density $\rho = 0.001$ and encounter radius $a = 1$.

and seabirds, many of which perform distinctly identifiable foraging behaviors on prey such as fish, squid and krill, whose movements are largely mysterious. A quantification of the predators' movement patterns - specifically, an identification of the length and time scales of movement, might suggest a rough figure for the size of a food patch and for the density of the items within it. This fundamental foraging strategy can be relevant at larger scales which characterize distant aggregation of foraging food as well. Thus, the hierarchical, complex foraging movements observed for albatross (Fritz et al., 2003; Grünbaum and Veit, 2003) can be explained by adapting their characteristic length scale of movement to the relevant hierarchical level of searching, whether between patch or within patch.

It should be noted that under the assumption that "reasonable" foraging movements will have a length scale of movement on the order of the neighbor distance, the second term in equation in 4.55 can be neglected, and a tractable, simple, useful encounter rate formula reduces to

$$E_r(\sigma) = 2 \, a \, \rho \frac{\sqrt{\sigma_a^2 + \sigma_b^2}}{\tau}. \tag{4.56}$$

A second principle of foraging that arises from these models is that *when searchers and targets are both moving, random encounter rates increase roughly linearly with their net velocities.* The consequences of this second principle is very case specific. Often, it will be to the advantage prey to minimize extraneous movements, unless it too is foraging. In some situations, it may be energetically sufficient for a predator to merely sit and wait for the flux of prey to encounter them. Certainly, that is the strategy of web-weaving spiders (various *spp.* in Aranae), who additionally increase their encounter rate by creating large and highly effective encounter radii. Among certain plankters, residing in an environment where ambient movement forces are often greater than their own ability to move, it has been shown that increased turbulence can significantly increase encounter rates (Marrasé et al., 1990; MacKenzie and

Kiorboe, 1995).

For plankton, larvae and other small aquatic organisms, encounter rates govern many aspects of survival in a highly dynamic environment which can often be legitimately considered random. They are, in some ways, ideal candidates for applying refined encounter rate theories. Higher-level predators, in contrast, rely on experience, subtle interpretation of environmental cues, adaptability and, often, strategy and tactics when foraging. However, any search must contain some component of randomness, and physiological and bioenergetic constraints and prey distribution invariably the kinds of movement. The mathematical principles behind encounter rates presented in this chapter can serve at the very least to guide the interpretation of observed movements and, perhaps, to inform higher-level questions of adaptation and evolution of organisms to successfully exploit various kinds of dynamic foraging niches.

Chapter 5

BEHAVIORAL CHANGES IN GAPPY ANIMAL MOVEMENT DATA

It may be but an idle whim, but it has always seemed to me, that the
extraordinary vacillations of movement displayed by some whales ...
indirectly proceeds from the helpless perplexity of volition.

Moby Dick

Hermann Melville (1851)

Summary

Individual animal movement data is collected at an increasing rate as remote-sensing
technology develops. At its best, analysis of the data can suggest mechanisms by
which organisms exploit a heterogeneous and variable environment. Unfortunately,
movement data are multidimensional, non-independent and almost always suffer from
the twin banes of measurement error and gappiness, making appropriate analyses far
from straightforward. While error in the data can be accounted for by state-space
models, an increasingly popular statistical approach, gappiness (i.e. irregular inter-
vals between times of measurement) has not been well-addressed in the literature.
Gappiness is particularly widespread in data on marine organisms, for which remote
sensing depends on a combination of transmitter exposure and satellite presence. I
suggest a method of dealing with gappiness by identifying a persistence component
of movement that can be modeled as a continuous auto-correlated stochastic process
with three parameters: a mean, a variance and a continuous auto-correlation coef-
ficient. I then develop several methods for identifying changes between behavioral

modes. A single breakpoint can be found with a maximum likelihood estimation over an entire gappy time-series, while multiple breakpoints can be found by sweeping a window and identifying which of the parameters change at any given breakpoint. After demonstrating the robustness of the method with simulation, I apply the routine to GPS data collected on a northern fur seal (*Callorhinus ursinus*) in the Kuril Islands in Russia. The resulting model suggests a complex behavioral profile that can potentially be analyzed against environmental covariates.

5.1 *Introduction*

In recent years, there has been a rapidly growing body of work devoted to the detailed study of animal movements in the wild, mirroring the increase in the technological ability to accumulate data. Most fundamental ecological processes are a direct result of movement, including foraging success, breeding success, migrations and dispersals. Furthermore, individual animal movements are a measurable behavioral response to a combination of internal states, physiological constraints and environmental factors. Informative analysis of movement data can yield sophisticated insights into the behavioral mechanisms that allow organisms to exploit temporally variable and spatially heterogeneous environments. This appears to be a common goal of many animal movement researchers.

Analysis of movement data, however, is far from straightforward, since the data is multi-dimensional and auto-correlated in space and time. A common approach to modeling a movement track is to apply some variety of a correlated random walk models (Skellam, 1951; Turchin, 1998; Okubo and Levin, 2001), which typically hypothesizes some distribution of step-lengths and turning angles clustered around zero degrees. Over a long enough time-scale, multiple behaviors can be captured by a single dataset and the properties of mixed random walk models have been explored (Grünbaum, 2000; Skalski and Gilliam, 2003). Recently, models have been constructed that successfully relate changes in correlated random walk parameters to environmen-

tal and landscape features (Morales et al., 2004; Forester et al., 2007; Haydon et al., 2008; Aarts et al., 2008).

There are two common features of movement data which complicate the straightforward application of correlated random walk models. The first is error in the measurement process. A fruitful body of research has emerged recently that addresses measurement error with the use of state-space models (SSM's) (Jonsen et al., 2003, 2005; Royer and Gaspar, 2005; Patterson et al., 2008). State-space models are efficient frameworks for parsing movement data into a process model of movement and an observation model that accounts for error. They have been, in particular, invaluable in the interpretation of ARGOS-satellite derived telemetry data.

The second significant issue with movement data is irregular timing of measurements, sometimes referred to as "gappiness". This issue is particularly important for tagged marine species such as pinnipeds, whales, penguins, sea turtles and large pelagic fish, where sending a signal requires that the tag be exposed to the air and that a connection be made with the receiving satellite. Clearly, a straightforward estimation of turning angle distributions and estimated velocities is compromised by the irregularity of the data. The solution to this problem is to model the movement data as an irregular subsampling of a continuous auto-correlated process (Johnson et al., 2008, *in press*).

In this paper, I present an efficient and robust method for identifying behavioral changes in a gappy movement dataset. The fundamental idea is to sweep an analysis-window over the movement data and identify the time and nature of any significant behavioral shifts, outputting estimates of descriptive parameters and aggregating them into a behavioral summary of the movement.

The complete analysis method involves several steps. I begin by suggesting a natural decomposition of velocity and turning-angle data into persistence and turning velocity components. These velocity components have the statistically attractive feature of being locally stationary and Gaussian, crucially allowing for the straight-

forward modeling of an auto-correlation structure. I consider these time-series to be samplings from a continuous auto-regressive process and present a straightforward likelihood-based method to determine a continuous auto-regression coefficient, a parameter that has an important biological interpretation. The movement can then be described by three parameters for each of the two velocity components: a mean, a variance and the continuous auto-correlation. Next, I present a maximum likelihood method to estimate the location of a single change point, i.e. the likeliest time at which one or more of the three parameters that describe the time-series changes values. It is of interest to be sure that the change is a "real" one and not an artifact of randomness, as well as to identify which of the parameters are changing. Both of these challenges are met by applying an information-based criterion to test all possible permutations of model parameters changing. A simulation study demonstrates that the Bayesian Information Criterion (BIC) does an excellent job of identifying which of the parameters are really changing their value at a given most likely changepoint. I provide code for all of the relevant pieces of the analysis algorithm in Appendix D.

The complete analysis machinery was applied to GPS-derived data on movements of an adult Northern fur seal female (*Callorhinus ursinus*) in the Kuril Islands in Russia. After giving birth, female fur seals nurse their pups for several weeks before beginning to take extended foraging trips up to seven days in length. They can wander up to a hundred kilometers from the rookery of birth in pursuit of sufficient forage to fulfill both their own energetic requirements and that of the nursing pup. These foraging trips show distinct modes of movement, including rapid travel, periods of frequent diving and semi-somnolent drifting. The application of the analysis method allows for the identification of a complex behavioral profile.

5.2 Models of movement

We take a "Lagrangian" approach to modeling movement, that is, we place the organism at the center of our analysis. Thus, rather than analyze the absolute positions

(X_i, Y_i) or compass orientation (Φ_i), we examine the two variables that the organism controls: speed V and turning angle Θ. Because we are specifically interested in analyzing irregularly sampled data, we have an additional vector of times of measurement T_i and define these variables as

$$V(T_i) = \sqrt{(X_i - X_{i-1})^2 + (Y_i - Y_{i-1})^2}/(T_i - T_{i-1}) \tag{5.1}$$

$$\Theta(T_i) = \Phi_i - \Phi_{i-1} \tag{5.2}$$

The statistical properties of these variables will depend on the time-interval between measurements, which we define as $\tau_i = T_i - T_{i-1}$.

A very common Lagrangian method for modeling movement data is the two-dimensional correlated random walk (CRW). A traditional CRW, which we might further specify as an *unoriented, homogeneous* correlated random walk (see Chapter 2). A CRW is defined as a discrete movement in which step-lengths R_t are independent and have some positive random distribution and the turning angles θ_t have some non-uniform distribution clustered around 0 degrees (Kareiva and Shigesada, 1983; Marsh and Jones, 1988). Common distributions for velocities include the Weibull and log-normal, while turning angles are often modeled with wrapped Cauchy distribution with zero mean and clustering parameter κ: When $\kappa = 0$ the angles are distributed uniformly between $-\pi$ and π, and when $\kappa = 1$, all the angles are concentrated at 0 (Fisher and Lee, 1994). A CRW can generate smooth tracks that do a good job of resembling real animal movement data, and the parameters of the model are readily estimable. For example, $\widehat{\kappa}$ is just the expected value of the cosine of the turning angles.

Whatever the distribution used for turning angles has been to model movement, the applications of the CRW are almost always drawn from independent distributions and auto-correlation between subsequent turning angles or step lengths are rarely discussed. Nonetheless, CRW's have proven to be useful and tractable models for

unoriented, homogeneous movements.

It is, however, not clear how to fit a CRW to gappy data, since neither $V(t_i)$ nor $\theta(t_i)$ are identically distributed when the interval between measurements changes. There are two feasible approaches for dealing with this problem. The first is to calculate or approximate the expected distributions of R and Θ as a function of the gap (see section 2.2). The second, and the one that is addressed in this report, is to transform the data by decomposing every step into orthogonal components of persistence velocity $V_p(t)$ and turning velocity $V_t(t)$:

$$V_p(t) = V(t)\cos(\Theta(t)) \tag{5.3}$$

$$V_t(t) = V(t)\sin(\Theta(t)) \tag{5.4}$$

V_p captures the tendency of a movement to persist in a given direction, and the velocity of that movement while V_t captures the tendency of movement to head in a perpendicular direction at a given interval. Thus, the primary descriptive features of movement, namely speed, directional persistence, and variability are captured in these variables. Empirical explorations of movement data via histograms or qqnorm plots suggest that both V_p and V_t are well approximated by mixed normal distributions, with V_t notably always having a mean very close to zero. This appealing statistical property allows for the application of an arsenal of analysis techniques for characterizing auto-correlated data.

Though geometrically orthogonal, these two variables are not necessarily independent. Because, however, our eventual technique is ultimately descriptive and the interpretation of each of these variables is somewhat unique, we choose to analyze them separately. An investigation of their correlation structure is the subject of future work.

Simulated Gappy Time Seris

Log−likelihood profile
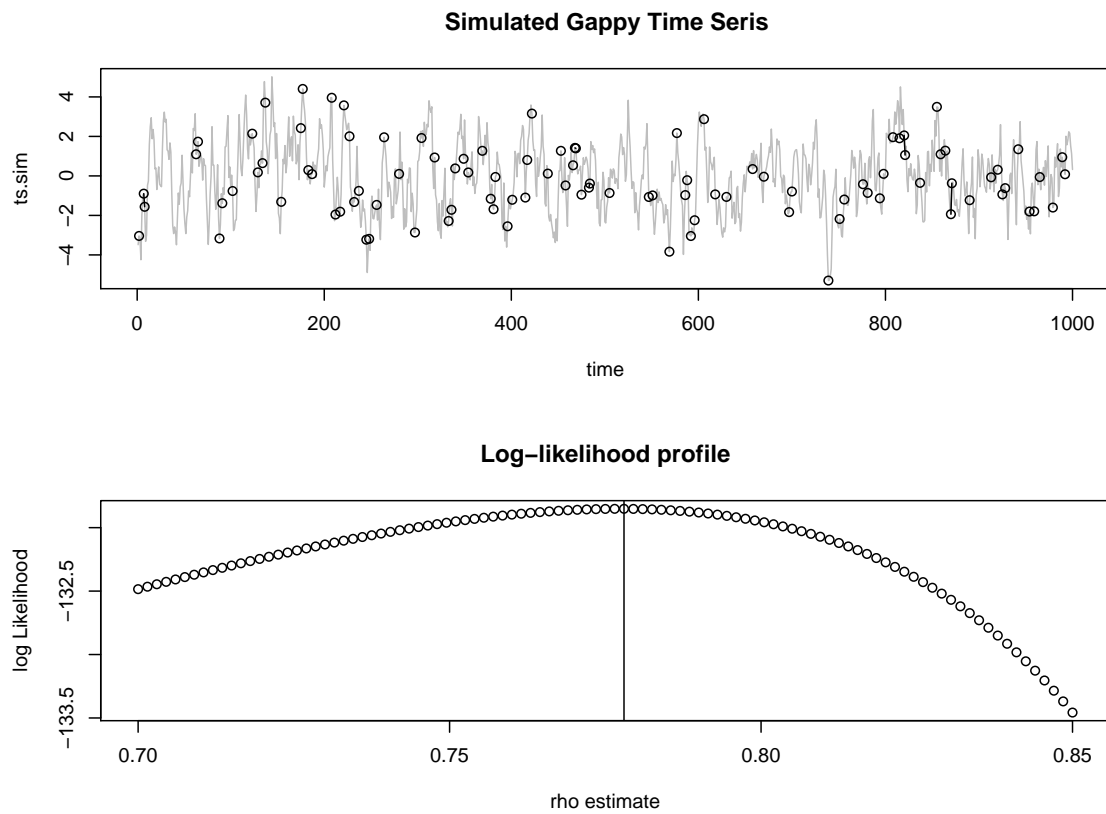
Figure 5.1: A simulated full AR(1) time series with auto-correlation coefficient $\rho = 0.8$ in plotted in gray on the top graph. The dots indicate a sampling of 100 points from this time-series. The lower plot shows is the log-likelihood profile over ρ for the sampled data. The estimated value for ρ in this simulation is 0.778

5.3 Defining correlation structure in gappy time-series

5.3.1 Gappy time-series model

We begin the analysis with the empirically supported assumption that the persistence velocity V_p (eq. 5.3) is a sample from a generic, stationary, Gaussian continuous space, continuous time process $X(t)$ which has the following properties:

$$
\begin{aligned}
X(0) &= X_0 \\
\mathrm{E}\left[X(t)\right] &= \mu \\
\mathrm{Var}\left[X(t)\right] &= \sigma^2 \\
\mathrm{Corr}[X(t), X(t-\tau)] &= \rho^\tau,
\end{aligned}
\tag{5.5}
$$

where $0 < \rho < 1$ is the first order auto-correlation at a time lag 1, i.e. at whatever units the time is measured. (Note, all subsequent discussion applies analogously to V_t, with the constraint that $\mu = 0$). Consider observations X_i that are made at times t_i, where $t_0 = 0$ and beginning at initial observation X_0. The series X_i can be described as

$$
X_i = \mu + \rho^{\tau_i}(X_{i-1} - \mu) + \epsilon_i,
\tag{5.6}
$$

where $i \in \{1, ..., n\}$, $\tau_i = t_i - t_{i-1}$ is the time interval between subsequent observations, ρ^{τ_i} is the auto-correlation as a function of the time gap and ϵ_i is a stochastic error term. Given the constraints in 5.5, ϵ_i can be shown to have mean 0 and variance $\sigma^2(1 - \rho^{2\tau_i})$. The variance is obtained as follows:

$$\begin{aligned}
\text{Var}\,[\epsilon_i] &= \text{Var}\,[X_i - \rho^{\tau_i} X_{i-1}] \\
&= \sigma^2 + \rho^{2\tau_i}\sigma^2 - 2\rho^{\tau_i}\text{Cov}[X_i, X_{i-1}] \\
&= \sigma^2(1 - \rho^{2\tau_i}) \tag{5.7}
\end{aligned}$$

In order to estimate the continuous correlation function ρ, we need to express the likelihood function for the entire process. Since any X_i depends only on the previous observed value X_{i-1}, a total conditional likelihood can be written

$$L(\rho|\mathbf{X}, \mathbf{T}) = \prod_{i=1}^{n} f(X_i|X_{i-1}, \tau_i, \rho), \tag{5.8}$$

where $\mathbf{X}$ and τ represent the vector of observations and times of observation respectively and the distribution function f is the probability density function of the conditional distribution $X_i|X_{i-1}$. For the process (5.6) with a Gaussian error structure, f is given by

$$f(X_i|X_{i-1}) = \frac{1}{\sigma\sqrt{2\pi(1 - \rho^{2\tau_i})}} \exp\left(\frac{(X_i - \rho^{\tau_i}(X_{i-1} - \mu))^2}{2\sigma^2(1 - \rho^{2\tau_i})}\right). \tag{5.9}$$

This likelihood (5.8) is smooth and easily maximized over the range of possible values for $0 < \rho < 1$. The estimate for ρ can be expressed as

$$\widehat{\rho} = \operatorname*{argmax}_{\rho} \ L(\rho|\mathbf{X}, \mathbf{T}) \tag{5.10}$$

To test this estimation routine, we simulated an AR(1) process with coefficient $\rho = 0.8$ and length $n = 1000$ (see figure 5.1). We then randomly sampled 100 observations from this process and estimate $\widehat{\rho}$ based on the gappy time-series. Performing this operation 100 times yielded a mean estimate $\widehat{\rho} = 0.797$ with standard error 0.052. The MLE estimate works excellently, even for extremely gappy data.

Penguin velocity time series

Log−likelihood profile
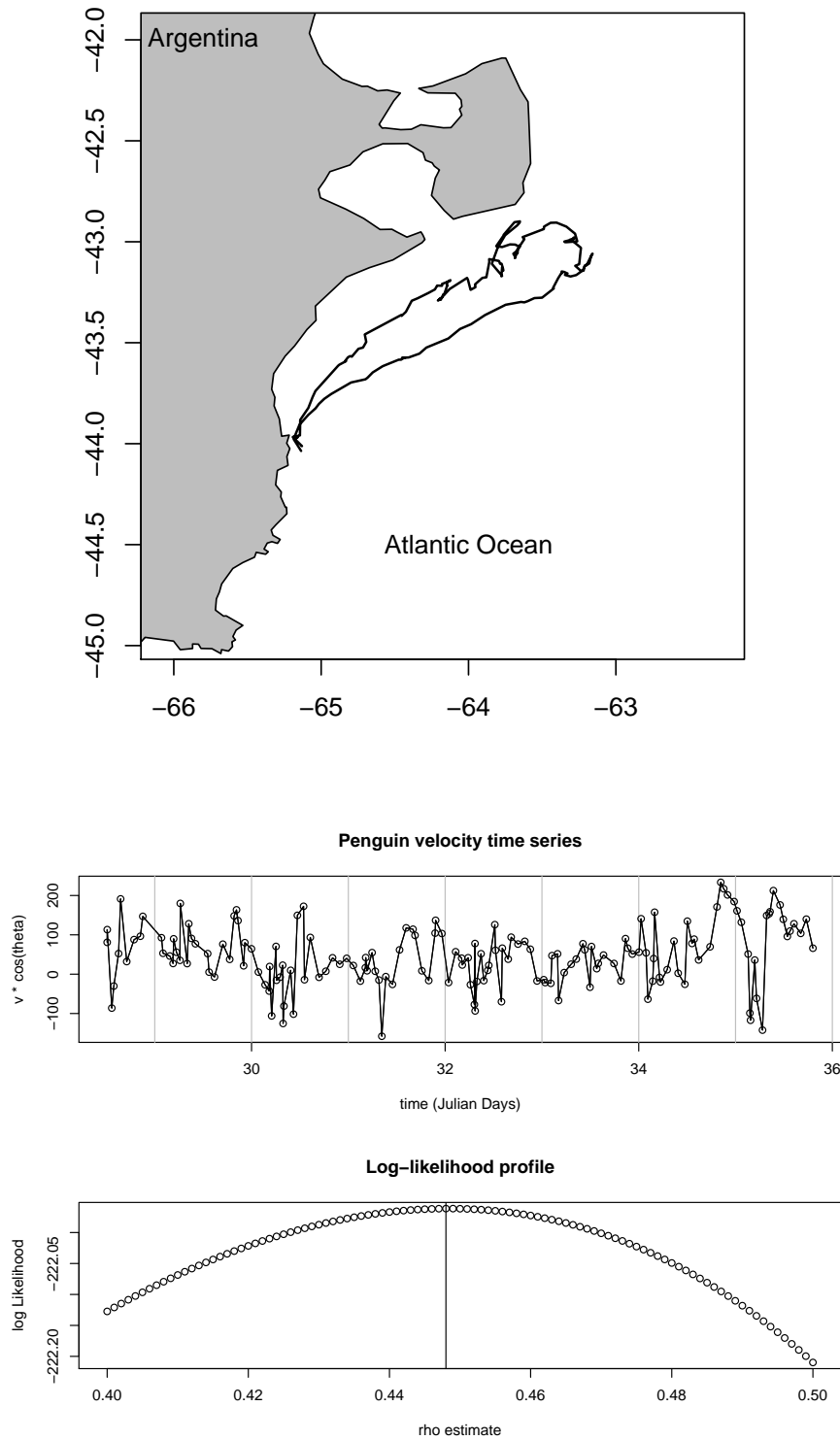
Figure 5.2:   Magellanic penguin foraging track (above), $v\cos(\theta)$ time series (middle) and log-likelihood profile for ρ estimation (below). The estimate $\widehat{\rho}$ is 0.448 1/hour.

5.3.2 Example with data

Magellanic penguins (*Spheniscus magellanicus*) in Argentina were tagged with satellite transmitters in early 2005 and their movement during foraging trips tracked using the ARGOS tracking system (figure 5.2). Here we analyze a single foraging trip of a penguin (P1) performed over eight days from January 28 to February 4. There were 163 positions recorded and the interval between the position fixes ranges from 1 to 278 minutes.

We extracted a 'velocity tendency' time-series, i.e. $V_p = V(t_i)\cos(\theta(t_i))$ where $V(t_i)$ is the measured speed (distance traveled divided by time interval) and $\theta(t_i)$ is the turning angle at $t(i)$ from the previous direction. We obtained $\widehat{\rho}$ using the MLE estimation routine.

The resulting estimates are $\widehat{\mu} = \bar{V}_P = 1.97$ km/h, $\widehat{\sigma} = 1.79$ km/day and $\widehat{\rho} = 0.448$ (figure 5.2). Thus, if the penguin's direction were sampled once an hour, the acf of the resulting time-series would be $\gamma(1) = 0.45$. A preliminary conclusion from this result is that the penguin exhibits multi-hour persistence in its movement.

5.4 Identifying structural shifts

We now address the question of how to expand the methods outlined in section 5.3 in order to identify structural shifts. A *structural shift* is a change in the underlying continuous process, for example a behavioral change from a goal-oriented traveling mode of movement to a search and forage kind of movement. The challenge is to infer where a structural shift occurs within a gappy and error ridden time series measurement of a process. We refer to the time at which the structural shift occurs as the *changepoint*.

Consider a continuous stochastic process $X(t)$ for $0 < t < T$ defined by parameter
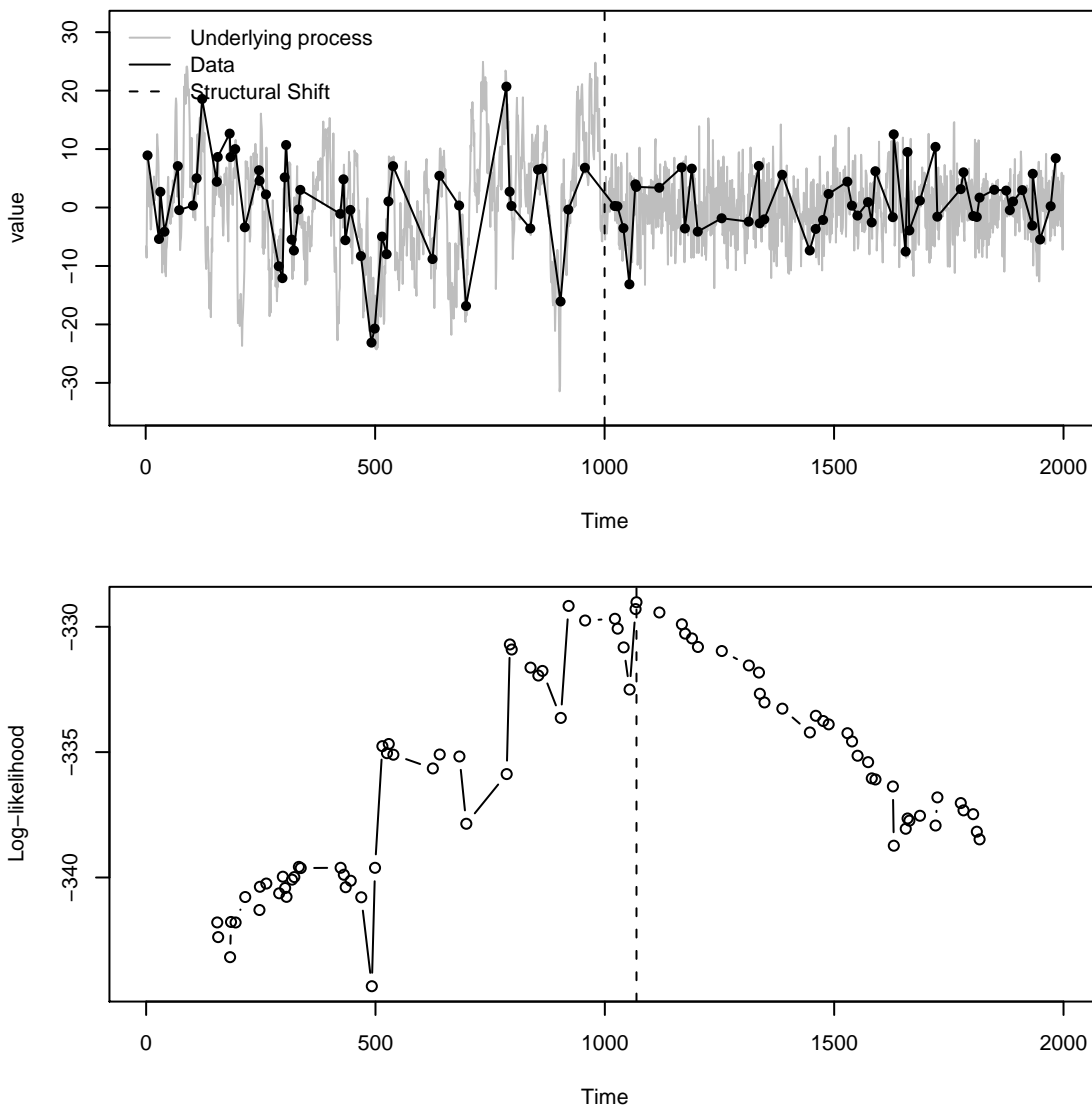
Figure 5.3: Simulation and estimation of gappy time-series with a single structural shift generated by sampling 100 values from a time-series of length 2000 with a breakpoint at $T_{br} = 1000$ represented by the vertical dotted line in the top graph. The true parameter values are $\mu_1 = 5$, $\mu_2 = 2$, $\sigma_1 = 10$ $\sigma_2 = 5$, $\rho_1 = 0.9$ and $\rho_2 = 0.2$. In this simulation, the resulting parameter estimates were: $\widehat{\mu}_1 = 4.6$, $\widehat{\mu}_2 = 0.24$, $\widehat{\sigma}_1 = 8.75$ $\widehat{\sigma}_2 = 4.73$, $\widehat{\rho}_1 = 0.91$, $\widehat{\rho}_1 = 0.35$, while the MLBP is at $t = 1060$.

set $\boldsymbol{\Theta}(t)$ whose values change at an unknown time point T_1 such that:

$$\boldsymbol{\Theta(t)} = \begin{cases} \boldsymbol{\Theta_1} & \text{if} \quad 0 < t \leq T_1 \\ \boldsymbol{\Theta_2} & \text{if} \quad T_1 < t \leq T \end{cases} \tag{5.11}$$

We now take N samples from the continuous process $X(t)$ to obtain a time series X_i at times T_i. If n is defined as the number of measurements within the first regime, such that $T_n \equiv \max(T_i < \tau)$, then the likelihood of the parametrization $(\theta_1, \theta_2, \tau_1)$ given data X_i is given simply by the product of the two likelihoods:

$$L(\boldsymbol{\Theta}|\mathbf{X}, \mathbf{T}) = \prod_{i=1}^{n} f(X_i|X_{i-1}, \boldsymbol{\Theta_1}) \prod_{j=n+1}^{N} f(X_j|X_{j-1}, \boldsymbol{\Theta_2}) \tag{5.12}$$

This likelihood can be maximized by sweeping over all possible values of n (from 1 to N), and obtaining the MLE's for the remaining parameters after the split using the method described in section 5.3.

Thus, the parameter estimates can be expressed as:

$$\widehat{n} = \underset{n}{\text{argmax}} \; L(\boldsymbol{\Theta}|\mathbf{X}, \mathbf{T}) \tag{5.13}$$

$$\hat{\mu}_j = \bar{X}_j \tag{5.14}$$

$$\hat{\sigma}_j = S_j \tag{5.15}$$

$$\hat{\rho}_j = \underset{\rho}{\text{argmax}} \; L(\rho|\mathbf{X_j}, \mathbf{T_j}, \hat{\mu}_j, \hat{\sigma}_j) \tag{5.16}$$

where $j \in (1, 2)$ indexes the two regimes, such that if $j = 1$ the data and estimates are taken in the range $i = (1, 2, ..., \widehat{n})$ and if $j = 2$, they are taken from the second regime $(i = \widehat{n} + 1, \widehat{n} + 2, ...N)$. We term the estimate for the changepoint $(t_{\widehat{n}})$ the *most likely changepoint*, MLCP.

The corresponding estimates for the orthogonal component of velocity V_t are simplified by the fact that the mean can safely be assumed to be 0. Indeed, analyses have

verified that when this assumption is relaxed, the mean is almost never significantly different from zero

5.4.1 Identifying models

Each of the three parameters that characterize a movement can change, and a change in each of the values corresponds to a different behavioral interpretation. Thus, for $V\cos(\theta)$, an increase in μ corresponds to a combination of faster and more directed movement. An increase in σ indicates more variable movement, e.g. more common stopping and moving or a greater increase in diving. A higher ρ indicates more directed and correlated movements, whether fast or slow. For $V\sin(\theta)$, higher values of σ indicate more turning or zig-zaggy movement, while higher values of ρ indicate longer turning radii. It is consequently of great interest to be able to identify which, if any, of the parameters actually change at the MLCP.

There are eight possible models to consider when analyzing a changepoint. These are defined as follows: M0 is the null hypothesis ($\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$ and $\rho_1 = \rho_2$); M1, M2 and M3 have one inequality each while the other two parameters remain constant ($\mu_1 \neq \mu_2$, $\sigma_1 \neq \sigma_2$, $\rho_1 \neq \rho_2$ respectively); M4, M5 and M6 have one equality each while two other parameters change (μ and σ, μ and ρ, σ and ρ respectively); and M7 is the most "alternate" hypothesis, in which all parameter values change at the MLBP.

Because the conditional likelihood is well-defined, we can apply a consistent criterion to select our models. We compared two criteria, AIC (Akaike's Information Criterion) and BIC (Bayesian Information Criterion) defined as:

$$I_A(\mathbf{X}, \mathbf{T}) = -2n \log\left(L(\hat{\theta}|\mathbf{X}, \mathbf{T})\right) + 2d \tag{5.17}$$

$$I_B(\mathbf{X}, \mathbf{T}) = -2n \log\left(L(\hat{\theta}|\mathbf{X}, \mathbf{T})\right) + d \log(n) \tag{5.18}$$
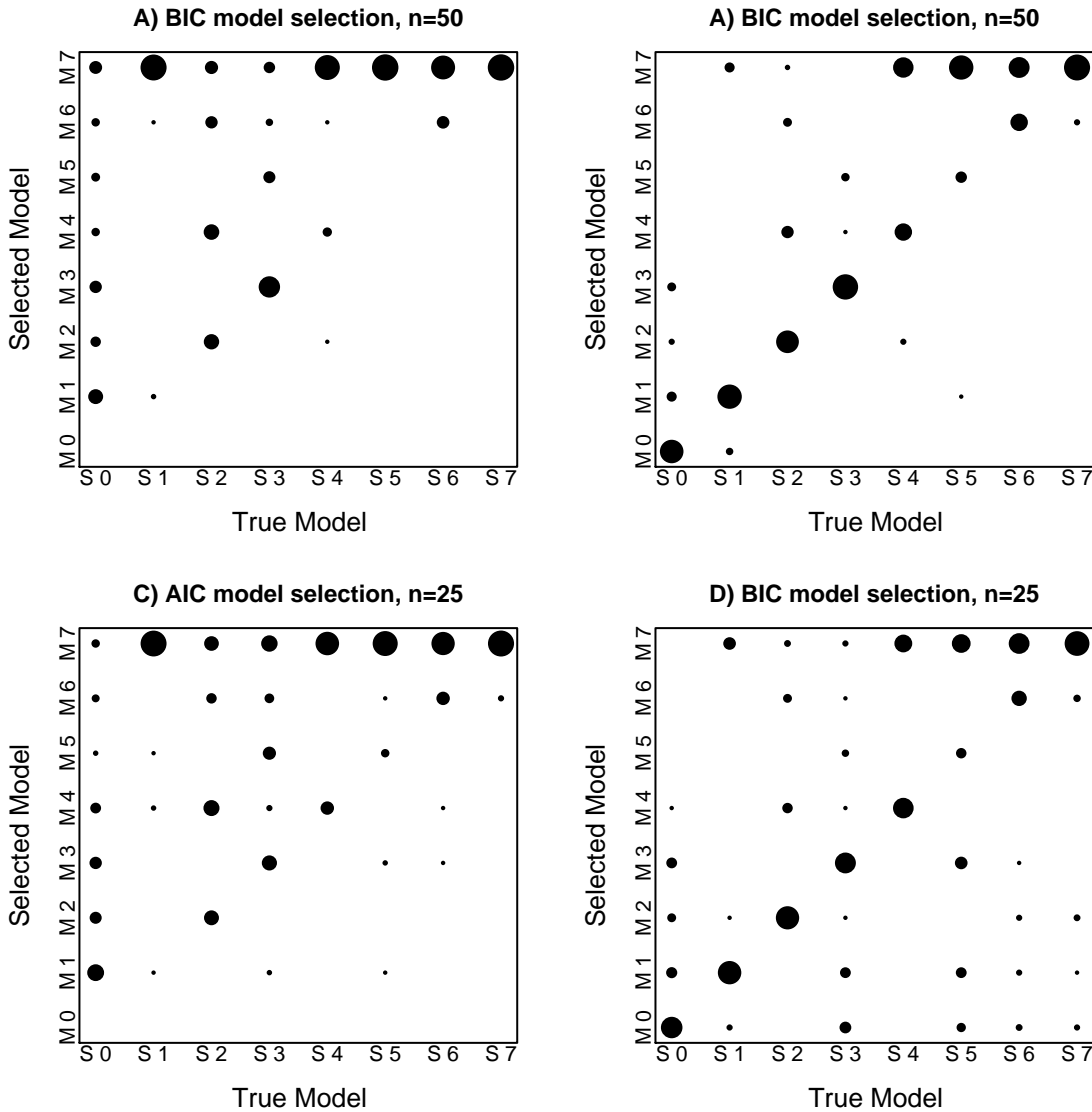
Figure 5.4: Results of model selection simulation using both AIC (A and C) and BIC (B and D) for breakpoints at $n = 25$ from a gappy time series of length $N = 50$ subsampled from a true process of length $T = 200$ (A and B) and $n = 50$, $N = 100$ and $T = 400$ (C and D). The true models are parameterized as in table 5.1, ranging from the null model (S0) of no difference before and after the gap to the extreme model (S7) with differences in all three parameters. 100 gappy time series were performed for each parameter set. The area of the circles is proportional to the number of time a model is selected for a given parameter set.

where $L(\hat{X}|\theta)$ is the likelihood as defined in equation 5.12, d is the number of parameters in each of the eight models (ranging from $d = 3$ for M0 to $d = 6$ for M7). The model with the lowest value of the information criterion is chosen. The BIC is more conservative than the AIC at selecting models, since it additionally "penalizes" the criterion according to the size of the dataset.

5.4.2 Simulation study

Figure 5.3 illustrates an example of the single changepoint estimation routine. The simulated data in this example comes from an underlying process of length $N = 2000$ in which a discrete structural shift occurs at $t = 1000$ in which the mean drops from 5 to zero, the variance decreases from 10 to 5, and the autocorrelation coefficient drops from 0.9 to 0.2. This simulation represents a shift from a highly variable, positively biased, correlated process associated with directed movement to a zero mean, lower variance, less correlated process that mimics random foraging and feeding. We randomly sampled 100 points from the complete time series and obtained the MLCP according to the method described above. The resulting estimated parameters are listed in the figure caption.

The plots of the log-likelihoods given in figure 5.3 give an idea of how precise the τ estimate might be. The log-likelihood profile seems adequately peaked around the correct value. It should be noted that the time series in this example is particularly gappy and the total dataset is relatively short.

We explored the properties of this experiment for a variety of parameter value changes by simulating 100 gappy processes ($N = 400$, $n = 50$, $T_{br} = 200$) for each of eight different models (see table 5.1A) ranging from the null model of no change in parameter values to the most alternate model of change in all parameter values. Resulting parameter estimates appear unbiased and relatively precise (table 5.1B), all falling well within one standard deviation of the true value. Perhaps most importantly, the MLCP estimates are highly accurate, with means between 197 and 203. We also

applied AIC and BIC to assess the eight models and report their results (table 5.1B and figure 5.4). The BIC performs far better than the AIC at identifying the true model, with, notably, a 78% rate of correctly identifying the null hypothesis and 72-94% probabilities of identifying parameters with a single parameter changing. For those models where two parameters shift, BIC tends to falsely favor the most complex model. In contrast, AIC performs miserably, never selecting the null model correctly and in other cases almost always choosing more complex models than necessary.

When analyzing actual data, a researcher may have greater interest in identifying the location and direction of significant, detectable structural shifts than in estimating the parameters themselves. For this reason, we prefer the more conservative criterion. A complete power analysis is very difficult to perform for the selection mechanism, since the ability of the model to correctly identify the model depends in complicated, non-independent ways on the magnitude of the difference between the parameters, the extent of the gappiness in the dataset and the length of the series. However, an effort was made in the simulation study to look at differences that are comparable if not smaller than those in the actual data. The ability of the BIC to pick the appropriate level of complexity inspires confidence in this method of model selection.

5.4.3 Multiple changepoints

Heretofore, we have only discussed the identification of a single most likely breakpoint. Within the a single track, an organism likely exhibits multiple behaviors. Rigorously, a continuous process $X(t)$ can be defined such that for any interval between $0 < t < \tau_m$ defined by parameter set $\boldsymbol{\Theta}(t)$ whose values change at unknown time points

$\mathbf{T} = \{\tau_1, ..., \tau_m\}$ such that:

$$
\Theta(\mathbf{t}) = \begin{cases}
\theta_1 & \text{if} & 0 < t \le \tau_1 \\
\theta_2 & \text{if} & \tau_1 < t \le \tau_2 \\
\vdots & \vdots & \vdots \\
\theta_m & \text{if} & \tau_{m-1} < t \le \tau_m
\end{cases}
\tag{5.19}
$$

Estimating the parametrization Θ for the multiple change-point model is non-trivial problem that has generated some literature, though only for regular (non gappy) time-series and applied primarily to financial markets (see, for example Chib, 1998). Furthermore, there are complications with model selection and inference in selecting the number of breakpoints in a longer dataset.

My approach to this problem is to pass a window over the complete time-series and apply the single changepoint analysis described above at each window. The algorithm can be summarized as follows:

1. Select a window of length $l << N$.

2. Find the most likely changepoint in a subsample of the data $X_1...X_l$.

3. Use some criterion to accept or reject the "significance" of the changepoint for each of the parameters μ, σ and ρ.

4. Based on the result of the test, log the location of the behavioral changepoint and the resulting estimated parameter values: $\hat{\mu}_1$, $\hat{\sigma}_1$, $\hat{\rho}_1$, $\hat{\mu}_2$, $\hat{\sigma}_2$, and $\hat{\rho}_2$.

5. Shift the window up one datapoint, and repeat steps 1-4.

Throughout the running of this algorithm, we output the estimates according to the model chosen by the BIC. Thus, if M0 is chosen, we estimate a single value for each of the parameters for the entire dataset. If M7 is chosen, we separately estimate

parameters at each side of the breakpoint. All of the estimates are logged and averaged after the sweep is performed.

This method has the advantage of being able to capture arbitrarily many behaviors. Furthermore, by returning and averaging the model estimated parameters at every timestep multiple times, we allow for gradual shifts in behavioral parameters between the discrete structural shifts, and potentially dramatic jumps where the behavior does change suddenly and significantly The aggregated information including both the parameter estimates, the significant breakpoints and their nature and direction (i.e. increasing μ, decreasing ρ, etc.) can be considered a distilled behavioral model of movement.

It should be noted that the one tunable variable in this method is the size of the window. The greater the size of the window, the more consistent the estimates provided will be and the higher the power of the model selection (figure 5.4), but the probability increases that shifts at smaller scales will be lost. A smaller window size will reveal finer-scale structure in the data, but the risk of spurious responses increases. Furthermore, issues can arise when estimating means and variances with small sample sizes, especially when the correlation is high. Because of the behavioral complexity of the dataset in the application that follows, I chose a window of size 30 which is probably near the lower limit of an acceptable window size (see Results).

All analysis was performed using the freely available `R` programming package.[1]

5.5 Data

I applied the movement analysis method to GPS data collected on several foraging trips taken by a nursing female northern fur seal (*Callorhinus ursinus*). Northern fur seals aggregate annually during the summer months at large rookeries. The females give birth usually in late June through July, and, after 8-10 consecutive days of

---

[1] http://www.R-project.org
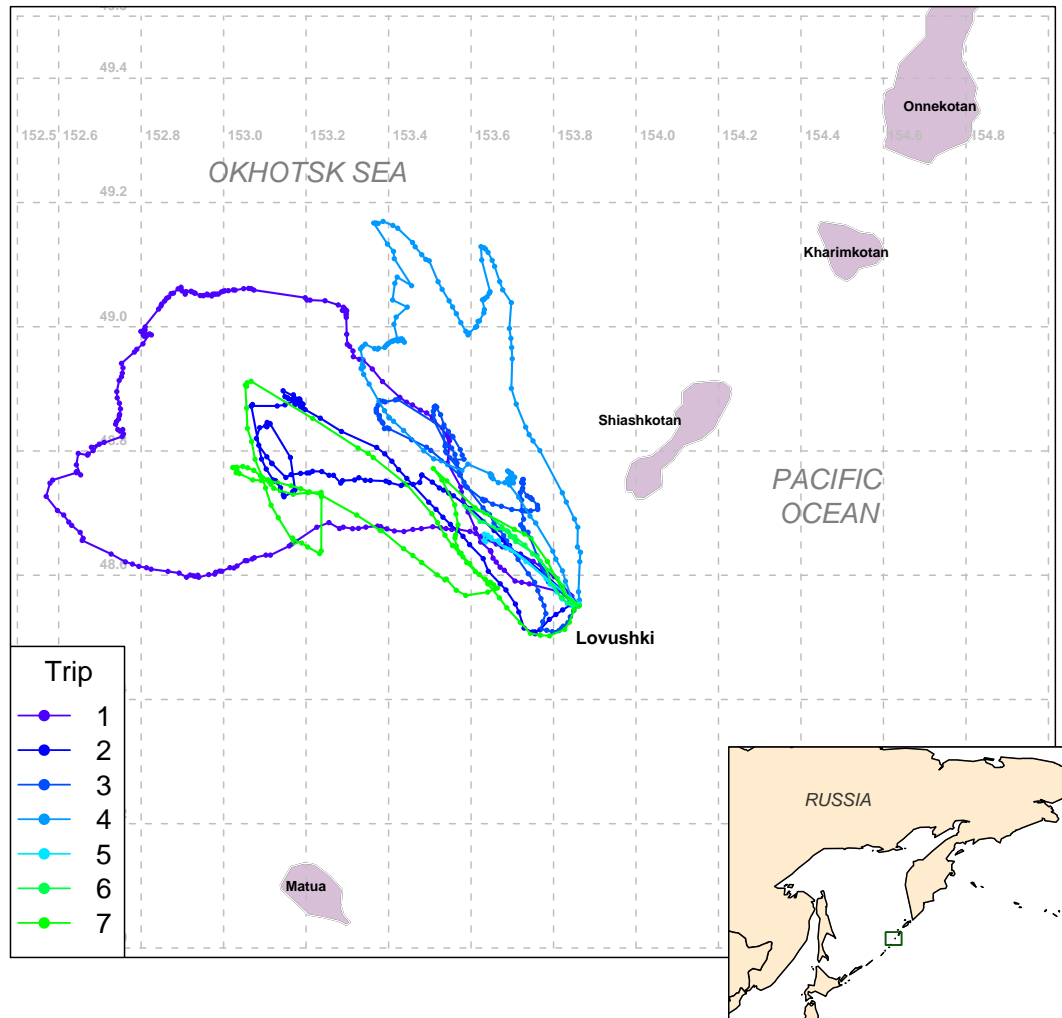
Figure 5.5: Map of all trips taken by NFS07-03 in Summer 2007. The square in the inset map indicates the study area.

nursing, begin to take foraging trips of up to a week in length to replenish energetic reserves (Gentry, 1998). The data analyzed here was collected from a female tagged in summer of 2007 at Dolgaya Rock, one of the Lovushki Islands - a small group in the central Kuril Island chain in Russia which runs between the Pacific Ocean to the southeast and the Sea of Okhotsk to the northwest (see figure 5.5). The Lovushki Islands are the location of one of the largest northern fur seal rookeries in the northwestern Pacific (Loughlin et al., 1984, Burkanov *pers. comm.*).

The female discussed in this paper, labeled NFS07-03, was captured on June 21, when her pup was one to three days old, and instrumented with a Fastloc®GPS data-logging device (MK10-F, Wildlife Computers Inc.) The tag allows for quick fixes and relatively high accuracy, with at least 60% of the locations within 100 m (Bryant, 2007). The MK10-F device was also equipped with dive depth, temperature and light sensors, and the animal was instrumented with a separate stomach temperature sensor. Here, we only consider the movement data; however, the existence of corroborating evidence for foraging behaviors, such as diving and prey capture, makes this an ideal system to explore what can be inferred from pure movement data.

The animal was monitored for 38 days in total, taking seven foraging trips in that time. The first trip occurred 9 days after tagging, on July 1, and lasted just over five days, while the last trip began on July 27 and lasted about two and a half days. On all trips, the fur seal headed in a northwesterly direction toward the Sea of Okhotsk (figure 5.5). A total of 763 position fixes were obtained over all seven trips, with individual trips ranging from 30 and 33 datapoints (trips 5 and 6) to 205 datapoints in length for trip 1. The time intervals between the fixes range widely from a less than a minute to 700 minutes, with the majority (over 80%) clustered around 15, 30 and 45 minutes. The data was filtered to exclude implausible swim speeds (> 11 km/h) and were georeferenced using standard methods. Velocities were estimated by dividing displacements by time intervals, and turning angles were calculated directly from the positions.
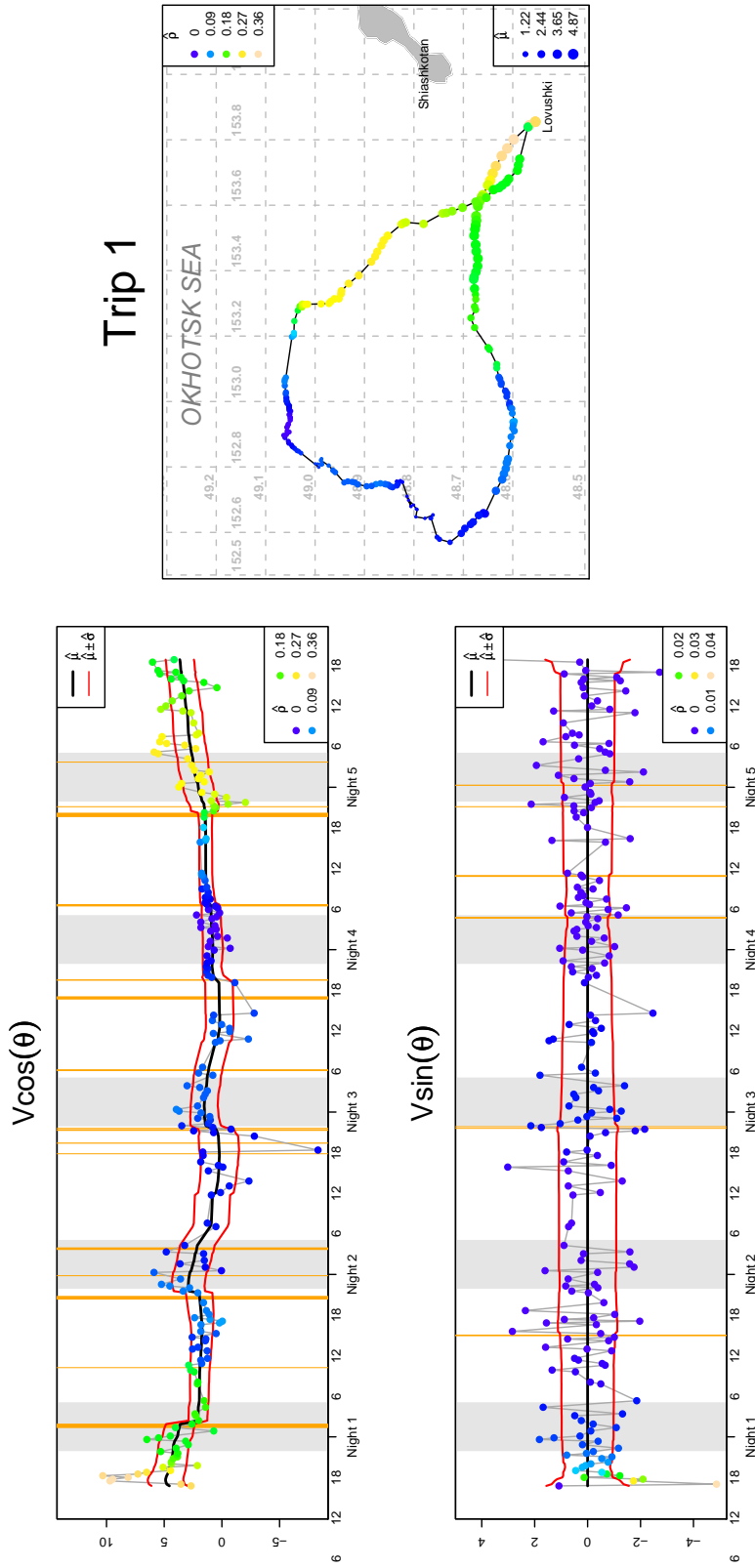
Figure 5.6: Decomposition analysis of NFS07–03 foraging trip #1. The plot on the left shows the time-series of velocity and turning angles decomposed into persistence ($V_p = V\cos(\theta)$) and turning ($V_t = V\sin(\theta)$) components. The black line is the estimate for the mean $\hat{\mu}$, the red line represents the estimate for the standard deviation $\hat{\sigma}$ around the mean, and the color reflects the auto-correlation $\hat{\rho}$: bluer colors indicate less auto-correlated movement, while yellow colors indicate more auto-correlated movement. Vertical orange lines indicate changepoints, where the estimation routine suggests there was a significant behavioral shift; thicker lines correspond to a higher number of selected shifts. Wide gray bands indicate nighttime. To the right is a mapping of the track itself. Colors again indicate the auto-correlation estimate for the persistence component (V_p), while the size of the dots is proportional to the estimated local mean value V_p. For this analysis, a window of size 30 was swept over the data.
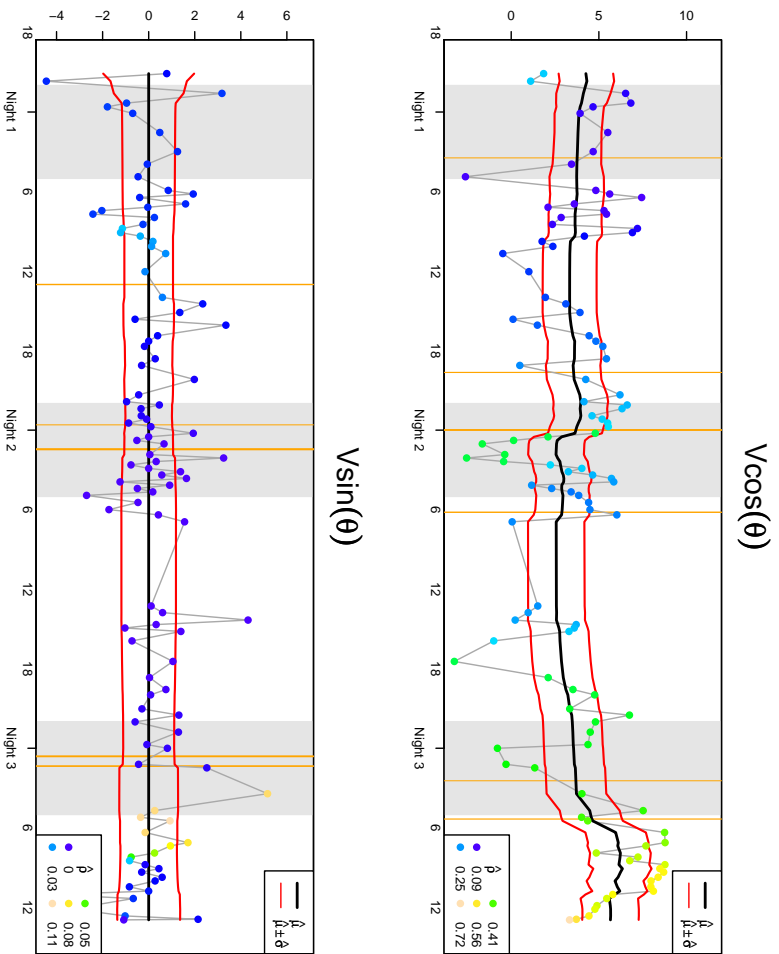
Figure 5.7: Decomposition plot of trip #7. See figure 5.6 for explanation.
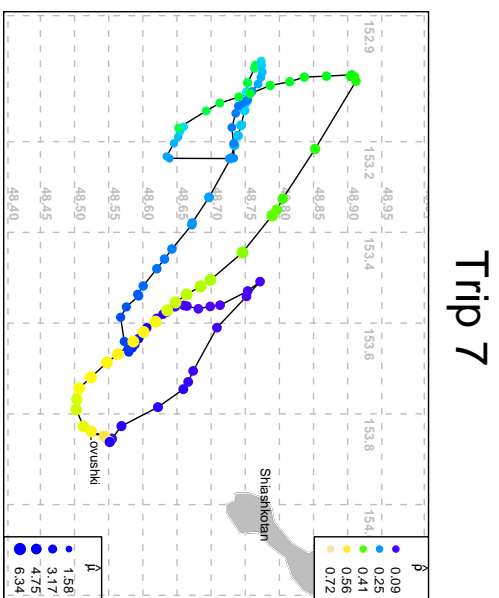
5.6   Results

I applied the analysis technique to five of the seven trips (those with more than 70 location fixes) with a window of size 30. At an average time interval between measurements of 37 minutes, the window covered on average a period of 18 hours, which seemed to allow for reasonable chance to pick up single behavioral shifts at an acceptable cost to power. The output of analyses with window size 50, corresponding to about a 24 hour period, seemed too insensitive since it is reasonable to expect that behaviors can change significantly more than once in a day.

Model outputs for two of the trips (1 and 7) are presented graphically in figures 5.6 and 5.7. A trace of in parameter space of the V_p estimates as they evolve in time is presented in figure 5.8.

Because the model output is somewhat complex, it is instructive to walk through a single track from beginning to end. The first trip is the longest (5.1 days long) and furthest (max. distance from rookery 96 km), as is typical for female fur seals taking their first feeding trip after a fasting period associated with birth. The fur seals's initial departure from the rookery is marked by high values of of V_p ($\widehat{\mu}$ around 5 km per hour) and a high estimated per-hour auto-correlation $\rho = 0.36$ (NB since the time data is in hours, the ρ is an estimate of what the first order auto-correlation coefficient at lag 1 would be if the movement data were collected once an hour exactly without gaps). The first significant changepoint, occurring at 01:53 in the middle of Night 1 was selected as the MLCP by every single one of the 25 windows that passed over it. The models chosen by the BIC were split about evenly between M4 (μ and σ change) and M7 (all three parameters change). The estimates for μ drop from near 4 km/h to around 2 km/h, estimates for σ drop from 1.5 to 0.7, estimates for ρ decrease more gradually from 0.3 to under 0.1. A similarly dramatic changepoint occurs at around 1900 before the fifth and last night of the trip, where fully 30 MLCP's were selected for at either 20:14 or 21:12. The BIC selected models were mixed between

M2 (only σ changes), M4 and M7. The track indicates that this final changepoint is associated with a sudden turn south, and a fairly correlated, moderately fast (around 3 km/hour) journey home.

The interim period between these two traveling bouts is marked most significantly by lower autocorrelation, essentially staying below 0.14 during the entire period and reaching near zero values during the fourth night. The mean persistence velocity and deviances, however, vary considerably in this period, and, while significant change-points are identified at evening/dusk and dawn on all three interim nights, the behavioral pattern at these times varies somewhat. Before Night 2 at 21:15 there is a significant increase in both μ and σ (11 of 15 MLCP's chose M4). At 00:20 there is a notable drop in ρ only (4 of 4 MLCP's chose M3), and at 4:18 there is a significant drop in μ only (11 of 11 MLCP's chose M1). The μ during the evening is fairly high, at 2.5-3.0 km/h. During the day the mean velocity plummets to 0-0.5 km/hour, while the deviance gradually increases until the evening before Night 3. Night 3 again shows an increase in μ and a decrease in σ, but to a lesser extent that Night 2, and again a drop in velocity during the day. Finally, Night 4 is marked by the lowest autocorrelations ($< .01$), low velocities (< 0.5 km/hour) but a moderate deviance (around $0.5 - 1.0$ km/hour). The final day shows the greatest tightening of the deviance with no appreciable increase in velocity, until the aforementioned burst home at 21:00.

The analysis of turning velocity V_t for the first track proves less informative than the analysis of V_p. Of the 175 MLCP's that were identified (one for each window), 121 returned M0, the null model of no significant changes in σ or ρ, (in contrast to only 23 of 175 M0's in V_p), 51 selected M2 (change in σ), 3 selected M3 (change in ρ), and none selected M6 (change in both). The value for ρ were all quite low, with a mean value of 0.002 and about half of the estimates equaling zero. It is notable, however, that at the very beginning of the trajectory, the ρ estimates are quite a bit higher (up to 0.04) than in the remainder of the trip, when the animal was moving at its highest velocity. The higher correlation here, and generally negative values for V_t near the
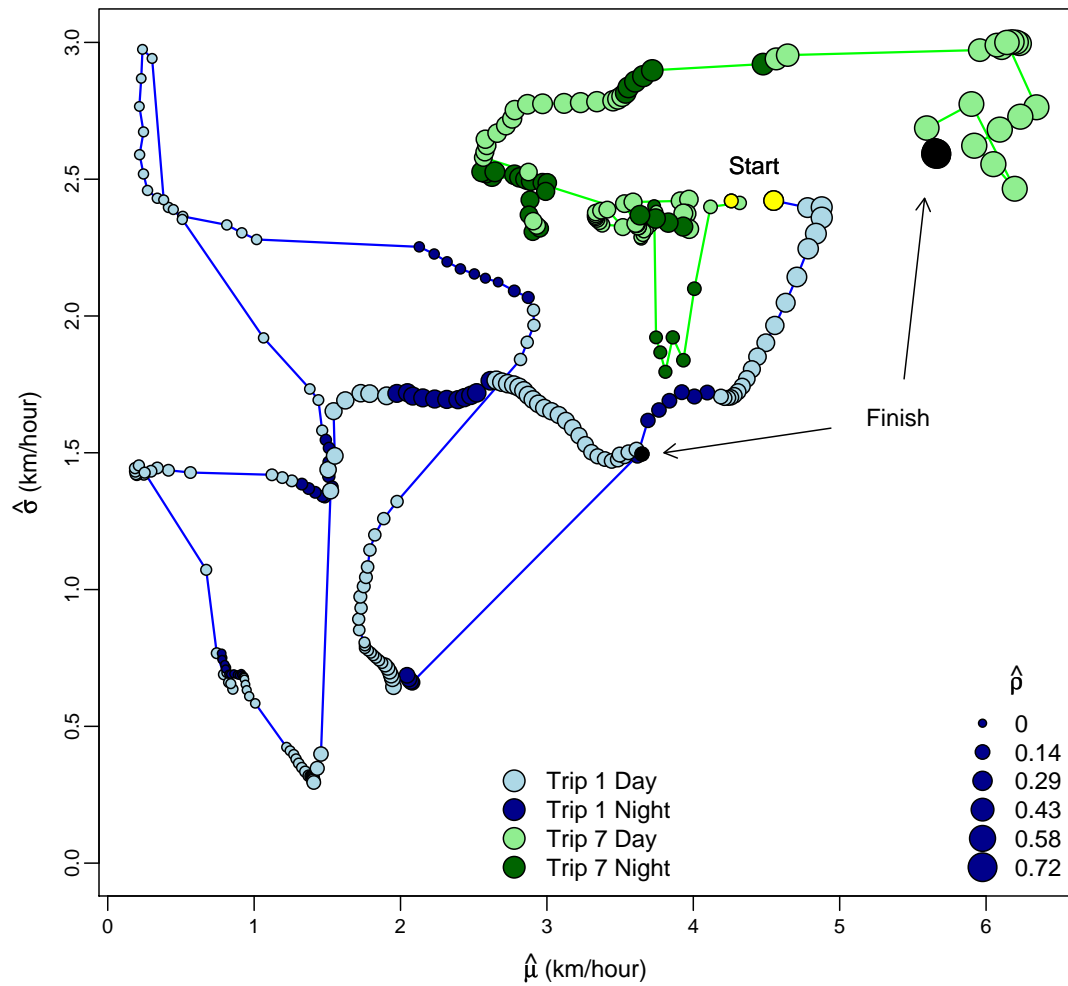
Figure 5.8: A mapping of all three parameters for V_p for trips number 1 (blue circles) and 7 (green circles). The size of the circles is proportional to $\hat{\rho}$. This connecting lines illustrate the progress of the walk through the parameter space, beginning at the yellow circles in the upper-right quadrant and finishing at the black dots. The long connections are associated with significant structural shifts.

beginning indicates that the animal is conducting a persistent, wide left-turning arc. The deviance remained fairly consistent throughout this trajectory, around 1 km/h, tightening most significantly to about 0.4 km/h during the intensive, low correlation, low V_p deviance fourth night of foraging.

Trip 7 displayed some features that were unique compared to the other analyzed trips. Notably, the velocities the fur seal moved at during Trip 7 were much, higher with an aggregate V_p mean of 3.8 km/h (s.e. 1.1) compared to 2.0 (s.e. 1.25) for Trip 1. Somewhat atypically, it left in the evening, moving quickly (around 5 km/h) but with low autocorrelation in its persistence velocity. It made a hairpin-like change in direction at dawn after its first night, appearing to turn back to the rookery, turns again in the afternoon to head NW. During Night 2, the fur seal exhibits considerable slowing, low autocorrelations and loops on its own track several times, appears to head out further to sea second time, spends a night moving at a relatively high mean velocity and auto-correlations, and at dawn after Night 3 performs one last hairpin turn and heads back at a very high pace, reaching the highest estimated persistence velocity of all trips at 6.4 km/h, likewise attaining the highest autocorrelations of any trips (0.73). It appears to be missing the rookery on its way home by several kilometers, but adjusts near the end. This return trip is likewise marked by some of the highest auto-correlations in the V_t time-series (up to 0.11), once again indicating a large and consistent turning radius of movement

A comparison of all the aggregated parameter outputs of the model for all of the trips are presented in figure 5.9. Violinplots are used in order to highlight the multimodal nature of many of the parameters.

5.7 *Discussion*

The method presented here displays the ability to robustly identify complex behavioral patterns in irregularly measured movement data. The maximum likelihood method is sensitive enough to identify changepoints even in very gappy or noisy
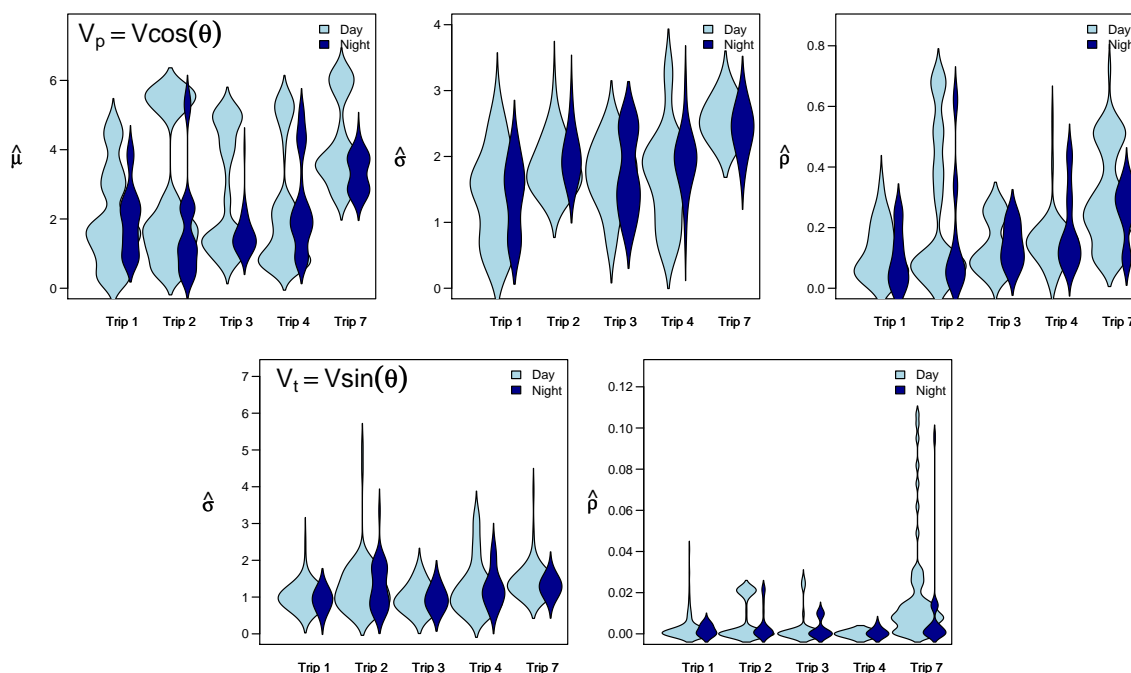
Figure 5.9: Violinplots of all parameter values for each of the five analyzed trips separated by daytime and nighttime estimates. The violin plot combined the information of a box-plot with a histogram by showing an estimated Gaussian kernel density of the data. It is particularly useful when the data is multimodal, as is the case for many of these estimates. The area of the "violins" is proportional to the number of datapoints in each sample.

data, and the BIC model selection method is conservative enough to protect against excessive complexity. The data considered in the application here is highly accurate positionally, but marked with a "typical" level of gappiness for a marine species. In general, however, the method is applicable to many kinds of data, including ARGOS, archival tags or radio telemetry data. Furthermore, the algorithms are computationally quite tractable. While the output is somewhat sensitive to the analysis window size, this variable is easily tuned according to the temporal resolution of the dataset and the consistency of the model output. Because the method is focused on identifying the location and direction of structural shifts, it is also robust to error, though this aspect is not explored in this paper.

An important innovation in the method presented here is the explicit and tractable analysis of the autocorrelation. This relies an analysis of persistence and turning transformations, which, while not as biologically intuitive as raw speed and turning angle data, have attractive statistical properties that allow for an explicit description of the autocorrelation structure. This autocorrelation is readily interpreted in terms of time-scales at which an organism's basic movement patterns (persistence and turning) change. For example, the most significant distinction between the traveling modes and non-traveling behavioral modes in the fur seal tracks was associated with a change in the auto-correlation. Often, this is an even more significant variable than mean velocities. The fur seal's return trips do not necessarily attain appreciably higher velocities than those during the "foraging" phase (see for example Trip 1, Night 2). Thus, the fur seal can have a high rate of persistence displacement in a given direction but have a much more erratic movement pattern, punctuated by feeding bouts, dives, turns and loops, whereas the steadiness of a directed movement is captured by a higher persistence velocity autocorrelation. A useful index to consider is the time-interval at which the first order correlation drops to some fixed number such as 0.5. Using the relationship $\tau_h = \log(0.5)/\log(\rho)$, we obtain an estimated "time to half-autocorrelation" $(\hat{\tau}_h)$. For trip 1, this value is around ranges between 30 and 40 minutes for the traveling modes, and between 4 and 15 minutes during the remainder of the trip. Similar patterns, broadly, hold for the remaining trips.

It is worth making a distinction between the information contained in the autocorrelations of persistence and turning and the parameters of standard CRWs. Although this commonly applied mechanistic model of movement is termed the "correlated random walk", in fact almost every application of the model assumes independent turning angles and independent velocities and the smoothness of a CRW path is a result of the clustering of the turning angles around zero degrees. In fact, there is a fundamental distinction between independent turning angles and correlated turning angles, which are manifested in larger turning radii. This is well-captured by the

auto-correlation coefficient on the turning component of velocity, as, for example, near the beginning or Trip 1 and the end of Trip 7, where the significant rise of the turning velocity auto-correlation is associated with a longer scaled arcs of movement, as contrasted with the smaller, independent directions shifts that dominate the bulk of the movements. Similarly, any autocorrelation in the velocities is largely accounted for in the persistence velocity correlation.

Returning to Trip 1, the analysis of the movement data over the three nights of foraging indicate a higher behavioral complexity that might have been expected. The primary prey items for fur seals are cephalopods and smaller pelagic fish that engage in vertical diel migrations, particularly during the crepuscular hours. It is known that the majority of feeding occurs at night (Goebel et al., 1991.) and the dive data associated with NFS07-03 corroborates this expectation, with most dives beginning at around 23:00 and continuing through the early morning (Andrews *pers. comm.*). During the day, northern fur seals are known to sleep at sea (Gentry, 1998). Thus, the movement during daylight hours, which tended to be the least autocorrelated in Trip 1 for the first three daylight periods is more likely to reflect a mixture of drifting with ambient currents and periods of directed swimming. The distinct changes identified before each of the evening bouts correspond to increased activity and changes in absolute orientation, implying an active pursuit of prey. Interestingly, the mean velocities were quite high, particularly during the first evening, implying perhaps that the prey patch itself was moving, or that the animal is actively seeking prey over a larger area. Again, the autocorrelations are relatively low throughout, indicating irregular movements and many changes. The behavior during the third night of foraging reflects the closest to the expectation of what active, successful foraging might look like, with the lower variance and the lowest autocorrelations. However, even here there is a distict, oriented, eastward movement.

The other important feature of the method presented here is the ability of the analysis not only to estimate values of parameters, but to identify significant shifts,

such that the model output provides both gradual changes in the parameter values and more discrete differences that are presumably associated with behavioral choices. The analysis of these behavioral shifts provides a window into the complexity of an animal's movements.

The violinplots of several of the aggregated parameter values do indicate that types of movement are often clustered into several groups. Notably, the estimates for the mean persistence velocity for Trips 2, 3 and 4 indicate a clustering of low velocities (around 2 km/h) and a separate clustering of high velocities (around 5 km/h). Generally, the higher velocities appear less often at night. This pattern is mirrored to a smaller extent in both the persistence and turning autocorrelation coefficients. However, the behavioral trace in figure 5.8 suggests a striking complexity in behavioral modes, especially as contrasted with the expectation of a few distinct behaviors. While a marine organism is free of many of the landscape-related obstacles that constraint terrestrial organisms, its movement, even at rest, can be confounded by ambient currents. However, the sudden and dramatic shifts in the parameter space trace strongly suggest that these signals are reflections of real behavioral choices.

Both a consideration of autocorrelations and the basic structure of behavioral shift models can be productively implemented on datasets which lack the unavoidable gappiness of marine organism tracking data.

5.8   *Conclusions*

A top predator like the fur seal is extremely well-adapted to exploiting the heterogeneous environment of the open ocean to fulfill its survival needs. Its movement is a complicated manifestation of an organism's internal state, access of information, physiological constraints and behavioral responses to environmental cues. The data collected on this movement is an irregular subsampling of this structurally complex, continuously auto-correlated, multi-dimensional process. The analysis method presented here is a purely descriptive attempt to capture and distill the dynamics of

the underlying process from noisy and gappy data. The output is a somewhat complex array of estimates, model selections and aggregated averages, but many of the nuanced patterns of movement behaviors emerge in a clear and tractable way.

The ultimate utility of the method, however, will be in its use for answering important biological questions. The greatest mystery is: How does an organism successfully exploit its environment? In order to approach this challenge, it is necessary to model whatever information about the behavior the analysis method produces against potentially informative covariates. In the study on fur seal movements in the Kuril Islands, for example, detailed data has been collected in parallel on diving events, ambient temperatures and foraging success via stomach temperature sensors. An analysis of these data against the movement model can yield higher level insights into the relationships between the estimated movement parameters and foraging strategies and success. On a larger scale, comparisons can be made between animals in different locations throughout the range, under varying environmental conditions, as well as between different groups (e.g. older and younger, male or female).

Both a consideration of autocorrelations and the basic structure of a behavioral shift models can be productively implemented on datasets which lack the unavoidable gappiness of marine organism tracking data. Indeed, the output of the analysis method as applied to terrestrial organisms can be modeled as a response landscape features which are more readily measurable on land than at sea.

Table 5.1: Results of a simulation study used to explore the efficiency of the estimation routine for a single structural break in a gappy time series. Underlying time series of length 400 were generated with a breakpoint at $T_{br} = 200$ for several models, ranging from a null model of no change in μ, σ and ρ, to a model where all three parameters change value. The parameters used in the simulation are tabulated in 5.1A, with those parameters which change at the breakpoint are boldfaced. From each timeseries, 50 datapoints were randomly selected to create "gappiness" and estimates were obtained using the estimation routine described in the text. Simulations were performed 100 times for each parameter set, and the resulting mean estimates and standard errors are reported in 5.1B. Essentailly all estimates are within one standard error of the true value. The last two columns (N_A and N_B) are the number of times that AIC and BIC respectively select the correct model out of 100 attempts (see figure 5.4).

A) True parameter values

	μ_1	μ_2	σ_1	σ_2	ρ_1	ρ_2
S0	0	0	1	1	0.5	0.5
S1	**-1**	**1**	1	1	0.5	0.5
S2	0	0	**0.5**	**2**	0.5	0.5
S3	0	0	1	1	**0.2**	**0.9**
S4	**-1**	**1**	**0.5**	**2**	0.5	0.5
S5	**-1**	**1**	1	1	**0.2**	**0.9**
S6	0	0	**0.5**	**2**	**0.2**	**0.9**
S7	**-1**	**1**	**0.5**	**2**	**0.2**	**0.9**

B) Estimates

	T_{br}	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	N_A	N_B
S0	206 (113)	0.0 (0.21)	0.0 (0.17)	1.0 (0.11)	1.0 (0.11)	0.48 (0.16)	0.47 (0.15)	0	78
S1	197 (13.2)	-1.0 (0.18)	1.0 (0.19)	1.2 (0.21)	1.2 (0.22)	0.59 (0.25)	0.63 (0.19)	1	84
S2	203 (10.3)	-0.03 (0.12)	-0.04 (0.27)	0.50 (0.04)	2.0 (0.19)	0.46 (0.16)	0.47 (0.16)	28	72
S3	201 (26.1)	-0.02 (0.22)	0.04 (0.43)	1.0 (0.08)	1.0 (0.12)	0.23 (0.21)	0.92 (0.056)	29	92
S4	199 (5.49)	-1.0 (0.08)	1.1 (0.33)	0.5 (0.04)	2.0 (0.17)	0.47 (0.2)	0.53 (0.17)	22	40
S5	199 (12.0)	-0.96 (0.17)	1.0 (0.51)	1.0 (0.12)	1.0 (0.14)	0.19 (0.18)	0.92 (0.033)	7	15
S6	200 (8.51)	0.0 (0.09)	0.02 (0.93)	0.49 (0.05)	2.0 (0.17)	0.20 (0.19)	0.92 (0.036)	22	40
S7	201 (15.6)	-0.99 (0.08)	0.94 (0.87)	0.50 (0.05)	2.0 (0.18)	0.19 (0.18)	0.93 (0.033)	97	97

Chapter 6

POPULATION-LEVEL HETEROGENEITY IN MODELS OF DISPERSAL AND MOVEMENT

One fish, two fish, red fish, blue fish...

Some are fast. And some are slow...

Not one of them is like another...

But I bet they have come a long, long way.

One Fish, Two Fish, Red Fish, Blue Fish

DR. SEUSS (1960)

Summary

[1]

Behavioral heterogeneity among individuals is a universal feature of natural populations. Most diffusion-based models of animal dispersal, however, implicitly assume homogeneous movement parameters within a population. Recent attempts to consider the effect of heterogeneous populations on dispersal distributions have been somewhat limited by the high number of parameters required to subdivide a population into several groups. A solution to this problem is to characterize the value of a movement parameter as continuously distributed within a population. We present several cases in which this method is useful and tractable, applying the framework both to spatial distribution data and closely related first passage times. The resulting models allow ecologists to identify the extent to which the variability in dispersal distributions can

be attributed to population level heterogeneity as opposed to intrinsic randomness. We apply the formulation to two very different cases of dispersal: resident organisms in a stream (freshwater chub *Nocomis leptocephalus*) and migrating organisms (juvenile salmonids *Oncorhynnchus spp.*). In both cases, accounting for heterogeneity provides novel insights into the behavioral mechanisms of movements.

6.1   Introduction

Animal movement models have been dominated by the concept of "biodiffusion", where the spread of organisms is a consequence of individual random movements that are commonly approximated as Brownian (Turchin, 1998; Okubo and Levin, 2001; Skellam, 1951; Zabel, 1994). Diffusion models owe their development in large part to the field of statistical mechanics in physics, which deals with large numbers of indistinguishable particles in a homogenous environment, such as gas molecules. Consequently, a common implicit assumption behind diffusion models is that the individuals that compose a population are identical. However, an important, indeed, almost defining, difference between molecules and animals is that there exits genotypic and phenotypic variability between organisms, and these differences can have an effect on dispersal rates. Furthermore, animals disperse in natural environments that are, as a rule, variable and heterogeneous both in space and time. Consequently, actual dispersal data rarely conforms with the essential Gaussian distribution predicted by idealized Brownian motion.

Commonly, in empirical dispersal studies, the data tend to show more distributions with positive kurtosis, i.e. sharp peak and fatter tails than expected from a Gaussian distribution (Price et al., 1994; Kot et al., 1996; Skalski and Gilliam, 2000; Coombs and Rodríguez, 2007). These deviations from normality can be accounted for by Lévy motion, where a power-scaled distribution of step lengths or waiting times between steps can yield leptokurtic distributions (Viswanathan et al., 1996; Ramos-Fernández et al., 2004; Uttieri et al., 2005; Fritz et al., 2003; Zhang et al., 2007) or with non-

Gaussian movement kernels which lead to faster rates of invasion (Kot et al., 1996; Coombs and Rodríguez, 2007). Nonetheless, these models still generally rely on the assumption that individuals within a population follow the same set of rules.

Some investigators have proposed heterogeneity as an explanation of leptokurtic distributions, suggesting that they arise from populations where two or more Gaussian dispersals are superposed (Fraser et al., 2001; Skalski and Gilliam, 2000, 2003; Coombs and Rodríguez, 2007). However, in these studies heterogeneity is limited to the number of discrete groups within which the behavior is homogeneous. In this framework, greater realism requires a consequently greater number of parameters, since each group requires it's own set of movement parameters and a proportion of the total population.

Here, we introduce a general framework for incorporating a continuous population-level heterogeneity in dispersal models and present several analytically tractable solutions. We demonstrate that population-level heterogeneity in the Brownian motion variance parameter yields leptokurtic distributions for population spread without requiring recourse to Lévy type behavior and loosely apply this result to observations made of freshwater fish dispersal observed in streams.

Since it is generally a resource-intensive endeavor to obtain a series of "snapshots" of a dispersing population over an entire spatial domain directly, an alternative method for is to study the temporal flux through a fixed site or boundary. Fagan (1997) surrounded mantids (*Mantidae* spp.) in a field with a sticky tape boundary and inferred dispersal rates by measuring fluxes. Similarly, Zabel and Anderson (1997) studied the migration and dispersal patterns of outmigrating juvenile salmonids (*Onchorhynchus spp.*) by obtaining first-passage time distributions and inferring the underlying advection diffusion process. We apply our heterogeneity framework to first passage time (FPT) processes. Our analyses are capable of separating the effects of intrinsic, Brownian randomness from population-level heterogeneity. The analysis allows us to identify behavioral differences between species which would otherwise be

masked.

6.2 Models

6.2.1 General framework

We confine the discussion to one-dimensional movement both for the sake of simplicity and because this conforms roughly with the stream-bound movement discussed in the applications.

Consider an organism whose spatial displacement in time $X(t)$ is expressed as a temporally evolving probability distribution function $f(x|t,\theta)$, where θ represents the vector of parameters that determine the nature of the movement. If we consider a population of N identical organisms, the spatial distribution function of the population $h(x)$ is the product $h(x,t) = N \times f(x|t,\theta)$. More explicitly, the expected number of organisms found in spatial interval Δx is:

$$h(x + \Delta x, t) = N \times \Pr[x < f(x|t,\theta) < x + \Delta x] \qquad (6.1)$$

This simple product is very commonly used as the transition between the description of an individual's movement and the distribution of an ensemble of individuals. It contains within it the often unstated assumption of homogeneous behaviors within a population.

Suppose now that some parameter θ is assigned individually to each individual in a population, such that the i'th organism's movement is determined by θ_i, and that we know the distribution function $g(\theta)$ from which the θ_i's are drawn. The total population distribution is given as the sum of all the individual distributions:

$$h(x|t,\theta) = \sum_{i=0}^{N} f(x|t,\theta_i) \qquad (6.2)$$

For large N, equation (6.2) is approximated as:

$$h(x|t) = \int_{\mathbf{D}} f(x|t, \theta) g(\theta) d\theta \tag{6.3}$$

where $\mathbf{D}$ is the domain of θ. Equation (6.3) is an explicit way to account for population-level heterogeneity in spatial distributions. This method applies equally well to boundary-flux or first-passage time problems, where the distance x is known and the arrival time t is the random variable. For this class of problems, equation (6.3) is expressed as:

$$h(t|x) = \int_{\mathbf{D}} f(t|x, \theta) g(\theta) d\theta \tag{6.4}$$

For two independently distributed parameters of movement, the expression becomes:

$$h(x|t) = \int_{\mathbf{D}_2} \int_{\mathbf{D}_1} f(x|t, \theta_1, \theta_2) g_1(\theta_1) g_2(\theta_2) d\theta_1 d\theta_2 \tag{6.5}$$

and the principle can be extended for any number of parameters.

In the following section, we present several important cases for which analytical results exist. Indeed, for many biologically reasonable distributions, population level distributions exist for parameters of interest $g(\theta)$. These are analogous to conjugate prior distributions for parameters encountered in Bayesian inference.

6.3 Examples

6.3.1 Deterministic movement with heterogeneous velocity

The simplest deterministic movement is a constant linear movement beginning at the origin, such that $X(t) = vt$, where v is velocity. This equation of movement can be
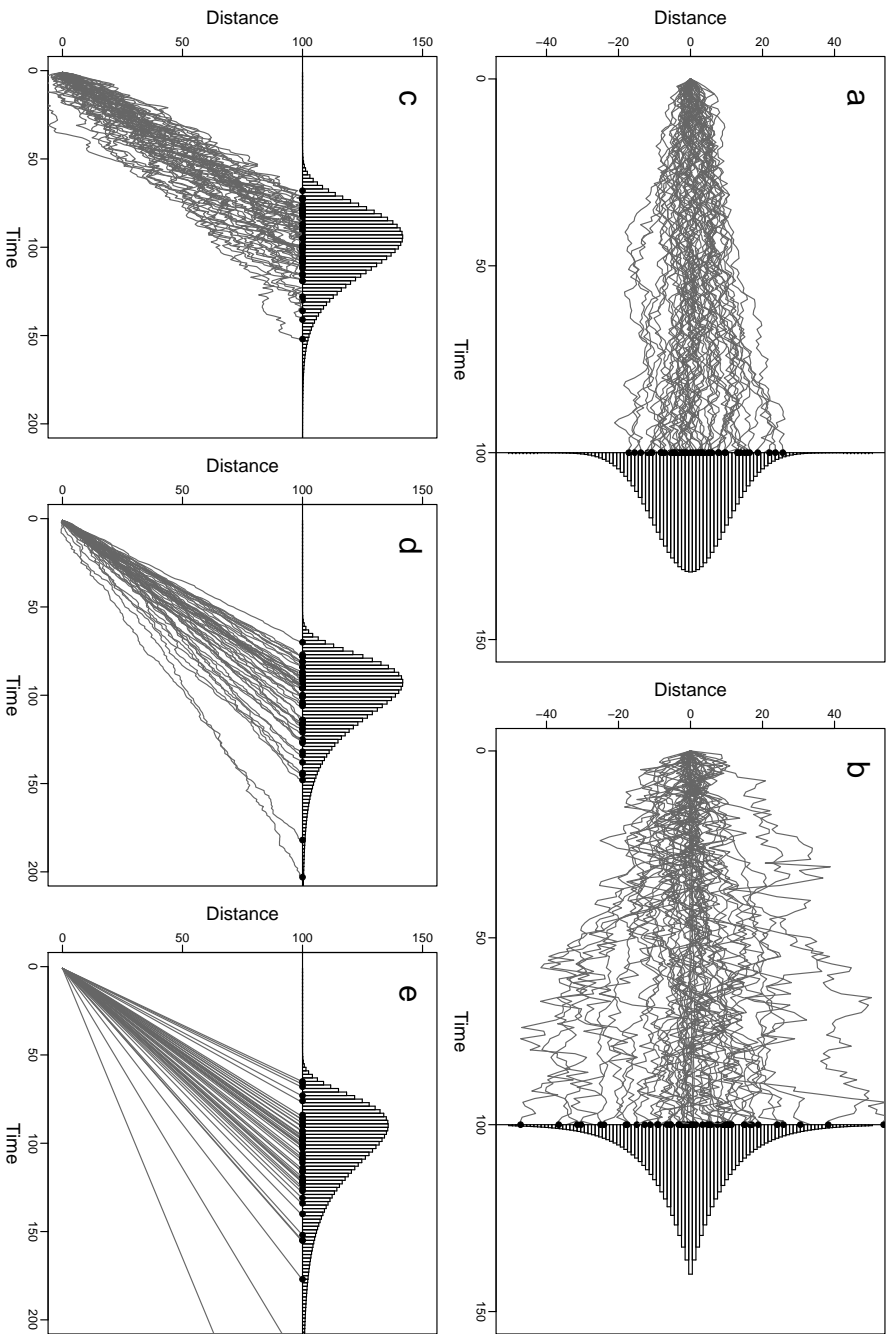
Figure 6.1: 50 simulated trajectories (gray lines) and theoretical distributions (bars). Spatial distribution at time 100 of (a) unbiased homogeneous random walkers ($\sigma_w = 1$) and (b) heterogeneous variance (σ_w) ($\sigma_w \sim$ Gamma $\sim \{\alpha = 1; \beta = 2\}$) walkers; arrival time distribution at fixed distance 100 for migrating walkers with mean velocity $\mu_v = 1$ of (c) homogeneous random walkers ($\mu_v = 1$, $\sigma_w = 2$, $\sigma_v = 0$); (d) a heterogeneous population of random walkers ($\sigma_w = 0.5$, $\sigma_v = 2$) and (e) a heterogeneous population of deterministic walkers ($\sigma_w = 0$, $\sigma_v = 2$). The theoretical distributions for these cases are presented in the text.

written as a probability density function (pdf):

$$f(x|t, v) = \delta(x - vt) \tag{6.6}$$

where $\delta(\cdot)$ is the Dirac delta function. Now, we assume a normal distribution for the velocities within the population:

$$g(v|\mu_v, \sigma_v) = \frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(\frac{(v - \mu_v)^2}{2\sigma_v^2}\right) \tag{6.7}$$

where μ_v and σ_v^2 are the mean and variance. Plugging equations (6.6) and (6.7) into (6.3) yields:

$$h(x|t) = \int_{-\infty}^{-\infty} \delta(x - vt)\frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(\frac{(v - \mu_v)^2}{2\sigma_v^2}\right) dv \tag{6.8}$$

Substituting $v = x'/t$ yields:

$$h(x|t) = \int_{-\infty}^{-\infty} \delta(x - x')\frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(\frac{x'/t - \mu_v^2}{2\sigma_v^2}\right)\frac{1}{t}dx' \tag{6.9}$$

The integrand is equal to zero everywhere except where $x' = x$, thus:

$$h(x|t) = \frac{1}{\sqrt{2\pi}\sigma_v t}\exp\left(\frac{(x - \mu_v t)^2}{2\sigma_v^2 t^2}\right) \tag{6.10}$$

The dispersal of this population increases in a Gaussian manner, with spatial mean $\langle x \rangle = \mu_v t$ and variance $\langle x^2 \rangle = \sigma_v^2 t^2$.

In the traditional biodiffusion literature, a Gaussian dispersal of organisms is a consequence of intrinsically random movements and the rate of dispersal scales with the square root of time. An important immediate result of this derivation is that

deterministically moving individuals also disperse in a Gaussian way, but that the rate of dispersal due to population heterogeneity is more rapid, scaling as t rather than $t^{1/2}$. Thus, the effects of intrinsic randomness and those of population-level heterogeneity can easily be confounded and are not easily separated when analyzing dispersal data.

In applications where there is a strong advective drift for moving organisms, data are more commonly obtained as first passage times at some fixed distance. A travel time distribution for this process is obtained in a similar fashion. The distribution of arrival times for a single constant velocity organism is given by:

$$f(t|x, v) = \delta \left(t - \frac{x}{v} \right) \tag{6.11}$$

Plugging equations (6.11) and (6.7) into (6.4) yields:

$$h(t|x) = \int_{-\infty}^{-\infty} \delta \left(t - \frac{x}{v} \right) \frac{1}{\sqrt{2\pi}\sigma_v} \exp \left(\frac{(v - \mu_v)^2}{2\sigma_v^2} \right) dv \tag{6.12}$$

Substituting $v = x/t'$:

$$h(t|x) = \int_{-\infty}^{-\infty} \delta(t - t') \frac{1}{\sqrt{2\pi}\sigma_v} \exp \left(\frac{x'/t - \mu_v^2}{2\sigma_v^2} \right) \frac{x}{t'^2} dt' \tag{6.13}$$

and solving at $t = t'$ yields:

$$h(t|x) = \frac{x}{\sqrt{2\pi}\sigma_v t^2} \exp \left(\frac{(x - \mu_v t)^2}{2\sigma_v^2 t^2} \right) \tag{6.14}$$

Some care needs to be taken in considering distribution (6.14). The integral of this function over the biologically meaningful range of $0 < t < \infty$ is less than 1. This is

because the distribution on velocities can be negative and consequently the domain on travel times ranges from $-\infty$ to ∞. In practice, when analyzing strongly advective processes the probabilities of negative velocities are very small.

It should be noted that the spatial distribution given by equation (6.10) can be obtained more directly by applying the well-known scalar transformation of the normal distribution (Casella and Berger, 1990), i.e. if $V \sim N\{\mu, \sigma^2\}$, then $X = tV \sim N\{t\mu, t^2\sigma^2\}$ (Casella and Berger, 1990). Similarly, the first passage time distribution (6.14) is the result of a straightforward change of variables: $T \sim x/V$, yielding the so-called reciprocal normal distribution. However, the derivations presented here demonstrate an implementation of population-level heterogeneity according to equations (6.3) and (6.4).

6.3.2   Stochastic movement with heterogeneous velocities

The paradigmatic stochastic process used to model animal movement is the so-called Brownian walk. In fact, a true Brownian walk implies infinite speeds at infinitesimal time intervals. A similar but more biologically meaningful approach is to model movement in terms of a Wiener process, which is a general Markovian process defined by the following conditions:

- $X(t)$ can be discretized at times $t_1, t_2, ...t_n$ with increment size Δt, such that $X(t_{i+1}) - X(t_i)$ and $X(t_{i+3}) - X(t_{i+2})$ are independent;

- $\Delta X = X(t_{i+1}) - X(t_i)$ is distributed with some constant mean $v\Delta t$ and variance σ_x^2.

The exact form of the distribution is irrelevant as the central limit theorem guarantees that at a reasonably large number of steps, the spatial distribution of the population is roughly normal, with mean vt and variance $\sigma_w^2 t$. The parameter σ_w is referred

to as the Wiener variance. The resulting probabilistic description of the organism's position is given by:

$$f(x|t, v, \sigma_w) = \frac{1}{\sqrt{2\pi\sigma_w^2 t}} \exp\left(-\frac{(x-vt)^2}{2\sigma_w^2 t}\right) \tag{6.15}$$

and the first passage time distribution at a fixed distance x is given by:

$$f(t|x, v, \sigma_w) = \frac{x}{\sqrt{2\pi t\sigma_w^2 t^3}} \exp\left(-\frac{(x-vt)^2}{2\sigma_w^2 t}\right) \tag{6.16}$$

Equation (6.16) is known to as the inverse Gaussian distribution and has been widely applied for travel time processes (Tweedie, 1957; Chhikara and Folks, 1989; Zabel and Anderson, 1997; Zabel, 2002).

We assume again the normal distribution for velocities (6.7) and apply equation (6.3) and (6.15):

$$h(x|t) = \frac{1}{2\pi\sigma_v\sigma_w t} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-vt)^2}{2\sigma_w^2 t} - \frac{(v-\mu_v)^2}{2\sigma_v^2}\right) dv \tag{6.17}$$

Solving the integral and simplifying yields:

$$h(x|t) = \frac{1}{\sqrt{2\pi(\sigma_v^2 t + \sigma_w^2)t}} \exp\left(-\frac{(x-\mu_v t)^2}{2(\sigma_v^2 t + \sigma_w^2)t}\right) \tag{6.18}$$

The basic form of this distribution, a traveling, widening Gaussian, is a straightforward mixture of the Wiener and reciprocal normal processes such that the spatial variance $\langle x^2 \rangle = \sigma_w^2 t + \sigma_v t^2$.

Analogously, the first passage time distribution (6.16) applied to (6.4):

$$h(t|x) = \frac{x}{2\pi\sigma_v\sigma_w t^{3/2}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-vt)^2}{2\sigma_w^2 t} - \frac{(v-\mu_v)^2}{2\sigma_v^2}\right) dv \tag{6.19}$$

can be solved:

$$h(t|x) = \frac{x}{\sqrt{2\pi(\sigma_v^2 t + \sigma_w^2)}t^{3/2}} \exp\left(-\frac{(x - \mu_v t)^2}{2(\sigma_v^2 t + \sigma_w^2)t}\right) \tag{6.20}$$

This distribution is similarly a mixture of the inverse Gaussian arrival time given by (6.16) and the reciprocal normal arrival time given by (6.14); each of these distributions are special cases of (6.20) where $\sigma_v = 0$ or $\sigma_w = 0$ respectively. Plots of this distribution for several parameter values are presented in figure 6.2.

6.3.3   Stochastic movements with heterogeneous Wiener variances

Consider a population of N individuals, the movement of each of which is characterized by a one-dimensional Wiener process with unique Wiener variance σ_i^2. Since variances are necessarily positive, an appropriate and flexible hypothesized distribution is the gamma distribution. Thus:

$$g(\sigma_w^2|a, b) = x^{a-1}\frac{e^{-x/b}}{b^a\,\Gamma(a)} \text{ for } x > 0 \tag{6.21}$$

where a and b are the shape and scale parameters respectively. Recall that the mean of the gamma distibution is ab and the variance is ab^2. Plugging (6.21) and the Wiener process (6.15) into the spatial equation (6.3) gives:

$$h(x|t) = \int_0^\infty \frac{1}{\sqrt{2\pi t}\Gamma(p/2)\,2^{p/2}}\sigma_w^{p-3} \exp\left(-\frac{(x - vt)^2}{2\sigma_w^2 t} - \sigma_w^2/2\right) d\sigma_w \tag{6.22}$$

which yields the analytical solution:

$$h(x|t) = \frac{1}{\Gamma(a)}\sqrt{\frac{2}{\pi tb}}\left(\frac{|x - vt|}{\sqrt{2tb}}\right)^{a-\frac{1}{2}} K_{\frac{1}{2}-a}\left(\frac{|x - vt|}{\sqrt{(bt)/2}}\right) \tag{6.23}$$

where $K_n(x)$ is the modified Bessel function of the second kind. The modified Bessel

functions exist in the positive domain and decreases monotonically; the absolute value in the argument leads to a peak at $x = vt$ with a symmetric decrease on both sides. Thus, (6.23) is a unimodal, symmetric pdf on x that displays advection at rate v and a characteristic widening typical of diffusion processes (fig. 6.3).

Distribution (6.23) has the following properties:

$$\text{Mean} \quad : \quad \mu = \mathrm{E}\,[X] = vt \tag{6.24}$$

$$\text{Variance} \quad : \quad \sigma^2 = \mathrm{E}\left[(X - \mathrm{E}\,[X])^2\right] = abt \tag{6.25}$$

$$\text{Skew} \quad : \quad \gamma = \frac{\mathrm{E}\left[(X - \mathrm{E}\,[X])^3\right]}{\sigma^3} = 0 \tag{6.26}$$

$$\text{Kurtosis} \quad : \quad \kappa = \frac{\mathrm{E}\left[(X - \mathrm{E}\,[X])^4\right]}{\sigma^4} - 3 = \frac{3a(1 + a)b^2 t^2}{(abt)^2} - 3 = \frac{3}{a} \tag{6.27}$$

The variance grows linearly with time as abt, an expected result since ab is the mean value for the Wiener variance σ_w. The kurtosis is positive, constant, and inversely proportional to the a parameter. This is an important result, since it provides an explanation for leptokurtic dispersals observed in plants (Kot et al., 1996; Clark, 1998), insects (Dobzhansky and Wright, 1943), fish (Skalski and Gilliam, 2000), and mammals (Sandell et al., 1991; Price et al., 1994) within the context of a Wiener process.

The gamma distribution for Wiener variance can also be applied to the inverse Gaussian first passage time (6.16), yielding:

$$h(t|x) = \frac{x}{\Gamma(a)} \sqrt{\frac{2}{\pi t^3 b}} \left(\frac{|x - vt|}{\sqrt{2tb}}\right)^{a - \frac{1}{2}} K_{\frac{1}{2} - a}\left(\frac{|x - vt|}{\sqrt{(bt)/2}}\right) \tag{6.28}$$

We have presented a few cases which are analytically tractable (table 6.1, figure 6.1) and intend to demonstrate some applications in which these exotic looking distributions can be applied. However, these cases are probably approaching the limit of what is useful or tractable to solve analytically. We were not, for example, able to

obtain a movement process with both normally distributed velocities and gamma distributed Wiener variances. However, where data are available, an analogous process can be applied to produce empirical distributions using numerical methods.

6.4   Model Applications

6.4.1   Spatial distribution of chub in a stream

Skalski and Gilliam (2000) performed a mark-recapture experiment on several freshwater fish species in a creek in Tennessee. In particular, 190 marked bluehead chub (Cyprinidae: *Nocomis leptocephalus*) were released at a single location and recaptured over 50 detection sites at monthly intervals and the moments of the spatial distribution were computed (table 6.2). The authors found that the dispersal of the chub displayed significant a slight mean shift increasing in time and a linearly increasing variance, consistent with Gaussian models of dispersal. The distributions also displayed a relatively constant positive kurtosis and skewness. The authors propose that the kurtosis is a consequence of a process in which the fish display two or more modes of movement: a 'fast' diffusion and a 'slow' diffusion (Skalski and Gilliam, 2000, 2003). Their model yields solutions that are essentially the sum of two Gaussians with different means and variances, which can be fitted to distribution with non-zero skew and kurtosis. The authors suggest that leptokurtosis of spatial dispersals essentially characterize the heterogeneity of the system.

While Skalski and Gilliam's model is generalizable to any number n of different modes of movement, the model is limited by the number of parameters that need to be estimated: a velocity, a diffusion parameter and a proportion for each mode of movement makes for $3n - 1$ estimates in the fully parametrized case, leading to what the authors refer to as the "spectre of parameter explosion".

Table 6.1: Table of underlying movement processes, heterogeneous parameters and resulting distributions for several simple one-dimensional cases.

Underlying process	Parameter distribution	Resulting Distribution
Deterministic movement: $f(x\|t,v) = \delta(x - vt)$	homogeneous population: constant v	1a) $h(x\|t) = \delta(x - vt)$
		1b) $h(t\|x) = \delta(t - x/v)$
	heterogeneous velocity: $g(v) = \mathrm{N}\{\mu_v, \sigma_v^2\}$	2a) $h(x\|t) = \frac{1}{\sqrt{2\pi}\sigma_v t}\exp\left(-\frac{(x-\mu_v t)^2}{2(\sigma_v t)^2}\right)$
		2b) $h(t\|x) = \frac{x}{\sqrt{2\pi}\sigma_v t^2}\exp\left(-\frac{(x-\mu_v t)^2}{2(\sigma_v t)^2}\right)$
Wiener process: $f(x\|t,v) = \mathrm{N}\{xt, \sigma_w^2 t\}$	homogeneous population: constant v and σ_w	3a) $h(x\|t) = \frac{1}{\sqrt{2\pi}\sigma_w t^{1/2}}\exp\left(-\frac{(x-\mu_v t)^2}{2\sigma_w^2 t}\right)$
		3b) $h(t\|x) = \frac{x}{\sqrt{2\pi}\sigma_w t^{3/2}}\exp\left(-\frac{(x-\mu_v t)^2}{2\sigma_w^2 t}\right)$
	heterogeneous velocity: $g(v) = \mathrm{N}\{\mu_v, \sigma_v^2\}$	4a) $h(x\|t) = \frac{1}{\sqrt{2\pi(\sigma_v^2 t+\sigma_w^2)t}}\exp\left(-\frac{(x-\mu_v t)^2}{2(\sigma_v^2 t+\sigma_w^2)t}\right)$
		4b) $h(t\|x) = \frac{x}{\sqrt{2\pi(\sigma_v^2 t+\sigma_w^2)t}^{3/2}}\exp\left(-\frac{(x-\mu_v t)^2}{2(\sigma_v^2 t+\sigma_w^2)t}\right)$
	heterogeneous Wiener variance $g(\sigma_w^2) = \mathrm{gamma}\{a, b\}$	5a) $h(x\|t) = \frac{1}{\Gamma(a)}\sqrt{\frac{2}{\pi t b}}\left(\frac{\|x-vt\|}{\sqrt{2tb}}\right)^{a-\frac{1}{2}}K_{\frac{1}{2}-a}\left(\frac{\|x-vt\|}{\sqrt{2tb}}\right)$
		5b) $h(t\|x) = \frac{x}{\Gamma(a)}\sqrt{\frac{2}{\pi t^3 b}}\left(\frac{\|x-vt\|}{\sqrt{(bt)/2}}\right)^{a-\frac{1}{2}}K_{\frac{1}{2}-a}\left(\frac{\|x-vt\|}{\sqrt{(bt)/2}}\right)$

A)

B)

Figure 6.2: Plots of the mixed inverse Gaussian and reciprocal normal (IGRN) distribution for different parameter values. In (A) IGRN distributions are plotted for several values of the velocity variance σ_v parameter; values for the remaining parameters are: $d = 10$, $v = 2$, $\sigma_w = 0.5$. A value of $\sigma_v = 0$ corresponds to a homogenous Wiener process. In (B) the IGRN is plotted for several values of the Wiener variance σ_w parameter. Values for the remaining parameters are: $d = 10$, $v = 2$, $\sigma_v = 0.5$. A value of $\sigma_w = 0$ corresponds to a heterogeneous, deterministic velocity process.
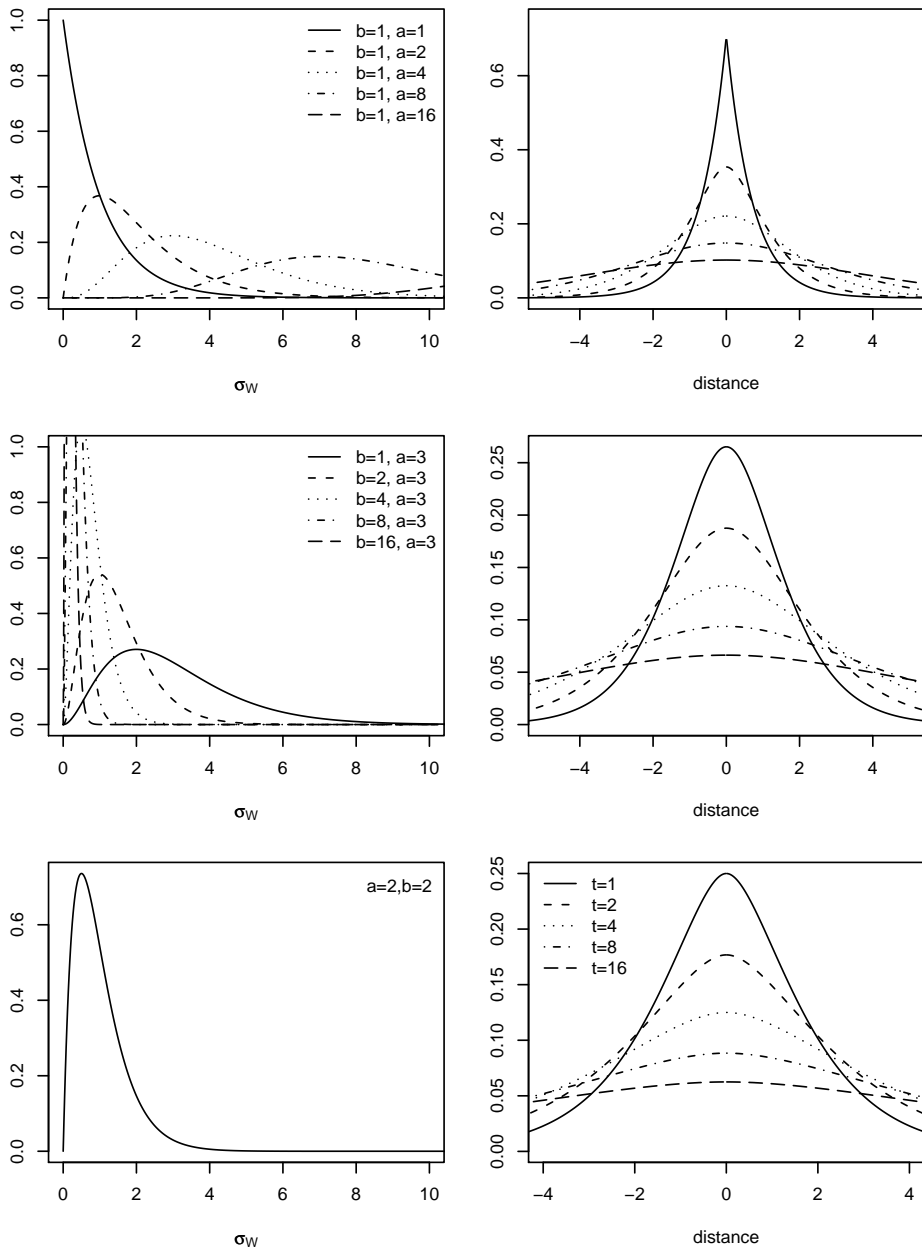
Figure 6.3: Plots of the gamma distribution and associated GVD distributions 6.23 for various values for a and b and times t. The velocity parameter v for all of these distributions is 0.
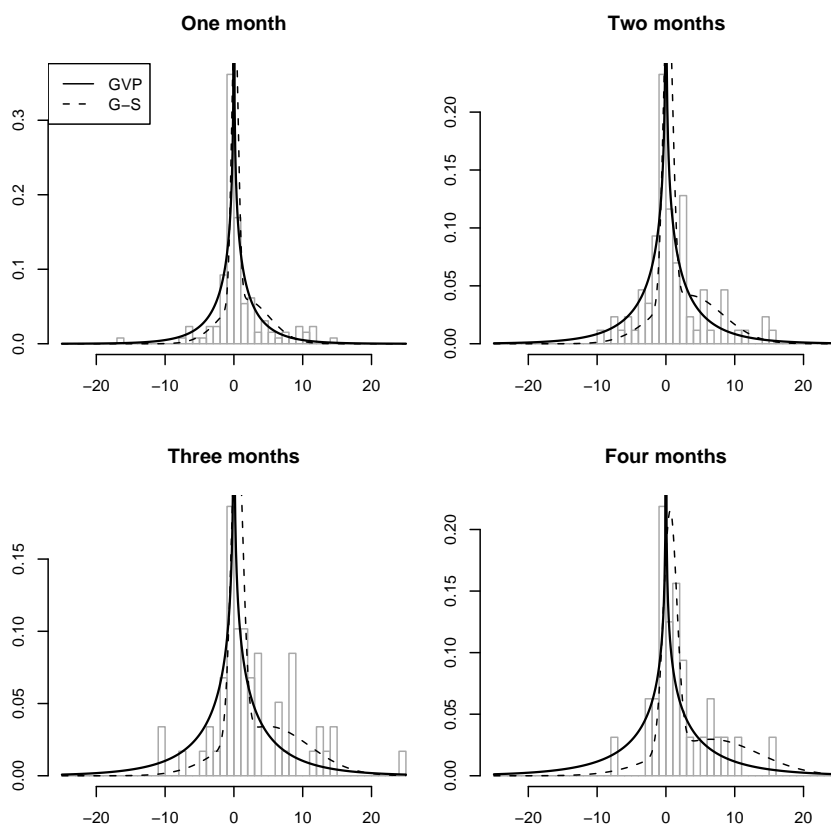
Figure 6.4: Comparison of the 3-parameter gamma-distributed variance model (G-V) and the five parameter Skalski-Gillam model (S-G) to histograms of bluehead chub disperal data at four months of observation.

Table 6.2: Mean, variance, skewness and kurtosis estimates for spatial distribution of bluehead chub (*Nocomis leptocephalus*) movements from (Skalski and Gilliam, 2000). The fish were marked and released at one site and monitored over a 4-mo period. Dispersal distance is measured in terms of number of sites.

Days	Mean	(se, n)	Variance	(se, n)	Skew	(se, n)	Kurt	(se, n)
0	0	(0,190)	0	(0,190)		na		na
30	1.13	(0.35,134)	22.29	(5.42,157)	0.44	(0.43,134)	7.34	(0.39,157)
60	1.57	(0.5,86)	27.40	(7.86,101)	0.95	(0.31,86)	6.37	(0.48,101)
90	3.19	(0.8,59)	45.64	(14.03,69)	0.87	(0.25,59)	4.58	(0.57,69)
120	2.44	(0.77,32)	66.62	(30.69,44)	1.08	(0.22,32)	7.51	(0.7,44)

Table 6.3: Parameter estimates for the gamma-distrubuted variance model for the chub dispersal data.

parameter	symbol (units)	estimate	(95% confidence interval)
shape parameter	a (unitless)	0.47	(0.426,0.506)
scale parameter	b (sites/day)	1.15	(0.497,1.809)
mean velocity	v (sites/day)	0.03	(0.0172,0.0352)

Estimating parameters

The gamma-variance process GVP (6.23) resolve this parameter explosion by summarizing all the possible kinds of movement of a population into a single two-parameter distribution. We estimated the parameters of the GVP by exploiting the simple expressions for the theoretical expected mean $\mu = vt$ (6.24), variance $\sigma^2 = abt$ (6.25) and kurtosis $\kappa = 3/a$ (6.27) where v is the velocity parameter and a and b are the shape and scale parameters of the gamma-distributed Wiener variance. Thus, we obtain $\hat{v}$ by regressing the mean measurements μ against time t, obtain an estimate for $a \times b$ by regressing the variance $\hat{\sigma^2}$ against time, and use the measured kurtosis $\hat{\kappa}$ to obtain a. Estimates of the standard error were obtained by performing these estimations 10,000 times over simulated measurements using Skalski and Gilliam's reported standard errors.

The results of our parameter estimates and model fits are presented in table 6.3. In figure 6.4, we compare our three parameter model with Skalski and Gilliams two Gaussian fish groups which fits five parameters: fast and slow fish diffusions (D_f, D_s), fast and slow fish advections (β_f, β_s) and a proportion of fast fish to slow fish (Z).

6.4.2   Migration times of outmigrating salmonids

Data

We applied the first passage time distribution with heterogeneous velocities (6.20) to data on seaward migration of juvenile salmon (*Oncorhynchus* spp.) in the Snake and Columbia rivers in the northwestern of the United States. Since the 1990's, hundreds of thousands of juvenile salmonids have been implanted with passive integrated transponder (PIT) tags and detected at hydroelectric projects downstream. PIT tags are small tags inserted into the body cavity and retained for life. Each tag contains a unique code that is recorded with the time of passage when the fish passes near detectors placed at fish bypass systems further downstream (Prentice et al., 1990). In this study, we analyze data from spring-run Chinook salmon (*O. tschawytcha*) and steelhead trout (*O. mykiss*) released in groups throughout their migratory season over a ten year period from 1996 through 2005. Details of the capture and tagging methods for each of the four groups are given by Marsh et al. (1997); Buettner and Brimmer (1998); Smith et al. (2000). The PIT-tag data were downloaded from a regional database available to the public (Pacific States Marine Fisheries Commission, 1996).

We chose to focus on these two species because they are of similar size (typically between 100 and 230 mm) and display similar migration timing. Both are listed as an threatened ESU's under the Endangered Species Act (USNMFS, 1992, 1998). We further focus our analysis on travel times between Lower Granite and Little Goose dams, a distance of 37.2 miles (59.9 km), and on fish traveling between April 10 and May 20, thereby capturing the largest number of both species in all years.

Flows were also measured at both dams and analyzed as a covariate in the parameter estimations. The average flow over the 40 day period varied from a maximum of 4530.7 m^{3}s^{-1} in 1997 to a minimum of 1868.9 m^{3}s^{-1} in 2004.

The numbers of fish released each year are quite large (mean 8457) but varied

widely between 254 Chinook detected in 1997 and 30460 steelhead detected in 2000. This variability reflects both the number of fish tagged and the detection probability, which could be compromised by high flows, as in 1997.

Statistical methods

We fit and assessed three models, the inverse Gaussian (IG) model (eq. 6.16), which assumes homogeneous organisms with Wiener-movement; the reciprocal normal (RN) model (eq. 6.14), model which assumes a fully heterogeneous population of deterministic organisms; and the combined IGRN model (eq. 6.20). Estimates for these models were obtained by using maximum likelihood estimation.

For the IG model, the maximum likelihood estimates for v and σ_w are given by:

$$\hat{v} = \frac{d}{\bar{t}} \tag{6.29}$$

$$\hat{\sigma}_w^2 = \frac{d^2}{N} \sum_{i=1}^{N} \left(\frac{1}{t_i} - \frac{1}{\bar{t}} \right) \tag{6.30}$$

There is a slight bias in the MLE estimator for σ_w. An unbiased uniform minumum variance estimator (Tweedie, 1957; Folks and Chhikara, 1978) is given by:

$$\hat{\sigma}_w^2 = \frac{d^2}{N-1} \sum_{i=1}^{N} \left(\frac{1}{t_i} - \frac{1}{\bar{t}} \right) \tag{6.31}$$

The RN model can be readily transformed into a normally distributed variable via the transformation $Y = 1/X$, such that the MLE estimates are:

$$\hat{v} = \frac{d}{\bar{\bar{t}}} \tag{6.32}$$

$$\hat{\sigma_v}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{d}{t_i} - \hat{v} \right)^2 \tag{6.33}$$

For the IGRN model $\hat{\mu}_v$ can be obtained in terms of the other parameters:

$$\hat{\mu}_v = \sum_{i=1}^{n} \frac{x}{\hat{\sigma}_w^2 t_i + \hat{\sigma}_v^2} \Big/ \sum_{i=1}^{n} \frac{t_i}{\hat{\sigma}_w^2 t_i + \hat{\sigma}_v^2} \tag{6.34}$$

There are, however, no analytical expressions for MLE's of the IGRN model variances, however these can be readily estimated using an optimization routine to maximize the log-likelihood function. All estimates were obtained using the R statistical package.

In order to obtain confidence intervals around the estimates for the IGRN model, we used a bootstrapping routine where estimates were generated for 1000 resamplings with replacement of each dataset and 95% empirical quantiles were obtained from the bootstrap distribution (Moore and McCabe, 2006). Quality of model fits were assessed used the Kolmogorov-Smirnov goodness of fit test, which measures the largest vertical distance between the empirical cumulative density and the theoretical density (Massey Jr., 1951). We used Akaike's information criteria (AIC) to compare models, which adjusts the likelihood function according to the number of parameters estimated (Akaike, 1974). Parameter estimates were regressed against mean flows using standard linear regression. Quality of distribution fits to data was further assessed visually using percentile (P-P) plots (Wilk and Gnanadesikan, 1968).

We summarize the role of heterogeneity in describing a migration process with a dimensionless index of heterogeneity ϕ defined as:

$$\phi = \sqrt{\frac{\sigma_v^2 v}{\sigma_v^2 v + \sigma_w^2 d}} \tag{6.35}$$

This index corresponds to the amount that population-level heterogeneity contributes to the total spatial variance of the migrating population when the spatial mean of the population distribution at some distance d. For these data, we choose the distance at which the data is collected (37.2 miles). For a homogeneous ($\sigma_v = 0$) population, $\phi = 0$; for a fully heterogeneous ($\sigma_w = 0$) population of deterministic travelers, $\phi = 1$.

Results

In almost all cases, the IGRN model was the best fitting and most parsimonious model according to the AIC index (table E.1, figure 6.5). The only exceptions were when the IGRN model estimates a Wiener variance of zero, essentially collapsing into the RN model, as occurs in 1997, 1999 and 2003 for the steelhead travel time data.

We present values of velocity estimates ($\hat{v}$) and heterogeneity index estimates ($\hat{\phi}$) with bootstrapped 95% confidence intervals and averaged flows in table 6.4 and plot these results in figure 6.6.

The velocity parameter estimates $\hat{v}$ were significantly higher for steelhead (mean 13.26, s.e. 3.71) than for Chinook (mean 7.64, s.e. 1.61, paired t-test p-value 0.001, table 6.4). Similarly $\hat{\phi}$ is much higher for steelhead (mean 0.665, s.e. 0.26) than for Chinook (mean 0.200, s.e. 0.16, p-value $\ll 0.001$). We performed simple linear regressions of these these estimates against flow. Steelhead showed significant responses in both $\hat{v}$ (slope 0.12 miles/day/kcfs, p-value $\ll 0.001$) and $\hat{\phi}$ parameters (slope 0.13 kcfs^{-1}, p-value 0.037), while Chinook showed no response (p-values 0.48 and 092 for $\hat{v}$ and $\hat{\phi}$ respectively, figure 6.6).

6.4.3 Seattle marathon analysis

Finally, we applied the IGRN model (6.20) to the results of the University of Washington Medical Center Seattle Marathon, run on Sunday, November 26, 2006. Participants in the marathon included women (n=741) and men (n=1314) aged 10 to 99 years, ranging from walkers to sub-elite competitive marathoners. Each participant
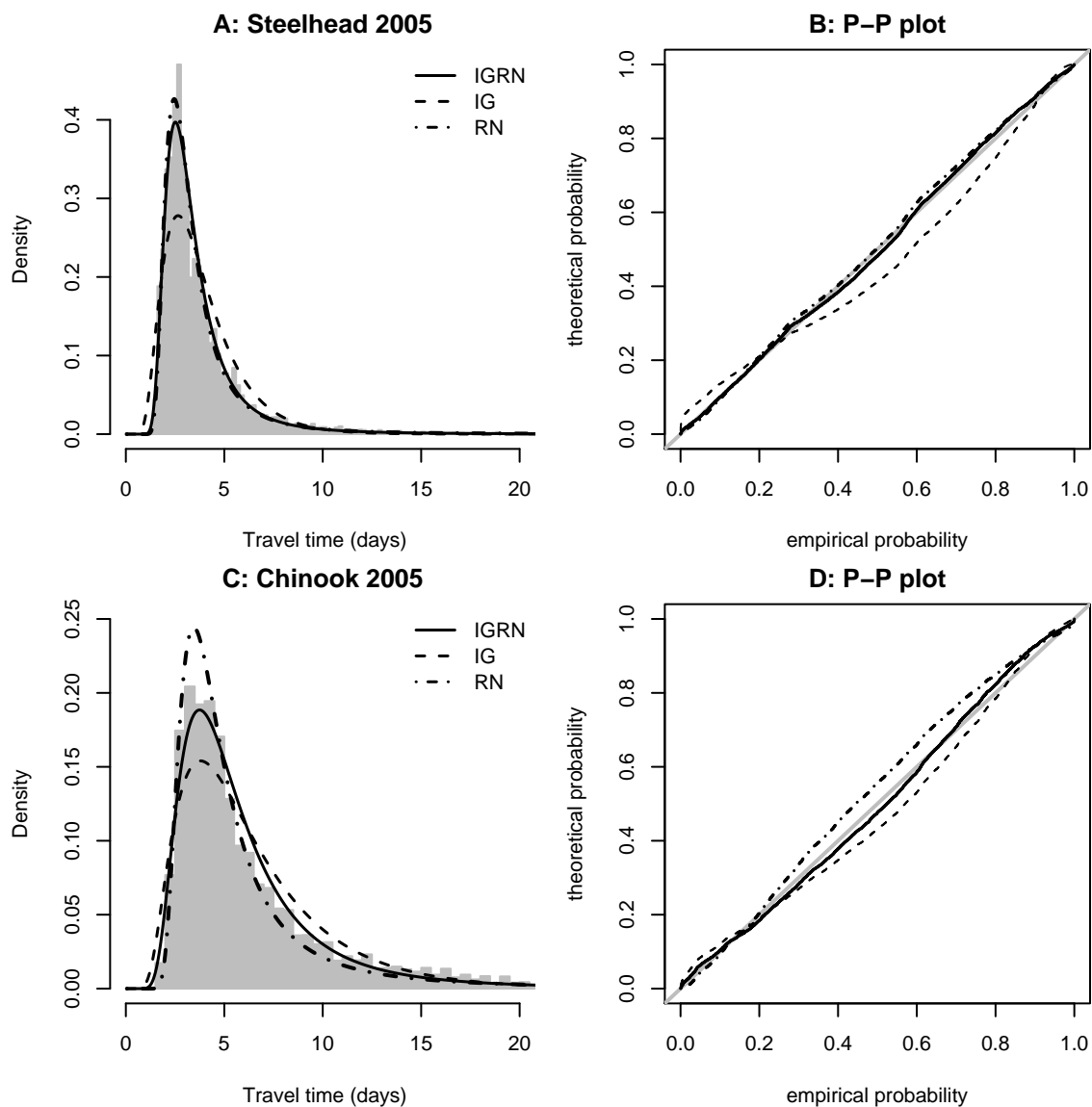
Figure 6.5: Examples of travel time model fits to travel times data. On all plots, the solid line is the IGRN model (6.20), the dashed line represents the homogeneous population of Wiener movers (IG model 6.16) and the dotted line represents the heteregeneous, deterministic fish model (RN model 6.14). The histograms represent travel times for (A) steelhead and (C) yearling chinook released at Lower Granite and detected at Little Goose dam, 37.2 km downstream, between May and June, 2005. The P-P plots (B) and (C) are a visual way to assess the fit of data to different theoretical distributions, with the 45 deg line representing a perfect fit. In all of these plots, the IGRN model is the best fit. For the steelhead the RN model is a very close second.
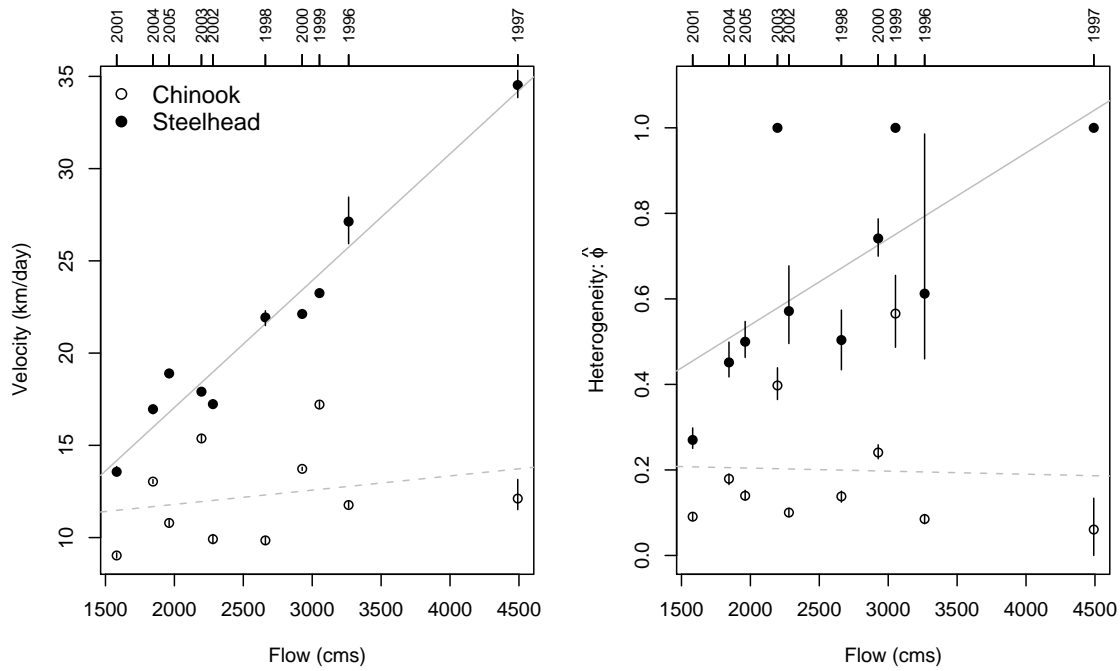
Figure 6.6: Plots of (A) velocity estimate ($\hat{v}$) and (B) heterogeneity index estimates ($\hat{\phi}$) over all years against mean flow (in kcfs). The crosses represent results for steelhead, the circles represent results for chinook. The solid and dashed lines represent linear regressions against flow for steelhead and chinook respectively. The vertical bats are bootstrapped 95% confidence intervals. Steelhead clearly show both higher values of $\hat{v}$ and $\hat{\phi}$ and stronger responses to flow.
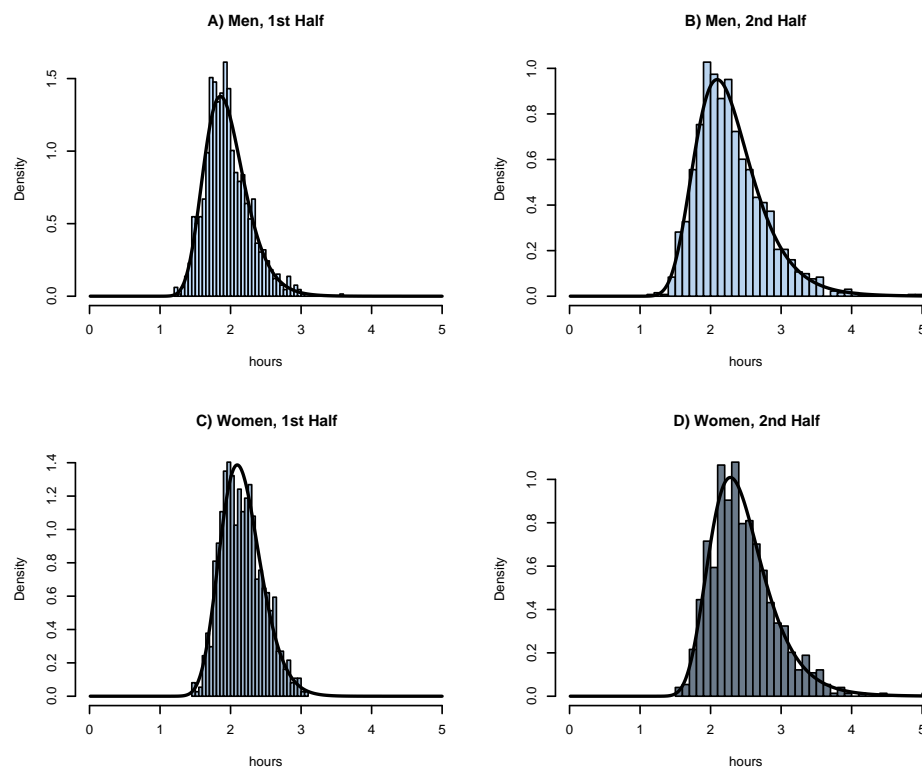
Figure 6.7: Histrograms of Seattle marathon results, broken up into two halves, with fitted IGRN model distributions.
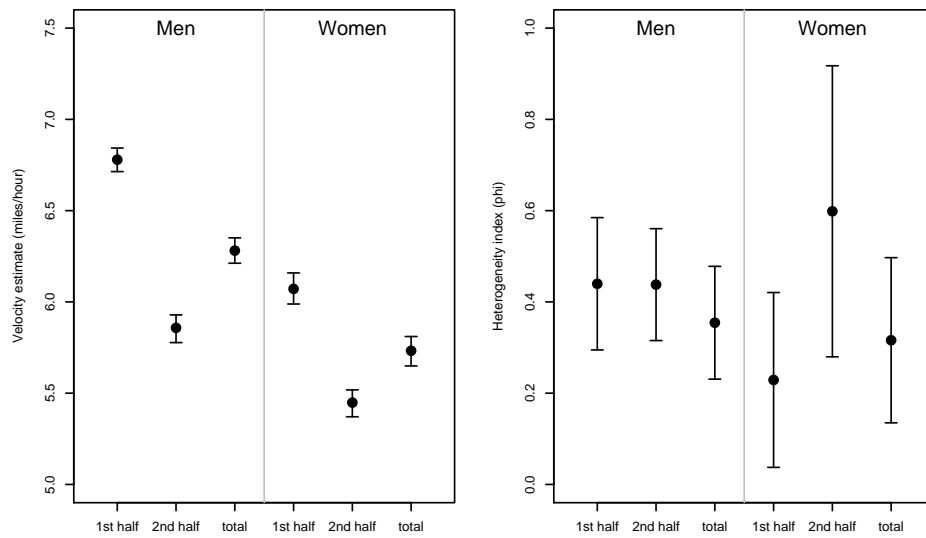
Figure 6.8: Plots of estimate results for (a) velocity v and (b) heterogeneity index ϕ. Vertical bars indicate estimates 95% confidence intervals. These are wider for the women because of smaller sample size.

was fitted with a timing chip that recorded the exact time of leaving the starting gate, passing the half-way mark (13.1 miles) and reaching the finish line (26.2 miles). The course is described as moderately hilly, beginning at 100 ft above sea level, dipping to 40 feet, and topping out at 200 ft. The first four miles are rather hilly, followed by a gradual downward slope and a long (11 mile) flat stretch, while the final eight miles is hilliest, including the steepest climb. The second half of the race is thus considerably more strenuous than the first.

All results for the 2006 Seattle marathon as well as a description of the course are publicly available on their website at http://www.seattlemarathon.org/.

Results

Distributions of race times for men and women are presented in figure 6.7. Race times ranged widely from 2.46 hours to 7.84 hours for men and 3.02 hours to 7.38 hours for women. Parameter estimates of the IGRN model fits are presented in table 6.5 and plotted in figure 6.7. The mean velocity estimates (presented in figure 6.8A) was significantly lower for the women's race time distribution ($\hat{v}$=5.7 miles/hour) than for the men ($\hat{v} = 6.2$) and for the second half of the race ($\hat{v}$=5.6) than for the first half of the race ($\hat{v}$=6.5), as is to be expected. Bootstrapped 95% confidence intervals for velocity estimates were fairly narrow (between 0.13 and 0.17).

Results of the heterogeneity index are plotted in figure 6.8B. Confidence intervals were much wider for women (from 0.36 to 0.64) than for men (0.24 to 0.29), due primarily to the smaller sample size. Men displayed a constant $\hat{\phi}$ around 0.40 (C.I.: {0.23,0.48}), while women showed a low ϕ (0.23, {0.04,0.42}) for the first half, but a much higher one for the second half of the race (0.6, {0.29,0.91}).

Discussion of marathon results

The participants in the Seattle marathon are clearly a heterogeneous group of racers with a wide range of abilities. If each runner were to pace themselves at a constant

pace throughout the race our model predicts that heterogeneity would dominate the dispersal of race times and ϕ would be around 1.0. In fact it tends to be much lower (mean estimates between 0.23 and 0.60). This result suggests that homogeneity and intrinsic randomness play a significant role in predicting race times. Homogeneity is reflected in the considerable clumping between racers, who tend to "buddy-up", especially in a pool where most participants are not running competitively. Intrinsic randomness is reflected in the considerable variability in the pacing itself, as individual racers speed up and slow down as a function of the difficulty of the run.

The most striking result is the dramatic increase in the heterogeneity parameter for women in the second half of the race. The much lower value of ϕ in the first half of the race for women might a reflection a greater tendency for clumping among women, while the great increase in the second might be interpreted as a greater differentiation or spreading in the field of women racers during the second, more arduous half of the course. Hopkins and Hewson (2001) present evidence suggesting that female runners are less variable in their race results than male runners. If this is interpreted as a better ability in general to pace oneself, this would be consistent with the high heterogeneity index identified for the female racers.

Again, a thorough analysis of the data would consider other covariates, such as age and experience, and consider alternative underlying assumptions behind the models. However, this application illustrates the wide applicability and potential informative nature of fitting models that explicitly incorporate population-level heterogeneity.

6.5 Summary and general discussion

6.5.1 Dispersing chub

The gamma-distributed variance process captures the important features of the blue-head chub data, namely linearly increasing spatial variance and positive, constant kurtosis, while neatly taking care of Skalski and Gilliam's potential 'parameter ex-

plosion'. It demonstrates how a direct incorporation of continuous population-level heterogeneity within the framework of a standard diffusive processes can yield leptokurtosis.

The major drawback of our model is that it can not account for the skewness, though a model built on the same framework that includes a continuous distribution of velocities for the population would. Unfortunately, we were not able to obtain an analytical distribution that could account for both a continuous distribution of mean velocities and Wiener variances.

On the other hand, a particularly attractive feature of our model is the simplicity of the parameterization, i.e. the surprisingly straightforward relationship between the a and b parameters which describe the distribution of the Wiener variance and the predicted spatial variance σ^2 and kurtosis κ. At no point did we require a fitting algorithm or a minimization routine.

It should be noted that we chose a flexible distribution with 'nice' properties to describe the variance parameter. It is, however, empirically unsupported. Within the heterogeneity framework, one might collect data on some parameter of the fish and empirically estimate the predicted distribution of a dispersion parameter. For example, lengths and weights of fish are routinely collected and might be used to motivate prior estimates of population-level heterogeneity. Our models provide a framework within which laboratory experiments can be extrapolated to field studies.

6.5.2 Migrating salmonids

The IGRN travel time model, which allocates the spatial dispersal of a migrating population to intrinsic randomness and population-level heterogeneity, fit the salmonid migration data significantly better than the inverse Gaussian model. Differences in the nature of the dispersal yield fresh insights into the nature of migration for migrating salmon. That steelhead trout migrate faster than yearling chinook salmon is a well-documented observation (Cada et al., 1997; Giorgi et al., 1997). However,

162
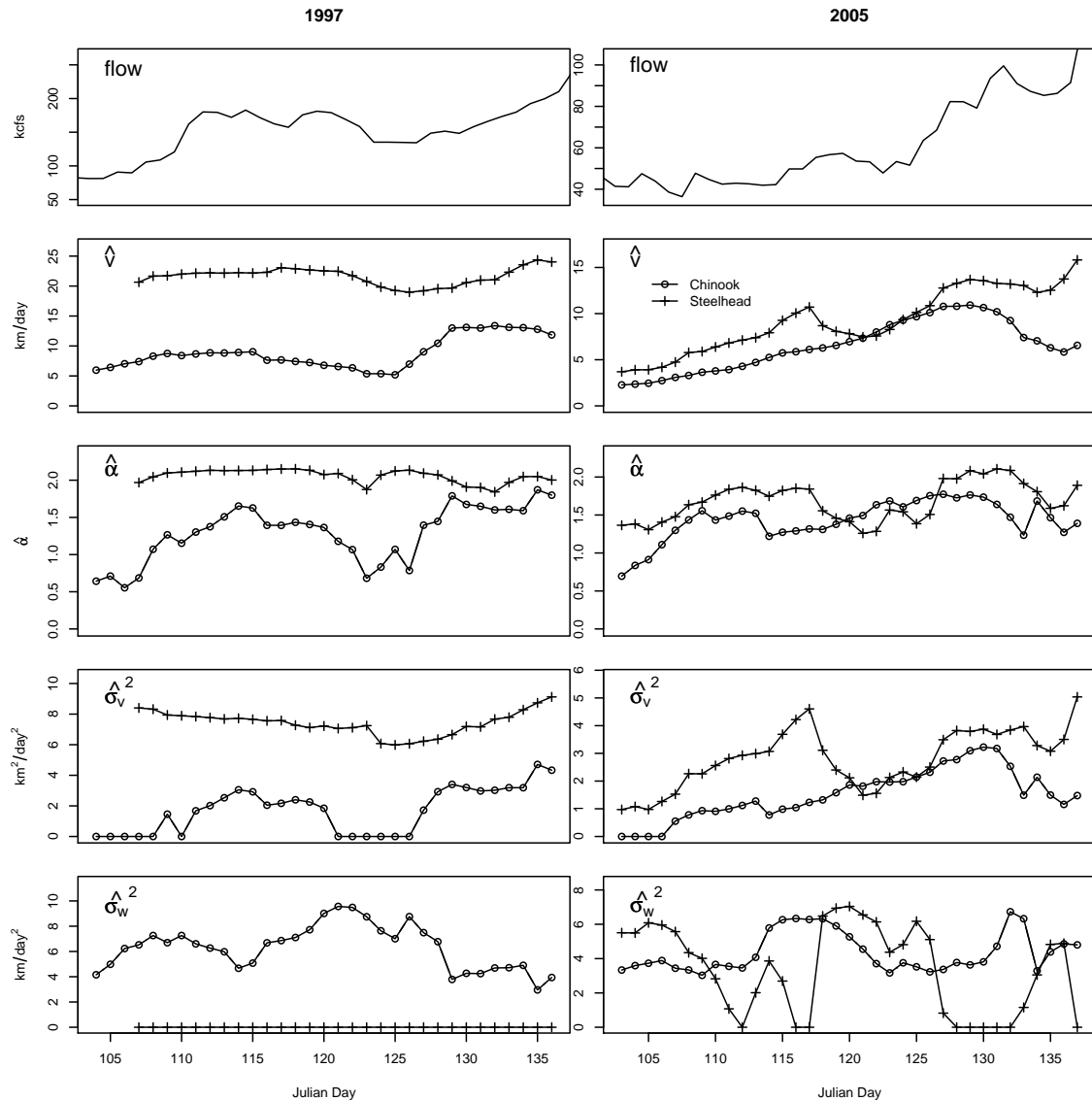
Figure 6.9: Comparison of parameter estimates as the season develops between 1997, an abnormally high flow year, and 2005, a relatively low flow year. The circles represent estimates for spring chinook; the pluses represent steelhead. In order to obtain sufficient sample sizes to produce reliable estimates, they are obtained from data pooled over a 7 and 5 day period around a given day respectively in 1997 and 2005.

the estimation of an additional parameter captures an additional behavioral signal. If the underlying assumptions of the dispersal models are valid, our results indicate that steelhead exhibit more heterogeneity-dominated migration ($\phi \approx 1$), i.e. each individual selects it's own velocity and moves in a relatively constant and directed way downstream with relatively little milling or randomness. Furthermore, steelhead appear to track very closely the flow of the river, with a strong linear interyear response to velocity. The relationship between the heterogeneity index ϕ and flows is most likely a result of greater variability in the flow regimes in higher flow years. This is consistent with the observation by Dauble et al. (1989) that steelhead in particular prefer the faster, deeper central channels of the river.

Chinook, in contrast, exhibit a more diffusion-like migration, likely engaging in diurnal feeding and station-holding behaviors during migration. The lack of a response to higher flows indicates that their behavior is less dictated by river velocity. A possible explanation might be that they cling closer to the slower moving waters near the riverbanks or have a stronger diel pattern of migration.

To further explore these hypotheses, the data needs to be scrutinized more closely. Notably, distributions within narrower migration windows need to be analyzed for more detailed responses to environmental variation. In figure 6.9, I present

For the purposes of this paper, however, it is most noteworthy that the construction of a model with the explicit incorporation of population level heterogeneity has provided a strong behavioral signal, reinforced by the consistency in the results and the robustness of the estimation routine. This signal is undetectable with standard diffusion models and inverse Gaussian arrival time distributions.

6.5.3   General Discussion

Though rarely stated, an implicit assumption of homogeneity underlies many applications of diffusion-based models of animal movements. This, despite the fact that one of the few safe generalizations one can make about populations of organisms is

that individuals are not identical and that natural environments are heterogeneous in space in time. Indeed, it would seem population-level heterogeneity should play at lease as central a role as random movement models in cases where dispersal processes are being modeled.

We have seen that a heterogeneity model is highly successful at capturing the main features of observed cyprinid fish dispersal in a stream observed by Skalski and Gilliam (2000). While certainly simplistic, the population level distribution assumed for the Wiener variance parameter is more realistically continuous and is very efficiently parameterized. Furthermore, a population level heterogeneity is in principle measurable in a laboratory environment. Thus, for example, swimming speeds can be assessed in terms of fish size, age or sex, as can less physical phenotypic traits such as 'boldness' (Fraser et al., 2001).

In the case of travel time data for directed, migrating organisms in a essentially one-dimensional geometry (e.g. migrating salmonids in a river) we can explicitly separate the contribution of intrinsic, random effects and population-level effects to the total dispersal with a single first-passage time dataset. The consistency of differences between populations which are experiencing similar environments and constraints provides strong evidence that a real behavioral difference is being observed.

It should be emphasized that there are many kinds of heterogeneity. An individual can display heterogeneous behaviors, a population can be heterogeneous in a given trait, and an environment can be spatially and temporally heterogeneous in some variable. In our formulation, heterogeneous behaviors are absorbed into the description or random movement, while heterogeneous populations are modeled using some simple assumptions about distributions within populations in parameters related to movement (velocities and Wiener variances). Appropriate ways of incorporating environmental heterogeneity is specific to the problem.

In the analysis of salmonid data, for example, while we have described a process with heterogeneous velocities, there is no way of knowing to what extent this hetero-

geneity is intrinsic to the population and to what extent it is a reflection of variable environmental conditions. Analysis of the data for narrower temporal windows is reflected by a decrease in heterogeneity, since organisms migrating in a narrower time window experience a narrower range of environments. By comparing the rate of decrease of the heterogeneity index to the size of a temporal window it might be possible to identify more exactly the 'intrinsic' heterogeneity of a population on the one hand, and to formulate more sophisticated models of immediate environmental response on the other. This approach is ultimately constrained by the data since parameter estimates for smaller cohorts are obtained at the cost of greater uncertainty.

At the smallest level, an individual organism responds in well-defined ways to it's immediate environment and internal state. From its perspective, there is presumably little that is truly 'random' about its movements. The more information that one has to explain a movement, the less random it becomes; a point quantified by Zabel (2002). When describing a system that consists of many different individuals in a complex environment moving in space and in time, notions of randomness or stochasticity, heterogeneity, variance and error have a tendency to blur together. However, when precisely defined they all refer to very distinct ideas. Here we have developed and presented a theoretical framework that explicitly separates stochasticity from heterogeneity. Under certain specific but biologically reasonable assumptions the framework yields analytical results in which parameters can be readily estimated. The framework, however, is very general. Where analytical results do not exist, it is tractable to empirically posit population-level distributions in parameters of movement and predict dispersal distributions and rates.

Table 6.4: Results of parameter estimation for fitting the IGRN model (6.20) to salmonid travel times. The confidence intervals were calculated using a bootstrapping procedure. The flow is the mean measured flow over the period at Lower Granite dam measured in kilo-cubic feet per second (kcfs). Mean velocity estimate are in miles per day. The ϕ parameter represents the contribution of population-level heterogeneity to the dispersion of the population at Little Goose dam, 37.2 miles downstream.

		Chinook				Steelhead			
Year	Flow	$\hat{v}$	95% C.I.	$\hat{\phi}$	95% C.I.	$\hat{v}$	95% C.I.	$\hat{\phi}$	95% C.I.
1996	116.58	7.31	(7.21,7.43)	0.09	(0.074,0.096)	16.86	(16.11,17.68)	0.61	(0.46,0.99)
1997	160.40	7.53	(7.15,8.17)	0.06	(0,0.134)	21.46	(21.03,21.95)	1.00	(1,1)
1998	95.01	6.12	(5.99,6.23)	0.14	(0.125,0.15)	13.63	(13.35,13.85)	0.50	(0.43,0.57)
1999	109.02	10.69	(10.59,10.79)	0.57	(0.487,0.655)	14.45	(14.32,14.58)	1.00	(1,1)
2000	104.55	8.53	(8.48,8.58)	0.24	(0.226,0.259)	13.75	(13.68,13.8)	0.74	(0.7,0.79)
2001	56.48	5.61	(5.53,5.7)	0.09	(0.083,0.1)	8.43	(8.31,8.61)	0.27	(0.25,0.3)
2002	81.41	6.16	(6.01,6.32)	0.10	(0.091,0.11)	10.71	(10.56,10.85)	0.57	(0.5,0.68)
2003	78.44	9.55	(9.46,9.67)	0.40	(0.365,0.439)	11.13	(11.04,11.24)	1.00	(1,1)
2004	65.88	8.10	(8.03,8.17)	0.18	(0.166,0.191)	10.54	(10.39,10.68)	0.45	(0.42,0.5)
2005	70.07	6.71	(6.6,6.83)	0.14	(0.13,0.152)	11.74	(11.6,11.88)	0.50	(0.46,0.55)

Table 6.5: Results of IGRN model (6.20) fitting to Seattle Marathon data. The confidence intervals were calculated using a bootstrapping procedure. Velocity (v) and velocity deviation (σ_v) are measured in miles per hour, Wiener deviation (σ_w) is measured in miles/hour^2.

Section	para-meter	Men ($n = 1314$)		Women ($n = 741$)	
		MLE	95% C.I.	MLE	95% C.I.
1st Half	$\hat{v}$	6.779	(6.71,6.84)	6.07	(5.99,6.16)
(13.1 miles)	$\hat{\sigma}_v$	0.724	(0.502,0.914)	0.375	(0,0.649)
	$\hat{\sigma}_w$	1.063	(0.695,1.304)	1.086	(0.763,1.249)
	$\hat{\phi}$	0.44	(0.294,0.586)	0.229	(0.04,0.42)
2nd Half	$\hat{v}$	5.858	(5.78,5.93)	5.45	(5.37,5.52)
(13.1 miles)	$\hat{\sigma}_v$	0.859	(0.641,0.998)	0.824	(0.521,0.953)
	$\hat{\sigma}_w$	1.18	(0.829,1.474)	0.711	(0,1.166)
	$\hat{\phi}$	0.438	(0.313,0.563)	0.599	(0.29,0.91)
Total	$\hat{v}$	6.281	(6.21,6.35)	5.73	(5.65,5.81)
(26.2 miles)	$\hat{\sigma}_v$	0.802	(0.57,0.951)	0.638	(0.337,0.824)
	$\hat{\sigma}_w$	1.466	(0.973,1.849)	1.268	(0.625,1.73)
	$\hat{\phi}$	0.354	(0.231,0.478)	0.316	(0.14,0.5)

Chapter 7

CONCLUDING THOUGHTS

Nil actum reputa si quid superest agendum.

Don't consider that anything has been done if anything is left to be done.

Lucan (39-65 AD)

This dissertation has been primarily methodological, covering thoroughly the "Quantitative" component of the degree I am hoping to have conferred upon me. The rest of the degree title, however, includes the words "Ecology" and "Resource Management",[1] so I conclude the dissertation with some thoughts about the role of movement models in asking ecological questions, and the role of movement studies in general in the very applied field of conservation.

7.1 Movement models and the science of ecology

Ecology as a field of scientific inquiry is highly empirical and hypothesis-driven. Typical studies in ecology might be broken down into the following generic steps: First, a question is asked (e.g. how does presence or absence of birds affect plant composition in tropical forests). This question is often formalized in the context of an existing theory or hypothesis (e.g. Janzen-Connell hypothesis of density dependence and distance-related survival for seeds). Second, An appropriate study system is selected (e.g. bird-free Guam versus unaffected neighboring islands) and measurements of some appropriate variable are obtained (e.g. measurements of dispersal distance

---

[1]as well as "Philosophy"!

and germination success of seeds).[2] Third, the data is distilled using some statistical summaries (e.g. means and variances of log-transformed dispersal distances and estimated probabilities of germination success). Fourth, rigorous hypothesis tests of the data are performed (e.g. factorial analysis of variances or generalized linear models mortality). Finally, depending on the results of the tests, support for or against the theory or prediction is obtained (e.g. the absence of birds in Guam significantly affects the reproductive success and dispersal of seeds[3]).

In the chapters of my dissertation, nowhere do I propose or test an ecological hypothesis or explore any underlying theories. Indeed, in the context of the model outlined above, most of my work has been confined to developing elaborate versions of the standard summary statistics. I supplant the typical double prong of means and variances with a veritable hydra of mean advections, measures of net variability, and time-length scales of significant autocorrelation, all further complicated by decomposition into orthogonal components, the inclusion of rotational tendencies, and/or population-level heterogeneities. Because these variables all have specific mathematical meaning and statistical properties, rigorous analyses can be performed allowing, for example, for the identification of discrete behavioral shifts or the quantification of significant population-level heterogeneity. However, interpretations of movement data in terms of actual behavior models, bioenergetics, individual responses to environmental covariates, or in an evolutionary context are largely absent. The only theoretical predictions are those that relate the encounter rate models to individual movement parameters. The grand goal of inferring the heuristic behavior function $f(\cdot)$ presented in the Introduction (equation 1.1) remains largely unfulfilled.

Clearly, informative mathematics and robust statistics are essential to the functioning of the generic version of the scientific method outlined above. However, the

---

[2]In practice, of course, it is the existence of a particular study system that dictates the question, but the logical structure of studies in ecology are rarely presented in that order of precedence.

[3]Example inspired by H. Rogers (*pers. comm*).

complexity of some of the methods begs the question of how these *mathematical* and *statistical* models can be applied to develop *behavioral* and *ecological* models. While there is no simple answer to this question, the first-order theory-derived hypothesis-testing framework outlined above might in practice be too narrow. In very general terms, I would advocate a sort of iterative process that boils down to the following steps: (1) develop informative movement models with a certain number of assumptions to distill the data; (2) use model-derived observations of unexplained phenomena of the system to inform further questions; (3) obtain data on potential explanatory variables; (4) construct further models based on the availability and quality of the explanatory variables, using an arsenal of statistical tools to identify, compare and parsimoniously choose the most informative refined models; (5) return, if necessary, to step (2).

Within this framework in mind, much of my work has been constrained to steps 1 and 2. For example, the heterogeneous migration model (chapter 6) reveals that different species of salmon reveal different extents of "heterogeneity domination", and that the response of this parameter tracks to flow regimes in different ways for different species. Furthermore, both travel times and survivals vary significantly within different reaches of a single reservoir (appendix C). These observations are essentially empirical in nature, though the require the filter of a movement model. They also immediately suggest several further directions of research: direct investigations of variability of individual movements, studies on predator distributions and movements within different reaches of a reservoir, statistical models that relate the value of various movement parameters to other measurable variables such as temperature and turbidity. In the Columbia River and other watersheds, many of these kinds of data are being collected. The mathematical movement model provides more than anything a structure and some guidance for integrating as much of the measurable variables as possible. Because there is a specific management goal in the river, namely the maximization of survival, the measure of success of a model is straightforwardly quantified

by its ability to fit survival estimates. Within the iterative modeling strategy, the study of salmon survival in the Columbia River is one that benefits from many, many iterations, as long as fresh kinds of explanatory data are being collected.

The implementation of the behavioral change-point analysis to fur seal data (chapter 5), revealed first and foremost that movement behaviors are extremely variable and show rather sudden and discrete shifts. Two consistent patterns emerge: the strongly autocorrelated signal during travel away and toward the home rookery and a increased activity between sunset and sunrise. Beyond these generalizations, it is difficult to know from the movement data alone what the fur seal is responding to. However, since the fur seal in the study was also tagged with stomach temperature and dive depth recorders, an analysis of the movement patterns with respect to foraging success will give extraordinary good information, and it is difficult to predict what kinds of further questions will emerge from even a cursory comparison. Measurable environmental covariates include depth and, perhaps, remotely sensed measures of sea surface temperature and primary productivity. Again, it is difficult to hypothesize what patterns may or may not emerge, but there exist several methodological options for making rigorous model comparisons. In the case of the analysis of individual movement tracks, experience with the complexity and variability of the data suggests to me that the most information can be obtained with the fewest preconceived hypotheses. Again, integrating a model of the foraging behavior of the fur seal to parallel information about its bioenergetic costs and diet can provide a very complete picture of foraging strategies. Furthermore, such an integrated model might provide some indirect infromation about the distribution and movement of prey objects, something that is very difficult to measure directly.

As a rule, human beings have terrible intuition about probabilities and statistics, but their ability to recognize and find meaning in patterns is remarkable.[4] It is

---

[4]This observation is paraphrased from the linguist Marc Rosenfelder, who addresses the many non-professionals that provide examples of remarkable resemblances between unrelated languages

the former unfortunate condition that requires sometimes intricate applications of probability theory and fancy mathematics when confronting data. But if the tools of statistical analysis can be used to filter away all but the most robust and significant patterns, than this shortcoming is more than redeemed by the latter gift. I believe that a great insights about the natural world can be obtained from a judicious awareness and manipulation of these very fundamental human qualities.

7.2 Movement ecology and conservation

By virtue of its incremental nature and excellent documentation, scientific process is inherently ratchet-like; forward and outward are essentially the only directions it knows. The related forward march of technological advances has contributed mightily to acquiring data on animal movements. The kinds of detailed information on locations and movements of organisms that are standard now were unimaginable twenty years ago. They have been made possible by the proliferation and improvement of satellites, the development of transistor circuitry, the epoxies for attaching the tags, the sedatives, the transmission and reception platforms, the exponential growth in the ability to access, process, organize, simulate the data. The mathematical language of interpreting movement data has moved forward in its own idiosyncratic pace, but ultimately it, too, is driven by the availability of the data and computational power. And so the science of animal movements has made significant steps forward, shedding light on questions as basic as "where does the organism go in winter?" to such sophisticated ones as "how does natural selection and environmental variability contribute to the evolution of the behavioral algorithms with which the organism optimizes it's exploitation of a heterogeneous environment?"

(e.g. Basque and Ainu, Welsh and Mandan, Hebrew and every other language) as evidence of deep connections between distant people. He presents a simple probabilistic model of phonemic combinations to show that, for example, the resemblance between the Japanese *gaijin* and the Hebrew *goyim* is easily explained by random chance. See `http://www.zompist.com/chance.htm` for details.

As irrevocably as science moves forward, global biodiversity appears to be moving in the opposite direction. The rate of extinctions now is generally agreed to be higher than at any point since the evolutionary arrival of humans, and virtually all of the pressures exacted on species are directly traceable to anthropogenic causes. The usefulness of animal movement studies to conservation is limited to the strength of the results and the application of the conclusions. While it rarely hurts to have the information obtained from movement studies, ultimately it is just information. No amount of hydroacoustic tagging and fractal analysis is going to help the seven or so Chinese river dolphin (*Lipotes vexillifer*) remaining in the Yangtze.[5]

Arguments can be made that some conservation efforts devote a disproportionate amount of resources to technological innovation and information gathering. The extraordinary amount of money devoted to Steller sea lion research ($120 million dollars since 2001) was the result of a successful effort in the U.S. Senate to suppress an attempt by a federal judge to temporarily shut down some areas in Alaska to fishing (Dalton, 2005). This windfall was a welcome boon to sea lion researchers, especially in light of the severity and remoteness of the sea lions' habitat and the high expense of doing research, and a flurry of projects were initiated. In a relatively short amount of time, a great amount of money was spent on obtaining information on the movement behavior, dispersal, reproductive physiology and behavior, diet, genetics, pathogens, pollutants and metabolism of Steller sea lions; but the science was no closer to obtaining a consensus on the effects of fisheries on sea lion populations. A subsequent panel of researchers did, however, make a concrete recommendation that the best way to test the hypothesis of fisheries interactions was to temporarily shut down fishing in some selected areas and follow up with intensified monitoring. Even this straightforward recommendation (guided, incidentally, in part by movement studies) was politically impossible to implement. In the study of Steller sea lions, politics and

---

[5]Two months after I first wrote that sentence in October 2006, the Yangtze river dolphin was declared "functionally extinct".

economics readily trumps the science when it comes to implementing conservation strategies.

While a single number is difficult to acquire (or comprehend), the resources invested in studying salmon in the Columbia river easily dwarf the Steller sea lion research budget, and the explosion of high-tech installations and data-acquisition tools and bypass systems and biological stations and propane-fueled avian predation dissuasion cannons has been commensurate. But ultimately the uncertainty behind the causes of salmon mortality in the river itself remains large, and the extremely important ocean phase remains even less well understood. These uncertainties means that whatever management recommendations are made can, are, and will be assailed in the courts. Certainly, the scientific body of knowledge becomes richer for the data obtained. But whatever the details of the court-mediated management strategy settled upon by the constellation of vested interests on the Columbia River at any given moment, little can or will be done about the bottom line: that the salmon themselves are still out- and inmigrating in a heavily impounded, predator-rich cascade of relatively warm and slow flowing reservoirs; that their populations are still going to rely heavily on hatchery supplementation; and that their survival is going to be largely dependent on the whim of highly unpredictable large-scale climatic conditions in the river and the ocean. In this context, the application of the few crude levers to assist outmigrating juveniles, such as augmenting water velocities or manipulating the ratio of powerhouse to spillway flow, can seem a bit like applying a bandage to a severed head.

As both of these cases and countless others demonstrate, conservation as an activity is driven first by political will, with all the associated entanglements with economic interests and cultural values, and only secondarily by science. Indeed, the science of movement often benefits more from a conservation agenda then the objects of conservation benefit from the science of movement.

Despite these caveats, there is no question that science is essential. All policy

must be predicated on knowledge, however inefficient the mechanics of that transition. Certainly, the most fundamental information to obtain for the protection of a rare species is the identification of habitat, and the most direct impact of movement studies has been the determination of critical habitats. Should the two Koreas reunite, it certainly behooves the both of them to be aware that their border is essential for the iconic red-crowned crane, a symbol (perhaps auspiciously) for good luck and marital bliss.

On a deeper if more nebulous level, the exploration of movement and behavior has consequences for a more complete understanding of the interactions between agents in ecological systems. In the scientific community, conservation is seen less and less in terms of isolated crises (like *Salmon!* or *Spotted Owl!* or *Panda Bear!*) and more in terms of a complex system of interactions. Note for example the increasing emphasis on Marine Protected Areas and so-called "Integrated Conservation" approaches to wildlife management. It seems that there is a tendency for public perception to follow suit. This phenomenon is surely encouraged by an increasing popular awareness of global warming, which many people appreciate as being a diffuse and highly uncertain threat, the effects of which shall be and already are being felt not by a single victim (like *Manatee!* or *Albatross!* or *Atlantic Cod!*) but by integrated ecological, human and economic systems.

Some of the most interesting work that has been done in the field of movement is that which is able to explain how organisms are successful at interacting with their complex, temporally and spatially heterogeneous environments; work that attempts to relate how the classical ecological variables like growth rates, carrying capacities and predation rates can be explained in terms of what they actually are: the product of movements and encounters and behaviors and the environment. While appearing 'fundamental' in nature, the success of such an endeavor and a deeper understanding of the mechanisms of ecological interactions will serve to inform not just the practice but the philosophy of conservation.

BIBLIOGRAPHY

Aarts, G., M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos, 2008. Estimating space-use and habitat preference from wildlife telemetry data. *Ecography* **31**:140–160.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**:716–723.

Alt, W., 1988. Modelling of Motility in Biological Systems. In J. McKenna and R. Temam, editors, *ICIAM '87: Proceedings of the First International Conference on Industrial and Applied Mathematics*, pages 15–30, Philadelphia. SIAM.

Alt, W., 1990. Correlation analysis of two-dimensional locomotion paths. In W. Alt and G. Hoffmann, editors, *Biological Motion: Proceedings of a workshop held in Königswinter Germany*, pages 254–268, Berlin. Springer-Verlag.

Anderson, J., 2002. An agent-based event driven foraging model. *Natural Resource Modeling* **15**:55–82.

Anderson, J., J. Hayes, P. Shaw, and R. Zabel, 1996. Columbia River salmon passage model (CRiSP.1.5). Technical report, University of Washington, School of Fisheries, WH-10.

Anderson, J. J., E. Gurarie, and R. W. Zabel, 2005. Mean free-path length theory of predator-prey interactions: application to juvenile salmon migration. *Ecological Modelling* **186**:196–211.

Andrews, R. D., 1998. Remotely releasable instruments for monitoring the foraging behaviour of pinnipeds. *Marine Ecology Progress Series* **175**:289–294.

Atkinson, R. P. D., C. J. Rhodes, D. W. Macdonald, and R. M. Anderson, 2002. Scale-free dynamics in the movement patterns of jackals. *Oikos* **98**:134–140.

Bartumeus, F., J. Catalan, U. L. Fulco, M. L. Lyra, and G. M. Viswanathan, 2002. Optimizing the encounter rate in biological interactions: Lévy verus Brownian strategies. *Physical Review Letters* **88**:article 097901–1.

Bartumeus, F., J. Catalan, G. M. Viswanathan, E. P. Raposod, and M. G. E. da Luze, 2008. The influence of turning angles on the success of non-oriented animal searches. *Journal of Theoretical Biology* **252**:43–55.

Bartumeus, F., M. G. E. Da Luz, G. M. Viswanathan, and J. Catalan, 2005. Animal search strategies: A quantitative random-walk analysis. *Ecology* **86**:3078–3087.

Bascompte, J. and C. Vilà, 1997. Fractals and search paths in mammals. *Landscape Ecology* **12**:213–221.

Bearon, R. N., D. Grünbaum, and R. A. Cattolico, 2004. Relating cell-level swimming behaviors to vertical population distributions in *heterosigma akashiwo* (raphidophyceae), a harmful alga. *Limnology and Oceanography* **49**:607–613.

Bearon, R. N., D. Grünbaum, and R. A. Cattolico, 2006. Effects of salinity structure on swimming behavior and harmful algal bloom formation in *heterosigma akashiwo*, a toxic raphidophyte. *Marine Ecology Progress Series* **306**:153–163.

Benhamou, S., 2004. On the expected net displacement of animals random movements. *Ecological Modelling* **171**:207–208.

Benhamou, S., 2007. How many animals really do the lvy walk? *Ecology* **88**:1962–1969.

Bergman, C. M., J. A. Schaefer, and L. S. N., 2000. Caribou movement as a correlated random walk. *Oecologia* **123**:364–374.

Berthold, P., W. Van den Bossche, M. Kaatz, and U. Querner, 2006. Conservation measures based on migration research in white stork *Ciconia ciconia, Ciconia boyciana*. *Acta Zoologica Sinica* **52**:211–214.

Block, B. A., H.Dewar, C. Farwell, and E. D. Prince, 1998. A new satellite technology for tracking the movements of atlantic bluefin tuna. *Proceedings of the National Academy of Science* **95**:9384–9389.

Blouin, F., J.-F. Giroux, J. Ferron, G. Gauthier, and G. J. Doucet, 1999. The use of satellite telemetry to track greater snow geese. *Journal of Field Ornithology* **70**:187–199.

Bovet, P. and S. Benhamou, 1988. Spatial analysis of animal's movements using a correlated random walk model. *Journal of Theoretical Biology* **131**:419–433.

Boyer, D., G. Ramos-Fernández, O. Miramontes, J. L. Mateos, G. Cocho, H. Larralde, H. Ramos, and F. Rojas, 2006. Scale-free foraging by primates emerges from their interaction with a complex environment. *Proceedings of the Royal Society Biological Sciences Series* **273**:1743–1750.

Brockmann, D., L. Hufnage, and T. Geisel, 2006. The scaling laws of human travel. *Nature* **439**:462–465.

Bryant, E., 2007. 2D Location Accuracy Statistics for Fastloc®Cores Running Firmware Versions 2.2 & 2.3, Technical Report TR01. Technical report, Wildtrack Telemetry Systems Ltd.

Buettner, E. W. and A. F. Brimmer, 1998. Smolt monitoring at the head of Lower Granite Reservoir and Lower Granite Dam. Annual Report 1996 to Bonneville Power Administration, Idaho Department of Fish and Game, Portland, Oregon.

Byers, J. A., 2001. Correlated random walk equations of animal dispersal resolved by simulation. *Ecology* **82**:1680–1990.

Cada, G. F., M. D. Deacon, S. V. Mitz, and M. S. Bevelhimer, 1997. Effects of water velocity on the survival of downstream-migrating juvenile salmon and steelhead: a review with emphasis on the Columbia River Basin. *Reviews in Fisheries Science* **5**:131–183.

Cant, E., A. Smith, D. Reynolds, and J. Osborne, 2005. Tracking butterfly flight paths across the landscape with harmonic radar. *Proc Biol Sci.* **272**:785790.

Casella, G. and R. L. Berger, 1990. Statistical Inference: Second Edition. Wadsworth & Brooks, Pacific Grove, CA.

Chernetsov, N., P. Berthold, and U. Querner, 2004. Migratory orientation of first-year white storks (*Ciconia ciconia*): inherited information and social interactions. *Journal of Experimental Biology* **207**:937–943.

Chhikara, R. S. and J. L. Folks, 1989. The Inverse Gaussian Distribution: Theory, Methods and Applications. Marcel Dekker Inc., New York.

Chib, S., 1998. Estimation and comparison of multiple change-point models. *Journal of Econometrics* **86**:221–241.

Clark, J. S., 1998. Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord. *The American Naturalist* **152**:204–224.

Clausius, R., 1859. On the mean length of the paths described by the separate molecules of gaseous bodies on the occurrence of molecular motion: together with some other remarks on the mechanical theory of heat. *Philosophical Magazine* **17**:81–91.

Collis, K., D. D. Roby, D. P. Craig, S. Adamany, J. Y. Adkins, and D. E. Lyons, 2002. Colony size and diet composition of piscivorous waterbirds on the lower Columbia River: Implications for losses of juvenile salmonids to avian predation. *Transactions Of The American Fisheries Society* **131**:537–550.

Collis, K., D. D. Roby, D. P. Craig, B. A. Ryan, and R. D. Ledgerwood, 2001. Colonial waterbird predation on juvenile salmonids tagged with passive integrated transponders in the Columbia river estuary: Vulnerability of different salmonid species, stocks, and rearing types. *Transactions Of The American Fisheries Society* **130**:385–396.

Coombs, M. F. and M. A. Rodríguez, 2007. A field test of simple dispersal models as predictors of movement in a cohort of lake-dwelling brook charr. *Journal of Animal Ecology* **76**:4557.

Coughlin, D. J., J. R. Strickler, and B. Sanderson, 1992. Swimming and search behaviour in clownfish *(Amphiprion perideraion)* larvae. *Animal Behavior* **44**:427–440.

Dalton, R., 2005. Conservation biology: Is this any way to save a species? *Nature* **436**:14–16.

Dauble, D. D., T. L. Page, and R. W. J. Hanf, 1989. Spatial distribution of juvenile salmonids in the Hanford Reach, Columbia River. *Fishery Bulletin* **87**:775–790.

de Iongh, H. H., B. J. Wenno, and E. Meelis, 1995. Seagrass distribution and seasonal biomass changes in realtion to dugong grazing in the Moluccas, East Indonesia. *Aquatic Botany* **50**:1–19.

Dicke, M. and P. A. Burroughs, 1988. Using fractal dimensions for characterizing tortuosity of animal trails. *Physiological Entomology* **13**:393–398.

Dobzhansky, T. and S. Wright, 1943. Genetics of natural populations. X: Dispersion rates in *Drosophila pseudoobsura*. *Genetics* **28**:304–340.

Doerr, V. A. J. and E. D. Doerr, 2004. Fractal analysis can explain individual variation in dispersal search paths. *Ecology* **85**:1428–1438.

Dunn, G. A. and A. Brown, 1987. A unified approach to analysing cell motility. *Journal of Cell Science - Supplement* **8**:81–102.

Edwards, A. M., R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, M. G. Buldyrev, S. V.and da Luz, E. P. Raposo, H. E. Stanley, and G. M. Viswanathan, 2007. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* **449**:1044–1048.

Evans, G. T., 1989. The encounter speed of moving predator and prey. *Journal of Plankton Research* **11**:415–417.

Fagan, W. F., 1997. Introducing a "boundary-flux" approach to quantifying insect diffusion rates. *Ecology* **78: 2**:579–587.

180

Fedak, M., P. Lovell, B. McConnell, and C. Hunter, 2002. Overcoming the constraints of long range radio telemetry from animals: Getting more useful data from smaller packages. *Integrative and Comparative Biology* **42**:3–10.

Firle, S., R. Bommarco, B. Ekbom, and M. Natiello, 1998. The influence of movement and resting behavior on the range of three carabid beetles. *Ecology* **79**:2113–2122.

Fisher, N. I., 1993. Statistical Analysis of Circular Data. Cambridge University Press, Cambridge, New York.

Fisher, N. I. and A. Lee, 1994. Time series analysis of circular data. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**:327–339.

Folks, J. L. and R. S. Chhikara, 1978. The inverse Gaussian distribution and its statistical application - a review. *Journal of the Royal Statistics Society. Series B (Methodological)* **40**:263–289.

Forester, J. D., A. R. Ives, M. G. T. andD. P. Anderson, D. Fortin, H. L. Beyer, D. W. Smith, and M. S. Boyce, 2007. State-space models link elk movement patterns to landscape characteristics in yellowstone national park. *Ecological Monographs* **77**:285–299.

Fortin, D., H. L. Beyer, M. S. Boyce, D. W. Smith, T. Duchesne, and J. S. Mao, 2005a. Wolves influence elk movements: behavior shapes a trophic cascade in Yellostone National Park. *Ecology* **86**:1320–1330.

Fortin, D., J. M. Morales, and M. S. Boyce, 2005b. Elk winter foraging at fine scale in Yellowstone National Park. *Oecologia* **145**:334–342.

Fraser, D. F., J. F. Gilliam, M. J. Daley, and A. N. L. andG. T. Skalski, 2001. Explaining leptokurtic movement distributions: intrapopulation variation in boldness and exploration. *The American Naturalist* **158**:124–135.

Fritz, H., S. Said, and H. Weimerkirsch, 2003. Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *Proceedings of the Royal Society of London B* **270**:1143–1148.

Geisel, T., 1960. One Fish, Two Fish, Red Fish, Blue Fish. Random House Children's Books.

Gentry, R., 1998. Behavior and Ecology of the Northern Fur Seal. Princeton University Press.

Gerritsen, J. and J. Strickler, 1977. Encounter probabilities and community structure in zooplankton: a mathematical model. *Journal of the Fisheries Reserve Board of Canada* **34**:73–82.

Giorgi, A. E., T. W. Hillman, J. R. Stevendon, S. G. Hays, and C. M. Pevin, 1997. Factors that influence the downstream migration rates of juvenile salmon and steelhead through the hydroelectric system in the mid-Columbia river basin. *North American Journal of Fisheries Management* **17**:268–282.

Goebel, M., J. Bengtson, R. DeLong, R. Gentry, and T. Loughlin, 1991. Diving patterns and foraging locations of female northern fur seals. *Fisheries Bulletin* **89**:171– 179.

Goodwin, R., J. Nestler, J. Anderson, L. Weber, and D. Loucks, 2006. Forecasting 3-D fish movement behavior using a Eulerian-Lagrangian-agent method (ELAM). *Ecological Modelling* **192**:197–223.

Gove, J. H., 2003. Moment and maximum likelihood estimators for weibull distributions under length- and area-biased sampling. *Environmental and Ecological Statistics* **10**:455–467.

Grünbaum, D., 2000. Advection-diffusion equations for internal state-mediated random walks. *SIAM (Society for Industrial and Applied Mathematics) Journal of Applied Mathematics* **61**:43–73.

Grünbaum, D. and R. Veit, 2003. Black-browed albatrosses foraging on Antarctic krill: density-dependence through local enhancement. *Ecology* **84**:3265–3275.

Grünbaum, D., S. Viscido, and J. K. Parrish, 2005. Cooperative Control, volume Volume 309 of *Lecture Notes in Control and Information Sciences*, chapter Extracting Interactive Control Algorithms from Group Dynamics of Schooling Fish, pages 103–117. Springer Berlin / Heidelberg.

Gurarie, E., J. J. Anderson, and R. W. Zabel, 2008. Incorporating population heterogeneity into analysis of animal dispersal and movement. *in review* .

Haydon, D. T., J. M. Morales, A. Yott, D. A. Jenkins, R. Rosatte, and J. M. Fryxell, 2008. Socially informed random walks: incorporating group dynamics into models of population spread and growth. *Proceedings of the Royal Society B-Biological Sciences* **275**:1101–1109.

Holley, D., 2006. Movement patterns and habitat usage of shark bay dugongs. Ph.D. thesis, Edith Cowan University.

Hopkins, W. G. and D. Hewson, 2001. Variability of competitive performance of distance runners. *Medicine & Science in Sports & Exercise* **33**:1588–1592.

Hughes, B. D., 1995. Random Walks and Random Environments. Clarendon Press, Oxford.

Hutchinson, J. M. C. and P. M. Waser, 2007. Use, misuse and extensions of the "ideal gas" models of animal movement. *Biological Reviews* **82**:335–359.

Johnson, D. S., J. M. London, M.-A. Lea, and J. W. Durban, 2008. Continuous-time correlated random walk models for animal movement data. *Ecology* **89**:1208–1215.

Johnson, D. S., D. L. Thomas, J. M. V. Hoef, and A. Christ, 2007. A general framework for the analysis of animal resource selection from telemetry data. *Biometrics* page Online Early Articles.

Jones, M. C. and A. Pewsey, 2005. A family of symmetric distributions on the circle. *Journal of the American Statistical Association* **100**:1422–1428.

Jonsen, I. D., R. A. Myers, and J. M. Flemming, 2003. Meta-analysis of animal movement using state-space models. *Wildlife Research* **21**:149–161.

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

Jouventin, P. and H. Weimerskirch, 1990. Satellite tracking of Wandering albatrosses. *Nature* **343**:746 – 748.

Kanai, Y., M. Ueta, N. Germogenov, M. Nagendran, N. Mita, and H. Higuchi, 2002. Migration routes and important resting areas of Siberian cranes (*Grus leucogeranus*) between northeastern Siberia and China as revealed by satellite tracking. *Biological Conservation* **106**:339–346.

Kareiva, P. M. and N. Shigesada, 1983. Analyzing insect movement as a correlated random walk. *Oecologia* **56**:234–238.

Keating, K., W. G. Brewster, and C. H. Key, 1991. Satellite telemetry: Performance of animal-tracking systems. *Journal of Wildlife Management* **55**:160–171.

Klafter, J., B. S. White, and M. Levandowsky, 1989. Microzooplankton feeding behavior and the Lévy walk. In G. Hoffmann and W. Alt, editors, *Biological motion*, page 281296. Springer-Verlag, Heidelberg, Germany.

Kot, M., M. A. Lewis, and P. v. d. Driessche, 1996. Dispersal data and the spread of invading organisms. *Ecology* **77**:2027–2042.

Laidre, K. L., M. P. Heide-Jorgensen, M. L. Logsdon, R. C. Hobbs, R. Dietz, and G. R. VanBlaricom, 2004. Fractal analysis of narwhal space use patterns. *Zoology* **107**:3–11.

Levandowsky, M., J. Klafter, and B. S. White, 1988. Feeding and swimming behavior in grazing zooplankton. *Journal of Protozoology* **35**:243246.

Levandowsky, M., B. S. White, and F. L. Schuster, 1997. Random movements of soil amebas. *Acta Protozoologica* **36**:237–248.

Lotka, A., 1924. Elements of Physical Biology. Williams and Wilkins, Baltimore.

Loughlin, T. R., D. J. Rugh, and C. H. Fiscus, 1984. Northern sea lion distribution and abundance: 1956-80. *The Journal of Wildlife Management,* **48**:729–740.

MacKenzie, B. and T. Kiorboe, 1995. Encounter rates and swimming behavior of pause-travel and cruise larval fish predators in calm and turbulent laboratory environments. *Limnology and Oceanography* **40**:1278–1289.

Mandelbrot, B., 1967. How long Is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* **156**:636–638.

Mårell, A., J. P. Ball, and A. Hofgaard, 2002. Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. *Canadian Journal of Zoology* **80**:854–865.

Marrasé, C., J. H. Costello, T. Granata, and J. R. Strickler, 1990. Grazing in a turbulent environment: Energy dissipation, encounter rates, and efficacy of feeding currents in *centropages hamatus*. *Proceedings of the National Academy of Sciences of the United States of America* **87**:1653–1657.

Marsh, D. M., J. R. Harmon, N. W. Paasch, K. L. Thomas, K. W. McIntyre, B. P. Sandford, and G. M. Mathews, 1997. Research related to transportation of juvenile salmonids on the Columbia and Snake Rivers. Report to U. S. Army Corps of Engineers, Fish Ecology Division, Northwest Fisheries Science Center, Seattle, Washington, USA.

Marsh, H., R. I. T. Prince, W. K. Saafeld, and R. Shepherd, 2005. The distribution and abundance of the dugong in Shark Bay, Western Australia . *Wildlife Research* **21**:149–161.

Marsh, L. and R. Jones, 1988. The form and consequence of random walk movement models. *Journal of Theoretical Biology* **133**:113–131.

Martell, M. S., C. J. Henny, P. E. Nye, and M. J. Solensky, 2001. Fall migration routes, timing, and wintering sites of North American ospreys as determined by satellite telemetry. *The Condor* **103**:715–724.

Massey Jr., F. J., 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* **46**:68–78.

184

Maxwell, J., 1860. Illustrations of the dynamical theory of gases: Part 1. on the motions and collisions of perfectly elastic spheres. *Philosophical Magazine* **19**:19–32.

Millspaugh, J. and J. M. Marzluff, editors, 2001. Radio Tracking and Animal Populations. Academic Press, San Diego, CA.

Moorcroft, P. R. and M. A. Lewis, 2006. Mechanistic Home Range Analysis. Princeton University Press, Princeton, NJ.

Moore, D. S. and G. P. McCabe, 2006. Introduction to the Practice of Statistics. W.H. Freeman and Company, New York, NY, fifth edition.

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

Muir, W. D., S. G. Smith, J. G. Williams, E. E. Hockersmith, and J. R. Skalski, 2001. Survival estimates for migrant yearling chinook salmon and steelhead tagged with passive integrated transponders in the lower Snake and lower Columbia Rivers, 1993-1998. *North American Journal of Fisheries Management* **21**:269–282.

Nams, V. O., 2005. Using animal movement paths to measure response to spatial scale. *Oecologia* **143**:179–188.

Nams, V. O. and M. Bourgeois, 2004. Fractal analysis measures habitat use at different spatial scales: an example with American marten. *Canadian Journal of Zoology* **82**:17381747.

Naughton, G. P. and D. H. Bennett, 2003. Diet composition of northern pikeminnow in the Lower Granite Reservoir system. *Northwest Science* **77**:19–24.

Okubo, A. and D. Grünbaum, 2001. Mathematical Treatment of Biological Diffusion. In A. Okubo and S. Levin, editors, *Diffusion and Ecological Problems*, pages 127–169. Springer, New York.

Okubo, A. and S. Levin, 2001. Diffusion and Ecological Problems: Modern Perspectives. Springer Verlag, New York.

Ovaskainen, O., 2004. Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. *Ecology* **85**:242–257.

Ovaskainen, O., H. Rekola, E. Meyke, and E. Arjas, 2008. Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data. *Ecology* **89**:542–554.

Pacific States Marine Fisheries Commission, 1996. PIT tag information system (PTAGIS). Gladstone, Oregon. Online database (Available through Internet: www.psmfc.org/ptagis).

Parrish, J. K. and L. Edelstein-Keshet, 1999. Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science* **284**:99–101.

Parrish, J. K., S. V. Viscido, and D. Grünbaum, 2002. Self-organized fish schools: An examination of emergent properties. *Biological Bulletin* **202**:296–305.

Patlak, C. S., 1953a. A mathematical contribution to the study of orientation of organisms. *Bulletin of Mathematical Biophysics* **15**:431–476.

Patlak, C. S., 1953b. Random walk with persistence and external bias. *Bulletin of Mathematical Biophysics* **15**:311–338.

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

Pearson, K., 1905. The problem of the random walk. *Nature* **72**:294.

Petersen, J. H., 2001. Density, aggregation, and body size of northern pikeminnow preying on juvenile salmonids in a large river. *Journal of Fish Biology* **58**:1137–1148.

Petersen, J. H. and D. L. Ward, 1999. Development and corroboration of a bioenergetics model for northern pikeminnow feeding on juvenile salmonids in the Columbia River. *Transactions of the American Fisheries Society* **128**:784–801.

Poe, T. P., H. C. Hansel, S. Vigg, D. E. Palmer, and L. A. Predergast, 1991. Feeding of predaceous fishes on out-migrating juvenile salmonids in John Day Reservoir, Columbia River. *Transactions of the American Fisheries Society* **120**:405–420.

Prentice, E. F., T. A. Flagg, C. S. McCutcheon, and D. F. Brastow, 1990. Pit-tag monitoring systems for hydroelectric dams and fish hatcheries. In N. P. e. al., editor, *Fish-marking techniques*, pages 323–334. American Fisheries Society Symposium 7, Bethesda, Maryland.

Price, M. V., P. A. Kelly, and R. L. Goldingay, 1994. Distances moved by Stephens' kangaroo rat (*Dipodomys stephensi Merriam*) and implications for conservation. *Journal of Mammalogy* **75**:929–939.

Quinn, T., 2005. The Behavior and Ecology of Pacific Salmon and Trout. American Fisheries Society, Bethesda, MD.

Ramos-Fernández, G., J. L. Mateos, O. Miramontes, H. Larralde, G. Cocho, and B. Ayala-Orozco, 2004. Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behavioral Ecology and Sociobiology* **55**:223–230.

Ransom, B. H., T. W. Steig, M. A. Timko, and P. A. Nealson, 2007. Basin-wide monitoring of acoustically tagged salmon smolts at hydropower dams in the mid-Columbia River basin, USA. In *Proceedings of Hydro 2007, Granada, Spain,*.

Raum-Suryan, K., M. Rehberg, G. Pendleton, K. Pitcher, and T. Gelatt, 2004. Development of dispersal, movement patterns, and haul-out use by pup and juvenile Steller sea lions (*Eumotopias jubatus*) in Alaska. *Marine Mammal Science* **20**:823850.

Real, L. A. and J. H. Brown, editors, 1991. Foundations of Ecology: Classic Papers with Commentaries. The University of Chicago Press, Chicago and London.

Rieman, B. E., R. C. Beamesderfer, and S. T. P. P. Vigg, 1991. Estimated loss of juvenile salmonids to predation by northern squawfish, walleyes and smallmouth bass in John Day Reservoir, Columbia River. *Transactions of the American Fisheries Society* **120**:448–458.

Roby, D., K. Collis, J. Adkins, Y. Suzuki, B. Courtot, D. Lyons, A. M. Myers, and C. Couch, 2005. Colony size, nesting success, and predation on salmon smolts by Caspian terns and double-crested cormorants in the lower Columbia River. *Pacific Seabirds* **32**:52–53.

Root, R. B. and P. M. Kareiva, 1984. The search for resources by cabbage butterflies (*Pieris rapae*): Ecological consequences and adaptive significance of Markovian movements in a patchy environment. *Ecology* **65**:147–165.

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

Sandell, M. J., J. Argell, S. Erlinge, and J. Nelson, 1991. Adult philopatry and disperal in the field vole *Microtus agrestis*. *Oecologia* **86**:284–292.

Seuront, L., F. G. Schmitt, M. C. Brewer, J. R. Strickler, and S. Souissi, 2004. From random walk to multifractal random walk in zooplankton swimming behavior. *Zoological Studies* **43**:498–510.

Sheppard, J., A. Preen, H. Marsh, I. Lawler, S. Whiting, and R. Jones, 2006. Movement heterogeneity of dugongs, *Dugong dugon* (Müller), over large spatial scales . *Journal of Experimental Marine Biology and Ecology* **334**:64–83.

Shigesada, N., 1980. Spatial distribution of dispersing animals. *Journal of Mathematical Biology* **9**:85–96.

Shimazaki, H. and H. Tamura, M.and Higuchi, 2004. Migration routes and important stopover sites of endangered oriental white storks (*Ciconia boyciana*) as revealed by satellite tracking. *Mem Natl Inst Polar Res Spec Issue* **58**:162–178.

Skalski, G. T. and J. F. Gilliam, 2000. Modeling diffusive spread in a heteregeneous population: a movement study with stream fish. *Ecology* **81**:1685–1700.

Skalski, G. T. and J. F. Gilliam, 2003. A diffusion-based theory of organism dispersal in heterogeneous populations. *The American Naturalist* **161**:441–458.

Skellam, J. G., 1951. Random dispersal in theoretical populations. *Biometrika* **38**:196–218.

Smith, S. G., W. D. Muir, S. Achord, E. E. Hockersmith, B. P. Sandford, J. G. Williams, and J. R. Skalski, 2000. Survival estimates for the passage of juvenile salmonids through Snake and Columbia River dams and reservoirs, 1998. Report to Bonneville Power Administration, Contract DE-AI79-93BP10891,s, Fish Ecology Division, Northwest Fisheries Science Center, Seattle, Washington, USA.

Speckman, S., D. Burn, M. Udevitz, and R. Benter, 2006. Estimating the size of the Pacific walrus population: the 2006 survey in Alaska. In V. Belkovich, editor, *Marine Mammals of the Holarctic*, pages 506–507, St. Petersburg, Russia. Marine Mammal Council.

Steig, T., P. Nealson, K. Kumagai, J. Horchik, C. Mott, and J. Sweet, 2006. Route specific passage of juvenile chinook, steelhead and sockeye salmon using acoustic tag methodologies at Rocky Reach and Rock Island dams in 2005, Draft Report prepared for Chelan County Public Utility District No. 1. Technical report, HTI.

Turchin, P., 1996. Fractal analyses of animal movement: a critique. *Ecology* **77**:2086–2090.

Turchin, P., 1998. Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants. Sinauer, Sunderland, Mass.

Tweedie, M. C. K., 1957. Statistical properties of Inverse Gaussian distributions. *The Annals of Mathematical Statistics* **28**:362–377.

Ulbrich, J., 1930. Die Bisamratte. Dresden: Heinrich.

USNMFS, 1992. Threatened status for Snake River spring/summer chinook salmon, threatened status for Snake River fall chinook salmon. Final Rule. Technical Report 57:78, U. S. National Marine Fisheries Service.

USNMFS, 1998. Endangered and threatened species: threatened status for two ESU's of steelhead in Washington, Oregon and California. Technical report, U. S. National Marine Fisheries Service.

Uttieri, M., E. Zambianchi, J. R. Strickler, and M. G. Mazzocchi, 2005. Fractal characterization of three-dimensional zooplankton swimming trajectories. *Ecological Modelling* **185**:51–63.

Viscido, S. V., J. K. Parrish, and D. Grnbaum, 2005. The effect of population size and number of influential neighbors on the emergent properties of fish schools. *Ecological Modelling* **183**:347–363.

Viscido, S. V., J. K. Parrish, and D. Grünbaum, 2004. Individual behavior and emergent properties of fish schools: a comparison between observation and theory. *Marine Ecology Progress Series* **273**:239–249.

Viswanathan, G. M., V. Afanasev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley, 1996. Lévy flight search patterns of wandering albatross. *Nature* **381**:413–415.

Viswanathan, G. M., S. V. Buldyrev, S. Havlin, M. G. E. d. Luz, E. P. Raposo, and H. E. Stanley, 1999. Optimizing the success of random searches. *Nature* **401**:911–914.

Wald, A., 1947. Sequential Analysis. J. Wiley & Sons.

Weimerskirch, H., T. Guionnet, J. Martin, S. A. Shaer, and D. P. Costa, 2000. Fast and fuel efficient? Optimal use of wind by flying albatrosses. *Proceedings of the Royal Society B.* **267**:1869–1874.

Weiss, G. H., 1994. Aspects and Applications of the Random Walk. North-Holland, Amsterdam, New York.

Wilk, M. B. and R. Gnanadesikan, 1968. Probability plotting methods for the analysis of data. *Biometrika* **55**:1–17.

Wilson, P., 1994. Spring FLUSH (Fish Leaving Under Several Hypotheses). Version 4.5. Technical report, Columbia Basin Fish and Wildlife Authority, Portland, OR.

With, K. A., 1994. Using fractal analysis to assess how species perceive landscape structure. *Landscape Ecology* **9**:25–36.

Yamamura, K., 2002. Dispersal distance of heterogeneous populations. *Population Ecology* **44**:93–101.

Zabel, R. and J. G. Williams, 2002. Selective mortality in chinook salmon: what is the role of human disturbance? *Ecological Applications* **12**:173183.

Zabel, R. W., 1994. Spatial and Temporal Models of Migrating Juvenile Salmon With Applications. Phd, University of Washington.

Zabel, R. W., 2002. Using "Travel Time" Data to Characterize the Behavior of Migrating Animals. *The American Naturalist* **159**:372–387.

Zabel, R. W. and J. J. Anderson, 1997. A model of the travel time of migrating juvenile salmon, with an application to Snake River spring chinook. *North American Journal of Fisheries Management* **118**:558–560.

Zhang, X., S. N. Johnson, J. W. Crawford, P. J. Gregory, and I. M. Young, 2007. A general random walk model for the leptokurtic distribution of organism movement: Theory and application. *Ecological Modelling* **200**:79–88.

Appendix A

R-CODE FOR ESTIMATING $\tau_{1/2}$ FOR WRAPPED CAUCHY MODEL WITH IRREGULAR INTERVALS.

A.1 Adjusted wrapped Cauchy distribution (AWCD, eq. 2.14)

```
dawrpcauchy <- function(theta,tau,halftau)
  # This function returns the distribution function for a continuous
  # wrapped-cauchy variable theta with half-time equal to "halftau" at
  # interval "tau".
  {
    rho<-(halftau/((sqrt(2)-1)*tau+halftau))^2
    return(dwrpcauchy(theta,0,rho))
  }
```

A.2 Generating numbers from AWCD

```
  rawrpcauchy <- function(n,tau,halftau)

  # This function provides random numbers generated from an adjusted
  # wrapped-cauchy variable with half-time equal to "halftau" at interval
  # "tau".
  {
  rho<-(halftau/((sqrt(2)-1)*tau+halftau))^2
  return(rwrpcauchy(n,0,rho))
  }
```

A.3 Obtaining MLE estimate for $\tau_{1/2}$ from data

```
awrpcauchy.mle <- function(thetas,gaps)
  # This function provides the mle estimate for "halftau" from data,
  # where "theta" is the vector of angles and "gaps" is the associated
  # gap intervals.
```

```
{
  LL <- function(halftau)
  {
    LL <- sum(log(dwrpcauchy.gap(thetas,gaps,halftau)))
    return(-LL)
  }
  o<-optimize(f=LL,interval=c(0,1000))
return(o$min)
}
```

Appendix B

R-CODE FOR IMPLEMENTING DUNN-BROWN-ALT MOVEMENT

B.1 Complex velocity auto-correlation function

```
CVAF <- function(V, p.lagmax=0.5){
# This function returns the CVAF of complex velocity vector "V".
# "p.lagmax" is the maximum lag as a proportion of the length of "V".
  lag <- 1:round(length(V)*p.lagmax)
  cvaf<-mean(Mod(V)^2)
  for(i in lag)
    cvaf[i+1] <- mean(V[-(1:i)] * Conj(V[-((length(V)-i+1):length(V))]))
  return(data.frame(lag, cvaf))}
```

B.2 Simulating a Dunn-Brown-Alt walk

```
DBAwalk <- function(muv,alpha,beta,N,scale){
# This function simulates a DBA walk of length "N" with the given
# parameter values.  Any of these parameters can be complex, but
# "beta" typically won't be.  The scaling parameter, typically an
# integer greater than 1, can be set as high as needed to guarantee
# a well-behaved solution.
  V <- muv
  for(i in 1:N){
    dV <- alpha*(muv-V[i])/scale +
          beta*complex(re=rnorm(1),im=rnorm(1))/sqrt(scale)
    V <- c(V,V[i]+dV)}
  return(V)}
```

B.3 Estimating DBA parameters from movement data

```
GetPnull <- function(V,re.cvaf=NA,l=0.5)
# This function returns the initial estimates of the parameters,
# i.e. p-tilde, from the velocity vector "V".
# "l" refers to the amount of the vector we use for analysis and is
# a tunable function that ultimately has little impact on the estimates.
{

  # Mean term
    Vbar.null <- mean(V)

  # use acf of real part to get "a2"
  # (related to the trough of the acf)
    myacf<-acf(re.cvaf,lag.max=600,plot=T)$acf
    t<-1:length(myacf)-1

    HalfPeriod <- t[myacf==min(myacf)]
    a2.null <- pi/HalfPeriod

   # An exponentially decaying cosine function ...
    theory <- function(p,t)
        return(exp(-p[1]*t) * cos(p[2]*t))

  # and its minimization...
    minme <- function(p)
       return(sum((theory(p,t)-myacf)^2))

  # are used to find an estimates of a1...
    p.null <- c(1,a2.null)
    p.hat <- optim(p.null,minme)$par
    a1.null <- p.hat[1]

  # which is used to get an initial estimate of b
    b1.null <- 2*a1.null*(mean(Mod(V)^2)-Mod(Vbar.null)^2)

  return(c(a1.null,a2.null,b1.null))
}


GetPhat <- function(V,l=0.5)
 # This function returns pinal estimates of the parameters,
```

```r
# (i.e. p-hat), from the velocity vector "V" using a fit
# to the Complex Velocity Autocorrelation Function
{
  mycvaf <- CVAF(V,l,plot=0)
  re.cvaf <- Re(mycvaf)
  t <- 1:length(re.cvaf)-1


  # obtain initial parameter estimates (from above)
  p.null <- GetPnull(V,re.cvaf)

  Vbar <- mean(V)

  theory <- function(p,tau=t)
  # Theoretical complex velocity correlation function
  {
    a1    <-  p[1]
    a2    <-  p[2]
    b.squared <- p[3]

    return(Mod(Vbar)^2 + (b.squared / (2 * a1)) * exp(-a1*tau) * cos(a2*tau))
  }

  MinimizeMe <- function(p)
    return(sum((theory(p,t)-re.cvaf)^2))

  p.hat <- optim(p.null, MinimizeMe)$par

  Vbar.hat <- Vbar
  a.hat <- complex(re=p.hat[1],im=p.hat[2])
  b.hat <- sqrt(p.hat[3])

  return(c(Vbar.hat,a.hat,b.hat))
}
```

Appendix C

ANALYSIS OF JUVENILE SALMON SURVIVAL IN MID-COLUMBIA RIVER RESERVOIRS

C.1 Introduction

The following is a brief analysis of travel times and survivals of salmon smolt in several reservoirs of the mid-Columbia River. The data was collected in 2005 by Chelan County Public Utility District 1 using hydroacoustic tags implanted in juvenile chinook (*O. tshawytscha*), sockeye (*O. nerka*) and steelhead (*O. mykiss*) salmon. Whereas the bulk of outmigrating salmon survival data comes from PIT tags which are detected only at dams, hydroacoustic detecting arrays were placed in several mid-reservoir locations, allowing for an analysis of migration timing and survival within a reservoir. Detection probabilities are quite high with hydroacoustic arrays, as they cover the entire width of the river. The data also provides longitudinal distribution of migration paths across the river and three dimensional tracks of the fish at the dams themselves, though these dimensions are not analyzed in this report.

The purpose of the analysis is to identify some of the inter-reservoir structure of the migration and survival patterns. The results are intended to motivate a more detailed and process oriented understanding and modeling of salmon migration, and can be at least conceptually applied to other reservoirs in the Columbia River. The ultimate goal is to guide management decisions that mitigate salmon survival.

Commonly, probability of survival is assumed to decay exponentially with time, a common model of time-dependent mortality. However, studies have shown that, in fact, distance traveled is possibly a stronger indicator of survival. This can be explained by the number of predators encountered depending more strongly on dis-

tance traveled downstream than time spent in a reservoir, a process referred to as the "gauntlet effect". Anderson et al. (2005) have proposed a simple model where the strength of the time and distance dependence of survival is determined by a parameter which reflects the net amount of randomness in the trajectories of both the smolt and their predators. Their model is referred to as the XT-model, reflecting dependence on space (x) and time (t). The XT-model, and its rigorous parameterization in chapter 4 is the conceptual framework that guides this analysis.

Figure C.1: Map of release and detection locations along the mid-Columbia river

C.2   Data

Hydroacoustic tags developed by HTI were implanted by Chelan County Public Utility District 1 in a total of 4476 juvenile chinook, sockeye and steelhead salmon during

spring outmigration between April 18 and June 9, 2005. The juveniles were released in groups of 24 individuals (chinook and steelhead) or 20 individuals (sockeye) from various locations on the Columbia River in Chelan county, WA: the tailrace of Wells dam (river mile 515.8), the juvenile salmon bypass system at Rocky Reach dam (r.m. 473.7), the tailrace of Rocky Reach dam, and the tailrace of Rock Island dam (r.m. 453.4). Detector arrays were located at Rocky Reach and Rock Island dams in a manner that allows for 3-d tracking of the salmon as it approaches either the bypass system, the turbines or the spillway at each dam. Arrays of detectors were also placed across the river a locations in the mid-reservoirs. These were: Beebe Bridge (r.m. 503.9) between Wells and Rocky Reach, Hydro Park station (r.m. 462.7) between Rocky Ridge and Rock Island and Crescent Bar (r.m. 442.52) and Sunland Estates (r.m. 431.08) downstream from Rock Island. The mid-reservoir arrays are designed to detect the passing fish, but do not provide 3-d tracking resolution. Each hydroacoustic tag emits a pulse at a frequency specific to each fish, thereby identifying fish singly, as well as providing the exact time of passage. Lengths and weights were also obtained for each individual. See Figure C.1 for detector locations, Table C.1 for details about the releases and Figures C.2 for passage tracks of all the fish.

C.3 Results

Table C.1: Table of hydroacoustically tagged salmon smolt releases in spring and summer 2005. R refers to the number of releases at each location. The smolt were released into the river typically every other day in the date interval indicated in the $Date$ column.

Species	Dates	Wells tailrace		R.R. bypass		R.R. tailrace		R.I. tailrace		N
		R	n	R	n	R	n	R	n	
Chinook	4/18-5/28	20	483	-	-	20	500	-	-	983
Steelhead	4/18-5/29	20	984	-	-	20	1000	20	501	2485
Sockeye	5/15-6/9	24	499	24	498	24	497	24	496	1990

Table C.2: Table summarizing mortality events and mean intensity of mortality. N refers to the number of smolt entering a reservoir and MI represents intensity of mortality presented as percent probability of mortality per river mile. Standard errors are measured using the variance from a standard binomial model $\widehat{\sigma} = \sqrt{p(1-p)/N}$ scaled to the length of the reach.

A: Chinook

Reach of river	Distance (miles)	N	Events	MI (%/mile)	SE
A: W to BB	11.6	483	8	0.143	0.050
B: BB to RR	30.12	475	23	0.161	0.033
C: RR to HP	16.73	952	48	0.301	0.042
D: HP to RI	3.55	904	7	0.218	0.082
E: RI to CB	10.88	897	37	0.379	0.061
F: CB to SE	11.44	860	35	0.356	0.059
TOTAL	84.32	984	158	M total:	0.161

B: Sockeye

Reach of river	Distance (miles)	N	Events	MI (%/mile)	SE
A: W to BB	11.6	499	25	0.432	0.084
B: BB to RR	30.12	474	7	0.049	0.018
C: RR to HP	16.73	1462	77	0.315	0.035
D: HP to RI	3.55	1385	4	0.081	0.041
E: RI to CB	10.88	1877	51	0.250	0.034
F: CB to SE	11.44	1826	34	0.163	0.028
TOTAL	84.32	1991	198	M total:	0.099

C: Steelhead

Reach	Distance (miles)	N	Events	MI (%/mile)	SE
A: W to BB	11.6	984	29	0.254	0.046
B: BB to RR	30.12	955	40	0.139	0.022
C: RR to HP	16.73	1915	66	0.206	0.025
D: HP to RI	3.55	1849	18	0.274	0.064
E: RI to CB	10.88	2332	147	0.579	0.046
F: CB to SE	11.44	2185	118	0.472	0.042
TOTAL	84.32	2845	418	M total:	0.147

C.4 Mortality intensity

When a fish ceases to be detected, a mortality event can be inferred. From these events, probabilities of mortality can be calculated, as can a mortality intensity, defined as the probability of death per mile traveled. In total, out of 5820 total implanted fish 774 mortality events occurred: 158 of 984 chinook (16%), 198 of 1991 sockeye (10%), and 418 of 2845 steelhead (15%). The distribution of mortality events over reservoirs is, however, not uniform (table C.2 and figure C.5).

In most reservoirs, the highest mortality intensity occurs in the reaches immediately after a dam. This effect is most striking for sockeye. In the Rocky Reach reservoir, the mortality intensity drops from 0.432 % per mile in the initial 11.6 mile reach between Wells dam and Beebe Bridge by nearly an order of magnitude to 0.048 % per mile in the remaining 30 miles to Rocky Reach dam. Chinook smolt on the other hand show an insignificant increase in mortality intensity within the first reservoir. In general chinook and steelhead, display a net increase in mortality intensity as they migrate downstream, with slight increases immediately after passing dams. Steelhead in particular show a significant rise in mortality after passing Rock Island dam.

It should be noted that specific causes of mortality are unknown. Presumably, in the mid-reservoir the bulk of mortality is due to predation. In the tailraces, however, it is difficult to separate the effects of mortality due to dam passage from the possibility of greater predation in the tailraces. A good explanation for the differences in the mortality distribution of the three species requires greater knowledge of the biology of the organisms and the mechanisms of mortality. It should be noted that the sockeye are the smallest of the three species ($\widehat{\mu}$ =117 mm, $\widehat{\sigma}$ =7.26 mm, compared to chinook: $\widehat{\mu}$ =163 mm, $\widehat{\sigma}$ =13.1 mm, and steelhead mean $\widehat{\mu}$ =184 mm, $\widehat{\sigma}$ =18.7 mm), possibly resulting in greater stress-induced mortality during dam passage. The large size of the steelhead might make them more susceptible in general to predation. Differences

in predator concentrations (pikeminnow, bass, walleye, birds) in the reservoirs, as well as differences in swimming behavior might further explain these differences. The later migration of the sockeye relative to the other species by a month might also be a factor.

The take-home message from this analysis is that when modeling survival for migrating salmon, one cannot assume that mortality is distributed uniformly between reservoirs, or even within a single reservoir. The results suggest that an aggregate model of down-migrating salmon survival would benefit from separating mortality at the dams from mortality in the mid-reservoir, with the general implication and the per mile mortality in the mid-reservoir is less than at the dam. This holds especially true for sockeye salmon.

The reaches of the river where velocities are highest tend to correspond with those where mortality intensity is also highest. This contradicts a widely held assumption that mortality is most closely correlated to amount of time spent in a reservoir, and provides some indirect support for a gauntlet-like process. A thorough analysis of the factors that determine the travel velocity of smolt can be applied more widely to reservoirs throughout the Columbia River.

There is little biological reason for migrating smolt to move with any significant randomness. This is supported both by theoretical arguments against increasing encounter probabilities with predators and with empirical evidence that suggests that survival of migrating smolt is dominated by a non-random 'gauntlet'-like process (in the context of the XT-model).

C.5 Concluding comments

There are two somewhat contradictory messages that emerge from this analysis regarding salmon smolt survival. The first is that, not surprisingly, the processes of migration and survival are complicated and heterogeneous. There are significant differences between the species, with sockeye in particular displaying the greatest re-

sponse to dam passage while simultaneously showing the highest overall survival. The significant differences in survival probabilities and travel velocities between the upper and lower reaches of the dams should be taken into account as most models of survival assume a homogeneous process, a constraint due in part to the lack previously of higher resolution migration data.

The second important message is that some things might actually be simpler than one might have guessed. For example, the differences in travel velocities are exceptionally consistent between species. This seems to indicate that migration might be less determined by differences in behavior and rather by the hydrological conditions of the river. This gives hope that a robust physical model of migration might be obtainable which could be generalizable to multiple reservoirs. The apparent lack of any kind of apparent selection on the fish, either for size or migration speed, similarly simplifies modeling of predation. It is clear from the distribution of mortality events that predation is an important source of mortality. The lack of any strong dependence of mortality on time spent in the reservoir is compelling evidence for a gauntlet-like process of survival. The juvenile salmon must necessarily move through a reservoir in a way that appears to be largely determined by the features of the reservoir. It is my own hunch that the spatial distribution of predators is the most significant factor in determining survival. If this is the case, then that is probably the most fruitful direction to focus work on: isolating the sources of mortality and separating the effects and distributions of the various predators.

An analysis of the 3d tracks of salmon approach to the dams might also be fruitful, if only to explore how mortality depends on the path the smolt chooses (bypass, turbine, spillway). It might also suggest the extent of mortality at the dam, though the power of the inference will be relatively diminished.

While PIT-tag data necessarily lacks the high detections and finer resolution of hydroacoustic data, it is far more voluminous and spatially wide-ranging. Elsewhere, efforts are being made to relate survival estimates of migrating smolt obtained from

PIT-tag data to factors such as migration distance, travel time, spill and flow at dams, temperature, turbidity and season. Hopefully, the information that can be obtained from hydroacoustic data can provide some process-based motivations for guiding the modeling that is done basin-wide.
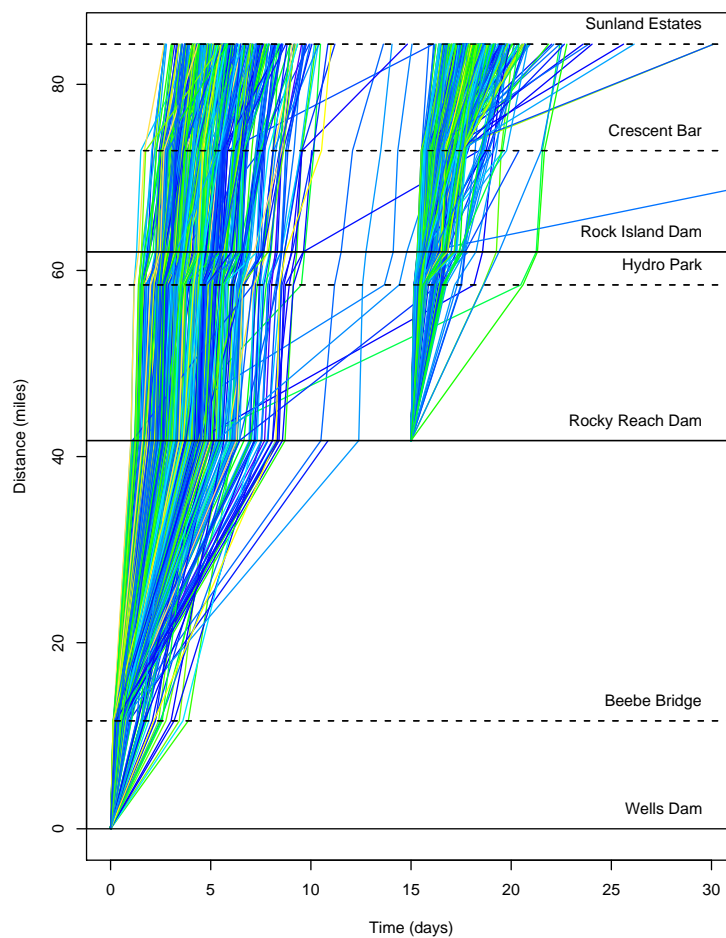
Figure C.2: Travel tracks for chinook juveniles. Darker blue colors represent smaller fish, while yellower colors represent larger fish. While the separate locations of release are separated for clarity, the actual date of release of is not reflected in the plot. Distance traveled is presented along the y-axis while days of travel are on the x-axis; a steeper track represents a fish that spends relatively little time in a given stretch. A track that terminates is a mortality event.
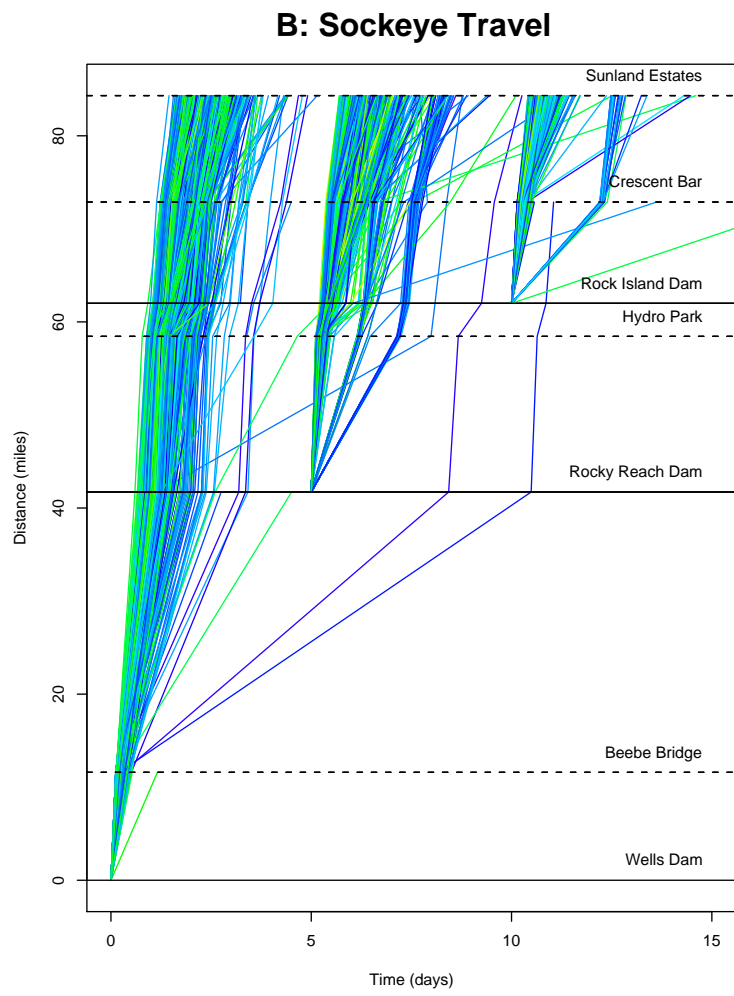
Figure C.3: ravel tracks for migrating sockeye.
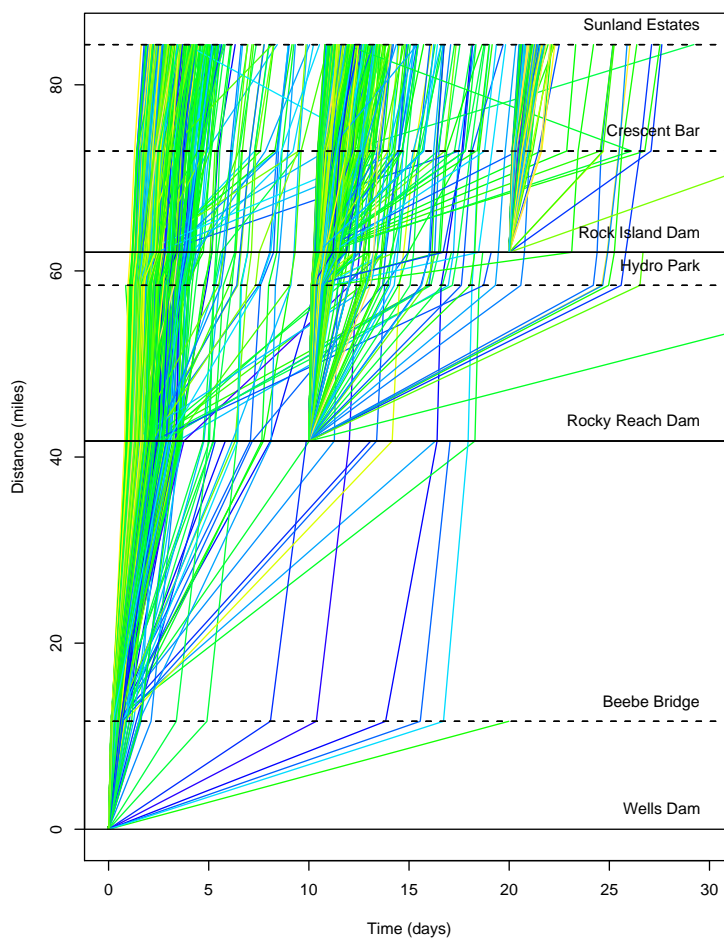
Figure C.4: ravel tracks for migrating steelhead.
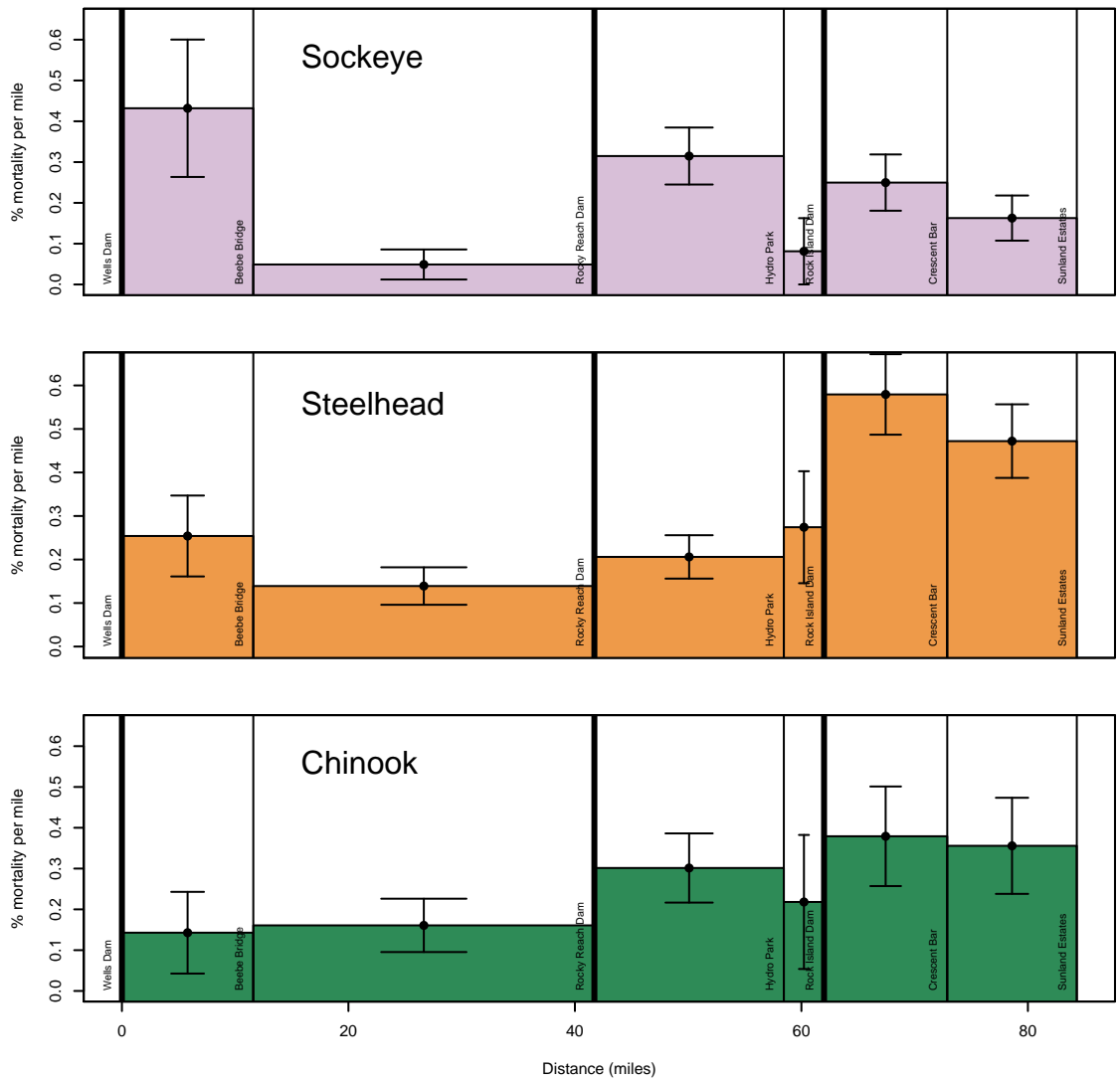
Figure C.5: Intensity of mortality per stretch of each species' travel. Bold faced vertical lines indicate dams. Error bars indicate 95% confidence intervals calculated using the estimated variance from a binomial mortality model (see Table C.2). The area of each rectangle is proportional to the total probability of mortality in the corresponding section.

Appendix D

R-CODE FOR ANALYZING GAPPY MOVEMENT DATA

D.1 Estimating continuous correlation coefficient ρ.

```
GetRho <- function(x,t)
# Obtains MLE estimate of rho for a gappy time-series
{
  # Negative log-Likelihood function
  getL <- function(rho)
  {
    dt <- diff(t)
    s <-sd(x)
    mu <- mean(x)

    n<-length(x)

    x.plus <- x[-1]
    x.minus <- x[-length(x)]

    Likelihood <- dnorm(x.plus,mean=mu+(rho^dt)*(x.minus-mu),
        sd=s*sqrt(1-rho^(2*dt)))
    logL <- sum(log(Likelihood))

    return(-logL)
  }

  # Optimisation routine
  o<-optim(0.5,getL,method="L-BFGS-B",lower=0,upper=0.999))
  return(data.frame(rho=o$par,LL=-o$value))}
```

D.2 Obtaining log-likelihood of single structural breakpoint τ.

```
GetDoubleL <- function(x,t,tbreak){
# This function obtains estimates for mu sigma and rho
```

```
# before and after a given break

  x1<-x[1:tbreak]
  t1<-t[1:tbreak]
  x2<-x[(tbreak+1):length(x)]
  t2<-t[(tbreak+1):length(t)]

  o1<-GetRho(x1,t1)
  o2<-GetRho(x2,t2)

  mu1 <- mean(x1)
  sigma1 <- sd(x1)
  rho1 <- o1$rho

  mu2 <- mean(x2)
  sigma2 <- sd(x2)
  rho2 <- o2$rho

  LL1 <- o1$LL
  LL2 <- o2$LL

  return(data.frame(mu=c(mu1,mu2),sigma=c(sigma1,sigma2),
           rho=c(rho1,rho2),LL=c(LL1,LL2)))}
```

Appendix E

TRAVEL TIME MODEL COMPARISONS: TABLE OF PARAMETER ESTIMATES

Table E.1: Parameter estimates for all three travel time distributions (IG, RN, and IGRN) fit to Snake River spring chinook and steelhead salmon detected at Lower Granite and Little Goose dams between April 10 and May 20 in all years from 1996 to 2005. The Akaike information criteria (AIC) summarizes the quality of the model, taking into account the amount of parameters estimated: a less negative number indicates a better fit.

		Spring Chinook					Steelhead				
Year	Model	N	$\hat{v}$	$\hat{\sigma}_v^2$	$\hat{\sigma}_w^2$	AIC	N	$\hat{v}$	$\hat{\sigma}_v^2$	$\hat{\sigma}_w^2$	AIC
1996	IG	5292	6.94		8.875	-24285	509	14.21		11.143	-1497
1996	RN	-	9.06	4.64		-25563	-	17.55	6.8		-1421
1996	IGRN	-	7.31	1.56	8.063	-24205	-	16.86	5.82	5.061	-1415
1997	IG	254	7.37		8.285	-1100	826	17.37		12.326	-2096
1997	RN	-	9.22	4.32		-1149	-	21.46	7.78		-1862
1997	IGRN	-	7.53	1.07	7.922	-1097	-	21.46	7.78	0	-1860
1998	IG	6287	5.36		7.805	-31970	4310	11.7		9.649	-14110
1998	RN	-	7	3.23		-32121	-	14.2	5.31		-13469
1998	IGRN	-	6.12	1.96	5.714	-31612	-	13.63	4.45	4.622	-13408
1999	IG	8584	9.46		7.201	-29377	4916	11.81		9.908	-16087
1999	RN	-	10.86	3.42		-27665	-	14.45	5.1		-14604
1999	IGRN	-	10.69	3.15	2.461	-27635	-	14.45	5.1	0	-14602
2000	IG	20499	7.79		6.836	-79813	30460	11.79		8.85	-94973
2000	RN	-	9.05	3.18		-78759	-	13.9	4.59		-87013
2000	IGRN	-	8.53	2.27	4.389	-77996	-	13.75	4.35	2.394	-86964
2001	IG	10949	5.11		7.894	-57267	8436	6.24		10.762	-41829
2001	RN	-	6.78	3.33		-58394	-	9.36	4.58		-40344
2001	IGRN	-	5.61	1.55	6.606	-56900	-	8.43	3.46	5.872	-40043
2002	IG	7185	5.34		9.969	-38421	7014	8.94		8.612	-26917
2002	RN	-	8.01	4.44		-39116	-	10.93	4.06		-25545
2002	IGRN	-	6.16	2.05	8.284	-38155	-	10.71	3.74	2.883	-25516
2003	IG	10609	7.94		8.617	-43918	5365	9.5		7.785	-18887
2003	RN	-	9.94	3.79		-41457	-	11.13	3.53		-17205
2003	IGRN	-	9.55	3.23	3.782	-41316	-	11.13	3.53	0	-17203
2004	IG	10431	7.29		7.999	-44851	7803	8.26		10.153	-33036
2004	RN	-	9.01	3.79		-45032	-	11.03	4.69		-31365
2004	IGRN	-	8.1	2.27	5.806	-43907	-	10.54	4.05	4.263	-31257
2005	IG	9925	5.81		8.891	-49898	9480	9.78		9.452	-35299
2005	RN	-	7.94	3.92		-50373	-	12.18	4.74		-33675
2005	IGRN	-	6.71	2.25	6.76	-49449	-	11.74	4.13	4.022	-33584

VITA

Eliezer (Eli) Gurarie was born in 1976 to Soviet emigrés in Jerusalem but did most of his growing up in Cleveland, Ohio, where he earned a Bachelor of Arts and Science degrees in languages, comparative literature and physics at Case Western Reserve University. He earned a Diplôme d'Etudes Approfondies, equivalent to a Master of Science, in environmental geosciences at the Centre Européen de Recherche et Enseignement en Géosciences de l'Environnement (CEREGE) at the Université d'Aix-Marseille in southern France in 2000. After a period calculating ecological footprints at a non-profit think-tank in Oakland, California, he enrolled in the interdisciplinary graduate program in Quantitative Ecology and Resource Management at the University of Washington in Seattle. During this time, he also participated actively in a joint Russian-American project studying Steller sea lions and other marine mammals on the Kuril Islands and the Bering and Okhotsk seas in northeast Asia. He graduated in 2008 with a PhD in Quantitative Ecology and Resource Management.