

Mathematical modelling and analysis of aspects of planktonic bacterial motility



Gabriel Aaron Rosser
St Anne's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2012

Contents

1	The biology of bacterial motility and taxis	8
1.1	Bacterial motility and taxis	8
1.2	Experimental methods used to probe bacterial motility	14
1.3	Tracking	20
1.4	Conclusion and outlook	21
2	Mathematical methods and models of bacterial motility and taxis	23
2.1	Modelling bacterial motility and taxis: a multiscale problem	24
2.2	The velocity jump process	34
2.3	Spatial moments of the general velocity jump process	46
2.4	Circular statistics	49
2.5	Stochastic simulation algorithm	52
2.6	Conclusion and outlook	54
3	Analysis methods for inferring stopping phases in tracking data	55
3.1	Analysis methods	58
3.2	Simulation study comparison of the analysis methods	76
3.3	Results	80
3.4	Discussion and conclusions	86
4	Analysis of experimental data	92
4.1	Methods	92

4.2	Results	109
4.3	Discussion and conclusions	124
5	The effect of sampling frequency	132
5.1	Background and methods	133
5.2	Stationary distributions	136
5.3	Simulation study of dynamic distributions	140
5.4	Analytic study of dynamic distributions	149
5.5	Discussion and conclusions	159
6	Modelling the effect of Brownian buffeting on motile bacteria	162
6.1	Background	163
6.2	Mathematical methods	166
6.3	A model of rotational diffusion in bacterial motility	173
6.4	Results	183
6.5	Discussion and conclusion	197
7	Discussion and conclusions	202
7.1	Further work	206
Appendices		
A	Mathematical methods	209
A.1	Algorithms for simulating random walks	209
A.2	Derivation of the one-dimensional Telegraph equation	211
A.3	Deriving expressions for the spatial moments of the general velocity jump process	213
A.4	Kernel density estimation	218
A.5	Conditional distribution of interarrival times in a Poisson process	219

B	Analysis of experimental data	220
B.1	Links to online video clips	220
B.2	Computational method for finding the minimum bounding circle	221
B.3	Results from the surface swimming dataset	221
B.4	Bayes factor for hypothesis testing	227
C	Modelling the effect of Brownian buffeting on motile bacteria	230
C.1	Implementation of the Euler-Maruyama algorithm	230
C.2	Derivation of second moments for the run-only model	230
	Bibliography	234

Acknowledgements

The past five years that I have spent at Oxford have seen some of the most rewarding, and most difficult, times of my life. There are so many people whom I would like to thank for their help and encouragement along the way. Many will regrettably remain nameless to prevent this thesis expanding by another chapter, but I would particularly like to acknowledge a few.

I am deeply grateful to my supervisors, Professor Philip Maini, Dr Ruth Baker, Dr Alexander Fletcher and Dr Mark Leake. Your enduring patience, support and commitment have spurred me on throughout my time here. Thank you for helping to make my project so interesting and enjoyable.

I also want to thank the Doctoral Training Centre for a fantastic first year, and the continued support after that. Without the DTC, I would never have found my way to such an excellent project and department. Thank you for doing things differently.

I am indebted to the many members of the Centre for Mathematical Biology who made it such a great place to work and socialise. I particularly wish to acknowledge the vast amount of help I have received from Kit, Louise, Aaron S and Aaron L, without which my project would be so much poorer. Thank you for your unerring altruism.

My involvement in a successful and fun collaboration played a positive formative part in my time at the CMB. I feel lucky to have been part of the ‘bacteria boyz’ with Kit, Trev and Dave, whom I count as friends in addition to collaborators. Thank you for showing

me the best face of interdisciplinary research.

To all of my friends, both those who have been with me from the start and those whom I have met along the way, your encouragement and/or derision have bolstered my spirits. Thank you for the laughs and adventures.

Finally, to my mum, dad and sister, whose contribution is beyond words. Thank you can never be enough.

Abstract

The motile behaviour of bacteria underlies many important aspects of their actions, including pathogenicity, foraging efficiency, and ability to form biofilms. In this thesis, we apply mathematical modelling and analysis to various aspects of the planktonic motility of flagellated bacteria, guided by experimental observations. We use data obtained by tracking free-swimming *Rhodobacter sphaeroides* under a microscope, taking advantage of the availability of a large dataset acquired using a recently developed, high-throughput protocol. A novel analysis method using a hidden Markov model for the identification of reorientation phases in the tracks is described. This is assessed and compared with an established method using a computational simulation study, which shows that the new method has a reduced error rate and less systematic bias. We proceed to apply the novel analysis method to experimental tracks, demonstrating that we are able to successfully identify reorientations and record the angle changes of each reorientation phase. The analysis pipeline developed here is an important proof of concept, demonstrating a rapid and cost-effective protocol for the investigation of myriad aspects of the motility of microorganisms. In addition, we use mathematical modelling and computational simulations to investigate the effect that the microscope sampling rate has on the observed tracking data. This is an important, but often overlooked aspect of experimental design, which affects the observed data in a complex manner. Finally, we examine the role of rotational diffusion in bacterial motility, testing various models against the analysed data. This provides strong evidence that *R. sphaeroides* undergoes some form of active reorientation, in contrast to the mainstream belief that the process is passive.

Statement of Originality

This thesis is submitted to the University of Oxford in fulfillment of the requirements of the degree of Doctor of Philosophy. This thesis contains no material which has been previously submitted for a degree or diploma at this University or any other institution. All experimental data presented in this thesis were obtained by Dr David A. Wilkinson whilst working in the laboratory of Prof. Judith P. Armitage at the Department of Biochemistry, University of Oxford. The specific datasets considered here and all results derived from their analysis have not previously been published at the time of writing. The algorithm used to obtain tracks from the data was written by Dr Trevor M. Wood. Otherwise, this thesis represents my own original work towards this research degree.

Gabriel Aaron Rosser

University of Oxford

October 2012

Foreword

In this thesis, we use mathematical modelling and analysis of experimental data to investigate the process of planktonic bacterial motility, also known as bacterial taxis. This widely-studied phenomenon is implicated in bacterial infection and industrial biofouling, which are responsible for severe health risks and vast financial losses. Furthermore, the motile behaviour of bacteria may provide important information on how they are adapted to inhabit a specific environmental niche. All of the present work is motivated by the availability of new, unpublished tracking data, obtained using a novel high-throughput experimental protocol. The subject organism is *Rhodobacter sphaeroides*, a rod-shaped bacterium that occurs naturally in soil, freshwater and marine environments. The biochemical systems underlying the motile behaviour of *R. sphaeroides* are more complex than those of the more widely studied model bacterium *Escherichia coli* [158].

The tracking data represent a rich source of information on the motility of *R. sphaeroides*, but robust analysis techniques are required to extract reliable statistics. We dedicate part of this thesis to the development of novel methodology to achieve this result. The remaining parts are concerned with modelling aspects of bacterial tracking data, including a study of the effect that sampling frequency has on the information we extract from tracking data, and an investigation of the effect of Brownian rotation on bacterial motility. The work is motivated, guided and verified throughout by comparison with the experimental data. This approach ensures that the problems we investigate in this thesis are biologically relevant and realistic.

The field of bacterial motility, though the subject of a great number of studies, contains many key areas in which further work is required. We now briefly identify the areas in which, to our knowledge, this thesis has made novel contributions. First, the state of the art in the analysis of bacterial tracking data remains, with a few exceptions, predominantly restricted to early work carried out by Berg and Brown [24] and contemporaries. The analysis methods developed in Chapter 3 are a novel approach to the quantitative study of tracking data. Our comparative assessment of their performance with simulated data is the first such study for tracking analysis methods. The application of our methods to real experimental data leads to novel insight into the motion of *R. sphaeroides*, as shown in Chapter 4. Furthermore, the analysis process presented in Chapters 3-4 demonstrates an important proof of principle for the new experimental protocol, constituting a new, low-cost, high-throughput methodology that is applicable to myriad biological investigations in which bacterial motility is a factor. Second, the study by Codling and Hill [50] on the effect of sampling frequency on tracking data suggests several further areas of investigation. We consider an extension of the model of motion used in this study in Chapter 5. In addition, we show how a novel microscopic description of bacterial motion permits a new analytic description of the role of sampling frequency. Finally, many models of bacterial motion ignore the role of noise, for example due to Brownian buffeting. In Chapter 6 we consider two extensions to an existing model by Hagen *et al.* [83], which describe the effect of Brownian rotation on motile bacteria. We test the predictions of our models against experimental data and show how this novel approach yields important insights regarding the reorientation mechanism in *R. sphaeroides*.

An overview of the contents of the thesis now follows. We begin by describing the biological background to bacterial motility in Chapter 1. In this chapter we also discuss the main experimental methods used to investigate bacterial taxis. Chapter 2 provides an overview of mathematical modelling approaches related to bacterial taxis and details the particular modelling framework used throughout the remainder of the thesis. In Chapter

3 we discuss the nature of the experimental data available to us and the information we wish to obtain. Novel methods for analysing the datasets to achieve this are motivated and proposed. We assess the performance of these methods with a simulation study. Having established the suitability of the methods, in Chapter 4 we proceed to present and analyse our experimental data. We study the effect of sampling frequency on the observed tracking data in Chapter 5, demonstrating that this is an important consideration. Finally, in Chapter 6 we address the role of Brownian rotation in bacterial motility, using mathematical modelling and experimental data to address the open biological question of the mechanism of reorientation in *R. sphaeroides*.

Chapter 1

The biology of bacterial motility and taxis

In this chapter, we discuss the biological background to bacterial motility, in particular the process by which bacteria swim through a liquid medium without any attachment to a solid surface. We review the relevant literature in this area and describe the experimental techniques with which researchers have probed bacterial motility.

1.1 Bacterial motility and taxis

The term *taxis* refers to the change in the direction of movement of an organism in response to an environmental cue. A related phenomenon is *kinesis*, which is also a response to an external stimulus, but without any directional bias. The more general term *motility* is used to describe movement that is not necessarily due to an external stimulus. There are many specific forms of taxis, exhibited by a range of living organisms, which differ in the nature of the environmental cue involved. For example, phototaxis is movement mediated by light sources, phonotaxis is motion guided by sounds, and chemotaxis is a response to specific chemicals detected in the surrounding environment. Many species of bacteria are able to respond to a wide range of signals from their surroundings, most

notably light levels and chemical concentrations [192]. In the latter category, chemotactic bacteria alter their direction of movement in order to move towards beneficial carbon sources such as certain sugars and amino acids, which are referred to as *chemoattractants* [5]. Bacteria have also been shown to bias their motion away from certain *chemorepellents*, such as phenol and various salts [198]. The term ‘adaptive klinokinesis’ is used in some articles, with the same interpretation as bacterial chemotaxis [113, 145], but this term is rarely used so we maintain the more common terminology.

Bacteria exhibit a wide range of motile behaviours [84, 94]. When attached to a surface by their cell body, they are denoted *sessile* and variously exhibit swarming, gliding and twitching motility, as well as combinations of these. Away from a surface, bacteria are known as *planktonic* and have been shown to float and swim. The evolution of planktonic motility in bacteria has occurred because it confers a selective advantage, enabling them to migrate towards optimal environments. However, not all planktonic bacteria are motile, as there exists a tradeoff between the energetic cost of generating the requisite motility apparatus and the benefits conferred.

Many bacteria are able to switch between sessile and planktonic phenotypes, with a resultant change in their motility [112]. Communities of sessile bacteria, known as biofilms, are of biomedical interest as they are known to cause many human diseases, including chronic lung infections, tooth decay and infectious endocarditis [95], in addition to being the primary cause of failure and infection in indwelling medical devices [10]. Furthermore, biofouling by biofilms in many industrial processes is believed to result in cumulative losses of billions of pounds each year [64].

Planktonic motility is an important factor in the development of biofilms, as this process controls the initial attachment of individual bacteria to a surface, prior to surface colonisation [148, 160]. Several studies have noted substantially different behaviour of

bacteria swimming near to a surface, compared with those swimming far from a surface [26, 27, 58, 74, 117, 201]. In particular, the trajectories of *Escherichia coli* tend to become approximately circular close to a surface. Several mechanisms for such behaviour have been proposed, involving hydrodynamic models [117, 165] and electrostatic and chemical interactions [201]. In general, the interactions between bacteria and solid surfaces are not well understood, and many aspects of altered metabolism and surface-bound motility at solid-liquid interfaces remain to be elucidated [84].

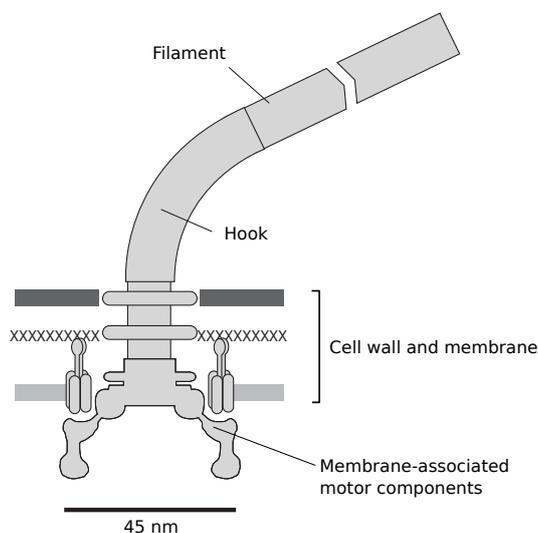


Figure 1.1: A simplified illustration of the components of the bacterial flagellar motor, adapted with permission from [38].

Throughout this thesis, we consider only planktonic bacteria, which swim freely through a liquid medium. With a few exceptions, such as the cyanobacterium *Synechococcus* [66], all motile planktonic bacteria propel themselves using one or more *flagella*. A flagellum is a structure comprising a long helical filament linked through a flexible joint to a rotary flagellar motor, which is mounted on the bacterial cell body [126]. A simple cartoon of the assembly is shown in Figure 1.1. The bacterial flagellar motor uses the transmembrane potential to generate torque, rotating the filament and propelling the cell forward. The mechanisms behind bacterial motility based on flagella have been well-studied in model organisms such as *E. coli*, *Salmonella enterica*, marine *Vibrio* and, more recently, *Rhodobacter sphaeroides* [22, 158]. Despite similarities in the flagellar-mediated propul-

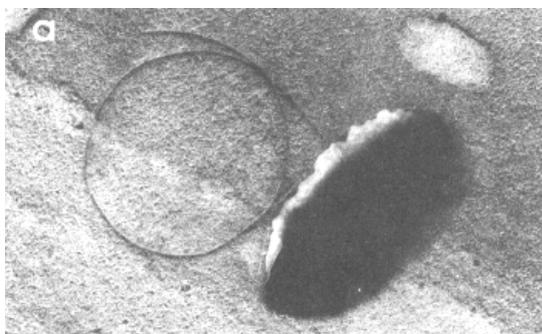
sive mechanisms, these and other species of bacteria differ in the number and arrangement of flagella, as summarised in Figure 1.2(a). Peritrichous and lophotrichous bacteria possess two or more flagella, differing in their points of attachment around the cell. Flagella emanate from the entire surface of the cell body in peritrichous bacteria and from a single pole in lophotrichous bacteria. Figures 1.2(b) and 1.2(c) show representative microscope images of *E. coli* and *R. sphaeroides* cells, respectively. These images demonstrate that the organisms are, respectively, peritrichous and monotrichous (with the flagellum attached medially).

peritrichous		<i>E. coli</i> , <i>S. typhimurium</i>
monotrichous (polar)		<i>P. aeruginosa</i> , <i>V. cholerae</i>
monotrichous (subpolar)		<i>R. sphaeroides</i>
lophotrichous		<i>H. pylori</i> , <i>B. bacilliformis</i>
amphitrichous		<i>S. undula</i>

(a)



(b)



(c)

Figure 1.2: (a) Schematic diagram of the ways in which flagella may be arranged around a bacterium (text, left), with specific examples of species which possess such rearrangements (text, right) [11, 55, 90, 99, 125, 142]. (b) Digital contrast-enhanced image of an *E. coli* cell stuck to a microscope coverslip. The cell is approximately $0.5\ \mu\text{m}$ in diameter along the short axis. Six flagella are visible, splayed out on the coverslip. Reprinted with permission from [29]. (c) Electron microscope image of a *R. sphaeroides* cell, with its single flagellum attached, measuring approximately $0.75\ \mu\text{m}$ in diameter along the short axis. Reprinted with permission from [11].

Bacteria are also known to exhibit a range of swimming behaviours. Three predominant modes of swimming are summarised in Figure 1.3. In run-and-twiddle motile behaviour

(also referred to as run-and-tumble motion by some authors), first proposed by Berg and Brown [24] to explain the observed motion of the peritrichous bacterium *E. coli*, bacteria swim in approximately straight lines interspersed by random, brief reorientation events. Swimming is achieved through counterclockwise rotation of all of the flagellar motors, causing the multiple flagella to form a coherent bundle [22]. Reversal of one or more of the flagellar motors causes the flagella to unbundle and leads to active reorientation [115]. Run-and-stop motion is similar, with the key difference that reorientation is achieved by relaxing the single flagellum [12]. The exact mechanism of reorientation in *R. sphaeroides* is still not understood [12]. We investigate this problem in Chapter 6. The stops observed in *R. sphaeroides* are of greater duration than the twiddles observed in *E. coli*. In the final mode of motion, run-and-reverse, bacteria are propelled forwards by a single flagellum, the reversal of which causes the bacterium to reverse its direction of travel, with the cell body being pulled by the flagellum. This type of motility is prevalent among marine bacteria [139]. It is listed for completeness, but we do not consider it further. Figure 1.3 is not intended as an exhaustive list of motile behaviour; for example, recent evidence suggests that some species, such as *Vibrio alginolyticus*, exhibit complex motility with aspects of both run-and-reverse and run-and-twiddle behaviour [212].

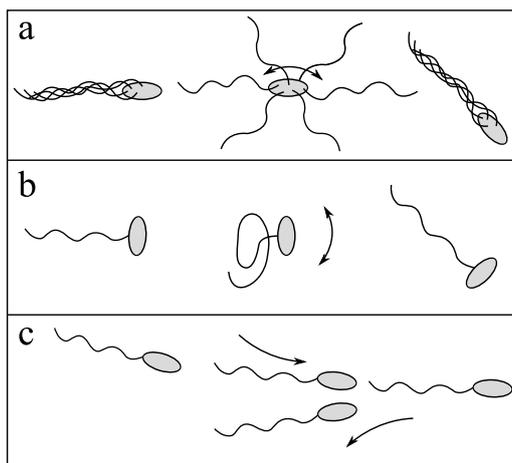


Figure 1.3: Schematic diagram of three modes of motility commonly found in bacteria. (a) Run-and-twiddle motion, exhibited by *E. coli* [24]. (b) Run-and-stop motion, exhibited by *R. sphaeroides* [11]. (c) Run-and-reverse motion, exhibited by *Shewanella putrefaciens* [139].

As a bacterium's one or more flagella rotate and propel it forwards, the cell body rotates in the opposite direction [44, 171]. This so-called 'body roll' occurs because the torque required to drive each flagellum is generated by biological components in the bacterial flagellar motor which are anchored to the cell body. In the case of *R. sphaeroides*, where the flagellum is located away from the cell poles (see Figure 1.2 and [11, 86, 205]), the rotation of the cell body is expected to be more visually pronounced than in monotrichous bacteria with a polar flagellum due to the larger radius of rotation. This effect becomes more significant as the length of the cell increases. The phenomenon of body roll is relevant to the experimental data that we will describe in Chapter 4.

Variations in the ratio of the torque produced by the bacterial flagellar motor to the flexibility of the connecting hook have been shown in recent mathematical models to affect the swimming behaviour of monotrichous bacteria [182]. A suboptimal ratio of these two quantities can lead to large oscillations in the position of the centre of mass of a cell during the course of a continuous swim, resulting in a helical trajectory. This can reduce the motility efficiency, and hence the swimming speed, in such situations. Again, this phenomenon is an important consideration when analysing data from experiments.

Various experimental studies of bacterial motility have noted large variations in swimming behaviour between individual bacteria in a monoclonal population [12, 45, 150]. Such phenotypic differences are common in bacteria, where significant stochastic variation exists in gene expression levels [194]. Therefore it is necessary to study the motion of many individuals from a bacterial population when carrying out an investigation into the phenomenon of planktonic motility, in order to gather data that accurately describe the whole population. Population-level experiments give information averaged over a population or subpopulation. Conversely, experiments involving visualising individual bacteria must be repeated a sufficient number of times.

1.2 Experimental methods used to probe bacterial motility

The earliest observations relating to bacterial motility were described by Engelmann in 1881 [61] and Pfeffer in 1884 [2]. Both used recently-developed microscopes to visualise bacteria, with Engelmann describing phototaxis and Pfeffer investigating chemotaxis. Adler and colleagues performed the first systematic research into bacterial motility [2], isolating genetic mutants from wildtype *E. coli* and studying the effect of mutations upon their motility and ability to undergo chemotaxis [152]. Much of this early work was carried out using swarm plate experiments, in which bacteria are inoculated on semi-solid agar plates and the rate at which they colonise the plate measured [206]. Another experimental method involves trapping bacteria in a capillary containing a chemical attractant and measuring their rate of accumulation [4]. Images of these two experiments are shown in Figure 1.4.

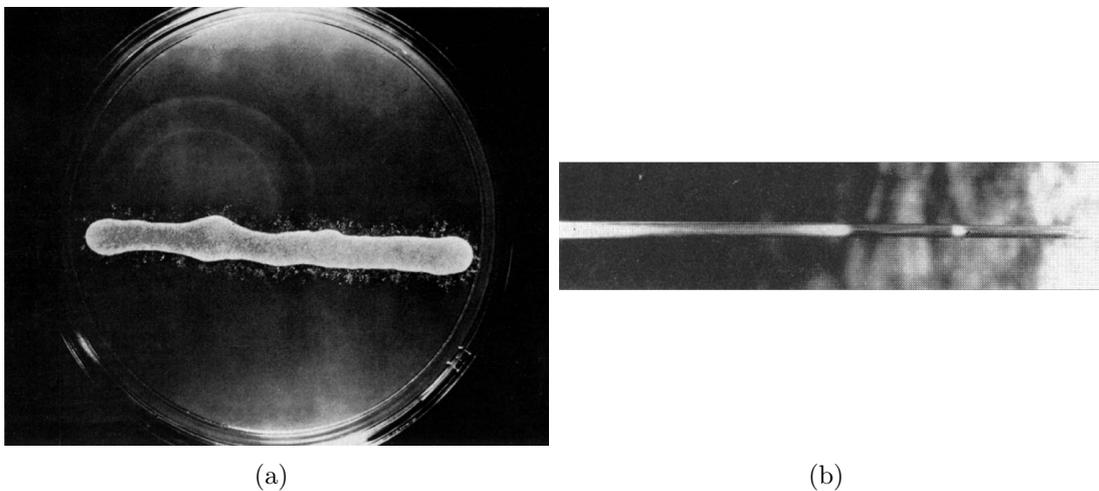


Figure 1.4: Images illustrating the early experimental methods used to probe bacterial chemotaxis. (a) Swarm plate assay carried out by Armstrong and Adler [13], reprinted with permission. This image was taken 20 hours after the agar plate was streaked with an inoculum containing *E. coli*. (b) Capillary assay carried out by Adler [3], reprinted with permission. Bacteria are introduced to the capillary on the left-hand side, immersed in a solution of nutrients, then the tube is sealed. This photograph is taken around 45 minutes after sealing, by which time two separate bands of motile bacteria have travelled along the tube.

Two prominent experimental methods were developed in the early 1970s that are still used in many contemporary studies of bacterial motility. The first is the tethered cell assay, in which cells are tethered to the surface of a microscope slide by a sheared flagellum [184]. The action of the flagellar motor causes the cell body to rotate, which is visualised using a microscope, as illustrated in Figure 1.5(a). This provides information about the speed and direction of rotation of the motor, in addition to any stops or reversals which may occur [28]. A drawback of this method is that the rotation of the cell body is often hindered through contact with the surface [85]. In addition, the motor is under high load in such a configuration, as it must rotate the whole cell body, which is subject to a great deal of viscous drag [38]. More recently, a related method has been developed where the bacterium is tethered to the surface by its cell body [174]. A visible marker, commonly a small latex bead, is attached to the sheared flagellum and observed using a microscope. The flagellum rotates with the attached marker, and the motion is recorded, as shown in Figure 1.5(b). The advantage of this method is that the load on the motor is reduced to approximately natural levels. This method has been successfully used in many studies to probe the detailed mechanisms underlying the bacterial flagellar motor [15, 167, 188].

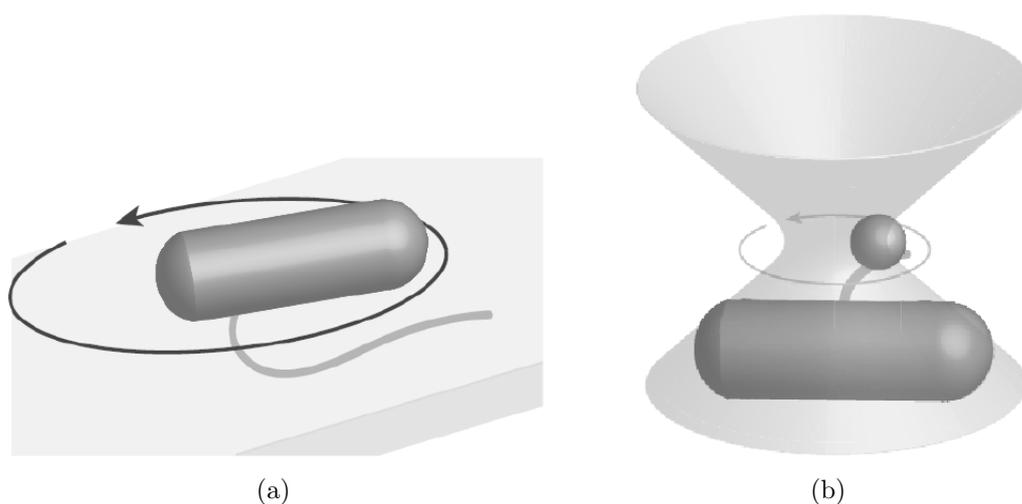


Figure 1.5: *Illustrations of the tethered cell assay, reprinted with permission from [38]. (a) The cell is tethered by its flagellum and the cell body rotates. (b) The cell is tethered by its body, a bead is attached to the filament, and the movement of the bead is tracked with a laser microscope. The yellow shaded cone illustrates the laser beam, which is focused upon the bead. Arrows denote the direction of rotation.*

A second experimental method, pioneered by Berg [20], involves tracking swimming bacterial cells using a microscope. The technology developed to achieve this was based on a microscope with a movable stage controlled by an electronic feedback loop from the detected light intensity at fixed points in the image. This enables the tracking of individual cells in three dimensions, and was used by Berg and Brown in the first systematic investigation of bacterial motility through tracking [24]. Similar experimental techniques have since been used by many researchers to investigate bacterial motility [37, 62, 74, 123, 201]. This approach enables the experimenter to track an individual bacterium for a long time, producing a detailed trajectory. In contrast with tethered cell assays, which monitor motor function, tracking experiments directly reveal the complete phenomenon of swimming behaviour. The method is however relatively laborious, producing a single track at a time, which limits the number of tracks that can be generated. In three separate studies, Duffy and Ford [62], Frymier *et al.* [74] and Vigeant and Ford [201] presented data for fewer than 100 cells, while Berg and Brown report tracking up to 200 cells per day [37]. A related experimental method uses computer software to track bacteria swimming through a fixed field of view [157, 175, 190]. All visible cells are tracked as they pass through the field of view. The process is automated, and therefore less labour-intensive. There exists a trade-off in the approaches discussed above. Methods based on a movable microscope stage are able to track cells in three dimensions at, or away from, a surface, but produce relatively small datasets. Software tracking methods typically operate in two spatial dimensions and generate larger datasets, for example the study by Poole *et al.* [157] involved around 500 tracks. These studies are limited, however, to tracking cells that are swimming at the surface of a microscope coverslip. Although tracking multiple cells in three dimensions has been demonstrated [193, 210], it has not yet been applied in any systematic studies, possibly due to the difficulty in reproducing the experimental protocol. For example, Wu *et al.* [210] developed a high-throughput three-dimensional tracking assay, but this relies on inducing cells to fluoresce using a plasmid, which requires additional experimentation and increases the complexity of the method.

An important distinction exists between the tracking of single cells and simultaneous tracking of all visible targets. Methods which require the researcher to manually select cells for tracking may involve a strong bias towards individual bacteria whose swimming behaviour is deemed ‘representative’ of the population. In reality, monoclonal wildtype bacterial populations often contain cells exhibiting a wide range of swimming behaviours, including a significant proportion of non-motile cells [152]. Studies involving manual selection may fail to represent this spread of motile behaviours. Although this is not explicitly mentioned in [24], for example, it is highly likely to have been manifest in the populations studied. When all bacteria within the microscope’s field of view are tracked, there is no possibility of such user selectivity, although it is conceivable that certain swimming patterns might be systematically under-represented if the tracking process fails more often for cells exhibiting them. As we shall see in Chapter 4, the dataset analysed in this thesis contains a wide variety of swimming behaviours.

The method used to gather the data presented in Chapter 4 represents an extension to the tracking protocols mentioned above. We track multiple targets simultaneously within a fixed field of view using phase-contrast microscopy, a similar approach to that followed by Xie *et al.* [211]. The tracking is performed in two dimensions. This method will be described in detail in Chapter 4; for now we restrict ourselves to discussing the main advantages and disadvantages associated with this approach. A major advantage of the method discussed here is the ability to track cells which are swimming away from a surface. This has not, to the best of our knowledge, previously been possible using traditional microscopy techniques available on standard commercial microscopes. Furthermore, the method is high-throughput in terms of the number of tracks generated, which is on the order of thousands for ten to twenty minutes of microscope footage. As with any fixed field of view method, a disadvantage of this method is that tracks are often short in comparison with moving stage methods. In addition, as we capture images

and track in two dimensions, we lose information relating to the depth (the perpendicular distance from the focal plane in the microscope) of the tracked bacteria.

Figure 1.6 highlights the differences between fixed field of view tracking and moving stage tracking. Figure 1.6(a) shows a representative sample of tracks from the dataset used in Chapter 4. These tracks vary in length from roughly $10\ \mu\text{m}$ to $100\ \mu\text{m}$. Several different swimming patterns are apparent, distinguished by the varying speeds and ‘jaggedness’ of the tracks. In stark contrast, the single track shown in Figure 1.6(b), obtained by Frymier *et al.* [74] using a moving stage tracking microscope, is several hundreds of microns long and three-dimensional.

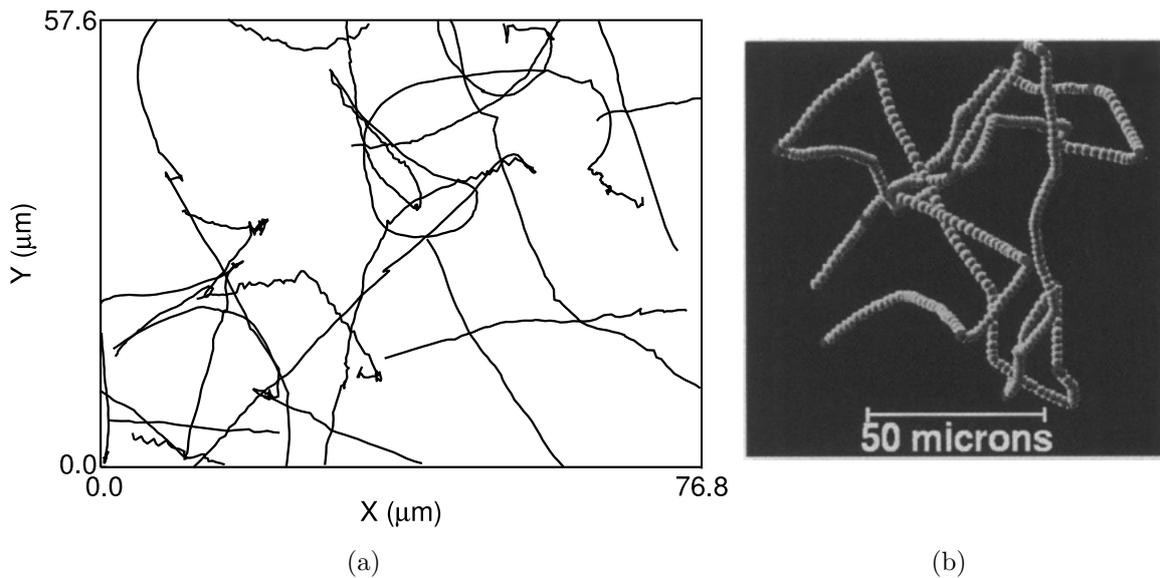


Figure 1.6: Representative figures showing tracks from (a) the dataset used in this thesis, described in Chapter 4, in which multiple cells are tracked simultaneously in two dimensions using a fixed field of view, and (b) Frymier *et al.* [74], in which single cells are tracked mechanically. Figure reprinted with permission.

Few studies have thus far demonstrated tracking of bacteria in a well-defined concentration gradient of chemoattractant, due to the technical difficulties involved with setting up such a gradient. Notably, Berg and Brown [24] successfully tracked bacteria swimming in a concentration gradient, although the exact nature of the concentration gradient is not well characterised. Macnab and Koshland [127] observed *Salmonella typhimurium* in

a concentration gradient set up using a peristaltic pumping device, but were not able to systematically obtain statistics relating to the observed motion. Other researchers have reported differences in bacterial motility observed in isotropic media with varying levels of chemoattractant [157, 175]. The study conducted by Sager suggests that bacteria respond to a stepwise change in concentration by altering their swimming patterns, before adapting to return to their original behavioural patterns. These findings, obtained using tracking experiments, are supported by other findings related to bacterial response and adaptation [8, 45, 111, 156]. In the method used to obtain our tracking data, the bacteria are in an time-invariant isotropic environment, hence we are considering the specific case of bacterial motility in the absence of environmental cues. This is an acceptable limitation because, as we shall see, there is much work to be done on the basal motility of bacteria before we are able to consider their chemotactic swimming patterns in gradients of chemoattractant.

1.2.1 Methods in development

Before moving on to describe in greater detail the process of tracking cells imaged under a microscope, we take the opportunity to discuss briefly some of the developments which are expected to lead to advances in experimentation in this field. Recent advances in three-dimensional holography promise to extend the tracking method used in the present work to three dimensions, in addition to improving the size of the field of view [217]. This technique should produce tracks which are, on average, of greater duration, as well as avoiding any issues associated with projecting tracks onto a two-dimensional plane. Advances in the field of microfluidics will also enable experimental studies of bacteria in well-defined steady gradients of chemoattractant or repellent [6]. Recent improvements to the original tethered cell assay have enabled high-throughput studies of the response of bacteria to stepwise changes in concentrations [111]. The modified tethered cell assay, in which bacteria are tethered by their cell body, has also seen a major improvement to

the experimental protocol through the use of gold nanoparticles in place of latex beads [189].

1.3 Tracking

We now focus on the problem of tracking bacterial cells that are imaged using a microscope. We first present an outline of the general methods by which tracking is achieved in a diverse range of contemporary research studies, before focusing on the specific case of bacterial tracking.

The aim of a tracking procedure is to associate targets across consecutive video frames. Automated tracking is a generic problem that is applicable to many research areas including biology [93, 173, 176, 180], fluid dynamics [149], security and surveillance [39, 91] and ecology [114]. There is a great deal of variation in the nature of the video images to which tracking algorithms may be applied, the appearance of the targets to be tracked and their motion. In many cases tracking is achieved in two stages: objects to be tracked are first detected in each frame of the video to create a list of targets (the object detection stage), then the targets are connected across frames to form a trajectory (the target association stage).

In most cases, the aforementioned object detection approaches require coupling to a data association algorithm in order to link targets across frames. The problem is usually complicated by the presence of spurious targets (caused by false positives from the detection stage), missed detections (false negatives from the detection stage), track crossing, object occlusion (full or partial overlapping of two or more objects) and track spawning (where an object splits into two). The majority of the studies discussed in Section 1.2 implement a data association method known as the nearest neighbour algorithm [149], which is sensitive to all of these complications, although this problem is rarely discussed in the

references. Berg [20] gives anecdotal evidence of the issue of occlusion; he describes a situation in which a motile bacterium strikes a non-motile bacterium whilst being tracked:

“the organism was seen . . . to collide with one which was not swimming; the microscope locked onto this bacterium and the other swam out of focus.”

Similarly, the presence of false detections can cause the data association algorithm to inadvertently incorporate spurious targets into a trajectory, rendering an incorrect track. This provides a motivation for studies in which the main focus is upon improving the accuracy of the object detection stage. In [210], for example, the bacteria to be tracked are genetically engineered to fluoresce, to improve the contrast in the microscope image. Similarly, in [190], the authors use dark field microscopy to improve image contrast. Xie *et al.* [211] develop a sophisticated cell detection algorithm to improve the detection of cells and reduce false detections. We describe the method used to acquire our datasets in more detail in Chapter 4.

1.4 Conclusion and outlook

In this chapter, we have briefly summarised the biology behind bacterial motility, and described the experimental methods which are used to investigate this phenomenon. The current state of the art in video microscopy for obtaining tracking data is little changed from some of the earliest methodologies developed. However, the methods used to process such data have seen many advances, made possible in part by the availability of powerful computers, in addition to the development of novel mathematical approaches. In our specific case, we use a recently developed tracking algorithm to improve the throughput of the experiment. A straightforward, inexpensive and readily available microscopy method is used to produce a large quantity of data, discussed in greater detail in Chapter 4.

The data we obtain must be interrogated in order to gain meaningful insight into the phenomenon of bacterial motility. Only by quantifying certain aspects of the observed

motion can we learn about the process. This represents a challenge: phenomena such as non-motile cells and short tracks were not an issue in experiments where cells were individually selected and tracked. We note that more recent studies into bacterial tracking such as [190, 193, 210] focus on the tracking methodology, but lack substantial further investigation into the nature of the tracks generated. This is in contrast with Berg and Brown's pioneering work [24], in which the authors present the characteristics relating to the swimming of *E. coli*.

Chapter 2

Mathematical methods and models of bacterial motility and taxis

In this chapter, we motivate and review the mathematical modelling approaches that have been applied in the area of bacterial motility and taxis. We also provide technical details of mathematical and modelling methods that will be used throughout the thesis. A large part of what we know today about the motile behaviour of planktonic bacteria is due to the efforts of the physicist Howard Berg and co-workers; see [23] for an overview of this work. Berg motivated the use of many physical models to explain the phenomena he observed, including a fluid dynamical explanation for the generation of forward thrust by the rotation of a helical filament [21], the effect of Brownian motion on swimming patterns [21], and the mechanism by which cells detect changes in chemical concentrations [25]. Thus, since the early 1970s, this field has benefited from the application of mathematical modelling.

The range and diversity of models relating to bacterial motility and taxis have grown a great deal over the last few decades. Such models have made many further contributions to our understanding of the biology underlying these processes. The benefits of modelling are numerous. First, models have in several cases helped to further under-

standing of specific mechanisms involved in chemotaxis and motility. A recent example is the work by Tindall *et al.* [196] on the chemotaxis pathway in *R. sphaeroides*, in which model predictions directly motivated further experimental work, which in turn elucidated new mechanistic details. Second, models enable us to make predictions about observed behaviour under certain assumptions. A key example is the effect of the image acquisition rate on observed tracking data, presented in Chapter 4, in which we use the underlying assumption of a particular model of motion to predict the effect of changing the imaging rate on observed statistics. It would be impossible to investigate this effect systematically without the application of a model. Third, mathematical modelling is crucial in formalising and quantifying the temporal and spatial scales that characterise a system. For example, Othmer and Hillen [147] discuss the scaling required to relate a microscopic model of motion to a diffusive process.

In Chapter 1, we presented a biological overview of planktonic bacterial motility and taxis, and discussed the experimental methods used to probe this process. An important conclusion is that novel data analysis methods are necessary to maximise the quantity of meaningful information we can extract from recently developed high-throughput experimental methods for cell tracking. This information, in the form of statistics relating to the observed motion, can then be used to parameterise models of bacterial motility. We remark that the flow of information suggested here, i.e. from data to statistics to models, is an oversimplification: the process of creating analysis tools and extracting statistics may itself be influenced by an assumed model of motion, as we shall see in Chapter 3.

2.1 Modelling bacterial motility and taxis: a multi-scale problem

The processes responsible for, and resulting from, bacterial motility and taxis cover a wide range of spatial and temporal scales, as illustrated in Figure 2.1. This represents

a challenge from a modelling perspective because it is difficult to incorporate multiple scales into an effective mathematical or computational model. Tindall *et al.* [197] provide a recent comprehensive review of mathematical models relating to bacterial chemotaxis. That no single model is able to span all time and length scales relating to chemotaxis is noted by the authors, who state,

“... while mathematical modeling has aided in understanding bacterial chemotaxis on the individual cell scale and guiding experimental design, no single model succeeds in robustly describing all of the basic elements of the cell.”

At the smallest spatial scale, underlying all observed motile and tactic behaviour, are the biochemical reactions occurring at the surface of the cell and inside it. These interactions, between proteins and other small molecules on the nanometre scale, are responsible for detecting chemoattractants and chemorepellents, transmitting a signal to the one or more motors and switching or stopping the motors. The timescales for these reactions vary over several orders of magnitude: chemoattractant molecules bind to chemoreceptors on a timescale of the order of milliseconds, adaptation occurs over seconds or minutes and accumulation of populations occurs over hours or days [197].

Throughout this thesis we focus on the phenomenological description of bacteria swimming in an isotropic environment, as the experimental tracking data that we consider is best described in this manner. The observed behaviour is well described by a modelling framework called the velocity jump (VJ) process, which describes the microscopic trajectories of individual bacteria. We give a detailed explanation of this model in Section 2.2. Before this, we give a brief overview of other successful modelling approaches that have been used to describe the various facets of bacterial motility and taxis, in order to show how our work relates to these descriptions. Wherever possible, we highlight the connections between the various modelling approaches and, where they exist, any attempts to create hybrid models spanning different behaviours or scales.

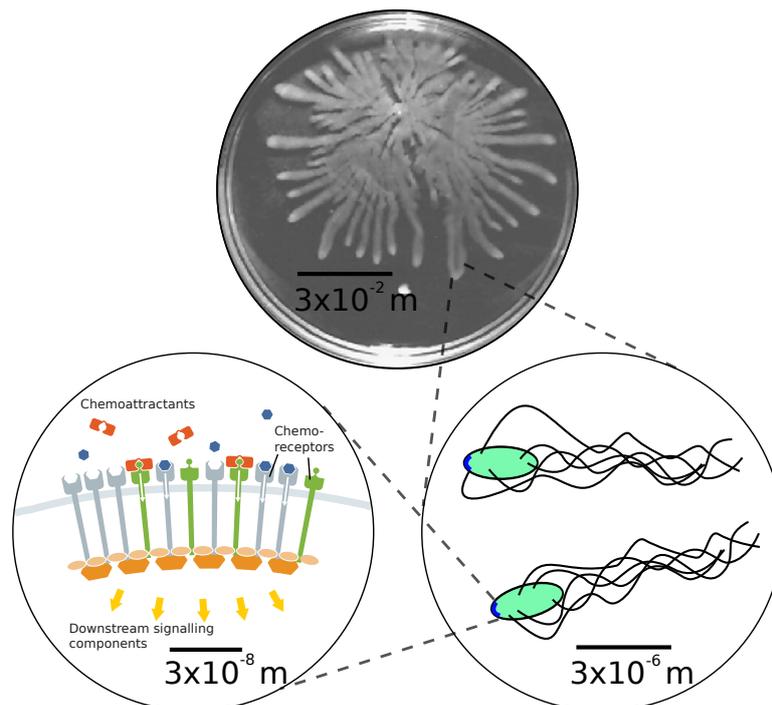


Figure 2.1: An illustration of the wide range of spatial scales involved in bacterial taxis. The uppermost circle depicts a swarm plate assay, reproduced with permission from [110]. A typical incubation period for this assay is one or more days. The right circle depicts individual bacteria swimming. The timescale for the rotation of the flagella is on the order of 10^{-2} s . The left circle illustrates the chemoreceptors responsible for response and adaptation in bacterial chemotaxis, adapted with permission from [199]. Diffusion and binding of small signalling and chemoattractant molecules occurs on a typical timescale of 10^{-3} s , an initial response to a stimulus takes around 10^{-1} s , and the cell adapts to changes in the environment in approximately $10^1 - 10^2 \text{ s}$ [197].

2.1.1 Models of response and adaptation

We first consider models of the biochemical reactions occurring within a bacterium, illustrated in the lower left of Figure 2.1. Some of the earliest studies of *E. coli* in chemoattractant concentration gradients showed that they are able to migrate towards sources of chemoattractants by modulating the frequency of their reorientations [2, 24]. Bacteria subjected to a stepwise change in chemoattractant concentration have been observed to respond to the change, before adapting and returning to their original motile behaviour. Response times for *R. sphaeroides* have been measured at around one second for stimulation by chemoattractant [111] and light [28]. Similar investigations revealed a response time of 0.2 s in *E. coli* [30]. Adaptation occurs on a longer timescale, from 5 s to 100 s in *R. sphaeroides* undergoing phototaxis [28], and over several minutes in *E. coli* [8].

The response and adaptation dynamics in bacterial chemotaxis have been the subject of a great deal of theoretical research, with most studies dealing with the model organism *E. coli* [197]. A large number of such models use the framework of ordinary differential equations (ODEs) together with the Law of Mass Action to describe how the concentrations of various intracellular proteins and chemoreceptors evolve in time. In these models, the cell is assumed to be a well-stirred reaction vessel and spatial details are ignored. More recently, a few authors have modelled the spatiotemporal variation of the same biological components using stochastic simulations [120, 121] and partial differential equations (PDEs) [196].

Other aspects of the chemotactic response observed biologically are the very high *sensitivity* to even small changes in the chemoattractant concentration [178, 187], which requires that an amplification of the signal (known as *gain*) occurs within the signalling pathway [19, 34], and the *robustness* of the system to variations between cells of the concentrations of various signalling components [8]. In particular, sensitivity and gain have been the focus of many models in which clustering of the chemoreceptors is considered, initially

proposed by Bray *et al.* [35]. Several models based on this premise have used the Ising model to describe the interactions between neighbouring chemoreceptors [63, 181].

Models of the internal chemotactic signalling pathways may, to some extent, be incorporated in higher-level models of cell motility. The first published example is a study by Erban and Othmer [70], in which the authors present a simplified two-state model of internal signalling dynamics in a VJ description of bacterial motion and use this to derive a macroscopic PDE governing the evolution of the system. They remark that the current state of the art in response and adaptation modelling involves more than 20 variables describing the various signalling components, which is at present too complex to incorporate into a VJ framework. The work in this area has since been extended, albeit still with a minimal two-state internal model, by Xue and Othmer [214].

2.1.2 Hydrodynamical models of swimming

The bacterial flagellar motor rotates approximately 100 times per second in both *R. sphaeroides* and *E. coli*, driven by the flow of protons through the motor complex [22, 156]. The observed rate of rotation varies substantially with the experimental conditions under which it is measured: when *R. sphaeroides* are tethered by their flagellum to a microscope slide and the rotating cell body imaged, the motor rotates approximately seven times per second due to the high load caused by the drag on the cell body [151]. A simple model for how a rotating helical filament (the flagellum) propels the cell, based on viscous drag, is described by Berg [21]. Of particular interest in the field of hydrodynamical modelling is the behaviour of bacteria swimming near solid surfaces. Ramia *et al.* [165] demonstrate that such models predict the circular swimming motion of bacteria at surfaces. More recently, Shum *et al.* [183] have performed a detailed numerical simulation of bacteria swimming near surfaces. This study considered the efficiency of swimming as a function of flagellum length, the wavelength of the flagellar helix, and the shape of the cell

body. The authors conclude that the efficiency of the swimming process is robust both to changes in these parameters and to the presence of a surface.

2.1.3 Population-level models

Whilst bacteria are unicellular organisms, and thus able to act independently of other bacteria, bacterial populations (monoclonal or otherwise) can show collective behaviour. Such behavioural traits occur due to interactions between bacteria and their environment (for example sensing and consuming chemoattractant) and interactions between bacteria (for example quorum sensing [135]). Mathematical models for such behaviour are broadly divided into two categories: those which seek to describe the behaviour of chemotactic bacterial populations, reviewed by Tindall *et al.* [195]; and those related to bacterial biofilms, reviewed by Klapper and Dockery [106]. As models of biofilms are beyond the scope of this thesis, we do not consider them further.

Keller and Segel [104] postulated the first model of population-level behaviour of chemotactic bacteria in an attempt to explain the bands of bacteria observed by Adler in his capillary assay [2]. They model the motion of bacteria as diffusive, with an additional ‘chemotactic’ drift term, describing the movement of bacteria up gradients of chemoattractant. The chemoattractant is a diffusing chemical species, which undergoes decay and consumption by the bacteria. The general form of the Keller-Segel (K-S) model is given by

$$\frac{\partial b}{\partial t} = \nabla \cdot (\mu(s)\nabla b) - \nabla \cdot (\chi(s)b\nabla s) + g(b, s) - h(b, s), \quad (2.1)$$

$$\frac{\partial s}{\partial t} = D_s \nabla^2 s - f(b, s),$$

where $b(\mathbf{x}, t)$ is the population density of bacteria at position \mathbf{x} and time t , $s(\mathbf{x}, t)$ is the concentration of chemoattractant, $\mu(s)$ is the bacterial diffusion coefficient (often assumed to be constant), $\chi(s)$ is the chemotactic coefficient, $f(b, s)$ describes degradation of the

chemoattractant, including both chemical degradation and consumption by the bacteria, D_s is the diffusion coefficient of the chemoattractant, and the functions g and h describe growth and death terms, respectively. Note that bacteria do not interact directly in the K-S model; all interactions are mediated by the chemoattractant, s , such that consumption of chemoattractant by one band of bacteria affects the motile behaviour of another spatially distinct band.

An important result of the K-S model is that it permits travelling wave solutions when the following simplifying assumptions are made: (1) bacteria neither reproduce nor die on the timescale of the experiment, so that $g = h = 0$; (2) the chemoattractant does not diffuse, $D_s = 0$; (3) χ takes the form $\chi(s) = \chi/s$; (4) a ‘speculative’ function is used for f , which the authors admit has little biochemical basis (see [104] for details). The travelling wave solutions thus generated agree reasonably well with the bands of bacteria observed by Adler. This influential model has since been used as the basis for many subsequent studies. Assumption (3) above is problematic, as it leads to a singularity in the chemotactic coefficient as the chemoattractant concentration approaches zero. This issue is discussed by Xue *et al.* [213], who emphasise that this form of χ leads to the model producing unrealistically high theoretical cell migration speeds at low levels of chemoattractant. Furthermore, more recent work based on the original K-S model has considered more biologically realistic functional forms for the degradation term, f , with similar outcomes: for examples, see the references in the review by Tindall *et al.* [195].

It has further been suggested that travelling waves solutions are obtained in extensions to the K-S equations that incorporate growth and death of the bacterial population [213]. Indeed, Lauffenburger *et al.* [116] investigate the possibility of achieving travelling wave solutions to a simple reaction-diffusion description of the bacterial population and chemoattractant in the absence of chemotactic drift, and show that travelling wave solutions can arise due to growth, death, consumption of chemoattractant and diffu-

sive behaviour. However, this result is unremarkable from a mathematical perspective: Fisher’s equation, which models the diffusion and growth of a population in the absence of chemotaxis, has long been known to give travelling wave solutions Murray [141], Xue *et al.* [213].

The VJ modelling framework with the inclusion of an external bias due to, for example, a chemoattractant gradient has been shown by Othmer and Hillen [147] to lead to a drift-diffusion equation in the appropriate scaling limit. Furthermore, Erban and Othmer [70] demonstrated a link between the VJ process with internal signalling dynamics and the macroscopic K-S equations in one spatial dimension. This analysis was subsequently extended to two- and three-dimensional models by the same authors [71]. This analysis illustrates an important link between the various scales of modelling bacterial motility and taxis. However, Erban and Chapman [69] used the VJ framework to model the more complex phenomenon of chemotaxis in crawling ameboid cells and showed that the resulting macroscopic equations differ from the K-S equations.

2.1.4 Random walk models of motion

The scale at which we wish to consider the process of bacterial motility is *phenomenological* and based on detailed knowledge of observed swimming patterns. We do not explicitly model the biochemical signalling pathways occurring within each individual bacterium, neither do we model interactions between individuals, or between bacteria and their environment. We restrict ourselves to this scale as it is the most appropriate when considering data from a microscopic tracking experiment in which bacteria are sealed in an isotropic and unvarying environment, as presented in Chapter 1, and analysed in Chapter 4. The VJ model of microscopic motion is a specific example of a more general class of models called *random walks*. We first briefly describe the random walk in full generality, before focusing on our specific case. The random walk was first described in the form of an open question by Pearson in 1905, entitled “The problem of the random walk” [153]:

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

The request was soon answered by Rayleigh, who provided a solution in the limit of very many steps [166]. There are two key properties that apply to random walks, *bias* and *correlation* [51]. Bias refers to a global preferred direction; in the context of Rayleigh’s problem, bias could be introduced by having the man turn more often in a northerly direction, or stepping further when he is travelling in this direction than when he is travelling in another direction. Correlation refers to the dependence between consecutive moves; Rayleigh’s wanderer would be exhibiting a correlated random walk (CRW) if, for example, the angle through which he turned between each move were not randomly chosen, but instead were selected relative to the angle of travel from a non-uniform distribution. If the mean cosine of the turning angle is greater than zero, the CRW exhibits the property of *persistence*, meaning that directional information is maintained over consecutive frames (circular statistics are discussed further in Section 2.4). This correlation will eventually vanish with increasing numbers of steps, leading to the concept of *persistence time*. The position of Rayleigh’s wanderer takes values in \mathbb{R}^2 , as the angle through which he turns is continuous on $[-\pi, \pi)$. This type of random walk is known as an *off-lattice* model. In contrast, the position of a wanderer in *on-lattice* models takes values defined on a discrete grid, for example a regular square array in two dimensions. These general concepts are illustrated in Figure 2.2 for some typical simulated tracks.

The simple random walk described by Pearson is known as a *position jump* process [146], because the moving object, which we henceforth refer to as ‘particle’ for generality, shifts position at random. The position of the particle is discontinuous between jumps; illustrations of position jump processes such as those in Figure 2.2 usually portray the process

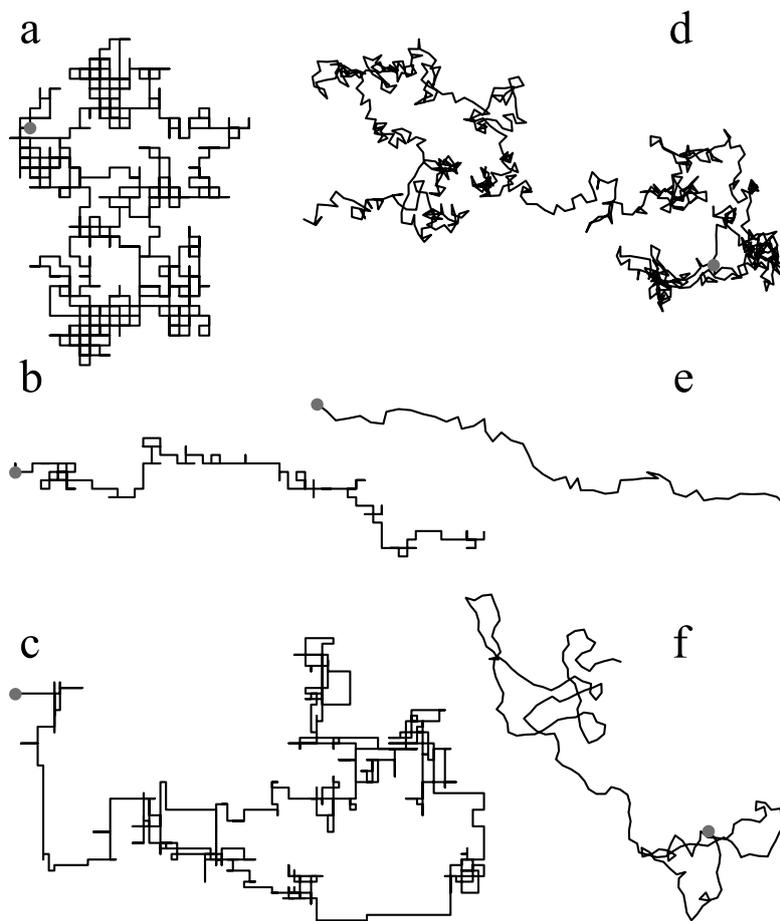


Figure 2.2: Simulated on-lattice (a)-(c) and off-lattice (d)-(f) random walks. (a), (b) Simple (unbiased and uncorrelated) random walk. (c), (d) Biased random walk, the off-lattice model incorporates sinusoidal reorientation as described by Hill and Häder [88]. (e), (f) Correlated random walk. In all simulated trajectories, the step length is constant. The dot indicates the starting point. A detailed description of the algorithms and parameters used is given in Appendix A.1.

as if it were continuous, but this is only for visual purposes. There is no direct correlation between consecutive steps in position jump models, though they may be biased. Pearson's constant step length position jump model is a reasonable approximation of the microscopic motion of particles undergoing Brownian diffusion [51]. This link has also been demonstrated in other position jump models of motion [69]. Position jump processes have been used to model a wide range of phenomena, including the dispersal of soil particles [133], the migration of tumour cells [57] and the spread of populations of oak trees and wingless beetles [185].

For the majority of living organisms, the assumption of uncorrelated moves is unrealistic as an individual has a tendency to continue moving in the same direction on some non-negligible timescale [51]. In such cases, the CRW is a more realistic model of the observed motion, since observed motions are temporally correlated. This process is also known as a VJ process, because it describes a particle that experiences random shifts in its velocity [146]. Henceforth, we will use the term VJ process. Similarly to position jump processes, a link between microscopic VJ models of motion and the macroscopic diffusion equation has been established [69]. In contrast to the position jump process, the position of the particle in the VJ process is continuous. As discussed, the VJ process is a natural framework to describe the observed motion of planktonic flagellated bacteria, and has been applied to model this phenomenon in previous studies [9, 49, 143, 146]. We shall use this mathematical framework throughout the remainder of this thesis. In the following section, we describe the VJ process in detail.

2.2 The velocity jump process

In describing the VJ process, we draw a distinction between a microscopic description of the trajectory traced out by a single velocity jumping particle, which is inherently stochastic, to a macroscopic and deterministic description of the temporal evolution of

the probability density function (pdf) governing the position and velocity of a population of particles. The two descriptions are complementary, providing access to different aspects of the VJ process. For example, a microscopic description of the process permits the definition of properties such as the observed distribution of step lengths and turning angles in a track, whereas a macroscopic description may be used to derive expressions for moments of the spatial distribution of a population. The distinction is in fact not clear cut; these two approaches are merely different mathematical descriptions of the same process, as we shall demonstrate below.

We first present a simple example of the VJ process in one spatial dimension in order to introduce the mathematical methods that are frequently applied in this context. This problem was first analysed by Goldstein [77] and considered later by Kac [100]. Much of the mathematical analysis is included in [47] and [146]. The one-dimensional model provides an illustrative example of the process of passing from a microscopic to a macroscopic description of the VJ process. In the following sections, we present details of the general VJ process in a higher number of spatial dimensions.

2.2.1 The velocity jump process in one spatial dimension

Consider a particle moving in one dimension with constant speed c . At random times the particle experiences an instantaneous jump in velocity, such that it reverses direction. This process is illustrated in Figure 2.3. The velocity jumping process is Poisson with rate λ [146]. Kac describes the model emotively in terms of a stochastic process on a one-dimensional lattice [100]:

“ I start a particle from ... $x = 0$ and the particle always moves with speed v . It can move either in a positive direction or in the negative direction. I flip a coin, let's say, to determine which ... So, what actually happens is that for a time you move in the direction you have chosen. And then, all of a sudden,

you flip over. For a time you move in the new direction. And then, all of a sudden, disaster overtakes you. And so you will oscillate.”

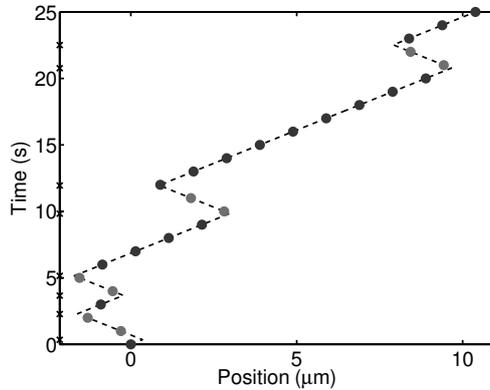


Figure 2.3: Illustration of the one-dimensional VJ process (the Telegraph process), computed using equations (2.2) and (2.3) with $\lambda = 0.2 \text{ s}^{-1}$, $c = 1 \text{ } \mu\text{ms}^{-1}$ and $\lambda\tau \ll 1$. A single particle moves in one dimension, reversing direction at random intervals. The times at which reversals occur are indicated by crosses on the vertical axis. The dashed line traces out one possible trajectory. Dots indicate the position of the particle recorded at regular time intervals.

Othmer *et al.* [146] discuss the process of passing from this microscopic, stochastic description of the movement process to a macroscopic, deterministic description of the evolution of the pdf of the particle’s position. We provide only the main results here; further details may be found in Appendix A.2.

Consider a very large number of non-interacting particles, undergoing the process described above. All particles start at $x = 0$ when $t = 0$. Denote the number density of individuals moving with velocity $v = c$ by $p^+(x, t)$, so that $p^+(x, t) dx$ gives the number of particles travelling in the positive x -direction located in the interval $[x, x + dx]$ at time t . The analogous number density of particles moving in the negative x -direction is denoted $p^-(x, t)$. We now consider the change in the distribution of the particles after a small time step τ . The distance moved by all particles in this time is $\delta = c\tau$. The probability that a particle reverses direction in this time interval is given by $\lambda\tau$, by the definition of a Poisson process, where $\lambda\tau \ll 1$ so that we may assume that at most one reversal occurs in the time interval $[t, t + \tau)$. Considering the microscopic motion of the particles over

the time step leads to the equations

$$p^+(x, t + \tau) = \lambda\tau p^-(x - \delta, t) + (1 - \lambda\tau)p^+(x - \delta, t), \quad (2.2)$$

$$p^-(x, t + \tau) = \lambda\tau p^+(x + \delta, t) + (1 - \lambda\tau)p^-(x + \delta, t). \quad (2.3)$$

Expanding equations (2.2) and (2.3) in the small parameters τ and δ and taking the limit $\tau, \delta \rightarrow 0$, such that $\delta/\tau = c$, we obtain after some manipulations

$$\frac{\partial^2 P}{\partial t^2} = c^2 \frac{\partial^2 P}{\partial x^2} - 2\lambda \frac{\partial P}{\partial t}, \quad (2.4)$$

where $P = p^+ + p^-$ denotes the total population of particles. Equation (2.4) is variously known as the *Telegraph equation*¹ and a *damped wave equation* [102]. The solution to equation (2.4) on the infinite, unbounded x domain, subject to the initial conditions

$$P(x, 0) = \delta(x), \quad \frac{\partial P}{\partial x}(x, 0) = 0, \quad (2.5)$$

is given by [146]

$$P(x, t) = \begin{cases} \frac{e^{-\lambda t}}{2} \left\{ \delta(x - ct) + \delta(x + ct) + \frac{\lambda}{c} \left[I_0(\Lambda) + \frac{\lambda t}{\Lambda} I_1(\Lambda) \right] \right\} & \text{for } |x| < ct, \\ 0 & \text{for } |x| \geq ct, \end{cases} \quad (2.6)$$

where $\Lambda = \lambda\sqrt{t^2 - x^2/c^2}$ and I_0 and I_1 denote modified Bessel functions of the first kind. This solution is plotted for various time points in Figure 2.4, overlaid with a histogram computed from the results of a stochastic simulation. Details of the stochastic simulation algorithm are given in Section 2.5. The density of particles in the wavefronts decreases with time (red bars), as fewer particles remain that have not undergone a single reversal event. The terms in equation (2.6) containing the delta function represent the propagation of a wavefront at $x = ct$ corresponding to particles which have not reversed by time

¹The Telegraph equation is so named because its original derivation by Heaviside in 1880 was to describe the transmission of signals along telegraph lines [215].

t . The wavefront decays exponentially, as expected, since the reversal process is Poisson. The solution is identically zero in front of this wavefront, as particles cannot travel faster than the fixed speed c .

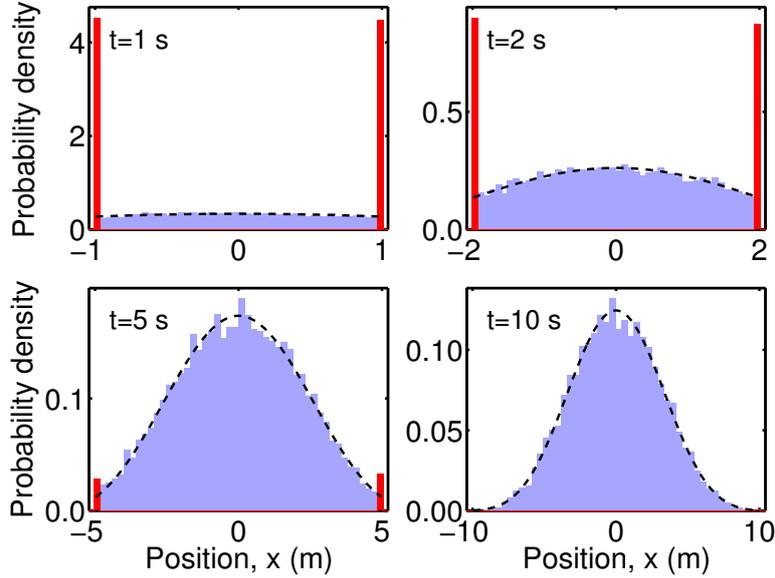


Figure 2.4: The solution to (2.4) on an infinite x domain with the initial conditions given in (2.5). Bars represent the histogram of a stochastic simulation of the microscopic process with 10^4 particles. Red bars denote particles at the wavefront, $|x| = ct$, and blue bars denote the remaining particles. The black dashed line is the analytic result in (2.6), plotted without delta functions. The x -axis is scaled to show the region $|x| \leq ct$ in each plot. The parameters used are $\lambda = 1 \text{ s}^{-1}$, $c = 1 \text{ ms}^{-1}$.

Equation (2.6) reduces to the solution of the diffusion equation for large t and $x \ll ct$ [102, 146]. To show this, we note that for large z , the modified Bessel functions $I_\nu(z)$ have the asymptotic expansion [1]

$$I_\nu(z) = \frac{e^z}{\sqrt{2\pi z}} + O\left(\frac{1}{z}\right). \quad (2.7)$$

Let $\xi = x^2/c^2t^2$, hence $\Lambda = \lambda t\sqrt{1-\xi}$. We now consider the solution far from the boundaries as $t \rightarrow \infty$, so that $|x| \ll ct$ and $\xi \ll 1$. In this regime, $\Lambda \approx \lambda t(1 - \xi/2)$ and equation (2.6) reduces to

$$P(x, t) \approx \sqrt{\frac{\lambda}{2\pi c^2 t}} \exp\left(\frac{-\xi t \lambda}{2}\right) = \sqrt{\frac{1}{4\pi D t}} \exp\left(\frac{-x^2}{4Dt}\right), \quad (2.8)$$

where $D = c^2/2\lambda$ is the diffusion coefficient. Equation (2.8) is the solution to the diffusion equation with the same delta function initial conditions as those in equation (2.5).

2.2.2 The velocity jump model in higher dimensions

The foregoing analysis was specific to the one-dimensional case. We now consider the general VJ process in d spatial dimensions, with particular emphasis on the case $d = 2$, since the experimental data available to us are two-dimensional. In addition, limiting the dimensionality simplifies some of the mathematical analysis. In the general VJ process, particles move in d spatial dimensions and undergo reorientations according to Poisson process with rate λ , hence the time intervals between reorientation events are exponentially distributed. During the reorientation process, particles undergo a random change in velocity, described by a specified pdf. Particles are assumed to maintain a constant velocity between reorientation events, so that the trajectory comprises straight line movements between the sites of reorientation. We consider two models, in which reorientations either take place instantaneously, called a *run-only* VJ process, or are accompanied by a stationary rest phase of finite duration, denoted the *run-and-stop* VJ process. These two processes approximate the observed motion of *E. coli* and *R. sphaeroides*, respectively (see Figure 1.3). The run-only VJ process is a reasonable description of the run-and-twiddle motion observed in *E. coli*, as the reorientations take place more rapidly than in *R. sphaeroides*. We assume that the duration of these reorientations is negligible. We do not discuss this further, as we mainly consider *R. sphaeroides* throughout this thesis.

As for the preceding one-dimensional case study, we may consider both macroscopic and microscopic descriptions of the general VJ process. Since the macroscopic approach is based on considering an appropriate large-number limit of the microscopic model, we start with the latter.

The microscopic run-only velocity jump process

With the exception of stochastic simulations (see Section 2.5), there are few examples of microscopic descriptions of the VJ process in the literature. Figure 2.5 illustrates the run-only VJ process in two dimensions. The trajectory of the particle comprises a series of constant velocity runs, which we denote running phases, with instantaneous jumps in velocity between phases. The duration of each running phase is random. Various species are believed to move in a manner that is well described by the VJ process, in the sense that their motility takes the form of approximately straight line movements, interspersed by stochastic reorientations. In addition to bacteria, the random motion of ovipositing butterflies [103], foraging bumblebees [33] and elk [73], and the clonal growth of the plant species *Solidago altissima* [41] are examples of such species. In the case of the butterfly and bumblebee flight, the trajectories are generated by only considering landing sites and joining these with straight lines. Elk foraging motion consists of approximately straight journeys between pauses. The plant growth was measured as a series of distances between branching points where new growth starts, separated by approximately straight rhizomes.

A slightly different application of the VJ process is to species whose movement is not well described by the model initially, but which conforms after spatial or temporal discretisation. The process of discretisation involves decomposing the continuous underlying trajectory into a series of straight line moves with constant time or length intervals [103]. Since any device used to observe animal motion or track animals can only record at discrete time points, this is performed automatically at the stage where data are gathered. Nonetheless, further discretisation may be required to achieve the best possible characterization of the movement pattern [103]. This approach has been used to model movement in algae [88], ants [33], dolphins [16] and zebra [36]. In such cases, the VJ model may have no biological relevance to the actual underlying motion. It is instead used as a method to aid mathematical analysis of the movement patterns [18].

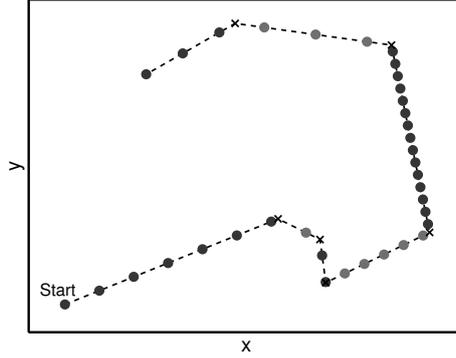


Figure 2.5: *Illustration of the two-dimensional run-only VJ process. A single particle moves in two dimensions (denoted by the black line), reorientating instantaneously at random intervals (indicated by black crosses). Filled circles show the location of the particle at regular time intervals. Reorientations, shown as black crosses, result in a change in the speed and angle of motion.*

An important early application of a VJ process to a biological problem was published in 1974 by Nossal and Weiss [144]. The authors derive expressions for the first two spatial moments of the position of a cell in two spatial dimensions after n reorientation events. This is based on a model of cell motion in two dimensions, equivalent to that illustrated in Figure 2.5, in which the rate of reorientation is a function of the direction of travel. Importantly, the distribution of initial angle of travel is assumed to be uniform on a circle. Kareiva and Shigesada [103] used a simplification of this representation, in which the turning rate is assumed to be constant, to model the movement of butterflies and caterpillars. The authors derive an expression for the mean squared displacement (MSD) from the starting location after n reorientations, which they propose as a test for whether observed data are compatible with the VJ model. This work was extended by McCulloch and Cain [131], who determine the analytic form of the variance in squared displacement.

An additional microscopic property of VJ trajectories is their *tortuosity*. This is a measure of how many twists and turns a trajectory contains. There are many different definitions used for calculating the tortuosity of a trajectory, which vary depending on the application [46]. Intuitively, we expect that more frequent reorientation events and a broader distribution of reorientation angles both lead to more tortuous trajectories in the

VJ process. This has been mathematically formalised in two studies. In the first, Bovet and Benhamou [33] define a property known as *sinuosity*, and establish a semi-empirical relationship between the sinuosity of a trajectory in terms of a fixed step length (the spatial rediscritisation interval). Later, Benhamou [18] derive a more general expression for the sinuosity of a VJ trajectory with a fixed time interval, in which steps are not necessarily of constant length. A key requirement of the derivation is that successive angle changes are independent, thus Benhamou asserts that it may be necessary to apply temporal rediscritisation to the data in order to avoid correlation between successive angle changes at high sampling rates. We shall return to the topic of tortuosity in Chapter 4.

Macroscopic equations governing the general velocity jump model

The continuum description of the VJ process in higher dimensions is analogous to the one-dimensional problem, however the full problem has no known analytic solution [146]. In the general model, particles run in a straight line with a given velocity $\mathbf{v} \in \mathbb{R}^d$ and switch velocity instantaneously as a Poisson process with rate λ . The density of particles at a position in the interval $[\mathbf{x}, \mathbf{x} + d\mathbf{x}]$, where $\mathbf{x} \in \mathbb{R}^d$, with a velocity in $[\mathbf{v}, \mathbf{v} + d\mathbf{v}]$ at time t is denoted $p(\mathbf{x}, \mathbf{v}, t)$. As for the one-dimensional case, we may interpret the quantity $p(\mathbf{x}, \mathbf{v}, t)$ as the number density of a large population of identical, non-interacting particles, or equivalently the pdf for the position and velocity of a single particle. Note that we require a continuous dependence on the velocity, in contrast with the fixed speed one-dimensional process, where it suffices to define two variables for the two possible particle velocities. For brevity, we henceforth use the more succinct notation $p \equiv p(\mathbf{x}, \mathbf{v}, t)$, except where the independent variables differ.

We now give a derivation of a macroscopic equation describing the evolution of the particle density. This was originally described by Othmer *et al.* [146]; here, we reproduce their results with additional explanation of the steps required. The VJ description of bacterial

chemotaxis has been generalised more recently to incorporate internal cellular dynamics [71]. We do not consider such models here, as the main focus of this thesis is to bacterial motility in a homogeneous environment. We begin by considering an expression for the total derivative of p with respect to t (also known as the material derivative). By the chain rule, we have

$$\frac{dp}{dt} = \frac{\partial p}{\partial t} + \frac{d\mathbf{x}}{dt} \cdot \nabla_{\mathbf{x}} p + \frac{d\mathbf{v}}{dt} \cdot \nabla_{\mathbf{v}} p. \quad (2.9)$$

Newton's laws of motion for a body of mass m subjected to a force \mathbf{F} are given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}, \quad (2.10)$$

$$\frac{d\mathbf{v}}{dt} = \frac{\mathbf{F}}{m}. \quad (2.11)$$

Substituting these into (2.9) yields

$$\frac{dp}{dt} = \frac{\partial p}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} p + \frac{\mathbf{F}}{m} \cdot \nabla_{\mathbf{v}} p, \quad (2.12)$$

which gives the total time derivative of the density of particles with individual mass m in the presence of an external force \mathbf{F} . The external force could describe, for example, an imposed external flow due to the environment. We will henceforth assume that no such external force exists, as none is present in the data that we will be considering.

Equation (2.12) is known as a *transport equation*. The total derivative of the density of particles is given by considering any reproduction and death processes and/or stochastic events relating to the motion. We neglect reproduction and death processes, which generally occur on a much longer timescale than that considered here, as discussed in Section 2.1. The total derivative is therefore equal to a stochastic component that describes re-orientation events and is fundamental to the VJ process. This aspect of the model differs between the run-only and run-and-stop variants.

Run-only velocity jump process

We define a turning kernel, $T(\mathbf{v}, \mathbf{v}')$, which gives the pdf for a change in velocity from \mathbf{v}' to \mathbf{v} . This is non-negative and normalised, such that

$$\int T(\mathbf{v}, \mathbf{v}') d\mathbf{v} = 1, \quad (2.13)$$

where the vector integral notation is shorthand for multiple integration over all d dimensions,

$$\int f(\mathbf{v}) d\mathbf{v} = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(v_1, \dots, v_d) dv_1 \dots dv_d. \quad (2.14)$$

The limits of integration will henceforth be over \mathbb{R}^d for all vector integrals unless otherwise stated. The normalisation condition (2.13) is required to ensure that number density is conserved across jumps, so that no particles are lost or gained during reorientation events. For a fixed initial velocity, \mathbf{v}' , we define the mean velocity after reorientation by

$$\bar{\mathbf{v}} = \int \mathbf{v} T(\mathbf{v}, \mathbf{v}') d\mathbf{v}. \quad (2.15)$$

Similarly, the mean speed after reorientation is given by

$$\bar{c} = \int \|\mathbf{v}\| T(\mathbf{v}, \mathbf{v}') d\mathbf{v}, \quad (2.16)$$

where $\|v\|$ denotes the Euclidean norm of vector \mathbf{v} . The mean cosine of reorientation is then defined as

$$\psi_d = \frac{\bar{\mathbf{v}} \cdot \mathbf{v}'}{\bar{c} \|\mathbf{v}'\|}, \quad (2.17)$$

where $\psi_d \in [-\pi, \pi)$. Note that $\bar{\mathbf{v}}$, \bar{c} and ψ_d may all be dependent upon the initial velocity \mathbf{v}' .

In the run-only model, the total derivative is equal to the change in p as a result of

stochastic reorientation, given by

$$\frac{dp}{dt} = -\lambda p + \lambda \int T(\mathbf{v}, \mathbf{v}') p(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}'. \quad (2.18)$$

The first term on the right-hand side of this equation represents the movement of particles out of the velocity space $[\mathbf{v}, \mathbf{v} + d\mathbf{v}]$ due to velocity jumping. The second term represents the reorientation of particles from all other velocities \mathbf{v}' to the interval $[\mathbf{v}, \mathbf{v} + d\mathbf{v}]$. Equating (2.12) and (2.18), we obtain

$$\frac{\partial p}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} p = -\lambda p + \lambda \int T(\mathbf{v}, \mathbf{v}') p(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}'. \quad (2.19)$$

The partial integro-differential transport equation (2.19) describes the evolution of the number density of particles undergoing a run-only VJ process.

Run-and-stop velocity jump process

We now consider the addition of a finite duration stationary phase in between runs, during which the particle reorientates. It is assumed in this model that particles maintain a perfect memory of their previous direction over the course of a stop of any duration, such that the same turning kernel applies as in the run-only model. Switches from the running phase into the stationary rest phase and the reverse transitions are modelled as Poisson processes with rate parameters λ and μ , respectively. The run-and-stop VJ process and the derivation of the governing transport equation were originally presented in [146]. The density of moving particles, $p(\mathbf{x}, \mathbf{v}, t)$, is defined as before. We also introduce a separate subpopulation of stopped particles. The density of stopped particles at a position in the interval $[\mathbf{x}, \mathbf{x} + d\mathbf{x}]$ with a previous velocity in $[\mathbf{v}, \mathbf{v} + d\mathbf{v}]$ at time t is denoted $r(\mathbf{x}, \mathbf{v}, t)$.

In this model, the equation describing the stochastic process is given by

$$\frac{dp}{dt} = -\lambda p + \mu \int T(\mathbf{v}, \mathbf{v}') r(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}'. \quad (2.20)$$

Equation (2.20) differs from the run-only equation (2.18) only in the integrand: in the run-and-stop model, particles switch out of the stationary phase into the running phase. We also require a description of the population of stationary particles, which is given by

$$\frac{\partial r}{\partial t} = \lambda p - \mu r. \quad (2.21)$$

From equations (2.12) and (2.20) we obtain

$$\frac{\partial p}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} p = -\lambda p + \mu \int T(\mathbf{v}, \mathbf{v}') r(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}'. \quad (2.22)$$

The transport equations describing the evolution of particles in the run-and-stop VJ process are thus given by equations (2.21)-(2.22).

2.3 Spatial moments of the general velocity jump process

When comparing mathematical models of motion to experimental data, the data often only reliably provide a few spatial moments of the observed motion [146]. In addition, the run-only and run-and-stop VJ models described above have no known closed analytic solution, and are computationally intensive to simulate numerically. Fortunately, it suffices in some cases to model the temporal evolution of the spatial moments of the particle density. The MSD is the second spatial moment of a movement process. It is commonly used in the fields of ecology [103] and biological physics [132, 134] as a measure of the rate at which a trajectory samples its environment. The MSD of a vector-valued random variable, $\mathbf{X} \in \mathbb{R}^d$, with pdf $f_{\mathbf{X}}(\mathbf{x})$ is defined by

$$\int f_{\mathbf{X}}(\mathbf{x}) \|\mathbf{x}\|^2 d\mathbf{x}.$$

Othmer *et al.* [146] show that it is possible to derive a closed analytic expression for the MSD of a VJ process directly from the transport equations. Both the run-only and run-and-stop variants are considered. We reproduce the relevant results here; full details of the derivations are given in Appendix A.3. In order to make analytic progress, we assume that the speed of a particle does not change over the course of a reorientation, only the angle. Furthermore, we assume, without loss of generality, that all particles are initially located at the origin.

With these assumptions, the MSD of the run-only VJ process is given by

$$\mathcal{D}_p^2(t) = \begin{cases} \frac{2\mathcal{S}_p^2}{\lambda_0} \left[t - \frac{1}{\lambda_0} (1 - e^{-\lambda_0 t}) \right] & \text{for } \psi_d \neq 1, \\ \mathcal{S}_p^2 t^2 & \text{for } \psi_d = 1, \end{cases}, \quad (2.23)$$

where $\lambda_0 = \lambda(1 - \psi_d)$, and \mathcal{S}_p^2 is the second moment of the velocity (mean squared speed). As we have assumed that the speed of a particle does not change following a reorientation event, \mathcal{S}_p^2 is constant and equal to the mean squared initial speed of the particles. The expression for the MSD in equation (2.23) exhibits different behaviour on different timescales when $\psi_d \neq 1$. For small t , a Taylor expansion of the exponential component shows that $\mathcal{D}_p^2(t) \approx \mathcal{S}_p^2 t^2$, which is indicative of a ballistic, or wave-like, process. In the limit $t \rightarrow \infty$, $\mathcal{D}_p^2(t) \approx 2\mathcal{S}_p^2 t$, which indicates a diffusive process. This is illustrated in Figure 2.6. The interpretation is that particles initially have a strong directional persistence, hence the population spreads out like a wave. This directional persistence is lost over time, and the process instead becomes diffusive. Othmer *et al.* [146] define a characteristic persistence time by $P = \frac{1}{\lambda_0}$. This is equal to the intercept of the limiting linear MSD regime with the x -axis, as indicated in Figure 2.6.

In the case of the run-and-stop VJ process, we have two populations, runners and stoppers. In reality, it is not practical to compute the MSD of running and stopped particles

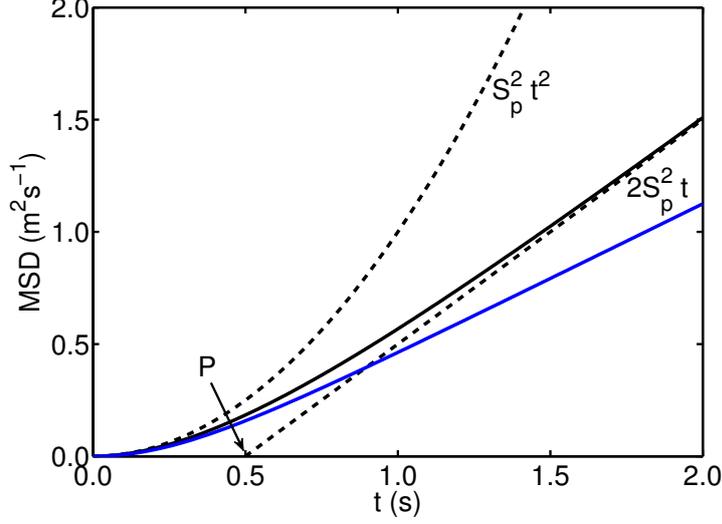


Figure 2.6: Theoretical MSD for a run-only (black solid line, equation (2.23)) and run-and-stop (blue line, equation (2.25)) VJ process with $\lambda = 4 \text{ s}^{-1}$, $\psi_d = 0.5$, $\mathcal{S}_p^2 = 1 \text{ m}^2 \text{ s}^{-1}$ and $\mu = 8 \text{ s}^{-1}$ for the run-and-stop case. The dashed lines show limiting behaviour, as marked.

separately; instead the MSD observed experimentally is that of the total population. We therefore solve the system of equations for the weighted MSD, defined by

$$\mathcal{D}^2(t) = \frac{N_p \mathcal{D}_p^2 + N_r \mathcal{D}_r^2}{N_0}, \quad (2.24)$$

where N_p and N_r are the total populations of runners and stoppers, respectively, and $N_0 = N_p + N_r$ is the total number of particles. We assume that the populations of running and stopping states are initially in equilibrium, which is the case in most experimental systems, providing the system is not deliberately perturbed immediately prior to observing it. As before, we assume that the speed of a particle does not change following a reorientation event, and that all particles are initially located at the origin. The solution subject to these conditions is then given by

$$\mathcal{D}^2(t) = \frac{2\mathcal{S}_p^2}{\lambda_0} \frac{\mu}{\lambda + \mu} \left\{ t + \frac{\lambda_+ - \lambda_0}{\lambda_- (\lambda_+ - \lambda_-)} (e^{-\lambda_- t} - 1) - \frac{\lambda_- - \lambda_0}{\lambda_+ (\lambda_+ - \lambda_-)} (e^{-\lambda_+ t} - 1) \right\}, \quad (2.25)$$

where the coefficients λ_{\pm} are defined by

$$\lambda_{\pm} = \frac{\lambda + \mu}{2} \left[1 \pm \sqrt{1 - \frac{4\lambda_0\mu}{(\lambda + \mu)^2}} \right]. \quad (2.26)$$

The values of λ_{\pm} are both real and positive, so the asymptotic behaviour of the weighted MSD as t becomes large is linear. This indicates that the run-and-stop VJ process becomes diffusive in the long time limit, as for the run-only case. The MSD for a run-and-stop VJ process is plotted in Figure 2.6. The parameters are chosen such that the mean duration of stops is half that of runs. The MSD has a similar form in both the run-only and run-and-stop cases, but is always lower for the run-and-stop case, since the stationary phase retards the propagation of particles.

2.4 Circular statistics

Reorientations in both the microscopic and macroscopic approaches in two or three spatial dimensions are naturally described in terms of a simultaneous change in angle and speed. Linear statistics cannot be applied to describe random changes in angle, due to the periodic nature of angular variables. Instead, we must apply circular statistics, which are extensively covered by Mardia and Jupp [130]. Throughout this thesis, we are only concerned with two-dimensional circular statistics, with all angles defined in the range $[-\pi, \pi)$. A pdf defined on a circle, denoted a *wrapped* pdf, must be non-negative and normalised, i.e.

$$f_{\Theta}(\theta) \geq 0 \quad \forall \theta \in [-\pi, \pi), \quad (2.27)$$

where asymmetric limits are included by convention, and

$$\int_{-\pi}^{\pi} f_{\Theta}(\theta) \, d\theta = 1. \quad (2.28)$$

The angular moments of a wrapped pdf[47] are defined by

$$a_n = \int_{-\pi}^{\pi} \cos(n\theta) f_{\Theta}(\theta) d\theta, \quad (2.29)$$

$$b_n = \int_{-\pi}^{\pi} \sin(n\theta) f_{\Theta}(\theta) d\theta, \quad (2.30)$$

where $n = 1, 2, \dots$. From these definitions, it is clear that $b_n = 0$ for all symmetric wrapped pdfs. Writing the angular moments in polar form, we obtain

$$\rho_n e^{i\phi_n} = a_n + ib_n, \quad (2.31)$$

where ρ_n and ϕ_n are the n^{th} length and angle moments, respectively. There are two definitions of the standard deviation of a wrapped pdf. The first is defined by analogy with the standard deviation in linear statistics [47]

$$\sigma = \sqrt{2(1 - \rho_1)}, \quad (2.32)$$

An alternative definition, arising from the wrapped normal distribution, is given by

$$\sigma' = \sqrt{-2 \ln(\rho_1)}. \quad (2.33)$$

It is in principle possible to wrap any linear pdf around a circle to obtain a circular distribution, however the resultant form is often difficult to work with analytically. We now describe three relevant wrapped distributions, following Codling [47].

2.4.1 The wrapped uniform distribution

The uniform distribution on a circle has pdf

$$f_{\Theta}(\theta) = \frac{1}{2\pi}. \quad (2.34)$$

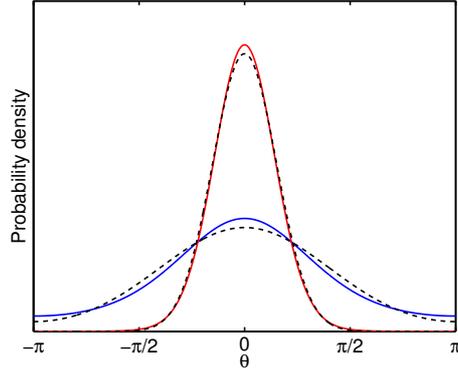


Figure 2.7: A linear plot of the von Mises distribution with zero mean for $\kappa = 1$ (blue line) and $\kappa = 5$ (red line), and the wrapped normal distributions with equivalent standard deviations (black dashed lines).

This distribution has mean length $\rho_1 = 0$ and standard deviation $\sigma = \sqrt{2}$. The alternative definition of standard deviation, σ' , and the mean angle, ϕ_1 , are both undefined.

2.4.2 The wrapped normal distribution

The wrapped normal distribution is defined by

$$f_{\Theta}(\theta; \mu, \sigma') = \frac{1}{\sigma' \sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp \left[-\frac{(\theta + 2\pi k - \mu)^2}{2\sigma'^2} \right], \quad (2.35)$$

where $\mu = \phi_1$ is the mean angle and σ' is the standard deviation. The mean length is related to standard deviation by $\rho_1 = e^{-\sigma'^2/2}$, from which we obtain equation (2.33).

2.4.3 The von Mises distribution

Owing to the complicated form of the wrapped normal distribution in equation (2.35), the von Mises distribution is often used as a good approximation to it [49]. The von Mises distribution is defined by

$$f_{\Theta}(\theta; \mu, \kappa) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}, \quad (2.36)$$

where I_0 denotes the modified Bessel function of the first kind with order 0, $\mu = \phi_1$ is the angular mean and $\kappa \geq 0$, denoted the concentration parameter, controls the ‘peakedness’ of the distribution [130]. Note that when $\kappa = 0$, the von Mises distribution is equivalent to the wrapped uniform distribution. As $\kappa \rightarrow \infty$, the distribution becomes sharply peaked around the mean angle [130]. The length moments for the von Mises distribution are given by

$$\rho_n = \frac{I_n(\kappa)}{I_0(\kappa)}, \quad (2.37)$$

from which we can equate the concentration parameter and the standard deviation of the wrapped normal using $\bar{R} = \rho_1$ and equation (2.33) [130]. The von Mises distribution is plotted in Figure 2.7 for two values of κ , overlaid with the wrapped normal distribution with equivalent standard deviation.

2.5 Stochastic simulation algorithm

In the absence of an analytic solution to the general VJ process, we must resort to numerical simulations to compute the outcome of the model for a given set of parameters. In two or more spatial dimensions, the deterministic numerical solution of equation (2.19), or equations (2.21)-(2.22), which govern the run-only and run-and-stop VJ processes, respectively, is a challenging problem. Taking position and velocity components into account, the equation is posed in four or more dimensions, rendering numerical integration extremely computationally intensive. Whilst some work has been conducted into finding numerical solutions for related processes [172], the results are not immediately generalisable to our system. We therefore use a stochastic computational simulation to generate realisations of the desired VJ process for a given set of parameters, a strong precedent for which has been set in the related literature [40, 49, 50, 70, 92, 97].

We simulate identical, non-interacting particles undergoing a VJ process in one or two spatial dimensions, though the algorithm has also been extended to three-dimensional

problems. The particles run in straight lines and undergo reorientation events as a Poisson process. Reorientation events are fully described by a specified run speed and angle change distribution. For example, many of the simulations carried out in this thesis are performed with a constant speed movement process, with turning angles drawn from the von Mises distribution. In the run-only simulation, reorientation events are instantaneous, with waiting times between events drawn independently from the exponential distribution with rate parameter λ . In the run-and-stop simulation, reorientations follow a stationary phase, whose duration is drawn from the exponential distribution with rate parameter μ . The stochastic trajectory of a single particle is simulated in this way until the time exceeds a predefined value, at which point the trajectory is ceased. This is repeated for the desired number of trajectories.

The underlying movement model described by successive reorientation events is continuous in time and space. In order to simulate the process of observing a moving particle, we discretise these trajectories using a defined constant sampling interval. The position of each simulated particle is calculated at sampling times separated by the sampling interval under the assumption that a constant velocity is maintained between reorientation events.

We are also able to incorporate the effect of measurement noise and Brownian buffeting into the simulations. This will be discussed in greater detail in Chapters 3 and 6. Figure 2.8 shows an illustration of the spatial density of a number of particles undergoing a run-only VJ process, computed using the stochastic simulation algorithm. Initially, all particles are located at the origin and travelling in the positive x -direction. In agreement with the results and discussion in Section 2.3, the evolution of the particle density displays ballistic and diffusive behaviour. The observed wavefront travels radially outwards in a ballistic manner, whilst simultaneously spreading and becoming less concentrated around a single density peak.

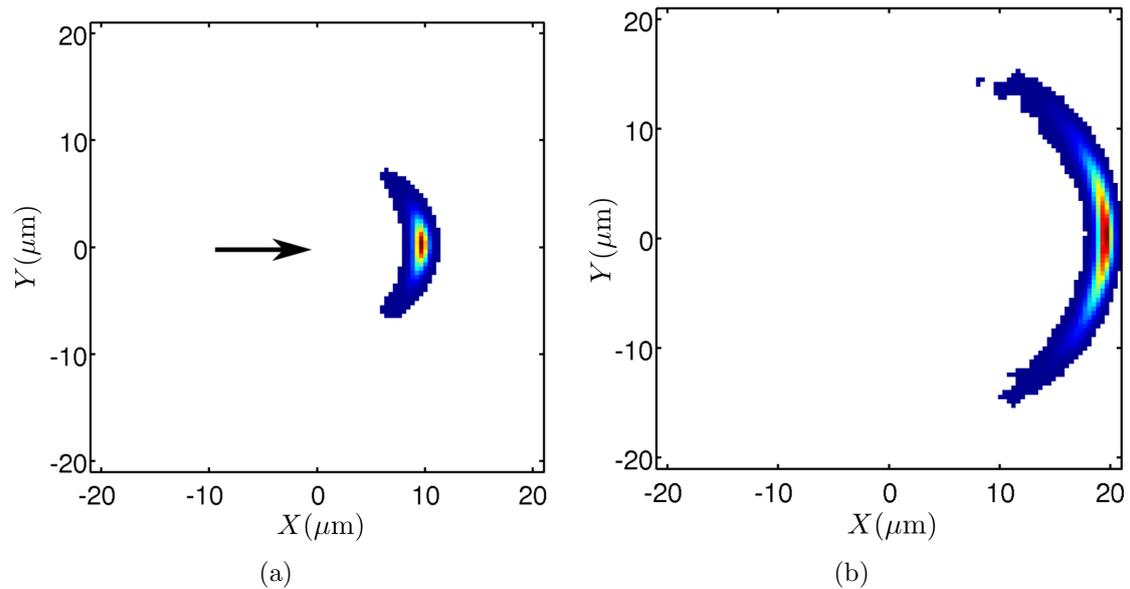


Figure 2.8: The density of particles computed using a stochastic simulation of 1000 particles undergoing a run-only VJ process with reorientation rate $\lambda = 1 \text{ s}^{-1}$, a constant run speed of $1 \text{ } \mu\text{m s}^{-1}$ and von Mises angular distribution ($\kappa = 100$), at times (a) $t = 10 \text{ s}$ and (b) $t = 20 \text{ s}$. Colour indicates density (arbitrary scale) from blue (low density) to red (high density). The arrow in (a) indicates the initial velocity of all particles.

2.6 Conclusion and outlook

In this chapter, we have reviewed the mathematical approaches relevant to modelling bacterial motility. We showed that the biological processes of bacterial motility and taxis occur over many orders of temporal and spatial scales. The bacterial tracking data presented in Chapter 4 are well-described by a VJ process in the absence of any gradient of chemoattractant. In Chapters 3, 5 and 6, we will use the VJ framework to describe the phenomenological process of bacterial motility, with comparisons drawn and reference made to these datasets. We therefore reviewed the relevant foregoing mathematical methods and analysis associated with the VJ process. Furthermore, we discussed the stochastic numerical method used throughout the remainder of this thesis.

Chapter 3

Analysis methods for inferring stopping phases in tracking data

In order to characterise the motile behaviour of *R. sphaeroides* (or indeed any other planktonic bacteria), various information relating to their motion is required. In the specific case of *R. sphaeroides*, we model the motion using a run-and-stop VJ process, as detailed in Chapter 2. Quantities of interest, which relate to the parameters in the VJ model, are (i) the rate of switching from a run to a stop (λ , measured in s^{-1}), (ii) the rate of switching from a stop to a run (μ , measured in s^{-1}), (iii) the run speed (c , measured in μms^{-1}), and (iv) the angle change between stops (θ , measured in radians). We denote this final quantity the *stopwise angle change* for brevity. These quantities are not necessarily constant and may be drawn randomly from a distribution, either across the whole population or for each individual. A further complicating factor is that interdependencies may exist between these quantities. For example, the stopwise turning angle may be correlated with the speed of swimming immediately before a stop.

In order to construct and parameterise a VJ model of the motility exhibited by *R. sphaeroides*, we should first consider what experimental data are available relating to the system, and how this informs the choice of our model. There is generally little advantage

in posing a model that contains more parameters than can be unambiguously assigned from available data. This imposes several restrictions. First, due to the limited field of view of the microscope, individual observed tracks are often short as bacteria swim out of view. In addition, in bulk swimming experiments, bacteria are lost when they swim out of focus, terminating the track. We cannot therefore reliably extract distributions relating to individual bacteria due to a paucity of data in many of the tracks, and are limited to population distributions. Hence we assume that the waiting times governing the durations of runs and stops, run speeds and angle changes made by bacteria are drawn from the same population-wide distribution for all individuals. In reality, there is generally a great deal of heterogeneity in the phenotype of clonal bacterial populations due, in part, to the stochastic nature of gene expression [67]. We discuss the validity of this assumption further in Chapter 4. A second restriction relates to correlation between variables. In order to incorporate correlation, we would need to measure the joint distribution of the variables of interest (for example the joint distribution of run speed and stopwise turning angle). There are generally insufficient data to extract joint distributions from the tracking data and we therefore assume independence of all the quantities listed above.

As discussed in Chapter 2, we assume for the purposes of the VJ model that switching occurs as a Poisson process with defined constant rate parameters. The waiting times between events are therefore exponentially distributed with a rate parameter, μ , for switching from stops to runs, and another rate parameter, λ , for switching from runs to stops. Each parameter is assumed to be constant for all bacteria within the population. Similarly, we make no attempt to measure the distribution of run speeds in individual bacteria, due to a paucity of data. We instead make the simplifying assumption that the observed ensemble distribution of run speeds represents the whole bacterial population. The same is true for the stopwise angle change distribution: each turning angle is assumed to be randomly drawn from a population ensemble distribution. Following the model proposed by Campos and Méndez [42], we assume that at each stop-to-run transi-

tion event both a new relative angle and a new running speed are drawn at random from the underlying distributions governing these quantities. Bacteria then run at a constant speed until the next reorientation event.

In order to infer the parameters for our VJ model with the assumptions made above, we seek to extract the following characteristics from the microscope videos:

1. the rate constants, λ and μ , governing the transitions between running and stopped states;
2. the run speed for each frame in which the cell is running, denoted the framewise speed;
3. the stopwise angle change, defined as the angle difference between the running direction immediately before a stop and the running direction immediately following a stop in a run-stop-run sequence.

The quantities of interest listed above have the common prerequisite that we are able to label the running and stopped phases in a track. Without this information, we are not able to extract quantities relating to differences between the two phases in an automated manner. This labelling process is a central theme to the present chapter, in which we present automated analysis methods to segment tracks into running and stopped phases. As we shall see in Chapter 4, in addition to tracking data from wildtype bacteria, crucial biological insight is gained from tracking data obtained from two mutant strains of *R. sphaeroides*, one of which is unable to stop (non-chemotactic) and the other of which is non-motile. Throughout this chapter, we assume that this information is available to us when analysing bacterial tracks. We take advantage of the availability of these additional datasets when analysing tracks, using them as a proxy for the two phases of movement in wildtype tracks, as discussed below. The process of generating tracks from microscope videos, details of experimental methods and additional analysis steps required when deal-

ing with real data are omitted here and discussed in detail in Chapter 4.

In this chapter, we present novel analysis methods, based on a hidden Markov model (HMM), to discern stopping and running phases in the observed tracks. There is no established ‘gold standard’ method for analysing bacterial tracks. Many experimental investigations of bacterial motility have made use of the Hobson Tracker [89], a piece of equipment originally designed to track spermatozoa that simultaneously performs tracking and identifies stopping phases in the motion. However, no literature is available explaining the internal workings of the device and it is known to suffer from hardware limitations in the number of bacteria that may be simultaneously tracked [169]. In absence of information relating to the the analysis methods employed in the Hobson Tracker, we therefore do not consider it further. The lack of a standard analysis method motivates a comparison of different analysis techniques in order to establish how suitable they are for this purpose. We use simulated tracking data to compare the novel analysis methods with an existing procedure that is based on heuristic arguments. This is, to our knowledge, the first example of a systematic comparison of analysis methods. We also show how the rate parameters mentioned above can be independently estimated when working with the approaches based on the HMM.

3.1 Analysis methods

The tracking algorithm used to generate tracks from experimental data, discussed in Chapter 4, in common with all other tracking methodologies applicable to this field, outputs data as a vector of the form $\mathbf{R} = (\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_T)$ for each track, where \mathbf{r}_t denotes a two-dimensional position vector at frame t , and the number of frames in the track is given by $T + 1$. Note that t is considered a discrete quantity throughout, as time is measured in numbers of frames. In characterising running and stopped phases, we are concerned not with the positions of cells in each frame, but with the motion of cells

between consecutive frames. The information of interest is thus the transitions between consecutive position vectors within a track. These form a vector of displacement vectors, $\mathbf{D} = (\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{T-1})$ with $\mathbf{d}_t = \mathbf{r}_{t+1} - \mathbf{r}_t$. The framewise speed is defined as the observed speed of travel between two consecutive frames, $\|\mathbf{d}_t\|/\Delta t$, where Δt is the time interval between frames. Note that we assume that consecutive images are acquired with a fixed time interval, as is the case in most microscope video applications.

It will also be useful to compute the framewise angle changes between consecutive vectors, which form a vector $\Theta = (\theta_0, \theta_1, \dots, \theta_{T-2})$, where $\theta_t \in [-\pi, \pi]$ is the difference in polar angle between \mathbf{d}_{t+1} and \mathbf{d}_t . Framewise angle changes are computed following the procedure in Algorithm 1, where $\arctan(y, x)$ is corrected for the quadrant in which the vector lies, computed using the built-in `atan2` function in Matlab. The representation of the data described above is illustrated in Figure 3.1.

Algorithm 1 Calculating framewise angle changes for a track

```

for  $i = 0$  to  $T - 2$  do
    Compute angle of  $\mathbf{d}_i$  relative to the horizontal,  $\phi' = \arctan(\mathbf{d}_{i,y}, \mathbf{d}_{i,x})$ .
    Compute angle of  $\mathbf{d}_{i+1}$  relative to the horizontal,  $\phi = \arctan(\mathbf{d}_{i+1,y}, \mathbf{d}_{i+1,x})$ .
    Compute angle difference,  $\theta_i = \phi - \phi'$ .
    if  $\theta_i > \pi$  then
         $\theta_i = \theta_i - 2\pi$ .
    else if  $\theta_i < -\pi$  then
         $\theta_i = \theta_i + 2\pi$ .
    end if
end for

```

Defining stopping periods in bacterial swimming tracks is complicated by various sources of noise in the data. Uncertainty in the position of the centroid of a cell in each image may cause a track to appear jagged, for example when cells tumble whilst swimming. Brownian buffeting may also cause deviations from straight-line swimming, and lead to stops that are not perfectly stationary. Tracking errors caused by incorrectly linking cells between consecutive frames, or by the disappearance of a cell for one or more frames, may

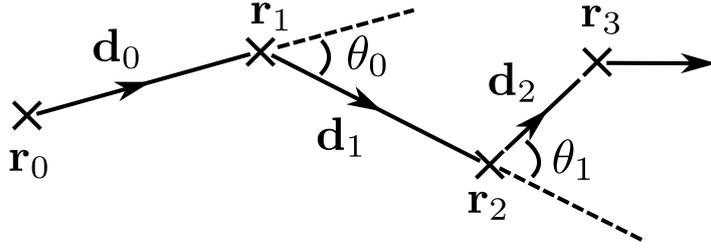


Figure 3.1: An illustration of the mathematical representation of the tracking data. The vectors \mathbf{r}_t denote the observed positions of the object being tracked at discrete sampling times t . The vectors \mathbf{d}_t are the displacement vectors between consecutive position vectors, and the angles θ_t are the angles between the displacement vectors.

also affect the appearance of a track. The observed framewise angle changes, θ_t , may be due to such sources of noise, or due to a genuine reorientation event. As a result of this added noise, the apparent motion exhibited during a stop phase may be misconstrued as a run and *vice versa*. The effect of noise will be explored in more detail in Chapter 6.

We assume that the displacement vectors of each track, \mathbf{d}_t , correspond to either a run state or stop state. The underlying state at frame t is denoted S_t , where we use the convention throughout that $S_t = 0$ corresponds to a stop and $S_t = 1$ corresponds to a run. For each track we define a state vector, $\mathbf{S} = (S_0, S_1, \dots, S_{T-1})$, describing the sequence of states. One aim of the analysis is to infer the underlying state sequence for each track. We therefore wish to assign to each displacement vector \mathbf{d}_t a probability of being in a run (or stop) phase,

$$P_t(i) = \mathbb{P}(S_t = i | \mathbf{d}_t),$$

where $i \in \{0, 1\}$.

Three methods for the determination of stops in tracks are now proposed, two of which are based on the HMM, followed by an additional post-processing method which may enhance the performance of all three methods. The analysis process is summarised in the flow diagram in Figure 3.2, the stages of which we shall now describe in detail.

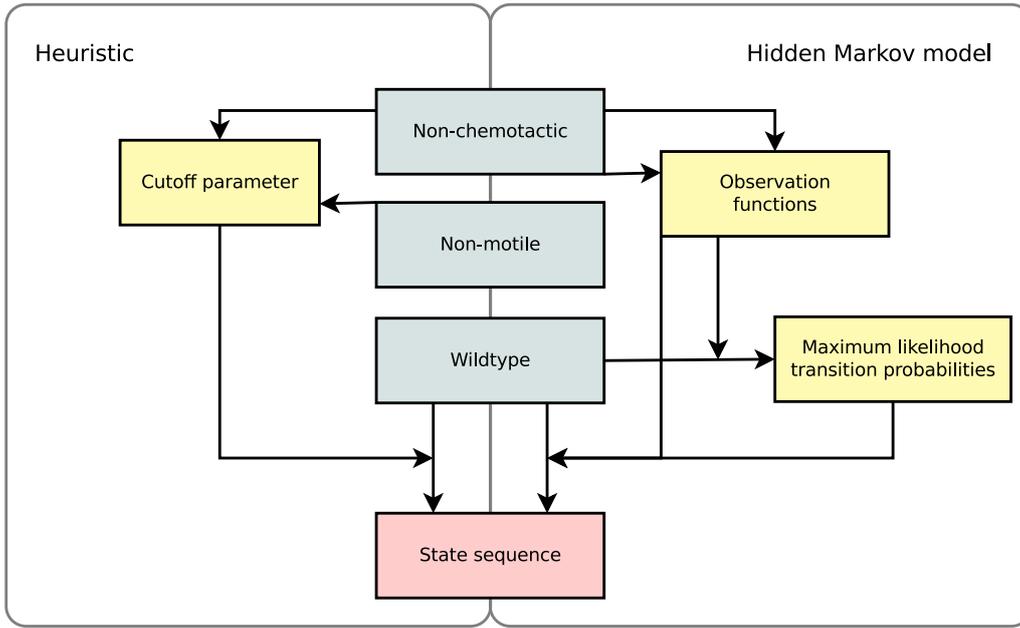


Figure 3.2: Flow diagram summarising the analysis process for the methods described in this chapter. Grey boxes denote inputs, in the form of simulated tracks, yellow boxes denote intermediate quantities required by the methods, and the red box denotes the output of the methods. Arrows indicate the flow of information in the process.

3.1.1 Heuristic analysis

The intuitive approach used by Taboada *et al.* [190] to identify stopping phases is to define a threshold parameter ρ_{CS} and denote each transition as a run if the framewise speed is greater than ρ_{CS} ,

$$P_t(1) = \begin{cases} 1, & \frac{\|\mathbf{d}_t\|}{\Delta t} \geq \rho_{CS}, \\ 0, & \frac{\|\mathbf{d}_t\|}{\Delta t} < \rho_{CS}. \end{cases} \quad (3.1)$$

We call this approach the *heuristic method*.

The single parameter to be determined in the heuristic analysis method is the threshold framewise speed, ρ_{CS} . The basis of the heuristic method is that there is a substantial difference between the distribution of speeds observed during framewise runs and stops.

If there is any overlap between the framewise speed distribution for run phases and stop phases, due to the effect of noise and/or population heterogeneity, then this approach will cause spurious inference in the crossover region. The value of ρ_{CS} should be selected to maximise the number of correctly inferred transitions. We discuss the process of optimising ρ_{CS} in Section 3.3.2.

3.1.2 Analysis using a hidden Markov model

A more sophisticated approach utilises a state space model to infer the state probabilities. We apply a HMM to the observed data. A brief discussion of the HMM follows; a more detailed tutorial on the HMM is given in [164] and a good reference for details of the implementation of the HMM is [161].

A Markov chain describes a directed graph linking M states with known transition probabilities. Each state is directly observable and the transitions are ‘memoryless’: the probability of a transition is determined only by the state immediately prior to that transition, a condition known as the Markov property. The system may be described in terms of a list of transition probabilities, the initial state and a vector of state configurations evolving over time. The transition probabilities give the probability of a transition occurring between two states in a time interval, dt , denoted $A_{ij} dt = \mathbb{P}(S_{t+dt} = j | S_t = i)$, where $i, j \in \{1, \dots, M\}$. We assume that the Markov chain is homogeneous, meaning that the transition probabilities are invariant with respect to time; the extension to time-dependent transition probabilities is straightforward. This model, with $M = 2$, is a reasonable representation of the motion of *R. sphaeroides*; the bacterial locomotion is a series of runs and stops with the frequency of both being a function of chemoattractant concentration. In the absence of any concentration gradient we assume that the probability of switching from a run to a stop (or *vice versa*) is constant per unit time and hence that the transitions are Markovian. In addition, as discussed previously, we consider

time in discrete units of frames, so that the transition probabilities may be rewritten as framewise quantities:

$$A_{ij} = \mathbb{P}(S_{t+1} = j | S_t = i). \quad (3.2)$$

A *hidden* Markov model is appropriate for a similar system in which each state is not directly observable. In the current work, for example, the true run or stop states are not directly measurable. Instead the observable is the vector of data, $\mathbf{Y} = (\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_{T-1})$, where each datum, \mathbf{Y}_t , is a random variable. The observed data are related in some way to the underlying state. This relationship is expressed by the observation function, defined by

$$b_i(\mathbf{y}_t) = f_{\mathbf{Y}_t | S_t}(\mathbf{y}_t | i), \quad (3.3)$$

which gives the pdf of observing the datum \mathbf{y}_t at time t , conditional on the system being in state i . The model, denoted Ψ , is fully described by $b_i(\mathbf{y}_t)$ and A_{ij} .

The aims of the analysis are twofold. First, we wish to maximise the likelihood of the model, defined by

$$\mathcal{L}(\Psi | \mathbf{Y}) = f_{\mathbf{Y}}(\mathbf{Y}; \Psi), \quad (3.4)$$

where $f_{\mathbf{Y}}(\mathbf{Y}; \Psi)$ is the pdf governing the observed data for a given realisation of the model. We henceforth denote the likelihood by \mathcal{L} for brevity. We maximise this likelihood over all of the parameters required in the model, which for our purposes amounts to finding the correct transition probability parameters for switching between runs and stops, as discussed below. Second, applying the maximum likelihood parameters, we evaluate $P_t(i)$, where $i \in \{0, 1\}$. In order to compute these probabilities and \mathcal{L} , two new quantities must first be evaluated. The cumulative forward probability, defined by

$$\alpha_t(i) = f_{\mathbf{Y}_0, \dots, \mathbf{Y}_t | S_t}(\mathbf{y}_0, \dots, \mathbf{y}_t | i), \quad (3.5)$$

is the pdf of observing all data up to and including time t , conditional on the system

being in state i at time t . This quantity may be calculated by summing over all possible combinations of states up to time $t - 1$,

$$\alpha_t(i) = \sum_{\mathbf{S}} b_{S_0}(\mathbf{y}_0) \left(\prod_{k=0}^{t-2} A_{S_k S_{k+1}} b_{S_{k+1}}(\mathbf{y}_{k+1}) \right) A_{S_{t-1} i} b_i(\mathbf{y}_t), \quad (3.6)$$

with initial condition $\alpha_0(i) = b_i(\mathbf{y}_0)$, where $\mathbf{S} \in \{0, 1\}^t$ denotes all possible state sequences S_0, \dots, S_{t-1} . Equation (3.6) may be interpreted as a series of transitions between discrete states with transition probabilities given by A_{ij} , with each state emitting an observation governed by the pdf $b_i(\mathbf{y}_t)$. Two implicit assumptions are that the state is independent of the observation and that the observations are independent of previous observations [17]. The number of configurations over which this summation is carried out is 2^T , where T is the number of observations (frames in our context). In this form, the cumulative forward probability is computationally intractable for our purposes as T is typically of the order of hundreds. A recursion formula that simplifies this computation is given by

$$\alpha_t(i) = \sum_{j=0}^1 \alpha_{t-1}(j) A_{ji} b_i(\mathbf{y}_t). \quad (3.7)$$

The likelihood of observing the data conditional upon Ψ is obtained by summing over all possible final states:

$$\mathcal{L} = \sum_{i=0}^1 \alpha_T(i). \quad (3.8)$$

Estimation of the optimal model is achieved by maximising this likelihood.

The quantity $\alpha_t(i)$ alone does not take all of the available data into consideration for $t < T$, only the data up to the time t . We therefore introduce a second quantity, the cumulative backward probability, defined by

$$\beta_t(i) = f_{\mathbf{Y}_{t+1}, \dots, \mathbf{Y}_T | S_t}(\mathbf{y}_{t+1}, \dots, \mathbf{y}_T | i). \quad (3.9)$$

This quantity is the pdf of all data in the future of time t , given that the state at time t is i . A similar definition to (3.6) applies, given by

$$\beta_t(i) = A_{iS_{t+1}} b_{S_{t+1}}(\mathbf{y}_{t+1}) \sum_{\mathbf{S}} \left(\prod_{k=t+1}^{T-1} A_{S_k S_{k+1}} b_{S_{k+1}}(\mathbf{y}_{k+1}) \right), \quad (3.10)$$

where $\mathbf{S} \in \{0, 1\}^{T-t}$ again denotes all possible state probabilities S_{t+1}, \dots, S_T . By convention, $\beta_T(i) = 1$, as there are no data to the future of time point T . Another useful recursion relation also exists, given by

$$\beta_t(i) = \sum_{j=0}^1 \beta_{t+1}(j) A_{ij} b_j(\mathbf{y}_{t+1}). \quad (3.11)$$

The cumulative forward and backward probabilities may be combined to obtain the probability of a cell being in a given state at each time point,

$$P_t(i) = \frac{\alpha_i(t) \beta_i(t)}{\sum_{j=0}^1 \alpha_j(t) \beta_j(t)}. \quad (3.12)$$

The normalisation factor in the denominator is required as the only possibilities allowed by the model are the two states, although the actual probability of observing a complete sequence of states is very low.

The observation function encodes prior information about the motion model for bacterial swimming. Its functional form varies between the two HMM-based analysis methods and is discussed in more detail below.

3.1.3 Speed-only model

For this approach we use the framewise speed, $\|\mathbf{d}_t\|/\Delta t$, as the observable data, which are hence scalars. From experiments with stopped and running-only bacterial strains we are able to define the observation function $b_i(y_t)$ empirically. This function expresses our

prior knowledge of the process of bacterial swimming, and is specified by reference to the non-chemotactic and non-motile mutant strains. The former is unable to reorientate, swimming continuously in approximately straight lines, with deviations due to noise processes as discussed. We assume that this mutant behaves as a wildtype bacterium in a running phase. Similarly, the non-motile strain is unable to propel itself at all, hence we assume that this mutant behaves as a wildtype bacterium in a stopped phase. These assumptions are verified experimentally in Chapter 4. We reiterate that the non-motile strain is not perfectly stationary under the microscope due to various sources of noise, hence we must also record the motion in this case. As we discuss in Section 3.4, a reasonable approximation exists to avoid the necessity for a non-motile strain. However, we only arrive at this conclusion by considering the non-motile strain, so its inclusion here is necessary.

In the speed-only model, we set the observation function $b_0(y_t)$ equal to the estimate of the pdf of framewise speeds in the non-motile mutant, and $b_1(y_t)$ equal to the same quantity in the non-chemotactic mutant. The estimation is carried out using a kernel density estimate (KDE), as discussed in Appendix A.4. An illustration of the form of the observation function is shown in Figure 3.3. By incorporating information from the mutant strains, we are in effect creating an empirical prior of the underlying motion model.

The only free parameters remaining in the model, Ψ , are the probabilities for state transitions, A_{ij} . It is assumed that switching between states occurs with a characteristic time τ_{ij} , where i and j are specified in the same manner as for A . The transition probabilities

for transitions between states, independent of observations, are therefore given by

$$A_{00} = 1 - \Delta t / \tau_{01},$$

$$A_{01} = \Delta t / \tau_{01},$$

$$A_{10} = \Delta t / \tau_{10},$$

$$A_{11} = 1 - \Delta t / \tau_{10},$$

where Δt denotes the duration of a single frame and $\Delta t \ll \tau_{ij}$ by design. We maximise \mathcal{L} over τ_{01} and τ_{10} .

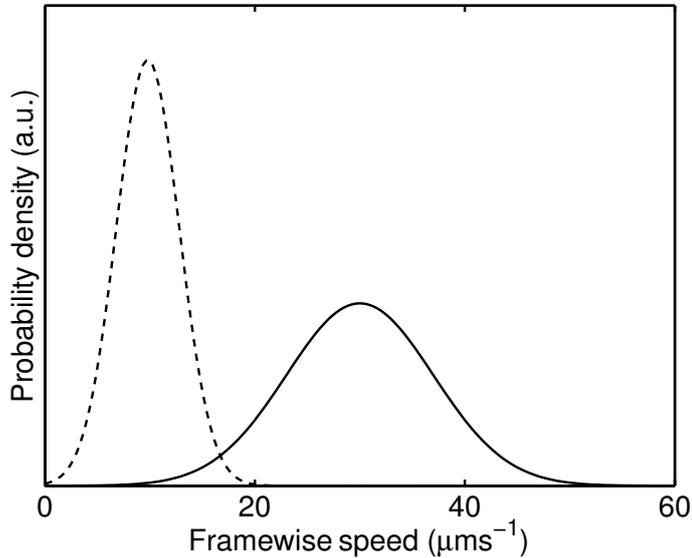


Figure 3.3: An illustration of the pdfs of framewise speeds for the running (solid line) and stopping (dashed line) phases. The observation functions b_1 and b_0 , respectively, are set equal to these in the speed-only HMM method.

The HMM formulation in equations (3.6) and (3.10) requires the assumption, discussed in Section 3.1.2, that observations do not affect the probability of transitions between states. In reality the probability of a cell being in a stopped phase beyond an experimentally determined maximum framewise speed is negligible. In the standard HMM form, this is implied by a very small value for the likelihood of the data given a stop relative to the probability given a run, $b_0(y_t) \ll b_1(y_t)$. In practice, however, both of these values become very small for extreme values of the framewise speed. In this case, numerical

rounding errors may cause the algorithm to fail by producing a result that does not predict correctly that the probability of a stop is negligibly small. Figure 3.3 illustrates this problem: the framewise run speed distribution and framewise stop speed distribution would have numerically indistinguishable values for a speed of $60 \mu\text{ms}^{-1}$. To avoid such numerical errors with high-speed outliers we use modified transition probabilities with a dependence on the observed speed, following Beausang and Nelson [17], given by

$$\begin{aligned}
A'_{00} &= 1 - \Delta t / \tau_{01}, \\
A'_{01} &= \Delta t / \tau_{01}, \\
A'_{10} &= (\Delta t / \tau_{10}) H, \\
A'_{11} &= 1 - (\Delta t / \tau_{10}) H,
\end{aligned}
\tag{3.13}$$

where $H \equiv H(\|\mathbf{d}_t\|/\Delta t - \rho)$ is a Heaviside step function and ρ is a maximum speed parameter, determined from the data. We let ρ equal the framewise speed below which 99.99% of the density of the non-chemotactic KDE lies. This prevents numerical errors, whilst simultaneously ensuring that we do not truncate the observation function $b_1(y_t)$ overly harshly at high framewise speeds. In the illustration in Figure 3.3, for example, $\rho \approx 50 \mu\text{ms}^{-1}$. Note that this modification is included purely to avoid numerical difficulties. The value of ρ has no significant effect on the outcome of the analysis providing it is set sufficiently high.

3.1.4 Full model

This approach uses both the framewise speed and the framewise angle between trajectories as the observable data. The observables are now vector quantities, $\mathbf{y}_t = (\|\mathbf{d}_t\|/\Delta t, \theta_t)$, equivalent to a plane polar representation of the tracks. By incorporating the extra angular information into our analysis method we hope to improve the sensitivity and specificity.

In order to incorporate angular information into the HMM some modifications are required. The first was introduced in Section 3.1.3; the same modified transition probabilities, A'_{ij} , are used here. Furthermore we seek an alternative formulation to that given in equation (3.6) which avoids the assumption of independence of each observation and the preceding state. The need for this modification is best explained by considering the observation function. A running cell exhibits angular directional persistence, thus the pdf of θ_t will be tightly centred on the origin in the case of a run-to-run transition. Conversely, the angular pdf in the case of a stop-to-run transition is uniform, as discussed below. The angular component of the observation function thus depends not only on the current state but also on the preceding state. A modified observation function is required to overcome this dependency. We denote this by $b_{ij}(\mathbf{y}_t)$, where $i \in \{0, 1\}$ refers to S_{t-1} , and $j \in \{0, 1\}$ refers to S_t .

The functional form of the observation function $b_{ij}(\mathbf{y}_t)$ is based on an empirical fit to experimental data with non-motile and non-chemotactic bacterial strains, similar to the case with the speed-only model. As for the speed-only model, a suitable approximation may make it possible to form a reasonable observation function in the absence of a non-motile strain. Again, this is discussed further in Section 3.4. The functional form differs from the that used in the case of the speed-only model, as we include the observed angle change in addition to the framewise speed. We assume independence of speed and angular distributions so that $b_{ij}(\mathbf{y}_t)$ is separable in these two components,

$$b_{ij}(\mathbf{y}_t) = f_j(\|\mathbf{d}_t\|/\Delta t)g_{ij}(\theta_t), \quad (3.14)$$

where both f_j and g_{ij} are normalised pdfs. This simplifying assumption is necessary as we do not have sufficient data to extract such a joint distribution at an acceptable resolution, as discussed previously. The speed component for each observation function, $f_j(\|\mathbf{d}_t\|/\Delta t)$, is independent of the previous state and is identical to that used in the

speed-only case (see Figure 3.3 for an illustration). The angular component, g_{ij} , is specified by comparison with the observed framewise angle changes in the two mutant strains. Since the observed angle changes are distributed approximately symmetrically in the non-chemotactic dataset, we take the absolute value of the observed angle changes¹. We fit the absolute observed angular data from the non-chemotactic dataset with an exponential function, as this provides a good fit, and because strongly peaked data are not suitable for use with the KDE. Note that, in the absence of noise, the angle change on going from a run to a run would always be identically zero. In reality, the distribution is concentrated around zero, but shows some departure due to various sources of noise, as discussed further in Chapter 6. We do not fit the observed angular data from the non-motile dataset, as this is known to follow a wrapped uniform distribution (see Chapter 5).

The dependence of the angular component on the previous state, in addition to the current state, introduces additional complexity. The first running frame after a stop has the same speed distribution as all subsequent running frames but a different angular distribution; the angle selected after a stop is assumed to be approximately uniformly distributed. This assumption is required in order to maintain the Markov property of the process; any directional persistence over the course of a stop would lead to a non-Markovian process. As is shown in Section 4.2.7, this assumption may not accurately reflect the true underlying process. It is, however, considered to be acceptable since the uniform distribution reflects the prior uncertainty as to the possibility of directional persistence. The functional form of the angular component is illustrated graphically in Figure 3.4(a). The complete observation function is illustrated in Figure 3.4 for the four possible combinations of i and j .

¹As we shall see in the next chapter, there is in fact a statistically significant departure from the symmetric distribution in the case of non-chemotactic bacteria tracked near a surface. However, this has little effect on the form of the observation function, so the assumption of symmetry is justified.

Implementing the modified observation function, equation (3.7) becomes

$$\alpha_t(i) = \sum_{j=0}^1 \alpha_{t-1}(j) A'_{ji} b_{ji}(\mathbf{y}_t), \quad (3.15)$$

with initial condition $\alpha_0(i) = b_{0i}(\mathbf{y}_t)$. Note that in the initial condition it is assumed that cells were stopped immediately before the track begins. This is a simple way to implement the uncertainty in the angle at the start of the track, since we do not know the direction of travel immediately prior to the start of the track. This assumption assures that the angular component is initially given by a wrapped uniform distribution. An alternative approach to deal with this uncertainty would be to take the initial state probabilities and direction of travel from the first displacement vector in the track, \mathbf{d}_0 . This would effectively shorten the track by one sample point at the start.

The modified form of equation (3.11) for β_t is

$$\beta_t(i) = \sum_{j=0}^1 \beta_{t+1}(j) A'_{ij} b_{ij}(\mathbf{y}_{t+1}), \quad (3.16)$$

with $\beta_T(i) = 1$, as for the speed-only method.

3.1.5 Determining the maximum likelihood HMM transition parameters

As discussed in Section 3.1.2, the two components describing the hidden Markov model are the observation function b_{ij} and the transition probabilities in A'_{ij} . The first component is independently determined, as discussed above. The transition probabilities A'_{ij} are specified by the three parameters in equation (3.13), namely τ_{01} , τ_{10} and ρ . The threshold parameter ρ is fixed, as discussed in Section 3.1.3. It is possible to obtain a maximum likelihood estimation of the final two parameters by maximising the likelihood

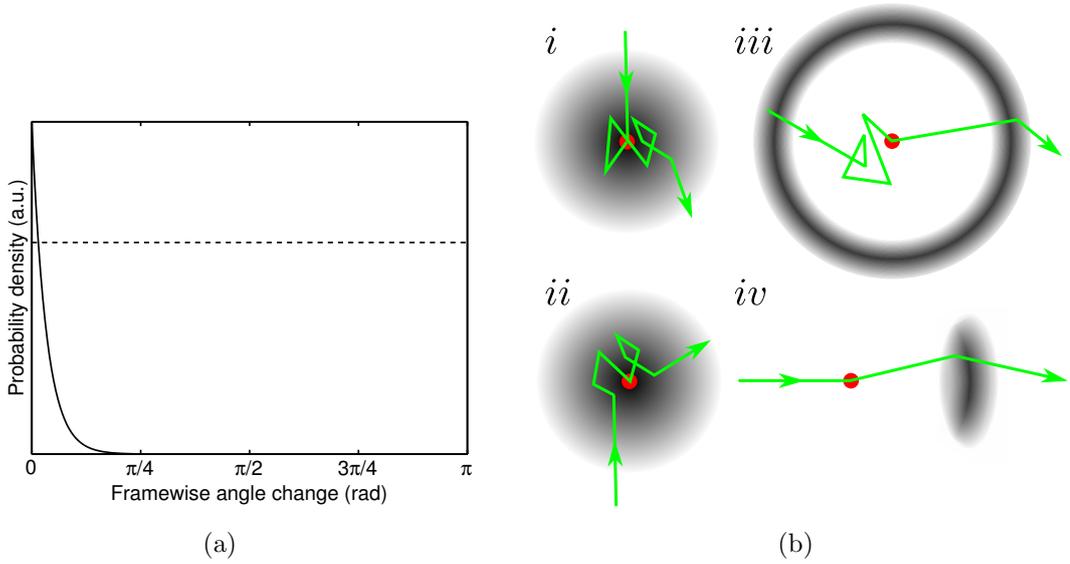


Figure 3.4: (a) Illustration of the pdfs of framewise angle changes for the running (solid line) and stopped (dashed line) phases. The angular components of the observation function are set equal to the running distribution (g_{11}) and the stopped distribution (g_{00} , g_{10} , and g_{01}). (b) Illustration of the functional form of the observation function, $b_{ij}(\mathbf{y}_t)$, for the full HMM method. The shaded area represents the pdf for observing the moving particle at a particular point in space at the next sampling time, given that the particle is at the point marked by the red dot at the current sampling time. Darker shades represent higher probabilities. The green line shows an example trajectory in each case. (i) Run-to-stop transition, b_{10} . (ii) Stop-to-stop transition, b_{00} . (iii) Stop-to-run transition, b_{01} . (iv) Run-to-run transition, b_{11} .

defined in equation (3.8).

The sampling time of the system, Δt , imposes a restriction on our ability to determine the parameters τ_{01} and τ_{10} as we require that $\Delta t \ll \tau_{ij}$. To take this limitation into account, we define two new variables, $p_{01} = \Delta t/\tau_{01}$ and $p_{10} = \Delta t/\tau_{10}$, which are interpreted as the probability of a transition from a stop to a run (p_{01}) and from a run to a stop (p_{10}) between two consecutive frame intervals. Since these new parameters are probabilities, they are restricted to take values in the range $[0, 1]$. We now focus on inferring p_{01} and p_{10} , bearing in mind that we can use these values to estimate the dwell times, τ_{ij} , providing that the above limitation on the sampling time is respected.

Das *et al.* [56] use a Markov chain Monte Carlo (MCMC) scheme to find the maximum likelihood estimate of their rate parameters in a similar application to that described here. In our case, the negative log-likelihood surface is always smooth and contains a single global minimum (see Figure 4.14 for an example), so that a deterministic optimisation routine is more computationally efficient. We use the Matlab function `fmincon` to carry out a constrained optimisation of the negative log-likelihood. The function to be minimised is defined by

$$h = - \sum_{i=1}^N \log \mathcal{L}^{(i)}, \quad (3.17)$$

where $\mathcal{L}^{(i)}$ denotes the likelihood of the data from the i^{th} track, as defined in equation (3.8), and N is the total number of tracks in the dataset. The logarithm operation is required to avoid numerical errors, as the likelihood can become extremely small. As the likelihood is a function of p_{01} and p_{10} , the minimisation is carried out over a two-dimensional vector space. The summation in equation (3.17) pools the results from all of the tracks in the dataset, so that the final likelihood is averaged over all tracks in the dataset. It is possible, in principle, to maximise the likelihood over each individual track, however the performance of this approach is poor when dealing with short tracks

(data not shown). The optimised parameters are subsequently used to compute the run probabilities using equations (3.7), (3.11) and (3.12).

3.1.6 Post-processing the output of the analysis methods

Each of the three methods presented returns a vector for each track processed containing the probability of a run between each observation point, $(P_1(1), \dots, P_{T-1}(1))$. In the case of the heuristic method, every value is either equal to 1 or 0, whereas the HMM methods return values in $[0, 1]$. In the latter case, if we wish to assign every between-frame motion to either a running or stopped state, we round all values to the nearest integer (0 or 1). The resulting vector can be considered to represent the run *status* (as opposed to run probability). This transformation is always carried out on the run probabilities computed using the HMM-based methods. In the case of the heuristic method, there is no distinction between the two properties.

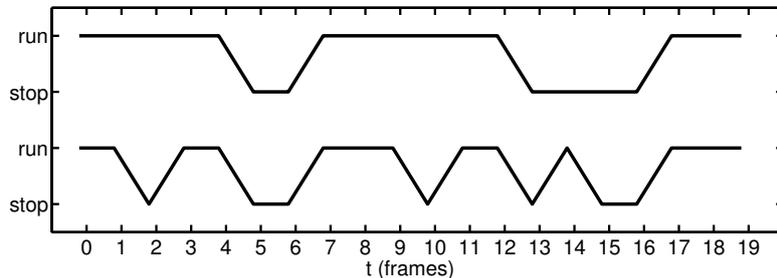


Figure 3.5: An illustration of the post-processing method applied to a run status vector. The bottom line represents the output from one of the heuristic analysis methods. The top line shows the result of applying post-processing techniques with $\tau_{0,min} = \tau_{1,min} = 2$. Running and stopped phases which only last one frame are filtered out by this process.

An additional heuristic step may be applied to the run status vector of each track, which in effect smooths the inferred path between the running and stopped phases. We define a run persistence parameter, $\tau_{1,min}$, and a stop persistence parameter, $\tau_{0,min}$, which correspond to the minimum duration, in frames, of a running and stopped phase, respectively. Running and stopped phases which have a duration shorter than $\tau_{1,min}$ and $\tau_{0,min}$, respectively, are relabelled. This proceeds in several stages, detailed in Algorithm

2. The process is also illustrated graphically in Figure 3.5. The decision to relabel short runs before short stops is arbitrary, and introduces a bias towards stops in the case of a series of rapid oscillations between the two states (the short run sections will first be converted to stops, resulting in a larger stopped section). Intuitively, a high density of between-frame stops detected within a region suggests that at some point in the region there was at least one stop. We choose to proceed in this fashion as we place greater importance on identifying every stop, possibly at the expense of including some false positives or inferring overly long stopping phases.

Algorithm 2 Post-processing of analysis method output

```

if Method is based on HMM then
    Convert run probability vector to a run status vector.
end if
for All run phases do
    Find the duration of the run phase.
    if Duration of run <  $\tau_{1,\min}$  then
        Convert all runs in this phase to stops.
    end if
end for
for All stopped phases do
    Find the duration of the stopped phase.
    if Duration of stop <  $\tau_{0,\min}$  then
        Convert all stops in this phase to runs.
    end if
end for

```

The post-processing method presented here was originally developed for use with the heuristic method, which often exhibits unrealistically rapid fluctuations between running and stopped states, as there is substantial overlap between the distribution of the magnitude of between-frame displacements for runs and stops. Smoothing the state path helps to remedy this situation. Nonetheless, as we shall see in Section 3.3, the application of this post-processing method following the HMM-based analysis can substantially improve the output when the noise level is high.

3.2 Simulation study comparison of the analysis methods

Before we can apply the analysis methods discussed above to experimental data, we must evaluate and compare their ability to correctly infer stop phases in tracks affected by varying levels of noise. A traditional means of evaluating this performance is to compare with the results of manual ('by hand') assignment of stopped phases in real tracks. This approach suffers from several key drawbacks, however. First, manual tracking is a time-consuming and often difficult process. As we shall see when the experimentally derived tracks are presented in more detail in Chapter 4, the stop phases in microscope videos are difficult to discern unambiguously by eye. A second limitation of manual analysis is the dependence on the necessarily subjective interpretation of the person processing the data.

An alternative approach to manual analysis, and the approach used here, is a simulation study. This is a common approach to assess the performance of automated analysis methods, and is exemplified in [17, 56, 122]. We assume that experimentally-obtained tracks are the result of a run-and-stop VJ process as described in Chapter 2. We define the sampling interval to be $\Delta t = 0.02$ s (= 1 frame) to match the frame capture rate of the microscope used to obtain experimental movies, and simulate each track for 5 s (250 frames). We use a simplified model of the noise in the system by adding a normally distributed perturbation to each coordinate of the recorded position in a track, with zero mean and standard deviation equal to $(2D_{\text{WN}}\Delta t)^{1/2}$, where D_{WN} is varied to modulate the level of noise applied to the system. This description is equivalent to uncorrelated Gaussian white noise; we note that the use of such a model to simulate the type of noise exhibited in real experimental systems may be oversimplified; this is discussed in more detail in Chapter 6. This simple implementation of noise is maintained for the present study, however, in order to avoid obfuscating the problem. The true underlying state sequence in the simulations, which is continuous in time, is recorded for later comparison

with the state inferred by the analysis methods.

We use two different simulations to demonstrate the effect of different underlying motion models on the performance of the analysis methods. In both simulations, the speed of each running phase is an independent and identically distributed (IID) Weibull distributed random variable with scale parameter 40.1 and shape parameter 4.1. This distribution and these parameters are chosen as they provide a good match with the distribution of framewise speeds seen in the experimental data (see Chapter 4). It is not clear *a priori* to what extent the assumption that running phases occur with a constant speed between each reorientation event is valid. For example, Packer *et al.* [151] observed large variations in flagellar motor rotation speeds between individuals in their study. Nonetheless, this choice of model is a reasonable approximation to the true underlying motion for the purposes of our simulation study. We discuss this further in Section 3.4. The first simulation is of a run-only VJ process with $\lambda = 2 \text{ s}^{-1}$, and the second is of a run-and-stop process with $\lambda = 2 \text{ s}^{-1}$ and $\mu = 10 \text{ s}^{-1}$. These parameters are two to fourfold higher than experimentally observed parameters [28], however all values cited in the literature for *R. sphaeroides* are measured using the tethered cell assay, which may well differ substantially from the values in free-swimming bacteria [157]. We discuss this in greater detail in Section 4.3. The values used here approximate an observed track reasonably well, and assure that the simulated tracks exhibit a sufficient number of stopping phases on average. Again, we return to this discussion in Section 3.4.

The simulation method used here was described in full detail in Chapter 2. The inclusion of a run-only simulation is intended as a control, to test whether the analysis methods considered infer a stopped phase when none are present in the tracks. The level of noise added was varied for each simulation to test the effect on the performance of the analysis methods. This was achieved by varying D_{WN} from $7.2 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ ($10^{-2} \text{ px}^2\text{s}^{-1}$, a very low level of noise) to $7.2 \mu\text{m}^2\text{s}^{-1}$ ($10 \text{ px}^2\text{s}^{-1}$, a very high level of noise). Tracks from

the run-and-stop simulated dataset with a number of different noise levels are shown in Figure 3.6.

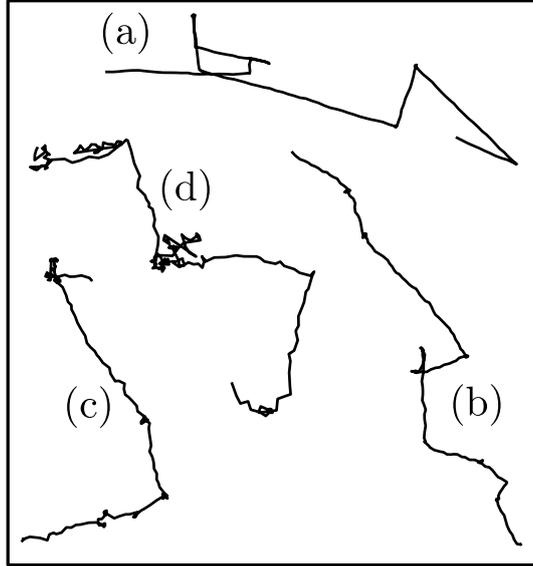


Figure 3.6: A representative selection of simulated tracks with different levels of added noise. (a) $D_{WN} = 7.2 \times 10^{-2} \mu\text{m}^2\text{s}^{-1}$, (b) $D_{WN} = 7.2 \times 10^{-1} \mu\text{m}^2\text{s}^{-1}$, (c) $D_{WN} = 1.4 \mu\text{m}^2\text{s}^{-1}$, (d) $D_{WN} = 2.8 \mu\text{m}^2\text{s}^{-1}$.

In carrying out the steps required to analyse the simulated datasets and compare their performance, we mimic the process that we use when analysing real data in Chapter 4. None of the parameters of the true underlying processes are assumed known to the analysis methods. In order to compute the empirical prior component of the HMM, we require a non-motile dataset and a non-chemotactic dataset. Since these are available in the real experimental dataset, we simulate them here. The non-motile dataset is simulated by setting the transition rate λ very high, and μ very low. In practice, values of 10^{12}s^{-1} and 10^{-12}s^{-1} , respectively, suffice to generate a track that is effectively stopped throughout its entire duration. The non-chemotactic dataset is simulated by setting λ very low, so that tracks exhibit no reorientations at all.

3.2.1 Discretising the underlying state sequence

The output of the analysis methods for each track is a vector of run status for the periods between discrete sampling time points. The true underlying motion is, by contrast, continuous in time. In order to facilitate a comparison between the output and the true state sequence, we discretise the latter over the same time domain. Any period between two consecutive sampling points that contains a stop phase is designated a stop in the discretised true underlying run status.

3.2.2 Assessing the rate of false positives and false negatives

Two common measures of statistical error in an inference process are the rates of false positives and false negatives, also known as type I and II errors, respectively. Here we consider the inferred state sequence as a series of stopping phases and running phases, where each phase consists of one or more consecutive framewise transitions identified as a stop or a run. A false positive corresponds to an inferred stopping phase where none is present in the true underlying state sequence. A false negative corresponds to an inferred running phase where none is present in the true underlying state sequence. Figure 3.7 illustrates the problem of determining the accuracy of the analysis methods in terms of the rates of false positives and negatives. The upper coloured bar indicates the true (discretised) state sequence and the bottom bar the inferred sequence. There are several discrepancies. A single frame stop has been inferred during the fourth running phase. This is a false positive; there is another false positive at the end of the trace. Conversely, during the fifth stopping phase the analysis method has inferred a single frame run. This is a false negative. All of the mistakes noted here would be fixed by the application of the post-processing method with τ_{01} and τ_{10} both greater than one (data not shown).

We assess the rate of false positives and false negatives as the ratio of the number of false positives and false negatives to the number of actual stop events in each track, respec-

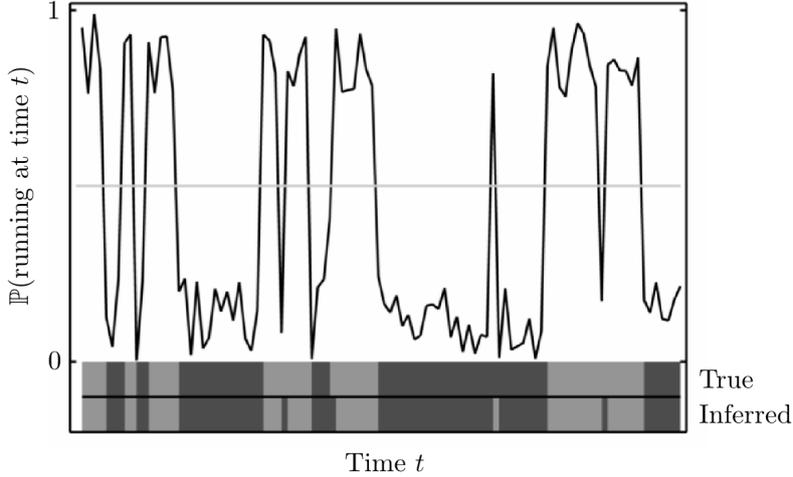


Figure 3.7: An illustration of the problem of assessing the accuracy of analysis methods in determining stopping phases. The trace represents the probability of running for each time interval, as inferred by the HMM method, i.e. $P_t(1)$. The grey line shows the value $P_t(1) = 0.5$, above which we identify the interval as a run. The top shaded bar shows the true underlying state sequence, where dark grey indicates a stop and light grey a run, and the bottom bar shows the states inferred by the analysis method.

tively, averaged over all tracks in the dataset. All simulated tracks are verified to contain at least one true stopping phase to avoid numerical singularities. Note that a value above one in either of these instances means that the number of inference errors is greater on average than the number of underlying stop events in each track.

3.3 Results

The previous section described in detail the methods we use to simulate tracks and subsequently compare the performance of the analysis methods. The general procedure is to simulate data, infer the necessary parameters for each analysis method, identify stops in the tracks using the analysis methods, then compare these to the true underlying trajectory. We carry out this process for a range of levels of added noise. In the case of the heuristic method, we must estimate the threshold speed parameter, ρ_{CS} , whereas for the HMM methods we must estimate the transition probabilities p_{01} and p_{10} . We now present the results of our simulation study, in terms of both the inferred parameters and

state sequences.

3.3.1 Maximum likelihood HMM transition parameters

Figure 3.8 shows the negative log-likelihood surface for the run-only and run-and-stop simulated datasets with $D_{\text{WN}} = 0.432 \mu\text{m}^2\text{s}^{-1}$, computed by sampling the likelihood over a range of values of the parameters p_{01} and p_{10} . In both cases, the surface is similar for all noise levels considered here (data not shown). Whilst the plots appear similar at first glance, the run-and-stop surface has a clear global minimum at $p_{01} \approx 0.1$, $p_{10} \approx 0.03$, whereas no such well-defined minimum exists for the run-only dataset. The run-only likelihood varies very little with the parameter p_{01} , compared with the run-and-stop dataset. This is as expected because the run-only tracks contain no stops, therefore the rate governing the stop to run transition has little effect. As a result of the flatness of the run-only surface, the optimisation routine often fails to converge to a global minimum (data not shown). We therefore no longer consider this dataset, since the absence of maximum likelihood transition rates impedes any further analysis. This result demonstrates the important point that the HMM is fundamentally defined by the underlying model, Ψ . If this model is incorrect, as is the case when we apply the HMM designed for run-and-stop motion to a run-only dataset, the inferences drawn may be spurious. We have determined that we should check the appearance of the negative log-likelihood surface before proceeding.

Figure 3.9 shows the maximum likelihood transition rates, inferred using the two HMM-based methods, for varying levels of noise. These are computed by multiplying the inferred values for p_{01} and p_{10} by a factor of fifty, to convert from units of frames to seconds. Recall that the true values are $\lambda = 2$, $\mu = 10$. Both methods estimate these values approximately correctly for low added noise, although there is still some discrepancy. In the case of μ , both methods increasingly underestimate the value as the level of noise increases, whereas

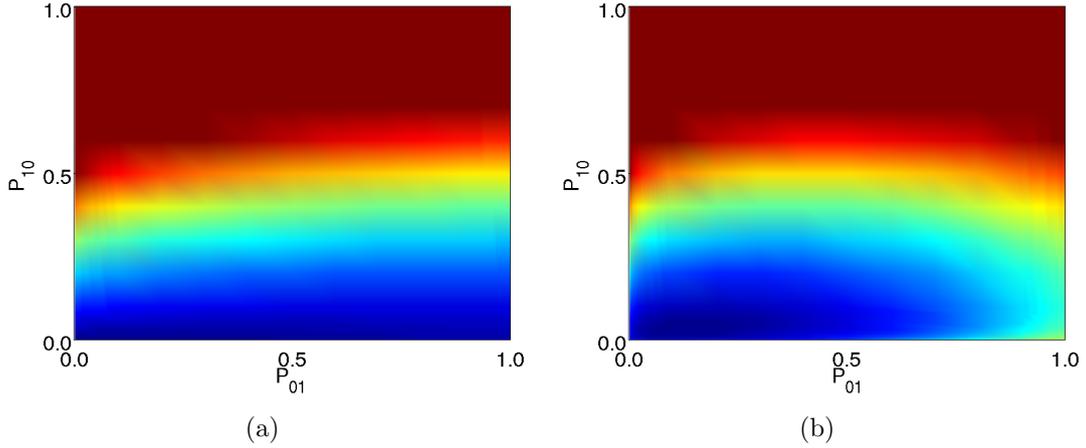


Figure 3.8: Negative log-likelihood surfaces for the run-only (a) and run-and-stop (b) datasets, computed using noise level $D_{WN} = 0.432 \mu\text{m}^2\text{s}^{-1}$. The false colour runs from low (blue) to high (red) positive values. Both plots are normalised in the same way.

for λ , the full HMM method generates overestimates as noise increases and the speed-only method produces underestimates. This result indicates that we should be cautious when interpreting the meaning of the maximum likelihood transition rates.

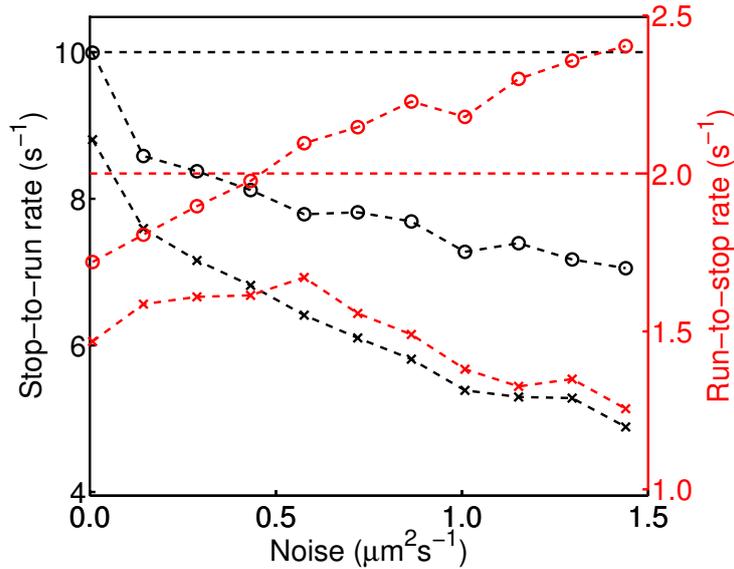


Figure 3.9: Inferred rates for stop to run transitions (μ , black) and run to stop transitions (λ , red), computed by minimising the negative log-likelihood. (\circ) HMM full; ($+$) HMM speed-only.

3.3.2 Heuristic cutoff parameter

The basis of the heuristic method is that there is a substantial difference between the distribution of speeds observed during framewise runs and stops. The value of ρ_{CS} should

be selected to maximise the number of correctly inferred transitions. We approach this problem by computing the observed framewise speeds for the non-chemotactic and non-motile strains. We estimate the true pdf of framewise speeds in each case using a KDE [32]. The parameter ρ_{CS} is then taken to be equal to the point at which the two estimated pdfs intersect. This does not in general guarantee that the measure of the crossover region is minimised, however for our purposes it is a reasonable approximation. Figure 3.10 shows the values of ρ_{CS} determined by this method for different levels of added noise. As expected, the cutoff level increases with noise level, since this causes the two distributions to broaden. This is demonstrated further by Figure 3.11, which shows the KDEs for two different noise levels. This figure also shows how the overlap region grows with noise level.

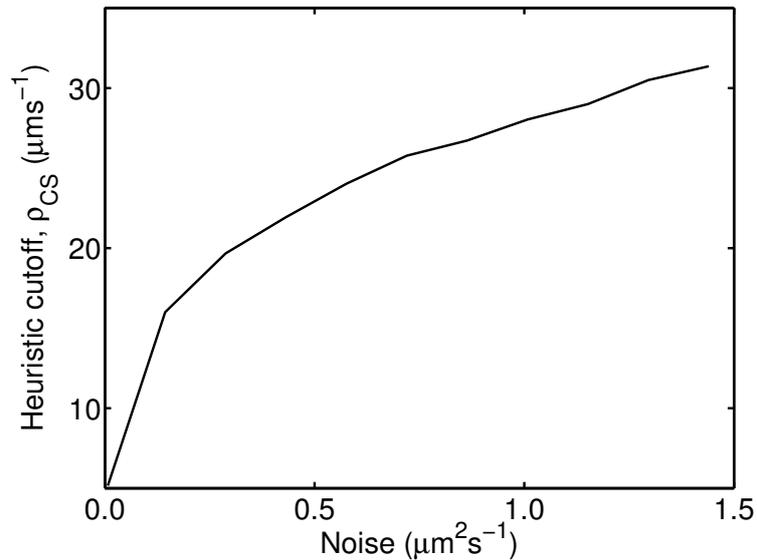


Figure 3.10: *The variation of heuristic cutoff speed, ρ_{CS} , with noise level, D_{WN} , for the simulated run-and-stop process.*

3.3.3 Assessing and comparing the performance of the analysis methods

Figure 3.12(a) shows the variation of false positive rate with noise strength. The heuristic method is highly sensitive to noise, producing a very large proportion of false positives

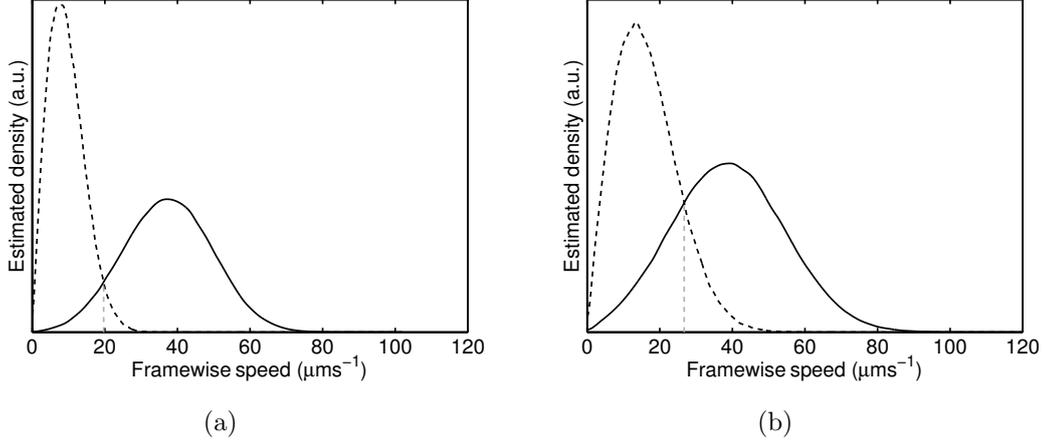


Figure 3.11: KDE of the observed framewise speeds in the non-chemotactic (black solid line) and non-motile (black dashed line) datasets, for the simulated run-and-stop process with (a) $D_{WN} = 0.288 \mu\text{m}^2\text{s}^{-1}$, and (b) $D_{WN} = 0.864 \mu\text{m}^2\text{s}^{-1}$. Grey dashed line indicates the automatically identified crossover region in both cases.

even for low levels of added noise, with rates rising to greater than one at intermediate levels of noise (data not shown), meaning that for every true stopping event a spurious event is inferred. This occurs because the heuristic method often produces tracks with rapid oscillations in the state sequence, as these are not penalised. This is in contrast to the HMM-based methods, in which the low values of the transition probabilities p_{01} and p_{10} penalise high-frequency switching. The heuristic false positive rate is reduced somewhat by application of the post-processing stage, however this method still performs significantly worse than the methods based on the HMM at most noise strengths. The full HMM method is also more sensitive to noise than the speed-only method, producing a higher number of false positives as D_{WN} is increased. This is rectified by the application of the post-processing stage.

The false negative rate varies less between analysis methods, as Figure 3.12(b) shows. In this case, the heuristic method is the most effective, rarely failing to detect a stop. The speed-only method, in contrast to the results of the false positive rate, has the highest number of false positives, and shows an approximately linear increase with noise.

The false positives generated by the different methods may bias the inferred distribution

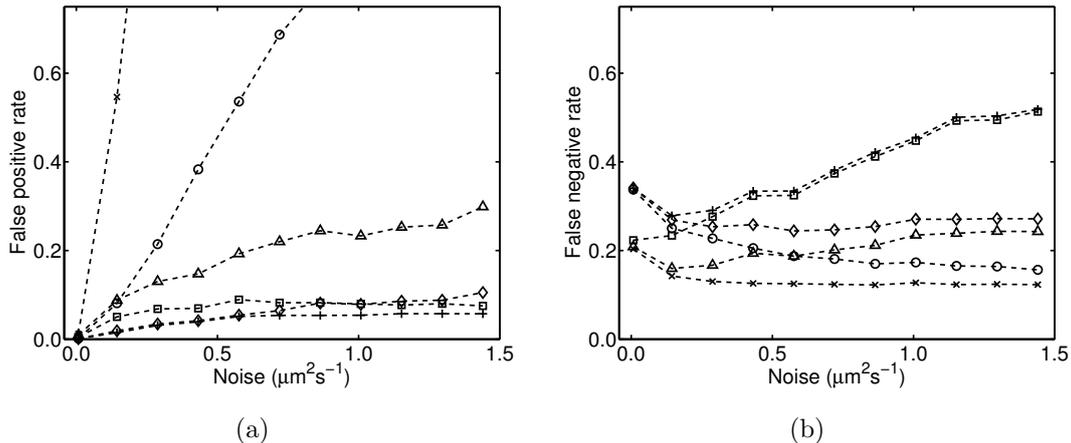


Figure 3.12: Variation with noise level of (a) the false positive rate, and (b) the false negative rate. Data are computed for the simulated run-and-stop process, with the following analysis methods: heuristic (\times); heuristic with post-processing (\circ); HMM speed-only (\square); HMM speed-only with post-processing ($+$); HMM full (\triangle); HMM full with post-processing (\diamond).

of stopwise angles. We quantify this bias using simulated data in Figure 3.13. Black bars indicate the inferred distribution of stopwise angle changes following the application of the relevant analysis method. The angle change distribution in the VJ process used to simulate our datasets is uniform, hence the inferred distribution should also be uniform. The heuristic method (Figure 3.13(a)) generates a distribution that is significantly skewed towards lower turning angles. This departure from the predicted uniform distribution is due to the presence of false positives, whose contribution is indicated by the yellow bars. The false positives bias the inferred distribution substantially. Application of the post-processing stage reduces the bias slightly, but it is still very marked (data not shown). The bias is far smaller in the case of the full HMM method, though the smaller number of false positives still bias the distribution towards low turning angles. The distribution generated by the speed-only HMM method, or either HMM-based method with post-processing, is not significantly different from the result for the full HMM method without post-processing (data not shown).

A second way in which the analysis methods may introduce bias is in the distribution of stop durations. Figure 3.14 shows how the ratio of false positives to total number of stopping phases (given by the sum of true stops and false positives) varies with the

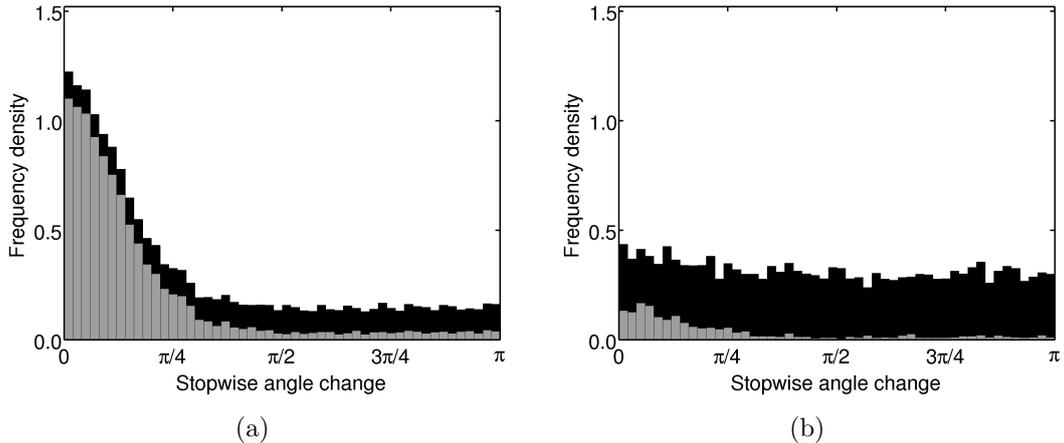


Figure 3.13: Histogram of the distribution of stopwise angle changes, as inferred by (a) the heuristic method, and (b) the full HMM method, for a run-and-stop process with $D_{WN} = 0.432 \mu\text{m}^2\text{s}^{-1}$. Black bars show all inferred stops, while grey bars indicate what proportion of these are due to false positives.

duration of the stop, for an intermediate level of added noise. As for the stopwise angle changes, the bias is most significant for the heuristic analysis method, although this is somewhat rectified by the application of the post-processing stage. The two HMM-based methods perform better than the heuristic method in all duration ranges, with the exception of the longest range considered, where the speed-only method and the heuristic method with post-processing perform similarly. We note that Figure 3.14 shows the fraction of false positives, not the absolute number; the expected number of stops decreases exponentially with stop duration.

3.4 Discussion and conclusions

In this chapter, we have investigated the problem of analysing tracks from bacteria undergoing a run-and-stop VJ process. We proposed three methods to infer sections in tracks that correspond to stopping events. The analysis methods were selected with a view to analysing tracking data from *R. sphaeroides*, as the data presented in Chapter 4 is from observations of this species. They are, however, generally applicable to tracks from any bacterium that undergoes approximately discrete reorientation events, of sufficient duration that it is possible to capture them with a video microscope. The heuristic method

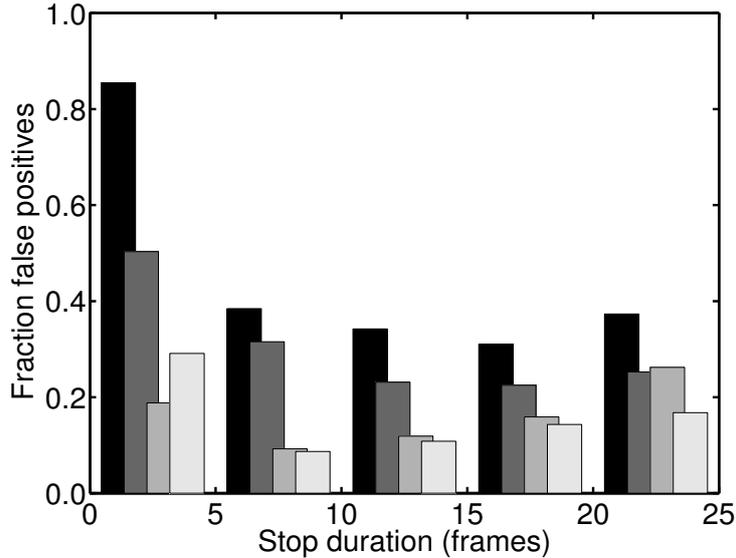


Figure 3.14: The fraction of false positives for different durations of stops, computed for a run-and-stop process with $D_{WN} = 0.432 \mu\text{m}^2\text{s}^{-1}$. A fraction of 1 means that all inferred stops with a duration in the range of interest are false positives. The four shades, darkest through to lightest, denote the analysis method used to infer stops. In order, these are: heuristic, heuristic with post-processing, speed-only HMM and full HMM.

is based on work by Taboada *et al.* [190]. Related analysis methods have been used in other studies [24, 62], however several separate heuristics are applied in these studies and it is unclear how the various parameters have been selected. The remaining two methods are variations of the HMM.

We proceeded to use simulated data to compare the performances of the three analysis methods for analysing tracks of freely swimming bacteria. This is, to the best of our knowledge, the first example of such a systematic study to compare analysis methods for tracking data. The simulation is based on the run-and-stop VJ process with simple additive white noise applied at varying strengths. Running speeds are drawn from the Weibull distribution, with parameters specified by a fit to the experimental data presented in Chapter 4. Following the model described by Campos *et al.* [43], we assume that the running speed remains constant for the duration of each running phase and that a new speed is selected at random upon each reorientation. We now discuss the validity of the way in which noise and reorientations are modelled.

The noise model used here may not capture all sources of noise in the system, however it approximates measurement error reasonably. One could also consider incorporating more sophisticated noise models, for example rotational Brownian motion, as considered in Chapter 6. However, for the purposes of a simple yet robust performance test the noise model considered here is adequate. We assume throughout that the VJ model is appropriate when applied to bacterial motion, following good evidence in the literature on the subject [24, 124].

The assumption that running speeds are maintained for the duration of a run has not been tested experimentally using tracking data. However, Packer *et al.* [150] observe that tethered cells exhibit a great deal of heterogeneity in their basal rotation rates, suggesting that the speed may vary over the course of a running phase. For the purposes of the simulation study, we opted for the present model, detailed by Campos *et al.* [43], as we have not characterised the variability of running speeds in single bacteria. Further work could involve extracting this information from bacterial tracking data. This is in principle possible by a straightforward analysis of single tracks, however we did not perform this analysis on the experimental data in Chapter 4 due to time constraints.

Our first test of the HMM-based methods was to apply them to inappropriate data from a run-only simulation, to test whether the analysis would produce nonsensical results. Numerical optimisation of the transition probability parameters was erratic, and the negative log-likelihood surface showed that the transition parameter p_{01} had little effect on the likelihood, as expected.

Further results presented in this chapter show that even low levels of noise cause the simple heuristic algorithm to fail and produce a very high incidence rate of false positives. Coupling the heuristic method with post-processing improves the performance somewhat,

but still does not make it competitive with the new HMM-based methods presented here. In particular, the heuristic method introduces significant bias into the inferred sequence of stopping phases, skewing the stopwise angle change and stopwise duration distributions towards lower values. In contrast, the HMM-based methods show very little bias in the resultant distributions.

The two HMM-based methods differ only in their treatment of the observed data, with the full model incorporating both framewise speed and angle change, whereas the speed-only model ignores angular information. As a result, the speed-only HMM method is less sensitive than the full HMM method, generating fewer false positives and more false negatives. The application of post-processing to either of these methods makes little difference to the level of such errors, with the exception of the false positive rate in the full HMM method, which is decreased significantly at higher levels of noise with post-processing.

Note that the results presented here are for a single choice of the parameters μ , λ , Δt and the Weibull speed distribution scale and shape factors. The Weibull parameters were selected to closely match the experimental data presented in Chapter 4, while the sampling rate, Δt is chosen to match that of the microscope camera used in Chapter 4. The transition rates are greater than those observed experimentally with tethered cell assays (see, for example, [28]), but we note that results obtained with this very different experimental method may not agree with tracking experiments [157]. Different values for these parameters will result in a different distribution of stop durations in the simulated data, which in turn will affect the performance of the analysis methods. For example, a slower sampling time (increased Δt) is expected to increase the number of false negatives across all methods, as short stops become impossible to discern on the observation timescale. Fixing Δt and increasing the value of λ is expected to have a similar effect. Further work is required to determine the nature of the sensitivity of the analysis methods to variations

in these parameters.

An important advantage to the methods based on the HMM is the ability to obtain a maximum likelihood estimate for the parameters p_{01} and p_{10} for the dataset being analysed. The maximum likelihood estimate of the value of these parameters allows us to estimate λ and μ , as $\lambda = p_{10}/\Delta t$ and $\mu = p_{01}/\Delta t$. In the case of high noise strength, we saw in Section 3.3.1 that the estimates for these parameters can differ from the true underlying values by a factor of up to two. The estimates are, however, reasonable for a range of intermediate noise levels. There is no straightforward method to estimate the same parameters using the heuristic method, hence we did not attempt this. However, it may be possible to fit an exponential distribution to the observed stop and run duration data. It is likely that the presence of false positives and significant bias would render this approach highly inaccurate, however.

A critical component of the HMM framework is the observation function, discussed in Section 3.1.2. This represents our prior knowledge of the nature of running and stopped phases in the observed data. In the simulation study, we form the observation function by reference to simulated non-motile and non-chemotactic datasets, which represent the two mutant strains that form part of the experimental *R. sphaeroides* dataset considered in Chapter 4. Analogous strains have been reported for many other bacterial species, for example *E. coli* [14], *Campylobacter jejuni* [191], and *Caulobacter crescentus* [68]. As mentioned in Section 3.1.2, if the non-motile mutant were not available for any reason, we propose that the observed motion of this strain may be approximated by a diffusive process. Within such an approximation, it is possible to calculate a theoretical value for the diffusion coefficient using Stokes' law, as we show in Chapter 6. We demonstrate in Chapter 6 that the observed motion of the non-motile *R. sphaeroides* strain does indeed agree well with such an approximation. Further work is needed to ascertain whether this constitutes a reasonable prior for other microorganisms of interest, however.

In Chapter 4 we shall analyse real experimental data, hence the present study is critical if we are to maximise the utility of our data and avoid false drawing false conclusions. From the study presented in this chapter, we conclude that the HMM-based methods are both far more reliable analysis methods than the heuristic approach. The simulation results indicate that the two HMM analysis methods perform similarly well, but that the full variant is more sensitive, and therefore more likely to generate false positives. It is not clear which method will be best suited to analysing the experimental data, hence we shall use both methods and compare the results. Furthermore, we shall consider the estimated values of the transition rates with caution, in light of the strong dependence of the accuracy on the noise level present in the data.

We are thus guided by the results of the simulation study, although, as we shall see, there are several features of the real tracking data that differ significantly from the simulated datasets considered here. The results from this chapter are furthermore applied as a guideline in Chapter 6, where it is important to quantify the extent of bias due to false positives from the analysis methods.

Chapter 4

Analysis of experimental data

Chapter 3 motivated the use of tracking data from swimming bacteria to extract various quantities necessary to characterise the motion of a bacterial swimmer. Analysis methods were presented in Chapter 3 that enable us to identify stopping phases in tracks and estimate the switching rates between stops and runs. These methods were tested and assessed using simulated tracks. In this chapter, we apply the methods of Chapter 3 to real experimental data. We use analysis methods based on the HMM throughout, as this was shown to perform substantially better than the proposed alternative heuristic method. As in previous chapters, we consider the model bacterium *R. sphaeroides*. The experimental setup used to obtain the data is described. We then discuss some additional considerations, required in the analysis stage, that were unnecessary when working with simulated data. Finally, we present and discuss relevant results obtained from our experimental datasets.

4.1 Methods

We now present the details of the experimental and analytical methods used to estimate quantities of interest relating to the swimming behaviour of *R. sphaeroides*. The process

Table 4.1: *Glossary of parameters used throughout this chapter.*

Parameter	Description	Value
ρ_P	Minimum cluster size to constitute a valid target	5 px
ρ_{FS}	Maximum framewise speed	$90 \mu\text{ms}^{-1}$
ρ_{EMS}	Minimum effective mean speed for non-chemotactic tracks	$15 \mu\text{ms}^{-1}$
ρ_E	Minimum extent for wildtype tracks	$6 \mu\text{m}$
ρ_{MC}	Maximum median curvature for wildtype tracks	$1.5 \text{ rad } \mu\text{m}^{-1}$
$\tau_{0,\text{min}}$	Post-processing minimum stop duration	2 frames
$\tau_{1,\text{min}}$	Post-processing minimum run duration	2 frames

of analysing tracks, which follows the imaging and cell tracking stages discussed below, is illustrated in the flow diagram shown in Figure 4.1. It differs from the process used in the simulation study in Chapter 3 (for comparison, see Figure 3.2) by the inclusion of a censoring process to filter certain tracks out of the experimental dataset, which we discuss in Section 4.1.6. We will also describe the process of selecting values for the parameters which control the censoring process. After censoring, the process of computing maximum likelihood transition parameters and state sequences is identical to that described in Chapter 3.

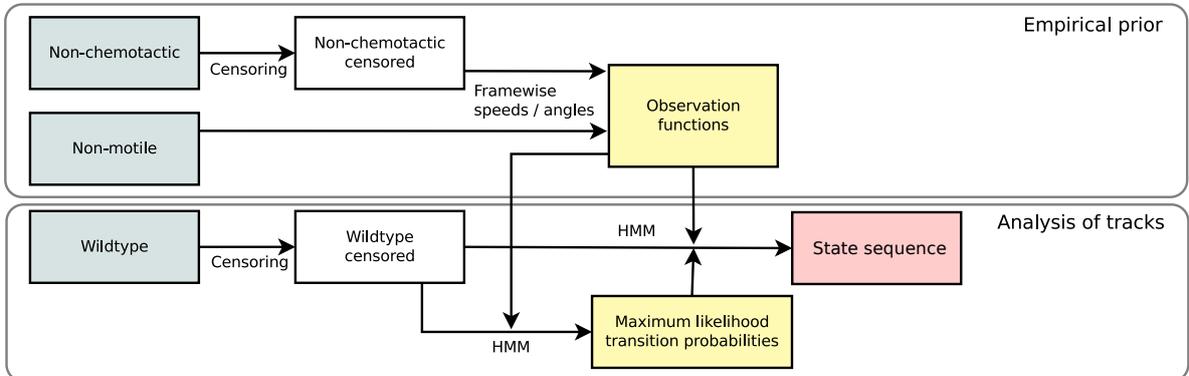


Figure 4.1: *Flow diagram of the stages involved in analysing the experimental tracking data. Tracking data inputs are shaded blue, yellow denotes quantities required in the analysis methods and red denotes the output quantity of primary interest.*

4.1.1 Experimental methods

All work pertaining to the experimental data used in this thesis was carried out by David Wilkinson, Department of Biochemistry, University of Oxford. Imaging and tracking were

performed on three different strains of *R. sphaeroides*: wildtype (WS8N), a non-motile mutant (467) and a non-chemotactic mutant that is incapable of stopping (1353). Details of the experimental protocol used to create the mutant strains, and the growth conditions, are given by Pilizota *et al.* [156]. Bacteria were transferred in motility buffer to a glass tunnel slide [155] for observation under the microscope. Imaging was performed at 50 frames per second using a phase contrast microscope with a $40\times$ magnification objective lens. The images are captured in 256 level greyscale, 640 pixels (px) wide and 480 px in height. The physical size of the microscope field of view was determined by imaging a calibration slide, which showed that a single pixel is equivalent to $0.12\ \mu\text{m}$. The field of view is hence a rectangle $76.8\ \mu\text{m}$ wide and $57.6\ \mu\text{m}$ high. For each strain, imaging was performed with the microscope focused both at the surface of the microscope coverslip and approximately $100\ \mu\text{m}$ below the top coverslip. In the case of surface swimming, most observed bacteria swim in the plane of the coverslip, as discussed in Chapter 1. In the case of focusing below the surface, which we denote the ‘bulk dataset’, the observed cells are swimming freely in the medium and may stray out of the focal plane. Typically between 10 and 20 minutes of footage are acquired in each of the six cases (three strains, both surface swimming and bulk swimming), from which we obtain between 3000 and 7000 tracks. Images from a surface swimming video and a bulk swimming video are shown in Figure 4.2. Short video clips of the raw data are available online, see Appendix B.1 for links. As we shall discuss in more detail in Section 4.3, the experimental method presented here has the advantage of being relatively cheap and straightforward to implement compared to traditional methods, and high-throughput in terms of the number of tracks acquired. In the case of the bulk dataset, the data obtained are a two-dimensional projection of the underlying three-dimensional bacterial trajectories; in the same section we also discuss this effect.

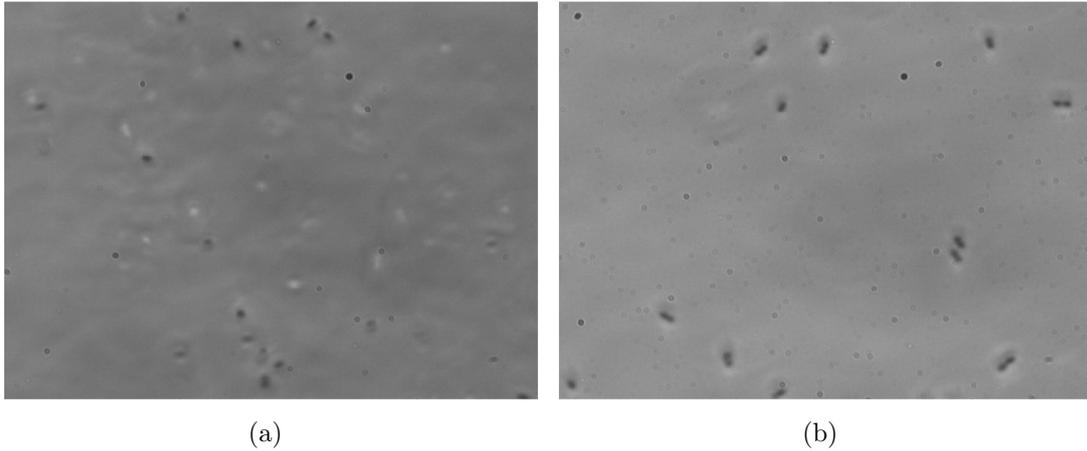


Figure 4.2: *Representative images obtained with the setup described in the main text. (a) The objective focus is approximately $100\ \mu\text{m}$ below the top coverslip (bulk swimming). (b) The objective focus is at the top coverslip (surface swimming).*

4.1.2 Tracking

The problem of tracking moving targets in videos was reviewed in detail in Section 1.3. We approach the problem as a two-stage process, involving object detection followed by data association. Representative images taken from our video dataset are shown in Figure 4.2 for surface and bulk swimming bacteria. Bacteria appear as light or dark objects upon a grey background, depending on their perpendicular distance from the focal plane. The majority of bacteria in the surface videos are observed as dark objects, indicating that they are in the focal plane. Light objects represent bacteria swimming out of the focal plane. The bulk videos exhibit greater variation in the appearance of cells, as they are distant from the surface and thus able to swim out of focus. Interference patterns are also visible in the forms of concentric rings of varying intensity, originating from cells which are out of the focal plane.

Similar microscopy data have been analysed previously by Xie *et al.* [211], who use a sophisticated kernel-based method of object detection followed by a data association method based on a modified nearest neighbour approach. Our approach differs fundamentally as our main focus is on the data association stage. In order to track bacteria in our videos, we employ a relatively simple object detection method followed by a recent

implementation of the PHD filter using a Gaussian mixture model (denoted GM-PHD) by Wood *et al.* [209]. The GM-PHD filter is a target association method that performs well even when the number of false positives is high, which is not the case in the method used by Xie *et al.* [211].

Object detection

The object detection stage is carried out in several steps:

1. compute the background as the mean intensity over all frames;
2. subtract the background from all frames;
3. find pixels in each frame with intensities after background subtraction above and below specified threshold parameters;
4. cluster groups of pixels that are 4-connected, meaning that every pixel in a cluster has another pixel in the same cluster in one of the four neighbouring sites around it. This is achieved using the Matlab function `bwlabel`;
5. discard any clusters containing fewer than a defined number of pixels, ρ_P ;
6. find the centroids (centre of mass) of each of the remaining clusters.

The centroids computed using this method now represent the targets present in each frame. The choice of threshold parameters was made separately for each video based on manual (by eye) verification that the regions correctly segmented bacterial cells in the images. The values of these parameters were chosen to minimise the number of false negatives, as the data association routine is robust to even high levels of spurious targets [209]. The minimum cluster size constraint was applied to the region data to remove spurious targets, which are often characterised by a small region size (typically fewer than 5 px). The minimum cluster size was fixed at 5 px, which removed a reasonable number of false positives whilst having no effect on true positives (real targets) as these

regions generally enclose greater than 10 px.

Target association

For the purposes of associating targets in consecutive video frames, we use a multitarget tracking method based on the multitarget Bayes filter. In contrast to the heuristic tracking algorithms based on the nearest neighbour scheme, this approach is probabilistic. The multitarget Bayes filter can be considered analogous to the standard (single target) Bayes filter, which gives a solution to the problem of updating our belief about the state of an unknown system by a sequence of noisy measurements. In our case, the measurements are the objects detected in each frame of the microscope movie, some of which correspond to real cells and others of which are false detections (clutter). The state of the system is the true position of a target in each frame. In most cases, the Bayes filter cannot be evaluated analytically or computationally. An important exception is the Kalman filter [101], which is used as the basis for many numerical approximation schemes in single particle tracking.

The *multitarget* Bayes filter introduces the additional complication that the number of targets in a given scene is unknown. Mahler [128] developed a mathematical framework to deal with such a problem, known as random finite set (RFS) theory. Just as the single target Bayes filter does not generally have an analytic or computationally tractable solution, neither does the multitarget filter. Instead the problem is simplified by only propagating the first moment of the filter through time, known as the probability hypothesis density (PHD), which gives the density of the expected number of targets at a point in space [129]. Whilst this simplifies the problem, it has no closed-form solution and further approximations are required to implement the PHD filter. The method developed by Wood *et al.* [209] uses a Gaussian mixture model to compute the PHD filter and achieve multitarget tracking, with the result denoted the GM-PHD filter.

The implementation of any tracker based on the Bayes filter requires the assumption of an underlying motion model. Wood’s tracker was originally developed to deal with sonar data and, as such, incorporates a straight-line motion model [207]. Changes of direction and stops, as seen in the *R. sphaeroides* microscope movies, are handled by the noise model. This incorporates Gaussian noise in the underlying velocity of each object being tracked. Providing that the variance of the Gaussian noise is set sufficiently high, changes in direction and pauses, such as those exhibited by wildtype *R. sphaeroides*, can be accommodated. Note that the tracking implementation can be extended to incorporate additional manoeuvring models [208]. The GM-PHD tracking approach copes with the type of clutter present in bacterial tracking scenarios substantially better than approaches based on the nearest neighbour algorithm [209].

We reiterate that the data output from the tracking algorithm is a vector of the form $\mathbf{R} = (\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_T)$ for each track, where \mathbf{r}_t designates a two-dimensional position vector at time t . The output of the tracking algorithm is next processed with a mean filter with a window width of three frames. The mean filter proceeds by replacing each position coordinate at frame t , \mathbf{r}_t , with the mean value of the three coordinates \mathbf{r}_{t-1} , \mathbf{r}_t , and \mathbf{r}_{t+1} . This stage smooths the tracks mildly, without significantly changing the appearance of stopping phases.

Figure 4.3 provides an illustration of typical tracking data. All notation is consistent with that introduced in Section 3.1. A bacterium swims in an approximately straight line, enters an approximately stationary stopped phase for some time, then swims off in a new direction. The crosses indicate observations made of the cell centroid at regular intervals of 0.02 s.

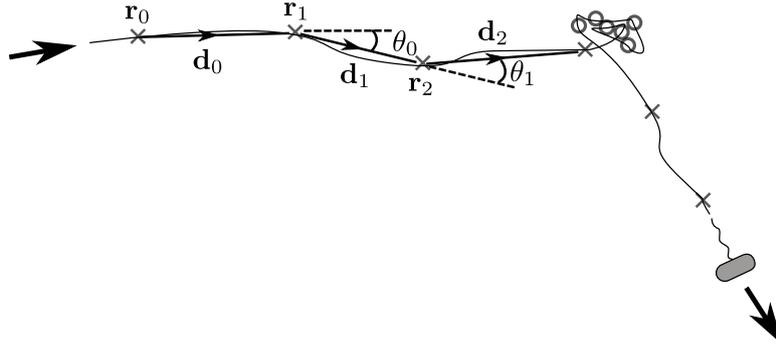


Figure 4.3: *Illustration of the representation of the tracking data from a motile bacterium. The black line indicates the true underlying trajectory of the bacterium, crosses correspond to observations made during a running phase, and circles correspond to observations made during a stopping phase.*

4.1.3 Estimating the distribution of framewise speeds and framewise angle changes

As discussed in Chapter 2, the estimated pdfs of framewise speeds and angle changes for all tracks in a dataset are quantities of interest, as they contain information about the motion of bacterial populations. Furthermore, we require the estimated pdf of framewise speeds when determining the parameters controlling the censoring process for the non-chemotactic dataset (indicated by the top left arrow in Figure 4.1). In order to estimate this pdf, we first compute all framewise speeds and angles for each track, as detailed in Section 3.1. In the case of framewise speeds, we next compute a KDE (see Section A.4) to obtain an unbiased estimate of the underlying pdf of framewise speeds. The KDE generally performs well on framewise speed data, since the distribution is often well described by a normalised sum of Gaussians. The same is not generally true of the distribution of framewise angle changes, which may, in the case of a track with predominantly running transitions, be strongly peaked around zero. In such cases, which we encounter when dealing with the non-chemotactic strain, plotting a histogram of the observed framewise angle change data indicates that the distribution is approximately exponential. The KDE fails to provide an accurate estimate of the pdf of the observed data here, so we instead resort to fitting the data with an exponential distribution.

4.1.4 Track tortuosity

Tortuosity is a measure of how tortuous a path is, meaning how many twists and turns it contains (see Section 2.2.2). In the context of bacterial taxis, a tortuous track may be produced by a bacterium that wobbles or rotates as it swims, so that it does not travel in a straight line, but rather along a winding route. Similarly, bacteria that move solely by Brownian motion (i.e. bacteria that are in a stopped phase), or that are stuck to a surface, trace out a highly tortuous trajectory as they change direction often. We are interested in estimating the tortuosity of a track as this is a key quantity used to characterise the motion of swimming bacteria, as well as being a useful property for the purposes of filtering the dataset (discussed in Section 4.1.6). Several methods have been proposed for estimating tortuosity, many of which are evaluated by Grisan *et al.* [81]. We choose to employ a method proposed by Lewiner *et al.* [118], in which a three-point estimator of the *curvature* of a track is used as a measure of the tortuosity. The curvature is defined for a given position, \mathbf{r}_i , $i = 1, \dots, T - 1$, in a track by

$$\kappa(\mathbf{r}_i) = \frac{\angle(\mathbf{r}_i - \mathbf{r}_{i-1}, \mathbf{r}_{i+1} - \mathbf{r}_i)}{\|\mathbf{r}_i - \mathbf{r}_{i-1}\| + \|\mathbf{r}_{i+1} - \mathbf{r}_i\|} = \frac{\angle(\mathbf{d}_{i-1}, \mathbf{d}_i)}{\|\mathbf{d}_{i-1}\| + \|\mathbf{d}_i\|}, \quad (4.1)$$

where the track notation is the same as defined in Chapter 3 and $\angle(\mathbf{d}_{i-1}, \mathbf{d}_i) \in [-\pi, \pi)$ denotes the turning angle between the vectors \mathbf{d}_{i-1} and \mathbf{d}_i . The curvature is undefined for the first and last points in a track, as we require three adjacent points to estimate it. Every track therefore has $T - 2$ curvature values associated with it. We use the median curvature value as a summary statistic, since this is more robust than the mean to extreme values of $\kappa(\mathbf{r}_i)$ that may occur when the denominator in (4.1) is close to zero. We use the value of the median curvature when filtering tracks, as we discuss in Section 4.1.6.

4.1.5 The minimum bounding radius and effective mean speed of a track

A spatial characteristic of each track in the experimental datasets is the radius of the smallest possible circle enclosing all points in the track, denoted the minimum bounding radius (MBR). This can be computed numerically in an iterative fashion, as described in Appendix B.2. The effective mean speed (EMS) of a track is defined as the MBR divided by the total duration of the track. This quantity is interpreted as the mean speed of the equivalent straight-line track; tracks that are approximately straight will have a true mean speed close to the EMS, whereas tracks with many twists and turns will have a higher true mean speed than the EMS.

4.1.6 Censoring tracks

When analysing experimental tracks from the wildtype or non-chemotactic datasets, there are three important considerations that were not necessary when we were working with simulated tracks in Chapter 3, namely: tracking errors, non-motile cells and highly tortuous tracks. First, the tracking process used to extract tracks from microscope videos (described in Sections 4.1.2 and 4.1.2) may fail with certain tracks and produce a purely artefactual result. This is a particular concern if the resultant failed track displays behaviour which differs substantially from correct tracks, since even a small number of failed tracks may affect the inferences we draw from the data. We aim to identify and remove these tracks before we draw any conclusions from the experimental datasets using a simple method to identify seriously anomalous behaviour based on the framewise speeds observed in each track. Any tracks containing one or more framewise speeds greater than a threshold value, denoted ρ_{FS} , are considered to be anomalous and discarded from the dataset. We take $\rho_{\text{FS}} = 90 \mu\text{ms}^{-1}$, a value significantly greater than the *mean* swimming speed of *R. sphaeroides*, which is approximately $40 \mu\text{ms}^{-1}$. We allow such a large margin for variation in the framewise speed as small errors in the object detection and associ-

ation stages can generate large fluctuations. For example, an inaccuracy in the process of detecting the centroid of an object in two consecutive frames (as depicted in Figure 4.4) could lead to the distance between the two centroids being exaggerated by 6 px ($0.72 \mu\text{m}$), resulting in an artificial increase to the reported framewise speed of $36 \mu\text{ms}^{-1}$ for a frame rate of 50 s^{-1} . We do not wish to discard tracks containing a few instances of such inaccuracies, since these quantities will not dominate the population average. This effect is expected to be minor when all tracks in a dataset are considered, and we note that an underestimation of the framewise speed is equally probable. Observed framewise speeds above the cutoff value of ρ_{FS} are unlikely to arise from such a source of noise; these are instead treated as a tracking error and the whole track is discarded.

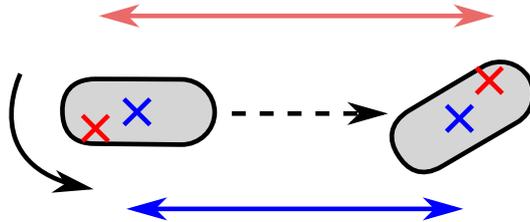


Figure 4.4: *An illustration of the error introduced into the measurement of framewise speeds due to inaccurate centroid detection. The figure depicts a swimming *R. sphaeroides* cell which is also rotating about its short axis, as observed over two consecutive frames. The true cell centre is shown in blue in each position, and the inaccurately-determined centroid is shown in red. The corresponding measurements of framewise distance are shown by the coloured arrows.*

In addition to tracker errors, a second consideration is the presence of a significant portion of non-motile tracked cells, as is usually observed in experiments of this kind [26, 152]. This is easily demonstrated by considering the tracks derived from the non-chemotactic strain. As this mutant is unable to stop, the derived tracks should (with some rare exceptions due to failure of the tracker or cell collisions) partition into two distinct groups, swimmers and non-motile cells. As Figure 4.6 and the related discussion in Section 4.2.2 demonstrate, the estimated distributions of framewise speeds and of median curvatures across all tracks in the non-chemotactic dataset are indeed both bimodal, with one mode corresponding to non-motile cells. Reasons for a lack of motility include cell death, a defective component in the cellular motility machinery, cell damage due to experimental

handling and (in the case of surface experiments) accumulation at the surface. In this thesis we consider the swimming behaviour of bacteria, hence the non-motile subpopulation is not relevant. It may play a role in the population-level description of bacterial motion, however we learn nothing about its motion (which is approximately Brownian) from the present analysis techniques.

The models used to simulate tracks (described in Section 3.2) incorporate a simple noise model which is unlikely to capture all of the effects that cause bacteria to swim in imperfectly straight lines. A third way in which the experimental data differ from simulated data is the wide range of tortuosities exhibited by real tracks, due to variation within the populations of *R. sphaeroides* being studied. Several tracks appear to be very tortuous, possibly as a result of discretization errors in the object detection stage when dealing with bacteria swimming in severely helical paths or with substantial cell body motion, as illustrated in Figure 4.4. This form of noise was noted by Hill and Häder [88] in their tracking study on the motion of algae. Other possible causes for tortuous tracks are bacteria with a damaged or defective flagellum, and two bacterial cells swimming whilst stuck together, prior to cell division. None of the methods discussed in Chapter 3 are able to cope with highly tortuous tracks, as these exhibit many large framewise angle changes and low framewise speeds in the running phase. It is therefore very challenging to discern stopping phases in such tracks, either automatically or by manual inspection.

Tortuous tracks are apparent in the non-chemotactic and wildtype datasets, both in the bulk and at the surface, and it is necessary to remove them from the dataset before performing any further analysis. The removal of these tracks in the present work is due to the inability of the presently available analysis methods to cope with the wide range of tortuosities observed in populations of motile *R. sphaeroides*, not because this behaviour is not itself of interest. We discuss this further in Section 4.3. Highly tortuous tracks constitute a small proportion of the total dataset: fewer than 10% of the tracks are

discarded for this reason, a similar fraction to that observed by Dowd and Matsumura [60] in their tracking study (see Table 4.2 and discussion in Section 4.2). Furthermore, the tracks arising from non-motile cells are highly tortuous, as they involve very many changes of direction, hence filtering to remove tracks from non-motile cells and filtering to remove tortuous tracks are closely related. Indeed, a single filtering stage is sufficient in the case of the non-chemotactic dataset to remove both tortuous and non-motile tracks.

The censoring processes applied to the non-chemotactic running-only strain and the wild-type strain differ, as the bacteria from these strains have different patterns of motility. We now discuss these processes for the two strains.

Filtering non-motile tracks from the non-chemotactic dataset

Non-chemotactic bacteria are expected to swim in approximately straight lines, with no sharp reorientations or pauses; any track exhibiting such events is likely to be due to a defective or non-motile cell and we wish to discard it. To identify such tracks, we assert a minimum EMS, as defined in Section 4.1.5. This threshold value is denoted ρ_{EMS} . Tracks with an EMS below this cutoff value are discarded.

In order to select an appropriate value for ρ_{EMS} , we consider the effect of the censoring stage on two key quantities, (i) the fraction of tracks remaining in the dataset after censoring, x_{remain} , and (ii) the estimated fraction of tracks originating from motile cells (as opposed to non-motile ones), x_{motile} . The latter quantity is computed using the method illustrated in Figure 4.5. We first identify the modal framewise speeds for motile and non-motile bacteria, which correspond to the global maximum in the estimated framewise speed pdf in the case of very severe censoring ('running peak' dominates) and no censoring ('non-motile peak' dominates). This process assumes that the pdf of framewise speeds is dominated by the density from non-motile cells, which we ascertain empirically (see Figure 4.6(a)). If this were not the case, a different method would be

required. The intermediate framewise speed is defined as the mean of these two values. In censoring cases between these two extremes, we identify the modal framewise speed for motile and non-motile cells, denoted s_m and s_{nm} , respectively, by searching for a maximum above and below the intermediate framewise speed. The corresponding values of the estimated framewise speed pdf at the motile and non-motile modes are denoted f_m and f_{nm} . We approximate the proportion of motile cells in the sample by

$$x_{\text{motile}} = \frac{f_m}{f_m + f_{nm}}. \quad (4.2)$$

We seek to maximise a weighted linear combination of the values x_{remain} and x_{motile} ,

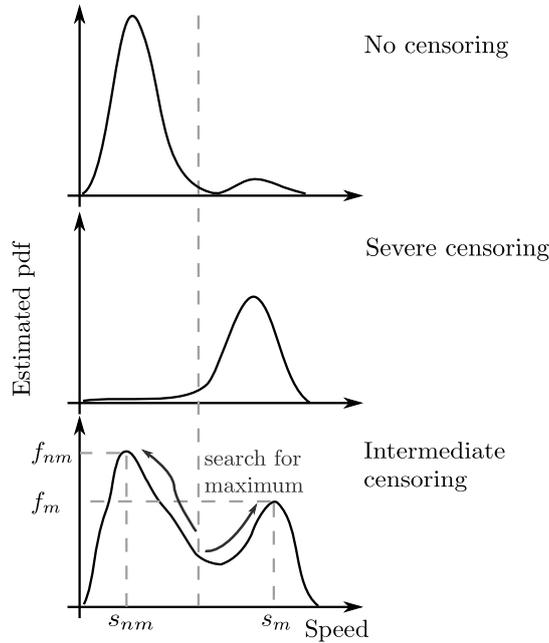


Figure 4.5: An illustration of the process used to estimate the proportion of tracks from motile cells based on the estimated pdf of the framewise speeds of the whole dataset. Two extreme censoring cases are first used to identify an intermediate framewise speed in-between the modal peaks for motile and non-motile cells (middle and top plots, respectively). In intermediate censoring cases, a local maximum is sought above and below the intermediate value, resulting in the values s_m and s_{nm} , with corresponding densities f_m and f_{nm} . The proportion of observed tracks which are motile is then approximated by $f_m/(f_m + f_{nm})$.

given by

$$\gamma x_{\text{remain}} + (1 - \gamma)x_{\text{motile}}, \quad (4.3)$$

where γ is the weighting coefficient, with $0 < \gamma < 1$. The parameter ρ_{EMS} is varied

over a range of values, and the corresponding values of x_{remain} and x_{motile} are computed. We compute the value of the weighted linear combination for several values of $\gamma \geq 0.5$. This constraint on γ is imposed as we are more concerned with filtering the majority of non-motile and highly tortuous tracks than preserving all tracks.

Filtering non-motile wildtype tracks

In the case of the wildtype *R. sphaeroides*, the censoring approach used for the non-chemotactic strain is not appropriate, as the cells are able to reorientate by stopping and therefore may not travel in an approximately straight line. The EMS of a valid wildtype track, which we wish to retain, may therefore be substantially lower than the true mean speed, so this is not a useful quantity for censoring the wildtype dataset. Since stopped phases in *R. sphaeroides* tracks typically make up a small proportion of the total track, the median curvatures of the tracks are not expected to be a great deal larger than those of the non-chemotactic bacteria. We therefore filter tracks from the wildtype bulk dataset based on their median curvature, with a maximum threshold value, denoted ρ_{MC} , chosen manually by considering the nature of tracks over a range of median curvatures. As described in Section 4.1.4, we use curvature as a quantitative measure of the tortuosity of a track. In addition, we assert a minimum value for the MBR, called the *minimum extent* and denoted ρ_{E} , and discard tracks whose MBR is lower than this cutoff value. We set $\rho_{\text{E}} = 6 \mu\text{m}$ (50 px), which ensures that the minimum bounding circle is required to cover around 1/5 of the height of the field of view (around 1/6 of the width), and approximately 1/40 of the area. Whilst this step may remove some short tracks that are acceptable in the sense that they have low tortuosity and arise from a motile bacterium, the majority of tracks covering such a small extent are either not of great interest (for example they contain no stopping phase and merely consist of a short run), or due to a non-motile cell.

4.1.7 Analysis methods

Three methods for analysing tracks were described and assessed in Chapter 3. We showed in the same chapter that there are several advantages to the HMM-based approaches. In particular, the heuristic method generates a very large number of false positives, in addition to producing highly biased statistics for stopwise angle changes. We therefore use only the full and speed-only HMM methods to analyse the experimental data. The analysis process is similar to that used on the simulated datasets in Chapter 3, with the addition of the censoring stage presented in the previous section. Wherever post-processing is applied, we use the same parameters ($\tau_{0,\min} = \tau_{1,\min} = 2$ frames) as were used in Chapter 3. These parameters mean that the post-processing stage removes any running phases or stopping phases that have a duration of a single frame from the run status of all tracks being considered.

4.1.8 Determining the observation function for the analysis methods based on the HMM

A key component of the HMM-based methods is the observation function, denoted $b_{ij}(\mathbf{y}_t)$ for the full method and $b_i(\mathbf{y}_t)$ for the speed-only method, which was previously introduced in Section 3.1.2. This function embodies our prior knowledge of the bacterial swimming motion. We now discuss the process by which we determine this function for the permitted values of i and j , both of which can take the values 0 and 1, corresponding to stopping and running, respectively. In the case of the full HMM method, the observation function depends on both the speed and the direction of motion relative to the previous direction of motion. The speed and angular components are assumed to be independent, as discussed in Section 3.1.4, so that the observation function is separable in the two independent variables. We now describe the process of estimating the observation function for the full HMM method. We show that the speed-only observation function is then calculated using a straightforward modification to the full observation function.

We assume that stopping phases in wildtype *R. sphaeroides* are similar in terms of the observed motion to the motion of the non-motile strain, and that running phases are similar to the motion of the non-chemotactic strain, as discussed in Chapter 3. This assumption is assessed in Section 4.2.3, with supporting experimental data. The speed component is identical for the stop-to-stop case, $(i, j) = (0, 0)$, and the run-to-stop case, $(i, j) = (1, 0)$, and is obtained using a KDE of the pdf of observed framewise speeds in the non-motile strain. Similarly, the speed component for stop-to-run, $(i, j) = (0, 1)$, and run-to-run, $(i, j) = (1, 1)$, cases is obtained by applying a KDE to the observed framewise speeds from the non-chemotactic tracks, after censoring.

The angular component is the same for the stop-to-stop, run-to-stop, and stop-to-run cases, all of which are assumed to follow a uniform distribution. We test this assumption in Section 4.2, by computing the observed angle change distribution for the non-motile dataset. We mentioned in Section 3.1.4 that the assumption of uniform angle changes in going from a stop to a run may be an oversimplification; if *R. sphaeroides* bacteria exhibit any angular persistence over the course of a stop, this assumption does not hold. Nevertheless, the use of the uniform distribution constitutes a weak prior in this case, meaning that the observed data are expected to influence the final probabilities to a greater extent than the observation function. We require such an assumption to maintain the Markov property of the movement process. The angular component of the run-to-run case, $(i, j) = (1, 1)$, is obtained by fitting an exponential distribution to the magnitude of the observed framewise angle changes. This is achieved by setting the exponential rate parameter to be equal to the reciprocal of the mean of the magnitude of observed framewise angle changes. In contrast to the situation with the stop-to-stop case, this is a relatively strong (*informative*) prior, since the exponential distribution will strongly influence the final estimated probability of being in a particular state.

In using the magnitude of observed framewise angle changes, we are assuming that the

distribution is symmetric. This was discussed previously in Section 3.1.4. The result shown in Figure 4.18 below indicates that the motion is not perfectly symmetric in the surface dataset, as the observed angle changes during running phases are biased towards clockwise rotation. Although the bias is statistically significant, the skewness of the observed angle change distribution is small, hence our assumption is justified.

The observation function for the speed-only HMM method may be determined by ignoring the angular component of the observation function in the full method. The stopping observation function, $b_0(y_t)$, is therefore equal to the stop-to-stop speed component of the full method, $b_{00}(\mathbf{y}_t)$. Similarly, the running observation function, $b_1(y_t)$ is set to be equal to the speed component of the run-to-run speed component of the full method, $b_{11}(\mathbf{y}_t)$.

4.2 Results

We now present the results obtained when the methods described in Section 4.1 are applied to tracks obtained from real experimental data. Unless otherwise stated, the results presented in this section are obtained from the bulk dataset. Where a comparison is drawn between the bulk and surface datasets, we make clear that both datasets are being considered. Selected analogous results for the surface dataset are presented in Appendix B.3. The number of tracks in each of the datasets considered in this section before and after the censoring stages is given in Table 4.2.

The estimated pdf of framewise speeds in the uncensored non-chemotactic tracks is shown in Figure 4.6(a), along with that for the non-motile tracks. The distribution of framewise speeds is in both cases dominated by a low speed peak due to non-motile cells. The similarity of the estimated pdfs at low speeds motivates treating the non-chemotactic tracks as comprising two distinct subpopulations, non-motile and motile, with the non-motile

Table 4.2: The number of tracks in each of the datasets considered, before and after censoring. NC denotes non-chemotactic, WT denotes wildtype.

Dataset	NC bulk	WT bulk	NC surface	WT surface
Number tracks	3773	6832	3171	4032
Number above ρ_{FS}	212	706	329	394
Number below ρ_E	–	3138	–	2652
Number below ρ_{EMS}	2055	–	2150	–
Number above ρ_{MC}	–	620	–	98
Number remaining	1506	2368	692	888

population behaving essentially identically to the non-motile strain. In reality, the situation is likely to be complicated by the presence of tracks exhibiting a range of tortuosities, which may contribute to intermediate framewise speeds between these two extremes, but we assume that such tracks do not contribute significantly to the distribution of framewise speeds for the non-chemotactic dataset. This view is supported by the observed distribution of tortuosities in the uncensored non-chemotactic dataset, shown in Figure 4.6(b), which appears to consist of two relatively well-separated clusters corresponding to highly tortuous non-motile tracks and motile tracks with low tortuosity. Few tracks lie between these two clusters.

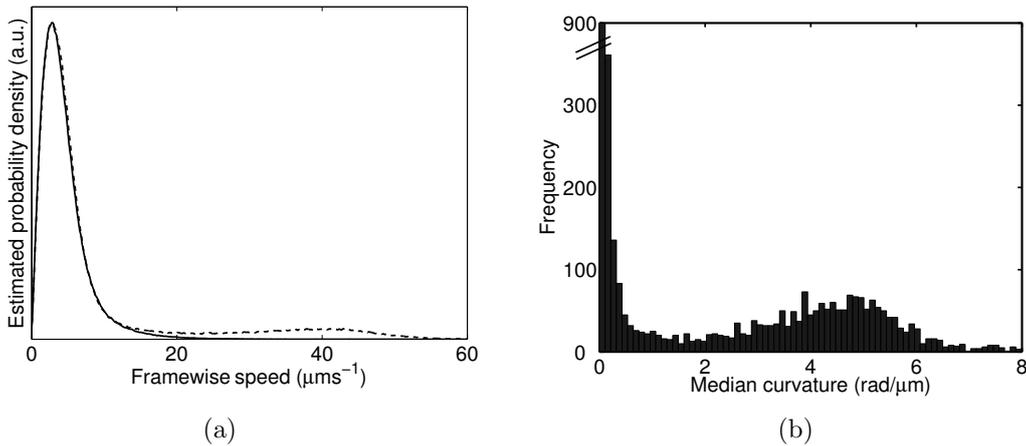


Figure 4.6: (a) Estimated probability distribution of framewise speeds for the non-chemotactic (black line) and non-motile (red line) bulk datasets. The distributions have been scaled so their maxima coincide. (b) Histogram of median curvature computed for all tracks in the non-chemotactic bulk dataset. All tracks are included, with the exception of tracks containing framewise speeds above $90 \mu\text{ms}^{-1}$. Note that the y -axis is broken; the density at low curvatures dominates the histogram. In both cases, the datasets are filtered to remove failed tracks.

4.2.1 Initial characterisation and censoring of wildtype tracks

In order to ascertain how the median curvature of a track relates to its appearance, and to its suitability for further analysis using the methods presented in Chapter 3, we plotted tracks with a range of median curvatures from the wildtype bulk dataset, shown in Figure 4.7. The tracks become increasingly jagged as the median curvature increases, indicating that this is a reasonable measure of tortuosity. There is also a tendency for tracks to have a smaller MBR (and hence cover a smaller area) as the median curvature increases. This is expected since tortuous tracks contain lots of high angle turns and therefore turn back on themselves more often on average than less tortuous tracks. Highly tortuous tracks also tend to have slower mean speeds. This is illustrated in Figure 4.8(a), which shows scatter plots of median curvature against mean track speed for the wildtype bulk dataset, after the removal of tracks exhibiting unrealistically high framewise displacements (blue points). In both cases, there is a relatively clear correlation between increasing median curvature and decreasing mean speed. This figure also shows the effect of removing tracks whose MBR is below $6 \mu\text{m}$ (red points). Many tracks with low mean speeds are removed, although a relatively large number of tracks remain in the dataset despite being highly tortuous.

In order to proceed, it is necessary to filter out the remaining tortuous tracks, which we achieve by discarding tracks with a median curvature above a specified cutoff value, denoted ρ_{MC} . Comparison of tracks (a)-(d) in Figure 4.7 demonstrates the effect of a stopping period in an otherwise smooth track. The median curvature of tracks (c) and (d) is higher than that of tracks (a) and (b) due to the inclusion of a stop. We must be careful to maintain such tracks, which contain useful information, when censoring by tortuosity. In practise, censoring by median curvature alone is not sufficient to discern all tracks with one or more genuine stops from those which are highly tortuous, however we attempt to minimise the number of useful tracks that are discarded when selecting ρ_{MC} . Guided by the appearance of tracks with varying median curvatures, similar to

those illustrated in Figure 4.7, we select $\rho_{MC} = 1.5 \text{ rad } \mu\text{m}^{-1}$. This is indicated in Figure 4.8(a) (grey dashed line).

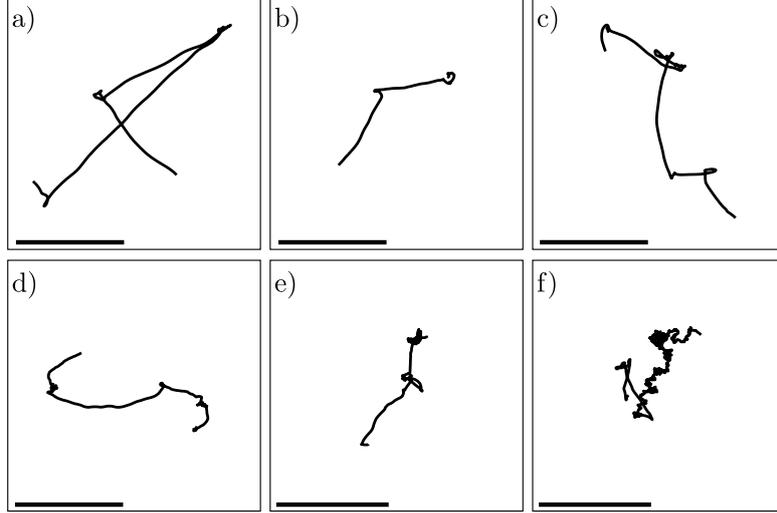


Figure 4.7: Tracks from the wildtype bulk dataset with different median curvatures. (a) $0.07 \text{ rad } \mu\text{m}^{-1}$, (b) $0.33 \text{ rad } \mu\text{m}^{-1}$, (c) $1.04 \text{ rad } \mu\text{m}^{-1}$, (d) $1.45 \text{ rad } \mu\text{m}^{-1}$, (e) $1.93 \text{ rad } \mu\text{m}^{-1}$, (f) $2.96 \text{ rad } \mu\text{m}^{-1}$. In each case, the black bar represents $20 \mu\text{m}$.

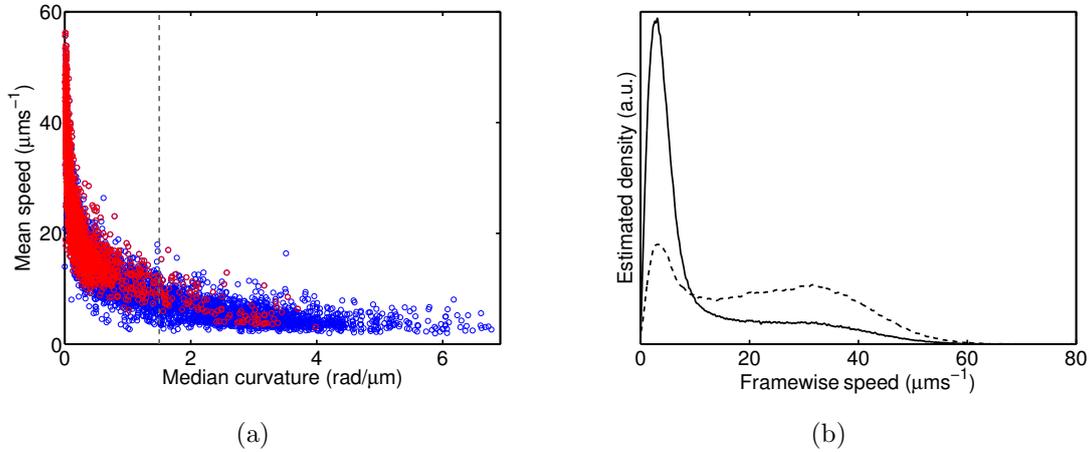


Figure 4.8: (a) Scatter plot of median curvature against mean track speed for the wildtype bulk dataset. Blue points come from all tracks in the dataset, after removal of those exhibiting unrealistically high framewise displacements. Red points are those remaining following censoring by applying a minimum extent of $6 \mu\text{m}$. Note that red points in all cases overlap with blue. The grey dashed line represents the maximum median curvature value; tracks with median curvature above this value will be discarded. (b) Estimated pdf of framewise speeds before censoring (solid line) and after censoring (dashed line) for the wildtype dataset.

A random sample of thirty tracks from the wildtype bulk dataset, before and after the application of the censoring process, is shown in Figure 4.9. Before censoring, the sample

of tracks displayed comprises a large proportion of small, non-motile tracks, in addition to several longer tracks, some of which are highly tortuous. Following removal of tracks which fail to meet the specified constraints ρ_{MC} and ρ_E , the selected tracks are all clearly due to motile cells. Several appear to contain reorientation events, indicating that we have not inadvertently filtered these out.

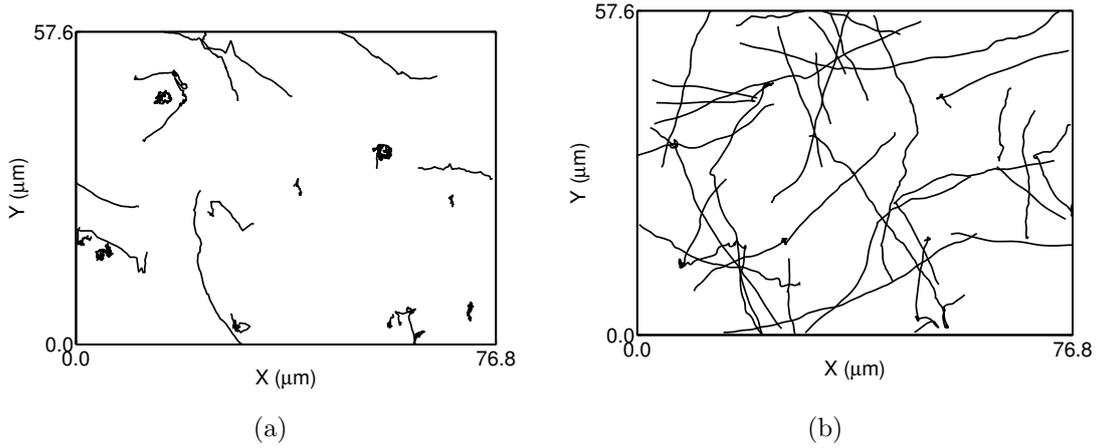


Figure 4.9: A random selection of 30 tracks from the wildtype dataset, (a) before and (b) after censoring. In both cases, the dataset was first processed to remove failed tracks, using the maximum framewise speed threshold. Censoring was carried out with $\rho_E = 6 \mu\text{m}$ and $\rho_{MC} = 1.5 \text{ rad } \mu\text{m}^{-1}$.

Figure 4.8(b) shows the estimated distribution of framewise speeds for the wildtype dataset, before and after the censoring process. The censoring process removes tracks which contribute substantially to the lower speed region, roughly below $15 \mu\text{ms}^{-1}$. This is consistent with the removal of non-motile or highly tortuous tracks with a low mean framewise speed. We cannot predict the expected distribution of framewise speeds *a priori*, as is the case when dealing with the non-chemotactic dataset, where we aim to minimise the density at low framewise speeds.

4.2.2 Optimising censoring parameters for the non-chemotactic strain

Figure 4.10 shows the effect of increasing the minimum EMS parameter, ρ_{EMS} , on the estimated pdf of framewise speeds in the non-chemotactic dataset; the contribution of the low speed peak diminishes considerably and a peak due to swimming cells appears centred on approximately $42 \mu\text{ms}^{-1}$. As ρ_{EMS} is increased, two contrasting effects are observed. First, an increasing number of tracks are discarded, as expected, thus the quantity x_{remain} decreases towards zero. Furthermore, the estimate of the fraction of the tracks which are due to motile bacteria, x_{motile} , increases towards one. Figure 4.11(a) shows these two effects. A weighted linear combination of these two quantities, defined in equation (4.3), is plotted in Figure 4.11(b) for several values of γ . The resulting estimate of the optimal value of the ρ_{EMS} , in the sense of maximising this linear combination, changes relatively little for $0.6 \leq \gamma \leq 0.9$. The final value of this threshold parameter, which is used for all subsequent analysis, is $\rho_{\text{EMS}} = 15 \mu\text{ms}^{-1}$, corresponding to $\gamma \approx 0.75$.

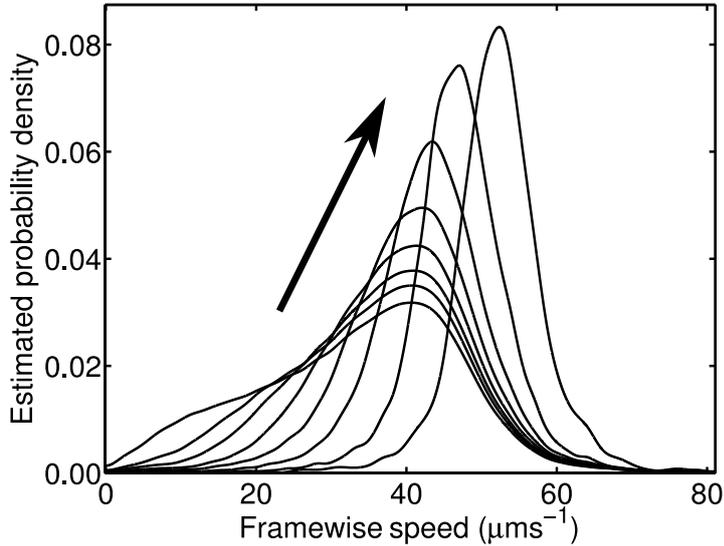


Figure 4.10: *Estimated pdf of framewise speeds for the non-chemotactic dataset after censoring, varying ρ_{EMS} from $12 \mu\text{ms}^{-1}$ to $48 \mu\text{ms}^{-1}$ in intervals of $6 \mu\text{ms}^{-1}$. The arrow shows the direction of change for the stated parameter variation.*

In contrast with the censoring process detailed in Section 4.2.1 for the wildtype dataset,

in which minimum MBR and maximum median curvature filters are required, only the minimum EMS criterion is required for effective censoring of the non-chemotactic dataset. This is because both tortuous tracks and non-motile tracks exhibit a significantly lower EMS than other tracks not in this category, hence a single constraint is sufficient.

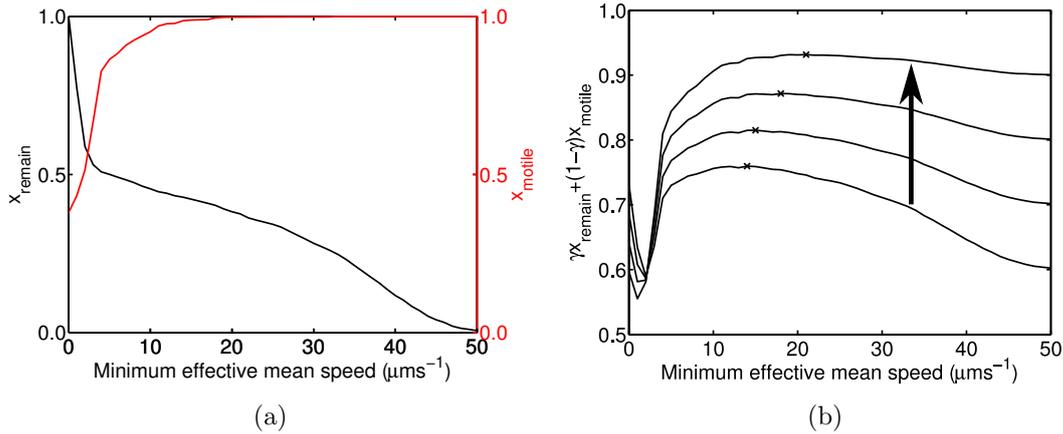


Figure 4.11: (a) The effect of varying ρ_{EMS} on x_{remain} (blue) and x_{motile} (green). (b) Weighted linear combinations of x_{remain} and x_{motile} for $\gamma \in \{0.5, 0.6, \dots, 0.9\}$. The black arrow shows the direction of increase of γ , and the black crosses indicate the position of the maximum. In the case where $\gamma = 0.5$, the maximum value is found when ρ_{EMS} is set to zero (no censoring).

4.2.3 Validating the run-and-stop model of motility

As we discussed in Chapter 3, a central assumption in the analysis methods based on the HMM is that the motion of non-chemotactic bacteria is similar to that of wildtype bacteria in a running phase, and likewise for the motion of non-motile bacteria and wildtype stopping phases. Having determined censoring parameters for the non-chemotactic and wildtype datasets, we are in a position to test this assumption. Figure 4.12(a) shows the KDE of the observed framewise speeds in the all three strains, with the non-chemotactic dataset censored as discussed in Section 4.2.2. Results from the wildtype strain are shown before and after censoring, which is carried out as described in Sections 4.1.6 and 4.2.1. The black dashed line in this figure and Figure 4.8(b) show the same distribution from the censored wildtype dataset on different y -axis scales. The distribution is bimodal, as

expected for tracks exhibiting running and stopping phases. The lower modal framewise speed, which is putatively due to bacteria in a stopping phase, corresponds very closely to that of the non-motile distribution. The agreement between the upper modal wildtype framewise speed, assumed to arise from bacteria in a running phase, and that of the non-chemotactic dataset is less good, but nonetheless qualitative. Figure 4.12(a) therefore provides initial evidence that our assumption is justified.

We test our assumption further by fitting the distribution from the wildtype, censored dataset with a linear combination of the distributions from the censored non-chemotactic and non-motile datasets. The fitting is performed by minimising the sum of squared residual errors using trust-region algorithm, implemented in the Matlab function `fmincon` [54]. The results are shown in Figure 4.12(b). Again, the agreement is qualitative, but several discrepancies are apparent. In particular, the non-chemotactic distribution is skewed to greater framewise speeds than the censored wildtype dataset, with a modal value around $10 \mu\text{ms}^{-1}$ higher. This discrepancy may be due to small differences in the swimming speeds of the non-chemotactic and wildtype strains. A further way in which the distributions differ is seen at intermediate framewise speed of $15 - 20 \mu\text{ms}^{-1}$. The linear combination distribution has a significantly lower density in this region than the wildtype distribution.

Having obtained the results shown in Figure 4.12, we are satisfied that our run-and-stop model of motility is reasonable for application to the *R. sphaeroides* dataset. The agreement in this figure, whilst not perfect, is encouraging considering that the distributions arise from three different mutants.

4.2.4 Observation function

The distributions of observed framewise speeds and angle changes for the non-chemotactic and non-motile strains are shown in Figure 4.13. The non-chemotactic data have been

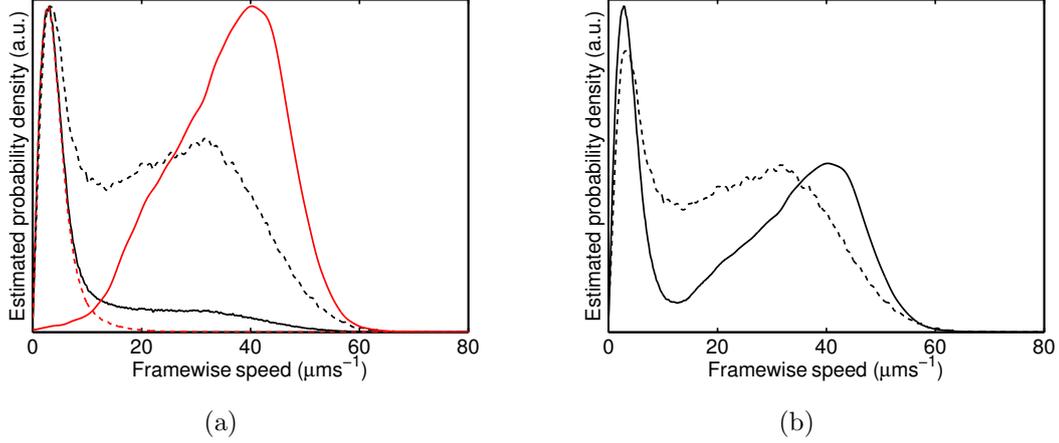


Figure 4.12: (a) Estimated pdf of framewise speeds for all three bulk datasets: (solid black) uncensored wildtype, (dashed black) wildtype, censored with $\rho_E = 6 \mu\text{m}$ and $\rho_{MC} = 1.5 \text{ rad } \mu\text{m}^{-1}$, (solid red) non-chemotactic, censored with $\rho_{EMS} = 15 \mu\text{ms}^{-1}$, and (dashed red) uncensored non-motile. All datasets have additionally been censored with the maximum framewise speed criterion with $\rho_{FS} = 90 \mu\text{ms}^{-1}$. The distributions have been scaled so their maxima coincide. (b) (Dashed black) estimated pdf of framewise speeds for the wildtype, censored dataset, (solid black) a linear combination of the censored non-chemotactic and non-motile distributions, optimised to give the closest fit to the wildtype, censored distribution.

censored to remove immotile cell tracks. As discussed in Section 4.1.8, the speed component of the observation function for run-to-run and stop-to-run transitions is taken to be equal to the KDE of the observed framewise speeds in the non-chemotactic dataset. Similarly, the speed component for stop-to-stop and run-to-stop transitions is equal to the KDE of the observed framewise speeds in the non-motile dataset. The angular component of the observation function for run-to-run transitions is taken to be equal to the exponential distribution that best fits the observed framewise angle changes in the non-chemotactic dataset. Finally, the angular component for all other transitions (stop-to-stop, run-to-stop, and stop-to-run) is assumed to be uniformly distributed on a circle, indicated by the dashed line in Figure 4.13(d). The wrapped uniform distribution is predicted from the theory of Brownian motion (see Chapter 6).

As Figure 4.13(d) shows, there is a discrepancy between the distribution of observed angle changes for the non-motile dataset and the assumed wrapped uniform distribution. The observed distribution is skewed towards small angles. This is likely due to a systematic

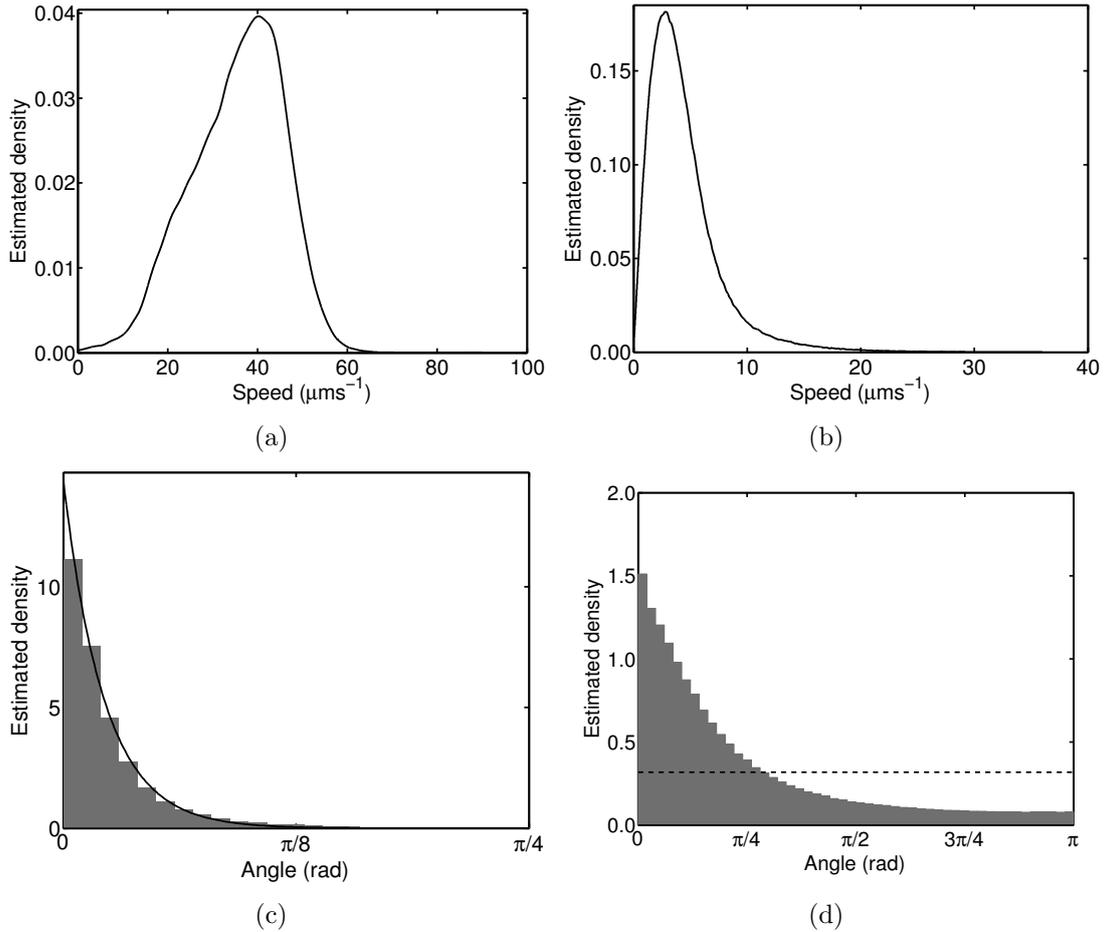


Figure 4.13: The functional forms of the observation functions, as derived from the estimated pdf of framewise speeds and angles for the non-chemotactic and non-motile bulk datasets. (a) Non-chemotactic strain estimated distribution of framewise speeds after censoring. (b) Non-motile strain estimated distribution of framewise speeds. (c) Non-chemotactic strain histogram of angle changes after censoring (grey), fit with an exponential distribution (black). (d) Non-motile strain histogram of observed angle changes (grey), with uniform approximation overlaid (black dashed line).

bias in the tracking algorithm, which is designed primarily for tracking objects that are moving in approximately straight lines. The tracker hence overestimates the number of transitions that are approximately straight line moves. In addition, the tracker ‘fills in’ broken tracks when detections are missing, again assuming straight line motion. This is reasonable for motile bacteria, but results in artefacts when applied to non-motile strains. We discuss this issue in greater detail in Section 4.3. As discussed previously, the assumption of a uniform angular distribution represents a weak prior. Faced with tracker artefacts, we prefer to select a weak prior that represents the underlying motion

reasonably, rather than tailoring our analysis to such artefacts. It is also possible to use a linearly interpolated fit to the histogram, rather than assuming a uniform distribution. We do not do so here due to time constraints, although we hypothesise that the results would not be significantly different.

4.2.5 Determining maximum likelihood transition parameters

We use trust-region minimisation [54] to minimise the negative log-likelihood of the wild-type bulk and surface datasets with respect to the transition probability parameters p_{01} and p_{10} , as described in Section 3.1.5. The negative log-likelihood surfaces are plotted in Figure 4.14 for the bulk and surface datasets, analysed using the full HMM method. The surfaces are qualitatively similar when computed using the speed-only method, though the likelihood values differ (data not shown). We convert the transition probabilities to rates for run-to-stop (λ) and stop-to-run (μ) transitions by dividing by Δt . We also compute the mean durations of stops ($\tau_{01} = 1/\mu$) and runs ($\tau_{10} = 1/\lambda$). The results are shown in Table 4.3. The results from both HMM-based methods are in reasonable agreement, with the speed-only method generating slightly higher estimates for both datasets. There is an apparent tendency towards longer stopping phases in the surface dataset relative to the bulk dataset. This is expected for cells swimming near a surface, as a fraction of the observed stops may actually be due to cells adhering temporarily to the surface. We note that the simulation study carried out in Chapter 3 showed that the transition probabilities are not necessarily an accurate reflection of the true underlying values when noise is present. We discuss these values and compare them with other literature values in Section 4.3.

Table 4.3: *Optimised values of the state duration and transition rate parameters for the two datasets and both HMM methods.*

Dataset	Method	τ_{01} (s)	τ_{10} (s)	μ (s^{-1})	λ (s^{-1})
bulk	full	0.29	0.63	3.45	1.59
bulk	speed-only	0.41	0.80	2.44	1.25
surface	full	0.49	0.65	2.04	1.54
surface	speed-only	0.67	0.75	1.49	1.33

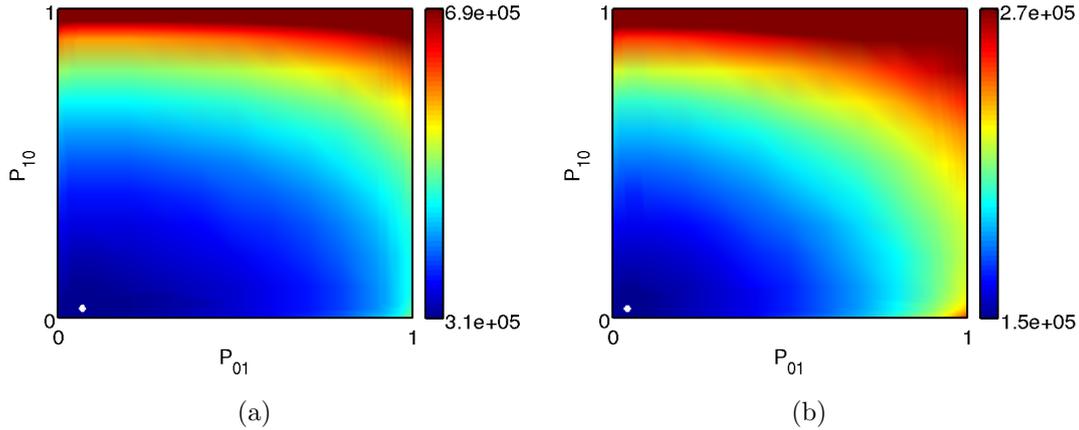


Figure 4.14: *Negative log-likelihood surfaces computed using the full HMM analysis method for the bulk (a) and surface (b) wildtype datasets. The white dot indicates the minimum, computed using a numerical optimisation algorithm.*

4.2.6 Assessing analysis method performance by manual inspection of tracks

When analysing tracks derived from experimental microscopy data, we have no ground truth with which to compare the locations of stopping phases inferred by the analysis methods. A robust and automated assessment of the analysis performance, similar to that carried out in Chapter 3 on simulated data, is therefore not possible. Nevertheless, a manual inspection of the inferred state sequence of tracks readily identifies a minority of tracks in which the analysis output is unrealistic, for example when there are a large number of rapid oscillations between running and stopping states. Figure 4.15 shows some individual tracks from the bulk wildtype dataset where the performance of the analysis methods may be assessed manually. In the left panel, several well-defined stopping regions within the tracks have been expanded for greater clarity. Note that, although the speed-only HMM method was used to compute the run probabilities in this figure, the results for these tracks are almost indistinguishable when the full HMM method is used (data not shown). On the right of the figure are shown a track from a bacterium swimming slowly in an exaggerated helical trajectory and a highly tortuous track. The helical track appears to contain a single genuine stopping event, however both analysis

methods incorrectly identify several of the helical turns as stopping phases, leading to an unrealistically rapidly oscillating state sequence. Application of post-processing to either analysis method circumvents this issue. It is not clear by looking at the tortuous track where the true stopping phases are, if indeed it contains any. All of the analysis methods infer multiple stopping phases along the length of the track, again leading to high frequency oscillations in the state sequence. The presence of such a track in the censored dataset motivated a manual examination of all tracks exhibiting either high median curvature or containing a large number of inferred stopping phases, which indicated that, of the 2780 tracks included in the bulk wildtype dataset, fewer than five are clearly identifiable as highly tortuous. Any effects from this minority of tracks, after pooling all analysed data, will be insignificant.

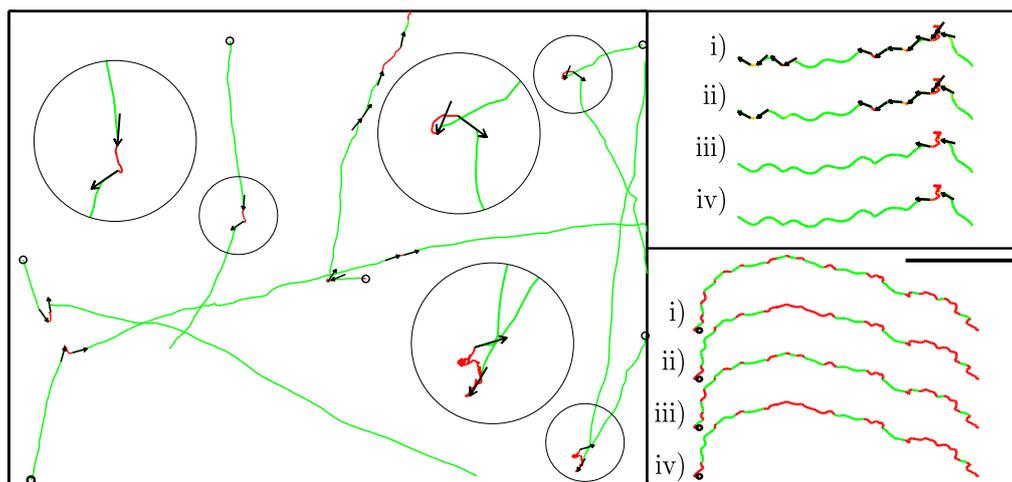


Figure 4.15: *Manual assessment of a selection of analysed tracks. Green indicates a running phase, red indicates a stopping phase, small circles indicate the starting position of the track, and pairs of arrows show the direction of travel of the bacterium immediately prior to and after a stop. (Left) A selection of tracks that were manually identified as having stopping phases correctly identified by the speed-only HMM method. Circled regions have been expanded for clarity. (Right) Track from a bacterium swimming in a helical trajectory (top), and a tortuous track with arrows removed for clarity (bottom). Lower case numerals denote analysis methods: (i) full HMM, (ii) speed-only HMM, (iii) full HMM with post-processing, and (iv) speed-only HMM with post-processing. The black bar is 10 μm long.*

4.2.7 Characteristics of the motion of wildtype bacteria

Figure 4.16 shows the estimated pdf of framewise speeds during all running and stopping phases identified in the wildtype bulk tracks. The estimated distributions were verified to change negligibly with the analysis method and/or inclusion of post-processing (data not shown). As expected, the framewise run speed distribution estimated from the wildtype dataset is similar to that estimated from the non-chemotactic dataset (see Figure 4.13(a) for comparison). Similarly, the estimated pdf of framewise speeds observed during stopping phases appears like that computed from the non-motile dataset (see Figure 4.13(b) for comparison).

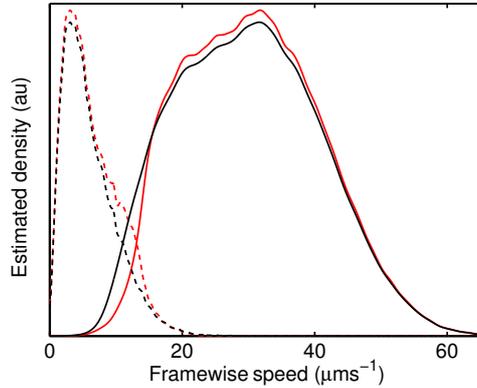


Figure 4.16: *Estimated framewise speed pdfs for running (solid line) and stopping (dashed line) phases in the bulk wildtype dataset, analysed using the full (black) and speed-only (red) HMM methods. Applying post-processing in either case made no significant difference to the estimation (data not shown).*

The estimated pdf of absolute stopwise angle changes (the angle change over the course of a run-stop-run transition) is shown in Figure 4.17(a) for the wildtype dataset. The two HMM-based methods produce almost indistinguishable results. Post-processing also makes little difference to the estimated distribution. Figure 4.17(b) shows an analogous plot for *E. coli* by Berg and Brown [24]. In contrast with our results, the mode of the distribution of stopwise angle changes in *E. coli* is not centred at zero. We discuss this further in Section 4.3.

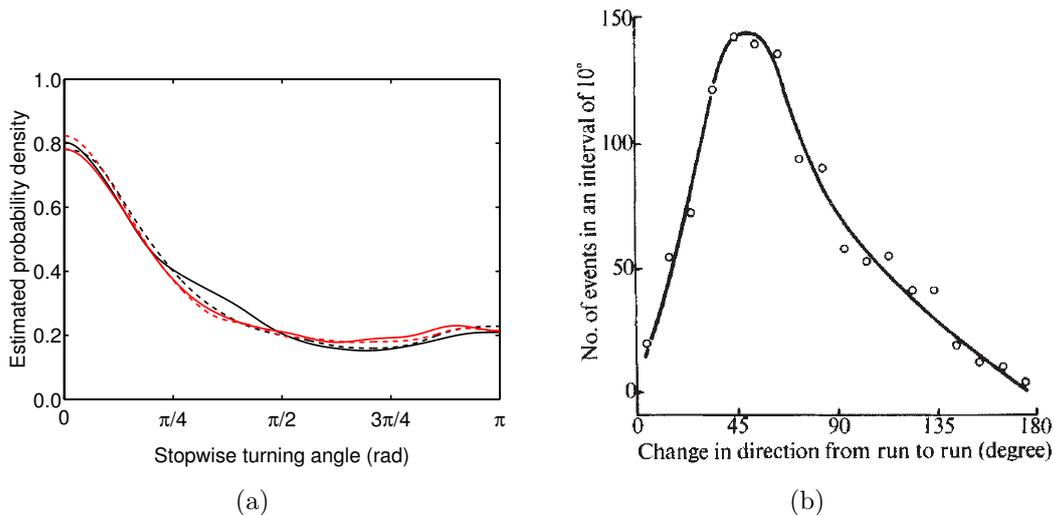


Figure 4.17: (a) Estimated pdf of the stopwise turning angles for the bulk wildtype dataset, computed using the full (black) and speed-only (red) HMM methods without (solid line) and with post-processing (dashed line). (b) Observed distribution of stopwise angle changes in *E. coli*, reprinted from [24] with permission.

The results shown in Figures 4.17(a) and 4.16, are encouraging, since they demonstrate excellent agreement between the two analysis methods, irrespective of the application of post-processing. The only tracks whose state sequences were found to vary significantly between the methods in our manual assessment in Section 4.2.6 were a small number of tortuous or helical tracks. The fact that the methods produce nearly identical results therefore supports our earlier findings that such tracks are a minority and do not discernibly affect the pooled statistics.

The protocol used to obtain the tracks considered in this chapter permits us to draw a comparison between bacteria swimming near and away from a surface. For example, many of the surface tracks shown in Figure B.4(b) appear to show a marked degree of curvature compared with the bulk tracks in Figure 4.9(b), in agreement with other studies of bacterial motion near a surface [74, 117, 119]. This is borne out in the framewise angle changes observed during a run, histograms of which are plotted in Figure 4.18. There is a significant degree of asymmetry evident in the surface tracks, with a strong bias towards clockwise rotation: 54.4% of the total number of non-zero framewise angle changes at

the surface were negative (due to a clockwise rotation), compared with 50.0% in the bulk. These values translate to highly significant evidence for unbiased swimming in the bulk, with a Bayes factor of 233, and swimming with an asymmetric bias at the surface, with a Bayes factor on the order of 10^{51} . An explanation of the Bayes factor, and the calculations required to obtain these results, are presented in Appendix B.4.

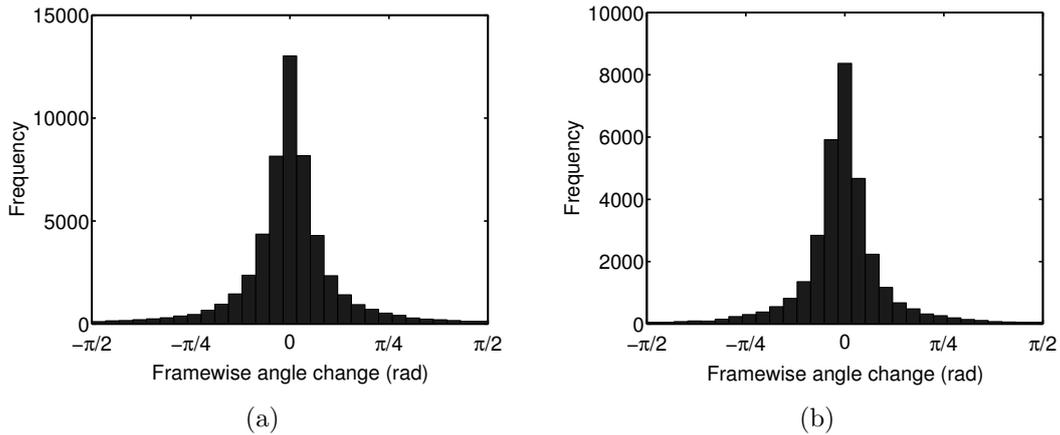


Figure 4.18: Histogram of the observed framewise turning angles during a run, computed using the speed-only HMM method without post-processing for (a) the wildtype bulk dataset, and (b) the wildtype surface dataset. The results upon inclusion of the post-processing process and from the full HMM method are very similar to the results shown here.

4.3 Discussion and conclusions

In this chapter, we have demonstrated the effective application of novel analysis methods to experimental data acquired using a relatively simple and inexpensive experimental protocol. The result is a high-throughput method to characterise bacterial motion. We showed that this method can be used to estimate certain key distributions, such as the pdf of stopwise angle changes, which is plotted in Figure 4.17(a). This pdf has previously been measured in *E. coli* (see Figure 4.17(b)), but has until now not been studied in *R. sphaeroides*. The distribution in *R. sphaeroides* is peaked about the origin, suggesting that angular persistence is present across a stop. As Figure 4.17 illustrates, the estimated pdf of stopwise angle changes measured here for *R. sphaeroides* differs from that recorded by Berg and Brown [24] for *E. coli*. We find a unimodal distribution centred at zero,

whereas Berg and Brown [24] find a bimodal distribution centred at approximately $\pm\pi/4$. This may be due to substantial differences in the reorientation mechanisms of the two species: *E. coli* undergoes rapid, active reorientation, achieved through the reversal of several flagellar motors, whereas *R. sphaeroides* reorientates more slowly, by an unknown mechanism that involves cessation of the rotation of the single flagellar motor [12, 156]. We return to this topic in Chapter 6. Alternatively, the discrepancy may be due to the substantially different experimental and analytic methods used in the two studies. For example, Berg and Brown track individual bacteria at a frame rate of 12.6 s^{-1} , whereas we simultaneously track multiple bacteria at a frame rate of 50 s^{-1} . In addition, Berg and Brown apply a series of heuristic arguments in order to identify reorientation phases, which may bias their findings towards larger angle changes, whereas we apply novel analysis methods based on the HMM.

Access to a sufficiently large dataset is important in order to elucidate more subtle effects, which might be missed if insufficient data are available. One such effect is the bias towards moving in arcs rather than straight lines, demonstrated in Figure 4.18. This bias has been predicted using a hydrodynamic model of bacterial swimming by Ramia *et al.* [165], and is consistent with experimental observations of *E. coli* [26, 58, 74, 117]. With the large dataset considered here, we are able to quantify this bias in *R. sphaeroides*. Furthermore, since the current method also enables the tracking of cells swimming freely away from a surface, we are able to show that such circular swimming is absent when bacteria are not near a surface, and demonstrate the statistical significance of the difference. To the best of our knowledge, this is the first systematic quantification of the turning bias.

The method used to obtain experimental tracks is crucially dependent upon the tracking algorithm, whose robustness to experimental noise makes it suitable for our application where other tracking systems fail [209]. The tracking algorithm is still prone to errors, however, and we have emphasised the necessity to identify and discard flawed tracks

based on applying a maximum plausible framewise speed. These were shown to occur in around 10% of tracks, for all of the datasets analysed. We have no reason to believe that the failure of the tracker biases the results, which would be the case if the failed tracks all exhibited a certain type of motile behaviour, and were hence underrepresented in the final dataset. Manual analysis of several failed tracks suggests that failures generally occur in complex scenes, such as when three or more bacteria approach one another closely (data not shown).

A disadvantage of the tracker is evident in Figure 4.13(d), which shows the histogram of observed framewise angle changes for the non-motile strain. The motion of non-motile bacteria is expected to be approximately diffusive, in which case the predicted angle change distribution is wrapped uniform (see Chapter 5 for a further discussion). Instead, the density of observed framewise angle changes is concentrated at small angles. We believe that this departure from uniformity is due to artefacts from the tracking algorithm, which is able to propagate tracks for several frames in the absence of detections. When this occurs, for example when a cell swims out of focus for one or more consecutive frames and is not identified in the object detection stage, the algorithm ‘fills in’ the missing sections of the track with straight-line motion estimates. This is expected to lead to a systematic bias towards small angle changes. This issue could be avoided in future work by disabling this feature of the tracker, as the HMM-based approaches may be extended to cope with missing data in a straightforward manner [177]. A further potential cause is the presence of systematic discretisation errors in finding the centroid of the cells (discussed in Section 4.1.6). Further work to address this issue could include incorporating a more sophisticated object detection stage, such as that used by Xie *et al.* [211]. Our use of a uniform pdf for the observation function (shown in Figure 4.13(d), dashed line) in the present work represents an uninformative prior (one which does not strongly affect the inferences drawn). Using a piecewise fit to the histogram of observed angle changes (Figure 4.13(d), red bars) for the observation function, instead of the uniform approxi-

mation, should not affect the results a great deal, as the observed distribution is still far broader than the equivalent framewise angle change distribution for the non-chemotactic dataset (compare Figures 4.13(d) and 4.13(c), noting that the x -axis scale differs).

We considered several important effects that must be taken into consideration before experimental tracks can be meaningfully analysed, and showed how censoring the datasets can help to mitigate such issues. In particular, the MBR and median curvature are very useful characteristics for censoring tracks that might otherwise lead to spurious inferences. A key example is the large proportion of non-motile cells, discussed in Section 4.1.6. A further, less prevalent but important concern are highly tortuous motile tracks, such as those shown in Figure 4.15. We reiterate that censoring the dataset in such a way to remove a significant proportion of tracks is not a desirable process, but is necessary for the current analysis approaches considered. Further work is necessary if tracks exhibiting a wide range of tortuosities are to be analysed. With a sufficiently large dataset, for example, a series of observation functions (see Figure 4.13) could be defined, with each corresponding to a range of tortuosities. The HMM-based methods might then be applied to many more recorded tracks than we are able to deal with in the present study.

An important question raised by the censoring process is how to censor unknown samples. For example, we may wish to analyse the motility of a mutant strain to determine whether it differs from the wildtype population. In such a case, we could not apply the threshold determination method in Section 4.2.1, as this would mean censoring the wildtype and unknown samples independently, and could introduce subjective bias. However, using the wildtype threshold parameters for unknown samples could result in removing bonafide tracks and falsely inferring that the motility in the unknown sample does not differ significantly from wildtype behaviour. Further work is required to determine the best practice in this situation, however we propose a method in which a fixed proportion of the most tortuous tracks are removed. A similar approach has previously been used

by Alon *et al.* [7], who sort tracks by the frequency of tumbles and discard the top and bottom 10%. This method avoids the need to select single or multiple threshold values.

The negative log-likelihood surfaces were plotted for the two HMM-based analysis methods in Figure 4.14. Whilst the surfaces are qualitatively similar, numerical optimisation to determine the minimum resulted in relatively different estimates for the transition rates, presented in Table 4.3. The results are in agreement with the findings of the simulation study in Section 3.3.1, in which the full HMM method generated larger estimates than the speed-only method for both transition rates (and thus smaller estimates for the mean duration). Furthermore, in the simulation study both methods generated estimates that were up to twofold lower than the true value over the range of noise levels considered. The discrepancy between the two methods in the inferred transition rates is thus an indication that any estimate of the rate parameters λ and μ should not be considered a reliable estimate of the true underlying rates. Nonetheless, we compare our findings with other values reported in the literature, shown in Table 4.4. It is immediately apparent that a wide range of transition rates have been recorded in the studies cited, despite the superficially similar experimental protocols. A few of the many possible explanations include the use of different wildtype strains, small differences in the composition of the motility buffer, and differences in the analysis methods. Comparing with our results, we see that the inferred value of μ is in reasonable agreement with the findings of Berry and Armitage [28] for the bulk dataset, and Brown [38] for the surface dataset. Our inferred value of λ is threefold greater than the next highest estimate for *R. sphaeroides*, however. The agreement for λ is better when compared with the study of *E. coli* by Berg and Brown [24]. We note that the tethered cell and tracking protocols differ a great deal. An important demonstration of the difference is the observation, by Poole *et al.* [157], that the use of antibody to tether *R. sphaeroides* cells to a microscope slide by their flagella substantially reduced their rotation speed and decreased the number of observed stops. This is consistent with our findings, as we estimate a larger value of λ than the tethered

cell studies listed in Table 4.4, corresponding to an increased number of stopping phases.

Table 4.4: Summarised literature values of transition rates in bacterial taxis. Standard deviations are given where they are available. The terms *bulk* and *surf.* refer to the dataset, while *full* and *speed-only* refer to the HMM method used to analyse the data.

Reference	Species	Method	μ (s^{-1})	λ (s^{-1})
[24]	<i>E. coli</i>	Single cell tracking	7.1	1.16
[28]	<i>R. sphaeroides</i>	Tethered cell	3.7	0.59
[38]	<i>R. sphaeroides</i>	Tethered cell	1.5	0.31 ± 0.19
[45]	<i>R. sphaeroides</i>	Tethered cell	0.97 ± 0.40	0.22 ± 0.10
[150]	<i>R. sphaeroides</i>	Tethered cell	0.96	0.22 ± 0.17
Section 4.2	<i>R. sphaeroides</i>	Tracking (bulk, full)	3.45	1.59
Section 4.2	<i>R. sphaeroides</i>	Tracking (bulk, speed-only)	2.44	1.25
Section 4.2	<i>R. sphaeroides</i>	Tracking (surf., full)	2.04	1.54
Section 4.2	<i>R. sphaeroides</i>	Tracking (surf., speed-only)	1.49	1.33

A further important consideration when considering the value of the inferred transition rates is the fact that they are computed for pooled data, so that individual variations between tracks are averaged over an entire dataset. Packer *et al.* [150] noted that there is considerable heterogeneity in switching rates within a bacterial population; this aspect was not included in the simulation study in Chapter 3 for simplicity. Considering each track separately would result in insufficient data being available for shorter tracks and those containing no run-stop-run transitions, and we do not consider that problem here. It may be possible to investigate population heterogeneity by applying the HMM-based methods to individual tracks obtained using single-cell tracking methods, as these tracks are generally longer (discussed in Section 1.2).

Unlike the simulation study in Chapter 3, the true state sequence is not known to us when working with real experimental tracks, and we are limited to manual inspection of the tracks to assess the performance of the analysis methods. A selection of tracks was shown in Figure 4.15 where the analysis methods appear to have performed well. These tracks were picked manually from the dataset because they appear easy to interpret, with clear running and stopping phases. Two tracks were also presented in the same figure,

for which one or more of the analysis methods had clearly failed to infer the correct run probabilities. One appeared to be the result of a cell swimming in a helical trajectory, while the other was highly tortuous and possibly arose from a bacterium with damaged motility apparatus. In the first instance, inclusion of post-processing helped to correct the inferred run probabilities. We note that the post-processing method is not included until *after* the maximum likelihood transition probabilities are computed. Therefore any tracks where the analysis methods fail could contribute towards a poor estimate for the transition probabilities. Manual assessment of the tracks remaining in the bulk dataset following censoring suggested that the failed tracks in Figure 4.15 were a minority, so this is unlikely to be a concern in the present study.

Using our experimental approach, we obtain two-dimensional position coordinates for the cell centroids. We have therefore implicitly projected the true three-dimensional motion of the bacteria swimming in the bulk onto the microscope’s image plane. It is possible to track bacteria in three dimensions, as first demonstrated by Berg in 1971 [20], and several groups have made use of three-dimensional tracking methods to investigate bacterial swimming [24, 37, 62, 74, 193, 201, 210]. The process for obtaining three-dimensional tracks is, however, generally more complex than the method we use and in many cases this leads to a reduced number of tracks available for analysis. In the case of [62, 74, 193, 201], the authors were able to extract tracks for fewer than 100 cells. Brown and Berg [37] report tracking up to 200 cells per day. The method used by Wu *et al.* [210] generated over 600 tracks, but relies upon inducing fluorescence via a plasmid into a subpopulation of bacteria to be tracked. In contrast, the datasets considered in this chapter were generated from approximately 20 minutes of microscope video footage. It is therefore perfectly foreseeable that the number of tracks in the present datasets could be increased by one or even two orders of magnitude. Furthermore, as this study is carried out in part as a proof-of-concept for the new tracking protocol, various aspects of the experimental setup are yet to be optimised. For example, the tracking process remains effective when a $20\times$

magnification microscope objective is used, increasing the number of tracks per unit time by approximately a factor of four [203].

The advantages of the experimental protocol in terms of simplicity and the large amount of data produced are offset by certain caveats that must be carefully considered in order to successfully interrogate the experimental observations. The tracking stage fails in approximately 10% of cases and the resulting tracks should be removed before processing. In addition, as the method relies on tracking all cells within a field of view, a large proportion of the cells may be non-motile for the various reasons discussed in Section 4.1.6, and must also be discarded. Finally, a large proportion of tracks are highly tortuous and we are currently unable to analyse these effectively, though further work on analysis methods could improve this situation. Methods that involve tracking individual cells do not suffer from the same pitfalls, since the experimentalist has control over which individual cells to track and will avoid such cases. Such methods can never be as high-throughput as the protocol presented here, however, as they require a great deal of intensive user involvement. Furthermore, our approach, which involves tracking all detected objects visible in any given frame, is not subject to the biases inherent in an experimental protocol where the experimentalist must select which cells to track. Although we must resort to discarding a large proportion of tracks, we are able to assess in a quantitative manner the contribution of those tracks to the whole population.

Chapter 5

The effect of sampling frequency

In the Chapter 4 we estimated the distribution of framewise run speeds and angle changes from experimental data. These estimated distributions may be affected by the sampling frequency of the digital camera used to acquire the original videos. In this chapter we investigate the effect of varying the sampling frequency on the observed data. Common examples of quantities of interest to experimentalists and modellers include:

1. pdf (how the cell density varies in space and time);
2. MSD;
3. angular distribution (for example, framewise angle change);
4. speed distribution (for example, framewise speed).

Of these, items 1 and 2 are not expected to vary with sampling frequency; in these cases the sampling frequency simply restricts the temporal resolution with which we may observe the underlying continuous behaviour. These quantities are all measured relative to the initial state. In contrast, quantities 3 and 4 are computed by considering the difference between consecutive observations of the system, and therefore vary with the sampling time. Figure 5.1 illustrates this point; although the underlying process has a fixed speed, the *observed* speed will vary when the sampled points do not correspond exactly with the jumps in velocity. Since jumps occur stochastically and sampling occurs

with a fixed interval, this will in general never be the case.

The specific focus of this chapter is the dependency of observed framewise speeds and angle changes on the sampling time. We first derive the analytic behaviour of these quantities in the limit of very long intervals between samples, based on results relating to diffusion. We then investigate the behaviour of the same quantities over a more realistic range of sampling frequencies, using a combination of analysis and simulation. A glossary of the mathematical notation used throughout this chapter is given in Table 5.1. Where possible, we have maintained the same notation as used in previous chapters.

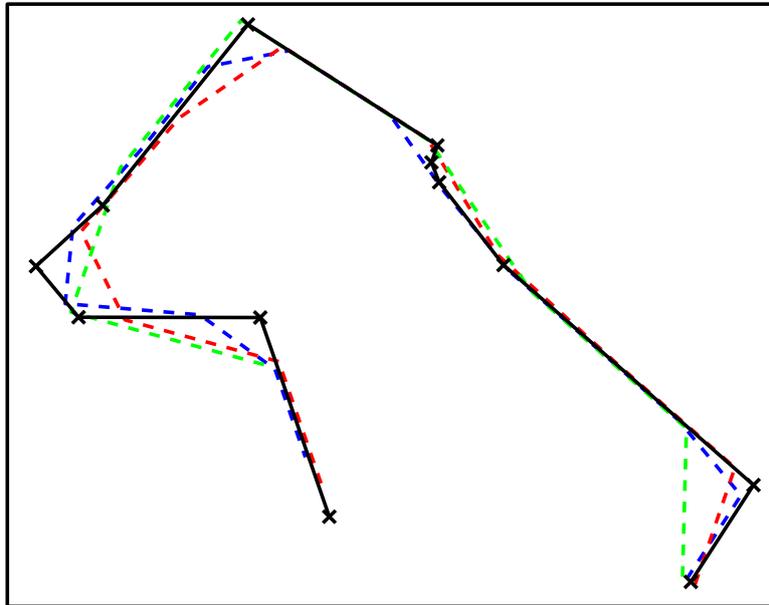


Figure 5.1: *Illustration of the effect of sampling frequency on the observed trajectory. The black solid line shows a simulated run-only, constant speed VJ trajectory, with reorientations at the locations marked with crosses. The dashed lines show the observed trajectory for different sampling intervals: (blue) $\tau/\tau_{10} = 0.7$, (red) $\tau/\tau_{10} = 1.0$, (green) $\tau/\tau_{10} = 1.3$.*

5.1 Background and methods

We assume throughout this chapter that bacteria undergo an unbiased VJ process that is sampled at regular time intervals. Each sample point records the position of the bacterium at that time. Both the run-only and run and stop variants are considered, as

Table 5.1: *Glossary of notation used throughout this chapter.*

Parameter	Description
τ	Sampling time interval.
τ_{10}	Mean time between reorientations ($\tau_{10} = 1/\lambda$).
c	Constant speed of runs.
\bar{c}	Mean observed speed.
σ_c	Standard deviation of observed speeds.
σ_c^*	Standard error of observed speeds.
σ_δ	True angular deviation.
σ_θ	Observed angular deviation.

described in Chapter 2. These models are applicable to a wide range of motile bacteria, including *R. sphaeroides* and *E. coli* (see Chapter 1 for details). The sampling time interval is denoted τ and the mean time between switching events is denoted τ_{10} , as per the notation introduced in Chapter 3. The parameter τ_{10} is a characteristic timescale defined by $\tau_{10} = 1/\lambda$, where λ is the rate of reorientation in the case of the run-only VJ process, and the rate of switching from a run to a stop in the case of the run and stop VJ process. In the run and stop process, the parameter τ_{01} denotes the mean stop duration, as in the previous two chapters. Following [50], we state our results in terms of the non-dimensional ratio τ/τ_{10} .

Codling and Hill [50] investigate a similar problem to the one considered here¹. They use stochastic simulations to investigate the effect of sampling time on three properties of an unbiased VJ process with constant run speed, c , in two dimensions: the observed angular standard deviation, σ_θ ; the mean observed speed, \bar{c} ; and the standard error of observed speeds, σ_c^* . All of these properties are expected to depend upon τ . The approach taken by the authors is strictly empirical. They show that the quantities listed above exhibit various patterns of behaviour, but do not derive the mathematical forms of the dependencies on sampling time. The main findings of Codling and Hill [50] are summarised below.

Angular deviation The authors find that $\sigma_\theta = 0.79\sigma_\delta\sqrt{\tau/\tau_{10}}$, where σ_δ is the true

¹Note that in [50], the symbol $\bar{\tau}$ is used instead of τ_{10} .

angular deviation. This relationship breaks down for $\sigma_\theta > 1.2$ rad.

Mean speed The observed mean speed for a VJ process with constant speed $c = 1$ is found to obey the relationship $\bar{c} = \exp(-0.074\sigma_\delta^2\tau/\tau_{10})$ for $\bar{c} < 0.75$. The explicit dependence on the underlying constant speed c is not stated, however it is straightforward to show that varying c leads to an equivalent multiplicative change in \bar{c} .

Standard error of speeds There is an approximately linear relationship between the standard error of observed speeds, σ_c^* , and the sampling time, τ . However, this only holds for relatively small values of τ/τ_{10} and true angular deviations, σ_δ . In addition, the authors note that there is “no obvious relation” between σ_δ and the slope of the linear fit. Note that σ_c^* quantifies the standard deviation of the mean observed speed, and therefore varies with the number of samples observed.

Standard error of speeds when the underlying process has variable speeds In this case the underlying process is assumed to have a normally distributed speed with mean 1 and standard deviation c_σ . The distribution is truncated, so that only positive speeds are permitted. The authors find that $\sigma_c^* = c_\sigma \exp(-0.069\sigma_\delta^2\tau/\tau_{10})$ for $\sigma_c^* > 0.33$.

Frequent reference in [50] is made to preceding work by Hill and Häder [88], in which the authors assume an underlying biased random walk on a circle to analyse experimental tracks of swimming algae. The authors use ‘rediscretization’ (varying the sampling time) to extrapolate to the true value for the quantities discussed above.

5.1.1 Simulation parameters

Where simulations are used in this chapter, the method used is that described in Chapter 2. The simulated data are from a two-dimensional VJ process with a constant run speed of $1 \mu\text{ms}^{-1}$. All particles are initialised at the origin, with initial velocity in the positive

x -direction. Note that the results are not affected if the initial location and velocity are randomised. Reorientations are drawn from the von Mises distribution with zero mean and a range of values of κ , corresponding to $\sigma_\delta \in \{0.1, 0.2, 0.4, 0.6, 0.8, 1.0\}$. The conversion between κ and σ_δ is found by equating (2.33) and (2.37), which yields

$$\sigma' = \sqrt{-2 \ln \left(\frac{I_1(\kappa)}{I_0(\kappa)} \right)}. \quad (5.1)$$

Table 5.2 lists the parameters used in the various simulations included in this chapter. Where a parameter is omitted from the table, its value is varied and will be stated in the main text of that section. Note that the parameters τ_{01} and c are chosen for simplicity; the results for more realistic parameters would simply be a linear transformation of those shown below. We do not incorporate noise in our simulations, as this would complicate the situation and make it difficult to decouple the effect of noise from the effect of varying the sampling rate. We consider the role of noise in Chapter 6.

5.2 Stationary distributions

When a VJ process is observed with large sampling intervals, many reorientation events occur between consecutive observations and the observed process is diffusive [48, 51, 154]. We may therefore use results relating to Fickian diffusion to describe the asymptotic behaviour of the VJ process in the diffusive limit. This provides a consistency check for the results that follow in Section 5.3.

5.2.1 Observed speeds

Consider a particle undergoing Brownian motion in n dimensions. Between each sample point, the particle moves through a vector denoted $\mathbf{d} = (d_1, \dots, d_n)$. The movements in

Table 5.2: *Simulation parameters used in this chapter. Here T_{total} is the total simulation time and N denotes the number of particles.*

Section	Topic	Simulation parameters	Figures
5.2	Long sampling time limit.	$\tau_{10} = 1 \text{ s}$, $T_{\text{total}} = 5000 \text{ s}$, $N = 1000$.	5.2, 5.3
5.3	Validation of our methodology by comparison with [50] .	$\tau_{10} = 1 \text{ s}$, $T_{\text{total}} = 200 \text{ s}$, $N = 1000$.	5.4
5.3	Simulation study of dynamic distributions.	$\tau_{10} = 1 \text{ s}$, $T_{\text{total}} = 200 \text{ s}$, $N = 1000$.	5.6, 5.7
5.4	Analytic study of dynamic distributions (forward problem).	$\tau_{10} = 5 \text{ s}$, $\tau = 1 \text{ s}$, $T_{\text{total}} = 200 \text{ s}$.	5.14, 5.15
5.4	Analytic study of dynamic distributions (inverse problem).	$\tau_{10} = 20 \text{ s}$, $\tau = 1 \text{ s}$, $T_{\text{total}} = 200 \text{ s}$.	5.17

each of the dimensions are IID according to a normal distribution,

$$d_i \sim \text{N}(0, \sqrt{2D\tau}), \quad (5.2)$$

where D is the macroscopic diffusion coefficient as before and $i \in \{1, \dots, n\}$. In Section 2.3 we discussed the results of [146], which show that the slope of the MSD of a VJ process is linear in the limit $t \rightarrow \infty$. Since the slope of the MSD in each dimension is equal to $2D$, we have $D = c^2\tau_{10}/(1 - \rho_1)$, where ρ_1 denotes the mean length of the angle change distribution (equivalent to the mean cosine for a symmetric distribution). The absolute displacement between each sample point is equal to the Euclidean norm of the vector \mathbf{d} . The distribution of the random variable $\|\mathbf{d}\|$ is closely related to the χ distribution, which describes the distribution of the square root of the sum of squared

standard normal variates. The variable Z , defined by

$$Z = \sqrt{\sum_{i=1}^k X_i^2}, \quad (5.3)$$

where X_1, \dots, X_k are standard normal variates, follows a χ distribution with parameter k . The χ distribution has mean

$$\sqrt{2} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)}$$

and variance $\sigma^2 = k - \mu^2$.

The variates being summed in our case, d_i , are not standard normal as they have a common standard deviation that depends upon the D and τ , as given in equation (5.2). We therefore define a new variable, $\mathbf{d}' = \mathbf{d}/\sqrt{2D\tau}$. The random variable $\|\mathbf{d}'\|$ now follows a χ distribution. In two dimensions, the mean of this distribution is $\sqrt{2}\Gamma(3/2)$ and the variance is $2(1 - \Gamma(3/2)^2)$. The mean of $\|\mathbf{d}\|$ is therefore given by $2\sqrt{D\tau}\Gamma(3/2)$ and the standard deviation is given by $2\sqrt{D\tau(1 - \Gamma(3/2)^2)}$. The *observed speed* of the diffusing particle between two sample points is given by $c(\tau) = \|\mathbf{d}\|/\tau$. Therefore the mean observed speed, denoted \bar{c} , is given by

$$\bar{c} = \sqrt{\frac{2}{\tau/\tau_{10}(1 - \rho_1)}} \Gamma(3/2) \quad (5.4)$$

Taking logarithms, we obtain

$$-\log(\bar{c}) = \frac{1}{2} [\log(\tau/\tau_{10}) + \log(1 - \rho_1) - \log(2)] - \log(\Gamma(3/2)). \quad (5.5)$$

A plot of $-\log(\bar{c})$ against $\log(\tau/\tau_{10})$ should therefore be asymptotically linear, with slope $1/2$.

We evaluate the standard deviation of observed speeds, denoted σ_c , by a similar process, giving

$$\sigma_c = \sqrt{\frac{2}{\tau/(\tau_{10}(1-\rho_1))}} (1 - \Gamma(3/2)^2). \quad (5.6)$$

Taking logarithms we obtain

$$-\log(\sigma_c) = \frac{1}{2} [\log(\tau/\tau_{10}) + \log(1-\rho_1) - \log(2)] - \log(1 - (\Gamma(3/2))^2). \quad (5.7)$$

Similar to the case for the mean observed speed, a plot of $-\log(\sigma_c)$ against $\log(\tau/\tau_{10})$ is asymptotically linear with slope 1/2.

Equations (5.5) and (5.7) are verified in Figure 5.2, using simulated data with the parameters listed in Section 5.1.1. In both cases, dashed lines indicate the asymptotic behaviour. The simulated dataset with the greatest true angular variance relaxes most rapidly to the diffusion limit (\circ). This result is intuitive; trajectories with small angular variance have a very long persistence time, as most reorientations incur only minor changes in direction, therefore such trajectories only appear diffusive at long timescales, where the persistence is lost. Conversely, a trajectory with high angular variance loses persistence rapidly, as reorientations are often drastic.

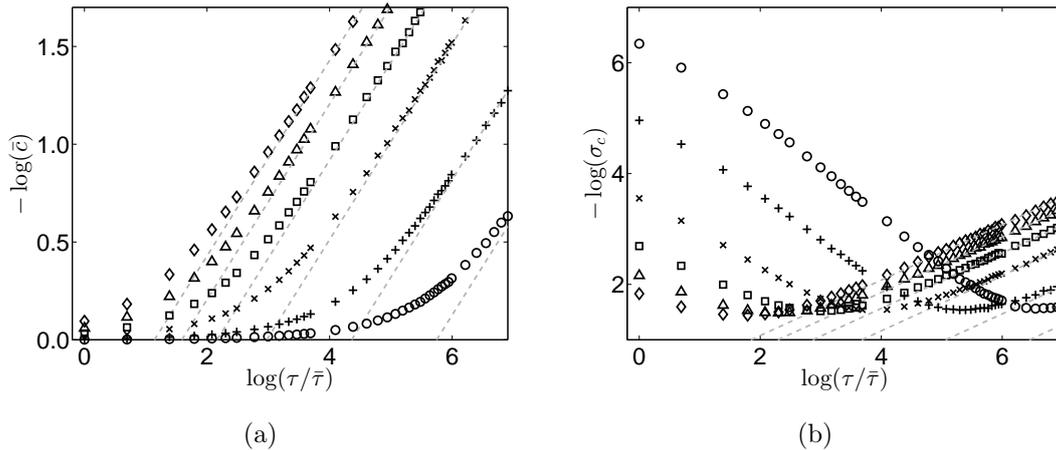


Figure 5.2: (a) Log-log plot of the observed mean speed against sampling time. (b) Log-log plot of the standard deviation of observed speeds against sampling time. The data are computed for a 2D run-only velocity jump process with $\sigma_\delta = 0.2(\Delta)$, $\sigma_\delta = 0.4(\square)$, $\sigma_\delta = 0.6(\times)$, $\sigma_\delta = 0.8(+)$ and $\sigma_\delta = 1.0(\circ)$. Dashed lines have slope 1/2.

5.2.2 Observed angle changes

For a particle undergoing Brownian motion, the observed angle changes between consecutive pairs of sampling points are distributed uniformly, as the solution to the diffusion equation in an infinite domain is radially symmetric [72]. As discussed in Section 2.4, the wrapped uniform distribution has a standard deviation of $\sqrt{2}$ if we use the definition given in equation (2.32). Note that the alternative definition, given in equation (2.33), is undefined for the same distribution. This result is verified in Figure 5.3. As for the mean observed speed, the greater the true angular variance, the more quickly the observed process relaxes to the diffusion limit.

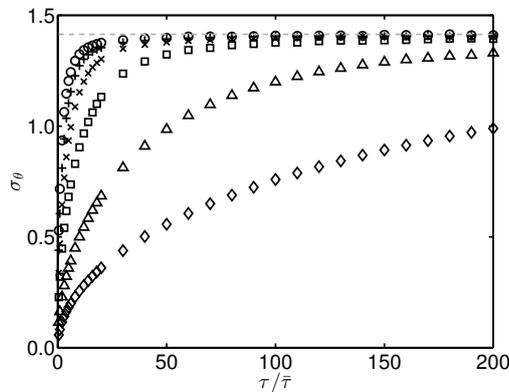


Figure 5.3: The observed angular variance, computed using the definition in equation (2.32), against sampling time for a run-only VJ process. Symbols indicate the values of σ_δ , as in Figure 5.2. The dashed line indicates the asymptotic value of $\sigma_\theta = \sqrt{2}$.

5.3 Simulation study of dynamic distributions

In the previous section we considered the long sampling time limit of the observed VJ process. The behaviour at shorter sampling times is, however, of greater practical interest. We now consider the effect of shorter sampling times, at which the observed process is not purely diffusive. As reviewed above, Codling and Hill [50] investigated the effect of variations in sampling time on the moments (mean and standard error) of the observed speed and angle using simulated data. In this section, we use a similar computational approach to investigate the observed *pdfs* of speeds and angle changes. The importance

of the effect of sampling time on the observed pdf of these quantities was demonstrated in Chapter 4: summary statistics such as the mean and standard deviation, whilst undoubtedly useful information, may be insufficient to distinguish the motion of different species of bacteria, for example, whereas the pdf contains additional information. Furthermore, we extend Codling and Hill’s study by considering a run and stop VJ process.

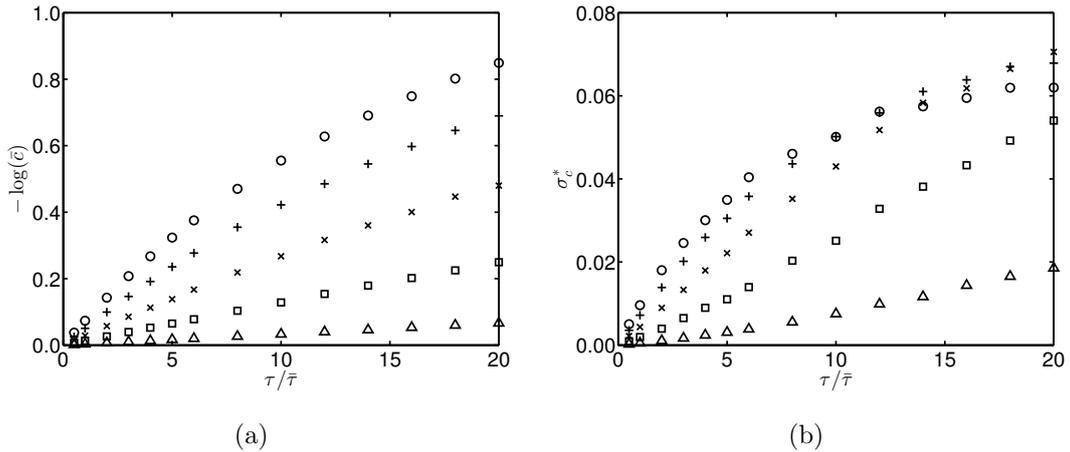


Figure 5.4: (a) Plot of the negative logarithm of the observed mean speed against sampling time. (b) Plot of the standard error of observed speeds against sampling time. The data are computed for a two-dimensional run-only VJ process with $\sigma_\delta = 0.2(\Delta)$, $\sigma_\delta = 0.4(\square)$, $\sigma_\delta = 0.6(\times)$, $\sigma_\delta = 0.8(+)$ and $\sigma_\delta = 1.0(\circ)$.

Before commencing, we verify that our simulation approach is able to recreate the results of Codling and Hill [50]. Figures 5.4(a) and 5.4(b) contain plots that are indistinguishable from Figures 2 and 3(a) in their study. A comparison of our results with those of Codling and Hill [50] is therefore permissible.

5.3.1 Run-only VJ process

A plot of the variation of particle density with position and time for two different values of the angular variance, σ_δ , is shown in Figure 5.5. The greater value of the angular variance leads to rapid spreading out of the particles from the origin in all directions compared with the lesser value, where directional persistence is evident for longer.

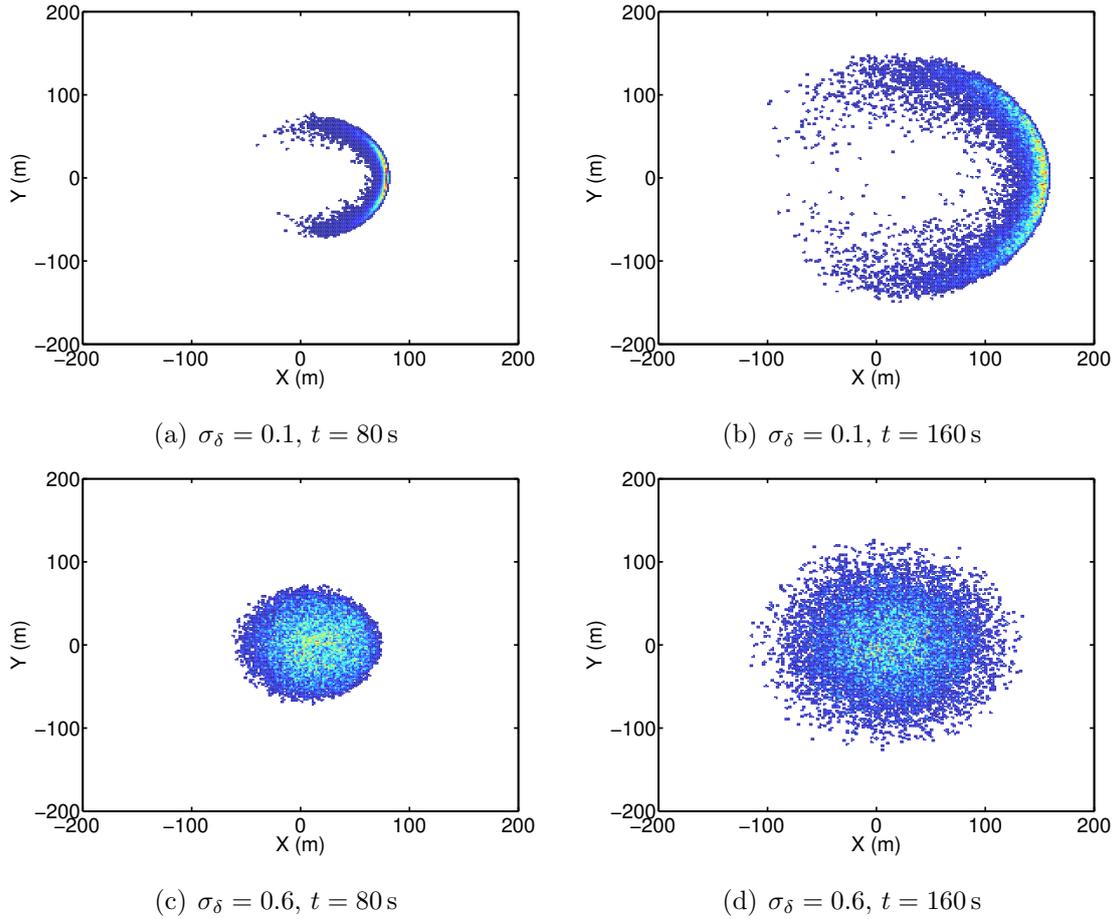


Figure 5.5: *The density of particles from simulated run-only data. Colour indicates density (arbitrary scale) from blue (low density) to red (high density). All particles are initially located at the origin and travelling in the positive x -direction.*

Observed speed distributions

The observed speed distribution for different sampling times is shown in Figure 5.6 for two different values of σ_δ . There is a greater spread of observed speeds in the case $\sigma_\delta = 0.6$ compared with the case $\sigma_\delta = 0.1$, because each reorientation event tends to result in a larger change of direction. As τ is increased, the observed speed distribution becomes broader. This occurs because each sampling interval has a greater probability of containing one or more reorientation events as the interval size increases. When a reorientation event occurs in a sampling interval, the observed speed is always lower than the true speed within that sampling interval.

The results in Figure 5.6 are in agreement with the findings of Codling and Hill [50]

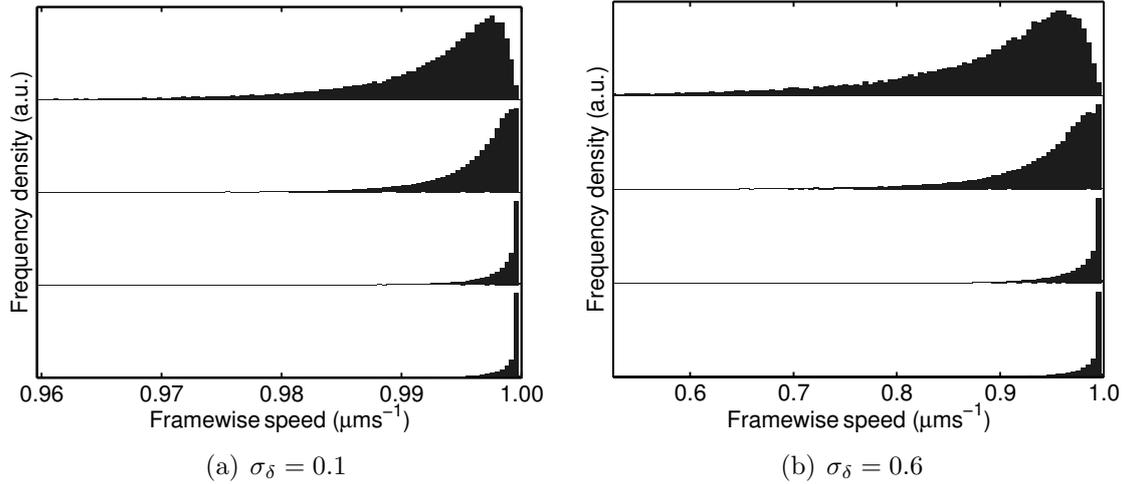


Figure 5.6: Observed speed distributions for different sampling times, computed with two different values of σ_δ . Stacked distributions correspond to different values of τ/τ_{10} : (bottom to top) 1, 2, 5, 10. The y -axis scale varies between stacked plots.

(see Figure 5.4). As we increase τ , the distribution of observed speeds is skewed towards lower speeds. This decreases the mean observed speed (Figure 5.4(a)), and increases the standard deviation of observed speeds (Figure 5.4(b)).

In the rapid sampling limit, $\tau/\tau_{10} \rightarrow 0$, the distribution of observed speeds converges to the true underlying distribution, as particles do not undergo reorientations between sampling intervals. For the constant speed process considered here, the distribution of observed speeds tends towards a delta function at the true underlying speed. This result is not restricted to a constant speed process; when there is an underlying distribution of speeds (for example the simulations used in Chapter 3), the observed speed distribution converges to the true underlying distribution (data not shown).

Observed angle change distribution

Figure 5.7 shows the variation of the distribution of observed angle changes with sampling time for two different values of σ_δ . Overlaid in red is the true underlying von Mises distribution used for reorientation. Note that this distribution is constant in each figure; the change in height is due to a rescaling of the y -axis. These results are in agreement with the study by Codling and Hill [50], as the broadening of the distribution of observed

angles results in an increase in σ_θ . In the short sampling limit, the distribution of observed angle changes tends towards a delta function at the origin, since, in this limit, all of the sampling intervals measure zero angle change because they contain no reorientation events. Thus the angular variance tends towards zero in this limit, as predicted by Codling and Hill [50].

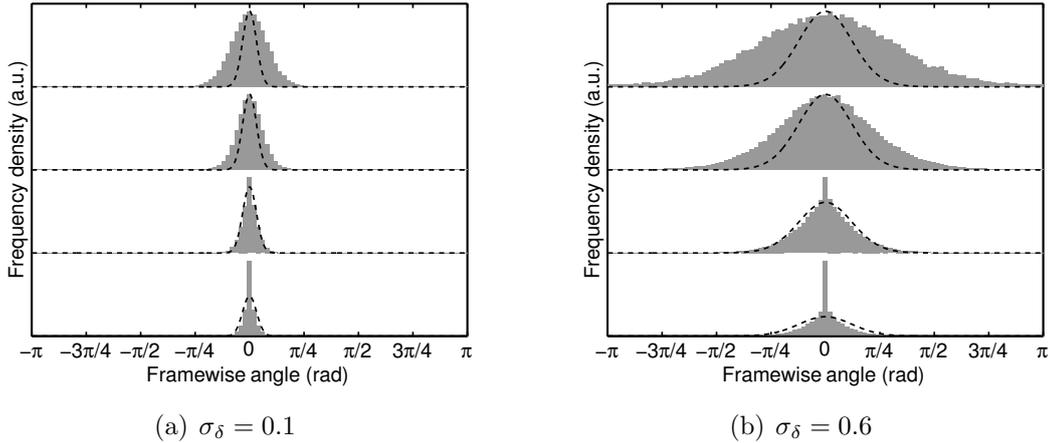


Figure 5.7: *Distributions of observed angle changes for different sampling times, computed with two different values of σ_δ . Blue bars show the binned stochastic data and the red dashed line shows the von Mises distribution used to simulate the process. Stacked distributions correspond to different values of τ/τ_{10} : (bottom to top) 1, 2, 5, 10. The y-axis scale varies between stacked plots.*

In contrast with the distribution of observed speeds, where the true underlying distribution is obtained in the limit $\tau/\tau_{10} \rightarrow 0$, at no sampling time does the distribution of observed angle changes match the true underlying angle change distribution. This is due to the nature of the fixed interval sampling; in order to regain the original distribution we would need to sample each particle at every reorientation event. However, in the absence of knowing when these jumps occur, we instead sample at regular intervals and therefore sample both across and in-between reorientations. This accounts for the large observed density at $\theta = 0$ for both values of σ_δ when $\tau/\tau_{10} = 1$; although the sampling interval matches the time between velocity jumps on average, there are many instances in which two consecutive sample points do not have a jump between them and thus an angle change of zero is recorded.

The above discussion raises the question of how we can infer the underlying turning kernel from the inferred data. We return to this problem in Section 5.4.

5.3.2 Run and stop VJ process

Figure 5.8 illustrates the evolution of particle density over time for simulated data of particles undergoing a run and stop VJ process with two values of σ_δ . Comparing this figure with Figure 5.5, which shows the equivalent plots for run-only data, it is apparent that the inclusion of a stopping phase retards the outward spread of particles. The particle density is also considerably more inhomogeneous within a circular region in the run and stop case than the equivalent run-only case, as particles have, on average, undergone fewer reorientations, thus the increased particle density at the travelling front is maintained for longer (compare Figures 5.5(b) and 5.8(b)).

Observed speed distributions

The simulated distribution of observed speeds for two values of σ_δ and τ_{01} are shown in Figure 5.9. As for the run-only case, in the limit of rapid sampling, the distribution of observed speeds relaxes to the true underlying distribution, this time with an additional peak at zero due to stoppers, shown in red. The relative densities of the stoppers and runners are determined by the ratio of the parameters τ_{01} and τ_{10} . The observed speed distributions have broader distributions than in the run-only case (compare with Figure 5.6), with larger values of τ_{01} shifting the distribution mean towards lower speeds, as expected. Furthermore, the observed distribution of run speeds is not smooth, as was the case for the run-only process. Instead, a sharp peak exists at a framewise speed of $1 \mu\text{ms}^{-1}$, corresponding to running particles that have not undergone any reorientations in a sampling interval. At sampling times $\tau/\tau_{10} \gtrsim 5$, this peak is no longer apparent.

In Figure 5.10, we plot \bar{c} and σ_c against sampling time for different values of τ_{01} . Note

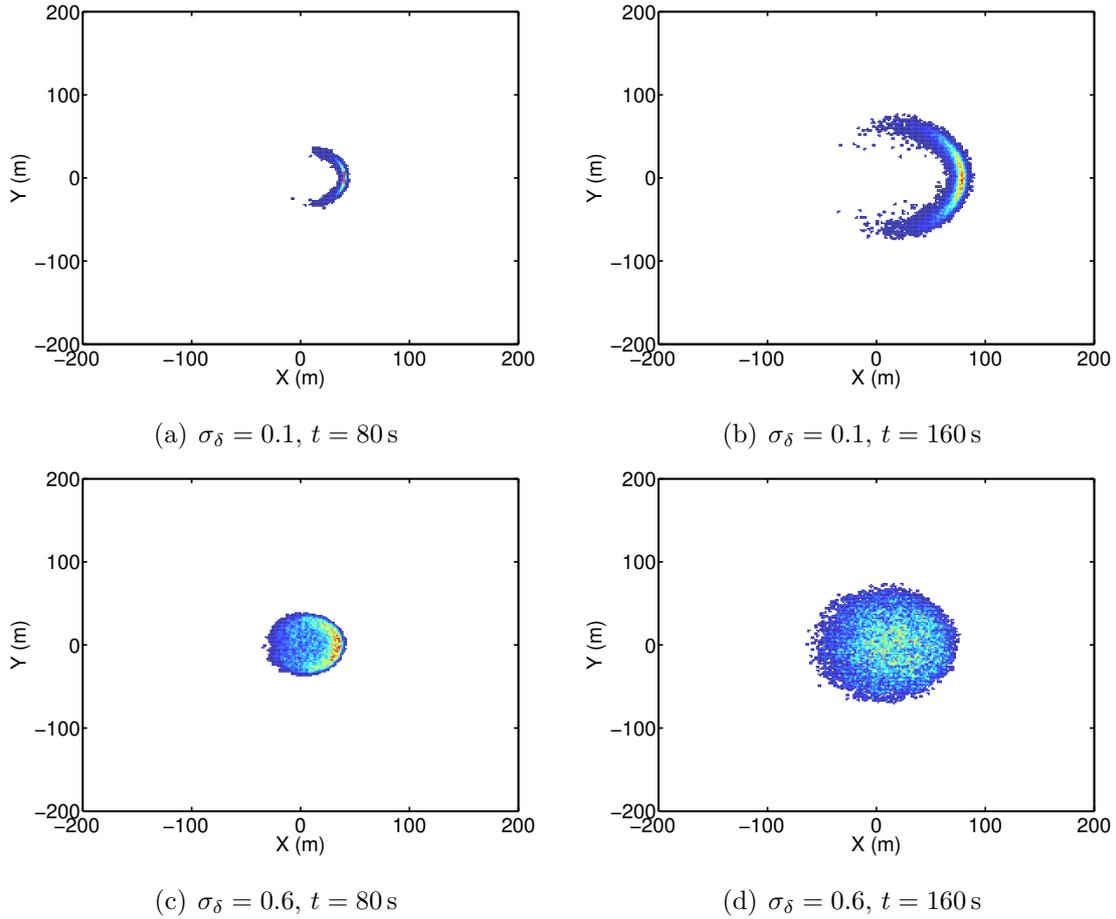


Figure 5.8: Histograms of the density of particles undergoing a run and stop VJ process with $\tau_{01} = \tau_{10} = 1$ s. Colour indicates density (arbitrary scale) from blue (low density) to red (high density).

that we do not include transitions that have an observed speed of exactly zero in the calculation of these quantities, meaning that the density shown in red in Figure 5.9 is ignored. In all cases, \bar{c} decays approximately exponentially with the sampling time, with the rate of decay proportional to the value of τ_{01} . Although the run-only case appears to exhibit linear decay, we showed in Section 5.2 that \bar{c} decays exponentially in the limit of long sampling times, so the appearance is due to the fact that we are considering sampling times shorter than this limit here. The variation of σ_c is more complex, monotonically increasing in the case of a run-only process and monotonically decreasing for all non-zero values of τ_{01} over the range of sampling times considered. The explanation is facilitated by comparison of Figures 5.6 and 5.9. In the run-only case, the distribution of observed framewise speeds is tightly clustered around the true underlying value of c at low τ . This

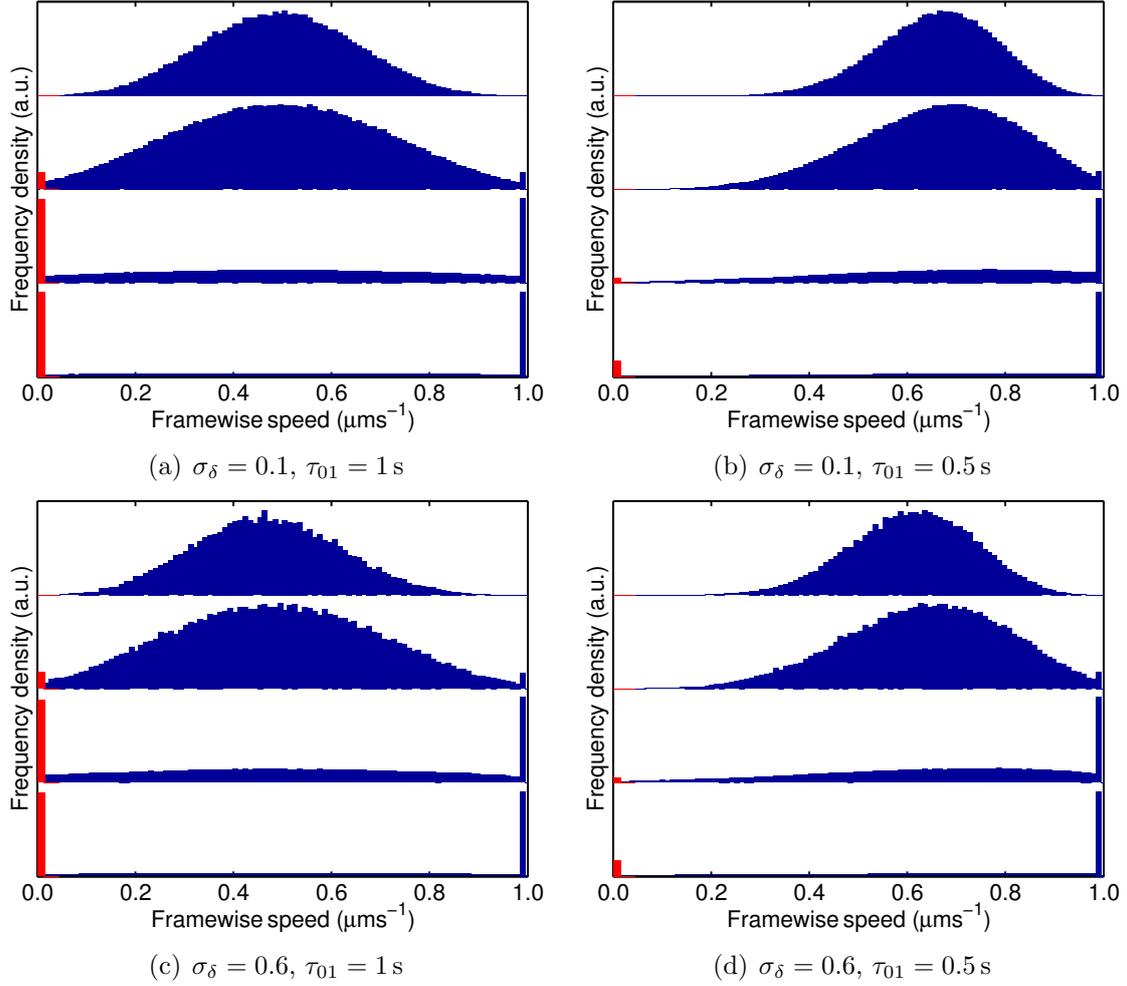


Figure 5.9: Observed speed distributions for different sampling times, computed with different values of σ_δ and τ_{01} . Blue bar indicates the speed distribution of runners. Red bar indicates relative proportion of stopped cells (i.e. cells with speed equal to zero). Stacked distributions correspond to different values of τ/τ_{10} : (bottom to top) 1, 2, 5, 10. The y-axis scale varies between stacked plots.

distribution broadens as τ increases. In contrast, there is a wide spread of speeds at low τ when $\tau_{01} \neq 0$, with most clustered around the true value of c , but a significant proportion distributed across the full range. The distribution effectively narrows as τ is increased, as the density of speeds at the extremes decreases.

Observed angle change distributions

The computed distribution of observed angle changes is shown for $\sigma_\delta = 0.6$ and $\tau_{01} = 1$ s in Figure 5.11. The distributions shown do not include angle changes for particles that are stopped for the entirety of a sampling interval, as the angle change in these cases is

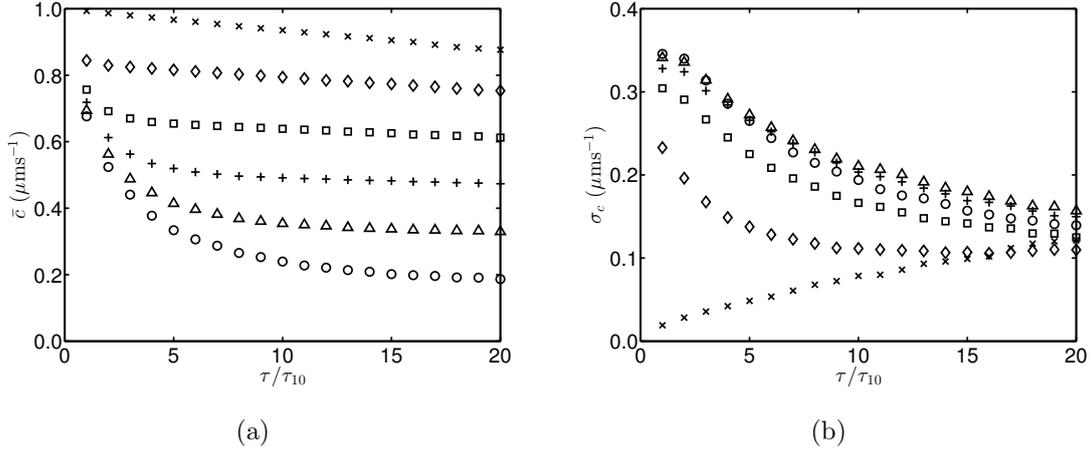


Figure 5.10: Variation of the mean (a) and standard deviation (b) of observed framewise speeds with τ/τ_{10} for a run and stop VJ process with $\sigma_\delta = 0.6$ and $\tau_{01} = 0$ s (\times , equivalent to a run-only process), $\tau_{01} = 0.2$ s (\diamond), $\tau_{01} = 0.5$ s (\square), $\tau_{01} = 1$ s ($+$), $\tau_{01} = 2$ s (\triangle) and $\tau_{01} = 5$ s (\circ).

not well defined. Comparing this figure with Figure 5.7, it is evident that the inclusion of a stopping phase causes the distribution of observed angle changes to become more concentrated about the origin. The shape of the distribution is, however qualitatively similar in both cases.

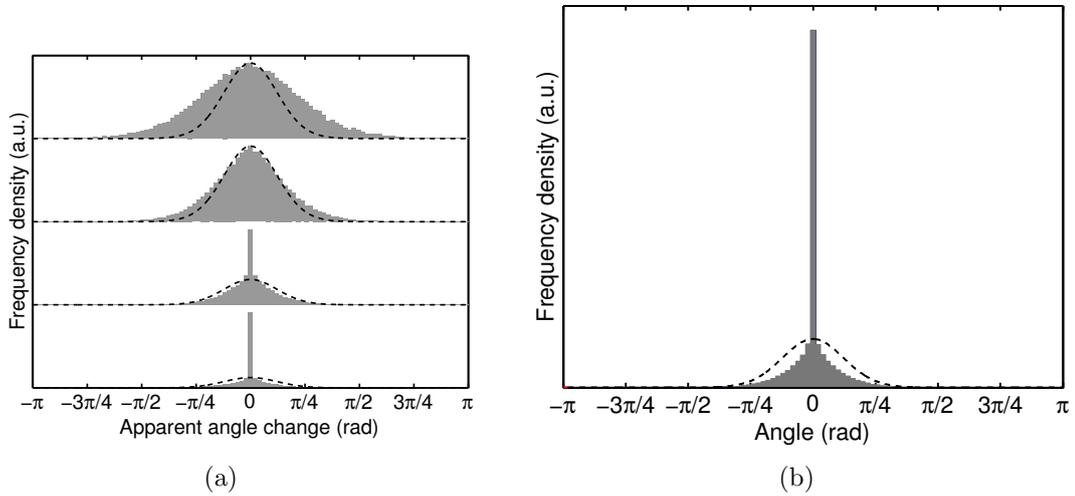
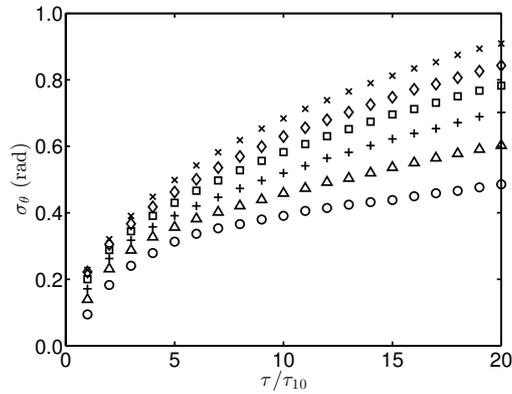


Figure 5.11: (a) Observed angle change distributions for different sampling times, computed with $\sigma_\delta = 0.6$ and $\tau_{01} = 1$ s for a run and stop VJ process. Grey bars indicate the distribution for running particles, dashed line shows the von Mises distribution used to simulate the process. Stacked distributions correspond to different values of τ/τ_{10} : (bottom to top) 1, 2, 5, 10. Note that the y-axis scale varies between plots. (b) Enlarged plot of the distribution when $\tau/\tau_{10} = 1$.

Figure 5.12 shows the dependence of σ_θ on sampling frequency. As the figure shows, this quantity increases monotonically with τ/τ_{10} over the range plotted. Further numerical

simulations (not shown) indicated that the angular deviation has an asymptotic value of $\sigma_\theta = \sqrt{2}$ as $\tau/\tau_{10} \rightarrow \infty$ in all cases, as predicted in Section 5.2. The value of τ_{01} affects how rapidly this asymptote is reached. As the mean duration of stopping phases increases, the angular deviation decreases for a given value of τ/τ_{10} . The underlying cause of this result is not obvious, but may be as follows. A larger value of τ_{01} means that, on average, fewer reorientation events will occur over a sample interval. Since the von Mises distribution is concentrated about the origin, small angle reorientation events occur with higher probability than more extreme angle changes. As a result, the distribution of observed angle changes will have greater density at low values, and the standard deviation of observed angle changes will be lower.



(a)

Figure 5.12: Standard deviation of observed framewise angle changes against τ/τ_{10} for a run and stop VJ process with $\sigma_\delta = 0.6$ and $\tau_{01} = 0$ s (\times , equivalent to a run-only process), $\tau_{01} = 0.2$ s (\diamond), $\tau_{01} = 0.5$ s (\square), $\tau_{01} = 1$ s ($+$), $\tau_{01} = 2$ s (\triangle) and $\tau_{01} = 5$ s (\circ).

5.4 Analytic study of dynamic distributions

Whilst computational simulations are useful for assessing the effect of sampling rate on the observed VJ process, they are time-consuming, as many particles must be simulated to obtain sufficiently smooth distributions. Furthermore, a new simulation is required for each new set of parameters, making it hard to draw general conclusions about the interplay between the parameters τ/τ_{10} and σ_δ . We now seek an analytic description of

the underlying VJ process, for application to the sampling rate problem.

In the following, we consider a run-only, constant speed VJ process. As before, the constant speed is denoted c . In contrast with the previous section, we require no assumptions on the underlying reorientation kernel; we instead denote this pdf $f_{\Phi}(\phi)$. We make the simplifying assumption that the sampling interval, τ , is sufficiently small relative to the mean time between reorientations, τ_{10} , that the probability of two or more reorientations occurring between consecutive sample points is negligible. Using the fact that reorientations in the underlying VJ process occur as a Poisson process, this probability is given by

$$\mathbb{P}(N \geq 2; \tau, \tau_{10}) = 1 - e^{-\tau/\tau_{10}}(1 + \tau/\tau_{10}), \quad (5.8)$$

where the random variable N denotes the number of reorientations in a time interval of duration τ . By way of a numerical example, choosing $\tau/\tau_{10} = 0.1$ gives $\mathbb{P}(N \geq 2; \tau, \tau_{10}) \approx 0.005$, so two or more reorientations are expected to occur within one sampling interval once per 500 intervals.

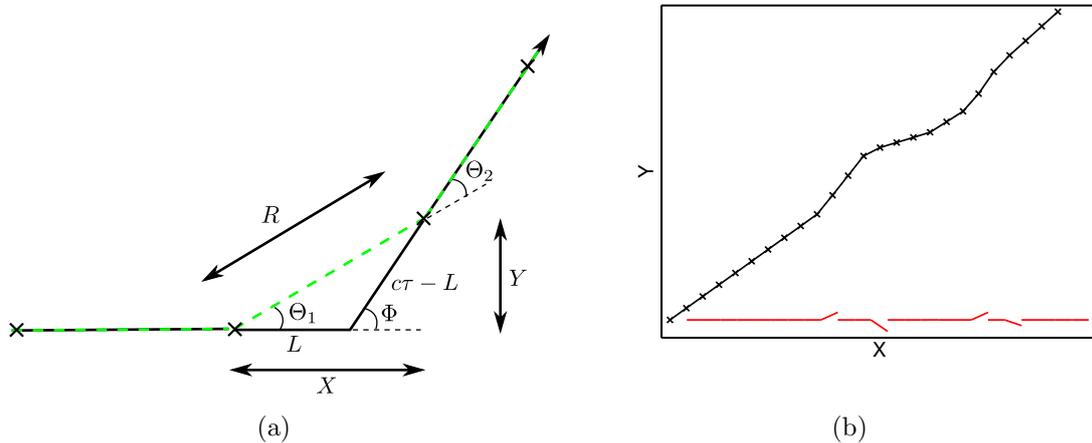


Figure 5.13: (a) Illustration of a portion of a VJ trajectory containing a single reorientation phase. Black lines show the true path and the crosses joined by the dashed green line indicate the observed path, after sampling. See main text for an explanation of the various labels. (b) A simulated track (black lines and crosses) and the resulting canonical transitions (red lines).

Figure 5.13(a) illustrates the problem considered here with an example trajectory. The

true underlying trajectory consists of two straight line sections (shown in black) separated by an instantaneous reorientation of angle Φ , given relative to the previous direction of travel. In contrast, the *observed* path (green dashed line) appears to contain two reorientation events, with angles Θ_1 and Θ_2 . The observed framewise displacement is denoted R . The difference between the observed and true paths is due to the discretisation of the continuous underlying trajectory. The length L is the distance between the last sampling point and the reorientation event. Since we consider a constant speed process, the distance travelled by a particle between over the course of a sampling interval is given by $c\tau$. The remaining distance between the reorientation event and the next sampling point is therefore given by $c\tau - L$.

Two alternative representations of the trajectory are the pairs of random displacements X and Y , where the axes are defined parallel and perpendicular to the previous direction of travel, respectively, and the polar representation R and Θ_i , where $i \in \{1, 2\}$. The mapping from the (L, Φ) plane to either of these representations is bijective. Since both alternative representations are defined relative to the previous direction of travel, we must rotate the transitions \mathbf{v}_t of tracks (either real or simulated) so that the previous direction of motion is effectively the positive x -direction, as in Figure 5.13(a). These rotated transition vectors are denoted *canonical transitions*. An illustration of a simulated track and its canonical transitions is shown in Figure 5.13(b). A complication with computing the canonical transitions arises immediately after a reorientation event. Referring to the final transition shown in Figure 5.13(a), it is evident that the *observed* prior direction of travel (with angle Θ_1) is not equal to the *actual* prior direction of travel (with angle Φ). It is therefore necessary to use the value of Φ , and not Θ_1 , to compute the final canonical transition.

We first solve the *forward problem*, which is the process of computing the joint pdf of R and Θ_1 , denoted $f_{R, \Theta_1}(r, \theta_1)$, assuming that the functional form of the angular pdf

$f_\Phi(\phi)$ and the value of c are both known. There is no need to specify the distribution of the random variable L since this is prescribed by the definition of the VJ process, as discussed below. Providing the resulting expression is integrable with respect to both of the independent variables, we are then able to solve for the marginal pdfs, $f_R(r)$ and $f_{\Theta_1}(\theta_1)$. The statistical properties of the VJ process are time-reversal invariant, meaning that the pdfs governing Θ_1 and Θ_2 must be identical. The forward problem therefore allows us to estimate the observed distribution of framewise speeds and angles for a given sampling time, τ , given knowledge of the true underlying distribution of angle changes.

In order to tackle this problem, we make two consecutive transformations by change of variables. We first transform to the (X, Y) plane, which requires the following relationships:

$$X = L + (c\tau - L) \cos \Phi, \quad (5.9)$$

$$Y = (c\tau - L) \sin \Phi. \quad (5.10)$$

The joint pdf of X and Y is then given by

$$f_{X,Y}(x, y) = \begin{cases} |\det J_{l,\phi}| f_{L,\Phi}(l(x, y), \phi(x, y)), & \text{if } (x, y) \in \mathcal{C} \\ 0, & \text{otherwise,} \end{cases}, \quad (5.11)$$

where the domain in which the transformation is valid is given by

$$\mathcal{C} = \{(x, y) : x^2 + y^2 \leq (c\tau)^2\}, \quad (5.12)$$

and the Jacobian matrix associated with the transformation is given by

$$J_{l,\phi} = \begin{pmatrix} \frac{\partial l}{\partial x} & \frac{\partial l}{\partial y} \\ \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} \end{pmatrix}.$$

Manipulating equations (5.9) and (5.10), we obtain

$$L = \frac{(c\tau)^2 - X^2 - Y^2}{2(c\tau - X)}, \quad (5.13)$$

$$\Phi = \arccos\left(\frac{Y^2 - (c\tau - X)^2}{Y^2 + (c\tau - X)^2}\right). \quad (5.14)$$

Differentiating equations (5.13) and (5.14) and evaluating the determinant yields

$$|\det J_{l,\phi}| = \frac{1}{c\tau - x}. \quad (5.15)$$

We now require an expression for the joint pdf $f_{L,\Phi}(l, \phi)$. Since the displacement and angle change are independent, we have $f_{L,\Phi}(l, \phi) = f_L(l)f_\Phi(\phi)$. Recall that stops occur as a Poisson process. It can be shown that for a Poisson process, the distribution of events within an interval, conditional on the number of events, is uniform [80]². The distribution of the displacement L is therefore given by

$$f_L(l) = \frac{1}{c\tau}. \quad (5.16)$$

Substituting into equation (5.11), we obtain

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{c\tau(c\tau - x)} f_\Phi\left(\arccos\left(\frac{y^2 - (c\tau - x)^2}{y^2 + (c\tau - x)^2}\right)\right) & (x, y) \in \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases} \quad (5.17)$$

We now make a second transformation to obtain the pdf $f_{R,\Theta_1}(r, \theta_1)$,

$$f_{R,\Theta_1}(r, \theta_1) = \begin{cases} |\det J_{x,y}| f_{X,Y}(x(r, \theta_1), y(r, \theta_1)), & \text{if } (r, \theta_1) \in \mathcal{C} \\ 0, & \text{otherwise,} \end{cases} \quad (5.18)$$

²A proof is reproduced in Appendix A.5.

where

$$J_{x,y} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta_1} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta_1} \end{pmatrix},$$

and the domain \mathcal{C} is reinterpreted as

$$\mathcal{C} = \{(r, \theta_1) : 0 \leq r \leq c\tau, \theta_1 \in [-\pi, \pi]\}. \quad (5.19)$$

Evaluating the determinant of the Jacobian matrix gives $\det J_{x,y} = r$. Combining this result with equations (5.9), (5.10), (5.17) and (5.18), we obtain

$$f_{R,\Theta_1}(r, \theta_1) = \begin{cases} \frac{r}{c\tau(c\tau - r \cos \theta_1)} f_{\Phi} \left(\arccos \left[\frac{r^2 \sin^2 \theta_1 - (c\tau - r \cos \theta_1)^2}{r^2 \sin^2 \theta_1 + (c\tau - r \cos \theta_1)^2} \right] \right), & (r, \theta_1) \in \mathcal{C}, \\ 0, & \text{otherwise.} \end{cases} \quad (5.20)$$

We may use equation (5.20) to compute the marginal pdfs governing the displacement R and the angle Θ_1 , by integrating the joint pdf:

$$f_R(r) = \int_{-\pi}^{\pi} f_{R,\Theta_1}(r, \theta_1) d\theta_1, \quad (5.21)$$

$$f_{\Theta_1}(\theta_1) = \int_0^{\infty} f_{R,\Theta_1}(r, \theta_1) dr. \quad (5.22)$$

We re-emphasise that Θ_1 and Θ_2 have the same marginal pdfs by symmetry, so that $f_{\Theta_2}(\theta_2)$ is also given by equation (5.22).

In order to verify the forward problem by comparing with a stochastic simulation, we must specify a turning angle pdf, $f_{\Phi}(\phi)$. As for previous sections in this chapter, we use the von Mises distribution with zero mean and angular deviation $\sigma_{\delta} = 0.6$. The remaining parameters are $\tau_{10} = 1 \text{ s}$, $\tau = 0.2 \text{ s}$ and $c = 1 \mu\text{ms}^{-1}$. Upon substituting the functional form of the von Mises distribution (given by equation (2.36)) into (5.20), the

integrals in (5.21) and (5.22) become too complex to perform analytically. We therefore use adaptive Simpson numerical integration [161], implemented in the Matlab function `quad`, to evaluate the pdf. The marginal pdf of framewise speeds requires a linear scaling of the pdf of framewise displacements, $f_R(r)$, by a factor of $c\tau$.

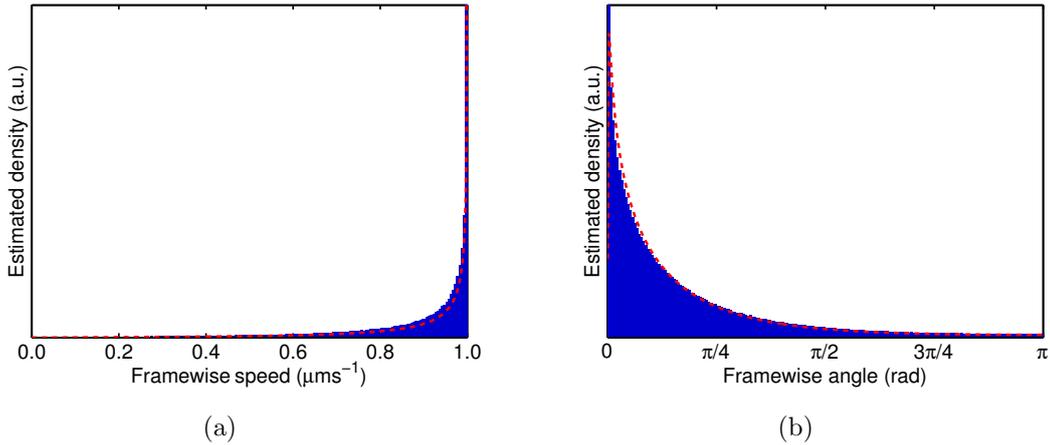


Figure 5.14: *Distribution of framewise speeds (a) and angle changes (b). In both cases the results from a stochastic simulation are shown in blue and the numerically computed analytic formulation plotted in red. Only positive angle changes are shown for clarity.*

Figure 5.14 shows the observed pdf of framewise speeds and angle changes, computed with the stochastic simulation algorithm, overlaid with the result of performing numerical integration of equation (5.20) as discussed above. The stochastic simulation results are first filtered to remove transitions in which no reorientations occur, as these result in a large number of framewise angle changes exactly equal to zero, and framewise speeds exactly equal to c . This is achieved by discarding transitions with a framewise angle change whose magnitude is less than a defined numerical tolerance. The agreement is good, and improves as the accuracy of the numerical integration is increased (data not shown). This comparison shows that we are able to describe analytically the effect of sampling a continuous VJ process on the observed data. By way of example, Figure 5.15 shows how the observed pdfs are affected by the underlying angular deviation, σ_δ .

We now turn our attention to the *inverse problem*, namely inferring the angle change

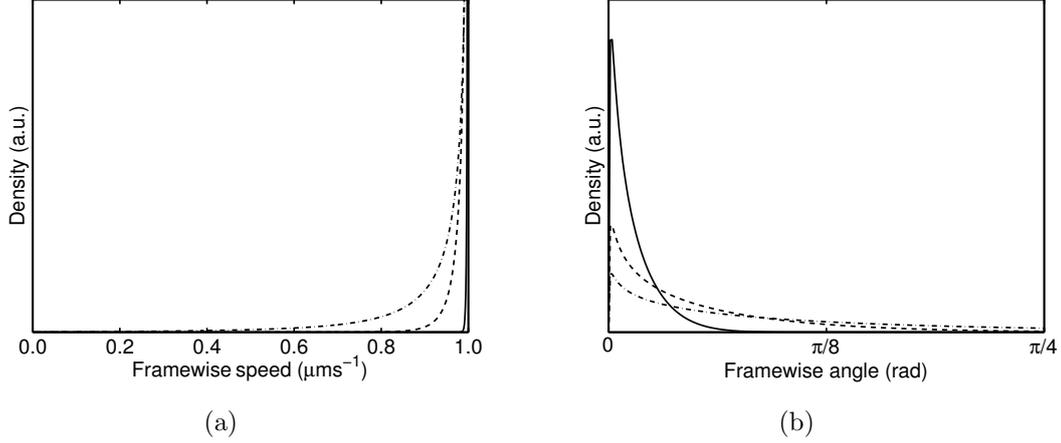


Figure 5.15: Distribution of framewise speeds (a) and angle changes (b) for $\sigma_\delta = 0.1$ (solid line), $\sigma_\delta = 0.6$ (dashed line), and $\sigma_\delta = 1.0$ (dashed-dotted line). Only positive angle changes are shown for clarity.

distribution $f_\Phi(\phi)$ from observed data. We assume that the value of c is known, as this may be estimated by alternative methods. We also assume that $\tau \ll \tau_{10}$, as discussed above. In order to estimate $f_\Phi(\phi)$ from data, we consider a discretised domain in ϕ , and use equation (5.14) to find the corresponding domain in the (x, y) plane. Note that the data must be in the canonical form, as described above. For a region, $\phi \in [\phi_1, \phi_2]$, where $\phi_2 \geq \phi_1$, we obtain the following inequalities

$$\begin{aligned} \phi_1 &\leq \arccos\left(\frac{y^2 - (c\tau - x)^2}{y^2 + (c\tau - x)^2}\right) < \phi_2 \\ \Rightarrow \sqrt{\frac{1 + \cos(\phi_1)}{1 - \cos(\phi_1)}} &\geq \frac{y}{c\tau - x} > \sqrt{\frac{1 + \cos(\phi_2)}{1 - \cos(\phi_2)}}. \end{aligned} \quad (5.23)$$

Equation (5.23) indicates that the area in (x, y) is bounded by two straight lines and the circle with radius $c\tau$, as shown in Figure 5.16. The area of one of the shaded regions is given by

$$A = \frac{1}{2}(\phi_2 - \phi_1 + \sin(\phi_1) - \sin(\phi_2)). \quad (5.24)$$

Rearranging equation (5.17) gives

$$f_\Phi(\phi(x, y)) = c\tau(c\tau - x)f_{X,Y}(x, y). \quad (5.25)$$

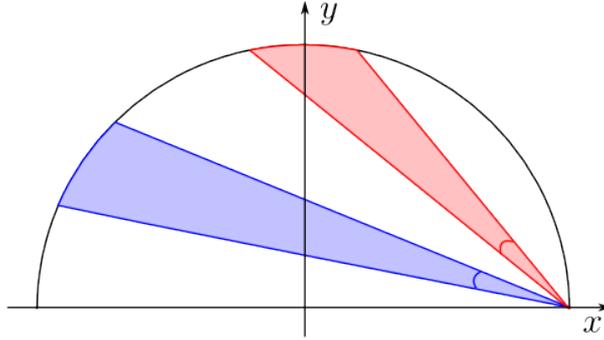


Figure 5.16: The regions in (x, y) space corresponding to two different ranges of ϕ shown in red and blue.

Integrating this quantity over a region, $\Omega = \{(x, y) : \phi(x, y) \in [\phi_1, \phi_2]\}$, we obtain

$$\iint_{\Omega} f_{\Phi}(\phi(x, y)) dx dy = c\tau \iint_{\Omega} (c\tau - x) f_{X,Y}(x, y) dx dy. \quad (5.26)$$

We are therefore able to estimate the probability density of ϕ within Ω as an approximation of the integral on the right hand side of equation (5.26), using the observed data. Extracting this value from the tracks produced by experiment or stochastic simulation is achieved by performing the following computations:

1. convert all transitions to canonical transitions;
2. discretise the turning angles ϕ , denoting each bin by ϕ_i ;
3. for each ϕ_i , find the corresponding region in (x, y) , denoted by Ω_i , using equation (5.23);
4. find all simulated data whose final position is within Ω_i . Compute the mean value of $c\tau - x$ for these matching results;
5. estimate $f_{X,Y}(x, y)$ in the region Ω_i by taking the number of matches divided by the area of the region, computed using equation (5.24);
6. the product of results from steps 4 and 5 gives an unnormalised estimate for the probability density of ϕ within the bin ϕ_i . Normalise the estimated distribution.

The result of performing this procedure on simulated data with varying numbers of tracks is shown in Figure 5.17(a) along with the true von Mises turning distribution. Tracks were simulated with $\tau_{10} = 20$ s, $\tau = 1$ s and $\sigma_\delta = 1.0$ for a total simulation time of 200 s. The inferred turning distribution is more accurate as the number of simulated trajectories increases. The agreement is already good when 100 tracks are included, and the improvement upon including 5000 tracks is minor. The estimate is less accurate at lower values of ϕ , because the area of the enclosing region in the (x, y) plane is small. There are therefore fewer data points available to estimate $f_{X,Y}(x, y)$ in the region, leading to a larger error in the estimate relative to regions at higher values of ϕ . Further numerical simulations show that this discrepancy at small values of ϕ increases with τ/τ_{10} (data not shown).

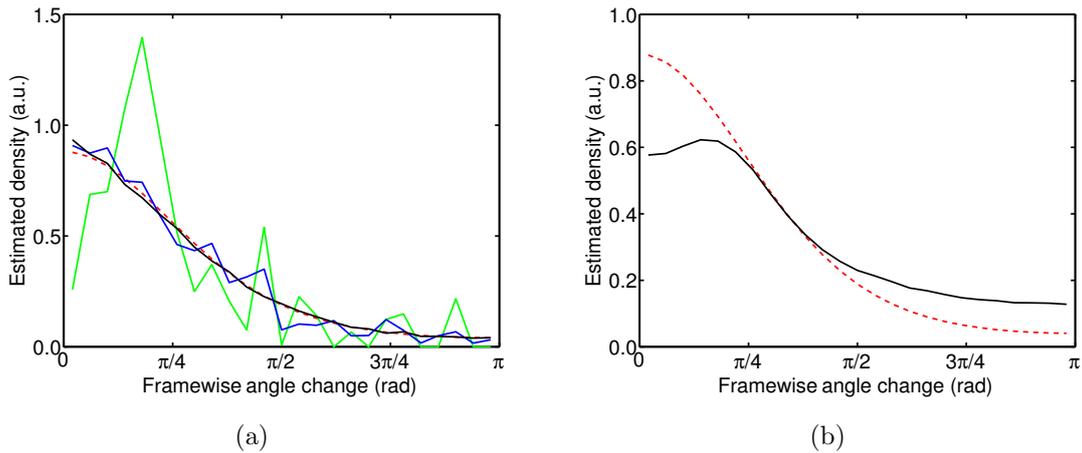


Figure 5.17: Plot of the inferred angle change distributions computed from simulated data. The true underlying turning kernel is von Mises with $\sigma_\delta = 1.0$, plotted in red. (a) Data simulated with no added noise and a varying number of particle trajectories: 10 (green), 100 (blue), and 5000 (black). (b) Simulated data comprising 5000 trajectories and incorporating Gaussian noise with $D = 1 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$.

A key difference between the simulated data used throughout this chapter and the experimental tracks in Chapter 4 is the lack of noise in the former. In order to assess the suitability of the present analysis to real tracks, we test our solution to the inverse problem on simulated data with a low level of added white noise. The tracks are simulated as described in Chapter 3, with noise strength $D = 1 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$. Figure 5.17(b)

shows the result of attempting to infer the underlying turning kernel with this noisy data. The algorithm performs poorly, with particularly large discrepancies at low and high values of ϕ . This failure occurs because transitions containing no reorientations, which are straightforward to filter out when no noise is present, may escape the filtering stage if the addition of noise makes them appear to be the result of a change in angle. Furthermore, the algorithm is highly sensitive to small perturbations around the region $(x, y) = (c\tau, 0)$. As Figure 5.16 illustrates, a small shift in position in this region can drastically change the computed value of ϕ .

5.5 Discussion and conclusions

In this chapter, we have considered the effect of the sampling frequency used to obtain data on an observed VJ process. We showed that this is an important factor to consider, as several of the quantities that we are interested in extracting from our tracking data are strongly affected by the sampling frequency. We initially derived results that are valid at long sampling times, based on the assumption of a diffusive process. These provide a useful consistency check for the ensuing work. We next proceeded to analyse the short and intermediate sampling time regimes using simulated tracks. In particular, we have built upon the study carried out by Codling and Hill [50] by considering the effect that sampling frequency has on the observed pdfs, rather than the summary statistics. We also incorporated a stopping phase into the underlying process, the effect of which has not been considered before. The results in Section 5.3 were shown to be consistent with the findings in [50]. An important result arising from the simulation study was the question of how to infer the underlying distribution of angle changes. Since sampling times do not in general coincide with reorientation events, we cannot simply ‘read off’ the result from a histogram of framewise angle changes. This provided the motivation for our analytic study in Section 5.4, in which we used mathematical modelling to describe the sampling process and its effect on the observed data. Our approach proved useful in

deriving the expected form of the pdf governing framewise speeds and angle changes for a given underlying angle change distribution (the forward problem), subject to a condition on the sampling time. Attempting to invert the process, in order to infer the underlying angle change distribution from observed data, proved moderately successful on simulated, noise-free data, although many tracks were required to ensure accurate inference.

The sampling times considered in Sections 5.2 and 5.3 are greater than or equal to the value of τ_{10} . This is in contrast to those considered in Section 5.4, in which we require $\tau \ll \tau_{10}$. The latter sampling regime is more realistic for the data that we presented in Chapter 4, from which we inferred that $\tau_{10} \approx 0.7$ s. The data were obtained with $\tau = 0.02$ s, hence the sampling rate is theoretically sufficient to allow for comparison with the results derived in Section 5.4. However, various complications, including the run-and-stop motion exhibited by *R. sphaeroides*, and the fact that bacteria do not move with a constant speed, prevent a detailed comparison at present. Indeed, this was not the main goal of our present work. Instead, we have completed a more general modelling study of the issue of sampling frequency, valid in different sampling time limits. Further work could include a more detailed study of the effect of sampling frequency when $\tau \ll \tau_{10}$, for example including stopping phases or a variable run speed.

A further important omission from the present work is the inclusion of noise in the tracks. A key example is the failure of our solution to the inverse problem when even low levels of noise were included in the simulated tracks. With the exception of Codling and Hill [50], who also neglect noise in their study, there are no investigations into the effect of sampling frequency on observed data. We therefore opted to focus on the noise-free case in order to address some of the open questions in this field. Further work could include a study of the interplay between sampling rate and measurement noise. This is necessary if we are to investigate the optimal sampling frequency, as measurement noise typically increases with increasing sampling rate in microscope cameras, due to the resulting reduction in

image intensity [202]. Another open question relates to the effect of process noise, for example Brownian buffeting, and how measurements extracted from such noisy tracks vary with the sampling frequency. This is the motivation for Chapter 6, in which we analyse the role of noise in bacterial motility.

Chapter 6

Modelling the effect of Brownian buffeting on motile bacteria

In the previous chapter, we saw that our analytic description of the effect of sampling frequency on observations of the idealised velocity jump (VJ) process was of limited application to experimental data due to the presence of noise. In particular, the VJ model of bacterial motion is one of straight line movements punctuated by reorientations; in reality various sources of noise cause a departure from this idealisation. Bacteria do not swim in perfectly straight lines, as is evident from, for example, Figure 4.9. The sources of noise that lead to such apparent departures from straight line motion include Brownian buffeting, measurement error, and fluctuations in the conformation of the flagellum and in the rotation rate of the flagellar motor. In the current chapter, we focus on the effect of rotational diffusion, caused by Brownian buffeting, on bacterial motility, since the physical and mathematical descriptions of such noise are well understood. We consider three models to analyse the role of rotational diffusion in the motility of *R. sphaeroides*, and other monoflagellates with similar motile behaviour. The first is a model developed by Hagen *et al.* [83] to describe the motion of a self-propelled particle that is subject to rotational diffusion. The two other models are novel; both are based on a combination of the self-propelled particle model and the VJ model described in Chapter 2. The

phenomenon of rotational diffusion is reconciled with the VJ process by describing the reorientation of a bacterium during a stopping phase in terms of rotational diffusion. A comparison of this model with experimental data demonstrates the important point that *R. sphaeroides* cannot reorientate by Brownian rotational motion alone. A simplified model of an alternative active reorientation mechanism is proposed.

6.1 Background

The small size and mass of a bacterium means that its motion is dominated by viscous forces; inertial forces are so small as to be negligible [21]. The relative importance of these two effects for a body moving in a fluid is quantified by the Reynolds number, defined by

$$R = \frac{vL\rho}{\eta}, \quad (6.1)$$

where v is the velocity of the particle relative to the fluid, L is the length scale of the particle, ρ is the density of the fluid and η is the fluid viscosity. For small bodies such as bacteria, R is very small (on the order of 10^{-5} [21]). A numerical example of the importance of the low Reynolds regime is given by Purcell [163]. Approximating a bacterium as a sphere with a diameter of $1 \mu\text{m}$ moving with a speed of $20 \mu\text{ms}^{-1}$, the coasting distance of a bacterium upon instantaneous cessation of the flagellar motor is equal to 4×10^{-12} m, or 4 millionths of the diameter of the bacterium. This quick calculation demonstrates that, at such low Reynolds numbers, viscous effects are dominant, and objects are brought to a halt almost immediately.

Whilst particles at low Reynolds number have extremely small coasting distances, they are never completely stationary as they exhibit Brownian motion, also known as diffusion. Brownian buffeting leads to two related random processes: translational and rotational diffusion. These processes are illustrated in Figure 6.1. Translational diffusion leads to a shift in the centre of mass of a particle, whereas rotational diffusion leads to a reorienta-

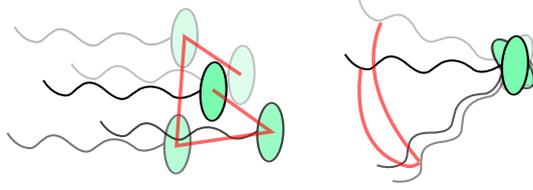


Figure 6.1: *Illustration of a bacterium undergoing translational (left) and rotational (right) diffusion. In each case, the cartoon is of a single cell. Transparent cells indicate historic position and orientation, with increasing transparency representing the more distant past. Red lines trace the trajectory of the cell centroid over time in the translational case, and trace a point on the flagellum to show the angle changes in the rotational case.*

tion of the particle about its centre of mass. The understanding of the role of Brownian buffeting on bacterial motility has been elucidated by Berg [21] using a theoretical model of a self-propelled particle that is subject to rotational diffusion. A bacterium is modelled as a particle propelled by a rigid helix. At random intervals, the cell undergoes an instantaneous reorientation event, in which a new direction is selected from the wrapped uniform distribution. Between these events, the bacterium is propelled forwards. This is the classic VJ model of Othmer *et al.* [146]. The inclusion of rotational diffusion leads to running phases in which the cell reorientates continuously, which is a departure from the VJ model. As a result of rotational diffusion, cells do not trace out perfectly straight runs.

Mitchell [136] uses the model of a self-propelled particle to analyse the effect of cell size and swimming speed on the efficiency and feasibility of bacterial motility in marine environments. In order for bacteria to perform chemotaxis and bias their movements towards sources of nutrients, they must be able to swim sufficiently quickly to detect a change in nutrient concentration before rotational diffusion randomises the direction of swimming. Purcell [163] describes this as the need to ‘out-swim diffusion’: nutrients undergo diffusion in a liquid medium, thus a bacterium must move a certain distance in order to arrive at a region with a detectably different nutrient concentration. Mitchell [136] equates this minimum distance with the mean distance a bacterium travels before rotational diffusion causes an orientation change of $\pi/2$. The extent of rotational diffusion scales with r^{-3} , where r is the radius of the bacterium, and is independent of speed [136]. Smaller cells

must therefore swim more rapidly in order to undergo chemotaxis, as Figure 6.2 shows.

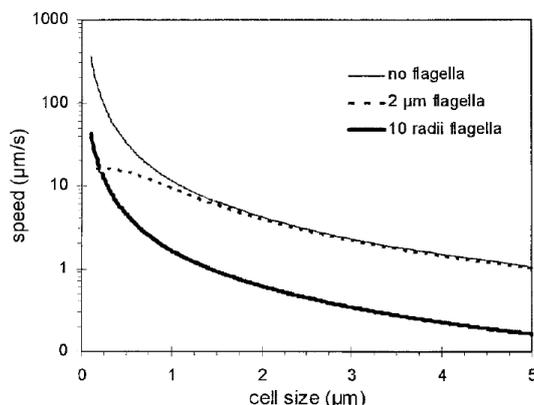


Figure 6.2: Numerical results from a theoretical model showing the dependence of minimum swimming speed on cell size for bacteria with or without a flagellum. Reproduced with permission from [137].

The extent of rotational diffusion is reduced when the frictional effects of the flagellum are incorporated [137], as this increases the viscous drag acting on the bacterium. This is termed ‘flagellar stabilization’. The dependence of minimum swimming speed on cell size for bacteria with or without a flagellum is shown in Figure 6.2. Dusenbery [65] carried out a similar analysis, and predicted that planktonic organisms with radii below around $0.6 \mu\text{ms}^{-1}$ are unlikely to gain any advantage from swimming motility. This prediction was borne out in a systematic investigation of known bacteria [65]. It is, however, unclear whether the more general prediction of Mitchell [137], that swimming speed should correlate with cell size, holds in nature. A diverse study of marine bacteria carried out by Johansen *et al.* [98] failed to find any such correlation.

Mitchell’s approach to considering the efficiency of bacterial chemotaxis also permits a discussion of the relative advantages of the various methods of reorientation by motile bacteria (see Figure 1.3) [138]. Of particular interest is the application of this model of rotational diffusion to *R. sphaeroides*: Mitchell and Kogure [138] point out that this bacterium is too large to reorientate efficiently by rotational Brownian diffusion alone. This is also noted by Armitage *et al.* [12], whose experimental study suggests that an

active reorientation mechanism may be present. However, bead assays in *R. sphaeroides* have provided contradictory evidence, which suggests that the motor does indeed stop [156]. We discuss this aspect, with novel supporting data and modelling, in more detail below.

Further outcomes of the self-propelled particle model arise from a more in-depth analysis of the relevant equations of motion, which take the form of coupled stochastic differential equations (SDEs) (see Section 6.2 for a detailed discussion). This permits the derivation of expressions for summary statistics such as the MSD, velocity auto-correlation function (VACF), skewness and kurtosis [82, 83, 216].

6.2 Mathematical methods

Brownian motion is the name given to the random motion of a small particle in a liquid medium, as first observed by the botanist Brown in 1827 while he was studying grains of pollen in water under a microscope [52]. The motion arises from the net force acting on the particle as a result of the very large number of collisions between the surrounding molecules in the liquid and the Brownian particle. The time average of these collisions is zero, however the instantaneous net force is a random property, causing the particle to exhibit ceaseless small movements. In this chapter, we use a mathematical description of Brownian motion called the Langevin equation. We present the details of the model in one dimension, before extending the description to two dimensions, to enable comparison of the model with our two-dimensional experimental data (see Chapter 2 for further discussion). The Langevin equation is a natural framework to use for the analysis of such stochastic systems. In particular, appropriate analysis of the relevant equations gives us access to the various moments mentioned previously. We now describe this equation, and provide detail on some relevant mathematical methods for analysing stochastic processes.

6.2.1 The Langevin equation

The Langevin equation is a SDE describing the forces experienced by a particle undergoing Brownian motion in a liquid. In one dimension, the translational Langevin equation is given by [52]

$$\frac{dx(t)}{dt} = v(t), \quad (6.2)$$

$$m \frac{dv(t)}{dt} = -\zeta v(t) + \xi(t), \quad (6.3)$$

where $x(t)$ is the time-dependent position of the particle, m is mass, ζ is the frictional drag coefficient and $\xi(t)$ is a random fluctuating force, discussed further below. The left-hand side of (6.3) is the force exerted on the particle, which is equal to the sum of a deterministic viscous drag term and a force arising from the random collisions of the surrounding molecules in the liquid. The random quantity $\xi(t)$ represents the net force exerted on the particle over a time interval that is short relative to the timescale of viscous relaxation, characterised by the quantity m/ζ , and long relative to the mean time between collisions of the molecules in the medium. The ensemble average of this force, denoted $\langle \xi(t) \rangle$, interpreted as the average over realisations of the random variable $\xi(t)$, is zero. The autocorrelation of the force $\xi(t)$ is given by

$$\langle \xi(t)\xi(t') \rangle = g\delta(t - t'), \quad (6.4)$$

where g is a constant, given by

$$g = 2kT\zeta, \quad (6.5)$$

where k is the Boltzmann constant and T is temperature in units of kelvin. Equation (6.5) is known as the Fluctuation-Dissipation Theorem [52]. This theorem thus links the strength of random fluctuations in the Langevin equation to physical properties of the system.

The property summarised in equation (6.4) is known as *delta correlation*, and is a characteristic of a *white noise* force [52]. We may interpret property (6.4) as a further statement of the separation of the viscous relaxation timescale and the timescale of the motion of the surrounding molecules. The number of collisions that occur on the timescale of the motion of the Brownian particle is very large, so that the motion of the particle is approximately independent of any previous motion. This is clearly an idealisation: there must be some timescale on which the motion of the molecules in the surrounding medium are correlated. However, this timescale is on the order of 10^{-12} s [52], whereas we are generally interested in timescales on the order of 10^{-3} s or greater (the timescale on which we observe bacterial motility). Furthermore, the viscous relaxation timescale for a Brownian particle is approximately 10^{-8} s for a spherical particle of radius $\approx 1 \mu\text{m}$ (see calculation below), so even if we were interested in this phenomenon, we could treat the oscillating force as uncorrelated. The idealised white noise force is therefore a suitable approximation to the true fluctuating force acting on a Brownian particle.

6.2.2 Overdamped Langevin equation

The left-hand side of equation (6.3) represents the inertial forces acting on the particle. This is significant for particles whose mass is large relative to the viscous drag coefficient ζ . In the case of a bacterium, however, viscous forces dominate so that $m/\zeta \ll 1$ s. We can quantify this statement by considering the approximate values of the two constants for a sphere of radius $1 \mu\text{m}$. By Stokes' Law, $\zeta = 6\pi\eta r$ for a sphere, which gives $\zeta \approx 10^{-7} \text{ kg s}^{-1}$ for a cell in water. Assuming that the density of a cell is approximately equal to that of water at room temperature, $m \approx 10^{-15} \text{ kg}$. The ratio m/ζ , which gives the characteristic viscous relaxation timescale, is hence on the order of 10^{-8} s.

Combining equations (6.2)-(6.3), we obtain

$$\frac{m}{\zeta} \frac{d^2x}{dt^2} + \frac{dx}{dt} = \frac{1}{\zeta} \xi(t). \quad (6.6)$$

The *overdamped* Langevin equation is obtained by neglecting inertial forces, which is justified as $m \ll \zeta$. We neglect the first term in equation (6.6) to obtain

$$\frac{dx}{dt} = \frac{1}{\zeta} \xi(t). \quad (6.7)$$

The overdamped Langevin equation may be used as a starting point for modelling bacterial motility (see, for example, [82, 83]).

6.2.3 The rotational Langevin equation

As illustrated in Figure 6.1, translational and rotational diffusion are related, but distinct, processes. The process described by equations (6.2)-(6.3) is translational Brownian motion. Consider an oriented sphere of uniform density with radius r and an orientation vector $\boldsymbol{\mu}(t)$, which is fixed relative to the sphere so that rotation of the sphere changes the orientation vector accordingly. The related equations for the rotational Brownian movement of the sphere in three dimensions are given by [52]

$$\frac{d\boldsymbol{\mu}(t)}{dt} = \boldsymbol{\omega}(t) \times \boldsymbol{\mu}(t), \quad (6.8)$$

$$I \frac{d\boldsymbol{\omega}(t)}{dt} + \zeta \boldsymbol{\omega}(t) = \boldsymbol{\xi}(t), \quad (6.9)$$

where I is the moment of inertia of the sphere, equal to $2mr^2/5$, $\boldsymbol{\omega}$ is the angular momentum and $\boldsymbol{\xi}(t)$ is a vector-valued white noise force, with each component IID with the same zero mean and delta correlation given in equation (6.4). As before, inertial terms are small compared with viscous terms, $I \ll \zeta$. Neglecting the first term in equation

(6.9) and substituting into equation (6.8) gives

$$\zeta \frac{d\boldsymbol{\mu}(t)}{dt} = \boldsymbol{\xi}(t) \times \boldsymbol{\mu}(t). \quad (6.10)$$

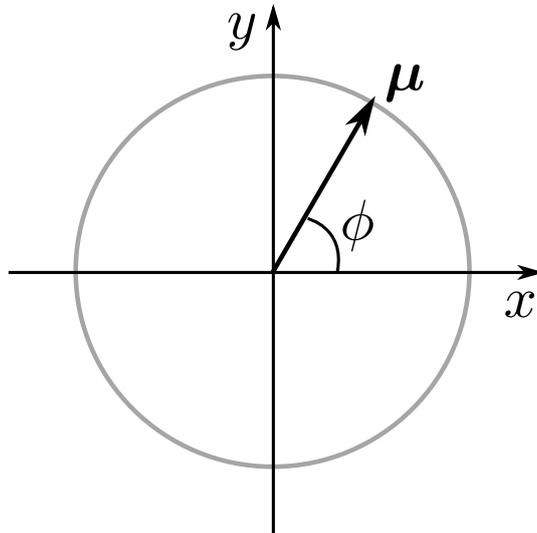


Figure 6.3: *Illustration of the orientation vector for a rotating sphere whose motion is confined to two dimensions. See text for details.*

Since we will be considering rotational diffusion in two dimensions, we now derive the simplified two-dimensional form of equation (6.10), as used in [83]. In this case, the orientation vector can be described in terms of a single degree of freedom, ϕ . This is the angle that the orientation vector makes with the horizontal, as illustrated in Figure 6.3.

The orientation vector becomes

$$\boldsymbol{\mu}(t) = \begin{pmatrix} \cos \phi \\ \sin \phi \\ 0 \end{pmatrix}. \quad (6.11)$$

Taking the dot product of equation (6.10) with the unit vector in the x -direction, we obtain

$$\frac{d\phi}{dt} = \frac{1}{\zeta} \xi(t), \quad (6.12)$$

where $\xi(t)$ is a single component of the white noise vector $\boldsymbol{\xi}(t)$.

6.2.4 The Wiener process

The white noise force $\xi(t)$ approximates the combined effect of very many collisions between the molecules in the surrounding medium and the Brownian particle. We have already seen that two important properties of this force are that it has an ensemble average of zero, and is delta correlated (see equation (6.4)). A further property of $\xi(t)$, often assumed but not strictly necessary for the ensuing analysis, is that it is a Gaussian process, i.e. it is normally distributed. This assumption may be justified by considering that $\xi(t)$ is the sum of very many identically distributed random collisions, hence by the Central Limit Theorem we expect it to follow a normal distribution.

The Wiener process, denoted $B(t)$, is best defined as the integral of the standard white noise process $X(t)$, which has zero mean and unit variance,

$$B(t) = \int_0^t X(s) ds. \quad (6.13)$$

For a more detailed discussion of the properties of the integral of a random process, see [107]. The interpretation of the Wiener process defined in equation (6.13) is as the position of a Brownian particle in one dimension at time t . From equation (6.13), it is clear that $B(0) = 0$. The integral in equation (6.13) is a Riemann integral, hence we may consider the Wiener process as the limit of a sum of Gaussian increments,

$$B(t) = \lim_{\delta_n \rightarrow 0} \sum_{i=1}^n X(t_i)(t_i - t_{i-1}), \quad (6.14)$$

where $0 = t_0 < t_1 < \dots < t_n = t$ is a partition of the domain $[0, t]$ into n intervals, and $\delta_n = \max_{1 \leq i \leq n} (t_i - t_{i-1})$ is the length of the largest interval. Since $B(t)$ is an infinite sum of IID random variables, it is a Gaussian process by the Central Limit Theorem. Non-intersecting increments of the Wiener process are stationary and independent, and are normally distributed with zero mean and mean squared value proportional to the

width of the increment,

$$\langle B(t) - B(s) \rangle = 0, \quad (6.15)$$

$$\langle [B(t) - B(s)]^2 \rangle = |t - s|, \quad \forall s, t \geq 0. \quad (6.16)$$

Setting $s = 0$ gives similar results for an increment starting at time zero,

$$\langle B(t) \rangle = 0, \quad (6.17)$$

$$\langle B^2(t) \rangle = t, \quad \forall t \geq 0. \quad (6.18)$$

As $B(t)$ is normally distributed, its characteristic function is given by

$$\varphi_{B(t)}(u) = \langle e^{iuB(t)} \rangle = e^{-u^2t/2}, \quad u \in \mathbb{R}. \quad (6.19)$$

In our case, the white noise process $\xi(t)$ has a second moment equal to $2kT\zeta$ (see equation (6.5)) hence $\xi(t) = \sqrt{2kT\zeta}X(t)$. Integrating (6.12) with the initial condition $\phi(0) = 0$, and applying the definition of the Wiener process, we obtain

$$\phi(t) = \frac{1}{\zeta} \int_0^t \xi(s) ds = \sqrt{\frac{2kT}{\zeta}} \int_0^t X(s) ds = \sqrt{\frac{2kT}{\zeta}} B(t). \quad (6.20)$$

Using equation (6.18), the second moment of $\phi(t)$ is then given by

$$\langle \phi^2(t) \rangle = \frac{2kT}{\zeta} t. \quad (6.21)$$

6.2.5 Integral of a stochastic process

In Section 6.3, we shall encounter integrals of the form

$$Z(t) = \int_0^t Y(s) ds,$$

where $Y(s)$ is a measurable stochastic process [80]. If

$$\langle |Y(s)| \rangle < \infty \quad \forall s \in [0, t], \quad (6.22)$$

then we can compute $\langle Z(t) \rangle$ using the following manipulation [59]:

$$\langle Z(t) \rangle = \int_0^t \langle Y(s) \rangle \, ds. \quad (6.23)$$

This equality follows from the application of Fubini's Theorem [162] to interchange the order of integration, since the process of taking the ensemble average is actually an integral over realisations of $Y(s)$.

6.3 A model of rotational diffusion in bacterial motility

In this chapter, we consider three models to analyse the role of rotational diffusion in flagellar-mediated bacterial motility. All three models are based on the overdamped Langevin description of a self-propelled particle. Following Hagen *et al.* [83], we consider only rotational diffusion, as translational diffusion has a relatively minor effect on the motion of a self-propelled particle when the rate of propulsion is sufficiently high, as is the case here. Figure 6.4 illustrates the geometry of the model. The bacterium is modelled as either a sphere or an ellipsoid that is propelled by a single flagellum, modelled as a rigid helix. When the cell body is ellipsoidal, the flagellum is attached at the midpoint of the long axis as is generally observed experimentally [11]. The unit vector in the axis that runs through the centre of the flagellum and the centre of the cell is the orientation vector $\boldsymbol{\mu}$, whose angle to the horizontal is denoted ϕ , as in Section 6.2.3. Since we consider only motion in two dimensions, we assume that $\boldsymbol{\mu}$ lies in the (x, y) plane. Note that the cell body is three-dimensional, however. In all models, the cell is assumed to be propelled

forward at a constant speed of $40 \mu\text{ms}^{-1}$ when the propulsive force is active. This is the approximate modal speed exhibited by wildtype *R. sphaeroides* (see Figure 4.16).

In the run-only model, the propulsive force is active at all times. This is a simplified representation of the *R. sphaeroides* non-chemotactic mutant strain (see Chapter 2), which is unable to stop. In the run-and-stop model, the propulsion force undergoes stochastic switching events between running and stopping states as a Poisson process. In line with the notation used in previous chapters, the rate of switching from a run to a stop is denoted λ , and that of switching from a stop to a run is μ . Rotational diffusion continues to act regardless of the the state of the propulsion force, hence the bacterium undergoes passive diffusional reorientation in the stopping phases. This model attempts to capture the details of a wildtype *R. sphaeroides* cell, assuming that stops occur by rotational diffusion alone. Finally, in the run-and-active-stop model, we incorporate an additional stochastic reorientating force in the stopping phases. This force acts in the same way as passive rotational diffusion, but with increased variance, so that reorientation occurs more rapidly in the stopping phase. This model represents a first step towards investigating the mechanism of reorientation in *R. sphaeroides*.

6.3.1 Differing geometries for the bacterial cell body

As illustrated in Figure 6.4, we consider the effect of varying the geometry of the bacterial cell body on the role of rotational diffusion. We model the bacterium as a prolate ellipsoid, which is a rugby ball-shaped ellipsoid whose two shorter semi-principal axes are of equal length. This approximation is commonly used when modelling bacteria [105, 108]. We denote the length of the axial semi-principal axis a , and the length of the equatorial semi-principal axes b and c . For a prolate ellipsoid, $a > b = c$. We quantify the geometry by the axial ratio, $\rho = a/b$, hence, for a prolate ellipsoid $\rho > 1$.

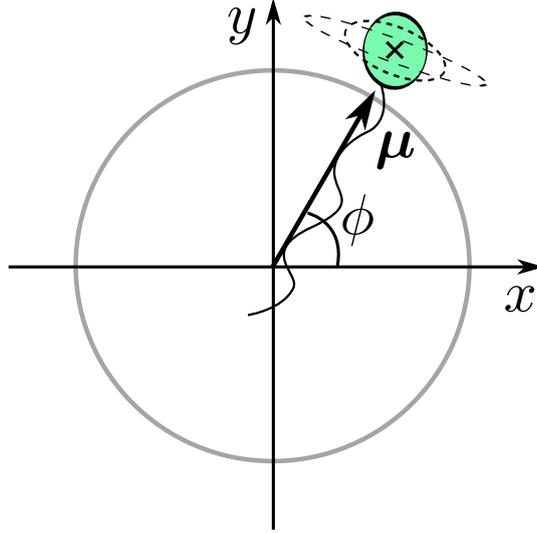


Figure 6.4: *Illustration of the model of rotational diffusion for a swimming bacterium. The three-dimensional bacterium is confined to travel in the (x, y) plane. Dashed lines represent alternative shapes for the cell body. The black cross indicates the centre of mass of the cell, which is calculated neglecting the effect of the flagellum.*

The translational and rotational diffusion coefficients of a general ellipsoid may be calculated using multiplicative adjustment to the values for a sphere of the equivalent volume, known as Perrin friction factors [109]. Berg [21] gives approximate expressions for these, however we use the exact forms here for greater accuracy when $\rho \approx 1$. For brevity, we define a new variable,

$$S = \frac{2\rho \tanh\left(\frac{\sqrt{\rho^2 - 1}}{\rho}\right)}{\sqrt{\rho^2 - 1}}. \quad (6.24)$$

Note that definition (6.24) is valid for prolate ellipsoids only. Since we consider a bacterium with a medially attached flagellum (see Figure 6.4), we are concerned with rotation about the *equatorial* semi-principal axis; rotation about the axial axis would cause the bacterium to swim out of the plane, breaking the assumptions of the model. The expression for Perrin's friction factor about the equatorial semi-principal axis is [109]

$$F_{eq} = \left(\frac{4}{3}\right) \frac{(1/\rho)^2 - \rho^2}{2 - S[2 - (1/\rho)^2]}. \quad (6.25)$$

To compute the rotational frictional drag coefficient for an ellipsoidal cell, we multiply the value of ζ for a sphere of equivalent volume by F_{eq} . The volume of a prolate ellipsoid is

given by $4\pi ab^2/3$. We assume that there is little variation in the length of the equatorial (shorter) semi-principal axis in bacteria, hence we fix b .

6.3.2 Run-only model

The governing Langevin equations for the run-only model are given by

$$\frac{d\mathbf{x}(t)}{dt} = c \begin{pmatrix} \cos \phi(t) \\ \sin \phi(t) \end{pmatrix}, \quad (6.26)$$

$$\frac{d\phi(t)}{dt} = \frac{1}{\zeta} \xi(t), \quad (6.27)$$

where $c = 40 \mu\text{ms}^{-1}$ is the speed of swimming. We assume, without loss of generality, that $\mathbf{x}(0) = \mathbf{0}$ and $\phi(0) = 0$. We introduce the rotational diffusion coefficient, D_r , given by

$$D_r = \frac{kT}{\zeta}. \quad (6.28)$$

The solution to equation (6.27) is the same as equation (6.20). Rewriting this in terms of the rotational diffusion coefficient, we obtain

$$\phi(t) = \sqrt{2D_r} B(t), \quad (6.29)$$

where $B(t)$ is a Wiener process as before. The angle $\phi(t)$ is therefore normally distributed with zero mean and variance equal to $2D_r t$. We now compute the first and second moments of the position vector $\mathbf{x}(t)$. For simplicity, we consider each element in equation (6.26) separately:

$$\frac{dx(t)}{dt} = c \cos \left(\sqrt{2D_r} B(t) \right), \quad (6.30)$$

$$\frac{dy(t)}{dt} = c \sin \left(\sqrt{2D_r} B(t) \right). \quad (6.31)$$

Integrating equation (6.30) gives

$$x(t) = c \int_0^t \cos\left(\sqrt{2D_r}B(s)\right) ds. \quad (6.32)$$

Rewriting the cosine function in complex form and applying the definition of the characteristic function in (6.19) yields

$$\langle x(t) \rangle = \frac{c}{D_r} (1 - e^{-D_r t}). \quad (6.33)$$

By symmetry, we have $\langle y(t) \rangle = 0$. This result is intuitive: particles initially have no y component in their velocity, and are equally like to travel in either the positive or negative y -direction.

We now state expressions for the second moments, denoted $\langle x^2(t) \rangle$ and $\langle y^2(t) \rangle$. A detailed derivation is presented in Appendix C.2. The second moments are given by

$$\langle x^2(t) \rangle = \left(\frac{c}{D_r}\right)^2 \left(\frac{2}{3}e^{-D_r t} + \frac{1}{12}e^{-4D_r t} - \frac{3}{4} + D_r t\right), \quad (6.34)$$

$$\langle y^2(t) \rangle = \left(\frac{c}{D_r}\right)^2 \left(\frac{4}{3}e^{-D_r t} - \frac{1}{12}e^{-4D_r t} - \frac{5}{4} + D_r t\right). \quad (6.35)$$

We note that equations (6.34) and (6.35) tend to zero as $t \rightarrow 0$, as expected. On a timescale that is short relative to the diffusion coefficient, $t \ll 1/D_r$, $\langle x^2(t) \rangle \approx (ct)^2$, which indicates that movement is ballistic, as is also the case for the VJ process over short timescales (see equation (2.23)). For times $t \gg 1/D_r$, both second moments vary as $c^2 t/D_r$. This linear behaviour of the second moment is characteristic of a diffusive process and, again, it is similar to that observed in the VJ process. The similarity of the run-only VJ process and the run-only rotational diffusion model has led to both models being considered by Campos *et al.* [43] to explain cell migration data. Several important differences exist between the models, however. The transition from ballistic

to diffusive motion in the VJ model occurs due to the discrete stochastic reorientation events, without which the process would be ballistic on all timescales¹. In contrast, the run-only model considered here has no reorientation events; the same transition is due to continuous rotational diffusion in this case. Even in the rapid turning limit of the run-only VJ process, $\lambda \rightarrow \infty$, so that the reorientation becomes approximately continuous, the two models are not equivalent. Turning angles in the VJ model are IID from an arbitrary circular distribution that is independent of λ , so that as the turning rate increases, persistence is lost increasingly rapidly (see equation (2.23) and related discussion). In contrast, the discrete approximation of the run-only model of rotational diffusion, which is the method we use to simulate the process (see Section 6.3.5), has turning angles whose pdf varies with the time step, and hence with the rate of turning.

6.3.3 Run-and-stop model

We now incorporate an additional stochastic process into the equations of motion of the self-propelled particle. This process, denoted $F(t)$, is a continuous time Markov process with two states, $F(t) \in \{0, 1\}$, $\forall t \geq 0$. The stochastic process F is identical to the underlying model used for the hidden Markov model in Chapters 3 and 4, such that the bacterium is in a stopping phase when $F(t) = 0$, and in a running phase when $F(t) = 1$. We assume that the bacterium is initially running, so that $F(0) = 1$. We require an expression for the ensemble average of this process, denoted $\langle F(t) \rangle$. We compute this by considering the evolution of the pdf $\boldsymbol{\pi}(t)$, defined by

$$\boldsymbol{\pi}(t) = \begin{pmatrix} p_0(t) \\ p_1(t) \end{pmatrix}, \quad (6.36)$$

¹This can be shown by taking the limit $\lambda_0 \rightarrow 0$ in equation (2.23).

where $p_i(t)$ is the probability that $F(t) = i$, for $i \in \{0, 1\}$. The evolution of $\boldsymbol{\pi}(t)$ is governed by the ODE [80]

$$\frac{d\boldsymbol{\pi}}{dt} = \begin{pmatrix} -\mu & \lambda \\ \mu & -\lambda \end{pmatrix} \boldsymbol{\pi}. \quad (6.37)$$

The solution of this linear ODE is given by

$$\boldsymbol{\pi} = \frac{\lambda}{\lambda + \mu} \left[\begin{pmatrix} 1 \\ \mu/\lambda \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-(\lambda+\mu)t} \right]. \quad (6.38)$$

Since $F = 1$ is the only allowable non-zero value, the ensemble average of F is simply given by

$$\langle F(t) \rangle = p_1(t) = \frac{1}{\lambda + \mu} (\mu + \lambda e^{-(\lambda+\mu)t}). \quad (6.39)$$

The Langevin equations governing the motion of the running and stopping particle are given by

$$\frac{d\mathbf{x}(t)}{dt} = cF(t) \begin{pmatrix} \cos \phi(t) \\ \sin \phi(t) \end{pmatrix}, \quad (6.40)$$

$$\frac{d\phi(t)}{dt} = \frac{1}{\zeta} \xi(t). \quad (6.41)$$

Integrating equation (6.41) gives the same solution as equation (6.29). Integrating equation (6.40) gives

$$x(t) = c \int_0^t F(s) \cos \left(\sqrt{2D_r} B(s) \right) ds, \quad (6.42)$$

$$y(t) = c \int_0^t F(s) \sin \left(\sqrt{2D_r} B(s) \right) ds. \quad (6.43)$$

We compute the ensemble average of $x(t)$ as follows:

$$\begin{aligned}\langle x(t) \rangle &= c \int_0^t \langle F(s) \rangle \langle \sin \left(\sqrt{2D_r} B(s) \right) \rangle ds \\ &= \frac{c}{\lambda + \mu} \left[\frac{\mu}{D_r} (1 - e^{-D_r t}) + \frac{\lambda}{D_r + \lambda + \mu} (1 - e^{-(D_r + \lambda + \mu)t}) \right],\end{aligned}\quad (6.44)$$

where the first equality holds by interchanging the order of integration, as in Section 6.3.2, and by the independence of $F(s)$ and $B(s)$. As for the run-only model, we have $\langle y(t) \rangle = 0$.

As before, we derive expressions for the second moments of $x(t)$ and $y(t)$. The manipulations are similar to the run-only case:

$$\begin{aligned}\langle x^2(t) \rangle &= c^2 \int_0^t \int_0^v \langle F(u) F(v) \rangle \left\{ \left\langle \cos \left[\sqrt{2D_r} (B(v) + B(u)) \right] \right\rangle \right. \\ &\quad \left. + \left\langle \cos \left[\sqrt{2D_r} (B(v) - B(u)) \right] \right\rangle \right\} du dv.\end{aligned}\quad (6.45)$$

The cosine terms are evaluated as before (see equations (C.2)-(6.34)). We require an expression for the autocorrelation $\langle F(u) F(v) \rangle$, where $u \leq v \leq t$. For a non-stationary continuous time Markov process $X(t)$, this is given by [76]

$$\langle X(s) X(t) \rangle = \sum_{x(t)} \langle X(s) | X(t) = x \rangle x \mathbb{P}(X(t) = x), \quad 0 \leq t \leq s, \quad (6.46)$$

where the summation is over all possible values of $X(t)$, and the first term in the summation denotes the average value of $X(s)$, conditional on $X(t)$ taking the value x . In our case, the summation only produces a single non-zero term,

$$\langle F(s) F(t) \rangle = \langle F(s) | F(t) = 1 \rangle \mathbb{P}(F(t) = 1). \quad (6.47)$$

The final term, $\mathbb{P}(F(t) = 1) \equiv p_1(t)$ is given by equation (6.39). The conditional mean,

$\langle F(s) | F(t) = 1 \rangle$, is found by solving for $p_1(s)$, subject to the modified initial condition $F(t) = 1$. The solution is given by

$$\langle F(s) | F(t) = 1 \rangle = \frac{1}{\lambda + \mu} (\mu + \lambda e^{-(\lambda+\mu)(s-t)}). \quad (6.48)$$

Substituting (6.39) and (6.48) into (6.47), we obtain

$$\langle F(s)F(t) \rangle = \left(\frac{\mu}{\lambda + \mu} \right)^2 \left(1 + \frac{\lambda}{\mu} e^{-(\lambda+\mu)t} \right) \left(1 + \frac{\lambda}{\mu} e^{-(\lambda+\mu)(s-t)} \right). \quad (6.49)$$

Substituting equation (6.49) into (6.45), and using the identities in equations (C.2) and (C.3) gives

$$\begin{aligned} \langle x^2(t) \rangle &= \left(\frac{c\mu}{\lambda + \mu} \right)^2 \int_0^t \int_0^v \left[e^{-D_r(3u+v)} + e^{-D_r(v-u)} \right] \\ &\quad \times \left[\left(1 + \frac{\lambda}{\mu} e^{-(\lambda+\mu)u} \right) \left(1 + \frac{\lambda}{\mu} e^{-(\lambda+\mu)(v-u)} \right) \right] du dv. \end{aligned} \quad (6.50)$$

The integral in equation (6.50) is straightforward to calculate but is not presented here for the sake of brevity. A similar expression may be derived for the second moment of $y(t)$:

$$\begin{aligned} \langle y^2(t) \rangle &= \left(\frac{c\mu}{\lambda + \mu} \right)^2 \int_0^t \int_0^v \left[e^{-D_r(v-u)} - e^{-D_r(3u+v)} \right] \\ &\quad \times \left[\left(1 + \frac{\lambda}{\mu} e^{-(\lambda+\mu)u} \right) \left(1 + \frac{\lambda}{\mu} e^{-(\lambda+\mu)(v-u)} \right) \right] du dv. \end{aligned} \quad (6.51)$$

6.3.4 Run-and-active-stop model

In this model, we incorporate an additional stochastic rotational force, which acts as a multiplier to the rotational diffusion coefficient in stopping phases only. We denote this

variable ω . The governing equations are given by

$$\frac{d\mathbf{x}(t)}{dt} = cF(t) \begin{pmatrix} \cos \phi(t) \\ \sin \phi(t) \end{pmatrix}, \quad (6.52)$$

$$\frac{d\phi(t)}{dt} = \frac{1}{\zeta} [F(t) + (1 - F(t))\sqrt{\omega}] \xi(t). \quad (6.53)$$

We do not solve for the moments in this case, as the result would be algebraically cumbersome and not particularly instructive.

6.3.5 Numerical implementation

In order to verify our analytic results, in addition to gaining quantitative insight where such results are not possible, we use a stochastic numerical simulation of the self-propelled particle model. Our numerical simulation uses the Euler-Maruyama (EM) method, a well-established method of generating a numerical approximation to the solution of SDEs [87]. The EM method is a discrete-time approximation to the underlying equations and therefore requires that we specify a simulation time step, denoted Δt . The error in the EM approximate solution decreases with Δt (it has a weak order of convergence equal to 1 [87]). In addition, the algorithm is only stable for sufficiently small time steps, hence we must set Δt sufficiently small to ensure that our results are accurate. The algorithm used to simulate the run-and-active-stop model is given in Algorithm 3 in Appendix C.1. The run-only model is simulated by setting $\lambda = 0$, and hence preventing stops. The run-and-stop model is simulated by setting $\omega = 1$, in which case $\sigma_{\text{run}} = \sigma_{\text{stop}}$ and there is no additional reorientation force.

All simulations carried out in this chapter had a time step $\Delta t = 0.02$ s, a total simulation time of 10 s and a temperature of $T = 300$ K. The viscosity of water at 300 K is 10^{-3} Pa s. We simulate 5000 tracks each time the algorithm is run. The results are found in Section 6.4. We state where the results are computed using simulated data.

6.4 Results

6.4.1 Comparison of the theoretical and estimated translational diffusion coefficient

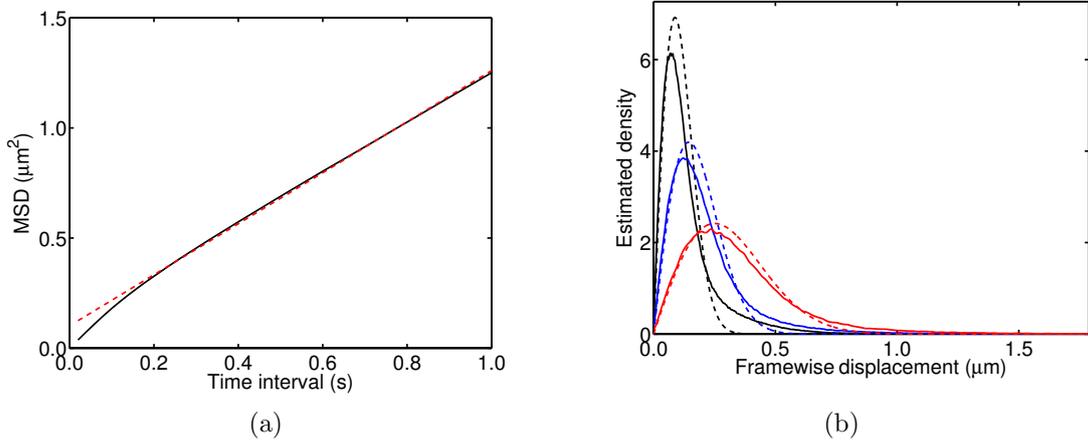


Figure 6.5: *Extracting characteristics of noise from the *R. sphaeroides* non-motile dataset. (a) The observed MSD with linear fit (red dashed line). The slope of the fit is $1.16 \mu\text{m}^2 \text{s}^{-1}$. (b) The estimated pdf of framewise displacements (solid line), with overlaid best fit to the χ distribution (dashed line). Black: $\tau = 0.02 \text{ s}$, blue: $\tau = 0.04 \text{ s}$, red: $\tau = 0.1 \text{ s}$.*

We first carry out a simple analysis of the data to estimate the translational diffusion coefficient, D_t . This is achieved by fitting the MSD with a linear function. According to the theory of translational diffusion, the MSD of a Brownian particle moving in two dimensions is given by [21]

$$\langle \|\mathbf{X}(\tau)\|^2 \rangle = 4D_t\tau, \quad (6.54)$$

where τ is the time step, following the notation used in Chapter 5. We perform this analysis on the non-motile *R. sphaeroides* dataset, since the motion observed in this dataset is solely caused by Brownian buffeting. The results are shown in Figure 6.5(a). The MSD does not appear linear for $\tau \lesssim 0.2 \text{ s}$, so we fit data outside of this region. We attribute this departure from the predicted linear behaviour of a diffusing particle to artefacts generated in the process of tracking. In particular, the tracking algorithm may fill in any missed detections by assuming that the intervening motion took place with a constant velocity, leading to an overestimation of the framewise displacements, as

discussed in Section 4.3. This would lead to a non-linear MSD with a steeper slope, as observed. This effect is only present at short time steps, as the tracking algorithm may only fill in short gaps.

We carry out an additional check that the observed motion is approximately diffusive by plotting the estimated distribution of displacements for various values of τ . We achieve this by downsampling the data. For example, downsampling by a factor of two entails discarding every even-numbered frame. The effective frame interval therefore increases by a factor of two, hence τ is also doubled. As discussed in Section 5.2, the observed displacements should follow a χ distribution, with 2 degrees of freedom. In each dimension, the displacement of a Brownian particle over a time step τ is distributed normally with zero mean and variance equal to $2D_t\tau$. Since the χ distribution gives the distribution of the square root of the sum of squared *standard* normal variates, we are required to rescale the distribution in order to fit to our data. We denote the displacement over a time step of length τ by $Z \equiv \|\mathbf{X}(\tau)\|$. The distribution of these displacements is given by

$$f_Z(z) = \frac{1}{\sqrt{2D_t\tau}} f_\chi\left(\frac{z}{\sqrt{2D_t\tau}}, 2\right), \quad (6.55)$$

where $f_\chi(x, k)$ denotes the pdf of the χ distribution with x as an independent variable and k denoting the number of degrees of freedom. We fit the observed data with this pdf and use the fitting parameter to extract an estimate of D_t . The observed and fitted distributions are shown in Figure 6.5(b) for three values of τ . The fit becomes progressively better as τ increases, although it shows good qualitative agreement for all time lags. The observed distributions further support our hypothesis that the departure from the theoretically predicted behaviour is due to an overestimation of step lengths by the tracking algorithm: the observed distributions show a heavier tail than predicted for all values of τ .

Table 6.1 displays the estimates of the translational diffusion coefficient calculated using

a linear fit to the MSD, fitting the χ distribution to the observed distribution of displacements, and the theoretical value. The latter is calculated by approximating a bacterium as a sphere of radius $1\ \mu\text{m}$. Stokes' Law for a sphere asserts that [21]

$$D_t = \frac{kT}{6\pi\eta r}, \quad (6.56)$$

from which we obtain the estimate $D_t = 0.22\ \mu\text{m}^2\ \text{s}^{-1}$. The various estimates are in good agreement with the theoretical value, particularly considering that the radius varies between cells in the population. For example, recalculating the theoretical value with $r = 0.75\ \mu\text{m}$, which is still a realistic value for a bacterium of slightly smaller size, we obtain $D_t = 0.29\ \mu\text{m}^2\ \text{s}^{-1}$.

Table 6.1: *Theoretical and estimated values of the translational diffusion coefficient. See text for details.*

Method	D_t ($\mu\text{m}^2\ \text{s}^{-1}$)
Theoretical	0.22
MSD linear fit	0.29
χ fit, $\tau = 0.02\ \text{s}$	0.19
χ fit, $\tau = 0.04\ \text{s}$	0.26
χ fit, $\tau = 0.1\ \text{s}$	0.31

6.4.2 Run-only model: comparison with experimental data

Using a similar analysis method as in the previous section, we may estimate the rotational diffusion coefficient, D_r . In this case, we consider the non-chemotactic strain of *R. sphaeroides*. Following appropriate censoring (detailed in Chapter 4), the remaining tracks in this dataset should exhibit running phases only. These tracks are approximately described by the run-only model, described in Section 6.3. The mean squared angle change (MSAC) in this model grows linearly with time, as given by (6.29). Furthermore, we expect that angle changes should follow a wrapped normal distribution

with zero mean and variance equal to the MSAC. Figure 6.6 shows the comparison of the non-chemotactic dataset with the predictions of the model. As for the experimentally observed MSD, the MSAC is not linear for $\tau \lesssim 0.2$ s, but is approximately linear outside of this region. From equation (6.29), we deduce that

$$\langle \phi^2(t) \rangle = 2D_r t. \quad (6.57)$$

Hence we may estimate the rotational diffusion coefficient from the data in Figure 6.6(a), giving $D_r = 0.13 \text{ rad}^2 \text{ s}^{-1}$.

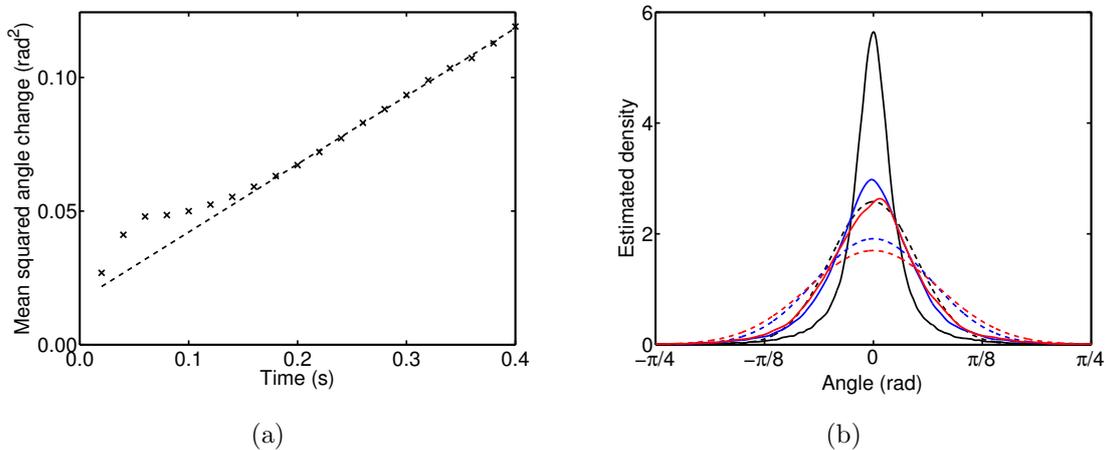


Figure 6.6: *Extracting characteristics of noise from the R. sphaeroides non-chemotactic dataset. (a) The observed MSAC (black crosses) with linear fit (dashed line). The slope of the fit is $0.255 \text{ rad}^2 \text{ s}^{-1}$. (b) The estimated pdf of angle changes (solid line), overlaid with the wrapped normal distribution with variance computed from the data (dashed line). Black: $\tau = 0.02$ s, blue: $\tau = 0.1$ s, red: $\tau = 0.2$ s.*

Figure 6.6(b) shows the observed distribution of angle changes for three values of τ , overlaid with the predicted wrapped normal distribution. There is a significant discrepancy between the observed and predicted distributions. Two explanations are proffered here: firstly, errors in finding the cell centroids of the bacteria, in addition to artefacts generated by the tracking algorithm, may contribute to an under-estimation of framewise angle changes, leading to a more peaked distribution, as observed; secondly, the method of computing ϕ is based on the assumption that framewise angle changes may be used

to approximate changes in the orientation of the cell. We return to the latter point in Section 6.4.3.

The theoretical value of the rotational diffusion coefficient is calculated using Stokes' Law for a spherical object [21], given by

$$D_r = \frac{kT}{8\pi\eta r^3}. \quad (6.58)$$

Approximating a bacterium as a sphere of radius $r = 1 \mu\text{m}$, we obtain the theoretical value $D_r = 0.16 \text{ rad}^2 \text{ s}^{-1}$, in close agreement with the value estimated from the data. We note, however, that the theoretical value is highly sensitive to the value of r ; choosing $r = 0.75 \mu\text{m}$ gives $D_r = 0.39 \text{ rad}^2 \text{ s}^{-1}$. We repeat this calculation in Section 6.4.4 for a bacterium modelled as an ellipsoid.

6.4.3 Run-only model: theoretical results

Figure 6.7(a) shows 20 trajectories simulated by solving equations (6.26) and (6.27) numerically. The mean and standard deviation of the x and y coordinates of an ensemble of particles, given in equations (6.33), (6.34) and (6.35) are shown in Figure 6.7(b). These agree well with the results from simulations (not shown). The estimated pdfs of the x and y coordinates of simulated particles are shown in Figure 6.8. There is no analytic description for the distributions of these quantities.

We commented in Section 6.4.2 that the method we use to measure angle change is not equivalent to the true underlying change in orientation. Figure 6.9(c) illustrates the difference between these two quantities. Figure 6.9(b) shows the MSAC of an ensemble of simulated particles, computed using the two different methods. The solid line is calculated using the true underlying orientation, $\phi(t)$. This line agrees well with equation (6.57). The dashed line represents the *measured* MSAC, calculated by estimating angle

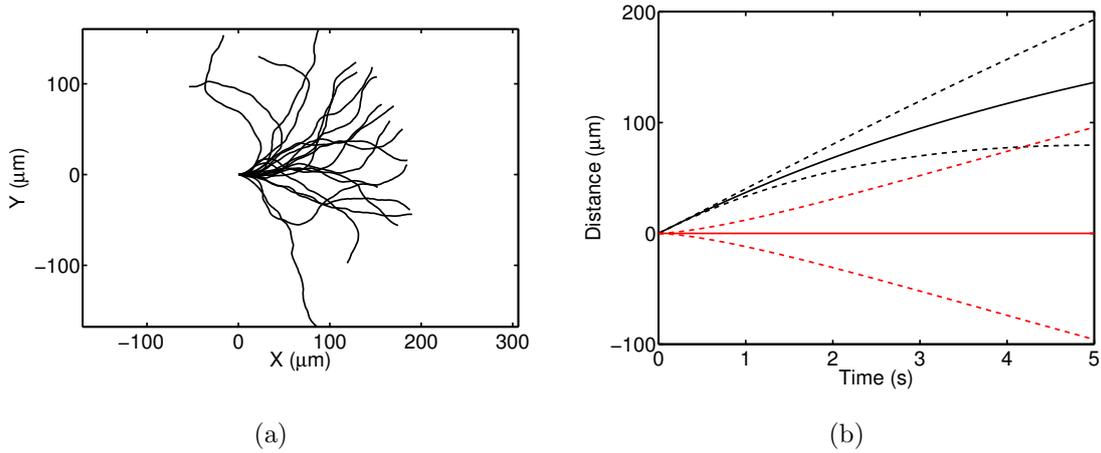


Figure 6.7: (a) 20 sample trajectories computed by solving the run-only model. (b) Mean (solid line) and mean \pm standard deviation (dashed lines) for x (black) and y (red) position of a particle in the run-only model.

changes from the position of the particles, denoted ϕ_{meas} . The measured MSAC also scales linearly with time, but with a constant of proportionality that is 60% of that in the true MSAC. Misuse of the MSAC, by assuming that $\phi(t) \sim \phi_{\text{meas}}$ and using the latter quantity to compute D_r , could therefore lead to a substantial underestimate of the rotational diffusion coefficient.

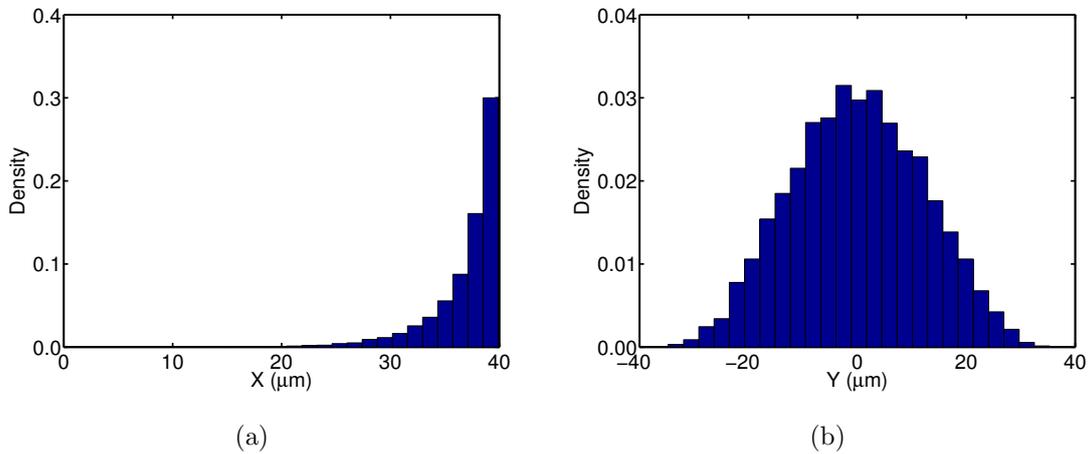


Figure 6.8: Histograms of the x and y position of simulated particles in the run-only model with $\tau = 0.1$ s ((a) and (b), respectively).

Just as for the MSD, it is possible to check not only the MSAC against data, but also the distribution of angle changes. The true distribution of $\phi(t)$ is a wrapped normal, with variance proportional to time. It is not, however, obvious how $\phi_{\text{meas}}(t)$ is distributed. In

order to answer this question, we simulate a particle undergoing a run-only process with rotational diffusion. Figure 6.9(a) shows the observed distribution of the measured angle changes for three different values of τ ; the distribution is indiscernible from the normal distribution in all three cases.

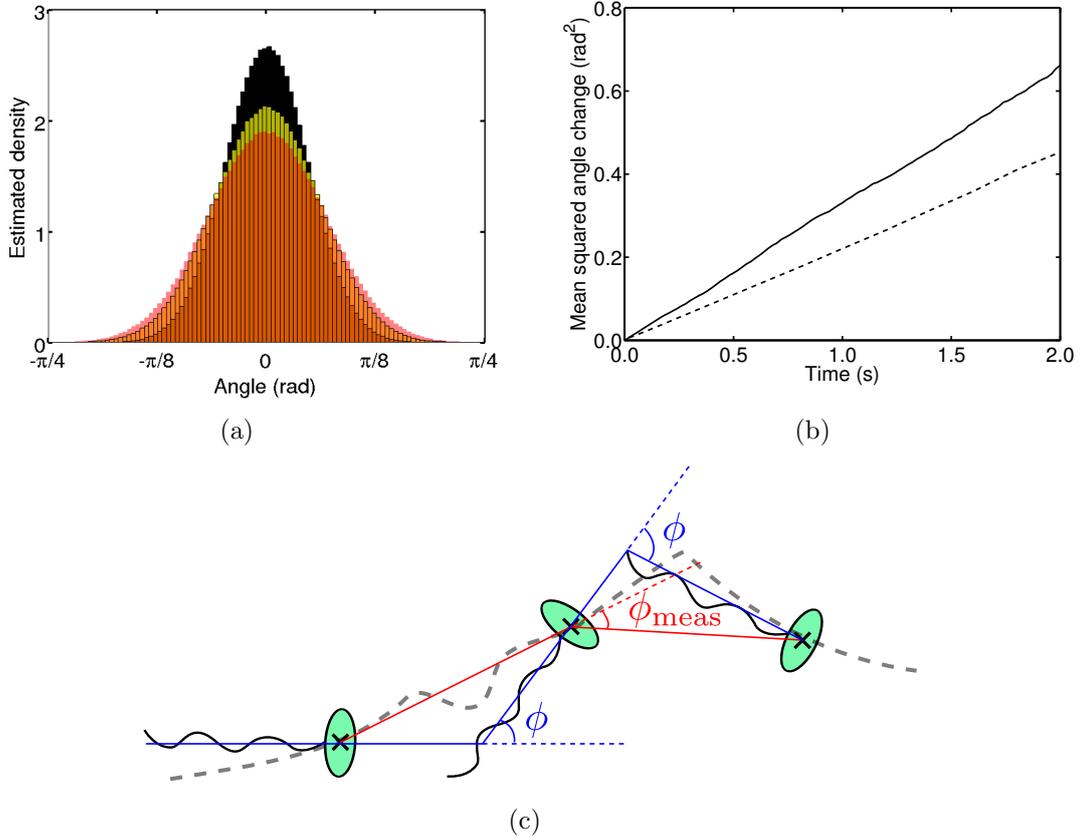


Figure 6.9: (a) Histograms of the measured angle changes, ϕ_{meas} , of simulated particles in the run-only model. (Black) $\tau = 0.1$ s, (yellow) $\tau = 0.16$ s, (red) $\tau = 0.2$ s. (b) The observed MSAC over time for simulated tracks. (Solid line) angle change computed using the true underlying orientation of the simulated particles, (dashed line) angle change computed by calculating the framewise angle change from particle positions. (c) Illustration of the difference between the true underlying angle change, ϕ (blue lines) and the measured angle change, ϕ_{meas} (red lines) of a swimming bacterium. The grey dashed line denotes the true trajectory of the bacterium.

6.4.4 The effect of cell geometry on rotational diffusion

We now consider how the geometry of the cell body affects the role of rotational diffusion on bacterial motility. We consider cells with a constant equatorial radius, $b = 1 \mu\text{m}$. We then vary the length of the axial radius a so that we consider a range of prolate ellipsoids (for which $a > b$). Figure 6.10(a) shows the effect of varying a on the rotational fric-

tional drag coefficient, ζ . The solid line corresponds to the drag coefficient for rotation in the equatorial axis, and the dashed line indicates the drag coefficient for the sphere of equivalent volume. This plot demonstrates that, as the axial ratio, ρ , increases, the difference between the rotational drag coefficient for a sphere and the ellipsoid of equal volume diverge. Ellipsoidal bacteria are therefore more stable to rotational diffusion than spherical bacteria. We illustrate this by repeating the numerical example given in Section 6.4.2, in which we calculated D_r for a spherical bacterium, for an ellipsoidal bacterium with the same volume. Figure 6.10(b) shows the variation of the diffusion coefficient for rotation in the equatorial axis with ρ for an ellipsoid of fixed volume. The values of D_r when $\rho = 1$ are the same as those calculated for a spherical cell in Section 6.4.2. The rotational diffusion coefficient decreases with ρ , again indicating that an ellipsoidal geometry stabilises the cell towards Brownian reorientation.

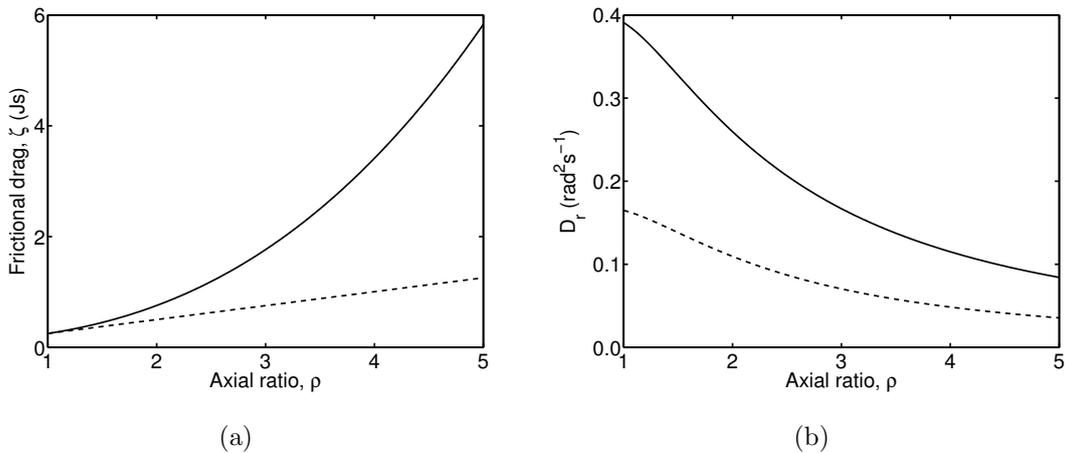


Figure 6.10: (a) The variation of the rotational drag coefficient ζ with ρ for a prolate ellipsoid with equatorial radius of length $1 \mu\text{m}$. The solid line represents the rotational drag coefficient for rotation in the equatorial semi-principal axis, while the dashed line shows the rotational drag coefficient for a sphere of equivalent volume. (b) The variation of the rotational diffusion coefficient for rotation in the equatorial semi-principal axis with ρ for a prolate ellipsoid of fixed volume. The radius of the equivalent sphere is $0.75 \mu\text{m}$ in the case of the solid line, and $1 \mu\text{m}$ in the case of the dashed line.

We further illustrate this point by returning to the calculations of Mitchell [137], who derives equations for the minimum useful swimming speed for bacteria in terms of the size of the bacterium (see Figure 6.2). Mitchell defines a characteristic length, L , over which

the change in the concentration of chemoattractant is sufficiently large to be detected. This is given by

$$L = \sqrt{D_m t}, \quad (6.59)$$

where $D_m \approx 1000 \mu\text{m}^2\text{s}^{-1}$ is the translational diffusion coefficient of a molecule of chemoattractant. Rearranging equation (6.57) for t , we obtain

$$t = \frac{\langle \phi^2 \rangle}{2D_r}. \quad (6.60)$$

The minimum useful speed is given by $v_{\min} = L/t$. Substituting (6.59)-(6.60) into this expression and rearranging yields

$$v_{\min} = \sqrt{\frac{2D_r D_m}{\langle \phi^2 \rangle}}. \quad (6.61)$$

We solve this equation with $\langle \phi^2 \rangle = (\pi/2)^2$, indicating that a cell must travel a distance L before its mean reorientation angle exceeds $\pi/2$. We note that this interpretation is an approximation, as the equality $\langle \phi^2 \rangle = (\pi/2)^2$ does not imply that $\langle |\phi| \rangle = \pi/2$. Figure 6.11(a) shows the variation of v_{\min} with axial ratio, ρ , for cells with various equatorial radii, b . Most wildtype *R. sphaeroides* cells have an equatorial radius in the range $0.5 < b \lesssim 1 \mu\text{m}$ [186, 204]. Figure 6.11(b) shows the mean swimming speed for each track in the *R. sphaeroides* non-chemotactic bulk dataset. It is not possible to draw firm conclusions from Figure 6.11, as we do not have detailed information about the sizes of the tracked cells. Nonetheless, the plots are consistent in the sense that most cells swim sufficiently quickly to meet the minimum useful speed, assuming that $0.5 \mu\text{m} \lesssim b \lesssim 1 \mu\text{m}$ and $\rho \gtrsim 1.5$.

6.4.5 Run-and-stop model: comparison with experimental data

We now compare the predictions of the run-and-stop model with experimental data, in order to assess how well the model captures the observed motile behaviour of *R. sphaeroides*. For this purpose, we use the wildtype bulk dataset, censored as described

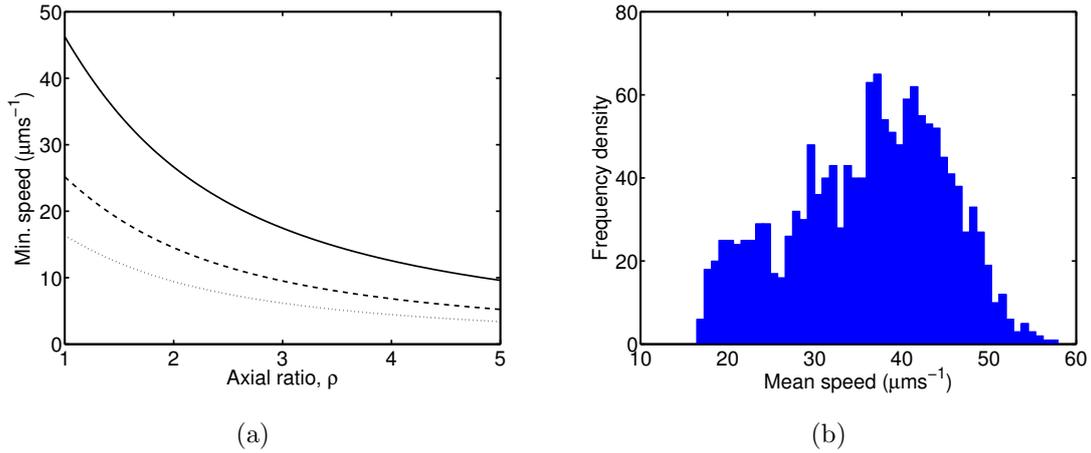


Figure 6.11: (a) The variation of minimum useful speed, v_{min} , with axial ratio for ellipsoidal bacteria with different fixed equatorial radii, $b = 0.5 \mu\text{m}$ (solid line), $b = 0.75 \mu\text{m}$ (dashed line), and $b = 1 \mu\text{m}$ (dotted line). (b) Histogram of the mean speeds for each track in the *R. sphaeroides* non-chemotactic bulk dataset.

in Chapter 4 and analysed using the full HMM method. We focus on the predicted variation of the variance of stopwise angle changes with stop duration. According to the run-and-stop model, the evolution of the orientation angle ϕ during both running and stopping phases is described by equation (6.29), which states that angle changes follow the wrapped normal distribution with zero mean and variance equal to $2D_r\tau$, where τ is the stop duration. The theoretical distribution of stopwise angle changes is shown in Figure 6.12(a) for several different stop durations.

A direct comparison of Figure 6.12(a) with the experimental data requires that we estimate the variance of the observed stopwise angle changes, denoted σ_θ^2 , for a variety of stop durations. This process is complicated by the presence of artefacts from the analysis procedure, as discussed in Section 3.3.3. The problem is summarised in cartoon form in Figure 6.12(b). The observed density of stopwise angle changes (solid line) has a substantial false positive component (dashed) line, which will skew our estimate of the variance. The number of artefacts is greater for short duration stops (see Figure 3.14). Estimation of the exact number and distribution of false positives is not possible, so we instead define an *acceptance region*, $|\theta| > a$, in which we assume that the density of false positives is negligible. This is supported by the simulation study in Chapter 3, in which we show

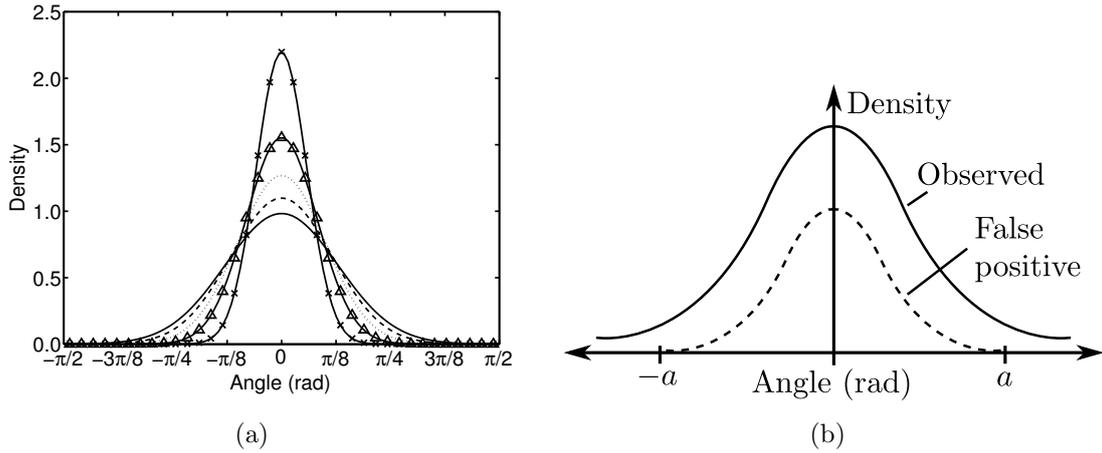


Figure 6.12: (a) The theoretical distribution of stopwise angle changes, as predicted by the run-and-stop model, for stops of 0.1 s (\times), 0.2 s (Δ), 0.3 s (dotted line), 0.4 s (dashed line), and 0.5 s (solid line). (b) An illustration of the problem of false positives due to analysis artefacts in the distribution of stopwise angle changes. The exact form of the dashed line distribution is unknown, but we can estimate the cutoff value, a , reasonably well.

that the density of false positives decreases with increasing stopwise angle change (see Figure 3.13). Based on our earlier simulation study, we choose $a = \pi/2$. We first bin the observed stops by their duration. For each group, we estimate the total density of stopwise angle changes in the acceptance region in two ways:

1. Assume no false positives, by taking the number of angle changes in the acceptance region divided by the total number of angle changes.
2. Use the simulated dataset from Chapter 3 to estimate the false positive level outside of the acceptance region, and hence to estimate the density within the acceptance region, taking false positives into account.

We next use the estimated total density in the acceptance region to compute σ_θ^2 , as described below. Method (1) provides a lower bound on the value of σ_θ^2 , since the presence of false positives outside the acceptance region is ignored. These false positives are clustered around the mean (i.e. zero), hence they artificially reduce the apparent variance. Method (2) constitutes an improved estimate that is corrected in an approximate fashion for false positives; note that this is not an upper bound, nor is it necessarily an accurate estimate, however it is likely to be more accurate than method (1). The estimated proportion of

false positives outside the acceptance region in the simulated dataset is shown in Table 6.2. We use the data simulated with the level of noise given by $D = 0.288 \mu\text{m}^2\text{s}^{-1}$, as this value is in close agreement with the estimates of D_t listed in Table 6.1.

Table 6.2: *The proportion of false positives outside the acceptance region for the simulated dataset in Chapter 3.*

Stop duration (s)	Proportion false positives (HMM speed only)	Proportion false positives (HMM full)
0 – 0.1	0.613	0.658
0.1 – 0.2	0.476	0.407
0.2 – 0.3	0.281	0.232
0.3 – 0.4	0.222	0.191
0.4 – 0.5	0.140	0.106

We now seek an expression linking the total density in the acceptance region to the variance of the wrapped normal distribution with zero mean. Recall that the pdf of this distribution is given by (equation (2.35))

$$f_{\Theta}(\theta; \sigma_{\theta}) = \frac{1}{\sigma_{\theta}\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} e^{-\theta^2/2\sigma_{\theta}^2}.$$

The total density in the acceptance region is then given by

$$F(a) = 2 \sum_{k=-\infty}^{\infty} \frac{1}{\sigma_{\theta}\sqrt{2\pi}} \int_a^{\pi} e^{-\theta^2/2\sigma_{\theta}^2} d\theta. \quad (6.62)$$

Each term in the summation is given by

$$I_k = \frac{1}{\sigma_{\theta}\sqrt{2\pi}} \int_a^{\pi} e^{-\theta^2/2\sigma_{\theta}^2} d\theta = \frac{1}{2} \left[\operatorname{erf} \left(\frac{\pi(2k+1)}{\sigma_{\theta}\sqrt{2}} \right) - \operatorname{erf} \left(\frac{a+2\pi k}{\sigma_{\theta}\sqrt{2}} \right) \right]. \quad (6.63)$$

Substituting (6.63) into (6.62), we obtain

$$F(a) = \sum_{k=-\infty}^{\infty} \left[\operatorname{erf} \left(\frac{\pi(2k+1)}{\sigma_{\theta}\sqrt{2}} \right) - \operatorname{erf} \left(\frac{a+2\pi k}{\sigma_{\theta}\sqrt{2}} \right) \right]. \quad (6.64)$$

Having obtained an estimate for $F(a)$ using either of the methods described above, we use trust-region constrained numerical optimisation [54] (implemented in the Matlab function `fminbnd`) to find σ_θ using equation (6.64). The results of this calculation, along with the theoretically predicted result for a spherical bacterium, are shown in Figure 6.13. The discrepancy between the lower bound estimate of σ_θ^2 and the theoretical value is striking. Even assuming the small cell radius of $0.5\ \mu\text{m}$, the theoretical variance is smaller than the lower bound estimated from the data by a factor of between five and seven. This demonstrates that *R. sphaeroides* does not reorientate passively by rotational Brownian diffusion: the bacterium is too large for this to be an effective mechanism for reorientation. To the best of our knowledge, the validity of the passive reorientation model in *R. sphaeroides* has not previously been tested. We discuss this novel result further in Section 6.5.

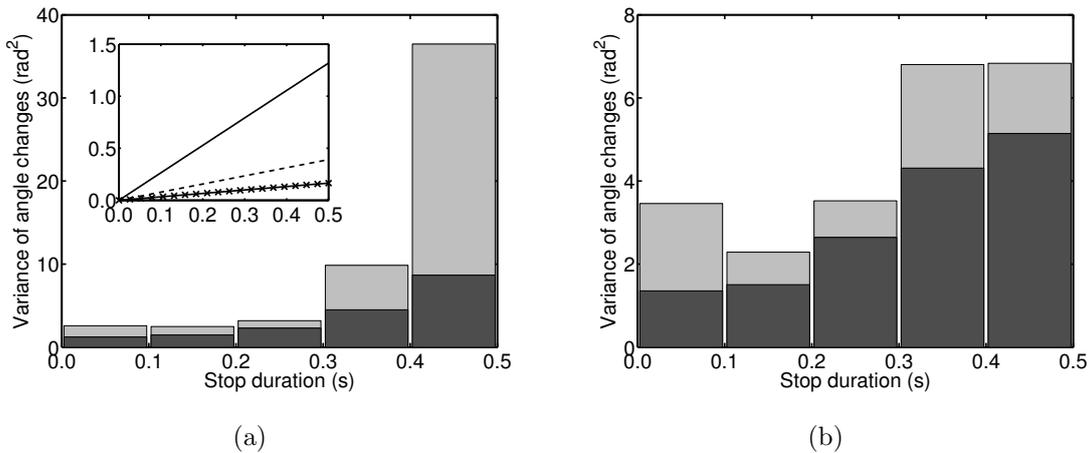


Figure 6.13: The lower bound (black bars) and revised estimates (grey bars) of the variance of stopwise angles, based on analysis of the wildtype *R. sphaeroides* bulk dataset with (a) the speed-only HMM method, (b) the full HMM method. The inset in (a) shows the theoretical variance, as predicted by the run-and-stop model, for a spherical cell with radius $r = 1\ \mu\text{m}$ (solid line), $r = 0.75\ \mu\text{m}$ (dashed line), $r = 0.5\ \mu\text{m}$ (crossed line).

6.4.6 Run-and-stop model: theoretical results

Figure 6.14 shows the analogous plots to those in Figure 6.7 for the run-and-stop model. Whilst the figures are similar, two key differences are notable. Firstly, the trajectories

are of varying lengths in Figure 6.14(a), since stopping phases cause the bacterium to halt transiently. Secondly, the mean and standard deviation of $x(t)$, and the standard deviation of $y(t)$, have lower magnitudes than the run-only analogues. This is as expected: stopping phases cause the cell to halt, during which time it continues to reorientate, but ceases to move. Figure 6.14(b) agrees well with simulated data (not shown).

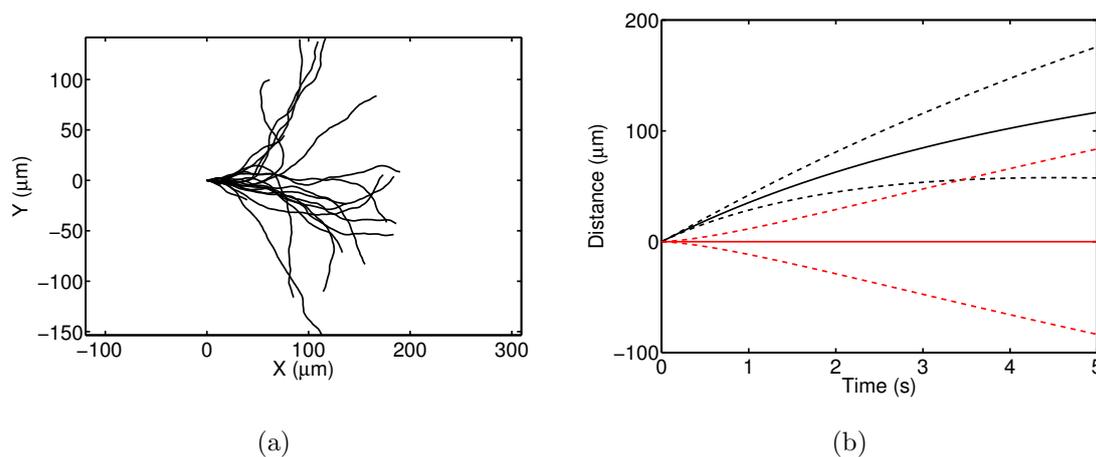


Figure 6.14: (a) 20 sample trajectories computed by solving the run-and-stop model. (b) Mean (solid line) and mean \pm standard deviation (dashed lines) for x (black) and y (red) coordinates of a particle in the run-and-stop model.

6.4.7 Run-and-active-stop model

Figure 6.15 shows the same estimates of the variance of stopwise angle changes, binned by stop duration, as used in Section 6.4.5, overlaid with predictions from the run-and-active-stop model. We model a bacterium as an ellipsoid (dimensions given in figure legend), as this approximately matches the true dimensions of *R. sphaeroides*. The multiplicative factor ω increases the variance of angle changes linearly, as expected. The data are too noisy to permit us to fit ω accurately, but the data suggest that wildtype *R. sphaeroides* reorientate $\sim 5 - 20$ times more rapidly than predicted by the run-and-stop model.

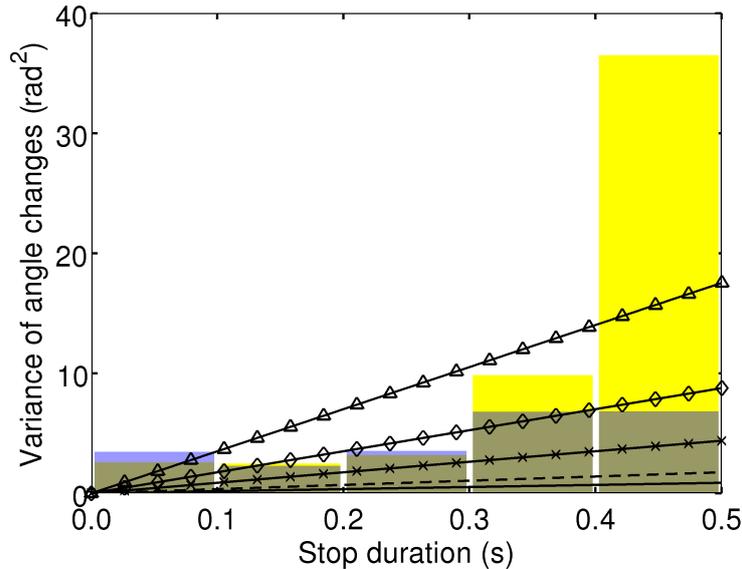


Figure 6.15: Estimates for the variance of stopwise angle changes overlaid with predictions from the run-and-active-stop model. The estimated values are the same as in Figure 6.13: speed-only HMM (yellow), and full HMM (blue). Theoretical results are for an ellipsoidal bacterium, with $b = 0.5 \mu\text{m}$, $\rho = 2$, and $\omega = 1$ (solid line), $\omega = 2$ (dashed line), $\omega = 5$ (\times), $\omega = 10$ (\diamond), and $\omega = 20$ (\triangle).

6.5 Discussion and conclusion

In this chapter, we have considered the role of Brownian diffusion in flagellar-mediated bacterial motility. We chose to focus on rotational diffusion, ignoring translational diffusion, since the former has a more significant effect. The motivation for this work is the need to reconcile the idealised VJ model, considered in Chapter 5, with the immediately apparent departure from the model in experimentally observed tracks. An important example of the outcome of studies of noise in bacterial motility is work by Mitchell [137] and Dusenbery [65], motivated by earlier work by Purcell [163]. These authors use a description of rotational diffusion to consider the physical size and swimming speed constraints present in chemotactic bacteria.

We used the overdamped rotational Langevin equation, in the context of a self-propelled particle model, to model rotational diffusion. The use of this framework is motivated by a related study by Hagen *et al.* [83]. We stated three minimal models for bacterial motility in two dimensions. The justification for only considering motion in two dimen-

sions is that our experimental datasets only capture bacteria moving in a plane, to a first approximation. These three models were chosen to describe different aspects of bacterial motility. The run-only model is chosen to describe the motion of a *R. sphaeroides* bacterium from the non-chemotactic strain. The run-and-stop and run-and-active-stop models are two closely-related variants, both of which aim to describe the motion of a wildtype *R. sphaeroides* bacterium.

We derived expressions for the first and second position moments in two of these models, demonstrating how the SDE modelling framework is a convenient representation for the random phenomenon of bacterial motility. Furthermore, we tested various fundamental predictions of the models directly against the experimental data, analysed using the methods presented in Chapter 4. Whilst examples exist of studies of the random motion of human dermal keratinocytes [179], and the amoeba *Dictyostelium discoideum* [31], to our knowledge this is the first example of such a comparative study in bacteria. Furthermore, in [31], the authors take a more generalised approach to their analysis, allowing departures from the standard Langevin formulation. It is not apparent how the more general approach relates to the underlying physical theory of motion, or the biological processes involved. In contrast, we robustly test the predictions of our minimal models.

Our results show good agreement between the theoretical and experimentally observed translational diffusion coefficients, extracted by analysing the non-motile strain. In particular, we show that the *distribution* of displacements match the theoretical predictions well; more commonly this test is ignored in favour of testing the MSD for linearity (see for example [53, 200]). When data are scarce, this is an acceptable compromise, however our analysis shows that a full comparison of the distribution, when possible, is a useful test.

Consideration of the effects of the geometry of bacteria in our model led to the conclu-

sion that the observed dimensions and shape of *R. sphaeroides* in our experiments is compatible with the mean speeds observed in the tracks. Agreement between the non-chemotactic dataset and the predictions of the run-only model was also reasonable based on the MSAC. The rotational diffusion coefficient estimated in this way agreed with the theoretically predicted value. However, we identified significant departures from the predicted distribution of angle changes. We attributed this discrepancy to errors in the image analysis and tracking procedures, in addition to the method used to estimate framewise angle changes. Expanding upon this latter point, we demonstrated the difference between the true orientation of a bacterium, which is not observable, and an approximation that is computed from the tracking data. Assuming that these two quantities are equal leads us to underestimate the rotational diffusion coefficient.

In order to compare the predictions of the two models incorporating stochastic stopping events with our wildtype data, we first needed to circumnavigate the issue of artefacts in the analysed data. The simulation study carried out in Chapter 3 was of great utility here, as it gave us an estimate of the level of false positives in the dataset. Guided by these estimates, we were able to demonstrate the important biological result that *R. sphaeroides* cannot reorientate purely passively by rotational diffusion. This was clear from the strong incompatibility between the predictions of the run-and-stop model and the observed data. However, we were subsequently unable to estimate accurately the stochastic rotational force, ω , to explain the observed data in terms of the run-and-active-stop model. Note that we did not include the additional frictional effects of the flagellum in the current work, which may be expected to reduce the reorientation rate even further [137]. Furthermore, we assumed a spherical cell body in Section 6.4.5; an ellipsoidal geometry would also further reduce the reorientation rate, as demonstrated in Section 6.4.4. Finally, using the run-and-active-stop model, we tentatively speculated that a random reorientational force may explain the observed data.

The issue of passive orientation in *R. sphaeroides* and related bacterial species has been discussed before in the literature, with no strong consensus. In one theoretical study, Mitchell [137] describes *R. sphaeroides* undergoing passive reorientation during a stopping phase. However, Mitchell and Kogure [138], referring to the run-and-stop model in a slightly later study, note that "... this method, while previously attributed to *R. sphaeroides*, is currently not reported for any species". No robust analysis of data has previously been carried out with reference to a well-stated model. We believe that this is an important and novel contribution to the field. The hypothesis that *R. sphaeroides* exhibits some form of active reorientation mechanism was briefly discussed by Armitage *et al.* [12]. The authors present a study in which they observed the various flagellar conformations of swimming and stopping *R. sphaeroides* cells. In particular, a conformation was observed in stopped cells in which "once coiled against the cell body, the flagellum often slowly rotated". In the context of the present work, we speculate that such a low-frequency movement of the flagellum during a stopping phase may lead to enhanced rotational diffusion. A further speculative mechanism involves the elastic relaxation of the flagellum upon entering a stop phase [75].

We contend that the present work represents an important contribution to the field of bacterial motility, as it demonstrates the utility of minimal models and comparison with experimental data in answering questions of fundamental biological interest. There are many opportunities for further investigation. We showed several departures from the theoretical predictions, particularly when comparing the non-chemotactic dataset with the run-only model. We suggested that these differences may be explained by artefacts introduced in the image analysis and tracking processes, or due to the method of calculating the angle change. Further studies could test whether these departures contain useful information about the biological processes underlying bacterial motility, or are indeed due to artefacts introduced in the various analysis stages. A related consideration is the variation in swimming speeds observed in a sample of bacteria. Several studies have

noted the highly heterogeneous nature of bacterial populations in terms of swimming speeds [12, 140, 157]. In our models, we fixed the swimming speed; an obvious extension is to replace this with a fluctuating quantity. Random velocities have been considered in the context of random walks [216], but have not been coupled with a model of rotational diffusion. It is possible that this may further explain the departure of the experimental results from the run-only model predictions. It is, however, unclear how the random velocities should be distributed: we measured a *population*-level distribution of velocities in Chapter 4, but a distribution on an *individual*-level would be required for this extension.

Chapter 7

Discussion and conclusions

In this thesis, we have investigated various aspects of the phenomenon of planktonic bacterial motility using mathematical modelling and analysis and comparison with experimental data. The motility of planktonic bacteria is a widely-studied process, not least because the process is an important paradigm in the field of bacteriology and has major implications in many health and industrial applications. However, due to the complexity of the phenomenon, the fact that it occurs on multiple spatial and temporal scales, and the complications involved in studying the process experimentally, there remain many open problems relating to bacterial taxis, spanning a wide range of research fields.

The experimental method that is used as a source of comparison throughout this thesis is the tracking of free-swimming bacteria under a microscope. The data generated by this technique contain a great deal of information about the motile behaviour of the bacteria being observed. This method is also less invasive than the popular technique of imaging tethered cells. We initially summarised the experimental tracking protocols described in the literature, and showed that there is currently no well-established, high-throughput approach. Since current methods are generally laborious or costly, bacterial tracking is not currently considered a standard experimental technique in this field. The availability of a non-invasive, inexpensive, robust experimental protocol that can quickly generate

large quantities of data would therefore represent an important advance. Our collaborators have recently developed a new tracking protocol that combines readily available microscopy techniques with a recently-developed software tracking algorithm [209], which promises to fulfill these requirements. Analysis of the tracks produced is, however, a challenge, since the datasets produced are approximately an order of magnitude larger than those generated by most existing methods.

Since many bacterial species, for example the model bacterium *R. sphaeroides*, exhibit run-and-stop motility, as discussed in Chapter 2, it is desirable to identify stopping phases in experimental tracking data. This permits us to extract important quantities of interest, such as the distribution of running and stopping durations, running speeds, and angle changes over the course of a stopping phase. The process of identifying stopping phases is complicated by the presence of various sources of noise in the experimental tracks, such as inaccuracies in estimating the location of the cell centre of a bacterium under the microscope. In Chapter 3, we approached this problem, presenting two novel analysis methods to identify stopping phases, based on the HMM. Both methods exploit the availability of mutant strains of bacteria to gain prior information on the appearance of stopping and running phases. Existing analysis methods are all heuristic and specific to the species being studied and experimental setup. In contrast, the methods described here are applicable to a wide range of species and experimental conditions with no need for modification, as the analysis is supported by reference data from mutant strains.

We next compared the performance of the new analysis methods with an established one in a systematic computational simulation study. This has not been attempted before, and represents an important development in the field. In order to create simulated tracks, we assumed that bacteria undergo a VJ process, as detailed in Chapter 2, with added noise. The simulation study demonstrated that the novel HMM-based methods perform substantially better than a heuristic approach for a wide range of added noise levels.

Furthermore, the study allowed us to quantify the extent to which bias is introduced in the analysed data due to the presence of false positives. This is an important result, as it allows us to state later conclusions with greater confidence, since we have an understanding of the expected level of bias.

In Chapter 4, we proceeded to analyse real experimental data, obtained for the model bacterium *R. sphaeroides* using the new experimental protocol, with the novel analysis methods from Chapter 3. In this chapter, we discuss several additional factors that must be considered when working with real data. An important example is the need to filter various tracks out of the dataset, as they arise from non-motile or defective cells. We present several methods for achieving this goal, accompanied with a discussion of how to determine the various parameters involved. The results from the analysis of the censored experimental datasets are very promising: the identified stopping phases were checked manually over many tracks and found to agree with the best ‘by-eye’ assessment of the tracking data. We also showed tracks where the analysis process had failed, and verified that these constitute a minority of the total dataset. Using analysed tracks, we finally presented several novel biological results, including the observed distribution of angle changes over the course of a stopping phase. Furthermore, we quantified the bias towards swimming in an arc for bacteria swimming at a surface, which has important biological implications. The work in Chapter 4 represents an important advance for two reasons. First, we have demonstrated that we are able to gain novel biological insight by the appropriate analysis of tracking data. More importantly this work represents a proof of the principle behind the new tracking protocol, paving the way for a host of further studies on bacterial motility (see Section 7.1).

The new datasets analysed in Chapter 4 are not necessarily optimised in terms of the experimental protocol, as we noted in Section 4.3. This provides a motivation for applying mathematical modelling to the problem of optimising the experimental design. In Chap-

ter 5, we investigated the effect of the sampling frequency of the microscope camera on the tracking data, by assuming that tracks arise from an underlying VJ process. This investigation is an example of the application of mathematical modelling to study a process that would be expensive and time-consuming to probe experimentally. Furthermore, the conclusions that we drew from the mathematical study of this process are more general than those arising from an experimental study would be. The role of sampling frequency on observations of motile bacteria remains a relatively open research area; Codling and Hill [50] have considered the problem previously, but did not include a stopping phase, as we did. Furthermore, in Section 5.4, we derived a novel analytic result for the observed distribution of framewise speeds and angle changes in the limit of rapid sampling. This result is novel; the study in [50] is based on empirical assessment of simulation results, rather than an analytic approach. We next attempted to take an alternative approach to inferring the distribution of stopwise angle changes, by inverting our previous analytic solution. We showed that this approach is successful when applied to noise-free simulated tracks, but that it fails completely when noise is included.

The limited applicability of our analytic work in Chapter 5 in the presence of noise provides the motivation for Chapter 6, in which we considered the role of Brownian buffeting in the phenomenon of bacterial taxis. We selected this particular type of noise for further investigation because it is the best understood; there are many other possible sources of noise in the experimental and analysis protocols, but the statistical properties of these fluctuations have not been studied. Furthermore, rotational Brownian motion has a vital effect on bacteria, as it imposes a lower limit on the speed at which bacteria must swim in order to benefit from their motility, as studied previously by Mitchell [137]. In this final chapter, we posed three simple models for the effect of Brownian motion on a motile bacterium. We described how the models are related to the VJ process used in Chapters 3-5 and made testable predictions. The predictions of the models were then compared with various aspects of the experimental data, such as the distribution of observed framewise

speeds, to assess how well the observed data agree with the models. To our knowledge, no studies exist in which simple model predictions relating to Brownian buffeting in bacteria have been tested against experimental data. Agreement with predictions from the theory of translational diffusion was good (see Section 6.4.1). In contrast, in Section 6.4.2, significant departures from the predictions of the rotational diffusion were observed. We proceeded to show in Section 6.4.3 that these may arise from an easily overlooked difference between the mathematical description of the orientation of a bacterium in the model, and the estimation of the same quantity from the experimental data.

An important conclusion drawn from Chapter 6 was that *R. sphaeroides* cannot reorientate in a passive fashion, mediated solely by rotational diffusion. In order to compare our model predictions with experimental observations in this case, we needed to revisit the results of the simulation study in Chapter 3, to assess the extent of bias in the distribution of experimentally observed stopwise angle changes. This is an important demonstration of the importance of fully characterising the analysis methods. The ensuing comparison in Section 6.4.5 is a novel result: studies have suggested that rotational diffusion may occur too slowly to reorientate a bacterium such as *R. sphaeroides* [138, 12], but this is the first time that the theory has been tested with real tracking data. A simple model was finally proposed to account for the observed data.

7.1 Further work

As stated previously, there are many open problems in the field of bacterial motility. The research in this thesis has made substantial progress towards filling some of these gaps, but it has equally opened many more avenues of enquiry, some of which are discussed here.

In Chapter 3, we noted that the need to pool all tracks in a dataset for the purposes of the

analysis stems from the short duration of each track. Any attempt to analyse individual tracks is unlikely to be successful, since many do not contain even a single stopping phase. This restriction would be lifted if data were made available in which tracks are of greater mean duration. Whilst traditional single cell tracking approaches produce this type of data, they typically produce smaller datasets, with greater potential for experimental bias (see Section 3.4). Recently, promising developments have been made in the field of digital holographic microscopy, for example, that may make sufficiently long tracks available in sufficient quantities [217]. These tracks may then be amenable to individual analysis, in order to infer the transition rates in each individual bacterium.

A related extension is to tracking in three spatial dimensions. The current protocol is limited to imaging in a plane, hence we are unable to obtain information about the depth of the observed bacteria. Hill and Häder [88] discusses the effect of projecting tracks from algae onto a plane, concluding that the effect is not of great significance for his application. Nonetheless, the HMM-based analysis methods may be readily modified to include three-dimensional data, which could also lead to longer tracks in the bulk dataset, as cells will not swim out of focus so often. Again, digital holographic microscopy is a promising method for acquiring such tracks.

A further possibility identified in Chapter 3 is to consider further representations of the tracking data, for example track curvature [159], or a recently proposed measure called spatiotemporal entropy [168].

The analysis of experimental data carried out in Chapter 4 suggested several possible modifications to improve the experimental approach. One important example is the possibility of imaging at a lower magnification, in order to visualise more bacteria in the field of view at any one time. There is a compromise to be found between extracting detailed microscopic trajectories (which require higher magnification) and maximising the number

and duration of tracks produced. The density of bacteria also affect the performance of the tracker; an upper limit remains to be determined.

A natural extension to Chapter 5 is a related question on the role of measurement error in the sampling process. Although we discounted noise for the purposes of the present study, a simple model of measurement error, as introduced by the imaging and object detection stage, would produce some valuable guidelines relating to the optimal sampling frequency. For example, as the sampling frequency increases, the observed track is expected to become more jagged, as the effect of noise plays an increasingly significant role. Thus we expect there to be some maximum recommended sampling rate, which depends on the extent of measurement error. A further investigation could predict the optimal magnification level by relating this property to the measurement error. Both suggestions are fine examples of the use of mathematical modelling to address questions that are experimentally inaccessible or costly.

We finally suggest an experimental study that would be ideally suited to the analysis methods developed in Chapter 3. As we have discussed, Chapter 4 represents a crucial proof of principle for the new, high-throughput tracking protocol. For example, a study of the effect of varying the viscosity or viscoelasticity of the surrounding medium on bacterial swimming may yield an insight into the way in which bacteria have evolved to occupy certain environmental niches. This study could be carried out across multiple species of bacteria. Providing that mutant strains are available (and, in some cases, even if they are not: see Section 4.3), the analysis methods are applicable to any bacterial species whose motion is well described as a run-and-stop VJ process. Such an investigation could probe the fundamental evolutionary differences between enteric bacteria, such as *E. coli*, and marine bacteria, such as *R. sphaeroides*.

Appendix A

Mathematical methods

A.1 Algorithms for simulating random walks

In this section, we provide details of the algorithms used to simulate the random walks shown in Figure 2.2. In all cases, we use the same constant step length, whose absolute value has no effect other than to scale the tracks. Similarly, the time step is constant across all simulations and does not affect the appearance of the figure. A full description of the stochastic simulation algorithm used to generate realisations of the velocity jump (VJ) process throughout the rest of the thesis is given in Section 2.5. In all cases, the algorithm begins with a single particle located at the origin.

Unbiased, uncorrelated on-lattice random walk

At each time step, the particle moves one step length up, down, left or right with equal probability.

Unbiased, uncorrelated off-lattice random walk

At each time step, the particle moves moves step length in a direction drawn from the wrapped uniform distribution (see Section 2.4 for details).

Biased on-lattice random walk

In this model, motion is biased towards the right with a strength quantified by the parameter $\epsilon_{\text{Bon}} \in [0, 0.75]$. At each time step, the particle moves one step length right with probability $1/4 + \epsilon_{\text{Bon}}$, and up, down or left with probability $1/4 - \epsilon_{\text{Bon}}/3$. We use $\epsilon_{\text{Bon}} = 0.2$ to generate the figure.

Biased off-lattice random walk

For this model of motion we use a model of sinusoidal reorientation proposed by Hill and Häder [88]. The strength of the bias is quantified by the parameter ϵ_{Boff} . The particle is biased towards rightward motion, in which direction the angle is defined as zero. At each time step, the particle moves one step length in a direction drawn from the von Mises distribution (see Section 2.4 for details) with a concentration parameter of 2 and a mean of $-\epsilon_{\text{Boff}} \sin(\theta)$, where θ is the previous direction of travel. Thus, the particle has a tendency to orient itself towards the right. We use $\epsilon_{\text{Boff}} = 0.1$ to generate the figure.

Correlated on-lattice random walk

In this model, the particle is restricted to move on a lattice by only permitting motion parallel to the x and y axes. The extent of correlation between successive time steps is quantified by the parameter ϵ_{Con} . The particle is initially travelling rightward. At each time step, the particle continues moving in the same direction with probability $1/4 + \epsilon_{\text{Con}}$, and turns through an angle of $-\pi$, $-\pi/2$, or $\pi/2$ with probability $1/4 - \epsilon_{\text{Con}}/3$. We use $\epsilon_{\text{Con}} = 0.2$ to generate the figure.

Correlated off-lattice random walk

In this model, the initial direction of travel is chosen from the wrapped uniform distribution. At each time step, the particle moves one step length in a direction given by rotating through an angle drawn from the von Mises distribution with a concentration

parameter of 2 and zero mean. Thus each new angle is chosen relative to the previous direction, which introduces correlation, quantified by the concentration parameter of the von Mises distribution.

A.2 Derivation of the one-dimensional Telegraph equation

We now provide full details of the steps required to derive the Telegraph equation, given in Section 2.2.1 by equation 2.4. We start with equations (2.2) and (2.3) in the same section. Dropping the explicit dependence of p^\pm on x and t for brevity, and expanding both equations as Taylor series in the small parameters τ and δ , we obtain

$$p^+ + \tau \frac{\partial p^+}{\partial t} + O(\tau^2) = \lambda\tau \left(p^- - \delta \frac{\partial p^-}{\partial x} + O(\delta^2) \right) + (1 - \lambda\tau) \left(p^+ - \delta \frac{\partial p^+}{\partial x} + O(\delta^2) \right), \quad (\text{A.1})$$

$$p^- + \tau \frac{\partial p^-}{\partial t} + O(\tau^2) = \lambda\tau \left(p^+ + \delta \frac{\partial p^+}{\partial x} + O(\delta^2) \right) + (1 - \lambda\tau) \left(p^- + \delta \frac{\partial p^-}{\partial x} + O(\delta^2) \right). \quad (\text{A.2})$$

Rearranging gives

$$\frac{\partial p^+}{\partial t} = \lambda(p^- - p^+) + \lambda\delta \left(\frac{\partial p^+}{\partial x} - \frac{\partial p^-}{\partial x} \right) - c \frac{\partial p^+}{\partial x} + O(\delta) + O(\delta^2) + O(\tau), \quad (\text{A.3})$$

$$\frac{\partial p^-}{\partial t} = \lambda(p^+ - p^-) + \lambda\delta \left(\frac{\partial p^+}{\partial x} - \frac{\partial p^-}{\partial x} \right) + c \frac{\partial p^-}{\partial x} + O(\delta) + O(\delta^2) + O(\tau), \quad (\text{A.4})$$

where we have used the fact that $\delta/\tau = c$. We now take the limit $\tau, \delta \rightarrow 0$, such that the ratio $\delta/\tau = c$ remains constant, to obtain

$$\frac{\partial p^+}{\partial t} = \lambda(p^- - p^+) - c \frac{\partial p^+}{\partial x}, \quad (\text{A.5})$$

$$\frac{\partial p^-}{\partial t} = \lambda(p^+ - p^-) + c \frac{\partial p^-}{\partial x}. \quad (\text{A.6})$$

Equations (A.5)-(A.6) describe the macroscopic evolution of the number density of right- and left-moving particles. The interpretation is straightforward: λ is the rate of switching directions, and hence states, which leads to the reaction-like terms on the right-hand side. The remaining terms describe advection due to the movement of the particles. These equations can be combined into a single equation by defining two new variables, $P(x, t) = p^+ + p^-$ and $j(x, t) = p^- - p^+$. These represent, respectively, the combined population of left- and right- moving particles, and the net leftward flux of particles. Summing equations (A.5)-(A.6) we obtain

$$\frac{\partial P}{\partial t} = c \frac{\partial j}{\partial x}, \quad (\text{A.7})$$

which we may differentiate to obtain

$$\frac{\partial^2 P}{\partial t^2} = c \frac{\partial^2 j}{\partial t \partial x}. \quad (\text{A.8})$$

Subtracting equation (A.5) from (A.6) gives

$$\frac{\partial j}{\partial t} = c \frac{\partial P}{\partial x} - 2\lambda j, \quad (\text{A.9})$$

which we may differentiate to obtain

$$\frac{\partial^2 j}{\partial x \partial t} = c \frac{\partial^2 P}{\partial x^2} - 2\lambda \frac{\partial j}{\partial x}. \quad (\text{A.10})$$

Substituting equation (A.10) into (A.8) and using (A.7) gives

A.3 Deriving expressions for the spatial moments of the general velocity jump process

In Section 2.3 we presented expressions for the mean squared displacement (MSD) of the run-only and run-and-stop VJ processes, as derived by [146]. We repeat the authors' analysis here to illustrate the steps required. The original paper contains only scant details of the mathematical analysis. Codling fills in the gaps in his thesis [47]; we combine and adapt these analyses.

A.3.1 Run-only velocity jump process

We take equation (2.19) as a starting point, repeated here for convenience:

$$\frac{\partial p}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} p = -\lambda p + \lambda \int T(\mathbf{v}, \mathbf{v}') p(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}'. \quad (2.19)$$

We assume that p has compact support in \mathbf{x} and \mathbf{v} for finite t , such that there exist regions $V, V' \in \mathbb{R}^d$ for which $p(\mathbf{x}, \mathbf{v}, t) = 0 \forall \mathbf{x} \notin V, \mathbf{v} \notin V'$. This assumption expresses the requirement that all particles must have a finite velocity and be located within a finite region of space, which is valid for all physically realistic distributions of particles. All vector integrals may thus be considered over regions V and V' , rather than \mathbb{R}^d . This will allow us to apply the Divergence Theorem later, which requires an integral over a compact space.

The total number of runners is defined as

$$N_p = \iint p(\mathbf{x}, \mathbf{v}, t) d\mathbf{v} d\mathbf{x}. \quad (\text{A.11})$$

Particle density is conserved as there are no birth or death processes, therefore N_p is constant and equal to the initial total density of runners.

The second position moment of p , equivalent to the MSD, is defined by

$$\mathcal{D}_p^2(t) = \frac{1}{N_p} \iint \|\mathbf{x}\|^2 p(\mathbf{x}, \mathbf{v}, t) \, d\mathbf{v} \, d\mathbf{x}, \quad (\text{A.12})$$

where we divide by N_p in order to normalise p . We may also define the mixed velocity-position moment and second speed moment (the mean squared speed) in a similar manner:

$$\mathcal{B}_p(t) = \frac{1}{N_p} \iint (\mathbf{x} \cdot \mathbf{v}) p(\mathbf{x}, \mathbf{v}, t) \, d\mathbf{v} \, d\mathbf{x}, \quad (\text{A.13})$$

$$\mathcal{S}_p^2(t) = \frac{1}{N_p} \iint \|\mathbf{v}\|^2 p(\mathbf{x}, \mathbf{v}, t) \, d\mathbf{v} \, d\mathbf{x}. \quad (\text{A.14})$$

Multiplying the transport equation (2.19) by $\|\mathbf{x}\|^2$ and integrating over \mathbf{x} and \mathbf{v} , we obtain

$$\begin{aligned} \iint \|\mathbf{x}\|^2 \frac{\partial p}{\partial t} \, d\mathbf{x} \, d\mathbf{v} + \iint \|\mathbf{x}\|^2 \mathbf{v} \cdot \nabla_{\mathbf{x}} p \, d\mathbf{x} \, d\mathbf{v} = \\ - \lambda N_p \mathcal{D}_p^2 + \lambda \iiint \|\mathbf{x}\|^2 T(\mathbf{v}, \mathbf{v}') p(\mathbf{x}, \mathbf{v}', t) \, d\mathbf{v}' \, d\mathbf{x} \, d\mathbf{v}. \end{aligned} \quad (\text{A.15})$$

The first term in this equation may be simplified by recognising that \mathbf{x} and \mathbf{v} are independent of t , hence using equation (A.12) we obtain

$$\iint \|\mathbf{x}\|^2 \frac{\partial p}{\partial t} \, d\mathbf{x} \, d\mathbf{v} = \frac{dN_p \mathcal{D}_p^2}{dt}. \quad (\text{A.16})$$

Rewriting the integrand in the second term and applying the Divergence Theorem gives

$$\iint \|\mathbf{x}\|^2 \mathbf{v} \cdot \nabla_{\mathbf{x}} p \, d\mathbf{x} \, d\mathbf{v} = -2N_p \mathcal{B}_p. \quad (\text{A.17})$$

We change the order of integration in the final integral in equation (A.15) using Fubini's

theorem [162] and use the normalisation property (2.13) to obtain

$$\iiint \|\mathbf{x}\|^2 T(\mathbf{v}, \mathbf{v}') p(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}' d\mathbf{x} d\mathbf{v} = N_p \mathcal{D}_p^2. \quad (\text{A.18})$$

Combining equations (A.16), (A.17) and (A.18), we obtain

$$\frac{d\mathcal{D}_p^2(t)}{dt} = 2\mathcal{B}_p(t). \quad (\text{A.19})$$

Multiplying equation (2.19) by $\mathbf{x} \cdot \mathbf{v}$ and integrating over \mathbf{x} and \mathbf{v} leads to

$$\begin{aligned} \frac{dN_p \mathcal{B}_p}{dt} + \iint (\mathbf{x} \cdot \mathbf{v}) \mathbf{v} \cdot \nabla_{\mathbf{x}} p d\mathbf{x} d\mathbf{v} = \\ - \lambda N_p \mathcal{B}_p + \lambda \iiint (\mathbf{x} \cdot \mathbf{v}) T(\mathbf{v}, \mathbf{v}') p(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}' d\mathbf{x} d\mathbf{v}. \end{aligned} \quad (\text{A.20})$$

Rewriting the left-hand integrand and applying the Divergence Theorem gives

$$\iint (\mathbf{x} \cdot \mathbf{v}) \mathbf{v} \cdot \nabla_{\mathbf{x}} p = -N_p \mathcal{S}_p^2. \quad (\text{A.21})$$

Changing the order of integration in the right-hand integral, we have

$$\iiint (\mathbf{x} \cdot \mathbf{v}) T(\mathbf{v}, \mathbf{v}') p(\mathbf{x}, \mathbf{v}', t) d\mathbf{v}' d\mathbf{x} d\mathbf{v} = \iint p(\mathbf{x}, \mathbf{v}', t) \mathbf{x} \cdot \bar{\mathbf{v}} d\mathbf{x} d\mathbf{v}', \quad (\text{A.22})$$

where the mean vector after reorientation, $\bar{\mathbf{v}}$, is defined in equation (2.15). Recall that $\bar{\mathbf{v}}$ may depend upon \mathbf{v}' . In order to make further progress, we must make some additional assumptions about the nature of the reorientation. Following Othmer *et al.* [146], we assume that the speed of a particle does not change with reorientation, only the direction. Furthermore, we assume that the turning kernel depends only upon the magnitude of the angle between \mathbf{v}' and \mathbf{v} , not on their absolute magnitude or direction. Under these assumptions, $\bar{\mathbf{v}}$ will always be parallel to \mathbf{v}' and $\bar{s} = \|\mathbf{v}'\|$. Equation (2.17) therefore

becomes

$$\psi_d = \frac{\|\bar{\mathbf{v}}\|}{\bar{s}} = \frac{\|\bar{\mathbf{v}}\|}{\|\mathbf{v}'\|}, \quad (\text{A.23})$$

hence ψ_d is independent of the direction of \mathbf{v}' . Since $\bar{\mathbf{v}}$ is parallel to \mathbf{v}' , we also have $\bar{\mathbf{v}} = \psi_d \mathbf{v}'$. Given these two assumptions, equation (A.22) may be rewritten:

$$\iint p(\mathbf{x}, \mathbf{v}', t) \mathbf{x} \cdot \bar{\mathbf{v}} \, d\mathbf{x} \, d\mathbf{v}' = \iint p(\mathbf{x}, \mathbf{v}', t) \mathbf{x} \cdot (\psi_d \mathbf{v}') \, d\mathbf{x} \, d\mathbf{v}' = N_p \mathcal{B}_p. \quad (\text{A.24})$$

Combining equations (A.20), (A.21) and (A.24), we obtain

$$\frac{dN_p \mathcal{B}_p}{dt} = N_p \mathcal{S}_p^2 - \lambda(1 - \psi_d) N_p \mathcal{B}_p. \quad (\text{A.25})$$

As particles are assumed to maintain their speed over reorientations, \mathcal{S}_p^2 is constant, so we can solve equation (A.25) subject to initial conditions on \mathcal{B}_p . An obvious choice for initial conditions is to assume that all particles start at the origin, hence $\mathcal{B}_p(0) = 0$. The solution in this case is given by

$$\mathcal{B}_p(t) = \begin{cases} \frac{\mathcal{S}_p^2}{\lambda_0} [1 - e^{-\lambda_0 t}] & \text{for } \psi_d \neq 1, \\ \mathcal{S}_p^2 t & \text{for } \psi_d = 1. \end{cases} \quad (\text{A.26})$$

Substituting this solution into equation (A.19), we solve subject to $\mathcal{D}_p^2(0) = 0$, to obtain the solution given in equation (2.23).

A.3.2 Run-and-stop velocity jump process

We start by considering equations (2.21) and (2.22), again repeated here for convenience:

$$\frac{\partial p}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} p = -\lambda p + \mu \int T(\mathbf{v}, \mathbf{v}') r(\mathbf{x}, \mathbf{v}', t) \, d\mathbf{v}', \quad (\text{2.22})$$

$$\frac{\partial r}{\partial t} = \lambda p - \mu r. \quad (\text{2.21})$$

Using similar mathematical analysis to that employed in the previous section, we can convert these two equations into a system of ordinary differential equations (ODEs). The analysis is complicated by the presence of two species, p and r , leading to a larger system of ODEs than was found in the previous section. Making the same assumptions as those required in the analysis of the run-only velocity jump process, in addition to further reasonable assumptions on the initial conditions, it is possible to solve for the MSD. Here we present a brief outline of the mathematical analysis, again adapted from [146].

Multiplying equations (2.21) and (2.22) by $\|\mathbf{x}\|^2$ and integrating over \mathbf{x} and \mathbf{v} , we obtain

$$\frac{dN_p \mathcal{D}_p^2}{dt} = 2N_p \mathcal{B}_p - \lambda N_p \mathcal{D}_p^2 + \mu N_r \mathcal{D}_r^2, \quad (\text{A.27})$$

$$\frac{dN_r \mathcal{D}_r^2}{dt} = \lambda N_p \mathcal{D}_p^2 - \mu N_r \mathcal{D}_r^2, \quad (\text{A.28})$$

where \mathcal{D}_r^2 is the second position moment of r , defined by

$$\mathcal{D}_r^2(t) = \frac{1}{N_r} \iint \|\mathbf{x}\|^2 r(\mathbf{x}, \mathbf{v}, t) d\mathbf{v} d\mathbf{x}, \quad (\text{A.29})$$

and N_r is the total population of stopped particles, given by

$$N_r = \iint r(\mathbf{x}, \mathbf{v}, t) d\mathbf{v} d\mathbf{x}. \quad (\text{A.30})$$

Multiplying equations (2.21) and (2.22) by $\mathbf{x} \cdot \mathbf{v}$ and integrating over \mathbf{x} and \mathbf{v} leads to

$$\frac{dN_p \mathcal{B}_p}{dt} = N_p \mathcal{S}_p^2 - \lambda N_p \mathcal{B}_p + \mu \psi_d N_r \mathcal{B}_r, \quad (\text{A.31})$$

$$\frac{dN_r \mathcal{B}_r}{dt} = \lambda N_p \mathcal{B}_p - \mu N_r \mathcal{B}_r, \quad (\text{A.32})$$

where \mathcal{B}_r is the mixed velocity-position moment, given by

$$\mathcal{B}_r(t) = \frac{1}{N_r} \iint (\mathbf{x} \cdot \mathbf{v}) r(\mathbf{x}, \mathbf{v}, t) d\mathbf{v} d\mathbf{x}. \quad (\text{A.33})$$

As before, we assume that particles maintain their speed upon reorientation, so that \mathcal{S}_p^2 is constant. We also assume that the populations of running and stopping particles are initially in equilibrium and that the total population remains constant, hence

$$N_p = \frac{\mu N_0}{\lambda + \mu}, \quad (\text{A.34})$$

$$N_r = \frac{\lambda N_0}{\lambda + \mu}, \quad (\text{A.35})$$

where N_0 is the initial number of particles. Under these assumptions, equations (A.27), (A.28), (A.31) and (A.32) constitute a closed linear system. In reality, it is not practical to compute the MSD of running and stopped particles separately; instead the MSD found experimentally is that of the total population. We therefore solve the system of equations for the weighted MSD, defined in equation (2.24). The solution subject to the initial condition $\mathcal{D}^2(0) = 0$ is given in equation (2.25).

A.4 Kernel density estimation

It is often desirable to estimate the underlying probability density function (pdf) of a one-dimensional process based on a finite number of observations. A typical approach involves generating a histogram of the observed data and using this as a piecewise linear approximation to the true underlying pdf. An alternative approach involves fitting the histogram with an analytic function, and using this as the approximation. Both methods suffer from the necessity of choosing a histogram bin width, which can have a strong effect on the resulting estimate. In a related method, an alternative estimate is obtained by assuming that each datum represents a Gaussian function with a mean at the observed value, and a common width amongst all of the observations. The normalised sum of these Gaussian functions then represents an estimate of the underlying pdf. Again, a width must be specified, which may significantly influence the estimated pdf.

The kernel density estimate (KDE) framework is a well-established, non-parametric method for estimating a pdf from a finite number of observed data [170]. The KDE method enables us to automatically select an optimal value for the common Gaussian width (denoted the bandwidth), based on the available data. Throughout this thesis, we estimate the bandwidth using the method proposed by Botev *et al.* [32], implemented in Matlab.

A.5 Conditional distribution of interarrival times in a Poisson process

Here, we prove that the distribution of interarrival times, conditional on the number of events, is uniform in a Poisson process. Consider events occurring as a Poisson process with rate λ . We denote the number of events that occur within the time interval $[0, t)$ by $N(t)$. Suppose that we know that exactly one event occurs during this interval, so that $N(t) = 1$. We seek an expression for the pdf of the time at which this event occurs, denoted T . We start by considering the conditional cumulative distribution function,

$$\begin{aligned} \mathbb{P}(T \leq x \mid N(t) = 1) &= \frac{\mathbb{P}(T \leq x \cap N(t) = 1)}{\mathbb{P}(N(t) = 1)} \\ &= \frac{\mathbb{P}(T \leq x \cap N(t) - N(x) = 0)}{\lambda t e^{-\lambda t}} \\ &= \frac{x}{t}, \end{aligned}$$

where the final equality follows since the probabilities of events occurring in the non-overlapping intervals $[0, x)$ and $[x, t)$ are independent. Differentiating the result with respect to x , we see that the distribution of event times within an interval, conditional on one event occurring, is uniform on that interval.

Appendix B

Analysis of experimental data

In this appendix, we present additional information to support the methods, results and conclusions in Chapter 4.

B.1 Links to online video clips

Short clips of the raw microscopy video data of three strains of *R. sphaeroides*, presented in Chapter 4, are available online at the following locations (all videos are hosted on Youtube):

1. Wildtype, bulk swimming: <http://youtu.be/NSr5ZTAq-BM>
2. Wildtype, surface swimming: <http://youtu.be/GU8woZGEDpQ>
3. Non-chemotactic, bulk swimming: <http://youtu.be/Hkoeofh2oP0>
4. Non-chemotactic, surface swimming: http://youtu.be/1_2RM-jaW0o
5. Non-motile, bulk swimming: <http://youtu.be/ye9sj8DJrMg>

B.2 Computational method for finding the minimum bounding circle

In order to implement the censoring methods described in Section 4.1.6, we use the Matlab code `minboundcircle`¹ to find the minimum bounding radius (MBR) of a list of two-dimensional coordinates. The algorithm initially discards those coordinates that do not lie on the edge of the convex hull of the supplied coordinate list. Next, three coordinates are initially selected and denoted the active set. The algorithm then enters a loop, in which the circle passing through the three coordinates in the active set is computed, and the remaining coordinates are tested for whether they lie inside this circle. If any do not, one of the coordinates in the active set is discarded and replaced with the coordinate lying furthest from the active set circle. The algorithm thus iterates, exchanging coordinates in the active set, until no coordinates are found outside of the active set circle. At this stage, the minimum bounding circle has been found.

B.3 Results from the surface swimming dataset

In this section we present results obtained by analysing the tracks from surface bacteria swimming near the surface of the microscope coverslip. These are analogous to the results obtained from the bulk dataset, which are found in Section 4.2.

The estimated pdf of observed framewise speeds for the non-chemotactic and non-motile surface datasets are plotted in Figure B.1. This plot is analogous to Figure 4.6(a), which illustrates the same plot for the bulk dataset. The data have been filtered to remove failed tracks. In addition, the non-chemotactic dataset has been filtered to remove tracks whose MBR is below the cutoff value $\rho_E = 1.2 \mu\text{m}$. This additional censoring step is required in order to observe separation of the two pdfs, which is not required in the case of the bulk

¹Code by John d'Ericco, available on the Matlab Central File Exchange website: <http://www.mathworks.com/matlabcentral/fileexchange/>.

dataset. We manually verified that many of the tracks that are removed by this censoring procedure appear to be stuck to the surface and spinning in circles. The contribution of these tracks to the observed framewise speed distribution differs substantially from that of the non-motile tracks. Following their removal, it is clear that the estimated pdfs are very similar for the bulk and surface datasets.

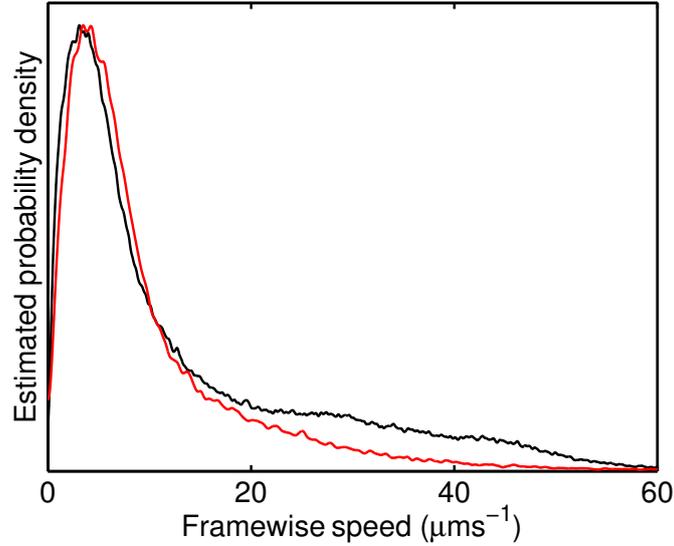


Figure B.1: *Estimated pdf of framewise speeds for the censored non-chemotactic (black line) and non-motile (red line) surface datasets. The distributions have been scaled in height so their maxima coincide.*

B.3.1 Initial censoring and characterisation of wildtype tracks

Figure B.2 shows tracks with a range of tortuosities from the wildtype surface dataset. The analogous plot for tracks from the bulk dataset is shown in Figure 4.7. The figure shows a qualitatively similar trend to that in the bulk dataset.

A scatter plot of median curvature against mean track speed is plotted in Figure B.3, analogous to Figure 4.8(a). The range of curvatures in the dataset with no censoring except for the removal of failed tracks is greater in the surface dataset. However, the effect of censoring by median curvature is similar in both cases. We therefore use the same threshold value to censor the wildtype surface dataset as that used for the wildtype

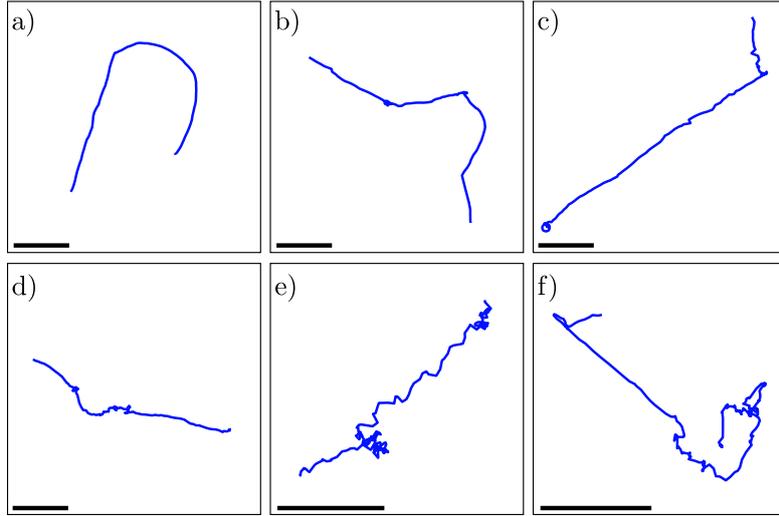


Figure B.2: Tracks from the wildtype surface dataset with different median curvatures. (a) $0.04 \text{ rad } \mu\text{m}^{-1}$, (b) $0.13 \text{ rad } \mu\text{m}^{-1}$, (c) $0.61 \text{ rad } \mu\text{m}^{-1}$, (d) $0.95 \text{ rad } \mu\text{m}^{-1}$, (e) $1.17 \text{ rad } \mu\text{m}^{-1}$, (f) $1.68 \text{ rad } \mu\text{m}^{-1}$. In each case, the black bar represents $12 \mu\text{m}$.

bulk dataset, $\rho_{\text{MC}} = 1.5 \text{ rad } \mu\text{m}^{-1}$.

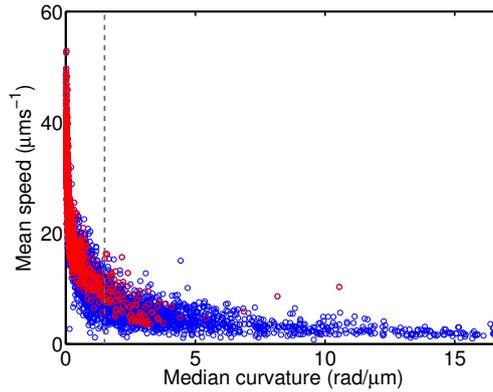


Figure B.3: Scatter plot of median curvature against mean track speed for the wildtype surface dataset. Blue points come from all tracks in the dataset, after filtering with ρ_{FS} . Red points are those remaining following censoring by applying a minimum extent of $6 \mu\text{m}$. The grey dashed line represents the maximum median curvature value; tracks with median curvature above this value will be discarded.

Figure B.4 shows a random sample of thirty tracks taken from the surface dataset, before and after censoring. The analogous plot for the bulk dataset is shown in Figure 4.9. Figure B.4(a) contains two tracks that roughly trace out circles. These are a common feature of the surface dataset, due to bacteria becoming stuck to the coverslip by their flagella and rotating on the spot. This also appears to occur transiently in one of the

post-censoring tracks in Figure B.4(b) (dark blue line).

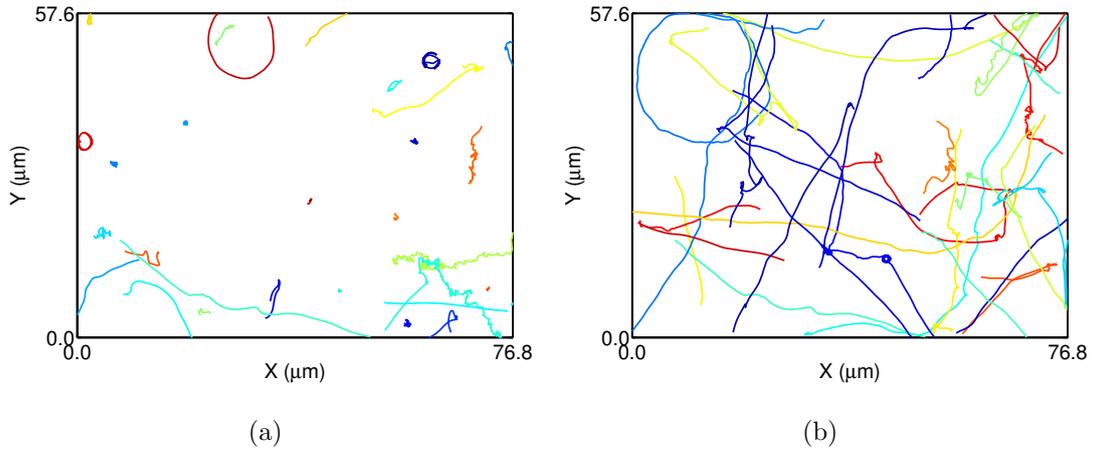


Figure B.4: A representative selection of 30 tracks from the wildtype surface dataset, (a) before and (b) after censoring. Censoring was carried out with $\rho_E = 6 \mu m$ and $\rho_{MC} = 1.5 \text{ rad} \mu m^{-1}$.

The estimated pdf of framewise speeds is shown in Figure B.5, before and after censoring. The analogous figure for the bulk dataset is 4.8(b). The uncensored distributions differ significantly between the bulk and surface datasets, with the modal speed appearing much lower in the uncensored surface data, while the distributions after censoring are qualitatively very similar. The difference in the uncensored distributions is likely to be due to the presence of stuck cells in the surface dataset, as discussed above and apparent in Figure B.4(a).

B.3.2 Censoring the non-chemotactic surface swimming dataset

Figures B.6 and B.7 are analogous to Figures 4.10 and 4.11, respectively, in Chapter 4. The results are qualitatively similar, with $\rho_{EMS} = 24 \mu m s^{-1}$.

B.3.3 Characteristics of the motion of wildtype bacteria

The estimated pdf of framewise speeds observed during stopped and running phases is shown for the wildtype surface dataset in Figure B.8(a). This plot is analogous to Figure 4.16. The discrepancy between the two hidden Markov model (HMM) methods is larger

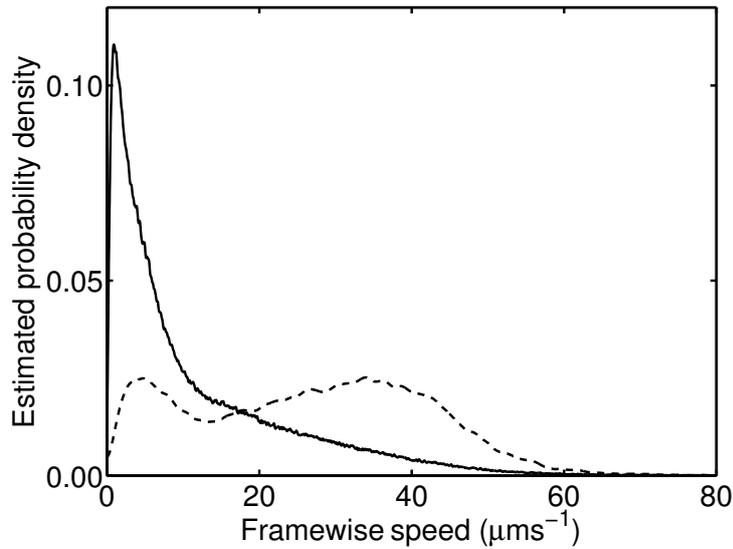


Figure B.5: *Estimated pdf of framewise speeds before censoring (solid line) and after censoring (dashed line) for the wildtype surface dataset.*

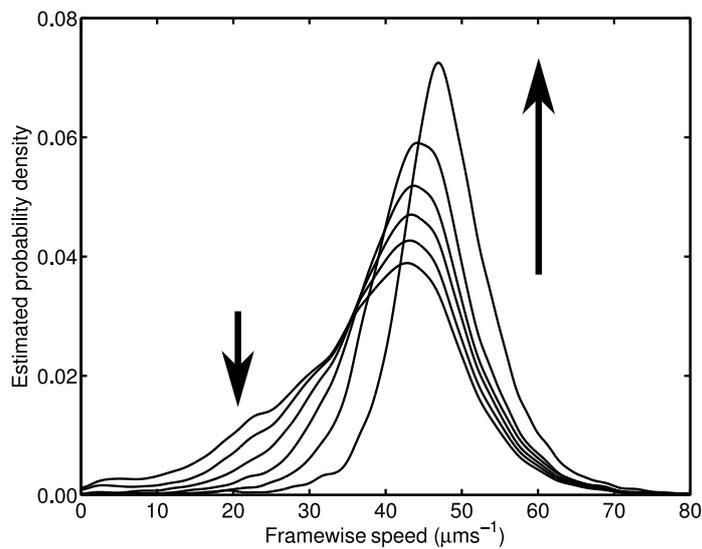


Figure B.6: *Estimated probability distribution of framewise speeds for the non-chemotactic strain swimming near a surface after censoring, varying the minimum effective mean speed parameter from $12 \mu\text{ms}^{-1}$ to $48 \mu\text{ms}^{-1}$ in intervals of $6 \mu\text{ms}^{-1}$. The arrows show the direction of change for the stated parameter variation.*

than it was for the bulk dataset, particularly for the running phase. Nonetheless, the inferred distributions do not differ a great deal.

Figure B.8(b) shows the estimated pdf of stopwise angle changes for the wildtype surface dataset. This is analogous to Figure 4.17(a) for the bulk dataset. The results are qual-

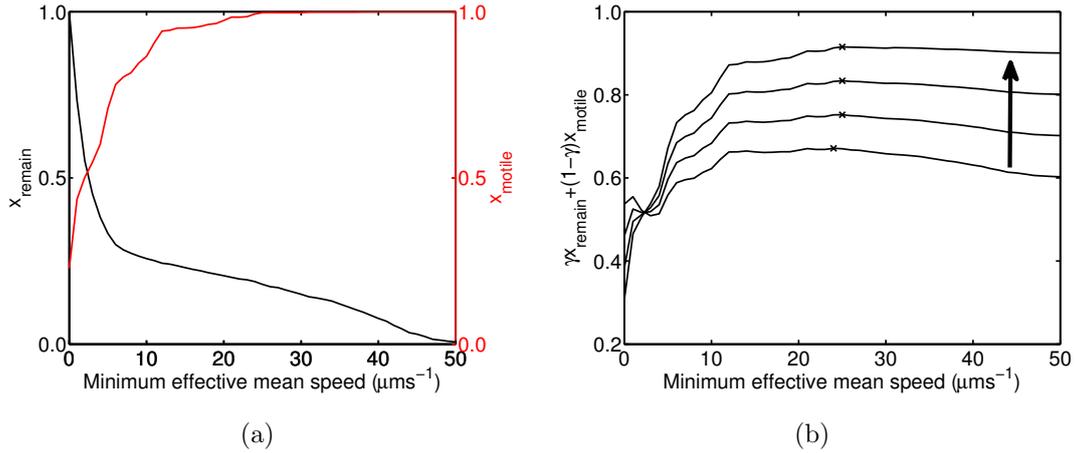


Figure B.7: (a) The effect of varying the minimum effective mean speed on x_{remain} (black) and x_{motile} (red). (b) Weighted linear combinations of x_{remain} and x_{motile} for $\gamma \in \{0.6, 0.7, 0.8, 0.9\}$. The black arrow shows the direction of increase of γ , and the black crosses indicate the position of the maximum.

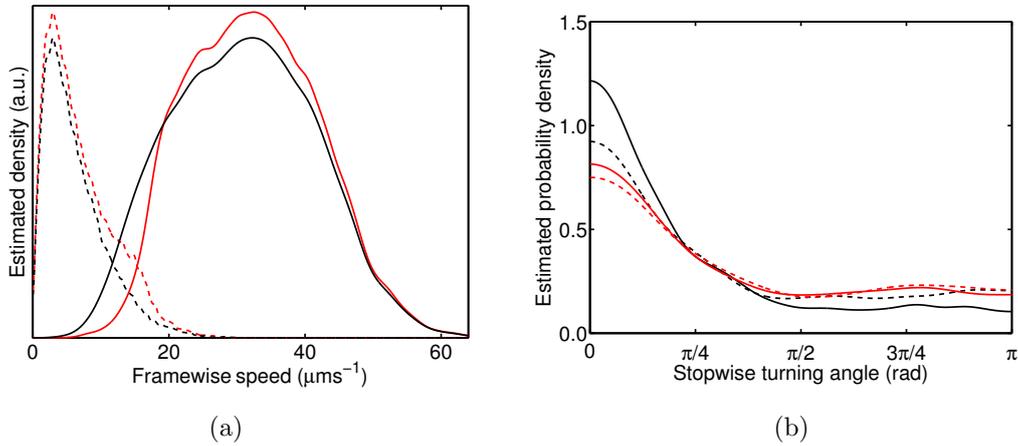


Figure B.8: (a) Estimated framewise speed pdfs for running (solid line) and stopping (dashed line) phases in the surface wildtype dataset, analysed using the full (black) and speed-only (red) HMM methods. Applying post-processing in either case made no significant difference to the estimation (data not shown). (b) Estimated pdf of the stopwise turning angles for the wildtype surface dataset, computed using the full HMM method (black) and speed-only HMM method (red), without (solid line) and with post-processing (dashed line).

itatively similar for the bulk and surface datasets, though the results from the surface dataset exhibit a more noticeable difference between the estimated pdf computed using the speed-only HMM method and that computed using the full HMM method with post-processing, with the full method producing a more peaked distribution. In addition, both methods show a more significant change when post-processing is applied than is the case when we consider the bulk dataset. These differences may be due to the greater sensitivity of the full method, compared to the speed-only method (see Section 4.3). Tracks

in the surface dataset are generally more tortuous than those in the bulk dataset, as shown by the mean median curvature after censoring, which is $0.19 \text{ rad } \mu\text{m}^{-1}$ for the bulk dataset and $0.25 \text{ rad } \mu\text{m}^{-1}$ for the surface dataset. The median curvatures also have a larger spread in the censored surface dataset, with a standard deviation of $0.31 \text{ rad } \mu\text{m}^{-1}$, compared with $0.25 \text{ rad } \mu\text{m}^{-1}$ for the bulk dataset.

B.4 Bayes factor for hypothesis testing

In this section we describe the method used to assess the statistical significance of the bias towards rotation in the surface dataset, discussed in Section 4.2.7. For this purpose, we use the Bayes factor, which is a ratio whose value allows us to compare how well two hypotheses, or models, match observed data [79]. One of these hypotheses is designated the null hypothesis, H_0 , which reflects a prior belief about the data. We seek to evaluate the probability of an alternative hypothesis, H_1 , in light of the observed data. The Bayes factor allows us to compare these hypotheses and thus to evaluate the model selection problem. Unlike the p-value that is used in frequentist statistical tests to achieve this comparison, the Bayes factor may not be used to represent the significance of the observed results (indeed, the p-value is arguably not meaningful in this sense either [78]). Instead, its value reflects the extent to which our *prior* belief in the null hypothesis is affected after observing the data. For example, a Bayes factor of ten means that our prior belief in the null hypothesis should be strengthened tenfold. The interpretation of the Bayes factor is therefore only meaningful relative to the strength of our prior belief in H_0 , and it is not possible to assign a significance level, as we may do when working with p-values. However, certain guidelines for the interpretation of the Bayes factor are found in [96].

The Bayes factor is defined as

$$B = \frac{\mathbb{P}(\text{Data}|H_0)}{\mathbb{P}(\text{Data}|H_1)}. \quad (\text{B.1})$$

Designating the set of parameters in model H_i as θ_i , the numerator in (B.1) is defined as

$$\mathbb{P}(\text{Data}|H_0) = \int \mathbb{P}(\theta_0|H_0)\mathbb{P}(\text{Data}|H_0, \theta_0) d\theta_0, \quad (\text{B.2})$$

and similarly for the denominator, with parameter set θ_1 . Note that the integrals involved may be multidimensional and that there is no requirement for the number of parameters in θ_0 and θ_1 to be equal.

Applying the above definition to the problem in Chapter 4 of assessing the likelihood of an asymmetric framewise turning distribution in the motion of *R. sphaeroides* images at the surface of a coverslip, the null hypothesis is that turns follow a symmetric distribution. We compare this to an alternative hypothesis that turns follow an asymmetric distribution. We make no assumptions about the nature of the distribution, other than the symmetry property; the framewise turning angles are simply considered as clockwise or anticlockwise, based on their sign. We now compare the two hypotheses using a binomial distribution; the null hypothesis is that the probability of a clockwise turn is exactly 0.5 and the alternative hypothesis is that the probability is anything other than 0.5. There are a total of N recorded framewise angle changes, of which n are clockwise. The null hypothesis has no parameters, so the numerator in equation (B.1) reduces to

$$\mathbb{P}(\text{Data}|H_0) = \binom{N}{n} \left(\frac{1}{2}\right)^N. \quad (\text{B.3})$$

The alternative hypothesis has a single parameter, the probability of a clockwise turn, q . For the denominator, we must integrate over all possible values of this parameter:

$$\mathbb{P}(\text{Data}|H_1) = \binom{N}{n} \int_0^1 q^n (1-q)^{N-n} dq = \frac{1}{N+1}. \quad (\text{B.4})$$

The Bayes factor is equal to the ratio of these values. Table B.1 shows the values computed for the bulk and surface wildtype datasets. In practice, the numerical values computed

Table B.1: Bayes factor and related quantities for assessing the probability of the null hypothesis (H_0) that bacteria turn with a symmetric distribution against the alternative hypothesis (H_1) that turns are asymmetrically distributed.

Dataset	N	n	$\log(\mathbb{P}(\text{Data} H_0))$	$\log(\mathbb{P}(\text{Data} H_1))$	Bayes factor
Wildtype bulk	87052	43507	-5.9212	-11.3	233.47
Wildtype surface	32113	14654	-128.07	-10.38	$\approx 10^{-52}$

using (B.3) and (B.4) are very small. We therefore compute the logarithm of the values in order to avoid numerical errors, and compute the logarithm of the Bayes factor, given by

$$\log(B) = \log(\mathbb{P}(\text{Data}|H_0)) - \log(\mathbb{P}(\text{Data}|H_1)). \quad (\text{B.5})$$

We present the computed values of B in Table B.1. The results show that the data *increase* our belief in the null hypothesis by a factor of over 200 in the case of the bulk dataset. This is very strong evidence for the symmetry of turns [96]. The data *decrease* our belief in the null hypothesis by a very large factor in the surface dataset, so that there is no significant probability that $q = 0.5$, regardless of our prior belief in the symmetry of turns.

Note that the Bayes factor is independent of the inferred value of q in the alternative hypothesis; we have shown that $q \neq 0.5$ in the surface dataset, but not by how much q differs from one half.

Appendix C

Modelling the effect of Brownian buffeting on motile bacteria

C.1 Implementation of the Euler-Maruyama algorithm

Algorithm 3 shows pseudo-code for simulating the run-and-active-stop process. The algorithm for the run-and-stop model is obtained by setting $\omega = 1$, and the run-only simulation is obtained by setting $\omega = 1$ and $\lambda = 0$.

C.2 Derivation of second moments for the run-only model

In this section, we show how we arrive at equations (6.34) and (6.35), which give expressions for the second position moments of the run-only model. The second moment of the run-only process in the x -coordinate is given by

$$\langle x^2(t) \rangle = c^2 \left\langle \int_0^t \cos(\sqrt{2D_r}B(u)) \, du \int_0^t \cos(\sqrt{2D_r}B(v)) \, dv \right\rangle$$

Algorithm 3 Numerical simulation of a single track of total duration $N\Delta t$ in the run and active stop model.

```

 $\zeta = 8\pi\eta r^3.$ 
 $D_r = kT/\zeta.$ 
 $\sigma_{\text{run}} = \sqrt{2D_r\Delta t}.$ 
 $\sigma_{\text{stop}} = \sqrt{2D_r\omega\Delta t}.$ 
 $\phi(t_0) = x(t_0) = y(t_0) = 0.$ 
 $F(t_0) = 1.$ 
for  $i = 1$  to  $N$  do
  Propagate particle:
   $x(t_i) = x(t_{i-1}) + F(t_{i-1})c \cos(\phi(t_{i-1})),$ 
   $y(t_i) = y(t_{i-1}) + F(t_{i-1})c \sin(\phi(t_{i-1})).$ 
  Update  $\phi$ ,  $F$ :
  Draw random variable  $u \sim U(0, 1).$ 
  if  $F(t_{i-1}) = 1$  then
    Draw random variable  $\xi(t_i) \sim N(0, \sigma_{\text{run}}).$ 
    if  $u < \lambda\Delta t$  then
       $F(t_i) = 0,$ 
    else
       $F(t_i) = 1.$ 
    end if
  else
    Draw random variable  $\xi(t_i) \sim N(0, \sigma_{\text{stop}}).$ 
    if  $u < \mu\Delta t$  then
       $F(t_i) = 1,$ 
    else
       $F(t_i) = 0.$ 
    end if
  end if
   $\phi(t_i) = \phi(t_{i-1}) + \xi(t_i).$ 
end for

```

$$\begin{aligned}
&= 2c^2 \left\langle \int_0^t \int_0^v \cos \left(\sqrt{2D_r} B(u) \right) \cos \left(\sqrt{2D_r} B(v) \right) du dv \right\rangle \\
&= c^2 \int_0^t \int_0^v 2 \left\langle \cos \left(\sqrt{2D_r} B(u) \right) \cos \left(\sqrt{2D_r} B(v) \right) \right\rangle du dv,
\end{aligned}$$

where, in the final step, we again interchange the order of operation using Fubini's Theorem. We now seek an expression for the ensemble average in the integrand. We first note that

$$\begin{aligned}
2 \left\langle \cos \left(\sqrt{2D_r} B(u) \right) \cos \left(\sqrt{2D_r} B(v) \right) \right\rangle &= \left\langle \cos \left[\sqrt{2D_r} (B(v) + B(u)) \right] \right\rangle \\
&\quad + \left\langle \cos \left[\sqrt{2D_r} (B(v) - B(u)) \right] \right\rangle.
\end{aligned} \tag{C.1}$$

The domain of integration is defined such that $u \leq v \leq t$. The Wiener increment $B(v) - B(u)$ is therefore normally distributed with zero mean and variance equal to $v - u$, using equation (6.16). It is not immediately apparent whether the sum $B(v) + B(u)$ is normally distributed, however we can demonstrate this property by rewriting the expression as follows:

$$B(v) + B(u) = 2B(u) + (B(v) - B(u)).$$

Both $2B(u)$ and the difference $B(v) - B(u)$ are normally distributed with zero mean, the former with variance $4u$, and the latter with variance $v - u$. It follows that $B(v) + B(u)$ is the sum of two normally distributed random variables and is therefore itself a normal random variable with zero mean and variance $3u + v$. Applying this to equation (C.1) we obtain

$$\left\langle \cos \left[\sqrt{2D_r} (B(v) + B(u)) \right] \right\rangle = e^{-D_r(v+3u)}, \tag{C.2}$$

$$\left\langle \cos \left[\sqrt{2D_r} (B(v) - B(u)) \right] \right\rangle = e^{-D_r(v-u)}. \tag{C.3}$$

Integrating these expressions gives

$$\langle x^2(t) \rangle = c^2 \int_0^t \int_0^v e^{-D_r(v+3u)} + e^{-D_r(v-u)} du dv, \quad (\text{C.4})$$

the evaluation of which yields equation (6.34). A similar process may be followed to obtain an expression for $\langle y^2(t) \rangle$, which is given in equation (6.35).

Bibliography

- [1] Abramowitz, M., Stegun, I. A., 1965. Handbook of Mathematical Functions, 1st Edition. Dover Publications.
- [2] Adler, J., 1966. Chemotaxis in bacteria. *Science* 153, 708–716.
- [3] Adler, J., 1966. Effect of amino acids and oxygen on chemotaxis in *Rhodobacter sphaeroides*. *J. Bacteriol.* 92, 121–129.
- [4] Adler, J., 1973. A method for measuring chemotaxis and use of the method to determine optimum conditions for chemotaxis by *Rhodobacter sphaeroides*. *J. Gen. Microbiol.* 74, 77–91.
- [5] Adler, J., Hazelbauer, G. L., Dahl, M. M., 1973. Chemotaxis toward sugars in *Rhodobacter sphaeroides*. *J. Bacteriol.* 115, 824–847.
- [6] Ahmed, T., Shimizu, T. S., Stocker, R., 2010. Microfluidics for bacterial chemotaxis. *Integr. Biol.* 2, 604–629.
- [7] Alon, U., Camarena, L., Surette, M. G., Aguera, Liu, Y., Leibler, S., Stock, J. B., 1998. Response regulator output in bacterial chemotaxis. *EMBO J.* 17, 4238–4248.
- [8] Alon, U., Surette, M. G., Barkai, N., Leibler, S., 1999. Robustness in bacterial chemotaxis. *Nature* 397, 168–171.
- [9] Alt, W., 1980. Biased random walk models for chemotaxis and related diffusion approximations. *J. Math. Biol.* 9, 147–177.

- [10] Archer, N. K., Mazaitis, M. J., Costerton, W. W., Leid, J. G., Elizabeth, M., Shirliff, M. E., 2011. *Staphylococcus aureus* biofilms: properties, regulation, and roles in human disease. *Virulence* 2, 445–459.
- [11] Armitage, J. P., Macnab, R. M., 1987. Unidirectional, intermittent rotation of the flagellum of *Rhodobacter sphaeroides*. *J. Bacteriol.* 169, 514–518.
- [12] Armitage, J. P., Pitta, T. P., Vigeant, M. A., Packer, H. L., Ford, R. M., 1999. Transformations in flagellar structure of *Rhodobacter sphaeroides* and possible relationship to changes in swimming speed. *J. Bacteriol.* 181, 4825–4833.
- [13] Armstrong, J. B., Adler, J., 1967. Genetics of motility in *Rhodobacter sphaeroides*: complementation of paralysed mutants. *Genetics* 56, 363–373.
- [14] Armstrong, J. B., Adler, J., Dahl, M. M., 1967. Nonchemotactic mutants of *Rhodobacter sphaeroides*. *J. Bacteriol.* 93, 390–398.
- [15] Bai, F., Branch, R. W., Nicolau Jr, D. V., Pilizota, T., Steel, B. C., Maini, P. K., Berry, R. M., 2010. Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch. *Science* 327, 685–689.
- [16] Bailey, H., Thompson, P., 2006. Quantitative analysis of bottlenose dolphin movement patterns and their relationship with foraging. *J. Anim. Ecol.* 75, 456–465.
- [17] Beausang, J. F., Nelson, P. C., 2007. Diffusive hidden Markov model characterization of DNA looping dynamics in tethered particle experiments. *Phys. Biol.* 4, 205–219.
- [18] Benhamou, S., 2004. How to reliably estimate the tortuosity of an animal’s path: straightness, sinuosity, or fractal dimension? *J. Theor. Biol.* 229, 209–220.
- [19] Berg, H., 2009. The gain paradox. *Prog. Biophys. Mol. Biol.* 100, 2–3.
- [20] Berg, H. C., 1971. How to track bacteria. *Rev. Sci. Instrum.* 42, 868–871.

- [21] Berg, H. C., 1993. Random Walks in Biology. Princeton University Press.
- [22] Berg, H. C., 2003. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* 72, 19–54.
- [23] Berg, H. C., 2006. Marvels of bacterial behavior. *P. Am. Philos. Soc.* 150, 428–442.
- [24] Berg, H. C., Brown, D. A., 1972. Chemotaxis in *Rhodobacter sphaeroides* analysed by three-dimensional tracking. *Nature* 239, 500–504.
- [25] Berg, H. C., Purcell, E. M., 1977. Physics of chemoreception. *Biophys. J.* 20, 193–219.
- [26] Berg, H. C., Turner, L., 1990. Chemotaxis of bacteria in glass capillary arrays. *Rhodobacter sphaeroides*, motility, microchannel plate, and light scattering. *Biophys. J.* 58, 919–930.
- [27] Berke, A. P., Turner, L., Berg, H. C., Lauga, E., 2008. Hydrodynamic attraction of swimming microorganisms by surfaces. *Phys. Rev. Lett.* 101, 038102–038105.
- [28] Berry, R. M., Armitage, J. P., 2000. Response kinetics of tethered *Rhodobacter sphaeroides* to changes in light intensity. *Biophys. J.* 78, 1207–1215.
- [29] Block, S. M., Fahrner, K. A., Berg, H. C., 1991. Visualization of bacterial flagella by video-enhanced light microscopy. *J. Bacteriol.* 173, 933–936.
- [30] Block, S. M., Segall, J. E., Berg, H. C., 1982. Impulse responses in bacterial chemotaxis. *Cell* 31, 215–226.
- [31] Bödeker, H. U., Beta, C., Frank, T. D., Bodenschatz, E., 2010. Quantitative analysis of random ameboid motion. *Europhys. Lett.*, 28005+.
- [32] Botev, Z. I., Grotowski, J. F., Kroese, D. P., 2010. Kernel density estimation via diffusion. *Ann. Stat.* 38, 2916–2957.

- [33] Bovet, P., Benhamou, S., 1988. Spatial analysis of animals' movements using a correlated random walk model. *J. Theor. Biol.* 131, 419–433.
- [34] Bray, D., 2002. Bacterial chemotaxis and the question of gain. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7–9.
- [35] Bray, D., Levin, M. D., Morton-Firth, C. J., 1998. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* 393, 85–88.
- [36] Brooks, C. J., Harris, S., 2008. Directed movement and orientation across a large natural landscape by zebras, *Equus burchelli antiquorum*. *Anim. Behav.* 76, 277–285.
- [37] Brown, D. A., Berg, H. C., 1974. Temporal stimulation of chemotaxis in *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 1388–1392.
- [38] Brown, M. T. Control of the unidirectional motor in *Rhodobacter sphaeroides*. D.Phil. thesis, University of Oxford.
- [39] Buch, N., Velastin, S. A., Orwell, J., 2011. A review of computer vision techniques for the analysis of urban traffic. *IEEE T. Intell. Transp.* 12, 920–939.
- [40] Byers, J. A., 2001. Correlated random walk equations of animal dispersal resolved by simulation. *Ecology* 82, 1680–1690.
- [41] Cain, M. L., 1990. Models of clonal growth in *Solidago altissima*. *J. Ecol.* 78, 27+.
- [42] Campos, D., Méndez, V., 2009. Superdiffusive-like motion of colloidal nanorods. *J. Chem. Phys.* 130, 134711–134714.
- [43] Campos, D., Méndez, V., Llopis, I., 2010. Persistent random motion: Uncovering cell migration dynamics. *J. Theor. Biol.* 267, 526–534.

- [44] Chattopadhyay, S., Moldovan, R., Yeung, C., Wu, X. L., 2006. Swimming efficiency of bacterium *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13712–13717.
- [45] Chernova, A., Armitage, J., Packer, H., Maini, P., 2003. Response kinetics of tethered bacteria to stepwise changes in nutrient concentration. *BioSystems* 71, 51–59.
- [46] Claussen, D. L., Finkler, M. S., Smith, M. M., 1997. Thread trailing of turtles: methods for evaluating spatial movements and pathway structure. *Can. J. Zool.* 75, 2120–2128.
- [47] Codling, E. A., 2003. Biased random walks in biology. Ph.D. thesis, University of Leeds.
- [48] Codling, E. A., Bearon, R. N., Thorn, G. J., 2010. Diffusion about the mean drift location in a biased random walk. *Ecology* 91, 3106–3113.
- [49] Codling, E. A., Hill, N. A., 2005. Calculating spatial statistics for velocity jump processes with experimentally observed reorientation parameters. *J. Math. Biol.* 51, 527–556.
- [50] Codling, E. A., Hill, N. A., 2005. Sampling rate effects on measurements of correlated and biased random walks. *J. Theor. Biol.* 233, 573–588.
- [51] Codling, E. A., Plank, M. J., Benhamou, S., 2008. Random walk models in biology. *J. R. Soc. Interface* 5, 813–834.
- [52] Coffey, W. T., Kalmykov, Y. P., Waldron, J. T., 2003. *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*, 2nd Edition. World Scientific Publishing Company.
- [53] Collins-Hooper, H., Woolley, T. E., Dyson, L., Patel, A., Potter, P., Baker, R. E., Gaffney, E. A., Maini, P. K., Dash, P. R., Patel, K., 2012. Age-related changes in

- speed and mechanism of adult skeletal muscle stem cell migration. *Stem Cells* 30, 1182–1195.
- [54] Conn, A. R., Gould, N. I. M., Toint, P. L., 1987. Trust-Region Methods (MPS-SIAM Series on Optimization). Society for Industrial and Applied Mathematics.
- [55] Correa, N. E., Klose, K. E., 2005. Characterization of enhancer binding by the *Vibrio cholerae* flagellar regulatory protein FlrC. *J. Bacteriol.* 187, 3158–3170.
- [56] Das, R., Cairo, C. W., 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, e1000556+.
- [57] Deroulers, C., Aubert, M., Badoual, M., Grammaticos, B., 2009. Modeling tumor cell migration: from microscopic to macroscopic. *Phys. Rev. E* 79, 031917–031930.
- [58] DiLuzio, W. R., Turner, L., Mayer, M., Garstecki, P., Weibel, D. B., Berg, H. C., Whitesides, G. M., 2005. *Rhodobacter sphaeroides* swim on the right-hand side. *Nature* 435, 1271–1274.
- [59] Doob, J. L., 1990. Stochastic Processes. Wiley-Interscience.
- [60] Dowd, J. P., Matsumura, P., 1997. The use of flash photolysis for a high-resolution temporal and spatial analysis of bacterial chemotactic behaviour: CheZ is not always necessary for chemotaxis. *Mol. Microbiol.* 25, 295–302.
- [61] Drews, G., 2005. Contributions of Theodor Wilhelm Engelmann on phototaxis, chemotaxis, and photosynthesis. *Photosynth. Res.* 83, 25–34.
- [62] Duffy, K. J., Ford, R. M., 1997. Turn angle and run time distributions characterize swimming behavior for *Pseudomonas putida*. *J. Bacteriol.* 179, 1428–1430.
- [63] Duke, T. A. J., Bray, D., 1999. Heightened sensitivity of a lattice of membrane receptors. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10104–10108.

- [64] Dürr, S., Thomason, J., 2009. Biofouling, 1st Edition. Wiley-Blackwell.
- [65] Dusenbery, D. B., 1997. Minimum size limit for useful locomotion by free-swimming microbes. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10949–10954.
- [66] Ehlers, K. M., Samuel, A. D., Berg, H. C., Montgomery, R., 1996. Do cyanobacteria swim using traveling surface waves? *Proc. Natl. Acad. Sci. U. S. A.* 93, 8340–8343.
- [67] Elowitz, M. B., Levine, A. J., Siggia, E. D., Swain, P. S., 2002. Stochastic gene expression in a single cell. *Science* 297, 1183–1186.
- [68] Ely, B., Gerardot, C. J., Fleming, D. L., Gomes, S. L., Frederikse, P., Shapiro, L., 1986. General nonchemotactic mutants of *Caulobacter crescentus*. *Genetics* 114, 717–730.
- [69] Erban, R., Chapman, S. J., 2007. Reactive boundary conditions for stochastic simulations of reaction-diffusion processes. *Phys. Biol.* 4, 16–28.
- [70] Erban, R., Othmer, H. G., 2004. From individual to collective behavior in bacterial chemotaxis. *SIAM J. Appl. Math.* 65, 361–391.
- [71] Erban, R., Othmer, H. G., 2005. From signal transduction to spatial pattern formation in *E. coli*: A paradigm for multiscale modeling in biology. Vol. 3. pp. 362–394.
- [72] Evans, L. C., 2010. Partial Differential Equations, 2nd Edition. American Mathematical Society.
- [73] Fortin, D., Morales, J. M., Boyce, M. S., 2005. Elk winter foraging at fine scale in Yellowstone National Park. *Oecologia* 145, 334–342.
- [74] Frymier, P. D., Ford, R. M., Berg, H. C., Cummings, P. T., 1995. Three-dimensional tracking of motile bacteria near a solid planar surface. *Proc. Natl. Acad. Sci. U. S. A.* 92, 6195–6199.
- [75] Gaffney, E. A. *personal communication*.

- [76] Gardiner, C., 2004. Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences, 3rd Edition. Springer.
- [77] Goldstein, S., 1951. On diffusion by discontinuous movements, and on the Telegraph equation. *Quart. J. Mech. Appl. Math.* 4, 129–156.
- [78] Goodman, S. N., 1999. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann. Intern. Med.* 130, 995–1004.
- [79] Goodman, S. N., 1999. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann. Intern. Med.* 130, 1005–1013.
- [80] Grimmett, G. R., Stirzaker, D. R., 2001. Probability and Random Processes, 3rd Edition. Oxford University Press, USA.
- [81] Grisan, E., Foracchia, M., Ruggeri, A., 2003. A novel method for the automatic evaluation of retinal vessel tortuosity. In: Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE. Vol. 1. IEEE, pp. 866–869.
- [82] Hagen, B. T., van Teeffelen, S., Löwen, H., 2009. Non-Gaussian behaviour of a self-propelled particle on a substrate. *arXiv:0906.3418 [cond-mat.soft]*.
- [83] Hagen, B. T., van Teeffelen, S., Löwen, H., 2011. Brownian motion of a self-propelled particle. *J. Phys.: Condens. Matter* 23, 194119+.
- [84] Harshey, R. M., 2003. Bacterial motility on a surface: many ways to a common goal. *Annu. Rev. Microbiol.* 57, 249–273.
- [85] Hashimoto, M., Mashimo, T., Hirano, T., Yamaguchi, S., Aizawa, S.-I., 2008. Functional roles of the hook in a rotating tethered cell. *J. Mol. Biol.* 375, 367–375.
- [86] Haya, S., Tokumaru, Y., Abe, N., Kaneko, J., Aizawa, S.-I., 2011. Characterization of lateral flagella of *Selenomonas ruminantium*. *Appl. Environ. Microbiol.* 77, 2799–2802.

- [87] Higham, D. J., 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* 43, 525–546.
- [88] Hill, N. A., Häder, D. P., 1997. A biased random walk model for the trajectories of swimming micro-organisms. *J. Theor. Biol.* 186, 503–526.
- [89] Hobson, G. Hobson tracker user manual. Hobson Tracking Systems Ltd., Sheffield, UK.
- [90] Holt, J. G., Sneath, P. H., 2005. Bergey’s Manual of Systematic Bacteriology, 2nd Edition. Vol. 2. Springer.
- [91] Hu, W., Tan, T., Wang, L., Maybank, S., 2004. A survey on visual surveillance of object motion and behaviors. *IEEE Tran. Syst. Man Cybern. C* 34, 334–352.
- [92] Jabbarzadeh, E., Abrams, C., 2005. Chemotaxis and random motility in unsteady chemoattractant fields: a computational study. *J. Theor. Biol.* 235, 221–232.
- [93] Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., Danuser, G., 2008. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* 5, 695–702.
- [94] Jarrell, K. F., McBride, M. J., 2008. The surprisingly diverse ways that prokaryotes move. *Nat. Rev. Microbiol.* 6, 466–476.
- [95] Jefferson, K., 2004. What drives bacteria to produce a biofilm? *FEMS Microbiol. Lett.* 236, 163–173.
- [96] Jeffreys, H., 1998. Theory of Probability, 3rd Edition. Oxford University Press, USA.
- [97] Jiang, L., Ouyang, Q., Tu, Y., 2010. Quantitative modeling of *Rhodobacter sphaeroides* chemotactic motion in environments varying in space and time. *PLoS Comput. Biol.* 6, e1000735+.

- [98] Johansen, J. E., Pinhassi, J., Blackburn, N., Zweifel, U. L., Hagström, A., 2002. Variability in motility characteristics among marine bacteria. *Aquat. Microb. Ecol.* 28, 229–237.
- [99] Josenhans, C., Eaton, K. A., Thevenot, T., Suerbaum, S., 2000. Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fliP*, a gene encoding a basal body protein. *Infect. Immun.* 68, 4598–4603.
- [100] Kac, M., 1974. A stochastic model related to the Telegrapher’s equation. *Rocky Mt. J. Math.* 4, 497–510.
- [101] Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45.
- [102] Karch, G., 2000. Selfsimilar profiles in large time asymptotics of solutions to damped wave equations. *Studia Mathematica* 143, 175–197.
- [103] Kareiva, P. M., Shigesada, N., 1983. Analyzing insect movement as a correlated random walk. *Oecologia* 56, 234–238.
- [104] Keller, E. F., Segel, L. A., 1971. Model for chemotaxis. *J. Theor. Biol.* 30, 225–234.
- [105] Kirby, B., 2010. Micro- and Nanoscale Fluid Mechanics: Transport in Microfluidic Devices. Cambridge University Press.
- [106] Klapper, I., Dockery, J., 2010. Mathematical description of microbial biofilms. *SIAM Rev.* 52, 221–265.
- [107] Klebaner, F. C., 1999. Introduction to Stochastic Calculus with Applications. World Scientific Publishing Company.
- [108] Koch, A. L., Blumberg, G., 1976. Distribution of bacteria in the velocity gradient centrifuge. *Biophys. J.* 16, 389–405.

- [109] Koenig, S. H., 1975. Brownian motion of an ellipsoid. A correction to Perrin's results. *Biopolymers* 14, 2421–2423.
- [110] Köhler, T., Curty, L. K., Barja, F., van Delden, C., Pechère, J.-C., 2000. Swarming of *Pseudomonas aeruginosa* Is dependent on cell-to-cell signaling and requires flagella and pili. *J. Bacteriol.* 182, 5990–5996.
- [111] Kojadinovic, M., Sirinelli, A., Wadhams, G. H., Armitage, J. P., 2011. New motion analysis system for characterization of the chemosensory response kinetics of *Rhodobacter sphaeroides* under different growth conditions. *Appl. Environ. Microbiol.* 77, 4082–4088.
- [112] Kolter, R., Greenberg, E. P., 2006. The superficial life of microbes. *Nature* 441, 300–302.
- [113] Lackie, J. M. (Ed.), 2007. The Dictionary of Cell and Molecular Biology, 4th Edition. Academic Press.
- [114] Lard, M., Bäckman, J., Yakovleva, M., Danielsson, B., Hansson, L.-A. A., 2010. Tracking the small with the smallest - using nanotechnology in tracking zooplankton. *PLoS One* 5, e13516+.
- [115] Larsen, S. H., Reader, R. W., Kort, E. N., Tso, W. W., Adler, J., 1974. Change in direction of flagellar rotation is the basis of the chemotactic response in *Rhodobacter sphaeroides*. *Nature* 249, 74–77.
- [116] Lauffenburger, D., Kennedy, C., Aris, R., 1984. Traveling bands of chemotactic bacteria in the context of population growth. *Bull. Math. Biol.* 46, 19–40.
- [117] Lauga, E., DiLuzio, W. R., Whitesides, G. M., Stone, H. A., 2006. Swimming in circles: motion of bacteria near solid boundaries. *Biophys. J.* 90, 400–412.
- [118] Lewiner, T., Gomes, J., Lopes, H., Craizer, M., 2005. Curvature and torsion estimators based on parametric curve fitting. *Comput. Graph.* 29, 641–655.

- [119] Li, G., Tam, L.-K. K., Tang, J. X., 2008. Amplified effect of Brownian motion in bacterial near-surface swimming. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18355–18359.
- [120] Lipkow, K., 2006. Changing cellular location of CheZ predicted by molecular simulations. *PLoS Comput. Biol.* 2, e39+.
- [121] Lipkow, K., Andrews, S. S., Bray, D., 2005. Simulated diffusion of phosphorylated CheY through the cytoplasm of *Rhodobacter sphaeroides*. *J. Bacteriol.* 187, 45–53.
- [122] Little, M. A., Steel, B. C., Bai, F., Sowa, Y., Bilyard, T., Mueller, D. M., Berry, R. M., Jones, N. S., 2011. Steps and bumps: precision extraction of discrete states of molecular machines. *Biophys. J.* 101, 477–485.
- [123] Lovely, P., Dahlquist, F. W., Macnab, R., Koshland, J., 1974. An instrument for recording the motions of microorganisms in chemical gradients. *Rev. Sci. Instrum.* 45, 683–686.
- [124] Lovely, P. S., Dahlquist, F. W., 1975. Statistical measures of bacterial motility and chemotaxis. *J. Theor. Biol.* 50, 477–496.
- [125] Macnab, R., Koshland, D. E., 1974. Bacterial motility and chemotaxis: light-induced tumbling response and visualization of individual flagella. *J. Mol. Biol.* 84, 399–406.
- [126] Macnab, R. M., 2003. How bacteria assemble flagella. *Annu. Rev. Microbiol.* 57, 77–100.
- [127] Macnab, R. M., Koshland, D. E., 1972. The gradient-sensing mechanism in bacterial chemotaxis. *Proc. Natl. Acad. Sci. U. S. A.* 69, 2509–2512.
- [128] Mahler, R. P. S., 2004. Statistics 101 for multisensor, multitarget data fusion. *Aerospace Elec. Sys. Magazine, IEEE* 19, 53–64.
- [129] Mahler, R. P. S., 2007. Statistical Multisource-Multitarget Information Fusion. Artech House Publishers.

- [130] Mardia, K. V., Jupp, P. E., 1999. *Directional Statistics*, 1st Edition. Wiley.
- [131] McCulloch, C. E., Cain, M. L., 1989. Analyzing discrete movement data as a correlated random walk. *Ecology* 70, 383–388.
- [132] Metzler, R., Tejedor, V., Jeon, J., He, Y., Wh, 2009. Analysis of single particle trajectories: from normal to anomalous diffusion. *Phys. Pol. B* 40, 1315–1331.
- [133] Meysman, F. J. R., Malyuga, V. S., Boudreau, B. P., Middelburg, J. J., 2008. A generalized stochastic approach to particle dispersal in soils and sediments. *Geochim. Cosmochim. Acta* 72, 3460–3478.
- [134] Michalet, X., 2010. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E* 82, 041914+.
- [135] Miller, M. B., Bassler, B. L., 2001. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* 55, 165–199.
- [136] Mitchell, J. G., 1991. The influence of cell size on marine bacterial motility and energetics. *Microb. Ecol.* 22, 227–238.
- [137] Mitchell, J. G., 2002. The energetics and scaling of search strategies in bacteria. *Am. Nat.* 160, 727–740.
- [138] Mitchell, J. G., Kogure, K., 2006. Bacterial motility: links to the environment and a driving force for microbial physics. *FEMS Microbiol. Ecol.* 55, 3–16.
- [139] Mitchell, J. G., Pearson, L., Dillon, S., 1996. Clustering of marine bacteria in seawater enrichments. *Appl. Environ. Microbiol.* 62, 3716–3721.
- [140] Mitchell, J. G., Pearson, L., Dillon, S., Kantalis, K., 1995. Natural assemblages of marine bacteria exhibiting high-speed motility and large accelerations. *Appl. Environ. Microbiol.* 61, 4436–4440.

- [141] Murray, J. D., 2003. *Mathematical Biology II: Spatial Models and Biomedical Applications*, 3rd Edition. Springer.
- [142] Murray, T. S., Kazmierczak, B. I., 2006. FlhF is required for swimming and swarming in *Pseudomonas aeruginosa*. *J. Bacteriol.* 188, 6995–7004.
- [143] Nicolau Jr, D. V., Armitage, J. P., Maini, P. K., 2009. Directional persistence and the optimality of run-and-tumble chemotaxis. *Comput. Biol. Chem.* 33, 269–274.
- [144] Nossal, R., Weiss, G. H., 1974. A descriptive theory of cell migration on surfaces. *J. Theor. Biol.* 47, 103–113.
- [145] Okubo, A., 1980. *Diffusion and Ecological Problems: Mathematical Models*. Vol. 10 of *Biomathematics*. Springer-Verlag.
- [146] Othmer, H. G., Dunbar, S. R., Alt, W., 1988. Models of dispersal in biological systems. *J. Math. Biol.* 26, 263–298.
- [147] Othmer, H. G., Hillen, T., 2002. The diffusion limit of transport equations II: chemotaxis equations. *SIAM J. Appl. Math.* 62, 1222–1250.
- [148] O’Toole, G. A., Kolter, R., 1998. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* 30, 295–304.
- [149] Ouellette, N., Xu, H., Bodenschatz, E., 2006. A quantitative study of three-dimensional Lagrangian particle tracking algorithms. *Exp. Fluids* 40, 301–313.
- [150] Packer, H. L., Gauden, D. E., Armitage, J. P., 1996. The behavioural response of anaerobic *Rhodobacter sphaeroides* to temporal stimuli. *Microbiol.* 142, 593–599.
- [151] Packer, H. L., Lawther, H., Armitage, J. P., 1997. The *Rhodobacter sphaeroides* flagellar motor is a variable-speed rotor. *FEBS Lett.* 409, 37–40.
- [152] Parkinson, J. S., 1987. Doing behavioral genetics with bacteria. *Genetics* 116, 499–500.

- [153] Pearson, K., 1905. The problem of the random walk. *Nature* 72, 294.
- [154] Pearson, K., 1905. The problem of the random walk. *Nature* 72, 342.
- [155] Pilizota, T., Bilyard, T., Bai, F., Futai, M., Hosokawa, H., Berry, R. M., 2007. A programmable optical angle clamp for rotary molecular motors. *Biophys. J.* 93, 264–275.
- [156] Pilizota, T., Brown, M. T., Leake, M. C., Branch, R. W., Berry, R. M., Armitage, J. P., 2009. A molecular brake, not a clutch, stops the *Rhodobacter sphaeroides* flagellar motor. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11582–11587.
- [157] Poole, P., Sinclair, D. R., Armitage, J. P., 1988. Real time computer tracking of free-swimming and tethered rotating cells. *Anal. Biochem.* 175, 52–58.
- [158] Porter, S. L., Wadhams, G. H., Armitage, J. P., 2008. *Rhodobacter sphaeroides*: complexity in chemotactic signalling. *Trends Microbiol.* 16, 251–260.
- [159] Postlethwaite, C. M., Brown, P., Dennis, T. E., 2013. A new multi-scale measure for analysing animal movement data. *J. Theor. Biol.* 317, 175–185.
- [160] Pratt, L. A., Kolter, R., 1998. Genetic analysis of *Rhodobacter sphaeroides* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol. Microbiol.* 30, 285–293.
- [161] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 2007. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press.
- [162] Priestley, H. A., 1997. Introduction to Integration. Oxford University Press.
- [163] Purcell, E. M., 1977. Life at low Reynolds number. *Am. J. Phys.* 45, 3–11.
- [164] Rabiner, L. R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–286.

- [165] Ramia, M., Tullock, D. L., Phan-Thien, N., 1993. The role of hydrodynamic interaction in the locomotion of microorganisms. *Biophys. J.* 65, 755–778.
- [166] Rayleigh, 1905. The problem of the random walk. *Nature* 72, 318.
- [167] Reid, S. W., Leake, M. C., Chandler, J. H., Lo, C. J., Armitage, J. P., Berry, R. M., 2006. The maximum number of torque-generating units in the flagellar motor of *Rhodobacter sphaeroides* is at least 11. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8066–8071.
- [168] Roberts, S., Guilford, T., Rezek, I., Biro, D., 2004. Positional entropy during pigeon homing I: application of Bayesian latent state modelling. *J. Theor. Biol.* 227, 39–50.
- [169] Rodriguez, A., Shotton, D. M., Guil, N., Trelles, O., 2000. Object tracking and event recognition in biological microscopy videos. In: 6th RIAO Conference on Content-Based Multimedia Information Access.
- [170] Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* 27, 832–837.
- [171] Rowe, A. D., Leake, M. C., Morgan, H., Berry, R. M., 2003. Rapid rotation of micron and submicron dielectric particles measured using optical tweezers. *J. Mod. Opt.* 50, 1539–1554.
- [172] Rupp, K., Grasser, T., Jüngel, A., 2012. Deterministic numerical solution of the Boltzmann transport equation. In: Günther, M., Bartel, A., Brunk, M., Schöps, S., Striebel, M. (Eds.), Progress in Industrial Mathematics at ECMI 2010. Mathematics in Industry. Springer Berlin Heidelberg, pp. 53–59.
- [173] Ruusuvuori, P., Aijö, T., Chowdhury, S., Garmendia-Torres, C., Selinummi, J., Birbaumer, M., Dudley, A. M., Pelkmans, L., Yli-Harja, O., 2010. Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics* 11, 248+.

- [174] Ryu, W. S., Berry, R. M., Berg, H. C., 2000. Torque-generating units of the flagellar motor of *Rhodobacter sphaeroides* have a high duty ratio. *Nature* 403, 444–447.
- [175] Sager, B. M., Sekelsky, J. J., Matsumura, P., Adler, J., 1988. Use of a computer to assay motility in bacteria. *Anal. Biochem.* 173, 271–277.
- [176] Saxton, M. J., Jacobson, K., 1997. Single-particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* 26, 373–399.
- [177] Schliep, A., Schönhuth, A., Steinhoff, C., 2003. Using hidden Markov models to analyze gene expression time course data. *Bioinformatics* 19, 255–263.
- [178] Segall, J. E., Block, S. M., Berg, H. C., 1986. Temporal comparisons in bacterial chemotaxis. *Proc. Natl. Acad. Sci. U. S. A.* 83, 8987–8991.
- [179] Selmeczi, D., Li, L., Pedersen, Nrelykke, S. F., Hagedorn, P. H., Mosler, S., Larsen, N. B., Cox, E. C., Flyvbjerg, H., 2008. Cell motility as random motion: A review. *Eur. Phys. J. Spec. Top.* 157, 1–15.
- [180] Serge, A., Bertaux, N., Rigneault, H., Marguet, D., 2008. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* 5, 687–694.
- [181] Shi, Y., 2001. Effects of thermal fluctuation and the receptor-receptor interaction in bacterial chemotactic signaling and adaptation. *Phys. Rev. E* 64, 021910+.
- [182] Shum, H., Gaffney, E. A., 2012. The effects of flagellar hook compliance on motility of monotrichous bacteria: A modeling study. *Phys. Fluids* 24, 061901+.
- [183] Shum, H., Gaffney, E. A., Smith, D. J., 2010. Modelling bacterial behaviour close to a no-slip plane boundary: the influence of bacterial geometry. *Proc. Roy. Soc. A* 466, 1725–1748.
- [184] Silverman, M., Simon, M., 1974. Flagellar rotation and the mechanism of bacterial motility. *Nature* 249, 73–74.

- [185] Skellam, J. G., 1951. Random dispersal in theoretical populations. *Biometrika* 38, 196–218.
- [186] Slovak, P. M., Wadhams, G. H., Armitage, J. P., 2005. Localization of MreB in *Rhodobacter sphaeroides* under conditions causing changes in cell shape and membrane structure. *J. Bacteriol.* 187, 54–64.
- [187] Sourjik, V., Berg, H. C., 2002. Receptor sensitivity in bacterial chemotaxis. *Proc. Natl. Acad. Sci. U. S. A.* 99, 123–127.
- [188] Sowa, Y., Rowe, A. D., Leake, M. C., Yakushi, T., Homma, M., Ishijima, A., Berry, R. M., 2005. Direct observation of steps in rotation of the bacterial flagellar motor. *Nature* 437, 916–919.
- [189] Sowa, Y., Steel, B. C., Berry, R. M., 2010. A simple backscattering microscope for fast tracking of biological molecules. *Rev. Sci. Instrum.* 81, 113704–113709.
- [190] Taboada, B., Poggio, S., Camarena, L., Corkidi, G., 2003. Automatic tracking and analysis system for free-swimming bacteria. In: Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE. Vol. 1. IEEE, pp. 906–909.
- [191] Takata, T., Fujimoto, S., Amako, K., 1992. Isolation of nonchemotactic mutants of *Campylobacter jejuni* and their colonization of the mouse intestinal tract. *Infect. Immun.* 60, 3596–3600.
- [192] Taylor, B. L., Zhulin, I. B., Johnson, M. S., 1999. Aerotaxis and other energy-sensing behavior in bacteria. *Annu. Rev. Microbiol.* 53, 103–128.
- [193] Thar, R., Blackburn, N., Kuhl, M., 2000. A new system for three-dimensional tracking of motile microorganisms. *Appl. Environ. Microbiol.* 66, 2238–2242.
- [194] Thattai, M., van Oudenaarden, A., 2004. Stochastic gene expression in fluctuating environments. *Genetics* 167, 523–530.

- [195] Tindall, M. J., Maini, P. K., Porter, S. L., Armitage, J. P., 2008. Overview of mathematical approaches used to model bacterial chemotaxis II: bacterial populations. *Bull. Math. Biol.* 70, 1570–1607.
- [196] Tindall, M. J., Porter, S. L., Maini, P. K., Armitage, J. P., 2010. Modeling chemotaxis reveals the role of reversed phosphotransfer and a bi-functional kinase-phosphatase. *PLoS Comput. Biol.* 6, e1000896+.
- [197] Tindall, M. J., Porter, S. L., Maini, P. K., Gaglia, G., Armitage, J. P., 2008. Overview of mathematical approaches used to model bacterial chemotaxis I: the single cell. *Bull. Math. Biol.* 70, 1525–1569.
- [198] Tso, W. W., Adler, J., 1974. Negative chemotaxis in *Rhodobacter sphaeroides*. *J. Bacteriol.* 118, 560–576.
- [199] Underbakke, E. S., Kiessling, L. L., 2010. Classifying chemoreceptors: quantity versus quality. *EMBO J.* 29, 3435–3436.
- [200] Uppaluri, S., Nagler, J., Stellamanns, E., Heddergott, N., Herminghaus, S., Engstler, M., Pfohl, T., 2011. Impact of microscopic motility on the swimming behavior of parasites: straighter trypanosomes are more directional. *PLoS Comput. Biol.* 7, e1002058+.
- [201] Vigeant, M. A., Ford, R. M., 1997. Interactions between motile *Rhodobacter sphaeroides* and glass in media with various ionic strengths, as observed with a three-dimensional-tracking microscope. *Appl. Environ. Microbiol.* 63, 3474–3479.
- [202] Waters, J. C., 2009. Accuracy and precision in quantitative fluorescence microscopy. *J. Cell Biol.* 185, 1135–1148.
- [203] Wilkinson, D. A. *personal communication*.
- [204] Wilkinson, D. A. Regulation of the expression and positioning of chemotaxis and motor proteins in *Rhodobacter sphaeroides*. D.Phil. thesis, University of Oxford.

- [205] Wilkinson, D. A., Chacko, S. J., Vénien-Bryan, C., Wadhams, G. H., Armitage, J. P., 2011. Regulation of flagellum number by FliA and FlgM and role in biofilm formation by *Rhodobacter sphaeroides*. *J. Bacteriol.* 193, 4010–4014.
- [206] Wolfe, A. J., Berg, H. C., 1989. Migration of bacteria in semisolid agar. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6973–6977.
- [207] Wood, T., Allwright, D., Bond, P., Long, S., Moroz, I., 2010. A new method for processing passive sonar data. In: 13th Conference on Information Fusion (FUSION). IEEE, pp. 1–7.
- [208] Wood, T. M., 2011. Interacting methods for manoeuvre handling in the GM-PHD filter. *IEEE T. Aerosp. Elec. Sys.* 47, 3021–3025.
- [209] Wood, T. M., Yates, C. A., Wilkinson, D. A., Rosser, G., 2012. Simplified multi-target tracking using the PHD filter for microscopic video data. *IEEE T. Circuits. Syst. for Video Technology* 22, 702–713.
- [210] Wu, M., Roberts, J. W., Kim, S., Koch, D. L., DeLisa, M. P., 2006. Collective bacterial dynamics revealed using a three-dimensional population-scale defocused particle tracking technique. *Appl. Environ. Microbiol.* 72, 4987–4994.
- [211] Xie, J., Khan, S., Shah, M., 2009. Automatic tracking of *Rhodobacter sphaeroides* in phase-contrast microscopy video. *IEEE T. Bio-Med. Eng.* 56, 390–399.
- [212] Xie, L., Altindal, T., Chattopadhyay, S., Wu, X.-L. L., 2011. Bacterial flagellum as a propeller and as a rudder for efficient chemotaxis. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2246–2251.
- [213] Xue, C., Ju, H., Painter, K. J., Erban, R., 2011. Travelling waves in hyperbolic chemotaxis equations. *Bull. Math. Biol.* 73, 1695–1733.
- [214] Xue, C., Othmer, H. G., 2009. Multiscale models of taxis-driven patterning in bacterial populations. *SIAM J. Appl. Math.* 70, 133–167.

- [215] Yavetz, I., 1995. From Obscurity to Enigma: The Work of Oliver Heaviside, 1872-1891, 1st Edition. Birkhäuser Basel.
- [216] Zaburdaev, V., Schmiedeberg, M., Stark, H., 2008. Random walks with random velocities. *Phys. Rev. E* 78, 011119+.
- [217] Zaid, I. M. *personal communication*.