Bayesian Inference Approaches for Particle Trajectory Analysis in Cell Biology

Bayesian inference approaches for particle trajectory analysis in cell biology

A dissertation presented by

Nilah Monnier

to the

Committee on Higher Degrees in Biophysics

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Biophysics

Harvard University
Cambridge, Massachusetts

April 2013

Dissertation advisor: Prof. Mark Bathe Nilah Monnier

Bayesian inference approaches for particle trajectory analysis in cell biology

Abstract

Despite the importance of single particle motion in biological systems, systematic inference approaches to analyze particle trajectories and evaluate competing motion models are lacking. An automated approach for robust evaluation of motion models that does not require manual intervention is highly desirable to enable analysis of datasets from high-throughput imaging technologies that contain hundreds or thousands of trajectories of biological particles, such as membrane receptors, vesicles, chromosomes or kinetochores, mRNA particles, or whole cells in developing embryos. Bayesian inference is a general theoretical framework for performing such model comparisons that has proven successful in handling noise and experimental limitations in other biological applications. The inherent Bayesian penalty on model complexity, which avoids overfitting, is particularly important for particle trajectory analysis given the highly stochastic nature of particle diffusion. This thesis presents two complementary approaches for analyzing particle motion using Bayesian inference. The first method, MSD-Bayes, discriminates a wide range of motion models—including diffusion, directed motion, anomalous and confined diffusion—based on mean-square displacement analysis of a set of particle trajectories, while the second method, HMM-Bayes, identifies dynamic switching between diffusive and directed motion along individual trajectories using hidden Markov models. These approaches are validated on biological particle trajectory datasets from a wide range of experimental systems, demonstrating their broad applicability to research in cell biology.

Contents

Acknowledgements

First I would like to thank my research advisor, Prof. Mark Bathe at MIT, for providing extensive guidance on the research topics and methodology presented in this thesis. Prof. Bathe is actively engaged in all of the research projects in his group and was always available for detailed discussions of the theory, methods, and applications presented here. He was also very supportive of activities that broadened my experience as a graduate student, from extended visits to collaborating laboratories to spending a semester as an exchange student at an institution across the country. I am also grateful to the other members of the Bathe group who provided substantial advice and support throughout the course of my thesis research. Syuan-Ming Guo and Jun He were instrumental in helping to develop the theory and methods presented in Chapter 2 for applying Bayesian inference to MSD curves. Keyao Pan instructed me in Monte Carlo methods and provided advice on applying Bayesian inference to hidden Markov models (HMMs). Arkajit Dey helped formulate the particle trajectory HMM presented in Chapter 3 and implemented an early version of that algorithm for his Masters thesis. Aprotim Mazumder, Zachary Barry, and Syuan-Ming Guo deserve special thanks for providing excellent feedback on my practice dissertation defense presentation. They and all the other current and former members of the Bathe group—including Philipp Diesinger, Reza Sharifi-Sedeh, Jeremy Bigness, Matthew Adendorff, Lun Yang, Do-Nyun Kim, Kirill Titievsky, Philip Bransford, Benjamin Steele, Ravi Charan, Pia Pal, and Nikita Kodali—have been great friends and mentors to me over the past few years.

I was very fortunate to have substantial interaction with and guidance from Dr. Péter Lénárt, a group leader at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, throughout the course of my thesis research. Dr. Lénárt and his former advisor Dr. Jan Ellenberg, also at EMBL, were active collaborators and generously hosted me in their groups on multiple occasions. Dr. Ellenberg also served as a member of my dissertation

advisory committee and in that role gave valuable advice on the overall direction of my thesis. I am especially grateful to Dr. Lénárt and his group members Masashi Mori and Kálmán Somogyi for spending considerable time teaching me experimental techniques for handling and imaging starfish oocytes, both in Heidelberg and during their summer visits to the Marine Biological Laboratory (MBL) in Woods Hole, MA. This thesis would not have been possible without all of the experimental datasets from starfish oocytes that were generated and shared by Drs. Lénárt and Mori, and a large portion of Chapter 4 was written in collaboration with them. I would also like to thank all the other members of the Lénárt and Ellenberg groups—particularly Joana Pinto, Wani Xiang, Nathalie Daigle, Aicha Metchat, and Sebastian Huet—for welcoming and befriending me while I was in Germany. Finally, my experience with the starfish oocyte system was greatly enhanced by the many weeks that I spent at the MBL working with Prof. Mark Terasaki. I am very grateful to Prof. Terasaki for all of the time he spent introducing me to this area of cell biology (and for giving me the opportunity to pursue a complementary interest in laser wounding experiments and cell membrane repair).

I also enjoyed a productive collaboration at MIT with Prof. Iain Cheeseman and his student Jens Schmidt to study the dynamics of the Ska1 protein complex on microtubules. Close interactions with Dr. Schmidt made an important contribution to improving the particle tracking algorithm described in Chapter 4 and demonstrating its generalizability across biological datasets. I also thank Prof. Cheeseman for many helpful discussions and for his time and effort in serving as a committee member at my dissertation defense. In addition, I had quite extensive interactions with the group of Prof. Adam Martin at MIT, particularly with Prof. Martin and his student Mimi Xie. I am grateful for all of their insights into the problem of ventral furrowing in *Drosophila* embryos and for the many cell trajectory datasets that they shared.

There are many other scientists who generously shared both published and unpublished particle trajectory datasets that were used to test the methods presented here, including Khuloud Jaqaman, Gaudenz Danuser, Tomoya Kitajima, Jennifer Lippincott-Schwartz, Prabuddha Sengupta, Daniel Choquet, Arnauld Sergé, W.E. Moerner, Nicolas Dray, Scott Holley, Paul Blainey, Börn Meier, Inhee Chung, Robert Singer, and Brian English. The methodological and computational work in this thesis would not be possible without the availability of such experimental datasets, and I thank all of these individuals for their willingness to share their hard-earned data.

Finally, I gratefully acknowledge the time and effort of the professors who have served as committee members during my time as a PhD student. In addition to Dr. Ellenberg and Prof. Cheeseman, Profs. Gaudenz Danuser and L. Mahadevan at Harvard served on my dissertation advisory committee and provided valuable feedback and advice that helped shape the work presented here. They also gave constructive comments on the written dissertation document. I am grateful to Profs. Timothy Mitchison, David Golan, and Katharina Ribbeck for serving on my qualifying examination committee and for their feedback on structuring and critically evaluating research proposals. I would especially like to thank Prof. James Hogle and Michele Jakoulov in the Harvard biophysics program for guiding me through the PhD process and for all of their work on behalf of the biophysics students.

Chapter 1

Introduction

1.1 Importance of particle trajectories in biological research

Biological systems rely on spatially and temporally regulated molecular interactions; however, the behavior of biological molecules has a highly stochastic component due to thermal fluctuations that give rise to random Brownian movements of small particles [Berg, 1993]. Robust biological processes and responses to stimuli require that cellular systems impose order and direction onto this naturally stochastic behavior of their constituent particles. Detailed examination of the motion of biological particles can reveal when and where this order is imposed and provide insights into the biological mechanisms by which such regulation is achieved. To analyze biological motion in detail, the positions of particles over time in biological systems must first be visualized using appropriate live-cell imaging technologies and then quantified using robust image processing and particle tracking algorithms (Section 1.2).

The resulting particle trajectories contain important information on the local environments with which the particles are interacting, on the spatial and temporal dynamics of active transport mechanisms, and on collective behavior and coordination between particles. Biological particle trajectories have provided insights into processes as varied as membrane protein dynamics, vesicle trafficking, directed transport along cytoskeletal

structures, chromosome segregation, and endocytosis [Saxton and Jacobson, 1997; Turner et al., 2000; Cremer and Cremer, 2001; Platani et al., 2002; Ehrlich et al., 2004; Gardner et al., 2005; Chuang et al., 2006; Brandenburg and Zhuang, 2007; Walter et al., 2009; Mori et al., 2011; Jaqaman et al., 2011; Kitajima et al., 2011]. Particles in cells can become fully confined due to physical corralling by cytoskeletal polymers [Saxton and Jacobson, 1997; Das et al., 2009; Cairo et al., 2010] or can be otherwise limited in their diffusive motion (subdiffusion or anomalous diffusion) due to a wide variety of underlying physical processes, such as the presence of obstacles or transient binding events to less mobile structures [Brangwynne et al., 2008; Weber et al., 2010; Wang et al., 2010; Rajani et al., 2011]. Particles in cells may also move in a directed manner (flow or drift), for example due to the action of molecular motors [Bormuth et al., 2009; Elting et al., 2011] or to cytoskeletal-driven cytoplasmic flows [Mori et al., 2011].

Evidence for all of these complex behaviors can be found in particle trajectories given appropriate methods and algorithms for analysis. Particle tracking and trajectory analysis have a long history of application across a number of fields, including fluid mechanics [Adrian, 1991] and the study of animal movements [Viswanathan et al., 1996; Benhamou, 2006], and have more recently become essential for tracking people and cars in automated analysis of surveillance videos [Stauffer and Grimson, 2000; Haritaoglu et al., 2000; Hu et al., 2004].

1.2 Imaging and tracking biological particles in living systems

Live cell imaging using light microscopy is essential for obtaining information about biological molecules, organelles, and cells in their native environments [Stephens and Allan, 2003; Meijering et al., 2006]. In particular, obtaining dynamic information on particle motions requires temporal information from time-lapse imaging. Recent work has

greatly expanded the available fluorescence tags for labeling and imaging biological particles [Lippincott-Schwartz and Patterson, 2003; Vaughan and Zhuang, 2011] and improved the spatial resolution of fluorescence microscopy past the diffraction limit [Klar et al., 2000; Betzig et al., 2006; Rust et al., 2006]. These and other advances are increasingly enabling the collection of detailed time-series data on the positions of biological particles over time within living cells and tissues [Seisenberger et al., 2001; Yildiz et al., 2003; Fernández-González et al., 2006; Westphal et al., 2008]. An important challenge in particle trajectory analysis is developing automated analysis methods to handle the large volume of high-throughput imaging data that is now being produced by these imaging technologies.

Before particle motion can be quantitatively characterized, the trajectories themselves (sequences of particle positions over time) must be extracted from the raw images [Meijering et al., 2006; Saxton, 2008]. Manual identification and tracking of particles is time-consuming and not scalable to high-throughput imaging datasets. Automating this process involves two steps: first, accurately determining the particle positions in each image ("segmentation"); and second, linking the particle positions over time into trajectories [Meijering et al., 2006]. Segmentation can be performed across all time frames prior to linking or can be alternated with the linking step; in the latter case, information from trajectories in previous time frames can be used to inform the segmentation of future time frames. Segmentation algorithms typically identify particle positions at sub-pixel resolution by fitting a 2D Gaussian distribution to diffraction-limited objects or by calculating an intensity-weighted centroid for non-Gaussian objects [Cheezum et al., 2001; Sergé et al., 2008]. Alternatively, approaches based on wavelet transforms [Olivo-Marin, 2002] or neural networks [Ouellette et al., 2006] have also been proposed.

Linking particle positions into trajectories can be more or less challenging depending on the density of particles in the image and the average distance that the particles move

from one frame to the next, which depends on the imaging rate [Meijering et al., 2006; Saxton, 2008]. At one extreme, the imaging rate may be high enough that particles in one frame always have some degree of overlap with their position in the previous frame and do not overlap other particles' positions. At the other extreme, particles may move an average distance between frames that is equal to or greater than the average spacing between particles. In the latter case, it is impossible to uniquely assign particle positions to trajectories. For intermediate cases, a number of optimization algorithms have recently been developed to obtain the most likely set of linked trajectories, using either the global optimum or locally optimal sub-trajectories that are subsequently joined into longer trajectories [Sage et al., 2005; Jaqaman et al., 2008; Jaensch et al., 2010]. These methods handle a variety of challenging situations such as fluorophore blinking [Jaqaman et al., 2008]. Alternatively, probabilistic tracking algorithms that output multiple possible trajectories and their associated confidence levels [Orton and Fitzgerald, 2002; Smal et al., 2007; Smal et al., 2008] or tracking algorithms that take into account information on particle modes of motion (for example using Kalman filtering) [Veenman et al., 2001; Comaniciu et al., 2003] can also be used.

1.3 Existing approaches for particle trajectory analysis

1.3.1 Overview

A particle trajectory consists of a sequence of N particle positions $\{\mathbf{r}_i\}_{i=1}^{N}$ observed at times $\{t_i\}_{i=1}^{N}$. Typically these time points will be separated by a constant time interval Δt. Each particle position $\mathbf{r}_i$ may have multiple components depending on the number of independent dimensions that were imaged. For a one-dimensional particle trajectory, the position is a scalar value $r_i = x_i$, while for two and three dimensions, the position is a vector $\mathbf{r}_i = [x_i, y_i]$ or $\mathbf{r}_i = [x_i, y_i, z_i]$, respectively. This sequence of particle positions can

be easily converted to a sequence of particle displacements by taking the vector difference between consecutive positions, $\{\Delta\mathbf{r}_i\}_{i=1}^{N-1} = \{\mathbf{r}_{i+1} - \mathbf{r}_i\}_{i=1}^{N-1}$. For a particle undergoing random Brownian motion, consecutive displacements are statistically independent due to stochastic interactions with surrounding molecules that randomly change the direction of motion [Berg, 1993; Saxton, 1997]. Analysis of these displacements forms the basis for most approaches to particle trajectory analysis.

The goal of particle trajectory analysis is to infer the type of motion that a particle is undergoing and to determine the associated parameters of motion. Because diffusion is a stochastic process, a single measured particle displacement is not sufficient to perform such inference. One must observe multiple displacements of the particle to assemble a more informative picture of its behavior. In any experimental system, such observations are always limited, both by available imaging technology and by the finite duration and number of particles involved in biological processes of interest. Therefore, analysis methods that enable inference even with limited sampling rate, acquisition time, and number of trajectories are particularly valuable [Jaqaman and Danuser, 2006]. One way to capture information from multiple displacements is to use ensemble average distribution functions; for example through moment analysis [Coscoy et al., 2007]. Because of the availability of closed-form analytical solutions for the dependence of the mean-square displacement (MSD) on time lag [Qian et al., 1991; Kusumi et al., 1993; Saxton and Jacobson, 1997], MSD is one of the most commonly-used metrics for characterizing particle motion.

1.3.2 Mean-square displacement analysis

The mean-square displacement along a given particle trajectory is computed for time lags τ according to,

$$MSD(\tau) \equiv \langle \Delta \mathbf{r}(\tau)^2 \rangle = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} |\mathbf{r}_{i+\tau} - \mathbf{r}_i|^2 \ . \tag{1.1}$$

Note that for each τ there are only $(N-\tau)$ samples of $\Delta \mathbf{r}(\tau)^2$ that can be derived from the trajectory. Thus the variance of this MSD estimate increases (the MSD curve gets more noisy) at higher τ [Qian et al., 1991]. The form of the MSD curve can be used to characterize the wide range of particle motions described in Section 1.1. The MSD is given in three-dimensions by the following closed-form analytical solutions for free diffusion (for which we use the abbreviation D), anomalous diffusion (DA), confined diffusion (DR), and flow or directed motion (V),

$$MSD_{\mathrm{D}}(\tau) = 6D\tau \ , \tag{1.2}$$

$$MSD_{\mathrm{DA}}(\tau) = 6D\tau^{\alpha} \ , \tag{1.3}$$

$$MSD_{\mathrm{DR}}(\tau) = R_c^2 \left(1 - \exp(-6D\tau / R_c^2)\right) \ , \tag{1.4}$$

$$MSD_{\mathrm{V}}(\tau) = v^2\tau^2 \ , \tag{1.5}$$

where v is the magnitude of the particle velocity, D is its diffusion coefficient, α is the anomalous exponent, and R_c is the radius within which the particle is confined [Saxton and Jacobson, 1997]. The form of the MSD curve for each of these four basic motion types is shown in Figure 1.1.

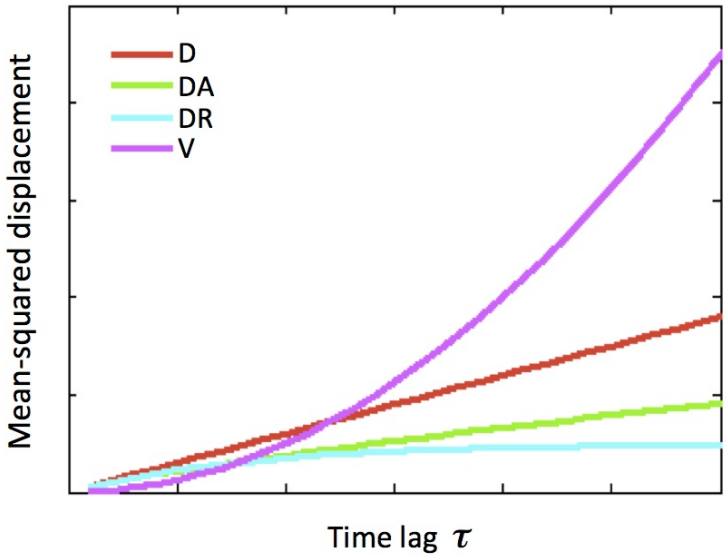

Figure 1.1: Example MSD curves for four motion models: pure diffusion (D), anomalous diffusion (DA), confined diffusion (DR), and pure directed motion (V). Curves are plotted in arbitrary units to show the form of the dependence on τ.

The three diffusive models in Equations 1.2-1.4 can also occur together with directed motion, yielding more complex motion models described by linear combinations of the above equations [Saxton and Jacobson, 1997], such as for free diffusion plus flow,

$$MSD_{\mathrm{DV}}(\tau) = 6D\tau + v^2\tau^2 \ . \tag{1.6}$$

Experimental particle position measurements typically contain a localization error characterized by a positional uncertainty with standard deviation σ_e, which adds a constant term of $6\sigma_e^2$ to the MSD [Michalet, 2010; Voisinne et al., 2010]. In some physical situations, such as confinement within a radius smaller than either the mean localization error or the mean diffusive step size given the sampling rate, the particle may appear stationary because the MSD curve is dominated by this constant term.

MSD curves are most commonly used to find the diffusion coefficient and anomalous exponent of a particle trajectory. The diffusion coefficient is typically obtained by linear regression using only the first few MSD measurements at small time lags [Kusumi et al., 1993; Huet et al., 2006]. The first points are used because they have the smallest variance [Qian et al., 1991] and also have the smallest contribution from directed or anomalous behavior, which are more pronounced at longer time lags. It has been shown that this estimator of the diffusion coefficient is unbiased and efficient for a particle undergoing pure diffusion, but can be biased if the true motion of the particle is more complex, such as confined diffusion [Masson et al., 2009; Voisinne et al., 2010]. The bias is smallest for the shortest time lags, so fitting is often performed by weighting the observed MSD values by the inverse of their variance [Huet et al., 2006]. The number of points used in the linear regression varies between one and 20 points in different studies [Lang et al., 2000; Johns et al., 2001; Daumas et al., 2003; Ng et al., 2003; Huet et al., 2006]. This method assumes that the sampling rate of the trajectory is high enough that directed or anomalous effects do not significantly contribute to the slope of the MSD curve at the earliest measured time lags. The diffusion coefficient value can be obtained either from a mean MSD curve across many trajectories in a dataset, or can be calculated independently for individual trajectories; in the latter case, variation within a population can be revealed by the distribution of diffusion coefficients [Kusumi et al., 1993; Cairo et al., 2006; Bomzon et al., 2006; Cairo et al., 2010].

Plotting the MSD curve versus time lag on a log-long scale is a common approach for visualizing the anomalous exponent α, which is the slope of the log-log plot,

$$\log MSD(\tau) = \log 6D + \alpha \log \tau \ . \tag{1.7}$$

The value of α varies with the magnitude of confined or directed motion present in a trajectory and has been used as a basis for model inference, as discussed below in Section 1.3.3.

Linear regression is typically performed to fit the anomalous exponent using Equation 1.7. An important caveat to this method of obtaining α, however, is that it is significantly affected by the presence of localization error in the particle position measurements. As mentioned above, localization error adds a constant offset to the MSD values. This constant is straightforward to take into account when fitting motion models using Equations 1.2-1.6; however, with a constant term the anomalous diffusion model no longer simplifies into a convenient log-log form as in Equation 1.7. On a log-log scale, an MSD curve with localization error will have a significantly reduced slope over the first few time lags [Martin et al., 2002]. If Equation 1.7 is used to fit α as the slope of the log-log plot in the presence of localization error, its value will be substantially underestimated and the motion will appear artificially subdiffusive [Martin et al., 2002].

1.3.3 Model selection using mean-square displacement

An important challenge for performing motion model selection or classification of particle trajectories is that purely random trajectories can appear to contain complex behaviors, such as confinement or directed motion, simply due to the stochastic nature of diffusion [Simson et al., 1995; Qian et al., 1991; Saxton, 1997]. Therefore, it is essential to rule out the null model of simple diffusion when classifying trajectories as arising from a more complex motion type [Saxton, 1993]. Ruling out the simple diffusion model has typically been done for particle trajectories by applying a proposed metric for model classification to a large number of simulated trajectories of simple diffusion to obtain a probability distribution for its value from these null trajectories. A threshold value of the metric is then chosen such that only a small percentage, typically 1-5 percent, of the simulated diffusive trajectories pass the threshold [Kusumi et al., 1993; Huet et al., 2006].

MSD curves have often been used to infer the type of motion that a particle is undergoing based on their deviation from linearity. Curves for directed motion generally have upward curvature while curves for anomalous or confined diffusion have downward curvature, and this curvature can be quantified based on the difference between the observed MSD curve at long time lags and a projection of a linear fit to the first few MSD points at small time lags [Kusumi et al., 1993; Huet et al., 2006]. As discussed above, the slope of these initial MSD points is assumed to reflect the underlying diffusion coefficient of the particle without significant contribution from directed or anomalous behavior. The linear fit is a projection of what the MSD curve would look like if the motion were purely diffusive at all time scales. Deviation from this projection is typically calculated as the ratio between the observed MSD value and the projected value, measured either at a single time lag [Kusumi et al., 1993] or averaged over a range of time lags [Huet et al., 2006]. Deviation ratios sufficiently greater than 1 indicate that the particle experiences directed motion, while ratios sufficiently less than 1 indicate that the particle experiences subdiffusion or confinement. The significance threshold is chosen based on simulations of simple diffusion, as described above.

Another approach to inferring motion type from an MSD curve is to fit the curve with the anomalous diffusion model using Equation 1.7 and classify the corresponding trajectory as simple diffusion, directed, or subdiffusive based on the fit value of α [Arcizet et al., 2008; Lawton et al., 2013]. Note that the MSD equations for pure diffusion and pure flow, Equations 1.2 and 1.5, are proportional to τ and τ^2, respectively. Thus, the fit value of α would theoretically be equal to 1 if the underlying motion were purely diffusive and equal to 2 if the motion were purely directed. If the motion has both diffusive and directional components, although the true MSD equation is given in Equation 1.6 above, fitting with the anomalous diffusion model will yield an intermediate value $1 < \alpha < 2$, with α approaching 2 as the relative contribution of directional motion increases. Finally, for a subdiffusive process such as confined or obstructed diffusion, the value of α will be less than 1. Therefore, the

value of α can be used as a continuous measure of the degree of directed or obstructed motion [Lawton et al., 2013] or used for classification by choosing threshold values of α [Arcizet et al., 2008].

1.3.4 Non-MSD approaches for detecting directed motion

MSD-based methods have limited application to short particle trajectories, because they require tens of points in the MSD curve to see a difference in curvature between complex motions and simple diffusion [Huet et al., 2006]. A different approach that has been proposed to detect directed motion in shorter trajectories is based on the asymmetry of a point cloud of particle position measurements [Saxton, 1993; Huet et al., 2006]. A point cloud can be described by a radius of gyration tensor and its principle moments, which are related to the spatial extent of the point cloud in different dimensions. Various metrics to measure asymmetry have been proposed based on the relative magnitudes of these principle radii of gyration, such as the ratio of the radii or the relative difference between them [Saxton, 1993; Huet et al., 2006]. However, it is important to note that trajectories of particles undergoing simple diffusion are also typically asymmetric due to the stochastic nature of diffusion [Saxton, 1993; Rudnick and Gaspari, 1987]. Therefore, it is not straightforward to distinguish directed motion from pure diffusion using an asymmetry metric without first examining the probability of observing particular values of asymmetry in simulations of random walks. A threshold for significance can then be set based on this probability distribution, as described in Section 1.3.3 above. Previous studies found that this asymmetry metric is applicable to trajectories with at least 10 points [Huet et al., 2006], shorter than what is needed for a reliable MSD-based analysis.

Another class of methods for distinguishing directed motion from random diffusion relies on correlations in the direction of motion of a particle at consecutive time points. For a purely random walk, consecutive directions are uncorrelated, whereas for pure directed motion they would be perfectly correlated. The degree of correlation can be measured by calculating the dot product between consecutive velocity vectors along the trajectory [Bouzigues and Dahan, 2007; Weber et al., 2012]. As with the displacement measurements used for MSD curves above, this correlation measurement can be calculated for velocity vectors over different time lags τ. For a flowing particle that also has a diffusive component or that has noisy position measurements, the observed correlation will be greater for longer time lags, due to the greater relative contribution of directed motion to the displacement vector. Velocity correlations can be analyzed for a single time lag or averaged over a range of time lags [Bouzigues and Dahan, 2007] or plotted versus time lag in an analogous manner to MSD curves [Weber et al., 2012]. For determining whether the observed correlation values are significant enough to classify a particle as having directed motion, again a threshold value can be chosen based on the distribution of correlation values obtained from simulations of a large number of purely diffusing trajectories [Bouzigues and Dahan, 2007].

1.3.5 Non-MSD approaches for detecting confinement

A natural approach to detect confinement within a trajectory is based on the fact that confinement, whether full or partial, reduces the distance that a particle travels over time in comparison to free diffusion. Therefore, the time that a particle stays within a circle or volume of a given radius centered at its initial position will be longer for confined particles than for freely diffusing particles with the same underlying diffusion coefficient [Condamin et al., 2008; Rajani et al., 2011]. This time is called the first-passage time (FPT) and has been used to study animal movements [McKenzie et al., 2009] and in the field of chemical kinetics [Condamin et al., 2007], and is also similar to the recently-proposed mean maximal

excursion method for trajectory analysis [Tejedor et al., 2010]. The probability that a free particle with diffusion coefficient D stays within a radius R for a time t can be derived analytically as a function of these three parameters [Saxton, 1993]. One implementation of FPT analysis [Simson et al., 1995] is to compare the observed first-passage time for some R with the expected first-passage time for a freely diffusing particle with a value of D estimated from the initial slope of the MSD curve, as described above. Alternatively, an appropriate radius for FPT analysis can be selected by measuring the variance in FPTs measured along a trajectory for different values of R. This variance is greatest when the tested radius is close to the size of the confinement zones within the trajectory [Rajani et al., 2011]. A potential complication is that a smaller peak in FPT variance is also present at a radius corresponding roughly to the mean single-step length of the trajectory, whether or not the particle is confined. Therefore, it is recommended that the results of FPT analysis on experimental trajectories be compared directly to results on simulated trajectories with the same diffusion coefficient [Rajani et al., 2011].

1.3.6 Detecting transient behavior along a trajectory

Since interpreting particle trajectories requires multiple observations of the particle positions and displacements, as discussed in Section 1.3.1, typically observations are pooled over an entire trajectory or set of trajectories assumed to undergo the same type of motion. However, particles in biological systems may experience different phases of motion even within a single observed trajectory, due to spatial or temporal regulation of the particle's environment or interactions with other molecules. For example, membrane proteins may pass through different compartments in the cell membrane that affect their behavior [Kusumi et al., 1993], and proteins that form complexes with other particles will show transient switching between diffusion coefficients governed by the on- and off-rates of binding [Das et al., 2009]. To test whether a trajectory contains multiple types of motion, the following

bootstrapping method has been proposed. A set of artificial trajectories is generated by sampling from the empirical step length distribution and the distribution of angles between consecutive displacements observed in the original trajectory [Rajani et al., 2011]. This set of bootstrapped trajectories represents the range of particle behaviors that would be expected assuming that the length and angle distributions are time-invariant. If the properties of the original trajectory, for example its MSD curve, fall outside the range of the bootstrapped trajectories, then the original trajectory is not well explained by time-invariant step and angle distributions and is likely to consist of multiple phases of motion [Rajani et al., 2011].

Detection of such transient phases of motion along a particle trajectory precludes the use of metrics that are averaged over the entire trajectory. However, the analysis methods described in the sections above can be adapted to detect transient motion by applying an analysis repeatedly to sub-trajectories of the original trajectory, typically using a sliding time window. Most published approaches for detecting transient motion in particle trajectories have relied on this type of sliding window approach [Simson et al., 1995; Huet et al., 2006; Arcizet et al., 2008; Rajani et al., 2011]. An important consideration is the width of the window, since there is a tradeoff between temporal resolution and obtaining sufficient samples for robust inference [Huet et al., 2006; Arcizet et al., 2008]. The window size is often selected based on prior knowledge of the typical timescales and motion parameters of the biological processes being investigated [Arcizet et al., 2008]. An alternative approach is to test a range of window widths at each position along a trajectory and select the width that gives the greatest difference from purely diffusive behavior [Huet et al., 2006]. Because stochastic fluctuations in diffusive motion may also produce occasional false-positive signals, these window-based methods typically enforce a minimum number of consecutive windows over which complex behaviors must persist in order to be deemed significant [Simson et al., 1995]. The MSD anomalous exponent α-based method for classification of directed motion, discussed in Section 1.3.3, has been applied in a sliding window approach

to characterize transport of endosomes and distinguish free diffusion from periods of active transport along microtubules [Arcizet et al., 2008]. The MSD-based deviation from linearly metric (Section 1.3.3) and the point-cloud asymmetry metric (Section 1.3.4) have been applied within sliding analysis windows along secretory vesicle trajectories to detect periods of constrained motion and directed motion, respectively [Huet et al., 2006]. This sliding-window approach is convenient in that it is very general and can be used in conjunction with most particle trajectory analysis metrics. However, it is limited in temporal resolution to the width of the time windows within which the metrics are applied, which must typically be tens of steps long to detect behavior that is significantly different from simple diffusion [Huet et al., 2006]. Thus, the maximum achievable temporal resolution of transient behavior using a window-based method does not approach the temporal resolution of the underlying trajectory.

An alternative approach that makes it possible to achieve single-step temporal resolution of motion switching along particle trajectories has recently been proposed [Das et al., 2009; Cairo et al., 2010; Chung et al., 2010; Persson et al., 2013]. The approach extends a maximum likelihood change-point detection approach for identifying switches in diffusion coefficient [Montiel et al., 2006] by using a hidden Markov model (HMM) to model multiple states of diffusion along a trajectory. HMMs are designed to analyze timeseries data and assume that the observation made at each time point—such as the observation of a particle's position or displacement from its previous position—derives from some unobserved (hidden) state of the system. As long as the relationship between the hidden states and the observations can be defined in terms of a probability distribution, then the series of observations can be used to infer the sequence of hidden states and assign a most likely state to each time point. HMMs have been applied successfully in other areas of biophysics [Venkataramanan and Sigworth, 2002; Bronson et al., 2009; Letinic et al., 2010]. For particle trajectories, the observations have been defined as either the

particle displacements [Das et al., 2009; Cairo et al., 2010; Persson et al., 2013] or squared displacements [Chung et al., 2010], and the hidden states have been defined as different values of the diffusion coefficient. While the majority of these studies made an a priori assumption that there were two diffusive states and used the HMM approach to fit the values of the two diffusion coefficients and find the sequence of diffusive states [Das et al., 2009; Cairo et al., 2010; Chung et al., 2010], the most recent study uses a Bayesian procedure to infer the number of diffusive states as well [Persson et al., 2013]. Their method is based on the principles of Bayesian inference described below.

1.4 Principles of model selection and Bayesian inference

A common theme in the sections above and in the field of particle trajectory analysis is the use of a simple model—typically pure diffusion—as a null hypothesis for detecting more complex forms of motion. Comparison to a null model is part of the more general problem of model inference and model selection when the models being compared have different complexity, or different numbers of parameters. It is important to avoid over-fitting or over-interpreting an observed dataset by inferring complexity that is not statistically justified. Inference of particle motion models is a particularly clear example of this problem, since the stochastic nature of simple diffusion leads to relatively high likelihoods of seeing behaviors in diffusive trajectories that may appear to be caused by more complex motion models [Kusumi et al., 1993; Michalet, 2010; Saxton, 1997]. As discussed in the context of thresholding-based motion model selection methods in Section 1.3, it is important to know the chance or likelihood that a particular level of apparent complexity can be generated from simple diffusion. This likelihood can be determined either analytically or through repeated simulations of diffusive trajectories, as described in Section 1.3.

A more general method of model selection is to compare the likelihoods of seeing a set of observations—such as a sequence of particle positions or values of a trajectory-derived metric—between different models. Typically the models will be parameterized by one or more parameters $\boldsymbol{\beta}$, and the likelihood of the observations can be computed at specific values of $\boldsymbol{\beta}$. For any given model M_k from a set of possible models $(M_1, ..., M_K)$, we can write the likelihood of seeing a set of observations $\mathbf{y}$ at a given value of the parameters $\boldsymbol{\beta}_k$ of that model as,

$$P(\mathbf{y}|\boldsymbol{\beta}_k, M_k) \ . \tag{1.8}$$

A maximum likelihood (ML) approach to model selection finds the parameters $\hat{\boldsymbol{\beta}}_{k,\mathrm{MLE}}$ (the maximum likelihood estimator, MLE) that maximize this likelihood for each model M_k [Casella and Berger, 2001],

$$\hat{\boldsymbol{\beta}}_{k,\mathrm{MLE}} = \arg\max_{\boldsymbol{\beta}_k} P(\mathbf{y}|\boldsymbol{\beta}_k, M_k) \ , \tag{1.9}$$

and then chooses the model with the maximum value of this maximum likelihood,

$$\hat{M} = \arg\max_{M_k} P(\mathbf{y}|\hat{\boldsymbol{\beta}}_{k,\mathrm{MLE}}, M_k) \ . \tag{1.10}$$

However, the maximum value of the likelihood will generally be larger for models with more parameters, due to the fact that more complex models have more degrees of freedom with which to fit the data [Casella and Berger, 2001; Sivia and Skilling, 2006]. The ML approach on its own does not penalize complex models for these extra parameters.

An alternative approach is to treat the parameters $\boldsymbol{\beta}_k$ not as fixed quantities but as random variables with their own probability distributions, following Bayesian statistics [Casella and Berger, 2001; Sivia and Skilling, 2006]. Model selection using Bayesian inference

considers the likelihood of the observations not at a single ML estimate of the parameters, but integrated over all possible parameter values, as follows,

$$P(\mathbf{y}|M_k) = \int P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)P(\boldsymbol{\beta}_k|M_k)d\boldsymbol{\beta}_k \ . \tag{1.11}$$

The total probability of the data given a model, $P(\mathbf{y}|M_k)$, is called the marginal likelihood or evidence [Bronson et al., 2009]. Note the presence of the term $P(\boldsymbol{\beta}_k|M_k)$ in Equation 1.11, which is a prior probability distribution over the values of the parameters. Prior probabilities on the models themselves can also be introduced using Bayes' theorem,

$$P(M_k|\mathbf{y}) = \frac{P(\mathbf{y}|M_k)P(M_k)}{P(\mathbf{y})} \ , \tag{1.12}$$

where $P(M_k|\mathbf{y})$ is the posterior probability of each tested model and the denominator $P(\mathbf{y})$ is a normalizing factor $P(\mathbf{y}) = \sum_k P(\mathbf{y}|M_k)P(M_k)$ such that the model probabilities sum to 1. If the prior model probabilities $P(M_k)$ are assumed equal for all k, which is suitable when no information is available to prefer one model over another, then the final model probabilities $P(M_k|\mathbf{y})$ are simply proportional to the marginal likelihoods $P(\mathbf{y}|M_k)$.

An important feature of the marginalization process in Equation 1.11 is that it introduces a inherent penalty on model complexity due to the fact that the prior probability of the parameters $P(\boldsymbol{\beta}_k|M_k)$ is spread over a higher-dimensional parameter space as the number of parameters in the model increases. Thus the value of the prior at the MLE parameters is significantly smaller for models with more parameters. In the words of Sivia and Skilling,

> Any likelihood gain from a better agreement with the data, allowed by the greater flexibility of a more complicated model, has to be weighed against the additional cost of averaging it over a larger parameter space. [Sivia and Skilling, 2006]

Thus the Bayesian approach to model selection automatically penalizes model complexity and identifies the simplest model consistent with the observed data [Sivia and Skilling, 2006; Raftery, 1995; Posada and Buckley, 2004; Carlin and Louis, 2009]. This approach is general in that it computes relative probabilities of an arbitrary set of competing motion models without any requirement on model form or nesting [Raftery, 1995]. Although Bayesian inference requires a choice of prior probabilities associated with each model and its parameters, this requirement objectifies the scientific process by formalizing and reporting these biases concisely in the mathematical form of a prior distribution [Raftery, 1995; Posada and Buckley, 2004]. Given a set of priors, Bayesian inference can be applied automatically, without user intervention.

Chapter 2

Bayesian inference for mean-square displacement analysis

Much of this chapter has been previously published in:

Monnier N, Guo SM, Mori M, He J, Lénárt P, Bathe M. Bayesian approach to MSD-based analysis of particle motion in live cells. *Biophysical Journal* 103(3):616-626 (2012)

2.1 Overview

Quantitative tracking of particle motion using live-cell imaging is a powerful approach to understanding the mechanism of transport of biological molecules, organelles, and cells. In most biological applications, the underlying mode of particle motion is unknown a priori and must be inferred using mathematical models. However, inferring complex stochastic motion models from single-particle trajectories in an objective manner is nontrivial due to noise from sampling limitations and biological heterogeneity. This chapter presents a systematic Bayesian approach to multiple-hypothesis testing of a general set of competing motion models based on particle mean-square displacements (MSDs) that automatically classifies particle motion, properly accounting for sampling limitations and correlated noise while appropriately penalizing model complexity to avoid over-fitting. We test the procedure rigorously using simulated trajectories for which the underlying physical process is known, demonstrating that it chooses the simplest physical model that explains the observed data. Further, we show that computed model probabilities provide a reliability test for the downstream biological

interpretation of associated parameter values. We subsequently illustrate the broad utility of the approach by applying it to disparate biological systems including experimental particle trajectories from chromosomes, kinetochores, and membrane receptors undergoing a variety of complex motions. This automated and objective Bayesian framework easily scales to large numbers of particle trajectories, making it ideal for classifying the complex motion of large numbers of single molecules and cells from high-throughput screens, as well as single-cell-, tissue-, and organism-level studies.

2.2 Correlated error in MSD curves

As discussed in Chapter 1, the stochastic nature of diffusion complicates the analysis of particle trajectories because particles undergoing simple diffusion can have trajectories that appear by eye to contain more complex behaviors, including periods of confinement or directed motion [Qian et al., 1991; Kusumi et al., 1993; Michalet, 2010; Saxton, 1997]. This complication exists not just for the particle trajectories themselves, but also for derived metrics such as mean-square displacement. Figure 2.1 shows multiple MSD curves derived from individual simulated trajectories of particles undergoing simple Brownian motion. Despite the fact that the analytical form of the MSD curve for this type of simple motion is linear in τ (Equation 1.2), the observed MSD curves are highly variable. Many of the curves have upward or downward curvatures, which are characteristic of directed motion and confined or subdiffusive motion, respectively; thus these MSD curves appear to suggest more complex particle motions than are actually present.

Note that the deviation of each MSD curve from its analytical form is highly correlated over the different time lags τ (Figure 2.1). When MSD curves are calculated from particle trajectories according to Equation 1.1, the same set of steps—grouped into different

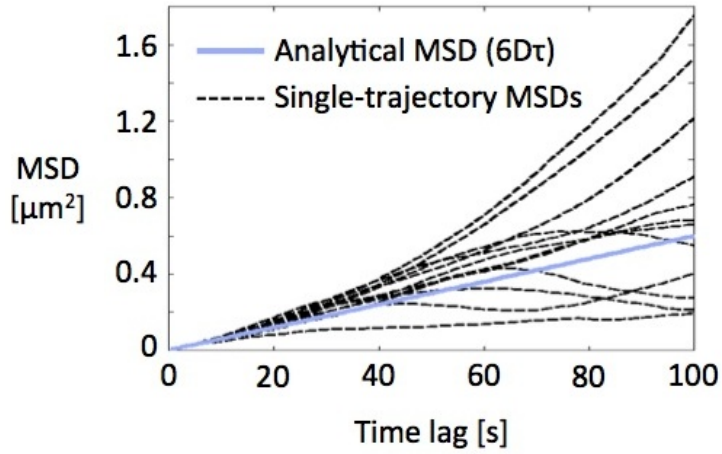

Figure 2.1: Example MSD curves (dashed lines) calculated from individual simulated particle trajectories undergoing simple diffusion with $D = 0.001$ μm^2/s, $\Delta t = 1$ s, and total time $T = 200$ s. The analytical form of the MSD curve for simple diffusion in three dimensions, $6D\tau$, is also shown (blue line). Figure previously published in [Monnier et al., 2012].

sized windows—is used to calculate the MSD value at each time lag τ. In other words, the same stochastic observations are reused for each calculation. This process leads to strong correlations in the difference between the calculated MSD values and their expected values over τ. The analytical form of the MSD is asymptotically correct only in the limit of infinite trajectory length. Averaging MSD curves from different trajectories improves the estimate of the mean MSD value at each time lag, but does not eliminate the correlations between time lags.

For a freely diffusing particle in the absence of localization error, an analytical solution for the covariance of these deviations from the analytical MSD curve has been derived [Qian et al., 1991; Michalet, 2010] and is shown in Figure 2.2. Note that the magnitude of the covariance increases significantly at larger values of τ, as discussed in Chapter 1. The presence of these correlated errors in MSD curves can result in fitting erroneous, overly-complex models [Qian et al., 1991; Kusumi et al., 1993; Michalet, 2010; Saxton, 1997]. To avoid this over-fitting problem, here we account for the correlations using

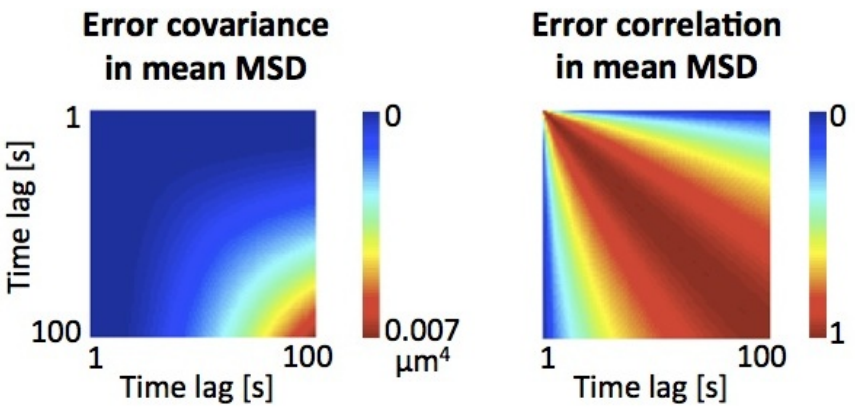

Figure 2.2: Analytical forms of the MSD error covariance and correlation matrices (from [Qian et al., 1991; Michalet, 2010]) for MSD curves from simple diffusion with parameters as in Figure 2.1. Figure previously published in [Monnier et al., 2012].

multiple independent MSD curves from independent particle trajectories or sub-trajectories to compute both a mean MSD curve and its associated error covariance matrix, as described below.

2.3 Application of Bayesian inference to MSD analysis

2.3.1 Classical regression and generalized least squares

Classical regression fits an observed series of data $\mathbf{y} = [y_1, y_2, ..., y_n]$ (in this case, the MSD values) with a model function $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta})$ (in this case, the motion model equations given in Chapter 1) according to $y_i = f(x_i, \boldsymbol{\beta}) + \epsilon_i$, where $\mathbf{x} = [x_1, x_2, ..., x_n]$ are the sample points (in this case, the time lags τ), $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_p]$ are the model parameters (including some combination of diffusion coefficient D, velocity v, anomalous exponent α, and/or confinement radius R_c), and ϵ_i are errors associated with the y_i measurements. In the case of particle trajectories, the errors arise primarily from finite sampling of the stochastic motion of the particle, as discussed above. The classical weighted least squares approach minimizes the

sum of the squared residuals, $\chi^2 = \sum_i (y_i - f(x_i, \boldsymbol{\beta}))^2/\sigma_i^2$, where the error terms ϵ_i are assumed to be uncorrelated and each is normally distributed with zero mean and standard deviation σ_i [Casella and Berger, 2001; Seber and Wild, 2003]. The chi-squared value χ^2 can then be used to test the goodness-of-fit of models conditioned on a null hypothesis.

Minimizing χ^2 is equivalent to maximizing the likelihood of observing the data points $\mathbf{y}$, where each y_i is normally distributed around its analytical value $f(x_i, \boldsymbol{\beta})$ with standard deviation σ_i. The likelihood is a product over the probability of observing each y_i,

$$\begin{aligned} P(\mathbf{y}) &= \prod_{i=1}^{n} \left[\frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(y_i - f(x_i, \boldsymbol{\beta}))^2}{2\sigma_i^2} \right) \right] \\ &= \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} \sigma_i} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - f(x_i, \boldsymbol{\beta}))^2}{\sigma_i^2} \right). \end{aligned} \tag{2.1}$$

The χ^2 value above is equal to the summation in the exponential term. For MSD curves, however, the y_i values are not distributed around the analytical $f(x_i, \boldsymbol{\beta})$ values according to independent normal distributions. As discussed in Section 2.2, the MSD errors ϵ_i for different time lags are highly correlated, making Equation 2.1 invalid. If the correlation matrix $\mathbf{C}$ of the MSD errors ϵ_i is known, then the likelihood can be rewritten in terms of the full multivariate normal distribution of the error vector $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\beta})$ as follows [Seber and Wild, 2003],

$$P(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2} \big(\mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\beta})\big)^T \mathbf{C}^{-1} \big(\mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\beta})\big) \right). \tag{2.2}$$

This approach is called generalized least squares and can be implemented by transforming the fitting equation $\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}$ with a matrix $\mathbf{A}$ equal to the inverse of the Cholesky decomposition of $\mathbf{C}$, such that the resulting equation $\mathbf{Ay} = \mathbf{Af}(\mathbf{x}, \boldsymbol{\beta}) + \mathbf{A}\boldsymbol{\epsilon}$ has uncorrelated errors $\mathbf{A}\boldsymbol{\epsilon}$ [Seber and Wild, 2003]. In this transformed coordinate system, the ordinary least squares approach that minimizes the sum of squared residuals can then be used to obtain

the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ for the parameters. Here we use the ordinary least squares algorithm in the "lsqcurvefit" function in MATLAB (The MathWorks, Inc., Natick, MA) to find the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ and the associated covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathrm{MLE}}$ of the fit parameters.

2.3.2 Bayesian approach to regression

We apply the Bayesian inference framework introduced in Chapter 1 to test multiple competing models for MSD curves. For K possible models $(M_1, ..., M_K)$, such as the motion models shown in Figure 2.3, the probability of each model given the observed data $\mathbf{y}$ can be expanded by Bayes rule as in Equation 1.12. We assume that the models all have equal prior probabilities $P(M_k)$, so the posterior model probabilities are simply proportional to the marginal data likelihood,

$$P(M_k|\mathbf{y}) \propto P(\mathbf{y}|M_k) \; . \tag{2.3}$$

$P(\mathbf{y}|M_k)$ is calculated for each model by marginalizing the likelihood $P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)$ over the model parameters $\boldsymbol{\beta}_k$ as in Equation 1.11,

$$P(\mathbf{y}|M_k) = \int P(\mathbf{y}|\boldsymbol{\beta}_k, M_k) P(\boldsymbol{\beta}_k|M_k) d\boldsymbol{\beta}_k \; . \tag{2.4}$$

In the case of the observed MSD values, the probability of observing the data $\mathbf{y}$ for any given realization of the parameters $\boldsymbol{\beta}_k$ of model M_k with model function $\mathbf{f}_k(\mathbf{x}, \boldsymbol{\beta}_k)$ is given by the general multivariate normal function in Equation 2.2 above, which includes $\mathbf{C}$, the covariance matrix of the errors ϵ_i, so Equation 2.4 becomes,

$$P(\mathbf{y}|M_k) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} \int \exp\left(-\frac{1}{2}\big(\mathbf{y} - \mathbf{f}_k(\mathbf{x}, \boldsymbol{\beta}_k)\big)^T \mathbf{C}^{-1} \big(\mathbf{y} - \mathbf{f}_k(\mathbf{x}, \boldsymbol{\beta}_k)\big)\right) P(\boldsymbol{\beta}_k|M_k) d\boldsymbol{\beta}_k \; . \tag{2.5}$$

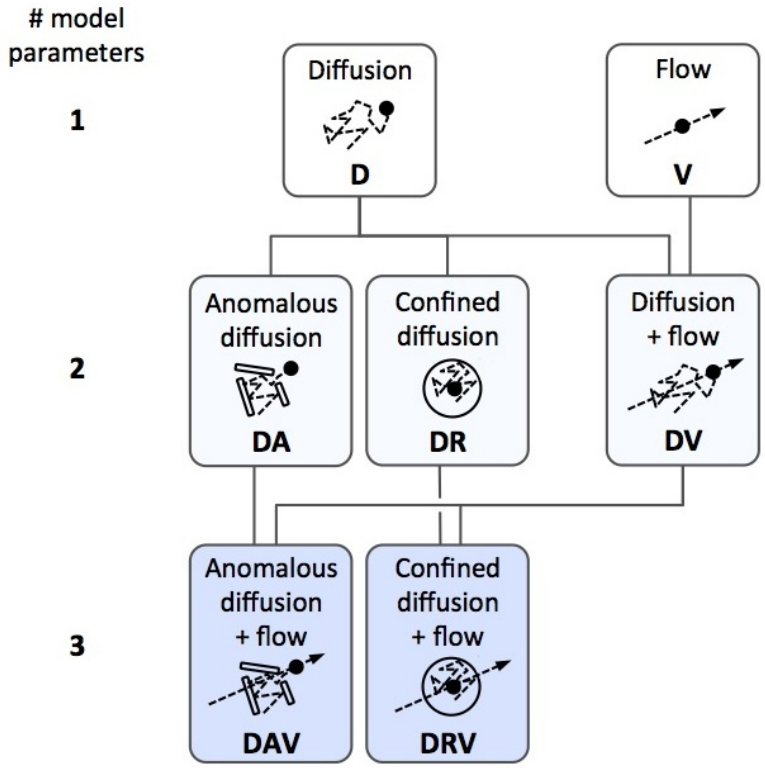

Figure 2.3: Models of particle motion. The simplest (single-parameter) models are shown in the top row, followed by the intermediate-complexity (2-parameter) models in the middle row and the most complex (3-parameter) models in the bottom row. Model abbreviations specify the parameters of each model; for example, the diffusion plus flow model (DV) has both a diffusion coefficient D and a velocity magnitude v as parameters. Lines connecting the models indicate nesting relationships; for example, both DR and DV are nested in DRV, but DAV and DRV are not nested one in the other. Figure previously published in [Monnier et al., 2012].

Implementation of the Bayesian approach thus requires the following steps to evaluate $P(\mathbf{y}|M_k)$. First, the error covariance matrix $\mathbf{C}$ is empirically estimated from the data (as described in Sections 2.3.3 and 2.3.4). Second, the integration in Equation 2.5 is performed for each model using the Laplace approximation (Section 2.3.5) and uniform parameter priors $P(\boldsymbol{\beta}_k|M_k)$ (Section 2.3.6) to obtain the final values of $P(\mathbf{y}|M_k)$ and the corresponding model probabilities. The overall procedure for applying Bayesian inference to MSD curves is summarized in Figure 2.4. Because this Bayesian framework inherently penalizes models with higher numbers of parameters, as discussed in Chapter 1, the set of competing models

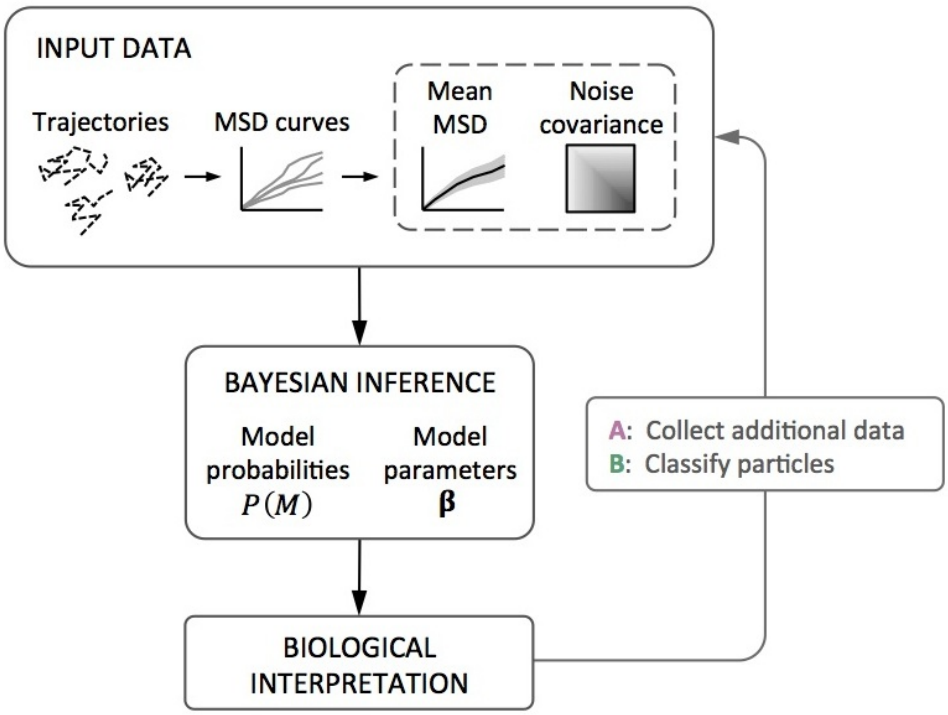

Figure 2.4: Sequence of steps for applying Bayesian inference to MSD-based analysis of particle trajectories. Starting from a set of particle trajectories, an MSD curve is calculated from each trajectory and then the set of MSD curves is used to calculate a mean MSD curve and its noise covariance matrix, which serve as inputs to the Bayesian inference procedure described in the text. The output model probabilities and parameters can be interpreted in the context of the biological system, and, if necessary to improve resolution of complex models, additional trajectories can be collected or existing trajectories can be classified into less heterogeneous subgroups (see Figure 2.11). Figure previously published in [Monnier et al., 2012].

evaluated by this method (Figure 2.3) can vary in complexity and can include nesting relationships that are difficult to treat using standard frequentist tests.

2.3.3 Empirical estimation of MSD correlated errors

As mentioned above, proper accounting for correlated errors is essential to the interpretation of MSD curves because correlated fluctuations around the analytical MSD values frequently leads to over-fitting of complex models. For example, the confined diffusion

model can fit correlated errors in pure diffusion MSD curves that causes the curve to trend downward at large time lags, and the diffusion plus flow model can fit correlated errors in pure diffusion MSD curves that causes the curve to trend upward at large time lags (examples of both of these cases can be observed in the diffusive MSD curves in Figure 2.1). We illustrate this point by fitting MSD derived from simulated trajectories undergoing simple diffusion with the motion models in Figure 2.3 using the weighted least squares form for the likelihood in Equation 2.1, which ignores correlations, versus the generalized least squares form in Equation 2.2. Significant overfitting is observed in the former but not the latter case (Figure 2.5). These results demonstrate that accounting for the effect of correlated errors is essential for properly analyzing and interpreting MSD curves, particularly when the true particle motion is a simple model such as pure diffusion. The degree to which correlated errors impact model selection varies depending on the true form of motion and the set of models included in the model selection process.

Although the analytical form of the MSD error covariance matrix has been derived for particles undergoing simple diffusion [Qian et al., 1991], the covariance matrix differs between motion models and its form has not been derived for the more complex models [Kusumi et al., 1993]. To avoid assuming any particular model of motion a priori, we use an empirical approach to estimate the error covariance matrix from multiple observations of the data $\mathbf{y}$, or multiple independent MSD curves, following recent work on fluorescence correlation spectroscopy (FCS) data [He et al., 2012; Guo et al., 2012]. The empirically estimated matrix is called the sample covariance matrix $\mathbf{S}$. The multiple MSD curves used to calculate $\mathbf{S}$ must be derived from independent, non-overlapping particle trajectories. Given J independent observations $\{\mathbf{y}^{(j)}\}_{j=1}^{J}$ of MSD values over the same set of time lags, whether from multiple or single trajectories, the residuals between each individual MSD curve and the mean MSD curve are used to estimate the variance and the covariance of the noise in the MSD estimates.

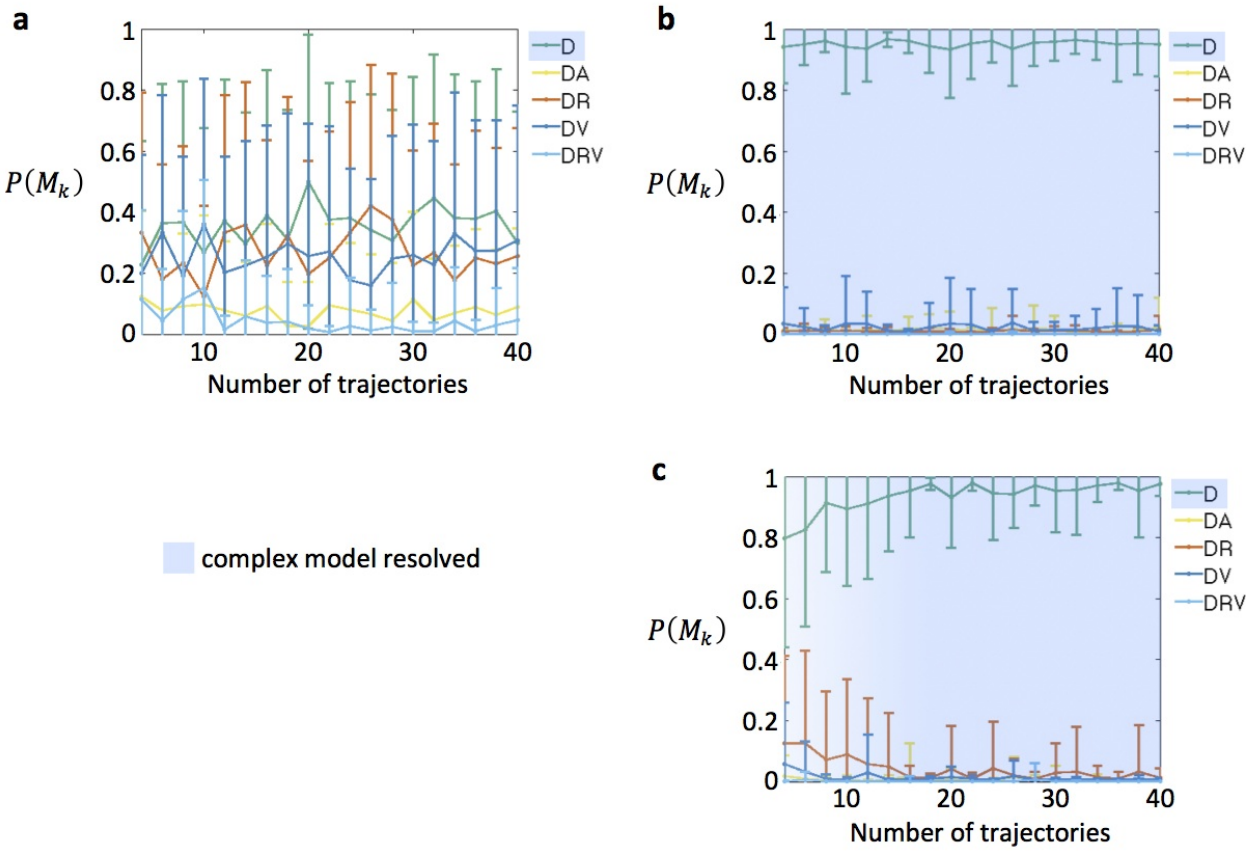

Figure 2.5: **(a)** Model probabilities obtained using weighted least squares fitting (ignoring noise covariance). Mean MSD curves with 100 points were calculated by averaging a variable number (shown on the x-axis) of independently simulated trajectories undergoing pure diffusion with the same parameters as in Figure 2.1. The resulting model probabilities are shown as means and standard deviations over 40 repetitions of the simulations and inference procedure. **(b)** Model probabilities obtained using generalized least squares (GLS) fitting (accounting for noise covariance) of the simulated trajectories undergoing pure diffusion as in (a), using the analytical form of the noise covariance [Qian et al., 1991; Michalet, 2010] (as shown in Figure 2.2) as the covariance matrix $\mathbf{C}$. Light blue shading indicates the range over which the true model (pure diffusion, D) can be resolved. **(c)** Model probabilities obtained using GLS fitting as in (b), but using the regularized sample noise covariance matrix (see Sections 2.3.3 and 2.3.4) as the covariance matrix $\mathbf{C}$. Figure previously published in [Monnier et al., 2012].

For each MSD curve j, the residual at each time lag τ_j is given by,

$$\epsilon_i^{(j)} = y_i^{(j)} - \bar{y}_i \ , \tag{2.6}$$

where $\bar{y}_i = \sum_j y_i^{(j)}/J$. The entries in the sample covariance matrix $\mathbf{S}$ for the mean MSD curve are then equal to,

$$S_{ii'} = \frac{1}{J(J-1)} \sum_{j=1}^{J} \epsilon_i^{(j)} \epsilon_{i'}^{(j)} \ . \tag{2.7}$$

The multiple observations of the MSD values are also used to calculate a mean MSD curve for the model fitting and model selection procedure.

The independent MSD curves used above can be obtained either from multiple particle trajectories or from splitting a single particle trajectory into multiple non-overlapping and independent sub-trajectories. In the latter case, the full trajectory of $N-1$ steps (N position measurements) is divided into J non-overlapping sub-trajectories of $\lfloor (N-1)/J \rfloor$ steps each. There is then a tradeoff between accurate estimation of the covariance matrix (which improves with the number of sub-trajectories J) and the time range spanned by the MSD curve, which is limited by the number of steps in the sub-trajectories. We note that in the case of anomalous diffusion as the true underlying motion model, the use of non-overlapping sub-trajectories may not be sufficient to ensure independence, as consecutive displacements along the trajectory may be correlated. In this case, the de-correlation time required to obtain independent sub-trajectories can be estimated using block-transformation [Guo et al., 2012; Flyvbjerg and Petersen, 1989] or a related approach.

2.3.4 Regularization of the error covariance matrix

Because the number of available independent observations of the MSD values from biological datasets is frequently less than the dimension of the covariance matrix, the sample covariance matrix $\mathbf{S}$ may be singular and require regularization [Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005]. We tested multiple regularization methods and found that a shrinkage approach to regularization [Schäfer and Strimmer, 2005] performs well when low numbers of MSD curves are available. The shrinkage estimator of the covariance matrix, $\mathbf{S}^*$ is a linear combination of the sample covariance matrix and a shrinkage target $\mathbf{T}$,

$$\mathbf{S}^* = \lambda \mathbf{T} + (1 - \lambda)\mathbf{S} \,, \tag{2.8}$$

where λ is the shrinkage weight, which is calculated from the uncertainty in $\mathbf{S}$ as described in [Schäfer and Strimmer, 2005]. As the number of independent curves J increases, the uncertainty in $\mathbf{S}$ decreases and the shrinkage weight also decreases so that $\mathbf{S}^*$ is closer to $\mathbf{S}$. This shrinkage estimator $\mathbf{S}^*$ is then used as the covariance matrix $\mathbf{C}$ in generalized least squares fitting of the mean MSD curve.

We found that shrinkage to a target that is a diagonal matrix with the mean MSD variance along the diagonal ("Target B" in [Schäfer and Strimmer, 2005]) performs best when low numbers of observed MSD curves are available. This method performs nearly as well as using the analytical covariance matrix in the case of normal diffusion (Figure 2.5c). It gives model preferences that are nearly indistinguishable from those obtained using the analytical covariance matrix when 10 or more MSD curves are used and on average continues to prefer the true pure diffusion model down to 4 independent MSD curves (Figure 2.5c), which we found was the minimum number of curves that could be used to reliably obtain a non-singular covariance matrix.

2.3.5 Laplace approximation for likelihood integration

While numerical integration is required to evaluate the integral in Equations 2.4 and 2.5 in general, the Laplace approximation [Kass and Raftery, 1995; He et al., 2012] can be used to perform this integration analytically by assuming the integrand $P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)P(\boldsymbol{\beta}_k|M_k)$ is well approximated by a multivariate Gaussian distribution around the estimated parameter values that maximize its value, $\hat{\boldsymbol{\beta}}_{k,\text{Bayes}} = \arg\max_{\boldsymbol{\beta}_k} P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)P(\boldsymbol{\beta}_k|M_k)$. Because we use a uniform prior parameter distribution $P(\boldsymbol{\beta}_k|M_k)$ (see Section 2.3.6), $\hat{\boldsymbol{\beta}}_{k,\text{Bayes}}$ is equal to the maximum likelihood point estimate $\hat{\boldsymbol{\beta}}_{k,\text{MLE}} = \arg\max_{\boldsymbol{\beta}_k} P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)$ and the Gaussian approximation applies to the likelihood $P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)$. Following an earlier treatment of fluorescence correlation spectroscopy datasets with Bayesian inference [He et al., 2012], we take the mean and covariance of the Gaussian approximation to be $\hat{\boldsymbol{\beta}}_{k,\text{MLE}}$ and $\hat{\boldsymbol{\Sigma}}_{k,\text{MLE}}$, respectively, as found by generalized least squares fitting as described in Section 2.3.1. Thus the form of the Laplace approximation is,

$$P(\mathbf{y}|\boldsymbol{\beta}_k, M_k) \approx P(\mathbf{y}|\hat{\boldsymbol{\beta}}_{k,\text{MLE}}, M_k) \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_{k,\text{MLE}}\right)^T \hat{\boldsymbol{\Sigma}}_{k,\text{MLE}}^{-1} \left(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_{k,\text{MLE}}\right)\right) , \quad (2.9)$$

and the integral in Equation 2.4 evaluates to,

$$P(\mathbf{y}|M_k) = (2\pi)^{p/2} |\hat{\boldsymbol{\Sigma}}_{k,\text{MLE}}|^{1/2} P(\mathbf{y}|\hat{\boldsymbol{\beta}}_{k,\text{MLE}}, M_k) P(\boldsymbol{\beta}_k|M_k) , \quad (2.10)$$

where p is the number of parameters in the model. The Laplace approximation is asymptotically exact in the limit of high amounts of data, which is not true of derived metrics such as the Aikake Information Criterion [Raftery, 1995]. The Bayesian Information Criterion is an alternative commonly-used special case of the Laplace approximation [Raftery, 1995; Kass and Wasserman, 1995], but does not sufficiently penalize model complexity in the case of small sample sizes [He et al., 2012].

2.3.6 Parameter priors

The prior probability $P(\boldsymbol{\beta}_k|M_k)$ of the parameters for each model is assumed to be uniform over a range $\beta_{kl}^{(\min)}$ and $\beta_{kl}^{(\max)}$ for each parameter β_{kl} in the set $\boldsymbol{\beta}_k$. Therefore $P(\boldsymbol{\beta}_k|M_k)$ is a constant, equal to the product of the uniform probabilities within each β_{kl} range,

$$P(\boldsymbol{\beta}_k|M_k) = \prod_l \frac{1}{\beta_{kl}^{(\max)} - \beta_{kl}^{(\min)}} \ . \tag{2.11}$$

We choose each range $\left[\beta_{kl}^{(\min)}, \beta_{kl}^{(\max)}\right]$ to be centered at the maximum likelihood estimate $\beta_{kl,\mathrm{MLE}}$ of each parameter and to span 200 times the uncertainty (standard deviation) in that parameter as obtained from the covariance matrix $\hat{\boldsymbol{\Sigma}}_{k,\mathrm{MLE}}$ [He et al., 2012]. Thus, parameters with higher uncertainties reduce the likelihood of a model more than parameters with smaller uncertainties.

2.4 Performance of MSD-Bayes on simulated trajectories

To evaluate the performance of the Bayesian procedure for MSD analysis (MSD-Bayes) in a controlled setting, we applied it to simulated trajectories of particles undergoing Brownian motion with flow (Figures 2.6 and 2.7) or within a confined spherical corral (Figure 2.8). Mean MSD curves obtained from simulated or experimental trajectories were fit with the models given in Equations 1.2-1.5 as well as additive combinations of Equation 1.5 with Equations 1.2-1.4 (e.g. Equation 1.6) using the Bayesian approach described in Section 2.3 above.

2.4.1 Implementation of particle trajectory simulations

Diffusive single-particle trajectories were simulated in three dimensions by drawing random step lengths in each of the three Cartesian directions from a Gaussian distribution with zero mean and standard deviation equal to $\sqrt{2D\Delta t}$, where D is the diffusion coefficient and Δt is the time interval for each step. Confinement of a diffusing particle was modeled as a reflecting spherical boundary of radius R_c centered at the initial particle position. Directed flow was modeled by adding a fixed displacement $\mathbf{v}\Delta t$ to the diffusive motion at each time step, where $\mathbf{v}$ is the velocity vector. While we use default simulation parameters comparable to the experimental conditions observed for starfish chromosomes in Chapter 4, namely D = 0.005 μm^2/s, Δt = 2.5 s, T = 300 s, $\tau_{\max}$= $T/4$, and n = 30 trajectories per dataset, we emphasize that the illustrated properties of the proposed multiple hypothesis testing procedure are general.

2.4.2 Effect of limited sampling of stochastic motion

An important source of error in MSD values is statistical sampling noise due to experimental limitations on the number, length, and sampling rate of available single particle trajectories (SPTs) [Saxton and Jacobson, 1997; Saxton, 1997]. Trajectories were simulated with the above default parameters (Figure 2.6b), with lower noise (higher T and n; Figure 2.6a), or with higher noise (lower T and n; Figure 2.6c) while systematically varying the value of a superimposed velocity $\mathbf{v} = [v, 0, 0]$. The relative contributions of diffusive and directed motion to the diffusion plus flow (DV) MSD equation (Equation 1.6) are of similar magnitude when $\tau \sim \tau_{\mathrm{DV}} \equiv 6D/v^2$. The Bayesian approach strongly prefers the DV model for v values corresponding to a timescale τ_{DV} that is comparable to the time lags covered by the MSD curve (Figure 2.6b). The simpler D and V models are preferred

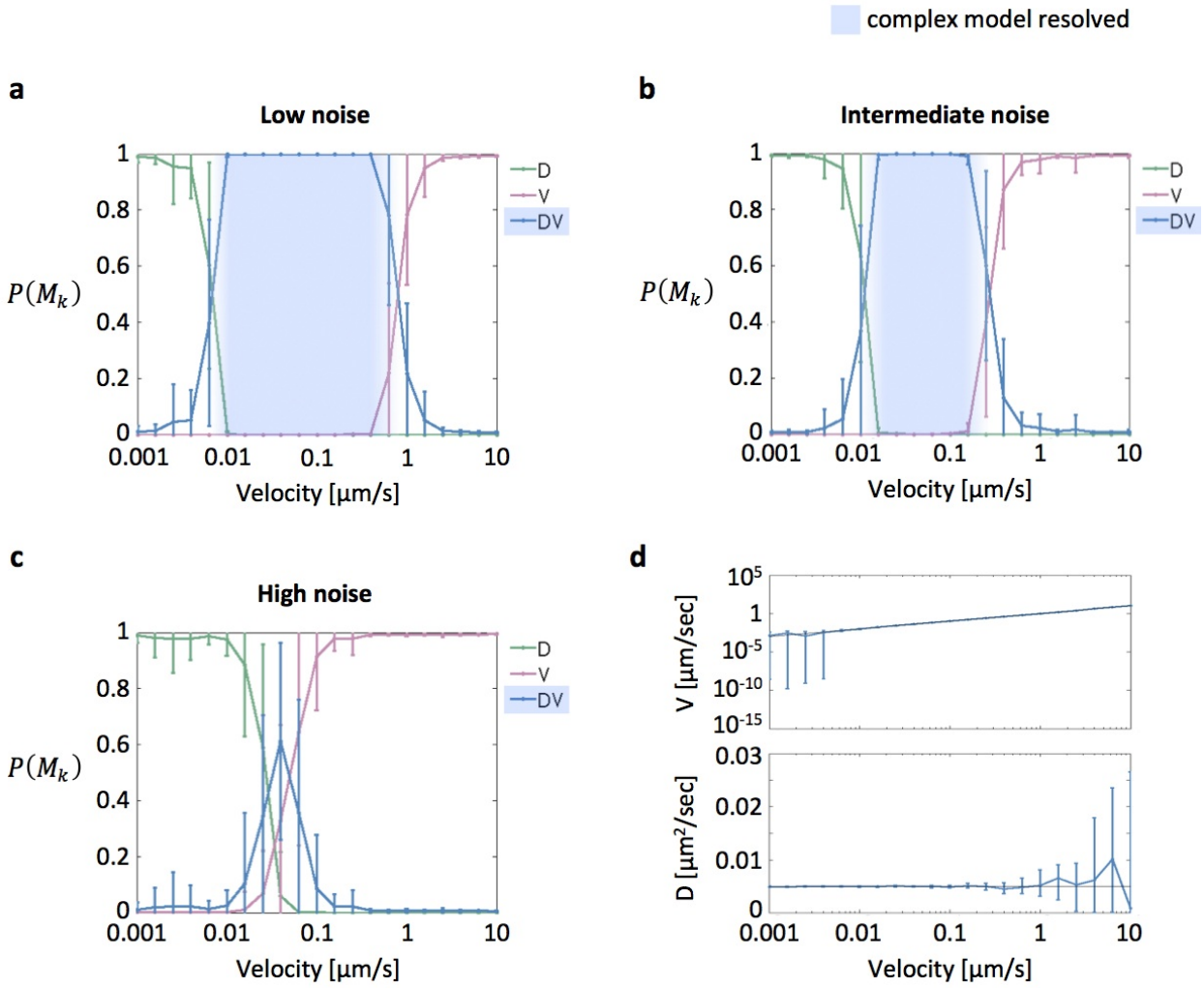

Figure 2.6: Model probabilities for simulated trajectories undergoing diffusion plus flow (DV) with different levels of noise due to sampling limitations. For all conditions, $D = 0.005$ μm^2/s, $\Delta t = 2.5$ s, and v varies as shown along the x-axis. The total time T of the trajectories varied from 600 s for the low noise condition **(a)** to 300 s for the intermediate noise condition **(b)** to 150 s for the high noise condition **(c)**. The number of independent observed trajectories n (corresponding to one simulated dataset) varied from 60 **(a)** to 30 **(b)** to 5 **(c)**. MSD curves with 30 points (up to $\tau_{\max} = 75$ s) were calculated for each of the n trajectories and used to calculate a mean MSD curve and error covariance matrix as input to the Bayesian inference procedure. The resulting model probabilities are shown as means and standard deviations over 50 repetitions of the simulations and inference procedure. Light blue shading indicates the range of velocity values over which the true model (diffusion plus flow, DV) can be resolved given the simulated experimental parameters. **(d)** Estimated values of D and v obtained from fitting the true diffusion plus flow (DV) model for the intermediate noise condition shown in (b) are plotted as medians and quartiles (blue lines) in comparison with the true values of D and v used in the simulation (black lines). Figure previously published in [Monnier et al., 2012].

at low and high v, respectively, where the contribution of the v or D parameter to the more complex DV model is not significant given the level of noise in the mean MSD curve. The locations of these crossovers to simpler preferred models at low and high v depend on the level of sampling noise (Figure 2.6). Examination of the fit parameter values for the true DV model (Figure 2.6d) shows that when the DV model probability is high both parameter values are well estimated, whereas their values become poorly estimated when the model probability is low. Thus, the Bayesian multiple hypothesis testing framework not only selects the appropriate model that is justified given the empirical level of noise, it also provides a pre-screening filter for downstream physical or biological interpretation of model parameter values, which are only reliable when the model to which they belong is strongly preferred.

We next independently varied three contributing factors to the sampling noise—trajectory number, trajectory length, and sampling rate—at fixed values of v (Figure 2.7). Starting with the default simulation parameters above and a fixed value of $v = 0.1$ μm/s near the right-hand crossover point in Figure 2.6b, decreasing the number of trajectories used to calculate each mean MSD curve and associated error covariance matrix from 30 to 4 results in loss of the ability to resolve the DV model over the simpler V model due to the increasing level of noise (Figure 2.7a). Decreasing T from 300 s to 40 s reduces the ability to resolve the v component of the motion (Figure 2.7b), while increasing Δt from 2.5 s to 15 s at a fixed T reduces the ability to resolve the D component of the motion (Figure 2.7c), due to the difference in the relative contributions of diffusion and flow to the MSD curve at high and low τ.

To test whether this Bayesian procedure applies generally to other motion models in addition to diffusion and flow, we repeated the above tests on simulations of confined diffusion (Figure 2.8). Here we also included the full set of competing models shown in Figure 2.3 to

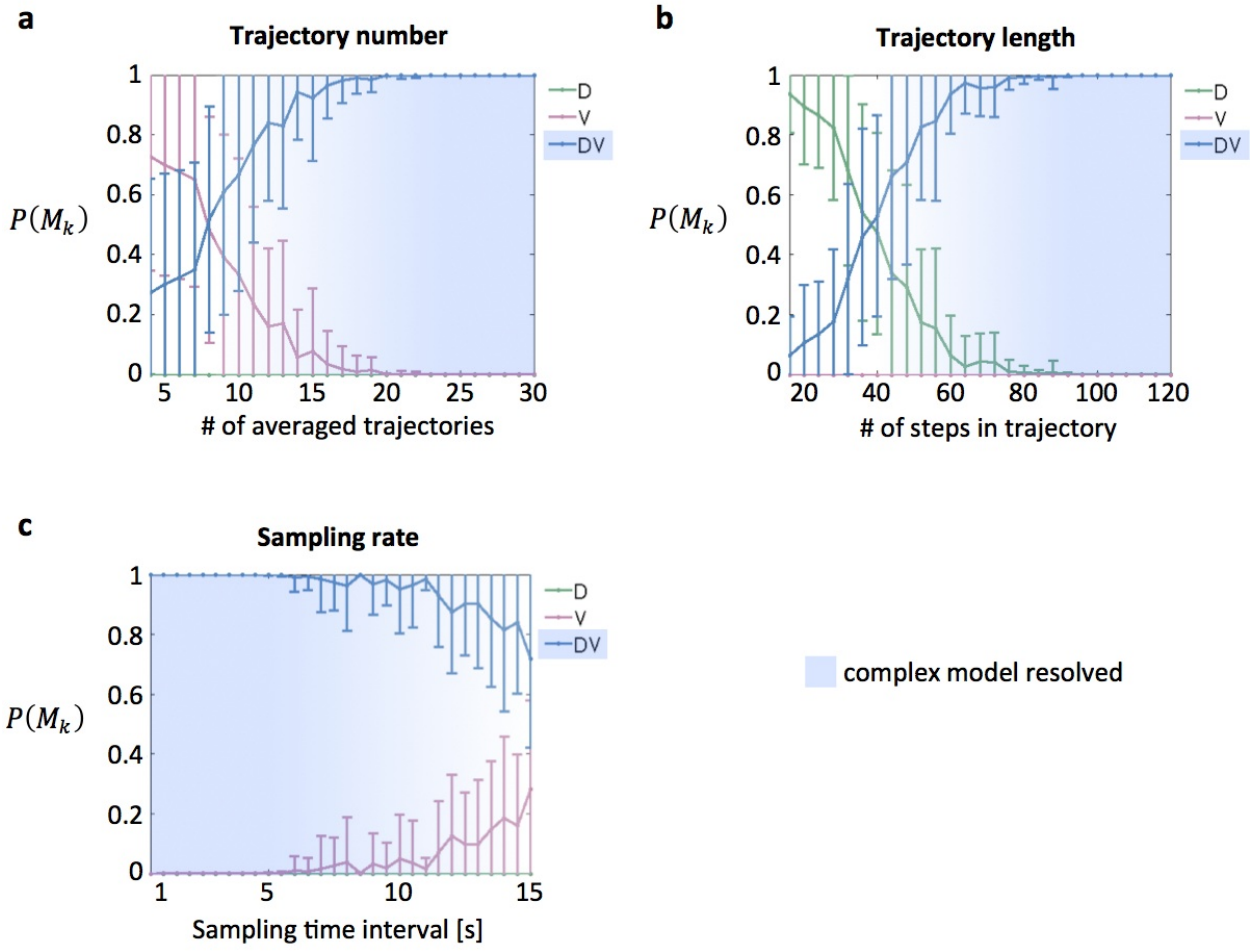

Figure 2.7: Model probabilities for trajectories simulated as shown in Figure 2.6b but at a fixed velocity and systematically varying one of the sampling parameters. **(a)** Velocity is fixed at $v = 0.1$ μm/s and the number of trajectories per dataset n is varied from 30 down to 4. **(b)** Velocity is fixed at $v = 0.02$ μm/s and the total trajectory time T is varied from 300 s down to 40 s (from 120 steps to 16 steps per trajectory). The number of points in the MSD curves is held constant at 1/4 of the number of steps in the trajectory. **(c)** Velocity is fixed at $v = 0.1$ μm/s and the sampling interval Δt is varied from 0.5 s up to 15 s with the total time of the trajectories held constant at 300 s. The number of points in the MSD curves is again held constant at 1/4 of the number of steps in the trajectory. Figure previously published in [Monnier et al., 2012].

test the robustness of the model selection procedure in the presence of both higher and lower complexity competing models. Confinement makes a significant contribution to the confined diffusion (DR) MSD equation (Equation 1.4) when the ratio $6D\tau/R_c^2$ is on the order of 1, or when $\tau \sim \tau_{\mathrm{DR}} \equiv R_c^2/6D$. The Bayesian approach strongly prefers the DR model when this ratio τ_{DR} is below the maximum τ in the MSD curve, whereas the simpler D model is preferred for larger confinement radii (Figure 2.8a). As above, the exact crossover point depends on the level of noise in the mean MSD curve. For a fixed value of R_c, increasing the trajectory time sampling interval Δt past the ratio τ_{DR} results in loss of the ability to resolve the diffusive component of the motion, making the particle appear stationary (Figure 2.8b), and decreasing n or T also reduces the ability to resolve the DR model (Figure 2.8c-d).

2.4.3 Effect of heterogeneity between particles

The above results demonstrate that the MSD-Bayes approach can be used to detect both confinement and directed motion in a systematic manner that accounts appropriately for the noise due to sampling limitations, avoiding over-fitting of complex models. However, heterogeneity in motion type between the particle trajectories used to calculate the mean MSD curve and the error covariance matrix may introduce additional noise and reduce the ability to resolve the underlying physical process. We note that variation in motion type within a single trajectory has a similar effect, and a Bayesian method to detect within-trajectory heterogeneity is explored in Chapter 3.

For particles undergoing directed motion, the contribution of directed motion to the MSD curve is $MSD_{\mathrm{V}}(\tau) = v^2\tau^2$ (Equation 1.5). If there is heterogeneity in the value of v for each particle, such that each particle j of J total particles has a velocity v_j, then the

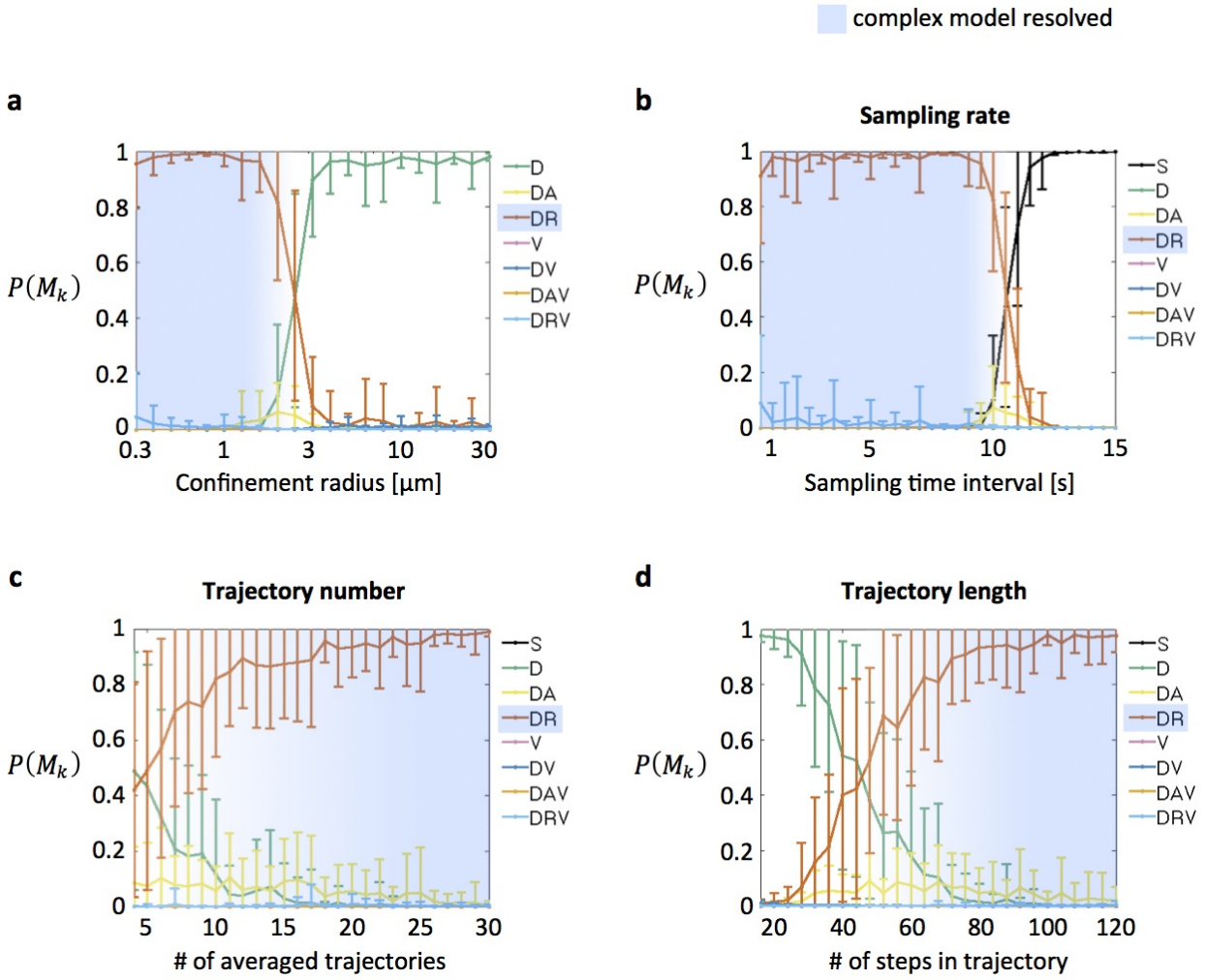

Figure 2.8: **(a)** Model probabilities for simulated trajectories undergoing confined diffusion (DR) inside a reflecting spherical boundary with $D = 0.005$ μm^2/s, time sampling interval $\Delta t = 2.5$ s, total time $T = 300$ s, and varying confinement radius R_c as shown along the x-axis. Analysis is performed as in Figure 2.6, but using the full set of motion models in Figure 2.3. **(b)** Trajectories are simulated as in (a) but at a fixed confinement radius $R_c = 0.4$ μm. The sampling interval Δt is varied from 0.5 s up to 15 s as in Figure 2.7c. S represents a stationary-particle model including only a constant term. **(c)** R_c is fixed at 1.5 μm, and the number of trajectories per dataset n is varied from 30 down to 4, as in Figure 2.7a. **(d)** R_c is fixed at 1.5 μm, and the total trajectory time T is varied from 300 s down to 40 s (from 120 steps to 16 steps per trajectory), as in Figure 2.7b. The number of points in the MSD curves is held constant at 1/4 of the number of steps in the trajectory. Figure previously published in [Monnier et al., 2012].

mean MSD curve takes the form,

$$MSD_{\mathrm{V}}(\tau) = \frac{1}{J}\sum_{j=1}^{J} v_j^2\tau^2 = \frac{\tau^2}{J}\sum_{j=1}^{J} v_j^2 \ . \tag{2.12}$$

This mean MSD curve has the same quadratic dependence on τ as the original MSD curve for directed motion, with an effective velocity for the heterogeneous population given by,

$$v_{\mathrm{eff}} = \sqrt{\frac{1}{J}\sum_{j=1}^{J} v_j^2} \ . \tag{2.13}$$

In this case, the directed motion model will still be a good fit to the heterogeneous MSD curve, but the v_{eff} parameter is biased toward the larger magnitude velocities in the particle population and will be greater than the mean velocity in the population (as is observed in Figure 2.9a below).

For particles undergoing simple diffusion, the contribution of their diffusive motion to the MSD curve is $MSD_{\mathrm{D}}(\tau) = 6D\tau$ (Equation 1.2). If there is heterogeneity in the value of D for each particle, then the mean MSD curve takes the form,

$$MSD_{\mathrm{D}}(\tau) = \frac{1}{J}\sum_{j=1}^{J} 6D_j\tau = \frac{6\tau}{J}\sum_{j=1}^{J} D_j \ . \tag{2.14}$$

This mean MSD curve has the same linear dependence on τ as the original MSD curve for pure diffusion, with an effective diffusion coefficient for the heterogeneous population given by,

$$D_{\mathrm{eff}} = \frac{1}{J}\sum_{j=1}^{J} D_j \ . \tag{2.15}$$

In this case, the diffusion model will still be a good fit to the heterogeneous MSD curve, and the D_{eff} parameter is equal to the mean diffusion coefficient of the population.

For particles undergoing confined diffusion, the contribution of their diffusive motion to the MSD curve is $MSD_{\mathrm{DR}}(\tau) = R_c^2(1 - \exp(-6D\tau/R_c^2))$ (Equation 1.4). If there is heterogeneity in the value of R_c for each particle, then the mean MSD curve takes the form,

$$MSD_{\mathrm{DR}}(\tau) = \frac{1}{J}\sum_{j=1}^{J} R_{c_j}^2 \left(1 - \exp(-6D\tau/R_{c_j}^2)\right) . \tag{2.16}$$

Unlike the models above, this mean MSD curve does not have the same τ dependence as the original MSD curve for confined diffusion and can no longer be described by the confined diffusion model with a single effective confinement radius. At low heterogeneity in R_c, the confined diffusion model may still describe this behavior better than the other tested models if the plateau region of the MSD curve is still present at long time lags, in which case this model may still be preferred by Bayesian inference (as in Figure 2.9b, discussed below), but with larger variability in model probabilities due to the increased apparent noise in the mean MSD.

To test these predicted effects of heterogeneity in a controlled setting, we simulated trajectories as above but allowed a single parameter (v or R_c) to vary randomly between particles according to a normal distribution. As a result, even perfectly-measured MSD curves from infinite trajectories would still vary between the different particles, introducing an apparent noise into the mean MSD curve estimate. As heterogeneity between particles is increased by increasing the standard deviation of the distribution of v or R_c values, the ability to resolve the true motion model diminishes in favor of simpler models due to this increase in apparent noise (Figure 2.9). In addition, for particles undergoing diffusion plus flow, the estimated value of v obtained from the DV model is systematically higher than the true mean velocity as predicted above (Figure 2.9a). For particles undergoing confined diffusion, heterogeneity in R_c changes the dependence of the mean MSD function on τ so that none of the models describes the resulting mean MSD curve satisfactorily (Figure 2.9b).

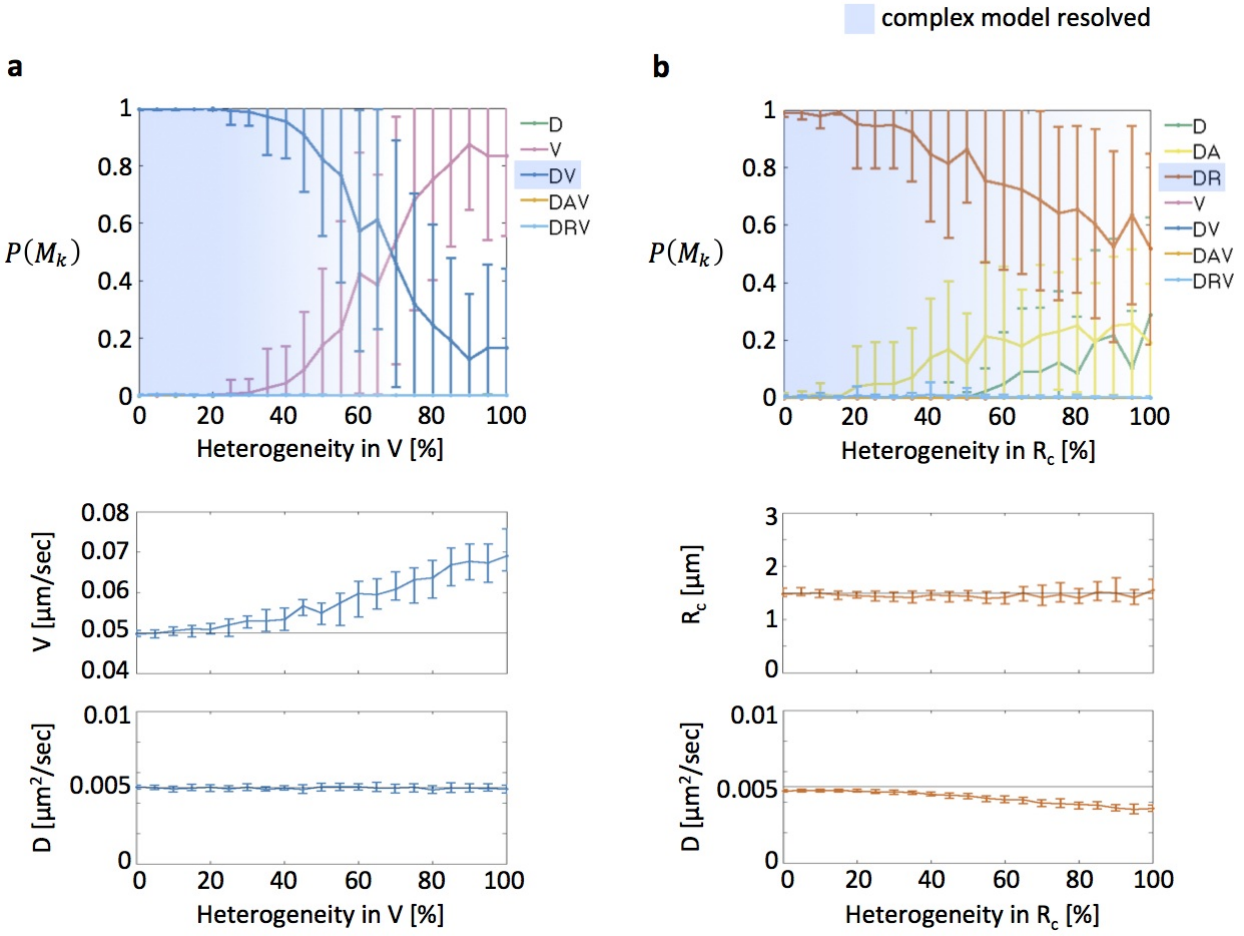

Figure 2.9: **(a)** Model probabilities for simulated trajectories as in Figure 2.6b but with the velocity of each particle drawn from a normal distribution centered on $v = 0.05$ μm/s with standard deviation (as a percentage of the mean) as shown on the x-axis. Estimated values of D and v obtained from fitting the true diffusion plus flow (DV) model are plotted as in Figure 2.6b. **(b)** Model probabilities for simulated trajectories as in Figure 2.8a but with the confinement radius of each particle drawn from a normal distribution centered on $R_c = 1.5$ μm with standard deviation (as a percentage of the mean) as shown on the x axis. Estimated values of D and R_c obtained from fitting the true confined diffusion (DR) model are plotted as in (a). Figure previously published in [Monnier et al., 2012].

The apparent diffusion coefficient decreases with increasing heterogeneity in R_c because the diffusion timescale is affected disproportionately by larger confinement radii.

2.5 Validation on experimental datasets

To test the performance of the MSD-Bayes procedure on experimental biological datasets, we first applied it to the motion of chromosomes during meiosis I in starfish oocytes. Chromosomes are transported towards the spindle at the animal pole (AP) of the oocyte (Figure 2.10a) by homogeneous contraction of a large actin network that forms in the nuclear region after nuclear envelope breakdown (NEBD) [Mori et al., 2011; Lénárt et al., 2005]. Chromosomes from 4 oocytes were imaged and tracked at 2.6-second time resolution, a more than 5-fold improvement in resolution over previous studies [Mori et al., 2011], during the 6-minute actin-dependent transport phase (Figure 2.10a). We analyzed the mean MSD curve over all 30 chromosome trajectories (Figure 2.10b) using the Bayesian inference approach to test the full set of motion models shown in Figure 2.3. The DV model is strongly preferred over the other models, consistent with the previously proposed hypothesis that chromosomes diffuse within the actin network as they are transported in a directed manner towards the spindle [Mori et al., 2011]. This result indicates that the chromosome trajectories provide significant evidence for both the diffusive and directed components of their motion but do not provide significant evidence for additional complexity such as confinement or anomalous diffusion, which could potentially result from steric interactions with the actin network structure [Mori et al., 2011] or the viscoelastic nature of the actin network [Wong et al., 2004]. These more complex motions are not necessarily ruled out by the above result, however, because the additional complexity of confined or anomalous diffusive models might be masked by sampling noise or heterogeneity as shown in the simulations above.

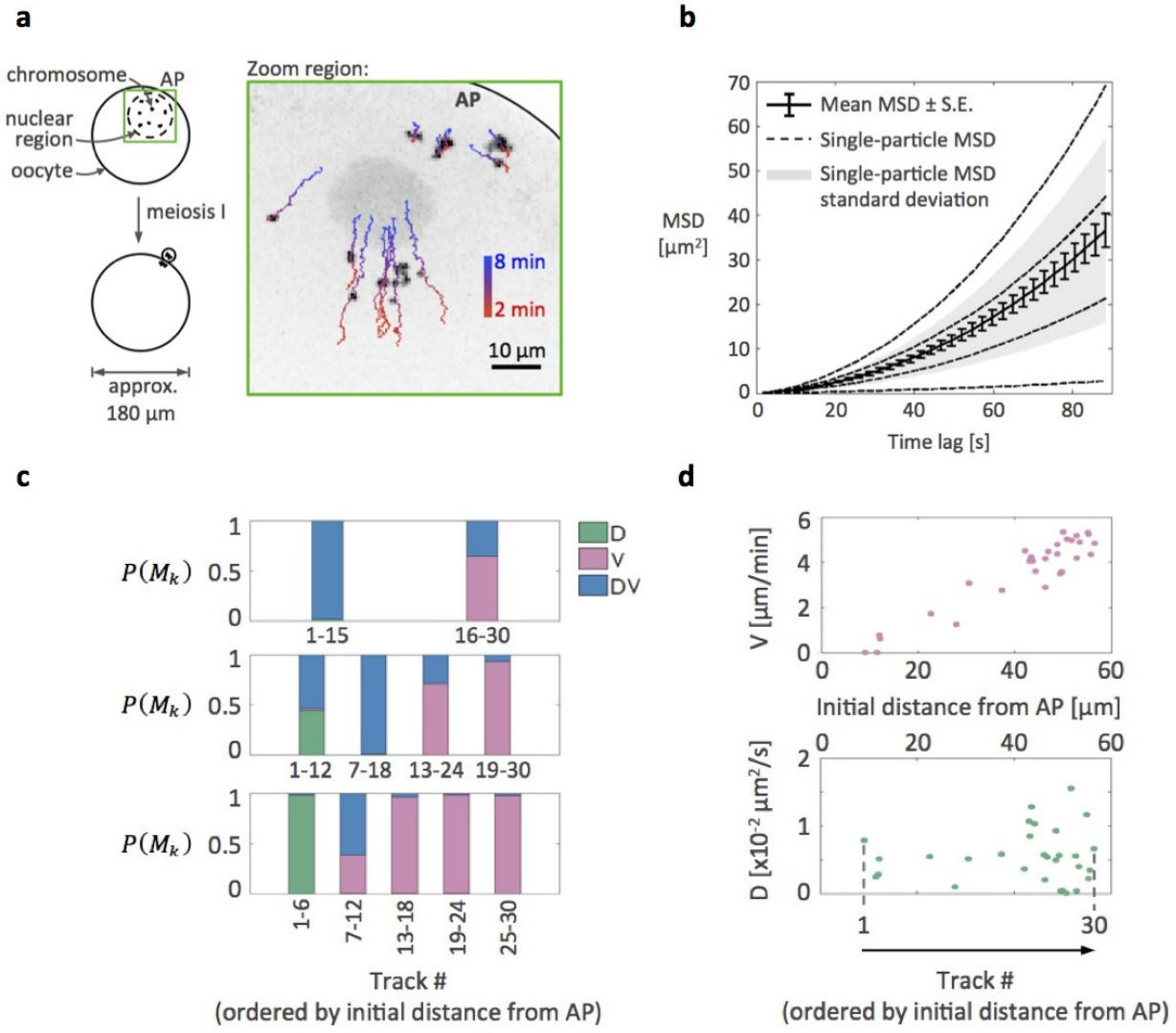

Figure 2.10: **(a)** *Left*: Cartoon of chromosome positions in the starfish oocyte at the start (top) and end (bottom) of meiosis I. AP indicates the animal pole of the oocyte towards which the chromosomes are congressing. *Right*: Maximum intensity Z-projection through a starfish oocyte nuclear region showing chromosomes labeled with H2B-GFP at 4 min after NEBD. Chromosome trajectories over the full actin transport phase are superimposed, colored from 2 min after NEBD (red) to 8 min after NEBD (blue). **(b)** Mean MSD curve with standard errors (solid black line) averaged over 30 chromosome trajectories from a total of 4 oocytes imaged at 2.6 sec time resolution for the 6 min period from 2-8 min after NEBD. Four example MSD curves from individual chromosome trajectories are shown (dashed black lines), as well as the standard deviation over all 30 of the individual-chromosome MSD curves (gray region). The preferred model by Bayesian inference is diffusion plus flow (DV) for the mean MSD curve. **(c)** Model probabilities obtained by fitting mean MSD curves over sub-groups of 15, 12 and 6 chromosomes (top to bottom), shown from left to right in order of increasing initial distance from the AP. Only the D, V, and DV model probabilities are shown (all other model probabilities were negligible). **(d)** Velocity and diffusion coefficient estimates obtained from the DV model fit to individual-chromosome MSD curves, showing the correlation of velocity with initial distance from the AP. Figure previously published in [Monnier et al., 2012].

Since the chromosomes were previously shown to have significant heterogeneity in their velocities, which are correlated with initial distance from the AP [Mori et al., 2011], we split the chromosome trajectories into equally-sized groups to reduce this heterogeneity and reanalyzed their motions (Figure 2.10c). An initial split into two groups revealed that the DV model is preferred for chromosomes closer to the AP, whereas the simpler V model is preferred farther from the AP (Figure 2.10c, top panel), confirming that there is heterogeneity along this biological coordinate. Splitting trajectories into less-heterogeneous sub-groups has a tradeoff (Figure 2.11) in that it reduces the number of trajectories per group, which was shown above to reduce the ability to resolve complex models. The effect of this tradeoff is apparent in the overall trend towards simpler models upon further sub-classification of the chromosome trajectories (Figure 2.10c). While the increase in sampling noise that results

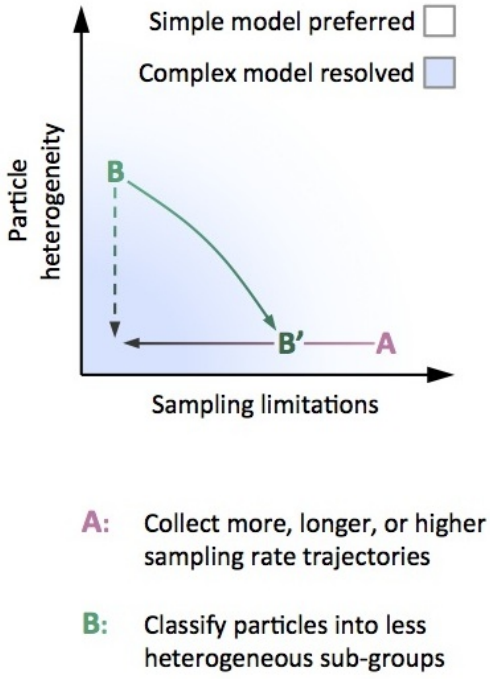

Figure 2.11: Complex models are most likely to be resolved when there is low heterogeneity between particles and low noise due to data collection limitations, such as the number, length, and sampling rate of the trajectories. The tradeoff between reducing particle heterogeneity and increasing sampling noise by splitting trajectories into smaller groups of fewer trajectories is illustrated by the transition from B to B'. Figure previously published in [Monnier et al., 2012].

from the reduction in number of SPTs per sub-group outweighs the reduction in heterogeneity in this case, additional oocytes could in principle be added to the total pool of data in order to again resolve the more complex DV model. Finally, the increasing probability of the simpler V model for chromosomes far from the AP and the simpler D model for chromosomes close to the AP is comparable to moving to the right and left, respectively, along the horizontal axes in Figure 2.6 because of the difference in velocities between these chromosomes (Figure 2.10d).

We next sought an alternate means of probing the starfish actin network that is not complicated by the networks directed motion. We examined the diffusion of 0.2-μm beads within the network by injecting them into the oocyte nucleus just prior to NEBD while simultaneously overexpressing mEGFP-UtrCH to stabilize actin bundles to prevent network contraction (Figure 2.12a). Bead trajectories have previously been used to characterize the density of obstacles, sizes of pores, and viscoelastic properties of cytoskeletal networks [Wong et al., 2004; Caspi et al., 2000]. We found that beads in the stabilized actin network exhibit a range of behaviors (Figure 2.12a) and that the mean MSD curve over multiple bead trajectories (Figure 2.12b) is best explained by the simple diffusion model, presumably due to this high heterogeneity. However, when individual bead trajectories are analyzed by splitting them into consecutive sub-trajectories (assumed to be independent) to estimate the mean MSD and noise covariance matrix for each bead (Figure 2.12c), then a variety of diffusive models are resolved, including the higher-complexity anomalous and confined diffusion models (Figure 2.12d). This dataset therefore provides an example in which heterogeneity between particles is high enough that moving from a mean MSD curve over all particles to individual-particle MSD curves improves the ability to resolve complex models despite the associated increase in sampling noise. A more detailed analysis of these heterogeneous bead dynamics will be the subject of future work.

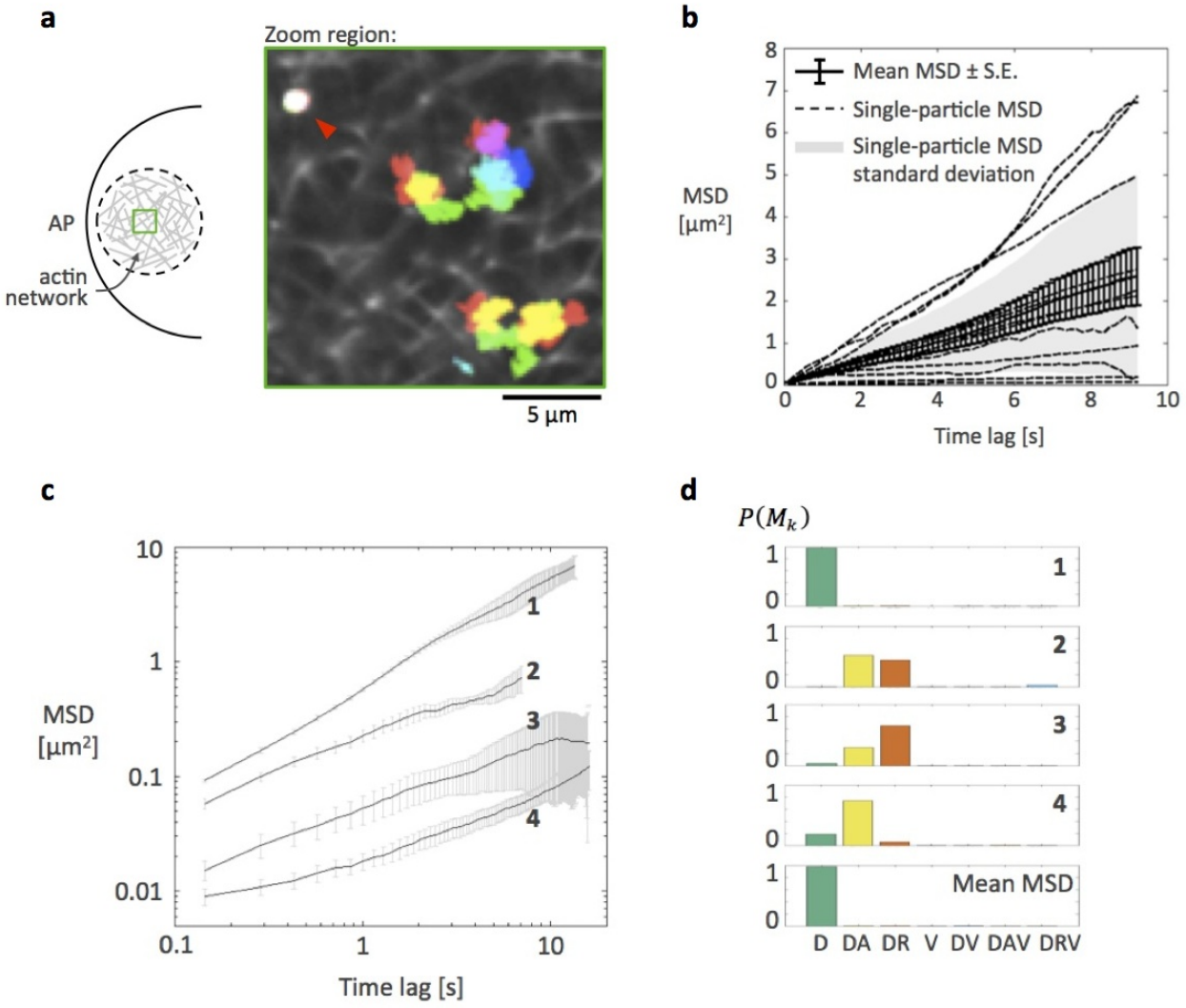

Figure 2.12: **(a)** *Left*: Cartoon of a stabilized actin network in the post-NEBD nuclear region of a starfish oocyte. *Right*: Time projection (red to blue) of the motions of 0.2 μm diameter beads in a utrophin-GFP-stabilized actin network. Some beads appear transiently immobilized (red arrowhead). **(b)** Mean and individual MSD curves as in Figure 2.10b from 12 bead trajectories in a utrophin-stabilized actin network. The preferred model by Bayesian inference is pure diffusion (D) for the mean MSD curve. **(c)** Four example MSD curves from individual beads in the actin meshwork, shown on a log-log scale. **(d)** Model probabilities for the 7 tested models fit to each of the four individual-bead MSD curves shown on the left, as well as to the mean MSD curve in (b). Figure previously published in [Monnier et al., 2012].

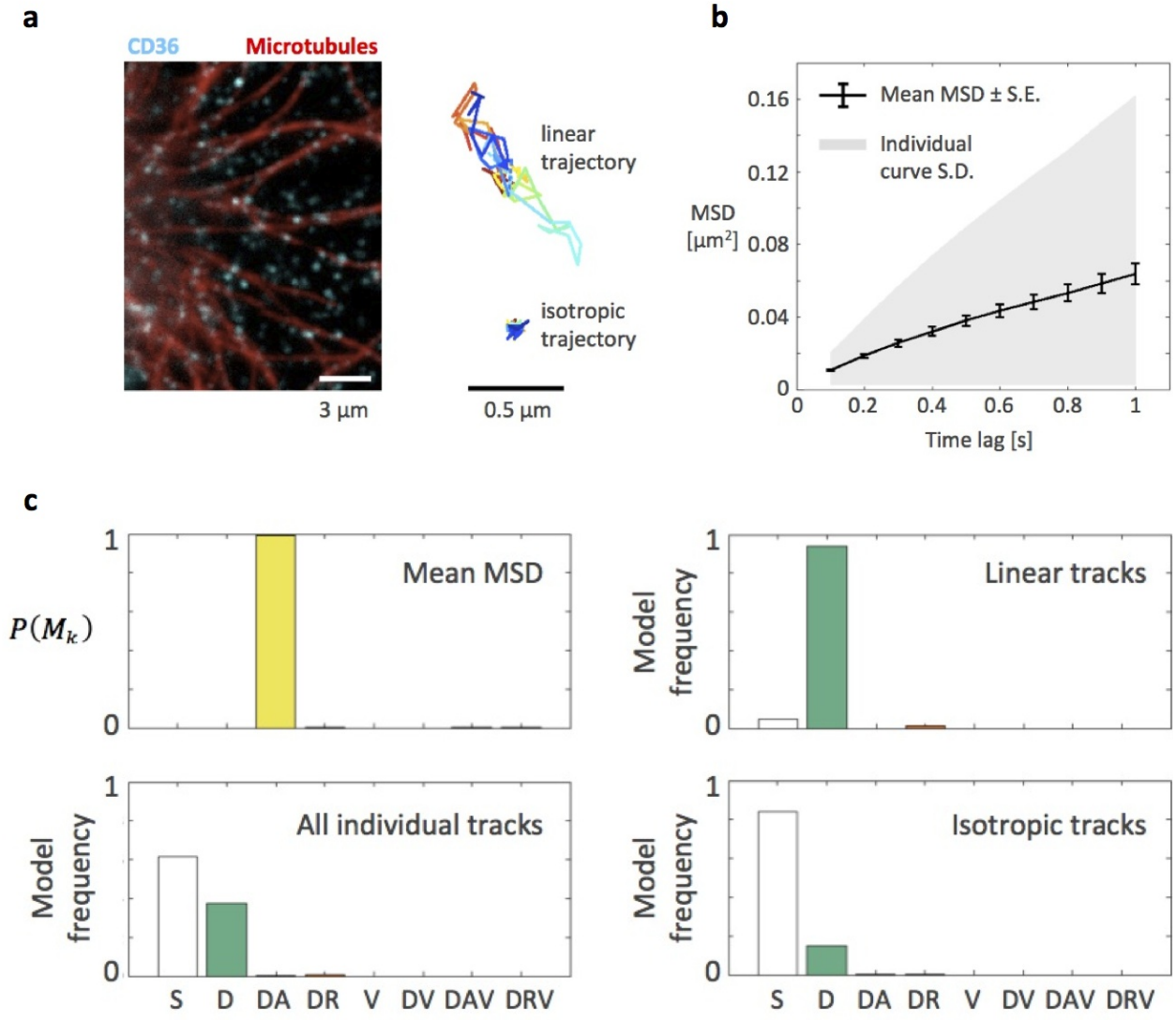

Figure 2.13: **(a)** *Left*: Image of the membrane receptor CD36 (blue) and microtubules (red) in a macrophage. Image reprinted from [Jaqaman et al., 2011] with permission from Elsevier. *Right*: Example trajectories classified as linear (top) and isotropic (bottom) by the asymmetry metric used in [Jaqaman et al., 2011], colored over time from red to blue. **(b)** Mean MSD curve over all of the CD36 trajectories that are at least 40 time steps in length (296 trajectories total). The preferred model by Bayesian inference is anomalous diffusion (DA) for the mean MSD curve. **(c)** Model probabilities for the mean MSD curve (top left). Frequency with which each of the 8 tested models (including a constant stationary-particle model as in Figure 2.8) is selected as the most probable model for all CD36 trajectories (bottom left), for the 84 linear CD36 trajectories (top right), and for the 212 isotropic CD36 trajectories (bottom right). Figure previously published in [Monnier et al., 2012].

As another example of detecting confinement in a very different biological system, we analyzed previously-published trajectories of the membrane receptor CD36 (Figure 2.13a), which exhibits a range of behaviors including linear motion, confined diffusion, and unconfined diffusion [Jaqaman et al., 2008; Jaqaman et al., 2011]. Testing the full set of motion models from Figure 2.3 with the Bayesian procedure reveals that the mean MSD curve over all CD36 trajectories (Figure 2.13b) is best fit by the anomalous diffusion model, but that individual CD36 trajectories are best explained by either pure diffusion or by the stationary-particle model described above (Figure 2.13c). The high probability of the stationary model suggests that these receptors are confined within a radius smaller than the mean diffusive step size of the trajectory, as in Figure 2.8b, or are attached to a stationary structure such as a cytoskeletal matrix [Shin et al., 2004], consistent with the confined diffusion classification in previous analysis of the trajectories [Jaqaman et al., 2011]. Pure diffusion is the preferred model for nearly all of trajectories previously classified as linear (Figure 2.13c), confirming that these motions are linear due to 1D diffusion (for example, along 1D tracks or within linear-shaped confinement zones), whereas the stationary model is preferred for most of the previously-classified isotropic trajectories (Figure 2.13c). Only a small fraction of receptors exhibit isotropic unconfined diffusion.

The above examples illustrate that a single automated Bayesian approach can be used to detect both directed motion and confinement or anomalous diffusion in a variety of biological systems. We next sought to detect both types of motion within a single biological dataset. Kinetochores in mouse oocytes (Figure 2.14a) were recently found to exhibit distinct complex motions during discrete time phases during meiosis [Kitajima et al., 2011]. Analyzing the mean MSD curve over the entire period of meiosis (Figure 2.14b) with the Bayesian procedure reveals that the highest-probability model for the mean behavior of the kinetochores is confined diffusion. However, sequentially dividing the kinetochore trajectories into time periods corresponding to the previously-described phases [Kitajima et al., 2011]

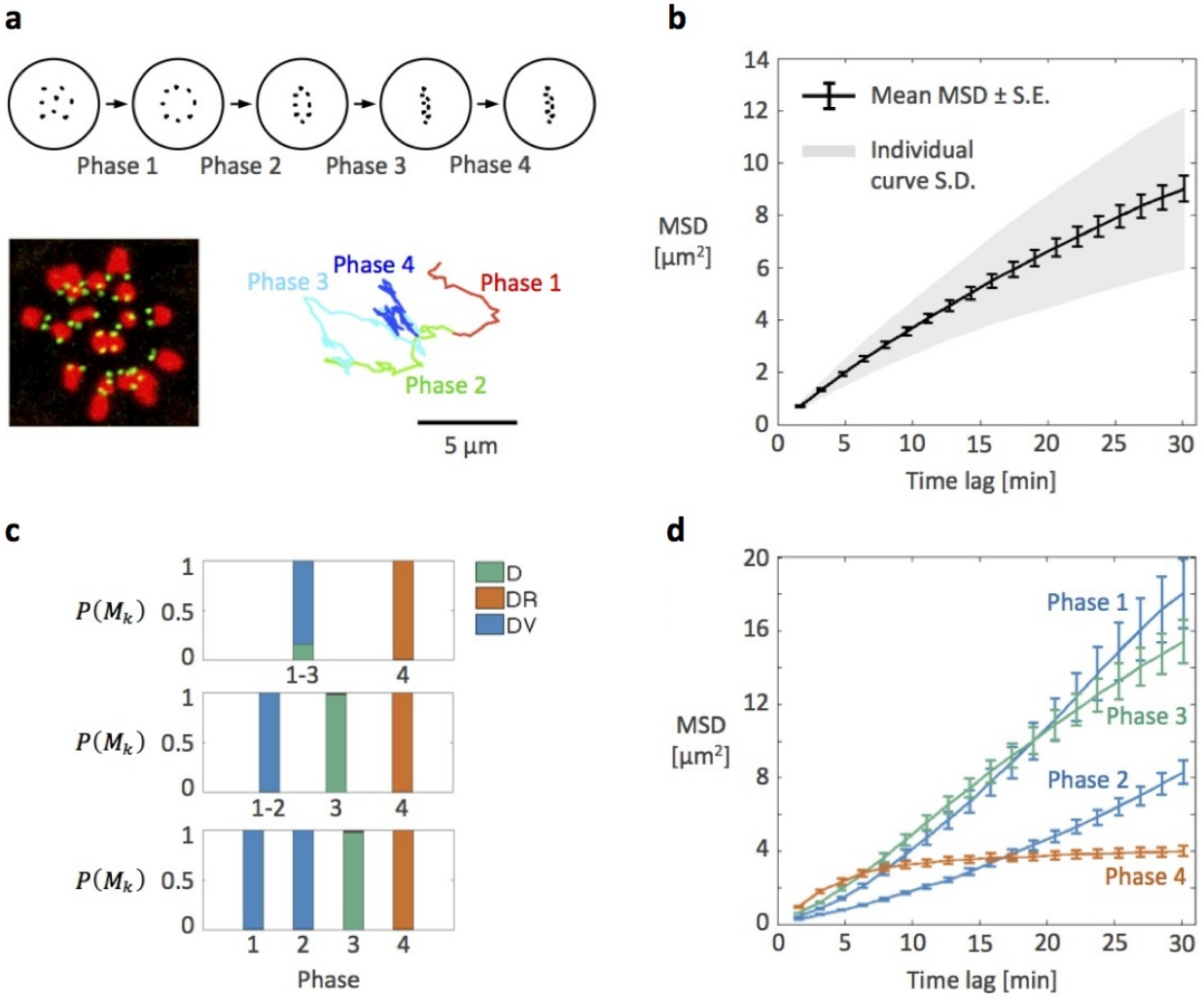

Figure 2.14: **(a)** *Top*: Cartoon of kinetochore motions during the different time phases defined in [Kitajima et al., 2011] leading up to the first meiotic division in mouse oocytes. *Bottom left*: Mouse kinetochores (green) and chromosomes (red) in a maximum intensity Z-projection through the spindle at the beginning of phase 2. Image reprinted from [Kitajima et al., 2011] with permission from Elsevier. *Bottom right*: Example kinetochore trajectory showing the four phases of motion. **(b)** Mean MSD curve over all 40 kinetochore trajectories from a single oocyte during the full 8.7-hour period of meiosis. The preferred model by Bayesian inference is confined diffusion (DR) for the mean MSD curve. **(c)** Model probabilities obtained by fitting mean MSD curves over all 40 kinetochore trajectories split into time phases as shown in (d). Only the D, DR, and DV model probabilities are shown (all other model probabilities were negligible). **(d)** Mean MSD curves over all 40 kinetochore trajectories for the individual time phases, corresponding to the model probabilities shown in the bottom plot on the left. Figure previously published in [Monnier et al., 2012].

reveals that confinement is localized to phase 4, whereas diffusion plus flow is preferred for phases 1 and 2 and pure diffusion is preferred for phase 3 (Figure 2.14c,d). Importantly, in future studies or screens, this Bayesian procedure may be used in an automated manner to discover the above phase boundaries a priori by systematic evaluation of boundary locations and number of phases.

2.6 Conclusion

The MSD-Bayes approach presented here handles multiple competing models for single-particle motion simultaneously, preferring simpler models when statistical noise and heterogeneity preclude the resolution of more complex models that are not justified by the data. Statistical noise due to sampling limitations and heterogeneity between particles limit the ability to resolve complex motion models. Sampling noise may be reduced by collecting more data, namely longer or more trajectories, in order to improve the statistical accuracy of estimates of the mean MSD and its correlated error (Figure 2.11, Case A). However, heterogeneity within a trajectory or across multiple trajectories may only be reduced by appropriately segmenting trajectories into smaller subsets along a relevant biological axis (Figure 2.11, Case B). Such segmentation typically comes at the cost of increasing sampling noise because the number of particle trajectories within in each sub-group is reduced, unless additional particle trajectories from the same system are acquired. Nevertheless, the present approach enables the systematic and automated analysis of information-rich particle trajectory datasets and can be applied to high-throughput screens involving cells, embryos, and whole animals by incorporation into automated screening platforms, such as Cell Profiler [Carpenter et al., 2006] and Cell Cognition [Held et al., 2010], or in-house analysis programs via download from http://msd-bayes.org.

Chapter 3

Hidden Markov model analysis of single particle trajectories

3.1 Overview

Many particle trajectory analysis methods, including the mean-square displacement (MSD) approach described in Chapter 2, involve metrics that are time averaged along an entire trajectory or a large portion of a trajectory. These approaches utilize multiple observations of a metric such as squared displacement to increase statistical power and improve estimation of its mean value; however, they sacrifice time resolution in order to perform such averaging. Rapid motion changes and short-lived phases of motion within a single trajectory are blurred or lost in the averaging process, as discussed in Section 1.3.6. Achieving single-step time resolution of transient motion phases is an important challenge in particle trajectory analysis. A promising recent approach has been the application of hidden Markov models (HMMs) to particle trajectories [Das et al., 2009; Cairo et al., 2010; Chung et al., 2010; Persson et al., 2013], in which different phases of motion can be modeled as different hidden states of the system, and switching between those hidden states gives rise to the time series of observed particle displacements. An important advantage of an HMM approach is that it can detect switching at single-step time resolution while still utilizing multiple observations of the displacement to perform inference within each motion state.

The four recent studies that have applied HMMs to detect switching within particle trajectories all assume that the motion model at each step of the trajectory is simple diffusion and look for switching between different values of the diffusion coefficient D [Das et al., 2009; Cairo et al., 2010; Chung et al., 2010; Persson et al., 2013]. Here we extend this HMM framework for particle trajectory analysis to infer switching between modes of motion—simple diffusion and diffusion plus directed motion—within a particle trajectory, in addition to switching between different diffusion coefficient or velocity values. Our extended method, HMM-Bayes, performs three levels of inference. First, the number of motion states—and, for each state, whether or not it includes a directed motion parameter—is inferred from the trajectory (Section 3.3), using Bayesian inference to penalize parameters and avoid overfitting as in Chapter 2. Second, the values of the parameters D and, if directed motion is present, v for each of the detected motion states are inferred using a Markov chain Monte Carlo approach to sample parameter space (Section 3.3.3). Third, the most likely sequence of hidden states along the trajectory is inferred using the Viterbi algorithm (Appendix A.2.1) to determine the time(s) at which the particle experiences a change in motion.

3.2 Modeling a particle trajectory with an HMM

3.2.1 Particle displacements depend on hidden motion parameters

As described in Section 1.3, a one-dimensional particle trajectory consists of a sequence of particle positions x_t separated by a time interval Δt. For a particle undergoing a simple random walk, the particle displacements $\Delta x_t = x_{t+1} - x_t$ are distributed according to a normal distribution with zero mean and a standard deviation that depends on the diffusion

coefficient D according to $\sigma = \sqrt{2D\Delta t}$ [Saxton, 1993],

$$\Delta x \sim \mathcal{N}(0, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\Delta x^2}{2\sigma^2}\right) = \frac{1}{\sqrt{4\pi D\Delta t}} \exp\left(-\frac{\Delta x^2}{4D\Delta t}\right). \tag{3.1}$$

The sequence of particle displacements over time can be thought of as a noisy signal centered at zero. If the particle undergoes a rapid change in diffusion coefficient, then the standard deviation of the noisy signal will change accordingly (Figure 3.1). In a biological system, such a change in diffusion coefficient could be caused by ligand binding or by a change in viscosity of the surrounding medium. If the particle experiences directed motion (drift or flow) in addition to random motion, the displacement distribution is still normal but now has a non-zero mean that depends on the velocity v_x according to $\mu = v_x\Delta t$,

$$\Delta x \sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\Delta x - \mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{4\pi D\Delta t}} \exp\left(-\frac{(\Delta x - v_x\Delta t)^2}{4D\Delta t}\right). \tag{3.2}$$

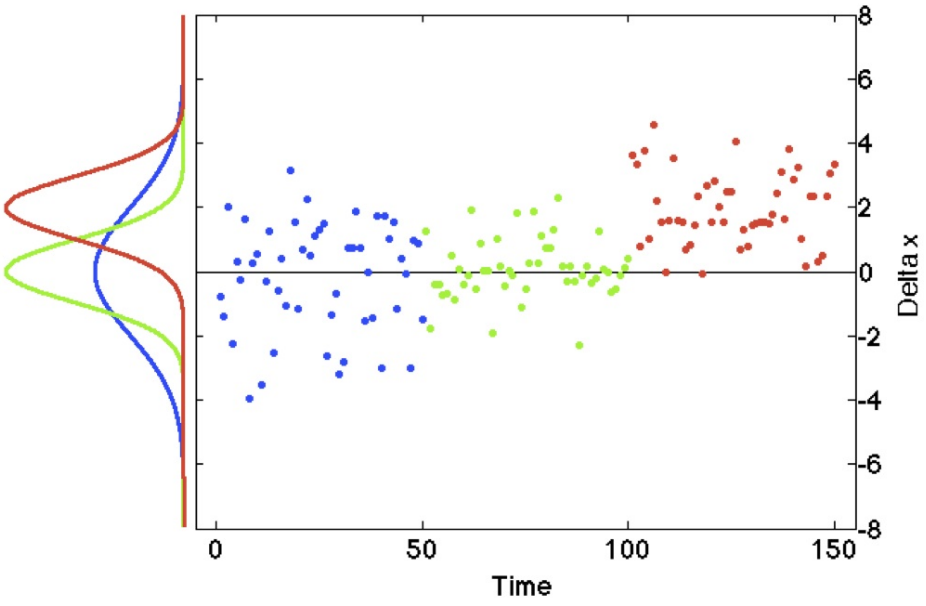

Figure 3.1: Simulated 1D displacements drawn from normal distributions with $\mu = 0$ and $\sigma = 2$ (blue), $\mu = 0$ and $\sigma = 1$ (green), and $\mu = 2$ and $\sigma = 1$ (red).

If the particle experiences a sudden onset of directed motion, the mean of the noisy displacements will shift (Figure 3.1). In a biological system, such a change in velocity might be caused by binding to cytoskeletal motor proteins or to an onset of cytoplasmic streaming.

Here we sought a method that would infer both the presence and timing of such transitions in velocity or diffusion coefficient from a sequence of particle displacements at single-step time resolution. Because we observe a time series of the displacement random variable Δx_t and we know the probability distributions relating this random variable to the motion parameters D and v_x, this problem lends itself to modeling in the framework of a hidden Markov model (HMM) in which the values of these two motion parameters are hidden states of the system.

3.2.2 Formulation of the particle trajectory HMM in one dimension

The discussion below assumes a basic knowledge of the structure and terminology of hidden Markov models, as described in Appendix A. Let the hidden states $\{S_i\}_{i=1}^{K}$ that are explored by the particle each be described by both a diffusion coefficient value D_i and a velocity value $v_{x,i}$,

$$S_i = [D_i, v_{x,i}] \; . \tag{3.3}$$

At every time point t, the particle is in one of these hidden states, but the specific state is unknown. Thus the sequence of hidden states that the particle passes through over time can be written as $\mathbf{s} = \{s_t\}_{t=1}^{T}$, where $s_t \in \{S_i\}_{i=1}^{K}$. The emissions $\mathbf{e} = \{e_t\}_{t=1}^{T}$ of the particle trajectory HMM are the observed time series of particle displacements,

$$e_t = \Delta x_t = x_{t+1} - x_t \; , \tag{3.4}$$

and each e_t is derived from the corresponding state of the particle s_t. A schematic of the HMM structure is shown in Figure 3.2.

With these definitions, the emission distribution $p(e)$ of the HMM is the normal distribution defined in Equation 3.2 above, parameterized by $\boldsymbol{\beta}_i = [D_i, v_{x,i}]$ for each state S_i,

$$p_i(e) = \frac{1}{\sqrt{4\pi D_i \Delta t}} \exp\left(-\frac{(e - v_{x,i}\Delta t)^2}{4D_i \Delta t}\right). \tag{3.5}$$

The starting probabilities π_i for the K states and the transition probabilities ϕ_{ij} between all pairs of states are scalar probabilities defined as in a standard HMM (Appendix A). Therefore, the full set of parameters $\boldsymbol{\theta}$ for a one-dimensional particle trajectory HMM with K states is,

$$\boldsymbol{\theta} = \left[\{\pi_i\}_{i=1}^K, \{\phi_{ij}\}_{i,j=1}^K, \{D_i, v_{x,i}\}_{i=1}^K\right]. \tag{3.6}$$

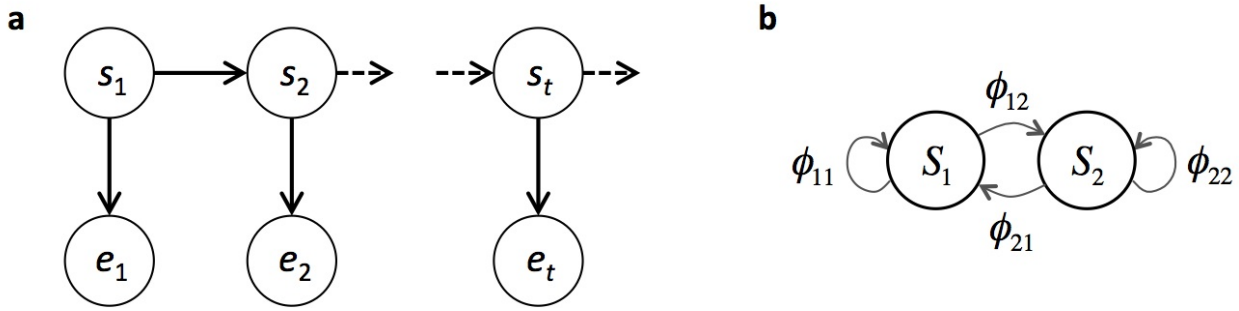

Figure 3.2: (a) Representation of an HMM as a probabilistic graphical model or Bayesian network in which circles indicate random variables and arrows indicate dependencies. Each emission e_t of an HMM depends only on the current state s_t (more formally, it is conditionally independent of all other variables given s_t), and each state s_t depends only on the previous state s_{t-1}. **(b)** State transition diagram for a 2-state HMM with states S_1 and S_2. The transition probability of going from state S_i to state S_j at any given time t is ϕ_{ij}.

3.2.3 Extension to two and three dimensions

For a two-dimensional particle trajectory with particle positions $\mathbf{r}_t = [x_t, y_t]$, the emissions become vectors,

$$\mathbf{e}_t = \Delta\mathbf{r}_t = [x_{t+1} - x_t, y_{t+1} - y_t] \,. \tag{3.7}$$

In addition, the velocity or drift of the particle is now a vector $\mathbf{v} = [v_x, v_y]$. We assume isotropic diffusion, such that the diffusion coefficient remains a single scalar value rather than a full two-dimensional covariance matrix. Under these conditions, the emission distribution of the particle trajectory HMM is the bivariate normal distribution,

$$p_i(\mathbf{e}) = \frac{1}{4\pi D_i \Delta t} \exp\left(-\frac{(e_x - v_{x,i}\Delta t)^2 + (e_y - v_{y,i}\Delta t)^2}{4D_i\Delta t}\right) . \tag{3.8}$$

Similarly, in three dimensions we have,

$$\mathbf{e}_t = \Delta\mathbf{r}_t = [x_{t+1} - x_t, y_{t+1} - y_t, z_{t+1} - z_t] \,, \tag{3.9}$$

and,

$$p_i(\mathbf{e}) = \frac{1}{(4\pi D_i \Delta t)^{3/2}} \exp\left(-\frac{(e_x - v_{x,i}\Delta t)^2 + (e_y - v_{y,i}\Delta t)^2 + (e_z - v_{z,i}\Delta t)^2}{4D_i\Delta t}\right) . \tag{3.10}$$

The full set of parameters for the two- and three-dimensional particle trajectory HMMs is,

$$\boldsymbol{\theta} = \left[\{\pi_i\}_{i=1}^K, \{\phi_{ij}\}_{i,j=1}^K, \{D_i, \mathbf{v}_i\}_{i=1}^K\right] . \tag{3.11}$$

Given knowledge of the number of states K and these parameters $\boldsymbol{\theta}$, it is straightforward to solve for the most likely sequence of hidden states that explains a given series of particle displacements using the Viterbi algorithm described in Appendix A.2.1.

3.3 Application of Bayesian inference to particle trajectory HMMs

3.3.1 Competing models

When analyzing a particle trajectory, we do not know either the number of motion states that the particle explores during the time it is observed or the motion parameters associated with those states. Any individual motion state may or may not include directed motion in addition to random diffusive motion. Therefore, there are two one-state models to be tested, one with a single parameter D_1 (model D) and one with both D_1 and a non-zero $\mathbf{v}_1$ parameter (model DV), where the number of independent components of $\mathbf{v}_1$ depends on the number of dimensions of the trajectory. Similarly, there are three two-state models to be tested, one in which both states have just D_i parameters (model D-D), one in which both states also have nonzero $\mathbf{v}_i$ parameters (model DV-DV), and one in which only one of the two states has a non-zero $\mathbf{v}_i$ (model D-DV). There are always $K+1$ of these possible models for each number of states K. Unlike in the case of MSD-Bayes in Chapter 2, here there is significant variation in the number of parameters associated with the competing models, as shown in Table 3.1. This variation has important consequences for model selection that will be discussed in more detail below.

3.3.2 Bayesian HMM framework

As in Chapter 2, we use a Bayesian inference approach to evaluate the relative probabilities of the competing models for the particle trajectory HMM, as shown in Table 3.1, up to some specified maximum number of states $K_{\max}$. Since there are $K+1$ models with different numbers of nonzero $\mathbf{v}_i$ parameters for each value of K from 1 to $K_{\max}$, the total number of models to test is $\kappa \equiv (K_{\max}^2 + 3K_{\max})/2$. Call this set of models $\{M_k\}_{k=1}^{\kappa}$.

Given an observed sequence of emissions $\mathbf{e}$, the likelihood of a model M_k can be expanded according to Bayes' rule as follows,

$$P(M_k|\mathbf{e}) = \frac{P(\mathbf{e}|M_k)P(M_k)}{P(\mathbf{e})} \propto P(\mathbf{e}|M_k) \;, \tag{3.12}$$

where the final proportionality holds if the prior probabilities of the models $P(M_k)$ are all equal. Thus, we only need to compute $P(\mathbf{e}|M_k)$, the likelihood of the data given a particular

Table 3.1: Comparison of the tested motion models with up to three states. Each state can either have a zero or nonzero velocity (D and DV states, respectively). The total number of independent parameters in each model is shown in the right-most column; note that the number of velocity parameters depends on the number of dimensions in which the trajectory was observed, since velocity is a vector quantity.

Model	# of states			# of independent parameters				# of	TOTAL
	total (K)	D (V=0)	DV	π	ϕ	D	V	dimensions	PARAMETERS
D	1	1	0	0	0	1	0	1	1
							0	2	1
							0	3	1
DV	1	0	1	0	0	1	1	1	2
							2	2	3
							3	3	4
D-D	2	2	0	1	2	2	0	1	5
							0	2	5
							0	3	5
D-DV	2	1	1	1	2	2	1	1	6
							2	2	7
							3	3	8
DV-DV	2	0	2	1	2	2	2	1	7
							4	2	9
							6	3	11
D-D-D	3	3	0	2	6	3	0	1	11
							0	2	11
							0	3	11
D-D-DV	3	2	1	2	6	3	1	1	12
							2	2	13
							3	3	14
D-DV-DV	3	1	2	2	6	3	2	1	13
							4	2	15
							6	3	17
DV-DV-DV	3	0	3	2	6	3	3	1	14
							6	2	17
							9	3	20

model. As in Chapter 2, the total likelihood of the data is found by marginalizing over the unknown parameter values $\boldsymbol{\theta}_k$ for that model as follows,

$$P(\mathbf{e}|M_k) = \int P(\mathbf{e}|\boldsymbol{\theta}_k, M_k) P(\boldsymbol{\theta}_k|M_k) d\boldsymbol{\theta}_k \ , \tag{3.13}$$

where $P(\boldsymbol{\theta}_k|M_k)$ is the prior probability of a particular realization of the parameter values for the model M_k.

In the case of a hidden Markov model, however, the model parameters $\boldsymbol{\theta}_k$ are not the only unknown variables of the system. The possible hidden states $\{S_i\}_{i=1}^K$ are defined by $\boldsymbol{\theta}_k$, but the temporal sequence of explored hidden states $\mathbf{s} = \{s_t\}_{t=1}^T$ is unknown. The likelihood of the observed emissions given the parameters, $P(\mathbf{e}|\boldsymbol{\theta}_k, M_k)$, must be marginalized over all possible hidden state sequences $\mathbf{s}_k$ for each model M_k,

$$P(\mathbf{e}|\boldsymbol{\theta}_k, M_k) = \sum_{\mathbf{s}_k} P(\mathbf{e}|\mathbf{s}_k, \boldsymbol{\theta}_k, M_k) P(\mathbf{s}_k|\boldsymbol{\theta}_k, M_k) \ , \tag{3.14}$$

so the marginal likelihood in Equation 3.13 becomes,

$$P(\mathbf{e}|M_k) = \int \left[\sum_{\mathbf{s}_k} P(\mathbf{e}|\mathbf{s}_k, \boldsymbol{\theta}_k, M_k) P(\mathbf{s}_k|\boldsymbol{\theta}_k, M_k) \right] P(\boldsymbol{\theta}_k|M_k) d\boldsymbol{\theta}_k \ . \tag{3.15}$$

As shown in Equations A.7 and A.8 in Appendix A.2.1, the term $P(\mathbf{e}|\mathbf{s}_k, \boldsymbol{\theta}_k, M_k)$ depends only on the emission probabilities $p_i(\mathbf{e})$, which are parameterized by the D_i and $\mathbf{v}_i$ values for each of the states in model M_k, and the term $P(\mathbf{s}_k|\boldsymbol{\theta}_k, M_k)$ depends only on the starting and transition probabilities π_i and ϕ_{ij}. Plugging in these equations (Equations A.7 and A.8) yields the full form of the marginal likelihood in terms of the model parameters,

$$P(\mathbf{e}|M_k) = \int \left[\sum_{\mathbf{s}_k} \left(\pi_{s_1} \prod_{t=2}^{T} \phi_{s_{t-1}, s_t} \prod_{t=1}^{T} p_{s_t}(\mathbf{e}_t) \right) \right] P(\boldsymbol{\theta}_k|M_k) d\boldsymbol{\theta}_k \ , \tag{3.16}$$

where $p_{s_t}(\mathbf{e}_t)$ is given in Equations 3.5, 3.8 and 3.10 above depending on the number of dimensions of the observed particle trajectory. For a given model M_k and its associated parameters $\boldsymbol{\theta}_k$, the summation over hidden state sequences $\sum_{\mathbf{s}_k} \left(\pi_{s_1} \prod_{t=2}^{T} \phi_{s_{t-1},s_t} \prod_{t=1}^{T} p_{s_t}(\mathbf{e}_t) \right)$ can be performed exactly using the forward algorithm [Ewens and Grant, 2005], which is described in detail in Appendix A.2.2.

3.3.3 Markov Chain Monte Carlo (MCMC) to sample parameter space

Although the integration in Equation 3.16 is intractable in general, the fact that the value of the integrand can be evaluated exactly at any given value of the variables of integration $\boldsymbol{\theta}_k$ lends itself to a Monte Carlo integration strategy, in which values of $\boldsymbol{\theta}_k$ are drawn at random from some sampling distribution $q(\boldsymbol{\theta}_k)$ [Robert and Casella, 2004]. As discussed in Appendix B.1, the best distribution $q(\boldsymbol{\theta}_k)$ for integrating an integrand $f(\boldsymbol{\theta}_k)$ is a probability distribution that is proportional to $f(\boldsymbol{\theta}_k)$ [Robert and Casella, 2004]. This technique is called importance sampling, because such a $q(\boldsymbol{\theta}_k)$ ensures that values of $\boldsymbol{\theta}_k$ are more likely to be sampled where the value of $f(\boldsymbol{\theta}_k)$ is large, corresponding to 'important' regions of $f(\boldsymbol{\theta}_k)$.

For the marginal likelihood integration in Equation 3.16, the integrand has two terms: the summation over hidden state sequences $\mathbf{s}_k$ and the parameter prior distribution $P(\boldsymbol{\theta}_k|M_k)$. As discussed in more detail in Section 3.3.5 below, we take the parameter prior to be a uniform distribution, which means it can be taken outside the integration in Equation 3.16. Therefore, we need only to evaluate the value of the integral,

$$I = \int \left[\sum_{\mathbf{s}_k} \left(\pi_{s_1} \prod_{t=2}^{T} \phi_{s_{t-1},s_t} \prod_{t=1}^{T} p_{s_t}(\mathbf{e}_t) \right) \right] d\boldsymbol{\theta}_k \ . \tag{3.17}$$

Renaming the integrand as $f(\boldsymbol{\theta}_k)$, such that,

$$f(\boldsymbol{\theta}_k) = \sum_{\mathbf{s}_k} \left(\pi_{s_1} \prod_{t=2}^{T} \phi_{s_{t-1},s_t} \prod_{t=1}^{T} p_{s_t}(\mathbf{e}_t) \right), \tag{3.18}$$

we seek a sampling distribution $q(\boldsymbol{\theta}_k)$ that is proportional to this $f(\boldsymbol{\theta}_k)$. Although we cannot sample directly from such a $q(\boldsymbol{\theta}_k)$, we can use a stochastic Markov chain Monte Carlo (MCMC) sampling approach [Robert and Casella, 2004; Grimmett and Stirzaker, 2001] to converge on a distribution of sampled $\boldsymbol{\theta}_k$ values that approaches the desired $q(\boldsymbol{\theta}_k)$.

Here we use the Metropolis MCMC algorithm [Gilks, 1995; Robert and Casella, 2004; Grimmett and Stirzaker, 2001] described in detail in Appendix B.2. Briefly, the value of $f(\boldsymbol{\theta}_k)$ is evaluated at a random starting point $\boldsymbol{\theta}_k^{(0)}$, then a new value $\boldsymbol{\theta}_k^{(1)}$ is proposed from a multivariate normal distribution in parameter space centered at $\boldsymbol{\theta}_k^{(0)}$, and the new value $\boldsymbol{\theta}_k^{(1)}$ is accepted with probability $\min\left(1, f(\boldsymbol{\theta}_k^{(1)})/f(\boldsymbol{\theta}_k^{(0)})\right)$ and rejected otherwise. By iteratively repeating this process, it can be shown that the distribution of sampled parameter values converges to a distribution $q(\boldsymbol{\theta}_k)$ that is proportional to $f(\boldsymbol{\theta}_k)$. See Appendix B.2 for details. This Metropolis MCMC approach is particularly suited to sampling from multidimensional distributions, as is the case for $\boldsymbol{\theta}_k$ (see Table 3.1 for the number of independent parameters in each model). The basic Metropolis MCMC algorithm described in Appendix B.2 has a number of features that must be tailored to the specific context in which it is applied, such as the generation of the initial guess $\boldsymbol{\theta}_k^{(0)}$ in parameter space, the width of the proposal distribution for generating new points in parameter space, and the selection of which parameter(s) in $\boldsymbol{\theta}_k$ will be updated at each step of the algorithm.

The full set of parameters to be sampled for a particle trajectory HMM is $\boldsymbol{\theta} = \left[\{\pi_i\}_{i=1}^{K}, \{\phi_{ij}\}_{i,j=1}^{K}, \{D_i, \mathbf{v}_i\}_{i=1}^{K}\right]$, as given in Equation 3.11, for a total number of states K assuming that all states have nonzero velocities. Since we also consider models with states

that have zero velocity, as described above in Section 3.3.1 and shown in Table 3.1, let K be the total number of states and K_{V} be the number of states with nonzero velocity for any given tested model M_k. Then the full set of parameters for the model M_k is,

$$\boldsymbol{\theta} = \left[\{\pi_i\}_{i=1}^{K}, \{\phi_{ij}\}_{i,j=1}^{K}, \{D_i\}_{i=1}^{K}, \{\mathbf{v}_{i'}\}_{i'=1}^{K_{\mathrm{V}}} \right] . \tag{3.19}$$

At each step in the MCMC iteration, one or more parameter(s) in this set could be selected for updating. We compared three selection methods—updating all parameters at once, updating only a single randomly-selected parameter, or updating a randomly-selected block of related parameters at each MCMC iteration—and found that the block method [Gilks, 1995] had the fastest and most robust convergence. The motivation for the block method is that parameters with correlated effects on the likelihood function $f(\boldsymbol{\theta}_k)$ should be updated at the same time to increase the chance of escaping from local maxima in the likelihood landscape [Gilks, 1995]. Specifically, we split the parameters into 3 blocks: the probability parameters $\left\{ \{\pi_i\}_{i=1}^{K}, \{\phi_{ij}\}_{i,j=1}^{K} \right\}$, the diffusion coefficients $\{D_i\}_{i=1}^{K}$, and the velocities $\{\mathbf{v}_{i'}\}_{i'=1}^{K_{\mathrm{V}}}$. At every iteration, one of these 3 blocks is selected randomly with equal probability, and updates are proposed for all of the parameters within that block. The new value of $f(\boldsymbol{\theta}_k)$ at the updated parameter values is computed using the forward algorithm, and if the update is accepted according to the Metropolis criterion, then the updates for all of the parameters within the block are retained and become the new starting point for the next MCMC iteration.

The velocity parameters $\{\mathbf{v}_{i'}\}_{i'=1}^{K_{\mathrm{V}}} = \{v_{x,i'}, v_{y,i'}, v_{z,i'}\}_{i'=1}^{K_{\mathrm{V}}}$ are theoretically unconstrained, but in practice the most likely values of $v_{x,i'}$, $v_{y,i'}$, or $v_{y,i'}$ for any state i' will not be greater than the maximum single-step velocity or less than the minimum single-step velocity observed in each of these dimensions over the full trajectory. Therefore, the initial guesses for the velocity parameters $v_{x,i'}$, $v_{y,i'}$, and $v_{y,i'}$ for each state i' are drawn from uniform distributions on this range. For example, an initial guess for the parameter $v_{x,i'}$ is drawn

from,

$$v_{x,i'}^{(0)} \sim \text{unif}\Big(\min_t \Big(\frac{e_{x,t}}{\Delta t}\Big), \max_t \Big(\frac{e_{x,t}}{\Delta t}\Big)\Big) \ . \tag{3.20}$$

Since the velocities are related to the means of the emission probability distributions $p_{i'}(\mathbf{e})$ by $\mu_{x,i'} = v_{x,i'}\Delta t$ as described in Section 3.2.1, this condition on the initial values of the velocities can be written as an equivalent condition on the means,

$$\mu_{x,i'}^{(0)} \sim \text{unif}\Big(\min_t \left(e_{x,t}\right), \max_t \left(e_{x,t}\right)\Big) \ . \tag{3.21}$$

Figure 3.3 illustrates an example of this range for a simple dataset of one-dimensional particle displacements. At every MCMC iteration for which the velocity parameter block is selected, new values for each $\mu_{x,i'}$, $\mu_{y,i'}$, and $\mu_{y,i'}$ are proposed from normal distributions centered at the previous values of these parameters and with standard deviation δ_μ. The values of these parameters are not constrained during the MCMC iterations and are thus able to move outside of the range imposed on the initial guesses above. The same δ_μ is used for each of these parameters, and its value is initialized and updated as described below.

The diffusion coefficients $\{D_i\}_{i=1}^K$ are theoretically constrained only to be greater than 0, but in practice the most likely values of the standard deviations σ_i of the emission probability distributions $p_i(\mathbf{e})$ will not be greater than the maximum width of the observed single-step emission distribution in each dimension, as illustrated for a single dimension in Figure 3.3. Since the diffusion coefficients are related to the emission standard deviations by $\sigma_i = \sqrt{2D_i\Delta t}$ as in Section 3.2.1, we can draw initial guesses for the parameters σ_i from the uniform distribution,

$$\sigma_i^{(0)} \sim \text{unif}\Big(0, \max_\xi \Big(\max_t \left(e_{\xi,t}\right) - \min_t \left(e_{\xi,t}\right)\Big)\Big) \ , \tag{3.22}$$

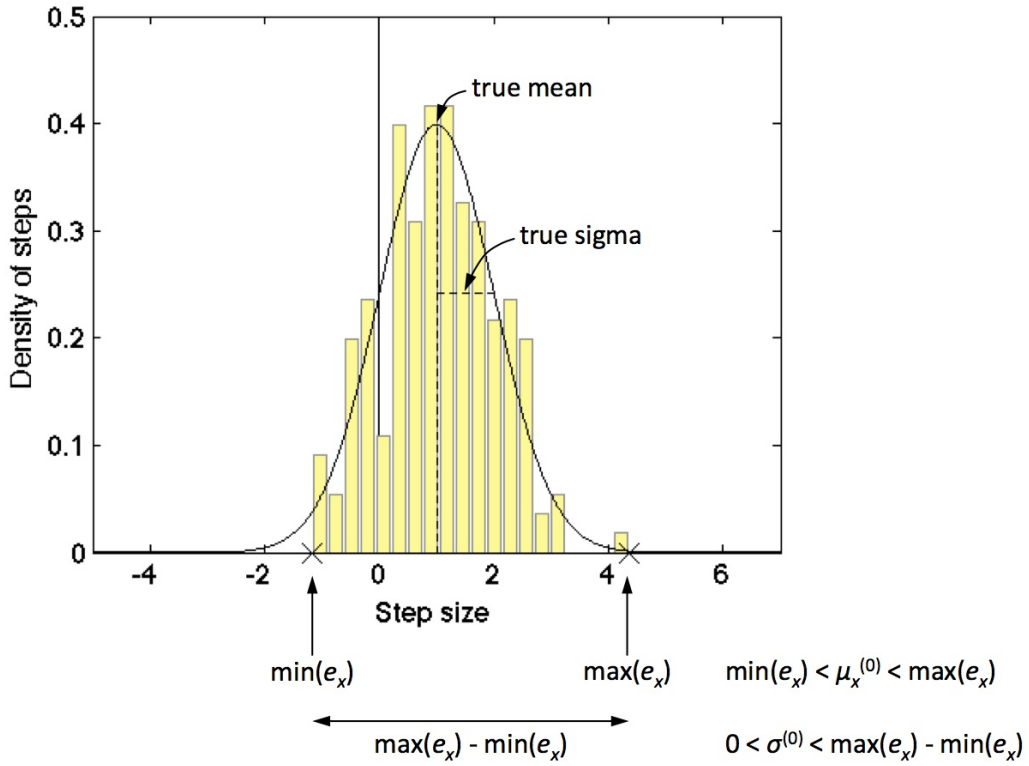

Figure 3.3: Initial guesses for the mean and sigma parameters at the start of each MCMC iteration are drawn from uniform distributions, whose ranges are set empirically based on the observed displacements (steps) in the particle trajectory. The initial mean in each dimension is no larger or smaller than the largest or smallest observed displacement in that dimension, respectively, and the initial sigma is no larger than the largest difference between the largest and smallest observed displacements observed in each dimension.

where ξ parameterizes the spatial dimensions of the particle trajectory; e.g., $\xi \in x, y, z$ for a three-dimensional trajectory. At every MCMC iteration for which the diffusion coefficient parameter block is selected, new values for each σ_i are proposed from normal distributions centered at the previous values of these parameters and with standard deviation δ_σ. If the proposed new value for any of the σ_i falls below 0, then the parameter update at this MCMC iteration is automatically rejected. However, the σ_i parameters are not otherwise constrained during the MCMC iterations and are thus able to grow larger than the maximum value imposed on the initial guesses above. The same δ_σ is used for each σ_i parameter, and its value is initialized and updated as described below.

The probability parameters $\left\{\{\pi_i\}_{i=1}^K, \{\phi_{ij}\}_{i,j=1}^K\right\}$ are all constrained to fall in the range $[0, 1]$, and must also satisfy the conditions $\sum_{i=1}^K \pi_i = 1$ for the starting probabilities and $\sum_{j=1}^K \phi_{ij} = 1$ for all i for the transition probabilities. Initial guesses for these probabilities are all set to the same value $1/K$, representing equal distributions over the number of states K. Specifically, $\pi_i^{(0)} = 1/K$ for all i and $\phi_{ij}^{(0)} = 1/K$ for all i and j. We found that these parameters fully explore their $[0, 1]$ range during the MCMC iterations even with these fixed initial conditions, and thus that they do not need to be randomly initialized. At every MCMC iteration for which the probability parameter block is selected, new values for each π_i and ϕ_{ij} are proposed from normal distributions centered at the previous values of these parameters and with standard deviations δ_π and δ_ϕ, respectively. If the proposed new value for any of these probabilities falls outside the range $[0, 1]$, then the parameter update at this MCMC iteration is automatically rejected. Once new values have been proposed for all of the probabilities $\left\{\{\pi_i\}_{i=1}^K, \{\phi_{ij}\}_{i,j=1}^K\right\}$, then the new vectors $\boldsymbol{\pi}$ and $\{\boldsymbol{\phi}_i\}_{i=1}^K$ are individually normalized to ensure that their component probabilities satisfy the above conditions $\sum_{i=1}^K \pi_i = 1$ and $\sum_{j=1}^K \phi_{ij} = 1$.

To ensure that the space of possible parameter values is fully explored and to minimize the chance that the MCMC chain becomes stuck in local maxima rather than the global maximum, multiple re-starts of the algorithm are run with different initial guesses for the velocity and diffusion coefficient parameters, randomly generated from the distributions described above. Figure 3.4 shows the convergence of the likelihood $f(\boldsymbol{\theta}_k) = P(\mathbf{e}|\boldsymbol{\theta}_k, M_k)$ and selected parameter values for 10 different initial guesses, where the observed emissions are from a simulated two-dimensional trajectory that explores just a single state with nonzero velocity (model DV in Table 3.1). Figure 3.5 shows similar convergence properties for a 2D trajectory that switches between two states (model D-DV in Table 3.1). Note that the MCMC chains for the two-state trajectory are more likely to be caught in local maxima than the one-state trajectory, due to the larger number of free parameters (7 vs. 3). However, the

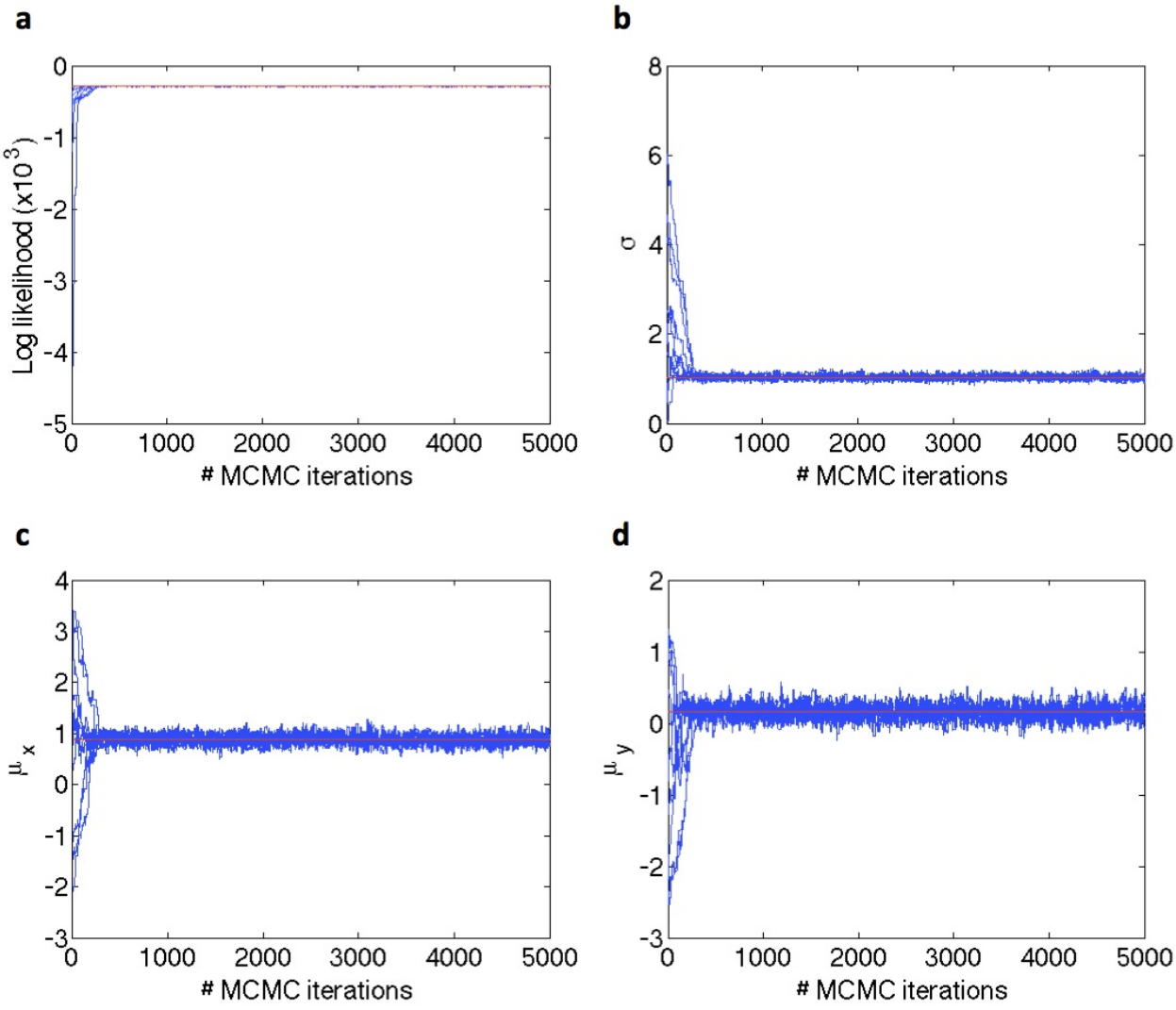

Figure 3.4: Convergence of the log likelihood **(a)** and the three parameter values σ **(b)**, μ_x **(c)**, and μ_y **(d)** of the one-state D model (see Table 3.1) for 10 re-starts of the MCMC iterations starting from different randomly-initialized values of these parameters. The fit trajectory was simulated with 100 displacements drawn from a normal distribution with mean (1,0) and a sigma of 1. The red lines indicate empirical values of the parameters over those 100 displacements. The MCMC runs all rapidly converge to the empirically-observed parameter values.

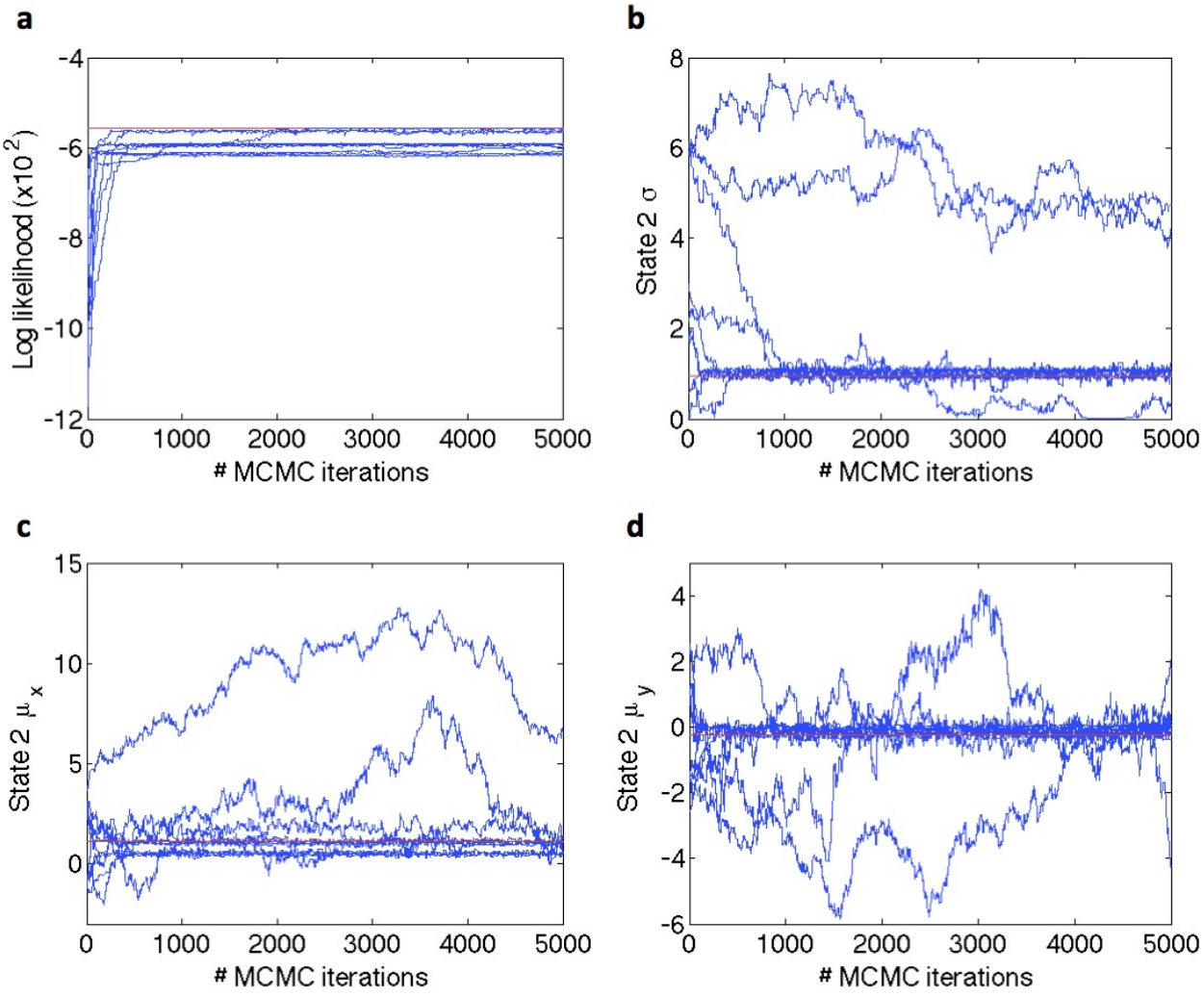

Figure 3.5: Convergence of the log likelihood **(a)** and three parameter values σ_2 **(b)**, $\mu_{x,2}$ **(c)**, and $\mu_{y,2}$ **(d)** of the DV state of a two-state D-DV model (see Table 3.1) for 10 re-starts of the MCMC iterations starting from different randomly-initialized values of the parameters. The fit trajectory was simulated with 200 displacements, switching from a normal distribution with mean (0,0) and a sigma of 1 to a normal distribution with mean (1,0) and a sigma of 1 after the first 100 displacements. The red lines indicate empirical values of the parameters over the 100 displacements generated from the second state. Note that some of the MCMC runs converge to the empirically-observed parameter values, while others are trapped in local minima in which some parameters have minimal effect on the likelihood and thus execute a random walk.

10 initialization restarts are sufficient for some of the chains to find the global maximum. In practice, we use 100 initialization restarts for models with more than a single state to ensure that the global maximum is found.

During the multiple MCMC initialization runs, the parameters δ_μ, δ_σ, δ_π, and δ_ϕ governing the width of the proposal distributions for new values of the various parameters in $\boldsymbol{\theta}_k$ are initialized and updated to maintain a target acceptance rate for the proposed parameter updates. The acceptance rate is defined separately for each block of parameters—probabilities, diffusion coefficients, and velocities—and is equal to the fraction of proposed updates that are accepted (according to the Metropolis criterion described above) relative to the total number of updates that are proposed for that parameter block. Acceptance rates that are too high indicate that the width of the proposal distribution is too small and the full space of parameter values may not be adequately explored. One the other hand, acceptance rates that are too low indicate that the width of the proposal distribution is too large and most proposals are moving far from the region(s) of high likelihood. In general, acceptance rates between 0.3 and 0.5 are considered the best compromise between these two effects, allowing sufficient exploration of the likelihood landscape [Roberts et al., 1997]. Here we initialize the values of δ_μ and δ_σ for the μ and σ parameters to 1/50 of the initial guess ranges given in Equations 3.21 and 3.22 above, and then update δ_μ and δ_σ at the end of each MCMC initialization run based on the acceptance rates calculated for that run. For the probability parameters, because they are constrained between [0,1], the values of δ_π and δ_ϕ should shrink as the probabilities approach either of the boundaries at 0 or 1 in order to maintain a consistent acceptance rate along the MCMC chain. Therefore, δ_π and δ_ϕ are updated at every MCMC iteration to be equal to 1/2 of the minimum distance of any of the π or ϕ probabilities, respectively, from either of the boundaries 0 or 1.

Following the MCMC initialization runs, a longer MCMC run—starting from the maximum likelihood parameters found during the initialization runs—is performed. These maximum likelihood parameters are assumed to have found the global maximum provided that sufficient initialization restarts were performed. Therefore, this longer MCMC run explores parameter values around the global maximum likelihood. Figure 3.6 shows the longer MCMC run for the above two-state D-DV trajectory, starting from the maximum likelihood parameters from the initialization runs in Figure 3.5.

3.3.4 Monte Carlo integration of the likelihood

The MCMC approach above results in a set of samples of $\boldsymbol{\theta}_k$ from a target distribution $q(\boldsymbol{\theta}_k)$ that is proportional to $f(\boldsymbol{\theta}_k)$, as explained in Appendix B.2. The Monte Carlo estimator $\hat{I}_k$ of our desired integral $I_k = \int f(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$ is equal to the mean value of the ratio of $f(\boldsymbol{\theta}_k)$ to $q(\boldsymbol{\theta}_k)$ [Robert and Casella, 2004],

$$\hat{I}_k = \left\langle \frac{f(\boldsymbol{\theta}_k)}{q(\boldsymbol{\theta}_k)} \right\rangle_q , \tag{3.23}$$

where the subscript q indicates that the mean is calculated over values of $\boldsymbol{\theta}_k$ that are sampled from the distribution $q(\boldsymbol{\theta}_k)$, as discussed in Appendix B.1. However, although we have such a sample of $\boldsymbol{\theta}_k$ from the MCMC iterations and also know the value of $f(\boldsymbol{\theta}_k)$ at each of those samples, the value of the probability density $q(\boldsymbol{\theta}_k)$ is unknown. Since $q(\boldsymbol{\theta}_k)$ is a probability distribution (which must integrate to 1) that is proportional to $f(\boldsymbol{\theta}_k)$, direct calculation of $q(\boldsymbol{\theta}_k)$ from $f(\boldsymbol{\theta}_k)$ would require knowledge of the normalization factor $\int f(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$, which is precisely the unknown integral that we are trying to evaluate. Therefore, it is necessary to find a probability distribution that approximates $q(\boldsymbol{\theta}_k)$ but has an analytical form that can be evaluated exactly at each point $\boldsymbol{\theta}_k$.

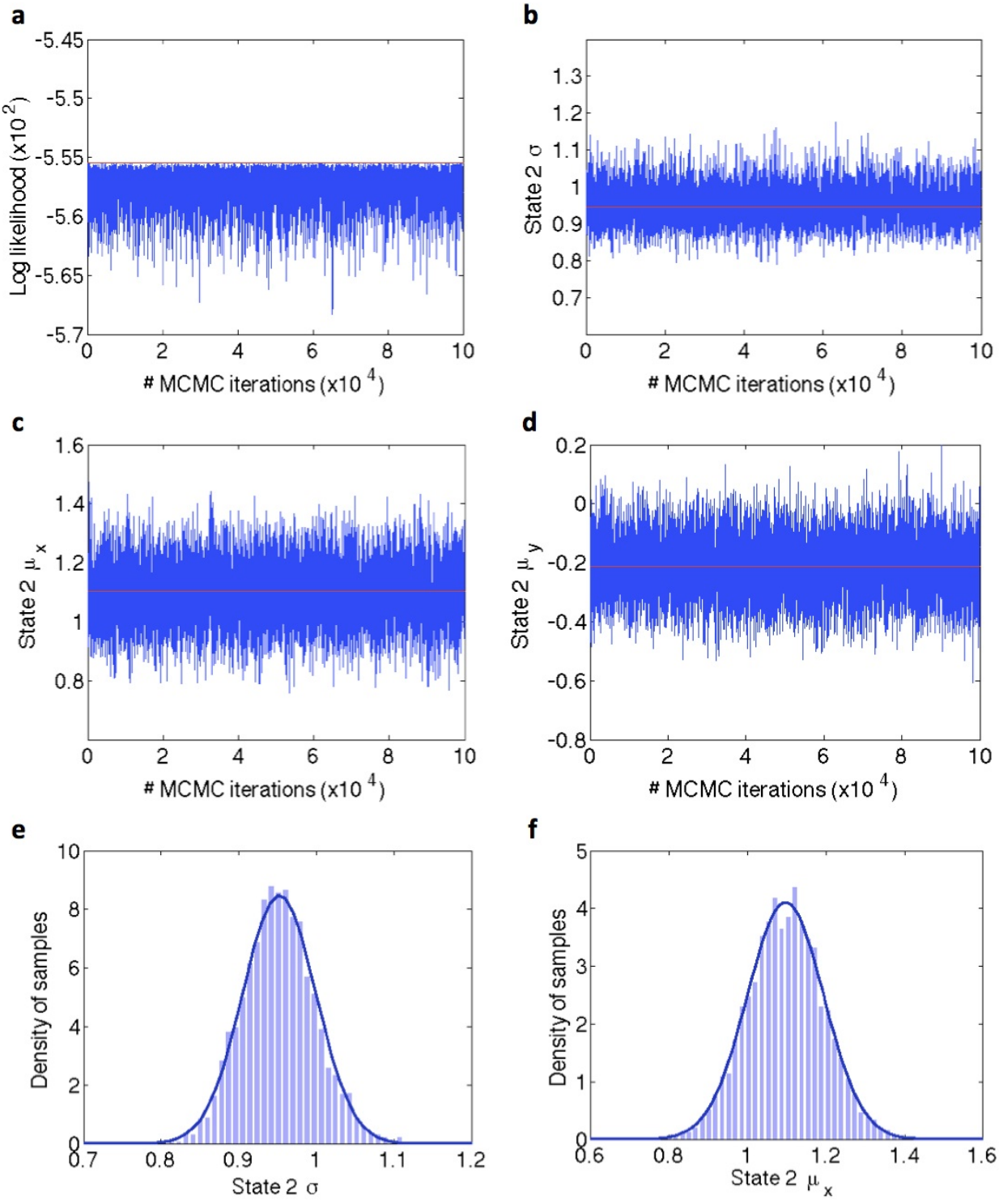

Figure 3.6: **(a)-(d)** Longer MCMC traces initialized from the maximum likelihood parameters found in the MCMC initialization runs in Figure 3.5. **(e)** and **(f)** show histograms (light blue) of the sampled values of σ_2 and $\mu_{x,2}$, respectively, from the traces in (b) and (c) and the Gaussian approximations (dark blue) described in Section 3.3.4

Here we approximate $q(\boldsymbol{\theta}_k)$ by a distribution $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$ that is multivariate normal over the emission mean and standard deviation parameters $\left\{\{\sigma_i\}_{i=1}^{K}, \{\boldsymbol{\mu}_{i'}\}_{i'=1}^{K_{\mathrm{V}}}\right\}$. For each parameter σ_i and $\mu_{\xi,i'}$ for $\xi \in x, y, z$, the mean and standard deviation of the normal approximation $q_{\mathcal{N}}$ along that parameter axis are set to the mean and standard deviation of the sampled values of that parameter during the second half of the long MCMC run, as shown in Figure 3.6e-f. The values of these parameters are sampled independently from these normal distributions during the Monte Carlo integration process. The probability parameters $\left\{\{\pi_i\}_{i=1}^{K}, \{\phi_{ij}\}_{i,j=1}^{K}\right\}$, on the other hand, are not independent of each other due to the conditions $\sum_{i=1}^{K} \pi_i = 1$ and $\sum_{j=1}^{K} \phi_{ij} = 1$. The only valid samples for the vectors $\boldsymbol{\pi}$ and $\{\boldsymbol{\phi}_i\}_{i=1}^{K}$ fall on K-dimensional simplexes, assuming that K is the number of states in the tested model. Here we sample these values uniformly from the simplexes, where the constant probability distribution over the simplex is equal to the inverse of the simplex length or area, given by $\sqrt{K}/(K-1)!$. Thus the approximate distribution $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$ has a multivariate normal marginal over the emission means and standard deviations and a uniform marginal over the starting and transition probability simplexes.

With this definition of $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$, its value can be calculated analytically at any given sampled point $\boldsymbol{\theta}_k$. Accordingly, numerical integration of the likelihood $f(\boldsymbol{\theta}_k) = P(\mathbf{e}|\boldsymbol{\theta}_k, M_k)$ for each model M_k is carried out by drawing samples of $\boldsymbol{\theta}_k$ from this analytical approximate distribution $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$ and calculating the values of $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$ and $f(\boldsymbol{\theta}_k)$ analytically and by the forward algorithm, respectively. The formula for the Monte Carlo estimator $\hat{I}_k$ given above is general for any sampling distribution; thus, an estimate of the integrated likelihood using the samples from $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$ is calculated as,

$$\hat{I}_k = \left\langle \frac{f(\boldsymbol{\theta}_k)}{q_{\mathcal{N}}(\boldsymbol{\theta}_k)} \right\rangle_q . \tag{3.24}$$

Because $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$ is close to the ideal distribution $q(\boldsymbol{\theta}_k)$, the variance of this estimator $\hat{I}_k$ is close to its minimum value for any given number of samples of $\boldsymbol{\theta}_k$. As expected, the variance decreases as the number of sampled points in parameter space increases (Figure 3.7a).

This integration approach results in an estimate of the value of $\int P(\mathbf{e}|\boldsymbol{\theta}_k, M_k) d\boldsymbol{\theta}_k$ for each model M_k. The remaining term in Equation 3.16 is the prior probability $P(\boldsymbol{\theta}_k|M_k)$, which is assumed constant wherever its value is nonzero (as discussed below in Section 3.3.5). Calling the value of this constant C_k for model M_k, the final estimate of the fully marginalized data likelihood $P(\mathbf{e}|M_k)$ is given by,

$$P(\mathbf{e}|M_k) = C_k \hat{I}_k \ . \tag{3.25}$$

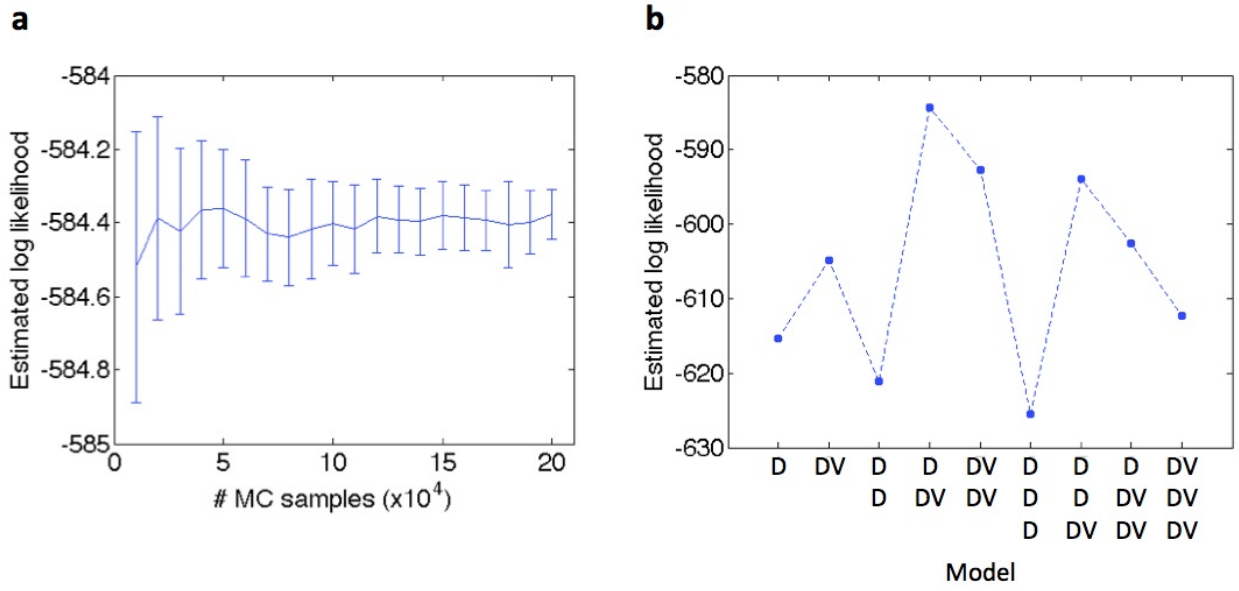

Figure 3.7: **(a)** Mean and variance of the estimator for the integrated log likelihood of the D-DV model as a function of the number of Monte Carlo samples used to obtain the estimate. **(b)** Integrated log likelihoods for the set of tested models with up to three states. Note that the inclusion of at least one nonzero velocity parameter greatly increases the likelihood because the true displacements have a nonzero mean in the second state. Inclusion of additional nonzero velocity parameters reduces the likelihood due to the increased penalty on complexity. Similarly, the 3-state models have lower likelihood than the corresponding 2-state models because they include additional parameters over those required to describe the displacements, which come from two states.

Because the model probabilities $P(M_k|\mathbf{e})$ are proportional to $P(\mathbf{e}|M_k)$ (Equation 3.12), the final model probabilities are calculated by normalizing these likelihoods $P(\mathbf{e}|M_k)$ across the set of tested models. The values of the likelihoods $P(\mathbf{e}|M_k)$ calculated using the full Bayesian HMM approach described above for the two-state D-DV trajectory from Figures 3.5 and 3.6 for a full set of tested models up to $K = 3$ (corresponding to the models in Table 3.1) are shown in Figure 3.7b. As expected, the true D-DV model used to simulate the trajectory is found to have the highest likelihood.

3.3.5 Parameter priors

As in the case of MSD-Bayes (Section 2.3.6), a non-informative parameter prior is imposed by setting $P(\boldsymbol{\theta}_k|M_k)$ equal to a uniform value over a bounded region in parameter space. For the probability parameters, which are inherently bounded on the simplexes described above, the prior is uniform over each simplex. For the emission mean and standard deviation parameters, the uniform range is centered on the mean of the sampled values of that parameter during the MCMC run above, with a width equal to a multiple of the standard deviation of those sampled values (the same mean and standard deviation that are used for the $q_{\mathcal{N}}(\boldsymbol{\theta}_k)$ distribution along that parameter in Section 3.3.4). The multiplication factor for the standard deviation is set to a value of 200 [He et al., 2012], the same as for MSD-Bayes.

3.4 Extensions of the particle trajectory HMM

3.4.1 Pooling multiple trajectories

The discussion above describes the application of a Bayesian framework for hidden Markov model analysis to a single particle trajectory. However, in many cell biological

applications, a large number of short trajectories are obtained from particles that are assumed to undergo the same dynamic processes and have the same motion parameters. If this assumption is valid, pooling the trajectories and analyzing them together should increase the inference power of the Bayesian HMM approach and improve its ability to resolve complex models with many states and parameters. Including multiple independent trajectories in the likelihood calculation is straightforward, as their individual likelihoods can be directly multiplied. For J trajectories, each with a set of observed displacements (emissions) $\mathbf{e}_j$, Bayes' rule in Equation 3.12 is modified to become,

$$P\left(M_k|\{\mathbf{e}_j\}_{j=1}^J\right) = \frac{P\left(\{\mathbf{e}_j\}_{j=1}^J|M_k\right)P(M_k)}{P\left(\{\mathbf{e}_j\}_{j=1}^J\right)} \propto P\left(\{\mathbf{e}_j\}_{j=1}^J|M_k\right) \ , \tag{3.26}$$

and the total likelihood of observing all the trajectories given a common underlying model M_k is,

$$P\left(\{\mathbf{e}_j\}_{j=1}^J|M_k\right) = \prod_{j=1}^{J} P(\mathbf{e}_j|M_k) \ . \tag{3.27}$$

The calculation of each individual $P(\mathbf{e}_j|M_k)$ in Equation 3.26 still follows Equation 3.16 above. Note that the $f(\boldsymbol{\theta}_k)$ discussed in Sections 3.3.3 and 3.3.4 now becomes the combined likelihood function $\prod_j P(\mathbf{e}_j|\boldsymbol{\theta}_k, M_k)$; thus, the parameters $\boldsymbol{\theta}_k$ for each model M_k are explored based on the full set of pooled trajectories rather than for each trajectory individually. The hidden state sequences $\mathbf{s}_j$ for each trajectory, on the other hand, are marginalized out during the model fitting and likelihood integration process, and the most likely state sequence can ultimately be obtained for each trajectory independently using the most likely parameters of the most likely model. Therefore, pooling trajectories requires the assumption that the trajectories all explore the same set of states with the same state parameters, but does not require that the trajectories have the same state sequence over time. In other words, the pooled trajectories can transition between the shared states at different times.

3.4.2 Detection of speed vs. velocity

As mentioned above, pooling trajectories assumes that the explored hidden states have the same parameter values for all trajectories. Recall that the states are parameterized by both a diffusion coefficient and a velocity vector, $S_i = [D_i, \mathbf{v}_i]$. If any of the motion states has a nonzero velocity, then in order for the components of the velocity vector to be the same for all pooled trajectories, the trajectories must all be flowing in the same direction. While this may be the case for some biological processes, such as large-scale cellular rearrangements during tissue morphogenesis or chromosomes transport during cell division, there are other processes in which particles may all have a directed component to their motion but be directed in different directions.

To extend the particle trajectory HMM approach described above to the case of pooled trajectories flowing in different directions, we developed a modified HMM in which the emissions are not the vector displacements of the particle but rather the squared displacement magnitudes. For a three-dimensional trajectory,

$$e_t = |\Delta \mathbf{r}_t|^2 = (x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2 + (z_{t+1} - z_t)^2 \ . \tag{3.28}$$

These emissions are now scalars instead of vectors, and their distribution is no longer a normal distribution as in Equation 3.5 above. Instead, note that the random variable e_t is the sum of squares of normally-distributed random variables $\Delta\xi_t = \xi_{t+1} - \xi_t$ for $\xi \in \{x, y, z\}$, each with mean $\mu_\xi = v_\xi \Delta t$ and standard deviation $\sigma = \sqrt{2D\Delta t}$. Sums of squares of normal random variables with nonzero means follow a noncentral chi-squared distribution, which is defined as follows [Grimmett and Stirzaker, 2001]. When the variables have different means μ_ξ but the same standard deviation σ, then the quantity $e_t = \sum_\xi \Delta\xi^2$ is distributed

according to,

$$e_t \sim \frac{1}{2\sigma^2} \exp\left(-\frac{e_t + \rho}{2\sigma^2}\right) \left(\frac{e_t}{\rho}\right)^{d/4-1/2} I_{d/2-1}\left(\frac{\sqrt{\rho e_t}}{\sigma^2}\right) , \tag{3.29}$$

$$\rho = \sum_{\xi} \mu_{\xi}^2 , \tag{3.30}$$

where d is the total number of summed variables, or in our case the number of dimensions, and $I_{d/2-1}$ is a modified Bessel function of the first kind. The parameter ρ corresponds to the squared magnitude of the mean vector across all the dimensions and can also be written in terms of the velocity magnitude $v = \sqrt{\sum_{\xi} v_{\xi}^2}$ as $\rho = (v\Delta t)^2$.

The new χ^2-HMM uses Equation 3.29 as the emission probability distribution, with emissions defined as squared displacement magnitudes as in Equation 3.28. Note that the emission probability distribution now depends on only two parameters, σ and ρ, that are related to the particle diffusion coefficient and velocity magnitude (or speed), respectively. Unlike for the original HMM above, the number of velocity parameters does not grow with the number of dimensions in which the particle was observed. The advantages of this approach, therefore, are that there is no dependence on d and that the pooled particle trajectories do not need to have the same directional velocity but only the same speed for each DV state that they explore.

3.4.3 Detection of convergent flow

The χ^2-HMM presented above is capable of detecting the presence of nonzero velocity states in a set of particle trajectories even when the direction of the velocity is completely uncorrelated between the different trajectories. However, more samples—in this case, more or longer trajectories—are required to resolve two noncentral chi-squared distributions than

two normal distributions that are based on the same underlying mean parameters. Therefore, it is preferable to take advantage of correlations in the individual-particle velocity directions if such correlations exist. For example, if particles are indeed flowing all in the same direction, then the original HMM would be preferable to the χ^2-HMM for resolving the presence of that directional flow.

Another type of correlated directional motion that can be found in biological systems is convergent flow towards a central point (or sink). Therefore, we developed a third variant of the particle trajectory HMM that both fits the position of such a sink and considers velocity towards or away from the sink by projecting each particle displacement onto a coordinate system defined by the sink position. Specifically, for a two-dimensional trajectory, two additional parameters $[x_0, y_0]$ defining the sink location are introduced into the parameter set for each model M_k. Only a velocity v_0 in the direction of the sink is considered as a nonzero parameter; velocities in the perpendicular direction are assumed to be zero. The full set of parameters for a model M_k with K total states and K_V states with nonzero velocity is thus,

$$\boldsymbol{\theta} = \left[\{\pi_i\}_{i=1}^{K}, \{\phi_{ij}\}_{i,j=1}^{K}, \{D_i\}_{i=1}^{K}, \{v_{0,i'}\}_{i'=1}^{K_\mathrm{V}}, \{x_0, y_0\} \right] . \tag{3.31}$$

During the MCMC iterations, the sink position $[x_0, y_0]$ is updated along with the other parameters, representing its own MCMC block as discussed in Section 3.3.3. After each update, the emissions are re-calculated from each particle trajectory using the trajectory displacements and the new sink position. Each original displacement vector $[\Delta x_t, \Delta y_t] = [x_{t+1} - x_t, y_{t+1} - y_t]$ is projected into components $[a_t, b_t]$ towards and perpendicular to the

sink, respectively. The unit vector $[u_x, u_y]$ defining the direction to the sink is,

$$[u_x, u_y] = \left[\frac{x_0 - x_t}{\sqrt{(x_0 - x_t)^2 + (y_0 - y_t)^2}}, \frac{y_0 - y_t}{\sqrt{(x_0 - x_t)^2 + (y_0 - y_t)^2}} \right] . \tag{3.32}$$

Then the component a_t towards the sink is equal to the dot product of the displacement with this unit vector,

$$a_t = [\Delta x_t, \Delta y_t] \cdot [u_x, u_y] , \tag{3.33}$$

and the component b_t perpendicular to the sink is equal to the dot product of the displacement with the perpendicular unit vector,

$$b_t = [\Delta x_t, \Delta y_t] \cdot [-u_y, u_x] . \tag{3.34}$$

All of the displacements from all of the particle trajectories are projected into these directions, and the projected displacements are the emissions of the new HMM. The emission distribution in this new coordinate system is still multivariate normal, with nonzero velocity v_0 only along the first component in the direction of the sink. For a state S_i with parameters $[D_i, v_{0,i}]$, the distribution of any individual observed emission $\mathbf{e}_t = [a_t, b_t]$ is,

$$p_i(\mathbf{e}_t) = \frac{1}{4\pi D_i \Delta t} \exp\left(-\frac{(a_t - v_{0,i}\Delta t)^2 + b_t^2}{4 D_i \Delta t} \right) . \tag{3.35}$$

With these changes, the remaining steps for model selection and parameter fitting are the same as for the original HMM described in Section 3.3.

3.5 Application to diffusive state switching

3.5.1 Performance on simulated trajectories

Trajectories were simulated by drawing displacements from normal distributions with prescribed means and standard deviations. We first tested whether the HMM-Bayes algorithm described above can correctly identify the presence of two different diffusive states with different diffusion coefficients in a single particle trajectory. Figure 3.8 compares the results of HMM-Bayes analysis on a simple trajectory with just a single diffusion coefficient and a trajectory that contains a single switch from a high D_1 to a low D_2, where $D_1/D_2 = 5$. HMM-Bayes correctly infers the number of states present in the trajectory and finds the correct sequence of states (sequence of D values) along the trajectory. The reported sequence of states is the maximum likelihood state sequence for the maximum likelihood parameter values found during the MCMC iterations. The model selection procedure works equally well for trajectories that switch back and forth between these states, either regularly (Figure 3.9a) or stochastically (Figure 3.9b). The presence of shorter-lived states in these trajectories does not have a significant impact on the model probabilities, but does increase the number of incorrectly-assigned states in the inferred state sequence (Figure 3.9).

As in Chapter 2, the ability to resolve complex motion models depends on the number of observations (in this case, the number of steps in the trajectory) and on the relative parameter values. We tested the performance of HMM-Bayes on trajectories with variable numbers of steps and with different D_1/D_2 ratios. As expected, as the number of steps decreases, the ability to resolve the two-state D-D model over the simpler one-state D model is reduced (Figure 3.10). The crossover between these two models occurs at a larger number of steps for a smaller D_1/D_2 ratio, meaning that longer trajectories are needed to resolve smaller

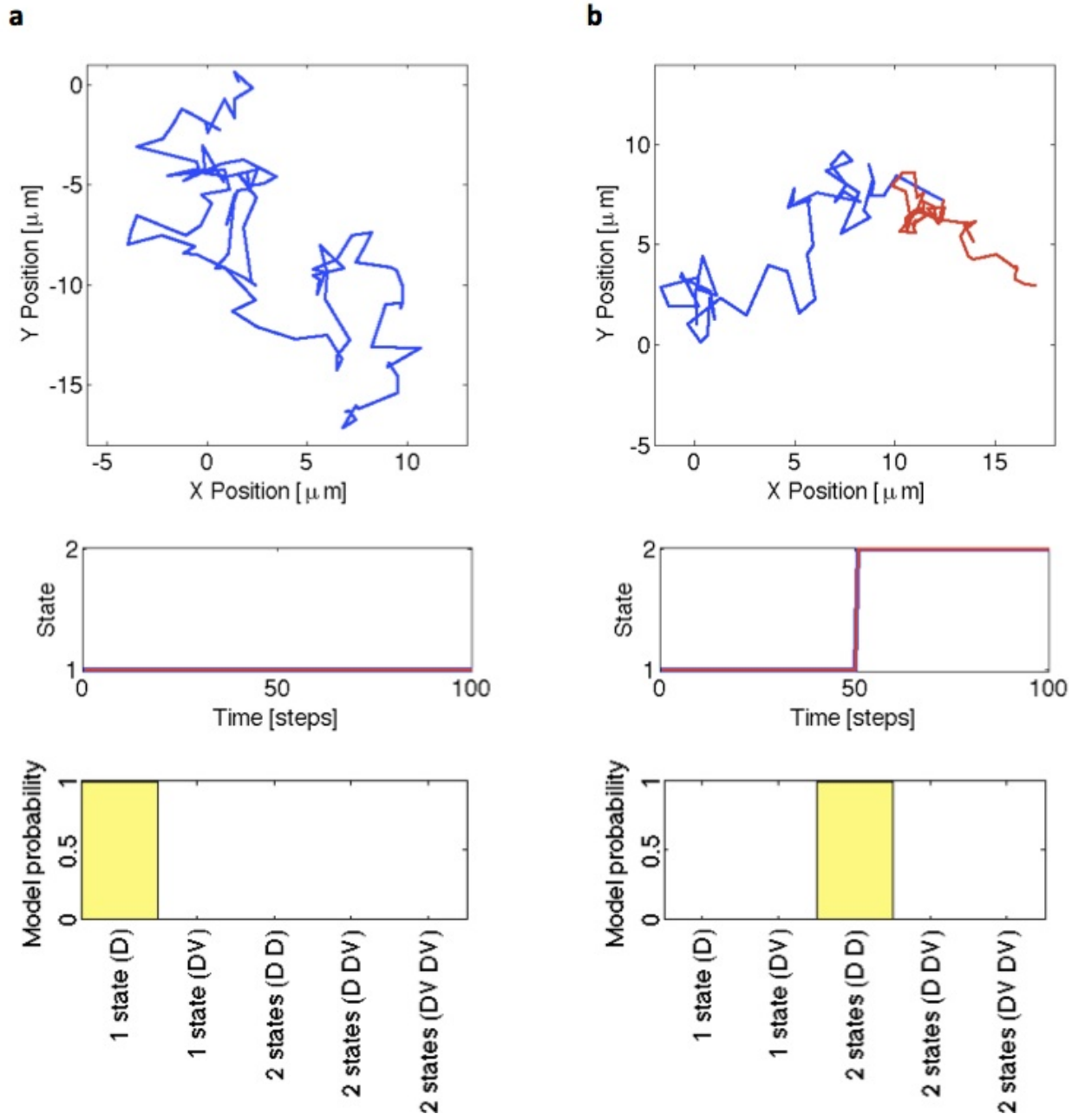

Figure 3.8: Results of HMM-Bayes on a trajectory with displacements drawn from a single state with zero mean and $\sigma = 1$ **(a)** and a trajectory with displacements drawn from two states with zero mean and $\sigma_1 = 1$ and $\sigma_2 = 1/\sqrt{5}$ **(b)**. *Top*: Simulated trajectories with displacements colored by state (blue = state 1, red = state 2). *Middle*: True state sequence (blue) and maximum likelihood state sequence (red) obtained from the highest-probability model. *Bottom*: Model probabilities for 1- and 2-state models.

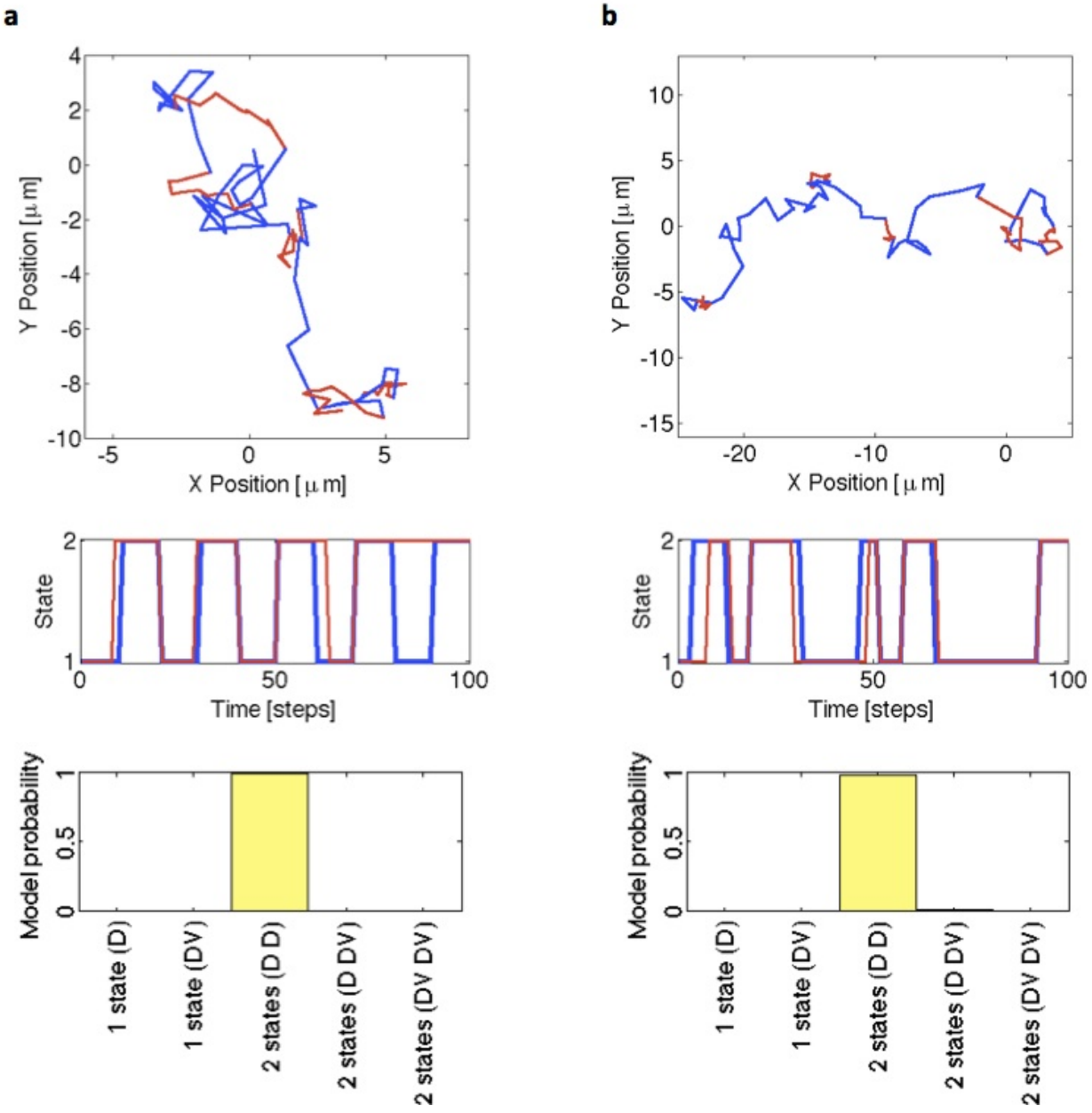

Figure 3.9: Results of HMM-Bayes on a trajectory with displacements drawn from the same two states as in Figure 3.8 with regular switching between states **(a)** and stochastic switching with a probability 0.1 of switching states at any step **(b)**. *Top*: Simulated trajectories with displacements colored by state (blue = state 1, red = state 2). *Middle*: True state sequence (blue) and maximum likelihood state sequence (red) obtained from the highest-probability model. *Bottom*: Model probabilities for 1- and 2-state models.

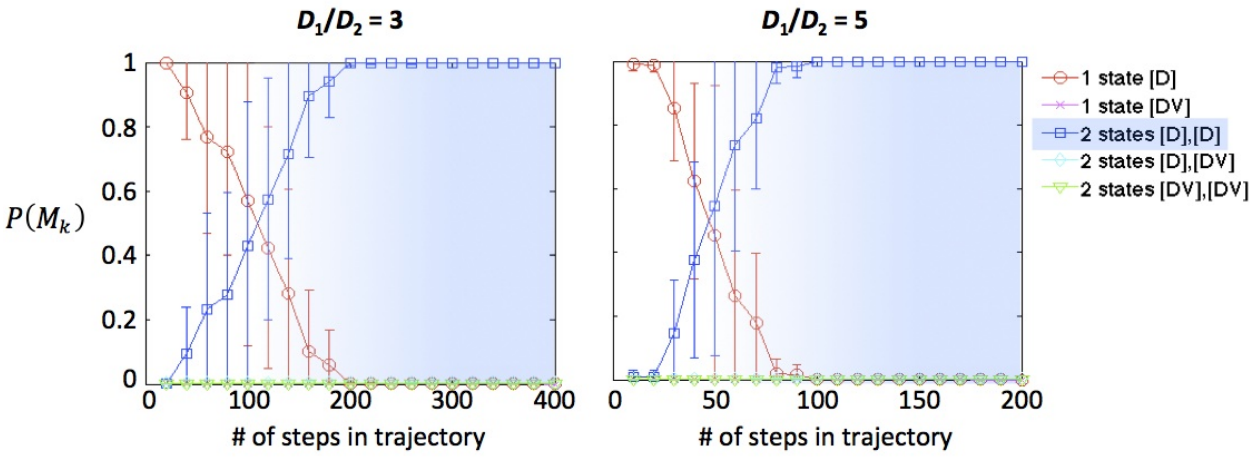

Figure 3.10: Model probabilities found by HMM-Bayes for trajectories simulated with two states as in Figure 3.8 with two different D_1/D_2 ratios (3, left, and 5, right) and with variable trajectory lengths as shown on the x-axis. Note that the ratio of the standard deviations σ_1/σ_2 used to generate the displacements is equal to the square root of the D_1/D_2 ratio. The model probabilities are shown as means and standard deviations over 40 repetitions of the simulations and inference procedure.

differences between diffusion coefficients. These results are consistent with the behavior of the MSD-Bayes method presented in Chapter 2.

The above results were all obtained using a single particle trajectory as the input to the HMM-Bayes algorithm; however, as described in Section 3.4.1, it is also possible to use information from multiple pooled trajectories to fit the particle trajectory HMM by multiplying the individual likelihoods. In theory, pooling trajectories should increase the number of available observations of the displacements such that the effective trajectory length is the sum of the lengths of all the pooled trajectories. For a D_1/D_2 ratio equal to 5, note that 100 steps in a single trajectory is sufficient to resolve the two different diffusive states, but 10 steps is not sufficient (Figure 3.10). Therefore, we tested whether 10 trajectories of 10 steps in length, each containing a single switch between the same two motion states, could be used to resolve the presence of the two diffusion coefficients. Figure 3.11a shows the control case of a single 10-step trajectory, for which the two motion states are not resolved,

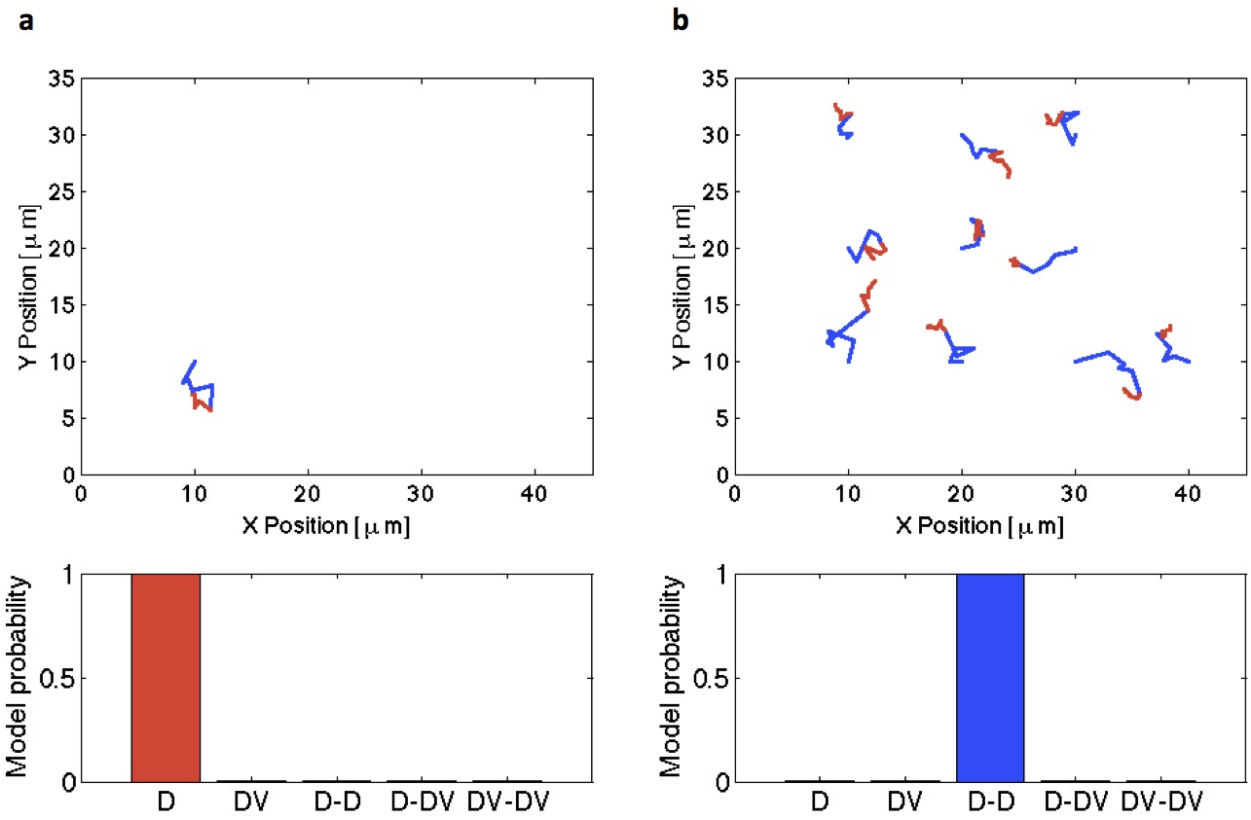

Figure 3.11: Results of running HMM-Bayes on a single trajectory **(a)** or 10 trajectories **(b)**, each with 10 displacements drawn from the same two states as in Figure 3.8b with a single switch between states. *Top*: Simulated trajectories with displacements colored by state (blue = state 1, red = state 2). *Bottom*: Model probabilities for 1- and 2-state models.

as expected. However, when 10 of these trajectories are pooled, the presence of the two different diffusive states is detected (Figure 3.11b). This result has important implications for the analysis of many types of cell biological particles for which only short trajectories are available. Assuming the particles are undergoing the same physical or chemical processes that define their available motion states, these short trajectories can be pooled to significantly increase the power of the inference process.

3.5.2 Validation on experimental datasets

To test the performance of the diffusive state detection capability of HMM-Bayes on experimental trajectories, we analyzed two high-resolution datasets containing hundreds to thousands of trajectories of two different membrane proteins. First, we analyzed trajectories of HIV Gag proteins, which assemble into oligomeric buds at the plasma membrane [Manley et al., 2008]. Figure 3.12a shows the cell surface distribution of HIV Gag trajectories that are between 20 and 100 steps long that were pooled and analyzed by HMM-Bayes. The algorithm detects the presence of three different diffusive states within these trajectories with diffusion coefficients of roughly 0.04, 0.005, and 0.0015 $\mu m^2/s$. Annotation of the individual trajectories with their maximum likelihood state sequences (Figure 3.12b-e) reveals that most of the Gag proteins switch once during the length of their trajectories, typically between the two lowest diffusion coefficient states, and that the three different diffusive states are not equally represented over the population, with the highest diffusion state being particularly rare (Figure 3.12f).

Next we analyzed trajectories of the AMPA receptor (AMPAR) in the membrane of hippocampal neurons. Previous analysis of these trajectories suggested that the receptors undergo different motions in dendrites versus dendrite spines [Hoze et al., 2012]. Therefore, we pooled trajectories in different regions of the cell, including the cell body, the dendrites, and dendrite spines of two different shapes (Figure 3.13a) and analyzed these regions separately. HMM-Bayes detects the presence of three different diffusive states in each of the analyzed regions. Interestingly, the values of the diffusion coefficients in the different regions are almost identical (Figure 3.13b), but the number of particles in each of the three states differs somewhat between regions. In particular, the state with the smallest diffusion coefficient is found least often in the cell body and most in the dendrite spines (Figure 3.13c-d), consistent with a possible role for this state in neuronal signaling [Hoze et al., 2012].

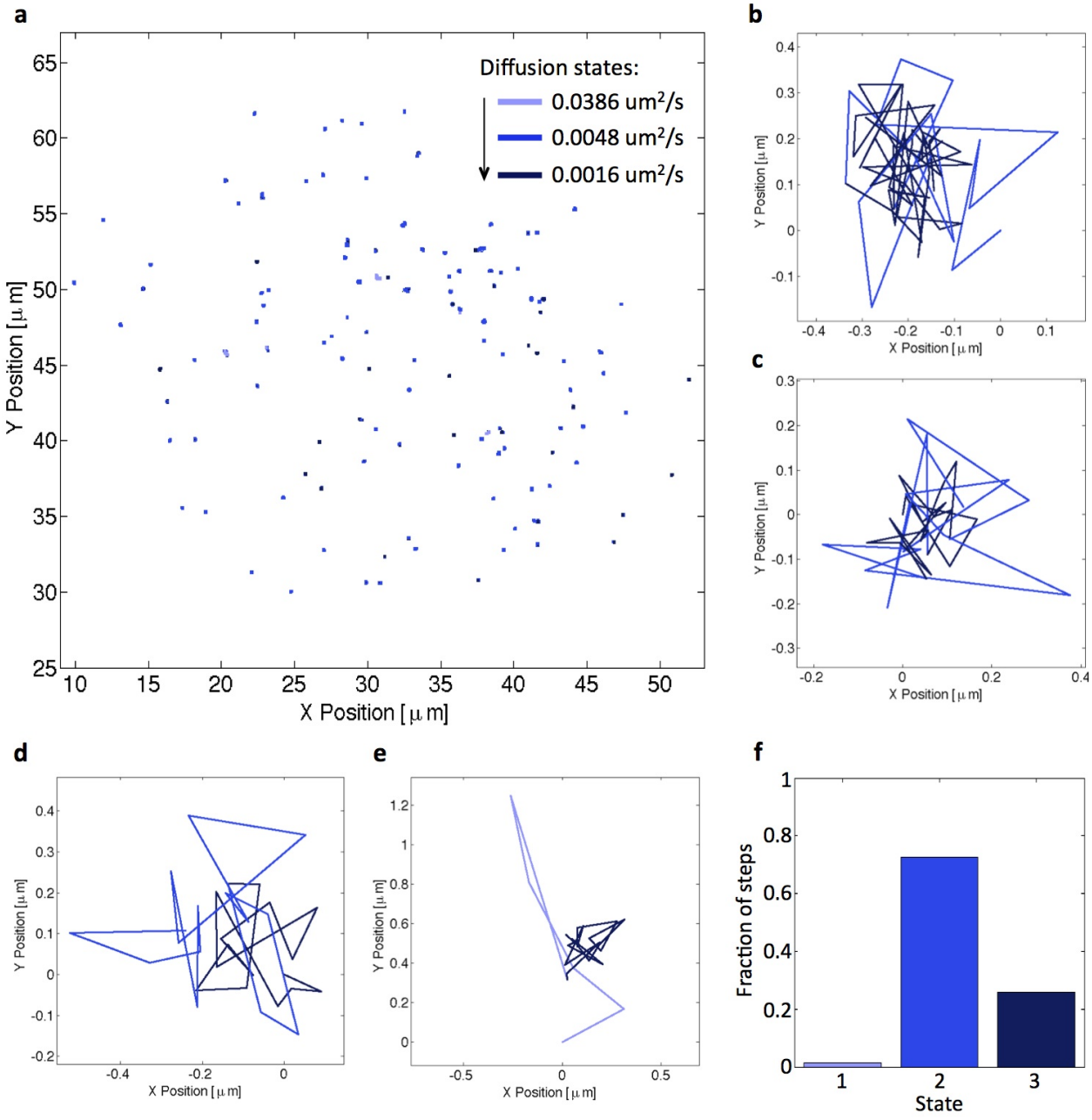

Figure 3.12: **(a)** Distribution of mEos-labeled HIV Gag protein trajectories on the surface of a HeLa cell. **(b)-(e)** Selected individual HIV Gag trajectories that show state-switching behavior. The trajectory steps are colored by their assigned state, corresponding to the diffusion coefficients listed in (a). **(f)** Fraction of the total set of particle displacements across all trajectories that are assigned to each of the three states.

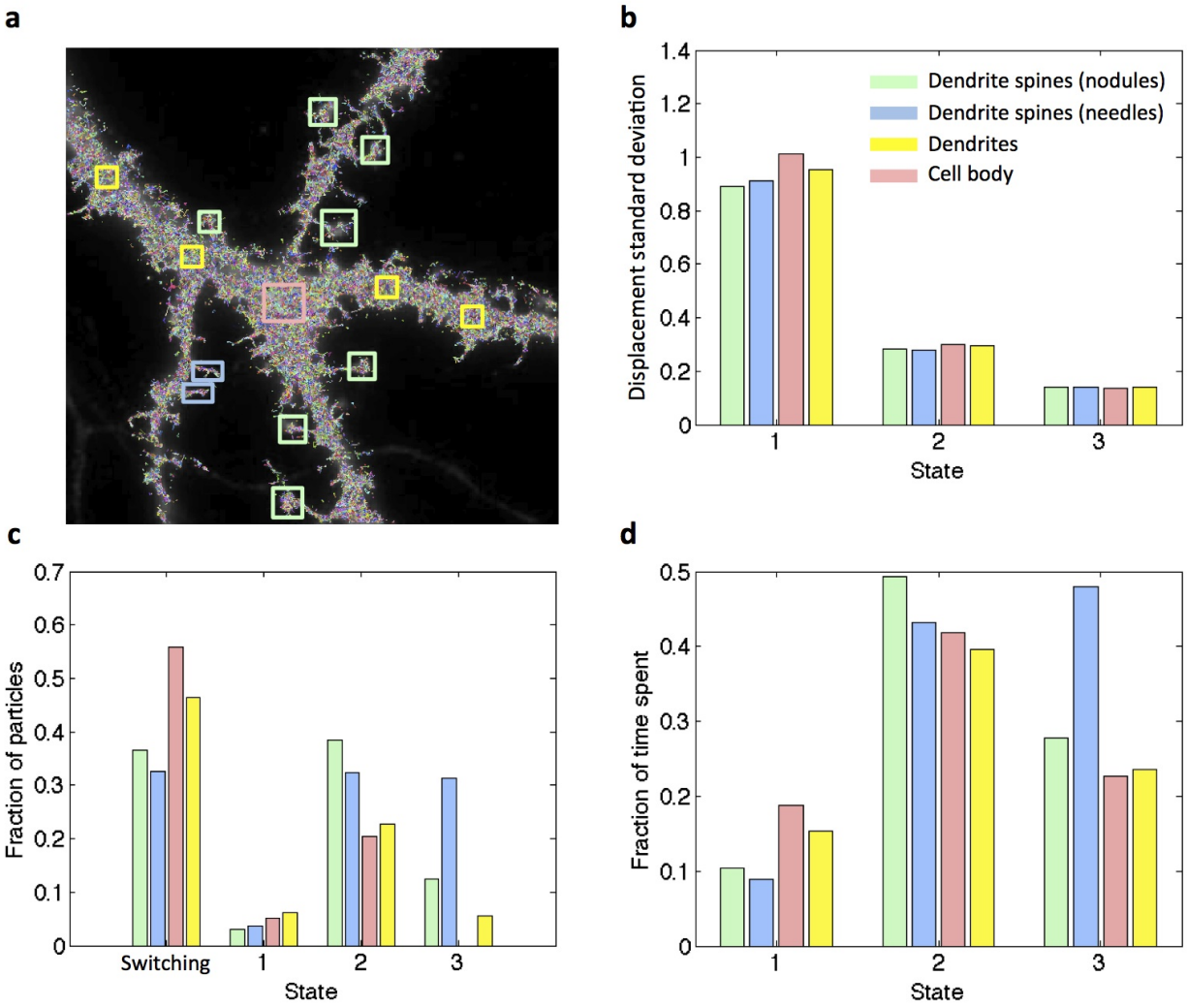

Figure 3.13: **(a)** AMPAR trajectories (randomly colored) superimposed on an image of the corresponding hippocampal neuron. Square boxes indicate the regions analyzed in (b)-(d). **(b)** Relative diffusion coefficients of the three detected states. **(c)** Fraction of particles that are entirely assigned to one of the three states or that switch between states. **(d)** Fraction of particle displacements assigned to each of the three states.

3.6 Application to directed motion

3.6.1 Single trajectories

The diffusive state detection results above demonstrate that the HMM-Bayes algorithm performs similarly to the recently-published diffusion detection HMM in [Persson et al.,

2013]. However, none of the published particle trajectory HMM methods have tested for the presence of directed motion in addition to diffusion. Here we show that the HMM-Bayes algorithm can correctly identify the presence of two different motion states, one diffusive and one directed, in single and pooled particle trajectories. Figure 3.14 shows the results of HMM-Bayes analysis on two simulated trajectories that switch between random and directed motion at different switching rates. HMM-Bayes correctly infers the number of states present in each trajectory and finds state sequences along the trajectories that are very similar to the true state sequences (Figure 3.14b).

An important quantity affecting the ability to resolve directed motion in the presence of stochastic diffusion is the ratio between the difference in displacement means and the sum of the displacement standard deviations, $|\mu_2 - \mu_1|/(\sigma_1 + \sigma_2)$. The simulations here are performed with an initial state of mean zero (no velocity) and the same standard deviation σ for each of the two states; thus, the relevant ratio of interest is μ_2/σ. We tested the performance of HMM-Bayes in inferring the presence of directed motion while varying the value of this ratio (Figure 3.15). As expected, as this ratio is decreased for a constant number of steps in the trajectory, the ability to resolve the two-state D-DV model over the simpler one-state D model is reduced (Figure 3.15).

3.6.2 Aligned vs. randomly-oriented flow

As in the case of diffusive state switching above, particle trajectories can be pooled to improve the resolution of directed motion. However, as discussed in Section 3.4.2, because the fit velocity for each state is a vector quantity, resolution will only be improved if the pooled trajectories have velocities in the same direction. To handle cases in which trajectories all have the same speed but different directions of velocity, we developed a variant of the

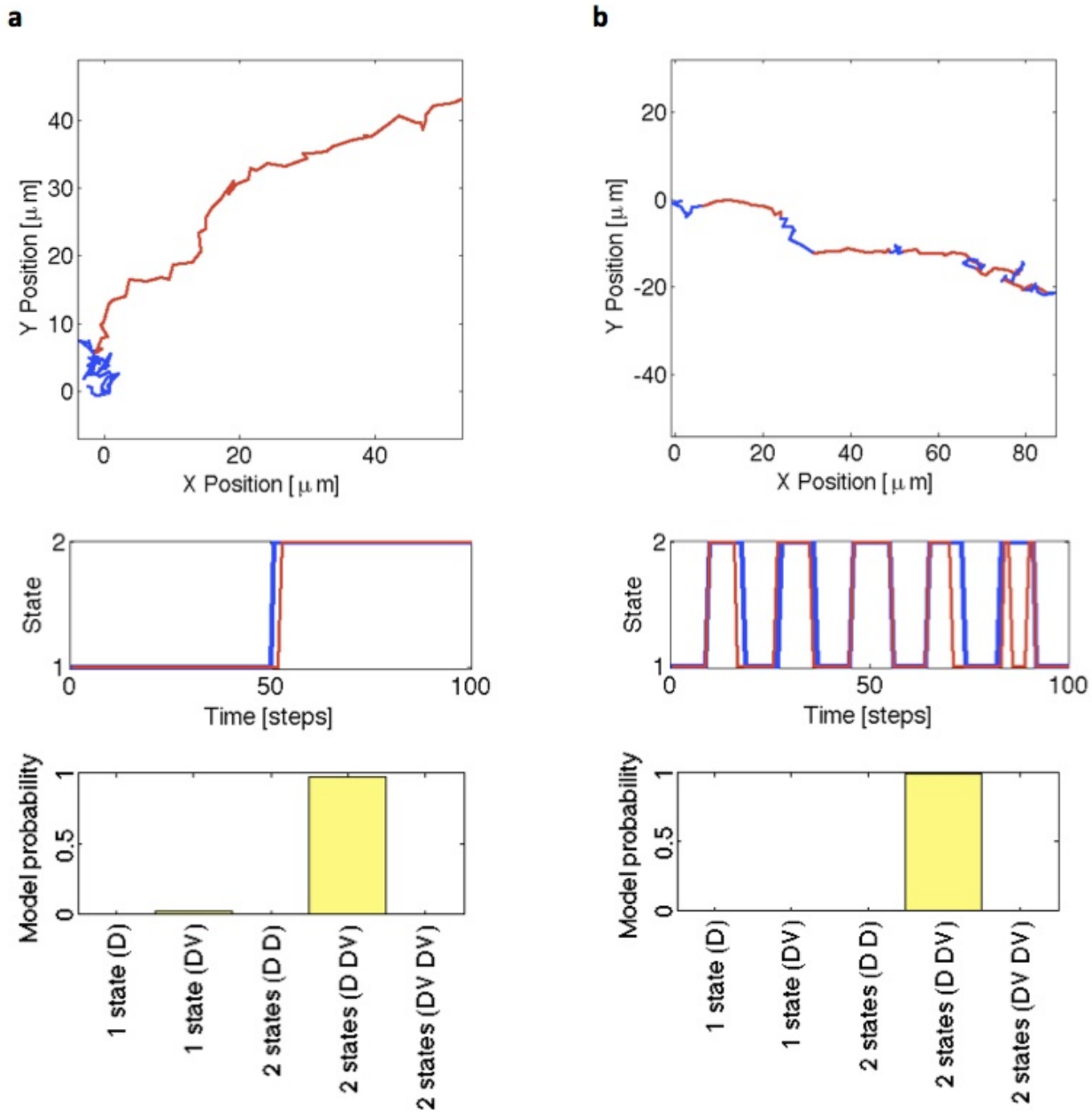

Figure 3.14: Results of HMM-Bayes on trajectories that switch between a state with displacement distribution with zero mean (blue) and a state with $\mu = 1$ (red) at two different switching intervals, once in the trajectory **(a)** or every 10 steps **(b)**. For both states, $\sigma = 1$. *Top*: Simulated trajectories with displacements colored by state (blue = state 1, red = state 2). *Middle*: True state sequence (blue) and maximum likelihood state sequence (red) obtained from the highest-probability model. *Bottom*: Model probabilities for 1- and 2-state models.

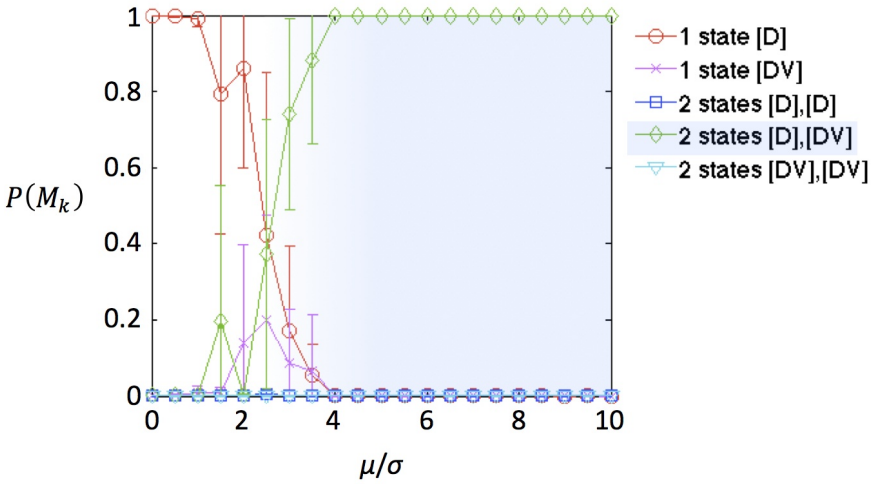

Figure 3.15: Model probabilities found by HMM-Bayes for trajectories simulated with two states as in Figure 3.14 but with only 20 steps per trajectory. The ratio of the mean of the displacement distribution in the second state to the sigma of the displacement distributions in both states is varied along the x-axis. The model probabilities are shown as means and standard deviations over 40 repetitions of the simulations and inference procedure.

HMM-Bayes algorithm that fits displacement magnitudes using chi-squared distributions instead of full displacements with normal distributions (Section 3.4.2). Here we compare the performance of the original HMM-Bayes algorithm and the χ^2-HMM-Bayes algorithm on simulated trajectories with directed motion in the same or different directions (Figure 3.16). When the particles are all flowing in the same direction, both algorithms correctly infer the presence of nonzero velocity (Figure 3.16a). However, when the same particle trajectories are rotated to random orientations, the original HMM no longer detects a nonzero mean for the displacement distributions, as expected. The χ^2-HMM, on the other hand, still detects the presence of directed motion in these randomly-oriented trajectories (Figure 3.16b).

The χ^2-HMM thus seems more generally applicable than the original HMM. However, this greater inference power for randomly-oriented trajectories has a trade-off in power when the trajectories do in fact have the same direction of velocity. The reason for this trade-off is that χ^2 distributions have more overlap than their normal distribution counterparts for

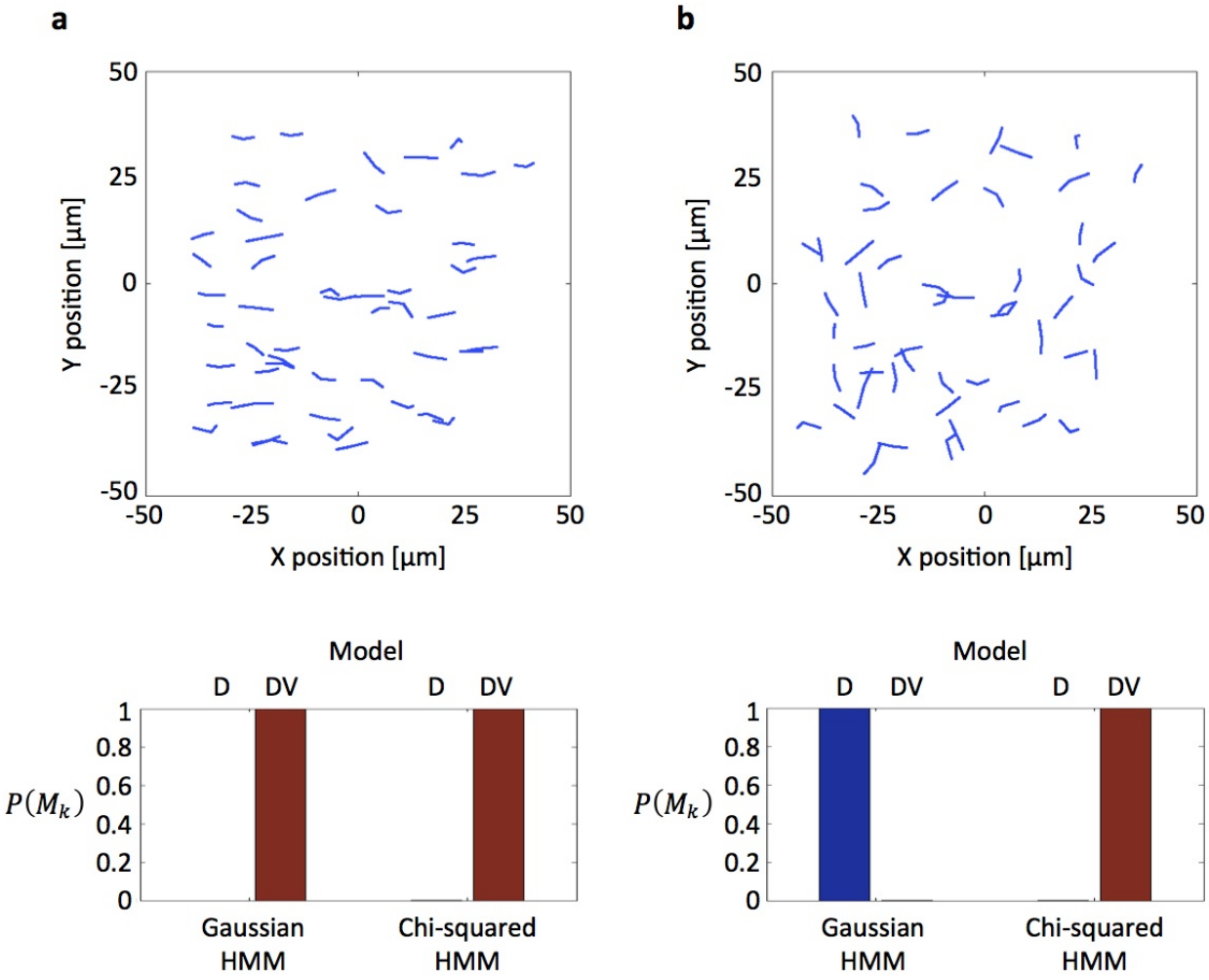

Figure 3.16: **(a)** *Top*: 50 simulated trajectories with 2 steps each, drawn from normal distributions with $\mu = 3$ and $\sigma = 1$. *Bottom*: Model probabilities for the one-state models obtained by the original HMM and the χ^2-HMM. **(b)** Same as (a), except that each trajectory has been rotated through an angle drawn at random from a uniform distribution on [0,2π].

the same underlying values of μ and σ in the different dimensions. In other words, a χ^2 distribution has lost any directional information that was present in a corresponding normal distribution; this loss of information is a benefit if the directional information is inconsistent between trajectories, but reduces inference power in the case where the particles are all aligned. To illustrate the decrease in performance for the case of consistent directional information, we compared the original and χ^2-HMM algorithms using individual trajectories undergoing flow in a constant direction and varying the μ/σ ratio as in Figure 3.15. The

χ^2-HMM requires a significantly higher value of μ/σ to resolve directed motion than the original HMM for a trajectory of the same length (Figure 3.17).

3.6.3 Convergent flow

Finally, we tested the third version of the HMM algorithm (Section 3.4.3) designed to detect the presence of convergent flow (directed towards a point-sink) in a pool of particle trajectories. This algorithm uses normal displacement distributions, not chi-squared distributions, so it does not suffer from the reduced sensitivity shown in Figure 3.17. Instead, it fits the position of a potential sink and considers velocity only along the direction towards or away from the sink at every step along the trajectories. Figure 3.18 explores the conditions under which such convergent flow can be detected. Reducing the magnitude of the convergent velocity by reducing the μ/σ ratio eventually leads to loss of the ability to resolve the

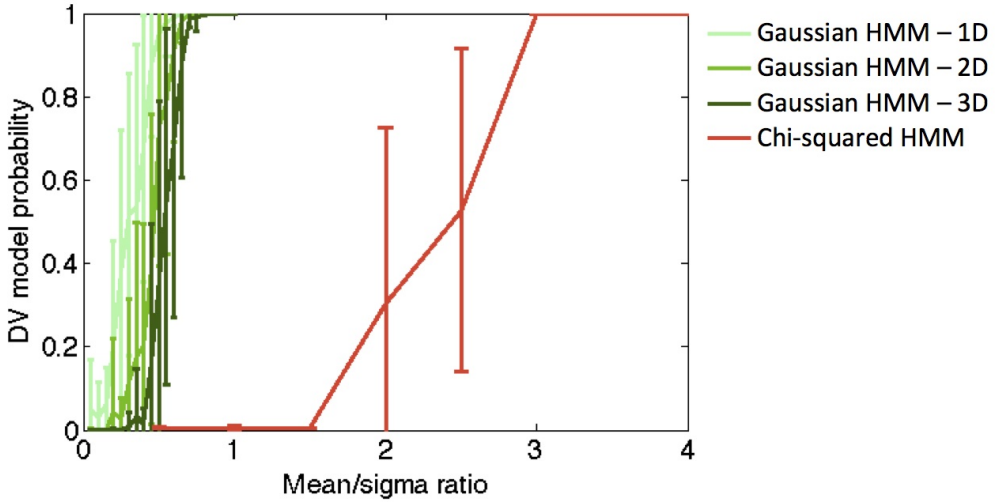

Figure 3.17: DV model (true model) probabilities found by the original HMM and the χ^2-HMM algorithms for 100-step trajectories simulated with a single diffusion plus flow state with $\sigma = 1$ and μ/σ ratio as shown along the x-axis. Note that the performance of the original Gaussian HMM varies slightly with the number of dimensions in which the particle trajectory is observed, due to the fact that the number of parameters in the fit velocity vector is equal to the number of dimensions.

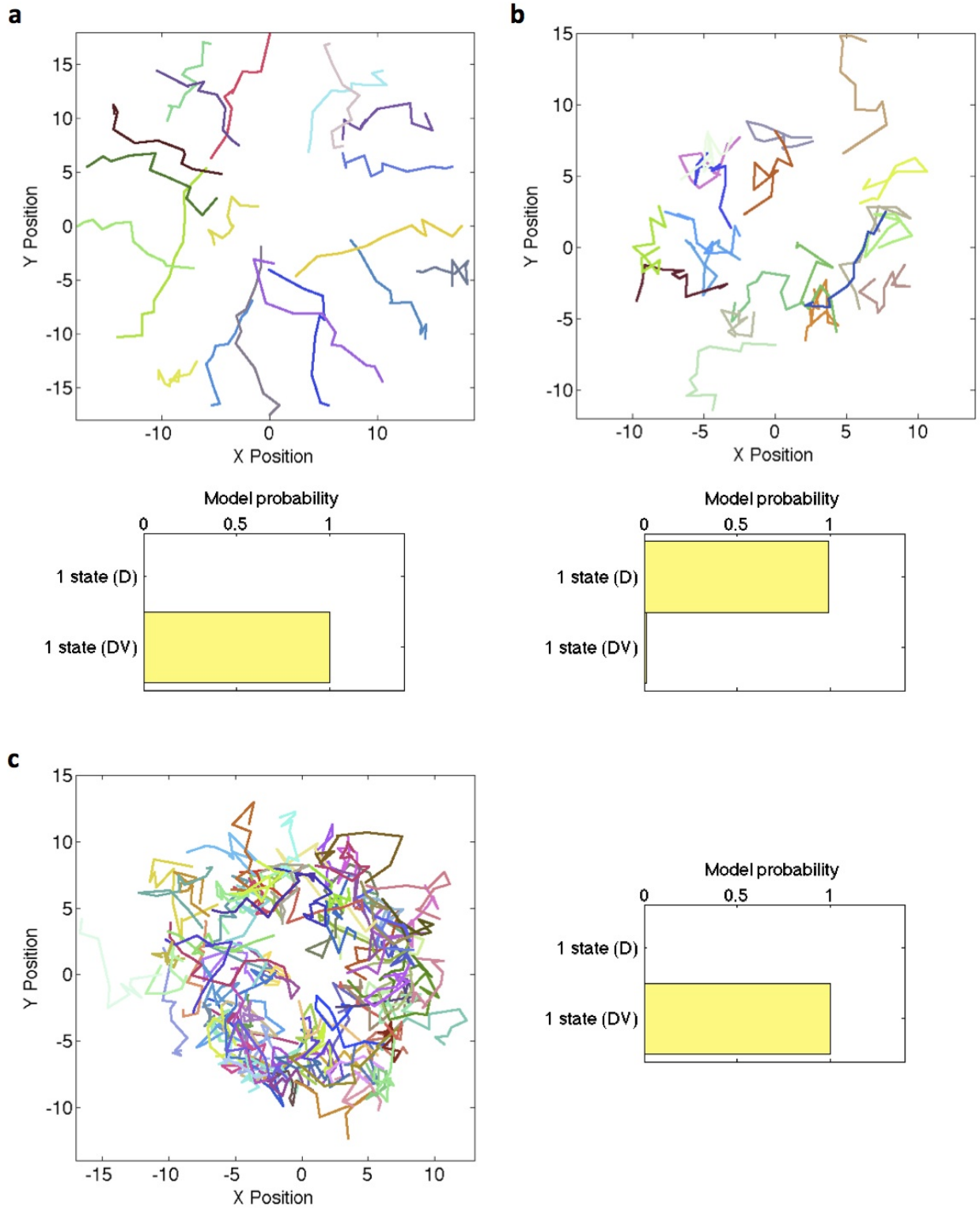

Figure 3.18: 10-step trajectories (randomly colored) simulated with $\sigma = 1$ and velocity only in the direction of the origin. **(a)** 20 trajectories with $\mu = 1$, **(b)** 20 trajectories with $\mu = 0.2$, and **(c)** 100 trajectories with $\mu = 0.2$. Model probabilities for the 1-state models calculated using the convergent flow HMM are shown in yellow.

convergent motion (compare Figure 3.18a and b), but subsequently raising the number of available trajectories restores the ability to resolve the convergent motion (Figure 3.18c). This behavior is analogous to the behavior of the other Bayesian particle trajectory analysis approaches presented above and in Chapter 2. The convergent flow variant of the HMM is likely to be applicable in cell biological systems in which particles come together to form complexes, to test whether they find each other stochastically or whether there is directed motion towards the assembly point.

3.7 Conclusion

Here we present a powerful approach based on hidden Markov models to extract transient motion states along single particle trajectories. Our approach builds on and extends recent HMM-based approaches to infer switching between motion models in addition to switching between parameter values, making it applicable to directed motion as well as diffusive motion. The approach is highly versatile in that it can be applied to individual particle trajectories or to pooled trajectories, depending on which is more appropriate to a given biological dataset. Because of the implications of trajectory pooling for velocity inference, we have developed three variants of the HMM-Bayes algorithm; one that is best suited to detecting directed motion within a single trajectory or in pooled trajectories that flow all in the same direction, a second that is best suited to detecting directed motion in pooled trajectories that flow in random or uncorrelated directions, and a third that is best suited to detecting convergent flow towards a common location. As all three of these cases occur frequently in cell biology, the HMM-Bayes algorithm presented here is a particularly promising approach for widespread and automated analysis of particle trajectories across biological systems.

Chapter 4

Analysis of chromosome transport mechanisms in cell division

Portions of this chapter has been previously published in:

Mori M*, **Monnier N***, Daigle N, Bathe M, Ellenberg J, Lénárt P. Intracellular transport by an anchored homogeneously contracting F-actin meshwork. *Current Biology* 21(7):606-611 (2011)
*These authors contributed equally

and:

Schmidt JC, Arthanari H, Boeszoermenyi A, Dashkevich NM, Wilson-Kubalek EM, **Monnier N**, Markus M, Oberer M, Milligan RA, Bathe M, Wagner G, Grishchuk EL, Cheeseman IM. The kinetochore-bound Ska1 complex tracks depolymerizing microtubules and binds to curved protofilaments. *Developmental Cell* 23(5):968-980 (2012).

4.1 Overview

This chapter focuses in detail on two biological systems to highlight the importance of quantitative particle trajectory analysis for elucidating mechanisms of intracellular transport; in particular, transport of chromosomes during mitosis and meiosis. The two systems described here—chromosome congression in starfish oocytes and chromosome segregation in human cells—involve very different cellular mechanisms for capturing and transporting chromosomes and rely on distinct sets of protein players. In the starfish oocyte system (Section 4.3), quantitative analysis of trajectories of the transported chromosomes reveals

novel design principles of the F-actin meshwork that drives their motion. In the case of chromosome segregation in human cells (Section 4.4), analysis of in vitro trajectories of one of the key components of the kinetochore, the Ska1 complex, reveals principles of its interaction with microtubules, an essential interaction for maintaining robust attachment between chromosomes and force-generating microtubules during chromosome transport. The particle trajectories used for analysis in both of these systems were generated using a novel tracking algorithm (Section 4.2) that handles both the amorphously-shaped starfish chromosomes and the diffraction-limited Ska1 particles.

4.2 Particle tracking algorithm for amorphous objects

4.2.1 Motivation

Tracking particle motion requires two basic steps, segmenting individual frames of a time-lapse image series and linking particle positions between frames. Segmentation refers to the image processing procedure that identifies particles of interest within a noisy image. As described in Section 1.2, segmentation is often aided by the fact that the particles of interest are labeled with fluorescent markers that are smaller than the point-spread function of the imaging system; as a result, the particles of interest all have the same shape, namely, a Gaussian intensity profile. These point particles can be identified within an image using methods that have been developed to detect Gaussian peaks [Sergé et al., 2008]. However, in some biological applications the particles to be tracked are larger than the point-spread function and are not reliably identified by these peak-detection methods. We found this to be the case for the starfish chromosomes that are analyzed in Section 4.3. The chromosomes are labeled using a fluorescently-tagged histone protein, H2B, which binds to locations all along the length of each chromosome; thus, the full amorphous shapes of these chromosomes are visible in the resulting images (Figure 4.1a). Therefore, we sought to develop a simple

but robust algorithm for segmentation and tracking that can identify amorphous structures in addition to point particles.

4.2.2 Segmentation by maximizing mean feature size

The segmentation algorithm presented here is based on the assumption that the particles of interest in an image are larger (in number of pixels or voxels) than background intensity fluctuations arising from stochastic noise in the imaging process. This assumption is generally valid, as background noise is typically on the single-pixel scale. Here we use this property to automatically choose the most appropriate level at which to threshold each image in a movie to identify particles of interest while excluding noise. Consider the effect of thresholding a noisy image at a range of threshold levels, as illustrated in Figure 4.1. At very high threshold levels, only the tips of the intensity peaks of the brightest particles will be above the threshold, with the resulting detected features being only a few pixels (or voxels) in size. Here we define a detected feature as a set of contiguous above-threshold pixels; each connected white object in the thresholded images in Figure 4.1 is therefore one feature. As the threshold is lowered, the number of pixels in each detected feature increases (compare levels 0.6 and 0.8 in Figure 4.1). However, when the threshold is lowered to the level of the background noise pixels, many single-pixel features start to be detected above threshold (level 0.4 in Figure 4.1). These noise features each contain significantly fewer pixels than the true particles. Finally, when the threshold level is reduced below the mean background noise level, the above-threshold pixels form one or more very large connected regions covering a substantial portion of the image (level 0.2 in Figure 4.1).

The behavior of the feature-detection process as a function of threshold allows us to automatically identify the threshold level corresponding to background noise in the image

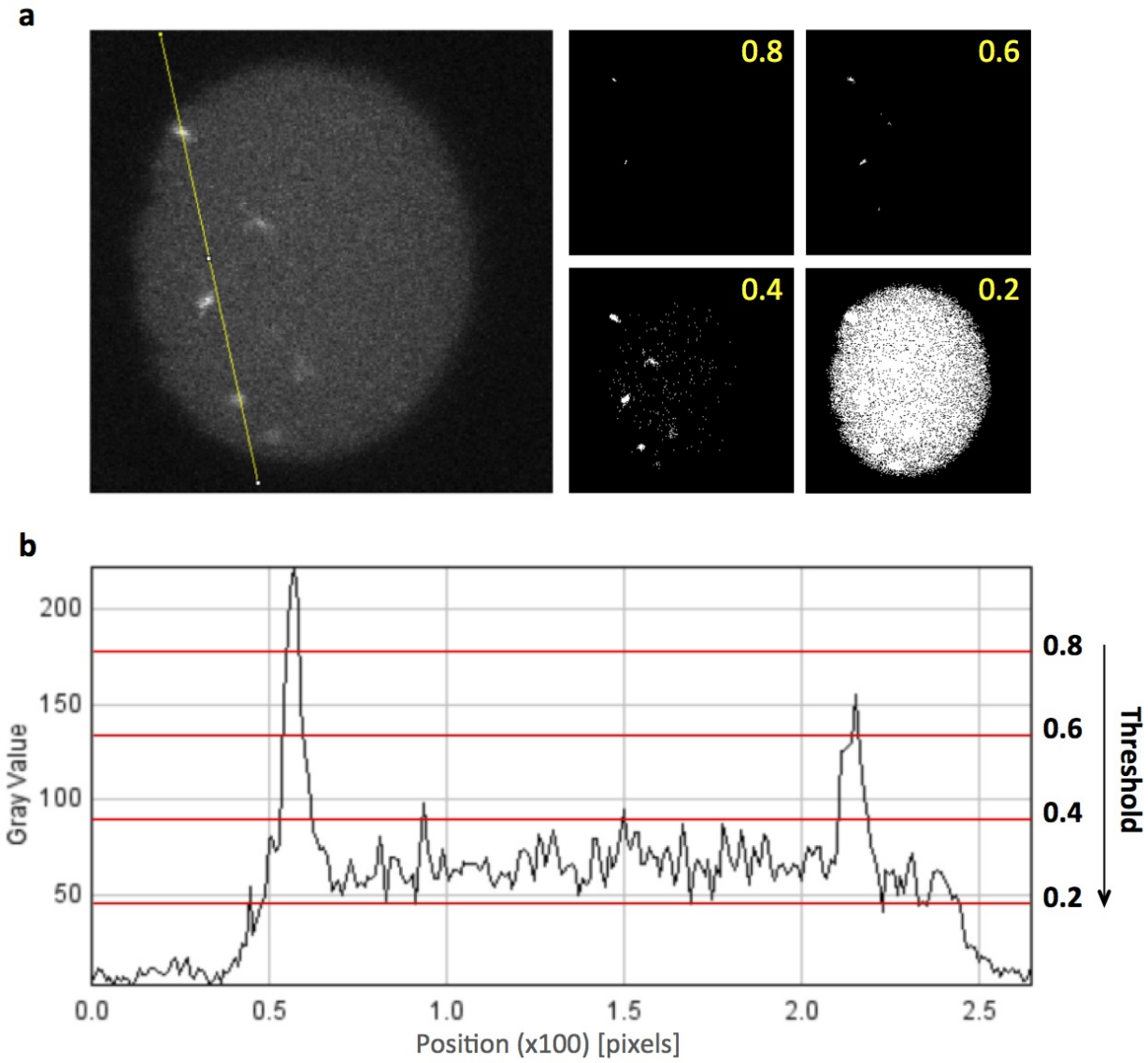

Figure 4.1: **(a)** *Left*: Single z-section through the nuclear region (visible as gray background fluorescence) of a starfish oocyte expressing H2B-mCherry. 27 z-sections were acquired to span the nucleus. Partially-condensed chromosomes are visible as brighter white spots. The yellow line indicates the location of the intensity profile in (b). *Right*: binary images obtained by thresholding the raw image at four different threshold levels, relative to the maximum raw intensity value. **(b)** Fluorescence intensity profile along the yellow line in (a). The threshold levels used in (a) are indicated as red lines.

by plotting the mean feature size (number of pixels or voxels per feature) against a range of tested threshold levels (Figure 4.2). Moving from right to left along this curve, the initial increase corresponds to the expected increase in size of detected features at lower thresholds, but the rapid drop occurs when the threshold reaches the highest level of background noise. The peak in this curve, therefore, corresponds to the threshold level at which only true particles are detected with almost no contamination from noise features. Once this threshold level has been identified, the user has some control over the noise tolerance by setting a multiplication factor for the threshold. This method for identifying the appropriate threshold level is automatic and generally applicable to a variety of images with different levels of noise or different overall intensity levels. The method is also independent of the shape of the particles, so it is equally applicable to amorphously-shaped particles as it is to Gaussian peaks.

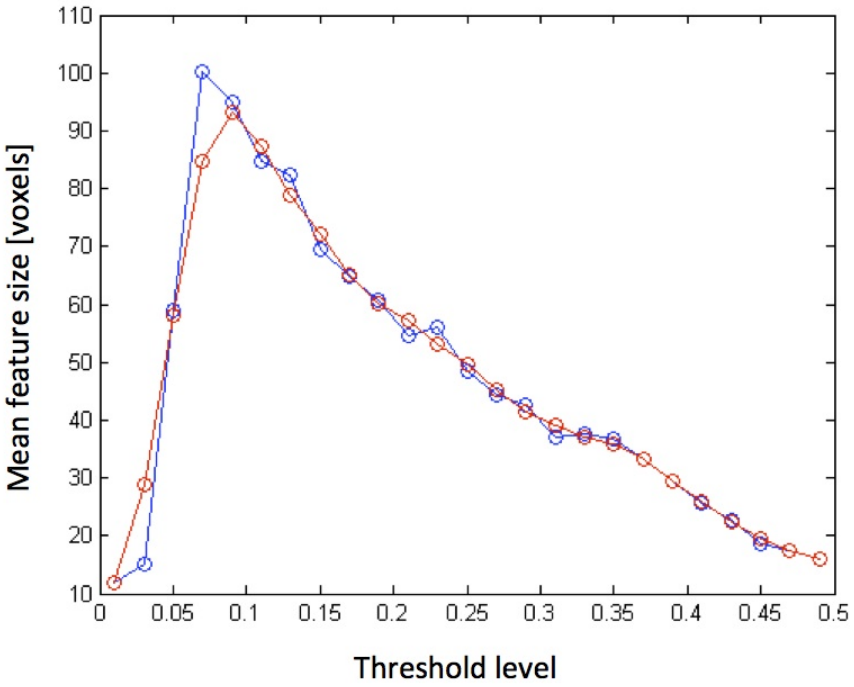

Figure 4.2: Plot of the mean size (in 3D voxels) of contiguous features detected at each of the threshold levels in a full z-stack through a starfish oocyte nucleus. The raw size measurements are shown in blue and a 3-point smoothed curve is shown in red.

In practice, there are a number of image processing steps that can enhance the raw images and improve the signal-to-noise ratio of the features of interest. For example, using a single-pixel Gaussian filter to reduce the intensity variations of the background noise and using local mean background subtraction to even out the noise intensities across the entire image can greatly enhance the signals from true particles (Figure 4.3). These initial image processing steps generally need to be tailored to specific imaging datasets prior to applying the maximum-mean-feature size segmentation procedure.

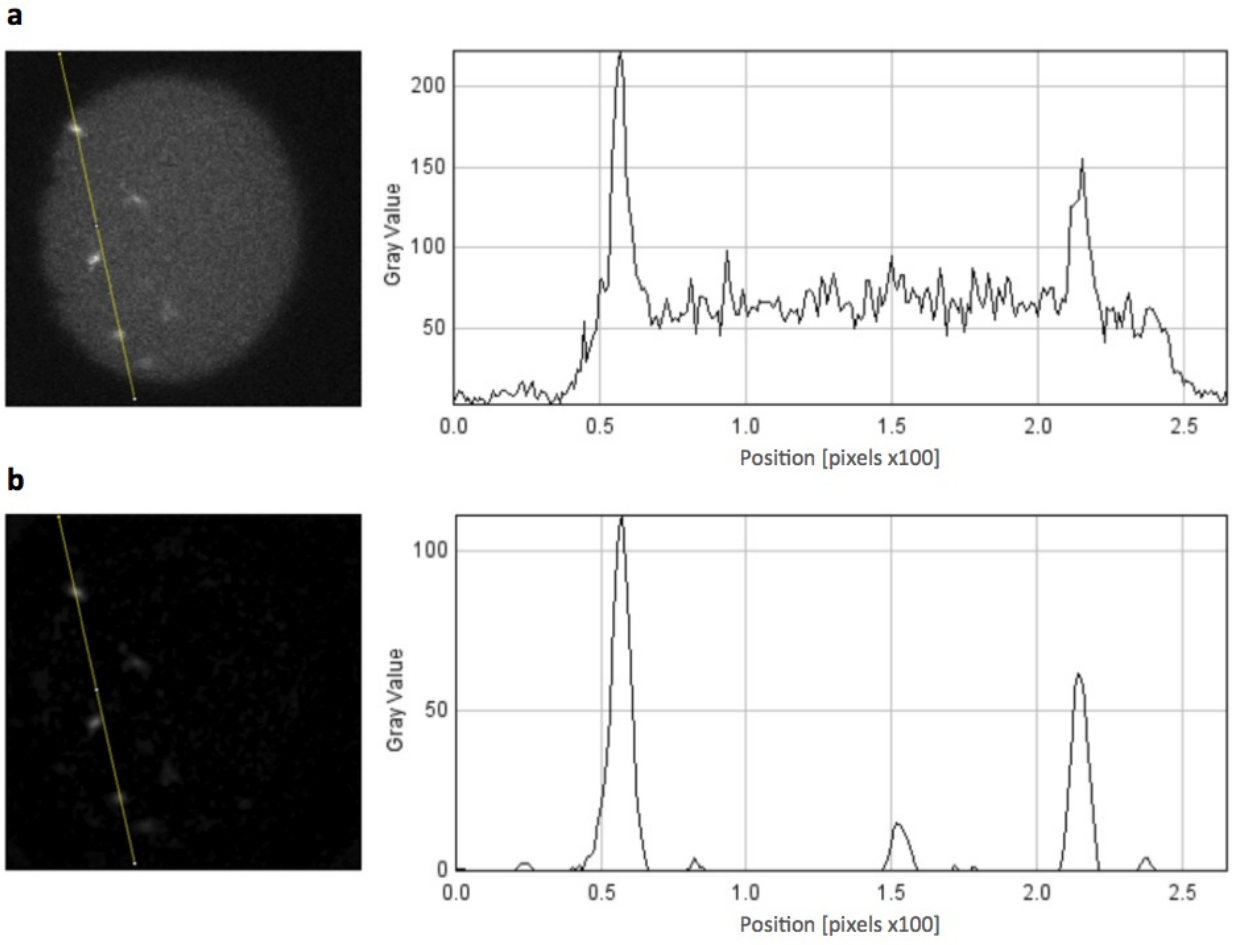

Figure 4.3: **(a)** Raw image and intensity profile from Figure 4.1. **(b)** Image and intensity profile after applying a Gaussian intensity filter with a radius of 1 pixel and subtracting a background image obtained by morphological opening on a scale larger than the size of the chromosomes.

4.2.3 Constructing and updating the assignment matrix

Assuming a reliable segmentation of the particles of interest from each image in a time series, the next step in tracking is to link particles between consecutive frames to form trajectories. We sought a simple procedure that would be more computationally efficient than a full optimization over possible tracking assignments but still take into account competition between trajectories that have high likelihoods of assignment to the same feature(s). Here we describe a method based on updating a track-feature assignment probability matrix by pooling information across the matrix. The theory below is written assuming a two-dimensional system, but the equations are analogous in three dimensions.

Initial values for the assignment probabilities are generated by comparing the position of each existing track T_i in the previous time frame, $\left[T_{i,x}^{(t-1)}, T_{i,y}^{(t-1)}\right]$, to the position of each detected feature F_j in the current time frame $\left[F_{j,x}^{(t)}, F_{j,y}^{(t)}\right]$. The previous position $\left[T_{i,x}^{(t-1)}, T_{i,y}^{(t-1)}\right]$ of a particle is the best predictor of its expected current position if the particle is freely diffusing. However, if the particle also has a directed motion component, then an expected current position can be calculated based on its velocity over the previous time frames. Using the average velocity over the previous τ frames, the prediction $\left[\hat{T}_{i,x}^{(t)}, \hat{T}_{i,y}^{(t)}\right]$ of the track's current position is,

$$\left[\hat{T}_{i,x}^{(t)}, \hat{T}_{i,y}^{(t)}\right] = \left[T_{i,x}^{(t-1)}, T_{i,y}^{(t-1)}\right] + \frac{\left[T_{i,x}^{(t-1)}, T_{i,y}^{(t-1)}\right] - \left[T_{i,x}^{(t-1-\tau)}, T_{i,y}^{(t-1-\tau)}\right]}{\tau} . \tag{4.1}$$

So as not to assume one motion model over the other, assignment probabilities are calculated using both the previous position and the predicted position of a track. The distance between a current feature position and a previous track position at $t-1$ is,

$$d_{\text{previous}}\left(T_i, F_j\right) = \sqrt{\left(F_{j,x}^{(t)} - T_{i,x}^{(t-1)}\right)^2 + \left(F_{j,y}^{(t)} - T_{i,y}^{(t-1)}\right)^2} . \tag{4.2}$$

Similarly, the distance between a current feature position and a predicted track position is,

$$d_{\text{predicted}}\big(T_i, F_j\big) = \sqrt{\Big(F_{j,x}^{(t)} - \hat{T}_{i,x}^{(t)}\Big)^2 + \Big(F_{j,y}^{(t)} - \hat{T}_{i,y}^{(t)}\Big)^2}\,. \tag{4.3}$$

Letting $\mathcal{A}_{ij} \equiv \mathcal{A}\,(T_i, F_j)$ be the relative probability of assignment between a track T_i and a new feature F_j, we note that the dependence of $\mathcal{A}_{ij}$ on the distances above can take different functional forms, the most natural being a Gaussian function because the displacements of diffusive particles are normally distributed, as discussed in Chapter 3. However, in practice we find that a Gaussian form for $\mathcal{A}_{ij}$ does not give sufficient weight to assignments between overlapping features in consecutive time frames, and that an exponential form gives better performance. Thus, we define $\mathcal{A}_{ij}$ as,

$$\mathcal{A}_{ij} = \exp\Big(-\lambda\big(d_{\text{previous}}(T_i, F_j) + d_{\text{predicted}}(T_i, F_j)\big)\Big)\,, \tag{4.4}$$

which factors into a product over the two distances above. With this exponential form for $\mathcal{A}_{ij}$, we find that the value of the constant factor λ does not have a significant effect on the tracking performance.

The relative assignment probabilities from Equation 4.4 are used to initially populate a track-versus-feature assignment matrix as in Figure 4.4. Using these values of $\mathcal{A}_{ij}$ alone, one could assign each track T_i to the feature F_j that maximizes $\mathcal{A}_{ij}$; this maximization would be performed over a single row in the matrix (Figure 4.4a). However, one would also like to take into account information from the rest of the matrix. For example, if there are two possible assignments for track T_i that have similar probabilities and if another track also has a high probability of being assigned one of those features, then it is more likely that the first track should be assigned to the other feature even if that assignment had a slightly lower initial probability. Such a situation often arises in practice when one particle moves towards the previous position of another particle. To take this type of competition for features into

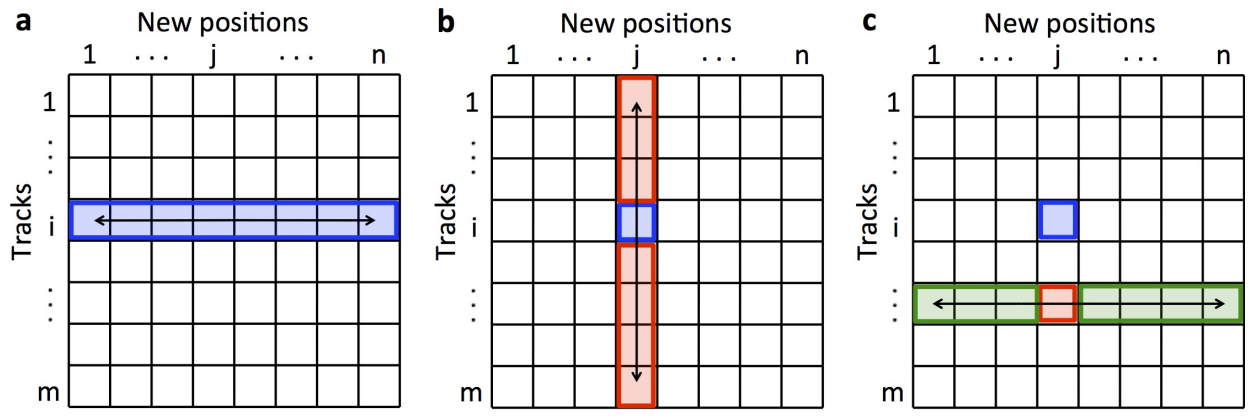

Figure 4.4: **(a)** Schematic of an assignment matrix containing relative assignment probabilities between a set of existing tracks and the positions of newly-detected features in the current frame of an image series. The simplest assignment for a track T_i would be to the feature with the maximum value in the relevant row of the matrix (blue). **(b)** The proposed update scheme would update each value in the original row based on the other values in the same column (red), which indicate the level of competition for feature F_j from the other tracks. **(c)** The two-layer update scheme first updates each value in the relevant column based on the other values in the same row (green), which indicate the level of competition for this particular competing track from the other features.

account, we use the following procedure to update the entries of the matrix. For each value $\mathcal{A}_{ij}$ in the assignment matrix, we update its value by weighting it relative to the sum of all probabilities in the same column of the matrix (Figure 4.4b), as follows,

$$\mathcal{A}'_{ij} \leftarrow \mathcal{A}_{ij} \cdot \frac{\mathcal{A}_{ij}}{\sum_k \mathcal{A}_{kj}} \,. \tag{4.5}$$

The relative probability of assignment to a particular feature F_j is now higher for features that have no other likely candidate tracks that could be assigned to them, and lower for features that do have other competing tracks. The set of new values $\mathcal{A}'_{ij}$ can then be maximized for each track T_i to choose a final assignment.

However, the relative assignment probabilities will improve further if we repeat this updating process by going a layer deeper into the matrix. Conceptually, if we consider two tracks T_i and $T_{i'}$ that have competing probabilities of assignment to a new feature F_j, the

degree to which each of these tracks competes for F_j should be reduced if the track has another feature that it could be assigned to with a similarly high probability. Therefore, before using the probabilities $\mathcal{A}_{kj}$ in Equation 4.5, we can update them using a similar procedure as above, weighting them relative to the sum of all probabilities in row k of the matrix (Figure 4.4c),

$$\mathcal{A}'_{kj} \leftarrow \mathcal{A}_{kj} \cdot \frac{\mathcal{A}_{kj}}{\sum_l \mathcal{A}_{kl}} \,. \tag{4.6}$$

The overall update to the initial probabilities $\mathcal{A}_{ij}$ is thus,

$$\mathcal{A}'_{ij} \leftarrow \mathcal{A}_{ij} \cdot \frac{\mathcal{A}_{ij}}{\sum_k \left(\frac{\mathcal{A}^2_{kj}}{\sum_l \mathcal{A}_{kl}} \right)} \,. \tag{4.7}$$

Note that this assignment probability updating scheme is nested and could be repeated additional times to capture more information on the competing tracks and features in the matrix. In practice, we find that the two-level update in Equation 4.7 is sufficient to correctly assign features to tracks in a range of tested datasets, including multiple movies of chromosomes and fluorescent beads in the starfish oocyte system (Section 4.3), in vitro movies of the kinetochore Ska1 complex diffusing on microtubules (Section 4.4), and other datasets including movies of kinetochore and spindle dynamics in live cells.

4.3 Analysis of a novel actin-based transport mechanism in starfish oocytes

4.3.1 Biological background

Actin-based contractility orchestrates changes in cell shape underlying cellular functions ranging from division to migration and wound healing [Pollard and Cooper, 2009; Pollard, 2010; Eggert et al., 2006; Insall and Machesky, 2009; Bement et al., 2007]. Actin

also functions in intracellular transport, with the prevailing view that filamentous actin (F-actin) cables serve as tracks for motor-driven transport of cargo [Pollard and Cooper, 2009; Ross et al., 2008]. An alternate mode of intracellular transport was recently discovered in starfish oocytes involving a contractile F-actin meshwork that mediates chromosome congression [Lénárt et al., 2005]. Oocytes have an exceptionally large nucleus (the germinal vesicle) that stores nuclear proteins for early embryonic divisions (Figure 4.5a) [Lénárt and Ellenberg, 2003]. As a consequence, specialized transport mechanisms are required to deliver chromosomes that are initially distributed throughout the nuclear space to the assembling meiotic spindle. In starfish oocytes, an actin-dependent process transports chromosomes that are initially scattered in the ∼80 mm diameter nucleus to within capture range of centrosomal microtubule asters at the cell cortex (or animal pole, AP) (Figure 4.5b). This process, which is essential to prevent chromosome loss and aneuploidy of the egg, involves an extensive F-actin meshwork (Figure 4.6a) that forms in the nuclear space and decreases in volume toward the AP during chromosome congression [Lénárt et al., 2005]. However, the

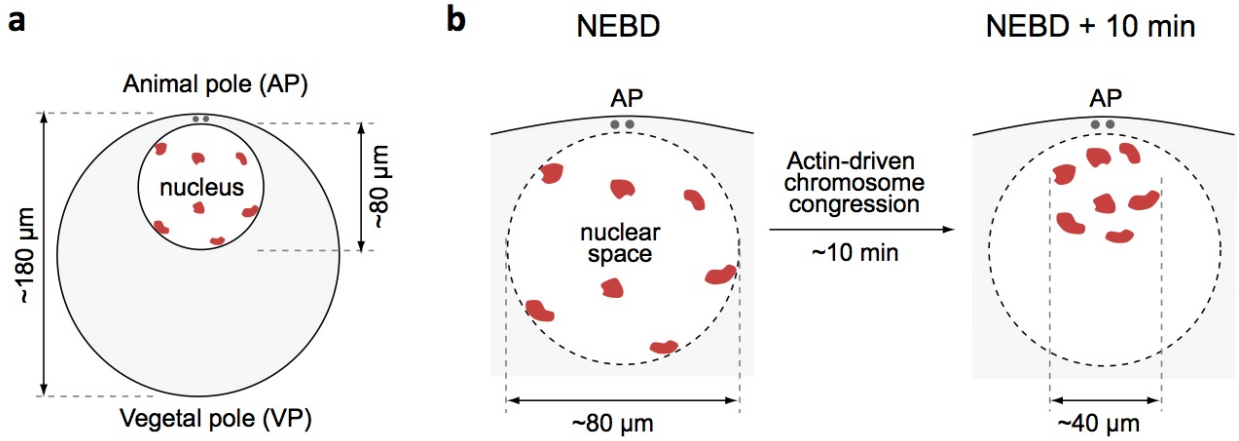

Figure 4.5: **(a)** Schematic of an immature starfish oocyte, showing the nucleus anchored at the animal pole (AP) and chromosomes scattered throughout the nuclear volume. **(b)** Schematic of the process of actin-driven chromosome congression that begins after nuclear envelope breakdown (NEBD), and during which the nuclear space occupied by chromosomes shrinks to approximately one-half of its original diameter. Dashed lines indicate the position of the nuclear envelope (NE) prior to NEBD. Figure previously published in [Mori et al., 2011].

specific mechanism of action of this novel mode of actin-based transport remained unknown. In particular, the organization and spatial distribution of contractile activity within the F-actin meshwork, the origin of the directionality of transport, and the mechanism by which this directed motion is transduced to chromosomal cargo remained important unanswered questions.

4.3.2 Quantitative analysis of chromosome motion

To resolve the mechanism of chromosome transport by F-actin, we first characterized chromosome trajectories during the actin-dependent phase of chromosome congression. High-resolution movies of fluorescently labeled chromosomes were tracked using the algorithm described in Section 4.2 (Figure 4.6b,c). Trajectories exhibited two previously identified

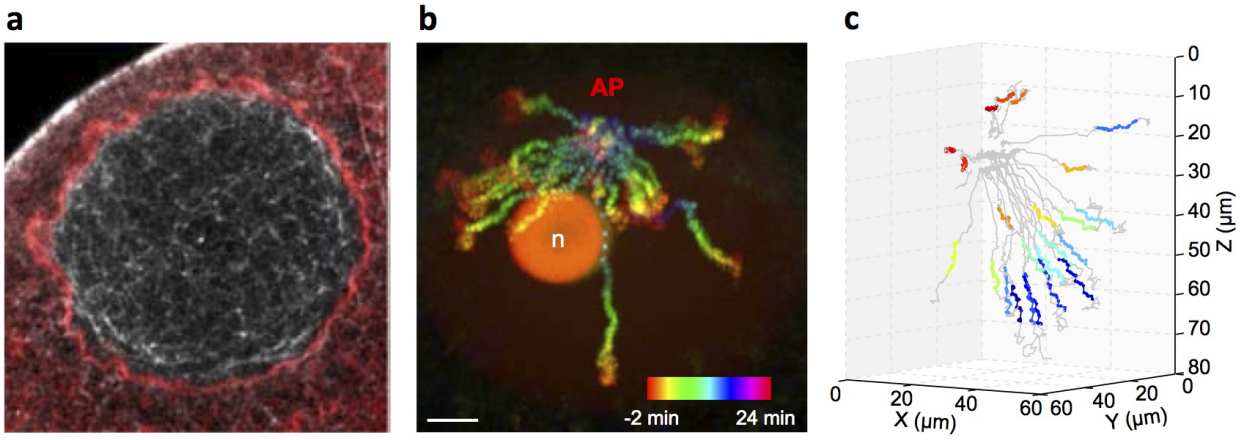

Figure 4.6: **(a)** Single confocal section through the nuclear region of an oocyte expressing UtrCH- 3mEGFP (to label F-actin; gray) and injected with $DiIC_{18}$ [Eggert et al., 2006] (to label endomembranes; red). The F-actin meshwork forms in the nuclear space defined by the remnant NE membranes. **(b)** Pseudo-colored time projection of all maximum intensity z-projections through the nuclear region of an oocyte expressing H2B-mCherry. 27 z-sections were acquired every 15 s. H2B-mCherry also labels the nucleolus (n) that disassembles after NEBD. Time is relative to NEBD, scale bar: 10 μm. **(c)** Trajectories of tracked chromosomes from the 3D dataset shown in (b). Trajectories are colored during the actin-driven transport phase (the color of each trajectory is consistent with Figure 4.7). Figure previously published in [Mori et al., 2011].

phases of poleward motion: an initial slow, actin-driven phase that begins just after nuclear envelope breakdown (NEBD) followed by a faster, microtubule-driven phase that begins ∼10 min after NEBD [Lénárt et al., 2005] (Figure 4.7a). Analysis of the actin-driven phase of motion using MSD-Bayes confirmed that the overall motion of the chromosomes contains both diffusive and direction components, as shown in Chapter 2. Therefore, the diffusion plus flow model (Equation 1.6) was used to analyze the individual chromosome trajectories independently to look for differences in motion parameters within the population. Fits of this model to MSD curves from three example actin-phase chromosome trajectories are shown in Figure 4.7b.

4.3.3 Chromosome velocities reveal homogenous meshwork contraction

Analysis of the chromosome velocity parameters revealed a novel and unexpected property of this actin-driven transport process: the constant poleward speed of each chromosome depends linearly on its initial distance from the AP (Figure 4.7c). This property rules out a number of models that could have explained contraction of the F-actin meshwork, including localized contractile activity at the AP, which would result in equal poleward speeds for all chromosomes. Instead, poleward chromosome speeds that depend linearly on initial distance from the AP imply that contractile activity is distributed homogeneously throughout the F-actin meshwork. An important corollary prediction of homogeneous contraction is that any two points in the meshwork (e.g., any pair of chromosomes) should exhibit a constant relative speed of travel toward one another during the congression process and that this speed should depend linearly on their initial separation distance. Analysis of pairwise chromosome approach velocities confirmed this prediction (Figure 4.7d). Furthermore, the fact that this is true for all pairs of chromosomes irrespective of their initial location in the nuclear space implies that the contraction is isotropic; i.e., it does not have an intrinsic, preferred directionality. A schematic of the homogeneous contraction model, illustrating

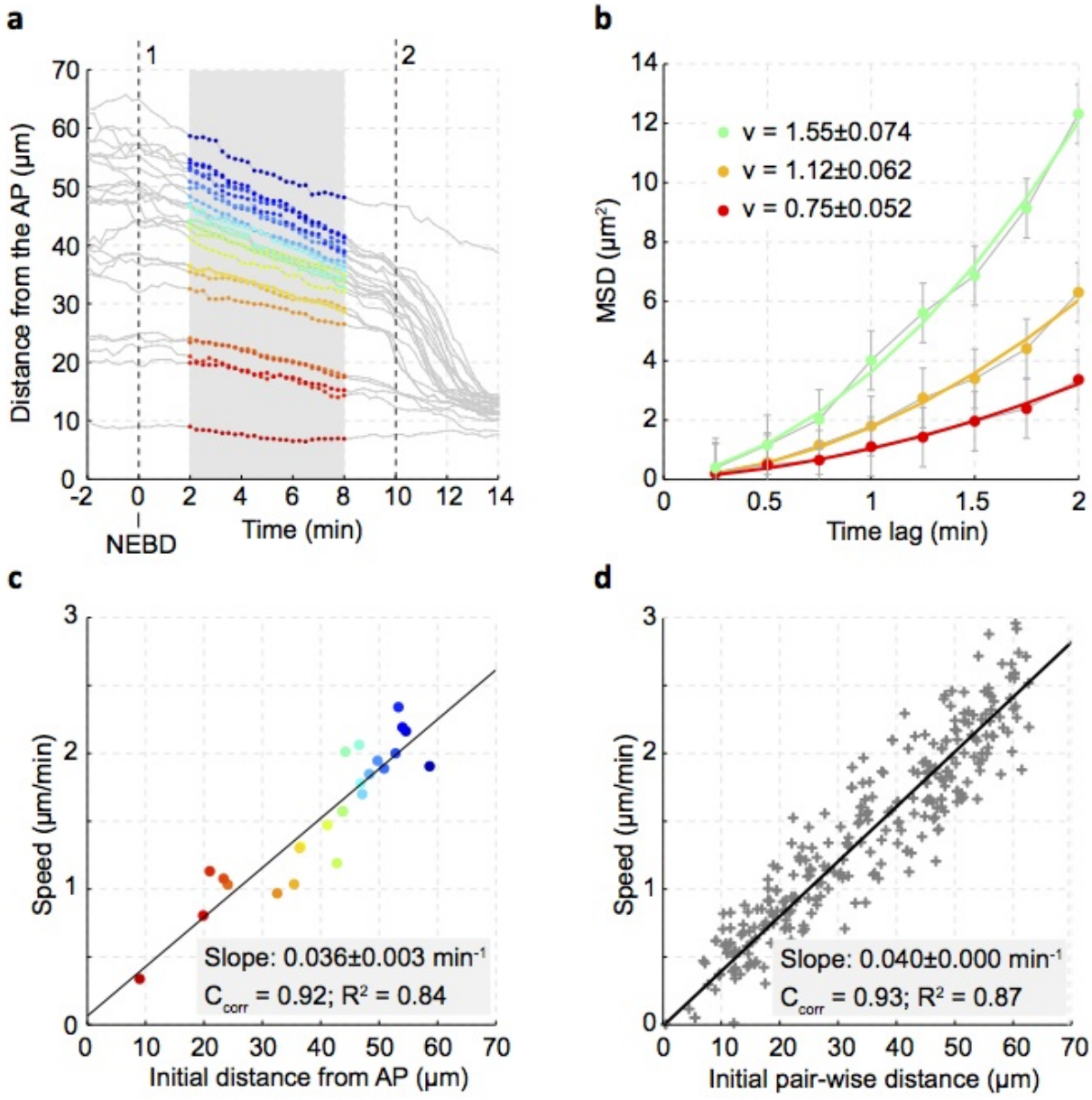

Figure 4.7: **(a)** Trajectories in Figure 4.6c plotted as distance from the AP versus time. (1) Start of slow actin-driven transport; (2) start of capture by microtubules. The shaded and colored portions of the trajectories (from 28 min after NEBD) are used in subsequent analyses of the actin-driven transport phase. **(b)** Mean-squared displacement curves from the trajectories of three example chromosomes. Raw MSD values (gray) were fit with the diffusion plus flow model (colored curves). **(c)** Dependence of pole-ward chromosome speeds on initial chromosome distance from the AP. C_{corr}: correlation coefficient. **(d)** Dependence of pair-wise chromosome approach speeds on the initial distance between each pair of chromosomes. Figure previously published in [Mori et al., 2011].

how it gives rise to speeds that scale with initial distance from an anchor point, is shown in Figure 4.8.

4.3.4 Actin dynamics support the homogenous contraction model

To test whether motion of the F-actin meshwork is consistent with the homogeneous contraction model suggested by chromosome trajectories, we imaged F-actin at high resolution in live oocytes by using the utrophin calponin homology domain (UtrCH [Burkel et al., 2007]). Kymograph analysis reveals that F-actin bundles form throughout the nuclear space 0-2 min after NEBD and subsequently begin to flow in a directed manner toward the AP (Figure 4.9a). As the meshwork flows toward the AP, it is continuously replenished by bundles originating at the nuclear envelope (NE) remnants (Figure 4.9a, arrowheads).

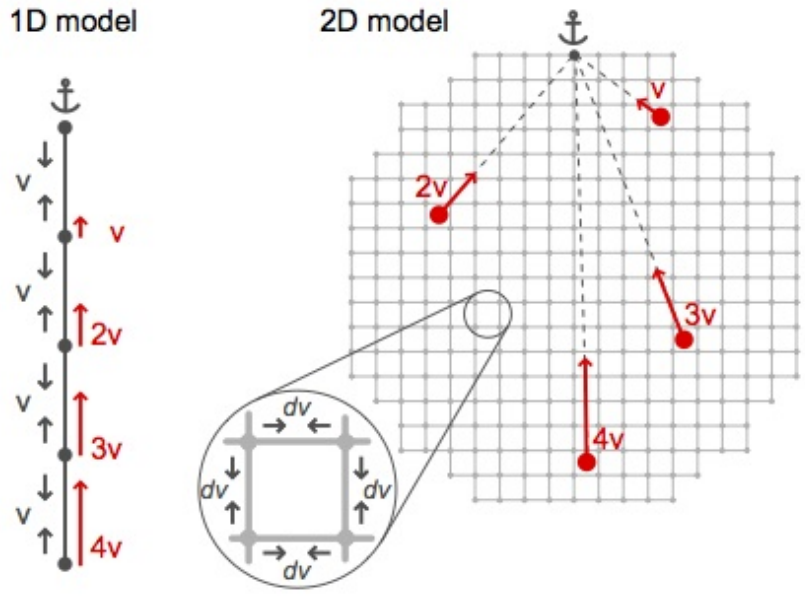

Figure 4.8: Schematic of the homogeneous contraction model, illustrating the dependence of speed towards a fixed point (anchor) on distance from the anchor. Homogeneously distributed contractile activity is represented as contractile elements (lines between nodes). The regular arrangement of the contractile elements in the 2D model is only for visualization purposes and is not a specific feature of the model. Figure previously published in [Mori et al., 2011].

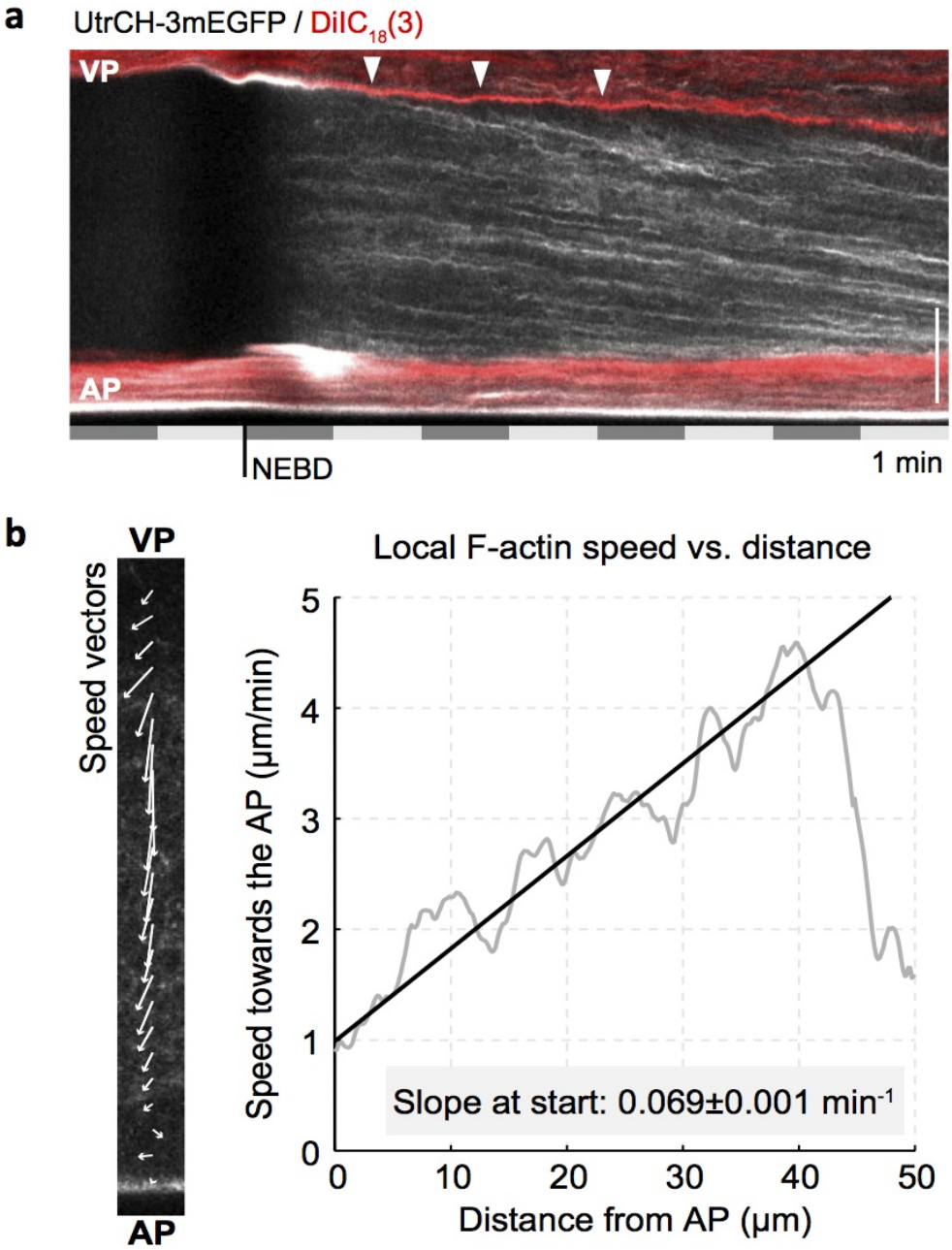

Figure 4.9: **(a)** Kymograph along the animal-vegetal axis of an oocyte expressing the same markers as in Figure 4.6a. Arrowheads mark some of the new actin structures produced at the membrane boundary. **(b)** Image correlation spectroscopy (ICS) analysis using a one-minute time interval during the middle of actin-driven transport (4-5 min after NEBD) and a sliding 50x50 pixel (7.5x7.5 μm) template. Left: white arrows show measured flow velocities overlaid on a selected frame from the time series. Right: pole-ward component of the measured velocities plotted as a function of distance from the AP. The slope of the linear correlation is extrapolated to the start of the actin-driven transport phase. Figure previously published in [Mori et al., 2011].

Local F-actin velocities along an axis from the ventral to animal pole measured using image correlation spectroscopy (ICS) confirm that the meshwork speed increases linearly with distance from the AP (Figure 4.9b). These quantitative observations are fully consistent with the homogeneous and isotropic meshwork contraction model. Furthermore, the slope of the velocity-distance dependence measured for the F-actin meshwork (0.069 min^{-1}) is within the range of values obtained from chromosome trajectories in different oocytes (0.067 +/- 0.025 min^{-1}), thus suggesting that this homogeneous meshwork contraction directly drives chromosome transport.

4.3.5 Transport direction is determined by cortical anchoring

Next, we asked how homogeneous, isotropic contraction of the F-actin meshwork is converted into the observed asymmetric, directional transport toward the AP. The center of mass of a homogeneously contracting meshwork will move in a directional manner if one side of the meshwork is attached to a fixed point (Figure 4.8). Thus, we hypothesized that mechanical anchoring of the F-actin meshwork to cortical F-actin would be sufficient to drive net meshwork transport toward the cortex. We tested this prediction by tracking chromosome motions in oocytes that had been centrifuged in order to relocate the nucleus away from the cortex [Matsuura and Chiba, 2004] and thereby remove any cortical anchoring (Figure 4.10). In support of the cortical anchoring model, relocating the nucleus to the center of the oocyte resulted in symmetric transport of chromosomes to the center of the nuclear region (Figure 4.10b). Interestingly, centrifugation of the nucleus to within ~5 mm of the cortex opposite the AP restored directionality: chromosomes were transported toward the nearest point on the cortex (Figure 4.10c). This result indicates that the AP, centrosomes, and microtubules are not required for asymmetric transport, which is determined solely by proximity of the nucleus to the cortex. A simple explanation is that the anchor consists of actin filaments that physically connect the meshwork to cortical F-actin. Importantly, in

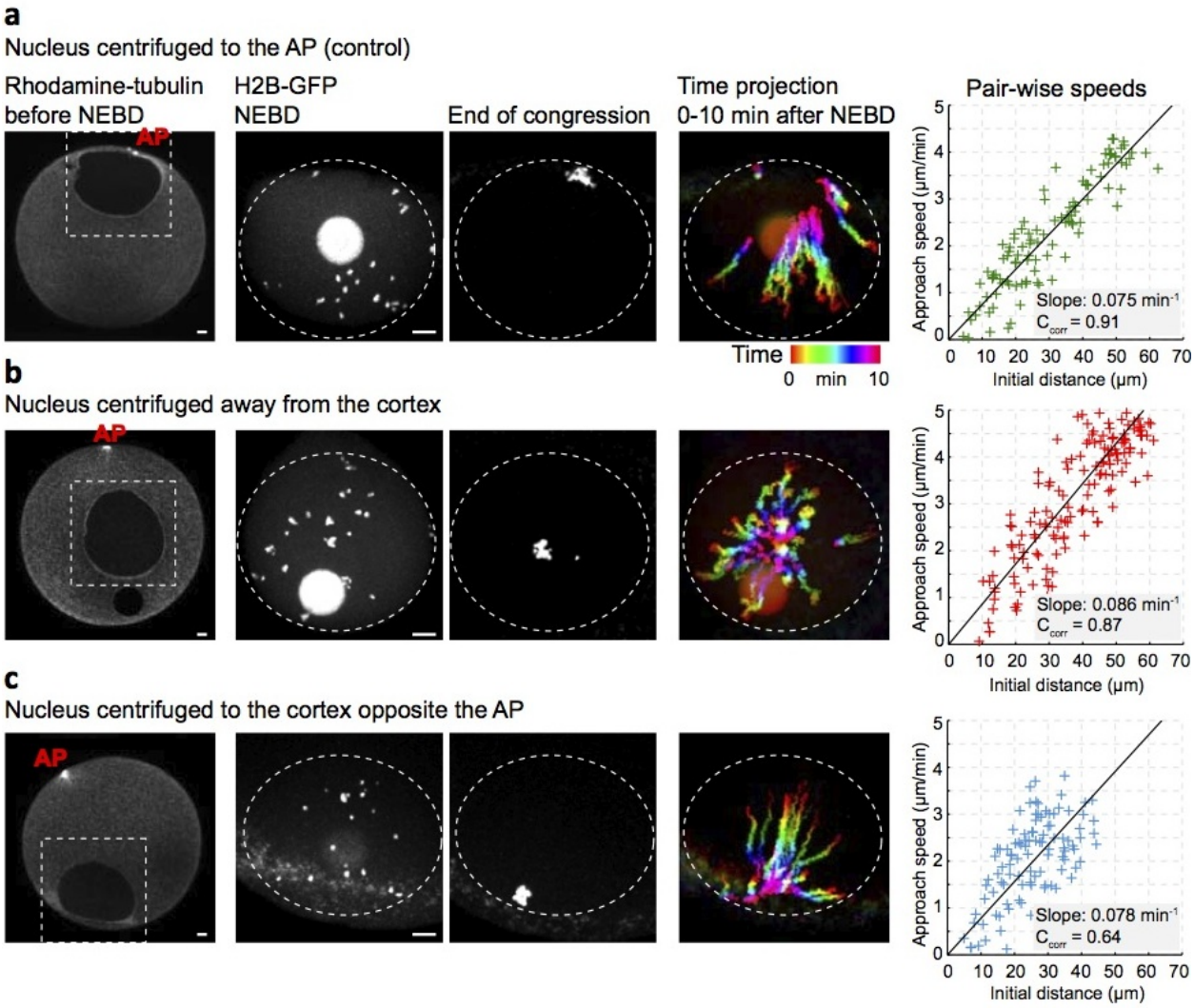

Figure 4.10: Oocytes expressing H2B-GFP and injected with rhodamine-tubulin for identification of the AP were centrifuged in order to relocate the nucleus either back towards the AP as a control **(a)**, to the center of the cell **(b)**, or to the opposite cortex **(c)**. From left: low magnification images of rhodamine-tubulin before NEBD (the smaller dark circle in (b) is the negative image of the oil drop used for injection); maximum intensity z-projections of the nuclear region, marked by a dashed rectangle in the left-most columns, showing H2B-GFP labeled chromosomes at NEBD and at the end of chromosome transport (dashed ellipses label the initial position of the NE); pseudo-colored time projections of z-projections during chromosome transport; and pair-wise chromosome approach velocities versus initial pair-wise separation distance, as in Figure 4.7d. Figure previously published in [Mori et al., 2011].

all of these centrifugation experiments, pairwise chromosome approach velocities obtained from automated tracking of the chromosomes are still linearly dependent on their initial separation distances and the slopes of this dependence are indistinguishable from control oocytes. Taken together, these results support the proposed model in which mechanical anchoring to the cell cortex converts homogeneous contraction of the F-actin meshwork into asymmetric, directed transport.

4.3.6 Transport of inert particles is size-dependent

Finally, we sought to answer the question of how the contracting meshwork transduces its motion to chromosomes in order to transport them to the AP. It was previously observed that chromosomes develop dense F-actin structures in their vicinity that could potentially serve to attach them to the meshwork via specific binding interactions [Lénárt et al., 2005]. However, our new high-resolution data reveal that these dense structures are specific to chromosomes located near the nuclear envelope, and many chromosomes scattered in the nuclear space are transported in the absence of any visible, chromosome-specific F-actin structures. This observation suggests that chromosomes may be transported without binding to F-actin simply by steric trapping within the meshwork.

To directly test this hypothesis, we injected a dense polydisperse mixture of inert fluorescent beads and bead aggregates into the oocyte nucleus and imaged them in 3D during chromosome congression. We found that the effective volume occupied by these inert particles decreased over time toward the AP (Figure 4.11a), indicating that they were transported by the F-actin meshwork. Moreover, the degree of the volume decrease was particle-size dependent: the effective radius of the space occupied by three different bead size groups decreased linearly over time at a rate that increased with particle size (Fig-

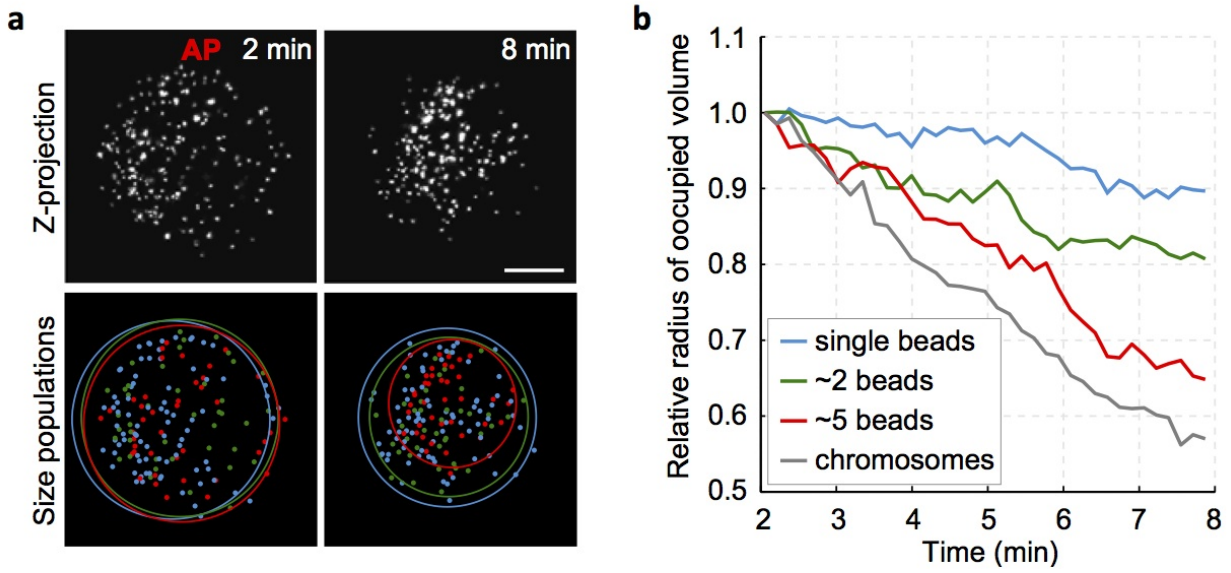

Figure 4.11: **(a)** Top: maximum intensity projections of 20 z-sections showing single and aggregated 0.4-μm diameter PEG-coated fluorescent beads injected into the nucleus of an oocyte. Time is given relative to NEBD. Bottom: particles identified and color-coded based on size categories (defined based on their relative intensities) and the effective radius of the occupied volume for each category (calculated as the radius of gyration of each set of particles). Scale bar: 20 μm. **(b)** Radius of the occupied volume of the set of particles in each size group plotted over time (each point is an average over five time frames). The gray line shows the same analysis performed on chromosomes that were labeled by H1-Alexa647 in the same experiment. Figure previously published in [Mori et al., 2011].

ure 4.11b). Importantly, the space occupied by the largest particle group (with an estimated average diameter of 0.7 μm) decreased with a rate approaching that of chromosomes ($\sim$2 μm diameter) imaged in the same cell. To analyze the transport of these large particles, we injected oocytes with large bead aggregates at sufficiently low density to track them using the tracking algorithm described in Section 4.2 above (Figure 4.12a). The resulting bead trajectories revealed a behavior strikingly similar to that of transported chromosomes (Figure 4.12b). Bead aggregates synchronously initiate directed poleward transport shortly after NEBD with poleward speeds that remain approximately constant during transport. In addition, their poleward and relative pairwise approach speeds are linearly dependent on their initial separation distance (Figure 4.12c) and have a slope that is indistinguishable

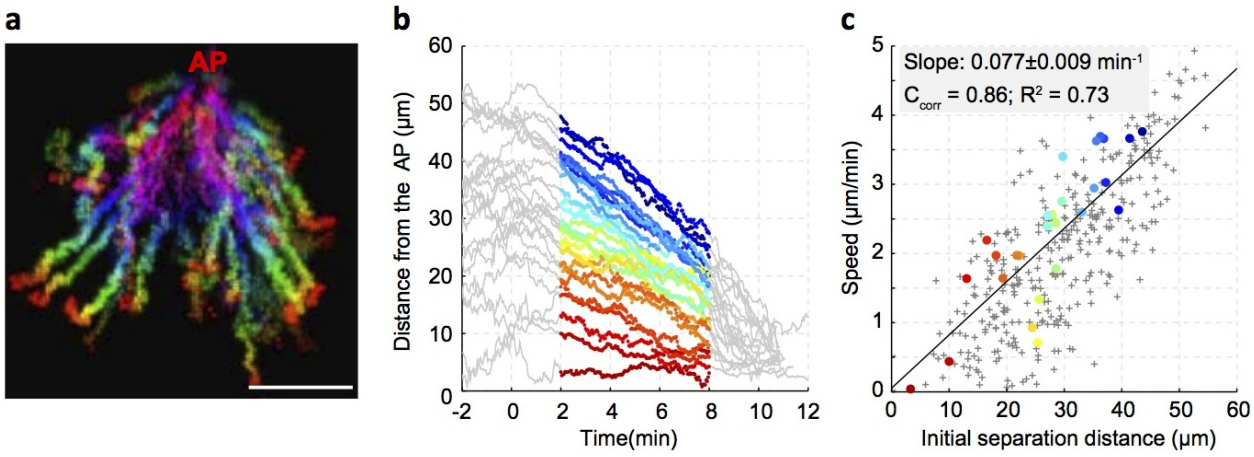

Figure 4.12: **(a)** Pseudo-colored time projection of z-projections of an oocyte injected with aggregates of 0.5 μm diameter PEG-coated fluorescent beads. Scale bar: 20 μm. **(b)** Distance from the AP versus time for the trajectories obtained from the dataset shown in (a). The actin-driven phase is highlighted; colors are consistent with (c). See Figure 4.7 for comparison with chromosomes. **(c)** Pole-ward (colored filled circles) and pair-wise (+) velocities versus initial distance from the AP and initial pair-wise distance, respectively. Figure previously published in [Mori et al., 2011].

from that of chromosomes within measurement error (0.077 min^{-1} compared to 0.067 +/- 0.025 min^{-1}).

These results demonstrate that the F-actin meshwork can effectively transport cargo without specific binding interactions. The fact that transport efficiency depends on cargo size implies that the force of meshwork contraction is transmitted to cargo at least in part by steric trapping within porous cages in the meshwork; smaller particles are more likely than large particles to escape through meshwork pores. The bead experiments indicate that efficient capture requires particles of roughly 1 μm diameter, defining an approximate effective mesh-size that is in good agreement with the spacing between bundles visible in high-resolution images of F-actin bundles that constitute the meshwork. Thus, these visible bundles are likely the structures mediating particle trapping and transport.

4.3.7 Conclusion

In this study we used live-cell imaging together with quantitative analysis of chromosome trajectories and meshwork velocities to show that the 3D F-actin meshwork present during meiosis in starfish oocytes contracts homogeneously and isotropically throughout the nuclear space. Although the intrinsic meshwork contraction lacks any specific directionality, centrifugation experiments revealed that anchoring of the meshwork to the cell cortex confers directionality to its large-scale motion. By injecting inert particles of different sizes, we showed that this directional transport activity is size-selective and transduced to chromosomal cargo at least in part by steric entrapment of particles larger than the effective mesh-size of the meshwork. Taken together, these results reveal mechanistic design principles of a novel and potentially versatile intracellular transport machine that is fundamentally distinct from previously observed mechanisms of F-actin-driven intracellular transport. In this system, force is generated by isotropic contractile activity that is distributed homogeneously throughout the F-actin meshwork, with a rate of contraction that is limited by physical tethering and filament production at remnant nuclear envelope membranes.

In starfish oocytes, this actin-meshwork homogeneous contraction mechanism is utilized for the essential function of transporting chromosomes to the AP. However, it is tempting to speculate that similar design principles may be used in other intracellular transport processes in various organisms and cell types. This transport system is inherently flexible in that anchoring to cellular structures other than the cortex may direct contractile motion to distinct subcellular locations, and tuning the effective mesh-size could allow for selective transport based on size. Alternatively, specific binding interactions between cargo and the meshwork could potentially enhance the rate and reliability of transport. Notably, in contrast to chromosome transport by microtubules, where re-establishment of severed microtubule-chromosome connections requires significant time and may result in

chromosome loss [Civelekoglu-Scholey and Scholey, 2010], transport by a space-filling F-actin meshwork may be considerably more robust because of physical entrapment within an extended meshwork.

Elucidating the detailed molecular mechanisms underlying each of the functional components that we have identified in this novel intracellular transport machine is an important goal of future work. In particular, our finding that contractile activity is homogeneously distributed throughout the actin meshwork suggests that the meshwork may be organized into quasi-independent contractile F-actin subunits, as recently also proposed for the contractile ring of *C. elegans* embryos [Carvalho et al., 2009] and stress fibers in cultured mammalian cells [Colombelli et al., 2009]. How the individual bundles forming the meshwork are organized to generate contractile force and whether this force is generated by acto-myosin contractility, depolymerization and/or bundling of actin filaments independent of motor activity [Carvalho et al., 2009; Sun et al., 2010; Shlomovitz and Gov, 2008; Zumdieck et al., 2007], or some combination of each remain open questions. It is an intriguing possibility that contractile units similar in structure and composition to those that form flat networks under the cell membrane to mediate cytokinesis or cell migration may alternatively organize into 3D F-actin meshworks to drive intracellular transport.

4.4 Analysis of Ska1-complex dynamics on microtubules

4.4.1 Biological background

As described in Section 4.3, the function of the F-actin meshwork during meiosis in starfish oocytes is to transport chromosomes through the large nuclear volume until they are within capture distance of spindle microtubules, which are ultimately responsible for

segregating the chromosomes into daughter cells. Transport of chromosomes by microtubules is mediated by a specific force-generating connection between the microtubules and the chromosomal DNA called the kinetochore complex, which involves more than 100 different proteins in human cells and assembles upon centromeric DNA [Cheeseman and Desai, 2008]. The driving force for chromosome segregation comes primarily from microtubule depolymerization, which releases potential energy stored in the polymerized tubulin subunits [McIntosh et al., 2010; Desai and Mitchison, 1997; Wang and Nogales, 2005; Grishchuk et al., 2005]. However, it is unclear how kinetochores remain associated with the depolymerizing ends of the spindle microtubules during this dynamic process.

To understand these kinetochore-microtubule interactions in detail, it is essential to characterize the components of the kinetochore complex that mediate the interactions. A central player is the conserved Ndc80 complex [Cheeseman et al., 2006]. Loss of Ndc80 function results in catastrophic defects in kinetochore-microtubule attachments [DeLuca et al., 2002]. The Ndc80 complex has been shown to remain associated with polymerizing and depolymerizing ends of microtubules when it has been artificially oligomerized on the surface of a microsphere [Powers et al., 2009]; however, monomeric Ndc80 complex lacks this activity. Therefore, the protein(s) responsible for maintaining the kinetochore-microtubule attachment during depolymerization remains an open question. Another recently-identified key mediator of kinetochore-microtubule interactions is the Ska1 complex [Daum et al., 2009; Gaitanos et al., 2009; Hanisch et al., 2006; Raaijmakers et al., 2009; Theis et al., 2009; Welburn et al., 2009]. The Ska1 protein contains a microtubule binding domain that is required for the formation of robust kinetochore-microtubule attachments in human cells [Schmidt et al., 2012], and depletion of the full Ska1 complex results in a checkpoint-dependent mitotic arrest with misaligned chromosomes. Here we analyze in detail the in vitro interactions of the Ska1 complex with microtubules.

4.4.2 Diffusion of Ska1 complex along microtubules in vitro

We visualized both the human and *C. elegans* Ska1 complexes containing GFP-Ska1 using total internal reflection fluorescence microscopy (TIR-FM). Both forms of the GFP-Ska1 complex readily diffused on microtubules, similar to the Ndc80-GFP complex (Figure 4.13). Note that the Ndc80 complex used here (Ndc80 "Broccoli") consists of well-behaved truncated versions of Ndc80 and Nuf2 and behaves identically to full-length Ndc80 complex [Schmidt et al., 2012]. The Ska1 and Ndc80 complexes in these images were identified and tracked using the tracking algorithm described in Section 4.2. Comparison of the intensity distributions between the two human complexes indicates that hNdc80 "Broccoli"-GFP is mostly monomeric when bound to microtubules, consistent with previous studies [Powers et al., 2009], while the GFP-hSka1 complex contains two distinct peaks in

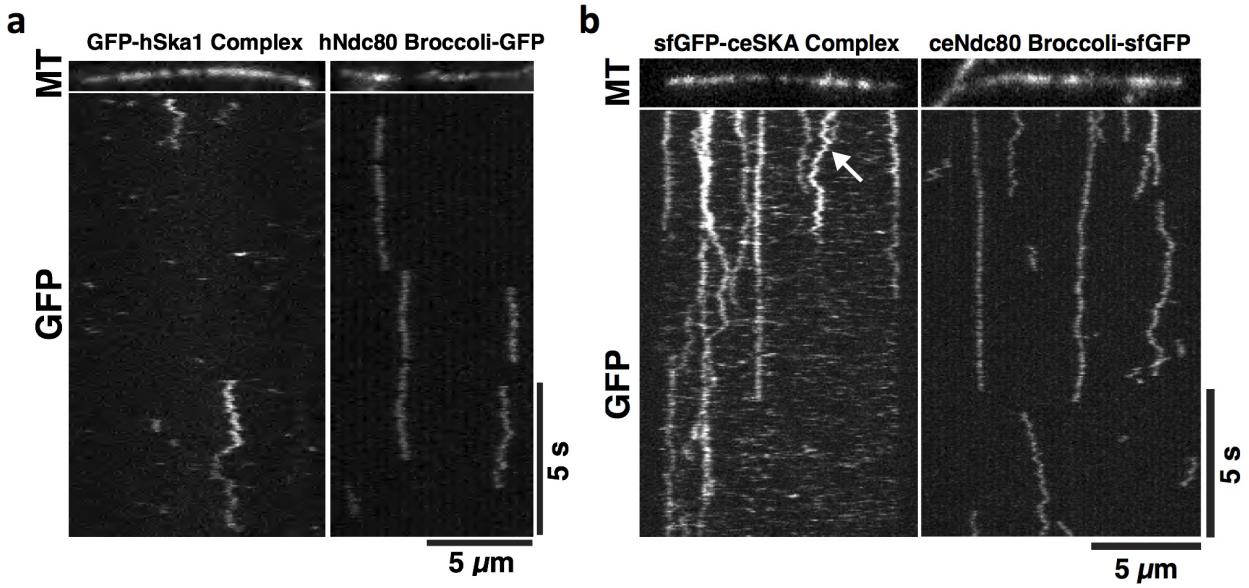

Figure 4.13: **(a)** Representative kymographs with microtubule position along the horizontal axis and time along the vertical axis showing one-dimensional diffusion of GFP-hSka1 complex (left, 100 pM) and Ndc80 "Broccoli"-GFP (right, 50 pM) on taxol-stabilized microtubules labeled with HiLyte 647. **(b)** Representative kymographs as in (a) showing one-dimensional diffusion of sfGFP-ceSka1 complex (left, 4 nM) and ceNdc80 "Broccoli"-sfGFP (right, 1 nM) on taxol-stabilized microtubules. Figure previously published in [Schmidt et al., 2012].

its intensity distribution (Figure 4.14a). These peaks correspond to particles with a mean brightness similar to the Ndc80-GFP complex and particles that have a 2-fold increased intensity, suggesting that GFP-hSka1 complex can bind to microtubules as either a monomer or dimer. The presence of bleaching steps with 2:1 intensity ratios in a number of GFP-hSka1 trajectories confirmed that these brighter complexes contain exactly two GFP-Ska1 molecules (Figure 4.14b). This result is consistent with previous biochemical analysis [Welburn et al.,

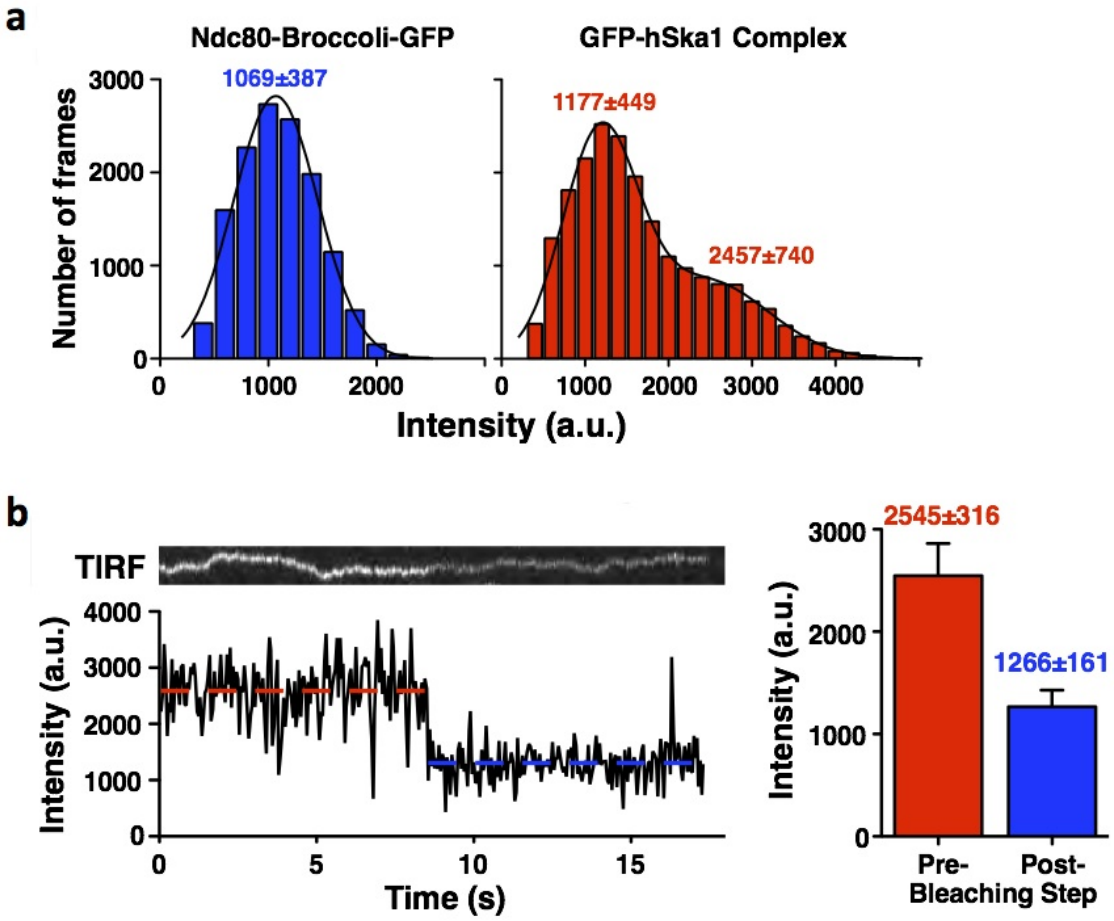

Figure 4.14: **(a)** Intensity distribution of 100 pM GFP-hSka1 complex (red, n = 20676) and 50 pM Ndc80 "Broccoli"-GFP (blue, n = 13411) at each time frame, determined by fitting a 2D Gaussian function to each spot identified by the tracking algorithm, ignoring first and last frames of tracks. **(b)** Photo-bleaching analysis of Ska1 complex particles. Left: kymograph and corresponding fluorescent intensity for one photobleaching event. Right: comparison of the average intensity before and after photobleaching events (n = 32). Figure previously published in [Schmidt et al., 2012].

2009] and recent structural work [Jeyaprakash et al., 2012] indicating that the hSka1 complex can dimerize to form a complex with two microtubule-binding sites. MSD curves generated from the GFP-hSka1 complex and Ndc80 "Broccoli"-GFP trajectories were found to be best explained by a simple diffusion model using the MSD-Bayes approach described in Chapter 2. The diffusion coefficients were 0.09 μm^2/s and 0.03 μm^2/s, respectively (Figure 4.15a), consistent with previous observations made for the Ndc80 complex [Powers et al., 2009].

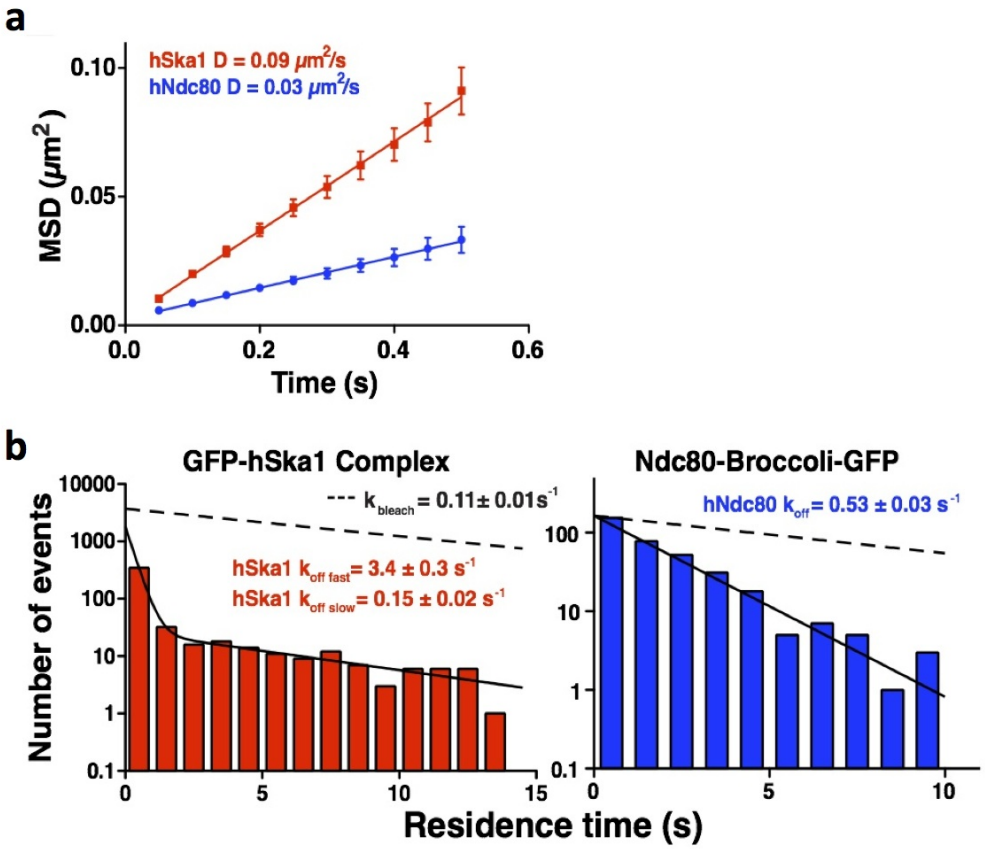

Figure 4.15: **(a)** Mean-squared displacement (MSD, mean and SEM) plotted against time for 100 pM GFP-hSka1 complex (red, n = 187) and 50 pM Ndc80 "Broccoli"-GFP (blue, n = 258). Diffusion coefficients were calculated from the slope of the linear fit of the MSD using the 1D pure diffusion model. **(b)** Distribution of the residence times of GFP-hSka1 complex (100 pM, n = 502) and hNdc80 "Broccoli"-GFP (50 pM, n = 359) trajectories on taxol stabilized microtubules. Figure previously published in [Schmidt et al., 2012].

The residence time distribution of the GFP-hSka1 complex trajectories on microtubules displays biphasic dissociation kinetics (Figure 4.15b): a rapid phase with a rate constant of $k_{\text{fast}} = 3.4$ +/- 0.3 s^{-1} and a slow phase with a rate constant of $k_{\text{slow}} = 0.15$ +/- 0.02 s^{-1}. The presence of these two phases supports the above observation that hSka1 exists in both monomer and dimer forms. The disassociation rate of the rapid phase is similar to the behavior of monomeric GFP-hSka1/2 ($k_{\text{off}} = 4.9$ +/- 0.4 s^{-1}), suggesting that the rapid phase represents hSka1 complex monomers. Indeed, analysis of short-lived (< 0.6 s) hSka1 complex binding events indicated that these were primarily monomeric based on their intensity distribution. Ndc80 Brocolli-GFP has a dissociation constant of $k_{\text{off}} = 0.53$ +/- 0.03 s^{-1} (Figure 4.15b). All measured dissociation rate constants were faster than the bleaching rate constant determined for immobilized Ndc80-GFP under identical imaging conditions ($k_{\text{bleach}} = 0.11$ +/- 0.01 s^{-1}).

In contrast to the hSka1 complex, the *C. elegans* GFP-ceSKA1 complex was primarily monomeric under the tested conditions. We note that sfGFP-ceSKA1 was used in these assays at 40-fold higher concentrations (4 nM) relative to GFP-hSka1 complex (100 pM); yet even at these concentrations, the ceSKA1 complex primarily contains one ceSKA1 subunit. The human (D = 0.09 $\mu m^2/s$) and *C. elegans* (D = 0.11 $\mu m^2/s$) Ska1 complex both diffused on microtubules with similar diffusion coefficients, whereas the *C. elegans* Ndc80 complex (D = 0.08 $\mu m^2/s$) diffuses more rapidly than its human counterpart (D = 0.03 $\mu m^2/s$).

4.4.3 Conclusion

The analyses above demonstrate that the human Ska1 complex associates with microtubules in vitro and suggest that a cooperative interaction between two Ska1-contaning complexes is required for this association to persist. The Ska1 complexes found on micro-

tubules consist of two populations, one with one copy of Ska1 and one with two copies of Ska1, which correlate with the biphasic dissociation rates observed for the Ska1 complex trajectories as they diffuse along microtubules. The long-lived associations of the Ska1 complex dimers with microtubules persist longer than the associations of the Ndc80 complex with microtubules, suggesting that Ska1 complex in the kinetochore in vivo may be a more robust mediator of kinetochore-microtubule interactions than Ndc80. In addition, we find that the Ska1 complex diffuses along the microtubule with a higher diffusion coefficient than the Ndc80 complex. These in vitro observations confirm the ability of the Ska1 complex to form persistent interactions with microtubules and point to an essential role for this complex in forming the robust kinetochore-microtubule interactions that are required for the force of microtubule depolymerization to be conveyed to chromosomes during cell division.

Appendix A

Theory of hidden Markov models

A.1 Formulation of a hidden Markov model

A hidden Markov model is a type of Bayesian network that consists of a Markov chain of hidden variables, denoted by a vector $\mathbf{s} = [s_1, ..., s_T] = \{s_t\}_{t=1}^{T}$, where each s_t can take on one of a discrete set of values (or states),

$$s_t \in \{S_i\}_{i=1}^{K} \ , \tag{A.1}$$

where K is the number of available states [Ewens and Grant, 2005]. The Markov property for the hidden states requires that the probability of being in a given state S_j at time t depends only on the previous state at time $t-1$; given knowledge of s_{t-1}, the probability distribution for s_t is conditionally independent of the previous states prior to time $t-1$. This conditional probability distribution is defined as,

$$\phi_{ij} \equiv P(s_t = S_j | s_{t-1} = S_i) \ , \tag{A.2}$$

and the probability of starting in state S_i at time $t = 1$ is,

$$\pi_i \equiv P(s_1 = S_i) \ . \tag{A.3}$$

Since these state variables are hidden, we do not directly observe their values but instead observe at each time t a secondary random variable e_t (emission) whose value depends only on the state at time t [Ewens and Grant, 2005]. The sequence of observed emissions is denoted by a vector $\mathbf{e} = [e_1, ..., e_T] = \{e_t\}_{t=1}^{T}$. When the emissions are continuous random variables, their distribution is given for each state as a continuous conditional probability distribution function,

$$p_i(e_t) \equiv P(e_t|s_t = S_i) \ , \tag{A.4}$$

parameterized by some set of parameters $\boldsymbol{\beta}$ that can take distinct values for the different states S_i. The full set of parameter values for an HMM, denoted $\boldsymbol{\theta}$, therefore includes the $K \times K$ matrix of transition probabilities ϕ_{ij}, the length-K vector of starting probabilities π_i, and the $K \times |\boldsymbol{\beta}|$ matrix of parameters for the emission probability distributions p_i, where the emission parameters for state S_i are denoted $\boldsymbol{\beta}_i$,

$$\boldsymbol{\theta} = \left[\{\pi_i\}_{i=1}^{K}, \{\phi_{ij}\}_{i,j=1}^{K}, \{\boldsymbol{\beta}_i\}_{i=1}^{K}\right] . \tag{A.5}$$

A.2 Inference on hidden Markov models

A.2.1 Maximum likelihood hidden state sequence (Viterbi algorithm)

Given a sequence of observed emissions $\mathbf{e}$, one would often like to infer the most likely sequence of hidden states $\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} P(\mathbf{s}|\mathbf{e}, \boldsymbol{\theta})$ that could have generated those emissions [Ewens and Grant, 2005]. The likelihood $P(\mathbf{s}|\mathbf{e}, \boldsymbol{\theta})$ can be expanded using Bayes' rule,

$$P(\mathbf{s}|\mathbf{e}, \boldsymbol{\theta}) = \frac{P(\mathbf{e}|\mathbf{s}, \boldsymbol{\theta})P(\mathbf{s}|\boldsymbol{\theta})}{P(\mathbf{e}|\boldsymbol{\theta})} \propto P(\mathbf{e}|\mathbf{s}, \boldsymbol{\theta})P(\mathbf{s}|\boldsymbol{\theta}) \ . \tag{A.6}$$

The probabilities $P(\mathbf{e}|\mathbf{s},\boldsymbol{\theta})$ and $P(\mathbf{s}|\boldsymbol{\theta})$ can be calculated for a given set of parameter values $\boldsymbol{\theta}$ using the recursive structure of the HMM,

$$P(\mathbf{s}|\boldsymbol{\theta}) = P(s_1|\boldsymbol{\theta}) \prod_{t=2}^{T} P(s_t|s_{t-1},\boldsymbol{\theta}) = \pi_{s_1} \prod_{t=2}^{T} \phi_{s_{t-1}s_t} \ , \tag{A.7}$$

$$P(\mathbf{e}|\mathbf{s},\boldsymbol{\theta}) = \prod_{t=1}^{T} P(e_t|s_t,\boldsymbol{\theta}) = \prod_{t=1}^{T} p_{s_t}(e_t) \ . \tag{A.8}$$

The full equation for $\hat{\mathbf{s}}$ then becomes,

$$\begin{aligned} \hat{\mathbf{s}} = \arg\max_{\mathbf{s}} P(\mathbf{s}|\mathbf{e},\boldsymbol{\theta}) &= \arg\max_{\mathbf{s}} P(\mathbf{e}|\mathbf{s},\boldsymbol{\theta}) P(\mathbf{s}|\boldsymbol{\theta}) \\ &= \arg\max_{\mathbf{s}} \left(\pi_{s_1} \prod_{t=2}^{T} \phi_{s_{t-1}s_t} \prod_{t=1}^{T} p_{s_t}(e_t) \right) . \end{aligned} \tag{A.9}$$

This maximization can be performed exactly using the Viterbi algorithm. Although with K states, there are K^T possible sequences of hidden states $\mathbf{s}$, the Viterbi algorithm takes advantage of the tree-like structure of an HMM to perform this maximization recursively in $O(K^2T)$ time [Ewens and Grant, 2005].

The Viterbi algorithm defines a message $V_t(i)$, calculated sequentially at each time point, which is equal to the maximum likelihood (over the possible hidden state sequences through time $t-1$) of ending in state S_i at time t and having observed all of the emissions from times $1, ..., t$ [Ewens and Grant, 2005]. Specifically,

$$V_t(i) = \max_{[s_1,...,s_{t-1}]} P(s_t = S_i, [e_1, ..., e_t] | [s_1, ..., s_{t-1}], \boldsymbol{\theta}) \ . \tag{A.10}$$

For the first time point there are no previous hidden states to maximize over, so $V_1(i)$ corresponds to the probability of being in state S_i and having observed e_1. The equation

above reduces to,

$$\begin{aligned} V_1(i) &= P(s_1 = S_i, e_1|\boldsymbol{\theta}) \\ &= P(e_1|s_1 = S_i, \boldsymbol{\theta})P(s_1 = S_i|\boldsymbol{\theta}) \\ &= p_i(e_1)\pi_i \ . \end{aligned} \tag{A.11}$$

Evaluating Equation A.10 for additional time points leads to the following recursive relationship for any t [Ewens and Grant, 2005],

$$V_t(i) = p_i(e_t) \max_{i'} \left(\phi_{i'i} V_{t-1}(i')\right) \ . \tag{A.12}$$

This message $V_t(i)$ can thus be computed recursively for each $t \in \{1, ..., T\}$.

At the end of the chain, the message $V_T(i)$ represents the probability of the most likely state sequence that ends with $s_T = S_i$ and that produced all of the observations $\mathbf{e} = [e_1, ..., e_T]$. So a final maximization over the states S_i for time T gives the probability of the maximum likelihood state sequence $\hat{\mathbf{s}}$,

$$P(\hat{\mathbf{s}}|\boldsymbol{\theta}) = \max_i V_T(i) \ . \tag{A.13}$$

To recover the sequence itself, the information on which state S_i was preferred in each of the maximization steps above is saved and can be traced back from the final maximization at time T to the first maximization at time 1 to construct the maximum likelihood state sequence [Ewens and Grant, 2005].

A.2.2 Maximum likelihood parameter values

The above calculation assumes that the parameter values $\boldsymbol{\theta}$ are known; however, the maximum likelihood parameter values $\hat{\boldsymbol{\theta}}$ can also be inferred from the observed sequence $\mathbf{e}$ using a similar application of Bayes' rule,

$$P(\boldsymbol{\theta}|\mathbf{e}) = \frac{P(\mathbf{e}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{e})} \propto P(\mathbf{e}|\boldsymbol{\theta})P(\boldsymbol{\theta}) \; . \tag{A.14}$$

If the prior on parameters $P(\boldsymbol{\theta})$ is uniform, then we only need to maximize $P(\mathbf{e}|\boldsymbol{\theta})$, the likelihood of the data given a set of parameters. However, the likelihood of the data must be marginalized over all possible state sequences, since the state sequence is unobserved, as follows,

$$P(\mathbf{e}|\boldsymbol{\theta}) = \sum_{\mathbf{s}} P(\mathbf{e}|\mathbf{s}, \boldsymbol{\theta})P(\mathbf{s}|\boldsymbol{\theta}) \; . \tag{A.15}$$

The probabilities $P(\mathbf{e}|\mathbf{s}, \boldsymbol{\theta})$ and $P(\mathbf{s}|\boldsymbol{\theta})$ are the same as given above in Equations A.7 and A.8. So the full equation for $\hat{\boldsymbol{\theta}}$ is,

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} \sum_{\mathbf{s}} P(\mathbf{e}|\mathbf{s}, \boldsymbol{\theta})P(\mathbf{s}|\boldsymbol{\theta}) \\ &= \arg\max_{\boldsymbol{\theta}} \sum_{\mathbf{s}} \left(\pi_{s_1} \prod_{t=2}^{T} \phi_{s_{t-1}s_t} \prod_{t=1}^{T} p_{s_t}(e_t) \right) \; . \end{aligned} \tag{A.16}$$

Evaluating $\hat{\boldsymbol{\theta}}$ involves two steps, summing the likelihood of observing the emissions over all possible hidden state sequences, and maximizing this sum over possible parameter values. The maximization over $\boldsymbol{\theta}$ is typically intractable and must be solved using approximate or numerical methods such as expectation-maximization (EM) or Monte Carlo (MC) sampling. The summation over hidden state sequences $\mathbf{s}$, however, can be performed exactly for a given set of parameters $\boldsymbol{\theta}$ using a variant of the Viterbi algorithm above called the forward

algorithm, which replaces the maximization steps with summations to compute the total marginal probability of observing the emissions $\mathbf{e}$.

The forward algorithm defines a message $F_t(i)$, calculated sequentially at each time point, which is equal to the total likelihood of ending in state S_i at time t and having observed all of the emissions from times $1, ..., t$ [Ewens and Grant, 2005]. Specifically,

$$F_t(i) = P(s_t = S_i, [e_1, ..., e_t]|\boldsymbol{\theta}) \ . \tag{A.17}$$

At the first time point, $F_1(i)$ is simply the probability that $s_1 = S_i$ and that e_1 was observed, and the equation above reduces to,

$$\begin{aligned} F_1(i) &= P(s_1 = S_i, e_1|\boldsymbol{\theta}) \\ &= P(e_1|s_1 = S_i, \boldsymbol{\theta})P(s_1 = S_i|\boldsymbol{\theta}) \\ &= p_i(e_1)\pi_i \ . \end{aligned} \tag{A.18}$$

As in the case of the Viterbi algorithm, the forward message follows a recursive relationship for any t [Ewens and Grant, 2005],

$$F_t(i) = p_i(e_t) \sum_{i'} \phi_{i'i} F_{t-1}(i') \ . \tag{A.19}$$

This looks the same as the Viterbi recursion in Equation A.12, except that the maximization is replaced by a summation over states. This message $F_t(i)$ can thus be computed recursively for each $t \in \{1, ..., T\}$. At the end of the chain, the message $F_T(i)$ represents the probability that $s_T = S_i$ and that all of the emissions $\mathbf{e} = [e_1, ..., e_T]$ have been observed. So a final

summation over the states S_i for time T,

$$P(\mathbf{e}|\boldsymbol{\theta}) = \sum_i F_T(i) \ , \tag{A.20}$$

gives the full probability of observing the emissions $\mathbf{e}$ marginalized over the hidden states $\mathbf{s}$. This method is used to evaluate the likelihood $P(\mathbf{e}|\boldsymbol{\theta})$ at any given value of the parameters $\boldsymbol{\theta}$.

Appendix B

Monte Carlo methods

B.1 Principles of Monte Carlo integration

Monte Carlo integration involves sampling the value of an integrand $f(\boldsymbol{\theta})$ at many different $\boldsymbol{\theta}$ values, which are distributed according to some sampling distribution $q(\boldsymbol{\theta})$. This sampling distribution can be uniform (the quadrature method) or non-uniform with higher density in regions where $f(\boldsymbol{\theta})$ is large (importance sampling). When $\boldsymbol{\theta}$ is sampled over any arbitrary distribution $q(\boldsymbol{\theta})$, the integral $\int f(\boldsymbol{\theta})d\boldsymbol{\theta}$ is equal to [Robert and Casella, 2004],

$$I \equiv \int f(\boldsymbol{\theta})d\boldsymbol{\theta} = \int \frac{f(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} q(\boldsymbol{\theta})d\boldsymbol{\theta} = \mathbb{E}_q\left[\frac{f(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right], \tag{B.1}$$

where the final equality is simply the definition of the expectation taken over the sampling distribution $q(\boldsymbol{\theta})$. Therefore our estimator for the integral I is simply the mean of the ratio $f(\boldsymbol{\theta})/q(\boldsymbol{\theta})$ calculated over a set of N sampled points in parameter space, where the sampled points are distributed according to $q(\boldsymbol{\theta})$,

$$\hat{I} = \left\langle \frac{f(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right\rangle_q = \frac{1}{N}\sum_{n=1}^{N} \frac{f(\boldsymbol{\theta}^{(n)})}{q(\boldsymbol{\theta}^{(n)})}. \tag{B.2}$$

It can be shown that the variance in the estimator $\hat{I}$ is minimized when the sampling distribution $q(\boldsymbol{\theta})$ is proportional to $f(\boldsymbol{\theta})$ [Robert and Casella, 2004].

B.2 Markov Chain Monte Carlo (MCMC) to sample a target distribution

An efficient method for generating $\boldsymbol{\theta}$ values from a distribution $q(\boldsymbol{\theta})$ proportional to $f(\boldsymbol{\theta})$ is to use a Markov chain with $q(\boldsymbol{\theta})$ as its stationary distribution [Robert and Casella, 2004; Grimmett and Stirzaker, 2001]. This approach requires generating sequential values of $\boldsymbol{\theta}$ with a transition probability $A(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}')$ of going from an old parameter set $\boldsymbol{\theta}$ to a new parameter set $\boldsymbol{\theta}'$ that is consistent with the target stationary distribution $q(\boldsymbol{\theta})$. At equilibrium, the detailed balance condition for Markov chains states that the average number of transitions from state $\boldsymbol{\theta}$ to state $\boldsymbol{\theta}'$ should be equal to the average number of transitions from state $\boldsymbol{\theta}'$ to state $\boldsymbol{\theta}$ [Grimmett and Stirzaker, 2001]. This condition is equivalent to,

$$q(\boldsymbol{\theta})A(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}') = q(\boldsymbol{\theta}')A(\boldsymbol{\theta}' \rightarrow \boldsymbol{\theta}) \; . \tag{B.3}$$

Since the target stationary distribution $q(\boldsymbol{\theta})$ is proportional to $f(\boldsymbol{\theta})$, we can re-write this condition as,

$$f(\boldsymbol{\theta})A(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}') = f(\boldsymbol{\theta}')A(\boldsymbol{\theta}' \rightarrow \boldsymbol{\theta}) \; . \tag{B.4}$$

The transition probability A can be split into two components [Robert and Casella, 2004; Grimmett and Stirzaker, 2001], a proposal (or trial) function $T(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}')$, which is the probability of proposing a trial move to $\boldsymbol{\theta}'$ from a current state $\boldsymbol{\theta}$, and an acceptance function $\alpha(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}')$, which is the probability of accepting a proposed move from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}'$,

$$A(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}') = \alpha(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}')T(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}') \; . \tag{B.5}$$

If we choose T to be symmetric, such that,

$$T(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}') = T(\boldsymbol{\theta}' \rightarrow \boldsymbol{\theta}) \; , \tag{B.6}$$

then the detailed balance condition becomes,

$$\alpha(\boldsymbol{\theta} \to \boldsymbol{\theta}')f(\boldsymbol{\theta}) = \alpha(\boldsymbol{\theta}' \to \boldsymbol{\theta})f(\boldsymbol{\theta}') \ , \tag{B.7}$$

which implies,

$$\frac{\alpha(\boldsymbol{\theta} \to \boldsymbol{\theta}')}{\alpha(\boldsymbol{\theta}' \to \boldsymbol{\theta})} = \frac{f(\boldsymbol{\theta}')}{f(\boldsymbol{\theta})} \ . \tag{B.8}$$

There are multiple options for $T(\boldsymbol{\theta} \to \boldsymbol{\theta}')$ and $\alpha(\boldsymbol{\theta} \to \boldsymbol{\theta}')$ that satisfy the above conditions. A common choice for these distributions is the random walk Metropolis approach [Grimmett and Stirzaker, 2001], in which the proposal function T defines a random walk along each parameter axis θ_i in parameter space. The random walk proposal function T can be either a uniform distribution,

$$\theta_i' \sim \text{unif}\big(\theta_i - \delta, \theta_i + \delta\big) \ , \tag{B.9}$$

or a normal distribution,

$$\theta_i' \sim \mathcal{N}\big(\theta_i, \delta^2\big) \ . \tag{B.10}$$

Either distribution is parameterized by some δ, which controls the magnitude by which a parameter θ_i can change at each step in the Markov chain. The Metropolis approach then chooses the acceptance probability α to satisfy the detailed balance condition above by setting it equal to [Gilks, 1995; Robert and Casella, 2004; Grimmett and Stirzaker, 2001],

$$\begin{aligned} \alpha(\boldsymbol{\theta} \to \boldsymbol{\theta}') &= \begin{cases} 1, & f(\boldsymbol{\theta}') \geq f(\boldsymbol{\theta}) \\ \dfrac{f(\boldsymbol{\theta}')}{f(\boldsymbol{\theta})}, & f(\boldsymbol{\theta}') < f(\boldsymbol{\theta}) \end{cases} \\ &= \min\left(1, \frac{f(\boldsymbol{\theta}')}{f(\boldsymbol{\theta})}\right) . \end{aligned} \tag{B.11}$$

This form for α means that all moves that increase $f(\boldsymbol{\theta})$ are accepted, while moves that decrease $f(\boldsymbol{\theta})$ are accepted with probability $f(\boldsymbol{\theta}')/f(\boldsymbol{\theta})$. The parameter δ in the proposal function above is typically chosen to give an overall acceptance rate $T(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}')\alpha(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}')$ of 30-50 percent [Roberts et al., 1997].

The steps of the Markov chain Monte Carlo (MCMC) integration approach can be summarized as follows:

1. Start with an initial guess of the parameters, $\boldsymbol{\theta}^{(0)}$.

2. Choose a parameter or parameters to move at each MCMC iteration. For example, all parameters can be moved at once, a single randomly-selected parameter can be moved as in the original Metropolis algorithm, or a subset (block) of parameters can be moved together [Gilks, 1995].

3. Propose a move for the selected parameter(s). For each selected parameter θ_i, choose the proposed new value θ_i' from either a uniform distribution (Equation B.9) or a normal distribution (Equation B.10). Choose δ such that the acceptance rate in step 4 is between 0.3 and 0.5 [Roberts et al., 1997].

4. Choose whether to accept or reject the proposed move by evaluating the function $f(\boldsymbol{\theta})$ at the old and new parameter values and accepting the move with probability, $\alpha = \min\left(1, \frac{f(\boldsymbol{\theta}')}{f(\boldsymbol{\theta})}\right)$. Note that any proposed move that takes a bounded parameter outside of its bounds is automatically rejected.

5. Repeat steps 2-4 until the distribution of sampled parameter values reaches equilibrium. This equilibrium distribution should approach the target distribution $q(\boldsymbol{\theta})$.

References

Adrian, R. (1991). Particle-imaging techniques for experimental fluid mechanics. *Annual Review of Fluid Mechanics*, 23(1):261–304.

Arcizet, D., Meier, B., Sackmann, E., Rädler, J. O., and Heinrich, D. (2008). Temporal analysis of active and passive transport in living cells. *Phys Rev Lett*, 101(24):248103.

Bement, W. M., Yu, H.-Y. E., Burkel, B. M., Vaughan, E. M., and Clark, A. G. (2007). Rehabilitation and the single cell. *Curr Opin Cell Biol*, 19(1):95–100.

Benhamou, S. (2006). Detecting an orientation component in animal paths when the preferred direction is individual-dependent. *Ecology*, 87(2):518–28.

Berg, H. (1993). *Random Walks in Biology*. Princeton University Press.

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–5.

Bomzon, Z., Knight, M. M., Bader, D. L., and Kimmel, E. (2006). Mitochondrial dynamics in chondrocytes and their connection to the mechanical properties of the cytoplasm. *J Biomech Eng*, 128(5):674–9.

Bormuth, V., Varga, V., Howard, J., and Schäffer, E. (2009). Protein friction limits diffusive and directed movements of kinesin motors on microtubules. *Science*, 325(5942):870–3.

Bouzigues, C. and Dahan, M. (2007). Transient directed motions of GABA(A) receptors in growth cones detected by a speed correlation index. *Biophys J*, 92(2):654–60.

Brandenburg, B. and Zhuang, X. (2007). Virus trafficking - learning from single-virus tracking. *Nat Rev Microbiol*, 5(3):197–208.

Brangwynne, C. P., Koenderink, G. H., MacKintosh, F. C., and Weitz, D. A. (2008). Cytoplasmic diffusion: molecular motors mix it up. *J Cell Biol*, 183(4):583–7.

Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, Jr, R. L., and Wiggins, C. H. (2009). Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys J*, 97(12):3196–205.

Burkel, B. M., von Dassow, G., and Bement, W. M. (2007). Versatile fluorescent probes for actin filaments based on the actin-binding domain of utrophin. *Cell Motil Cytoskeleton*, 64(11):822–32.

Cairo, C. W., Das, R., Albohy, A., Baca, Q. J., Pradhan, D., Morrow, J. S., Coombs, D., and Golan, D. E. (2010). Dynamic regulation of CD45 lateral mobility by the spectrin-ankyrin cytoskeleton of T cells. *J Biol Chem*, 285(15):11392–401.

Cairo, C. W., Mirchev, R., and Golan, D. E. (2006). Cytoskeletal regulation couples LFA-1 conformational changes to receptor lateral mobility and clustering. *Immunity*, 25(2):297–308.

Carlin, B. and Louis, T. (2009). *Bayesian methods for data analysis*. Chapman & Hall/CRC.

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., Golland, P., and Sabatini, D. M. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol*, 7(10):R100.

Carvalho, A., Desai, A., and Oegema, K. (2009). Structural memory in the contractile ring makes the duration of cytokinesis independent of cell size. *Cell*, 137(5):926–37.

Casella, G. and Berger, R. (2001). *Statistical Inference*. Duxbury Press, 2nd edition.

Caspi, A., Granek, R., and Elbaum, M. (2000). Enhanced diffusion in active intracellular transport. *Phys Rev Lett*, 85(26 Pt 1):5655–8.

Cheeseman, I. M., Chappie, J. S., Wilson-Kubalek, E. M., and Desai, A. (2006). The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. *Cell*, 127(5):983–97.

Cheeseman, I. M. and Desai, A. (2008). Molecular architecture of the kinetochore-microtubule interface. *Nat Rev Mol Cell Biol*, 9(1):33–46.

Cheezum, M. K., Walker, W. F., and Guilford, W. H. (2001). Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophys J*, 81(4):2378–88.

Chuang, C.-H., Carpenter, A. E., Fuchsova, B., Johnson, T., de Lanerolle, P., and Belmont, A. S. (2006). Long-range directional movement of an interphase chromosome site. *Curr Biol*, 16(8):825–31.

Chung, I., Akita, R., Vandlen, R., Toomre, D., Schlessinger, J., and Mellman, I. (2010). Spatial control of EGF receptor activation by reversible dimerization on living cells. *Nature*, 464(7289):783–7.

Civelekoglu-Scholey, G. and Scholey, J. M. (2010). Mitotic force generators and chromosome segregation. *Cell Mol Life Sci*, 67(13):2231–50.

Colombelli, J., Besser, A., Kress, H., Reynaud, E. G., Girard, P., Caussinus, E., Haselmann, U., Small, J. V., Schwarz, U. S., and Stelzer, E. H. K. (2009). Mechanosensing in actin stress fibers revealed by a close correlation between force and protein localization. *J Cell Sci*, 122(Pt 10):1665–79.

Comaniciu, D., Ramesh, V., and Meer, P. (2003). Kernel-based object tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):564–577.

Condamin, S., Bénichou, O., Tejedor, V., Voituriez, R., and Klafter, J. (2007). First-passage times in complex scale-invariant media. *Nature*, 450(7166):77–80.

Condamin, S., Tejedor, V., Voituriez, R., Bénichou, O., and Klafter, J. (2008). Probing microscopic origins of confined subdiffusion by first-passage observables. *Proc Natl Acad Sci U S A*, 105(15):5675–80.

Coscoy, S., Huguet, E., and Amblard, F. (2007). Statistical analysis of sets of random walks: how to resolve their generating mechanism. *Bull Math Biol*, 69(8):2467–92.

Cremer, T. and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2(4):292–301.

Das, R., Cairo, C. W., and Coombs, D. (2009). A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. *PLoS Comput Biol*, 5(11):e1000556.

Daum, J. R., Wren, J. D., Daniel, J. J., Sivakumar, S., McAvoy, J. N., Potapova, T. A., and Gorbsky, G. J. (2009). Ska3 is required for spindle checkpoint silencing and the maintenance of chromosome cohesion in mitosis. *Curr Biol*, 19(17):1467–72.

Daumas, F., Destainville, N., Millot, C., Lopez, A., Dean, D., and Salomé, L. (2003). Confined diffusion without fences of a g-protein-coupled receptor as revealed by single particle tracking. *Biophys J*, 84(1):356–66.

DeLuca, J. G., Moree, B., Hickey, J. M., Kilmartin, J. V., and Salmon, E. D. (2002). hNuf2 inhibition blocks stable kinetochore-microtubule attachment and induces mitotic cell death in HeLa cells. *J Cell Biol*, 159(4):549–55.

Desai, A. and Mitchison, T. J. (1997). Microtubule polymerization dynamics. *Annu Rev Cell Dev Biol*, 13:83–117.

Eggert, U. S., Mitchison, T. J., and Field, C. M. (2006). Animal cytokinesis: from parts list to mechanisms. *Annu Rev Biochem*, 75:543–66.

Ehrlich, M., Boll, W., Van Oijen, A., Hariharan, R., Chandran, K., Nibert, M. L., and Kirchhausen, T. (2004). Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell*, 118(5):591–605.

Elting, M. W., Bryant, Z., Liao, J.-C., and Spudich, J. A. (2011). Detailed tuning of structure and intramolecular communication are dispensable for processive motion of myosin VI. *Biophys J*, 100(2):430–9.

Ewens, W. and Grant, G. (2005). *Statistical Methods in Bioinformatics*. Springer, 2nd edition.

Fernández-González, R., Muñoz-Barrutia, A., Barcellos-Hoff, M. H., and Ortiz-de Solorzano, C. (2006). Quantitative in vivo microscopy: the return from the 'omics'. *Curr Opin Biotechnol*, 17(5):501–10.

Flyvbjerg, H. and Petersen, H. (1989). Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91(1):461–466.

Gaitanos, T. N., Santamaria, A., Jeyaprakash, A. A., Wang, B., Conti, E., and Nigg, E. A. (2009). Stable kinetochore-microtubule interactions depend on the Ska complex and its new component Ska3/C13Orf3. *EMBO J*, 28(10):1442–52.

Gardner, M. K., Pearson, C. G., Sprague, B. L., Zarzar, T. R., Bloom, K., Salmon, E. D., and Odde, D. J. (2005). Tension-dependent regulation of microtubule dynamics at kinetochores can explain metaphase congression in yeast. *Mol Biol Cell*, 16(8):3764–75.

Gilks, W. (1995). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, 1st edition.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press, 3rd edition.

Grishchuk, E. L., Molodtsov, M. I., Ataullakhanov, F. I., and McIntosh, J. R. (2005). Force production by disassembling microtubules. *Nature*, 438(7066):384–8.

Guo, S.-M., He, J., Monnier, N., Sun, G., Wohland, T., and Bathe, M. (2012). Bayesian approach to the analysis of fluorescence correlation spectroscopy data II: application to simulated and in vitro data. *Anal Chem*, 84(9):3880–8.

Hanisch, A., Silljé, H. H. W., and Nigg, E. A. (2006). Timely anaphase onset requires a novel spindle and kinetochore complex comprising Ska1 and Ska2. *EMBO J*, 25(23):5504–15.

Haritaoglu, I., Harwood, D., and Davis, L. (2000). W4: real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830.

He, J., Guo, S.-M., and Bathe, M. (2012). Bayesian approach to the analysis of fluorescence correlation spectroscopy data I: theory. *Anal Chem*, 84(9):3871–9.

Held, M., Schmitz, M. H. A., Fischer, B., Walter, T., Neumann, B., Olma, M. H., Peter, M., Ellenberg, J., and Gerlich, D. W. (2010). CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat Methods*, 7(9):747–54.

Hoze, N., Nair, D., Hosy, E., Sieben, C., Manley, S., Herrmann, A., Sibarita, J.-B., Choquet, D., and Holcman, D. (2012). Heterogeneity of AMPA receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging. *Proc Natl Acad Sci U S A*, 109(42):17052–7.

Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 34(3):334–352.

Huet, S., Karatekin, E., Tran, V. S., Fanget, I., Cribier, S., and Henry, J.-P. (2006). Analysis of transient behavior in complex trajectories: application to secretory vesicle dynamics. *Biophys J*, 91(9):3542–59.

Insall, R. H. and Machesky, L. M. (2009). Actin dynamics at the leading edge: from simple machinery to complex networks. *Dev Cell*, 17(3):310–22.

Jaensch, S., Decker, M., Hyman, A. A., and Myers, E. W. (2010). Automated tracking and analysis of centrosomes in early Caenorhabditis elegans embryos. *Bioinformatics*, 26(12):i13–20.

Jaqaman, K. and Danuser, G. (2006). Linking data to models: data regression. *Nat Rev Mol Cell Biol*, 7(11):813–9.

Jaqaman, K., Kuwata, H., Touret, N., Collins, R., Trimble, W. S., Danuser, G., and Grinstein, S. (2011). Cytoskeletal control of CD36 diffusion promotes its receptor and signaling function. *Cell*, 146(4):593–606.

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., and Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods*, 5(8):695–702.

Jeyaprakash, A. A., Santamaria, A., Jayachandran, U., Chan, Y. W., Benda, C., Nigg, E. A., and Conti, E. (2012). Structural and functional organization of the Ska complex, a key component of the kinetochore-microtubule interface. *Mol Cell*, 46(3):274–86.

Johns, L. M., Levitan, E. S., Shelden, E. A., Holz, R. W., and Axelrod, D. (2001). Restriction of secretory granule motion near the plasma membrane of chromaffin cells. *J Cell Biol*, 153(1):177–90.

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kass, R. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.

Kitajima, T. S., Ohsugi, M., and Ellenberg, J. (2011). Complete kinetochore tracking reveals error-prone homologous chromosome biorientation in mammalian oocytes. *Cell*, 146(4):568–81.

Klar, T. A., Jakobs, S., Dyba, M., Egner, A., and Hell, S. W. (2000). Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc Natl Acad Sci U S A*, 97(15):8206–10.

Kusumi, A., Sako, Y., and Yamamoto, M. (1993). Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of

calcium-induced differentiation in cultured epithelial cells. *Biophys J*, 65(5):2021–40.

Lang, T., Wacker, I., Wunderlich, I., Rohrbach, A., Giese, G., Soldati, T., and Almers, W. (2000). Role of actin cortex in the subplasmalemmal transport of secretory granules in PC-12 cells. *Biophys J*, 78(6):2863–77.

Lawton, A. K., Nandi, A., Stulberg, M. J., Dray, N., Sneddon, M. W., Pontius, W., Emonet, T., and Holley, S. A. (2013). Regulated tissue fluidity steers zebrafish body elongation. *Development*, 140(3):573–82.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Lénárt, P., Bacher, C. P., Daigle, N., Hand, A. R., Eils, R., Terasaki, M., and Ellenberg, J. (2005). A contractile nuclear actin network drives chromosome congression in oocytes. *Nature*, 436(7052):812–8.

Lénárt, P. and Ellenberg, J. (2003). Nuclear envelope dynamics in oocytes: from germinal vesicle breakdown to mitosis. *Curr Opin Cell Biol*, 15(1):88–95.

Letinic, K., Sebastian, R., Barthel, A., and Toomre, D. (2010). Deciphering subcellular processes in live imaging datasets via dynamic probabilistic networks. *Bioinformatics*, 26(16):2029–36.

Lippincott-Schwartz, J. and Patterson, G. H. (2003). Development and use of fluorescent protein markers in living cells. *Science*, 300(5616):87–91.

Manley, S., Gillette, J. M., Patterson, G. H., Shroff, H., Hess, H. F., Betzig, E., and Lippincott-Schwartz, J. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat Methods*, 5(2):155–7.

Martin, D. S., Forstner, M. B., and Käs, J. A. (2002). Apparent subdiffusion inherent to single particle tracking. *Biophys J*, 83(4):2109–17.

Masson, J.-B., Casanova, D., Türkcan, S., Voisinne, G., Popoff, M. R., Vergassola, M., and Alexandrou, A. (2009). Inferring maps of forces inside cell membrane microdomains. *Phys Rev Lett*, 102(4):048103.

Matsuura, R.-k. and Chiba, K. (2004). Unequal cell division regulated by the contents of germinal vesicles. *Dev Biol*, 273(1):76–86.

McIntosh, J. R., Volkov, V., Ataullakhanov, F. I., and Grishchuk, E. L. (2010). Tubulin depolymerization may be an ancient biological motor. *J Cell Sci*, 123(Pt 20):3425–34.

McKenzie, H. W., Lewis, M. A., and Merrill, E. H. (2009). First passage time analysis of animal movement and insights into the functional response. *Bull Math Biol*, 71(1):107–29.

Meijering, E., Smal, I., and Danuser, G. (2006). Tracking in molecular bioimaging. *Signal Processing Magazine, IEEE*, 23(3):46–53.

Michalet, X. (2010). Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys Rev E Stat Nonlin Soft Matter Phys*, 82(4 Pt 1):041914.

Monnier, N., Guo, S.-M., Mori, M., He, J., Lénárt, P., and Bathe, M. (2012). Bayesian approach to MSD-based analysis of particle motion in live cells. *Biophys J*, 103(3):616–26.

Montiel, D., Cang, H., and Yang, H. (2006). Quantitative characterization of changes in dynamical behavior for single-particle tracking studies. *J Phys Chem B*, 110(40):19763–70.

Mori, M., Monnier, N., Daigle, N., Bathe, M., Ellenberg, J., and Lénárt, P. (2011). Intracellular transport by an anchored homogeneously contracting F-actin meshwork. *Curr Biol*, 21(7):606–11.

Ng, Y.-K., Lu, X., Gulacsi, A., Han, W., Saxton, M. J., and Levitan, E. S. (2003). Unexpected mobility variation among individual secretory vesicles produces an apparent refractory neuropeptide pool. *Biophys J*, 84(6):4127–34.

Olivo-Marin, J. (2002). Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989–1996.

Orton, M. and Fitzgerald, W. (2002). A Bayesian approach to tracking multiple targets using sensor arrays and particle filters. *IEEE Transactions on Signal Processing*, 50(2):216–223.

Ouellette, N., Xu, H., and Bodenschatz, E. (2006). A quantitative study of three-dimensional Lagrangian particle tracking algorithms. *Experiments in Fluids*, 40(2):301–313.

Persson, F., Lindén, M., Unoson, C., and Elf, J. (2013). Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat Methods*, 10(3):265–9.

Platani, M., Goldberg, I., Lamond, A. I., and Swedlow, J. R. (2002). Cajal body dynamics and association with chromatin are ATP-dependent. *Nat Cell Biol*, 4(7):502–8.

Pollard, T. D. (2010). Mechanics of cytokinesis in eukaryotes. *Curr Opin Cell Biol*, 22(1):50–6.

Pollard, T. D. and Cooper, J. A. (2009). Actin, a central player in cell shape and movement. *Science*, 326(5957):1208–12.

Posada, D. and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol*, 53(5):793–808.

Powers, A. F., Franck, A. D., Gestaut, D. R., Cooper, J., Gracyzk, B., Wei, R. R., Wordeman, L., Davis, T. N., and Asbury, C. L. (2009). The Ndc80 kinetochore complex forms load-bearing attachments to dynamic microtubule tips via biased diffusion. *Cell*, 136(5):865–75.

Qian, H., Sheetz, M. P., and Elson, E. L. (1991). Single particle tracking. analysis of diffusion and flow in two-dimensional systems. *Biophys J*, 60(4):910–21.

Raaijmakers, J. A., Tanenbaum, M. E., Maia, A. F., and Medema, R. H. (2009). RAMA1 is a novel kinetochore protein involved in kinetochore-microtubule attachment. *J Cell Sci*, 122(Pt 14):2436–45.

Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25:111–163.

Rajani, V., Carrero, G., Golan, D. E., de Vries, G., and Cairo, C. W. (2011). Analysis of molecular diffusion by first-passage time variance identifies the size of confinement zones. *Biophys J*, 100(6):1463–72.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, 2nd edition.

Roberts, G., Gelman, A., and Gilks, W. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120.

Ross, J. L., Ali, M. Y., and Warshaw, D. M. (2008). Cargo transport: molecular motors navigate a complex cytoskeleton. *Curr Opin Cell Biol*, 20(1):41–7.

Rudnick, J. and Gaspari, G. (1987). The shapes of random walks. *Science*, 237(4813):384–9.

Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods*, 3(10):793–5.

Sage, D., Neumann, F., Hediger, F., Gasser, S., and Unser, M. (2005). Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 14(9):1372–1383.

Saxton, M. J. (1993). Lateral diffusion in an archipelago. single-particle diffusion. *Biophys J*, 64(6):1766–80.

Saxton, M. J. (1997). Single-particle tracking: the distribution of diffusion coefficients. *Biophys J*, 72(4):1744–53.

Saxton, M. J. (2008). Single-particle tracking: connecting the dots. *Nat Methods*, 5(8):671–2.

Saxton, M. J. and Jacobson, K. (1997). Single-particle tracking: applications to membrane dynamics. *Annu Rev Biophys Biomol Struct*, 26:373–99.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol*, 4:Article32.

Schmidt, J. C., Arthanari, H., Boeszoermenyi, A., Dashkevich, N. M., Wilson-Kubalek, E. M., Monnier, N., Markus, M., Oberer, M., Milligan, R. A., Bathe, M., Wagner, G., Grishchuk, E. L., and Cheeseman, I. M. (2012). The kinetochore-bound Ska1 complex tracks depolymerizing microtubules and binds to curved protofilaments. *Dev Cell*, 23(5):968–80.

Seber, G. and Wild, C. (2003). *Nonlinear Regression*. Wiley Series in Probability and Statistics. Wiley.

Seisenberger, G., Ried, M. U., Endress, T., Büning, H., Hallek, M., and Bräuchle, C. (2001). Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science*, 294(5548):1929–32.

Sergé, A., Bertaux, N., Rigneault, H., and Marguet, D. (2008). Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat Methods*, 5(8):687–94.

Shin, J. H., Gardel, M. L., Mahadevan, L., Matsudaira, P., and Weitz, D. A. (2004). Relating microstructure to rheology of a bundled and cross-linked F-actin network in vitro. *Proc Natl Acad Sci U S A*, 101(26):9636–41.

Shlomovitz, R. and Gov, N. S. (2008). Physical model of contractile ring initiation in dividing cells. *Biophys J*, 94(4):1155–68.

Simson, R., Sheets, E. D., and Jacobson, K. (1995). Detection of temporary lateral confinement of membrane proteins using single-particle tracking analysis. *Biophys J*, 69(3):989–93.

Sivia, D. and Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial*. Oxford University Press, 2nd edition.

Smal, I., Draegestein, K., Galjart, N., Niessen, W., and Meijering, E. (2007). Rao-Blackwellized marginal particle filtering for multiple object tracking in molecular bioimaging. *Inf Process Med Imaging*, 20:110–21.

Smal, I., Meijering, E., Draegestein, K., Galjart, N., Grigoriev, I., Akhmanova, A., van Royen, M. E., Houtsmuller, A. B., and Niessen, W. (2008). Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering. *Med Image Anal*, 12(6):764–77.

Stauffer, C. and Grimson, W. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757.

Stephens, D. J. and Allan, V. J. (2003). Light microscopy techniques for live cell imaging. *Science*, 300(5616):82–6.

Sun, S. X., Walcott, S., and Wolgemuth, C. W. (2010). Cytoskeletal cross-linking and bundling in motor-independent contraction. *Curr Biol*, 20(15):R649–54.

Tejedor, V., Bénichou, O., Voituriez, R., Jungmann, R., Simmel, F., Selhuber-Unkel, C., Oddershede, L. B., and Metzler, R. (2010). Quantitative analysis of single particle trajectories: mean maximal excursion method. *Biophys J*, 98(7):1364–72.

Theis, M., Slabicki, M., Junqueira, M., Paszkowski-Rogacz, M., Sontheimer, J., Kittler, R., Heninger, A.-K., Glatter, T., Kruusmaa, K., Poser, I., Hyman, A. A., Pisabarro, M. T., Gstaiger, M., Aebersold, R., Shevchenko, A., and Buchholz, F. (2009). Comparative profiling identifies C13orf3 as a component of the Ska complex required for mammalian cell division. *EMBO J*, 28(10):1453–65.

Turner, L., Ryu, W. S., and Berg, H. C. (2000). Real-time imaging of fluorescent flagellar filaments. *J Bacteriol*, 182(10):2793–801.

Vaughan, J. C. and Zhuang, X. (2011). New fluorescent probes for super-resolution imaging. *Nat Biotechnol*, 29(10):880–1.

Veenman, C., Reinders, M., and Backer, E. (2001). Resolving motion correspondence for densely moving points. *Transactions on Pattern Analysis and Machine Intelligence*, 23(1):54–72.

Venkataramanan, L. and Sigworth, F. (2002). Applying hidden Markov models to the analysis of single ion channel activity. *Biophysical journal*, 82(4):1930–1942.

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

Voisinne, G., Alexandrou, A., and Masson, J.-B. (2010). Quantifying biomolecule diffusivity using an optimal Bayesian method. *Biophys J*, 98(4):596–605.

Walter, T., Held, M., Neumann, B., Hériché, J.-K., Conrad, C., Pepperkok, R., and Ellenberg, J. (2009). Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. *Journal of Structural Biology*.

Wang, H.-W. and Nogales, E. (2005). Nucleotide-dependent bending flexibility of tubulin regulates microtubule assembly. *Nature*, 435(7044):911–5.

Wang, X., Wohland, T., and Korzh, V. (2010). Developing in vivo biophysics by fishing for single molecules. *Dev Biol*, 347(1):1–8.

Weber, S. C., Spakowitz, A. J., and Theriot, J. A. (2010). Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys Rev Lett*, 104(23):238102.

Weber, S. C., Thompson, M. A., Moerner, W. E., Spakowitz, A. J., and Theriot, J. A. (2012). Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function. *Biophys J*, 102(11):2443–50.

Welburn, J. P. I., Grishchuk, E. L., Backer, C. B., Wilson-Kubalek, E. M., Yates, 3rd, J. R., and Cheeseman, I. M. (2009). The human kinetochore Ska1 complex facilitates microtubule depolymerization-coupled motility. *Dev Cell*, 16(3):374–85.

Westphal, V., Rizzoli, S. O., Lauterbach, M. A., Kamin, D., Jahn, R., and Hell, S. W. (2008). Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science*, 320(5873):246–9.

Wong, I. Y., Gardel, M. L., Reichman, D. R., Weeks, E. R., Valentine, M. T., Bausch, A. R., and Weitz, D. A. (2004). Anomalous diffusion probes microstructure dynamics of entangled F-actin networks. *Phys Rev Lett*, 92(17):178101.

Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E., and Selvin, P. R. (2003). Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science*, 300(5628):2061–5.

Zumdieck, A., Kruse, K., Bringmann, H., Hyman, A. A., and Jülicher, F. (2007). Stress generation and filament turnover during actin ring constriction. *PLoS One*, 2(8):e696.