University of Alberta

Quantitative Analysis of Single Particle Tracking Experiments:
Applying Ecological Methods in Cellular Biology

by

Vishaal Rajani

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Applied Mathematics

Department of Mathematical and Statistical Sciences

Fall 2010
Edmonton, Alberta

Examining Committee

Dr. Gerda de Vries, Mathematical and Statistical Sciences

Dr. Thomas Hillen, Mathematical and Statistical Sciences

Dr. Christopher W. Cairo, Chemistry

Dr. Gustavo Carrero, Centre for Science, Athabasca University

Abstract

Single-particle tracking (SPT) is a method used to study the diffusion of various molecules within the cell. SPT involves tagging proteins with optical labels and observing their individual two-dimensional trajectories with a microscope. The analysis of this data provides important information about protein movement and mechanism, and is used to create multistate biological models. One of the challenges in SPT analysis is the variety of complex environments that contribute to heterogeneity within movement paths. In this thesis, we explore the limitations of current methods used to analyze molecular movement, and adapt analytical methods used in animal movement analysis, such as correlated random walks and first-passage time variance, to SPT data of leukocyte function-associated antigen-1 (LFA-1) integral membrane proteins. We discuss the consequences of these methods in understanding different types of heterogeneity in protein movement behaviour, and provide support to results from current experimental work.

Proem and Poem

Throughout my journey in mathematical biology, I have witnessed a side of science whose objective is not to *declare*, but to *discover*. While some parts of science are focused on finding answers, others are devoted to asking questions. I recommend readers to view this thesis as not only a case study of movement analysis or a collection of insights into the mechanism of a specific biomolecule, but also an example of mathematical biology performing one of its greatest roles; building bridges at the frontier of experimental and theoretical science, by asking questions.

As a message to my future self; remember that this was only the beginning of an adventure of *A Noiseless Patient Spider*.

> A noiseless, patient spider,
> I mark'd, where, on a little promontory, it stood, isolated;
> Mark'd how, to explore the vacant, vast surrounding,
> It launch'd forth filament, filament, filament, out of itself;
> Ever unreeling them – ever tirelessly speeding them.
>
> And you, O my Soul, where you stand,
> Surrounded, surrounded, in measureless oceans of space,
> Ceaselessly musing, venturing, throwing, – seeking the spheres, to connect them;
> Till the bridge you will need, be form'd – till the ductile anchor hold;
> Till the gossamer thread you fling, catch somewhere, O my Soul.
>
> - Walt Whitman (1819-1892). Leaves of Grass. 1900.

Acknowledgements

I would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Pacific Institute for the Mathematical Sciences (PIMS) for providing funding for this thesis.

There are many individuals I would like to thank for their contribution to my work and motivation. I have been very fortunate to have a trio of supervisors who have all played a large role in my growth as a scholar, teacher, and human being. I am thankful for the mentoring of Gerda de Vries, who guided my work with a watchful eye. An ideal supervisor, she granted me the freedom to explore the problem with my own efforts, providing helpful advice and insights into all of my ideas and reasoning. I am also grateful for her zest for interdisciplinary science and for giving me opportunities to TA alongside her distinct and effective teaching style. I was given the amazing opportunity to learn from the best. I would like to thank Gustavo Carrero for his excitement which sparked my own. His perspective on the universality of movement models helped motivate the initial stages of the project. Our on-going discussions sustained this motivation, bringing the project to fruition. I was very fortunate to have the guidance of Chris W. Cairo. In addition to supplying the data for this work, he shared his vast experimental knowledge of chemistry and biology. I would like to thank him for his unique perspective and ambitious attitude which spurred my interest as a budding scientist in an interdisciplinary world.

I would like to thank all the members of the CMB and the Department of Mathematical and Statistical Sciences, both academic and administration, for providing a community of learning and enrichment. For their best supporting roles, I am grateful to Thomas Hillen, Mark Lewis, Adriana Dawes, Cecilia Hutchinson, Alex Potapov, Jaime Ashander, John Simpson, Jason Martyn, Steven Taschuk, and Ross Lockwood. All the experiences that I shared with individual professors and students acted as a positive inspiration to achieving my goals. In addition, I would like to thank members of the Biol 633: *Advanced Techniques in Biology* seminar, taught by Dr. Mark Lewis and Dr. Evelyn Merrill at the University of Alberta in Fall 2009, for helpful discussions that triggered major progress in this thesis. I would also like to extend my gratitude to Daniel Coombs, Raibatak Das, and Omer Dushek for their helpful discussions.

I would like to thank my family: my sister, brother, and parents, for providing all their love and patience. Their support was invaluable in enabling me to follow my dreams. A great deal of gratitude goes to my friends Chase Kantor and Ravin Bastiampillai for making sure I didn't forget how to have fun along the way. Lastly, I would like to thank Jessica Nicoll, for her love and devotion. In times of frustration, excitement, and discovery, she provided me with a channel to cherish and reflect on my experiences.

A final thanks goes to Roland for many thought-provoking walks, and to W. W. for the inspiration.

Thank you,
Vishaal Rajani

Contents

List of Tables

List of Figures

List of Abbreviations

CRW	Correlated Random Walk
FPT	First-Passage Time
LFA-1	Leukocyte Function-Associated Antigen-1
MSD	Mean Square Displacement
NSOM	Near-Field Scanning Optical Microscopy
PMA	phorbol-12-myristate-13-acetate
SPT	Single Particle Tracking
TCZ	Transient Confinement Zones
TEM	Transmission Electron Microscopy
1-D	One Dimension
2-D	Two Dimensions

1 Introduction

Movement data is ubiquitous. In physics, movement data is used to describe the movement of particles as a result of complex interactions between different forces. This movement takes place within fluids, waves, or chemical reactions, varying in scale from subatomic particles to planetary bodies. The analysis of movement data is widely used in ecology for determining habitat preferences, wildlife conservation, migration patterns and predator-prey relationships. Biochemical interpretations of movement data are usually significant for describing inter- and intra-cellular interactions, which are fundamental for communicating information throughout the human body. In public health and social sciences, understanding the interactions between human beings and their spatial environment requires significant data collection, usually requiring a vast number of sources and funding. Even in economics, the stock market is sometimes viewed as a randomly walking entity, for which movement data is essential for analysis and prediction.

It has thus become a vast and popular problem in research to understand movement data. A number of techniques have been developed to extract patterns and descriptive properties from movement data. However, it is a challenge to appropriately and effectively utilize these techniques, so as to accurately infer the movement behaviour of individuals. Due to the richness of movement data, a combination of different techniques is required to fully understand convoluted movement processes, going beyond the extraction of a simple diffusion rate. While detailed observational studies strive to piece together our understanding of individual biological components or species, a number of situations arise where direct observation is inefficient or unavailable. These difficulties are often due to the rate of movement, size, or distance covered of the individual or particle. For example, a microscopic object traveling very quickly over a large distance would be very difficult to observe with the naked eye. On the other hand, observing a large object that moves very slowly over a small distance would also be challenging. However, in view of recent developments in global positioning and microscopy, the convenience of following the movement of an individual indirectly has drawn interest to the phenomenon of associating movement with behaviour. As a result, analysis relies on movement data collected by technological instruments such as microscopes or satellites. This data is often in the form of movement *trajectories*, or, sequences of time-separated position coordinates. These indirect methods of observation have increased the need for mathematical modeling and data analysis in quantitatively decrypt-

ing movement trajectories to aid in detecting patterns, testing hypotheses, and building biological models.

Throughout this project we explore and evaluate methods that are currently used in movement data analysis, and attempt to provide insight into the utility of ecological movement analysis techniques within a cellular biology framework. Many cellular components travel within and around the cell, carrying out important basic functions that allow the cell to play its larger physiological role. We focus on understanding the behaviour of integrins, cell surface protein receptors, which diffuse across the cellular membrane and are fundamental in performing intercellular signaling.

1.1 Problem Description

In this thesis, we focus on the data analysis of Single-Particle Tracking (SPT). SPT is a microscopy technique used to visualize the movement of biomolecules, by labeling individual particles with optical labels. Often, a single SPT data set can contain several types of non-Brownian motion, which we recognize to occur in two main forms, *macroheterogeneity* and *microheterogeneity*.

Macroheterogeneity, or heterogeneity between individual trajectories, is important in understanding the variety of different movement behaviours in a population. In its simplest form, macroheterogeneity consists of different diffusion rates across a population. These differences are often found by measuring the variance of individual parameters across a population (Saxton, 1997). Analyzing macroheterogeneity aids in understanding dominant and subdominant behaviours of moving individuals under different conditions.

We define microheterogeneity as heterogeneity within individual trajectories. Understanding microheterogeneity can give insight into the exact diffusion processes and can significantly contribute to understanding movement behaviour. Microheterogeneity is usually characterized by regions of concentrated movement, changes between diffusion rates, or complicated autocorrelation between individual steps, aspects that cannot always be understood by a clear analysis of macroheterogeneity.

Heterogeneity is the source of much difficulty in SPT data analysis. However, since many proteins individually exhibit multiple movement behaviours based on their conformation, affinity, or environment (Cairo and Golan, 2007), many experimentalists believe that breaking down movement heterogeneity can lead to a better understanding of movement

behaviours of an individual protein.

Our protein of interest for this work is the leukocyte function-associated antigen (LFA-1) integrin found on immune cells. LFA-1 is theorized to have different conformations that exhibit a variety of functions and movement behaviours (van Kooyk and Figdor, 2000). The distributions of diffusion rates provide evidence of rich heterogeneity among LFA-1 populations and evidence for multi-state models (Cairo et al., 2006). However, these conclusions are based solely on the initial rate of diffusion of each trajectory, and do not include information about the specific differences between diffusion mechanisms from one state to the next.

Microheterogeneity in SPT data is widely sought after as being key in understanding the movement mechanisms of individual proteins. LFA-1 integrins are affected by cytoskeletal and external binding with variable affinities, cell-surface features such as corrals and confinement zones, lipid rafts, and varying diffusion rates, all which may change throughout a single trajectory. As a result, a vast number of diffusion models have been thought to explain movement on cellular membranes, including combinations of transient confinement zones (random appearances of confined diffusion) and hop diffusion (diffusion between spatial regions of different shapes and sizes). In addition, LFA-1 proteins have been observed to form receptor clusters (Cambi et al., 2006; van Zanten et al., 2010). A variety of methods have been developed to investigate different elements of microheterogeneity. Simson et al. (1995) have developed a method to detect non-random transient confinement zones by comparing the amount of time spent in a particular region to that of Brownian diffusion, based on probability calculations. Variants of this method have been applied to a broader range of confinement shapes and hop diffusion (Meilhac et al., 2006; Dietrich et al., 2002). Other studies have relied on Markov Chain Monte Carlo methods to study microheterogeneity in terms of switching between diffusion rates (Das et al., 2009). However, these methods are often computationally difficult or rely on a number of data-specific parameters.

In view of this, we develop an approach to analyzing heterogeneity in SPT data, by means of using mathematical techniques commonly applied in an ecological context. Using this approach, we hope to:

1. better understand the macroheterogeneity of LFA-1 movement by testing SPT data for different modes of motion (Brownian and non-Brownian),

2. detect and determine possible mechanisms that might affect the microheterogeneity

of individual LFA-1 movement trajectories,

3. evaluate the utility of methods commonly used in ecology in the context of cellular biology, and

4. recommend further directions for SPT experimentation and data analysis.

In this thesis, the analysis of heterogeneity stated in the above objectives are embodied by two major hypotheses.

A. Individual SPT trajectories of LFA-1 data follow a correlated random walk model.

B. Individual SPT trajectories of LFA-1 data contain areas of confined diffusion.

These hypotheses are tested by ecological methods, and the results are discussed in terms of macro and microheterogeneity.

We believe that a clear understanding of these concepts could build on the multi-state LFA-1 model described by Cairo et al. (2006). In addition, we anticipate these ideas could be applied to other types of SPT experiments, and provide insights for more general studies in movement modeling.

1.2 Thesis Overview

In Chapter 2, we begin by covering the fundamentals of movement analysis, which are helpful in understanding the context of specific models. In particular, many current methods of SPT data analysis rely on an understanding of simple random walks in one and two dimensions, and their relationship to the diffusion coefficient, a parameter describing the rate of diffusion. We discuss the different ways this parameter is estimated from the mean squared displacement, and how the spread of diffusion coefficients is commonly used to characterize diffusion in the SPT environment. In addition, we introduce the ecological concepts of the correlated random walk (Patlak, 1953), and variance first-passage time (Fauchald and Tveraa, 2003), and how they are applied to data.

In Chapter 3, we apply the fundamentals in testing individual SPT trajectories using the CRW model (Patlak, 1953) which has been adapted to many different ecological applications (Kareiva and Shigesada, 1983; Turchin, 1998). In this way, we are able to understand macroheterogeneity by classifying trajectories by their diffusion mechanism, rather than only by their diffusion rate. In the context of microheterogeneity, we suggest using variance

first-passage time, another ecological approach commonly used to identify regions of high foraging or concentrated movement. We apply this method to LFA-1 protein to detect and estimate the size of transient confinement zones, providing evidence for LFA-1 clustering in accordance with Cambi et al. (2006) and van Zanten et al. (2010).

A common problem in interpreting movement data is the occurrence of error or bias. Error often has detrimental effects on movement metrics (Jerde and Visscher, 2005; Hurford, 2009), that are often used in mathematical modeling. As a result, many different ecological studies make note of behavioural, environmental, and human errors affecting their research (Obbard et al., 1998; Ryan et al., 2004; Moen et al., 1996; Bowman et al., 2000) while other studies are completely devoted to error analysis (D'Eon et al., 2002). Throughout this thesis, we encounter a variety of different aspects of error and bias occurrence in SPT data, and we have left the discussion of these features to Chapter 4.

Finally, in Chapter 5, we summarize the conclusions and discuss recommendations for future work.

2 The Fundamentals of Movement Analysis [1]

In this chapter, we will discuss some of the basics of movement analysis. A clear understanding of these concepts is important, as they will be used throughout the thesis for random walk simulations, model testing, and the characterization of different types of movement. In section 2.1, we will describe a couple of different preliminary models applied to movement data. In section 2.2, we will describe a method for simulating random walks, commonly used in model comparison and testing. Traditional methods used in SPT analysis will be described in section 2.3. Finally in section 2.4, we will discuss some methods applied to ecological movement data, which will set up our approach to analyzing SPT data.

2.1 Random Walk Models

Understanding the fundamentals of random walk modeling is important. First and foremost, it provides an intuition of the underlying process by which a particle moves. There is no better way to understand particle movement than to break it down, step by step. Secondly, it provides a preliminary "recipe" for simulation. It gives an idea of parameters that can be manipulated such that random walks of different types can be simulated. Lastly, analyzing the characteristics of random walks can provide insights for estimating parameters from experimental data, such that the underlying process of a given experimental group of moving particles can be better understood. In this section, we attempt to provide enough fundamental knowledge of random walks so that we can successfully simulate random walks, compare statistics between experimental data and simulated data, and develop and test hypotheses for related data sets. In sections 2.1.1 and 2.1.2, we derive the simple random walk (SRW) in one and two dimensions, which are important for many applications. We introduce the correlated random walk (CRW) model in section 2.1.3.

2.1.1 The Simple Random Walk in 1-D and Mean Square Displacement

The simple random walk (SRW) is summarized in countless papers and textbooks. In this section, we use the 1-D lattice definition as described by Berg (1983).

The SRW can be described by a population of M particles starting at the origin, each deciding to move left or right with equal probability ($p_l = p_r = \frac{1}{2}$) and a step length

[1] Portions of this chapter have been submitted for publication.
Rajani V, Carrero G, Golan D, de Vries G, Cairo C 2010. Analysis of molecular diffusion by first-passage time variance identifies the size of receptor clusters. *Biophysical Journal*, 36 manuscript pages.

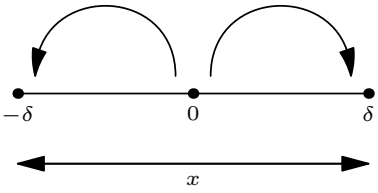


Figure 2.1: Movement rules for a SRW on a 1-D lattice. Positions on the lattice are length δ apart, and the probabilities of moving left p_l and right p_r are equal.

of δ (see Figure 2.1). Let $x_i(n)$ be the position of the i^{th} particle after the n^{th} step. Thus,

$$x_i(n) = x_i(n-1) \pm \delta. \tag{2.1}$$

From this formula, we can calculate the mean displacement of a particle. The mean displacement is given by

$$\begin{aligned}
\langle x(n) \rangle &= \frac{1}{M} \sum_{i=1}^{M} x_i(n) \\
&= \frac{1}{M} \sum_{i=1}^{M} [x_i(n-1) \pm \delta] \\
&= \frac{1}{M} \left(\sum_{i=1}^{M} x_i(n-1) + \sum_{i=1}^{M} \pm\delta \right) \\
&= \frac{1}{M} \sum_{i=1}^{M} x_i(n-1) \\
&= \langle x(n-1) \rangle.
\end{aligned} \tag{2.2}$$

Here, in the limit as $M \to \infty$, $\pm\delta$ will average out to zero. This indicates that the mean displacement remains zero between steps. In other words, particle spread remains centered around the origin. Given this feature, it requires that we develop an expression for the variance, or standard deviation, to measure particle spread. This quantity is conveniently described through the Mean Square Displacement (MSD), a characteristic of random walks that is typically used in many different applications. Using equation (2.1), we calculate

$$x_i^2(n) = (x_i(n-1) \pm \delta)^2 = x_i^2(n-1) \pm 2\delta x_i(n-1) + \delta^2, \tag{2.3}$$

yielding

$$\begin{aligned}\langle x^2(n)\rangle &= \frac{1}{M}\sum_{i=1}^{M} x_i^2(n) \\ &= \frac{1}{M}\sum_{i=1}^{M}[x_i^2(n-1) \pm 2\delta x_i(n-1) + \delta^2] \\ &= \langle x^2(n-1)\rangle + \delta^2. \end{aligned} \tag{2.4}$$

From this recurrence relation, we see that for a particle starting at the origin,

$$\begin{aligned}\langle x^2(0)\rangle &= 0, \\ \langle x^2(1)\rangle &= \delta^2, \\ \langle x^2(2)\rangle &= 2\delta^2, \\ &\vdots \\ \langle x^2(n)\rangle &= n\delta^2. \end{aligned} \tag{2.5}$$

Thus, the MSD increases linearly with step number. After executing n steps, each at a frame-rate of τ, time t is given by $t = n\tau$. Thus, the linear growth of MSD in (2.5) can be rewritten as

$$\langle x^2(n)\rangle = n\delta^2 = \left(\frac{t}{\tau}\right)\delta^2 = 2\left(\frac{\delta^2}{2\tau}\right)t. \tag{2.6}$$

Defining $D := \frac{\delta^2}{2\tau}$, we obtain

$$\langle x^2(n)\rangle = 2Dt. \tag{2.7}$$

Here, D is the diffusion coefficient $[length^2/time]$ and can be used as a single-parameter descriptor of particle movement. We will talk about how the value of D can be estimated from data in section 2.3, and how it traditionally utilized to describe the movement of proteins. We will first describe the context of the diffusion coefficient in relation to the variance of a probability density function. This relationship will help illustrate one method of simulating diffusing particles, and will elucidate the meaning of the diffusion coefficient in terms of particle spread.

The probability that a particle takes k steps to the right in n trials is given by the binomial distribution

$$P(k; n, p) = \frac{n!}{k!(n-k)!}p^k q^{n-k}, \tag{2.8}$$

where p is the probability that the particle moves to the right, and $q = 1-p$ is the probability that the particle moves to the left. Here, it can be shown that the mean and variance of $P(k; n, p)$ are given by (Berg, 1983)

$$\mu = np, \tag{2.9}$$

$$\sigma^2 = npq. \tag{2.10}$$

When n is large, the binomial distribution (2.8) approximates the normal distribution,

$$P(k; n, p) \approx P(k)dk = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(k-\mu)^2}{2\sigma^2}} dk, \tag{2.11}$$

where $P(k)dk$ is the probability that the number of steps taken to the right lies between k and $k + dk$. Rewriting any given position, $x(n)$, as the difference between the total distance moved to the right ($k\delta$) and the total distance moved to the left ($(n-k)\delta$), we have

$$x(n) = [k - (n-k)]\delta = (2k - n)\delta. \tag{2.12}$$

Solving for k, (2.12) becomes

$$k = \frac{x}{2\delta} + \frac{n}{2}, \tag{2.13}$$

yielding

$$dk = \frac{dx}{2\delta}. \tag{2.14}$$

With the identity $2Dt = n\delta^2$ (combining (2.5) and (2.7)), we substitute (2.9), (2.10), (2.13) and (2.14) into (2.11) yielding

$$P(x)dx = \frac{1}{\sqrt{4\pi Dt}} e^{\frac{-x^2}{4Dt}} dx, \tag{2.15}$$

where the standard devation is given by

$$\sigma = \sqrt{2Dt}. \tag{2.16}$$

2.1.2 The Simple Random Walk in 2-D

Many of the previous results can be generalized to two and three dimensional lattices. For two dimensions, we start with a similar structure as in Figure 2.1, but assume that a particle

can move in four directions with step size $\sqrt{2}\delta$, starting at the origin (see Figure 2.2). Each particle moves according to the following two independent movement equations

$$\begin{aligned} x_i(n) &= x_i(n-1) \pm \delta, \\ y_i(n) &= y_i(n-1) \pm \delta. \end{aligned} \tag{2.17}$$

If we measure displacement in terms of

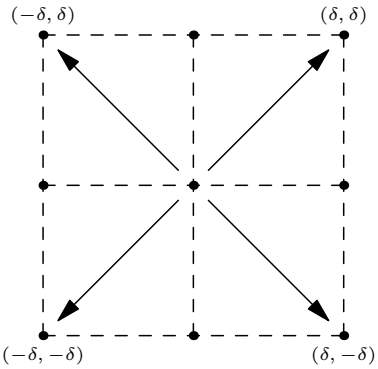


Figure 2.2: Movement rules for a SRW on a 2-D lattice. A particle starting at the origin can move in four directions of equal step length $\sqrt{2}\delta$ with equal probability.

$$r^2 = x^2 + y^2, \tag{2.18}$$

then using (2.4), the 2-D MSD calculation is simply

$$\begin{aligned} \langle r^2(n) \rangle &= \langle x^2(n) + y^2(n) \rangle \\ &= \langle x^2(n) \rangle + \langle y^2(n) \rangle \\ &= \langle x^2(n-1) \rangle + \langle y^2(n-1) \rangle + 2\delta^2 \\ &= \langle x^2(n-1) + y^2(n-1) \rangle + 2\delta^2 \\ &= \langle r^2(n-1) \rangle + 2\delta^2. \end{aligned} \tag{2.19}$$

As a result, we have a pattern similar to the 1-D case:

$$\begin{aligned}
\langle r^2(0)\rangle &= 0, \\
\langle r^2(1)\rangle &= 2\delta^2, \\
\langle r^2(2)\rangle &= 4\delta^2, \\
&\vdots \\
\langle r^2(n)\rangle &= 2n\delta^2.
\end{aligned}$$

Once again, with $t = n\tau$ and $D = \delta^2/(2\tau)$,we obtain

$$\langle r^2(n)\rangle = 2n\delta^2 = \left(\frac{t}{\tau}\right) 2\delta^2 = 4\left(\frac{\delta^2}{2\tau}\right) t = 4Dt. \tag{2.20}$$

Since x and y positions are independently generated we have

$$P(x,y)dxdy = P(x)dx \cdot P(y)dy, \tag{2.21}$$

and thus, $P(x,y)$ is given by the 2-D normal distribution,

$$P(x,y) = \frac{1}{4\pi Dt} e^{\frac{-(x^2+y^2)}{4Dt}}, \tag{2.22}$$

where we have

$$\sigma = \sqrt{4Dt}. \tag{2.23}$$

The final equations (2.22) and (2.23) provide us with a very important relationship between a probability distribution and the diffusion coefficient. In section 2.3, we will discuss in detail how to estimate this value.

2.1.3 The Correlated Random Walk

The SRW is the simplest random walk model for describing basic physical and chemical processes. However, this model is not always sufficient for describing basic movement. Animals are often observed to move with a Correlated Random Walk (CRW) that allows for some degree of *persistence* or correlation between subsequent steps (Turchin, 1998). For example, most particles have the ability to move in one direction, and then move in the opposite direction almost instantaneously. Alternatively, to make this same change in

direction, animals require a number of turns, taking place over a series of moves. As a result, each individual step, when compared with the preceding step, will only show small changes in direction. This short-term correlation will disappear on larger time scales (Codling et al., 2008).

A popular approach to modeling CRWs is by using the Patlak model (Patlak, 1953). The Patlak model assumes correlation between subsequent moves, described via turning angles and step lengths. Turning angles are representations of changes in direction that are measured as the clockwise angle differences between every pair of consecutive steps (Figure 2.3). A description of the complete calculation is given in Appendix A.2. Step lengths complement the description of move direction with information about the extent of mobility. Larger move lengths in a movement path often imply a higher rate of movement, while smaller step lengths often imply slow or restricted motion. As a result, for a given

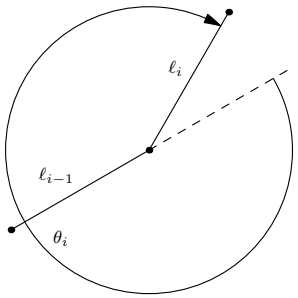


Figure 2.3: The definition of the turning angle. The turning angle θ_i is measured as the clockwise angle difference between two subsequent steps, ℓ_{i-1} and ℓ_i. A full description of the calculation is given in Appendix A.2.

movement path, the characteristics of the collected turning angle distribution and length distribution from empirical data can be used to describe the underlying movement process. The Patlak model is

$$\frac{\partial u}{\partial t} = \frac{1}{2}\nabla \cdot \left[\frac{1 + c\left(2\frac{m_1^2}{m_2} - 1\right)}{1-c} \nabla\left(\frac{m_2}{2\tau}u\right) - \frac{cm_1^3}{\tau m_2 (1-c)} \nabla\left(\frac{m_2}{m_1}\right) u \right], \tag{2.24}$$

where

$$m_1 = \int_0^\infty \ell q(\ell) d\ell, \tag{2.25}$$

$$m_2 = \int_0^\infty \ell^2 q(\ell) d\ell, \tag{2.26}$$

$$c = \int_{-\pi}^{\pi} \cos\theta p(\theta) d\theta, \tag{2.27}$$

and $p(\theta)$ and $q(\ell)$ are the turning angle and step length distributions described above. The extensive use of this model by experimentalists is largely due to the ease of applying this model to discrete data. With the purpose of analyzing insect movement, Kareiva and Shigesada (1983) derived a formula for Net Squared Displacement:

$$\overline{R_n^2} = nm_2 + 2m_1^2 \left[\frac{(c - c^2 - s^2)n - c}{(1-c)^2 + s^2} + \frac{2s^2 + (c + s^2)^{(n+1)/2}}{((1-c)^2 + s^2)^2} \gamma \right], \tag{2.28}$$

where

$$\gamma = ((1-c)^2 - s^2)\cos((n+1)\alpha) - 2s(1-c)\sin((n+1)\alpha),$$

and m_1, m_2, c and s can be estimated from data as

$$m_1 = \frac{1}{N}\sum_{i=1}^{N} \ell_i, \tag{2.29}$$

$$m_2 = \frac{1}{N}\sum_{i=1}^{N} \ell_i^2, \tag{2.30}$$

$$c = \frac{1}{N}\sum_{i=1}^{N} \cos(\theta_i), \tag{2.31}$$

$$s = \frac{1}{N}\sum_{i=1}^{N} \sin(\theta_i). \tag{2.32}$$

Here, m_1, m_2, c and s are discretely calculated by averaging over every value, $\theta_i \in p(\theta)$ and $\ell_i \in q(\ell)$, in the data set. Note that in the case that the turning angle distribution is uniform, $c = 0$ and $s = 0$, and (2.28) reduces to

$$\overline{R_n^2} = nm_2, \tag{2.33}$$

which is the same as the MSD expression (2.5) for a SRW.

Given (2.28), we can estimate the theoretical MSD curve for CRWs. In section 2.4, we will discuss how to use this formula in CRW testing. In the following section, as a prerequisite for understanding movement data, we will discuss the different types of methods used for simulating random walks. These methods will be important for model testing and investigating different analytical methods used to analyze data.

2.2 Random Walk Simulations

A bulk of the computational work in this project is in simulating random walks. Unfortunately, due to the lack of accepted framework, there is no single correct way of generating random walks (Turchin, 1998). This causes random walk definitions and terminology to vary from study to study. To prevent any misinterpretation, we will fully describe the process by which we simulate different types of random walks.

Random walk simulations are important for a variety of reasons. They help in the understanding of the movement rules of a given random walk process. As a result, it is common to compare random walk data to different types simulated random walks, to see which model could potentially describe the movement data. An example of such a comparison will be discussed in section 2.4.1. First, we will discuss the simulation methods used for this project. In section 2.2.1, we introduce our choice for simulating an unbiased, two-dimensional simple random walk, without the confines of a lattice. Later, in section 2.2.2, we make modifications to the algorithm to introduce areas of confined diffusion.

2.2.1 SRW with Variable Move Lengths

In sections 2.1.1 and 2.1.2, we developed SRW models on 1-D and 2-D lattices. In reality, moving individuals are not limited to a lattice, but are able to move in any direction on the unit circle and with a distribution of step lengths (Codling et al., 2008). From equations (2.22) and (2.23), given a diffusion coefficient, we could generate a distribution of all the possible step lengths that might occur around a given position, allowing variable step lengths, rather than a fixed step length. To preserve the realism of moving without a lattice, our SRW simulations will utilize uniform angle and normal length distributions. As a result, we define this as an unbiased, uncorrelated, brownian motion model. The i^{th} position for a random walker, (x_i,y_i), starting from a given point, (x_0,y_0), is given by the recurrence relation

$$x_i = x_{i-1} + \cos(\alpha_i)\ell_i, \tag{2.34}$$

$$y_i = y_{i-1} + \sin(\alpha_i)\ell_i, \tag{2.35}$$

where the step length ℓ_i is chosen from a normal distribution $N(\mu, \sigma)$ with zero mean ($\mu = 0$) and standard deviation $\sigma = \sqrt{4D\tau}$ (where $\tau = 0.001s$ is the sampling rate), and an angle

α_i is chosen from a uniform distribution ($\alpha \in [-\pi, \pi]$). Here, α_i is related to the turning angle θ_i via the relationship

$$\alpha_i = \alpha_{i-1} - \theta_i \tag{2.36}$$

Equations (2.34)-(2.35) thus allow us to simulate random walks according to a prescribed diffusion coefficient. In section 2.3.3, we discuss how to recover this diffusion coefficient from MSD.

Turning angle and length distributions calculated from a typical random walk as generated via equations (2.34)-(2.35) (see Appendix A) are shown in Figure 2.4. As expected, both distributions simply reflect the input. Note that the resulting length distribution is normal, since calculating the resulting length based on the changes in x and y yields

$$\sqrt{(cos(\alpha_i)\ell_i)^2 + (sin(\alpha_i)\ell_i)^2} = \ell_i \in N(0, \sqrt{4D\tau}). \tag{2.37}$$

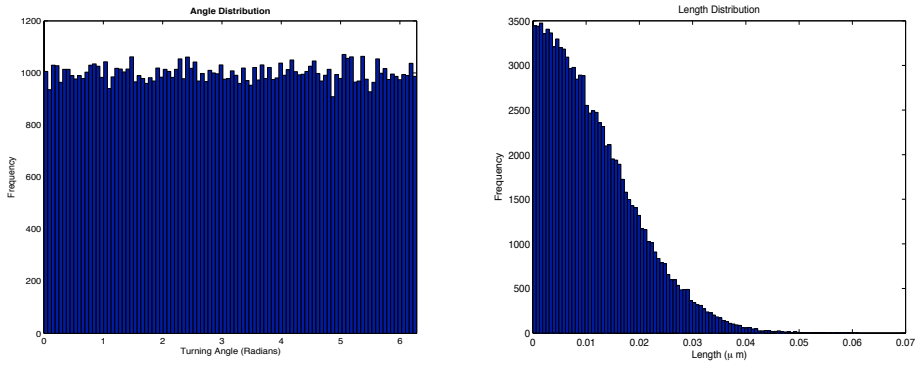


Figure 2.4: Angle and length distributions of an SRW with variable move lengths. The walk is 100,000 steps long and $\ell_i \in N(0, 4D\tau)$ with $D = 5$ x 10^{-10} cm^2/s .

2.2.2 Transient Confinement Zones

Building on this model, we can add parameters to make more complicated models of anomalous diffusion. We simulate transient confinement zones (TCZs), or confined movement within a small circular area, by allowing a particle to move into a confinement zone of radius r_c based on a fixed probability p_i (see Figure 2.5). After a particle enters a confinement zone, diffusion remains constant, but steps can only be taken within the radius of the confinement zone r_c. If a step is taken out of the confinement zone, then the particle will leave with probability p_o. Otherwise, the extraneous step is retaken. The pseudocode in

Algorithm 1 illustrates the method we used to simulate TCZs.

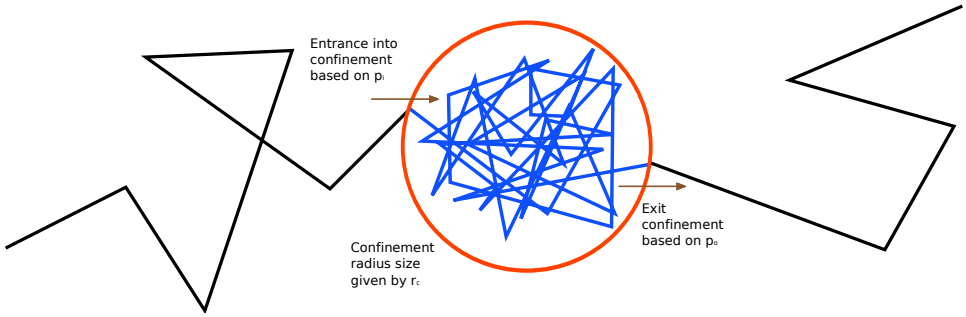


Figure 2.5: Transient confinement zone simulation.

2.3 Traditional Analysis of SPT Data

The most common way used to represent movement data is by $N+1$ position coordinates,

$$\begin{aligned} p_0 &= (x_0, y_0), \\ p_1 &= (x_1, y_1), \\ p_2 &= (x_2, y_2), \\ &\vdots \\ p_N &= (x_N, y_N), \end{aligned} \tag{2.38}$$

for a total number of N timesteps. Without loss of generality, we can assume that $p_0 = (0, 0)$ (this can be done by a very simple translation). The most elementary picture representations of this type of data are trajectory plots, which are created by plotting each coordinate on an (x,y) coordinate plane and by joining subsequent points with a line. While trajectory plots are the most intuitive way of visualizing movement paths, the stochastic nature of random data can be misleading. For example, the SRW in Figure 2.6 is perceived to have some degree of directional bias and regions of confinement, but neither are involved in the actual diffusion process. Due to this visual bias, it is often required that analysis be pushed to more unbiased quantitative methods where artifacts of randomness can be discounted by robust techniques. For this reason, a wide range of tools for understanding movement have been developed. While all of these tools can be applied to most data types, it is important to understand the underlying assumptions involved in the use of each. Although the data output is in a format very intuitive to understand, the analysis and interpretation of this

Algorithm 1 Transient Confinement Zone Algorithm. The following algorithm will generate an N-step trajectory of a particle diffusing at rate D, moving in and out of confinement zones of radius R based on probabilities p_o and p_i

$x_0 = 0$, $y_0 = 0$
$c = (x_0, y_0)$
$\sigma = \sqrt{4D\tau}$
while $i \leq N + 2$ **do**
  $i = i + 1$
  $r_1 = randn$ (normally distributed random number with $\mu = 0$ and $\sigma = 1$)
  $r_2 = 2\pi(rand) - \pi$ (uniformly distributed random number between $-\pi$ and π)
  $x_i = x_{i-1} + \sigma * r_1 * \cos(r_2)$
  $y_i = y_{i-1} + \sigma * r_1 * \sin(r_2)$
  $s = ||(x_i - c(1)), (y_i - c(2))||$
  if $s \leq R$ **then**
    $r_3 = rand$ (uniformly distributed random number in [0,1])
    if $r_3 \geq p_o$ **then**
      $i = i - 1$
    else
      $r_4 = 1$
      while $r_4 \geq p_i$ **do**
        $i = i + 1$
        if $i \geq N + 1$ **then**
          break loop
        end if
        $r_1 = randn$ (normally distributed random number with $\mu = 0$ and $\sigma = 1$)
        $r_2 = 2\pi(rand) - \pi$ (uniformly distributed random number between $-\pi$ and π)
        $x_i = x_{i-1} + \sigma * r_1 * \cos(r_2)$
        $y_i = y_{i-1} + \sigma * r_1 * \sin(r_2)$
        $c = (x_i, y_i)$
        $r_4 = rand$
      end while
    end if
  end if
end while

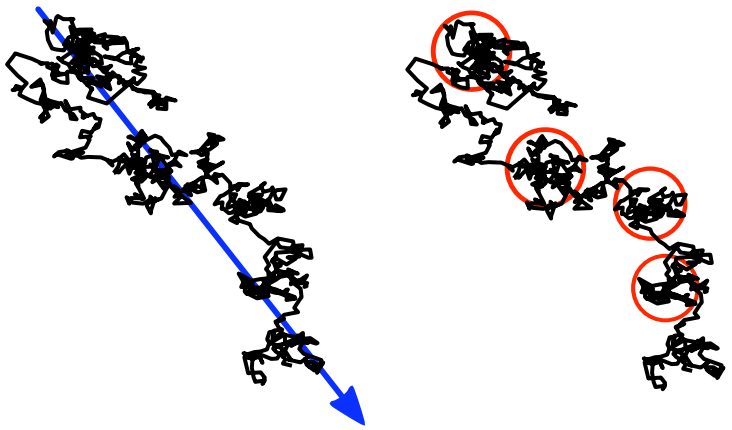


Figure 2.6: The visual bias of a random walk. A SRW which can be misconstrued as having anomalous properties. This SRW ($N = 1000$, $D = 2$ x 10^{-20} cm^2/s) can be perceived to have directional bias (left) or regions of confinement (right).

movement data remains to be mastered. We begin by discussing some of the more common measures of SPT movement data.

Single particle tracking is a technique that is used to monitor the movement of individual proteins along cellular membranes with computer-enhanced video microscopy. Single particle tracking involves the tagging of individual proteins with colloidal gold, latex beads or fluorescent particles, and recording the position of the label at fixed time step increments within a spatial resolution of nanometers and a time resolution of milliseconds (Saxton and Jacobson, 1997). This spatial resolution is the main advantage of this method, over other approaches that measure the movement of entire populations (such as Fluorescence Recovery After Photobleaching (FRAP)) as it allows to ask questions specifically regarding the macroheterogeneity of different subpopulations, and the microheterogeneity within the individual mechanisms of diffusing proteins.

In sections 2.3.1 and 2.3.2, we start by introducing an estimate for MSD, a quantity that was theoretically described in section 2.1.2. Assuming linearity for MSD, one can estimate diffusion coefficients, a method commonly used for breaking down macroheterogeneity in movement data. This will be discussed in section 2.3.3. With this information, we can

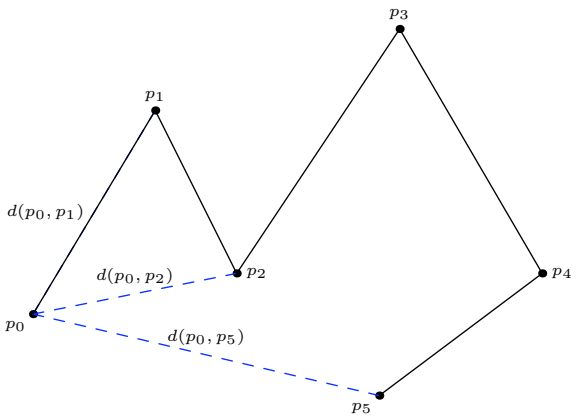


Figure 2.7: Distances used in the calculation of MSD with no internal averaging.

begin to explore macroheterogeneity with a diffusion profiling method, which we explain in section 2.3.4 Note that the estimation of diffusion coefficients from MSD is based on the underlying assumptions of a SRW, which we derived in section 2.1.

2.3.1 Mean square displacement without internal averaging

The square displacement $d^2(p_i, p_j)$ between two positions is defined as the squared Euclidean distance between the two points p_i and p_j, with coordinates (x_i, y_i) and (x_j, y_j), respectively,

$$d^2(p_i, p_j) = (x_i - x_j)^2 + (y_i - y_j)^2. \tag{2.39}$$

Thus, with respect to the initial point, the square displacement $r^2(t)$ at a given time $t = n\tau$ is given by

$$r^2(t) = r^2(n\tau) \approx d^2(p_n, p_0) = (x_n - x_0)^2 + (y_n - y_0)^2. \tag{2.40}$$

Given M trajectories, the simplest way of estimating MSD is by calculating the square displacement for each time interval with respect to the initial starting point (see Figure 2.7) in each individual trajectory. The MSD estimation is then achieved by averaging over all the trajectories in the population. In particular, for $n = 0, 1, 2, \ldots, N$,

$$\rho(n\tau) \approx \frac{1}{M} \sum_{m=1}^{M} r_m^2(n\tau),$$

where $r_m^2(n\tau)$ describes the square displacement $r^2(n\tau)$ of the m^{th} trajectory.
It is important to note that this formulation of MSD is biased towards the initial position as a reference point.

2.3.2 Mean square displacement with internal averaging

The formulation of MSD with internal averaging discussed by Qian et al. (1991) and Saxton (1997) allows for a change in reference point, and is no longer biased towards the initial starting point. Internal averaging refers to the averaging of $n\tau$ time step increments throughout each individual trajectory before averaging over the entire population of trajectories. However, for large n, the averages are based on fewer time step increments, and thus the MSD calculation is less reliable (Saxton, 1997).

Using the Euclidean distance formula in $\mathbb{R}^2$ as in section 2.3.1, we average the squared distance between position pairs separated by $n\tau$ for $n = 1, 2, \ldots, N$ (see Figure 2.8). Thus the formula is given by (Qian et al., 1991),

$$\overline{r^2(t)} = \overline{r^2(n\tau)} \approx \sum_{i=0}^{N_p(n)-1} \frac{d^2(p_i, p_{i+n})}{N_p(n)} = \sum_{i=0}^{N_p(n)-1} \frac{(x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2}{N_p(n)},$$

where $N_p(n)$ is the number of position pairs separated by $n\tau$. Since $N_p(n) = N - n + 1$ for a trajectory of $N + 1$ steps, the formula takes on a more convenient form (Saxton, 1997):

$$\overline{r^2(n\tau)} \approx \sum_{i=0}^{N-n} \frac{(x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2}{N - n + 1} \tag{2.41}$$

Note that this calculation includes lengths that overlap each other (see Figure 2.8). For example, for time separation 2τ, squared distances $d^2(p_0, p_2)$ and $d^2(p_1, p_3)$ are both included in the average, even though they overlap. The alternative to this method is averaging only over independent, non-overlapping pairs (see Figure 2.9). Instead of using Equation (2.41), MSD is then estimated by

$$\overline{r^2(t)} = \overline{r^2(n\tau)} \approx \sum_{i=0}^{N_i(n)} \frac{d^2(p_{ni}, p_{ni-n})}{N_i(n)} = \sum_{i=0}^{N_i(n)} \frac{(x_{ni} - x_{ni-n})^2 + (y_{ni} - y_{ni-n})^2}{N_i(n)}, \tag{2.42}$$

where $N_i(n) = [\frac{N}{n}]$. Here $[\cdot]$ denotes the greatest integer function (Saxton, 1997). The reasoning for using the method of overlapping lengths, as opposed to the one with non-overlapping lengths, is that averaging over all pairs utilizes all the data, and that the data points are all weighted equally (Saxton, 1997). Similar to calculating MSD without internal

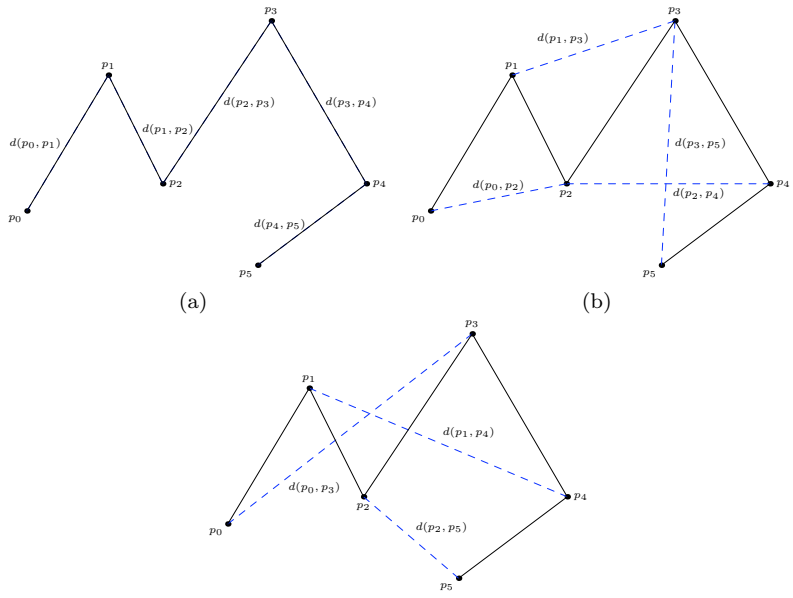


Figure 2.8: Distances used in the calculation of MSD with internal averaging and overlapping pairs. (a) Position pairs obtained with time separation τ. (b) Position pairs obtained with time separation 2τ. (c) Position pairs obtained with time separation 3τ.

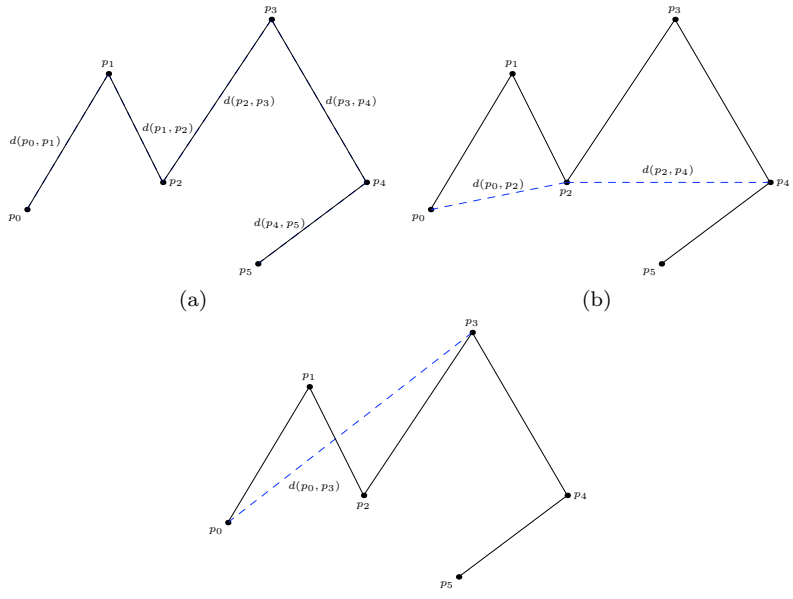


Figure 2.9: Distances used in the calculation of MSD with internal averaging and non-overlapping pairs. (a) Position pairs obtained with time separation τ. (b) Position pairs obtained with time separation 2τ. (c) Position pairs obtained with time separation 3τ.

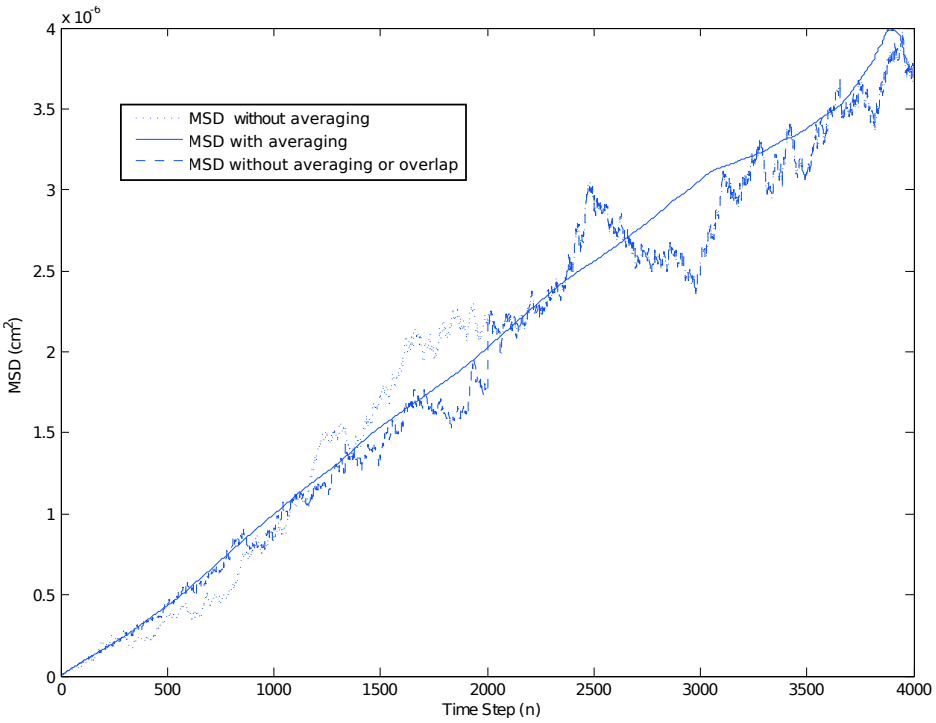


Figure 2.10: MSD estimation using no internal averaging and internal averaging with and without pair overlap. The MSD was calculated from a population of 10 simulated trajectories with $N = 4000$, $\tau = 0.001$ s, and a diffusion coefficient of $D = 1$ x 10^{-6} cm^2/s.

averaging, given M trajectories, averaging over the population yields an estimate for MSD

$$\rho(n\tau) \approx \frac{1}{M} \sum_{m=1}^{M} \overline{r_m^2(n\tau)}, \tag{2.43}$$

where $r_m^2(n\tau)$ describes the average square displacement (square displacement with internal averaging) $r^2(n\tau)$ of the m^{th} trajectory.

To compare the different methods, we simulated a SRW with variable step lengths as discussed in section 2.2.1, and calculated the MSD using each of the three methods. See Figure 2.10 for the resulting plot. Notice that the smoothest MSD curve is given by the calculation with internal averaging and overlap, due to the autocorrelation between overlapping pairs. Also note that half way through the calculation, the MSD calculation without internal averaging and the calculation without overlapping pairs both merge into the same curve.

2.3.3 Estimating Diffusion Coefficients

The diffusion coefficient, D, is estimated via a least squares fit of expression (2.20) to the data. Assuming a zero y-intercept, the diffusion coefficient is calculated, by convention, using one third of the MSD data.

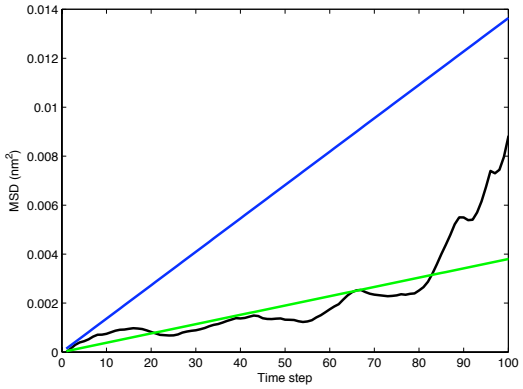


Figure 2.11: The definition of the diffusion coefficient. This figure depicts the fitting of (2.20) to raw data for different definitions of D. The MSD of a simulated SRW ($N = 100$, $D = 5$ x 10^{-16} cm^2/s) is shown in black, and the fits for $D_m = D(0:4)$ and $D_M = D(0:\frac{N}{3})$ are shown in blue and green respectively.

This scalar value depends on the number of tracks M, track length N and the definition of the diffusion coefficient (Saxton, 1997). The diffusion coefficient can be described in various ways, depending on how many time points of the MSD data are included in the least squares fit. Often two different definitions for D are implemented. A short-term diffusion coefficient (often termed D_{micro} or D_m) and long-term diffusion coefficient (termed D_{macro}, or D_M) are calculated by fitting (2.20) to different proportions of the MSD curve. While specific definitions vary, D_m is found by fitting to the first few points of the MSD (in our case, the first four points $D_m = D(0:4)$), while D_M is found by fitting to a larger fraction of the MSD (in our case, $D_M = D(0:\frac{1}{3}N)$.

Different definitions for D yield different forms of information. Since D_m is only measured over the first few points, it provides an indication of the diffusion independent of any anomalous traits which may only appear after long time periods. As such, D_m is usually independent of directed motion, obstacles and corral boundaries (Saxton, 1997). D_M can yield more information; however, a clear framework for classifying different types of behaviour in terms of D_M has not been developed.

The MSD curve for cases of anomalous diffusion can usually result in a positive or negative deflection from linearity for longer time scales. Positive deflection is likely due to

the presence of flow or advection, while negative deflection is usually the result of a source of hindrance; the presence of obstacles or confined diffusion (Qian et al., 1991). This is reflected in the fact that for small N, sources of anomalous diffusion may not be detected. To counter this issue, Saxton and Jacobson (1997) mention the formula

$$\rho(t) = 4Dt^{\alpha}, \tag{2.44}$$

where a new parameter α, estimated by least squares fitting, is used to describe positive concavity for $\alpha > 1$ or negative concavity for $\alpha < 1$. This MSD formula is used to describe the stochastic process known as a *Lévy flight*, which is the product of a heavy-tailed probability distribution of step lengths (Benhamou, 2007). In addition to its wide use in SPT analysis for analyzing macroheterogeneity in movement paths, (2.44) has also been applied to determine the foraging behaviour of a number of different species (Edwards et al., 2007).

In the following section, we will discuss how the spread of diffusion coefficients can be used to understand heterogeneity in a population.

2.3.4 Diffusion Profiling

To measure the degree of macroheterogeneity in movement data, the usual process is to measure the diffusion profile, or spread of diffusion coefficients. Measuring the spread of all the parameters in a population can be indicative of the extent of the heterogeneity in diffusion mechanism (Saxton, 1997). For example, for a data set of M trajectories, where some trajectories are simulated to diffuse with one rate D_1 while others with D_2, the resulting distributions of estimated diffusion coefficients will exhibit macroheterogeneity.

Without the *a priori* knowledge of the simulation, one might hypothesize that there are two distinct groups of diffusing populations, or one population with probabilities of switching between two diffusion coefficients. In experimental data, this may be even more complicated. While some differences can be seen in population distributions (Cairo et al., 2006), without additional statistical testing, it is possible that conclusions are reliant on the method of data representation rather than on the data itself. In addition, while differences can be seen between diffusion coefficients, this tells little about the actual movement mechanism for the particle.

However, the advantages of this procedure do outweigh its shortcomings. While conclusions are obscured in qualitative observation, this procedure supplies aspects of great

importance to experimentalists in the field. As a result, there have been effective studies into parameters that effect the scatter of D. Saxton (1997) describes the effect of M, N and the definition of D on distributions estimated from simulated data. In addition, calculating and analyzing distributions of diffusion coefficients is very simple to do. This method supplies an efficient, effective and widely accepted and understood method for measuring population structure in SPT data. The benefits of this method will be illustrated in application to LFA-1 proteins in Chapter 3.

2.4 Ecological Approaches to Movement Data

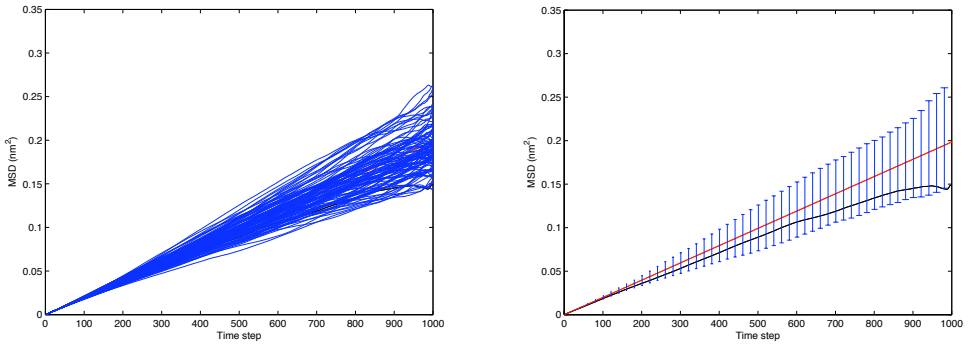
Although the data format is very similar, the preliminary approaches to analyzing movement data are different in ecology. In this section, we focus on introducing some methods from an ecological context, and discuss how they are applied to movement data. In section 2.4.1, we describe a bootstrapping procedure used to identify CRWs. From the perspective of microheterogeneity, we describe the method of variance first-passage time in section 2.4.2, which is used to understand areas of concentrated movement, a feature of random walks that cannot be depicted by simpler models. The context of these methods in application to SPT data will be described in the next chapter.

2.4.1 Correlated Random Walk Testing

The MSD calculations are important in a variety of different schemes throughout single-particle tracking (as were be discussed in section 2.3). However, the calculations are also useful in CRW model testing. Due to the stochastic nature of random walks, we use a bootstrapping procedure described by Turchin (1998).

To test if data from a movement path (or a set of movement paths) could be created under the assumptions of a CRW model, we begin by collecting all of the turning angle and lengths from the data into distributions $\hat{\Theta}$ and $\hat{\Lambda}$. Appendix A describes how to calculate these from data. Ideally, a movement path following a CRW could be generated by merely picking angles and lengths from these distributions at random, and generating a random walk as described by equations (2.34) and (2.35). As prescribed, we simulate many groups of trajectories, termed *pseudopaths*, or *pseudotrajectories* (Turchin, 1998). The number of pseudotrajectories in a group will be the same as the number of trajectories in the data set being tested. The next step in the process is to calculate the MSD (using (2.41)) of each

group of pseudotrajectories, plotting each MSD curve on the same graph. The result will be a family of curves, with gradually increasing variance with respect to time (see Figure 2.12a).



(a) (b)

Figure 2.12: Bootstrapping procedure prescribed by Turchin (1998). (a) The MSD of the simulated groups of pseudotrajectories (blue) surrounding the MSD of the experimental data (black) and theoretical net squared displacement (red). (b) The MSD of the pseudotrajectory groups are shown as a pseudotrajectory envelope for viewing purposes.

The highest and lowest values from the MSD curves comprise the pseudotrajectory envelope (Figure 2.12b), the basis for rejecting or accepting the CRW model. A large number of pseudotrajectory groups is preferable, as the larger the number, the more refined the resolution of the envelope. Note that the theoretical net square displacement (2.28) from the data is a curve approximating the mean of this envelope (Figure 2.12b). The last step in this process is to finally decide whether to accept or reject the CRW model for the data. For this final step, we compare the MSD curve (using (2.41)) of original data set to the MSD envelope obtained from the pseudotrajectories. The MSD curve of the original data set will either lie within the pseudotrajectory envelope, or parts (or all) of the trajectory will lie outside the envelope. In the latter case, we reject the CRW model as being a process capable of explaining the data.

As a word of caution, trajectories should not be grouped together in the testing of the model unless they have been first tested for macroheterogeneity. CRWs grouped with non-CRWs may result in a model rejection, but this does not mean that each individual trajectory cannot be explained by a CRW. Due to this problem, in Chapter 3, we will prescribe the use of bootstrapping for individual trajectories.

2.4.2 Variance First-Passage Time

Variance of first-passage time is a method that has been developed to detect regions of heterogeneity within paths, punctuated by changes in turning rates or movement speeds (Johnson et al., 1992). First-passage time (FPT) refers to the number of steps that an individual takes within a circle of a given radius r. Recording the first-passage time for a circle with radius r centered on each step of a movement path yields a distribution, and the variance of this distribution is a measure of the amount of heterogeneity at the spatial scale of radius r (Fauchald and Tveraa, 2003). By varying r, one can determine the spatial scale that exhibits the most heterogeneity. This significance of this spatial scale has been described with respect to the searching and foraging behaviour of a number of different species, such as the wandering albatross (Weimerskirch et al., 2007), the bottlenose dolphin (Bailey and Thompson, 2006) and the Antarctic petrel (Fauchald and Tveraa, 2003), as the scale at which animals change their movement most frequently. In effect, this measurement provides an estimate for the size of confined or concentrated regions of movement.

The first-passage time $T_r(n)$ of each path is calculated by counting the number of steps taken within a circle of radius r, centered at each point (x_n, y_n) of the trajectory (Fauchald and Tveraa, 2003). Increasing the radius size of the circle allows for more of the trajectory to be captured within its frame of reference, and thus the more tortuous the path segment, the larger the value of $T_r(n)$ will be for that range. Measuring the variance of these values over the entire trajectory, denoted by

$$S(r) = var[log(T_r(n))], \tag{2.45}$$

gives an indication of the degree by which the particle changes its movement behavior from linear to more tortuous movement. The log-transform is prescribed to make $S(r)$ independent of the mean first-passage time (Fauchald and Tveraa, 2003; Weimerskirch et al., 2007). Since the value of $T_r(n)$ is sensitive to changes in movement behavior, peaks in $S(r)$ describe spatial scales at which tortuous movement is concentrated.

To test this method, we decided to run the variance of FPT algorithm on a series of simulated trajectories simulated with TCZs. We modified p_o, p_i and r_c to see the effect on $S(r)$. Figure 2.13 summarizes the findings of this test. When we varied r_c for simulated trajectories with the inability to leave confinement zones, we found that the location of peaks in $S(r)$ closely predicted the confinement radius for r_c= 50, 100, and 200 nm. (Figure

2.13A). However, for large confinement zones (r_c = 500 nm, 1000 nm), we found that the movement track was too short to detect confinement effects, thus the $S(r)$ curves showed no peak at this location.

Varying the probabilities p_o and p_i, while keeping r_c fixed, mainly changed the height and width of the $S(r)$ peak. The simulations showed that increasing p_i resulted in a broader peak, whose location slightly overestimated the actual confinement size of r_c (Figure 2.13B). Conversely, we found that lower p_o resulted in the individual spending more time in confinement, yielding higher peaks in $S(r)$ (Figure 2.13C).

A few patterns are clear. The location of peaks in $S(r)$ seems to give a rough estimation of the size of confinement zones. While exact size may be difficult to predict, spotting a peak in $S(r)$ can be a useful way of detecting heterogeneity. In addition, we find that the height of the peaks correlates with the amount of time spent in confinement zones. For example, high p_i and low p_o, both implying increased confinement, yield higher, more intense peaks in $S(r)$.

It is interesting to note, as has been done in other studies with various types of confined diffusion (Fauchald and Tveraa, 2003), that the peak of $S(r)$ can give an estimate of the size of the simulated confinement zone within the same order of magnitude. When particles are confined to spatial regions where the probability of leaving is $p_o = 0$, the location of the peak in $S(r)$ accurately reflects the size of the simulated confinement zone (Figure 2.13A). If the spatial confinement zone radius is large with respect to the diffusion rate, then the confinement zone may not be detected. As a result, a trajectory diffusing in a very large confinement zone may be categorized as a CRW. When p_o and p_i are manipulated with r_c fixed, we see changes in peak intensity, sometimes causing variations in the peak position. The characteristics of peak intensity and peak width have yet to be explored in relation to simulation parameters. In the next chapter, this technique will be applied to SPT data obtained for the LFA-1 protein in order to determine the presence and size of transient confinement zones.

In the next chapter, we will suggest using variance first-passage time as a tool in detecting transient confinement behaviour in LFA-1 protein movement.

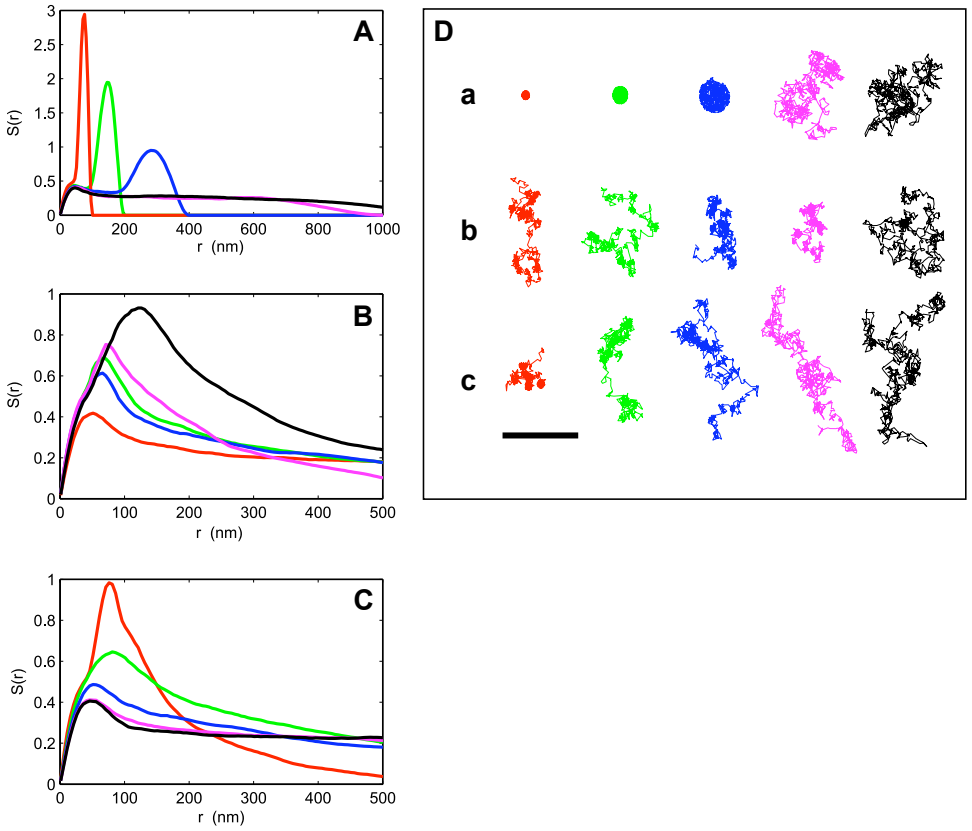


Figure 2.13: First-passage time analysis of simulated trajectories. Each $S(r)$ curve is averaged over 20 trajectories, simulated with $D = 5$ x 10^{-9} cm^2/s. The confinement radius, r_c, probability of leaving a confinement zone, p_o, and probability of entering a confinement zone p_i were varied. (A) $p_i = 1$ and $p_o = 0$ while $r_c = 50$ nm (red), 100 nm (green), 200 nm (blue), 500 nm (magenta) and 1000 nm (black). (B) $p_o = 0.1$ and $r_c = 50$ nm while $p_i = 0.05$ (red), 0.2 (green), 0.4 (blue), 0.8 (magenta) and 1 (black). (C) $p_i = 0.1$ and $r_c = 50$ nm while $p_o = 0.05$ (red), 0.2 (green), 0.4 (blue), 0.8 (magenta) and 1 (black). (D) Sample trajectories from (A)-(C). The scale bar is 1 μm.

3 Detecting Heterogeneity in Biomolecular Diffusion [2]

In this chapter, we will discuss the use of the ecological methods in the context of SPT data. In particular, we focus on a particular cell surface adhesion receptor, the leukocyte function-associated antigen (LFA-1) protein. We begin by giving a short biological overview of adhesion receptors in section 3.1. In section 3.2, we test LFA-1 SPT data for correlated random walk behaviour, to determine the underlying movement mechanism of LFA-1 proteins. To detect the presence of confined diffusion, in section 3.3 we apply the method of variance first-passage time to LFA-1 data sets. As a result, we are able to better understand the population structure (in terms of both macro and microheterogeneity) of LFA-1 proteins. Lastly, in section 3.4, we suggest future applications for first-passage time in SPT data. We finish with an evaluation of our methods and findings in section 3.5.

3.1 Biological Framework

For animals, it makes sense that movement and behaviour are naturally linked. Therefore, by analyzing the movement of animals, we can make predictions about the way animals react with their environment to better understand their behaviour.

On the other hand, biomolecules such as proteins do not intuitively possess this relationship between the way they move and behave. However, with SPT, we have been able to view different forms of macro and microheterogeneity for biomolecules, resulting in a variety of movement patterns. In the case of cell surface protein receptors, this movement is thought to be strongly linked to protein function. In section 3.1.1, we give a brief overview of the role of integrin protein receptors in the cell, and current biological models of how receptor diffusion is connected to their function. In section 3.1.2, we specifically discuss current biological models of LFA-1, developed from heterogeneity observed in SPT experiments.

3.1.1 Cell surface adhesion receptors

Cell surface adhesion receptors (integrins) help regulate activity between cells and the extracellular matrix (Springer, 1990). They play a fundamental role in cell-cell signaling, and allow individual cells to communicate with a variety of different environments. Integrin receptors carry out cellular functions via a receptor-ligand relationship. These functions

[2] Portions of this chapter have been submitted for publication.
Rajani V, Carrero G, Golan D, de Vries G, Cairo C 2010. Analysis of molecular diffusion by first-passage time variance identifies the size of receptor clusters. *Biophysical Journal*, 36 manuscript pages.

are accomplished by conformational changes and chemical affinities of individual receptors (Cairo et al., 2006), spatial organization (van Zanten et al., 2010), and receptor expression.

As their name suggest, in addition to relaying information between cells, adhesion receptors are important in regulating adhesive structures which promote binding between cells and their environment (Cairo and Golan, 2007). The spatial distribution of these receptors is critical in facilitating receptor adhesion and cellular binding, and thus, there is much interest in the lateral diffusion of adhesion receptors. A large number of experiments have been conducted to understand different characteristics of receptor lateral mobility, in particular, the hypothesized relationship between mobility and receptor affinity (Cairo et al., 2006).

In the investigation of this hypothesis, anomalous diffusion has been observed in adhesion receptor SPT data. Specifically, diffusion is observed to be slower than expected rates, which can be attributed to any number of complex processes (Saxton and Jacobson, 1997). In the case of cellular adhesion receptors, anomalous diffusion could be due to a variety of static and dynamic processes including receptor clustering (van Kooyk and Figdor, 2000), microdomain formation (Edidin, 2003), cytoskeletal attachment (van Kooyk et al., 1999; Cairo et al., 2006), membrane compartmentalization (Ritchie et al., 2005), and receptor-ligand interactions (Cairo and Golan, 2007). On T-cells of the immune system, these complex processes can be generalized to any combination of four major mobility regulating mechanisms: *reorganization* of lateral receptor association with microdomains and other components, *recruitment* or clustering of receptors, *dispersion* of receptors to evenly distribute binding sites, and *anchoring* by either cytoskeletal or cytoplasmic proteins (Cairo and Golan, 2007). The large number of possible mechanisms proves anomalous diffusion to be very complicated to decipher.

3.1.2 The Leukocyte Function-Associated Antigen

The leukocyte function-associate antigen (LFA-1) is an important adhesion receptor in the immune system. It is an integrin protein; a transmembrane heterodimer made from non-covalently linked α and β protein chains (van Kooyk et al., 1999) and is located on most leukocytes, particularly on T-cells. LFA-1 is fundamental in triggering the immune response. By binding with its ligand, the intercellular adhesion molecule (ICAM-1), LFA-1 facilitates the movement of T-cells across the endothelium, and the formation of the immunological synapse.

As with other adhesion receptors, although LFA-1 is known to undergo anomalous diffusion, the complete relationship between receptor affinity and lateral mobility is not entirely known. During the formation of immune synapses, LFA-1 receptors undergo a complex combination of the four processes listed above; *reorganization*, *recruitment*, *dispersion* and *anchoring* (Cairo and Golan, 2007).

The model of LFA-1 is shown in Figure 3.1. The model illustrates four different conformations of LFA-1: closed, open, intermediate, and ligated, in two different cellular states: resting and activated. It was found that different conformations were either primarily freely diffusing (high mobility) or bound to the cytoskeleton (low mobility) when in different cellular states. The dominant state of mobility is indicated by the filled grey boxes in either the resting or activated cell. This model was proposed by Cairo et al. (2006), based on the diffusion profiling of different epitopes. Epitopes, or antigenic determinants, are different structural configurations of proteins that are recognized by antibodies in the immune system. Cairo et al. (2006) ran a series of SPT experiments by employing different epitopes of LFA-1. Each epitope labels a different conformation of LFA-1. A diffusion coefficient was estimated from the MSD of each movement track, and the diffusion coefficients were viewed as a distribution (for details of this procedure, see section 2.3.4). The results of the study are listed in Table 3.1. The activation of receptors was induced by the addition of phorbol-12-myristate-13-acetate (PMA). PMA is a chemical that is known to stimulate protein receptor activity (Constantin et al., 2000). The mobility of each configuration was based on the observed population diffusion profiles. The results in Table 3.1 give much insight into the

Type	**Conformation**	**Resting Mobility**	**Activated Mobility**
TS1/18 epitope	open and closed	primarily immobile	primarily mobile
HI111 epitope	closed	primarily mobile	not visible
MEM148 epitiope	open	primarily mobile	primarily immobile
ICAM-1 ligated	ligated	primarily immobile	totally immobile

Table 3.1: Configurations and mobility of LFA-1, as observed by Cairo et al. (2006). It is important to note that the TS1/18 epitope blocks ICAM-1 adhesion, while HI111 and MEM148 allow LFA-1 to remain open to ICAM-1 adhesion.

relationship between LFA-1 protein affinity and lateral mobility. The observation of multiple subpopulations in diffusion profiling motivated a multiple state model for LFA-1 under both resting and activated conditions. The link between affinity and mobility motivates further study into heterogeneity both at the population and the individual level. In section 3.2,

we apply a CRW model to help make the distinction between different types of LFA-1 movement. For our study, we make use of four of the same data sets utilized by Cairo et al. (2006); TS1/18 and MEM148 labelled LFA-1 in both resting and PMA activated forms.[3]

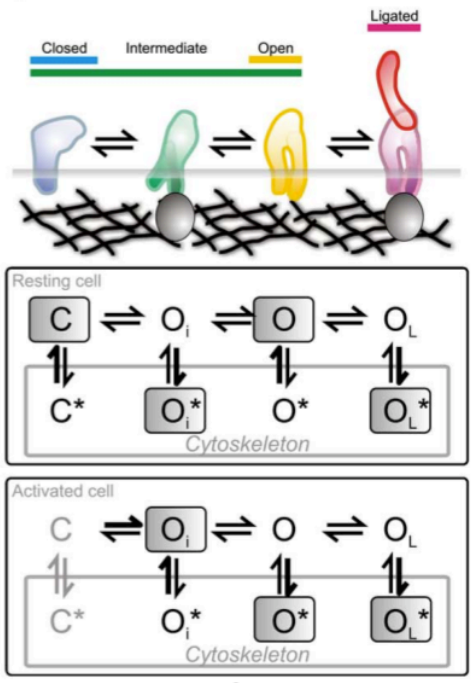


Figure 3.1: LFA-1 model proposed by Cairo et al. (2006). The model displays four conformations of LFA-1: closed, open, intermediate, or ligated. Each conformation was tagged using different epitopes, on both resting and activated cells. Based on diffusion profiles, receptors were classified as either primarily mobile (detached from cytoskeleton) or primarily immobile (attached to cytoskeleton), shown by the filled grey boxes. (Figure from Cairo et al. (2006)).

3.2 Rejecting the CRW Model for LFA-1 Diffusion

As discussed in the previous chapter, there are a number of methods that can be used to understand heterogeneity in SPT tracks. However, to our knowledge, the CRW model and FPT analysis have not been applied to SPT data. In this chapter, we establish the role that these methods can play in determining the mobility mechanisms of LFA-1 receptors. The idea that cell surface receptors can encounter different types of spatial heterogeneity (Cairo and Golan, 2007), and that LFA-1 proteins exhibit anomalous diffusion Cairo et al. (2006), suggests that more complex models and data analysis methods should be used to understand LFA-1 movement mechanisms. In order to test proteins for a random Brownian

[3]SPT data was acquired at 1000 FPS with a Fastcam Super 10K Camera (Photron USA, Inc., San Diego, CA). Video data were processed with Metamorph (Universal Imaging, Downington, PA). See Cairo et al. (2006) for full experimental details.

motion model, we decided to start by using an approach commonly used in ecology; testing for a CRW model.

Since the most popular descriptors for describing the diffusion of individual proteins are given by diffusion coefficients, we decided to utilize the CRW model to evaluate the appropriateness of a linear fit of (2.20) to the MSD. If a non-uniform turning angle distribution was collected from data, we would expect that the single parameter D would be insufficient to completely describe the diffusion process of LFA-1. In Figure 3.2, turning angle and length distributions are shown for TS1/18 and MEM148 labelled LFA-1, in both resting and PMA activated states. It is clear that the turning angle distributions for none of the experiments of LFA-1 are uniform, indicating perhaps some degree of persistence or anti-persistence.

We would like to know if these distributions, by selecting angles and lengths at random, can recreate the diffusion process observed in the LFA-1 SPT data sets. To answer this question, we test the CRW model by the bootstrapping procedure described by Turchin (1998) (described in section 2.4.1), to see if this algorithm for creating movement tracks is limited to a two-step correlation (via a turning angle). Each data set was tested as a population, using 500 groups of pseudotrajectories. Figure 3.3 shows that none of the individual populations could be described by a CRW model. From the failure of the CRW test for each data set, we can assume that most trajectories do not follow a CRW, and instead, there is a more complicated process involved. More troubling is that the generality of the CRW (as previously discussed) would not support this data being subjected to diffusion coefficient estimation, unless higher correlations between steps were ignored or removed.

Due to the possibility of macroheterogeneity, we decided to repeat the CRW test for individual trajectories, in the hope that some individual trajectories might be classified as a CRW. The results are given in Table 3.2. Since trajectories which pass the CRW test have higher rates of diffusion (increased MSD slopes to fit into the pseudotrajectory envelope), we were able to successfully filter each population of trajectories into two separate groups, identifying a *dispersive* class as those trajectories that followed the CRW model. It is interesting to note that the activated TS1/18 trajectories had an increased number of trajectories that passed the CRW model. Comparing with the results of Cairo et al. (2006) shown in table 3.1, we find that this agrees with seeing a primarily mobile proportion of LFA-1 trajectories with the addition of PMA. In addition, for MEM148 labeled LFA-1 activated with PMA, the CRW test indicates that fewer trajectories can fit the pseudotrajectory

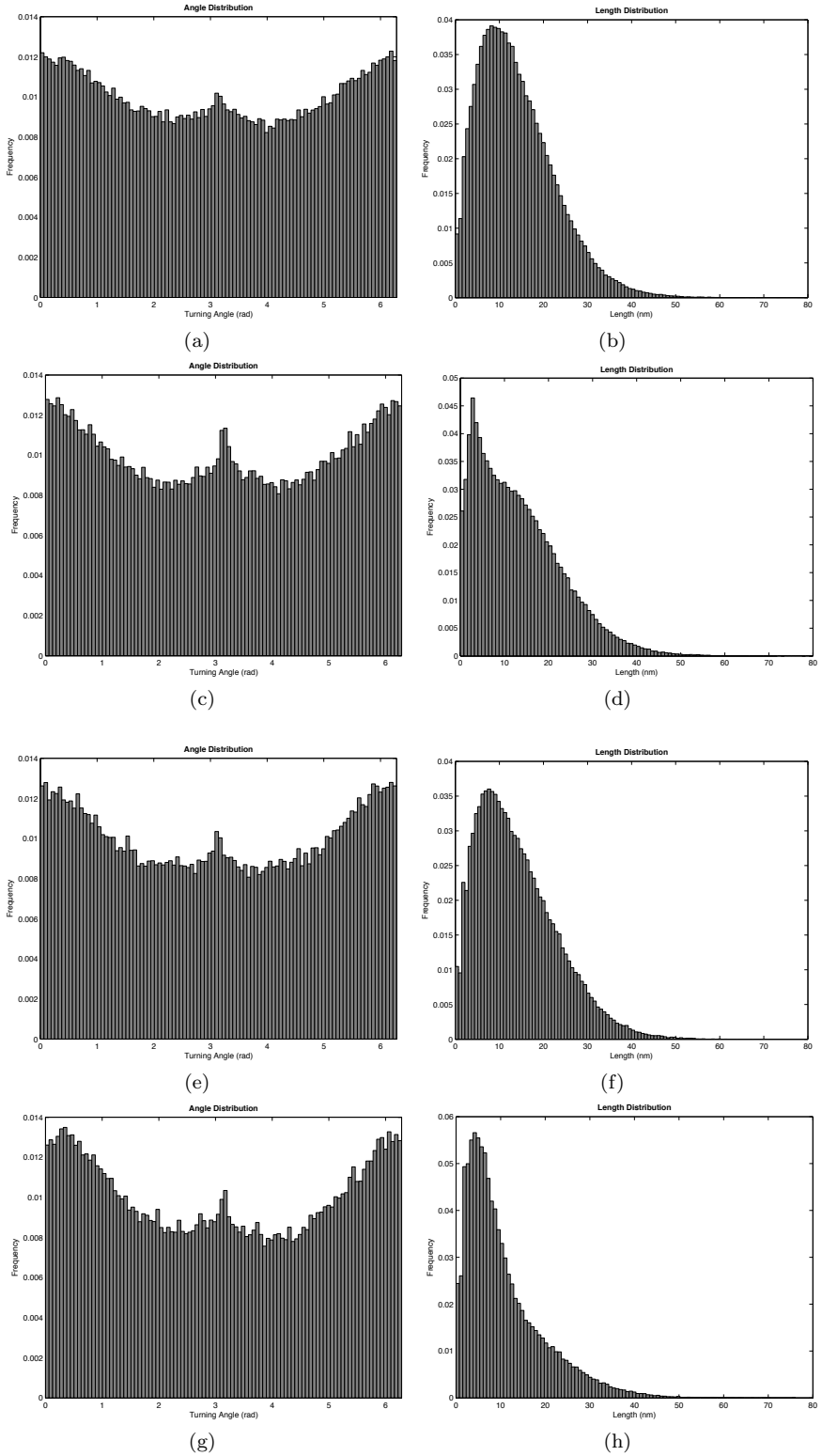


Figure 3.2: Turning Angle and Length Distributions collected from SPT data. (a) Angle distribution for TS1/18 labeled LFA-1 (control) (b) Length distribution for TS1/18 labeled LFA-1 (control) (c) Angle distribution for TS1/18 labeled LFA-1 (PMA treated) (d) Length distribution for TS1/18 labeled LFA-1 (PMA treated) (e) Angle distribution for MEM148 labeled LFA-1 (control) (f) Length distribution for MEM148 labeled LFA-1 (control) (g) Angle distribution for MEM148 labeled LFA-1 (PMA treated) (h) Length distribution for MEM148 labeled LFA-1 (PMA treated).

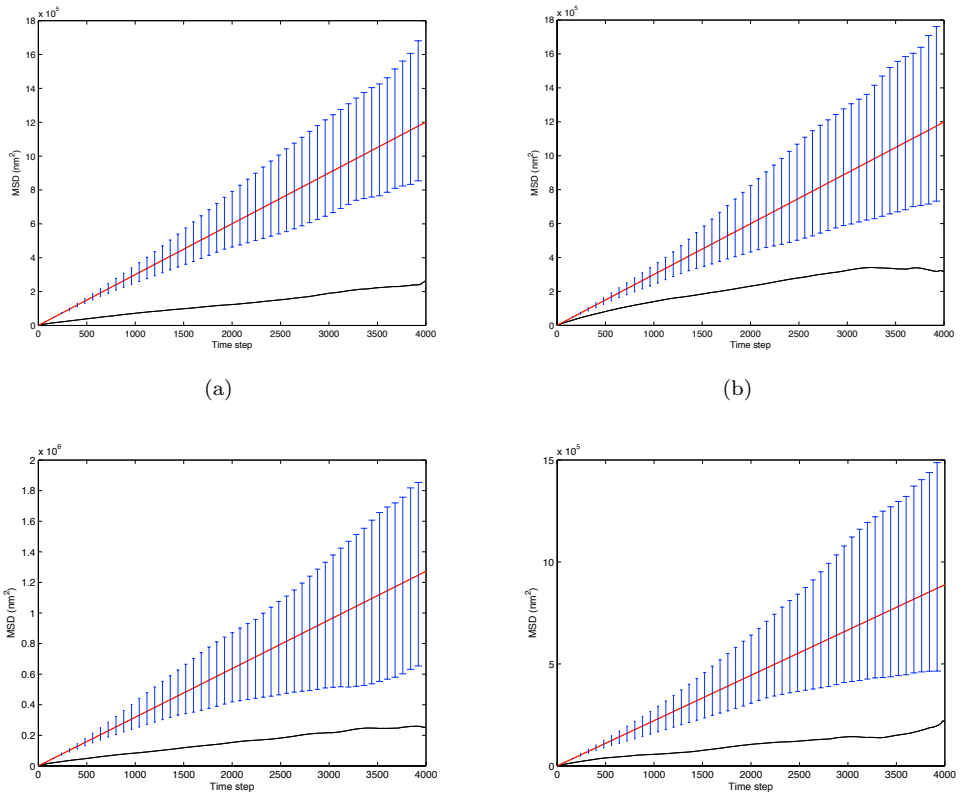


Figure 3.3: Rejecting the CRW model for population data sets. The CRW model was tested for each data set, with 500 groups of pseudotrajectories. The black line in each figure depicts the observed MSD of the population, the red line represents the net squared displacement according to the CRW model, and the pseudotrajectory envelope is represented by the blue error bars. (a) TS1/18 labeled LFA-1 (control); (b) TS1/18 labeled LFA-1 (PMA treated); (c) MEM148 labeled LFA-1 (control); (d) MEM148 labeled LFA-1 (PMA treated).

envelope, admitting slower diffusion. This is in agreement with Cairo et al. (2006), which shows a decreased *dispersive* class (decreased mobility).

By using the CRW model, we were able to explore another way of measuring macro-heterogeneity. While other methods such as diffusion profiling only break down heterogeneity with respect to scalar values estimated from the slope of the MSD, the basis for separating trajectories using CRW has much deeper context. Instead of separating trajectories by the magnitude of their corresponding diffusion coefficient, trajectories are grouped according to their diffusion process. There are shortcomings of this approach. Although we were able to rule out a two-step correlation via a turning angle, as well as random brownian motion, there is very little information revealed about the trajectories that fail the CRW test.

In the next section, we apply another ecological technique, variance first-passage

Data set	N_T	CRW Accepted (%)	CRW Rejected (%)
TS1/18 labeled (control)	75	13	87
TS1/18 labeled (PMA treated)	39	36	64
MEM148 labeled (control)	39	10	90
MEM148 labeled (PMA treated)	31	3	97

Table 3.2: Proportion of trajectories that passed/failed the CRW model in each data set.

time, to gain more insight into trajectories for which the CRW model was rejected.

3.3 Variance First-Passage Time Analysis of LFA-1 Receptors

While most macroheterogeneity was taken care of by the filtering procedure discussed in section 3.2, the larger problem was understanding the microheterogeneity in trajectories for which the CRW model was rejected. Motivated by the non-uniform angle distributions (Figure 3.2), we decided to see if LFA-1 particles exhibited properties of the area restricted search (ARS), by using the variance first-passage time (FPT) analysis (see section 2.4.2 for details). A peak at π in the turning angle distribution indicates high direction reversal, common in regions of confinement. The knowledge that LFA-1 receptors undergo immobile states, as a result of cytoskeletal adherence, ICAM-1 binding or receptor clustering (recruitment) justifies the use of variance FPT in LFA-1 data. Using variance FPT, we scanned each trajectory for confined regions. Confined motion for a given receptor often yields heterogeneous areas of highly correlated turning angles and lengths, which are analogous to animal searching in high prey density or resource rich areas. Due to these high correlations, confined motion should cause the rejection of the CRW model.

With biological motivation for searching for regions of confined movement in SPT data, we applied the variance FPT calculation to each individual trajectory within its respective population. We calculated $S(r)$ for each trajectory (see section 2.4.2), letting r vary from 0 nm to 1500 nm. Each curve reflects the changes in variance FPT over different spatial scales. Peaks in $S(r)$ represent the occurrence of large changes in microheterogeneity, indicative of some degree of confined or concentrated diffusion, which transiently occur throughout the trajectory. In addition to indicating higher correlation, the position of the peak gives a good indication of relative confinement size (Fauchald and Tveraa, 2003). We note that the trajectories which were identified as CRWs in the previous section do not contain any higher correlation and lack peaks in the FPT curve.

In section 3.3.1, we discuss in detail how the results of the variance FPT analysis allow us to make conclusions about LFA-1 diffusion behaviour in response to PMA activation. Following this discussion, in section 3.3.2, we speculate about the relevance of peaks in $S(r)$ with respect to receptor clustering.

3.3.1 Understanding Population Structure of LFA-1 Data Sets

In Figures 3.4 and 3.5, we summarize the results of applying the variance FPT method to the four data sets of interest, namely TS118 labeled LFA-1 (control and PMA activated) and MEM148 labeled LFA-1 (control and PMA activated). For each data set, we separated the trajectories based on their peak locations using the `peakdet` algorithm written by Billauer (2008). This algorithm detects peaks in curves by checking for a threshold difference between a maximum and its surroundings. By measuring the height of a peak relative to adjacent troughs, the algorithm is able to pick out meaningful maxima and minima in data series. After individual CRW testing, each $S(r)$ curve was binned into one of four categories based on peak location: 0-50 nm, 50-150 nm, >150 nm, and those that were not rejected as a CRW (little or no peak). For Figures 3.4(a)(b) and 3.5(a)(b), A-D show the $S(r)$ curves for each of these four categories, while F displays sample trajectories from each. Figure E shows the average $S(r)$ curve from each category. While the bin values for peak location were chosen by convenience, the method is very useful for understanding the full range of probable movement behaviours. It is interesting to note that the sample trajectories from group D (CRWs) are much more diffusive and larger in size than the sample trajectories from any other group. Figure 3.6C provides the overall percentages of each population within each category. From this visualization, it is useful to see the population structure change for different labels and under different conditions.

While the separation of curves based on peak location is revealing, the averaged curve over the peaked $S(r)$ curves (see Figure 3.6) demonstrates the method's usefulness in coordination with the model proposed by Cairo et al. (2006). Figures 3.6A and 3.6B both provide a population view of the macroheterogeneity structure of LFA-1 proteins in different data sets. Averaging over all the peaked $S(r)$ curves allows us to view the overall effect of PMA activation on LFA-1 protein receptors. For TS1/18 labelled LFA-1, in addition to the overall increase in mobility with activation based on the CRW test (discussd in section 3.2), we observe a shift to the right of the average variance FPT peak position upon activation. This shift is indicative of an increase in the relative size of confinement zone, probably

as a result of the decreased cytoskeletal interaction by the intermediate state of LFA-1 as described by Cairo et al. (2006) in Figure 3.1. In addition to the decreased mobility observed (discussed in section 3.2) for MEM148, there is a large increase in trajectories with peaks in $S(r)$ in the 50 nm range. In relation to the model in Figure 3.1, this provides evidence for cytoskeletal binding, and provides a rough estimate for the spatial region of diffusion when LFA-1 proteins undergo cytoskeletal adherence or clustering behaviour.

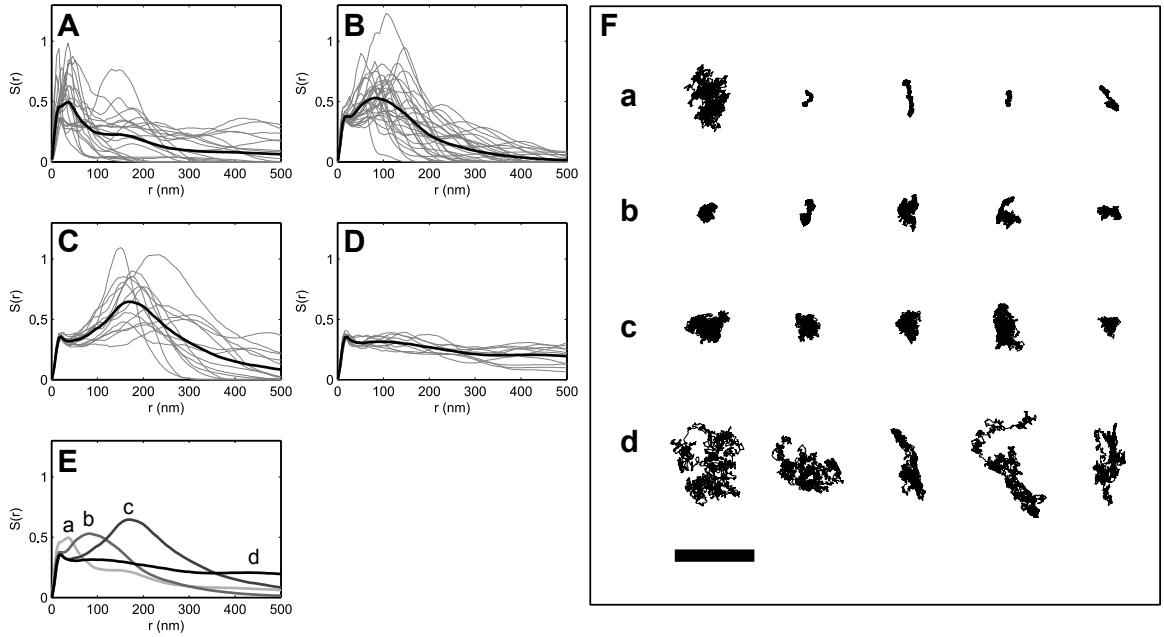
3.3.2 Estimating Relative Size of Confinement Zones

LFA-1 activation has been proposed to be a product of recruitment or 'clustering' of multiple receptors within a spatial region (Cairo and Golan, 2007). However, there is a lack of experimental evidence for the formation of these clusters on activated cells (Cambi et al., 2006).

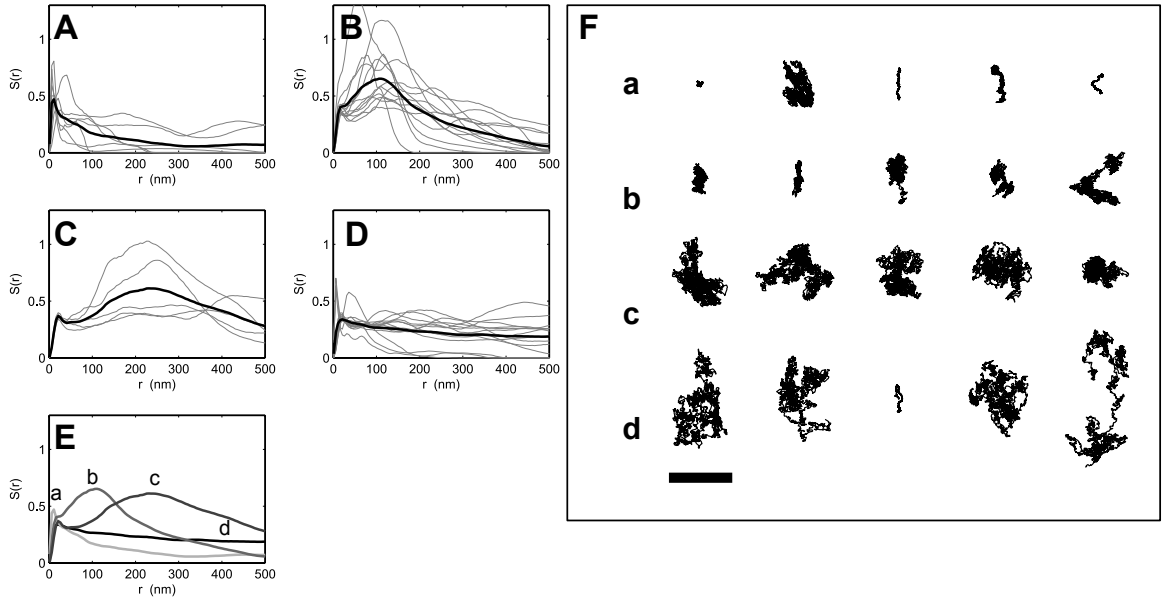
Motivated by the observation that monocytes readily bind to cells with ICAM-1, while dendritic cells do not, Cambi et al. (2006) explored the membrane-receptor organization of LFA-1 on these cells via Transmission Electron Microscopy (TEM). As a result, they found three levels of avidity for LFA-1 proteins: (i) randomly distributed inactive molecules, (ii) ligand-independent nanometer sized clusters, and (iii) ligand-triggered micrometer sized clusters. By using a nearest neighbor algorithm, Cambi et al. (2006) divided distances into three different classes, 0-50 nm, 50-100 nm and >100 nm, where clusters were defined when two proteins were within 50 nm of each other. Cluster sizes were observed to be much larger in size (with increased occurrence) for monocytes, than for dendritic cells.

In an alternate experiment, near-field scanning optical microscopy (NSOM) was used to identify cluster domain sizes of LFA-1 within a resolution of 30 $\pm$ 6 nm (van Zanten et al., 2010). The results of this study are shown in Figure 3.7. A bimodal distribution of domain size was created based on 52 individually observed fluorescent spots, with the main distribution centered at 72 $\pm$ 21 nm, and a smaller distribution around 130 nm. These studies lend support to the hypothesis that spatial configuration and receptor density plays a role in LFA-1 activation.

The size of confinement zones observed via the variance FPT method by the position of the peak in $S(r)$ have been predicted within multiple different ranges (see Figures 3.4 and 3.5). In particular, we observe that MEM148 labeled LFA-1, upon activation have a large population of confined trajectories in the 50 nm range. Although this measurement does not necessarily describe clustering, it describes the spatial scale at which single particles

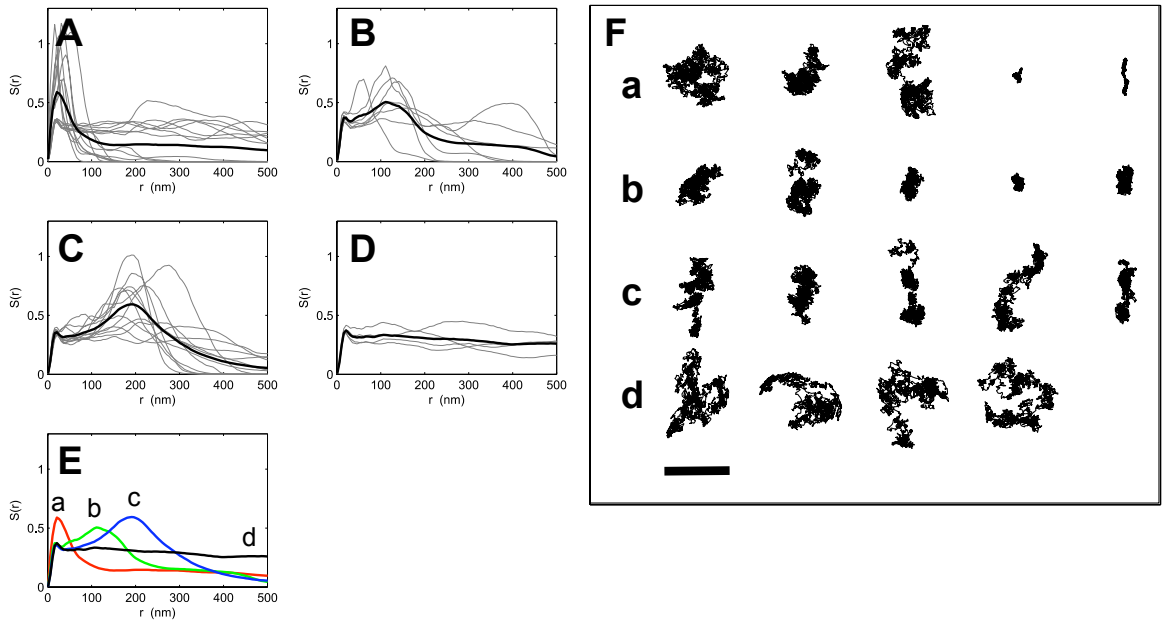


(a)

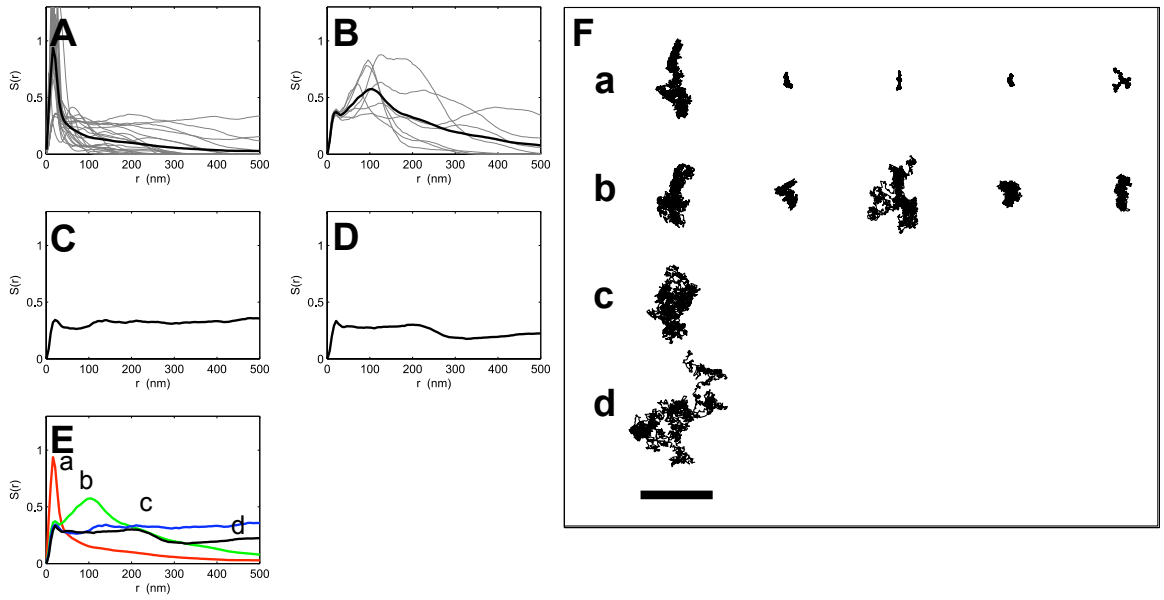


(b)

Figure 3.4: Variance FPT analysis for TS1/18 labeled LFA-1, (a) control and (b) PMA treated. Each $S(r)$ curve was categorized into one of four groupings based on peak location (detected using Billauer (2008)): (A) peaks in the $S(r)$ curve that occurred in the 0 - 50 nm range, (B) peaks in the $S(r)$ curve that occurred in the 50 - 150 nm range, (C) peaks that occurred > 150 nm or (D) trajectories that were not rejected as a CRW (no peak). (E) The average $S(r)$ curves for each category are overlaid for viewing. (F) Sample trajectories from each category. The scale bar is 1 μm.



(a)



(b)

Figure 3.5: Variance FPT analysis for MEM148 labeled LFA-1, (a) control and (b) PMA treated. Each $S(r)$ curve was categorized into one of four groupings based on peak location (detected using Billauer (2008)): (A) peaks in the $S(r)$ curve that occurred in the 0 - 50 nm range, (B) peaks in the $S(r)$ curve that occurred in the 50 - 150 nm range, (C) peaks that occurred > 150 nm or (D) trajectories that were not rejected as a CRW (no peak). (E) The average $S(r)$ curves for each category are overlaid for viewing. (F) Sample trajectories from each category. The scale bar is 1 μm.

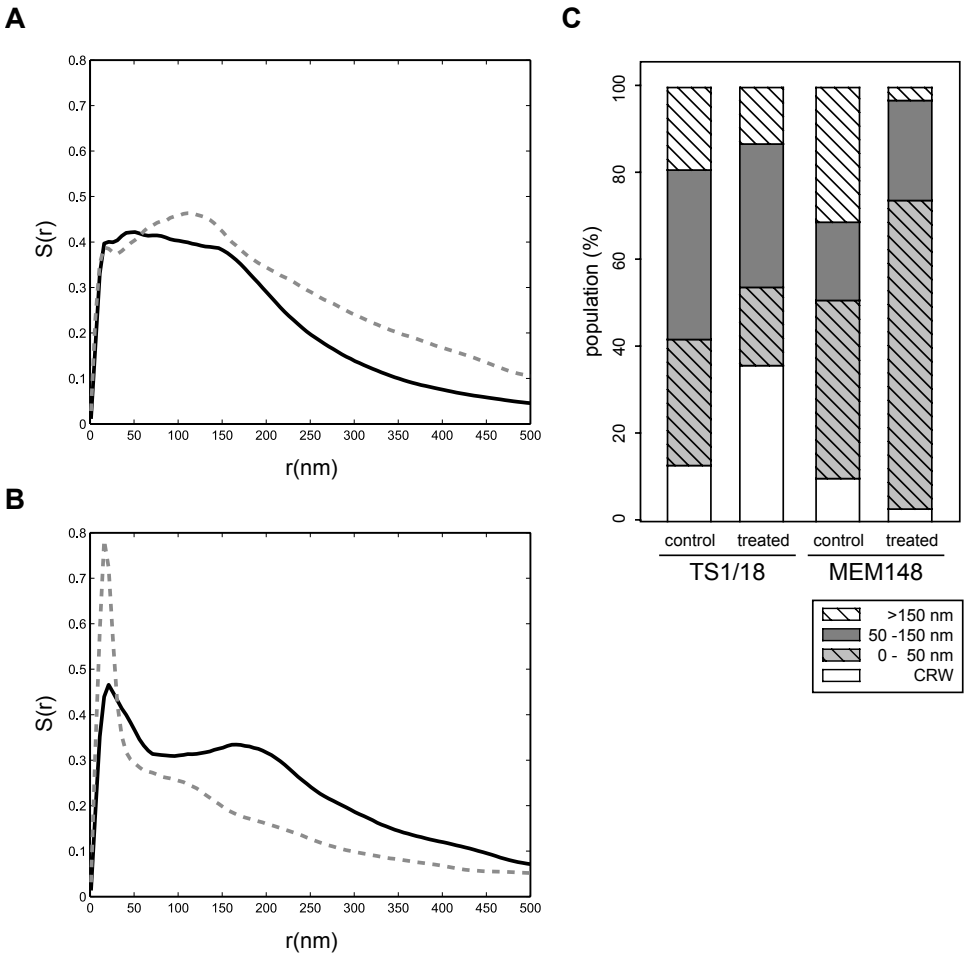


Figure 3.6: Variance FPT analysis for LFA-1 populations. Average $S(r)$ curves were calculated for all the trajectories in the TS1/18 labeled (A) and MEM148 labeled (B) data sets that were rejected by the CRW model. The control data sets are depicted by the black solid line, while PMA activated data sets are shown by the dotted grey line. (C) Shows the overall percentages for the category classifications made in Figures 3.4 and 3.5.

experience the greatest heterogeneity. Given that the clusters as observed with TEM and NSOM are on spatial scales 20-60 nm, it is not surprising that the particle's spatial scale as measured with variance FPT is within the same range. Thus, our analysis supports nanoscale diffusion changes within the range of the estimated cluster size, especially in PMA-induced cell activation. Specifically, the variance FPT analysis agrees with the findings of van Zanten et al. (2010), that the majority of cluster sizes lie in the range of 20-60 nm with a smaller proportion of larger scale clusters.

This pinnacle of our findings admits an important role for variance FPT, as it is a simple analysis that provides macroheterogeneity filtering as well as supplies microheterogeneity information, regarding cluster and confinement size.

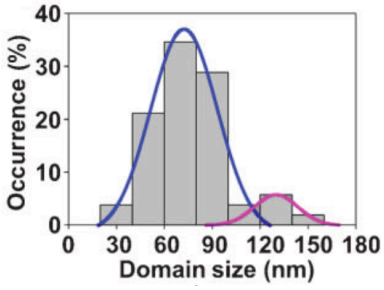


Figure 3.7: The occurences of cluster sizes of LFA-1 proteins as determined by van Zanten et al. (2010) using NSOM. The main distribution is centered at 72 ± 21 nm, and the smaller distribution around 130 nm. (Figure from van Zanten et al. (2010)).

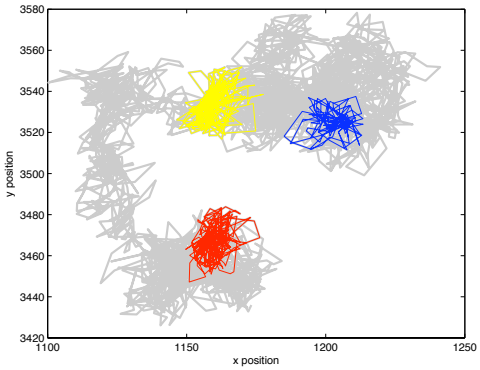
3.4 Trajectory Colouring

We believe that with further research and numerical analysis, the FPT curve can tell us much more about the microheterogeneity of diffusing particles. Since peaks in $S(r)$ supply an important parameter regarding spatial scale, we can review the discrete FPT distribution at the spatial scale of the peak maximum r, to discern areas of confinement, or anomalous diffusion.

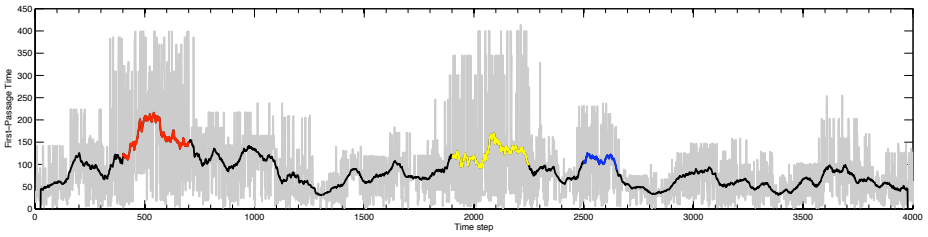
A quick exploration of a discrete FPT distribution is shown in Figure 3.8. Areas of high variability and high FPT in relation to other areas of the distribution were selected by eye and coloured. The timespan of colouring was also identified within the trajectory plot. We found that specific areas of the discrete distribution (those with a sustained high FPT) correlated heavily with diffusion constrained to a small spatial region. Although tools to automatically identify these regions need to be developed, we see a great potential for analyzing anomalous diffusion on an individual scale to estimate parameters such as time spent, frequency and functional response to clustering or binding to the cytoskeleton. These fundamental parameters have not been easily estimated from other methods.

3.5 Evaluation of Methods and Results

In this chapter, we have demonstrated the utility of the CRW model and variances FPT as ecological approaches to SPT data analysis. The CRW model was useful in filtering trajectories which were limited to short term or no correlation between steps. However, it was not useful in being able to detect more complicated movement patterns in behaviour. This became much more concerning when we observed that there was a large majority of trajectories that individually failed the CRW model. As a preliminary step for filtering het-



(a)



(b)

Figure 3.8: Trajectory colouring for trajectory 22 from MEM148 labelled LFA-1 data (control). (a) The trajectory plot with coloured regions chosen from areas of sustained high FPT. (b) The FPT vs time curve with an overlying average curve (calculated with a window of 50 points). The FPT was calculated with $r = 21$ nm, selected based on the location of the $S(r)$ peak of the trajectory. Regions of high FPT were chosen and coloured by visual inspection.

erogeneity, however, the CRW model supplied us with two groups of trajectories, described as *dispersive* (for those accepted as a CRW) and *non-dispersive* (those that are rejected as a CRW), which reflected the results based on diffusion profiling found in Cairo et al. (2006).

In diffusion profiling (discussed in section 2.3.4), there are a number of parameters that need to be correctly defined, thus making the method purely qualitative and difficult to compare across different lab environments, where conventions might be different. For example, varying definitions of D_m and D_M can cause large discrepancies between different data sets or experiments. Also, in the case of detecting multiple populations of closely positioned diffusion coefficient distributions (D_{slow} and D_{fast}), it can be difficult to determine in which population certain proteins lie. While diffusion profiling has been very successful in proposing biological models (Cairo et al., 2006; Saxton, 1997), it seems to fall short of being able to accurately describe heterogeneity. On the other hand, the CRW model has a larger underlying theoretical framework in which trajectories are explicitly separated based on their diffusion process. Trajectories can not only be distinguished as CRWs based on

their MSD curve positioning within the pseudotrajectory envelope, but the confidence of accepting a trajectory or population of trajectories can be quantitatively determined by the relative distance of the MSD curve to the edges of the envelope. As a result, we can accurately characterize individual trajectories, not only based on their qualitative grouping in relation to the single parameter D_m, but based on angle and length correlations throughout the entire trajectory. As such, important information about anomalous diffusion that would not otherwise be noted, is included in the analysis. Trajectories that pass the CRW bootstrapping test (either as individual trajectories or as an entire population) contain further information about the diffusion process via their turning angle and length distributions. Trajectories that fail the test, however, give a strong indication for anomalous diffusion that would not be detected in diffusion profiling.

As previously discussed, the CRW test is only really useful as a first step in analysis for SPT data. For our particular SPT LFA-1 data sets, the CRW model fails for the majority of the trajectories. This has property has also been noted in other data sets (Cairo CW, U of A, pers. comm.). Whether this is due to more complex diffusion processes, or merely due to the viscous and obstructed environment of the plasma membrane, is not known; further analysis is required to assess microheterogeneity.

Variance FPT analysis seems to pick up where CRW leaves off. Through a simple and intuitive analysis, we can determine the spatial scale at which heterogeneity occurs, post data collection. This spatial scale is useful in a number of ways, providing insight into the level and relative location of concentrated diffusion throughout an individual trajectory. The location of peaks in the variance of FPT curve ($S(r)$) provide confinement size estimates accurate within an order of magnitude. For a long enough SPT track length, the specificity of this type of information has only previously been estimated by complex experimental procedures (Cambi et al., 2006; van Zanten et al., 2010).

However, the novelty of this method has left a large number of unanswered questions. The structure of $S(r)$ curves has not been intensely explored in relation to different parameters (r_c, p_o, p_i) or to different types of diffusion processes and provides opportunities for further work.

4 Errors in Movement Data [4]

Error in movement data is a research topic which has increased in popularity with the development of new techniques and methods for data acquisition. Lab experiments and field studies involve methods that rely on the measurement of position, state, or location of an individual, which are translated via observational measurement into data. It is in this translation that errors and biases are likely to occur from a variety of different causes.

The importance of being aware of error in data is largely due to the impact that error can have on movement measurements, causing repercussions that can affect the validity of derived biological or mathematical models that are used in resource, conservation, epidemiological or pharmaceutical planning. From an ethical standpoint, errors could also lead to incorrect conclusions about cellular and physiological functions, spreading inaccurate information throughout the scientific community. For this reason, the fear that small errors in data could cause false conclusions and large miscommunications, error is no longer an unwanted byproduct, but a commonly studied phenomenon.

In general, measurement error can often be described as either *random* or *systematic* error (Taylor, 1997). Random error is defined as the reduction in precision of a single measurement. It can cause natural variation around a true value, so that the mean of repeated measurements converges to that true value. For example, position error, or uncertainty in the measurement of a single position due to noise is a common form of random error. Often, statistical methods can be used to reliably estimate the structure of random error (Taylor, 1997). Systematic error, on the other hand, can be defined as the reduction in accuracy of a measurement. Thus, it is much more difficult to manage and identify. Rather than causing variation around a single value, systematic error causes a non-random deviation in all the estimated values of a true value, shifting the entire mean, thereby making the true value unattainable. Since repeated measurements would also undergo the same amount of systematic error, the true location cannot be identified by repeated measurements. Thus, systematic error can lead to much more serious consequences than random error. While random error naturally occurs due to experimental or environmental noise, systematic error is usually a product of improper experimental design, incorrect sampling (Turchin, 1998), or biases involving missed data points (Frair et al., 2004). As a result, the structure of

[4] Portions of this chapter were submitted in participation for a seminar course Biol 633: *Advanced Techniques in Biology* taught by Dr. Mark Lewis and Dr. Evelyn Merrill at the University of Alberta, in the Fall semester of 2009.

systematic error can only be understood by evaluating measurement methods, technological factors, and understanding the variables and specific nature of the data set.

Throughout this project, a common theme has been the link between models pertaining to ecology and cellular biology. In this chapter, we discuss how error in movement data, present in both GPS and SPT data, contributes heavily to this theme. Specifically, we address movement error in relation to *positional error*, or error associated with measuring a particular location (section 4.2), and *sampling error*, or false impressions created by using the incorrect sampling interval to observe a movement track (section 4.3).

4.1 Causes of Error in Movement Data

We begin by discussing the context of movement error in terms of its causes. Due to their similarity in form, the consequences of error in GPS and SPT data can be comparable. However, the congruence in error causation within these environments is not as obvious. In the next three sections, we discuss the sources of error that are shared by both types of experiments; behavioural, environmental and human/technological.

4.1.1 Behavioural Causes of Error

The complex biological factors affecting the movement of individuals result in a variety of sources of measurement error in GPS and SPT data. Often, the source of this error can be attributed to the collared or labelled individual itself.

In ecological situations, animal behaviour can often disrupt the clarity of data acquisition from the GPS collars or radio tags leading to missing data points. These missing point biases have been documented in countless studies (D'Eon et al., 2002). Specifically, the precise orientation of a collar or tracker on an animal can manipulate the accuracy or precision of a true position. For example, foraging black bears often have a reduced GPS fix rate (chance of successfully obtaining a position by satellite) due to the orientation of their GPS collars while digging (Obbard et al., 1998), resulting in missing data points. In other cases, moose and deer have been known to cause reduced fixed rates while in bedding position (Moen et al., 1996; Bowman et al., 2000). Even animal movement speed can modify the amount of error affecting movement data. It has been observed that animals, such as deer, that roam more freely with higher motility have less position error than slower moving animals (Bowman et al., 2000). Reasons for this particular error are related to sampling

rates, and will be discussed later in section 4.3.

Funnily enough, many of these examples are very similar to situations that can occur in SPT data. Protein receptors undergo a number of different diffusion behaviours that can affect a clear SPT signal. Analogous to animal bedding or foraging, receptors can undergo anomolously slow diffusion. If the slow diffusion is on the scale of the position error, position error can largely affect measurements. Various types of confined motion such as transient confinement zones, or obstructed or restricted motion due to cytoskeletal adherence or obstacles in the plasma membrane can yield anomalously slow diffusion.

These behavioural challenges can cause a variety of problems in estimating relationships to spatial features. For example, if lost fixes and high position error are associated closely with bear foraging or deer bedding, there would be large biases in estimating the importance of resources patches or habitats associated with these behaviours (Bowman et al., 2000). In SPT, one might see similar problems for slowly moving protein receptors with high position error. It would be difficult to understand the specific spatial aspects of confined diffusion, e.g., confinement size or location, for a given receptor if position error was very high.

4.1.2 Environmental Causes of Error

Environmental components, external to animal or receptor behaviour, can also cause large error problems. Due to the widespread use of GPS for a variety of different animals in very diverse and heterogeneous environments, the effects of terrain and canopy cover have been studied thoroughly. D'Eon et al. (2002) showed that studies done in steep mountainous terrain with mature coniferous forests admitted higher GPS error than studies done in open clearings. It was also shown that the combined effects of terrain and canopy cover could lead to a more serious error problem than in areas with only one of the two factors (D'Eon et al., 2002). Without doing an environmental study of the organism, it is difficult to understand changes in diffusion in behaviour which could be classified as erroneous. However, modern studies involving spatial geographic information systems (GIS) allow the overlay of a movement path with landscape features such as hills, mountains and various types of forests, to supplement the knowledge of GPS error in these locations.

The cellular environment can also cause different types of error and biases. The plasma membrane for example, is hardly an ideal medium for diffusion. It is full of twists and turns, involving strange naturally occurring corrals or craters that greatly modify the

diffusion behaviour of the protein. Without a way to capture an image of the surface of the cell (akin to GIS), we have no way of determining the cause of these changes in diffusion, and often these create a bias in determining a biological model for receptor behaviour. It is for this reason that synthetic membranes admit higher rates of diffusion than naturally occurring plasma membranes (Cairo, pers. comm.). In addition, the effect of surface roughness on SPT has been noted (Hall, 2008). Other instances of environmental error include the loss of SPT labels as protein receptors diffuse to the edge of the membrane only to continue diffusing on the other side of the cell, out of our field of view. This can result in smaller length tracks which, from previous discussion, can reduce the ability to detect confinement zones, or can result in a loss of precision when detecting a diffusion coefficient D from the MSD curve.

4.1.3 Human and Technical Causes of Error

Apart from the causes of error closely associated with the moving individual in question, there are a number of sources of error that are caused directly by the methods of observation, or biases that are directly a result of human interaction. In ecology, the malfunctioning of GPS collars (Bowman et al., 2000) or short-lived battery life of GPS loggers (Ryan et al., 2004) can cause missed point biases in movement data. Even until May 2000, the accuracy of GPS had been intentionally degraded by the policy of Selective Availability, enforced by the US Department of Defense, to prevent potential enemy use.

Among the different types of technical and microscopy sources of error, a large source of unexplored bias in single-particle tracking exists in the selection and tagging of individual protein receptors. Often faster particles are chosen based on their visibility (Cairo CW, U of A, pers. comm.), biasing the macroheterogeneity by underestimating the presence of a slower receptor population. Other errors and biases in SPT analysis are due to the effects of SPT labels themselves, whether they are colloidal gold, latex beads or fluorescent particles (Saxton and Jacobson, 1997). Issues include drag from interaction with the extra-cellular matrix and cross-linked binding sites due to multivalent labels (Saxton and Jacobson, 1997). Cross-linking often reduces the diffusion rate and causes hindrance for moving through spatial features (Saxton and Jacobson, 1997). Problems with the loss of a clear SPT signal using green fluorescent protein labels has also been noted (Dushek O, U of Oxford, pers. comm.).

In the specific case of the data analyzed in this thesis, the size difference between

the LFA-1 protein and polystyrene bead can also be a potential source of position error. While the extracellular portion of LFA-1 is only approximately 15 nm (across) x 21 nm (high) (van der Merwe et al., 2000), the label is substantially larger at 1 μm (Cairo et al., 2006). Thus, it would reasonable to assume the presence of error attributed to detecting small movements (at a resolution of tens of nanometers) of the center of the polystyrene bead.

4.2 The Effects of Position Error

Position error in movement data affects each position throughout a movement trajectory with a slight deviation. This may not seem to change the overall demeanor of a movement track, however, it can cause large deviations in the calculation of movement metrics used in the parameterization of mathematical models. Due to the widespread use of CRW models (Turchin, 1998; Kareiva and Shigesada, 1983), the effects of error on step length and turning angle distributions have been thoroughly explored (Hurford, 2009; Jerde and Visscher, 2005).

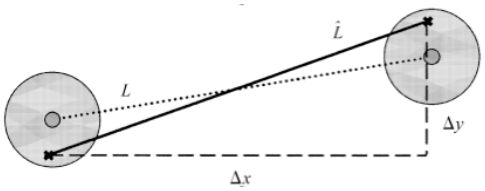


Figure 4.1: The effect of error on step lengths. In particular, for small step lengths, position error may cause an overestimation or underestimation of the true step length (Jerde and Visscher, 2005). In this figure, a measured step length $\hat{L}$ is shown as an overestimation in relation to the true step length L. (Figure from Jerde and Visscher (2005)).

Position error can cause overestimation or underestimation in step lengths (Jerde and Visscher, 2005) based on the extent of error at each location. Figure 4.1 illustrates the discrepancies that can occur, specifically for overestimation. For small step lengths, the error distributions can overlap between locations, allowing for a larger chance for overestimation. As a result, for stationary or slow moving animals (Jerde and Visscher, 2005; Ryan et al., 2004), overestimation can be a serious problem.

For small step lengths within the magnitude of the position error, turning angle calculations can also undergo large quantities of inconsistency. Hurford (2009) observed that turning angle measurements for small step lengths are measured as spurious 180 degree turns, which might imply that an animal is doing a large quantity of direction reversal commonly characterized in foraging or encamped behaviour (Morales et al., 2004), or in terms of

protein receptors, undergoing confined diffusion, clustering, or diffusion with adherence to the cytoskeleton.

In view of the effects of error on step lengths and turning angles and our use of these distributions in testing for CRWs, the existence of error is an important issue in the analysis of SPT data of LFA-1. The turning angle and length distributions for the SPT data of protein receptors exhibit features which could be interpreted as products of position error and sampling errors. It should be noted that the turning angle distributions of LFA-1 data (see Figure 3.2) contain two distinct peaks; a peak at 0 or 2π, indicative of persistence (or false persistence caused by oversampling) and a peak at π, which could be spurious turning angles caused by high levels of position error (Hurford, 2009). After our implementation of the variance FPT analysis however, it could be understood that turning angles with a peak around π are due to a process of confined diffusion. The element of persistence in the turning angle distributions (peak at 0 or 2π) may be an artifact of oversampling, though it may be that proteins naturally show some small degree of persistence in their movement, which may be of interest to researchers in receptor membrane interactions. In addition, the length distributions look to be very skewed from a normal distribution, as if smaller lengths are less detected, which could be due to the fact that smaller lengths are overestimated in the presence of error. However, it very well could be that the smaller step lengths are not observed due to the minimum length size observable by microscopy.

The issue of strange peaks in the angle distributions was somewhat alleviated when we inspected the turning angle distributions of individual trajectories rather than the entire population. Results are shown in Figure 4.2. We found that the angle distributions for trajectories experienced peaks only at π or 2π (Fig. 4.2(a) and (b)), while other distributions had neither (Fig. 4.2(c)) or both peaks (Fig. 4.2(d)) in the distribution. Due to the variety of characteristics for individual distributions, we were satisfied with the existence of macroheterogeneity across distributions. This provided context and continued justification for our work; to better understand the large variety of movement behaviours in LFA-1 data.

Measuring the structure of position error can provide the ability to detect and quantify its effects on movement data. In section 4.2.1, we will discuss how the structure of position error can be characterized. In section 4.2.2, we will apply this characterization to the better understand the role that position error can play in SPT data.

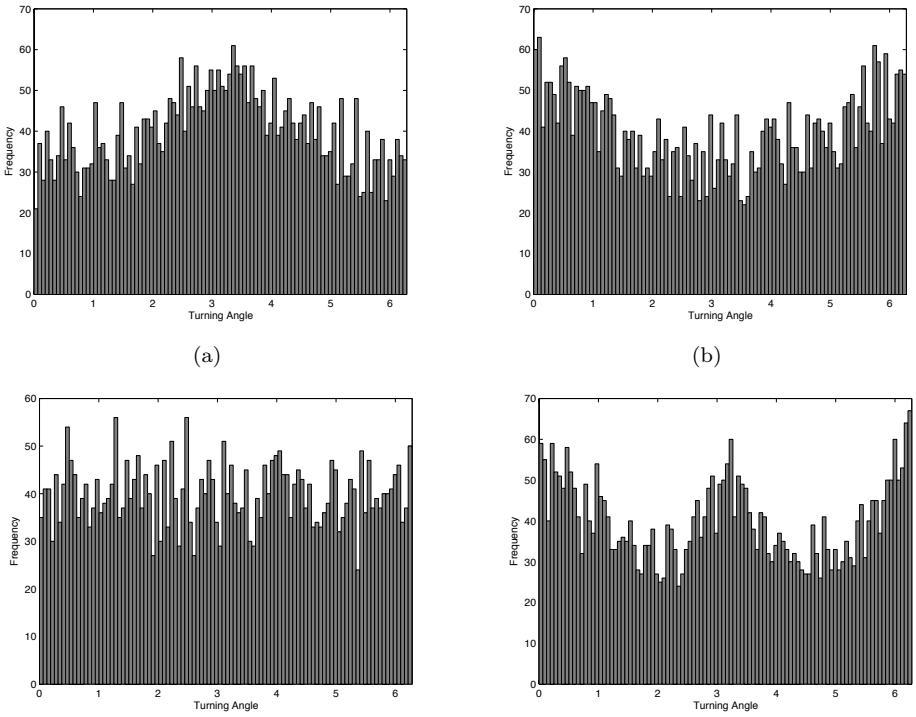


Figure 4.2: Turning angle distributions for individual trajectories. Angles are measured in radians, as described in Appendix A.2. (a) Distribution with a peak at π; (b) Distribution with a peak at 2π; (c) Distribution with a peak at both π and 2π (d) Distribution with no peak.

4.2.1 Describing the Structure of Position Error

After the reluctant admission of the possibility of the existence of position error in movement data, often the next step is to attempt to quantify the error and to describe structure of error in movement data. Since this has been realized in ecology for many years, a number of ways to describe error structure have been established for GPS technology that can be adapted to SPT data.

A common technique involved in measuring the structure of position error, is by measuring the distribution of measured points around a stationary transmitter (GPS collar or optical bead). A mathematically convenient way to categorize or quantify position error is by fitting a suitable probability density function to this error distribution. Common distributions used to fit the spread of error positions are the Normal distribution,

$$f(r) = \frac{r}{\sigma^2} e^{\frac{-r^2}{\sigma^2}} \tag{4.1}$$

Laplace distribution,

$$f(r) = \frac{r}{\beta^2} e^{\frac{-r}{\beta}} \tag{4.2}$$

or Bessel distribution,

$$f(r) = r\rho^2 K_0(\rho r) \tag{4.3}$$

where $f(r)$ is the probability of finding a point at a distance r from the true value, K_0 is a modified Bessel function of the second kind, and σ, β and ρ are estimated from maximum likelihood methods (Hurford, 2009). The usefulness of this method is that it grants the ability to generate randomly distributed position errors, so that they can be added to simulated movement paths and can be studied more rigorously.

4.2.2 Position Error in SPT Data

To explore the effect of position error in the SPT environment, we used the calculation of MSD to understand, given a fixed magnitude of error, the spectrum of diffusion coefficients that would be most vulnerable to spurious peaks at π in the turning angle distribution. From a fixed protein label, we first connected the measured locations in the order which they were recovered (separated by a time step), and calculated the MSD based on the method by Qian et al. (1991), discussed in section 2.3.2. As a result, we obtained an MSD curve, from which we estimated a diffusion coefficient, D_{error}. From this diffusion coefficient, using formula

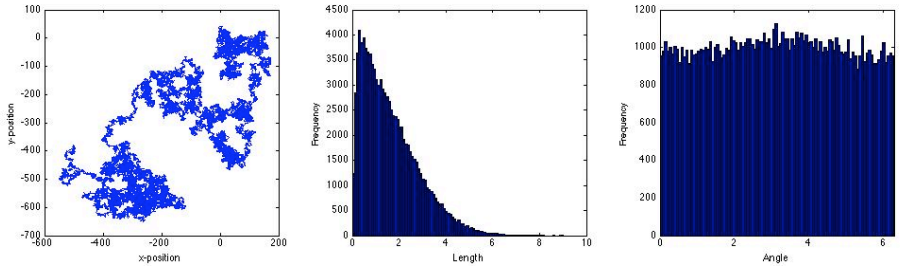
2.22, we were able to generate a distribution of "error steps" that were added to each step of a random walk simulated by a diffusion coefficient D_{sim}. Then, each step of the original random walk was resampled to become the new measured position.

We ran simulations for different magnitudes of the diffusion coefficient and observed an emergence in a central peak of the turning angle distribution, illustrating the possible magnitude of error that might cause bias in estimating model parameters from the distribution. The results of this simulation are shown in Figure 4.3. We show the trajectory plot, turning angle distribution and length distribution for trajectories with varying magnitudes of error represented by the ratio $\frac{D_{error}}{D_{sim}}$. We see that as we increase the amount of position error in the simulated trajectory, the peak at π is more pronounced. This verified, through an alternate means than described by Hurford (2009), that spurious turning angle peaks could be caused by position error. However, it was difficult to determine if this was the case for the LFA-1 data sets.

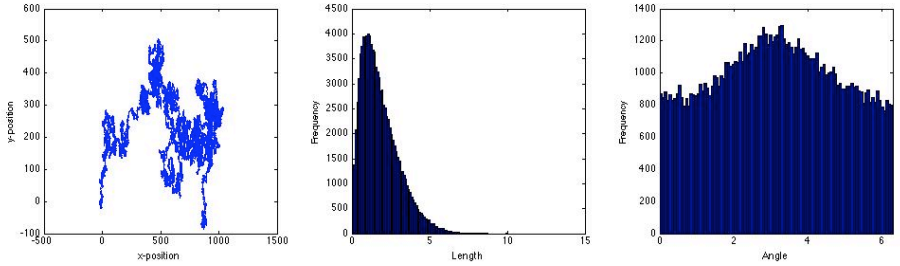
4.3 Sampling Error

A common feature of all types of movement lies in the resolution of the movement path, which is usually set by the sampling interval of GPS or SPT measurements. The length of a sampling interval is usually determined by a number of factors. The shortest length of a sampling interval is limited by technological ability, i.e., the transmission speed of a GPS collar, or the shutter speed of the camera on a microscope. Although there is no real limitation to the longest sampling interval, the interval must be short enough to capture meaningful movement patterns. It is with this dichotomy that experimentalists are concerned, as choosing a correct sampling interval can be challenging. If not appropriately chosen, an incorrect sampling rate can result in biases due to sampling error.

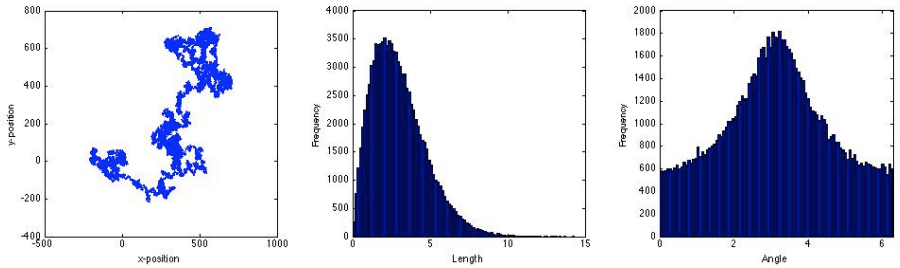
A high sampling rate, though it might seem preferable, can cause an autocorrelation in step lengths and can give rise to a spurious turning angle peaks at 0 degrees (Turchin, 1998), giving an impression of false persistence. In ecology, this type of oversampling can cause a large underestimation in home range patterns (Swihart and Slade, 1985). However, Moorcroft and Lewis (2006) proposed that this can be easily corrected by sampling data at longer intervals until positive correlation is removed. Although this may seem like a good idea, especially in studies where persistence is deemed not important for the biological questions being asked, it can cause many problems for studies where measuring possible



(a)



(b)



(c)

Figure 4.3: The emergence of a spurious peak at π with an increased magnitude of position error in simulated trajectories. Trajectories were simulated with a uniform turning angle distribution and normal length distribution derived from a diffusion coefficient (D_{sim}) as described in section 2.2. Error was then added to each position, with a deviation selected from a normal distribution generated by a diffusion coefficient at a different magnitude (D_{error}). The trajectory plots, length distributions and angle distributions are shown for different magnitudes of error, (a) $\frac{D_{error}}{D_{sim}} = 0.001$, (b) $\frac{D_{error}}{D_{sim}} = 0.01$, (c) $\frac{D_{error}}{D_{sim}} = 0.1$.

persistence can be deemed a point of interest, as in SPT receptor movement. While animals can be assumed to have a degree of persistence, the ability of receptors to under go persistent movement is still debatable. Therefore, we cannot discount persistent peaks in the angle distribution (around 0 or 2π) as artifactual or not important. In addition, for very small step lengths or slow moving individuals, a high sampling rate can amplify the effect of GPS or SPT position error, increasing the rate and quantity of false readings.

On the other hand, a low sampling rate, yielding longer sampling intervals, can also result in biases from sampling error. For highly tortuous movement tracks, for example, a low sampling rate can underestimate the distance traveled (Ryan et al., 2004). For these reasons, Jerde and Visscher (2005) suggest that the choice of sampling interval should be based on the motility of the individual, to prevent errors or biases in model parameterization. While this may seem reasonable for animal studies, where observations about animal motility can be made based on environment and resource surveys, there is still much debate surrounding the appropriate sampling interval for different protein receptors. As a result, a common practice in SPT has become to sample as much as possible (within the range of the technological equipment) such that any lower sampling rate can be chosen for analysis. But the question remains, which sampling rate allows for the best interpretation.

The issue of oversampling was explored in LFA-1 SPT data, by decreasing the sampling rate so that less position points were used in bootstrapping procedures. Although the turning angle peak at 0 disappeared after the data was resampled by taking only every 2^{nd} step, the CRW model was still rejected (see Figure 4.4). In fact, resampling the trajectories by 100 and even 500 steps still resulted in a model rejection.

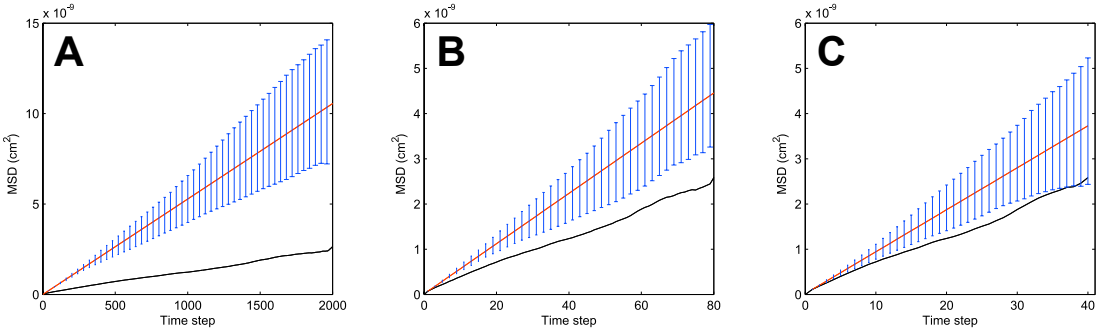


Figure 4.4: CRW test on resampled data. Despite the resampling of the TS1/18 labeled LFA-1 data set by (A) 2 steps (B) 50 steps and (C) 100 steps, the CRW test was still rejected. Each oversampled data set was tested with 500 groups of pseudotrajectories, forming the pseudotrajectory envelope (blue). The observed MSD of the data (black) is significantly different from the expected MSD of the model (red) in all cases.

4.4 Summary of Error in LFA-1 SPT Data

In this section, we summarize our investigation of the existence of position and sampling error in LFA-1 data. The difficult identification and removal of systematic error rely on a better understanding of SPT and the cellular environment. Although there are no direct methods derived from data analysis to deal with this type of error, we propose that improved experimental techniques should reveal previously unnoticed factors causing systematic error.

In section 4.2, we identified some characteristics of the LFA-1 turning angle distributions that could imply position and sampling error. In particular, we identified the existence of turning angle peaks at π which we verified can be caused by position error (Hurford, 2009). However, when we looked at the turning angle distributions of individual trajectories, we noted that despite possible position error, the richness of heterogeneity still existed. In addition, we discovered the possibility of confined diffusion, which could also be a reason for turning angle peaks at π. The peak at 0 or 2π could signify that the data is being oversampled. However, the resampling of data points for the TS118 labeled LFA-1 (control) by every two steps reduced the data to half the size without yielding any new information. Even undersampling by 100 steps resulted in CRW model rejection.

To counteract the high elements of error in movement data, a number of simple filtering methods have been developed to reduce bias in conclusions made by mathematical modeling. Since small step lengths are associated with spurious turning angles, methods have been developed to filter step lengths below a certain threshold (Hurford, 2009). However, the overall result is a net loss in movement information. Removing pairs of points from a data set can be costly. Where precision is increased by removing inaccurate measurements, losing data can be also detrimental to modeling. In the case of LFA-1 SPT data, where the mechanism of movement is still widely unknown, we decided against pursuing data filtering. In addition, removing small step lengths would be detrimental to any diffusion process involving the hypothesized "stationary" or "adhered state" for protein receptors. As a result, we concluded that the filtering of data points was not worth the loss of information and lack of benefits received by doing so.

However, once more information can be specifically observed about the cellular environment and receptor behaviour, we propose that there could be other ways to appropriately manage different types of error in SPT data. On the ecological front, studies on elephant seals are prone to errors based on diving and ocean surface temperature gradients. However,

new bootstrapping procedures have been developed that are based on a variety of characteristics such as animal speed, direction, and landscape features (Tremblay et al., 2009). As a result, these methods are used to more accurately recover movement paths. With more technological improvements and better developed data acquisition techniques, it is possible that additional helpful data (like cell surface characteristics) could be collected for protein receptors and used for more accurately acquiring SPT trajectories.

5 Discussion

In this thesis, we concerned ourselves with the analysis and interpretation of SPT movement data of LFA-1 integrin protein receptors. The regulation of LFA-1 integrin plays an essential role in the formation of the immune synapse and triggering the immune response. In view of the observation that the formation of LFA-1 clusters is closely linked to LFA-1 binding (van Kooyk and Figdor, 2000), many studies have focused on the diffusion properties of LFA-1, to discern different molecular conformations and affinities (Cairo et al., 2006). Due to the rich macro and microheterogeneity in LFA-1 movement processes (Cairo et al., 2006; Das et al., 2009), we applied ecologically driven mathematical models to detect and understand different types of movement mechanisms in LFA-1 data, and used this understanding to build on the multi-state model of LFA-1. In section 5.1, we summarize our results and conclusions in the context of our objectives. Lastly, in section 5.2, we recommend future directions for research.

5.1 Summary and Conclusions

Understanding heterogeneity in movement data is key in building biological models of both ecological and cellular environments. It is important to realize that the role of mathematical movement models is not restricted to a single research area, but can be generalized as tools for different scenarios, both macroscopic and microscopic. In chapter 2, we described the general nature of mathematical movement models, and how current ecological and cellular techniques are rooted on the same simple diffusion process. Since previous work on SPT analysis mainly relied on the spread of diffusion coefficients (Saxton, 1997; Cairo et al., 2006), based on the assumption that each individual trajectory followed a simple random walk processes, we introduced more general techniques that allowed for persistence and confined diffusion. We discussed the utility of the CRW model developed by Patlak (1953), adapted for application by Kareiva and Shigesada (1983), as a first step in analysis to understand the underlying movement process. In addition, we discussed variance first-passage time, and how it can be used to detect and estimate the size of transient confinement zones.

In Chapter 3, we built on work done by Cairo et al. (2006) by developing a fresh approach to analyzing single-particle tracking data. By using models primarily utilized in ecology, we were able to view properties of LFA-1 diffusion that are not discernible by other mathematical methods. In particular, we used the CRW model to detect non-Brownian

motion and high degrees of autocorrelation, suggesting more complicated movement mechanisms. As a result, we were able to understand population structure based on diffusion mechanism rather than by diffusion rate. We found each data set could not be collectively described as following the CRW model, and that the majority of individual trajectories failed the CRW bootstrapping test due to high degrees of correlation in turning angles and step lengths. While trajectories that individually passed the CRW bootstrapping test are described as *dispersive*, those that failed the test are thought to be a product of microheterogeneity within individual trajectories.

This microheterogeneity was explored further using variance first-passage time, to detect areas of concentrated diffusion and provide evidence for transient confinement zones in LFA-1 data. Many individual trajectories were classified as having confined diffusion, and we were able to analyze population structure based on the size of confinement zones as determined by the location of peaks in $S(r)$. By viewing an increase in intensity of small confinement zones ($\leq$50 nm) following PMA activation in MEM148 labeled LFA-1, we conclude that this is possibly a result of cytoskeletal interactions inducing recruitment, or the formation of clusters. This supports the observations made of LFA-1 clustering (Cambi et al., 2006; van Zanten et al., 2010). The relative size of confinement zones as predicted by peaks in the variance FPT curves, $S(r)$, predict a similar spatial scale as size of clusters observed by microscopy (Cambi et al., 2006; van Zanten et al., 2010).

In Chapter 4, we led a discussion of the role of error and bias in movement data. Although we were not able to provide a solution for particular instances of error in SPT data, we were able to provide an overview of different types of error and common solutions, raising awareness of the universality of error and bias across ecological and cellular platforms.

In summary, this thesis demonstrates that some data analysis techniques from ecology can be applied fruitfully in the context of cell biology. As a result, we provide a novel approach into understanding macroheterogeneity and microheterogeneity of SPT data, and insights into the mechanism of LFA-1 diffusion.

5.2 Future Work

In view of our results, there are a variety of future directions for understanding SPT data, particularly for investigating the specific mechanism of LFA-1 movement.

For trajectories that were classified as CRWs, we recommend extensive analyses to be

done on the angle and length distributions of the individual movement tracks. This would enable further understanding of the dispersive class of LFA-1 proteins (Turchin, 1998). In addition, it is possible that more could be learned from the spread of parameters (such as the mean, variance, etc.) from the angle and length distributions.

We suggest that the use of variance FPT can be better adapted and refined for SPT. Parameters involved in $S(r)$ peaks such as peak height and width have not yet been explored, and could be important in understanding exact diffusion processes. For specific processes such as confined diffusion, it would be interesting to develop an analytical form of $S(r)$ and parameters of the diffusion process. Then, given a variance FPT curve, one could search parameter space for all the parameter sets that could fit that particular curve. This could help refine estimates of r_c, p_o and p_i (in the case of transient confinement zones).

Another approach would be to develop a colouring procedure more extensive than the one provided in Figure 3.8. This technique could provide new information about the relative location and time spent in confinement zones, by using numerical methods to spot time intervals of high variance in FPT. These parameters would enrich our knowledge of the LFA-1 protein mechanism and add to the model developed by Cairo et al. (2006). If effective, these techniques could be applied to other types of SPT data.

Since there are many sources of error and bias still at large in microscopy, we believe that further work can be done to better deal with to experimental error. While this has been extensively researched in terms of GPS error and ecological systems, we see a large motivation for study at the cellular level.

Lastly, our approach of using ecological methods to understand cellular phenomena is far from exhaustion. There are a large number of techniques used in ecological data analysis that could be applied at the cellular level. In addition, some instances in protein movement, while currently unique to the cellular environment, could result in insights at the ecological level. Due to the vastly different environments and varying methodologies, many new beneficial techniques are being developed in each of the individual research areas that could be shared between disciplines. There are helpful resources about movement modeling that have crossed disciplines (Codling et al., 2008), and we hope that this understanding can be adapted to future studies of movement.

Bibliography

H. Bailey and P. Thompson. Quantitative analysis of bottlenose dolphin movement patterns and their relationship with foraging. *Journal of Animal Ecology*, 75:456–465, 2006.

S. Benhamou. How many animals really do the Lévy walk? *Ecology*, 88(8):1962–1969, 2007.

H. C. Berg. *Random Walks in Biology*. Princeton University Press, 1983.

Eli Billauer, 2008. URL `http://www.billauer.co.il/`.

J. Bowman, C. Kochanny, and S. Demarais. Evaluation of a GPS collar for white-tailed deer. *Wildlife Society Bulletin*, 28(1):141–145, 2000.

C. W. Cairo and D. E. Golan. T cell adhesion mechanisms revealed by receptor lateral mobility. *Biopolymers*, 89(5):409–419, November 2007.

C. W. Cairo, R. Mirchev, and D. E. Golan. Cytoskeletal regulation couples LFA-1 conformational changes to receptor lateral mobility and clustering. *Immunity*, 25:297–308, 2006.

A. Cambi, B. Joosten, M. Koopman, F. de Lange, I. Beeren, R. Torensma, J. A. Fransen, M. Garcia-Parajo, F. N. van Leeuwen, and C. G. Figdor. Organization of the integrin LFA-1 in nanoclusters regulates its activity. *Molecular Biology of the Cell*, 17:4270–4281, October 2006.

E. A. Codling, M. J. Plank, and S. Benhamou. Random walk models in biology. *J. R. Soc. Interface*, 5:813–834, 2008.

G. Constantin, M. Majeed, C. Giagulli, L. Piccio, J. Y. Kim, E. C. Butcher, and C. Laudanna. Chemokines trigger immediate $\beta 2$ integrin affinity and mobility changes: Differential regulation and roles in lymphocyte arrest under flow. *Immnunity*, 13:759–769, December 2000.

R. Das, C. W. Cairo, and D. Coombs. A hidden markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. *PLOS Computational Biology*, 5:e1000556, 2009.

R. D'Eon, R. Serrouya, and G. Smith. GPS radiotelemetry error and bias in mountainous terrain. *Wildlife Society Bulletin*, 30(2):430–439, 2002.

C. Dietrich, B. Yang, T. Fujiwara, A. Kusumi, and K. Jacobson. Relationship of lipid rafts to transient confinement zones detected by single particle tracking. *Biophys. J.*, 82: 274–284, January 2002.

M. Edidin. The state of lipid rafts: From model membranes to cells. *Annual Review of Biophysics and Biomolecular Structure*, 32:257–283, 2003.

A. Edwards, R. Phillips, N. Watkins, M. Freeman, E. Murphy, V. Afanasyev, S. Buldyrev, M. da Luz, E. Raposo, H. Stanley, and G. Viswanathan. Revisiting lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165):1044–1048, 2007.

P. Fauchald and T. Tveraa. Using first-passage time in the analysis of area restricted search and habitat selection. *Ecology*, 84(2):282–288, 2003.

J. L. Frair, S. E. Nielsen, E. H. Merril, S.R. Lele, M. S. Boyce, R. H. M. Munro, G. B. Stenhouse, and H. L. Beyer. Removing GPS collar bias in habitat selection studies. *Journal of Applied Ecology*, 41:201–212, 2004.

D. Hall. Analysis and interpretation of two-dimensional single-particle tracking microscopy measurements: Effect of local surface roughness. *Analytical Biochemistry*, 377:24–32, 2008.

A. Hurford. GPS measurement error gives rise to spurious 180 ° turning angles and strong directional biases in animal movement data. *PLoS ONE*, 4(5):e5632, 2009.

C. Jerde and D. Visscher. GPS measurement error influences on movement model parameterization. *Ecological Applications*, 15(3):806–810, 2005.

A.R. Johnson, J.A. Wiens, B.T. Milne, and T.O. Crist. Animal movements and population dynamics in heterogeneous landscapes. *Landscape Ecology*, 7(1):63–75, 1992.

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

N. Meilhac, L. Le Guyader, L. Salomé, and N. Destainville. Detection of confinement and jumps in single-molecule membrane trajectories. *Physical Review E*, 73:011915, 2006.

R. Moen, J. Pastor, Y. Cohen, and C. Schwartz. Effects of moose movement and habitat use on GPS collar performance. *The Journal of Wildlife Management*, 60(3):659–668, 1996.

P. Moorcroft and M. Lewis. *Mechanistic Home Range Analysis*. Princeton University Press, 2006.

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

M. Obbard, B. Pond, and A. Perera. Preliminary evaluation of GPS collars for analysis of habitat use and activity patterns of black bears. *Ursus*, 19:209–217, 1998.

C. S. Patlak. Random walk with persistence and external bias. *Bull. Math. Biophys.*, 15 (3):311–338, 1953.

H. Qian, M. P. Sheetz, and E. L. Elson. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys. J.*, 60:910–921, 1991.

K. Ritchie, X. Shan, J. Kondo, K. Iwasawa, T. Fujiwara, and A. Kusumi. Detection of non-brownian diffusion in the cell membrane in single molecule tracking. *Biophys. J.*, 88: 2266–2277, 2005.

P. Ryan, S. Petersen, G. Peters, and D. Grémillet. GPS tracking a marine predator: the effects of precision, resolution and sampling rate on foraging tracks of african penguins. *Marine Biology*, 145:215–223, 2004.

M. J. Saxton. Single-particle tracking: The distribution of diffusion coefficients. *Biophys. J.*, 72:1744–1753, 1997.

M. J. Saxton and K. Jacobson. Single-particle tracking: Applications to membrane dynamics. *Annual Review of Biophysics and Biomolecular Structure*, 26:373–399, 1997.

R. Simson, E. D. Sheets, and K. Jacobson. Detection of temporary lateral confinement of membrane proteins using single-particle tracking analysis. *Biophys. J.*, 69:989–993, September 1995.

T. A. Springer. Adhesion receptors of the immune system. *Nature*, 346(6283):425–434, 1990.

R. Swihart and N. Slade. Influence of sampling interval on estimates of home-range size. *The Journal of Wildlife Management*, 49:1019–1025, 1985.

J. R. Taylor. *An introduction to error analysis*. University Science Books, 1997.

Y. Tremblay, P. W. Robinson, and D. P. Costa. A parsimonious approach to modeling animal movement data. *PLoS ONE*, 4(3):e4711, 2009.

P. Turchin. *Quantitative Analysis of Movement*. Sinauer Associates, 1998.

P. Anton van der Merwe, S. J. Davis, A. S. Shaw, and M. L. Dustin. Cytoskeletal polarization and redistribution of cell-surface molecules during t cell antigen recognition. *Seminars in Immunology*, 12(1):5–21, 2000.

Y. van Kooyk and C. G. Figdor. Avidity regulation of integrins: the driving force in leukocyte adhesion. *Curr. Opin. Cell. Bio.*, 12(5):542–547, 2000.

Y. van Kooyk, S. J. van Vliet, and C. G. Figdor. The actin cytoskeleton regulates LFA-1 ligand bindig through avidity rather than affinity changes. *The Journal of Biological Chemistry*, 274(38):26869–26877, 1999.

T. S. van Zanten, M. J. Lopez-Bosque, and M. F. Garcia-Parajo. Imaging individual proteins and nanodomains on intact cell membranes with a probe-based optical antenna. *Small*, 6 (2):270–275, 2010.

H. Weimerskirch, D. Pinaud, F. Pawlowski, and C. Bost. Does prey capture induce area-restricted search? A fine-scale study using GPS in a marine predator, the wandering albatross. *The American Naturalist*, 170(5):734–743, November 2007.

Appendices

A Calculating Lengths and Turning Angles

A.1 Length Calculation

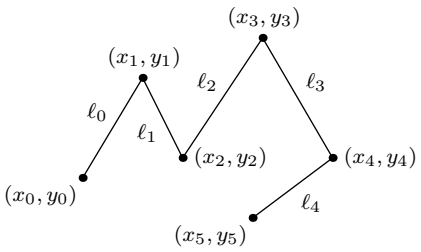


Figure A.1: Sample trajectory path. Each length ℓ_i is calculated by the formula for Euclidian distance.

Given a trajectory as in Figure A.1, step length, ℓ_i is calculated via the formula for Euclidean distance:

$$\ell_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

A.2 Turning Angle Calculation

To find the turning angle of a particular step length ℓ_i, we first calculate α_i^x, the angle of the step to the nearest x-axis, where (x_i, y_j) is given by the origin (see Figure A.2). This

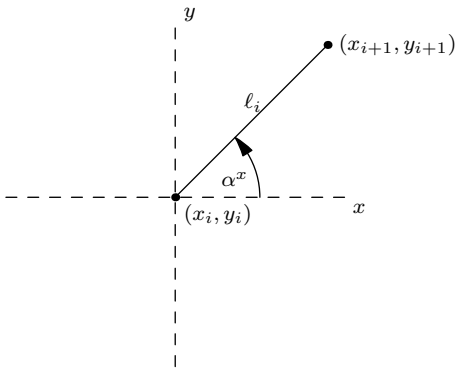


Figure A.2: Angle of the step to the nearest x-axis.

is calculated by

$$\alpha_i^x = \cos^{-1}\left(\frac{|x_{i+1} - x_i|}{\ell_i}\right)$$

Now, we transform this angle to an angle measured counterclockwise from the positive x-axis, denoted by α. Depending on quadrant, we make the following calculations as shown by Figure A.3.

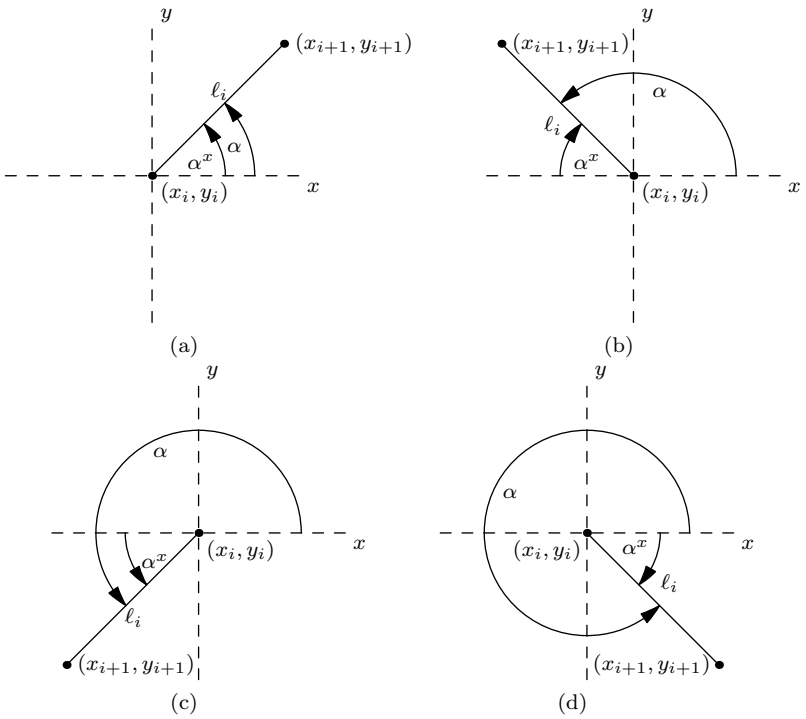


Figure A.3: Calculation of α given α_x, specific to each quadrant. (a) α_x in Quadrant 1. (b) α_x in Quadrant 2. (c) α_x in Quadrant 3. (d) α_x in Quadrant 4.

Quadrant 1:

$$\alpha_i = \alpha_i^x$$

Quadrant 2:

$$\alpha_i = \pi - \alpha_i^x$$

Quadrant 3:

$$\alpha_i = \pi + \alpha_i^x$$

Quadrant 4:

$$\alpha_i = 2\pi - \alpha_i^x$$

We also have special cases of moving left, right, up or down. It is easily seen that:

- if the particle moves left, $\alpha_i = \pi$,
- if the particle moves right, $\alpha_i = 0$,

- if the particle moves up, $\alpha_i = \frac{\pi}{2}$, or,
- if the particle moves down, $\alpha_i = \frac{3\pi}{2}$.

Lastly, we calculate the turning angle in mod 2π: θ_i, given three cases (Figure A.4).

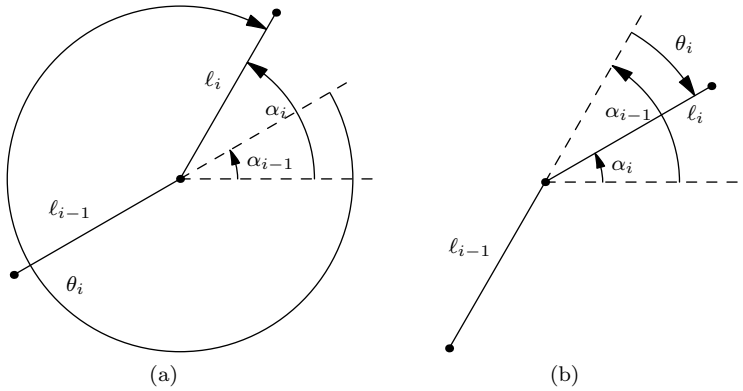


Figure A.4: Calculation of the turning angle θ_i. (a) Calculation of θ when $\alpha_i > \alpha_{i-1}$ (b) Calculation of θ when $\alpha_i < \alpha_{i-1}$

1. $\alpha_i > \alpha_{i-1}$

$$\theta_i = 2\pi + \alpha_{i-1} - \alpha_i$$

2. $\alpha_i < \alpha_{i-1}$

$$\theta_i = \alpha_{i-1} - \alpha_i$$

3. $\alpha_i = \alpha_{i-1}$ (Trivial Case)

$$\theta_i = 0$$

From these calculations we can create distributions of lengths and turning angles from the experimental data to be used for random walk simulation or further anlaysis.