# Time Series Analysis of Particle Tracking Data for Molecular Motion on the Cell Membrane

Wenxia Ying, Gabriel Huerta Department of Mathematics and Statistics
University of New Mexico
Albuquerque NM 87131-1141 USA

Stanly Steinberg
Cancer Research and Treatment Center
Department of Mathematics and Statistics
University of New Mexico
Albuquerque NM 87131-1141 USA

Martha Zúñiga
225 Sinsheimer Laboratories
University of California
Santa Cruz, CA 95064 USA

July 11, 2008

# Contents



# List of Tables

# List of Figures




**Abstract**

Biophysicists use single particle tracking (SPT) methods to probe the dynamic behavior of individual proteins and lipids in cell membranes. The mean squared displacement (MSD) has proven to be a powerful tool for analyzing the data and drawing

conclusions about membrane organization, including features like lipid rafts, protein islands and confinement zones defined by cytoskeletal barriers. Here we implement time series analysis as a new analytic tool to analyze further the motion of membrane proteins. The experimental data track the motion of 40nm gold particles bound to Class I major histocompatibility complex (MHCI) molecules on the membranes of mouse hepatoma cells.

Our first novel result is that the tracks are significantly autocorrelated. Because of this, we developed linear autoregressive models to elucidate the autocorrelations. Estimates of the signal to noise ratio for the models show that the autocorrelated part of the motion is significant. Next we fit the probability distributions of jump sizes with four different models. The first model is a general Weibull distribution that shows that the motion is characterized by an excess of short jumps as compared to a normal random walk. We also fit the data with a chi distribution which provides a natural estimate of the dimension $d$ of the space in which a random walk is occurring. For the biological data, the estimates satisfy $1 < d < 2$, implying that particle motion is not confined to a line but also does not occur freely in the plane. The dimension gives a quantitative estimate of the amount of nanometer scale obstruction met by a diffusing molecule. We introduce a new distribution and use the generalized extreme value distribution to show that the biological data also have an excess of long jumps as compared to normal diffusion. These fits provide novel estimates of the microscopic diffusion constant.

Previous MSD analyzes of SPT data have provided evidence for nanometer-scale confinement zones that restrict lateral diffusion, supporting the notion that plasma membrane organization is highly structured. Our demonstration that membrane protein motion is autocorrelated and is characterized by an excess of both short and long jumps reinforces the concept that the membrane environment is heterogeneous and dynamic. Autocorrelation analysis and modeling of the jump distributions are powerful new techniques for the analysis of SPT data and the development of more refined models of membrane organization.

The time series analysis also provides several methods of estimating the diffusion constant in addition to the constant provided by the mean squared displacement. The mean squared displacement for most of the biological data shows a power law behavior rather the the linear behavior of Brownian motion. In this case, we introduce the notion of an instantaneous diffusion constant. All of the diffusion constants show a strong consistency for most of the biological data.

# 1 Introduction

Single particle tracking methods (SPT) are used by biophysicists to visualize the motion of macromolecules on the cell membrane [33]. The observed motion is erratic and is consequently modeled as a random walk [43, 46, 56, 48, 50, 3]. The dynamic data acquired by particle tracking are used to provide insight into the dynamic organization of the membrane in living cells [24, 27, 55, 32, 42, 16, 17, 19, 58, 14, 54]. Understanding signal transduction is a fundamental problem in cell biology [60, 44]. Results from SPT are integrated with other measurements to develop spatiotemporal models of signal transduction. The tracking requires the molecule be labeled with a probe. A standard label is a gold nanoparticles 40nm in diameter. Other labels used are particles of about the same size made from other materials (e.g. polystyrene) or fluorescent molecules [21, 59]. Some recent tracking techniques emphasize the use of quantum dots [2, 37, 35, 15, 36, 11], which are available in several colors, allowing several particles to be tracked simultaneously. With gold labels, the centroids of the particles can be located to within 30nm (nanometers) [50, 20]. The data are commonly taken at video rate (1/30 second) but can be taken much faster [34]. Biophysicists typically analyze SPT data using the mean squared displacement (MSD), which for random walks generated by mean-zero, independent and identically distributed (IID) jumps, is proportional to time. Usually the estimates of the MSD for biological data are not proportional to time, and consequently the diffusion is viewed as anomalous [43]. A recent discussion of the difficulties in analyzing SPT data is given in [13].

In this paper we view SPT data as a time series [52] and then use time-series analysis to better understand the implications of some experimental data. We focus on understanding the motion of particles at the finest nanometer spatial scale appropriate for the data. Because the mean squared displacement of particle tracks grows with time, the particle positions are not a *stationary* process. In this case, the usual strategy in statistical analysis is to study the time series obtained by differencing the original time series which, in our context, is the time series of jumps generating the random walk. It is the jumps that we analyze.

The experimental biological data are from the paper *Short Class I Major Histocompatibility Complex Cytoplasmic Tails Differing in Charge Detect Arbiters of Lateral Diffusion in the Plasma Membrane* by G.G. Capps, S. Pine, M. Edidin and M.C. Zúñiga [7, 18]. This paper studies the directed and Brownian movement of the wild type and seven mutants of the class I major histocompatibility complex (MHCI) molecule on the membranes of mouse hepatoma cells. MHCI is a transmembrane protein with a single membrane spanning domain and an extracellular accessory subunit, $\beta_2$-microglobulin. It plays a central role in antigen presentation for the development of immune responses. In the biology paper, the MHCI molecules are labeled with 40nm gold particles. The authors assume the the dynamic behavior of the membrane molecule is not seriously affected by the presence of the gold tag. The data consist of 34 gold particle paths with between 623 and 2117 points in a path. The coordinates of the positions are in nanometers and the time step is given by the standard video rate (1/30 second). Herein we refer to these data as *the biological data*. The paths for these data are displayed in appendix B. We do not attempt to reanalyze completely the

whole data set; we use a few selected data sets to illustrate the techniques being presented. We supplemented the data with a 10,000 step random walk whose jumps have mean zero normally distributed components. The artificial data were used to test our algorithms and codes.

In this introductory section, we set up the notation for modeling the biological data and introduce a null hypothesis that the components of the jumps between particle locations are IID and have a normal distribution. We then give the results of some preliminary analyses of the biological data. Graphically, the biological data analyzed appear to have normally distributed components, which is equivalent to the jump sizes having a simple Weibull distribution (see [8] equation (1)) while the angles are uniformly distributed. However, essentially none of the biological data pass rigorous statistical tests for having these properties, that is, the null hypothesis is rigorously shown to be false.

Section 2 provides a series of statistical tests for the time series of the jumps that help elucidate the nature of the molecular motion. We then apply these test to the biological data. First we show that the jumps are a statistically *stationary* process so that it is appropriate to apply standard time series analysis. Our first important observation is that the time series of the jumps is autocorrelated, meaning that some significant part of the motion is deterministic. This is surprising, given that the time steps in the biological data are 1/30 of a second. The results imply that the walks are autocorrelated over significant times and distances. To characterize the autocorrelated part of the motion, we produce linear vector multi-step autoregression models for the biological data. These models isolate the autocorrelated (deterministic) part of the motion from the random part (white noise) of the motion. Essentially all of the biological data have a non-trivial deterministic model. We estimate the importance of the deterministic motion using the signal to noise ratio. However, the autocorrelations are not so strong that it is possible to predict the motion. This section ends by showing the null hypothesis is false in a very strong sense. Consequently analysis of the paths using techniques adapted to autocorrelated data will be more accurate than those based on the assumption that the steps are *independent*.

In section 3, we propose four models for the marginal probability distribution of the lengths of the jumps. The first is a generalized Weibull distribution. This analysis implies that the biological data have excessive short jumps and possibly excessive long jumps as compared to the normal diffusion where the jumps have a simple Weibull distribution. The chi distribution also generalizes the simple Weibull distribution to spaces of arbitrary dimensions. This model implies that the particles are diffusing on a space of dimension $d$ where $1 < d < 2$. This provides an estimate of the amount of fine scale obstructions to the diffusion that can be interpreted as a fractal dimension [29]. The fractal dimension significantly affects reaction kinetics [4, 47].

Next we introduce a novel distribution to test simultaneously if there are an excessive number of both long and short jumps. All of the biological data have excess long jumps in the sense that the distributions have polynomially decaying tails, not exponentially decaying tails as in normal diffusion. Finally we introduce the generalized extreme value (GEV) distribution studied by two of the authors [62, 9] that also shows that there are excessive

long jumps in the biological data. All four of these distributions fit the data well in the sense that the residual of the least squares fits is small. Additionally, the $p$ values for the fits indicate that the fits are good for all but two experiments. At the end of section 3, we estimate the probability of the zero length jumps and then use maximum likelihood to estimate the parameter for the GEV distribution. These results also show there are excessive long jumps.

Because the biological data are autocorrelated, using the idea that the mean squared displacement is proportional to time to estimate the diffusion constant may not be accurate. In section 4, we introduce seven methods of estimating the diffusion constant. The most basic method computes the diffusion constant in terms of the second moment of the jump sizes. This estimates the diffusion constant on the finest possible scale. The diffusion constant can also be computed from the second moment of models for the jump distribution for which we report the results for the chi distribution. An important estimate uses the developed autoregression models to decompose the jumps into a deterministic part and a white-noise part and then uses the white noise to estimate the diffusion constant. We introduce four methods of fitting the MSD, two linear models and two power law models, each of which gives an estimate of the diffusion constant. These models indicate that the autocorrelations do modestly affect the estimates. All of these methods produce modestly different estimates, but overall, the results are consistent. All of the methods produce accurate results for our artificial data. The MSD remains the tool of choice for the analyses of larger scale effects such as cytoskeletal obstructions to free diffusion [24], but such analyzes should take into account any autocorrelations in the data.

The last section gives some conclusions and some critical open problems that need to be solved before these methods can be used for a full analysis of biological data.

## 1.1 Background

In this paper we will use the units nanometers (nm $= 10^{-9}$m) to measure length and microseconds ($\mu = 10^{-6}$s) to measure time. Let

$$J_n = (\Delta X_n, \Delta Y_n)\,, \quad 1 \le n \le N\,, \tag{1.1}$$

be vector random variables, that is let $\Delta X_n$ and $\Delta Y_n$ be random real variables. We will *not* assume that the $J_n$ are independent or identically distributed (IID). The positions of particles in the plane will be modeled by a random walk in the plane given by

$$P_0 = (X_0, Y_0)\,, \quad P_n = (X_n, Y_n) = J_1 + J_2 + \cdots + J_n\,, \quad 1 \le N\,, \tag{1.2}$$

where $(x_0, y_0)$ is a point in the plane. The data to be analyzed gives the positions $P_n$ of the walkers, so then the jumps are

$$J_n = P_n - P_{n-1} = (\Delta X_n, \Delta Y_n)\,, \quad 1 \le n \le N\,. \tag{1.3}$$

It is convenient to use polar coordinates to represent the jumps. So the lengths of the jumps $L_n$ and the angle $\Theta_n$ between the jump vector and the $x$-axis are

$$L_n = \|J_n\| = \sqrt{\Delta X_n^2 + \Delta Y_n^2}\,, \quad \Theta_n = \arctan(\Delta X_n, \Delta Y_n)\,, \tag{1.4}$$

where arctan gives a value in $[-\pi, \pi]$ such that if $L_n \neq 0$ then $\cos(\Theta_n) = \Delta X_n / L_n$ and $\sin(\Theta_n) = \Delta Y_n / L_n$, and consequently $\tan(\Theta_n) = \Delta Y_n / \Delta X_n$ if $\Delta X_n \neq 0$. If $J = (0, 0)$ then $\Theta = 0$ (in MatLab).

If the probability distribution function (PDF) of the jump lengths is given by a radial distribution $p(r)$ then the moments of the distribution are given by

$$M_i = \int_0^\infty r^i p(r)\, dr\,, \quad 0 \leq i\,. \tag{1.5}$$

For a PDF, $M_0 = 1$. The moments are estimated using

$$M_i = \frac{1}{N} \sum_{n=1}^{N} L_n^i\,. \tag{1.6}$$

An important null hypothesis is that the Cartesian coordinates $\Delta X$ and $\Delta Y$ are independent and each is IID with mean zero and standard deviation $\sigma$. If we use the fact that $dx\, dy = r\, dr\, d\theta$ for polar coordinates, then the joint probability measure for $X$ and $Y$ is

$$\frac{1}{\sigma\sqrt{2\,\pi}} e^{-\frac{x^2}{2\,\sigma^2}}\, dx\, \frac{1}{\sigma\sqrt{2\,\pi}} e^{-\frac{y^2}{2\,\sigma^2}}\, dy = \frac{1}{\sigma^2\, 2\,\pi} e^{-\frac{x^2+y^2}{2\,\sigma^2}}\, dx\, dy \tag{1.7}$$

$$= \frac{r}{\sigma^2} e^{-\frac{r^2}{2\,\sigma^2}}\, dr\, \frac{d\theta}{2\,\pi}\,. \tag{1.8}$$

Consequently, $L$ and $\Theta$ are independent, $\Theta$ is uniformly distributed in $[-\pi, \pi]$, and $L$ has a simple Weibull probability distribution

$$w(r, \sigma) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\,\sigma^2}}\,. \tag{1.9}$$

Reversing the argument shows the converse is also true. The three dimensional analog of this argument produces the Maxwell-Boltzman distribution commonly used in thermodynamics.

The mean squared displacement, which is frequently used to analyze random walks, is the expected value of the square of the lengths of the paths:

$$\mathrm{MSD}_n = E(\|P_n\|^2)\,. \tag{1.10}$$

In the analysis of a path $P_n$ the MSD is estimated by

$$\mathrm{MSD}_n = \frac{1}{N-n} \sum_{m=1}^{m=N-n} \|P_{n+m} - P_m\|^2\,, \quad 1 \leq n \leq N-1\,. \tag{1.11}$$

The estimate is only meaningful for values of $n << N$.

In the null hypothesis, if we do not assume that the components of the jumps are normally distributed, but only that the first and second moments of the jumps are finite, then

$$\mathrm{MSD}_n = E\left(\|P_n\|^2\right) = E\left(\|\sum_{k=1}^{n} J_k\|^2\right) = \sum_{k=1}^{n} E\left(\|J_k\|^2\right) = \sum_{k=1}^{n} E\left(L_k^2\right) = M_2\, n\,, \qquad (1.12)$$

so the MSD grows linearly with the time step $n$. If the time step in the walk is $dt$ and $t = n\, dt$, then

$$\mathrm{MSD}(t) = \mathrm{MSD}_n = M_2\, n = \frac{M_2}{dt}\, t\,. \qquad (1.13)$$

In the case that the components of the jumps are normally distributed with mean zero and standard deviation $\sigma^2$ or equivalently, the length of the jumps have a simple Weibull distribution with second moment $M_2 = 2\,\sigma^2$ then

$$\mathrm{MSD}_n = 2\,\sigma^2\, n\,, \quad \mathrm{MSD}(t) = \frac{2\,\sigma^2}{dt}\, t\,. \qquad (1.14)$$

The analysis of the data is performed either using MatLab programs that we have developed or utilities in the statistical software **R**. These are available on the web. One test we performed on our software was to generate 10,000 jumps with mean zero normally distributed components and then use these to generate an IID random walk that we analyzed.

## 1.2 Preliminary Analysis of the Biological Data

The results of all of our analyses are tabulated in appendix A. The data are divided into 8 groups that correspond to the experimental conditions given in [7]. The data in this appendix are frequently reported with fewer digits than in the paper.

The results of the preliminary analysis of the biological data are given in table A.3. The first column in this table labels the data set with a number that corresponds to the experiment label in table A.7. The number $N$ of the point in the track is given in column 2. The first tests that we apply to data are simple *sanity* checks, the most important of which is computing the largest jump (see column 3 of of table A.3) and the mean jump size (see column 4). The error in measuring the position of the particle is less than 30nm, so column 5 gives the % of jumps less that 5nm, an indication of the number of jumps so small that they could easily be zero. All of the tracks have a significant number of small jumps. Any data set with a large percentage of small jumps cannot satisfy the null hypothesis of having a Weibull distribution of the jump sizes because this distribution is zero for zero length jumps. Column 6 of table A.5 gives the percentage of jumps that are exactly zero as this can cause problems in some of the statistical tests. About half of the biological tracks had zero jumps.

We plotted the cumulative distribution functions (CDF) of $X$, $Y$, $L$ and $\Theta$ (not shown). For $X$ and $Y$, the plots look normal, while for $L$ the plots appear similar to the CDF for the Weibull, and for $\Theta$ the plots look linear. However, careful statistical analysis shows that none of these "null hypotheses" about the distributions of $X$, $Y$, $L$ and $\Theta$ are correct.

# 2 Time Series Tests

The application of time series tests [52] requires the time series data to be *ergodic* and *stationary*. Intuitively, ergodic means that the statistics of many paths at some fixed time is equivalent to looking at one path for a long time. Stationary means that the mean and standard deviation over a large number of series do not depend on time. We will assume that the data to be analyzed are ergodic. The positions $P_n$ in a random walk are not stationary because their standard deviation, which is proportional to the mean squared displacement, commonly grows with time. In such a situation, the standard statistical approach is to study the time series of the differenced data, which for particle tracking data is just the jumps $J_n$. Note that $P_n$ and $J_n$ are vector valued where the vectors have a physical meaning of positions in the plane. The statistical software **R** has packages for analyzing vector time series, but the underlying assumption is that these are vectors of data. A common application would be to check if the columns of data are correlated. For the biological data this would mean that the $x$ and $y$ components of the time series are correlated, which turns out to not be very relevant for the biological data.

The book [52] is an excellent reference for the statistical tests that we apply to the biological data. We use the statistical software **R** to do the testing. Unfortunately **R** does not contain a direct stationarity test for vector time series. However, the size of the mean of the jumps for each of the experiments is less than 1nm (see column 6 of table A.3), so the data can be modeled as having mean zero. Because the jumps are modeled as mean zero, the standard deviation of the jumps $J_n$ is the same as the second moment $M_2$ of the lengths of the jumps $L_n^2 = X_n^2 + Y_n^2$ (see columns 7 of table A.3). As a result, it is reasonable to test the scalar time series $L_n$. There are several test available to check if a scalar time series is stationary; we use two tests that are described below. For the biological data, the scalar time series $L_n$ are stationary so we can now apply the standard time series analysis.

Surprisingly, the jumps have a significant autocorrelation, so they do not satisfy the commonly used hypothesis of being *independent*. The first two autocorrelation coefficients are displayed in columns 8 and 9 of table A.3. We confirm this conclusion by plotting the autocorrelations coefficients with a 95% confidence interval for being zero. Next we determine linear vector autoregression models with lag $p$ (VAR($p$) models) for the data. All of the biological data sets have a nontrivial model.

Even though the steps are not independent, the marginal probability distributions $\Delta X$, $\Delta Y$, $L$ and $\Theta$ still make sense. For essentially all of the biological data, $\Delta X$ and $\Delta Y$ are not normally distributed, $L$ does not have a simple Weibull distribution, and $\Theta$ does not have a uniform distribution. Consequently, the null hypothesis (1.7) is definitely false.

## 2.1 Autocorrelations

The assumption that the time series of jumps $J_n$ is stationary implies that the mean and standard deviation do not depend on the position in the series, so it then makes sense to

define the mean, standard deviation, and autocorrelations of the series by

$$\mu = E(J_n)\,, \tag{2.1}$$

$$\sigma^2 = E\left(\|J_n - \mu\|^2\right) \tag{2.2}$$

$$\rho_p = E\left(\frac{(J_{n+p} - \mu) \circ (J_n - \mu)}{\sigma^2}\right)\,. \tag{2.3}$$

where $E$ is the expectation operator. Note that $\mu$ is a vector while $\sigma$ is a positive real and $\rho_p \in [-1, 1]$ and $\circ$ is the scalar product of two vectors. If we assume that the time series is ergodic then the mean, standard deviation, and autocorrelations can be estimated by

$$\mu = \frac{1}{N}\sum_{n=1}^{N} J_n\,,$$

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N} \|J_n - \mu\|^2\,,$$

$$\rho_p = \frac{1}{N-m}\sum_{n=1}^{N-m} \frac{(J_{n+m} - \mu) \circ (J_n - \mu)}{\sigma^2}\,, \quad m < N\,.$$

If $j(r)$ is the PDF of the jump sizes, the moments of the jump sizes are defined by

$$M_k = \int_0^\infty r^k\, j(r)\, dr\,, \quad k \geq 0\,. \tag{2.4}$$

The moments are estimated by

$$M_k = \frac{1}{N}\sum_{n=0}^{N} L_n^k\,. \tag{2.5}$$

If we assume the jumps are mean zero, then the standard deviation of the jumps $J_n$ is the second moment of the jump lengths $L_n$:

$$\sigma^2 = M_2\,. \tag{2.6}$$

These are given in column 7 of table A.3.

The autocorrelation coefficients $\rho_1$ and $\rho_2$ are given in columns 8 and 9 in table A.3. Importantly, all of the data are significantly autocorrelated. Another way to see the autocorrelations in the data is to use **R** to plot the autocorrelation function (ACF) $\rho_p$ as a function of the lag $p$ along with two dashed horizontal lines that give the 95% confidence interval for the ACF values being zero. As an example, we show a plot of the ACF for the data from experiment 12 in Figure 2.1. Eight of the positive spikes fall outside the confidence interval, while there are only a few small negative spikes, indicating that the data have significant positive autocorrelation. All of the biological data show significant autocorrelation using this test (not shown).

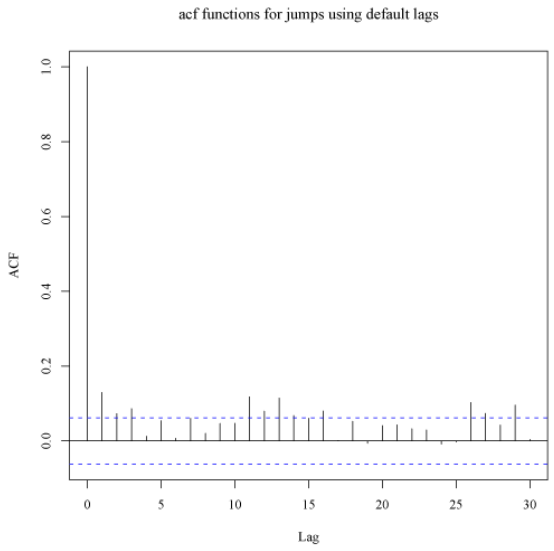


Figure 2.1: The Autocorrelation Function of the Jumps for Experiment 12.

## 2.2 Testing the Stationarity of Time Series

To correctly apply the standard statistical tools to the jumps we need to test if the lengths of the jumps are stationary. The augmented Dickey-Fuller (ADF) test assumes that the data $L_n$ can be modeled as an auto regressive process of order $p$, that is, as an AR($p$) model. First, let

$$\Delta L_n = L_n - L_{n-1}\,, \tag{2.7}$$

and then estimate the autoregression

$$\Delta L_n = \mu + \beta L_{n-1} - \sum_{i=1}^{p} \alpha_i\, \Delta L_{n-i} + \epsilon_n\,. \tag{2.8}$$

where $p \geq 0$, $\mu$, $\beta$ and the $\alpha_i$ are real constants, and $\epsilon_n$ are scalar white noise random variables. If the coefficient $\beta$ is not zero, then the regression doesn't have a unit root, and the series is stationary.

Applying the ADF test in **R** to the biological data with the null hypothesis that the data are not stationary (or $\beta = 0$) gives $p$-values that are all less than 0.01, implying that the null hypothesis is false, that is, implying that the $L_n$ series do not have unit roots. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which tests the null hypothesis that $L_n$ is trend stationary gives results consistent with the results of the ADF test.

## 2.3 Linear Autocorrelation Models

To better understand the structure of the autocorrelations, we use **R** to construct the linear vector autoregression VAR($p$) models, using the ordinary least squares method, of the time

series of jumps:

$$J_n = \mu + \sum_{i=1}^{i=p} A_i \, J_{n-i} + \epsilon_n \, , \tag{2.9}$$

where $p$ is the number of lags, $\mu$ is a vector and $A_i$ are $2 \times 2$ matrices. The $\epsilon_n$ terms are white noise random variables. The model (2.9) without the white noise term is called the deterministic model. The model (2.9) can be computed in the statistics software **R**, using the **ar** command.

For the biological data, we constructed VAR($p$) models for the jumps for all of the PTs. The lag $p$ (not $p$-values) for the model are listed in column 10 of table A.3. We find that $1 \leq p \leq 33$ with most values of $p < 10$. In **R** we can also compute the signal to noise ratio for scalar models of the x-component and the y-component of the model and report the norm of this vector in column 11 of table A.3. These are between 2% and 53% and have a mean of 16%. Experiments with a small ratio also have a small first autocorrelation coefficient $\rho_1$. The correspondence for large ratios is not as good as for small. High percentages are an indication of strong non-random motion. We also analyzed the eigenvector and eigenvalues of the coefficient matrices $A_i$, but did not find anything useful.

Here are some examples for the $p = 2$ model (see equation (2.9)). For experiment 8, the VAR(2) model is

$$\begin{bmatrix} \Delta X_n \\ \Delta Y_n \end{bmatrix} = \begin{bmatrix} -.01 \\ .01 \end{bmatrix} + \begin{bmatrix} .37 & .18 \\ .02 & .05 \end{bmatrix} \begin{bmatrix} \Delta X_{n-1} \\ \Delta Y_{n-1} \end{bmatrix} + \begin{bmatrix} -.10 & .01 \\ -.00 & -.11 \end{bmatrix} \begin{bmatrix} \Delta X_{n-2} \\ \Delta Y_{n-2} \end{bmatrix} . \tag{2.10}$$

For experiment 17, the VAR(2) model is

$$\begin{bmatrix} \Delta X_n \\ \Delta Y_n \end{bmatrix} = \begin{bmatrix} -.09 \\ .13 \end{bmatrix} + \begin{bmatrix} -.05 & -.04 \\ -.09 & -.12 \end{bmatrix} \begin{bmatrix} \Delta X_{n-1} \\ \Delta Y_{n-1} \end{bmatrix} + \begin{bmatrix} .11 & -.02 \\ -.02 & -.01 \end{bmatrix} \begin{bmatrix} \Delta X_{n-2} \\ \Delta Y_{n-2} \end{bmatrix} . \tag{2.11}$$

For experiment 28, the VAR(2) model is

$$\begin{bmatrix} \Delta X_n \\ \Delta Y_n \end{bmatrix} = \begin{bmatrix} -.00 \\ -.01 \end{bmatrix} + \begin{bmatrix} -.14 & .04 \\ .16 & -.12 \end{bmatrix} \begin{bmatrix} \Delta X_{n-1} \\ \Delta Y_{n-1} \end{bmatrix} + \begin{bmatrix} -.08 & .00 \\ .03 & -.08 \end{bmatrix} \begin{bmatrix} \Delta X_{n-2} \\ \Delta Y_{n-2} \end{bmatrix} . \tag{2.12}$$

The expected value for the jumps size ranges between 12 and 76 and the coefficients in the linear models (see 2.10, 2.11 and 2.12) are significantly smaller than the expected value of the jumps, so the linear model can only account for a modest part of the motion.

## 2.4 Marginal Probability Distributions

As noted in the introduction, from plots of the cumulative distribution functions of $\Delta X_n$, $\Delta Y_n$, $L_n$ and $\Theta_n$ it appears that the jumps $\Delta X_n$ and $\Delta Y_n$ are normally distributed. Furthermore, the $L_n$ appear to have a simple Weibull distribution and the $\Theta_n$ appear uniformly distributed. Thus, these marginal distributions appear to satisfy the null hypothesis described in section 1.1, equation (1.7). Note that because the $J_n$ are not independent, the

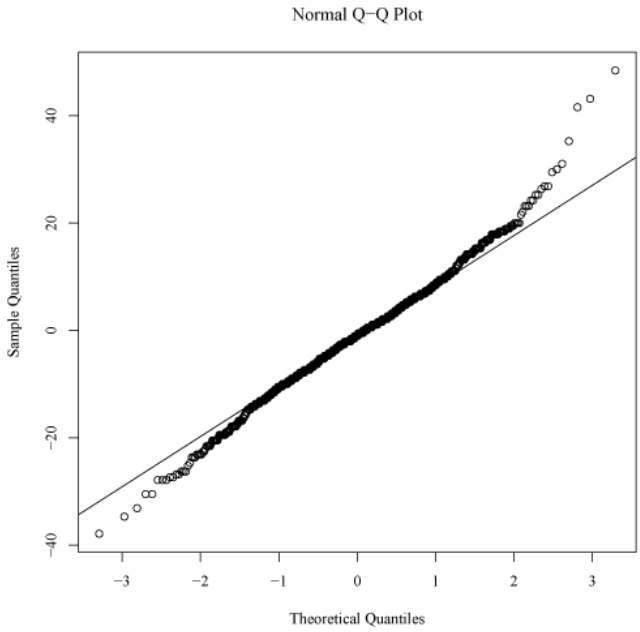


Figure 2.2: The Q-Q plot for $\Delta X$ From Experiment 12.

equivalence given in equation (1.7) doesn't apply, so we took these distributions as hypotheses to test.

To check normality we use two tests: the Shapiro-Wilk (SW) [51] test and Anderson-Darling (AD) [1] test. When we use the SW test in the statistical software **R** we need $p < .05$ to reject the null hypothesis that $\Delta X_n$ and $\Delta Y_n$ are normally distributed. For the biological data the values of $p > .05$ for $\Delta X$ (in descending size) are 0.7, 0.3, 0.2 for experiments 9, 33, 21. Otherwise $p < .05$ with most $p$ values being near zero. For $\Delta Y$, the $p > .05$ are 0.4, 0.3, 0.2, 0.2, 0.1 for experiments 17, 33, 12, 23, 30. So only experiment 33 has both $\Delta X$ and $\Delta Y$ normally distributed. The data for experiment 3 appear unusual in several other tests. The AD test gives results consistent with SW test. The AD test is nice as it can be used to see if a sample of data comes from several specific distributions: normal, lognormal, Weibull, and extreme value type I. (We will use this later.) Both the AD and SW tests are powerful, so we have confidence that the components of the jumps are not normally distributed.

We used the normal Q-Q (Q stands for quantile) plot as a graphical tool for diagnosing differences between the biological distributions and the normal distribution to confirm that $\Delta X$ and $\Delta Y$ are not normal. For a sample from a normal distribution, the Q-Q plot should approximate a straight line, especially near the center (0.0). In Figure 2.2, we give a sample Q-Q plot for $\Delta x$ for experiment 12 which clearly deviates from the normal distribution. The plots (not presented) for $\Delta X$ and $\Delta Y$ for the remaining experiments support that these variables are not normally distributed.

We use the Kolmogorov-Smirnov (KS) goodness-of-fit test to test whether the angles $\Theta_n$

are uniformly distributed. This test is an alternative to the AD test, but the KS test is non-parametric and distribution free in the sense that the critical values do not depend on the specific distribution being tested. If the assumption about the distribution can be validated, then the AD test is more powerful and preferred, but the null hypothesis will be seen not to be valid, so we use the KS test instead. If $p < .05$, then the null hypothesis is rejected, the angles are not uniformly distributed. The large $p$-values for the KS test of the biological data are .95, .88, .82, .64, .46, .44, .38, .21, .12, .05 for experiments 29, 12, 3, 14, 33, 4, 21, 27, 5, 19. This matches our impression from the cumulative distribution plots (CDF) plots that the angles are close to uniformly distributed – as this is true for 10 out of 33 data sets. Note that the angles for experiment 33 are uniformly distributed, as suggested by formula (1.7) and the fact that $\Delta X$ and $\Delta Y$ are normally distributed for this experiment.

Finally, we test if the jump lengths have a (simple) Weibull distribution (1.9), that is, $w(r/s)/s$ where

$$w(r) = r\, e^{-\frac{r^2}{2}}\,, \tag{2.13}$$

as conjectured in equation (1.7). For the biological data, the $p$-values are 0.25, 0.198, 0.103, 0.076, 0.065, for experiments 8, 21, 17, 6, 33. So most of the jump size data does not have a simple Weibull distribution.

We see that only experiment 33 passes all four tests for being a normal distribution; experiment 21 passes three tests. However, experiment 33 has a nontrivial autoregressive model with 14 lags and a signal to noise ratio of 52%. All but one of the data sets deviate significantly from a IID random walk where the components of the jumps come from a normal distribution.

# 3 Modeling the Distribution of the Jumps

A standard approach for understanding a data set is to find a probability distribution for the data. In this section we study four models for the probability distribution function $p = p(r)$ of the jump sizes: the general Weibull (GW), the chi, the long-short (LS), and the generalized extreme value (GEV) distributions. To help keep our results organized, the PDFs and CDFs of the distributions are labeled as follows: the GW uses $w$ and $W$; the chi uses $c$ and $C$; the LS uses $p$ and $P$; and the GEV uses $h$ and $H$. Plots of the PDFs for these models are shown in figure 3.3 for biologically relevant values of their parameters. The plotted PDFs are normalized so that their first moment is one and the simple Weibull PDF (SW) is given to assist in comparing the distributions. We know that the simple Weibull PDF (2.13) does not represent the biological data accurately, but the general Weibull (GW) PDF fits the data well. The GEV PDF [9] is good for detecting long jumps. Both the general Weibull and chi distribution contain the simple Weibull as a special case and thus can easily detect normal diffusion.

The main concern is to understand if there are more or fewer long or short jumps in a PDF $p$ than would be predicted by the simple Weibull PDF. Note that for a uniform distribution particles in a finite region in $d$ dimensional region containing the origin and for a sufficiently small $R$, the probability that the particle distance $r$ from the origin is less than $R$ is proportional to the volume of the sphere of radius $R$:

$$\text{Prob}(r < R) = P(r) \propto \Omega_d \, R^d \,, \tag{3.1}$$

where $\Omega_d$ is the volume of the $d$-dimensional unit sphere. Consequently the probability distribution $p(r)$ must satisfy

$$p(r) \propto r^{d-1} \,, \tag{3.2}$$

for $r$ small. For two dimensional space, this is exactly what the simple Weibull distribution gives, that is, $d = 2$. So in two dimensional space, if a probability distribution $p$ satisfies

$$p(r) \propto r^{d-1} \text{ for } r \to 0 \,, \quad d > 0 \,, \tag{3.3}$$

then $p$ has excessive short jumps if $0 < d < 2$ and a paucity of short jumps if $d > 2$. If the $1 < d < 2$, then $d$ has a natural interpretation as a fractal dimension [29]. From figure 3.3 we see that the General Weibull, chi and long-short distributions have excessive short jumps. The GEV distribution is unusual, so it will be discussed later. For the biological data, $1 < d < 2$.

If the PDF decays exponentially for $r \to \infty$, then there are no excessive long jumps, which is the case for GW and chi distributions. If

$$p(r) \propto \frac{1}{r^{k+1}} \text{ for } r \to \infty \,, k > 0 \,, \tag{3.4}$$

then there are excessive long jumps and the number increases as $k$ decreases. Note that only the first $k$ moments of $p(r)$ are finite. If $k \leq 2$ then the second moment of $p(r)$ is infinite as is

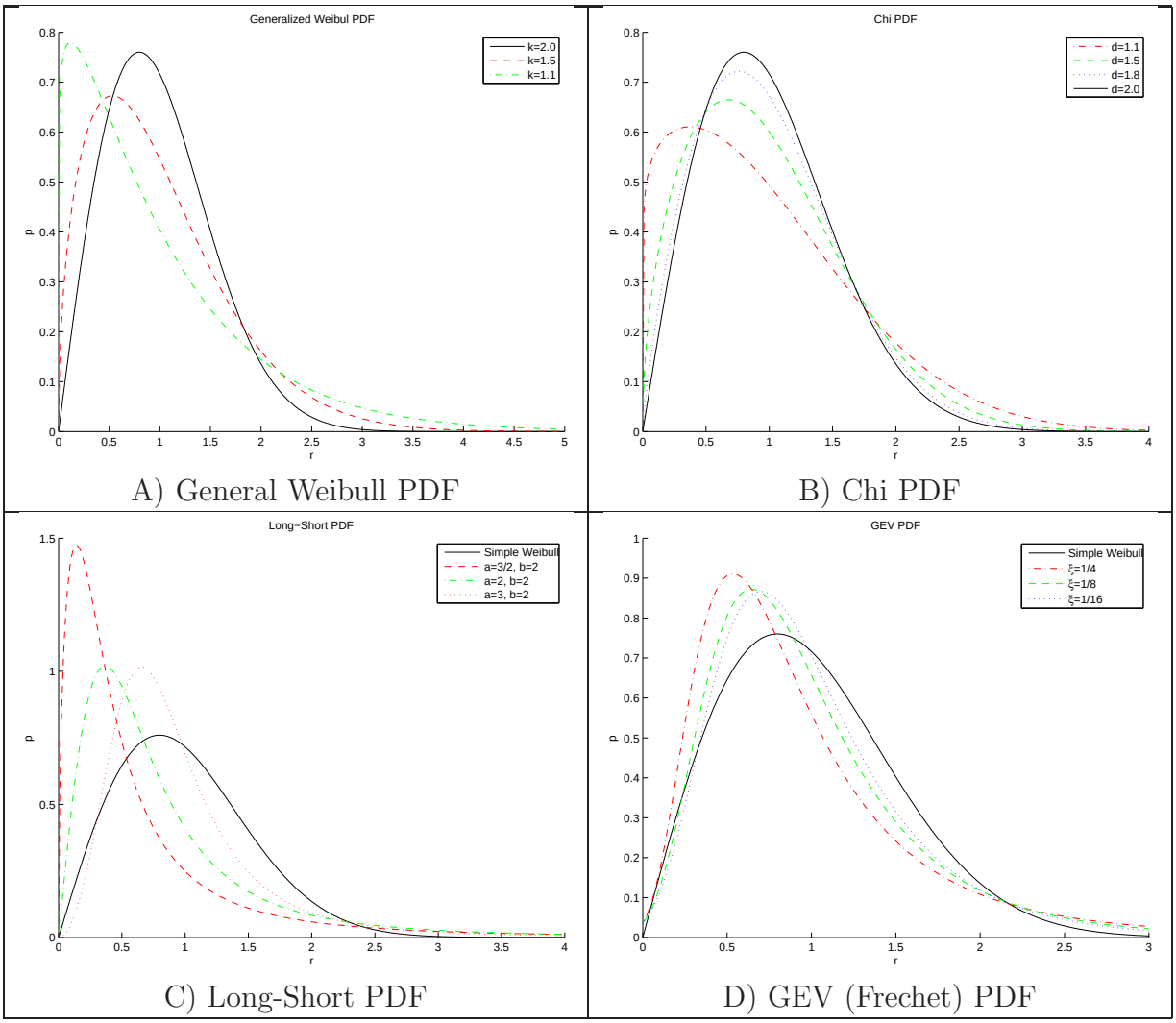


Figure 3.3: PDFs for Modeling the Jump Size Distribution.

.

| | residual | $p$-value |
|---|---|---|
| Weibull | $0.16 \times 10^{-3}$ | $.11 < p < .93$ |
| Chi | $0.23 \times 10^{-3}$ | $.06 < p < .82$ |
| L-S | $0.07 \times 10^{-3}$ | $.28 < p < .94$ |
| GEV | $0.09 \times 10^{-3}$ | $.27 < p < .96$ |

Table 3.1: The Maximum Residuals and $p$-Values for the Weibull, Chi, Long-Short and GEV Distributions.

.

the diffusion constant for the related random walk. Such diffusion are known as *anomalous* in the mathematical literature [22, 41]. The long-short and GEV distributions have excessive long jumps (fat tails).

To simplify our notation, we give PDFs and CDFs for the distribution of jump sizes with scale parameter equal to one. If $p(r)$ and $P(r)$ are a PDF and a CDF, then the scale parameter $s$ can be introduced using $p(r, s) = p(r/s)/s$ and $P(r, s) = P(r/s)$.

There are several methods of estimating the parameter of a PDF so that it fits the biological data. A commonly used method is maximum likelihood. However, this method is designed for distributions that are nowhere zero. All distributions have a zero value, so this is not a good choice. Other possibilities are to make a least squares (LS) fit to either the PDF or CDF. We use the PDF and for this we binned the jumps. Because the position data $P_n$ has measurement errors on the order of 10nm, we used 5nm bins. The maximum residual for fitting the four PDFs to the data are given in table 3.1. They are small indicating good fit. The $p$ values for the fits were estimated using a two sample Kolmogorov-Smirnov (K-S) test. The null hypothesis is that the data are distributed according to the proposed distribution. The K-S test compares the jump length data to simulated data from the proposed distribution with estimated parameters. A $p$-value with $p > .05$ means we accept the null hypothesis that the data come from the proposed distribution. For the Weibull and Chi distributions the $p$ values for experiments 2 and 19 are essentially zero, indicating poor fits. However, plots of the data for these experiments are given in figure 3.4 and they look good so the test is measuring a subtle difference not detected by the residuals. In all other cases the $p > .05$, so the fits are statistically significant.

At the end of this section, we separately analyze the zero and non-zero jumps. Then we can apply the maximum likelihood (ML) method to estimate the parameter in the GEV distribution for the non-zero jumps. We have also used the method of moments (MOM) to check some of our estimates (not presented).

## 3.1 The General Weibull Distribution

The general Weibull PDF and CDF are $w(r,\lambda,k) = w(r/\lambda,k)$ and $W(r,\lambda,k) = W(r/\lambda,k)$ where

$$w(r,k) = k\, r^{k-1} e^{-r^k}\,, \quad W(r,k) = 1 - e^{-r^k}\,, \tag{3.5}$$

and $k > 0$, $\lambda > 0$ and $r > 0$. We will estimate $\lambda$ and $k$ for fitting the biological data. The moments of $w(r,\lambda,k)$ are

$$M_1 = \lambda\,\Gamma(1+\frac{1}{k})\,, \quad M_2 = \lambda^2\,\Gamma(1+\frac{2}{k}) \tag{3.6}$$

For the plots in figure 3.3, the value of $\lambda$ is computed by choosing $M_1 = 1$. When $k = 2$, the general Weibull distribution is just the simple Weibull distribution, while, for $k = 1$, it is the simple exponential distribution. For the biological data, we use this distribution for $1 \le k \le 2$, and plots for this range of values are given in figure 3.3. Also, in this range of values, $w(r,\lambda,k) \propto r^{k-1}$ and thus has an excessive number of small jumps. The distribution decays exponentially at infinity and so gives no indication of excessive long jumps.

For the biological data the values of $\lambda$ and $k$ are given in columns 2 and 3 of table A.4. The two smallest values of $k$ are 1.3829 and 1.4752, while the two largest values are 1.9781, 2.0145 for experiments 19, 2, 9 and 25. We see that $1 < k < 2$ for all but experiment 25, so essentially all of the data has excess small jumps.

## 3.2 The Chi Distribution

The chi PDF and CDF are $c(r,\sigma,d) = c(r/\sigma,d)/\sigma$ and $C(r,\sigma,d) = C(r/\sigma,d)/\sigma$ where

$$c(r,d) = \frac{2}{2^{d/2}\Gamma(d/2)}\, r^{d-1}\, e^{-\frac{r^2}{2}}\,, \quad C(r,d) = 1 - \frac{\Gamma\left(\frac{d}{2},\frac{r^2}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\,, \tag{3.7}$$

$d \ge 1$, $\sigma > 0$ and the $\Gamma$ of two arguments is the incomplete gamma function. We will estimate $\sigma$ and $d$ for the fitting biological data. The *chi* distribution is a special case of the *Nakagami* distribution and can be transformed to either the *chi-square* distribution or the *gamma* distribution. The moments of $c(r,\sigma,d)$ are

$$M_1 = \sigma\,\frac{\sqrt{2}\,\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\,, \quad M_2 = \sigma^2\, d\,. \tag{3.8}$$

Plots of $c(r,\sigma,d)$ with $\sigma$ chosen so that $M_1 = 1$ are given in figure 3.3 for the biologically relevant cases of $1 \le d \le 2$. For $d < 2$, the distributions have both excessive short jumps and exponentially decaying long jumps.

Extending the reasoning used to derive equation (1.9) for the distribution of the jump length of a normal IID random walk in two dimensions to a space of $d$ dimensions implies that the jump sizes should be distributed according to $c(r,\sigma,d)$ where $d$ is and integer.

However, this distribution makes sense for all positive real $d$. Consequently, when analyzing the distribution of jumps by fitting them with $c(r, \sigma, d)$ we can interpret $d$ as the dimension of the space in which the diffusion is occurring. For the biological data, the dimension satisfies $1 < d < 2$. If $d = 2$ then the diffusion is in the plane while if $d = 1$, the motion is confined to a curve. For $1 < d < 2$ a plausible interpretation is that the diffusion is restricted to a nanometer scale fractional part of the membrane.

For the biological data the values of $\lambda$ and $d$ are given in columns 4 and 5 of table A.4. The two smallest values of $k$ are 1.2 and 1.3, while the two largest values are 1.97 and 2.02 for experiments 19, 2, 25 and 9. These are the same experiments as for the general Weibull distribution. In figure 3.5 we show the paths for these data, while in figure 3.4 we show the fits to the PDFs. For essentially all of experiments the values for $d$ imply there is significant obstruction to the diffusion. The consistency of the results for the general Weibull and chi distributions supports this conclusion.

## 3.3 The Long-Short Distribution

To better understand the nature of both the long and short jumps we introduce the PDF $p(r, s, \alpha, \beta) = p(r/s, \alpha, \beta)/s$, which has the CDF $P(r, s, \alpha, \beta) = P(r/s, \alpha, \beta)/\sigma$ where

$$p(r, \alpha, \beta) = \frac{\alpha\,(\beta - 1)\, r^{\alpha-1}}{(1 + r^{\alpha})^{\beta}}\,, \quad P(r, \alpha, \beta) = 1 - \left(r^{\alpha} + 1\right)^{1-\beta}\,, \tag{3.9}$$

and $r \geq 0$, $s > 0$, $\alpha > 0$ and $\beta > 1$. We will estimated $s$, $\alpha$ and $\beta$ for the biological data. The first two moments of this distribution are

$$M_1 = \frac{s(\beta - 1)\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\beta - \frac{1}{\alpha} - 1\right)}{\alpha\Gamma(\beta)}\,, \quad \alpha(\beta - 1) > 1\,,$$
$$M_2 = \frac{s^2(\beta - 1)\Gamma\left(\frac{\alpha+2}{\alpha}\right)\Gamma\left(\beta - \frac{2}{\alpha} - 1\right)}{\Gamma(\beta)}\,, \quad \alpha(\beta - 1) > 1\,.$$

Because this distribution looks like $r^{\alpha-1}$ near $r = 0$, the discussion on the general Weibull distribution indicates that we are interested in $1 < \alpha < 2$. When $r$ is large, the distribution has polynomial decaying tail: $p(r) \propto 1/r^{1+\alpha(\beta-1)}$. The values of $s$, $\alpha$, $\beta$ and $k = 1 + \alpha(\beta - 1)$ are given in columns 6, 7, 8, and 9 of table A.4. The two smallest values of $\alpha$ are 1.66 and 1.81, while the two largest values are 2.18 and 2.19 for experiments 23, 20, 16 and 13. The mean of the values of $\alpha$ is 1.98, so this distribution is indicating that there are only a modest number of experiments with excesses short jumps. The two smallest values of $k$ are 6.50 and 6.57, while the two largest values are 35.67 and 37.37 for experiments 1, 22, 9 and 12. These experiments do not correspond to those for the general Weibull or the chi PDFs that had some special property. The data sets do have fat tails, but $k >> 2$ so none of the experiments show *mathematically anomalous* diffusion.

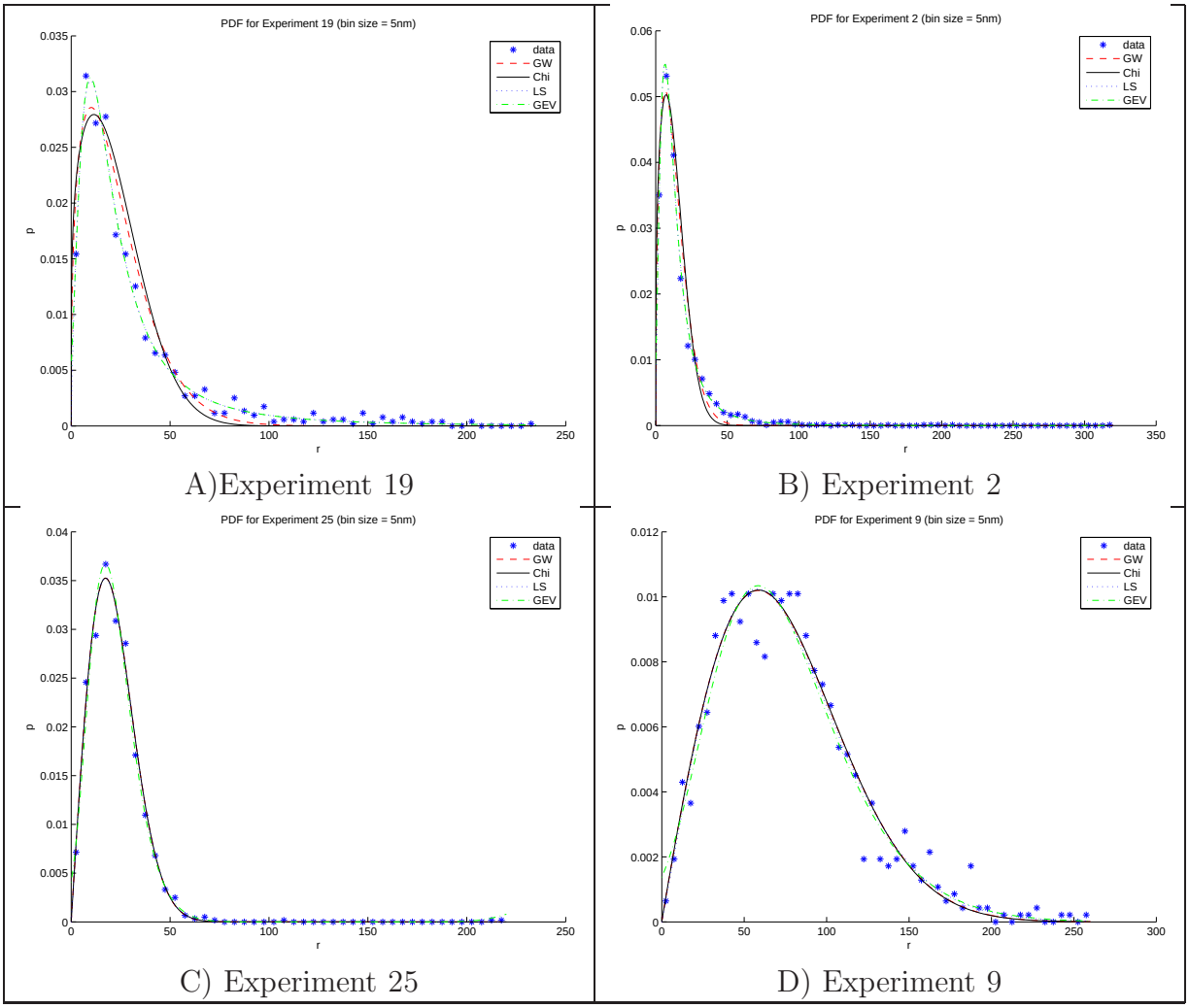


Figure 3.4: PDF Fits for the Experiments Extreme Values of $k$ and $d$. Experiments 19 ($k = 1.4$, $d = 1.2$) and 2 ($k = 1.5$, $d = 1.3$) had the smallest values, while 25 ($k = 2.0$, $d = 2.0$) and 9 ($k = 2.0$, $d = 2.0$) had the largest values.

.

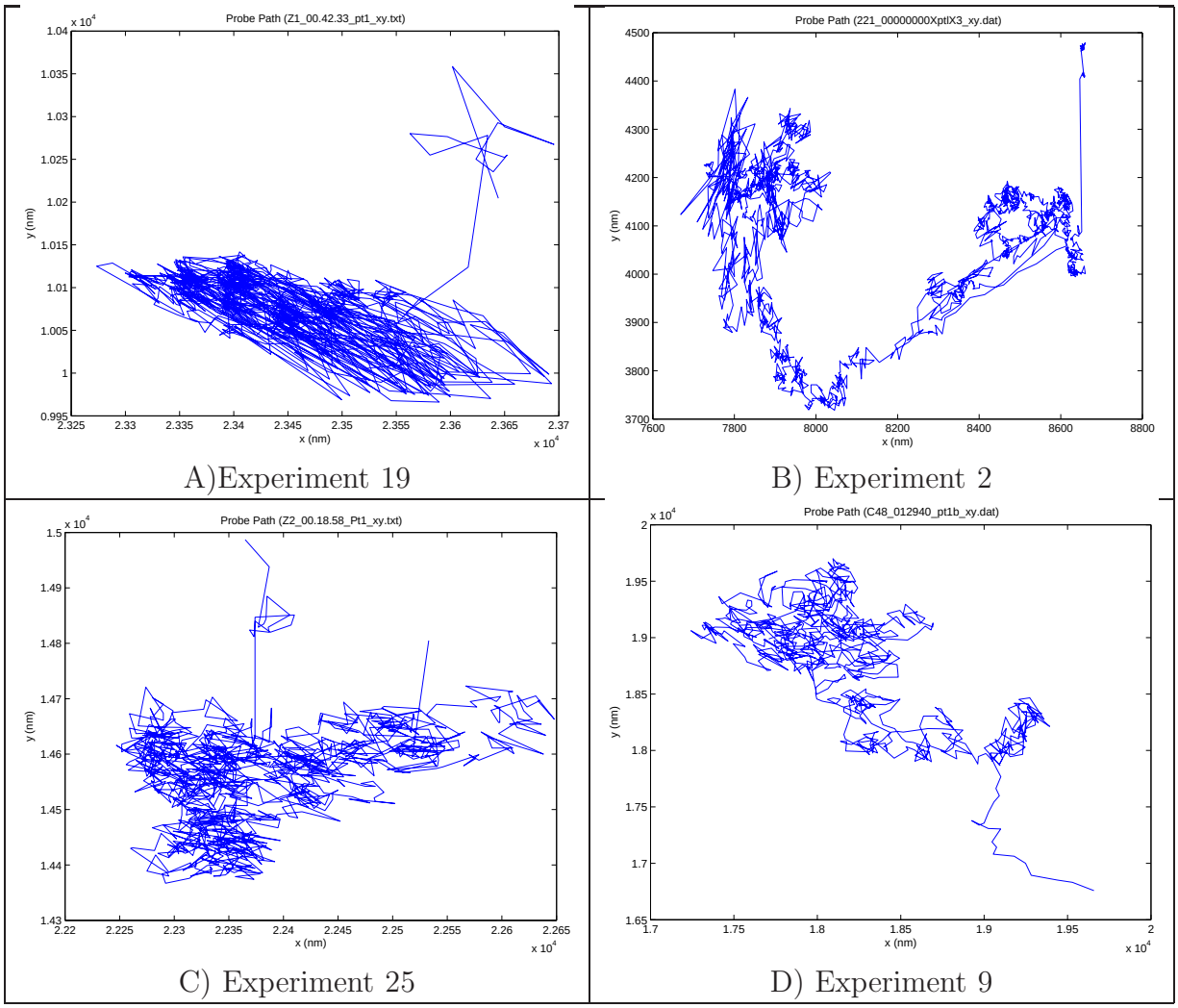


Figure 3.5: Paths for the Experiments With the Smallest $k$ and $d$. Experiments 19 ($k = 1.4$, $d = 1.2$) and 2 ($k = 1.5$, $d = 1.3$) had the smallest values, while 25 ($k = 2.0$, $d = 2.0$) and 9 ($k = 2.0$, $d = 2.0$) had the largest values.

.

## 3.4 Generalized Extreme Value Distribution

It is important to confirm the conclusion that the jumps have fat tails. Two of the authors [62] have used the generalized extreme value (GEV) distribution to study such problems, so we also do that here. The GEV PDF and CDF [9] are $h(x,\mu,\sigma,\xi) = h((x-\mu)/\sigma,\xi)/\sigma$ and $H(x,\mu\,\sigma,\xi) = H((x-\mu)/\sigma,\xi)$ where

$$h(z,\xi) = [1+\xi\,z]_+^{-1/\xi-1}\,e^{-[1+\xi\,z]_+^{-1/\xi}}\,, \quad H(z,\xi) = e^{-[1+\xi\,z]_+^{-1/\xi}}\,. \tag{3.10}$$

The parameter spaces are $-\infty < x,\mu,\xi < \infty$ and $\sigma > 0$ We estimate the parameters $\mu$, $\sigma$ and $\xi$ for fitting the biological data. and the + sign in the model denotes the positive part. The first two moments of $h(x,\mu,\sigma,\xi)$ are

$$M_1 = \mu + \frac{\sigma}{\xi}\,(\Gamma(1-\xi)-1)\,, \quad M_2 = \left(\mu - \frac{\sigma}{\xi}\right)^2 + \frac{2\,\sigma}{\xi}\left(\mu - \frac{\sigma}{\xi}\right)\Gamma(1-\xi) + \frac{\sigma^2}{\xi^2}\Gamma(1-2\xi)\,. \tag{3.11}$$

This distribution contains 3 subfamilies as special cases:

| | name | support |
|---|---|---|
| $\xi < 0$ | Weibull | $x < \mu - \sigma/\xi$ |
| $\xi = 0$ | Gumbel | $-\infty < \mu < \infty$ |
| $\xi > 0$ | Fréchet | $x > \mu - \sigma/\xi$ |

Note that the Weibull case is not the same as what we have studied before. More precisely, when $\xi < 0$, $h$ is a translation of the previous general Weibull distribution. Because the biological data have support for $x > 0$, we had hoped that our fits would have values of $\xi > 0$ so that our model would have support for $x > x_0$ with $x_0 = \mu - \sigma/\xi$. Typically $x_0 < 0$, which we interpret as meaning there are excess short jumps. When $\xi > 0$ the distribution decays at infinity like $1/x^k$ with $k = 1/\xi$. The values of $\sigma$, $\mu$, $\xi$ and $k = 1/\xi$ in columns 2, 3, 4 and 5 of table A.4. Unfortunately, experiments 3, 9, 21, 25, 27, 30, and 33 have negative $\xi$ and so cannot be analyzed with this method. For the remaining experiments, the two smallest values of $k$ are 1.9663 and 2.7941, while the two largest values are 69.6 and 120.5 for experiments 19, 2, 12 and 5. These result do not agree with those for the long-short distribution.

To improve these result, we first analyze the zero jumps and then remove them and then fit the remaining non-zero jumps using the GEV distribution. In table A.5 we give the percentage of zero jumps, which, if we interpret as a probability $q$, then our full model will have the form

$$q\,\delta_0(x) + (1-q)\,h(x,\mu,\sigma,\xi)\,. \tag{3.12}$$

Now we use ML to estimate the parameters $\sigma$, $\mu$ and $\xi$, which are listed in columns 10, 11, 12, and 13 of table A.5. Now all of the data sets have $\xi > 0$ except for experiments 9 and 33. For all of the experiments for which both methods produced $\xi > 0$, $0.9 < \xi(LS)/\xi(ML) < 3.2$ so the two methods of estimating $\xi$ (or $k$) are close. For the data sets for which $\xi > 0$,

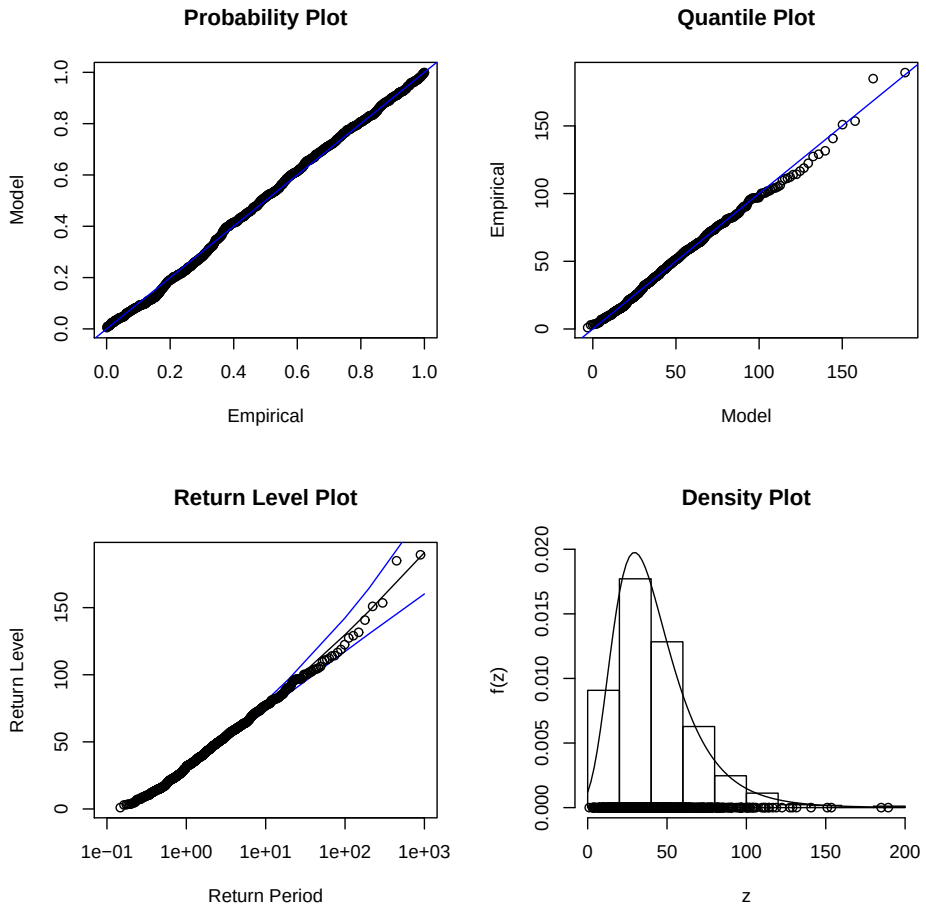


Figure 3.6: Diagnostic Plots for GEV Fit to the Jump Data for Experiment 6.

$1.9 < k < 37$. However, for all but experiment 19, $k < 2$, so only this data set can possibly have slightly mathematically anomalous diffusion.

We also used maximum likelihood to estimate the parameters and $p$-values for the general Weibull and chi fits. The estimated values are close to those for the least squares method and, except for one experiment, the $p$ values indicate that the fits are good.

Sometimes conclusions can be sensitive to the fitted model. To check that the model fits data well, we applied two test to all of the GEV ML fits. In figure 3.6 we present the diagnostic plots for assessing the accuracy of the GEV model for experiment 6. The plots in the first row of figure 3.6 are a probability plot and a quantile plot. Both of these plots provide techniques for assessing whether or not a data set follows a given distribution. The data are plotted against a theoretical distribution in such a way that the points should form a straight line. Departures from this straight line indicate departures from the specified distribution. We can see each set of plotted points is near-linear, so neither plot gives cause to doubt the validity of the fitted model.

The lower two plots are return level plots and density estimate plots. The return level plot shows the return period against the return level, and shows an estimated 95% confidence interval. The return level is the level (in this case for jumps) that is expected to be exceeded, on average, once every $m$ time points (in this case steps). The return period is the amount of time expected to wait to exceed a particular return level. For example, the biological data show that one would expect the maximum jumps to exceed 150nm on average every 200 steps. The density plot shows that the density estimate is consistent with the histogram of the data. Thus all these diagnostic plots lend support to the fitted model.

| ratio | min | mean | max | |
|---|---|---|---|---|
| $D^{\chi}/D$ | 0.36 | 0.52 | 0.80 | |
| $D^{W}/D$ | 0.73 | 0.91 | 0.98 | |
| $D_1/D$ | 0.07 | 0.68 | 1.29 | $B \neq 0$ |
| $D_2/D$ | 0.27 | 0.77 | 1.28 | $B = 0$ |
| $D_3/D$ | 0.31 | 0.91 | 1.99 | $B \neq 0$ |
| $D_4/D$ | 0.24 | 0.83 | 1.35 | $B = 0$ |

Table 4.2: The Ratios of the Alternate Diffusion Constants to the Fine Scale Diffusion Constant $D$ (4.4).

# 4 Diffusion Constants

In applications to cell biology, it is important to be able to estimate the diffusion constant for molecules in the cell membrane. Time series analysis and modeling the probability distribution of the jump sizes offer several different ways of estimating the diffusion constant. In this paper we focus on finest time and spatial scales supported by the data. Consequently our fundamental diffusion constant $D$ is computed directly from the jump sizes. The models for the jump size distributions provide alternate estimates. Additionally, the white noise in the VAR($p$) (2.9), models can be used to estimate the diffusion constant for only the random part of the motion.

Visual inspection of the plots of the MSD for the biological data indicates that the MSD is often given by a power law, at least for few time steps. Sample plots are given in figure 4.7, which also contains plots of the data fitting results obtained below. For our 33 data sets, 19 appear concave down, 9 appear linear, 3 appear concave up, and 2 are not easy to classify. Consequently we fit the MSD with both linear and power-law functions for 10 time steps. Small changes in the number of of steps do not affect our results significantly. Also, we fit the data in a way that should provide estimates of the measurement error, which is on the order of 30nm. Fitting with linear functions is a standard way of estimating the diffusion constant [50]. It is also common to fit the MSD with a power law [49, 48], but to our knowledge, in this context, no one has introduced a diffusion constant that is truly constant. Instead, the diffusion constant varies with time. To facilitate comparing the power law fits to our other results, we introduce an "instantaneous" diffusion constant for such power-law fits. Because we only look at diffusion for short times, we do not estimate the cross-over time to normal diffusion as was done in [49]. However the instantaneous diffusion constant that we introduce is equal to $D^*(C)$ discussed in this paper.

All of our estimates of the diffusion constants are given in table A.6. They are reported in units of

$$10^{-2}\text{nm}^2/\mu\text{s} = 10^{-2}\mu\text{m}^2/\text{s} = 10^{-10}\text{cm}^2/\text{s}\,, \tag{4.1}$$

(see figure 2 of [7]). These results are reasonably consistent, but vary enough to make the analysis of biological data more difficult, but more accurate. With the exception of a few

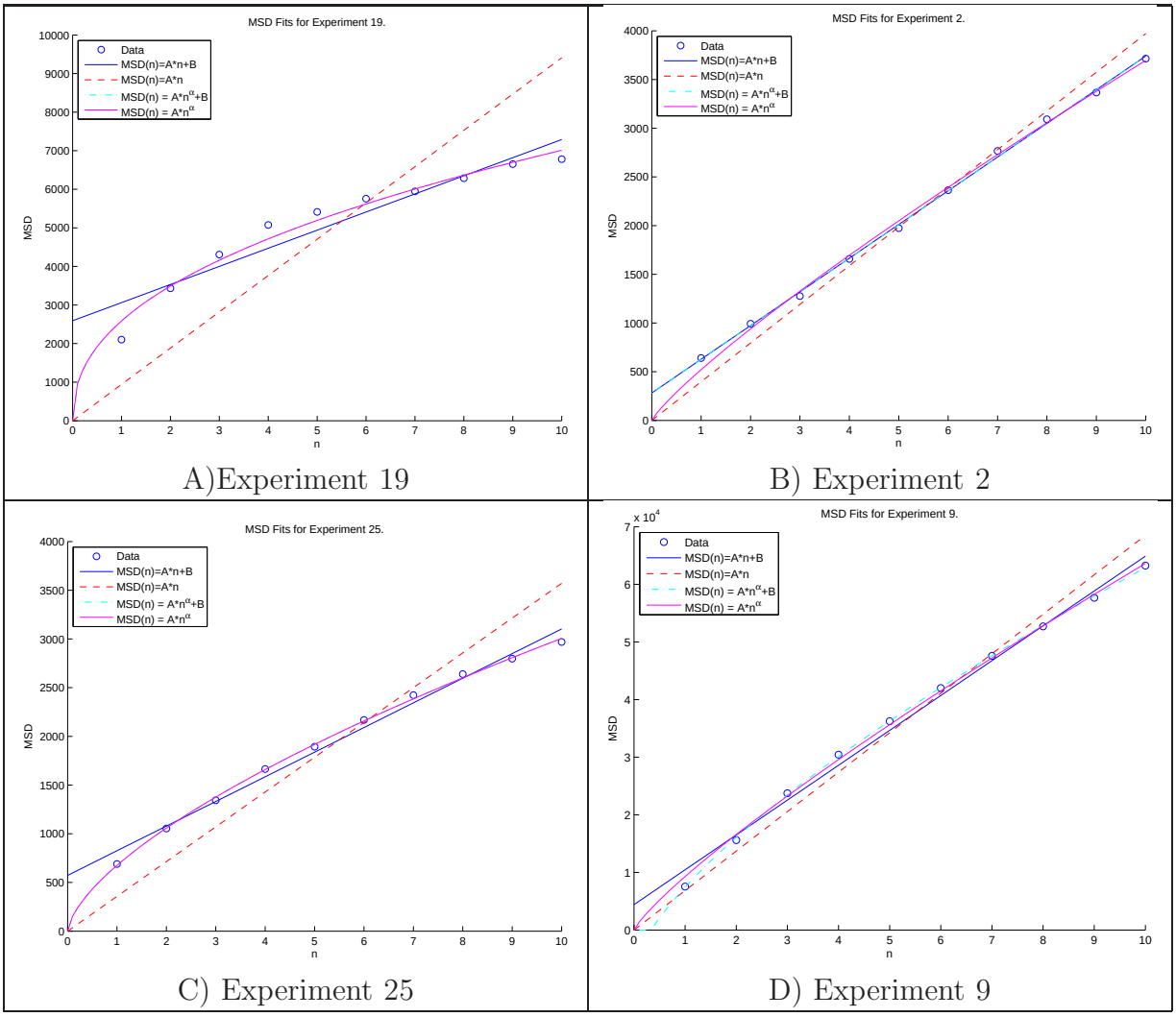


Figure 4.7: MSD Fits for the Experiments With the Smallest $k$ and $d$.

.

experiments, the fact that the ratios used to derive table 4.2 vary between 1/4 and 5/4 quantifies the essential consistency of the estimates. Importantly, the diffusion constant computed directly from the jump sizes and those computed from fitting the jump size distribution are independent of the order of the jumps and thus are not strongly affected by the autocorrelations.

Our fine scale diffusion constant $D$ is computed from the second moment $M_2$ (1.6) of the jump sizes. From (1.13),

$$\text{MSD}_n = M_2\, n \text{ or } \text{MSD}(t) = \frac{M_2}{\Delta t}\, t\,, \tag{4.2}$$

where $M_2$ is the second moment of the jump sizes. The standard definition [50] of the diffusion constant is given via

$$\text{MSD}(t) = 4\, D\, t\,. \tag{4.3}$$

Consequently

$$D = \frac{\text{MSD}(t)}{4\, t} = \frac{M_2}{4\, dt}\,, \tag{4.4}$$

For the biological data, the diffusion constants computed using this formula are given in column 2 of table A.6. We use this diffusion constant as a basis for comparing definitions of the diffusion constant.

The chi, Weibull and GEV models for the probability can also be used to compute the diffusion constant. Because the chi distribution can be used to estimate the dimension of the space in which the diffusion is occurring, it is the most interesting. In this case

$$D^\chi = \frac{M_2^\chi}{4\, dt}\,, \tag{4.5}$$

where $M_2^\chi$ is the second moment (3.8) of $c(r, \sigma, d)$ computed using (1.5). For the biological data, the values of $D^\chi$ are given in column 3 of table A.6. Comparisons to other diffusion constants is given in table 4.2. The chi distribution produces diffusion constants $D^\chi$ that are about one-half of the diffusion constant $D$ obtained from the second moment of the jump size data. We do not report diffusion constants using the Weibull and GEV distributions, but the results are similar.

Given that the linear models given in section 2.3 decompose the jumps into a deterministic part and a white noise $\epsilon_n$ random part (2.9), it makes good sense to estimate the diffusion constant using the white noise:

$$D^W = \frac{E(\epsilon_n^2)}{4\, dt}\,. \tag{4.6}$$

This definition has the advantage that it directly eliminates the effects of the autocorrelations. For the biological data, the values of $D^W$ are given in column 4 of table A.6 and comparisons with the other diffusion constants are given in table 4.2. The diffusion constant $D^W$ is about 10% lower than $D$, which is in reasonable agreement with our signal to noise ratio estimates given in column 10 of table A.3 that have a mean of 16%.

To motivate the form of the fits we use [38, 14] we return to the null hypothesis that the jumps $J_i$ are mean zero IID and jump lengths have a PDF that has a finite second moment $M_2$. Also assume that there is an error in the measurements of the positions that can be modeled by mean-zero IID random variables $E_n$ with a finite second moment $\mu$. In addition, if the errors and jumps are independent, then equation (1.10) becomes

$$\text{MSD}_n = E(\|P_n + E_n\|^2) = E(\|P_n\|^2) + E(\|E_n\|^2)\,. \tag{4.7}$$

Then, as in (1.12)

$$\text{MSD}_n = M_2\, n + \mu\,. \tag{4.8}$$

Consequently, for linear fits we use the form

$$\text{MSD}_n = A\, n + B\,, \quad 1 \le n \le N\,. \tag{4.9}$$

After some experimentation, we chose $N = 10$ so we are looking at a scale an order of magnitude greater than the finest scale. We also fit the data assuming that $B = 0$. Samples of these fits are shown in figure 4.7. For the biological data, the fits with $B \neq 0$ are significantly different from those with $B = 0$ and those with $B \neq 0$ are visually better and have smaller residuals than fits that assume $B = 0$. In both cases the diffusion constant is defined by

$$D^{\text{linear}} = \frac{A}{4\,\Delta t}\,. \tag{4.10}$$

For the biological data, comparisons of these diffusion constants to $D$ are given in table 4.2 where $D_1 = D^{\text{linear}}$ with $B \neq 0$ and $D_2 = D^{\text{linear}}$ with $B = 0$. The values of these diffusion constants are given in table columns 5 and 6A.6. So, on average, $D_1$ and $D_2$ are about about two thirds of $D$. Assuming $B = 0$ on average increases the diffusion constant. This can easily be seen in the plots of the MSD.

On the other hand, the values of $B$ are not close to 30nm and in fact many are negative. Perhaps the fact that the biological data are autocorrelated is a source of this discrepancy? The parameter $r = |B|/(A\,dt)$ is dimensionless and, for the biological data, satisfies $0.005 < r < 15.436$ with the mean of $r$ being 1.9055. So $B$ is always significant. Using exactly 10 steps is not important in these estimates.

The discussion so far suggest the use of a power law of the form

$$MSD_n = A\, n^{\alpha} + B\,, \tag{4.11}$$

to model the MSD. First write (4.11) in the form

$$MSD(t) = A\left(\frac{t}{\Delta t}\right)^{\alpha} + B \tag{4.12}$$

and then set

$$D(t) = \frac{1}{4}\frac{d}{dt}\text{MSD}(t) = \frac{\alpha\, A}{4\,\Delta t}\left(\frac{t}{dt}\right)^{\alpha-1} \tag{4.13}$$

and

$$D_n = \frac{\alpha\, A}{4\,\Delta t}\, n^{\alpha-1}\,. \tag{4.14}$$

Consequently, we define the instantaneous diffusion constant as

$$D' = \frac{\alpha\, A}{4\,\Delta t}\,. \tag{4.15}$$

In the case that $\alpha = 1$ this is consistent with the definitions using linear fits. When we assume $B \neq 0$ then $D_3 = D'$, while if $B = 0$, we set $D_4 = D'$. For the biological data if we set $\alpha_3 = \alpha$ when $B \neq 0$ and $\alpha_4 = \alpha$ when $B = 0$, then

$$-0.9801 \leq \alpha_3 - 1 \leq 0.2409\,, \quad -0.7769 \leq \alpha_4 - 1 \leq 0.1369 \tag{4.16}$$

In both cases there are 5 $\alpha - 1$ that are positive, indicating that the power-law fits are concave up and the track displays at least some super diffusion. The remaining 28 tracks are concave down. For 6 of the MSD plots, both $\alpha$ satisfy $|\alpha - 1| < .01$ indicating that these MSDs are nearly linear. This is close to what we guess by looking at the plots of the MSD.

For most of the biological data, the plots of the two fits are very close. Comparisons of these diffusion constants to $D$ are given in table 4.2 and the values of these diffusion constants are given in columns 7 and 8 of table A.6. So, on average, $D_3$ is 91% of $D$ and $D_2$ is about 83% of $D$.

There are several advantages to instantaneous diffusion constant. First, it separates the diffusion "constant" $D(t)$ into two parts, a true constant $D' = D(1)$ that has the correct units for a diffusion constant and a dimensionless power law determined by $\alpha$. If in (4.9) we choose $N = 3$ and use (4.11) to interpolate the data (not just fit the data) and make the approximation

$$D(1) \approx \frac{\mathrm{MSD}_2 - \mathrm{MSD}_1}{4\,\Delta t}\,, \tag{4.17}$$

then (4.2) gives

$$D' \approx \frac{M2}{4\,\Delta t} = D\,. \tag{4.18}$$

So this definition is compatible with our basic definition.

As mentioned in the introduction, to check our algorithms, we supplemented our data with an artificial data set generated using a random walk of 10,000 steps whose jumps have components that are normally distributed with mean zero and standard deviation of 10nm. For such a walk the diffusion coefficient is .0015 $\mathrm{nm}^2/\mu\mathrm{s}$. This would be reported as .15 in

table A.6. For this artificial data, the diffusion constant computed by our methods are

$$\begin{aligned} D &= 0.1522 \\ D^{\chi} &= 0.1569 \\ D^{W} &= 0.1522 \\ D_1 &= 0.1584 \\ D_2 &= 0.1563 \\ D_3 &= 0.1498 \\ D_4 &= 0.1535 \end{aligned}$$

This is very strong support for the correctness of our algorithms and codes.

# 5 Conclusions and Future Work

We have introduced some time series analysis methods, confirmed that they are appropriate for analyzing jump data and have used these methods to analyze some high quality data that track the movement of individual molecules in cell membranes. The tracked molecules are MHCI expressed on rat hepatoma cells. The analysis has led to two important results. First, the jump data have significant autocorrelations. To understand the correlations, we provide linear autoregressive models for the data and, based on these models, use the signal to noise ratio to quantify the importance of the correlations. Second, for the size of jumps data, we give four statistical models that fit the data well. These models suggest that the there are excessive short and long jumps compared to normal diffusion. The fit with the chi distribution provides an estimate of the fine-scale obstructions in the membrane. Based on these results, we introduced four alternate methods for estimating diffusion constants and applied these to the biological data. The diffusion constants are different, but these differences are consistent for the biological data set.

To facilitate the use of our results we provide MatLab and **R** code on our web page. In the future, we will start our analysis of SPT data with time series analysis and use this to characterize the properties of the membrane at the finest time and spatial scales reasonable for the data. We will then follow this with MSD analysis to quantify larger scale structures. If others can replicate our results for other types of data, then time-series analysis should be incorporated into the next generation of analysis tools for the analysis of SPT data.

There is still significant work to be done to fine tune the techniques presented.

- A critical need is to combine this work with an algorithm to decompose the paths into pieces that display the same type of motion. In [7], this is done by assigning the motion to be confined if the MSD for a segment of the path grows below linear, and to directed if the MSD grows faster than linear along with visual examination. An alternate methods is described in [23]. Once the decomposition is accomplished, the estimates of the diffusion constant given here can be combined with the information from the MSD to investigate larger scale structures in the membrane [40, 24, 25, 5, 46, 12].

- We are in the process of extending this analysis to particle tracking data from experiments in which the labeling is done using one or more sizes of quantum dots [36, 2]. An important problem is that the quantum dots blink and the length of the time that a dot can be off can be quite large. This makes the analysis more complicated. For the quantum dot analysis we are comparing the dynamic behavior of membrane molecules from experimental situations in which the labeling is assumed not to alter the dynamic behavior of the tagged molecule (as in the present paper) and also where the labeling is known to alter rector dynamics (as in [36]).

- There has been some work done in understanding how correlations in the successive jump angles affect the MSD [61, 39]. These models should be studied using time series analysis.

- Models of the deviation of the jump angles from uniform should provide additional insight into the fine scale structure of the membrane [57].
- Correlated random walks have been used to study the movement of several types of animals [26, 6, 53]. These studies have included the use of state-space models and Kalman filters. The correlations studied are significantly simpler than the ones found in our data, but the methods should be able to be extended to membrane data.
- It should be possible to refine the modeling of jump distributions using higher moments of the jump sizes [10].
- To better understand the membrane organization, it is important to build models that help explain the biological data. Models of membrane motion that include diffusion, transport, and attractive harmonic forces have been studied in [10]. We also have been studying similar models, but in our models the jump lengths have a Weibull distribution rather than the normal distribution used in [10]. We also use a force generated by a Lennard-Jones potential rather then a harmonic potential so that there are no long range forces between molecules.
- Autocorrelations (two-time correlations) of a particle confined to a box in the plane were studied in [13] and then used this to estimate the size of the confining box and then the results were used to estimate confinement zones in SPT data. In [28, 30] models containing two confinement zones have been used that capture some of the effects of cytoskeletal fence confinement zones (see also [47, 45, 31]).

# 6 Acknowledgments

We would like to thank Abigail McCarthy, Research Fishery Biologist, NOAA for providing references on the motion of animals and Flor A. Espinoza Hidalgo for corrections. This work was supported by NIH grants P20 GM067594 R01 GM49814, and R01 AI051575.

# References


[1] T. W. Anderson and D. A. Darling. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23:193–212, 1952.

[2] Nicholas L. Andrews, Keith A. Lidke, Janet R. Pfeiffer, Alan R. Burns, Bridget S. Wilson, Janet M. Oliver, and Diane S. Lidke. Actin restricts FcεRI diffusion and facilitates antigen-induced receptor immobilization. Nature Cell Biology, in press.

[3] Howard C. Berg. *Random Walks in Biology*. Princeton University Press, Princeton NJ, USA, 1993.

[4] Hugues Berry. Monte Carlo simulations of enzyme reactions in two dimensions: Fractal kinetics and spatial segregation. *Biophys. J.*, 83(4):1891–1901, 2002.

[5] Thomas Bickel. A note on confined diffusion. *Physica A: Statistical Mechanics and its Applications*, 377(1):2007, 2007.

[6] S. N. Luttich C. M. Bergman, J. A. Schaefer. Caribou movement as a correlated random walk. *Oecologia*, 123(3):364–374, 2000.

[7] G. George Capps, Samuel Pine, Michael Edidin, and Martha C. Zúñiga. Short class I major histocompatibility complex cytoplasmic tails differing in charge detect arbiters of lateral diffusion in the plasma membrane. *Biophys. J.*, 86(5):2896–2909, 2004.

[8] Richard J. Cherry, Patricia R. Smith, Ian E. G. Morrison, and Nelson Fernandez. Mobility of cell surface receptors: a re-evaluation. *FEBS letters*, 430(1):88–91, 1998.

[9] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, New York, USA, 2001.

[10] Sylvie Coscoy, Etienne Huguet, and Francois Amblard. Statistical analysis of sets of random walks: How to resolve their generating mechanism. *Bull. Math. Bio.*, 69(8):2467–2492, 2007.

[11] Maxime Dahan, Sabine Levi, Camilla Luccardini, Philippe Rostaing, Beatrice Riveau, and Antoine Triller. Diffusion dynamics of glycine receptors revealed by single-quantum dot tracking. *Science*, 302(5644):442–445, 2003.

[12] Frederic Daumas, Nicolas Destainville, Claire Millot, Andre Lopez, David Dean, and Laurence Salome. Confined diffusion without fences of a G-protein-coupled receptor as revealed by single particle tracking. *Biophys. J.*, 84(1):356–366, 2003.

[13] Nicolas Destainville and Laurence Salome. Quantification and correction of systematic errors due to detector time-averaging in single-molecule tracking experiments. *Biophys. J.*, 90(2):L17–19, 2006.

[14] Christian Dietrich, Bing Yang, Takahiro Fujiwara, Akihiro Kusumi, and Ken Jacobson. Relationship of lipid rafts to transient confinement zones detected by single particle tracking. *Biophys. J.*, 82(1):274–284, 2002.

[15] Mara M. Echarte, Luciana Bruno, Donna J. Arndt-Jovin, Thomas M. Jovin, and La I. Pietrasanta. Quantitative single particle tracking of NGFreceptor complexes: Transport is bidirectional but biased by longer retrograde run lengths. *FEBS letters*, 581(16):2905–2913, 2007.

[16] Michael Edidin. Lipids on the frontier: a century of cell-membrane bilayers. *Nat. Rev. Mol. Cell Biol.*, 4:414–418, 2003.

[17] Michael Edidin. The state of lipid rafts: From model membranes to cells. *Annual Review of Biophysics and Biomolecular Structure*, 32(1):257–283, 2003.

[18] Michael Edidin, Martha C. Zuniga, and Michael P. Sheetz. Truncation mutants define and locate cytoplasmic barriers to lateral mobility of membrane glycoprotiens. *Proc. Acad. Nat. Sci.*, 91(8):3378–3382, 1994.

[19] Takahiro Fujiwara, Ken Ritchie, Hideji Murakoshi, Ken Jacobson, and Akihiro Kusumi. Phospholipids undergo hop diffusion in compartmentalized cell membrane. *Cell Biology*, 157(6):2002, 2002.

[20] RN Ghosh and WW Webb. Automated detection and tracking of individual and clustered cell surface low density lipoprotein receptor molecules. *Biophys. J.*, 66(5):1301–1318, 1994.

[21] Ben N. G. Giepmans, Stephen R. Adams, Mark H. Ellisman, and Roger Y. Tsien. The fluorescent toolbox for assessing protein location and function. *Science*, 312(5771):217–224, 2006.

[22] R. Gorenflo and F. Mainardi. Fractional diffusion processes: Probability distributions and continuous time random walk. *Springer-Verlag*, LNP 621:148–166, 2003.

[23] Ryan Gutenkunsta, Nathaniel Newlands, Molly Lutcavage, and Leah Edelstein-Keshet. Inferring resource distributions from atlantic bluefin tuna movements: An analysis based on net displacement and length of track. *J. Theor. Biol.*, 245(2):243–257, 2006.

[24] S. Jin and A.S. Verkman. Single particle tracking of complex diffusion in membranes: Simulation and detection of barrier, raft, and interaction phenomena. *Journal of Physical Chemistry B*, 111(14):3625–3632, 2007.

[25] Songwan Jin, Peter M. Haggie, and A. S. Verkman. Single-particle tracking of membrane protein diffusion in a potential: Simulation, detection, and application to confined diffusion of CFTR Cl- channels. *Biophys. J.*, 93(3):1079–1088, 2007.

[26] Ian D. Jonsen, Ransom A. Myers, and Joanna Mills Flemming. Meta-analysis of animal movement using state-space models. *Ecology*, 84(11):30553063, 2003.

[27] Panagiotis Kabouridis. Lipid rafts in T cell receptor signalling (review). *Molecular Membrane Biology*, 23:49–57(9), 2006.

[28] Z. Kalay, P. E. Parris, and V. M. Kenkre. Effects of disorder in location and size of fence barriers on molecular motion in cell membranes. *J. Phys. Condens. Matter*, 20, 2008.

[29] N. C. Kenkel and D. J. Walker. Fractals in the biological sciences. *COENOSES*, 11:77–100, 1996.

[30] V. M. Kenkre, L. Giuggioli, and Z. Kalay. Molecular motion in cell membranes: analytic study of fence-hindered random walks. *Phys. Rev. E*, 77, 2008.

[31] Sharon Khan, Andy M. Reynolds, Ian E. G. Morrison, and Richard J. Cherry. Stochastic modeling of protein motions within cell membranes. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 71(4):041915, 2005.

[32] A. Kusumi, H. Murakoshi, K. Murase, and T. Fujiwara. Single-molecule imaging of diffusion, recruitment, and activation of signaling molecules in living cells. In S. Damjanovich, editor, *Biophysical Aspects of Transmembrane Signaling*. Springer-Verlag, Heidelberg, 2005.

[33] Akihiro Kusumi, Hiroshi Ike, Chieko Nakada, Kotono Murase, and Takahiro Fujiwara. Single-molecule tracking of membrane molecules: plasma membrane compartmentalization and dynamic assembly of raft-philic signaling molecules. *Semin Immunol*, 17(1):3–21, 2005.

[34] Akihiro Kusumi, Chieko Nakada, Ken Ritchie, Kotono Murase, Kenichi Suzuki, Hideji Murakoshi, Rinshi S. Kasai, Junko Kondo, and Takahiro Fujiwara. Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: High-speed single-molecule tracking of membrane molecules. *Annual Review of Biophysics and Biomolecular Structure*, 34(1):351–378, 2005.

[35] Diane S. Lidke and Donna J. Arndt-Jovin. Imaging takes a quantum leap. *Physiology*, 19(6):322–325, 2004.

[36] Diane S. Lidke, Keith A. Lidke, Bernd Rieger, Thomas M. Jovin, and Donna J. Arndt-Jovin. Reaching out for signals: filopodia sense EGF and respond by directed retrograde transport of activated receptors. *J. Cell Biol.*, 170(4):619–626, 2005.

[37] D.S. Lidke, P. Nagy, R. Heintzmann, D.J. Arndt-Jovin, J.N. Post, H. Grecco, E.A. Jares-Erijman, and T.M. Jovin. Quantum dot ligands provide new insights into receptor-mediated signal transduction. *Nature Biotechnology*, 22:198–203, 2004.

[38] Douglas S. Martin, Martin B. Forstner, and Joseph A. Kas. Apparent subdiffusion inherent to single particle tracking. *Biophysical Journal*, 83(4):2109–2117, 2002.

[39] C. E. McCulloch and M. L. Cain. Analyzing discrete movement data as a correlated random walk. *Ecology*, 70(2):383388, 1989.

[40] N. Meilhac, L. Le Guyader, L. Salomé, and N. Destainville. Detection of confinement and jumps in single-molecule membrane trajectories. *Phys. Rev. E*, 73:011915–9, 2006.

[41] Ralf Metzler and Joseh Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, 339:1–77, 2000.

[42] Kotono Murase, Takahiro Fujiwara, Yasuhiro Umemura, Kenichi Suzuki, Ryota Iino, Hidetoshi Yamashita, Mihoko Saito, Hideji Murakoshi, Ken Ritchie, and Akihiro Kusumi. Ultrafine membrane compartments for molecular diffusion as revealed by single molecule techniques. *Biophys. J.*, 86(6):4075–4093, 2004.

[43] Jr. Nicolau, Dan V., John F. Hancock, and Kevin Burrage. Sources of anomalous diffusion on cell membranes: A Monte Carlo study. *Biophys. J.*, 92(6):1975–1987, 2007.

[44] J. M. Oliver, J. R. Pfeiffer, Z. Surviladze, S. L. Steinberg, K Leiderman, M Sanders, C Wofsy, J Zhang, HY Fan, N Andrews, S Bunge, TJ Boyle, P Kotula, and BS Wilson. Membrane receptor mapping: the membrane topography of Fc$\epsilon$RI signaling. In P.J. Quinn, editor, *Subcellular Biochemistry 37: Membrane Dynamics and Domains*, pages 3–34. Kluwer Academic/Plenum Publishers, 2004.

[45] A.M. Reynolds. On the anomalous diffusion characteristics of membrane-bound proteins. *Physics Letters A*, 342(5-6):439–442, 2005.

[46] Ken Ritchie, Xiao-Yuan Shan, Junko Kondo, Kokoro Iwasawa, Takahiro Fujiwara, and Akihiro Kusumi. Detection of non-brownian diffusion in the cell membrane in single molecule tracking. *Biophys. J.*, 88(3):2266–2277, 2005.

[47] M. J. Saxton. Anomalous diffusion due to obstacles: a Monte Carlo study. *Biophys. J.*, 66(2):394–401, 1994.

[48] Michael J. Saxton. Anomalous subdiffusion in fluorescence photobleaching recovery: A Monte Carlo study. *Biophys. J.*, 81(4):2226–2240, 2001.

[49] MJ Saxton. Anomalous diffusion due to binding: a monte carlo study. *Biophys. J.*, 70(3):1250–1262, 1996.

[50] M.J. Saxton and K. Jacobson. Single particle tracking: Applications to membrane dynamics. *Ann. Rev. Biophys. Biomol. Struct.*, 26:373–399, 1997.

[51] S.S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3 and 4):591–611, 1965.

[52] R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications With R Examples*. Springer Verlag, New York, 2006.

[53] John R. Sibert, Michael K. Musyl, and Richard W. Brill. Horizontal movements of bigeye tuna (Thunnus obesus) near Hawaii determined by Kalman filter analysis of archival tagging data. *Fisheries Oceanography*, 12(3):141151, 2003.

[54] Patricia R. Smith, Ian E. G. Morrison, Keith M. Wilson, Nelson Fernandez, and Richard J. Cherry. Anomalous diffusion of major histocompatibility complex class I molecules on hela cells determined by single particle tracking. *Biophys. J.*, 76(6):3331–3344, 1999.

[55] Kenichi Suzuki, Ken Ritchie, Eriko Kajikawa, Takahiro Fujiwara, and Akihiro Kusumi. Rapid hop diffusion of a G-protein-coupled receptor in the plasma membrane as revealed by single-molecule techniques. *Biophys. J.*, 88(5):3659–3680, 2005.

[56] Qing Tang and Michael Edidin. Lowering the barriers to random walks on the cell surface. *Biophys. J.*, 84(1):400–407, 2003.

[57] Concepcion Tojo and Panos Argyrakis. Correlated random walk in continuous space. *Phys. Rev. E*, 54(1):58–63, Jul 1996.

[58] Marija Vrljic, Stefanie Y. Nishimura, Sophie Brasselet, W. E. Moerner, and Harden M. McConnell. Translational diffusion of individual class II MHC membrane proteins in cells. *Biophys. J.*, 83(5):2681–2692, 2002.

[59] Stefan Wieser, Manuel Moertelmaier, Elke Fuertbauer, Hannes Stockinger, and Gerhard J. Schutz. (un)confined diffusion of CD59 in the plasma membrane determined by high-resolution single molecule microscopy. *Biophys. J.*, 92(10):3719–3728, May 2007.

[60] Bridget S. Wilson, Stanly L. Steinberg, Karin Liederman, Janet R. Pfeiffer, Zurab Surviladze, Jun Zhang, Lawrence E. Samelson, Li-hong Yang, Paul G. Kotula, and Janet M. Oliver. Markers for detergent-resistant lipid rafts occupy distinct and dynamic domains in native membranes. *Mol. Biol. Cell*, 15(6):2580–2592, 2004.

[61] Mingming Wu, John W. Roberts, Sue Kim, Donald L. Koch, and Matthew P. DeLisa. Collective bacterial dynamics revealed using a three-dimensional popul ation-scale defocused particle tracking technique. *Appl. Environ. Microbiol.*, 72(7):4987–4994, 2006.

[62] Wenxia Ying. *Modeling Non-stationary Extremal Events Via Bayesian Method*. PhD thesis, University of New Mexico, Albuquerque, NM, USA, expected in 2008.

# A Results of the Analysis of the Biological Data

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| # | N | max $L$ | mean $L$ | % small $L$ | $\|\mu\|$ | $M_2$ | $\rho_1$ | $\rho_2$ | $p$ | % $S/N$ |
| 1 | 1007 | 56 | 13 | 14.5 | 1.915 | 226.7 | -0.19 | -0.03 | 9 | 11.4 |
| 2 | 2113 | 316 | 17 | 17.5 | 0.413 | 640.2 | -0.23 | -0.05 | 31 | 22.8 |
| 3 | 1300 | 142 | 31 | 2.5 | 0.889 | 1326.1 | 0.07 | -0.02 | 4 | 5.7 |
| 4 | 1447 | 197 | 33 | 3.3 | 1.360 | 1687.0 | 0.16 | -0.02 | 2 | 5.7 |
| 5 | 1229 | 170 | 47 | 1.0 | 0.264 | 2990.3 | 0.21 | -0.07 | 5 | 12.8 |
| 6 | 893 | 189 | 43 | 1.0 | 1.648 | 2461.3 | 0.12 | -0.00 | 9 | 10.0 |
| 7 | 2018 | 241 | 41 | 1.0 | 2.038 | 2374.5 | 0.08 | 0.02 | 8 | 8.1 |
| 8 | 1269 | 217 | 34 | 2.1 | 0.107 | 1777.1 | 0.25 | -0.01 | 2 | 15.0 |
| 9 | 932 | 257 | 76 | 0.3 | 3.182 | 7530.1 | 0.03 | 0.01 | 4 | 5.4 |
| 10 | 1126 | 123 | 29 | 3.1 | 0.511 | 1143.0 | -0.18 | -0.07 | 6 | 16.5 |
| 11 | 1171 | 315 | 62 | 0.8 | 1.792 | 5368.8 | 0.04 | 0.04 | 3 | 5.8 |
| 12 | 1004 | 263 | 76 | 0.5 | 1.350 | 7865.4 | -0.22 | 0.02 | 1 | 9.8 |
| 13 | 1013 | 215 | 41 | 1.3 | 0.118 | 2333.1 | -0.19 | -0.02 | 9 | 19.9 |
| 14 | 789 | 245 | 45 | 1.5 | 0.217 | 3023.6 | 0.10 | 0.01 | 1 | 6.0 |
| 15 | 832 | 296 | 70 | 1.0 | 1.239 | 6855.3 | -0.05 | -0.02 | 1 | 3.6 |
| 16 | 1012 | 215 | 41 | 1.3 | 0.100 | 2334.7 | -0.19 | -0.02 | 9 | 19.9 |
| 17 | 912 | 204 | 55 | 0.3 | 3.990 | 3936.0 | -0.08 | 0.06 | 2 | 2.8 |
| 18 | 1059 | 312 | 65 | 0.3 | 0.635 | 5903.6 | 0.07 | 0.05 | 16 | 12.5 |
| 19 | 1039 | 231 | 31 | 7.7 | 0.322 | 2101.2 | -0.18 | -0.11 | 16 | 42.4 |
| 20 | 977 | 122 | 23 | 5.6 | 0.293 | 799.7 | -0.13 | 0.00 | 1 | 4.5 |
| 21 | 916 | 116 | 17 | 7.6 | 0.419 | 374.4 | -0.22 | -0.12 | 19 | 32.6 |
| 22 | 1569 | 118 | 14 | 10.0 | 0.155 | 277.3 | -0.18 | -0.09 | 28 | 26.6 |
| 23 | 623 | 128 | 34 | 2.2 | 1.307 | 1623.2 | -0.32 | 0.06 | 17 | 37.6 |
| 24 | 1196 | 242 | 44 | 1.2 | 0.701 | 2946.2 | -0.19 | -0.01 | 10 | 18.0 |
| 25 | 1206 | 217 | 22 | 3.6 | 0.205 | 689.2 | -0.21 | -0.05 | 18 | 19.7 |
| 26 | 660 | 306 | 66 | 0.1 | 0.470 | 6027.2 | -0.14 | 0.02 | 4 | 14.0 |
| 27 | 791 | 409 | 46 | 1.5 | 0.939 | 3104.7 | -0.06 | -0.05 | 3 | 5.0 |
| 28 | 1673 | 131 | 30 | 2.6 | 0.295 | 1230.4 | -0.16 | -0.05 | 2 | 5.8 |
| 29 | 1379 | 282 | 49 | 1.2 | 0.995 | 3507.2 | 0.10 | -0.13 | 3 | 5.0 |
| 30 | 1381 | 204 | 53 | 0.7 | 0.200 | 3793.0 | -0.32 | -0.01 | 5 | 29.7 |
| 31 | 2117 | 170 | 37 | 2.5 | 0.116 | 1908.2 | -0.20 | -0.10 | 9 | 15.3 |
| 32 | 1044 | 262 | 20 | 5.8 | 0.224 | 603.4 | -0.28 | -0.03 | 8 | 25.8 |
| 33 | 1311 | 163 | 43 | 1.1 | 0.348 | 2416.2 | -0.37 | -0.08 | 14 | 52.3 |

Table A.3: Results of the Statistical Analysis of the Biological Data.

| | Weibull | | Chi | | Long-Short | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| # | $\lambda$ | $k$ | $\sigma$ | $d$ | $s$ | $\alpha$ | $\beta$ | $k$ |
| 1 | 14 | 1.7 | 11 | 1.6 | 25 | 1.9 | 4.3 | 7.3 |
| 2 | 15 | 1.5 | 13 | 1.3 | 12 | 2.0 | 2.0 | 2.9 |
| 3 | 34 | 1.9 | 25 | 1.9 | 105 | 2.0 | 10.3 | 19.4 |
| 4 | 34 | 1.7 | 28 | 1.6 | 43 | 2.0 | 3.0 | 4.8 |
| 5 | 51 | 1.8 | 39 | 1.7 | 137 | 1.9 | 7.8 | 13.8 |
| 6 | 46 | 1.9 | 34 | 1.8 | 92 | 2.0 | 5.5 | 10.0 |
| 7 | 43 | 1.8 | 32 | 1.8 | 70 | 2.0 | 4.1 | 7.3 |
| 8 | 35 | 1.8 | 27 | 1.7 | 48 | 2.0 | 3.2 | 5.6 |
| 9 | 84 | 2.0 | 60 | 2.0 | 276 | 2.0 | 12.9 | 25.2 |
| 10 | 31 | 1.8 | 24 | 1.7 | 46 | 2.0 | 3.6 | 6.2 |
| 11 | 66 | 1.8 | 52 | 1.7 | 145 | 1.9 | 5.8 | 10.1 |
| 12 | 83 | 1.8 | 65 | 1.7 | 222 | 1.9 | 7.6 | 13.3 |
| 13 | 43 | 1.8 | 32 | 1.8 | 51 | 2.2 | 2.9 | 5.1 |
| 14 | 47 | 1.8 | 36 | 1.7 | 63 | 2.1 | 3.2 | 5.6 |
| 15 | 74 | 1.8 | 56 | 1.8 | 93 | 2.1 | 3.0 | 5.3 |
| 16 | 43 | 1.8 | 32 | 1.8 | 51 | 2.2 | 2.9 | 5.1 |
| 17 | 60 | 1.8 | 45 | 1.8 | 114 | 2.0 | 5.0 | 9.0 |
| 18 | 69 | 1.8 | 53 | 1.7 | 144 | 1.9 | 5.5 | 9.5 |
| 19 | 26 | 1.4 | 24 | 1.2 | 18 | 1.8 | 1.9 | 2.6 |
| 20 | 25 | 1.6 | 21 | 1.5 | 38 | 1.8 | 3.6 | 5.7 |
| 21 | 18 | 1.9 | 14 | 1.8 | 149 | 1.9 | 54.6 | 103.0 |
| 22 | 15 | 1.9 | 11 | 1.9 | 20 | 2.2 | 3.2 | 5.7 |
| 23 | 38 | 1.6 | 33 | 1.5 | 135 | 1.7 | 9.5 | 15.0 |
| 24 | 46 | 1.7 | 37 | 1.6 | 71 | 1.9 | 3.7 | 6.0 |
| 25 | 24 | 2.0 | 17 | 2.0 | 229 | 2.0 | 93.6 | 188.1 |
| 26 | 68 | 1.8 | 51 | 1.8 | 94 | 2.1 | 3.4 | 6.0 |
| 27 | 49 | 1.8 | 37 | 1.8 | 141 | 1.9 | 8.9 | 16.1 |
| 28 | 32 | 1.8 | 25 | 1.7 | 54 | 2.0 | 4.1 | 7.1 |
| 29 | 51 | 1.7 | 41 | 1.6 | 71 | 1.9 | 3.3 | 5.4 |
| 30 | 58 | 1.9 | 44 | 1.8 | 172 | 1.9 | 9.5 | 17.6 |
| 31 | 40 | 1.7 | 31 | 1.6 | 86 | 1.9 | 5.6 | 9.6 |
| 32 | 21 | 1.8 | 16 | 1.7 | 25 | 2.1 | 2.8 | 4.8 |
| 33 | 48 | 1.9 | 35 | 1.9 | 135 | 2.0 | 9.5 | 18.2 |

Table A.4: The Mean Square Fit Parameters.

| | GEV (LS) | | | | zeros | GEV (ML) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| # | $\sigma$ | $\mu$ | $\xi$ | $k$ | % | $\sigma$ | $\mu$ | $\xi$ | $k$ | $p$ |
| 1 | 6.5 | 8.9 | 0.05 | 19.9 | 0.20 | 9.1 | 5.8 | 0.07 | 13.9 | 0.7 |
| 2 | 7.1 | 8.6 | 0.36 | 2.8 | 0.47 | 8.4 | 6.8 | 0.42 | 2.4 | 0.8 |
| 3 | 14.9 | 22.6 | -0.02 | -59.2 | 0.00 | 22.9 | 13.8 | 0.04 | 25.6 | 0.4 |
| 4 | 15.7 | 21.7 | 0.18 | 5.6 | 0.14 | 21.8 | 14.9 | 0.17 | 5.7 | 0.6 |
| 5 | 23.0 | 32.9 | 0.01 | 120.5 | 0.00 | 33.7 | 21.3 | 0.04 | 23.8 | 0.9 |
| 6 | 20.0 | 30.2 | 0.04 | 27.5 | 0.00 | 30.6 | 18.6 | 0.06 | 16.4 | 0.5 |
| 7 | 18.8 | 28.5 | 0.08 | 12.2 | 0.00 | 28.6 | 17.8 | 0.11 | 9.3 | 0.7 |
| 8 | 15.7 | 22.9 | 0.14 | 7.3 | 0.00 | 22.8 | 15.0 | 0.16 | 6.1 | 1.0 |
| 9 | 35.7 | 55.8 | -0.07 | -13.4 | 0.21 | 56.9 | 33.2 | -0.00 | -333.3 | 0.8 |
| 10 | 13.7 | 20.2 | 0.11 | 8.9 | 0.18 | 20.4 | 12.8 | 0.08 | 13.0 | 0.9 |
| 11 | 29.7 | 42.5 | 0.05 | 18.8 | 0.00 | 43.2 | 27.6 | 0.09 | 11.5 | 0.9 |
| 12 | 37.4 | 52.5 | 0.02 | 59.5 | 0.40 | 54.2 | 34.1 | 0.06 | 16.9 | 0.9 |
| 13 | 18.5 | 28.2 | 0.13 | 7.8 | 0.00 | 28.4 | 17.7 | 0.11 | 9.1 | 0.5 |
| 14 | 20.6 | 30.5 | 0.12 | 8.6 | 0.13 | 30.7 | 19.6 | 0.15 | 6.7 | 0.8 |
| 15 | 32.3 | 48.9 | 0.13 | 7.8 | 0.00 | 49.4 | 31.0 | 0.09 | 11.2 | 0.6 |
| 16 | 18.5 | 28.2 | 0.12 | 8.1 | 0.00 | 28.4 | 17.7 | 0.11 | 9.1 | 0.9 |
| 17 | 25.9 | 39.1 | 0.09 | 11.6 | 0.00 | 39.9 | 24.1 | 0.03 | 37.0 | 0.8 |
| 18 | 30.9 | 44.3 | 0.06 | 17.9 | 0.09 | 45.2 | 28.9 | 0.10 | 10.2 | 0.9 |
| 19 | 13.2 | 14.6 | 0.51 | 2.0 | 0.10 | 14.6 | 12.3 | 0.51 | 1.9 | 0.6 |
| 20 | 11.6 | 15.2 | 0.21 | 4.7 | 0.10 | 15.3 | 10.6 | 0.16 | 6.3 | 0.9 |
| 21 | 8.2 | 11.9 | -0.09 | -11.7 | 0.00 | 12.0 | 7.3 | 0.04 | 25.6 | 0.8 |
| 22 | 6.6 | 10.1 | 0.07 | 14.5 | 0.32 | 9.9 | 6.2 | 0.09 | 11.1 | 0.5 |
| 23 | 18.0 | 23.0 | 0.17 | 5.9 | 0.00 | 24.0 | 16.2 | 0.05 | 18.9 | 0.7 |
| 24 | 21.0 | 28.6 | 0.17 | 6.0 | 0.00 | 28.8 | 19.6 | 0.18 | 5.5 | 0.9 |
| 25 | 10.1 | 16.4 | -0.09 | -11.4 | 0.08 | 16.2 | 9.4 | 0.06 | 18.2 | 0.8 |
| 26 | 29.7 | 45.2 | 0.08 | 12.3 | 0.00 | 45.8 | 28.4 | 0.11 | 9.0 | 0.5 |
| 27 | 21.9 | 31.9 | -0.01 | -76.3 | 0.25 | 32.3 | 20.3 | 0.10 | 10.3 | 0.9 |
| 28 | 14.3 | 20.7 | 0.10 | 9.7 | 0.12 | 21.0 | 13.3 | 0.08 | 12.3 | 0.9 |
| 29 | 23.3 | 32.5 | 0.17 | 5.8 | 0.00 | 33.0 | 21.9 | 0.14 | 7.0 | 0.8 |
| 30 | 25.5 | 38.2 | -0.00 | -208.3 | 0.14 | 39.0 | 23.6 | 0.03 | 34.5 | 0.4 |
| 31 | 18.0 | 24.9 | 0.06 | 16.3 | 0.00 | 25.5 | 16.8 | 0.09 | 11.5 | 0.7 |
| 32 | 9.2 | 13.6 | 0.16 | 6.2 | 0.29 | 13.7 | 8.6 | 0.13 | 7.8 | 0.4 |
| 33 | 20.5 | 31.6 | -0.05 | -21.5 | 0.08 | 32.2 | 19.2 | -0.02 | -62.5 | 0.9 |

Table A.5: The GEV Parameters by Least Squares and Maximum Likelihood.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| | PDF | | White | Linear | | Power | |
| | | | | $B \neq 0$ | $B = 0$ | $B \neq 0$ | $B = 0$ |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| # | $D$ | $D^{\chi}$ | $D^{W}$ | $D_1$ | $D_2$ | $D_3$ | $D_4$ |
| 1 | 0.17 | 0.16 | 0.16 | 0.12 | 0.12 | 0.13 | 0.13 |
| 2 | 0.48 | 0.17 | 0.41 | 0.26 | 0.30 | 0.26 | 0.33 |
| 3 | 0.99 | 0.90 | 0.96 | 0.92 | 0.99 | 1.37 | 1.09 |
| 4 | 1.27 | 0.92 | 1.22 | 1.51 | 1.51 | 1.54 | 1.54 |
| 5 | 2.24 | 2.01 | 2.06 | 1.81 | 2.22 | 4.46 | 2.65 |
| 6 | 1.85 | 1.60 | 1.72 | 2.34 | 2.23 | 2.10 | 2.10 |
| 7 | 1.78 | 1.42 | 1.68 | 2.14 | 2.07 | 2.05 | 1.99 |
| 8 | 1.33 | 0.95 | 1.20 | 1.67 | 1.71 | 2.38 | 1.79 |
| 9 | 5.65 | 5.24 | 5.47 | 4.54 | 5.14 | 7.27 | 5.82 |
| 10 | 0.86 | 0.74 | 0.79 | 0.45 | 0.54 | 0.51 | 0.62 |
| 11 | 4.03 | 3.40 | 3.86 | 4.12 | 4.22 | 5.08 | 4.40 |
| 12 | 5.90 | 5.25 | 5.57 | 2.60 | 3.40 | 3.97 | 4.09 |
| 13 | 1.75 | 1.38 | 1.61 | 0.74 | 1.01 | 1.23 | 1.23 |
| 14 | 2.27 | 1.67 | 2.20 | 2.73 | 2.64 | 2.62 | 2.54 |
| 15 | 5.14 | 4.18 | 5.00 | 3.49 | 4.10 | 5.94 | 4.75 |
| 16 | 1.75 | 1.38 | 1.61 | 0.74 | 1.01 | 1.34 | 1.23 |
| 17 | 2.95 | 2.72 | 2.89 | 3.29 | 3.04 | 2.22 | 2.67 |
| 18 | 4.43 | 3.64 | 4.09 | 5.71 | 5.37 | 5.37 | 5.01 |
| 19 | 1.58 | 0.54 | 1.26 | 0.35 | 0.71 | 0.84 | 0.84 |
| 20 | 0.60 | 0.49 | 0.58 | 0.49 | 0.49 | 0.49 | 0.49 |
| 21 | 0.28 | 0.26 | 0.21 | 0.07 | 0.12 | 0.15 | 0.15 |
| 22 | 0.21 | 0.18 | 0.17 | 0.08 | 0.12 | 0.14 | 0.14 |
| 23 | 1.22 | 1.17 | 0.94 | 0.51 | 0.63 | 0.74 | 0.74 |
| 24 | 2.21 | 1.62 | 2.01 | 1.20 | 1.43 | 1.50 | 1.64 |
| 25 | 0.52 | 0.45 | 0.42 | 0.19 | 0.27 | 0.33 | 0.33 |
| 26 | 4.52 | 3.52 | 4.20 | 2.88 | 3.30 | 3.72 | 3.72 |
| 27 | 2.33 | 1.86 | 2.26 | 1.37 | 1.71 | 2.58 | 2.04 |
| 28 | 0.92 | 0.79 | 0.89 | 0.51 | 0.60 | 0.68 | 0.68 |
| 29 | 2.63 | 2.04 | 2.52 | 1.94 | 2.29 | 3.35 | 2.65 |
| 30 | 2.84 | 2.60 | 2.35 | 0.73 | 1.19 | 1.24 | 1.47 |
| 31 | 1.43 | 1.21 | 1.29 | 0.44 | 0.71 | 1.09 | 0.87 |
| 32 | 0.45 | 0.33 | 0.39 | 0.07 | 0.16 | 0.18 | 0.18 |
| 33 | 1.81 | 1.72 | 1.33 | 0.13 | 0.50 | 0.57 | 0.43 |

Table A.6: The Diffusion Constants.

| # | name |
|---|---|
| 1 | 221_00000000XptlX1_xy |
| 2 | 221_00000000XptlX3_xy |
| 3 | 221_000092806XptlX1a_xy |
| 4 | 221_00061418XptlX2_xy |
| 5 | 911_001255104XptlX4_xy |
| 6 | 911_00542314rtXptlX6_xy |
| 7 | 911_01183800_pt4_xy |
| 8 | 911_01353510ptlX4_xy |
| 9 | C48_012940_pt1b_xy |
| 10 | C48_013812_pt3_xy |
| 11 | C48_014210_pt1_xy |
| 12 | C48_015022_pt2a_xy |
| 13 | wildtypeLd_00435612XptlX1_xy |
| 14 | wtLd_00212.50_xy |
| 15 | wtLd_003650.10_xy |
| 16 | wtLd_00435612Xptl1_xy |
| 17 | Z1_003224_pt1_xy |
| 18 | Z1_003224_pt2_xy |
| 19 | Z1_00.42.33_pt1_xy |
| 20 | Z1_00.46.00_Pt1_xy |
| 21 | Z2_00.09.20_Pt1_xy |
| 22 | Z2_00.10.11_Pt1_xy |
| 23 | Z2_00.13.23_Pt2_xy |
| 24 | Z2_00.14.45_pt2_xy |
| 25 | Z2_00.18.58_Pt1_xy |
| 26 | Z3_004731_pt1b_xy |
| 27 | Z3_004950_pt1_xy |
| 28 | Z3_005430_pt2_xy |
| 29 | Z3_010100_pt1_xy |
| 30 | Z3_010444_pt1_xy |
| 31 | Z4_0001_pt1_xy |
| 32 | Z4_0003_pt1_xy |
| 33 | Z4_001600_pt1_xy |

Table A.7: Experiment Number and Name.

# B The Particle Paths

The figure of the particle paths are grouped according to the experimental conditions under which the paths were observed. There are from 3 to 5 figures per group.

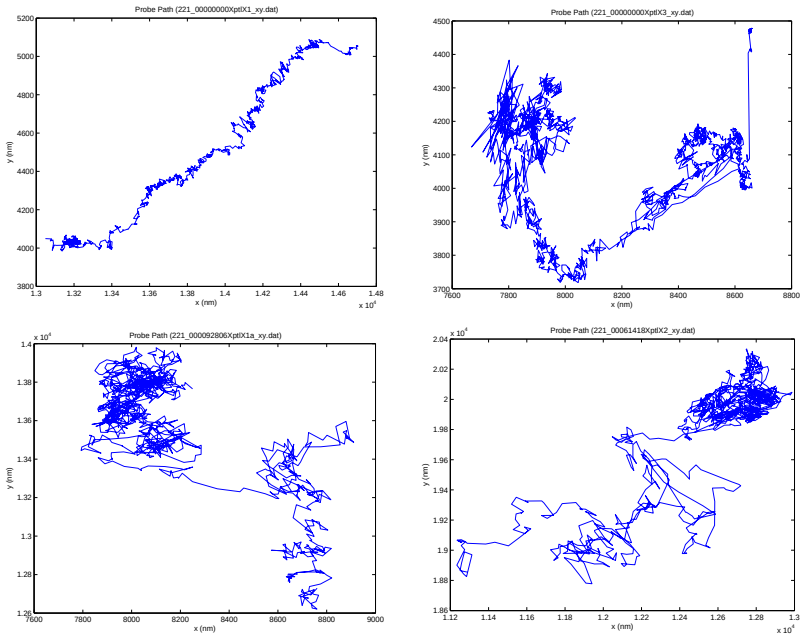


Figure B.8: Paths for Experiments: 1, 2, 3, 4

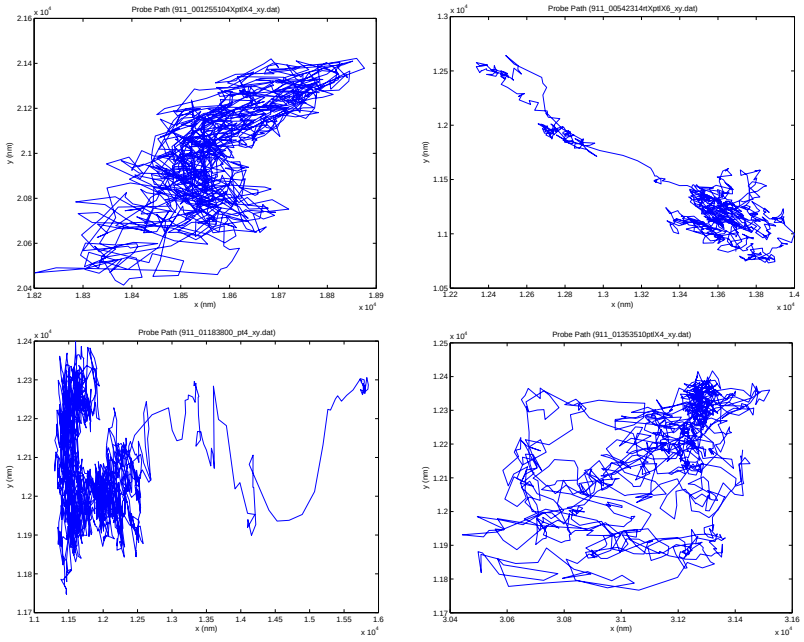


Figure B.9: Paths for Experiments: 5, 6, 7, 8

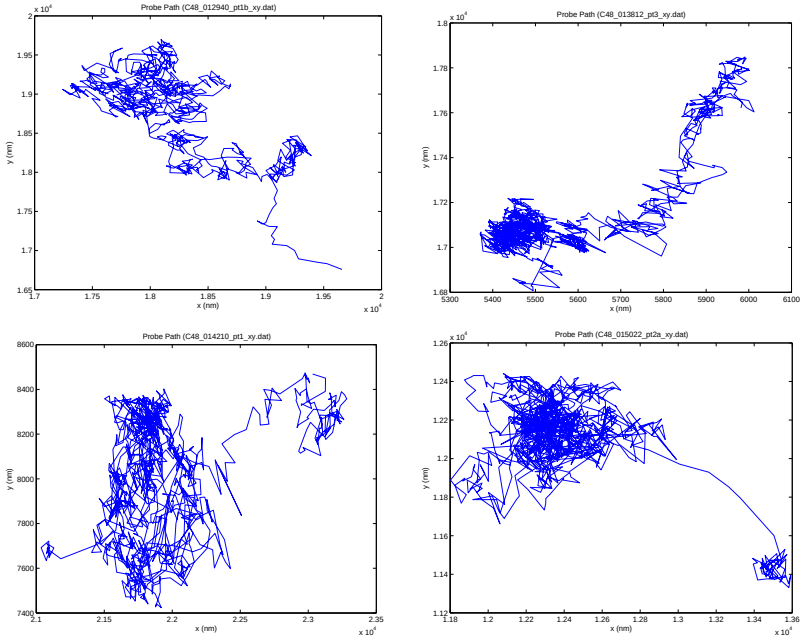


Figure B.10: Paths for Experiments: 9, 10, 11, 12

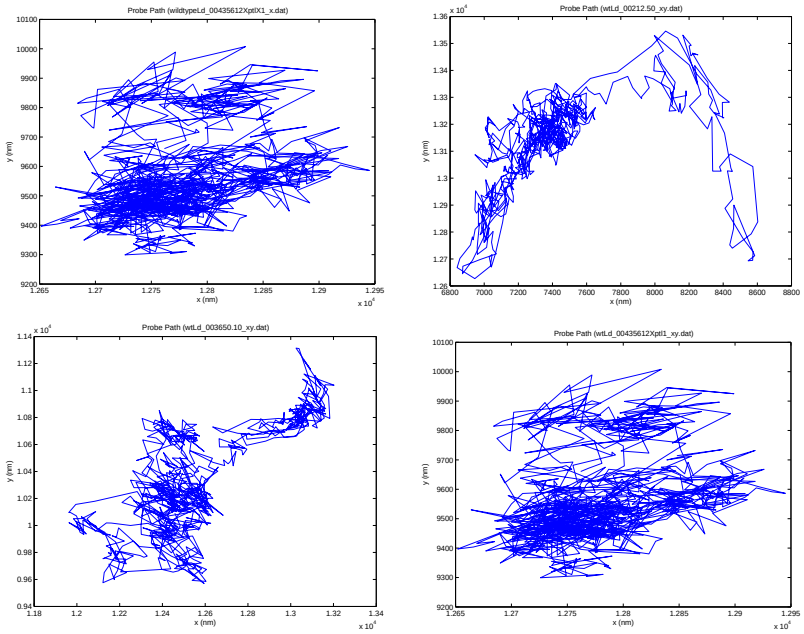


Figure B.11: Paths for Experiments: 13, 14, 15, 16

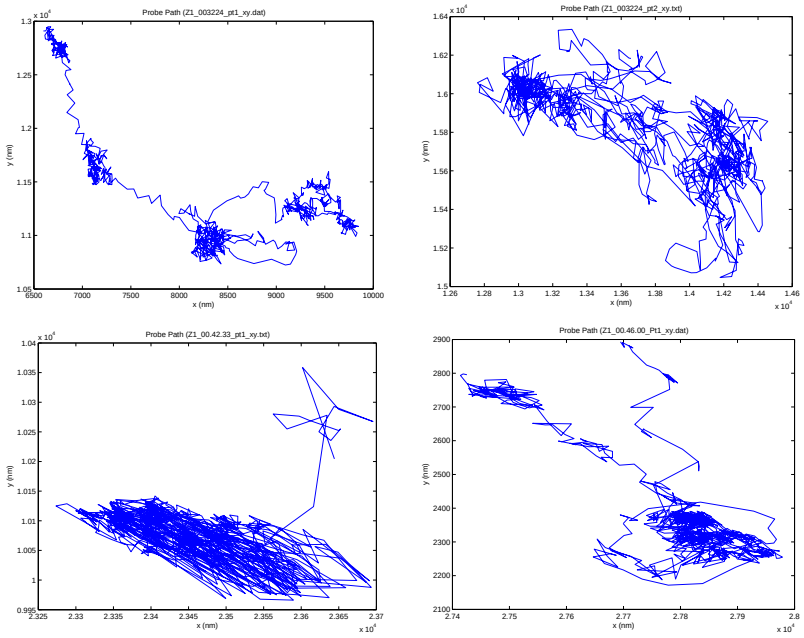


Figure B.12: Paths for Experiments: 17, 18, 19, 20

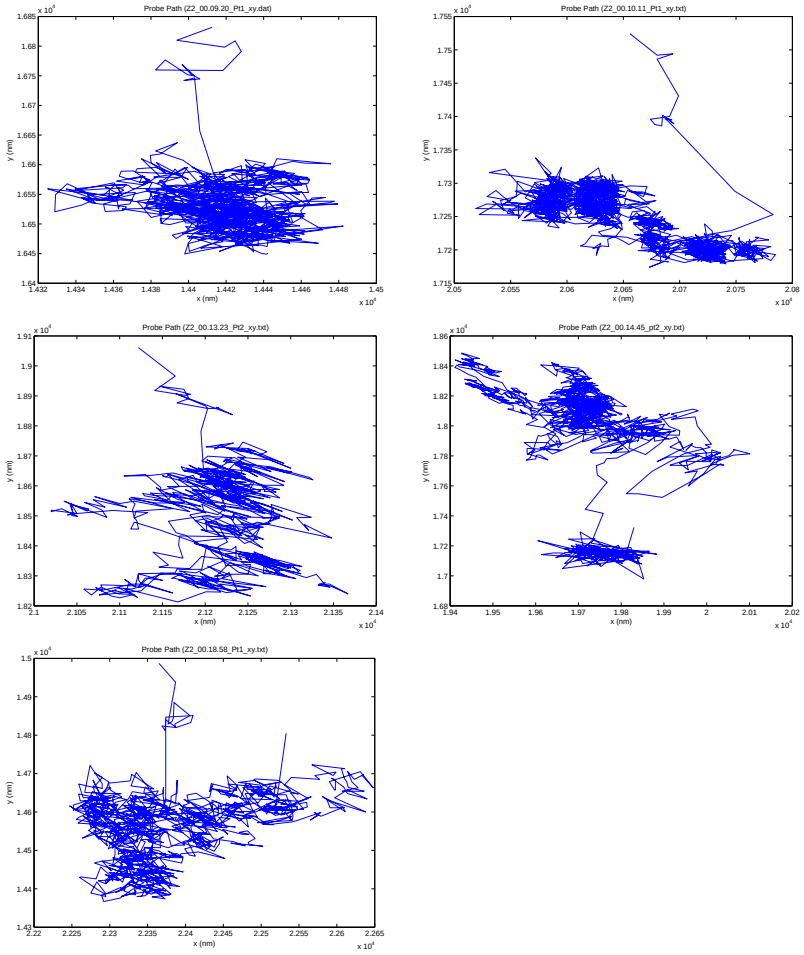


Figure B.13: Paths for Experiments: 21, 22, 23, 24, 25

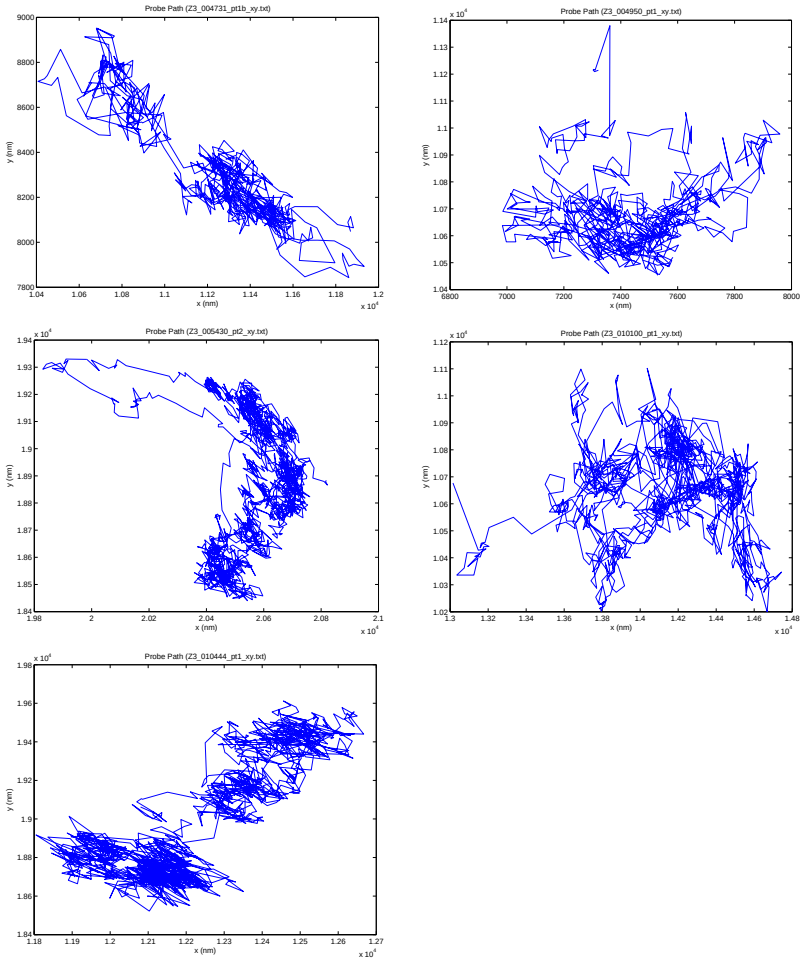


Figure B.14: Paths for Experiments: 26, 27, 28, 29, 30

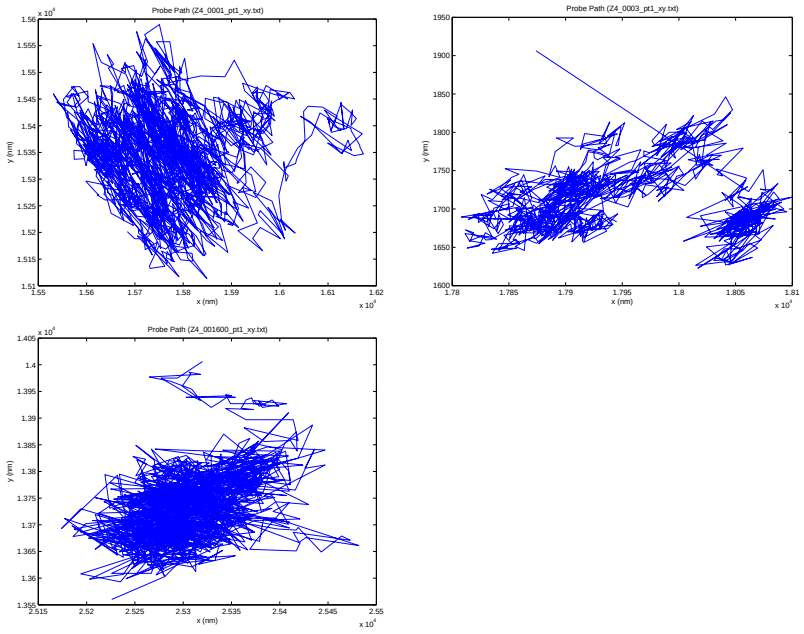


Figure B.15: Paths for Experiments: 31, 32, 33

# C The PDF fits

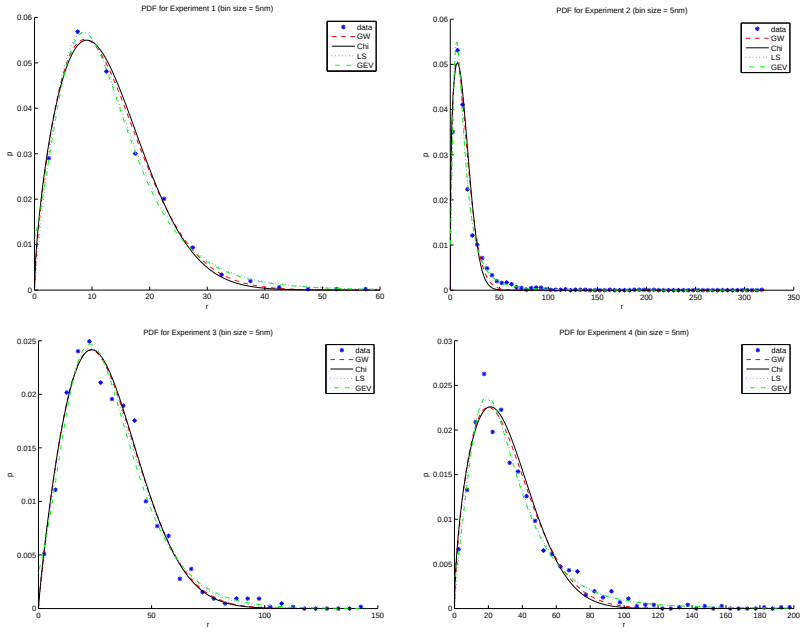


Figure C.16: PDF Fits for Experiments: 1, 2, 3, 4

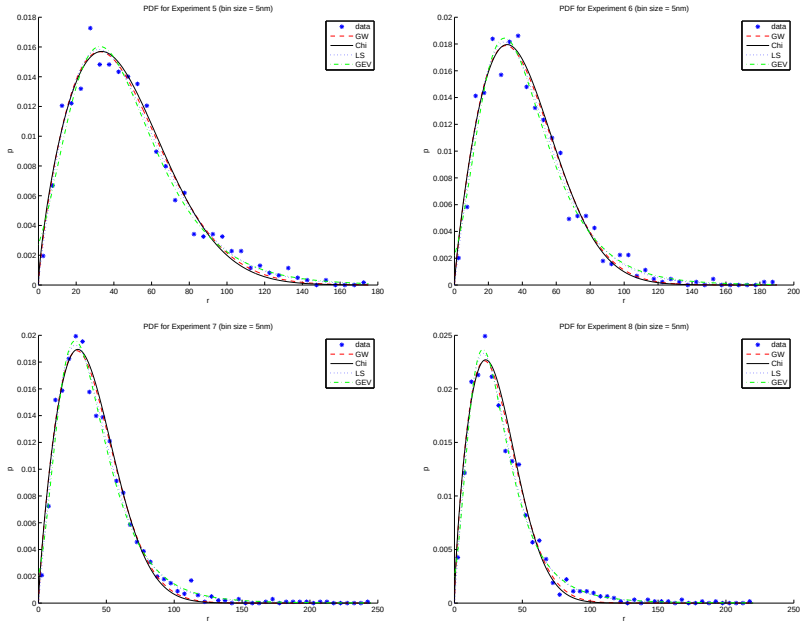


Figure C.17: PDF Fits for Experiments: 5, 6, 7, 8

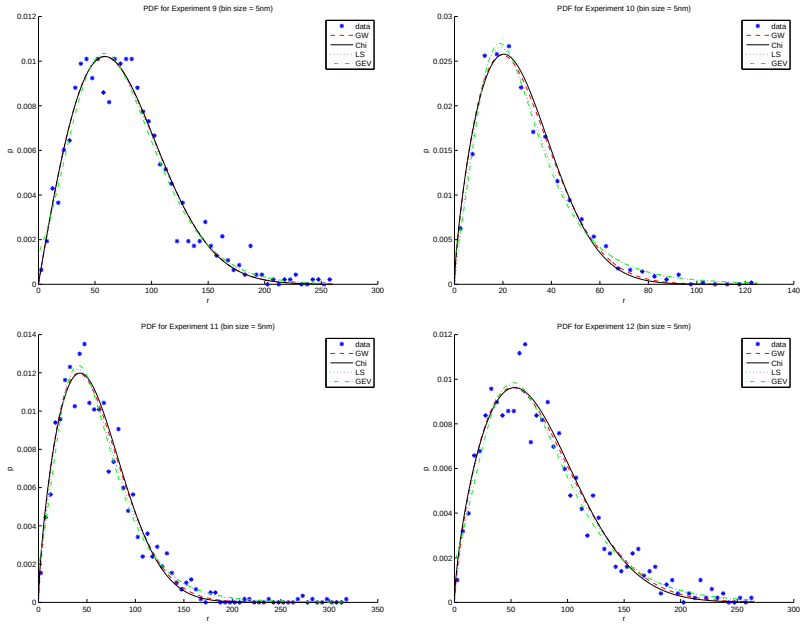


Figure C.18: PDF Fits for Experiments: 9, 10, 11, 12

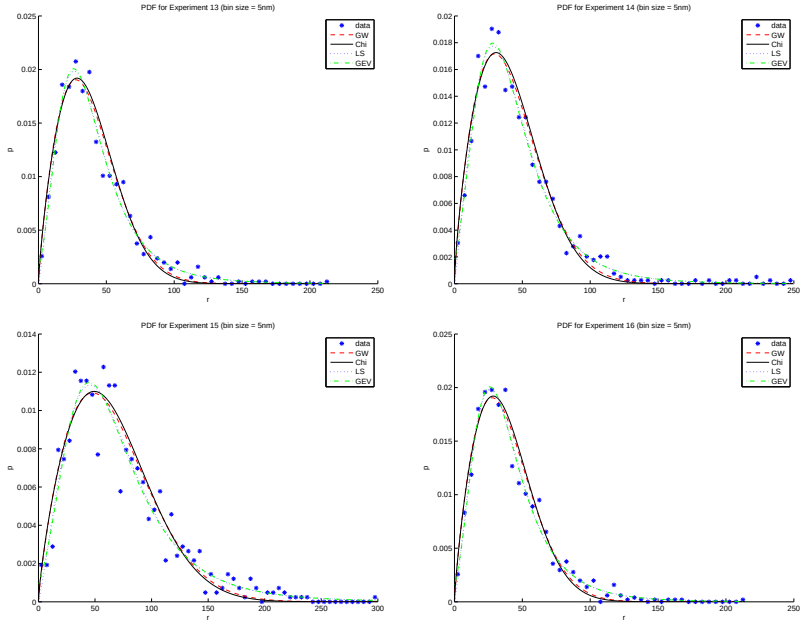


Figure C.19: PDF Fits for Experiments: 13, 14, 15, 16

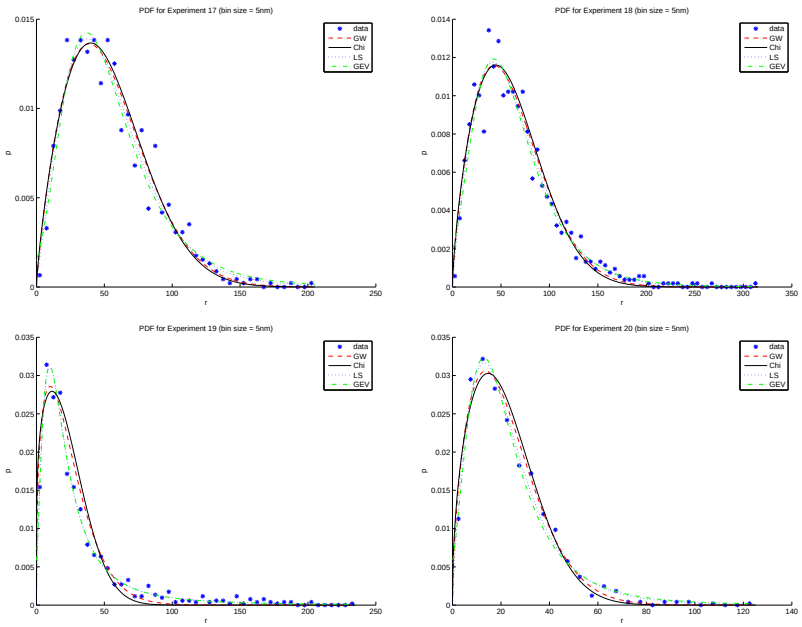


Figure C.20: PDF Fits for Experiments: 17, 18, 19, 20

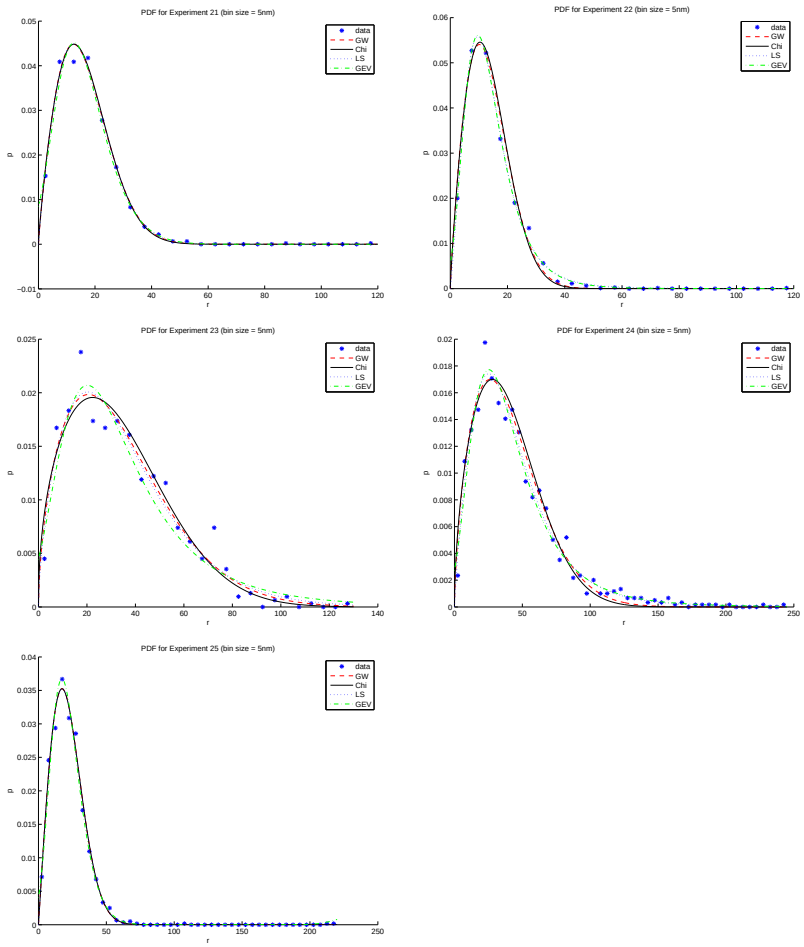


Figure C.21: PDF Fits for Experiments: 21, 22, 23, 24, 25

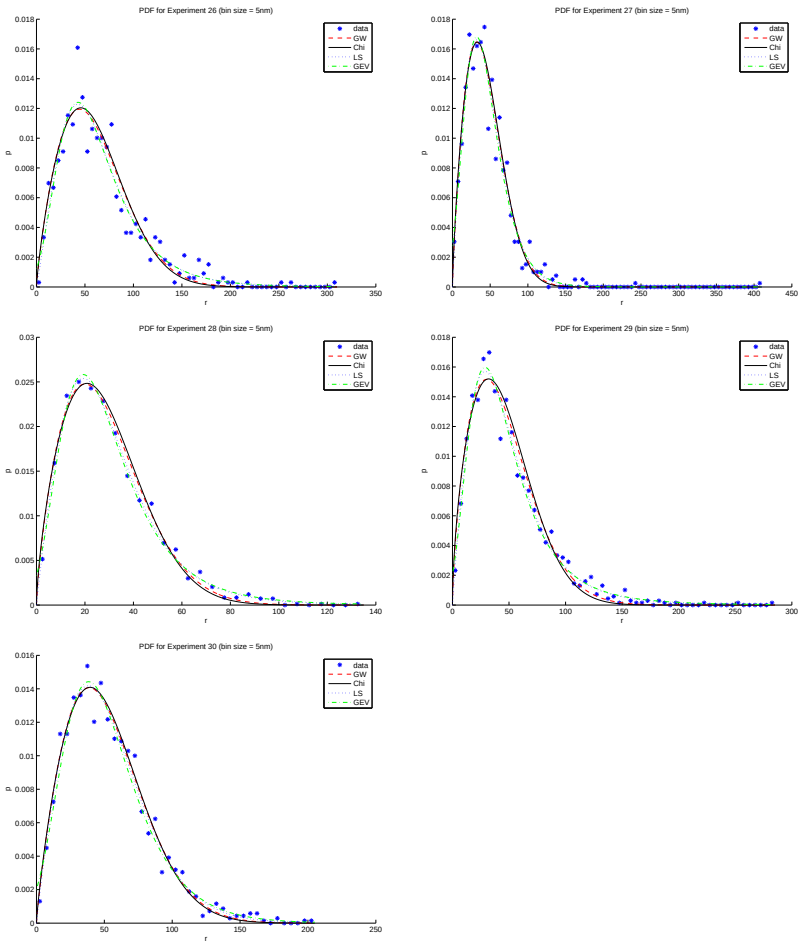


Figure C.22: PDF Fits for Experiments: 26, 27, 28, 29, 30

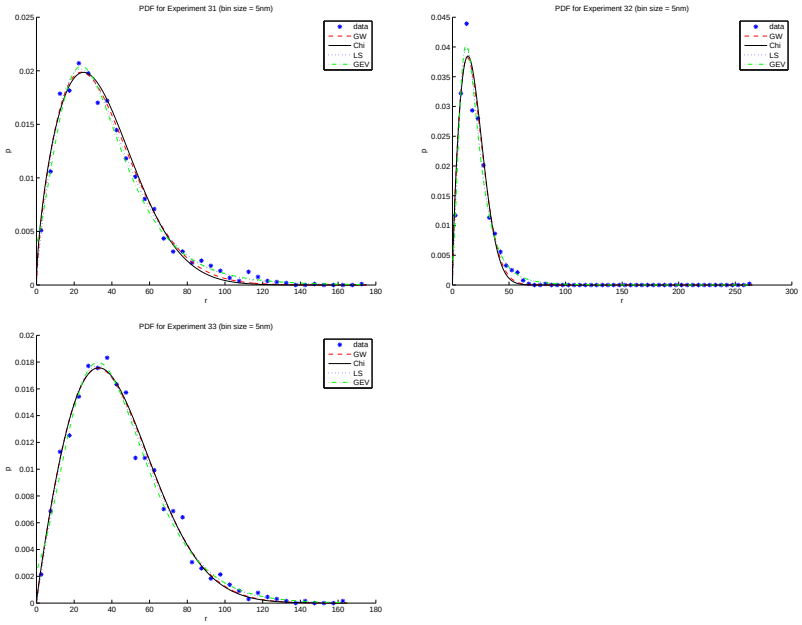


Figure C.23: PDF Fits for Experiments: 31, 32, 33

# D The MSD fits

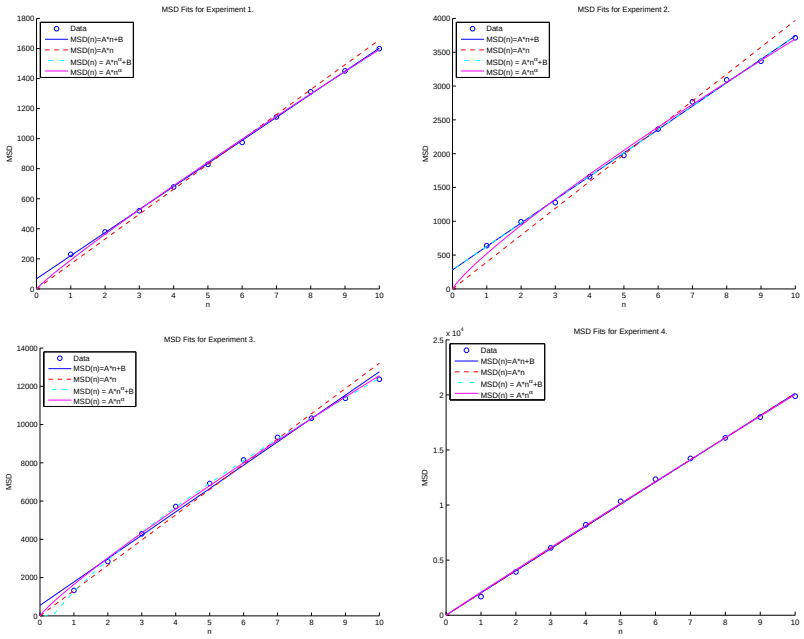


Figure D.24: MSD Fits for Experiments: 1, 2, 3, 4

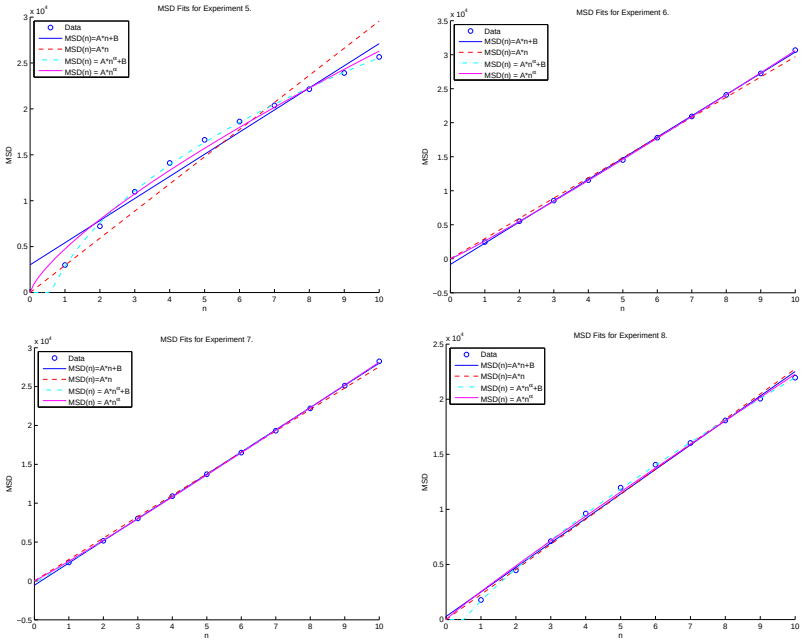


Figure D.25: MSD Fits for Experiments: 5, 6, 7, 8

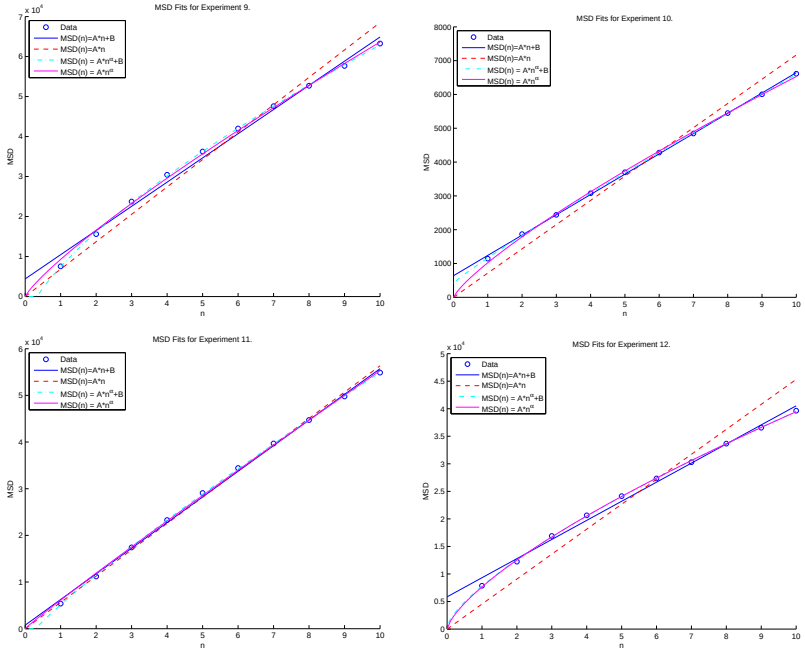


Figure D.26: MSD Fits for Experiments: 9, 10, 11, 12

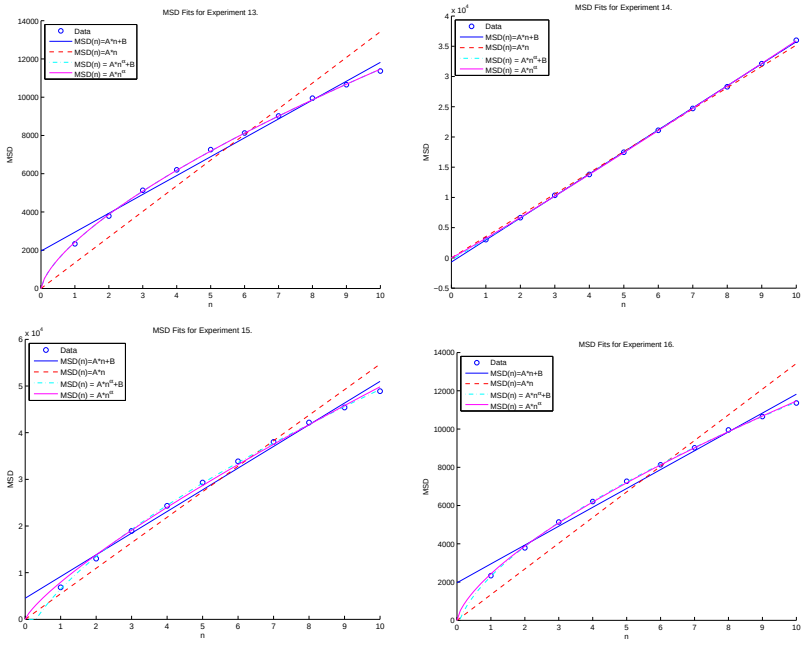


Figure D.27: MSD Fits for Experiments: 13, 14, 15, 16

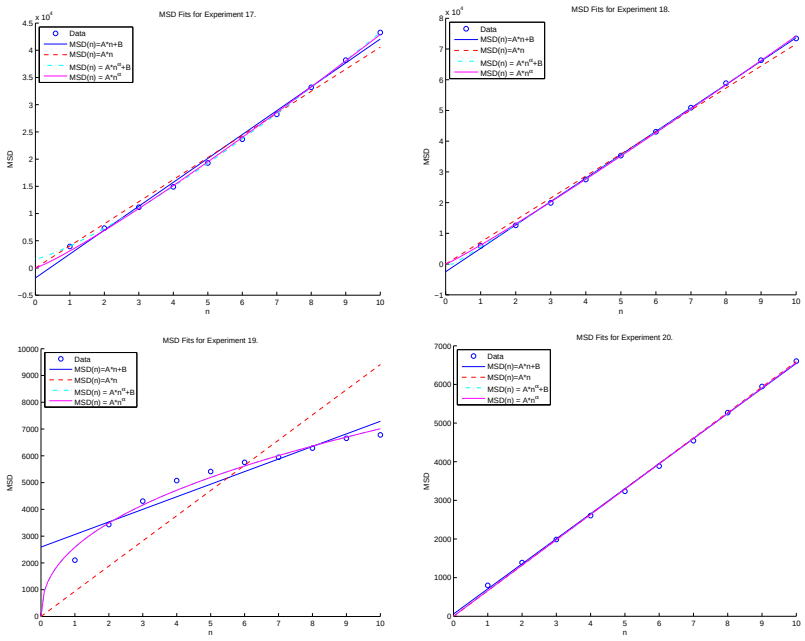


Figure D.28: MSD Fits for Experiments: 17, 18, 19, 20

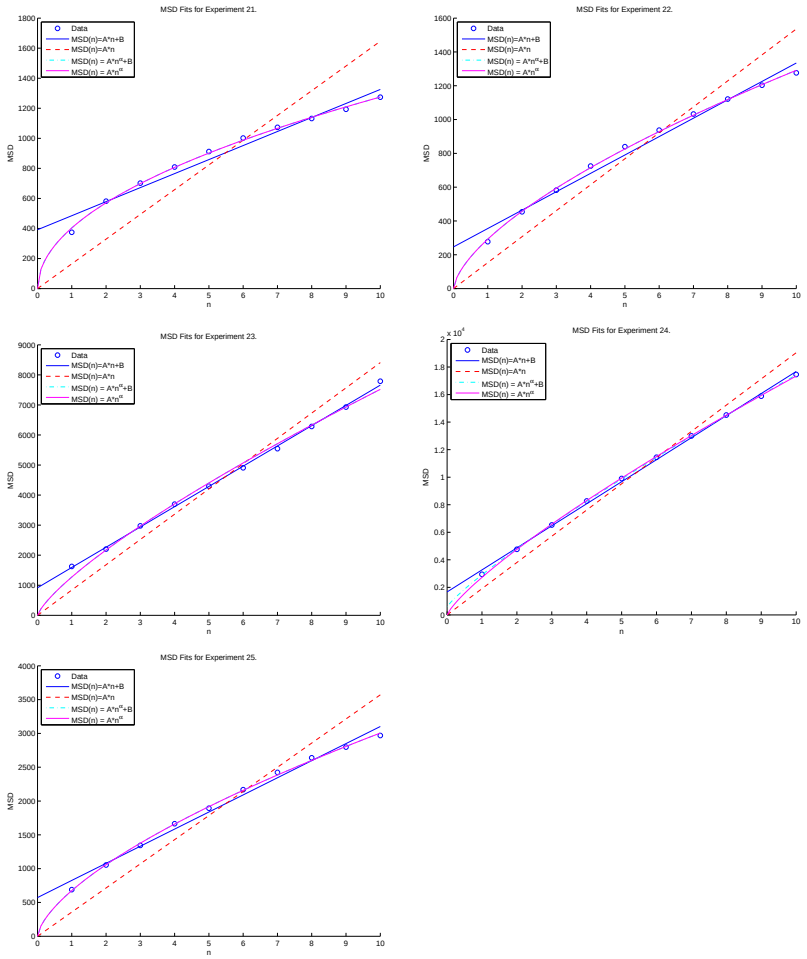


Figure D.29: MSD Fits for Experiments: 21, 22, 23, 24, 25

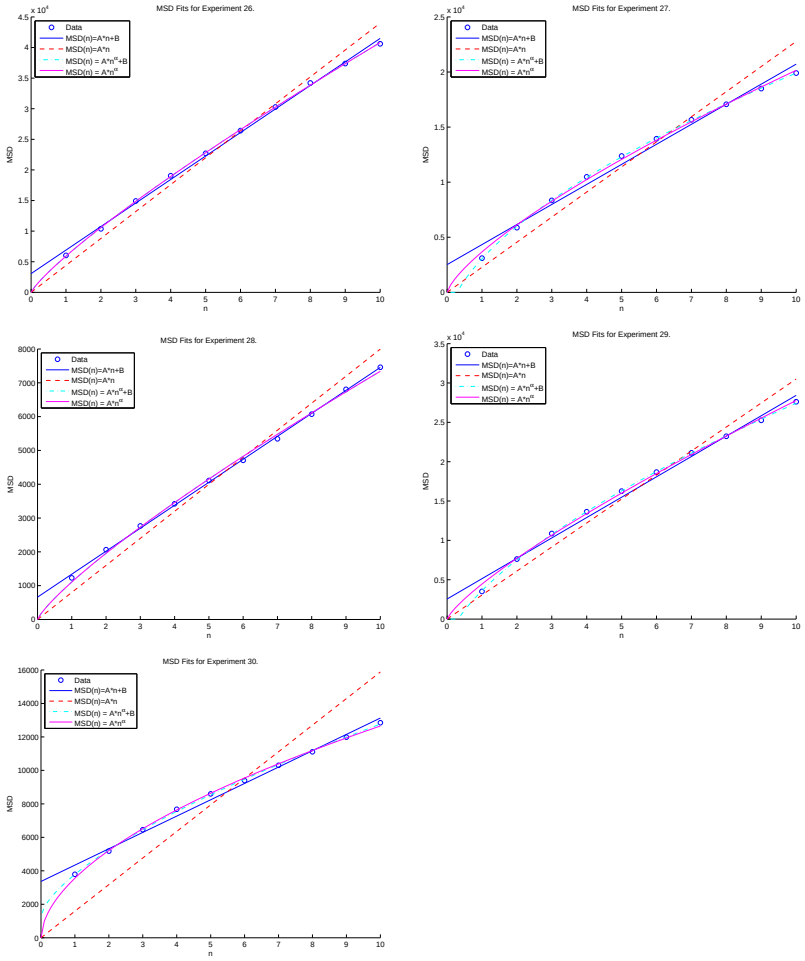


Figure D.30: MSD Fits for Experiments: 26, 27, 28, 29, 30

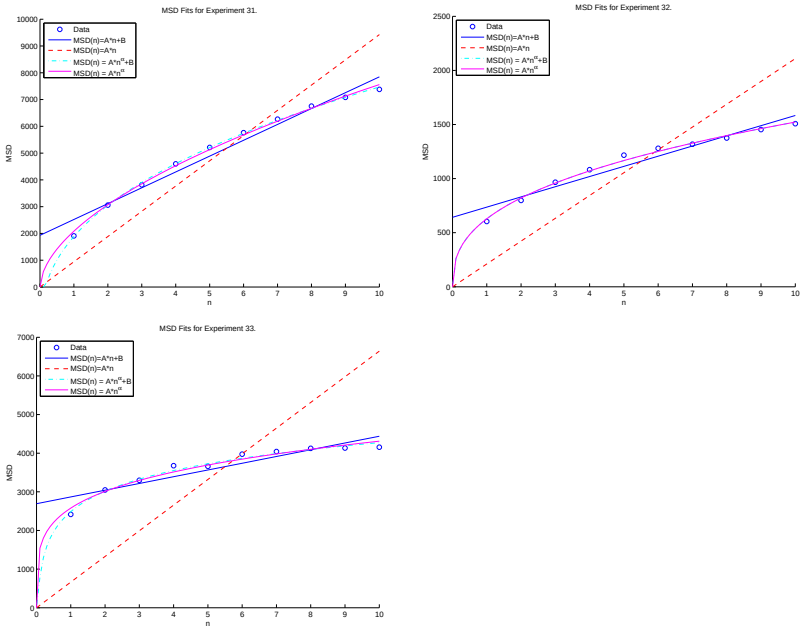


Figure D.31: MSD Fits for Experiments: 31, 32, 33