Modelling collective movement across scales: from cells to wildebeest

Elaine A. Ferguson

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Biodiversity, Animal Health and Comparative Medicine

College of Medical, Veterinary & Life Sciences

University of Glasgow

April 2018

Abstract

Collective movements are ubiquitous in biological systems, occurring at all scales; from the sub-organismal movements of groups of cells, to the far-ranging movements of bird flocks and herds of large herbivores. Movement patterns at these vastly different scales often exhibit surprisingly similar patterns, suggesting that mathematically similar mechanisms may drive collective movements across many systems. The aims of this study were three-fold. First, to develop mechanistic movement models capable of producing the observed wealth of spatial patterns. Second, to tailor statistical inference approaches to these models that are capable of identifying drivers of collective movement that could be applied to a wide range of study systems. Third, to validate the approaches by fitting the mechanistic models to data from diverse biological systems. These study systems included two small-scale *in vitro* cellular systems, involving movement of groups of human melanoma cells and *Dictyostelium discoideum* (slime mould) cells, and a third much larger-scale system, involving wildebeest in the Serengeti ecosystem.

I developed a series of mechanistic movement models, based on advection-diffusion partial differential equations and integro-differential equations, that describe changes in the spatio-temporal distribution of the study population as a consequence of various movement drivers, including environmental gradients, environmental depletion, social behaviour, and spatial and temporal heterogeneity in the response of the individuals to these drivers. I also developed a number of approaches to statistical inference (comprising both parameter estimation and model comparison) for these models that ranged from frequentist, to pseudo-Bayesian, to fully Bayesian. These inference approaches also varied in whether they required numerical solutions of the models, or whether the need for numerical solutions was bypassed by using gradient matching methods. The inference methods were specifically designed to be effective in the face of the many difficulties presented by advection-diffusion models, particularly high computational costs and instabilities in numerical model solutions, which have previously prevented these models from being fitted to data. It was also necessary for these inference methods to be able to cope with data of different qualities; the cellular data provided accurate information on the locations of all individuals through time, while the wildebeest data consisted of coarse ordinal abundance categories on a spatial grid at monthly intervals.

By applying the developed models and inference methods to data from each study system, I drew a number of conclusions about the mechanisms driving movement in these systems. In all three systems, for example, there was evidence of a saturating response to an environmental gradient in a resource or chemical attractant that the individuals could deplete locally. I also found evidence of temporal dependence in the movement parameters for all systems. This indicates that the simplifying assumption that behaviour is constant, which has been made by many previous studies that have modelled movement, is unlikely to be justified. Differences between the systems were also demonstrated, such as overcrowding affecting the movements of melanoma and wildebeest, but not *Dictyostelium*, and wildebeest having a much greater range of perception than cells, and thus being able to respond to environmental conditions tens of kilometres away.

The toolbox of methods developed in this thesis could be applied to increase understanding of the mechanisms underlying collective movement in a wide range of systems. In their current form, these methods are capable of producing very close matches between models and data for our simple cell systems, and also produce a relatively good model fit in the more complex wildebeest system, where there is, however, still some room for improvement. While more work is required to

make the models generalisable to all taxa, particularly through the addition of memory-driven movement, inter-individual differences in behaviour, and more complex social dynamics, the advection-diffusion modelling framework is flexible enough for these additional behaviours to be incorporated in the future. A greater understanding of what drives collective movements in different systems could allow management of these movements to prevent the collapse of important migrations, control pest species, or prevent the spread of cancer.

Acknowledgments

First and foremost, I am grateful to my supervisors, Prof. Jason Matthiopoulos, Prof. Dirk Husmeier, Prof. Robert Insall, and Dr. Grant Hopcraft, for giving me this opportunity and for all of their support over the last four years. Being part of such an enthusiastic, friendly, and academically diverse team has been a joy. Thanks for putting up with this eternal pessimist – I swear I was actually enjoying myself most of the time!

This PhD thesis would not have been possible without the financial support provided by the University of Glasgow's Lord Kelvin/Adam Smith PhD scholarship scheme.

Dr. Luke Tweedy kindly provided the *Dictyostelium* data for analysis. The wildebeest distribution dataset, and the associated rainfall and canopy cover datasets, were made available to me by Prof. Ricardo Holdo, who was extremely helpful in answering my many questions about these data. Dr. Barbara Helm's annual helpful discussions and encouragement were much appreciated.

Thanks to the occupants of room 303, past and present, for always being available when a group moan or celebration was required. The Matthiopoulos collective and the Spatial Interest Group have also been a wonderful source of interesting conversation, and I am indebted to my fellow members of The Wildebeest Action Team (including Lacey Hughey and Colin Torney) for ideas, discussions, and shared experiences in the Serengeti. The Institute of Biodiversity, Animal Health and Comparative Medicine has been a very welcoming and intellectually stimulating place to work, and I look forward to many more collective movements to the pub of a Friday.

My family have offered a huge amount of support throughout, and I am very grateful to Chris for the steady supply of 'real' food and pep talks when things got a little tough. Izzy kept me sane and drove me mad in equal measure during the writing of this thesis – there are lots of long walks coming your way.

Table of Contents

List of Tables

List of Figures

Author's Declaration

I declare that the work presented in the thesis is the result of my own work, except where explicit reference is made to the contribution of others. No part of this thesis has been submitted for any other degree at the University of Glasgow, or any other institution.

The work was funded by a University of Glasgow Lord Kelvin/Adam Smith PhD scholarship, and conducted by myself under the supervision of Jason Matthiopoulos, Dirk Husmeier, Robert Insall and Grant Hopcraft. Chapters 2, 3 and 4 have been published at the following references:

- Ferguson, E.A., Matthiopoulos, J., Insall, R.H. & Husmeier, D., 2016. Inference of the drivers of collective movement in two cell types: Dictyostelium and melanoma. Journal of the Royal Society Interface, 13(123), 20160695.
- Ferguson, E.A., Matthiopoulos, J., Insall, R.H. & Husmeier, D., 2017. Statistical inference of the mechanisms driving collective cell movement. Journal of the Royal Statistical Society. Series C: Applied Statistics, 66(4), pp.869–890.
- Ferguson, E.A., Matthiopoulos, J. & Husmeier, D., 2017. Constructing wildebeest density distributions by spatio-temporal smoothing of ordinal categorical data using GAMs. In 32nd International Workshop on Statistical Modelling. Groningen, Netherlands, pp. 70–75.

None of the data analysed were collected by myself; references to the researchers involved in the data collection and to previous publications of these data are provided in the text. To summarise, Luke Tweedy (Tweedy et al. 2016) provided the *Dictyostelium* data analysed in chapters 2-3. The melanoma data of chapter 2 were collected by Muinonen-Martin et al. (2014). Ricardo Holdo provided access to the wildebeest distribution dataset analysed in chapters 4-5, which was collected by the Serengeti Ecological Monitoring Programme (SEMP)). In chapter 5, the rainfall data was collected by Holdo et al. (2009), the grass nitrogen data by Hopcraft et al. (2014) and the canopy cover data by Norton-Griffiths (1979), Reed et al. (2009), and Frankfurt Zoological Society and Harvey Maps (2010)

Elaine Ferguson

Elaine Ferguson

April 2018

1. Introduction to collective movement

Collective movement is widespread at all scales in biological systems; from the sub-organismal movements of groups of cells in the body during the processes of embryonic morphogenesis, wound healing and cancer metastasis (Friedl and Gilmour 2009, Rorth 2009), to the movements of herds of large herbivores around whole ecosystems (Fryxell and Sinclair 1988) and the bird migrations that traverse continents (Hahn et al. 2009). Despite the vastly different scales at which these movements occur, they exhibit some surprisingly similar patterns. For example, a phase transition from disordered movement to aligned, directional movement as the density of interacting individuals increases has been observed in systems ranging from locusts and glass prawns (Buhl et al. 2006, Mann et al. 2013), to bacteria and fish tissue cells (Szabó et al. 2006, Sokolov et al. 2007). Such similarities in movement behaviour, despite very different social and cognitive abilities, beg the question of common causality. In this introductory chapter, I introduce the various types of mechanisms that have been proposed as drivers of collective movement, and the methods that have been used to model these mechanisms and infer their presence in various biological systems. I conclude by outlining the aims and structure of this thesis.

1.1. Mechanisms driving collective movement

The precise mechanisms leading to collective movement behaviour may vary from system to system, but four broad categories of such mechanisms can be distinguished; environmental variability, environmental depletion, interactions between individuals, and memory. Many systems may involve mechanisms from more than one of these categories. The different mechanisms or their relative contributions to the emergent patterns of movement may also change temporally or spatially, as has been indicated by many examples of seasonal or state-based movement (Bonner 1982, Morales et al. 2004, Hopcraft et al. 2014).

1.1.1. Environmental Variability

Collective movement may emerge as a result of individuals responding in similar ways to spatiotemporally varying environments, such that each individual tracks the most favourable conditions and thus increases its fitness. In some ecosystems, certain resources vary predictably in time and space along environmental gradients. For example, in the Serengeti ecosystem, a declining rainfall gradient from north to south occurs alongside an opposing gradient of declining plant nutritional quality from south to north. Wildebeest respond to these gradients by following the nutritional gradient south for the wet season and then following the rainfall gradient back north for the dry season, when conditions in the south deteriorate (Holdo et al. 2009). Tracking of environmental gradients is also observed in some zooplankton species, which move down gradients of ultraviolet radiation and predation risk to deeper waters during the day, and then follow the gradient in algal food abundance back up the water column at night, when ultraviolet radiation and predation risk at the surface are lower (Hansson and Hylander 2009). Situations like these, where the environment varies predictably, and all organisms have a similar response to this variation, produce predictable and periodic migratory patterns. In cases where the spatiotemporal distribution of resources varies unpredictably, however, tracking of favourable conditions can lead to nomadic movement patterns that are irregular in time and space (Jonzén et al. 2011). Nomadism is exhibited by a number of bird species in arid environments, where rainfall above a critical level triggers the arrival of nomadic birds in an area. These nomads exploit the temporarily abundant local resources to breed before moving on (Dean et al. 2009). Tracking of environmental gradients is also widely observed in cell systems through the process of chemotaxis, which involves cells detecting and biasing their direction of movement in response to gradients in certain chemicals known as chemoattractants (Insall 2010, Coburn et al. 2013).

The environment can also drive movement in cases where two essential resources, such as breeding sites and foraging sites, are geographically separated (Börger et al. 2011). In some taxa, such as seabirds moving between foraging sites and nesting colonies, this scenario can lead to animals making regular commutes (daily, or every few days) over relatively long distances (Dingle and Drake 2007). In other cases, migrations between sites that offer alternative resource types occur over much longer time scales. Minke whales, for example, are capital breeders that build up energy reserves during the summer in their northern feeding grounds, before migrating to their less productive breeding grounds in equatorial waters during the winter (Christiansen et al. 2013). Seasonally utilised sites may be separated by unsuitable habitat, so that following an environmental gradient in the favourability of conditions is no longer a viable strategy for reaching the destination. Navigation between such sites may instead involve individuals responding to the sun, stars and Earth's magnetic field (Cochran et al. 2004), once cues, such as day length, have informed them that the time to switch sites has arrived (Gwinner 1996).

1.1.2. Environmental Depletion

While, as described in section 1.1.1, organisms may move in response to spatio-temporally varying environmental conditions that are generated externally (for example, by weather patterns), they may also cause these variations in habitat favourability themselves through local depletion of resources. Depletion-driven movements will be amplified by the presence of conspecifics, since multiple individuals exploiting the same resource are more likely to deplete it than a single animal, forcing all the individuals to move on. If there are only a limited number of alternative habitat patches to exploit, or if interactions between the individuals occur (see section 1.1.3), these onward movements are likely to be collective (Börger et al. 2011). An example of resource depletion driving movement is found in Mormon crickets, which show a greater propensity to move when they are protein deprived (Simpson et al. 2006). There is also evidence that wildebeest in the Serengeti move further when at higher densities, perhaps suggesting greater depletion of resources by larger concentrations of animals as a factor in their movement behaviour (Thirgood et al. 2004, Harris et al. 2009, Hopcraft et al. 2014, 2015). Local depletion of chemoattractants, leading to the creation of detectable chemical gradients, has also been identified as a driver of movement in a number of cellular systems. Gradients in the chemical LPA (lysophosphatidic acid), which can be broken down by and is attractive to melanoma cells, for example, were recently discovered *in vivo* around cutaneous tumours (Muinonen-Martin et al. 2014). It is likely that such gradients are only able to develop when the cells reach high enough densities (i.e. when the tumour reaches a large enough size), and this may be the reason that the probability of melanoma recurrence or metastasis following surgical removal of a tumour has been found to depend heavily on the thickness of the tumour removed (Breslow 1970, Owen et al. 2001); the larger the tumour, the stronger the LPA gradient around it, and, thus, the greater the drive for cells to migrate out from the tumour and cause subsequent metastases. A similar role of self-generated chemical gradients has been proposed for the migration of the cells of the lateral line primordium during zebrafish embryonic development (Donà et al. 2013, Venkiteswaran et al. 2013).

1.1.3. Interactions between individuals

Environmental variability has the potential to drive organisms seeking the same environmental conditions to move collectively in the absence of interactions between the individuals, while environmental depletion can drive such movements using just indirect density-dependent interactions. However, direct interactions between individuals can also be an important component of movement behaviour, with studies indicating that simple attraction, repulsion or alignment rules occur between individuals in a wide range of taxa. In locusts, for example, a spontaneous switch from solitary to gregarious behaviour occurs when conditions become crowded, leading to scarce resources (Simpson et al. 2001). Since locusts also become cannibalistic under conditions of limited resources (hence the draw of conspecifics as a food source), this attraction is coupled with a tendency for individuals to be repelled by any conspecifics approaching from behind, and to align their direction of movement with neighbouring individuals, in an effort to avoid being bitten, resulting in directional collective movement (Bazazi et al. 2008). Attraction and repulsion dynamics have also been inferred for golden shiners, where each fish is repulsed by conspecifics that come too close, but attracted to conspecifics that are further away, allowing maintenance of an inter-individual distance that is large enough to prevent collisions and small enough to produce a cohesive school (Katz et al. 2011). The combination of these forces of attraction and repulsion also leads to the coordinated and aligned movement of the school. Similar short-range repulsion and longer-range attraction has been observed in flocking surf scoters, which also exhibit explicit alignment interactions at intermediate distances (Lukeman et al. 2010). In cell systems, movement-inducing interactions can result from the release and receipt of chemical signals by the cells, which may be of the same or different types. For example, breast tumour cells release colony-stimulating factor 1, which attracts macrophages, and the macrophages in turn release epidermal growth factor, which stimulates movement of the tumour cells, potentially facilitating tissue invasion and metastasis (Wyckoff et al. 2004). Cell-cell interactions may also occur through direct contact, as in the case of contact inhibition of locomotion, where moving cells that come into contact will collapse the protrusions that they use to produce movement at the site of contact, and then move off in a new direction (Mayor and Carmona-Fontaine 2010).

A commonly observed movement phenomenon in systems of interacting individuals is a phase transition from disordered to ordered directional movement as density increases. This transition, and the critical density at which it occurs, has been observed under experimental conditions in locusts (Buhl et al. 2006), keratocytes (Szabó et al. 2006) and glass prawns (Mann et al. 2013). Sokolov et al. (2007) also observed a shift from individual to collective movement behaviour in swimming bacteria, though in this case the transition was more gradual, possibly due to random noise in the orientation of individuals. Simulations from self-propelled particle (SPP) models (see section 1.2.2 for a description of this class of models) have indicated that this transition to ordered movement can be replicated through simple attraction, repulsion or alignment rules between individuals (Vicsek et al. 1995, Buhl et al. 2006, Szabó et al. 2006), like those in the systems described in the previous paragraph This emergence of coordinated directional group movement from simple interaction rules at high densities occurs even in the absence of directional environmental cues.

When a directional environmental cue is present and each individual has some degree of error in detecting this cue, SPP models have also indicated that the average ability of an individual to accurately follow the most favourable conditions or reach a target location is improved when it interacts with its neighbours through attraction and alignment relative to when it navigates without such interactions, using only its own flawed assessment of the environment (Grünbaum 1998, Codling et al. 2007). This improved navigation in the presence of interactions occurs as a result of an averaging of the individuals' imperfect directional preferences, an effect known as the 'many-wrongs principle' (Simons 2004). An experiment by Berdahl et al. (2013) provides evidence of this effect in golden shiners, which are more successful in tracking their preferred patches of low light level across a tank when in larger groups. Similarly, shoals of lake whitefish have been found to be more responsive than individual fish in their avoidance of a toxic cadmium gradient (McNicol et al. 1996). In cases where certain individuals are better able to determine the appropriate direction than others (for example, due to greater age and experience), and the group members are able to recognise these differences in ability, then each individual may not be weighted equally in the choice of group direction. An example of this is found in whooping cranes, where the accuracy of an individual during migration is dependent on the age of the oldest individual in its flock, but not on the individual's own age, or the group size (Mueller et al. 2013), despite the fact that accuracy is expected to increase with group size when individuals are given equal weighting (Grünbaum 1998, Codling et al. 2008, Berdahl et al. 2013). This suggests that the directional preference of the flock is dictated by the most experienced individual. Interactions between individuals also play a role in groups reaching consensus decisions on the direction of movement in cases where there are two subsets of individuals that have different preferences. In baboons, it has been observed that if the difference in preferred directions is small, the group tends to follow a trajectory that is an average of these preferences. However, if there is a large difference in the preferred directions, the baboons will typically choose the direction preferred by the majority or, if there is equal support for both directions, will choose one of them at random (Strandburg-Peshkin et al. 2015). These findings for decision-making in baboons agreed with predictions previously made by simulations from SPP models (Couzin et al. 2005).

1.1.4. Memory

A final driver that is likely to be important in shaping movement patterns is memory. Memories may be obtained through experience by an initially naïve individual passing through different locations and remembering their quality and position. Bison, for example, remember the location and quality of patches of meadow within forest habitat, and use this knowledge to select meadows that they have previously visited and that are of a higher profitability than those they have visited in their most recent foraging efforts (Merkle et al. 2014). In other cases, memories are genetic, being passed through generations, and causing individuals to be pre-programmed to move to a certain location seasonally, even if they have never previously visited that location. This is observed in some bird species that are able as juveniles to successfully migrate to the correct area at the correct time, without guidance from experienced individuals, due to their genetic knowledge of the direction in which they should fly and the distance for which they should continue (Helbig 1996). Other species may need a combination of genetic and learned memory. Juvenile whooping cranes that had been reintroduced into a location with no experienced adults had to learn their first southwards migration by following an ultralight aircraft. However, the same juvenile cranes were then able to initiate their first successful northwards migration independently the following spring, suggesting at least some genetic influence (Urbanek et al. 2005, Mueller et al. 2013). The use of memory to guide movement is expected to be most advantageous in cases where the landscape does not change rapidly over time, rendering memories useless, and where the landscape is of intermediate complexity, since memories are unnecessary in a homogeneous environment and costly to maintain in a highly heterogeneous one (Fagan et al. 2013). Single cells, unlike vertebrates do not have a well-developed brain in which to store memories, so we might not expect to observe any influences of learned memory in cellular systems. Cells do exhibit behaviours that could be considered to represent types of memory, however, such as their tendency to persist in their direction of travel (i.e. perform a correlated random walk; see section 1.2.1), even in the absence of any directional cue (Bosgraaf and Van Haastert 2009). Another example of cell memory occurs during cell differentiation, where a precursor cell exposed to short-term signals permanently becomes more specialised, as though it retains a memory of the conditions that caused the specialisation (Ajo-Franklin et al. 2007)

Memories may be retained by an individual for long periods of time. Genetic memories in particular will be retained for generations beyond the lifetime of an organism, but learned memories can also lead to persistent behaviours, such as breeding and foraging site fidelity, where an individual will return to the same location year after year. Turtles and salmon, for example, are believed to imprint on signatures (such as the magnitude and inclination) of the Earth’s magnetic field at their natal sites and then use this imprint to navigate back to this natal site as breeding adults years later (Lohmann et al. 2008). Returning to a site that has proven successful in previous years is advantageous in that it prevents unnecessary energy expenditure on searching for new sites. However, if the target site changes in some way, and individuals do not adapt to these changes, simply continuing to move to the same location at the same time every year, these movements can become maladaptive. Such failure to alter migratory behaviour has been observed in a number of bird species, where populations have been unable to adjust the timing of their migration in response to climate change-induced changes in the timing of peak resource abundance (Visser and Both 2005). In other cases, memories are much more short-lived. Glass prawns, for example, remember and may change direction in response to other conspecifics that they encountered travelling in the opposite direction, but these memories have a half-life of only around one second (Mann et al. 2013).

1.2. Models and inference

A wide range of models have been proposed for describing how collective movement emerges from the mechanisms described in section 1.1. Many of these models have been shown to produce movement behaviour that is at least qualitatively similar to that observed in real systems, and a growing number of studies are also attempting to statistically fit these models to data and use model comparison techniques to infer the drivers of movement in these systems. Here, I introduce a selection of movement model classes that have been particularly popular, and which are general and flexible enough that they can be applied to different systems. Consequently, I exclude from my presentation highly specialised movement models describing, for example, the dynamics of cell protrusions (Neilson et al. 2011, Coburn et al. 2013), which may be apt for investigating cell movement, but are unsuitable for application to the movement of large mammals.

1.2.1. Random walk models

Random walks are among the most commonly used methods for modelling movement in a diverse range of settings, as is evident by their use to describe movements of cells (Hall 1977, Tweedy et al. 2016), mice (Blackwell 1997) and various large ungulates (Morales et al. 2004, Hopcraft et al. 2014, Langrock et al. 2014), to name but a few. This popularity results, in part, from the way that they intuitively describe movements of individuals through time as a stochastic series of steps. Given that individual-based movement data typically take the form of a series of locations at different points in time, between which steps can be inferred, random walk models are particularly suited to the analysis of such data. These models are also very flexible. At their simplest, they can describe Brownian motion (or diffusion), where movement is uncorrelated (the direction is not influenced by the direction at past time points) and unbiased (there is no preference for a particular direction). However, animals and cells typically do not move via pure diffusion, and a number of extensions to this basic model have been developed that allow description of more realistic movement patterns through combined processes of diffusion and drift (or advection) (Codling et al. 2008). A few of these extensions include the correlated random walk (each step tends to be in a similar direction to the previous one), the biased random walk (movement is biased in a particular direction) (Codling et al. 2008), and the Ornstein-Uhlenbeck process (a form of biased random walk where movement is biased towards a particular point, with the strength of attraction to this point increasing with distance) (Blackwell 1997).

Studies using random walks to model movements of large herbivores (Morales et al. 2004, Haydon et al. 2008, Hopcraft et al. 2014, Langrock et al. 2014) have described the movement between each pair of successive time points in terms of the step length and the turning angle relative to the previous step. These step lengths and turning angles are drawn from specified distributions; e.g. gamma or Weibull distribution for step length and von Mises or wrapped Cauchy distribution for turning angle (Langrock et al. 2012). While the various extensions of the basic random walk have improved our ability to describe short-term movement patterns, applying a single random walk with non-changing step length and turning angle distributions is unlikely to be realistic in the long term, since animals tend to change their movement behaviour as they move between different habitats and interact with other individuals (Morales et al. 2004). To address this problem, models composed of mixtures of random walks have been developed, where each walk within the mixture may have different step length and turning angle distributions, and different sources of bias. These walks each describe an unobserved behavioural state underlying the observed movement pattern, e.g. foraging vs. ranging or grouped vs. solitary (Morales et al. 2004, Haydon et al. 2008, Langrock et al. 2014). Individuals can switch between any two behavioural states with specified probabilities, described using a transition matrix (Morales et al. 2004, Langrock et al. 2014). Since these changes in behavioural state are likely driven by factors such as habitat type or interactions between individuals, the probability of switching can be expressed as a function of these factors (e.g. the probability that a migratory animal becomes encamped increases with the quality of the habitat) (Morales et al. 2004, Haydon et al. 2008). Responses to the environment and conspecifics can also be introduced through biases in the direction of movement (Langrock et al. 2014). An alternative to imposing a small number of discrete movement states is to allow the turning angle and step length distribution parameters to vary continuously with environmental variables (Hopcraft et al. 2014). In addition, realistic individual-level variation in movement behaviour can be introduced by allowing individuals to vary in the parameter values describing their step lengths, turning angles and responses to environmental factors (Hopcraft et al. 2014).

Several studies have carried out parameter inference and model comparison for random walk models using movement data from real systems. In some cases parameter inference has been achieved through likelihood maximisation, with comparison of different candidate model for movement being achieved using AIC (Akaike Information Criterion) (Langrock et al. 2012, 2014). AIC, like all information criteria, favours models with a high goodness of fit, while imposing a penalty for the number of parameters required to achieve this fit (Akaike 1974). In other studies, Bayesian approaches to model inference based on MCMC (Markov Chain Monte Carlo) algorithms have been adopted (Blackwell 1997, Morales et al. 2004, Hopcraft et al. 2014). Bayesian inference has two main advantages: 1. it allows prior information about values of the model parameters to be accounted for; 2. by estimating the full posterior probability distributions, it gives a better description of the uncertainty around the estimated parameter values. Model comparison for random walk models in a Bayesian framework has been achieved using DIC (Deviance Information criterion) (Spiegelhalter et al. 2002, Morales et al. 2004, Hopcraft et al. 2014).

Random walk models are a potentially very flexible modelling approach. Social, environmental and memory-based drivers can be incorporated through effects on behavioural switching, step lengths and turning angles, and biases in direction (movement could for example be biased up a local gradient in environmental quality). However, previous random walk models have typically only included one or two of these three movement drivers. Environmental depletion mechanisms have rarely been incorporated into random walk models; I am aware of just one example that described cellular movement in response to a gradient in chemoattractant that is self-generated through depletion, and this model was used for qualitative comparison with data, rather than being formally fitted (Tweedy et al. 2016). Studies that account for potential seasonal changes in the parameters of random walk models, and not just state-switching behaviour at short time intervals, are also uncommon, though there are some examples: Hopcraft et al. (2014) illustrated differences in the movement decisions of wildebeest between their wet and dry season ranges by fitting a random walk model to data from each of the two ranges separately. One limitation of random walk models for studying collective movement is that fitting them requires data where the movement of the individuals has been followed through time. If we want to use these models to describe a field system involving a large number of interacting individuals, collecting the necessary data may be infeasible. GPS tags, for example, are too expensive to deploy in large numbers (Hebblewhite and Haydon 2010), and fitting them to all the individuals in a group would be time consuming and highly disruptive. Large numbers of moving individuals could be recorded simultaneously by hovering small, inexpensive drones with video cameras attached over collectively moving groups, but flight times for such drones are typically in the region of minutes (Anderson and Gaston 2013), which in many cases will not be long enough to give an accurate description of the full spectrum of movement behaviour.

1.2.2. Self-propelled particle models

Self-propelled particle (SPP) models could technically be considered a sub-class of biased random walk models, where the direction of the bias of each individual in a group is informed by the position and/or heading of its neighbours. However, here I consider these models as a separate class due to the large volume of collective movement literature that has built up around them, and because these models typically consider interactions between individuals that are more complex than in the average random walk model, with each individual being able to affect the movement of every other individual using a set of rules at every time step. SPP models are typically described by systems of difference equations, where each individual’s location at a particular time step is determined from its position and velocity at the previous time step. The direction of movement of an individual at each time step was described in the original Vicsek model (Vicsek et al. 1995) as resulting from an alignment rule, whereby the individual moves in the average direction taken by the other individuals within an interaction radius in the previous time step, with the addition of a random noise term. Later models have included additional interaction rules by modelling concentric zones of repulsion, alignment and attraction around each individual (Couzin et al. 2002). While SPP models have most commonly described an individual’s neighbours as all other individuals occurring within these fixed spatial zones, sometimes with an assumption that repulsive interactions take precedence to avoid collisions (Couzin et al. 2002, Szabó et al. 2006, Lukeman et al. 2010), alternatives, such as the restriction of interactions to a fixed number of nearest neighbours have also been developed (Ballerini et al. 2008, Mann 2011, Mann et al. 2013). Blind zones can also be incorporated, so that an individual only responds to individuals within its field of vision (Couzin et al. 2002, Lukeman et al. 2010). Traditionally, SPP models have been Markovian, assuming that an individual’s choice of direction is dependent only on information derived from individuals encountered during the current time step, but more recently, non-Markovian models have been used to incorporate the influence of memories of past neighbour encounters at earlier time steps (Mann et al. 2013). SPP models have also tended to use the simplification that all individuals travel at the same constant speed, but following on from the observation that interactions between individuals can involve changes in speed, as well as changes in direction (Katz et al. 2011), variable speed models have also been considered (Mishra et al. 2012, Berdahl et al. 2013). As mentioned in section 1.1.3, SPP models have been used to demonstrate that simple attraction, repulsion and/or alignment dynamics can produce the phase transition from disordered to ordered movement that has been observed with increasing density of interacting organisms in a wide range of systems (Vicsek et al. 1995, Buhl et al. 2006, Szabó et al. 2006, Sokolov et al. 2007, Mann et al. 2013).

The effect of an environmental gradient or a remembered location can easily be added to an SPP model as an additional bias on the direction of movement. The direction chosen by an individual at a given time point is then a summation of the preferred direction based on the gradient/memory, the preferred direction based on interactions with neighbours, and a random noise term. A weighting can also be applied to each directional preference, describing the priority that an organism gives each movement cue (Couzin et al. 2005, Codling et al. 2007, Lukeman et al. 2010). Alternatively the environmental cue can be assumed to alter the speed of the individual (Berdahl et al. 2013). Simulations from such models have been used to demonstrate the ‘many wrongs’ principle (Grünbaum 1998, Codling et al. 2007, Berdahl et al. 2013), to indicate that a small number of informed individuals can accurately lead a large group of uninformed individuals (Couzin et al. 2005), and to show the different consensus decisions that arise in cases where individuals within a group differ in their directional preferences and in the strength of these preferences (Couzin et al. 2005, 2011). Crucially these patterns of group movement behaviour predicted from simulations have also been observed in data from lab and field systems (Reebs 2000, Couzin et al. 2011, Strandburg-Peshkin et al. 2015).

While simulations from SPP models have shown them to be able to qualitatively reproduce the dynamics of group movement behaviour in a wide range of systems, statistical inference has rarely been used to fit these models to data, so that the validity of the proposed underlying interaction mechanisms has not been fully tested. A small number of studies, however, have begun to tackle this problem. Lukeman et al. (2010) fitted a subset of the parameters of a set of candidate SPP models for describing the behaviour of surf scoter flocks using an optimisation approach. This approach minimised the difference between characteristic functions calculated both from the data and from simulations from a particular model with a given parameter set. Mann et al. (2013) carried out model selection to identify the social interaction mechanisms underlying changes in the direction of glass prawns moving clockwise or anti-clockwise around a ring-shaped arena. In this case, a Bayesian approach was used to calculate the marginal likelihoods of the data given each model; the marginal likelihood is a statistic that inherently accounts for model complexity, and so can be used to select a best model. The most probable parameter values from the posterior were then used to simulate from each model and calculate the Kullback-Leibler divergence (Kullback and Leibler 1951) of the distribution of the proportion of prawns moving clockwise obtained in the simulations from that for the data. The marginal likelihood and Kullback-Leibler divergence based tests both supported the same best model. A third study involving model inference for SPP models is Mann (2011), where a Bayesian approach was used to select the correct model for simulated datasets based on Bayes factors.

The focus of research using SPP models has typically been on the social drivers of movement behaviour, with the description of environmental effects being kept at a very basic level, often just involving a single fixed gradient or target location. To my knowledge, the literature does not currently contain any SPP models that have included environmental depletion or temporal changes in behaviour. Individual variation in behaviour can be easily introduced to these models, but, as a result of their individual-based nature, they share the limitation of random walk models in that fitting them requires individual-based data. Much of the SPP model literature has been focussed on finding qualitative agreement between simulation outcomes and behaviour in real systems, and there is a need for further development of formal statistical inference for this class of models.

1.2.3. Advection-Diffusion Models

Advection-diffusion (also known as convection-diffusion) equations are a type of partial differential equation (PDE) that describe changes in the density of moving organisms in space and time, as a result of the combined processes of advection (directional movement) and diffusion (random movement). They are essentially the deterministic counterpart of random walks; the output of a simulation of one or more individuals from a particular random walk is a stochastic movement path or set of stochastic movement paths, while the output of a simulation from the corresponding advection-diffusion PDE is the density of individuals that we expect to see at every point in space and time given the random walk (if we simulated the individuals from the random walk many times, we would converge on the distribution from the advection-diffusion model) (Moorcroft and Lewis 2006). Advection-diffusion models can, therefore, incorporate all the environmental, social and memory biases on movement direction and speed that we can incorporate into random walk models (see section 1.2.1) via their advection and diffusion coefficients (see section 2.3 for a mathematical description of these models), making them similarly flexible. They have been widely used to describe movement behaviour in systems of, for example, cells (Keller and Segel 1970, Hillen and Painter 2009), coyotes (Moorcroft et al. 2006), caribou (Fortin et al. 2013) and tuna (Sibert et al. 1999).

Advection-diffusion models are a population-based modelling approach, and this gives them an advantage over the individual-based modelling approaches described in sections 1.2.1-1.2.2, in that the computational cost of simulating a density surface from these models does not increase with group size, while every additional individual in an individual-based approach requires computation of a new movement path. Introducing individual variation in behaviour, however, is less intuitive in a population-based framework, though behavioural differences between group members could be incorporated by splitting the modelled population into sub-populations that are each described by their own advection-diffusion PDE with its own movement parameters and mechanisms. Movement of organisms between these groups, representing changes in individual state could also be included in such a framework; an approach that is widely used in compartmental models in epidemiology to describe susceptible, infected and recovered groups within a population (Ross 1911; Kermack & McKendrick 1927; see Brauer (2008) for a more recent introduction).

The role of environmental depletion in driving collective movement has been considered in advection-diffusion models far more often than in the individual-based modelled approaches described above. This has primarily been through studies in the cellular literature. The popular Keller-Segel model, which describes the aggregation of cells in response to a spatial gradient in a chemoattractant that they can both release into the environment and deplete from the environment through the release of an enzyme that breaks down the attractant (Keller and Segel 1970). The depletion mechanism is incorporated by modelling the concentrations of the chemoattractant and the enzyme that breaks it down using additional PDEs. The enzyme's equation includes a term describing how it increases with cell density, and the chemoattractant's equation has a term describing how it decreases with increasing enzyme concentration.

Rarely have studies attempted to fit advection-diffusion models to data. This is, in part, a result of the need to solve these models numerically for each parameter set for which we wish to calculate a likelihood during parameter optimisation or MCMC sampling. Numerical integration can be computationally costly, and in the case of advection-diffusion equations is also hampered by instabilities in the model solution that occur when advection dominates over diffusion and can halt inference procedures prematurely (Sibert et al. 1999). However, the development of methods in the statistical literature, such as gradient matching (Macdonald & Husmeier 2015; Xun et al. 2013; chapter 5 of this thesis), that bypass the need for numerical solution is promising. Despite the numerical difficulties, a small number of ecological studies have been successful in carrying out inference for advection-diffusion models. A maximum likelihood approach was used to infer the parameters of a model describing tuna movement (Sibert et al. 1999). This model also accounted for spatial and temporal variation in the parameters describing the rates of advection and diffusion, something that has been relatively rare in other studies. Inference of the mechanism driving the distribution of coyote packs was also achieved using maximum likelihood, with AIC being used to select the best candidate model (Moorcroft et al. 2006). Hierarchical Bayesian approaches to inference have also been demonstrated on data describing the invasion of North America by the Eurasian collared-dove (Wikle and Hooten 2006, Cressie and Wikle 2011).

1.3. Aims and structure of this thesis

The preceding review of the mechanisms producing collective movement and the methods used to model them, highlights a number of areas where further work is required. First, while efforts to understand collective movement are increasing in number, the majority of studies have not used formal statistical inference (comprising parameter optimisation and model selection) to infer movement drivers in real systems, instead just presenting results from simulations or showing that models are qualitatively capable of reproducing the patterns observed in a particular system. Second, of the models that have been fitted to data, most have only looked at one or two of the potential types of movement drivers described in section 1.1., while I anticipate that many systems, particularly in the field, will involve a greater range of these mechanisms. Third, environmental depletion mechanisms have been particularly poorly studied as a movement driver, with few studies even simulating from models involving such effects. Finally, the majority of modelling studies have not included temporal or spatial variation in the parameters and mechanisms describing movement behaviour (though we recognise the work that has been done to incorporate state-switching behaviour into random walk models), despite the fact that, outside of the lab, it is likely that most systems are subjected to seasonal conditions that could heavily impact behaviours. I aimed to address these issues by:

1. Developing models that incorporate a wide range of movement mechanisms, including environmental depletion, and temporal and spatial variation in movement behaviour
2. Developing and adapting methods for fitting these models to data
3. Applying these fitting methods to infer parameter values for a range of candidate models, using data from a range of study systems at different scales (see below)
4. Using model selection to select the most parsimonious model (i.e. the model that best balances quality of fit to the data and the number of parameters) for each study system, thus inferring the movement mechanisms most likely to be influencing these systems

The models presented in this thesis are based on advection-diffusion partial differential equations, largely because this population-based approach does not require individual-based data for inference, and such data were unavailable for one of the study systems considered (wildebeest, see below). Additional reasons for selecting advection-diffusion models included the potential computational gains made by not having to compute a movement path for every individual when simulating from the model, and also the need for the development of new inference methods for these models that can be effective in the face of numerical stability issues.

The aims outlined above are addressed with respect to data from three study systems (which are described fully in subsequent chapters). These include two lab-based cellular systems involving the slime mould *Dictyostelium discoideum* and human melanoma. *Dictyostelium* is an organism that exists as both a single celled amoeba and a multicellular aggregate at different stages in its development (Bonner 1982), and that has emerged as a model organism for eukaryotic cell movement (Carnell and Insall 2011). Melanoma is a particularly aggressive cancer as a consequence of the rapidity with which it can spread (Balch et al. 2009), making an understanding of the mechanisms by which it moves crucial. The raw data in both of the cellular study systems is in the form of time series of microscopy images. The third system considered is wildebeest (*Connochaetes taurinus*) movement in the Serengeti. The data for this system takes the form of ordinal categorical wildebeest abundance categories, which were recorded on a spatial grid on a monthly basis for three years in an effort to observe the changing distribution of animals in space and time. Large field systems are seldom used in studies of collective movement, which predominantly focus on simple easily observed lab systems, and I hope to start readdressing this bias using this large complex system in which the movement drivers are likely to be many and varied. By fitting the mechanistic movement models developed during this study to data from systems at such vastly different scales as single cells and large ungulates, I hope to demonstrate that the model framework can be widely applied to understand movement in many systems.

A summary of the subsequent chapters in this thesis is as follows. In chapter 2, I developed a series of candidate advection-diffusion models for describing the movement behaviour in the two cellular systems. The mechanisms considered in these models included a response to a chemical gradient that is self-generated by local depletion, attraction to or repulsion from conspecifics, and an overcrowding effect. Time-varying parameters were also considered. These models were fitted to each of the cellular datasets using parameter optimisation-based approaches that maximised the likelihood, and model selection using various information criteria was used to infer the best model. In chapter 3, I investigated extensions to the framework for inference of cellular movement drivers, including spatial variation in parameter values, and a Bayesian approach to model inference. Chapter 4 describes a method for obtaining smooth density surfaces in time and two-dimensional space from large ordinal categorical datasets using GAMs. This method was applied to the wildebeest dataset as a prerequisite for the work carried out in the next chapter. In chapter 5, I extended the modelling framework developed for analysing the cell data, so that it could be used for the inference of movement drivers in the wildebeest system. Parameter inference in this chapter involved a gradient matching approach that made use of the wildebeest density surface developed in chapter 4. Finally, I discuss the results and future directions in chapter 6.

2. Inference of the drivers of collective movement in two cell types: *Dictyostelium* and melanoma

The work presented in this chapter has been published at the following reference:

Ferguson, E.A., Matthiopoulos, J., Insall, R.H. & Husmeier, D., 2016. Inference of the drivers of collective movement in two cell types: Dictyostelium and melanoma. Journal of the Royal Society Interface, 13(123), 20160695. Available at: http://rsif.royalsocietypublishing.org/content/13/123/20160695

2.1. Introduction

Collective movements are important in many cell systems, affecting processes of considerable medical interest, including wound healing, the immune response and the spread of cancers. Cell movements can have both random (diffusive) and directional components. Chemotaxis, the movement of cells up or down spatial gradients in the concentrations of chemicals (chemoattractants or chemorepellants), is the process underlying many of the directional cell movements that we observe (Majumdar et al. 2014). The chemical gradients to which cells respond can result from chemicals diffusing from a local source, which is typically formed by either the cells themselves or nearby cells of a different type releasing chemicals into the environment. An example of local source gradient generation is the suggested mechanism by which macrophages promote metastasis of breast tumours; the tumour cells release an attractant for macrophages, which chemotax towards the tumour and release an attractant for the tumour cells, encouraging their migration away from the primary tumour (Wyckoff et al. 2004). Chemical gradients may also result from local sinks, which are typically caused by cells depleting a chemical from their environment. Recent studies have suggested that cell movements caused by chemotactic gradients that cells self-generate by depletion may be common to a wide range of cell types (Scherber et al. 2012, Donà et al. 2013, Venkiteswaran et al. 2013, Muinonen-Martin et al. 2014, Tweedy et al. 2016). Cell movements resulting from diffusion and chemotaxis may additionally be influenced by density-dependent effects. If cells are in a tightly-packed environment, then they may restrict each other's abilities to move in response to stimuli. The process of contact inhibition of locomotion, which occurs in many cell types and forces cells to change direction when they contact one another (Mayor and Carmona-Fontaine 2010), also has a more pronounced effect at high density.

Identification of the drivers of movement in a particular cell system is a crucial step in understanding how we might influence that system through new medical interventions, such as the use of chemical-releasing implants to disrupt chemotactic gradients responsible for cancer cell migration (Fleming and Saltzman 2002, Deisboeck and Couzin 2009). However, without any prior knowledge, identifying movement drivers experimentally can be a long process. Mathematical models offer a potential solution. By fitting sets of candidate cell movement models to data from cell systems, and then carrying out model comparison to identify the best model, we can get an indication of what mechanisms are most likely to be driving movement in those systems. This information could then be used to guide experimental work, to confirm the existence of these mechanisms.

Since the development of the Keller-Segel model to describe the aggregation of *Dictyostelium discoideum* cells (Keller and Segel 1970), a large body of work has emerged on the modelling of cell movement mechanisms using partial differential equations (PDEs); see Hillen and Painter (2009) for a guide to these cellular models. However, I am unaware of any attempts to formally fit these models to cell movement data and infer movement drivers through model comparison. A possible reason for this is the computational expense. The PDEs involved are of the advection-diffusion-reaction type, describing spatio-temporal changes in the distribution of cells as a result of random cell movements (diffusion), directional movements through chemotaxis (advection) and changes in the numbers of cells through cell division and death (reaction). PDEs with the level of complexity and flexibility required to simulate realistic cell movements typically have to be solved and optimised numerically due to a lack of analytical solutions and closed-form likelihoods, which incurs high computational costs. Numerical solution of the models also introduces error, and when advection is strong relative to diffusion, this error can manifest as oscillations in the modelled cell density. When severe, these instabilities can cause the model solver to fail, halting parameter optimisation prematurely (Sibert et al. 1999). Inference for these models is further complicated by the presence of local likelihood optima that can trap optimisation algorithms before the global optimum is reached. Finally, adequate data on all important variables are not always available; cells may be affected by unidentified chemicals in their environment, and concentrations of even known important chemicals may be impossible to obtain at sufficiently high spatiotemporal resolution. In such cases, these latent variables must be inferred from the information provided by the observed variables. Overcoming these difficulties in model fitting would be an important step towards helping us understand cell movement in a wide range of systems.

In this chapter, I describe six candidate models for cell movement that incorporate various biological hypotheses, including chemotaxis up self-generated gradients, repulsive and attractive interactions between the cells, and interference effects due to cell crowding. Temporal changes in the weightings given to these different movement drivers were also considered within the models. I then develop an inference method that involves the application of maximum likelihood estimation to many bootstrap samples of the data and aims to overcome the challenges associated with model fitting outlined above. This method is tested on data from movement assays for cells of two different types; *Dictyostelium discoideum* and human melanoma. *Dictyostelium* is an amoeba that can exist in both unicellular and multicellular forms (the data used in this study are from cells in the unicellular phase), and is frequently used as a model organism for eukaryotic cell movement (Carnell and Insall 2011). When in their solitary form, *Dictyostelium* cells feed on bacteria, which are located by climbing up gradients in bacteria-produced chemicals. They also respond to chemical gradients under conditions of starvation, when they produce waves of the chemoattractant cAMP, attracting other nearby cells to form a multicellular aggregate (Bonner 1982). Melanoma is a skin cancer, made particularly aggressive by the rapidity with which it spreads, with the risk of metastasis increasing sharply with the thickness of the tumour (Breslow 1970, Balch et al. 2009). Given that metastasis is the primary cause of human cancer deaths (Steeg 2006), understanding why these cells move is important. Recent work has suggested that migration of melanoma cells away from the primary tumour is driven by the tumour becoming large enough to create a local gradient in the chemoattractant lysophosphatidic acid (LPA) through depletion (Muinonen-Martin et al. 2014). Here, I attempt to draw conclusions about the drivers of movement in these cell types, under the conditions of certain movement assays, by fitting the candidate models to data from these assays and carrying out model comparison. Note that the major driver of movement in the two datasets, a self-generated gradient in attractant, has already been determined experimentally

(Muinonen-Martin et al. 2014, Tweedy et al. 2016), so that the ability to identify this key mechanism provides a useful test for the inference scheme developed here. Self-generated gradients are important in driving movement in a range of systems (Scherber et al. 2012, Donà et al. 2013, Venkiteswaran et al. 2013, Muinonen-Martin et al. 2014, Tweedy et al. 2016), and the development of model selection methods that can detect drivers of this type is, therefore, particularly desirable. Other processes that could be playing a more minor role in producing the movement patterns observed in the data, such as overcrowding or chemical interactions between the cells, have been less exhaustively tested for, and so I also test for these within the set of candidate models.

2.2. Data

Data on the collective movement of *Dictyostelium* cells during an under-agarose assay (Laevsky and Knecht 2001) were collected by Tweedy et al. (2016). The agarose under which the cells moved contained folate, a chemoattractant that the cells can deplete from their environment, at an initially homogeneous concentration of 10μM. Under these conditions, *Dictyostelium* cells create a gradient in folate through depletion, and collectively move up this gradient (Tweedy et al. 2016).

A similar dataset on the collective movement of melanoma cells was collected by Muinonen-Martin et al. (2014). Here the migration of the cells was observed between two wells connected by a bridge in a direct visualisation chamber (Muinonen-Martin et al. 2010) that was homogeneously filled with 10% FBS (foetal bovine serum). It was previously determined experimentally that collective movement in this case is primarily driven by a self-generated gradient in LPA, a component of FBS that can be depleted by the melanoma cells (Muinonen-Martin et al. 2014).

Dictyostelium cells move more rapidly than melanoma cells, so the *Dictyostelium* dataset covers both a larger spatial distance (~2500μm compared to ~400μm), and a shorter time frame (5.5 hours compared to 50 hours) than the melanoma dataset. Supplementary videos 2.1 and 2.2 (Appendix A.8) show microscopy images that were captured during these time periods, for *Dictyostelium* and melanoma respectively. I extracted the cell coordinates manually from these images at half-hour time intervals for *Dictyostelium* and ten-hour intervals for melanoma. The cells were initialised in a linear group along the *y*-axis in both assays. Since little variation in movement behaviour is expected in the y-direction as a result of this initial distribution, the datasets are effectively one-dimensional, and I reduced the data to one spatial dimension (*x*) for the analyses. One-dimensional logspline density estimates (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) were used to visualise the spread of the cells up the spatial axis for both *Dictyostelium* and melanoma.

Spatio-temporal variation in the concentration of the chemoattractants, folate and LPA, was unmeasurable during the assays. Therefore, I treated these concentrations as latent variables during model fitting.

2.3. Models

All of the cell movement models considered in this study involve one-dimensional advection-diffusion-reaction PDEs of the form:

$$\frac{\partial C(x,t)}{\partial t}=\underbrace{-\frac{\partial}{\partial x}\{a(x,t)C(x,\mathrm{t})\}}_{\text{advection}}+\underbrace{\frac{\partial}{\partial x}\left\{D_C(t)\frac{\partial C(x,t)}{\partial x}\right\}}_{\text{diffusion}}+\underbrace{\nu C(x,t)}_{\text{reaction}} \tag{2.1}$$

where *t* is time, *x* is space and $C(x,t)$ is cell density. A positive or negative value of the advection coefficient $a(x,t)$ leads to directional movement towards higher or lower *x* respectively. The diffusion coefficient $D_C(t)\geq 0$ describes the rate at which cells spread out from high to low density areas via randomly directed movements, and the reaction term describes exponential growth of the cell population through cell division at rate $\nu \geq 0$.

I investigated six different advection coefficients, each representing a hypothesis for the drivers of cell movement. The **diffusion model** assumes that cell movement is simply random, with no directional movement component, i.e.:

$$a(x,t)=0 \tag{2.2}$$

Directional movement up a spatial gradient in the concentration of an attractant $A(x,t)$ is described in the **basic model**:

$$a(x,t)=\alpha(t)\frac{\partial A(x,t)}{\partial x} \tag{2.3}$$

Here the rate of advective cell movement depends both on the strength of the gradient in $A(x,t)$ and the magnitude of the parameter $\alpha \geq 0$. The attractant concentration is modelled through a second PDE:

$$\frac{\partial A(x,t)}{\partial t}=-\gamma(t)C(x,t)A(x,t)+D_A\frac{\partial^2 A(x,t)}{\partial x^2} \tag{2.4}$$

This function allows the cells to create self-generated gradients in $A(x,t)$ through local depletion in proportion with their density and the remaining level of attractant, at a rate determined by $\gamma \geq 0$. The parameter D_A describes the constant rate at which attractant diffuses in the medium.

While the basic model (equation (2.3)) assumes that the ability of the cells to chemotax up a gradient in attractant is influenced only by the steepness of the gradient, it has been shown that chemotaxis also depends on the concentration of chemoattractant in a cell's local environment (Tweedy et al. 2013). This dependency is a result of receptor saturation. Cells detect spatial gradients in chemicals through the resulting gradients in the occupancy of their surface receptors for those chemicals. When the background chemoattractant concentration is high, a cell's receptors can become saturated, so that an underlying chemotactic gradient fails to produce a detectable gradient in receptor occupancy, preventing accurate chemotaxis. In the **receptor saturation model**, I replace the chemoattractant gradient of the basic model (equation (2.3)) with a gradient in receptor occupancy, calculated according to Michaelis-Menten kinetics, where K_d is the dissociation constant that describes the folate concentration at which half the cells' folate receptors are occupied, as follows:

$$a(x,t)=\alpha(t)\frac{\partial}{\partial x}\left(\frac{A(x,t)}{A(x,t)+K_d}\right) \tag{2.5}$$

Cell movement may be influenced by attractive or repulsive chemical interactions between the cells. In the **receptor saturation and interaction model**, I incorporate these behaviours by allowing the cells to move directionally in response to gradients in their own density, in addition to the gradient in receptor occupancy for $A(x,t)$:

$$a(x,t)=\alpha(t)\frac{\partial}{\partial x}\left(\frac{A(x,t)}{A(x,t)+K_d}\right)+\frac{\eta(t)}{1+\lambda C(x,t)}\frac{\partial C(x,t)}{\partial x} \tag{2.6}$$

Here, a negative η indicates repulsion between the cells and a positive η indicates attraction. The strength of the interaction is reduced at high cell densities through the parameter $\lambda \geq 0$. This feature is intended to mimic the effect of saturation of the cell receptors for the chemical involved in the interaction; at high cell density, higher concentrations of the chemical released by the cells are expected, leading to saturation effects that reduce the ability of the cells to detect and migrate in response to the conspecific density gradient. Keller and Segel (1970) previously proposed a method for modelling cell interactions, in which the cells respond directly to the interaction chemical, the production and decay of which is modelled through an additional PDE. The more indirect approach I use here, where the cells instead respond to their own density gradient, has the advantages that it requires fewer new parameters, which simplifies model fitting, and it avoids the need to make an assumption about the unknown initial distribution of the interaction chemical.

It is expected that the ability of cells to move freely will be reduced at high density, both because tight packing of cells means that there is physically less space for them to move into, and because more contact between cells occurs at high density, meaning that the effects of contact inhibition of locomotion (Mayor and Carmona-Fontaine 2010) will be more evident. I incorporate these effects into the receptor saturation model (equation (2.5)) to produce the **receptor saturation and overcrowding model**:

$$a(x,t)=\alpha(t)\frac{\partial}{\partial x}\left(\frac{A(x,t)}{A(x,t)+K_d}\right)\left(1-\frac{C(x,t)}{C_{\max}}\right) \tag{2.7}$$

The new term in the advection coefficient, which is derived in Hillen and Painter (2009), causes advection up the gradient in receptor saturation to slow as cell density approaches its maximum value $C_{\max}$.

Finally, the **full model** combines the effects of receptor saturation, cell interactions and overcrowding, with the advection coefficient:

$$a(x,t)=\left(\alpha(t)\frac{\partial}{\partial x}\left(\frac{A(x,t)}{A(x,t)+K_d}\right)+\frac{\eta(t)}{1+\lambda C(x,t)}\frac{\partial C(x,t)}{\partial x}\right)\left(1-\frac{C(x,t)}{C_{\max}}\right) \tag{2.8}$$

Note that all of the models presented here are nested within the full model as illustrated in the model relational graph of Fig. 2.1.

Four of the model parameters α, D_C, γ and η, which relate to cell advection and diffusion rates, and the rate of depletion of chemoattractant, are permitted to vary in time to allow for changes in cell behaviour over the course of the assays. These temporal dependencies were introduced by modelling the parameters as polynomial functions of time, which were exponentiated for those parameters that were restricted to positive values (α, D_C and γ). The degrees of the polynomial functions were selected as described in section 2.5.

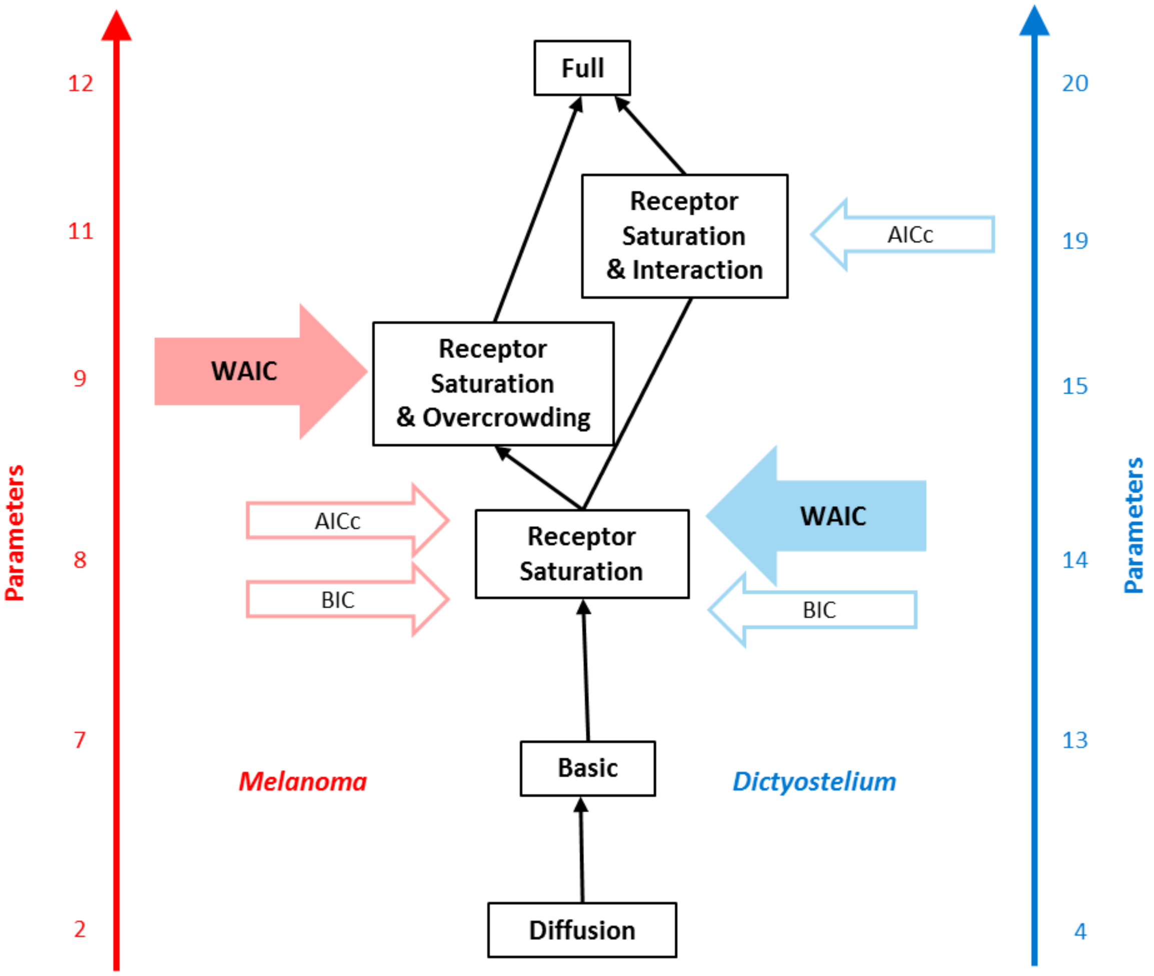

Figure 2.1: Graph illustrating the relationships between the candidate models. Wherever two of the models (described in section 2.3) occupy adjacent nodes, it is possible for the more complex model (with the greater number of parameters) to be reduced to the less complex one by constraining parameters. The number of parameters given for each model is based on a degree of one for the polynomials describing the time-varying parameters for melanoma, and a degree of three for *Dictyostelium* (see Tables A.6.1-2 in Appendix A.6). For each dataset, the models preferred by WAIC, AICc and BIC are indicated with arrows.

2.4. Likelihood calculation

For a given dataset, model and set of parameters $\boldsymbol{\theta}$, I obtained spatiotemporally varying functions describing cell density $C(x,t)$ and attractant concentration $A(x,t)$ by solving the PDEs numerically using the method of lines (Schiesser and Griffiths 2009, Soetaert et al. 2010) (see Appendix A.1.1 for details). For melanoma, there were no cells in the observation region at $t=0$, so I used initial conditions of $C(x,0)=0$ and $A(x,0)=1$ (100% of the initial concentration of the attractant (LPA) remaining in the serum). For *Dictyostelium*, where some cells had already moved into the observation area at $t=0$ (which was around an hour after the cells were actually introduced to the experiment) as a consequence of their having a more rapid movement rate than melanoma cells, the initial distribution of cells was obtained by applying logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) to the cell location data. I assumed a sigmoidal function for the unobserved initial distribution of the attractant for *Dictyostelium* (folate), the parameters of which were estimated along with the model parameters. Increases in the total number of cells due to cell division were relatively minor over the time period of interest for *Dictyostelium*, so I set ν to zero. For melanoma, the value of ν was estimated from the data as described in Appendix A.1.3. In both datasets, large numbers of cells moved into the observation region via the left boundary, and I captured these movements by introducing a cell flux across this boundary, which was equal to the rate of change in the number of cells observed in the region minus the rate of change in cell numbers due to cell division. Full details on the choices of boundary and initial conditions can be found in Appendix A.1.

The models were fitted to the cell locations at the T time points for each dataset. The raw observations $(y_1,...,y_n)$ were, thus, each referenced by both a spatial location and time, i.e. $y_i=(x_i,t_i)$. The total number of cells observed over the T time points was given by

$$n=\sum_{j=1}^{T} n_j \tag{2.9}$$

where n_j was the number of cells observed at time point $j\in(1,...,T)$.

Following numerical integration of the model, the likelihood of $\boldsymbol{\theta}$ can be calculated for each (x_i,t_i) as:

$$P(x_i\,|\,t_i,\boldsymbol{\theta})=\frac{C(x_i,t_i)}{\int_0^l C(x,t_i)\,dx} \tag{2.10}$$

Division by $\int_0^l C(x,t_i)\,dx$ normalises the cell density to convert it into a probability density in space. By summing over the y_i, the total log-likelihood could then be obtained as:

$$\log L=\sum_{i=1}^{n}\log\left\{P(x_i\,|\,t_i,\boldsymbol{\theta})\right\} \tag{2.11}$$

However, since the number of cells observed increases over time for both datasets, this standard log-likelihood will be biased towards producing a good fit at the end of the time period considered; potentially leading to a poorer match between model and data at the beginning of the time period. An alternative method that corrects for this bias is to weight each $\log\left\{P\left(x_i \mid t_i, \boldsymbol{\theta}\right)\right\}$ according to the total number of cells observed at the corresponding time point as follows:

$$\log \tilde{L} = \frac{n}{T}\sum_{j=1}^{T}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}\log\left\{P\left(x_i \mid t_i, \boldsymbol{\theta}\right)\right\}\right] \tag{2.12}$$

In this weighted log-likelihood calculation, the multiplication by n/T returns the value to the scale of the standard log-likelihood. Weighted likelihoods have frequently been used to remove bias by down-weighting observations believed to be of a lower quality (Hu and Zidek 2002, Agostinelli and Greco 2013). Here, I down-weight observations not because they are of a lower quality, but because they provide less new information, given that there are already many other observations at the same time point.

2.5. Model inference

For all models considered, it was necessary to infer both the model parameters and, for *Dictyostelium*, also the parameters of the sigmoidal distribution describing the unknown initial distribution of folate (see Appendix A.1.2). During inference, I used a lower bound of zero for the diffusion coefficient D_A of LPA in the melanoma assay, while for *Dictyostelium*, I used literature values for the diffusion coefficient of folate (Kalimuthu and John 2009, Ershad et al. 2013) to introduce more restrictive upper and lower bounds of 200μm^2/s and 150μm^2/s respectively for D_A. For both datasets, I set a lower bound for $C_{\max}$ that was equal to the maximum cell density value observed in the logspline density estimates obtained from the cell location data (blue lines in Figs 2.2-3). I bounded the parameters K_d and λ below by zero, leaving them unbounded above. The parameters describing the initial folate distribution were given upper and lower bounds that prevented initial distributions known to be unrealistic (see Appendix A.1.2). The remaining parameters (α, γ, η and D_C) were modelled as polynomial functions of time, which for α, D_C and γ were exponentiated to bound the functions below by zero. The coefficients of the polynomial functions were unbounded during model inference.

It was necessary to select the degrees of the polynomial functions used to describe the time-varying parameters. Ideally, this would be achieved by carrying out inference for each model on each dataset using a range of polynomial degrees for each of the parameters, and then applying model comparison to select the best combination of polynomial degrees for each model. However, inference for these models is computationally expensive, making such an exhaustive model comparison infeasible. I instead proceeded by fitting the most complex model (the full model, equation (2.8)) to each of the two datasets by maximising the weighted log-likelihood (equation (2.12); see Appendix A.2 for details on the maximisation procedure), and gradually increasing the degree of the polynomials, always keeping the degree the same for all time-varying parameters in the model. I stopped increasing the polynomial degree when there was no further improvement in the values of two model comparison statistics; AICc (the Akaike Information Criterion corrected for small sample sizes (Akaike 1974, Hurvich and Tsai 1989)) and BIC (Bayesian Information Criterion (Schwarz 1978)). Once I had used this maximum weighted log-likelihood approach to obtain the optimal polynomial degree for the temporal variation of the parameters for each dataset, I carried out inference for the full set of six candidate models, using the more computationally costly, but more reliable, pseudo-Bayesian approach described below, always using the polynomial degree selected based on AICc and BIC.

The use of a Bayesian approach to obtain a posterior distribution of the parameters provides access to WAIC (Widely Applicable Information Criterion (Watanabe 2010)); a recently developed model comparison statistic that makes fewer assumptions than those commonly calculated from maximum likelihood estimates (including AICc and BIC). The key improvement offered by WAIC is that it allows for the fact that some parameters might be poorly determined by the data (for details see Chapter 7 of Gelman et al. (2013)). However, Markov Chain Monte Carlo (MCMC) algorithms, the standard approach to obtaining a sample from the posterior, are intrinsically sequential, making them unable to exploit parallel computer clusters. This sequential nature of MCMC presents further problems for advection-diffusion models, as chains can break down or become trapped in regions of parameter space where unstable numerical solutions cause model solving algorithms to fail (Sibert et al. 1999). I avoided these issues by using the following method to obtain a pseudo-posterior for each of the models and datasets.

The cell location data were sampled with replacement for each time point involved in the fitting process to obtain many bootstrap datasets of the same size as the original ones. A maximisation of the weighted log-likelihood (equation (2.12)) was then carried out for each model on each bootstrap dataset using an optimisation algorithm (I found that the quasi-Newton BFGS algorithm performed well for the *Dictyostelium* data, while the Nelder-Mead algorithm was more effective at reaching high-likelihood parameter regions for the melanoma data). By optimising on many re-samples of the data, I obtained many parameter sets that could be used as a proxy for a sample from the posterior distribution of the parameters, where there is an assumption of uniform prior distributions. This pseudo-posterior is similar to a true posterior in that it describes uncertainty in the parameter values, with the variance of the pseudo-posterior being driven by the uncertainty in the data, which is introduced through the bootstrapping procedure. Similar approaches to obtaining a pseudo-posterior have previously been applied by other authors; see for example Friedman et al. (2000). Note that this approach to inference is computationally costly, due to the need to run many optimisations per model (I used 3,000), but has advantages in being easily automated and parallelised. Additionally, any optimisations that fail due to numerical instability can simply be discarded and reinitialised, though, as discussed in section 2.7, this leads to certain regions of parameter space being under-represented in the pseudo-posterior distribution.

As a result of the optimiser becoming trapped on local optima, I found that, for both datasets, the pseudo-posteriors obtained by this method tended to be multi-modal. I removed all but the highest-likelihood peak in the pseudo-posteriors, as described in Appendix A.3, prior to using the pseudo-posteriors to calculate WAIC as:

$$\begin{aligned}\mathrm{WAIC} = &-2\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\\ &+2\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right\}\end{aligned} \tag{2.13}$$

where m is the number of optimisations, $\left\{y_j = \left(x_j, t_j\right)\right\}_{j=1,n}$ are the cell location data, and $\boldsymbol{\theta}_i$ are the optimised parameter sets. To verify that WAIC approximated using a pseudo-posterior obtained by bootstrap sampling gives comparable results to the standard WAIC calculated by direct sampling from the true posterior, I carried out a test study that used both methods to select the order of a polynomial model fitted to independent benchmark data (Appendix A.4). There was very close agreement between the WAIC values obtained using the two methods, suggesting that, at least in this simple test case, the pseudo-posterior is practically equivalent to the true posterior.

2.6. Results

Based on AICc and BIC, I selected a degree of three for the polynomial function describing the temporal variation in the parameters for *Dictyostelium*, and a degree of one for melanoma (Tables A.6.1-2 in Appendix A.6), suggesting that the *Dictyostelium* cells are changing their behaviour more rapidly than the melanoma cells.

For Dictyostelium, WAIC selects the receptor saturation model as the best model, while, for melanoma, the slightly more complex receptor saturation and overcrowding model is preferred (Table 2.1). While there are known issues with AICc and BIC – AICc can select overly complex models, while BIC typically selects overly simple models (Ripplinger and Sullivan 2008), and neither accounts for parameter uncertainty – that make them less reliable than WAIC, I also compared the models based on these simpler statistics to check for consistency (Tables A.6.3-4 in Appendix A.6). The difference between the model selected by WAIC and the models selected by AICc and BIC never exceeds a graph distance of one (Fig. 2.1).

For both datasets, the selected models produce very good visual agreement with the data (Figs 2.2-3). These fits are a vast improvement over those produced by the simple diffusion model (Figs A.7.1-2 in Appendix A.7), and also provide a clear improvement over the basic model (Figs A.7.3-4 in Appendix A.7); the inclusion of the receptor saturation effect appears to allow the models to better replicate the peaked cell front, which the basic model tends to smooth over.

Table 2.1: Selection of best model for each cell type based on WAIC. WAIC values (equation (2.13)) are given for the six candidate models for both datasets. Standard errors (in brackets) were calculated as described in Appendix A.5. The best model for each dataset (i.e. the model with the lowest WAIC value) is indicated *.

Model	**WAIC**	
	Dictyostelium	**Melanoma**
Diffusion	88367.1 (0.10)	5985.5 (0.03)
Basic	87970.7 (0.77)	5736.2 (7.70)
Receptor Saturation	87631.2 (0.39)*	5719.9 (3.10)
Receptor Saturation & Interaction	87636.8 (0.44)	5743.1 (2.08)
Receptor Saturation & Overcrowding	87648.0 (0.44)	5712.2 (1.85)*
Full	87646.3 (0.47)	5739.6 (2.25)

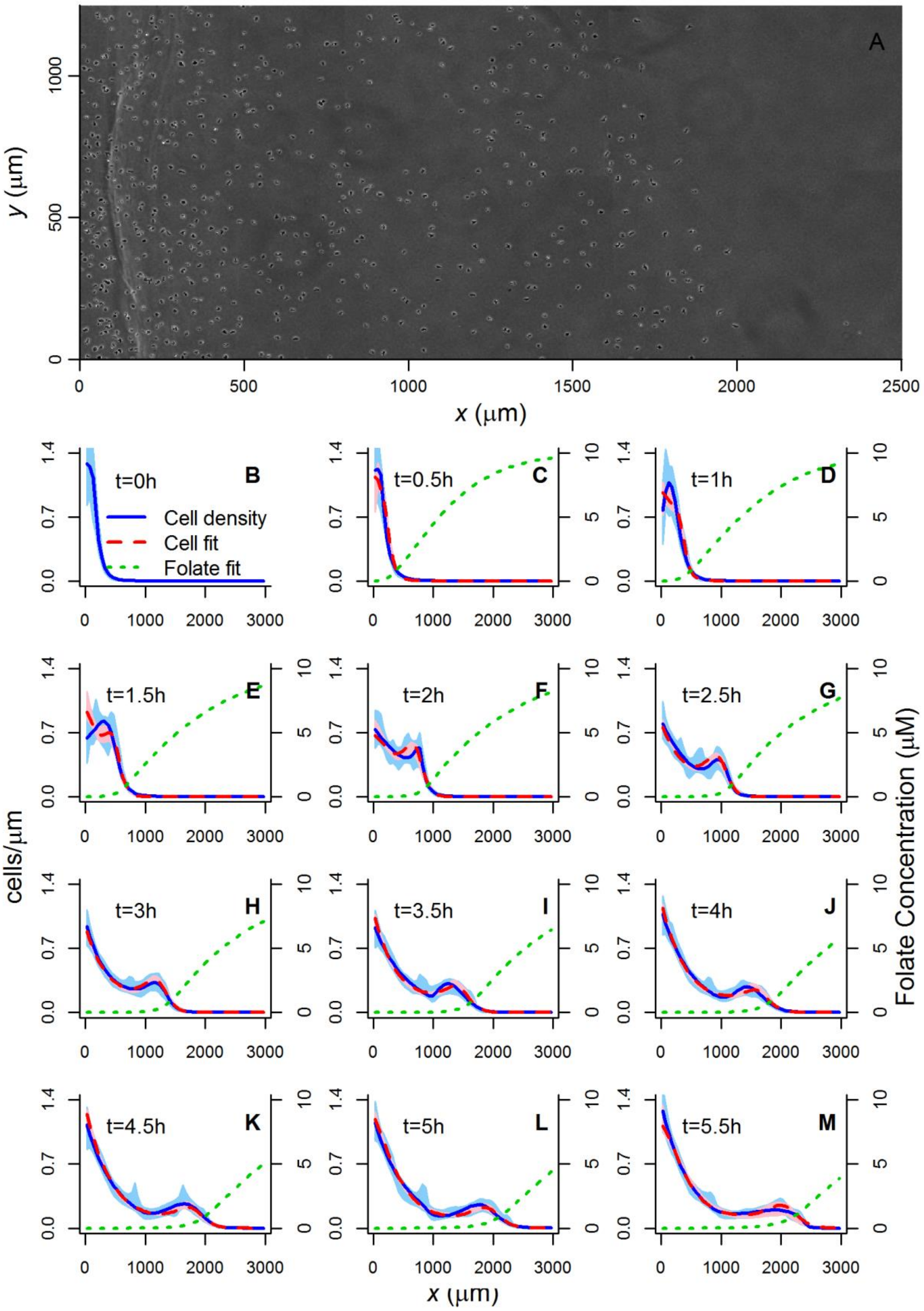

Figure 2.2: ***Dictyostelium* data and fitted best model.** **A)** Image taken 4 hours into the *Dictyostelium* cell movement assay (see **J** for corresponding cell density estimate). **B-M)** Cell distributions obtained every half hour using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) in the *x* dimension are shown by blue lines, with 95 percentile intervals obtained using 10,000 bootstrap samples of the data indicated by blue shaded areas. Cell distributions produced by the best model (the receptor saturation model, Table 2.1) for this dataset, using the optimised parameters from the bootstrap optimisation that gave the highest value of the maximum weighted log-likelihood (equation (2.12)), are shown by dashed red lines. The corresponding folate distributions predicted by this model are indicated by green dotted lines. Pink shaded areas show the 95 percentile interval for the modelled cell density, based on 200 samples from the pseudo-posterior.

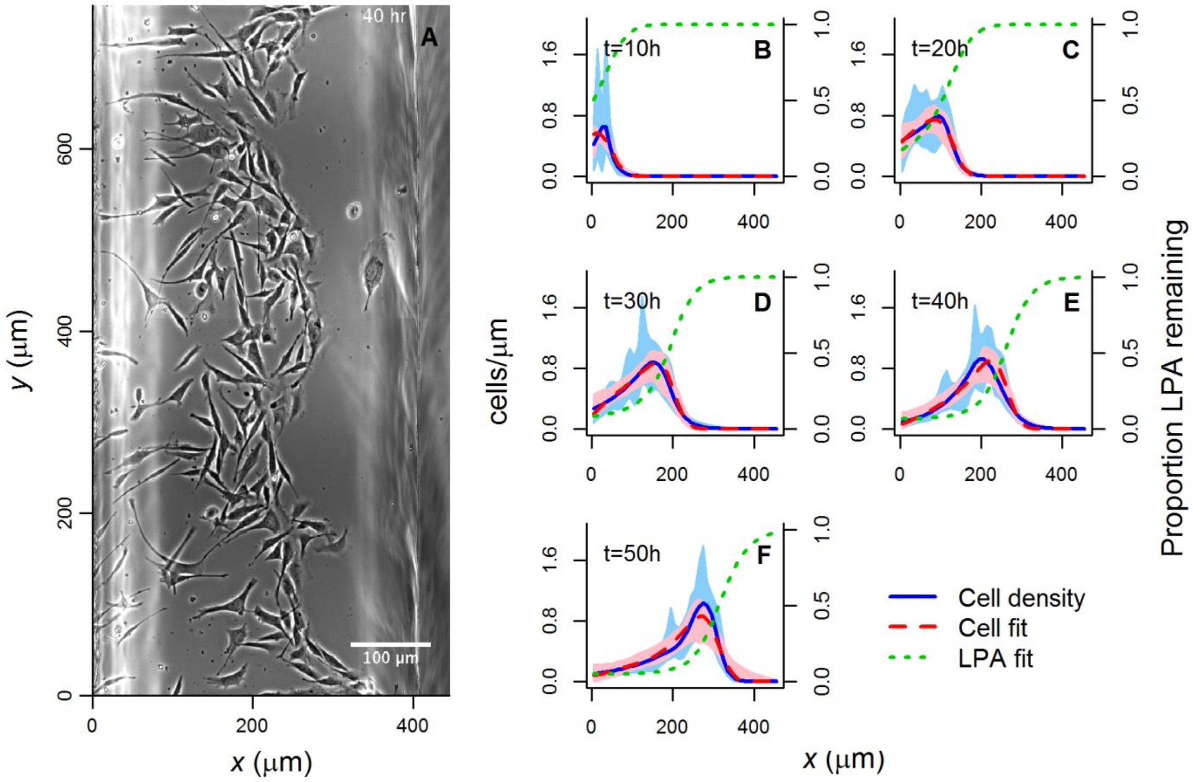

Figure 2.3: Melanoma data and fitted best model. A) Image taken 40 hours into the melanoma cell movement assay (see **E** for corresponding cell density estimate). **B-M)** Cell distributions obtained every 10 hours using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) in the *x* dimension are shown by blue lines, with 95 percentile intervals obtained using 10,000 bootstrap samples of the data indicated by blue shaded areas. Cell distributions produced by the best model (the receptor saturation and overcrowding model, Table 2.2) for this dataset, using the optimised parameters from the bootstrap optimisation that gave the highest value of the maximum weighted log-likelihood (equation (2.12)), are shown by dashed red lines. The corresponding LPA distributions predicted by this model are indicated by green dotted lines. Pink shaded areas show the 95 percentile interval for the modelled cell density, based on 200 samples from the pseudo-posterior.

For *Dictyostelium*, the diffusion rate of the cells, D_C, is estimated to first increase with time and then to decline again towards the end of the time period (Fig. 2.4A). The responsiveness of the *Dictyostelium* cells to the folate gradient, α, tends to increase over time (Fig. 2.4B), and the rate at which the cells deplete folate, γ, shows no clear trend (Fig. 2.4C). To investigate the importance of the temporal variation in each of these parameters in improving the fit of the selected model, I refitted the model multiple times by maximum weighted log-likelihood (see Appendix A.2), gradually replacing the time-varying parameters with constants, and comparing these simplified models based on AICc and BIC (Table A.6.5 in Appendix A.6). I found that BIC selects only α and D_C to be time-varying parameters, suggesting that γ can be left time-invariant. The difference in AICc score between the model with all three time-varying parameters and the model with time-invariant γ is small. These findings are consistent with the trends in Fig. 2.4.

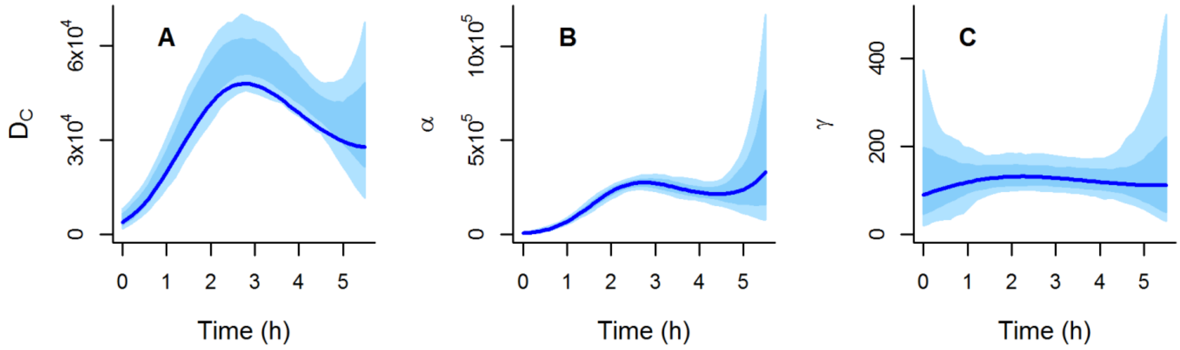

Figure 2.4: Time-varying parameters for the best *Dictyostelium* model (the receptor saturation model, Table 2.1). Lines show the mean of the pseudo-posterior obtained by many optimisations of the weighted log-likelihood (equation (2.12)) on bootstrap samples of the data. Shaded areas indicate 95 and 66 percentile intervals obtained from 200 samples from the pseudo-posterior.

Carrying out a similar model selection for melanoma (Table A.6.6 in Appendix A.6), both AICc and BIC consistently suggest that the time dependence of D_C and γ can be removed, and that α should be retained as the only time-varying parameter. A plot of the time dependence of α is given in Fig. 2.5, which shows a monotonically decreasing trend. There is a large amount of uncertainty in the value of α, particularly at the beginning of the time series. However, as was suggested by the AICc and BIC results (Table A.6.6), making α time invariant leads to a visible reduction in the quality of the fit of the model to the data (compare Fig. 2.3 and Fig. A.7.5).

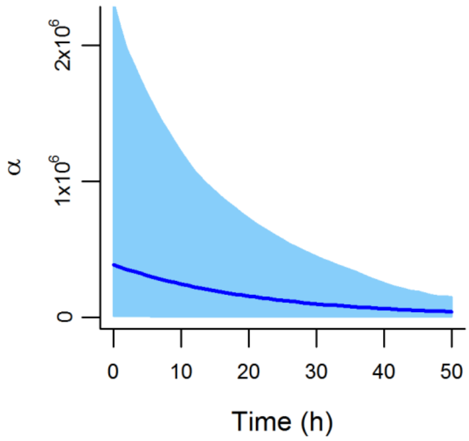

Figure 2.5: Temporal variation in α for the best melanoma model (the receptor saturation and overcrowding model, Table 2.1), based on the mean of the pseudo-posterior obtained by many optimisations of the weighted log-likelihood (equation (2.12)) on bootstrap samples of the data. The shaded area indicates the 66 percentile interval obtained from 200 samples from the pseudo-posterior.

2.7. Discussion

Despite several decades of work developing mathematical models for collective cell movement, surprisingly little has been done to confront these models with data. Recent developments in both microscopy techniques and computer-intensive statistics are gradually removing the obstacles in this area. Here, I have begun exploring the technical challenges associated with carrying out statistical inference (comprising both parameter estimation and model selection) for PDE models using microscopy data on collective cell movement.

The novel inference method presented here, which involves running independent parameter optimisations on many bootstrap replicates of the data, was motivated by Friedman et al. (2000), where it was referred to as a "poor man's" approximation of the posterior distribution. In comparison to MCMC, this bootstrapping approach is easily automated, and can be parallelised to spread the high computational cost over many processors. By generating a pseudo-posterior distribution, the bootstrapping approach also allows computation of WAIC, which accounts for parameters that are poorly determined when penalising model complexity, making it a more powerful and reliable model comparison statistic than AICc and BIC. This reduced penalty for poorly-defined parameters may be why, in the melanoma case, WAIC selects a more complex model than AICc and BIC (Fig. 2.1)). I showed in a test study that obtaining WAIC from a pseudo-posterior can give good correspondence with the standard WAIC calculated by sampling from the true posterior (Appendix A.4). This test study involved a polynomial regression problem, where the goal was to identify the optimal degree of a polynomial describing the relationship between two variables using model comparison. The polynomial models considered in the test study had numbers of estimated parameters ranging from 2-10, which is comparable to the cell movement models considered in this chapter (see Fig.2.1). However, despite this similar model complexity, it is possible that the complexity of the likelihood surfaces differs between this test study and the cell movement study. If the likelihood surfaces in the cell movement study were less smooth and had more local optima than occurred in the test study, then more of the parameter optimisations could have become trapped on these local optima. This would have led to the shapes of the pseudo-posteriors being different to those of the true posteriors, making a WAIC comparison based on these pseudo-posteriors less reliable. Further testing of the bootstrapping method on problems known to produce complex likelihood surfaces is, therefore, required to determine how robust the approach is under these conditions. It should be noted that, in the test study, AICc and BIC based approaches were just as accurate in selecting the correct model as WAIC calculated from the pseudo-posterior (Table A.4.1), so that the more computationally expensive and non-standard bootstrapping approach was unnecessary. However, given that in the melanoma movement study WAIC estimated through bootstrapping selected a different model to that identified by AICc and BIC (Fig. 2.1), it could be argued that, assuming the WAIC estimate is a good representation of the true value, this represents a case in which the extra work to obtain an estimate of WAIC was valuable.

An issue arose during fitting of the cell movement models that could have led to a certain distortion in the approximations of the posterior distributions. This was that some optimisations failed due to instability in the numerical model solution at certain parameter combinations, which could have meant that certain areas of parameter space were under-represented. These numerical instabilities are a known issue for advection-diffusion models that become evident when the Péclet number (the ratio of the advection coefficient to the diffusion coefficient, multiplied by the box length used when discretising the PDE in space during numerical solution (Soetaert and Herman

2009)) exceeds one. The pseudo-posteriors, therefore, are limited to those regions where the numerical solutions of the models are relatively stable, and this may have led to them being different to the pseudo-posteriors that would have been obtained with accurate analytical solutions. If the method were to be applied in a case where the majority of the posterior probability density was located in an unstable region of parameter space, the majority of optimisations would fail. As a result, it may be computationally infeasible to obtain sufficient optimised parameter sets, and the resulting parameter estimates would be highly inaccurate anyway. In such cases, methods for fitting differential equations that bypass the need for numerical solution may be the only option (Macdonald & Husmeier 2015; Xun et al. 2013; see also chapter 5 of this thesis).

A further potential problem with the inference methodology outlined in this chapter lies in the application of bootstrapping to a dataset where the points (cells) may be interacting with one another. Bootstrapping leads to some individuals appearing multiple times in a bootstrap sample at a particular time point, while other individuals are omitted from the sample entirely. Since this leads to changes in the cell density at each point in space relative to the original data, and individuals in the models that include cell interactions are responding to this rearranged density, the shape of the posterior distributions of the parameters associated with cell interactions could be affected. This, in turn, could lead to a wrong conclusion being drawn as to whether cell interactions do or do not affect movement behaviour in the system in question. Simulation studies to test whether bootstrapping affects our ability to determine the correct model in cases where cell interactions are or are not present would be a useful avenue for future work.

The decision to model cell movement in one spatial dimension, rather than two, could also have influenced the inference results. This choice was made in order to make the computational cost of solving the PDEs numerically during inference feasible, and was justified on the basis that the cells began the experiment in linear group lying parallel to the *y*-axis, limiting the amount of variation in the cell distribution along this axis. However, it is acknowledged that there was some variation in how quickly the cell fronts advanced at different points in *y*, as can be observed in both Fig. 2.2A and Fig. 2.3A. This could have resulted from the initial cell density not being exactly constant in *y*, which would have led to variation in the rate of development of the chemoattractant gradient through depletion, and, therefore, variation in the rate at which the cells moved in response to this gradient. As a consequence of the slight *y*-variation in the rate of progression of the cell front along *x*, collapsing the data onto the single *x*-dimension is likely to have resulted in a cell front that appears broader and less sharp than that that would have been observed if the data from a narrower window in *y* were considered. This distortion of the cell distribution when the cell data are reduced to one dimension may have had an influence on the parameter values estimated. The appearance of a broader cell front than was actually present may, for example, have led to overestimation of the diffusion coefficient, since increased diffusion leads to the cells spreading out more.

In addition to these specific limitations, I acknowledge that statistical methods, on their own, are not able to identify a model with absolute certainty, as has been discussed, for example, in Burnham and Anderson (2002). This is a consequence of both sampling uncertainty, and the reliance of these methods on heuristic approximations (as discussed in the previous paragraph, or in Appendix A.2). However, statistical methods can identify those models that are most likely given current data, filtering out those that are unlikely to be correct, and thus guiding future targeted experimental work to confirm the statistical findings. This makes model inference a useful tool, as narrowing down hypotheses using experiments alone is often made infeasible by the number and complexity of these hypotheses, and the cost of such experiments. The reliability of the novel statistical procedures used here has been critically assessed in two ways. First, I compared the model selection scheme, based on WAIC estimated from the pseudo-posterior, with two established asymptotic model selection criteria (AICc and BIC), and found that the models selected by these different statistics are never separated by a graph distance of more than one. This agreement between statistics is reassuring; WAIC is expected to provide a slight improvement on, but not a complete deviation from, the asymptotic results. Second, while complete *a priori* knowledge of the processes affecting cell movement in the two datasets is lacking, partial knowledge with which to validate the statistical results is available, as discussed below.

Through model inference and comparison, I have drawn a number of conclusions about the drivers of collective movement in assays for both *Dictyostelium* and melanoma cells. In both systems, the simple diffusion model is rejected as a description of the observed movement patterns in favour of more complex models that incorporate directional movement in response to attractant gradients that are self-generated through depletion. This indication of the importance of the self-generated gradient mechanism shows agreement with experimental findings for melanoma (Muinonen-Martin et al. 2014), and experimental and simulation model results for *Dictyostelium* (Tweedy et al. 2016), that this mechanism is a key driver of the direction of chemotaxis in these systems. Confidence in the ability of the inference methods to identify the correct movement mechanisms is further increased by the fact that, for both cell types, a substantial improvement of the receptor saturation model over the basic model is observed (Table 2.1). This agrees with the widely-accepted concept that connection between extracellular signals and the intracellular mechanisms that drive cell migration occurs through cell-surface receptors. These receptors communicate to the inside of the cell by adopting two states, unoccupied and occupied; thus the only information seen by the motility machinery is the fractional occupancy of the receptors. At high receptor saturation there can be very little difference in receptor occupancy between the front and rear of the cell. Incorporating receptor saturation led to the models being better able to replicate the form of the peak in cell density that marks the moving cell front. The receptor saturation effect causes this peak to become more defined, by causing the cells at the very front of the distribution, where attractant is most concentrated, to move more slowly than those directly behind, leading to a build-up of cells where the faster moving individuals meet the slower front-runners. The inference methods also allowed prediction of how the gradients in folate and LPA concentration, on which no directly measured data were available, changed over the course of the assays. For *Dictyostelium*, the form of the predicted folate distribution gives a relatively close visual match to that measured experimentally by Tweedy et al. (2016), using the same assay but with a higher initial folate concentration.

In addition to providing insights into the self-generated gradient mechanism, model comparison suggests that an effect of cells blocking each other’s movement when at high density (described in those models with an overcrowding effect) was evident in the melanoma data, but not in the *Dictyostelium* data. The primary reason for this difference may be that the cell densities in the *Dictyostelium* dataset never became high enough for overcrowding effects to exert an effect that the inference methods could detect; a visual comparison of images from the two datasets indicates that there are less direct contacts between the *Dictyostelium* cells (Fig. 2.2A) than between the melanoma cells (Fig. 2.3A). It is not completely clear how contact inhibition of locomotion (CIL) would be expected to modify cells moving in a self-generated gradient, but this process is known to occur in neural crest cells (Scarpa et al. 2015). Since the melanocytes that mutate into melanoma cells develop from neural crest cells (Parichy et al. 2007), it is likely that melanoma cells will also exhibit CIL, which may be a contributing factor to the selection of the receptor saturation and overcrowding model for the melanoma dataset. Previous results simulated from an individual-based cell movement model suggested that CIL may also play a role in *Dictyostelium* movements in the system investigated here (Tweedy et al. 2016). The inability to detect this effect in *Dictyostelium* here through a preference for the receptor saturation and overcrowding model over the receptor saturation model may be a result of the loss of information incurred in moving from an individual-based modelling approach, where the movement path of each cell is known, to the population-based approach used in this study, where individual movement paths are not analysed.

The model comparison found no evidence for direct attractive or repulsive interactions between the cells for melanoma; a finding that is backed up by a lack of evidence for such conspecific interactions in the literature. For *Dictyostelium*, AICc suggests that such interactions may be important, but the other two comparison statistics (including the more reliable WAIC) place the receptor saturation and interaction model second to the receptor saturation model (Table 2.1, Table A.6.3 in Appendix A.6). Thus, while there may be some chemical communication between the *Dictyostelium* cells, its effect on the observed behaviour is not strong enough to be reliably detected. Vegetative *Dictyostelium* cells are known to secrete and respond to chemorepellents, but these appear to act over short time scales (minutes rather than hours) and ranges, so that repulsive interactions are not found to be important over the time-frame and distances involved in the assay investigated here (Keating and Bonner 1977, Kakebeeke et al. 1979). Since *Dictyostelium* is well known for exhibiting aggregative interactions when exposed to prolonged starvation conditions (Bonner 1982), a shift in preference towards the receptor saturation and interaction model may have been observed had the cell movement assay been run for a longer time period, or used *Dictyostelium* cells that were at a later stage in their development.

I found evidence in both datasets for changes in cell behaviour over time (Figs 2.4-5, Tables A.6.1-2 in Appendix A.6). The diffusion coefficient for *Dictyostelium* is estimated to be low at the beginning of the assay (Fig. 2.4A), which may be a result of most of the cells still being in the process of transitioning under the gel at this stage. During this transition, the cells experience resistance from the gel (Laevsky and Knecht 2001), which will reduce the speed of diffusion. The diffusion rate increases once the cells have successfully moved under the gel, but then declines again towards the end of the time period, which may be a result of both starvation (Chubb et al. 2000) and the cells changing their mode of motility from predominantly random movement towards chemotaxis, which is strong at the end of the time period (Fig. 2.4B). The chemotactic response of the *Dictyostelium* cells to the folate gradient increases over time. Slow initial chemotaxis may again be a result of the cells still adapting to move under the gel, while starvation may contribute to the subsequent increase in the efficiency of chemotaxis; starvation results in increasing polarity of the cells, leading to greater persistence in their direction of movement (Zhang et al. 2002). It is also possible that the decreased random movement and increase in chemotaxis is caused by repression of macropinocytosis, which is important for feeding but incompatible with chemotaxis (Veltman et al. 2014). The production of folate deaminase (the enzyme responsible for folate depletion) by *Dictyostelium* has previously been found to increase over time in response to folate exposure (Bernstein et al. 1981). However, I found no evidence for this trend in the rate with which the *Dictyostelium* cells analysed here deplete folate (Fig. 2.4C). It is possible that this increase in enzyme production had already occurred by the time the first image was obtained, over an hour after the cells were added to the system, and was, therefore, not detectable in the data. For melanoma, only the chemotactic responsiveness of the cells shows a temporal trend, declining over the course of the assay (Fig. 2.5). This decline could be caused by cells being imperfectly maintained during the longer assay conditions, or by endoctyosis and degradation of the LPA receptor (LPAR1), which is a universal behaviour (Donà et al. 2013).

To conclude, I have developed an inference methodology that overcomes many of the computational difficulties associated with fitting a set of candidate PDE models for cell movement to data. I have applied these methods to data from two systems, one involving *Dictyostelium*, a well-studied model organism in this field, and the other involving human melanoma, a cancer made particularly aggressive by its rapid spread. Through model comparison, I have drawn conclusions about the drivers of movement in these systems, many of which are in agreement with previous experimental and modelling work, and, thus, offer a validation of the inference methods applied. The study systems examined here are relatively simple in comparison with the levels of complexity often observed *in vivo*, where multiple cell types may be interacting within a considerably more complex environment. However, they are nonetheless examples of real cell movement behaviour, one of which is of great medical relevance, in which I have been able to detect the presence of self-generated chemotactic gradients; a movement driver recently found to be important in many systems, including *in vivo* (Scherber et al. 2012, Donà et al. 2013, Venkiteswaran et al. 2013, Muinonen-Martin et al. 2014, Tweedy et al. 2016). This success is an encouraging first step, indicating that model inference has the potential to support targeted experimental work in increasing our understanding of collective cell movement in a range of systems.

3. Bayesian inference of the spatio-temporal mechanisms driving collective *Dictyostelium* movement

The work presented in this chapter has been published at the following reference:

Ferguson, E.A., Matthiopoulos, J., Insall, R.H. & Husmeier, D., 2017. Statistical inference of the mechanisms driving collective cell movement. Journal of the Royal Statistical Society. Series C: Applied Statistics, 66(4), pp.869–890. Available at: http://dx.doi.org/10.1111/rssc.12203

3.1. Introduction

In the preceding chapter, I developed a pseudo-Bayesian approach to inference that involved running parameter optimisations on many bootstrap samples of the cell movement data being analysed, to produce a pseudo-posterior. Using the pseudo-posteriors obtained for a set of candidate models and a cell movement dataset, I was able to calculate WAIC (widely applicable information criterion (Watanabe 2010)) for each model, so as to select the optimal model and draw conclusions about the drivers of cell movement in the data. As discussed in chapter 2 (section 2.5), this pseudo-Bayesian approach has an advantage over a frequentist approach based on maximum likelihood in that it allows use of WAIC, rather than less reliable model comparison statistics like AIC (Akaike 1974) and BIC (Schwarz 1978), which tend to select overly complex and overly simple models respectively (Ripplinger and Sullivan 2008). The many parameter optimisations to be run on the bootstrap samples of the data are computationally costly, but they can easily be run in parallel to limit the time cost, also providing this pseudo-Bayesian approach an advantage over fully Bayesian approaches based on MCMC (Markov chain Monte Carlo) sampling, which are inherently sequential and cannot be parallelised. Despite this, there are two key disadvantages of the pseudo-Bayesian approach over fully Bayesian methods. The first is that it does not allow full advantage to be taken of prior information about the parameters; an assumption of uniform prior distributions is made, meaning that upper and lower bounds for a parameter can be specified, but more nuanced prior distribution information, for example, about the mode or skewness, cannot be accommodated. The second issue with the pseudo-Bayesian approach is that it has thus far undergone fairly limited testing to prove that it can adequately approximate the true posterior distribution. I carried out an initial test of the method in Appendix A.4, which showed that WAIC values calculated for various polynomial models fitted to a test dataset using samples from the true posterior versus using samples from a pseudo-posterior generated through bootstrapping were closely correlated, giving some confidence in the validity of the method. However, further testing of the bootstrapping approach on more complex models, with more complex likelihood surfaces, is needed before it can be established that the method is comparable to a fully Bayesian approach in all cases.

For the two reasons outlined above, development of a Bayesian method for fitting advection-diffusion PDE models for cell movement based on MCMC sampling is desirable. Such methods for fitting PDE models of the spatio-temporal distribution of organisms using a hierarchical Bayesian framework have previously been proposed in the literature (Wikle and Hooten 2006, Cressie and Wikle 2011), but have typically been applied to much simpler models than the cell models with complex advection coefficients describing a range of movement mechanisms that I outlined in section 2.3. For these complex advection-diffusion models, which must be solved numerically at each step in an MCMC simulation at great computational cost, achieving convergence of MCMC chains may not be feasible using traditional approaches.

One possible influence on cell movement behaviour that was not considered in the models described in section 2.3 is that of spatial features in the environment. The experiments used to produce the datasets analysed in chapter 2 were purposefully set up in such a way that there were as few spatial effects on movement as possible. However, the trough in the agarose gel in which the *Dictyostelium* cells began the experiment is one spatial feature that is known to affect movement rates, since the cells experience resistance as they transition from the trough to the narrow gap under the gel (Laevsky and Knecht 2001). Similar effects could also have affected the melanoma cells as they transitioned between the central bridge and the troughs to the left and right of their spatial region. For this reason spatial variation in the cell movement parameters should be considered, in addition to the temporal variation that I have already considered in chapter 2.

In this chapter, I developed a Bayesian inference scheme that uses the delayed rejection adaptive Metropolis algorithm (DRAM; Haario et al. (2006)), with some changes to the standard protocol for achieving convergence that allow this inference approach to be feasible in the face of computationally costly numerical solutions of complex advection-diffusion models. I applied this inference scheme to fit a set of candidate advection-diffusion models for cell movement – including the 6 models previously described in section 2.3 and 3 additional models, which in this chapter all incorporated both spatial and temporal dependencies in the parameters – to data on the movement of *Dictyostelium* cells. Model selection was then carried out on the basis of WAIC. The main *Dictyostelium* dataset analysed in this chapter was collected in a repeat of the experiment used to produce the data analysed in Chapter 2 (Tweedy et al. 2016), allowing an assessment of the repeatability of the inference results between different groups of cells of the same species.

3.2. Data

Two datasets on the movement of groups of *Dictyostelium* cells were utilised in this chapter, both collected by Tweedy et al. (2016). The first of these was obtained using the same experimental conditions that produced the *Dictyostelium* data described in section 2.2, where the cells moved under agarose containing folate at an initially homogeneous concentration of 10μM. The second dataset was collected using the same procedure, but with agarose containing 0μM folate. The cell movement was imaged under a microscope (Fig. 3.1A) over 5.5 hours for the 10μM folate dataset and 3.5 hours for the 0μM folate dataset, and I manually extracted the coordinates of the cells from the images at half-hourly intervals. As in chapter 2, I collapsed the dataset along the *y*-axis for the analyses, considering only the *x* coordinates of the cells (an additional analysis supporting this simplifying assumption of one-dimensional movement is presented in Appendix B.1). For the 10μM folate dataset, one-dimensional density estimates obtained from the cell location data show a very similar pattern to that observed for the dataset collected under the same conditions that I described in chapter 2 (see Fig. 2.2B-M); a gradual spread of the group of cells up the spatial axis, and the development of a bimodal cell distribution, with one peak indicating the progressing cell front and a second peak indicating the cells' point of origin in a trough cut into the agarose along the far left of the region (Fig. 3.1B-M). Note, however, that the peaked cell front emerges later (at around 3 hours, compared to 2 hours) and is less pronounced in the dataset described in this chapter than in the dataset described in chapter 2

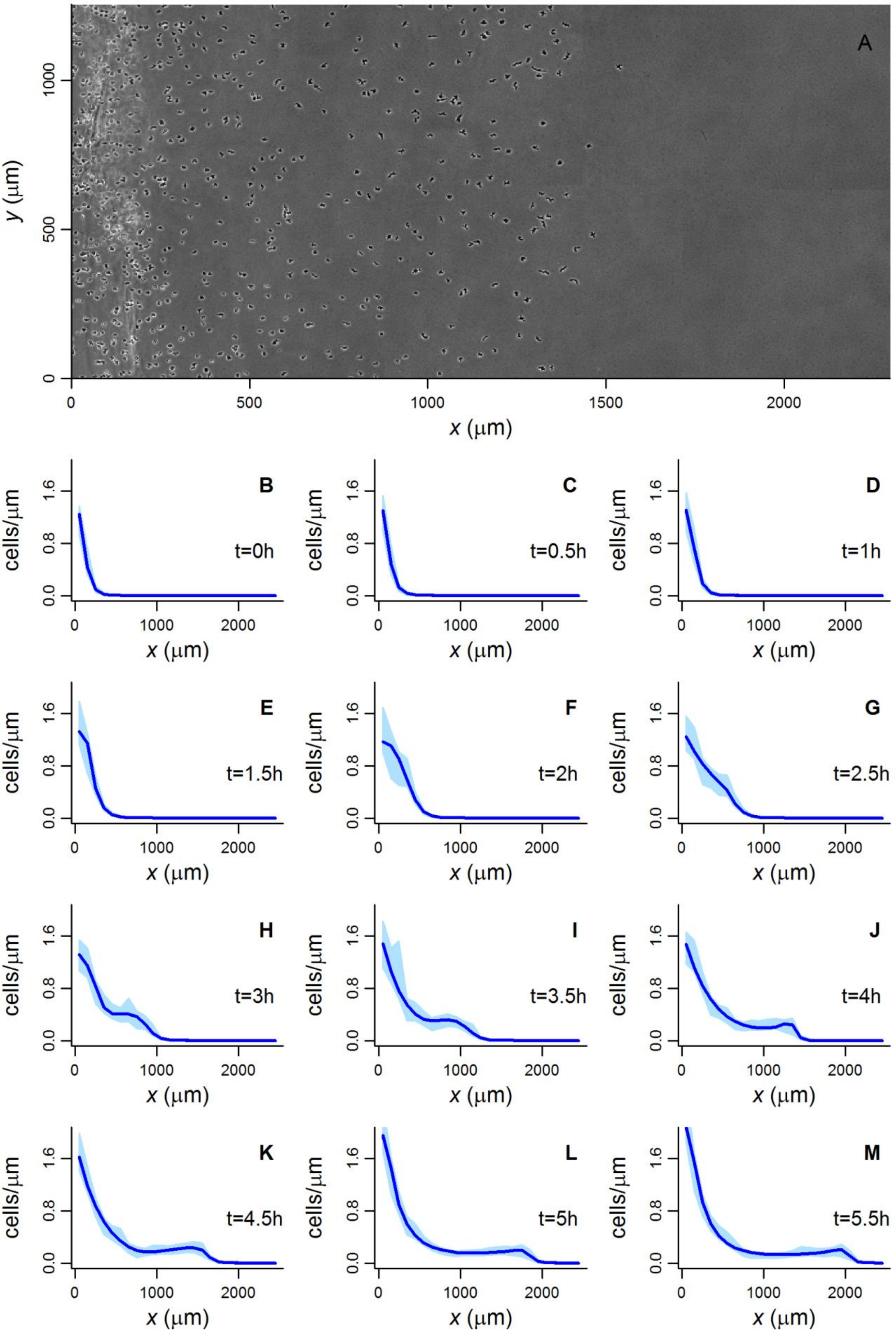

Figure 3.1: 10μM folate *Dictyostelium* data. A) Example image from the *Dictyostelium* cell movement dataset with 10μM of folate in the gel. This image was obtained 4 hours into the experiment (compare with **J**). The edge of the trough from which the cells originated is visible at the far left. **B-M)** One-dimensional logspline density estimates (Stone et al. 1997) showing the cell distribution at half-hour intervals. 95 percentile intervals were obtained by non-parametric bootstrapping, using 10,000 samples of the data.

(compare Fig. 3.1B-M to Fig. 2.2B-M). There is no pronounced peaked cell front in the 0μM folate data, where cells move out from the trough more slowly and in lower densities (Fig. 3.2).

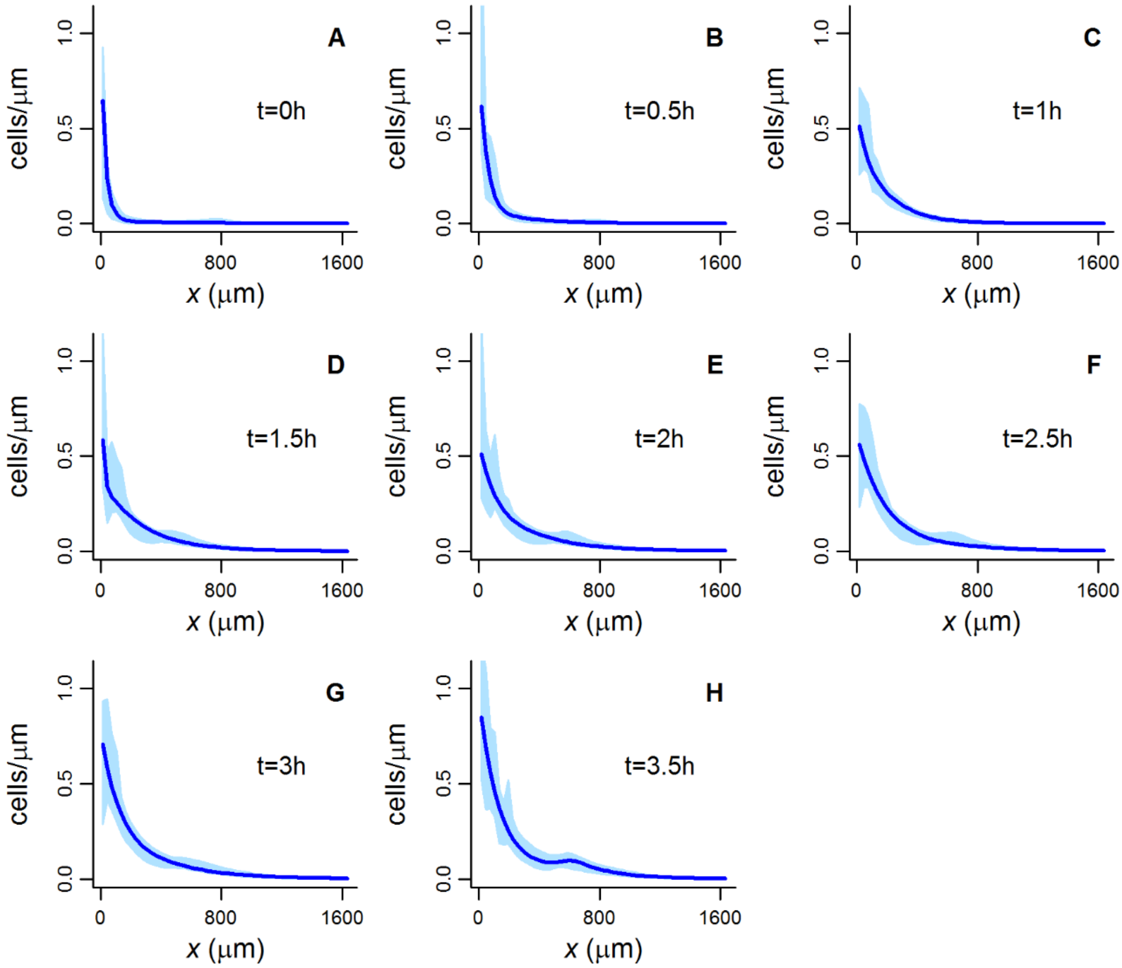

Figure 3.2: 0μM folate *Dictyostelium* data. One-dimensional logspline density estimates (Stone et al. 1997) showing the cell distribution at half-hour intervals from the 0μM folate dataset. 95 percentile intervals were obtained by non-parametric bootstrapping, using 10,000 samples of the data.

3.3. Models

3.3.1. Model descriptions

In this chapter, I again consider the diffusion, basic, receptor saturation, receptor saturation and interaction, receptor saturation and overcrowding, and full models described in chapter 2 (section 2.3, equations (2.1-8)), with the cell division rate ν assumed to be zero (since cell division is anticipated to be only a very minor contributor to changes in cell density over the experimental time periods). I also add three additional models containing further possible combinations of the movement drivers considered in the advection coefficients of the original set of model. The names and advection coefficients of these new models are as follows:

- **interaction model**:

$$a(x,t) = \alpha(x,t)\frac{\partial A(x,t)}{\partial x} + \frac{\eta(x,t)}{1+\lambda C(x,t)}\frac{\partial C(x,t)}{\partial x} \tag{3.1}$$

- **overcrowding model**:

$$a(x,t) = \left(1 - \frac{C(x,t)}{C_{\max}}\right)\left(\alpha(x,t)\frac{\partial A(x,t)}{\partial x}\right) \tag{3.2}$$

- **interaction and overcrowding model**:

$$a(x,t) = \left(1 - \frac{C(x,t)}{C_{\max}}\right)\left(\alpha(x,t)\frac{\partial A(x,t)}{\partial x} + \frac{\eta(x,t)}{1+\lambda C(x,t)}\frac{\partial C(x,t)}{\partial x}\right) \tag{3.3}$$

As in chapter 2, the parameters α , D_C , γ and η are permitted to vary in time to account for changes in cell state. In this chapter, I also consider spatial variation in the parameters. Spatial effects on the parameters are expected to be limited due to the experimental set-up; the cells are moving under a gel, the structure and initial composition of which do not vary throughout the majority of the modelled region. However, the trough cut into the agarose gel in which the cells began the experiment is one major spatial feature in the cells' environment that could affect movement rates, as the cells will experience resistance as they move from the trough and under the gel (Laevsky and Knecht 2001). The parameters directly controlling cell movement rates (α, D_C, and η) are therefore allowed to vary in space in addition to time. It is anticipated that the depletion rate of folate could increase over time as the cells, induced by their exposure to folate, release more and more folate deaminase (the enzyme responsible for breaking down folate) into their environment (Bernstein et al. 1981). However, there are no spatial features present in the environment of the cells that could influence folate deaminase production (it will be unaffected by the presence of the trough for example). Hence, the folate depletion rate γ (equation (2.4)) is allowed to vary in time, but not in space. Spatial and temporal dependence in η was implemented through the description:

$$\eta(x,t) = E + F(x) + G(t) \tag{3.4}$$

where E is a constant, and $F(x)$ and $G(t)$ are polynomials, with zero intercepts, in space and time respectively. For α, D_C, and λ, which are constrained to values ≥ 0, I exponentiated the right hand side of equation (3.4); taking D_C as an example:

$$D_C(x,t) = \exp\left(E + F(x) + G(t)\right) \tag{3.5}$$

Note that for λ, the coefficients of $F(x)$ were set to zero. The degrees of the polynomials $F(x)$ and $G(t)$ were chosen through statistical model selection, as described in section 5.

I formally adopt the hierarchical Bayesian modelling framework proposed in Cressie & Wikle (2011), page 114, and specify probability distributions at three tiers of a basic hierarchy:

1. Data model: p(data|process,parameters)

2. Process model: p(process|parameters)
3. Parameter model: p(parameters)

At the bottom level of this hierarchy are the prior distributions of the parameters $\boldsymbol{\theta}$, which I describe in section 3.3.2 below. The time-varying probability distribution $p(x\,|\,t,\boldsymbol{\theta})$, given by the solution of the PDEs based on $\boldsymbol{\theta}$, forms the second tier in the hierarchy, and provides the likelihood of each observation (x_i, t_i) given $\boldsymbol{\theta}$ (equation (2.10)). The distribution at the top level of the hierarchy corresponds to the observational noise model. I could distinguish between the measured cell locations $\tilde{x}_i$ and the unknown true cell locations x_i, with the observational noise model:

$$p(\tilde{x}_i\,|\,x_i) = N(\tilde{x}_i\,|\,x_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-1}{2\sigma^2}(\tilde{x}_i - x_i)^2\right) \tag{3.6}$$

Given a data set of cell locations $D = \{x_i\}_{i=1,n_j}$ at time t_j, the terms entering the likelihood would then be of the form:

$$\begin{aligned} p(\tilde{x}_i\,|\,t_j,\boldsymbol{\theta}) &= \int p(\tilde{x}_i, x_i\,|\,t_j,\boldsymbol{\theta})dx = \int p(\tilde{x}_i\,|\,x_i)p(x_i\,|\,t_j,\boldsymbol{\theta})dx_i \\ &= \int N(\tilde{x}_i\,|\,x_i,\sigma^2)\,p(x_i\,|\,t_j,\boldsymbol{\theta})dx_i \end{aligned} \tag{3.7}$$

which is a convolution of the previous density with a Gaussian kernel of variance σ^2. To get an initial estimate of σ^2, I manually extracted the cell locations twice, with a year elapsing between the extractions, ensuring independence. The cell locations from one extraction were paired with their nearest neighbours from the second extraction. Fig. 3.3 shows a scatter plot of these paired locations, and indicates very good agreement. A reasonable initial estimate for σ^2 is:

$$\hat{\sigma}^2 = \frac{1}{\sum_{j=1}^{T} n_j} \sum_{j=1}^{T}\sum_{i=1}^{n_j}\left(\tilde{x}_i(t_j) - \tilde{\tilde{x}}_i(t_j)\right)^2 \tag{3.8}$$

where $\tilde{x}_i$ and $\tilde{\tilde{x}}_i$ are the two independent localisations of cell i at time t_j. In this way I find $\hat{\sigma}^2 = 52.29\mu m^2$, which implies $\hat{\sigma} = 7.23\mu m$. The spatial discretisation involved in numerically solving the partial differential equations in this chapter is based on a spatial grid size of $30\mu m$ for the 0µM folate dataset and $100\mu m$ for the 10µM folate dataset. Consequently, the estimated standard deviation of the observational noise is smaller by one or two orders of magnitude than the spatial resolution of the numerical discretisation, and accounting for it would have no practical effect. I, therefore, discard the observation model, and assume that $\tilde{x}_i = x_i$.

3.3.2. Prior distribution

I was able to obtain literature values for two of the model parameters; the dissociation constant K_d (De Wit et al. 1986) and the diffusion coefficient D_A (Kalimuthu and John 2009, Ershad et al. 2013) of folate. For D_A, where I had confidence in the literature values due to their high level of consistency, I specified a rescaled beta prior, with mode positioned at the literature value and cut-offs positioned close to this value. For K_d, I specified a gamma prior with a mode of the literature value and scale chosen such that the probability fell to practically zero within an order of magnitude. These priors enforce the required positivity constraint.

As for the *Dictyostelium* dataset produced with 10μM folate in chapter 2, I used knowledge of the experimental conditions to set sensible boundaries on the values of the parameters describing the initial sigmoidal distribution of folate, δ and ε (equation (A.1.8) in Appendix A.1; Appendix B.3), so that the priors for these parameters could be described using rescaled beta distributions.

For the remaining parameter priors, I used simulations from the models to identify values of the parameters beyond which the cell distributions differed substantially from those observed. Priors were then defined on the basis of these extreme values as either Gaussian distributions with mode zero or exponential distributions, with scales chosen such that the probability of extreme values was close to zero. Full details of the priors applied in this study can be found in Appendix B.3.

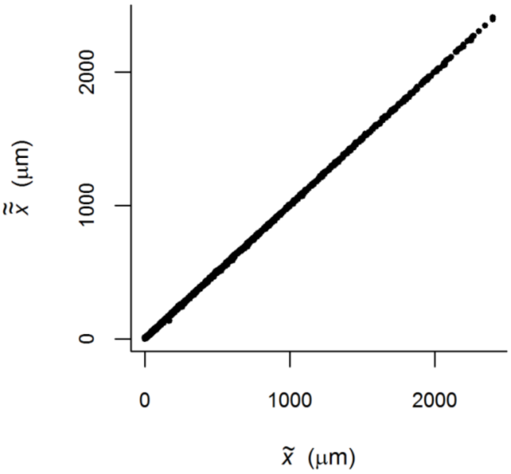

Figure 3.3: Observation noise. Two independent extractions $\tilde{x}_i$ and $\tilde{\tilde{x}}_i$ of the location of each cell plotted against one another. Note the close agreement between the two values, indicating minimal observation noise.

3.4. Bayesian model inference

As in chapter 2, numerical solution of the PDEs was carried out using the method of lines (Schiesser and Griffiths 2009, Soetaert et al. 2010; see Appendix A.1.1 for details) to obtain spatiotemporally varying functions describing cell density $C(x,t)$ and attractant concentration $A(x,t)$. The initial cell density distribution $C(x,0)$ was obtained for each dataset from the cell locations at t=0 using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) as before, and, for the 10μM folate dataset, the unknown initial folate distribution was again assumed to follow a sigmoidal distribution (Appendix A.1.2). The form of the boundary conditions was also assumed to be the same as before (Appendix A.1.4), with cell fluxes of zero on the right boundary of the region and $N'(t)$, the rate of change in the number of cells in the region with time, on the left boundary; see Appendix B.3 for the form of this left boundary function for the two datasets examined in this chapter. Calculation of the standard and weighted log-likelihoods (equations (2.11-12)) of a set of parameters **θ** given one of the datasets (composed of the locations of all the cells in the region of interest at half hourly time points) was achieved using the cell density curve *C*(*x*,*t*) as outlined in section 2.4.

I followed a Bayesian approach to inference and sampled parameters from the posterior distribution with MCMC. The key question was what kind of MCMC scheme to use. I attempted inference with standard random walk Metropolis MCMC, but this proved to be too slow in mixing. Advanced schemes, such as Hamiltonian Monte Carlo, which require repeated likelihood computations along the proposal path, are computationally inefficient, due to the high computational costs of the numerical solution of the PDEs. A reasonable compromise is the delayed rejection adaptive Metropolis (DRAM) algorithm, proposed by Haario et al. (2006). This is an MCMC algorithm with a multivariate proposal distribution that is automatically adapted to allow for posterior correlations among the parameters and to identify the directions of principal change along the ridges in the posterior landscape. The acceptance rate is improved by the delayed rejection part of the algorithm where, instead of immediately advancing the chain following rejection of a parameter set, a second proposal is made that depends both on the current position of the chain and the rejected parameter set. Multiple additional proposals can be implemented if desired. I implemented DRAM using the function modMCMC in the FME package (Soetaert and Petzoldt 2010) in R (R Core Team 2015), using one delayed rejection step, and updating the proposal distribution every 10 iterations.

The absence of any attractant in the experimental conditions that produced the 0μM folate dataset meant that I could immediately rule out all the models described in sections 2.3 and 3.3 that included a response to a chemoattractant gradient, leaving only the diffusion model (equations (2.1-2, 3.5)). It is, however, acknowledged that, just as for the 10μM folate data, responses to the conspecific density gradient and overcrowding could have been present in the 0μM folate data. I discuss the potential consequences of the failure to consider these behaviours in section 3.6. The 0μM folate dataset was used to determine the appropriate degrees of the polynomials describing the dependencies of the cell diffusion parameter D_C on space and time (equation (3.5)). A possible approach is to use RJMCMC (Green 1995). However, convergence is typically slow, which is aggravated by the high computational costs of the numerical solution of the PDEs, and the sequential nature of the process. An alternative approach is the separate computation of marginal likelihoods; see e.g. Friel & Pettitt (2008). However, in combination with the numerical solution of the PDEs, the computational costs are unrealistically high. The method can in principle be parallelised, but in practice the parallel processing capacity is already used up by the parallel tempering scheme on which the method is based. An alternative approach, which is computationally less expensive, and promoted in Gelman et al. (2013), Chapter 7, is WAIC (Watanabe 2010), calculated as:

$$\begin{aligned}\mathrm{WAIC} = &-2\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\\ &+2\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right\}\end{aligned} \tag{3.9}$$

where m is the number of parameter sets sampled from the posterior, $\left(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m\right)$ are these parameter sets, and $\left\{y_j=\left(x_j,t_j\right)\right\}_{j=1,n}$ are the cell observations. This score can be directly computed from the MCMC trajectory, and the computation is straightforward to parallelise, as the MCMC trajectories for different models can run on different processors simultaneously. I, therefore, fit versions of the diffusion model with polynomial degrees for the dependencies of D_C on time and space ranging from zero to six, and select the best combination of polynomial degrees as that giving the lowest WAIC. Two chains were run from random parameters for each model variation, and I assessed within-chain convergence using the Geweke diagnostic (Geweke 1991) and between-chain convergence using the Gelman-Rubin statistic (Gelman and Rubin 1992).

For the 10μM dataset, I first took the degrees of the polynomials describing spatial and temporal dependencies in D_C from the 0μM folate dataset, and then carried out a local readjustment of these degrees using the diffusion model applied to this new dataset (see Appendix B.5 for details). I then ran MCMC simulations for the remaining eight candidate models using the 10μM folate data. To keep the approach computationally feasible, I used the same polynomial degrees in space and time as were selected for D_C using the diffusion model for all four of the parameters with spatial and temporal dependencies (α, η, γ and D_C) in the other more complex models.

Note that the advection terms entering all models other than the diffusion model are complex nonlinear functions that model the processes of cell-cell interaction, cell-molecule interaction, receptor saturation, etc. This has two consequences that affect MCMC convergence: (1) the additional nonlinear complexity changes the topology of the log-likelihood, leading to a higher degree of multi-modality, and (2) the system of coupled nonlinear differential equations is stiff, leading to a substantial reduction in the numerical integration step size (for numerical stabilisation). The second aspect is particularly dramatic. I found that by including the advection term, the numerical solution of the differential equations slowed down by a whole order of magnitude as a mere consequence of the step size adjustment. Since the numerical solution of the differential equations is required in every step of the MCMC simulation, the impact on the overall runtime is substantial: for the models other than the diffusion model, no indication of convergence was found despite a month of run time.

With the computational resources available, I could typically carry out 100,000 MCMC steps per week for the diffusion-only model, but only 10,000 MCMC steps per week for many of the more complex models with the nonlinear advection terms included. To obtain a reasonable degree of convergence, quantified in terms of the Gelman-Rubin statistic obtained from independent simulations started from hyperdispersed starting points, I would have required far in excess of 100,000 MCMC steps for the models with the advection term included, which was computationally infeasible.

To deal with this problem, I adopted the following approximation. I started with repeated maximisations of the log-likelihood (more accurately: the log unnormalised posterior), to obtain a good approximation of the MAP (maximum a posteriori parameter configuration). This exploits the fact that optimisation is parallelisable, and that approximating the MAP by the best local optimum from several independent initialisations is common practice in complex systems science. I then started two independent MCMC simulations of a minimum 80,000 MCMC steps from the MAP, and checked for convergence based on consistency of the WAIC scores obtained from two sections (the middle and end thirds of the MCMC chains, discarding the first third of steps as burn-in) from two independent MCMC runs (hence giving 4 WAIC scores overall). In this way, I restricted the exploration of the configuration space to the area around the MAP. The justification of this approach is discussed in section 3.6, and a test of the performance of the approach on simulated data is provided in Appendix B.4. I repeated this procedure twice, using both the standard (equation (2.11)) and weighted (equation (2.12)) log-likelihoods.

3.5. Results

WAIC values were obtained for fits of the diffusion model to the 0μM folate dataset with different combinations of polynomial orders for the dependencies of the diffusion rate on space and time (equation (3.5)). I found that a polynomial degree of two in space and four in time was associated with the smallest WAIC values, both for the standard likelihood (equation (2.11); Table B.5.1) and the weighted likelihood (equation (2.12); supplementary Table B.5.2). The cell distributions produced by this model show good agreement with those estimated directly from the data (Fig. 3.4). The patterns of change in cell diffusion in time and space predicted by this model are illustrated in Fig. 3.5 (see also Fig. B.6.1 in Appendix B.6). Cell diffusion is slowest at the beginning and end of the time period of interest, with two peaks in diffusion occurring in the middle. In space, diffusion is slowest at the edges of the region of interest, with a single peak in the centre.

I fitted the diffusion model with a polynomial degree of four in time and two in space (as suggested by model selection on the 0μM folate dataset) to the 10μM folate dataset, and then carried out a local readjustment of the polynomial degrees using this dataset. This involved identifying polynomial coefficients where the posterior distribution was focussed around zero (Fig. B.5.3), and using this information as a guide to which polynomial degrees might be reduced to prevent unnecessary model complexity. I tried different adjustments of the polynomial degrees, and selected the best degrees based on WAIC. This gave a degree of three in time for the standard likelihood and two for the weighted likelihood (Table B.5.7). I maintain a polynomial degree of two in space for both the standard likelihood and weighted likelihood, as suggested by Fig. B.5.3.

WAIC values calculated from the mid and end sections of the two chains for the eight models that include an advection component are closely grouped by model (Fig. 3.6), and the ranking of the models based on these values is consistent across the standard and weighted likelihoods (Fig. 3.6 and Table 3.1). The diffusion model gives a much poorer WAIC value than the other models (Table 3.1), which all include an interaction of the cells with the chemoattractant (folate) in their environment, suggesting that this interaction is necessary for achieving a good fit to the data. For both the standard likelihood and weighted likelihood, the interaction model produces the best mean WAIC value (Table 3.1), but there is a similar level of support for the model that includes both interaction and overcrowding terms, as indicated by the standard errors of the mean WAIC values (Table 3.1), and the large degree of overlap between the four individual WAIC values for these models (Fig. 3.5). On examination of the parameters, I found that the estimated value of C_{max} (the maximum cell density), which implements the overcrowding effect described in equation (3.2)), was very large. A large value of C_{max} essentially causes the interaction and overcrowding model to revert to the interaction model, explaining the similarity in WAIC for these models. I, therefore, select the interaction model as the optimal model for explaining these data. In addition to concluding that the correction for overcrowding has, at most, a very small effect, I also find that the effect of receptor saturation does not improve model fit.

Model outputs from the interaction model show very good agreement with the 10μM folate data (Fig. 3.7), successfully reproducing the steep cell front, which the simpler diffusion model fails to capture (Appendix B.8). A residual analysis finds no significant mismatch between the selected model and the data (see Appendix B.9).

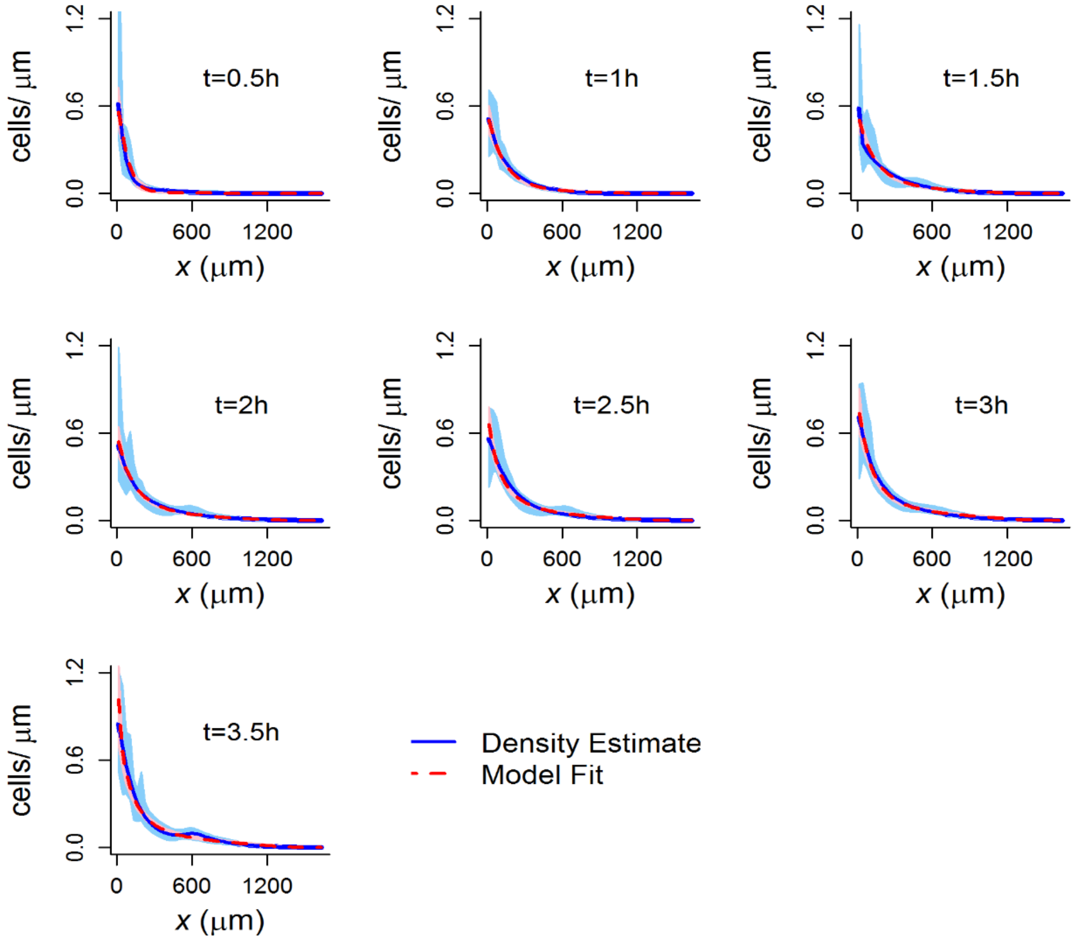

Figure 3.4: Fit of the diffusion model to the 0μM folate data. Plots of the cell distributions at half-hourly intervals simulated (using the posterior mean parameters) from the diffusion model fitted to the 0μM folate data using the standard likelihood (equation (2.11), with polynomial degrees of four and two describing the temporal and spatial dependencies of the diffusion coefficient respectively. Direct density estimates from the data, obtained using logspline density estimation (Stone et al. 1997), are included for comparison. 95 percentile intervals for the density estimates (blue shaded areas) were obtained by non-parametric bootstrapping, using 10,000 samples of the data. 95 percentile intervals for the model (pink shaded areas) were obtained from 500 samples from the posterior distribution.

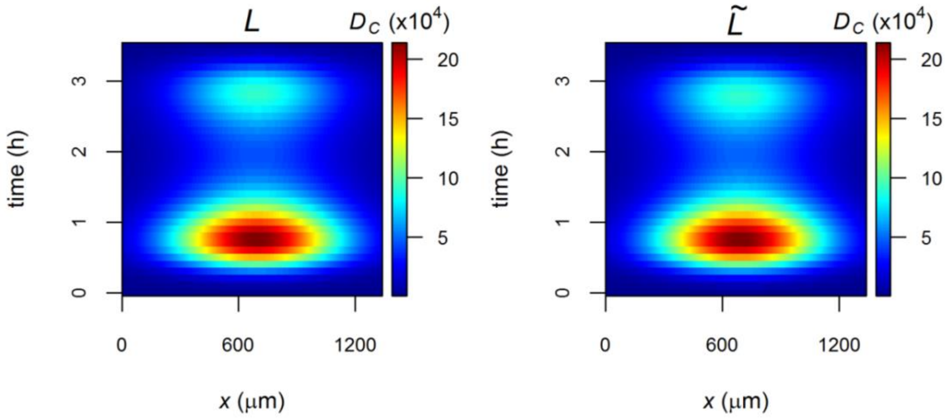

Figure 3.5: Heat maps of spatial and temporal dependencies of the cell diffusion coefficient D_C obtained by fitting the diffusion model to the 0μM folate dataset. Plots show the value of the diffusion coefficient D_C in time and space as calculated in equation (3.5) and estimated using both the standard and weighted likelihoods (L and $\tilde{L}$).

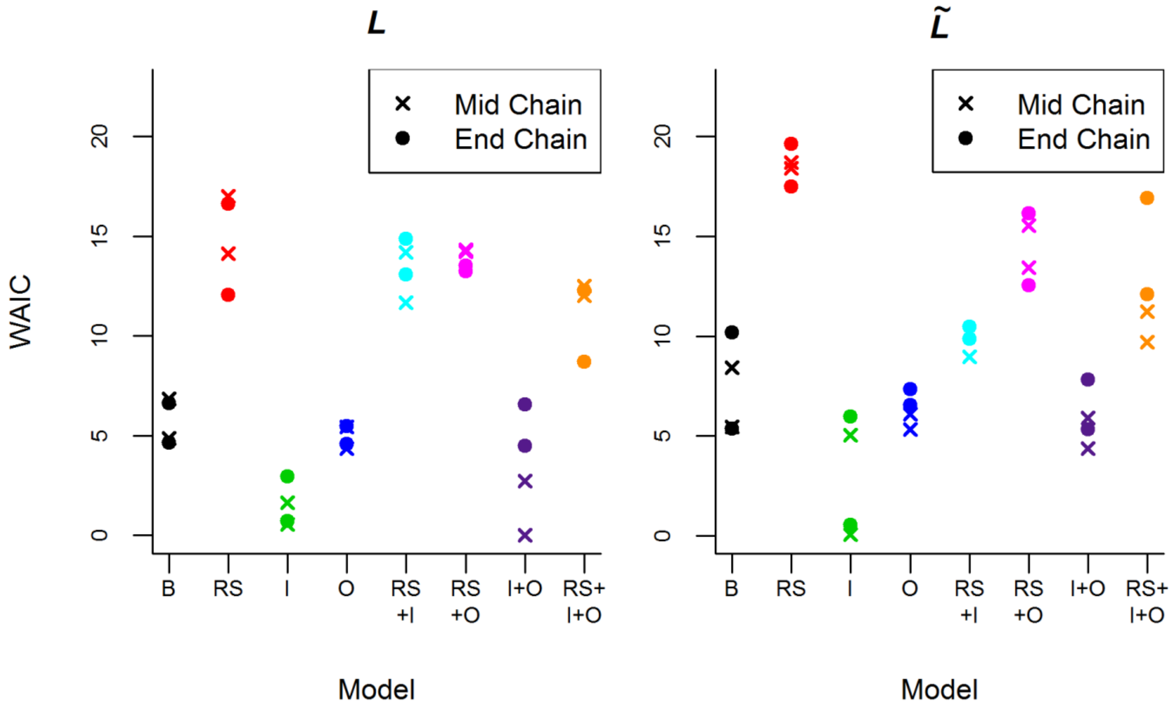

Figure 3.6: Consistency in WAIC values. Plots of the four WAIC values calculated for each of the models fitted to the 10μM dataset using the standard likelihood and the weighted likelihood, L and $\tilde{L}$. For each model, I obtained two MCMC chains, and calculated the WAIC (equation (3.9)) separately for the middle third (crosses) and the end third (points) of each chain. Note that the minimum WAIC value has been subtracted from all values to aid comparison. Model abbreviations: B=Basic, RS=Receptor Saturation, I=Interaction, O=Overcrowding.

Table 3.1: WAIC-based comparison of the candidate models for the 10µM folate data. WAIC values for each model fitted to the 10µM folate dataset, using both the standard (equation (2.11)) and weighted (equation (2.12)) likelihoods, L and $\tilde{L}$. The values for the diffusion model, which was the only model for which I achieved formal convergence of MCMC chains based on the Geweke and Gelman-Rubin diagnostics, were obtained using equation (3.9), with the standard errors (in brackets) being calculated as described in Appendix A.5. The values for all other models were obtained as the means of the 4 WAIC values calculated from the mid and end sections of the chains for those models (Fig. 3.5, Appendix B.7). The best model for each of L and $\tilde{L}$ is marked *. Model abbreviations: RS=Receptor Saturation, I=Interaction, O=Overcrowding.

Model	**WAIC**	
	L	$\tilde{L}$
Diffusion	702.0 (0.1)	605.9 (0.09)
Basic	4.3 (0.58)	4.5 (1.18)
RS	13.5 (1.16)	15.6 (0.44)
I	0 (0.55)*	0 (1.52) *
O	3.5 (0.29)	3.4 (0.42)
RS+I	12.0 (0.69)	6.7 (0.37)
RS+O	12.4 (0.26)	11.5 (0.85)
I+O	2.0 (1.39)	2.9 (0.73)
Full (RS+I+O)	9.9 (0.9)	9.6 (1.55)

Illustrations of the spatial and temporal dependencies of the parameters of the interaction model fitted to the 10µM data can be found in Figs 3.8-9 (see also Fig. B.6.2 in Appendix B.6). The standard and weighted likelihoods gave good agreement in their estimates of the parameter trends in time and space with one exception. This single case of disagreement occurred for the parameter η (describing attraction/repulsion between cells), which was found to decrease with x when fitted with the standard likelihood (Fig. 3.8B) and to increase with x when fitted with the weighted likelihood (Fig.3.9B). There is also a slight trend for this parameter to increase over time for both the standard and weighted likelihoods. The response of the cells to the folate gradient is estimated to become stronger with time and weaker with increasing x (Figs 3.8A, 3.9A). Cell diffusion is slow initially, peaks at around 3.5h and then starts to decline again. It also tends to decrease in x (Figs 3.8C, 3.9C). The rate of folate depletion increases with time (Figs 3.8D, 3.9D).

3.6. Discussion

In this chapter, I developed a detailed protocol for Bayesian inference in PDE models of cell migration and interaction. Hierarchical Bayesian frameworks have previously been proposed for fitting advection-diffusion PDE models describing spatio-temporal distributions of organisms (Wikle and Hooten 2006, Cressie and Wikle 2011). However, these frameworks have typically been applied to models that are relatively simple, including few movement mechanisms, and the key advance of this work is in the consideration of a range of processes relating to the way cells sense and interact with their environment, leading to complex non-linear advection terms. This leads to stiff PDEs, for which the numerical integration step size must be taken to be very small to stabilise the numerical solution, substantially increasing computational costs. Consequently, adequate adaptations are required to render statistical inference computationally viable.

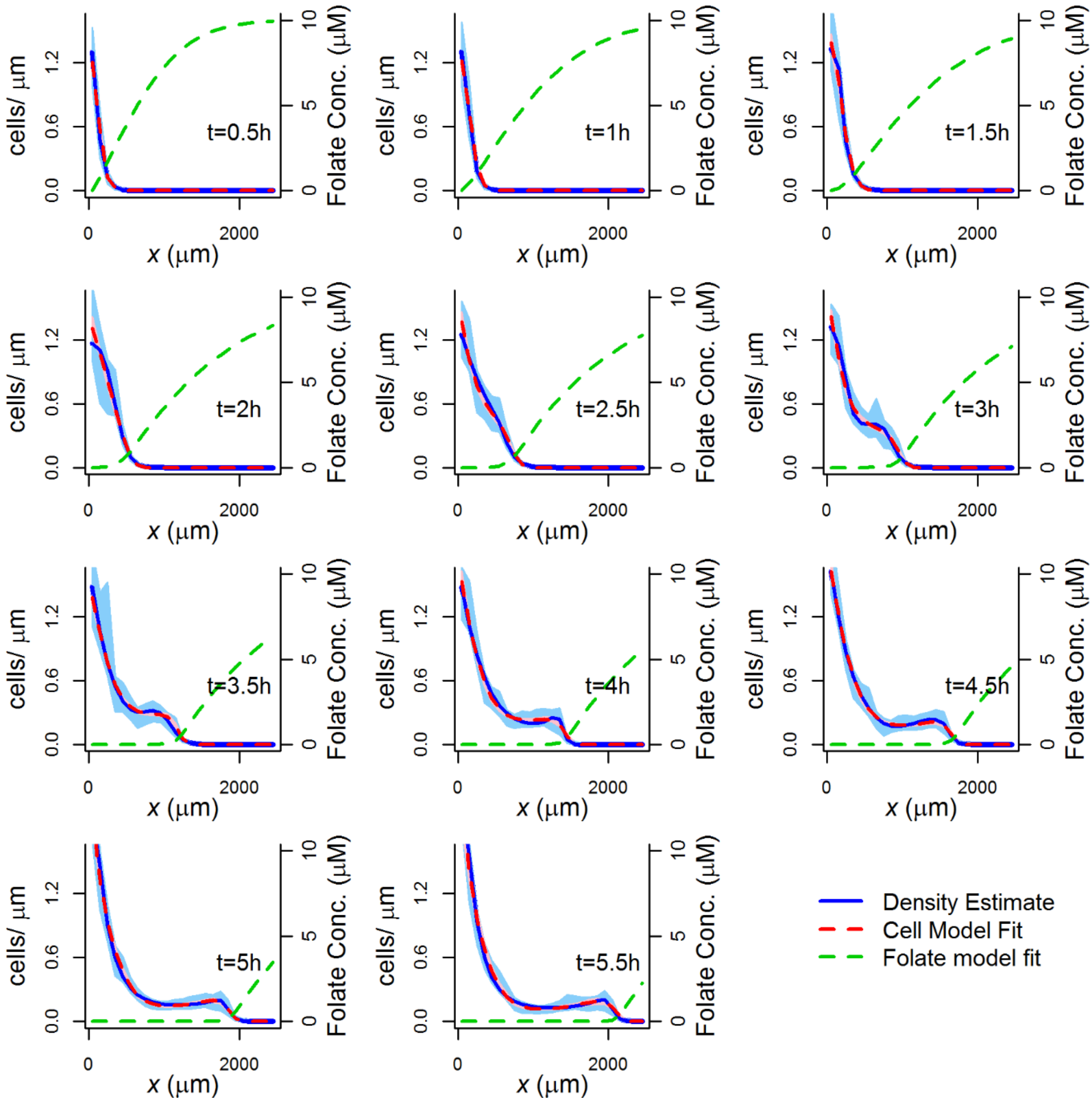

Figure 3.7: Fit of the interaction model to the 10µM folate data. Plots of the cell distributions at half hourly intervals simulated from the interaction model fitted to the 10µM folate data using the standard likelihood (equation (2.11)). I used the MAP (maximum a posteriori parameter configuration) of the model to produce the model fit lines. 95 percentile intervals for the model (pink shaded area) were obtained from 250 parameter sets sampled evenly from the latter two thirds of the two MCMC chains for this model. Direct density estimates from the data, obtained using logspline density estimation, are included for comparison. 95 percentile intervals for the density estimates (blue shaded area) were obtained by non-parametric bootstrapping, using 10,000 samples of the data.

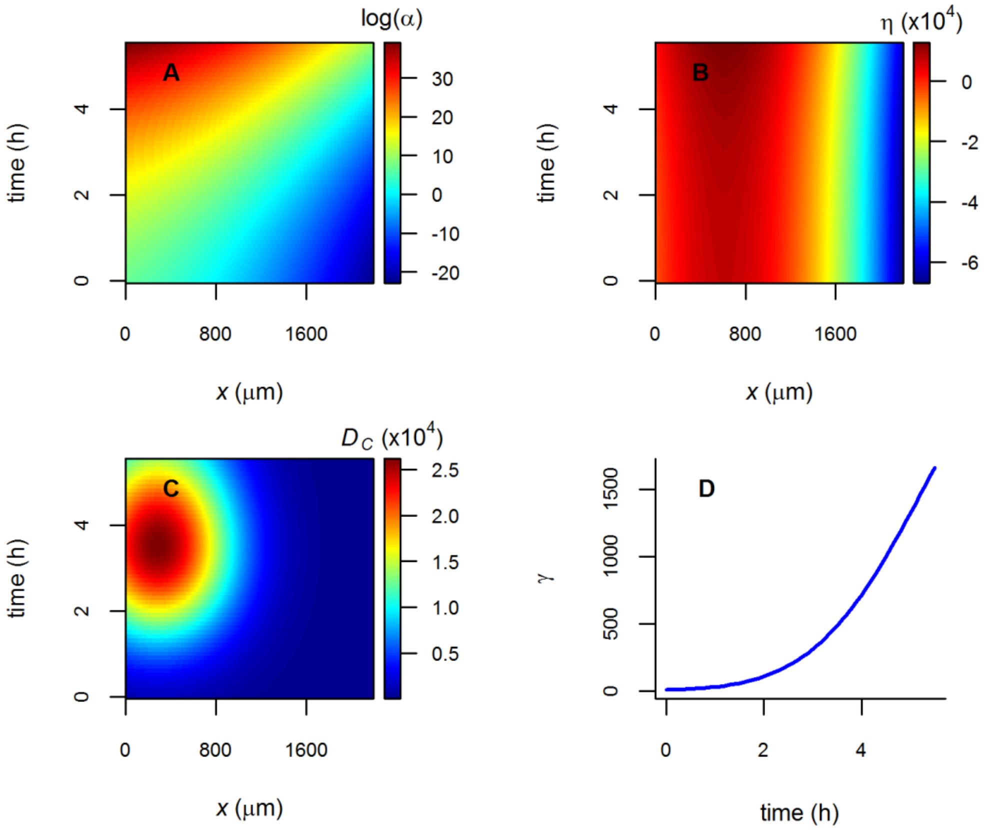

Figure 3.8: Spatial and temporal dependencies of the parameters of the interaction model fitted to the 10μM folate dataset using the standard likelihood. Plots show the cell advection rates in response to folate (**A**) and conspecific density (**B**) gradients, the cell diffusion rate D_C (**C**), and the temporal dependence of the folate depletion rate (**D**). Function values were calculated as described in equations (3.4-5).

The approach to Bayesian inference I have adopted here has a particular focus on model selection: given a set of hypotheses for the mechanisms driving cell migration, which ones are most consistent with the data? Model selection via Bayes factors, either directly estimated via parallel tempering (Friel and Pettitt 2008), or indirectly by RJMCMC (Green 1995), is computationally intractable due to the need to solve a stiff system of PDEs in every step of the Markov chain. Classical information criteria, on the other hand, such as AIC or BIC, rely on asymptotics that are hardly met in practice, especially not for the high degree of nonlinear complexity inherent in the models considered here. As a compromise between numerical tractability and accuracy, I have adopted an approach based on WAIC (Watanabe 2010). This approach is similar to DIC (Spiegelhalter et al. 2002) in spirit, but has been shown to be more “widely applicable” in the sense that it is not restricted to non-singular likelihood functions (as opposed to DIC). WAIC has been favourably reviewed in Gelman et al. (2013), Chapter 7. A recent study suggests that for model selection in complex nonlinear systems, WAIC clearly outperforms DIC and is on a par with Bayes factors (Aderhold et al. 2017).

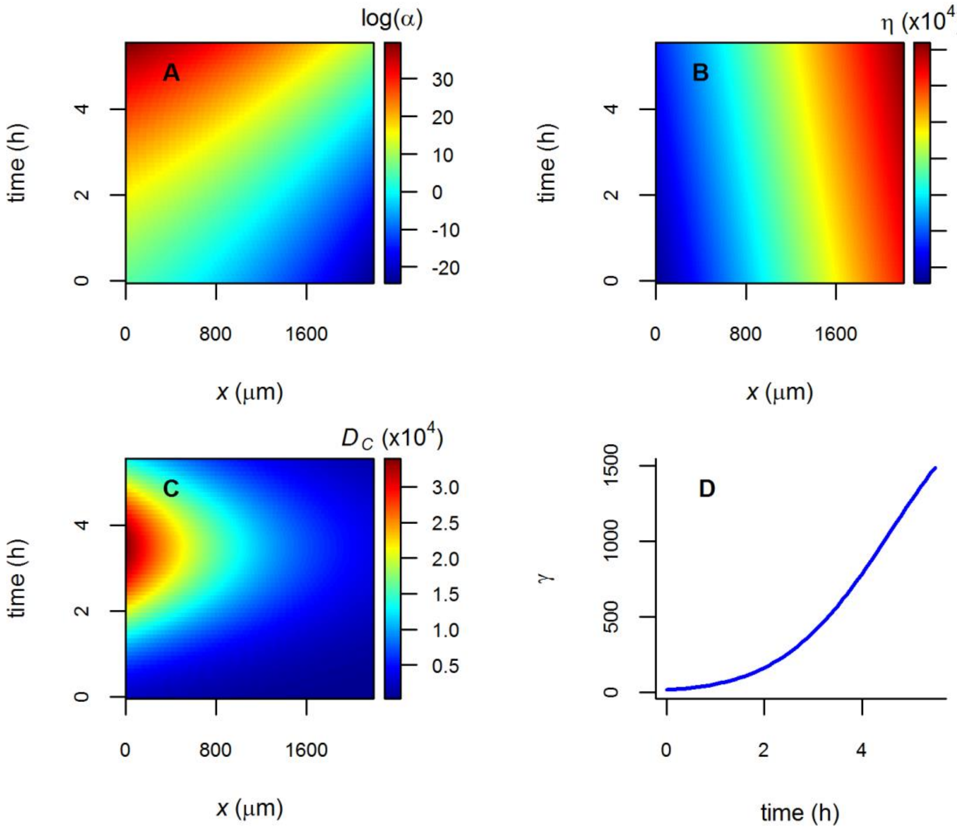

Figure 3.9: Spatial and temporal dependencies of the parameters of the interaction model fitted to the 10µM folate dataset using the weighted likelihood. Plots show the cell advection rates in response to folate (**A**) and conspecific density (**B**) gradients, the cell diffusion rate D_C (**C**), and the temporal dependence of the folate depletion rate (**D**). Function values were calculated as described in equations (3.4-5).

I have found that the application of the outlined procedure to a diffusion model, e.g. as investigated in Wikle & Hooten (2006) and Cressie & Wikle (2011), is computationally tractable. However, when including a complex advection term, MCMC run times increase substantially as a consequence of the stiffness of the PDEs. This does not allow MCMC simulations of a sufficient length to satisfy established convergence criteria to be run. The method I have proposed to deal with this difficulty is effectively a restriction of the configuration space. Rather than initialising independent MCMC simulations from starting points sampled from a hyperdispersed distribution, I started all MCMC simulations from an estimate of the MAP (maximum a posteriori parameters). I ran independent MCMC simulations over a minimum 80,000 iterations (the first third of which were discarded as burn-in) and computed the WAIC scores in a variety of ways: for different sections (middle versus end) of the same MCMC trajectory, for different MCMC trajectories, and for different objective functions (the standard versus the weighted log likelihood). The results show that the model selection results are consistent (Fig. 3.6). This suggests convergence in the actual WAIC scores, providing confidence in the model selection results.

This inference method has the following justification: (1) Approximating the posterior distribution by the area around the MAP is akin to the Laplace approximation, which is widely applied to complex systems for which MCMC simulations are computationally too expensive (as evidenced by the large number of applications using INLA (Rue et al. 2009)). The method described here is less restrictive than the Laplace approximation, in that it does not require a second-order truncation of the Taylor series expansion. (2) Approximating the posterior distribution by a unimodal model distribution from a standard function family is also commonly done in variational inference, which is another alternative method for systems that are too complex for MCMC (e.g. Bishop (2006)). Again, the approximation I use here is less restrictive than variational inference, in that it does not restrict the approximation to any a priori chosen functional form. In an empirical investigation using simulated data (see Appendix B.4 for details), I found that the level of accuracy and precision of my approach is the same as for model selection with Bayes factors calculated using population MCMC (Girolami et al. 2010).

The only alternative approach that could achieve a degree of MCMC convergence that meets established convergence criteria is to resort to gradient matching (Xun et al. 2013). Here, the computational costs of the individual MCMC steps are substantially reduced by bypassing the need for a numerical solution of the PDEs. However, gradient matching is an approximate method, and the current state of the art incurs a potentially substantial loss in model accuracy (Macdonald et al. 2015). Facing the choice between approximate modelling (gradient matching) and sound inference (standard MCMC convergence) versus accurate modelling (numerical integration) and approximate inference (MCMC around the MAP) I have here opted for the latter alternative. This is in line with the frequently cited proposition by John W. Tukey (1915-2000) that “the approximate answer to the right problem is worth a good deal more than an exact answer to the approximate a problem”. However, an interesting topic for future research is to put this proposition to the test and systematically compare both paradigms empirically.

By applying the proposed Bayesian inference method and model selection using WAIC to a set of nine candidate models, I have drawn a number of conclusions about the mechanisms that drive the *Dictyostelium* movements in the 10μM folate data. These conclusions can be compared to those drawn in chapter 2 by applying the pseudo-Bayesian bootstrapping inference approach to a second 10μM folate dataset obtained in a repeat of the assay used to produce the dataset analysed in this chapter. In this chapter, as in chapter 2, I was able to successfully determine that a self-generated gradient in folate has a significant role in producing the observed movement patterns, as previously reported by Tweedy et al. (2016). This self-generated gradient mechanism is responsible for the sharp, dense moving cell front that is characteristic of these data, and which simple diffusion models fail to replicate. Interest in self-generated gradients is growing rapidly, as studies have suggested that they may play an important role in embryonic development (Donà et al. 2013) and the spread of cancers (Muinonen-Martin et al. 2014). Many other examples of self-generated gradients likely remain to be discovered throughout biomedical science, and the inference framework I have described here, provides a further useful tool for detecting these gradients. This framework also allows estimation of how the form of the latent chemical gradient develops over time, which is generally not possible experimentally; measurement of the chemical gradient requires destruction of the gel under which the cells are moving and ends the experiment, making repeated measurements over time impossible (Tweedy et al. 2016).

Despite its known influence on cell movement behaviour (Tweedy et al. 2013), and the fact that it was determined to be an important part of the mechanism by which the cell interact with the self-generated folate gradient in chapter 2, I did not obtain an improvement in model performance on inclusion of the receptor saturation term in this chapter. This surprising difference in results between chapters could be a consequence of the groups of cells used to produce the two datasets being in slightly different states at the start of the experiments, leading to differences in movement behaviour. Indeed, it is observed that the peaked cell front, which is believed to result from receptor saturation making cells at the very front of the distribution move more slowly than those immediately behind, emerges later and is less pronounced in the dataset analysed in this chapter than in the dataset analysed in chapter 2 (compare Figs 2.2 and 3.1). This could have made an effect of receptor saturation harder to detect in this chapter. Additionally, as discussed further below, the choice of a model without receptor saturation is likely to be, in part, a consequence of the models without the receptor saturation term having enough flexibility to mimic the effect of receptor saturation (slower movement of the cells at the very front of the cell distribution than of the cell directly behind) through temporal and spatial variation in the parameter describing the basic gradient-following mechanism. This leads to there being little improvement in model fit on inclusion of an explicit receptor saturation term. Since spatial variation in parameters was not considered in the models in chapter 2, the basic model was there unable to adequately mimic receptor saturation, and the explicit saturation term was correctly selected for.

In this chapter, I find that including direct interactions between the cells, allowing them to attract or repel one another, provides an improvement in model performance, as indicated by a reduction in WAIC. This result differs from that in chapter 2, where an interaction effect was not incorporated in the best model selected by WAIC (though it should be noted that a model containing cell interactions was placed second by WAIC, and was selected as the optimal model by AICc). I suspect that this slight change in the level of support for cell interactions between this chapter and the previous one is a consequence of the cells being in a slightly different condition or stage in their development. The *Dictyostelium* cells studied in both chapters were vegetative, and therefore lack most of the complex cell-cell interactions of aggregating cells (Varnum and Soll 1981, Bonner 1982). However, vegetative cells can still exhibit weaker interactions, including short-range cell-cell repulsion driven by autorepellents (Keating and Bonner 1977, Kakebeeke et al. 1979). Additionally, lack of nutrients in the environment could cause the cells to starve progressively over the 5.5 hour time period. During starvation, cells go through different phases of development, during which they produce cell surface molecules that affect movement by altering cell-cell interactions. Contact sites A (csA), for example, is induced within hours of starvation (Eitle and Gerisch 1977). CsA mediates cell-cell adhesion, and while aggregation was not visually obvious in the data analysed here, low levels of csA could still modify interactions between the cells. Changes in csA and similar proteins could promote small repulsion and attraction effects, explaining why the interaction model was preferred. It is clear, however, that cell-cell interactions are not the primary driving mechanism of the observed movements in this chapter; the improvement in WAIC obtained by including the interaction effect is smaller by a factor of 100 than that obtained by inclusion of the self-generated folate gradient (Table 3.1).

In this chapter, I estimated functions for three model parameters (α, η and D_C) that varied in both time and space, and one model parameter (γ) that varied in time only. The finding of temporal dependencies in the movement behaviour (see Figs 3.5 and 3.8-9) is in agreement with the results of chapter 2. Temporal dependence in the diffusion parameter of the 10µM folate data is very similar in form between the two chapters, but there are differences for the folate depletion rate and the response to the folate gradient, which could be a consequence of either differences between the cells studied in each chapter, or of the differences in the model structures (compare Fig. 2.4A and Fig. B.6.2D). The finding of spatial variation in the parameters is a new result for this chapter, and is in line with experimental work that has shown that the movement behaviour of cells can be affected by features of their physical environment, such as the rigidity of the substrate

(Lo et al. 2000, Ng et al. 2012). In the cell movement assay used to produce the data analysed in this chapter, the edge of the trough within which the *Dictyostelium* cells were seeded is known to provide resistance to movement (Laevsky and Knecht 2001), potentially causing some of the spatial dependencies that were observed. The estimated spatial and temporal dependencies in the parameter values shown in Fig. 3.5 and Figs 3.8-9 are relatively complex. There are two possible explanations for this complexity, both of which are discussed below: 1. it is caused by real biological processes that affect cell behaviour; 2. it is a consequence of the models being overly flexible due to deficiencies in the approach used to select the degrees of the polynomials describing parameter variation in time and space.

Experimental work has revealed many potential biological causes of changes in cell behaviour, some of which have opposing effects, leading to highly complex and variable patterns. Including mathematical descriptions of all of these details would lead to a model that is overcomplex and computationally intractable. For that reason, I chose a more abstract level of description, describing changes in the relevant coefficients in space and time using smooth polynomial functions. The estimated forms and complexity of these functions (Figs 3.5 and 3.8-9) are not dissimilar in form and complexity to those that have determined empirically (e.g. Bernstein et al. (1981), Chubb et al. (2000)). Potential biological causes of the variation seen in each of the parameters are as follows.

For the 0μM dataset, I found that diffusion was slower both at the start and the end of the time period, but increased in the middle; showing a double peaked profile (Figs 3.5 and B.6.1B). A similar pattern is seen for the 10μM data, but with just a single diffusion peak in the middle of the time period (Figs 3.8-9C and B.6.2D). A low diffusion rate early in the assays is explained by the fact that most of the cells are still positioned in the trough in the gel. Movement is restricted in the trough area as the cells must flatten themselves in order to make their transition under the gel (Laevsky and Knecht 2001). Later in the time period many cells will have moved clear of the trough, leading to an increase in diffusion. The decline in diffusion at the end of the time period may be related to cell starvation, since cells in the early stages of starvation can show a decline in motility (Chubb et al. 2000). Changes in the rate of diffusion in space were also observed. For the 0μM dataset, I found that diffusion was slower at the beginning and end of the spatial axis, with a peak in the middle (Figs 3.5 and B.6.1A). The low diffusion rate at the far left of the spatial region was most likely caused by the presence of the trough in that area. The decline in diffusion at the far right may have been a consequence of the low cell densities in this area. *Dictyostelium* cells in the vegetative state, like those studied here, can release a repellent which will enhance movement at high densities, relative to low densities (Keating and Bonner 1977). In the 10μM data, a decline in diffusion was again observed at the far right of the region, but there was little evidence for reduced diffusion near the trough (Figs 3.8-9C and B.6.2C). The reason for this is unclear, but, since the cells lift the gel slightly as they pass under it (Laevsky and Knecht 2001), the larger number of cells moving under the gel in this dataset compared to the 0μM folate dataset may mean that the transition under the gel becomes easier over time, so that the effect of the trough only occurs early in the time period.

The responsiveness to the folate gradient (α) estimated for the 10μM data decreases in space and increases in time (Figs 3.8-9A and B.6.2A-B). Taken separately these two patterns are hard to explain, but if they are examined together and compared with plots showing the changing cell distribution in time (Fig. 3.1), it can be seen that they result in the value of α always being around log(10) at the position of the moving cell front, and declining towards the edge of the cell front. As a result, the model is able to mimic the effect of receptor saturation, despite the fact that the receptor saturation term (equation (2.5)) was not included. It does this by causing the cells at the very front of the distribution, where folate is less depleted and receptor saturation is, therefore, higher, to advect more slowly than those behind, where folate has been more depleted and receptor saturation is lower. I suspect that this ability of the model to mimic receptor saturation (equation (2.5)) through the flexibility of α in space and time is the reason why an improvement in model fit was not found on inclusion of the receptor saturation term (Table 3.1), despite the fact that this effect has been reported in the literature for cell chemotaxis (Tweedy et al. 2013).

There was only one case where disagreement occurred between the standard and weighted likelihoods in the trend of one of the estimated parameter functions in space or time. This disagreement occurred for the function describing spatial dependence of the parameter η (describing attraction/repulsion between cells), which tended to decrease with x when estimated using the standard likelihood (though with considerable uncertainty; Fig. B.6.2E), but increased with x when estimated using the weighted likelihood (Figs 3.8-9B and B.6.2E). This ambiguity means that the trends for η are difficult to interpret biologically, since clearly at least one of the opposing estimates of the spatial trends is wrong. Furthermore, the weighted likelihood estimated a positive value of η over almost the entire spatial region and time period (Fig. 3.9B), indicating attraction between the cells, while, for the standard likelihood, there is a large region to the right of the spatial region where η is negative (Fig. 3.8B), indicating repulsion. It should be noted that *Dictyostelium* cells are capable of exhibiting both attractive and repulsive dynamics, and can switch between the two based on their developmental state and environmental conditions (Keating and Bonner 1977, Bonner 1982), so that either of the estimated scenarios for η is biologically plausible. However, as both scenarios clearly cannot be true for the same dataset, this is indicative of inaccuracies in parameter inference.

The temporal variation in the folate depletion rate γ is the pattern that can most easily be explained. This rate shows a monotonic increase (Figs 3.8-9D) as a result of the cells' exposure to folate in their environment, which enhances their production of folate deaminase, the enzyme responsible for breaking down folate (Bernstein et al. 1981). The longer the cells are present in the medium, the more enzyme they will have released, leading to faster folate depletion.

A limitation of the model inference scheme that could have influenced the model selection result was that, for the 10µM dataset, the level of flexibility of all parameters of all models in space and time was chosen on the basis of the flexibility chosen by WAIC for the diffusion parameter in the diffusion-only model (see section 3.5). It would have been preferable to fit the degree of each polynomial function in space and time separately for each parameter in each model, but this would have been prohibitively computationally expensive. The diffusion model was the only model considered for which I was able to achieve relatively fast (within a few days) convergence of MCMC chains, allowing me to test a range of polynomial degrees. However, since the diffusion-only model is mechanistically very simple, it is likely that in the inference scheme's attempt to get this unrealistic model to capture the features of the data, the flexibility of the diffusion parameter in space and time will have been exploited to compensate. This may have led to the polynomial degrees fitted based on the diffusion model to be higher than those required by the more mechanistically complex models. This could have caused the patterns in the parameters in space and time shown in Figs 3.8-9 being more complex than those that would have been estimated had the polynomial degrees producing those patterns been selected on a model by model basis. This is likely to have been the cause of the receptor saturation effect not being included in the final model. The inclusion of the single constant parameter K_d in the models with receptor saturation, provides behaviour similar to that provided by the flexibility in α in the model without receptor saturation

(Figs 3.8-9A). However, since the polynomials for α haven't been simplified by removing multiple coefficients that are now unnecessary in the models with K_d the receptor saturation term, which is known to be biologically realistic, is not selected for by WAIC. This problem of computational constraints limiting the number of alternative model structures considered may be alleviated by faster inference approaches, such as gradient matching (Xun et al. 2013). However, as discussed above, this approximate inference method comes with its own limitations.

A related effect to that discussed above that could have influenced the estimated patterns of spatial and temporal dependence in the parameters selected for each dataset is a failure to consider all potential mechanisms driving cell movement behaviour. Missing mechanisms could lead to the existing ones trying to compensate by increasing their complexity in space and time. For the 10µM data, for example, I assumed that at a given point in space and time all the cells are either attracting or repulsing one another. It is, however, possible that the cells exhibit both short-range repulsion through a mechanism such as contact inhibition of locomotion (Mayor and Carmona-Fontaine 2010) and attraction through a longer distance mechanism, such as chemical signalling. Combined effects of attraction and repulsion could offer an explanation for the ambiguous results for the time and space dependence in the interaction parameter η. Consideration of models that incorporate these complex interaction behaviours (Mogilner and Edelstein-Keshet 1999) should, therefore be a goal for future work. For the 0µM data, I only considered the simple diffusion model, and failed to address any potential interactions between the cells through attraction, repulsion and overcrowding. This could be the cause of the relatively complex patterns observed in the diffusion coefficient (Fig. 3.5). Given that a small effect of cell interactions was detected for the 10µM data, there is a chance that such an effect would have been found in the 0µM data had it been tested for, and, if I were to analyse these datas again, I would include such a test. The diffusion-only model does provide a good fit to the data, however, with the model producing a cell density distribution that remains with the 95 percentile interval of the distribution estimated from the data (Fig. 3.4). One small feature of the data that is not replicated by the diffusion model is the slight second peak in cell density that forms at 3.5 hours (Fig. 3.4). This feature could have been captured by a model with the interaction term included if η exhibited some relatively complex behaviour around the spatiotemporal location of the second peak, with higher repulsion or attraction occurring on one side of the peak than the other, but only at the end of the time period. There is, however, no obvious biological reason why such behaviour should occur.

In conclusion, I have presented a framework that allows effective Bayesian inference and model comparison for complex PDE models, despite the serious computational costs incurred in solving these models numerically. Like the pseudo-Bayesian approach to inference discussed in chapter 2, this has allowed the identification of mechanisms driving the movement of *Dictyostelium*. There were some changes in the optimal model between this chapter and chapter 2, but these differences can be easily explained as being a consequence of differences in the state of the cells used to produce the data used in each study, and the increased model flexibility that was introduced in this chapter. In both chapters, model selection was able to clearly identify the self-generated gradient mechanism known to be the main movement driver in the experimental system (Tweedy et al. 2016). The fully Bayesian approach discussed in this chapter, however, introduces a key advantage over the previous pseudo-Bayesian approach in its ability to make full use of prior information about the parameter values. It also doesn't rely on any manipulations of the data that could affect our ability to detect cell interactions, and is a more standard approach than bootstrapping, with less need for further validation. In an extension of the modelling framework previously outlined in chapter 2, I have now investigated spatial variation, in additional to temporal variation, in the parameter values, and found both sources of changing movement behaviour to be important in this cellular system. However, high computational costs limited the number of candidate models we were able to consider, and it is acknowledged that this may have led to spurious patterns in the parameters in space and time.

4. Constructing wildebeest density distributions by spatio-temporal smoothing of ordinal categorical data using GAMs

The work presented in this chapter has been published at the following reference:

Ferguson, E.A., Matthiopoulos, J. & Husmeier, D., 2017. Constructing wildebeest density distributions by spatio-temporal smoothing of ordinal categorical data using GAMs. In 32nd International Workshop on Statistical Modelling. Groningen, Netherlands, pp. 70–75. Available at: https://iwsm2017.webhosting.rug.nl/IWSM_2017_V1.pdf

4.1. Introduction

Spatio-temporal smoothing of species distribution data has many potential uses in ecology; for example, to examine changes in home range over time, or to provide a smooth density function that can be used with gradient matching approaches (Xun et al. 2013) to fit advection-diffusion PDE models of animal movement. A wide range of smoothing methods – including kernel density estimation, splines, generalised additive models (GAMs), Gaussian processes, etc. – have been developed in the statistical literature. However, the practicalities and expense involved in collecting species distribution data over large areas in the field can mean that such data are not in a form that most user-friendly implementations of these smoothing methods can readily be applied to. Ordinal categorical data, for example, are common in ecology (Guisan and Harrell 2000), and may be collected when it is infeasible to accurately count all individuals in a population, so that the abundance at each point in space and time is instead estimated as belonging to a broader abundance category. A relatively small number of approaches have been developed for smoothing data of this type, where we need to recover the underlying true density of individuals from the categories. Chu & Ghahramani (2005), for example presented a method for fitting Gaussian processes (also known as kriging) to ordinal categorical data, while Wood et al. (2016) describe ordinal categorical methods for GAMs.

The computational costs of smoothing can rise quickly with the size of the dataset and the number of dimensions in which the smoothing is to be implemented. Large datasets are a particular problem for Gaussian processes (like that of Chu & Ghahramani (2005)), since the computational complexity is cubic in the number of data points (section 19.2 of Barber (2012)). If the data describe a complex pattern in space and time, so that a complex smoother with large numbers of parameters is required to adequately describe this pattern, costs rise even more rapidly. This can again cause problems for ecological data, which may cover large, complex landscapes over long time periods. Methods that allow smoothing of these datasets even when computational resources are limited would therefore be very useful.

In this chapter, I present an application of a GAM-based approach for applying spatio-temporal smoothing to a large ordinal categorical dataset on the distribution of wildebeest in the Serengeti ecosystem of Tanzania and Kenya. I chose a GAM-based approach for two reasons. First, GAMs are a more computationally feasible approach for dealing with large datasets than Gaussian processes. Second, the ordinal categorical GAM method of Wood et al. (2016) has been implemented in the well-documented and user-friendly mgcv package (Wood 2011) in R (R Core Team 2015).

The work I carried out in this chapter was an important pre-requisite for the study presented in chapter 5. The purpose of chapter 5 was to fit advection-diffusion PDE models of wildebeest movement to the data introduced in section 4.2. The fitting approach used was based on gradient matching; a method that bypasses the need for computationally costly numerical PDE solutions by first obtaining a smooth interpolation of the state variable that is described by the data, and then optimising the PDE parameters such that the difference between the partial derivatives of the state variable with respect to time obtained directly from the interpolant and from the PDE (using a given parameter set and information about the partial derivatives with respect to space from the interpolant) is minimised (Xun et al. 2013, Macdonald and Husmeier 2015). The wildebeest density surface in space and time that is the product of this chapter provides the interpolant required for gradient matching in chapter 5.

4.2. Data

The wildebeest distribution data that I applied the smoothing approach developed in section 4.3 to have previously been described and utilised in a number of studies (Maddock 1979, Norton-Griffiths 1979, Boone et al. 2006, Holdo et al. 2009). These data were obtained from aerial surveys of the Serengeti ecosystem between August 1969 and August 1972 on a roughly monthly basis (surveys were carried out on 33 of the 37 months in this period). During these surveys each cell in a grid of 25km^2 cells was estimated as belonging to one of five ordinal wildebeest abundance categories, which are described as containing 0, 1-25, 26-250, 251-2,500 and >2,500 individuals per 25km^2. The original grid was irregularly shaped, covering the area that encompasses the range of the wildebeest migration (Maddock 1979). To simplify the analysis, I worked with the dataset on a 56x46 rectangular grid that was just large enough to contain the original irregular grid, and assumed that any of the cells in this new grid that were not included in the original one contained zero wildebeest. Thus, the dataset involved 2,576 cells making up the rectangular spatial grid, all of which were sampled at 33 time points, resulting in a large dataset with a total 85,008 data points. The entire time series of 33 maps showing the ordinal wildebeest abundance categories can be viewed in supplementary video 4.1 (see Appendix C.1.1), and a subset of three of these maps can be seen in Fig. 1A-C.

4.3. Methods

To smooth the wildebeest distribution data in time t and the two spatial dimensions (x, y), I fitted GAMs (generalised additive models) with a tensor product between these three variables using the mgcv package in R. This tensor product allows for interactions between the three variables, and was composed of cubic regression spline smooths, where overfitting (excessive curvature) was prevented by penalisation of the integral of the squared second derivatives. I used the ordinal categorical GAM method described in Wood et al. (2016), where the linear predictor gives the value of a latent variable, here representing the wildebeest density underlying the ordinal categories. The cut-off points that demarcate the five ordinal categories were specified, and the probability that a point in space and time belongs to a given category equals the probability that the latent variable lies between the corresponding category cut-offs at that point.

In Wood et al. (2016), the latent function can range from $-\infty$ to ∞, but this is unrealistic for the current problem, since wildebeest density has a known minimum of zero and a finite maximum $W_{\max}$. These constraints can be introduced by applying a sigmoidal transformation to the unbounded latent function L after the GAM has been fitted, giving a preliminary estimate of wildebeest density $\hat{W}$ as follows:

$$\hat{W}\left(x,y,t\right)=\frac{W_{\max}}{1+e^{-L(x,y,t)}} \tag{4.1}$$

No transformation needed to be applied to the ordinal categorical data prior to the fitting of the GAM to get L, but it was necessary to apply the following inverse sigmoid transform to the category cut-offs $\mathbf{c}$ that were used to inform the GAM fitting procedure:

$$\hat{\mathbf{c}}=-\log\left(\frac{W_{\max}}{\mathbf{c}}-1\right) \tag{4.2}$$

I estimated $W_{\max}$ by first assuming that the wildebeest densities in the grid cells assigned to the lower four ordinal categories, which had known upper and lower bounds, were equal to the mid-points of those categories. The sum of the densities in these lower category cells for each month was then subtracted from the total number of wildebeest W_T known to be in the region from a population count that took place in 1971, during the time period that the distribution data were collected (Norton-Griffiths 1973). The remaining wildebeest for each month were assumed to be divided evenly between the cells in the highest ordinal category (which was unbounded above) for that month. I took $W_{\max}$ to be the largest wildebeest density estimated for grid cells in the highest abundance category over all months. This led to $W_{\max}=332{,}355$ wildebeest/25km^2, which corresponds to up to 46% of the total population W_T being present in a single grid cell.

Even after applying sensible upper and lower bounds to the latent function, large fluctuations in the area under $\hat{W}$ (which represents the total number of wildebeest in the region) could occur over time, and, at time points where multiple cells were assigned values of $\hat{W}$ that were close to $W_{\max}$, the estimated total number of wildebeest in the region at those times could greatly exceed W_T, sometimes by an order of magnitude. At other time points, the situation was reversed, and the estimated number of wildebeest in the region was much less than W_T. This behaviour was undesirable, since wildebeest numbers are expected to remain relatively stable at W_T over the time period of interest. I therefore considered the normalised wildebeest density $\bar{W}$, where the total number of animals was maintained at W_T by normalising $\hat{W}$ as follows:

$$\bar{W}=\frac{\hat{W}\left(x,y,t\right)W_T}{\iint\hat{W}\left(x,y,t\right)dxdy} \tag{4.3}$$

Due to computational time and memory constraints, a sufficiently flexible GAM could not be fitted to the entire large dataset simultaneously. I therefore divided the time series into three contiguous intervals and fitted a GAM in (x,y,t) to each interval separately. Each GAM had 20 knots in the marginal smooth in each spatial dimension, and a number of knots in the marginal smooth in time that was equal to the number of time points present in the data subset to which the GAM was fitted (12 for the first subset, 11 in the second and 12 in the third). This resulted in the effective degrees of freedom, which are determined by the degree of penalization (selected during fitting) applied to the integral of the squared second derivatives, being considerably lower than the maximum number available, suggesting that the number of knots was sufficient (Wood 2006). The three GAMs were joined together to form a single continuous function by averaging at the link times l_i $\left(i \in \{1,2\}\right)$, and allowing the influence of each GAM on the others to decline smoothly, according to the parameter σ, as distance from the point of joining increased. For a given point $\left(\bar{x},\bar{y},\bar{t}\right)$, therefore, I obtain a final estimate of wildebeest density W by:

$$W\left(\bar{x},\bar{y},\bar{t}\right)=\bar{W}_{GAM\,j}\left(\bar{x},\bar{y},\bar{t}\right)+\sum_{i=1}^{2}a_i \exp\left(\frac{-\left(\bar{t}-l_i\right)^2}{2\sigma^2}\right)m_i\left(\bar{t}\right) \tag{4.4}$$

Here $\bar{W}_{GAM\,j}$ is the normalised wildebeest density obtained from the GAM fitted to time interval j, where:

$$j=\begin{cases}1 \text{ if } \bar{t}\le l_1\\ 2 \text{ if } l_1<\bar{t}\le l_2\\ 3 \text{ if } \bar{t}>l_2\end{cases} \tag{4.5}$$

The a_i are given by:

$$a_i\left(\bar{x},\bar{y},l_i\right)=\frac{\bar{W}_{GAM_i}\left(\bar{x},\bar{y},l_i\right)-\bar{W}_{GAM_{i+1}}\left(\bar{x},\bar{y},l_i\right)}{2} \tag{4.6}$$

and the m_i, which ensure that the adjustments to $\bar{W}$ are made in the correct direction on either side of each link point, are:

$$m_i\left(\bar{t}\right)=\begin{cases}-1 \text{ if } \bar{t}\le l_i\\ \ \ 1 \text{ if } \bar{t}>l_i\end{cases} \tag{4.7}$$

If the influence of the adjoining GAMs declines too slowly with distance from the link points, relative to the rate at which changes occur in $\bar{W}_{GAM_i}$ (i.e. σ is too large), unrealistic negative values of W can occur. I therefore tuned σ by starting with a relatively large value and gradually decreasing it until no negative values of W occurred.

4.4. Results

Application of the method described above to the ordinal categorical wildebeest distribution data, produces a smooth density function in space (Fig. 4.1D-F) that, when categorised into the same ordinal categories as the original data, shows a pattern that resembles that in the original data, but with some added smoothness (compare Fig. 4.1A-C with Fig. 4.1G-I). The resulting function is also smooth throughout the time period, with the exception of at the link points, where it is continuous, but small kinks occur as a consequence of the fact that the procedure described in equation (4.4) to link the three time intervals together forces the GAMs to have the same value at these link points, but not the same derivative. Nevertheless, no visually obvious distortions to the wildebeest density function are observed around the link points; wildebeest density does not appear to change either more slowly or more rapidly around the GAM link times than it does elsewhere in the time period (Fig. 4.2). Supplementary video 4.2 (see Appendix C.1.2) shows the changes in wildebeest density across the spatial region for the full time period from August 1969 to August 1972, as estimated using the GAM-based approach described in this chapter.

4.5. Discussion

In summary I have presented an application of a GAM-based method for recovering realistic density estimates in space and time from coarse ordinal categorical data. The approach is able to reduce the high computational costs of smoothing large datasets in multiple dimensions, by fitting models to subsets of the data and linking them together. Such a method could be very useful in ecology, where deficient data of this type may frequently be collected due to feasibility constraints in large field systems. Conversion of these ordinal categorical data into detailed densities is required for use of many of the analytical techniques that are commonly applied to animal distribution data, for example, to estimate home ranges (Worton 1987).

There are a number of limitations to the methodology described, however. In order to maintain a realistic population size, I found that it was necessary to normalise the estimates of the density function obtained from the GAMs based on the known size of the wildebeest population around the time the data were collected (equation (4.3)). Failure to do so led to a wildebeest population that fluctuated wildly in size over the time period, suggesting that the ordinal categorical GAM method was giving a poor recreation of the wildebeest densities underlying the data. The fluctuations in the estimated wildebeest population are likely to have been a consequence of the coarseness of the data. In binning wildebeest density into such broad categories, substantial losses of information are incurred, and it becomes very challenging for the GAM to retrieve an accurate density. Additionally, gaps of around a month occurred between observations, during which the GAM had little information with which to estimate the size and distribution of the population. The method may be more effective when applied in cases where the abundance categories are narrower, and the data is less sparse. However, it is acknowledged that data of this quality are likely to be infeasible to collect in many ecological systems of interest. A possible alternative to the method applied that would potentially have ameliorated the problem of a fluctuating population size would have been to assume for each time point that the number of wildebeest in each grid cell in one of the lower four abundance categories was equal to the mid-point of the associated category. The remaining animals in the population could then be distributed evenly among cells in the highest abundance category. A GAM with a gamma distributed response could have been fitted to these estimated wildebeest densities. Since each data point is then a defined number, and the GAM is not as free to interpret it as potentially lying at any point within a broad category, there should be a greater tendency for the area under the estimated surface through time to remain close to the known population size. However, this approach would not fully account for the uncertainty present in the original data, and the estimated numbers of wildebeest, particularly for cells in the highest abundance category, may have substantial amounts of error. A second alternative, if we were prepared to give up on retrieving an accurate description of the density distribution, would be to instead consider the functions in space and time describing the probability of being in a particular abundance category. While the exact number of animals at a point in space and time may be difficult to estimate with any accuracy, estimating the most likely

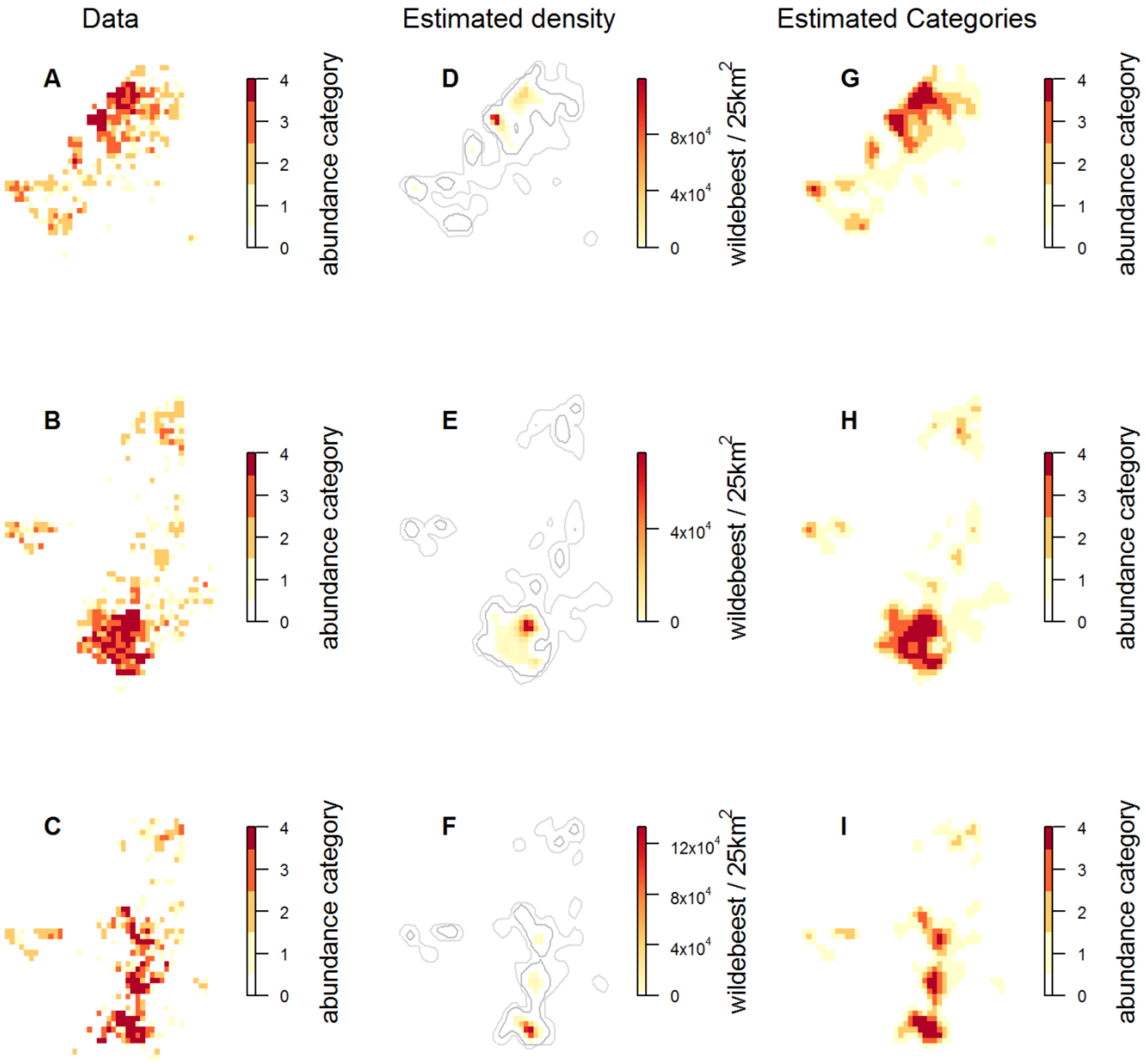

Figure 4.1: GAM-based interpolant fit to the wildebeest distribution data in space at three different time points. A-C) The wildebeest spatial distribution data for months 1, 18 and 35. **D-F)** The smooth wildebeest density distribution estimated in space by the model for months 1, 18 and 35. The two contours indicate the boundaries between abundance categories 0, 1 and 2 (which respectively contain 0, 1-25, and 26-250 wildebeest/25km^2). **G-I)** Estimated wildebeest abundance categories based on **D-F**.

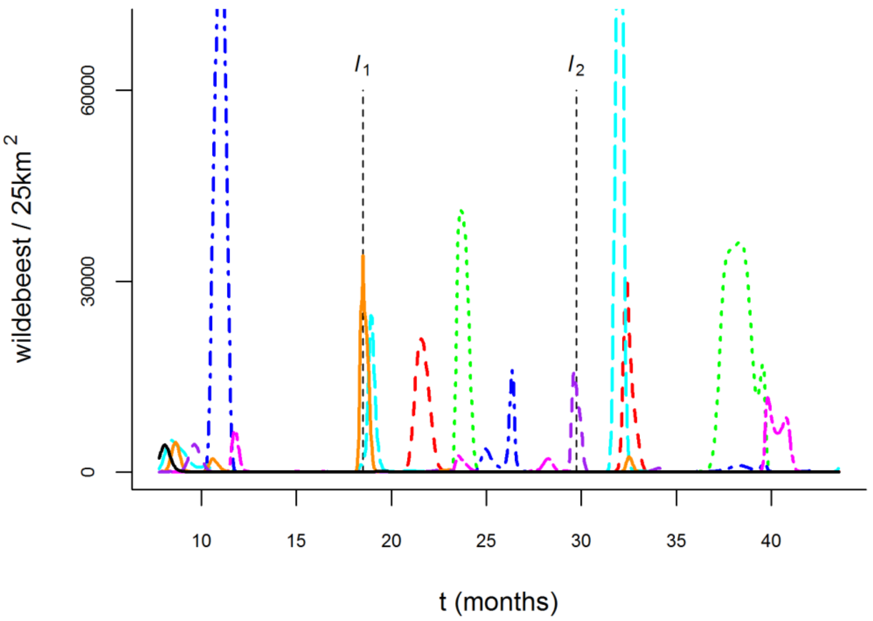

Figure 4.2: Changes in estimated wildebeest density in eight grid cells over the time period of interest. Different cells are indicated by different colours/line types. This particular set of cells was selected haphazardly, with the aim being to include cells that contained non-trivial numbers of wildebeest at at least one point in time, and that experienced these large numbers of wildebeest at a range of time periods. Changing wildebeest numbers in cells at the link times between the GAMs can thus be compared to those at other times. The link times between the three GAMs are indicated by dashed vertical lines.

category for this point may be a less difficult problem. The functions describing the probability of each category are an additional output of the ordinal categorical GAM method in mgcv. Since, however, my aim in this chapter was to obtain a surface from which wildebeest density gradients could be estimated for comparison with those produced by the PDE model of the next chapter, this option would not have been appropriate.

A second possible issue with the method is that it simply smooths the data without any knowledge of the rate at which the species of interest can actually move. In the density surfaces output by the final model, it can be seen that on a small number of occasions spikes in the density of wildebeest occur rapidly in an area, seemingly out of nowhere (supplementary video 4.2, Appendix C.1.2), suggesting that the GAM is rapidly drawing density from more distant areas. This may or may not be a realistic description of the way these animals move; wildebeest move an average of 4.25km per day, but are capable of rapidly moving much longer distances, with daily movements of up to 58km being recorded (Hopcraft 2010). Regardless, further development of the method to account for the maximum speed of movement of the focal species would be desirable.

As mentioned in section 4.4 above, the density function in space and time produced by equation (4.4) is smooth throughout space and time, except at the link times between the three GAMs used to produce it. The lack of smoothness at the link times could be resolved by forcing the GAMs to have the same derivative, in addition to the same value, at the link points. This is similar to the approach used to produce splines by joining together polynomials at knots, where the functions must have the same value and derivatives (see section 5.2 of Hastie et al. (2009)). Binding together the derivatives at the link points is necessary if the density function is to be used to calculate analytical values of the derivatives of density. However, this is unnecessary if, as is the case here, the derivatives are to be estimated using finite difference approximation (see chapter 5, section 5.5).

Finally, I note that the value estimated for $W_{\max}$ may have been a little conservative. While it allows for 46% of the total population (332,355 individuals) to occur within a single cell, which is reasonable, it is likely that the wildebeest density at certain points within the grid cell will have densities higher than this value, which allows 70m^2 for each individual. An alternative way of estimating a less restrictive value for $W_{\max}$ would have been to determine the minimum amount of space taken up by an animal based on average body size. It should be noted that the normalisation step used to maintain the population at the correct size meant that wildebeest densities had the potential to rise above $W_{\max}$. However, values in excess of $W_{\max}$ were not observed in practice (see supplementary video 4.2).

In the next chapter, I implement a gradient matching approach (Xun et al. 2013, Macdonald and Husmeier 2015) to fit advection-diffusion PDE models of wildebeest movement (similar to those I previously used to describe movement in cellular systems (chapters 2-3)). This inference approach makes use of wildebeest density distributions obtained from the ordinal categorical distribution data using the methods described in this chapter.

5. Inference of the mechanisms driving the Serengeti wildebeest migration

5.1. Introduction

The annual wildebeest migration in the Serengeti ecosystem involves the movement of around 1.2 million individuals, each of which covers an average of 1,550km/year (Hopcraft et al. 2015) as they move between their wet season range on the south-eastern short-grass plains and their dry season range in the woodlands and savannah to the west and north of the region (Maddock 1979, Thirgood et al. 2004, Boone et al. 2006). This mass movement of animals is an important driver of the dynamics of the entire ecosystem. Changes in the size of the wildebeest population lead to changes in grazing pressure, which have previously had impacts on plant composition, the frequency of fires (due to changes in dry grass biomass), and the abundance of other herbivore species (Sinclair 1979). At a time when migrations of large ungulates are collapsing globally, mainly as a result of human activity (Bolger et al. 2008, Harris et al. 2009), understanding the movement of these ecosystem engineers, so that risks to the continuance of the migration can be managed, may be critical for preserving the ecosystem as a whole.

Many studies have suggested potential environmental drivers of the movement of the Serengeti wildebeest. There is wide consensus (McNaughton 1979, Boone et al. 2006, Holdo et al. 2009) that movement to the north and west of the region in the dry season is primarily a result of the spatial gradient in rainfall, which declines from the north-west to the south-east (Fig. 5.1A) and alters the quality and abundance of forage and the availability of water. The low annual rainfall and shallow soil horizon on the south-eastern plains leads to rapidly deteriorating grazing conditions at the end of the wet season, such that green grass production stops (McNaughton 1979) and water quality declines (Wolanski and Gereta 2001). This forces wildebeest to move on to the wetter western and northern areas, where low quality green forage is available through the dry season. The reasons for the movement back to the south-eastern plains when the wet season returns are less obvious, since the dry season range consistently has greater higher rainfall and, therefore, grass biomass throughout the year (Holdo et al. 2009). The current hypothesis hinges on differences in the grass quality between the north-west and south-east. Wilmshurst et al. (1999), for example, suggested that wildebeest distribute themselves on the plains during the wet season as part of an energy maximisation strategy, since the grasses are shorter, greener and more digestible than the taller, more mature grasses that dominate the dry season range. Since tall grasses typically have a lower nitrogen concentration than shorter ones, an increasing gradient in plant nitrogen concentration runs from north-west to south-east, in the opposite direction to the rainfall gradient (Fig. 5.1B). Previous modelling work has identified this nitrogen gradient as a driver of the wet season wildebeest distribution (Holdo et al. 2009). Murray (1995) additionally found that concentrations of both calcium and phosphorus (two elements that are important for lactating females) in grass were higher in the wet season range than the dry season range. In particular, concentrations of phosphorus on the dry season range were insufficient to support lactation, potentially explaining the movement to the more phosphorus-rich plains.

Two pieces of evidence suggest that gradients in grass availability that are produced (self-generated) by the wildebeest themselves through local depletion are important in determining wildebeest movement patterns. The first is that individual wildebeest move further each day during the wet season, when they are on the southern plains and grass is at its peak quality and abundance, than they do at any other time (Hopcraft et al. 2014). This was an unexpected result (animal movement is typically expected to become slower and more tortuous in high quality foraging areas; see, for example, Morales et al. (2004)), and may be a consequence of density dependence; the high densities of wildebeest (and other grazers) on the southern plains mean that local grass resources are rapidly depleted and the animals are forced to keep moving. The second piece of evidence for self-generated resource gradients in wildebeest is the observation that when rinderpest reduced wildebeest numbers in the Serengeti to ~15% of their current level, the annual migration was shorter in length. The subsequent increase in migration distance was probably a result of the recovering wildebeest numbers causing greater depletion of resources, forcing the herds to move further to maintain year-round access to sufficient forage (Thirgood et al. 2004, Harris et al. 2009, Hopcraft et al. 2015).

Given that direct interactions between individuals have been found to be important in driving the movement behaviour of other ungulate species, including reindeer (Langrock et al. 2014) and elk (Haydon et al. 2008), we might expect that such social interactions would also be influential in modifying the movement patterns of the Serengeti wildebeest. However, studies that include the effects of intraspecific social interactions on wildebeest movement are limited. Gueron and Levin (1993) developed a theoretical model describing how the wave-like patterns of dense wildebeest migration fronts might develop through the effects of neighbours on the movement speed of leading individuals. However, since this model only examined the behaviour of a subset of individuals in a specific type of herd formation, and was not formally fitted to data, the insights it offers for the migration as a whole are limited. There is some evidence for social interactions influencing the distribution of wildebeest in other regions. In Amboseli, Kenya, an effect of aggregation on wildebeest distribution was found in the dry season, but not the wet season (Mose et al. 2013), while in Karongwe Game Reserve, South Africa, wildebeest were found to form larger groups when in areas of open scrub where the probability of encountering lions was greatest (Thaker et al. 2010).

A number of previous wildebeest movement models have been fitted to data from the Serengeti ecosystem in an attempt to infer movement drivers (Boone et al. 2006, Holdo et al. 2009, Hopcraft et al. 2014). However, none of these have included all three types of movement driver described above (gradients in environmental covariates, environmental depletion through grazing, and interactions between individuals). In this chapter, I apply a model with all of these components, based on the PDE model framework introduced in Chapter 2, to data on the distribution of the Serengeti wildebeest population. This required a number of extensions to the methodology that I previously applied to cell movement. First, unlike in the simple experimental cell systems, movement had to be modelled in 2D (rather than 1D) space. Second, because of the far greater computational expense of numerically solving the PDE model in 2D, and the issues encountered with instability in the numerical solutions, I implemented a new method, known as gradient matching (Xun et al. 2013, Macdonald and Husmeier 2015), to fit the movement models to the data. This method removes the need to numerically solve the PDEs during model fitting. To further reduce computational costs, I used gradient matching within a frequentist parameter optimisation setting, rather than within the pseudo-Bayesian and Bayesian approaches developed in chapters 2-3. Finally, I also extended the models to account for the ability of wildebeest to sense their wider environment using visual, auditory and olfactory cues, giving them a greater range of perception than cells. By carrying out model selection over a set of candidate wildebeest movement models, containing different combinations of movement drivers, I aimed to draw conclusions about the drivers of wildebeest movement in the Serengeti ecosystem.

5.2. Data

Here, I used the same dataset describing the distribution of the Serengeti wildebeest population that was introduced and described in detail in the previous chapter (section 4.2; see also supplementary video 4.1 and its description in Appendix C.1). These data – on a 46x56 spatial grid of 25km^2 cells, where each cell was assigned to one of five ordinal abundance categories – were collected, roughly monthly, at 33 time points between August 1969 and August 1972 (Norton-Griffiths 1973, Maddock 1979). Like other recent studies (Boone et al. 2006, Holdo et al. 2009), I also use this relatively old dataset, because there are no more recent data on the distribution of the entire population at multiple time stages of the annual migration. The current number of wildebeest in the ecosystem is approximately double the 1971 population estimate of 720,769 animals (Norton-Griffiths 1973, Hopcraft et al. 2015). This increase in wildebeest density over time may be responsible for changes in the annual migration route, such as the herd's tendency to migrate further north in the dry season (Thirgood et al. 2004, Harris et al. 2009, Hopcraft et al. 2015). However, the behavioural mechanisms and parameters underlying these migration routes (such as the rate of grass consumption, or the strength of conspecific interactions) are unlikely to have changed over the last 50 years (evolutionary changes in behaviour are likely to be negligible during this time period). Hence, conclusions drawn from these data should still be applicable to the larger present-day Serengeti wildebeest population.

I also use datasets on three environmental variables – rainfall, grass nitrogen concentration and tree canopy cover – to try to explain the wildebeest movement behaviour. Rasters of monthly rainfall, produced by Holdo et al. (2009), from rain gauge data from the region were available for the period from January 1969 to December 1972. A raster of grass nitrogen concentration was created by Hopcraft et al. (2014), who applied kriging to data obtained from 148 sites across the region between 2006 and 2008, using the mean NDVI (Normalized Difference Vegetation Index; a measure of vegetation greenness) at the sites as an explanatory variable. Note that these nitrogen data were collected ~35 years after the collection of the wildebeest data, and it is possible nitrogen concentrations may have changed over this period, particularly since the wildebeest population roughly doubled (Hopcraft et al. 2015). The associated increase in grazing could have affected the typical age and thus nitrogen content of the grasses. However, while precise values of nitrogen concentration may have changed, it has previously been noted that broad spatial patterns in nitrogen concentration in the region have remained constant over long periods of time (Holdo et al. 2009). I produced a raster of tree canopy cover by ordinary kriging of information from different sources (Norton-Griffiths 1979; Reed et al. 2009; Frankfurt Zoological Society and Harvey Maps 2010) ; see Appendix D.1 for details. These environmental variables are illustrated in Fig. 5.1.

5.3. Grass dynamics model

Detailed data on grass abundance in the Serengeti region were not available for the time period of interest, so to include a response of the wildebeest to grass availability in the movement models to be considered (see section 5.4), it was necessary to simulate this environmental variable using a model of grass dynamics. Here, I used the grass model outlined in Holdo et al. (2009), which is a two-compartment ordinary differential equation (ODE) model, describing changes in the densities (in g/m) of green and dry grass (denoted G and D respectively) in the proportion of the local area that is available to grass growth as follows:

$$\frac{dG}{dt} = \psi R_{day}\left(G+\sigma\right)\left(1-\frac{G+\rho D}{K_G}\right) - \delta_G G - \frac{I_G W}{(1-T)} \tag{5.1}$$

$$\frac{dD}{dt} = \delta_G fG - \delta_D D - \frac{I_D W}{\left(1-T\right)} \tag{5.2}$$

The first term of equation (5.13) describes green grass growth, which increases in response to daily rainfall R_{day}, according to rate parameter ψ. The parameter σ ensures that grass density does not recover unrealistically slowly if it drops to a value near zero. Grass growth is increasingly limited as the abundance of green grass (plus a shading effect of dry grass on green grass, moderated by parameter ρ) approaches the green grass carrying capacity K_G, given by:

$$K_G = \mu_0 + \mu_1 R_{ann} \tag{5.3}$$

where R_{ann} is the mean annual rainfall (Fig. 5.1A). Green grass decay occurs at rate δ_G and consumption of green grass by wildebeest W occurs at rate

$$I_G = \min\left(MVI_G, \frac{\alpha_w G}{\beta_w + G + D}\right) \tag{5.4}$$

where MVI_G is the maximum daily voluntary intake of green grass, α_w is the maximum rate at which wildebeest can crop grass, and β_w is the grass abundance at which the intake rate reaches 50% of its maximum value. Biologically, I_G is the maximum intake rate of green grass, which is either a function of the cropping rate and grass availability, thereby accounting for hungry animals eating as much as they can, or the maximum daily voluntary intake rate, accounting for satiation. The amount of grass consumed by wildebeest based on I_G in equation (5.1) is divided by the proportion of the immediate area that contains grass, $(1-T)$. This leads to the impact of wildebeest being greatest in cells that contain a low proportion of grass, based on the assumption that the animals in that cell will be focussed on the grassed portion of the cell.

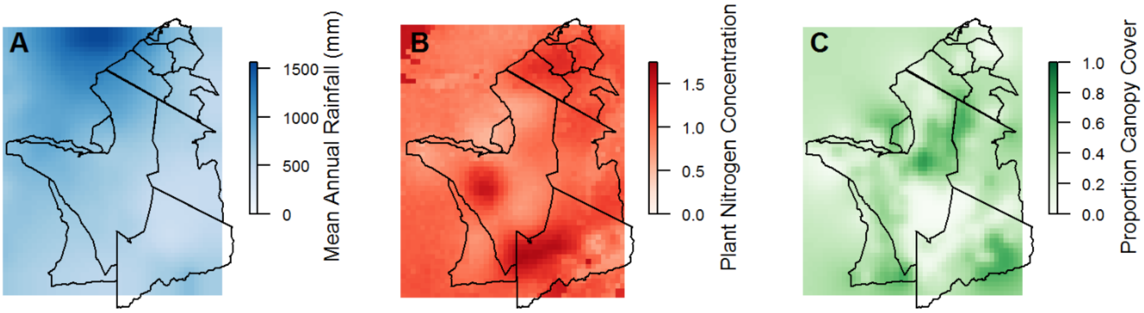

Figure 5.1: Maps of environmental variables. A) Mean annual rainfall across the Serengeti ecosystem over the years 1969-1972 (calculated from data monthly rainfall maps; Holdo et al. (2009)). **B)** Plant nitrogen concentration across the Serengeti ecosystem from data collected in the period 2006-2008 (Hopcraft et al. 2014). **C)** Proportion of tree canopy cover across the Serengeti ecosystem based on data from Norton-Griffiths (1979) and Frankfurt Zoological Society and Harvey Maps (2010) (see Appendix D.1).

In equation (5.2), a fraction f of decaying green grass becomes dry grass. Dry grass also decays at rate δ_D, and wildebeest consume dry grass at the rate

$$I_D = \min\left(MVI_D, \frac{\alpha_w D}{\beta_w + G + D} \right) \tag{5.5}$$

where MVI_D is the maximum voluntary intake of dry grass. The amount of dry grass consumed by wildebeest is again divided by the proportion of the cell that is grass.

To obtain values of G over the spatial region and time period of interest with which to inform the wildebeest movement model, I numerically solved the grass dynamics model (equations (5.1-5)) for each point of interest in space, using the lsodes integrator from the R package deSolve (Soetaert et al. 2010). As I had no data on G and D with which to initialise the model during the numerical integration, I instead initialised the system with zero grass in January 1967 and allowed G and D to develop towards realistic distributions prior to the point where the first wildebeest distribution data were collected in August 1969, more than 2.5 years later. The numerical integration was then continued until August 1972, which was the end of the time period covered by the wildebeest distribution data. I obtained values for all parameters in the grass model (see Table 5.1 for a summary) from Holdo et al. (2009), who developed this model and gathered these parameter values from the literature.

To integrate the grass dynamics model I required information on two covariates; rainfall and wildebeest density. As noted in section 5.2, maps of monthly rainfall were available from January 1969 to December 1972. R_{ann}, the mean annual rainfall, was calculated from these data to give the map shown in Fig. 5.1A. R_{day} was calculated for a given time point by taking the monthly rainfall associated with that time point, and dividing by 30. For each month of the year in the period of January 1967 to January 1969, where I did not have rainfall data, I took the monthly rainfall to be the average rainfall for that month of the year during the four years for which data were available. Wildebeest abundances for the period August 1969 to August 1972 were obtained from GAMs (Generalised Additive Models) fitted to the ordinal categorical wildebeest distribution data (see section 5.5). For each month of the year in the period prior to August 1969, a wildebeest abundance map was obtained by averaging daily estimates from the GAM for the same month in the three subsequent years.

The changing abundances of green and dry grass estimated over the spatial region by numerical integration of equations (5.1-5) under the conditions outlined above, and using the least complex of the GAMs considered (section 5.5, Table 5.2, Fig.5.4F) to provide the wildebeest abundance estimates, are illustrated in Fig. 5.2 and supplementary video 5.1 (see Appendix D.2.1 for video description). I observed that grass abundances outside the protected areas (shown by the black outlines in Fig. 5.1), which are the areas most used by the wildebeest, were sometimes higher than anticipated, particularly in the north-east of the region. These levels of grass abundance outside the protected areas are likely to be unrealistic, because the grass model does not account for the livestock grazing and other human-related activities on which I lack data but expect will be reducing grass availability in these areas. Since any grass outside of the protected areas is largely inaccessible to the wildebeest, I prevent it from having an unrealistic impact on wildebeest movement in the models by assuming that the grass abundances outside the area encompassing the range of the wildebeest migration (Maddock 1979) are zero (see Fig. 5.2G-L and supplementary video 5.2 (video description can be found in Appendix D.2.2)). To also prevent the wildebeest being unrealistically driven out of their normal range by inaccessible areas of high plant nitrogen concentration, I additionally set plant nitrogen in these outer areas to zero when using this covariate (shown in Fig. 5.1B) in the models. The values of green grass density G estimated by the grass model were converted into green grass intakes I_G using equation (5.4) before being used in those wildebeest models that included a response to I_G.

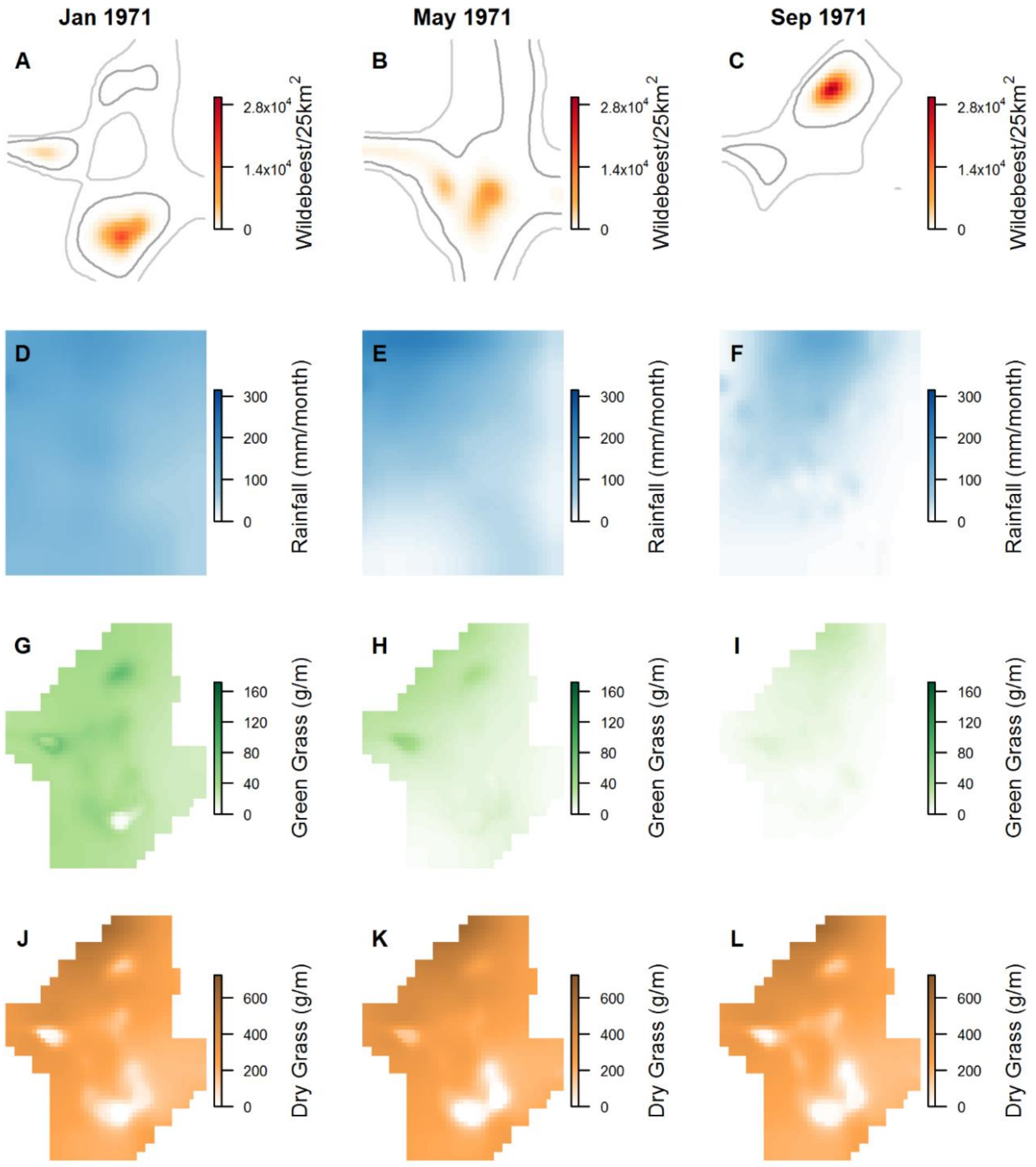

Figure 5.2: Simulated grass biomass. Green and dry grass abundance outputs, alongside wildebeest and rainfall inputs, from the grass dynamics model (section 5.3) across the spatial region at 3 time points. **A-C)** Wildebeest density estimated from the least complex GAM fitted to the ordinal categorical distribution data (section 5.5). **D-F)** Monthly rainfall. **G-I)** Green grass abundance estimated from the grass model. **J-L)** Dry grass abundance estimated from the grass model. Note that the abundances of green and dry grass outside the area encompassing the range of the wildebeest migration have been set to zero. A video of the changing grass abundances over the full time period of interest can be observed in Supplementary video 5.2 (Appendix D.2.2).

Table 5.1: Summary of model parameters. Values of the fixed parameters were taken from Holdo et al. (2009).

Parameter	Description	Inferred or fixed?	Time-varying?
D_W	wildebeest diffusion coefficient (equation (5.6))	Inferred	Yes
η	strength of wildebeest movement response to gradient in green grass intake rate I_G (or green grass abundance G) (equations (5.7-8,5.16-17)	Inferred	Yes
ε	strength of wildebeest movement response to gradient in grass nitrogen concentration N (equations (5.7-8,5.16-17)	Inferred	Yes
γ	strength of wildebeest movement response to gradient in conspecific density W (equations (5.7-8,5.16-17)	Inferred	Yes
$W_{\max}$	maximum conspecific density tolerated by wildebeest (equations (5.7-8,5.16-17)	Inferred	Yes
r	wildebeest range of perception (only for non-local models) (equation (5.9))	Inferred	No
ψ	strength of rainfall effect on grass growth rate (equation 5.1)	Fixed	No
σ	prevents unrealistically slow grass regrowth from near zero values (equation (5.1))	Fixed	No
ρ	effect of shading by dry grass D on the growth of green grass G (equation (5.1))	Fixed	No
δ_G	decay rate of G (equation (5.1))	Fixed	No
δ_G	decay rate of D (equation (5.2))	Fixed	No
f	fraction of decaying G that becomes D (equation (5.2))	Fixed	No
μ_0	intercept of linear model describing effect of annual rainfall on grass carrying capacity (equation (5.3))	Fixed	No
μ_1	slope of linear model describing effect of annual rainfall on grass carrying capacity (equation (5.3))	Fixed	No
α_w	maximum rate at which wildebeest can crop grass (equations (5.4-5))	Fixed	No
β_w	grass abundance at which wildebeest intake rate is 50% of its maximum (equations (5.4-5))	Fixed	No
MVI_G	maximum daily voluntary intake of green grass by wildebeest (equation (5.4))	Fixed	No
MVI_D	maximum daily voluntary intake of dry grass by wildebeest (equation (5.5))	Fixed	No

5.4. Wildebeest movement models

I considered advection-diffusion partial differential equation (PDE) models of wildebeest movement of the form:

$$\frac{\partial W}{\partial t} = \underbrace{-\frac{\partial}{\partial x}\left\{a_x W\right\} - \frac{\partial}{\partial y}\left\{a_y W\right\}}_{\text{advection}} + \underbrace{\frac{\partial}{\partial x}\left\{D_W \frac{\partial W}{\partial x}\right\} + \frac{\partial}{\partial y}\left\{D_W \frac{\partial W}{\partial y}\right\}}_{\text{diffusion}} \tag{5.6}$$

which describe spatio-temporal changes in wildebeest density W. Note that this is similar to the general model form used to describe cell movement in chapters 2-3 (equation (2.1)). The major change between this model and the previous, cell-based one is that movement is now being modelled in two-dimensional rather than one-dimensional space. As a result, there are now two advection terms and two diffusion terms, describing movements along each of the two spatial axes, denoted x and y. I did not include a reaction term like that in equation (2.1) to describe changes in wildebeest density resulting from births and deaths, instead assuming (as in the previous chapter) that the population size remained constant at 720,769 individuals, the population size estimated in 1971 (Norton-Griffiths 1973). The decision not to include a reaction term was made primarily because the coarse ordinal categorical data analysed here did not provide accurate enough information on the population size over time to estimate a rate of population change (as was possible in our melanoma analysis (Appendix A.1.3)). While there is evidence that the population was increasing during the three-year period when the wildebeest data were being collected (Hopcraft et al. 2015), it is not anticipated that the population increase over just three years would have had a major impact on the migratory patterns observed; even at the maximum growth rate observed for this population (~10% per annum (Mduma et al. 1999)), wildebeest numbers at the end of the time period of interest would still only be at ~60% of the current population size.

In equation (5.6), I assumed a wildebeest diffusion coefficient D_W that is constant over x and y, and does not depend on any environmental variables. I considered the following functions for the advection coefficients in x and y, denoted a_x and a_y:

$$a_x = \left(1 - \frac{W}{W_{\max}}\right)\left(\eta \frac{\partial\left(G(1-T)\right)}{\partial x} + \varepsilon \frac{\partial N}{\partial x} + \gamma \frac{\partial W}{\partial x}\right) \tag{5.7}$$

$$a_y = \left(1 - \frac{W}{W_{\max}}\right)\left(\eta \frac{\partial\left(G(1-T)\right)}{\partial y} + \varepsilon \frac{\partial N}{\partial y} + \gamma \frac{\partial W}{\partial y}\right) \tag{5.8}$$

where G is the density of green grass (in g/m) in the proportion of the spatial location that is available to grass (which develops as previously described in section 5.3), T is the proportion tree canopy cover (Fig. 5.1C), and N is the grass nitrogen concentration (Fig. 5.1B). These advection coefficients describe preferential movement of the animals up the gradients with respect to x and y in G multiplied by the proportion of the immediate area that is grass, assumed to be $(1-T)$. It is anticipated that wildebeest will focus their grazing effort on the proportion of the local area that is grass (as described in section 5.3 and equation (5.1)), so that their intake rate is not impeded by the presence of tree canopy cover. Still, it seems reasonable to assume that if a wildebeest were offered two locations, both with the same value of G, but one of which is 100% grass, and the other only 50% grass, the location with 100% grass should be twice as attractive, hence why G is multiplied by $(1-T)$. The speed that the wildebeest move in response to this gradient in $G(1-T)$ is mediated by the parameter $\eta > 0$. It is assumed that the wildebeest are responding just to green grass, and not both green and dry grass, since evidence from previous studies suggests that these animals follow an energy maximisation strategy by focussing on the more nutritious young, green grass (Wilmshurst et al. 1999, Boone et al. 2006). We may alternatively want to consider that the wildebeest are moving in response to their green grass intake rate, which takes account of digestive and food handling constraints, and the additional presence of dry grass, as modelled in equation (5.16), rather than simply responding to green grass density. Therefore, I also tested versions of these advection coefficients where G was replaced by I_G, to determine whether the animals would be content to remain stationary once their intake had been maximised, or whether they would continue to seek out areas of higher grass abundance, despite there being no immediate benefit to doing so. Movement of wildebeest towards regions of higher grass nitrogen concentration N is also incorporated into the model, with the strength of the response to the nitrogen gradients being described by the parameter $\varepsilon > 0$. I incorporate the effects of conspecifics on movement in two ways, the first being through movement in response to the conspecific gradient (mediated by parameter γ), and the second being an overcrowding effect, which is identical to that used in the cell movement models (equation (2.7)) and reduces advection along both axes to zero as the wildebeest density approaches the parameter $W_{\max}$; the maximum conspecific density that the animals will tolerate. See Table 5.1 for a summary of the model parameters.

One potential problem with using the basic advection-diffusion PDE outlined above to describe wildebeest movement is that it assumes that the animals move based only on local information about the exact point in space at which they are currently located – they cannot perceive and respond to non-local environmental conditions associated with positions at a distance from themselves. While this assumption is justifiable for cell behaviour, where detection of movement driving chemicals in the environment is typically achieved only via cell surface receptors, this may not be the case for wildebeest, which have superior methods of sensing their wider environment using visual, auditory and olfactory cues. I, therefore, also considered non-local versions of the advection coefficients in equations (5.7-8), which were derived by first defining a function for the perceptive field $P(x)$, which declines with distance from the current location:

$$P(x,y)=\begin{cases} r-\sqrt{x^2+y^2}, & \text{if } x^2+y^2<r^2 \\ 0, & \text{if } x^2+y^2\geq r^2 \end{cases} \tag{5.9}$$

where r is the wildebeest radius of perception. This function is depicted in Fig.5.3A, which shows wildebeest perception at its highest at the animal's location $(0,0)$, and then linearly declining with increasing distance from this point until it hits zero at a distance of r. I chose to use this function rather than (say) a smooth bivariate Gaussian, both for computational efficiency and to allow us to estimate the radius of perception r, a parameter that is perhaps more intuitive than the standard deviation of perception that would be estimated for the Gaussian function. In practice, using a Gaussian function would have made little difference to the qualitative results (Mogilner and Edelstein-Keshet 1999). The next step was to take the convolutions of $P(x)$ with the partial derivatives of the variables driving movement (i.e. $I_G(1-T)$, N and W) with respect to x and y in equations (5.7-8) as follows (taking $\partial W/\partial x$ as an example):

$$\begin{aligned}\left(\frac{\partial W}{\partial x}*P\right)(x,y,t)&=\iint_{\mathbb{R}^2}\frac{\partial W}{\partial x}(x',y',t)P(x-x',y-y')dx'dy'\\&=\lim_{b\to\infty}\left(\iint_A\frac{\partial W(x',y',t)}{\partial x}P(x-x',y-y')dA\right)\end{aligned}\qquad(5.10)$$

where A is the circular area $(x'-x)^2+(y'-y)^2<b$, with radius b and centre (x,y) at time t. By integration by parts:

$$\left(\frac{\partial W}{\partial x}*P\right)(x,y,t)=\lim_{b\to\infty}\left(\int_B W(x',y',t)P(x-x',y-y')\hat{v}dB-\iint_A W(x',y',t)\frac{\partial P(x-x',y-y')}{\partial x}dA\right)\qquad(5.11)$$

where B is $(x'-x)^2+(y'-y)^2=b$, the boundary of A, and $\hat{v}$ is the outward unit surface normal to B. I then obtained:

$$\left(\frac{\partial W}{\partial x}*P\right)(x,y,t)=-\iint_C W(x',y',t)\frac{\partial P(x-x',y-y')}{\partial x}dC\qquad(5.12)$$

where C is the circle centred on (x,y) at t with radius r, since r is finite and $P(x-x',y-y')=0$ for all $(x-x',y-y')$ where $(x-x')^2+(y-y')^2\geq \mathrm{r}^2$ (see equation (5.9)). The convolution of $\partial W/\partial y$ and P could similarly be found to be:

$$\left(\frac{\partial W}{\partial y}*P\right)(x,y,t)=-\iint_C W(x',y',t)\frac{\partial P(x-x',y-y')}{\partial y}dC\qquad(5.13)$$

and the partial derivatives of P with respect to x and y (illustrated in Fig. 5.3B-C) are:

$$\frac{\partial P(x,y)}{\partial x}=\begin{cases}-\dfrac{x}{\sqrt{x^2+y^2}}, & \text{if } x^2+y^2<r^2\\ 0, & \text{if } x^2+y^2>r^2\end{cases}\qquad(5.14)$$

$$\frac{\partial P(x,y)}{\partial y}=\begin{cases}-\dfrac{y}{\sqrt{x^2+y^2}}, & \text{if } x^2+y^2<r^2\\ 0, & \text{if } x^2+y^2>r^2\end{cases}\qquad(5.15)$$

I could then replace the gradients driving movement in equations (5.7-8) with the convolution integrals obtained as described in equations (5.12-13), with equations (5.14-15) substituted in, to give the advection coefficients of the non-local model:

$$\begin{aligned}a_x=\left(1-\frac{W(x,y,t)}{W_{\max}}\right)\Bigg(&\eta\iint_C G(x',y',t)\big(1-T(x',y')\big)\frac{x'-x}{\sqrt{(x-x')^2+(y-y')^2}}dC\\&+\varepsilon\iint_C N(x',y')\frac{x'-x}{\sqrt{(x-x')^2+(y-y')^2}}dC+\gamma\iint_C W(x',y',t)\frac{x'-x}{\sqrt{(x-x')^2+(y-y')^2}}dC\Bigg)\end{aligned}\qquad(5.16)$$

$$a_y = \left(1 - \frac{W(x,y,t)}{W_{\max}}\right)\left(\eta \iint_C G(x',y',t)\left(1-T(x',y')\right)\frac{y'-y}{\sqrt{(x-x')^2+(y-y')^2}}\,dC \right. \\ \left. + \varepsilon \iint_C N(x',y')\frac{y'-y}{\sqrt{(x-x')^2+(y-y')^2}}\,dC + \gamma \iint_C W(x',y',t)\frac{y'-y}{\sqrt{(x-x')^2+(y-y')^2}}\,dC\right) \tag{5.17}$$

The consequence of these new advection coefficients is that, when making movement decisions in response to one of the three movement drivers $G(1-T)$, N and W, the animals now consider the values of these variables over their entire range of perception r. Note that this non-local model reverts to the local model in the case where $r=0$ and P becomes the δ function. Similar non-local models have previously been proposed for modelling swarm behaviour based on various types of social interactions (Mogilner and Edelstein-Keshet 1999, Topaz and Bertozzi 2004, Miller et al. 2012).

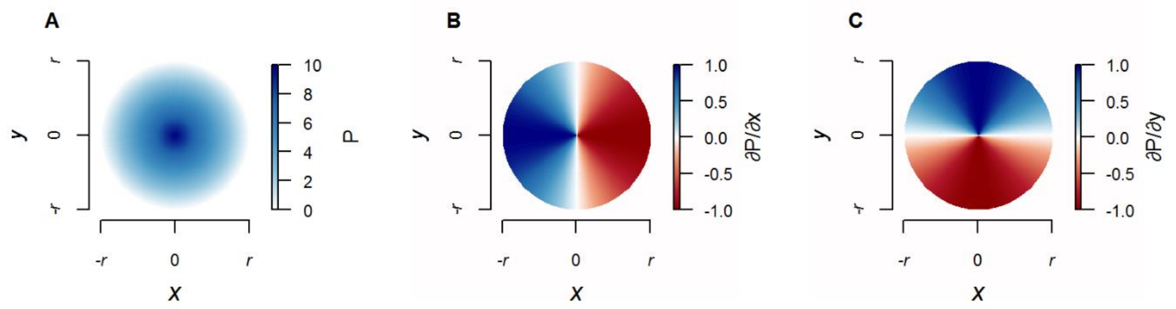

Figure 5.3: Wildebeest perceptive field. A) Depiction of the perceptive field of a wildebeest P, as defined in equation (5.9). **B)** Partial derivative of P with respect to x (equation (5.14)). **C)** Partial derivative of P with respect to y (equation (5.15)).

In addition to the two models described by the advection coefficients in equations (5.7-8) and (5.14-15), I investigated models where the green grass density G in these advection coefficients had been replaced by the green grass intake rate I_G. During model selection, I also considered models with advection coefficients that were nested within equations (5.7-8) and (5.16-17) by removing the effect of $G(1-T)$ (or $I_G(1-T)$), N, W, or $W_{\max}$ on wildebeest movement.

When fitting advection-diffusion models to the cellular movement datasets, I found that the movement parameters had to be time-varying in order to obtain a good model fit (see section 2.6). Here, I tested whether the same was true in the wildebeest system by assessing versions of the models described above where the parameters η, ε, γ, $W_{\max}$ and D_W were constant in time, and versions where these parameters were time-varying. I assumed that the radius of perception r was constant, since it describes the sensory capabilities of the animals, which are not anticipated to change seasonally. Since I did not have any *a priori* reason to assume any particular functional form for the time-variance of these parameters, and initial attempts to fit the parameters as simple polynomial functions of time (as was done for the cell behaviour parameters in chapter 2) failed to effectively describe the observed wildebeest distribution patterns, I fitted values of these parameters separately to each of the 33 time points present in the wildebeest dataset. Model fitting and comparison methods are described in section 5.5.

5.5. Model Inference

In Chapters 2-3, I carried out model inference by numerically integrating the PDE models many times with different parameter values within an optimisation or MCMC algorithm. These numerical PDE solutions were computationally costly even for the 1D cell movement models, but for the 2D wildebeest movement models described in section 5.4, the costs of numerically integrating on a large grid are even greater, particularly for the non-local models that require various integrals to be calculated (equations (5.16-17)). Gradient matching is an inference approach that has been used for both ODEs (Macdonald and Husmeier 2015) and PDEs (Xun et al. 2013) in order to bypass the need for expensive numerical solutions. It is a two-step process that involves first obtaining a smooth interpolation of the state variable (in this study, wildebeest density) in time and space using the data, and then optimising the PDE parameters such that the difference between the partial derivatives of the state variable with respect to time obtained directly from the interpolant and from the PDE (using a given parameter set and information about the partial derivatives with respect to space from the interpolant) is minimised.

For the interpolation step, I used the method described in chapter 4 to obtain continuous wildebeest density surfaces in space and time from the ordinal categorical wildebeest distribution data. This involved splitting the large dataset into three contiguous time intervals, fitting a GAM to each, and then connecting the three GAMs together at the time points where they overlapped (the ‘link points’). The three GAMs I fitted in chapter 4 included a tensor product between time and the two spatial dimensions, using cubic regression splines with 20 knots in each of the x and y marginal bases, and either 12 or 11 knots (equal to the number of time points present in the data subset to which the GAM was being fitted) in t. In this chapter, I used the method described in chapter 4 to produce further sets of linked GAMs in which the number of knots were reduced, with the aim being to fit the PDEs to each of these interpolants of different complexities, and find the optimum combination of PDE and interpolant (as outlined at the end of this section). These reduced-knot linked GAMs all provided poorer fits to the data in terms of the number of cells on the original grid that were assigned to the wrong abundance category by the GAM (Table 5.2, Fig 5.4). AICc (Akaike 1974, Hurvich and Tsai 1989) and BIC (Schwarz 1978) values were calculated for the GAMs as follows:

$$\mathrm{AICc}_{GAM} = -2\ln\left(L\right) + 2\hat{k} + \frac{2\hat{k}\left(\hat{k}+1\right)}{n-\hat{k}-1} \tag{5.18}$$

$$\mathrm{BIC}_{GAM} = -2\ln\left(L\right) + \hat{k}\ln\left(n\right) \tag{5.19}$$

where L is the likelihood and $\hat{k}$ is the effective number of parameters of the fitted GAM, and n is the number of data points. AICc_{GAM} suggests that the original GAM with the most knots should be preferred, while the more conservative BIC_{GAM} selects one of the new, less complex GAMs (Table 5.2).

Table 5.2: Comparison of sets of linked GAMs with different numbers of knots in the spatial and temporal marginal cubic spline bases. The GAMs are compared based on the percentage of grid cells that they assigned to a different wildebeest abundance category than in the original wildebeest distribution data, and also based on the comparison statistics AICc and BIC.

Knots in x and y marginals	**Knots in t marginal**	**% Grid cells in wrong category**	**AICc$_{GAM}$**	**BIC$_{GAM}$**
20	12/11	12.5	57783*	99652
12	12/11	17.5	66550	87744
10	10	20.1	70545	86307*
8	8	27.2	77717	86927
6	6	39.9	87528	92375

From the linked GAMs fitted to the wildebeest distribution data I was able to obtain estimates of W at any point in the spatio-temporal domain of interest. I could similarly obtain values of G and I at any point in time and space from the grass dynamics model, and of N and T (see Fig. 5.1B-C) at any point in space (these variables are assumed to be constant in time) by kriging. Kriging was carried out using the autoKrige function from the automap package (Hiemstra et al. 2009) in R (R Core Team 2015), which tests a range of variogram models and selects the one giving the lowest residual sum of squares with the sample variogram. From the estimates of W, it was possible to use finite differencing to approximate the partial derivative of wildebeest density with respect to time $\partial W/\partial t$ at the centre of each cell in the spatial grid, at each time point at which the wildebeest distribution data were collected, as follows:

$$\frac{\partial W(\bar{x},\bar{y},\bar{t})}{\partial t} \approx \frac{W(\bar{x},\bar{y},\bar{t}+0.5h_t)-W(\bar{x},\bar{y},\bar{t}-0.5h_t)}{h_t} \tag{5.20}$$

where h_t is a constant step parameter. The partial derivatives with respect to x and y that populate the right hand side of equation (5.6) could similarly be approximated for a given set of movement model parameters using the estimates of the state variables W, G (or I_G), T and N. As h_t (or, similarly, h_x or h_y) moves closer to zero the approximation of the partial derivative at $(\bar{x},\bar{y},\bar{t})$ should become gradually more accurate. However, at very small values of these step parameters, rounding errors that occur during computation start to dominate the estimate and accuracy decreases again. This is illustrated in Fig. 5.5, where it can be seen that the value of the estimates of the partial derivatives of W with respect to x, y and t at a specific point $(\bar{x},\bar{y},\bar{t})$, obtained from the most complex GAMs (Fig. 5.4B), stabilise at relatively small values of h_t, h_x and h_y in the range of 10^{-2} to 10^{-6} kilometres/days, but then start to show large fluctuations at very small values. This pattern appears to be consistent across points in time and space. Based on this finding, I calculated the partial derivatives using $h_t = 10^{-4} days$ and $h_x = h_y = 10^{-4} km$. However, I also tested a more arbitrary alternative scheme with larger step sizes of $h_t = 10 days$ and $h_x = h_y = 1km$ to check whether the model selection results were consistent over alternative step sizes. The larger step sizes chosen here also reduce the impact of any potential under-smoothing between data points in the GAMs by averaging any sharp changes in W over a greater distance.

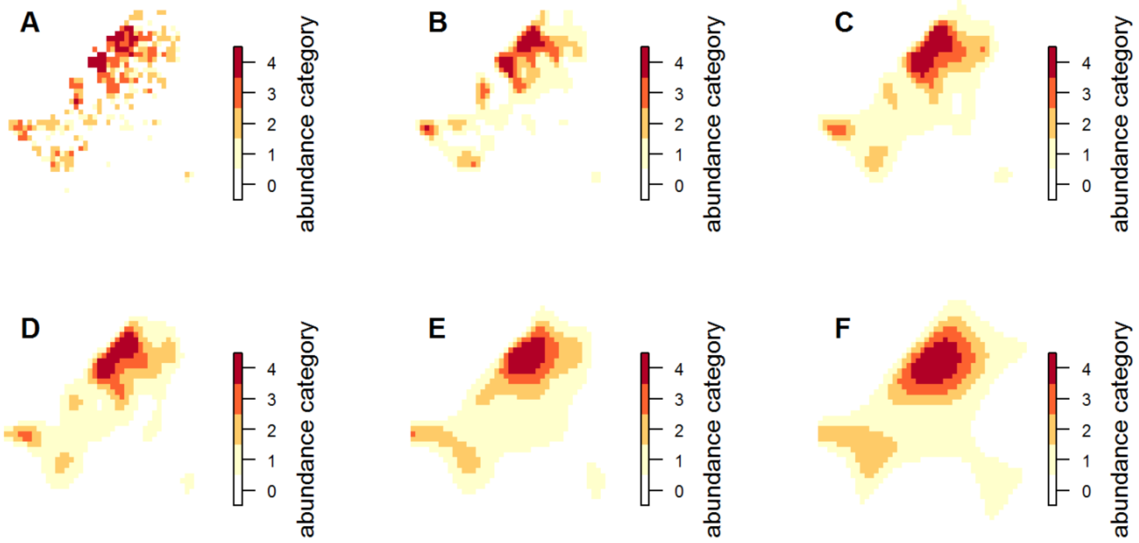

Figure 5.4: Comparison of GAMs of varying complexity. A) Ordinal categorical wildebeest distribution data from August 1969 over the spatial region of interest. **B-F)** Ordinal categories estimated for August 1969 from five different GAMs fitted to the wildebeest distribution data. These GAMs are of successively decreasing flexibility, with the numbers of knots in the spatial and temporal marginal cubic spline bases declining as described in Table 5.1.

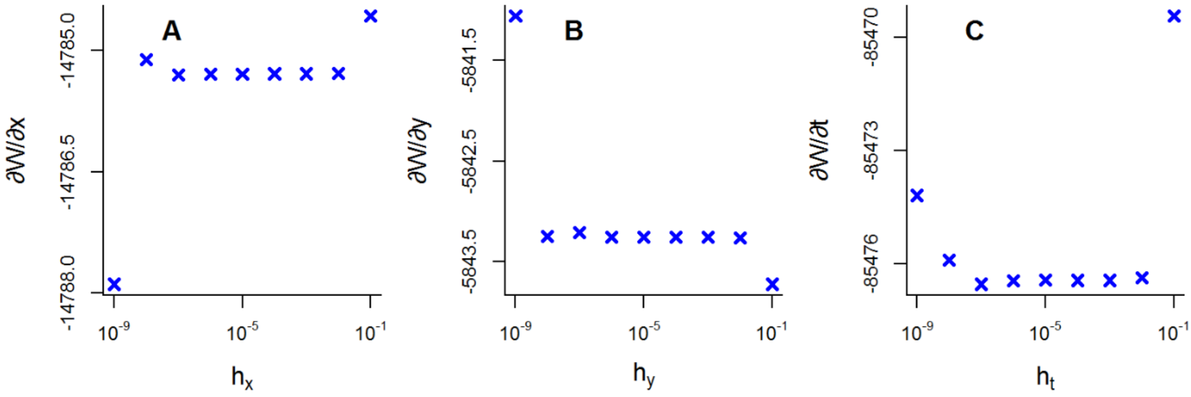

Figure 5.5: Comparison of finite differencing step size parameter values. Estimates of the partial derivatives of W with respect to x (**A**), y (**B**) and t (**C**) at a specific point in space and time $\left(\bar{x},\bar{y},\bar{t}\right)$ that were obtained from the most complex GAMs at a range of values of h_x, h_y and h_t. Units for h_x and h_y are kilometres and for h_t are days.

For the integro-differential equations, which use the advection coefficients given in equations (5.16-17), approximation of the partial derivatives in the right hand side of equation (5.1) also required estimation of the integrals in these advection coefficients. I did this for 26 potential values of r, the wildebeest range of perception, spread evenly at $5km$ intervals between $5km$ and $130km$. Estimation of the integral $\iint_C W(x',y',t)\left((x'-x)\Big/\sqrt{(x-x')^2+(y-y')^2}\right)dC$, for example, at a point (x,y,t) involved first obtaining values of W at time t for every grid cell whose centre lay within r of (x,y). Each of these values of W was then multiplied by the value of $(x'-x)\Big/\sqrt{(x-x')^2+(y-y')^2}$, where (x',y') is the location of the centre of the grid cell where the W value was obtained. By summing each value $W(x',y',t)\left((x'-x)\Big/\sqrt{(x-x')^2+(y-y')^2}\right)$, I then obtain an estimate of $\iint_C W(x',y',t)\left((x'-x)\Big/\sqrt{(x-x')^2+(y-y')^2}\right)dC$.

By obtaining estimates of the partial derivatives with respect to space on the right hand side of equation (5.1) at each point in space and time that occurred in the original wildebeest distribution dataset, as described above, I was able to use equation (5.1) to calculate associated estimates of $\partial W/\partial t$ for a given set of movement model parameters. The sum of squared residuals (SSR) between these estimates of $\partial W/\partial t$ and the alternative estimates obtained directly from the GAMs using equation (5.20) could then easily be obtained. Fitting of the wildebeest movement model parameters by gradient matching was achieved by minimising the SSR using the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm within the R function optim(). It would also have been possible to use a Bayesian scheme based on MCMC sampling for parameter estimation; the choice of a frequentist approach was made because of time constraints, given that achieving parameter convergence with the optimisation algorithm was less computationally costly than with MCMC sampling. The parameter optimisation was repeated 50 times for each model using different sets of randomly selected initial parameter values, after which only the optimised parameter set that gave the lowest SSR out of the 50 was retained. By running multiple optimisations from different starting points, I reduced the risk that the optimised parameters represented a local rather than a global optimum. During the optimisations, I set lower and upper bounds on a number of the model parameters. A diffusion coefficient cannot be negative, so D_W was bounded below at zero, but left unbounded above. Similar zero lower bounds were set for η and ε, since I can think of no biological reason why the wildebeest should be repelled by greater abundances of green grass or grass that is of a higher quality in terms of nitrogen concentration. I leave γ unbounded, allowing for either attraction or repulsion between conspecifics. The parameter $W_{\max}$ was given a minimum value of the maximum wildebeest density obtained from the GAMs, and a maximum value of 6.7×10^5 wildebeest per km^2. For the non-local models (equations (5.11-12)), I fitted each model with each of the 26 values of the parameter r between $5km$ and $130km$; a value of $130km$ allows a wildebeest standing in the centre of the region of interest to determine the environmental conditions across almost the entire region. The values considered for r may appear to include unrealistically large values, but this range was selected on the basis of a previous study that estimated the wildebeest radius of perception to be at least 80km (Holdo et al. 2009).

I used a backward selection approach to determine which of the proposed mechanisms affecting wildebeest movement (movement up gradients in green grass intake/abundance, conspecific density and grass nitrogen concentration, and movement being limited by overcrowding) should be retained in the optimal model. Comparison of the PDE models was achieved using two information criteria; AICc and BIC. These statistics were obtained from the SSR as follows:

$$\text{AICc}_{PDE} = n\ln\left(\frac{SSR}{n}\right) + 2k + \frac{2k(k+1)}{n-k-1} \tag{5.21}$$

$$\text{BIC}_{PDE} = n\ln\left(\frac{SSR}{n}\right) + k\ln(n) \tag{5.22}$$

where n is the number of data points and k is the number of parameters estimated during fitting. The best model is indicated by the lowest value of each of these comparison statistics. Note, however, that these statistics are describing the fit of the PDE models to the GAM-based model that was fitted to the data, not directly to the data itself. This means that if the GAM model is unrealistic in some way, due to it being under- or over-fitted to the data for example, then it may not be possible to get a good fit of the PDEs to the GAM, or, if it is possible to get a good fit of the PDEs, then the fit of the GAM to the data may be so poor that the PDE is not actually giving a good description of the true movement behaviour. As a way of balancing the quality of fit of the PDE with the quality of fit of the GAM, I fitted all of the PDE models using each of the five GAM models outlined in Table 5.2, which vary in their complexity (as indicated by the number of knots used in each), such that both AICc_{GAM} and BIC_{GAM} (equations (5.18-19)) suggest that the least complex GAMs may be underfitted to the data, whilst BIC suggests that the most complex are perhaps overfitted. AICc_{PDE} and BIC_{PDE} values were calculated using equations (5.21-22) for all of these PDE model fits and I used the following equations to penalise these statistics based on the AICc_{PDE} and BIC_{PDE} values (equations (5.21-22)) for the GAMs from which each PDE fit was produced:

$$\text{pAICc} = (1-\lambda)\text{AICc}_{PDE} + \lambda\text{AICc}_{GAM} \tag{5.23}$$

$$\text{pBIC} = (1-\lambda)\text{BIC}_{PDE} + \lambda\text{BIC}_{GAM} \tag{5.24}$$

Here, $0 \le \lambda \le 1$ is a weighting parameter that describes how much weight is to be put on the fit of the GAM versus the fit of the PDE. Ideally the value for λ would be selected using cross-validation, but given the high computational costs of this, I instead compute the pAICc and pBIC values across the range of possible values of λ and take the best PDE/GAM combination to be the one that gives the lowest value of these statistics over the greatest range of λ.

5.6 Results

The AICc_{PDE} and BIC_{PDE} values calculated from the fits of all variations of the wildebeest movement PDE models to the GAMs describing the spatio-temporal distribution of wildebeest (see Appendix D.3; equations (5.21-22)) show three general patterns. The first is that the models fitted using finite difference approximations of the partial derivatives that were obtained using the small step size scheme, where $h_t = 8.64s$ and $h_x = h_y = 10cm$ (see equation 5.20), typically had poorer (larger) AICc_{PDE} and BIC_{PDE} values than the same models fitted using finite difference approximations obtained with the large step size scheme, where $h_t = 10days$ and $h_x = h_y = 1km$. Since the larger step size scheme effectively adds an extra degree of smoothing to the partial derivatives of wildebeest density obtained from whichever GAM the model was fitted to, this trend may indicate that the GAMs are under-smoothing the data. The second pattern observed is that PDE models fitted to GAMs of lower complexity typically have improved AICc_{PDE} and BIC_{PDE} values compared to the same PDE models fitted to higher complexity GAMs, again indicating that a greater degree of smoothing of the data is desirable. Third, I find that, despite the large increase in the total number of fitted parameters in the time-varying parameter models compared to the constant parameter models, the time-varying parameter models had consistently lower AICc_{PDE} and BIC_{PDE} values. This strongly indicates changes in wildebeest movement behaviour over time.

The best model for every GAM complexity based on both AICc_{PDE} and BIC_{PDE} (the two statistics were always in agreement; see tables in Appendix D.3) included the spatial gradients in grass nitrogen concentration N and conspecific density W as drivers of movement, along with an overcrowding effect mediated by $W_{\max}$ (Table 5.3). Each of these best models also included a green grass-based movement response, but there was some disagreement between GAM complexities over whether this response should be to green grass abundance G or green grass intake I_G. A non-local version of each best model was always selected over a local version, with estimates for r ranging from 30km to 50km (Table 5.3).

Table 5.3: Optimal wildebeest movement PDE model for each GAM complexity. The drivers of wildebeest movement in the optimal model selected for each of the five GAM complexities (indicated by the different numbers of knots in the marginal bases in x, y and t) based on AICc_{PDE} and BIC_{PDE} (see tables in Appendix D.3). All of these best models were non-local, had time-varying parameters, and were obtained using the larger step size scheme considered, where $h_t = 10days$ and $h_x = h_y = 1km$ (see equation (5.20)). Note that all of the AICc_{PDE} and BIC_{PDE} values (equations (5.21-22)) recorded here have had the minimum value subtracted for ease of comparison. The distance estimated for the wildebeest range of perception r is given for each GAM complexity. G=green grass abundance; I_G=green grass intake; N=plant nitrogen concentration; W=wildebeest density; W_{max}=maximum tolerated wildebeest density.

Knots in x and y marginals	**Knots in t marginal**	**Best Model**	**Range of perception r (km)**	**AICc_{PDE}**	**BIC_{PDE}**
6	6	$I_G + N + W + W_{max}$	50	0	0
8	8	$G + N + W + W_{max}$	35	90629	90629
10	10	$G + N + W + W_{max}$	30	186794	186794
12	12/11	$G + N + W + W_{max}$	40	173381	173381
20	12/11	$I_G + N + W + W_{max}$	30	220452	220452

Calculating pAICc and pBIC (equations (5.23-24)) for the best models fitted to each GAM complexity over the full range of values for λ indicates that for the vast majority of this range (88% for pAICc and 94% for pBIC), the best PDE model fitted to the lowest GAM complexity (6 knots in each of the 3 marginal bases) would be selected as the best PDE/GAM combination (Fig. 5.6). A comparison of the values of $\partial W/\partial t$ calculated from this PDE, with its optimal parameter values, and the values of $\partial W/\partial t$ calculated from the GAM over the spatial region for the first four time points observed in the data is provided in Fig. 5.7 (see supplementary video 5.3 and its description in Appendix D.2.3 for a comparison over the full time series). It is observed that the closeness of the match between the two estimates of the $\partial W/\partial t$ values varies between time points and over space, with some features in the GAM being accurately represented by the PDE, while other features are not. I provide a similar comparison of the temporal gradients for the best of the constant parameter PDE models fitted to the same GAM (Fig. 5.8). The performance of the constant parameter PDE in replicating the patterns observed in the GAM is much poorer, with the values of $\partial W/\partial t$ from the PDE never reaching the magnitudes observed from the GAM.

The parameter values estimated for the best PDE/GAM combination fluctuate over time, with no obvious patterns (Fig. 5.9). I checked for the presence of seasonality in these parameters by fitting a GAM containing a cyclic cubic regression spline smooth in time with a period of a year to the time series for each parameter using the mgcv package in R (Wood 2006). This analysis found no evidence of seasonality in any of the five parameters (P≥0.05 for all cyclic spline smooths in time). It is observed that both positive and negative values occur for the parameter γ, suggesting attractive interactions between the animals at some time points, but repulsive interactions at others.

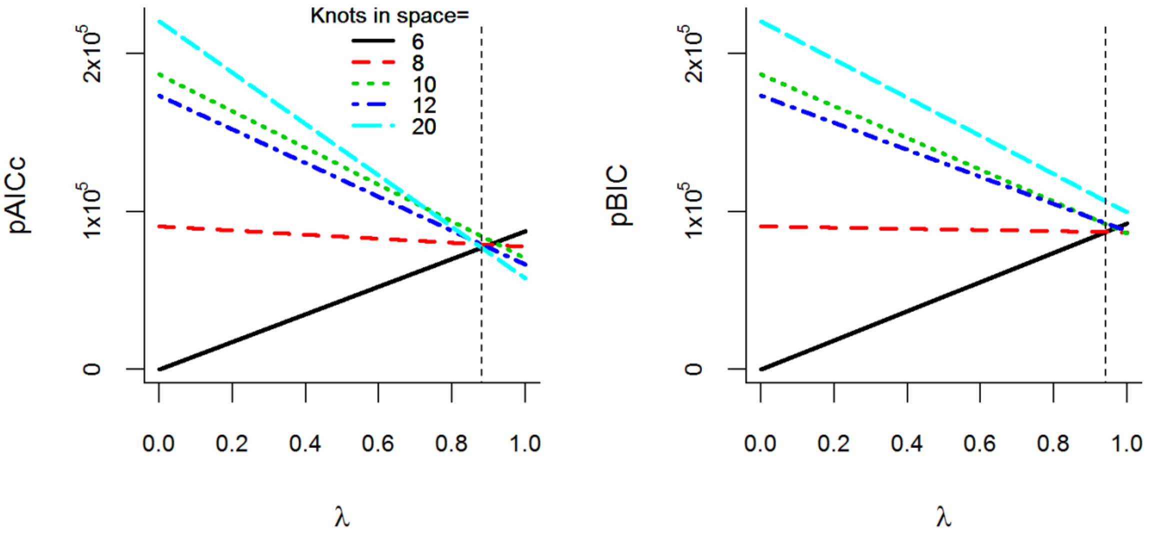

Figure 5.6: Selection of the optimal wildebeest PDE/GAM combination. Plots of the changing values of the pAICc (**A**) and pBIC (**B**) (equations (5.23-24)) for the best PDE model (based on AICc and BIC; see tables in Appendix D.3) fitted to each of the five GAM complexities considered (here indicated by the different colours and line types), as the weighting parameter λ is increased from zero to one. The vertical dashed black line indicates the value of λ at which the PDE model fitted to the least complex GAM (with 6 knots in space) ceases to have the best value of each of the two model comparison statistics.

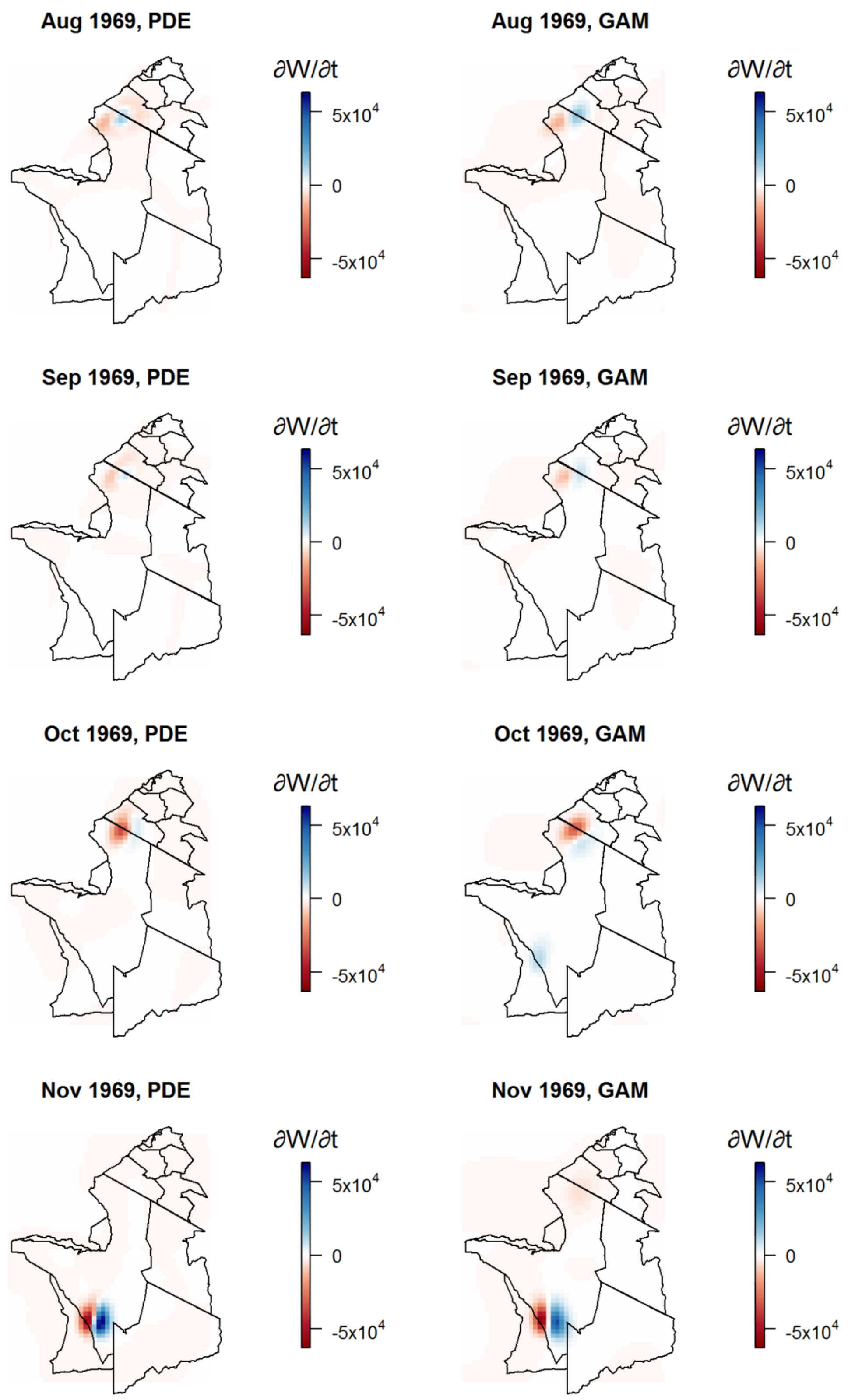

Figure 5.7: Best time-varying parameter wildebeest PDE model fitted to least complex GAM. Comparison of $\partial W/\partial t$ as estimated from the best PDE fitted to the least complex GAM (Table 5.2; suggested to be the best PDE/GAM combination based on pAICc and pBIC (Fig.5.6)) using the optimised parameters (left plots) and $\partial W/\partial t$ as estimated directly from the least complex GAM by finite differencing (right plots) across the spatial region at the first four time points present in the original wildebeest data (rows). Comparison plots for all 33 time points can be observed in supplementary video 5.3 (for video description see Appendix D.2.3).

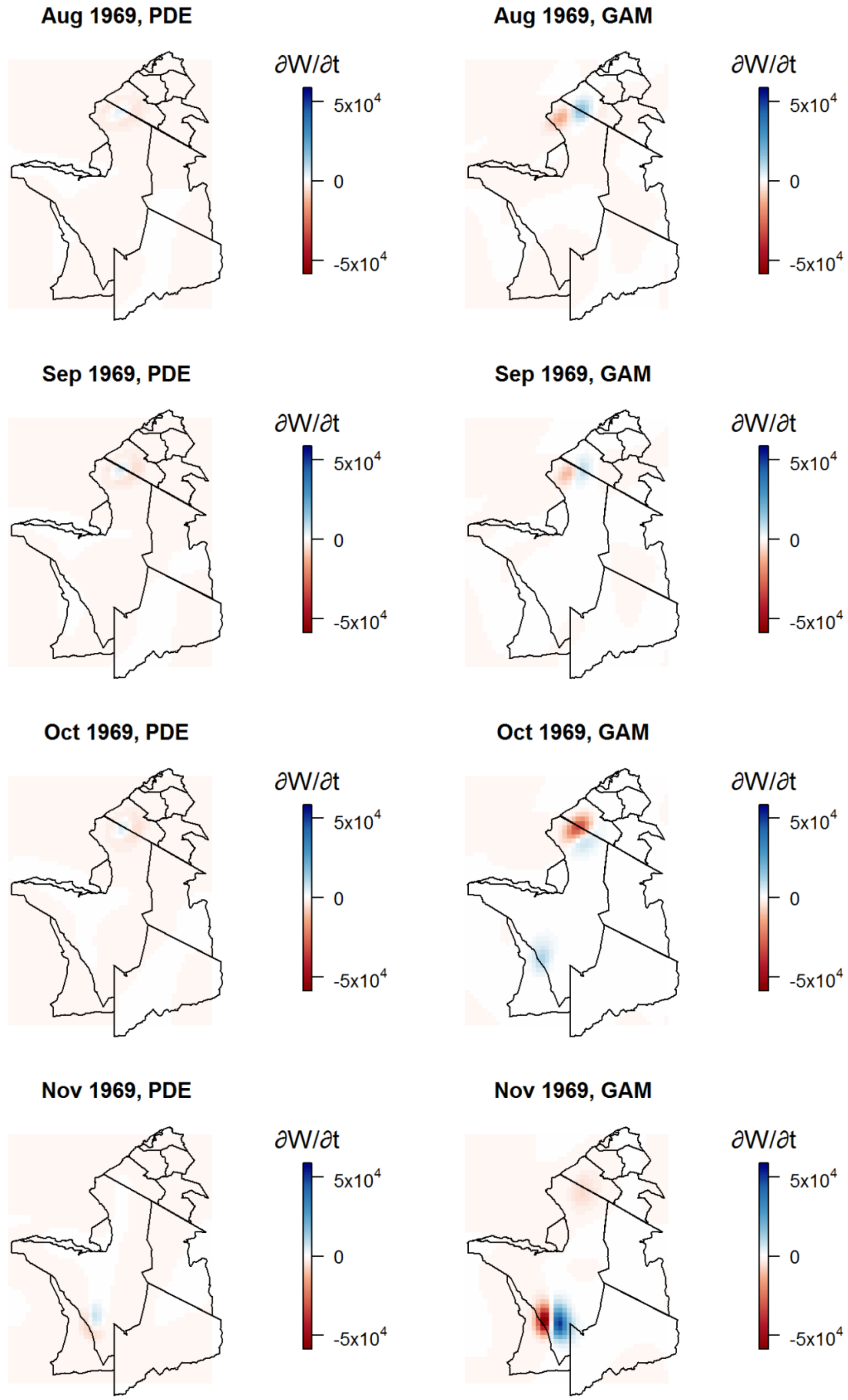

Figure 5.8: Best constant-parameter wildebeest PDE model fitted to least complex GAM. Comparison of $\partial W/\partial t$ as estimated from the best constant parameter PDE fitted to the least complex GAM using the optimised parameters (left plots) and $\partial W/\partial t$ as estimated directly from the least complex GAM by finite differencing (right plots) across the spatial region at the first four time points present in the original wildebeest data (rows). Comparison plots for all 33 time points can be observed in supplementary video 5.4 (for video description see Appendix D.2.4).

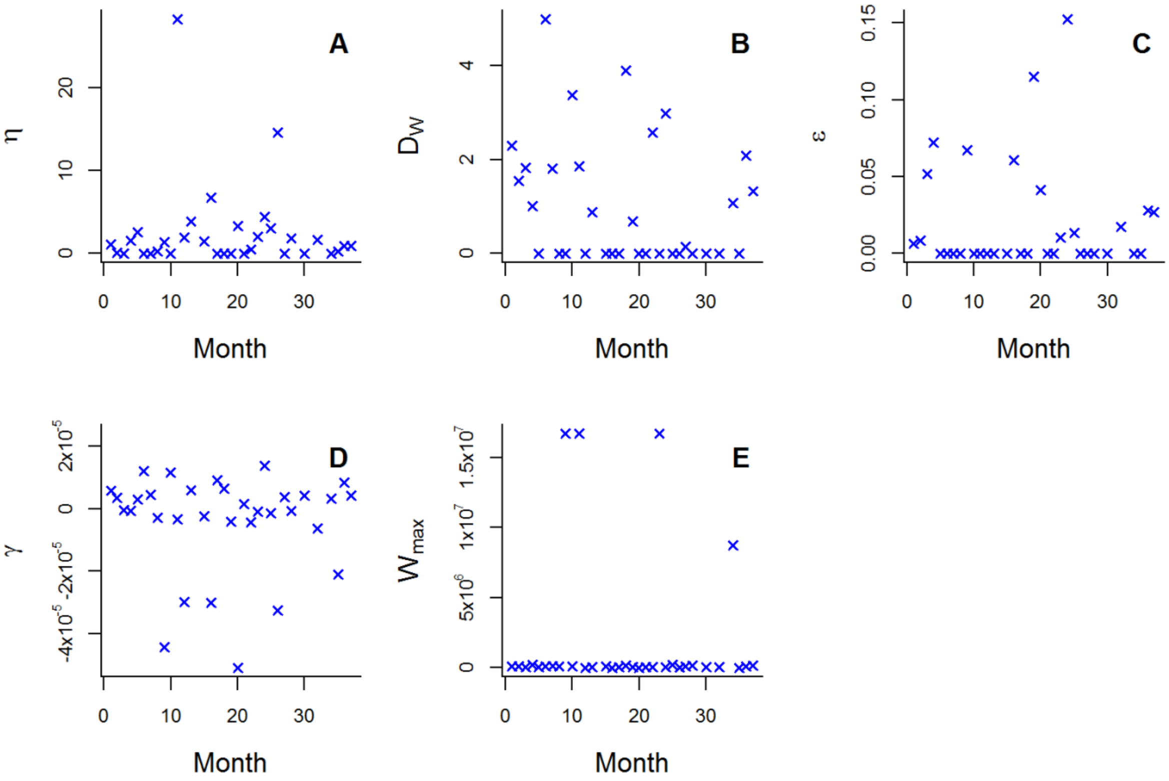

Figure 5.9: Fitted values of each of the time-varying PDE model parameters (Table 5.1) at each time point present in the wildebeest distribution data, from the best model fitted to the least complex of the GAMs (Table 5.3).

5.7 Discussion

I have used a gradient matching approach to fit advection-diffusion PDE models of wildebeest movement to GAM-based wildebeest density surfaces, which had in turn been fitted to ordinal categorical wildebeest distribution data. Model comparison statistics were used to identify the best model for each of five complexities of the wildebeest density surface, and to give an indication of the overall best combination of PDE model and GAM-based density surface. The best PDE/GAM combination included influences of green grass intake (which is determined by factors including rainfall and depletion by grazing), grass nitrogen concentration and conspecific density (both through overcrowding and attraction/repulsion interactions) on wildebeest movement. I also found evidence that the responsiveness of wildebeest to these movement drivers changes over time.

The PDE in the best PDE/GAM combination suggested that wildebeest move in the direction that maximises their green grass intake, which is a result that agrees with previous modelling work applied to the same dataset (Holdo et al. 2009). However, the best models fitted to some of the alternative GAM complexities disagreed with this conclusion, indicating that wildebeest are responding directly to green grass abundance. Intuitively, we might expect that the animals should be seeking to maximise their intake of green grass, which is limited by cropping and digestive constraints (Wilmshurst et al. 1999, Holdo et al. 2009), rather than simply seeking out the location with the highest possible green grass abundance, since, if their intake is already at its maximum, continuing to move to areas of higher grass abundance appears to be a waste of energy. However, given that these animals are typically moving as part of a large herd, leading to rapid depletion of local resources, it may still be a good strategy to readily move to other areas with more green grass. It is also possible that green grass availability in the ecosystem over the time period of interest was such that green grass intake often could not reach its maximum, so that maximising green grass intake and maximising green grass abundance resulted in similar movement patterns. These variables could have led to the observed difficulties in distinguishing between the two alternative strategies.

The finding that grass also directs movement via its nitrogen concentration agrees with results reported in previous modelling studies. Holdo et al. (2009) found that wildebeest had a significant preference for areas with higher grass nitrogen concentration, and proposed that this was the factor driving the southwards migration to the nitrogen-rich plains in the wet season. A tendency for individual wildebeest to move further each day and change direction more frequently when close to or within high-nitrogen patches on the plains, but move shorter distances each day when close to or within high nitrogen patches in the woodlands, was reported by Hopcraft et al. (2014). The suggested explanation for the increased daily movement in response to high nitrogen on the wet season range was that very high densities of grazers congregate on these high-quality patches, causing rapid depletion and forcing more onward movement. The typically more dispersed distribution of animals in the woodlands during the dry season may allow individuals to linger for longer in high-nitrogen patches before resources are depleted. The results presented here provide further confirmation of the attractiveness of areas of high grass nitrogen concentration to wildebeest as they attempt to meet protein requirements.

Attractive and repulsive interactions between individuals have never previously been considered in models of the Serengeti migration, and here I find evidence that such interactions may be important in determining movement patterns. The finding of repulsive interactions in some months was surprising, because previous studies have indicated that aggregatory behaviour is important in this herding species (Thaker et al. 2010, Mose et al. 2013). There are a number of possible explanations for this, some biological, and others that relate to the model being a poor description of wildebeest behaviour or to inaccuracies in inference. Biologically, it is possible that under certain conditions, when resources are particularly scarce, the animals prefer to move further away from each other to reduce competition while grazing. In cases where such repulsive interactions are occurring, a herd may still be maintained by the fact that the individuals all require the same resources and are forced to congregate on limited suitable habitat despite these repulsive tendencies. The mix of attractive and repulsive behaviours observed over different months could also be a consequence of the model wrongly assuming that only one of these interaction types can be occurring at any one time, when in fact both may play a role in driving the dynamics of the herd. Other studies of collective movement have found evidence that individuals are repulsed by conspecifics that come too close, but attracted to conspecifics that are further away; leading to aligned movement and maintenance of a stable inter-individual distance (Lukeman et al. 2010, Katz et al. 2011). Differential equation models that incorporate these short-range repulsion and long-range attraction dynamics have been proposed, and I hope to investigate such models in future work (Mogilner and Edelstein-Keshet 1999, Topaz and Bertozzi 2004, Miller et al. 2012). It should be noted, however, that the inclusion of the overcrowding effect in the PDE models should partially account for such behaviours, as, similarly to short-range repulsion, it prevents densities in excess of a maximum $W_{\max}$. A failure to include other important wildebeest movement drivers, not involving conspecific interactions (see discussion of possible mechanisms below), in the models considered could also have led to the complex inferred patterns of attraction and repulsion. A model that is mechanistically too simple to describe the observed behaviour may try to compensate for missing mechanisms through the flexibility allowed in those mechanisms that are present, leading to spurious inferences. It is also possible that the somewhat messy pattern observed for conspecific interactions is a consequence of identifiability issues between the parameter mediating these effects γ and the diffusion parameter D_W . Both diffusion and conspecific repulsion tend to lead to individuals moving from areas of high to low density, and, given that the months in which the strongest repulsive interactions were observed correspond to months where diffusion was relatively low (Fig. 5.9), it is possible that the parameters are compensating for one another. Finally. as discussed in more detail below, errors in inferring the parameters of the PDE model may arise due to the GAM to which the PDE is fitted not providing an accurate description of the changing wildebeest distribution.

The best PDE/GAM combination suggested a wildebeest range of perception of 50km, with the best PDEs fitted to the alternative GAM complexities suggesting values in the range 30-40km. These distances are shorter than that estimated in a previous model by Holdo et al. (2009), who found that a range of perception of at least 80km was required to produce realistic migration patterns, but are still surprisingly long. There are four possible explanations for this: 1. A large range of perception is being estimated to compensate for some other mechanism that is missing in the model, or as a consequence of inaccuracies in inference resulting from the reliance of the gradient matching technique on an imperfect interpolant; 2. Wildebeest are not actually perceiving information about environmental quality over these large distances in real time, but as a result of their learned or genetic memory; 3. Wildebeest are able to use distant cloud formations, or other long-range cues such as wind, to determine where rainfall, and the resulting new grass growth, is occurring; 4. Wildebeest are only able to directly perceive information about the environment over relatively short distances, but receive information about distant locations indirectly as it is passed through interacting individuals in dispersed herds. Previous modelling studies have shown that such interactions can allow individuals with imperfect knowledge of the environment to more accurately navigate up noisy gradients in environmental quality as part of a group than would be possible in isolation (Grünbaum 1998, Couzin et al. 2005). By allowing for interactions between individuals in our model, I have accounted for this fourth explanation of the long ranges of perception, and it is suspected this is the reason that the presented estimates were somewhat shorter than that of Holdo et al. (2009), who did not account for interactions. Including the effects of memory and cloud cover in the models is a goal for future work, which could allow further narrowing down of the cause of the long estimated range of perception.

The finding that wildebeest movement behaviour changes over time agrees with previous studies that suggest that the behaviour of this species differs between different habitats and seasons (Thaker et al. 2010, Mose et al. 2013, Hopcraft et al. 2014). Such changes in behaviour could be a result of changes in the nutritional requirements of the animals at different times of year as a consequence of events such as calving and the rut. However, I was unable to find a clear seasonal pattern in the changes of the movement parameters that could be explained by these seasonal events. This could be a consequence of the coarse temporal resolution of the data; calving and the rut are events that happen in a roughly 2-3 week period (Hopcraft et al. 2015), so these events could easily be missed by the monthly data collections. A lack of seasonality could also be a result of the way in which the time-varying parameters were fitted, with a separate value of each parameter being estimated independently for each time point. Since there is no dependence of a parameter value at one time point on the values of that parameter at previous or subsequent time points, there is no incentive for smoothness over time in the parameter time series, resulting in the volatile patterns observed in Fig. 5.9. Smoothness could be enforced by assuming a smooth functional form for the parameters over time; a methodology that I previously implemented for the temporal changes in the cell movement parameters in chapters 2-3. The low-order polynomials that were found to be suitable for changes in cell behaviour over hours or days, however, are unlikely to be appropriate for describing changes in wildebeest movement over a period of years, hence why I chose a more flexible approach in this study. A possible alternative of intermediate flexibility would be to model the parameters as cyclic cubic regression splines, which are cubic regression splines where the start and end points (which would here be assumed to be the start and end of a year) are forced to have the same value, and first and second derivatives, resulting in a smooth function that repeats annually (Wood 2006). Two final possible explanations for the lack of pattern observed in the time-varying parameters, both of which are discussed in more detail below, are: 1. Important wildebeest movement drivers are still missing from the current best PDE model, forcing the values of the other parameters to try and compensate in unpredictable ways for these missing drivers; 2. The GAM fitted to the data is a poor description of wildebeest movement patterns, and, as a result, the PDE model requires wildly fluctuating parameters to imitate it.

While the PDE model from the best PDE/GAM combination was able to produce a gradient surface that was similar to that obtained from the GAM for many of the time points, this match was poorer in other time points (see Fig. 5.7, supplementary video 5.3 and Appendix D.2.3). This could, again, be a result of a failure to account for all important wildebeest movement drivers in this PDE model. Including the effects of gradients in the concentration of additional elements in the grass, such as sodium and phosphorus, which have both previously been suggested as drivers of the migration (Murray 1995, Hopcraft et al. 2015), could be a next step in improving the movement model. Additional improvements could be made in the way in which the wildebeest respond to these nutrients in the grass. Currently, it is assumed that the movement response to the nitrogen gradient is unaffected by the density of grass. However, it seems likely that the animals will actually become less responsive to the nitrogen gradient when grass is low. An individual in a location with dense grass of a moderate nitrogen concentration is likely to reach a higher total nitrogen intake over the course of a day than an individual in a location with a high plant nitrogen concentration but very low grass density, which will additionally make it vulnerable to starvation. In addition, the grass nitrogen gradient may become difficult for the animals to detect when grass density is low. Such an effect could be incorporated by having the animals respond to the gradient in $G(1-T)N^{\beta}$ (where β is a parameter that alters the relative weighting given to nitrogen and grass density) rather than to the two separate gradients in $G(1-T)$ and N. The presence of other species in the ecosystem is another potentially important factor that we have not considered. Hopcraft et al. (2014) found only a weak influence of perceived predation risk on individual-level movement of wildebeest, but identified a stronger response to human presence. Other grazing species could also affect wildebeest movement patterns through the increased grazing pressure created by their presence. Data on the distribution of two additional grazers, zebra and Thompsons gazelle, were collected at the same time as the wildebeest data considered in this study (Maddock 1979), so the effects of the distribution of these species on grass biomass and the distribution of wildebeest could be incorporated into the model. Memory is another effect that may be considered in the future, as wildebeest may use information obtained during migrations in previous years to guide their movement decisions in subsequent years. It has previously been noted that the Serengeti wildebeest migration route regularly changes in response to the population size or environmental conditions in a particular year (Pennycuick 1975, Thirgood et al. 2004, Harris et al. 2009, Hopcraft et al. 2015), which suggests that memory is not the over-riding mechanism by which these animals move. However, evidence from the Tarangire-Manyara ecosystem, where there are multiple alternative wet season ranges and individuals show high fidelity to a particular range between years (Morrison and Bolger 2012), suggests that memory may still have some influence on movement decision in this species. A final potential model improvement that I have already discussed above, is the incorporation of more complex interactions between individuals, such as short-range repulsion and long-range attraction.

Another possible cause of the poorer fit of the PDE model to the GAM-based wildebeest density surface at certain time points is that it is not the PDE model, but the GAM-based model that is failing to provide an accurate description of the changing wildebeest distribution over the whole region and time period of interest. The data to which the various GAM models were fitted were collected on a roughly monthly basis, and, given that the average distance a wildebeest covers in just a day is 4.25km (Hopcraft 2010), this temporal resolution can be considered to be relatively coarse. As a fitted GAM simply aims to produce a smooth interpolant of the data, any fine-scale changes in the wildebeest distribution over time that are not observed in the data due to the coarseness of the temporal resolution, will also be missing in the smooth GAM surface. Additionally, given that the data took the form of ordinal categories rather than accurate densities, there was substantial information loss during data collection. The GAM has to attempt to recover this information, inevitably with some error. If the GAM to which the PDE model is fitted does not realistically describe wildebeest movement behaviour, then, even if the correct PDE model for describing the real movement process was known, we might fail to get an accurate match between the temporal gradients in wildebeest density estimated from the fitted PDE model and the GAM. This reliance of the results of inference on the quality of the interpolating surface is an important potential issue when using a gradient matching approach where the fitting of the interpolant to the data and the fitting of the differential equation model to the interpolant are carried out as two separate processes. A more robust alternative is to regularise the interpolant with the differential equations (see for example Dondelinger et al. (2013), Ramsay et al. (2007), Xun et al. (2013)). This involves fitting both the interpolant and the differential equation model simultaneously, with an objective function that both rewards an improved fit of the interpolant to the data and penalises a decreasing quality of fit of the differential equation model to this interpolant, thus providing a better balance between the two fits. Such an approach was not feasible in this study, as fitting the GAM-based interpolants was highly expensive in terms of computational time and memory due to the large size of the wildebeest distribution dataset (85,008 data points) and the complexity of the GAMs, which were required to smooth these data in three dimensions. Repeatedly adjusting the fit of these GAMs based on the fit of the PDEs would not have been possible with the time and resources available. I did, however, implement a less expensive, alternative approach, where I fitted GAMs of five different complexities to the data, fitted the PDEs to each of these different complexity GAMs, and then selected the best PDE/GAM combination by balancing the model comparison statistics calculated from the fits of the GAMs and the fits of the PDEs. In doing so, I found that the least complex interpolant, which had a relatively poor fit to the data as a result of its increased degree of smoothing, but produced the best fitting PDEs, had the greatest degree of support. This preference for the least complex GAM surface suggests that the more complex GAMs that provided better values of AICc_{GAM} and BIC_{GAM} were actually overfitting to the data.

Despite the issue of the reliance of gradient matching on the quality of the interpolation, discussed above, this method of fitting the wildebeest movement models provides two key advantages. First, parameter inference was far less computationally costly using this approach than it would have been using the methods described in chapters 2-3, where the movement models were numerically integrated for each new parameter combination tested. I was able to run 50 optimisations of one of the wildebeest models using the gradient matching approach in less computational time than it often took to run one optimisation of one of the cell models using numerical integrations, despite the larger size of the wildebeest dataset and the more complex, two-dimensional nature of the wildebeest models. Second, by avoiding numerical integration of the PDEs, gradient matching also allows avoidance of the instabilities that are inherent to numerical solutions of advection-diffusion equations in certain regions of parameter space, which presented difficulties when I fitted the cell movement models in chapter 2-3. These numerical instabilities can cause attempts at parameter optimisation to halt prematurely (Sibert et al. 1999), and, if the true movement parameters are in an unstable region of parameter space, inference through numerical integration to get the correct model parameters becomes not just slow, but computationally impossible. It appears that inference through a numerical integration-based method would not have been appropriate in this study, since I have thus far been unable to solve the best wildebeest movement model numerically using the parameters estimated by gradient matching without instabilities causing the model solver to fail. Unfortunately, this inability to numerically solve the final model makes it difficult to obtain wildebeest density surfaces from the model that can be compared to the original data. A possible method for achieving such a comparison would be to convert the PDE model into an SDE (stochastic differential equation) model, which could be used to simulate the movement of lots of individuals. This would be computationally expensive, but it would only be necessary to carry out such a simulation once.

In conclusion, I have carried out inference on PDE models of wildebeest movement to identify a number of drivers of the Serengeti wildebeest migration. These drivers include gradients in environmental covariates, depletion of resources, and interactions between conspecifics. No previous model of this migration has included all of these movement mechanisms, and, indeed, very few models fitted to data from any system exhibiting collective movement have considered all three of these factors. In the process of developing these wildebeest models, I have further extended the framework introduced for modelling cellular movement in chapters 2-3, by modelling in two-dimensional space and considering responses to non-local information. These extensions make the framework applicable to a much wider range of systems, but meant that it was no longer feasible to use numerical model solutions during parameter inference, forcing the use of gradient matching. I have found this method to be a promising approach to decreasing computational costs and allowing inference for advection-diffusion equations in regions of parameter space where it would not otherwise be possible. However, there are some worrying features of the inference results, such as the highly erratic changes in parameter values through time and the poor fit of the PDE model to the GAM interpolant at certain time points, that may be a consequence of inaccuracies in the gradient matching methods. As a result, tests of the approach's ability to retrieve parameter values from datasets of various qualities simulated from movement models of varying complexity are required to assess under which conditions it allows accurate inference. This future work could determine whether gradient matching is really a good solution to the problem of inferring movement behaviour from real, inevitably imperfect, ecological data from complex field systems.

6. Discussion and conclusion

The aim of this thesis was to develop mathematical models and tailored statistical inference methodologies that could help determine which from a comprehensive list of mechanisms are driving collective movement behaviour in a range of study systems at different scales. Therefore, in the preceding chapters, I have developed a range of advection-diffusion PDE models. These described changes in the distribution of populations as an outcome of movement responses to gradients in environmental variables, which in some cases are self-generated by the organisms through depletion, and responses to conspecifics, through attraction, repulsion or overcrowding effects. I have also developed three alternative approaches to inference for these models, and applied them to draw conclusions about the drivers of movement in three systems, two involving small scale cellular movements, and the third involving the large-scale movement of wildebeest around the Serengeti ecosystem. In this chapter, I discuss the key results and developments in methodology arising from this work, before considering some limitations and directions for addressing these in future work.

6.1. A comparison of three study systems

As discussed in chapter 1, collective movements are ubiquitous in biology, exhibiting similar patterns in a range of systems, at often very different scales, suggesting that commonalities in the behaviours that drive movement exist between disparate systems. During the development of models for three study systems, involving *Dictyostelium* cells, human melanoma cells, and wildebeest, I found that similar mathematical features could be used for all three systems to describe hypotheses for movement (for instance, conspecific attraction/repulsion and movements up spatial gradients in chemoattract/grass). I used model inference to draw conclusions about the types of mechanism driving movement in each study system, allowing an investigation of the common causality question.

In chapters 2-3, state-of-the-art methods from computational inference and statistical model selection were applied to show that the movement patterns of cells in both the *Dictyostelium* and the melanoma movement assays were primarily a consequence of the cells depleting a chemical from their environment, and then moving up the resulting gradient in that chemical, as was already known from previous work on these study systems (Muinonen-Martin et al. 2014, Tweedy et al. 2016). The chemicals that the cells were responding to (folate for *Dictyostelium* and lysophosphatidic acid for melanoma) differed between the systems, but the mechanism by which this response occurred was modelled using the same mathematical functions in both cases. A similar mechanism was also identified for the wildebeest system in chapter 5, where the selected best model included preferential movement towards areas with a higher green grass intake rate. This intake rate was determined by green grass abundance, which, in turn, was influenced by depletion due to wildebeest grazing. These results suggest that self-generated gradient mechanisms may be important for generating movement in a range of systems, as is also suggested by previous experimental work in additional systems (see, for example, Scherber et al. (2012), Simpson et al. (2006), Donà et al. (2013)), and that the methods presented in this thesis provide an effective means of detecting these behaviours when they occur in a system.

I found evidence in chapters 2-3 for receptor saturation affecting movement in both *Dictyostelium* and melanoma. This mechanism results in cells being unable to detect and therefore less responsive to chemoattractant gradients when the local concentration of the chemoattractant is high (Tweedy et al. 2013). I also found some evidence of a similar effect in wildebeest, where the PDE in the best PDE/GAM combination involved a response to the green grass intake rate, in preference to a direct response to green grass abundance (as was also previously found by Holdo et al. (2009)). The animals can only consume a limited amount of grass in a day due to food handling and digestive constraints (Wilmshurst et al. 1999), so that once there is enough grass available locally to maximise the intake rate, there is no incentive to continue to move up the grass abundance gradient. Thus, a saturating response to attractive resources in the environment appears to be common across the systems considered.

I found no evidence for attractive or repulsive interactions between the cells in the melanoma data analysed in chapter 2. For *Dictyostelium*, I found evidence for such interactions in one of the two repeats of the experiments analysed (chapter 3), but only limited evidence in the other (chapter 2), where only one of three model comparison statistics (AICc) supported a model with interactions. This indicates that the importance of these interactions varies not just between systems, but also within systems, perhaps based on the state of the particular groups of cells being considered (this particular cell species is well-known for changing its interaction behaviour in response to starvation conditions, as discussed in section 3.6). There was also evidence for attractive and repulsive interactions between wildebeest, suggesting some similarities between *Dictyostelium* and this large ungulate species. Interactions with conspecifics through overcrowding, whereby an individual’s ability to move is inhibited at high density, were found to be important in both melanoma and wildebeest, but not in *Dictyostelium*, possibly because the *Dictyostelium* cells never reached high enough densities for such effects to be detected by the inference schemes.

In all three systems, temporal changes in movement behaviour over time were found to be important in describing the observed movement patterns. It is acknowledged that some of the observed temporal patterns may be spurious consequences of imperfect models (with overly flexible parameters and/or important missing mechanisms) or inaccurate inference as discussed in sections 3.6 and 5.7. However, the presence of behavioural changes is supported by previous work in many systems, including *Dictyostelium* and wildebeest (Bonner 1982, Hopcraft et al. 2014). While there are exceptions, such as random walk models where the animals can switch between movement states (Haydon et al. 2008, Langrock et al. 2014), and the advection-diffusion model of Sibert et al. (1999), the majority of studies that have modelled collective movement have not considered temporal changes in movement parameters. The results presented in this thesis indicate that such simplifications may not be justified. I also found evidence for spatial variation in the movement parameters of *Dictyostelium* cells in chapter 3, but did not test for similar spatial effects in wildebeest, both to reduce computational costs and because the wildebeest herds tend to be focussed in certain areas at certain times of year, so that the effects of space and time on the parameters are likely to be closely correlated, potentially leading to parameter identifiability issues during inference. As a result, no comparison could be made between the cell and wildebeest systems in terms of spatial variation in behaviour.

In addition to changing the approach to inference, discussed in section 6.2, I made two key changes to the original advection-diffusion models used for cell movement before they were applied to wildebeest movement. These were the switch from modelling movement in one spatial dimension to movement in two spatial dimensions, and the introduction of movement responses to non-local information, which allowed for the fact that wildebeest can perceive environmental conditions at locations that are at a distance from their current spatial location. Neither of these changes makes the models unsuitable for modelling cell systems, and, in fact, they should also make the models a more accurate description of cellular behaviour. The assumption of one-dimensional cell movement was a convenient simplification (justified by two statistical hypothesis tests (Appendix B.1)) to reduce computational costs, and was enabled by the particular experimental set-ups, which were spatially two-dimensional, but had little variation in movement behaviour along one spatial axis. Such a simplification is unlikely to be plausible in any non-experimental cellular system, for example during *in vitro* movements of melanoma cells out from a tumour (Muinonen-Martin et al. 2014), or for *Dictyostelium* cells in their natural soil environment (Bonner 1982); indeed, movements in these cases may even require modelling in three spatial dimensions. The local models that I fitted to the cell movement data, which assume that cells respond to the conditions at a point location, are likely to be a close enough approximation to the truth to get good agreement between models and data in many cases (as I found in chapters 2-3). However, cells are able to detect the presence of chemicals across the entire length of their membranes, so, while the range of perception described by the length of a cell is tiny in comparison to the 50km that I estimated for the wildebeest range of perception in chapter 5, the non-local models would be a technically more accurate (if far more computationally costly) description of cell behaviour.

Perhaps the biggest difference between the cellular and wildebeest systems studied in this thesis was that the optimal PDE model fitted to the wildebeest dataset was more complex than those that were selected for the cell datasets. The wildebeest PDE model included various environmental and social effects on movement, which were allowed substantial flexibility in the way that they changed over time, more so than with the low order polynomials used to model temporal dependence in the cell systems. Yet, this model still didn’t appear to be complex enough to capture all of the features of the changing wildebeest distribution as detailed in even the least complex GAM fitted to the data, suggesting that there are still effects missing. This higher complexity in the large-scale wildebeest system, where the data came from a natural, fully-functioning ecosystem, is probably more a consequence of the cell data being from highly simplified and controlled lab systems, rather than an indication that cellular movements are inherently less complex than large mammal movements. Cells moving within the complex ecosystem of the body, for example, have far more scope for complex interactions with a range of different cell types (see for example Wyckoff et al. (2004)), and may move through much more diverse habitats (consider cancer cells moving from tumour, to surrounding tissues, to bloodstream, to other distant tissues (Steeg 2006)), than do cultured *Dictyostelium* cells moving under a gel in a petri dish for a few hours.

6.2. A comparison of three inference methods

As discussed in chapter 1, a majority of studies describing models of collective movement have not formally fitted these models to data or used model comparison techniques to identify the most likely movement drivers. In this thesis, I have developed and trialled three different approaches to inference for advection-diffusion PDE models, where statistical inference of models with realistic levels of complexity has been particularly limited as a consequence of high computational costs and numerical instability issues (see Sibert et al. (1999) and chapter 6 of Soetaert & Herman (2009)). A summary of these three approaches is as follows:

1. A pseudo-Bayesian scheme during which a ‘posterior’ distribution was developed by running parameter optimisations (using maximum likelihood) on many bootstrap samples of the data (see section 2.5). During parameter optimisation, the PDE model was solved numerically for each new parameter set. Model selection was achieved using WAIC (Watanabe 2010).
2. A Bayesian scheme using the delayed rejection adaptive Metropolis algorithm (DRAM; (Haario et al. 2006)), initialised with parameters that were a good approximation of the MAP (maximum a posteriori parameter configuration); see section 3.4 for details. The models were solved numerically for each parameter set tested in both the initial optimisations and the MCMC chains. Model selection again made use of WAIC.
3. A frequentist approach, where the PDE models were fitted by optimising the parameters such that the difference between temporal gradients in wildebeest density estimated from the PDE and from a GAM-based interpolant fitted to the original data was minimised (a method known as gradient matching; see section 4.3 for the interpolant fitting methodology and 5.5 for details on the gradient matching procedure). This method did not require numerical PDE solutions for each parameter set tested. Model selection was achieved using AICc and BIC values calculated for both the GAM and fitted PDE models.

Which of these methods is most suitable for advection-diffusion model inference will vary between cases based on the considerations discussed below.

Given the potentially high computational costs of frequently solving advection-diffusion models numerically, consideration of the computational resources and time that are available for a particular study is important in the choice of inference scheme. If both computational resources and time are limited (or the models are of such complexity that inference schemes involving many numerical solutions are infeasible even with generous resources), then the best option may be to avoid numerical solution entirely and pursue a gradient matching approach. However, the greatly decreased computational cost offered by such methods comes at the price of a potential reduction in the accuracy of the fitted model (Macdonald et al. (2015)). This reduced accuracy is a consequence of not fitting the PDE directly to the data, but instead to an interpolant, which may not be an accurate description of the true density surface from which the data are a sample (see section 5.7 for further discussion of this point). For this reason, if it is feasible to use an inference method that involves numerical model solutions, then it may be advisable to do so. If a computer cluster is available then the first inference method I described, involving the development of a pseudo-posterior through multiple parameter optimisations on bootstrap samples of the data, which can easily be parallelised, has an advantage over the second approach based on MCMC sampling, which is inherently sequential and thus cannot fully exploit the computational resources. In a case where parallel processing capacity is limited, MCMC-based inference may be preferred to the pseudo-Bayesian approach, as it is the more traditional and thoroughly tested option.

Another issue that will influence the choice of inference methodology is the availability of prior information. If detailed priors are available, then a fully Bayesian approach based on MCMC sampling is best able to exploit this information. If only basic information on upper and lower bounds of parameters is available, then any of the three approaches – frequentist, pseudo-Bayesian or fully Bayesian – can make use of this prior information.

In addition to allowing use of detailed prior information, Bayesian approaches also allow parameter uncertainty to be taken into account during model selection. If posterior distributions are available from MCMC sampling it is possible to use this information on parameter uncertainty to calculate and compare WAIC scores for different models, rather than the less reliable comparison statistics, including AIC and BIC, that are offered in a frequentist scheme. The use of advanced comparison statistics like WAIC is not just restricted to fully Bayesian approaches. I used a simple test study to determine whether WAIC scores calculated using the pseudo-Bayesian scheme were comparable to those calculated using a true posterior (Appendix A.4), which indicated good agreement between the two. However, it should be noted that this test involved a very different model system to the cell movement one, and it is unclear whether it involved a likelihood surface of a comparable complexity to that being explored in the real problem (see discussion in section 2.7). Indeed, more basic inference approaches using AICc and BIC were just as effective as the WAIC obtained by bootstrapping in this test case. Therefore, provisional to more extensive testing, the pseudo-Bayesian scheme may offer a promising alternative the fully Bayesian one for taking parameter uncertainty into account.

A further consideration that must be made is the stability of numerical model solutions. Instabilities in the numerical solution of advection-diffusion PDEs under certain parameter regimes, particularly when advection is dominant over diffusion (i.e. the Péclet number is high) (Leonard 1979, Soetaert and Herman 2009), can mean that it is not possible to explore certain regions of parameter space using inference methods that involve numerical solution of the PDEs (Sibert et al. 1999). This may not be a problem if the optimal parameters lie in a stable region of parameter space – as appeared to be the case for the models fitted to the cell systems in this thesis. However, if the true movement parameters are in an unstable region (as seems to have been the case for the wildebeest system), then the only option for parameter inference may be a gradient matching approach, which does not require numerical model solutions. It should be noted that, while a frequentist approach to inference was used with the gradient matching approach in chapter 5, it would be equally possible to combine gradient matching with a Bayesian scheme (see, for example, Macdonald et al. (2015), Xun et al. (2013)), and so take advantage of prior information and parameter uncertainty as discussed above. I used a frequentist approach alongside the gradient matching methods in chapter 5 primarily to remain within time constraints; achieving convergence of MCMC chains is typically more time consuming than achieving convergence of an optimisation algorithm. I would not advocate combining gradient matching with the pseudo-Bayesian approach unless the fitting of an interpolant to the data was a low cost procedure (which was not the case for the interpolant I fitted to the wildebeest data; chapter 4), since an interpolant would have to be fitted to every bootstrap sample of the data.

The inference methods developed in this thesis are able to cope with data of different qualities. This was demonstrated in the contrast between the cellular data, which provided accurate information on the locations of all individuals through time, and the wildebeest data, which consisted of coarse ordinal abundance categories on a spatial grid at monthly intervals. The high quality cell data were used to calculate a likelihood from the numerical PDE solutions as described in section 2.4. However, they could also be smoothed in space and time to allow them to be used within a gradient matching approach. The GAM-based method described in chapter 4 was used both to enable recovery of realistic wildebeest densities from the ordinal categories into densities and to produce a spatio-temporal interpolant that could be used for inference of the PDE models using gradient matching. It would also have been possible to use these data in an inference method based on numerical PDE solutions by optimising the PDE model parameters such that the difference (as quantified, for example, by the Kullback-Liebler divergence) between the wildebeest density surfaces obtained from the GAM-based method and the numerical PDE solution was minimised.

6.3. Limitations and future directions

As discussed in section 6.1, the models and inference methods developed in this thesis have been successfully used to identify drivers of movement in three systems. The selected models for the cellular systems appear to have produced movement patterns that give very good matches to the data for the cell systems. However, while the best model in the wildebeest system seems to be successfully capturing many of the main features of the population distribution in time and space, it is still missing some details, suggesting that further model development might be required, as discussed in section 5.7. It must also be acknowledged that the three study systems studied here, while representing movement at two very different scales, do not cover the huge diversity of taxa in which collective movements are observed (see the examples given in fish, insects, birds, etc. in chapter 1). Ideally, the modelling framework would be able to generalise to describe the movement behaviour in any of these various systems. Below, I briefly outline three areas where the models discussed in this thesis could be extended to potentially improve their ability to explain movement in the systems studied (particularly wildebeest), or to make them more applicable to additional systems, where their application could be useful in the future. I also discuss some limitations to the inference approaches used (see section 6.2) and how these limitations might be overcome.

6.3.1. Memory-driven movement

In chapter 1, I identified and described four key types of movement driver (see section 1.1); environmental variability, environmental depletion, interactions between individuals, and memory. Incorporating a range of these movement drivers into collective movement models was one of the aims of this thesis, and, of the four types of driver, the only one that I have not yet considered in the advection-diffusion models that I developed is memory.

The reason for memory being given a lower priority than the other three movement drivers in this work was primarily a consequence of the particular study systems investigated. The lack of a brain may mean that memories are relatively unimportant in cellular movement decisions. However, the tendency of cells to persist in their movement direction, even in the absence of any directional cue could be considered a type of memory (Bosgraaf and Van Haastert 2009), as could events in cell differentiation, where a precursor cell exposed to short-term signals permanently becomes more specialised (Ajo-Franklin et al. 2007). For wildebeest, the plasticity observed in the migration route between years suggests that movement decisions are primarily made in response to the current environmental conditions rather than in response to memories from previous years (Pennycuick 1975, Thirgood et al. 2004, Harris et al. 2009, Hopcraft et al. 2015). A study on wildebeest in another region, however, has shown high wet season fidelity (Morrison and Bolger 2012), so memory may be a minor movement driver that could be introduced to further improve the fit of the wildebeest model in the future. Memory could be incorporated into the models through a bias in the movement of wildebeest at a particular time point towards the location where they were most densely focussed at the same time in the previous year.

In addition to perhaps providing a better description of wildebeest behaviour, the inclusion of learned or genetic memory in the models is likely to be essential for explaining the movement of many other species that are believed to rely much more heavily on this driver to accurately navigate to distant locations; for example, salmon, sea turtles, and many bird species (Helbig 1996, Lohmann et al. 2008, Mueller et al. 2013). In such memory-driven systems, the earth's magnetic field is most often credited as the guiding mechanism. This movement behaviour could be incorporated into the models by assuming that the individuals bias their movements up or down gradients in the magnitude and inclination of the magnetic field, until they reach the remembered signature of their target location (Lohmann et al. 2008). The target location of the individuals could be switched based on seasonal environmental cues to simulate back and forth migratory movements.

6.3.2. Individual differences in movement behaviour

A second feature that was not included in the models presented here, possibly limiting their applicability to certain systems, was individual variation. Differences between individuals have frequently been identified as an important driver of movement patterns. In partial migration, for example, different subsets of the population follow different movement patterns, with one part of the population migrating seasonally, while the other part remains in the same region year-round. Membership of these population subsets is often determined by the competitive ability or state of individuals. In European blackbirds, for example, it is typically the less competitive females and juveniles that migrate for the winter, while the adult males are able to remain and monopolise the limited resources (Lundberg 1985). A similar dynamic is observed in roach, where individuals that have been able to attain a larger size while feeding in lakes over the summer are more likely to migrate to streams, where food is low but there is less risk of predation, over the winter (Broderson et al. 2008). Differences in individual behaviour are also important in systems where there are subsets of leaders and followers within a population, such as in the case of whooping cranes, where more experienced individuals appear to have a larger influence over the movement of their flocks than do less experienced individuals (Mueller et al. 2013). Other studies have identified distributions of movement parameter values across the individuals in a population (see, for example, Hopcraft et al. (2014)).

Individual-based models, such as the random walk and self-propelled particle models discussed in section 1.2, may be a more flexible framework for incorporating differences in the behaviour of individuals than the advection-diffusion PDEs that are the main focus of this thesis. This is particularly true if, for whatever reason, every individual in the population must have its own personal set of movement parameters. However, the advection-diffusion models could be extended to describe the movement of different sub-groups within the population, where the individuals within a sub-group share a set of movement parameters. This would require that each sub-group be modelled using a separate equation, similar to the approach used by compartmental models in epidemiology (Brauer 2008). Individuals could even be allowed to switch between sub-groups at a given rate to allow for the switches in behavioural state that have been identified as being important in a number of species using models based on mixtures of random walks (Morales et al. 2004, Langrock et al. 2014).

While it is known that individual differences in movement behaviour exist in wildebeest (Hopcraft et al. 2014), it is unclear whether incorporating individual variation would have led to an improved model. In general, more work is required to determine under what scenarios differences in individual behaviour are likely to have important consequences for the emergent population movement patterns, and when such individual differences can reasonably be averaged over in models to still adequately replicate the whole population movement.

6.3.3. Complex social behaviour

The models developed in this work allow attractive or repulsive interactions between conspecifics, but not both at the same time. As previously discussed in section 5.7, this may be somewhat at odds with the more complex situation found in a number of systems, where individuals may be simultaneously be repulsed by individuals that are too close to them, attracted to individuals that are further away, and perhaps also be actively trying to align their direction of motion with individuals at intermediate distances (Lukeman et al. 2010, Katz et al. 2011). Such complex interactions have typically been modelled using self-propelled particle models (Couzin et al. 2002), but a smaller number of studies have also incorporated these dynamics into advection-diffusion models using integro-differential equations similar to those I used to describe non-local responses to environment conditions in the wildebeest system (Mogilner and Edelstein-Keshet 1999, Topaz and Bertozzi 2004, Miller et al. 2012). To my knowledge, there have been no attempts to fit these advection-diffusion models to data, and this would be an interesting avenue for future work. Extending the model I used to describe the distribution of the wildebeest population in chapter 5 to include complex social behaviours of this type may improve the fit of the model to the data; particularly since the estimated parameters currently indicate a difficult to interpret mixture of both attractive and repulsive interactions between conspecifics.

6.3.4. Limitations of the inference methodologies

A number of limitations of the various inference strategies I developed during this work have already been touched upon in section 6.2. For the methods involving numerical solution of the models, a major issue is the instabilities in certain parameter regimes. A finite differencing scheme that reduces these issues with instabilities, such as ‘upwind’ differencing, can be chosen to numerically solve the models, but this does not always completely remove the issue; some parameter combinations can still produce instabilities that can prevent accurate inference (Sibert et al. 1999). Thus, inference using these schemes is limited to cases where the true model parameters are within a stable region. This is most likely to be the case in systems where advective movement does not overpower diffusion (Leonard 1979).

The alternative to numerical solution-based methods is gradient matching. This approach also has a limitation in that the accuracy of inference is dependent on the accuracy of the interpolant used to describe the data (Macdonald and Husmeier 2015). As discussed in more detail in section 5.7, the quality of the interpolant could be improved by regularising it with the differential equations (see Dondelinger et al. (2013), Ramsay et al. (2007), Xun et al. (2013)). Implementation of a regularisation scheme is, therefore, recommended when using gradient matching, but if (as was the case for the application in chapter 5) such an approach is not feasible due to computational costs, the method described in equations (5.19-20) offers a less expensive alternative. This cheaper approach, which involved selecting the best combination of various alternative interpolants and PDE models by balancing model comparison statistics calculated for both interpolants and PDEs, still requires some validation, however. A study to compare whether the results of this approach are generally similar to those of a proper regularisation scheme would be useful.

Further validation studies for the inference methodologies developed in chapters 2-3 could also be an area for future work. Both the pseudo-Bayesian scheme based on optimisations on many bootstrapped datasets, and the Bayesian approach using MCMC chains initialised at the MAP (rather than random initial parameters from a hyperdispersed distribution) are, non-standard techniques that were necessary as a consequence of the high computational costs of numerical solutions and of reaching convergence of MCMC chains for the particular models that were the focus of this work. I have presented short validation studies for both of these techniques (see Appendices A.4 and B.4), which indicate that they produce model inference results that are comparable to more standard approaches in at least one test example. In fact, MCMC sampling around the MAP was found to be more precise than standard Metropolis sampling, and similarly precise to population MCMC, in identifying the correct model (Appendix B.4). Additionally, as discussed in section 3.6, this method is less restrictive than the Laplace approximation (Rue et al. 2009), which is a more standard approach to approximating the posterior in the face of computational difficulties. More extensive testing with a range of models and simulated datasets may still be advisable, however, particularly for the bootstrapping technique of chapter 2, to increase confidence in these newly developed inference approaches.

6.4. Conclusion

Over the course of this thesis, I have developed advection-diffusion models of collective movement behaviour that incorporate a wide range of movement drivers – including environmental variation, environmental depletion and conspecific interactions – and that account for spatial and temporal variation in the response of individuals to these drivers. I have also developed a range of inference methods that can be applied to determine the drivers of collective movement from data for a given system. These methods have been specifically designed to allow effective inference in the face of the many difficulties presented by advection-diffusion models, particularly high computational costs and numerical instabilities, which have previously led to these models rarely being fitted to data. These models and inference techniques have been applied to data from three study systems to successfully allow conclusions to be drawn about the drivers of movement in these systems, and thus show that collective movements in systems at opposite ends of the scale spectrum can be influenced by surprisingly similar dynamics. More work is required in making the models generalisable to the full range of collective movements observed in biological systems, particularly through the addition of memory mechanisms, inter-individual differences in behaviour, and more complex social dynamics, but the advection-diffusion modelling framework is flexible enough for these additional behaviours to be incorporated in future work. In short, the techniques presented in this thesis represent a toolbox that I hope will be used for increasing understanding of the mechanisms underlying collective movement in a wide range of systems. An improved understanding of what drives collective movements could allow these movements to be managed, for example, to prevent the collapse of important migrations, to control pest species, or to prevent the mass movement of cancer cells around the body.

7. Bibliography

Aderhold, A., D. Husmeier, and M. Grzegorczyk. 2017. Approximate Bayesian inference in semi-mechanistic models. Statistics and Computing 27:1003–1040.

Agostinelli, C., and L. Greco. 2013. A weighted strategy to handle likelihood uncertainty in Bayesian inference. Computational Statistics 28:319–339.

Ajo-Franklin, C. M., D. A. Drubin, J. A. Eskin, E. P. S. Gee, D. Landgraf, I. Phillips, and P. A. Silver. 2007. Rational design of memory in eukaryotic cells service Rational design of memory in eukaryotic cells. Genes & Development 21:2271–2276.

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

Anderson, K., and K. J. Gaston. 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Frontiers in Ecology and the Environment 11:138–146.

Balch, C. M., J. E. Gershenwald, S. J. Soong, J. F. Thompson, M. B. Atkins, D. R. Byrd, A. C. Buzaid, A. J. Cochran, D. G. Coit, S. Ding, A. M. Eggermont, K. T. Flaherty, P. A. Gimotty, J. M. Kirkwood, K. M. McMasters, M. C. Mihm, D. L. Morton, M. I. Ross, A. J. Sober, and V. K. Sondak. 2009. Final version of 2009 AJCC melanoma staging and classification. Journal of Clinical Oncology 27:6199–6206.

Ballerini, M., N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. 2008. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. Proceedings of the National Academy of Sciences of the United States of America 105:1232–1237.

Barber, D. 2012. Bayesian Reasoning and Machine Learning. Cambridge University Press, Cambridge, UK.

Bazazi, S., J. Buhl, J. J. Hale, M. L. Anstey, G. A. Sword, S. J. Simpson, and I. D. Couzin. 2008. Collective motion and cannibalism in locust migratory bands. Current biology 18:735–739.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B 57:289–300.

Berdahl, A., C. J. Torney, C. C. Ioannou, J. J. Faria, and I. D. Couzin. 2013. Emergent sensing of complex environments by mobile animal groups. Science 339:574–576.

Bernstein, R. L., C. Rossier, R. Van Driel, M. Brunner, and G. Gerisch. 1981. Folate deaminase and cyclic AMP phosphodiesterase in Dictyostelium discoideum: their regulation by extracellular cyclic AMP and folic acid. Cell Differentiation 10:79–86.

Bishop, C. M. 2006. Pattern Recognition and Machine Learning. Springer.

Blackwell, P. G. 1997. Random diffusion models for animal movement. Ecological Modelling 100:87–102.

Bolger, D. T., W. D. Newmark, T. A. Morrison, and D. F. Doak. 2008. The need for integrative approaches to understand and conserve migratory ungulates. Ecology Letters 11:63–77.

Bonner, J. T. 1982. Comparative biology of cellular slime molds. Pages 1–33 *in* W. F. Loomis, editor. The development of Dictyostelium discoideum. Academic Press, New York.

Boone, R. B., S. J. Thirgood, and J. G. C. Hopcraft. 2006. Serengeti wildebeest migratory patterns modeled from rainfall and new vegetation growth. Ecology 87:1987–1994.

Börger, L., J. Matthiopoulos, R. M. Holdo, J. M. Morales, I. Couzin, and E. McCauley. 2011. Migration quantified : constructing models and linking them with data. Pages 111–128 *in* E. J. Milner-Gulland, J. M. Fryxell, and A. R. E. Sinclair, editors. Animal Migration: A Synthesis. Oxford University Press, New York.

Bosgraaf, L., and P. J. M. Van Haastert. 2009. Navigation of chemotactic cells by parallel signaling to pseudopod persistence and orientation. PloS one 4:e6842.

Bowman, A. W., and A. Azzalini. 2014. R package “sm”: nonparametric smoothing methods.

Brauer, F. 2008. Compartmental models in epidemiology. Pages 19–79 *in* F. Brauer, P. van den Driessche, and J. Wu, editors. Mathematical epidemiology. Springer.

Breslow, A. 1970. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. Annals of surgery 172:902–908.

Broderson, J., P. A. Nilsson, L.-A. Hansson, C. Skov, and C. Bronmark. 2008. Condition-dependent individual decision-making determines cyprinid partial migration. Ecology 89:1195–1200.

Buhl, J., D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. 2006. From disorder to order in marching locusts. Science 312:1402–1406.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Page Ecological Modelling. 2nd edition. Springer, New York.

Carnell, M. J., and R. H. Insall. 2011. Actin on disease - Studying the pathobiology of cell motility using Dictyostelium discoideum. Seminars in cell & developmental biology 22:82–88.

Christiansen, F., G. A. Víkingsson, M. H. Rasmussen, and D. Lusseau. 2013. Minke whales maximise energy storage on their feeding grounds. The Journal of experimental biology 216:427–436.

Chu, W., and Z. Ghahramani. 2005. Gaussian processes for ordinal regression. Journal of Machine Learning Research 6:1019–1041.

Chubb, J. R., A. Wilkins, G. M. Thomas, and R. H. Insall. 2000. The Dictyostelium RasS protein is required for macropinocytosis, phagocytosis and the control of cell movement. Journal of cell science 113:709–719.

Coburn, L., L. Cerone, C. Torney, I. D. Couzin, and Z. Neufeld. 2013. Tactile interactions lead to coherent motion and enhanced chemotaxis of migrating cells. Physical biology 10:46002.

Cochran, W. W., H. Mouritsen, and M. Wikelski. 2004. Migrating songbirds recalibrate their magnetic compass daily from twilight cues. Science 304:405–408.

Codling, E. A., J. W. Pitchford, and S. D. Simpson. 2007. Group navigation and the “many-wrongs principle” in models of animal movement. Ecology 88:1864–1870.

Codling, E. A., M. J. Plank, and S. Benhamou. 2008. Random walk models in biology. Journal of the Royal Society Interface 5:813–834.

Couzin, I. D., C. C. Ioannou, G. Demirel, T. Gross, C. J. Torney, A. Hartnett, L. Conradt, S. A. Levin, and N. E. Leonard. 2011. Uninformed individuals promote democratic consensus in animal groups. Science 334:1578–1580.

Couzin, I. D., J. Krause, N. R. Franks, and S. A. Levin. 2005. Effective leadership and decision-making in animal groups on the move. Nature 433:513–516.

Couzin, I. D., J. Krause, R. James, G. D. Ruxton, and N. R. Franks. 2002. Collective Memory and Spatial Sorting in Animal Groups. Journal of Theoretical Biology 218:1–11.

Cressie, N., and C. K. Wikle. 2011. Stististics for spatio-temporal data. John Wiley & Sons.

Dean, W. R. J., P. Barnard, and M. D. Anderson. 2009. When to stay , when to go : trade-offs for southern African arid-zone birds in times of drought. South African Journal of Science 105:24–28.

Deisboeck, T. S., and I. D. Couzin. 2009. Collective behavior in cancer cell populations. BioEssays 31:190–197.

Dingle, H., and V. A. Drake. 2007. What Is Migration? BioScience 57:113–121.

Donà, E., J. D. Barry, G. Valentin, C. Quirin, A. Khmelinskii, A. Kunze, S. Durdu, L. R. Newton, A. Fernandez-Minan, W. Huber, M. Knop, and D. Gilmour. 2013. Directional tissue migration through a self-generated chemokine gradient. Nature 503:285–289.

Dondelinger, F., M. Filippone, S. Rogers, and D. Husmeier. 2013. ODE parameter inference using adaptive gradient matching with Gaussian processes. Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics 31:216–228.

Eitle, E., and G. Gerisch. 1977. Implication of developmentally regulated Concanavalin A binding proteins of Dictyostelium in cel adhesion and cyclic AMP regulation. Cell Differentiation 6:339–346.

Ershad, S., K. Dideban, and F. Faraji. 2013. Synthesis and application of polyaniline/multi walled carbon nanotube nanocomposite for electrochemical determination of folic acid. Analytical & Bioanalytical Electrochemistry 5:178–192.

Fagan, W. F., M. a Lewis, M. Auger-Méthé, T. Avgar, S. Benhamou, G. Breed, L. Ladage, U. E. Schlägel, W.-W. Tang, Y. P. Papastamatiou, J. Forester, and T. Mueller. 2013. Spatial memory and animal movement. Ecology letters 16:1316–1329.

Fleming, A. B., and W. M. Saltzman. 2002. Pharmacokinetics of the carmustine implant. Clinical pharmacokinetics 41:403–419.

Fortin, D., P.-L. Buono, A. Fortin, N. Courbin, C. T. Gingras, P. R. Moorcroft, R. Courtois, and C. Dussault. 2013. Movement responses of caribou to human-induced habitat edges lead to their aggregation near anthropogenic features. The American naturalist 181:827–36.

Frankfurt Zoological Society, and Harvey Maps. 2010. Serengeti: Official map and visitor guide.

Friedl, P., and D. Gilmour. 2009. Collective cell migration in morphogenesis, regeneration and cancer. Nature reviews. Molecular cell biology 10:445–457.

Friedman, N., M. Linial, I. Nachman, and D. Pe'er. 2000. Using Bayesian Networks to Analyze Expression Data. Journal of Computational Biology 7:601–620.

Friel, A. N., and A. N. Pettitt. 2008. Marginal likelihood estimation via power posteriors. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 70:589–607.

Fryxell, J. M., and A. R. E. Sinclair. 1988. Causes and consequences of migration by large herbivores. Trends in Ecology & Evolution 3:237–241.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian Data Analysis. 3rd edition. CRC Press.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–472.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. Pages 169-193 *in* J. Bernado, J. Berger, A. Dawid, and A. Smith, editors. Bayesian Statistics 4. Oxford University Press, Oxford, UK.

Girolami, M., B. Calderhead, and V. Vyshermirsky. 2010. System identification and model ranking: the Bayesian perspective. Pages 201–230 *in* N. D. Lawrence, M. Girolami, M. Rattray, and G. Sanguinetti, editors. Learning and Inference in Computational Systems Biology. 1st edition. MIT Press, Massachusetts.

Goodwin, B. C. 1965. Oscillatory behavior in enzymatic control processes. Advances in enzyme regulation 3:425–438.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732.

Grünbaum, D. 1998. Schooling as a strategy for taxis in a noisy environment. Evolutionary Ecology 12:503–522.

Gueron, S., and S. A. Levin. 1993. Self-organization of Front Patterns in Large Wildebeest Herds. Journal of Theoretical Biology 165:541–552.

Guisan, A., and F. E. Harrell. 2000. Ordinal response regression models in ecology. Journal of Vegetation Science 11:617–626.

Gwinner, E. 1996. Circannual clocks in avian reproduction and migration. Ibis 138:47–63.

Haario, H., M. Laine, A. Mira, and E. Saksman. 2006. DRAM: Efficient adaptive MCMC. Statistics and Computing 16:339–354.

Hahn, S., S. Bauer, and F. Liechti. 2009. The natural link between Europe and Africa - 2.1 billion birds on migration. Oikos 118:624–626.

Hall, R. L. 1977. Amoeboid movement as a correlated walk. Journal of mathematical biology 4:327–335.

Hansson, L.-A., and S. Hylander. 2009. Size-structured risk assessments govern Daphnia migration. Proceedings of the Royal Society B: Biological Sciences 276:331–336.

Harris, G., S. Thirgood, J. G. C. Hopcraft, J. P. G. M. Cromsight, and J. Berger. 2009. Global decline in aggregated migrations of large terrestrial mammals. Endangered Species Research 7:55–76.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd edition. Springer, New York.

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

Hebblewhite, M., and D. T. Haydon. 2010. Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. Philosophical transactions of the Royal Society B: Biological Sciences 365:2303–2312.

Helbig, A. J. 1996. Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palaearctic warblers (Aves: Sylviidae). The Journal of Experimental Biology 199:49–55.

Hiemstra, P. H., E. J. Pebesma, C. J. W. Twenhöfel, and G. B. M. Heuvelink. 2009. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. Computers and Geosciences 35:1711–1721.

Hillen, T., and K. J. Painter. 2009. A user’s guide to PDE models for chemotaxis. Journal of

Mathematical Biology 58:183–217.

Holdo, R. M., R. D. Holt, and J. M. Fryxell. 2009. Opposing rainfall and plant nutritional gradients best explain the wildebeest migration in the Serengeti. The American Naturalist 173:431–445.

Hopcraft, J. G. C. 2010. Balancing food and predation risk: Ecological implications for large herbivores in the Serengeti. University of Groningen.

Hopcraft, J. G. C., R. M. Holdo, E. Mwangomo, S. Mduma, S. J. Thirgood, J. M. Fryxell, M. Borner, H. Olff, and A. R. E. Sinclair. 2015. Why are wildebeest the most abundant herbivore in the Serengeti ecosystem? Pages 125–174 *in* A. R. E. Sinclair, K. L. Metzger, S. A. R. Mduma, and J. M. Fryxell, editors. Serengeti IV: Sustaining biodiversity in a coupled human-natural system. University of Chicago Press, Chicago.

Hopcraft, J. G. C., J. M. Morales, H. L. Beyer, M. Borner, E. Mwangomo, A. R. E. Sinclair, H. Olff, and D. T. Haydon. 2014. Competition, predation, and migration: Individual choice patterns of Serengeti migrants captured by hierarchical models. Ecological Monographs 84:355–372.

Hu, F., and J. Zidek. 2002. The weighted likelihood. The Canadian Journal of Statistics 30:347–371.

Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

Insall, R. H. 2010. Understanding eukaryotic chemotaxis: a pseudopod-centred view. Nature Reviews Molecular cell biology 11:453–458.

Jonzén, N., E. Knudsen, R. D. Holt, and B. Sæther. 2011. Uncertainty and predictability : the niches of migrants and nomads. Pages 91–109 *in* E. J. Milner-Gulland, J. M. Fryxell, and A. R. E. Sinclair, editors. Animal Migration: A Synthesis. Oxford University Press, New York.

Kakebeeke, P. I. J., R. J. W. De Wit, S. D. Kohtz, and T. M. Konijn. 1979. Negative chemotaxis in Dictyostelium and Polysphondylium. Experimental Cell Research 124:429–433.

Kalimuthu, P., and S. A. John. 2009. Selective electrochemical sensor for folic acid at physiological pH using ultrathin electropolymerized film of functionalized thiadiazole modified glassy carbon electrode. Biosensors and Bioelectronics 24:3575–3580.

Katz, Y., K. Tunstrøm, C. C. Ioannou, C. Huepe, and I. D. Couzin. 2011. Inferring the structure and dynamics of interactions in schooling fish. Proceedings of the National Academy of Sciences of the United States of America 108:18720–18725.

Keating, M. T., and J. T. Bonner. 1977. Negative chemotaxis in cellular slime molds. Journal of Bacteriology 130:144–147.

Keller, E. F., and L. A. Segel. 1970. Initiation of slime mold aggregation viewed as an instability. Journal of theoretical biology 26:399–415.

Kermack, W. O., and A. G. McKendrick. 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 115:700–721.

Kooperberg, C. 2015. logspline: Logspline density estimation routines.

Kooperberg, C., and C. J. Stone. 1992. Logspline density estimation for censored data. Journal of Computational and Graphical Statistics 1:301–328.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. The Annals of Mathematical Statistics 22:79–86.

Laevsky, G., and D. A. Knecht. 2001. Under-agarose folate chemotaxis of Dictyostelium discoideum amoebae in permissive and mechanically inhibited conditions. BioTechniques 31:1140–1149.

Langrock, R., J. G. C. Hopcraft, P. G. Blackwell, V. Goodall, R. King, M. Niu, T. A. Patterson, M. W. Pedersen, A. Skarin, and R. S. Schick. 2014. Modelling group dynamic animal movement. Methods in Ecology and Evolution 5:190–199.

Langrock, R., R. King, J. Matthiopoulos, L. Thomas, D. Fortin, and J. M. Morales. 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. Ecology 93:2336–2342.

Leonard, B. P. 1979. A stable and accurate convective modelling procedure based on quadratic upstream interpolation. Computer methods in applied mechanics and engineering 19:59–98.

Lo, C.-M., H.-B. Wang, M. Dembo, and Y.-L. Wang. 2000. Cell movement is guided by the rigidity of the substrate. Biophysical Journal 79:144–152.

Lohmann, K. J., N. F. Putman, and C. M. F. Lohmann. 2008. Geomagnetic imprinting: A unifying hypothesis of long-distance natal homing in salmon and sea turtles. PNAS 105:19096–19101.

Lukeman, R., Y.-X. Li, and L. Edelstein-Keshet. 2010. Inferring individual rules from collective behavior. PNAS 107:12576–12580.

Lundberg, P. 1985. Dominance behaviour, body weight and fat variations, and partial migration in European blackbirds Turdus merula. Behavioural Ecology and Sociobiology 17:185–189.

Macdonald, B., C. Higham, and D. Husmeier. 2015. Controversy in mechanistic modelling with Gaussian processes. Proceedings of The 32nd International Conference on Machine Learning 37:1539–1547.

Macdonald, B., and D. Husmeier. 2015. Gradient matching methods for computational inference in mechanistic models for systems biology: a review and comparative analysis. Frontiers in bioengineering and biotechnology 3:180.

Maddock, L. 1979. The "migration" and grazing succession. Pages 104–129 *in* A. R. E. Sinclair and M. Norton-Griffiths, editors. Serengeti: dynamics of an ecosystem. University of Chicago Press, Chicago.

Majumdar, R., M. Sixt, and C. A. Parent. 2014. New paradigms in the establishment and maintenance of gradients during directed cell migration. Current opinion in cell biology 30:33–40.

Mann, R. P. 2011. Bayesian inference for identifying interaction rules in moving animal groups. PloS one 6:e22827.

Mann, R. P., A. Perna, D. Strömbom, R. Garnett, J. E. Herbert-Read, D. J. T. Sumpter, and A. J. W. Ward. 2013. Multi-scale inference of interaction rules in animal groups using Bayesian model selection. PLoS computational biology 9:e1002961.

Mayor, R., and C. Carmona-Fontaine. 2010. Keeping in touch with contact inhibition of locomotion. Trends in cell biology 20:319–328.

McNaughton, S. J. 1979. Grassland-herbivore dynamics. Pages 46–81 *in* A. R. E. Sinclair and M. Norton-Griffiths, editors. Serengeti: dynamics of an ecosystem. University of Chicago Press, Chicago.

McNicol, R. E., E. Scherer, and J. H. Gee. 1996. Shoaling enhances cadmium avoidance by lake whitefish, Coregonus clupeaformis. Environmental Biology of Fishes 47:311–319.

Mduma, S. A. R., A. R. E. Sinclair, and R. Hilborn. 1999. Food regulates the Serengeti wildebeest:

a 40-year record. Journal of Animal Ecology 68:1101–1122.

Merkle, J. A., D. Fortin, and J. M. Morales. 2014. A memory-based foraging tactic reveals an adaptive mechanism for restricted space use. Ecology Letters 17:924–931.

Miller, J. M., A. Kolpas, J. P. J. Neto, and L. F. Rossi. 2012. A Continuum Three-Zone Model for Swarms. Bulletin of Mathematical Biology 74:536–561.

Mishra, S., K. Tunstrøm, I. D. Couzin, and C. Huepe. 2012. Collective dynamics of self-propelled particles with variable speed. Physical Review E 86:11901.

Mogilner, A., and L. Edelstein-Keshet. 1999. A non-local model for a swarm. Journal of Mathematical Biology 38:534–570.

Moorcroft, P. R., and M. A. Lewis. 2006. Mechanistic Home-Range Analysis. Princeton University Press, Princeton.

Moorcroft, P. R., M. A. Lewis, and R. L. Crabtree. 2006. Mechanistic home range models capture spatial patterns and dynamics of coyote territories in Yellowstone. Proceedings of the Royal Society B 273:1651–1659.

Morales, J. M., D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell. 2004. Extracting more out of relocation data: Building movement models as mixtures of random walks. Ecology 85:2436–2445.

Morrison, T. A., and D. T. Bolger. 2012. Wet season range fidelity in a tropical migratory ungulate. Journal of Animal Ecology 81:543–552.

Mose, V. N., T. Nguyen-huu, D. Western, P. Auger, and C. Nyandwi. 2013. Modelling the dynamics of migrations for large herbivore populations in the Amboseli National Park, Kenya. Ecological Modelling 254:43–49.

Mueller, T., R. B. O'Hara, S. J. Converse, R. P. Urbanek, and W. F. Fagan. 2013. Social learning of migratory performance. Science (New York, N.Y.) 341:999–1002.

Muinonen-Martin, A. J., O. Susanto, Q. Zhang, E. Smethurst, W. J. Faller, D. M. Veltman, G. Kalna, C. Lindsay, D. C. Bennett, O. J. Sansom, R. Herd, R. Jones, L. M. Machesky, M. J. O. Wakelam, D. A. Knecht, and R. H. Insall. 2014. Melanoma cells break down LPA to establish local gradients that drive chemotactic dispersal. PLoS biology 12:e1001966.

Muinonen-Martin, A. J., D. M. Veltman, G. Kalna, and R. H. Insall. 2010. An improved chamber for direct visualisation of chemotaxis. PloS one 5:e15309.

Murphy, K. P. 2012. Machine Learning: A Probabilistic Perspective. MIT Press, Cambridge, Massachusetts.

Murray, M. G. 1995. Specific nutrient requirements and migration of wildebeest. Pages 231–256*in* A. R. E. Sinclair and P. Arcese, editors.Serengeti II: Dynamics, management, and conservation of an ecosystem. University of Chicago Press, Chicago.

Neilson, M. P., D. M. Veltman, P. J. M. van Haastert, S. D. Webb, J. A. Mackenzie, and R. H. Insall. 2011. Chemotaxis: a feedback-based computational model robustly predicts multiple aspects of real cell behaviour. PLoS biology 9:e1000618.

Ng, M. R., A. Besser, G. Danuser, and J. S. Brugge. 2012. Substrate stiffness regulates cadherin-dependent collective migration through myosin-II contractility. Journal of Cell Biology 199:545–563.

Norton-Griffiths, M. 1973. Counting the Serengeti migratory wildebeest using two-stage sampling. East African Wildlife Journal 11:135–149.

Norton-Griffiths, M. 1979. Influence of grazing, browsing and fire on vegetation dynamics. Pages 310–352 *in* A. R. E. Sinclair and M. Norton-Griffiths, editors. Serengeti: dynamics of an ecosystem. University of Chicago Press, Chicago.

Owen, S. A., L. L. Sanders, L. J. Edwards, H. F. Seigler, D. S. Tyler, and J. M. Grichnik. 2001. Identification of higher risk thin melanomas should be based on Breslow depth not Clark level IV. Cancer 91:983–991.

Parichy, D. M., M. V. Reedy, and C. A. Erickson. 2007. Regulation of melanoblast migration and differentiation. Pages 108–139 *in* J. J. Nordland, R. E. Boissy, V. J. Hearing, R. A. King, and J. P. Ortonne, editors. The pigmentary system and its disorders. 2nd edition. Oxford University Press, Oxford.

Pennycuick, L. 1975. Movements of the migratory wildebeest population in the Serengeti area between 1960 and 1973. East African Wildlife Journal 13:65–87.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O., G. Hooker, D. Campbell, and J. Cao. 2007. Parameter estimation for differential equations: a gen- eralized smoothing approach. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 69:741–796.

Reebs, S. G. 2000. Can a minority of informed leaders determine the foraging movements of a fish shoal? Animal Behaviour 59:403–409.

Reed, D. N., T. M. Anderson, J. Dempewolf, K. Metzger, S. Serneels, and D. Bowman. 2009. The spatial distribution of vegetation types in the Serengeti ecosystem: The influence of rainfall and topographic relief on vegetation patch characteristics. Journal of Biogeography 36:770–782.

Ripplinger, J., and J. Sullivan. 2008. Does choice in model selection affect maximum likelihood analysis? Systematic biology 57:76–85.

Rorth, P. 2009. Collective cell migration. Annual Review of Cell and Developmental Biology 25:407–429.

Ross, R. 1911. The prevention of malaria. John Murray, London.

Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 71:319–392.

Scarpa, E., A. Szabo, A. Bibonne, E. Theveneau, M. Parsons, and R. Mayor. 2015. Cadherin Switch during EMT in Neural Crest Cells Leads to Contact Inhibition of Locomotion via Repolarization of Forces. Developmental Cell 34:421–434.

Scherber, C., A. J. Aranyosi, B. Kulemann, S. P. Thayer, M. Toner, O. Iliopoulos, and D. Irimia. 2012. Epithelial cell guidance by self-generated EGF gradients. Integrative Biology 4:259–269.

Schiesser, W. E., and G. W. Griffiths. 2009. A compendium of partial differential equation models: method of lines analysis with Matlab. Cambridge University Press, New York.

Schwarz, G. 1978. Estimating the dimension of a model. The annals of statistics 6:461–464.

Sibert, J. R., J. Hampton, D. A. Fournier, and P. J. Bills. 1999. An advection–diffusion–reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (Katsuwonus pelamis). Canadian Journal of Fisheries and Aquatic Sciences 56:925–938.

Simons, A. M. 2004. Many wrongs: the advantage of group navigation. Trends in ecology & evolution 19:453–455.

Simpson, S. J., E. Despland, B. F. Hägele, and T. Dodgson. 2001. Gregarious behavior in desert locusts is evoked by touching their back legs. Proceedings of the National Academy of Sciences 98:3895–3897.

Simpson, S. J., G. a Sword, P. D. Lorch, and I. D. Couzin. 2006. Cannibal crickets on a forced march for protein and salt. Proceedings of the National Academy of Sciences of the United States of America 103:4152–6.

Sinclair, A. R. E. 1979. Dynamics of the Serengeti ecosystem. Pages 1–30 *in* A. R. E. Sinclair and M. Norton-Griffiths, editors. Serengeti: dynamics of an ecosystem. University of Chicago Press, Chicago.

Soetaert, K., and P. M. Herman. 2009. A practical guide to ecological modelling: using R as a simulation platform. Springer.

Soetaert, K., and T. Petzoldt. 2010. Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. Journal Of Statistical Software 33:1–28.

Soetaert, K., T. Petzoldt, and R. W. Setzer. 2010. Solving Differential Equations in R: Package deSolve. Journal of Statistical Software 33:1–25.

Sokolov, A., I. S. Aranson, J. O. Kessler, and R. E. Goldstein. 2007. Concentration Dependence of the Collective Dynamics of Swimming Bacteria. Physical Review Letters 98:158102.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B: Statistical Methodology 64:583–616.

Steeg, P. S. 2006. Tumor metastasis: mechanistic insights and clinical challenges. Nature medicine 12:895–904.

Stone, C. J., M. H. Hansen, C. Kooperberg, and Y. K. Truong. 1997. Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture. The Annals of Statistics 25:1371–1470.

Strandburg-Peshkin, A., D. R. Farine, I. D. Couzin, and M. C. Crofoot. 2015. Shared decision-making drives collective movement in wild baboons. Science 348:1358–1361.

Szabó, B., G. Szöllösi, B. Gönci, Z. Jurányi, D. Selmeczi, and T. Vicsek. 2006. Phase transition in the collective migration of tissue cells: Experiment and model. Physical Review E 74:61908.

Thaker, M., A. T. Vanak, C. R. Owen, M. B. Ogden, and R. Slotow. 2010. Group dynamics of zebra and wildebeest in a woodland savanna: effects of predation risk and habitat density. PloS one 5:1–6.

Thirgood, S., A. Mosser, S. Tham, G. Hopcraft, E. Mwangomo, T. Mlengeya, M. Kilewo, J. Fryxell, A. R. E. Sinclair, and M. Borner. 2004. Can parks protect migratory ungulates? The case of the Serengeti wildebeest. Animal Conservation 7:113–120.

Topaz, C. M., and A. L. Bertozzi. 2004. Swarming patterns in a two-dimensional kinematic model for biological groups. SIAM Journal on Applied Mathematics 65:152–174.

Tweedy, L., D. A. Knecht, G. M. Mackay, and R. H. Insall. 2016. Self-generated chemoattractant gradients: Attractant depletion extends the range and robustness of chemotaxis. PLoS biology 14:e1002404.

Tweedy, L., B. Meier, J. Stephan, D. Heinrich, and R. G. Endres. 2013. Distinct cell shapes determine accurate chemotaxis. Scientific Reports 3:2606.

Urbanek, R. P., L. E. A. Fondow, C. D. Satyshur, A. E. Lacy, and S. E. Zimorski. 2005. First cohort of migratory whooping cranes reintroduced to eastern North America: the first year after release. North American Crane Workshop Proceedings 9:213–224.

Varnum, B., and D. R. Soll. 1981. Chemoresponsiveness to cAMP and folic acid during growth, development, and dedifferentiation in Dictyostelium discoideum. Differentiation 18:151–160.

Veltman, D. M., M. G. Lemieux, D. A. Knecht, and R. H. Insall. 2014. PIP3-dependent macropinocytosis is incompatible with chemotaxis. Journal of Cell Biology 204:497–505.

Venkiteswaran, G., S. W. Lewellis, J. Wang, E. Reynolds, C. Nicholson, and H. Knaut. 2013. Generation and dynamics of an endogenous, self-generated signaling gradient across a migrating tissue. Cell 155:674–687.

Vicsek, T., A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. 1995. Novel type of phase transition in a system of self-driven particles. Physical Review Letters 75:1226–1229.

Visser, M. E., and C. Both. 2005. Shifts in phenology due to global climate change: the need for a yardstick. Proceedings of the Royal Society B: Biological Sciences 272:2561–2569.

Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11:3571–3594.

Wikle, C. K., and M. B. Hooten. 2006. Hierarchical Bayesian spatio–temporal models for population spread. Pages 145–169 *in* J. S. Clark and A. E. Gelfand, editors. Hierarchical Modelling for the Enviromental Sciences: Statistical Methods and Applicatons. Oxford University Press.

Wilmshurst, J. F., J. M. Fryxell, B. P. Farm, A. R. E. Sinclair, and C. P. Henschel. 1999. Spatial distribution of Serengeti wildebeest in relation to resources. Canadian Journal of Zoology 77:1223–1232.

De Wit, R. J., R. Bulgakov, T. F. Rinke de Wit, and T. M. Konijn. 1986. Developmental regulation of the pathways of folate-receptor-mediated stimulation of cAMP and cGMP synthesis in Dictyostelium discoideum. Differentiation 32:192–199.

Wolanski, E., and E. Gereta. 2001. Water quantity and quality as the factors driving the Serengeti ecosystem , Tanzania. Hydrobiologia 458:169–180.

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society. Series B: Statistical Methodology 73:3–36.

Wood, S. N., N. Pya, and B. Säfken. 2016. Smoothing parameter and model selection for general smooth models. arXiv:1511.03864v2.

Worton, B. J. 1987. A review of models of home range for animal movement. Ecological Modelling 38:277–298.

Wyckoff, J., W. Wang, E. Y. Lin, Y. Wang, F. Pixley, E. R. Stanley, T. Graf, J. W. Pollard, J. Segall, and J. Condeelis. 2004. A Paracrine Loop between Tumor Cells and Macrophages Is Required for Tumor Cell Migration in Mammary Tumors. Cancer Research 64:7022–7029.

Xun, X., J. Cao, B. Mallick, A. Maity, and R. J. Carroll. 2013. Parameter Estimation of Partial Differential Equation Models. Journal of the American Statistical Association 108:1009–1020.

Zhang, H., D. Wessels, P. Fey, K. Daniels, R. L. Chisholm, and D. R. Soll. 2002. Phosphorylation of the myosin regulatory light chain plays a role in motility and polarity during Dictyostelium chemotaxis. Journal of Cell Science 115:1733–1747.

Appendix A: Additional information for chapter 2

A.1. Numerical model solution

A.1.1. The method of lines

I numerically solved the partial differential equations on which the models in section 2.3 were based using the method of lines (Schiesser and Griffiths 2009, Soetaert et al. 2010). This involved discretising the spatial region of interest of length l into equal-sized boxes, so that changes in cell density and attractant concentration in these boxes through time could be described as a system of ordinary differential equations (ODEs).

The basic form of the cell movement PDEs (equation (2.1)) can be rewritten as:

$$\begin{aligned}\frac{\partial C(x,t)}{\partial t} &= -\frac{\partial}{\partial x}\left\{a(x,t)C(x,t) - D_C(t)\frac{\partial C(x,t)}{\partial x}\right\} + \nu C(x,t)\\ &= -\frac{\partial Flux_C(x,t)}{\partial x} + \nu C(x,t)\end{aligned} \tag{A.1.1}$$

where $Flux_C$ is the cell flux, which describes the net movement of cells up the spatial axis if positive and down the spatial axis if negative. The one-dimensional spatial regions of interest were divided into boxes of length $\Delta x = 50\mu m$ for *Dictyostelium* and $\Delta x = 10\mu m$ for melanoma, allowing cell density changes in box $i \in (1,...,B)$ to be described:

$$\frac{dC^i(t)}{dt} = -\frac{Flux_C^{i,i+1}(t) - Flux_C^{i-1,i}(t)}{\Delta x} + \nu C^i(t) \tag{A.1.2}$$

where $Flux_C^{i-1,i}$ describes the cell flux between boxes $i-1$ and i. The cell fluxes across the boundaries of the modelled region were specified as described below in Appendix A.1.4. Fluxes over the region's internal box boundaries were obtained by approximating the spatial derivatives by finite differencing. For example, given equations (2.3, A.1.1), $Flux_C$ for the basic model is:

$$Flux_C(x,t) = \alpha(t)\frac{\partial A(x,t)}{\partial x}C(x,t) - D_C(t)\frac{\partial C(x,t)}{\partial x} \tag{A.1.3}$$

and the $Flux_C^{i-1,i}$ are estimated by:

$$Flux_C^{i-1,i}(t) = \alpha(t)\frac{A^i(t) - A^{i-1}(t)}{\Delta x}C^{i-1}(t) - D_C(t)\frac{C^i(t) - C^{i-1}(t)}{\Delta x} \tag{A.1.4}$$

For those models incorporating the attractant $A(x,t)$, the additional attractant PDE (equation (2.4)) can, like the cell PDE (equation (A.1.1)), be rewritten in terms of fluxes:

$$\begin{aligned}\frac{\partial A(x,t)}{\partial t} &= -\gamma(t)C(x,t)A(x,t) - \frac{\partial}{\partial x}\left\{-D_A\frac{\partial A(x,t)}{\partial x}\right\} \\ &= -\gamma(t)C(x,t)A(x,t) - \frac{\partial Flux_A(x,t)}{\partial x}\end{aligned} \tag{A.1.5}$$

Changes in attractant levels in a particular box i in the discretised spatial region are then be described by:

$$\frac{dA^i(t)}{dt} = -\gamma C^i(t)A^i(t) - \frac{Flux_A^{i,i+1}(t) - Flux_A^{i-1,i}(t)}{\Delta x} \tag{A.1.6}$$

where $Flux_A^{i-1,i}$ describes the attractant flux between boxes $i-1$ and i. Attractant fluxes across the internal box boundaries were approximated in the same way as the cell fluxes (equation (A.1.4)), using finite differences:

$$Flux_A^{i-1,i}(t) = D_A\frac{A^i(t) - A^{i-1}(t)}{\Delta x} \tag{A.1.7}$$

The attractant fluxes across the external boundaries of the modelled region were specified as described in Appendix A.1.4.

Numerical solutions of the models were obtained by numerical integration of the system of ODEs described in equations (A.1.2, A.1.6). Numerical integration was achieved using the R package deSolve (function ode.1D) (Soetaert et al. 2010).

A.1.2. Initial conditions

In the melanoma dataset, there were no cells in the observation region at $t = 0$. I, therefore, expect that no depletion of the attractant LPA had occurred in this region by $t = 0$, so that LPA remained at 100% of its initial concentration throughout the region at this point. Appropriate initial conditions from which to solve the models are, thus, $C(x,0) = 0$ and $A(x,0) = 1$.

In the *Dictyostelium* dataset, cells were already present in the left of the observation region at $t = 0$. The initial cell density distribution was, therefore, obtained from the cell locations at *t*=0 by first obtaining a probability density function by logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015). This probability density function was then rescaled to ensure that the integral of $C(x,0)$ over the modelled region equalled the number of cells in the observation region at $t = 0$.

The folate in the *Dictyostelium* assay was homogeneously distributed in the gel at a concentration of 10µM prior to the addition of the cells to a folate-free trough that was cut into the gel (the edge of this trough is visible to the left of the image in Fig. 2.2A). However, there were no data on the folate distribution at the time point $t = 0$ where the first cell observations were made. Given that some cells have already moved under the gel at the left side of the region of interest at $t = 0$, it seems likely that some depletion of the folate will have occurred in this region. I, therefore, expect the folate distribution at $t = 0$ to be roughly sigmoidal in form, with low concentrations occurring near the initially folate-free trough, and a smooth increase in concentration to a maximum of 10μM occurring as we move to the right, away from the trough and the folate-depleting cells. Such a distribution of attractant at $t = 0$ can be obtained by assuming the sigmoidal functional:

$$A(x,0) = \frac{10}{1+e^{-\delta(x-\varepsilon)}} \qquad \text{(A.1.8)}$$

The parameters δ and ε respectively describe the steepness of the increase in folate as we move to the right of the region, and the location in x at which half the folate is remaining. Since the precise values of these parameters were unknown, they were inferred during model fitting. I set realistic maximum and minimum values for both of these parameters (δ_{min}=0.002, δ_{max}=1, ε_{min}=0 and ε_{max}=700) by comparing the cell distribution at $t = 0$ to folate distributions obtained from equation (A.1.8) with a range of parameter values, and selecting those values giving the realistic extremes that the attractant distribution at $t = 0$ could take (Fig. A.1.1). There is little change to the folate distribution if δ is increased above the selected δ_{max}, hence the choice of this bound. Decreasing δ below δ_{min} causes folate to be depleted too far in advance of the cell front, or to extend too far into the initially folate-free trough area. An ε value of more than ε_{max} will also lead to too extensive a depleted region, while a value below ε_{min} results in high levels of folate in the trough area.

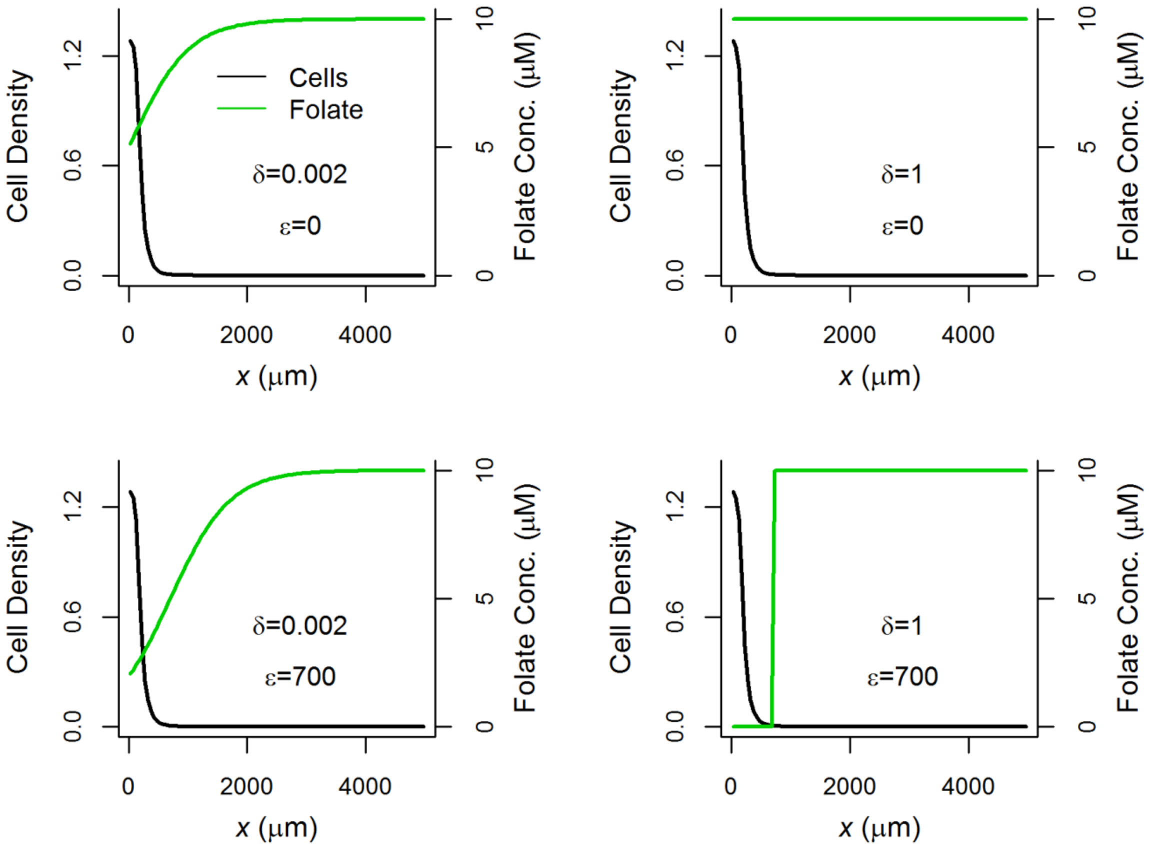

Figure A.1.1: Extremes that the initial folate distribution was permitted to take during model fitting. Green lines show the initial attractant distributions calculated from equation (A.1.8) using each combination of the maximum and minimum values of the parameters δ and ε. Black lines show the initial cell distribution obtained by logspline density estimation.

A.1.3. Cell Division

In both the *Dictyostelium* and melanoma datasets, the number of cells in the observation region increased substantially over time, primarily as a result of cells moving into the region across the left boundary (Figs A.1.2A & A.1.3A). A second contributor to increasing cell numbers is cell division. For *Dictyostelium*, where the assay was run over a relatively short time interval (5.5 hours), cell division is a very minor contributor, and can reasonably be ignored. I, therefore, set the cell division rate ν of *Dictyostelium* to zero, and assumed that all increases in cell number were a result of cell movements across the left boundary (see Appendix A.1.4 for details). For melanoma, however, where the time interval of interest spanned 50 hours, cell division had a larger impact on the cell distribution, such that ignoring it did not give good agreement between models and data; attributing all changes in cell number to movements led to modelled cell densities that were too high at the boundary of the region. From the microscopy images, it was observed that the influx of cells over the region's left boundary ceased by $t = 30$, and, since any subsequent increases in cell number can be assumed to result from cell division, I estimated $\nu = 0.004$ for melanoma by fitting an exponential curve to the data from $t = 30$ onwards (Fig A.1.3A).

A.1.4. Boundary conditions

For both of the datasets it was necessary to account for movements of cells into the regions of interest across the left boundaries by incorporating appropriate boundary conditions into the models. To achieve this, I first took the time series:

$$S = \left\{ n_j : j \in (1, \ldots, T) \right\} \tag{A.1.9}$$

for each dataset, where n_j is the number of cells observed at time point $j \in (1,...,T)$. I then used these data, as outlined in Figs A.1.2-3, to estimate smooth functions $N'(t)$ describing the rates at which the numbers of cells in the regions of interest increased over time. It can be assumed that these increases in cell numbers resulted from just two processes; movements across the region's left boundary (all cells began the assays to the left of the observation region) and cell division (see Appendix A.1.3). A reasonable left boundary condition would, therefore, be:

$$Flux_C^{0,1} = N'(t) - \nu \int_0^l C(x,t)\,dx \tag{A.1.10}$$

where $Flux_C^{0,1}$ is the cell flux across the left boundary of the region. For *Dictyostelium*, given the choice of $\nu = 0$, equation (A.1.10) reduces to:

$$Flux_C^{0,1} = N'(t) \tag{A.1.11}$$

while, for melanoma, as no cells cross the left boundary after $t = 30$, we have:

$$Flux_C^{0,1} = \begin{cases} N'(t) - \nu \int_0^l C(x,t)\,dx & \text{if } t \in (0,30) \\ 0 & \text{if } t > 30 \end{cases} \tag{A.1.12}$$

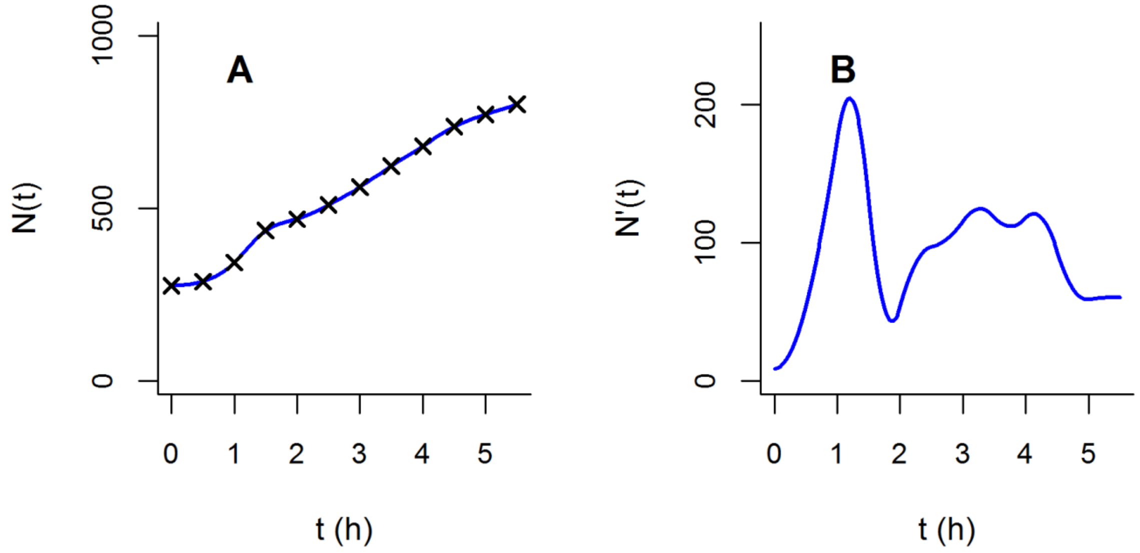

Figure A.1.2: Changes in the number of *Dictyostelium* cells in the region of interest over time. A) Numbers of *Dictyostelium* cells observed in microscopy images at half-hourly intervals (black crosses), interpolated using a cubic spline *N*(*t*) (blue line). **B)** Derivative of the spline fitted in **A**. This curve was used to define realistic boundary conditions for the cells (see Appendix A.1.4).

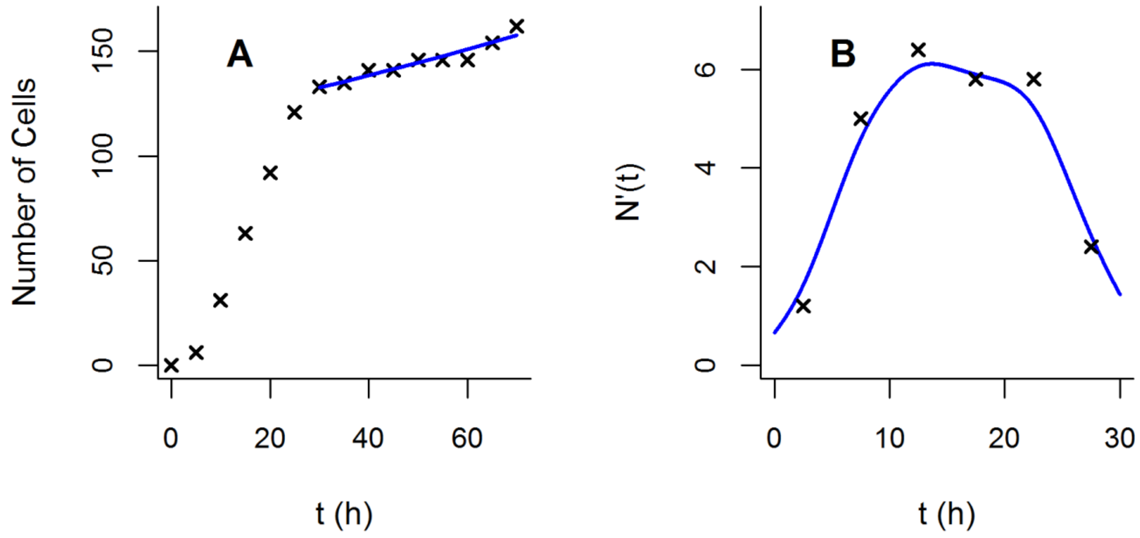

Figure A.1.3: Changes in the number of melanoma cells in the region of interest over time. A) Numbers of melanoma cells observed in microscopy images at five-hourly intervals (black crosses). The blue line shows the exponential curve fitted to the data from $t = 30$ in order to estimate the rate of population growth through cell division ν. **B)** Crosses show finite difference approximations of the rate of change in cell numbers during the interval from $t = 0$ to $t = 30$, calculated from the data shown in **A.** The blue line shows the nonparametric regression curve *N'*(*t*) fitted to the points using the sm package in R (Bowman and Azzalini 2014). This curve was used to define realistic boundary conditions for the cells (see Appendix

A.1.4). As no new cells entered the region across the left boundary after $t = 30$, extending $N'(t)$ beyond this point was unnecessary.

In both datasets, no cells crossed the right boundary during the time period considered, so I applied a zero-flux boundary condition:

$$Flux_C^{B,B+1} = 0 \tag{A.1.13}$$

where B is the total number of boxes making up the discretised spatial region. This condition prevents any loss or gain of cell density across this boundary.

A reasonable assumption that I make for the boundary conditions for the attractants (folate and LPA) is that the flux across each region boundary equals the flux across the nearest internal box boundary in the spatial discretisation:

$$Flux_A^{0,1}(t) = Flux_A^{1,2}(t) \tag{A.1.14}$$

$$Flux_A^{B,B+1}(t) = Flux_A^{B-1,B}(t) \tag{A.1.15}$$

A.2. Weighted log-likelihood maximisation

Numerical solution of the PDE models using the method of lines (Appendix A.1.1) introduces error through discretisation of the models in space and time. This numerical error in the model solution leads to noise in the computation of the derivatives of the weighted log-likelihood (equation (2.12)) with respect to the parameters (via difference quotients). If the parameter difference is sufficiently large, corresponding to a low resolution representation, this numerical noise tends to average out and the weighted log-likelihood appears to be smooth (top row of Fig. A.2.1). However, if the difference is small, corresponding to a higher resolution representation, the numerical noise does not average out and spurious low-magnitude high-frequency oscillations are observed (bottom row of Fig. A.2.1). These numerical artefacts in the weighted log-likelihood surface cause problems for parameter inference by trapping optimisation algorithms that seek to maximise this function.

When fitting the models by the maximum weighted log-likelihood, I introduced steps to deal with the problem of numerical instabilities leading to optimisers becoming trapped on local optima. These involved first attempting to get close to the global optimum for each model by running 200 optimisations from random initial parameter sets using an optimiser (I found that the quasi-Newton BFGS algorithm performed well for the *Dictyostelium* dataset, while the Nelder-Mead algorithm was more effective at reaching high weighted log-likelihood parameter regions for the melanoma dataset). From these 200 optimisations I retained only the one that gave the highest weighted log-likelihood. One-dimensional profile weighted log-likelihood plots around these best parameter sets (using a low enough resolution for each parameter to obtain a smooth weighted log-likelihood profile) were then used to determine whether the weighted log-likelihood was actually at a maximum at the optimised value for each parameter. If the parameters had not been fully optimised, I adjusted one of the parameters that was furthest from its optimal position (selected based on the weighted log-likelihood plots) to an improved position. A re-optimisation of the full parameter set was then implemented. This process of parameter adjustment and re-optimisation was continued until re-plotting the weighted log-likelihood profiles showed that a maximum had been reached for all parameters (Fig. A.2.2), indicating that the maximum weighted log-likelihood had been reached. Model comparison using AIC$_\mathrm{C}$ (the Akaike Information Criterion corrected for small sample sizes (Akaike 1974, Hurvich and Tsai 1989)) and BIC (Bayesian Information Criterion (Schwarz 1978)) could then be carried out by calculating these statistics for each model as:

$$\mathrm{AIC_C} = -2\log\tilde{L}^* + 2k + \frac{2k(k+1)}{n-k-1} \qquad \text{(A.2.1)}$$

$$\mathrm{BIC} = -2\log\tilde{L}^* + k\log n \qquad \text{(A.2.2)}$$

where $\log\tilde{L}^*$ is the maximum weighted log-likelihood and k is the number of model parameters. These statistics reward models based on their fit to the data, indicated by $\log\tilde{L}^*$, and apply a complexity penalty based on k, on the assumption that all parameters are well-determined by the data.

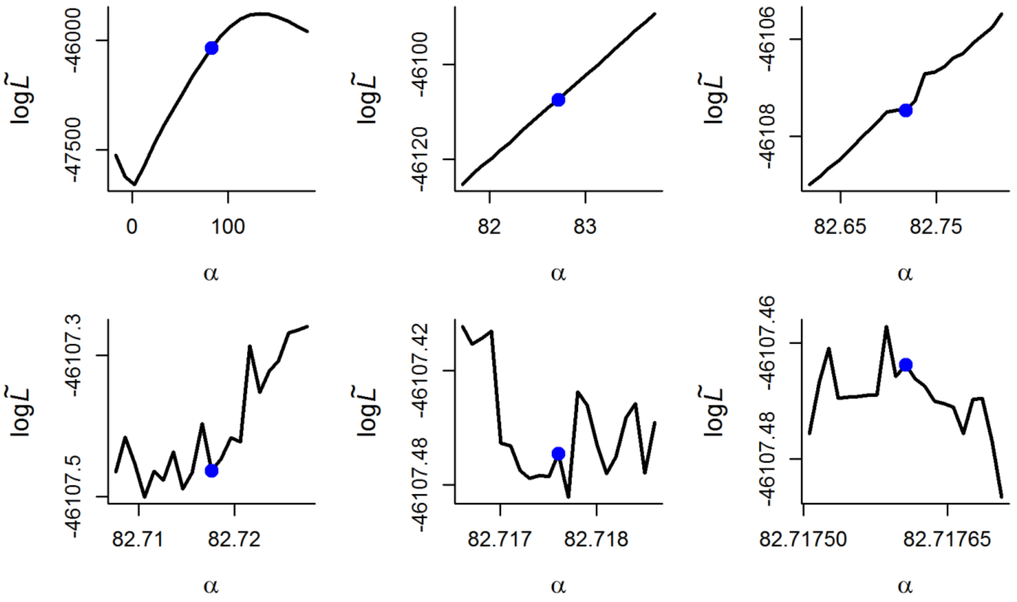

Figure A.2.1: Numerical error in the likelihood surface. One-dimensional plots of the weighted log-likelihood (equation (2.12)) against a parameter α at different resolutions. The value to which the parameter was optimised on one run of the quasi-Newton BFGS optimisation algorithm is marked with a point. Note that the optimiser has failed to reach the maximum likelihood value, and become trapped on a local optimum instead. These local optima are artefacts of the numerical noise inherent in the discretisation of the PDEs, and only appear at high resolution (i.e. when making small changes in the parameter values).

This weighted log-likelihood maximisation procedure is very effective for obtaining a reliable estimate of the optimal parameters. However, the reliance of this method on visual inspections of the profile weighted log-likelihood and manual parameter adjustments make it labour intensive. In addition, this method does not produce an estimate of the posterior distribution of the parameters, making it difficult to assess parameter uncertainty, and restricting access to more advanced model comparison statistics like WAIC (Watanabe 2010). For these reasons, I only relied on model inference using the maximum weighted log-likelihood during selection of the degrees of the polynomial functions describing the time-varying parameters (Tables A.6.1-2), and when determining the relative importance of the time-variance in each parameter in the best model for each dataset (Tables A.6.5-6). When carrying out the more important task of comparing the full set of candidate models for each dataset, I applied the inference scheme described in section 2.5 of the main text, which involved the development of a pseudo-posterior through multiple optimisations on bootstrap samples of the data, and thus allowed the calculation of WAIC. While this bootstrapping method allows a more advanced model comparison, it does incur high computational costs, which is why, in the face of limited cluster resources with which to parallelise this procedure, I resorted to the computationally cheaper weighted log-likelihood maximisation for the more minor model comparisons. While WAIC should be preferred as the more reliable statistic, I did also compare the full set of models using AICc and BIC to check for consistency between these statistics (Tables A.6.3-4).

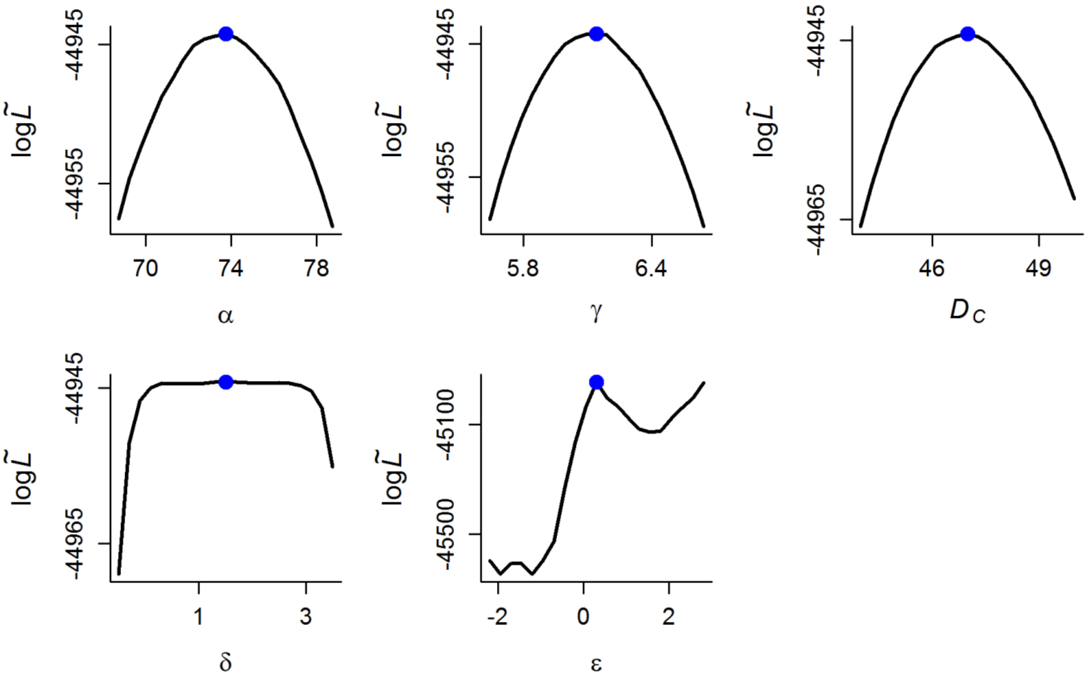

Figure A.2.2: Sufficient optimisation of the model parameters. One-dimensional plots of the weighted log-likelihood (equation (2.12)) landscape around the parameters for one of the models following sufficient optimisation. For each parameter, the resolution was selected to be low enough to give a visually relatively smooth likelihood surface. Note that all parameters have now been optimised to a true peak in the likelihood surface (compare with Fig. A.2.1).

A.3. Eliminating bimodality in the pseudo-posterior

The inference method involving multiple optimisations on many bootstrap samples of the data (see section 2.5) resulted in the production of a pseudo-posterior for each model. For both datasets, bimodality was observed in the pseudo-posteriors for all models except the simple diffusion model (Figs A.3.1-2). This bimodality is a result of the presence of local optima, which cause some optimisations to become trapped before they reach the maximum likelihood parameters. For both datasets, the positions of the lower-likelihood peaks in the posteriors of the more complex models roughly correspond to the position of the single likelihood peak that occurs for the diffusion model. This suggests that these lower-likelihood peaks are made up of optimisations that failed to properly fit the parameters describing the self-generated attractant gradient mechanism; a suggestion that is backed up by the fact that model outputs obtained by sampling from these lower peaks closely resembled those obtained from a diffusion-only scenario (shown in Figs A.7.1-2). The presence of these low-likelihood peaks in the pseudo-posteriors will affect the values of model comparison statistics calculated from these pseudo-posteriors, potentially influencing model rankings. I, therefore, chose to isolate and use only the highest-likelihood peak when evaluating the models. This was achieved for each dataset by introducing a cut-off value in the log-likelihood for all the models except the diffusion model, which was positioned in the trough between the two peaks. Any optimisations that achieved a log-likelihood that was lower than this cut-off were discarded, and only the remaining optimisations (indicated by the blue shaded areas in Figs A.3.1-2) were used in the calculation of model comparison statistics (see section 2.5 of the main text).

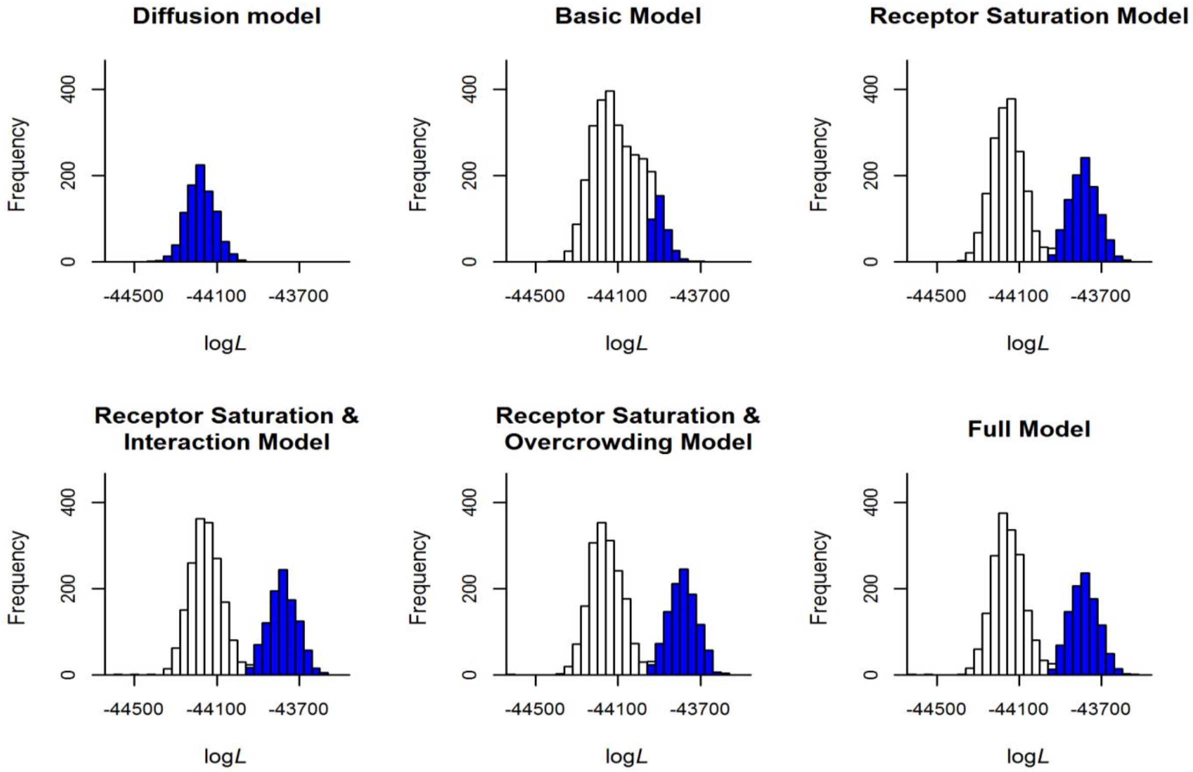

Figure A.3.1: Histograms of the pseudo-posteriors produced by multiple optimisations of each model on bootstrap samples of the *Dictyostelium* data. Note that all pseudo-posteriors except that for the diffusion model exhibit bimodality (though the two peaks are fused in the case of the basic model). For all models except the diffusion model, I introduced a cut-off of $\log L = -43940$ to isolate the upper peak in the likelihood. The blue shaded areas illustrate the shapes of the pseudo-posteriors after imposing this cut-off.

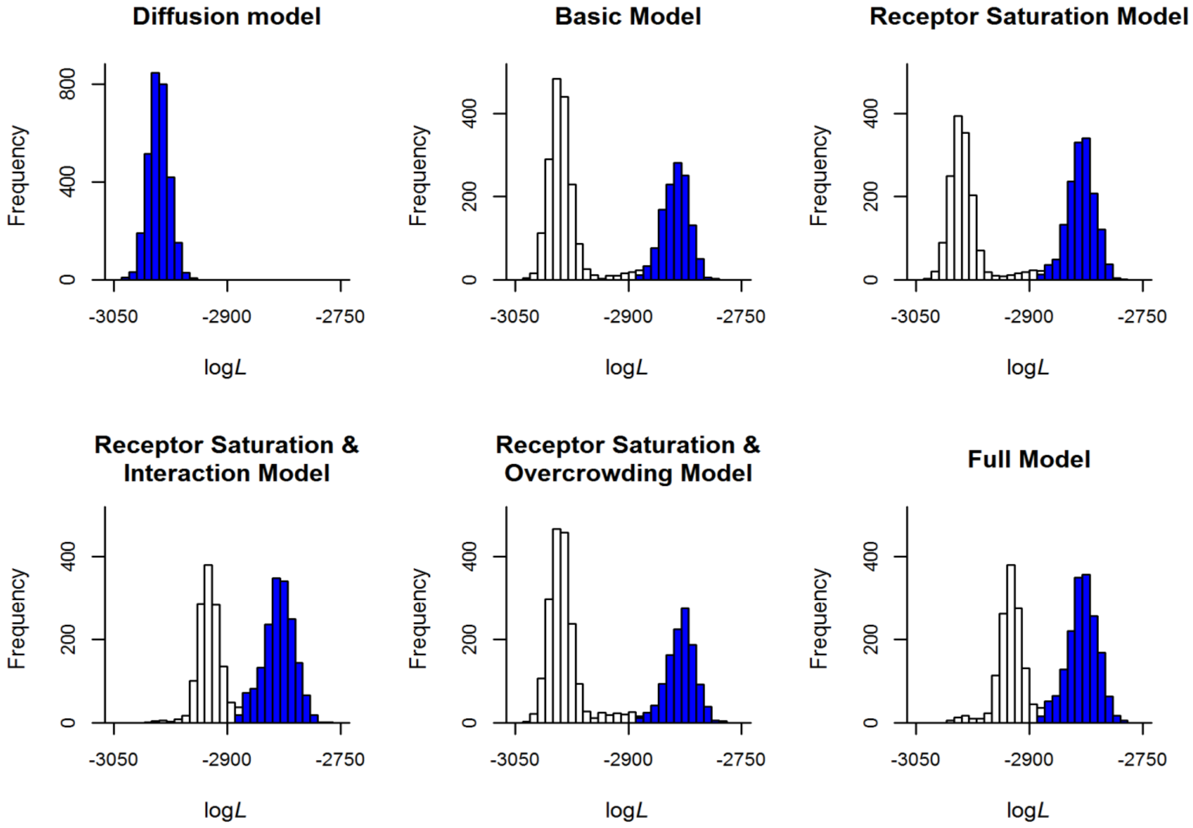

Figure A.3.2: Histograms of the pseudo-posteriors produced by multiple optimisations of each model on bootstrap samples of the melanoma data. Note that all pseudo-posteriors except that for the diffusion model exhibit bimodality. For all models except the diffusion model, I introduced a cut-off of $\log L = -2885$ to isolate the upper peak in the likelihood. The blue shaded areas illustrate the shapes of the pseudo-posteriors after imposing this cut-off.

A.4. Validation of WAIC calculated using a pseudo-posterior

A.4.1. Background

The various candidate cell movement models described in section 2.3 were compared using WAIC values calculated using a pseudo-posterior that was obtained by fitting the models to many bootstrap datasets (section 2.5). To verify whether this method produces results comparable to sampling from a true posterior, I carried out an additional study using the radiocarbon dataset from the sm package in R (Bowman and Azzalini 2014), which describes the radiocarbon age of Irish oak in comparison to its true calendar age. This involved comparing the fits of polynomial models of degrees one to nine (Fig. A.4.1) using DIC (Deviance Information Criterion (Spiegelhalter et al. 2002)) and WAIC values calculated from the true posterior and from the pseudo-posterior obtained by the bootstrapping method. Note that I have not compared the models based on DIC in the main text, since I encountered issues with negative values being estimated for the effective number of parameters (a known issue with this comparison statistic), rendering DIC less reliable than the more recently developed WAIC.

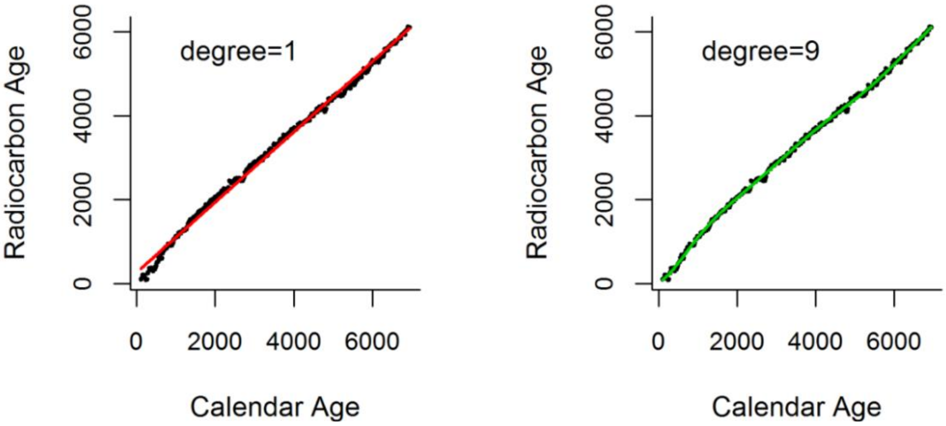

Figure A.4.1: Fits of polynomials of degrees one and nine to the radiocarbon dataset

A.4.2. Calculation of DIC and WAIC from the true posterior

The polynomial models fitted to the data take the form:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{A.4.1}$$

where $\mathbf{y} = (y_1, ..., y_n)^T$ is the vector of radiocarbon age observations (*n*=343), $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^T$ is the vector of coefficients (*k* is equal to the degree of the polynomial plus one) and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^T$ is iid (independent and identically distributed) Gaussian error, with mean zero and variance σ^2. For each model considered, σ^2 was estimated by fitting to the data and calculating the variance of the residuals. The design matrix $\mathbf{B}$ is given by:

$$\mathbf{B} = \begin{pmatrix} x_1^{\,0} & \cdots & x_1^{\,(k-1)} \\ \vdots & \ddots & \vdots \\ x_n^{\,0} & \dots & x_n^{\,(k-1)} \end{pmatrix} \tag{A.4.2}$$

where $\mathbf{x} = \left(x_1, \dots, x_n\right)^T$ is the calendar age covariate.

Gaussian priors with mean zero and variance ζ^2 were applied to each of the parameters. I specified vague prior distributions where $\zeta^2 = 1 \times 10^6$. The likelihood is given by:

$$P\left(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\beta}, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left\{\frac{-\left(\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\right)^T \left(\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\right)}{2\sigma^2}\right\} \tag{A.4.3}$$

As the priors and likelihood are Gaussian distributions, the posterior is Gaussian also, and is given by:

$$P\left(\boldsymbol{\beta} \middle| \mathbf{x}, \mathbf{y}, \sigma^2, \zeta^2\right) = N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \tag{A.4.4}$$

where

$$\boldsymbol{\mu} = \left(\mathbf{B}^T\mathbf{B} + \frac{\sigma^2}{\zeta^2}\mathbf{I}\right)^{-1}\mathbf{B}^T\mathbf{y} \tag{A.4.5}$$

$$\boldsymbol{\Sigma} = \sigma^2\left(\mathbf{B}^T\mathbf{B} + \frac{\sigma^2}{\zeta^2}\mathbf{I}\right)^{-1} \tag{A.4.6}$$

I drew a sample of m=20,000 parameter sets $(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)$ from the posterior distribution for each model and, using the likelihood function stated above (equation (A.4.3)), calculated the DIC as:

$$\mathrm{DIC} = \frac{2\sum_{i=1}^{m} -2\log\left\{P\left(\mathbf{y}\,|\,\mathbf{x},\boldsymbol{\beta}_i,\sigma^2\right)\right\}}{m} + 2\log\left\{P\left(\mathbf{y}\,|\,\mathbf{x},\bar{\boldsymbol{\beta}},\sigma^2\right)\right\} \tag{A.4.7}$$

where $\bar{\boldsymbol{\beta}}$ are the mean values of the parameters, and the WAIC as:

$$\begin{aligned}\mathrm{WAIC} = &-2\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(y_j\,|\,x_j,\boldsymbol{\beta}_i,\sigma^2\right)\right\}\\ &+2\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(y_j\,|\,x_j,\boldsymbol{\beta}_i,\sigma^2\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(y_j\,|\,x_j,\boldsymbol{\beta}_i,\sigma^2\right)\right\}\right]^2\right\}\end{aligned} \tag{A.4.8}$$

A.4.3. Calculation of DIC and WAIC from bootstrap samples

The data were sampled with replacement to generate m=20,000 bootstrap datasets of the same dimension n as the original dataset, each consisting of a vector of radiocarbon age observations $\mathbf{r}_i = (r_{i,1},...,r_{i,n})^T$ and the associated calendar ages $\mathbf{q}_i = (q_{i,1},...,q_{i,n})^T$, where $i \in (1,\ldots,m)$. Since I chose a vague prior for the regression parameters (i.e. with a large value of the variance hyperparameter ζ^2 in equations (A.4.4-6), maximum likelihood parameter estimates will be effectively the same as maximum a posteriori estimates. I, therefore, fitted the nine polynomial models to each of the bootstrap datasets using maximum likelihood to obtain a sample of parameter sets $(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m)$ that are taken as an approximation of a posterior distribution. This ‘pseudo-posterior’ can be used to estimate the DIC and WAIC in two alternative ways. The first uses the parameter sets obtained from the bootstrap data (and their mean $\bar{\boldsymbol{\theta}}$), with only the true un-bootstrapped data as follows:

$$\mathrm{DIC}_{\mathrm{A}} = \frac{2\sum_{i=1}^{m} -2\log\left\{P\left(\mathbf{y}\,|\,\mathbf{x},\boldsymbol{\theta}_i,\sigma^2\right)\right\}}{m} + 2\log\left\{P\left(\mathbf{y}\,|\,\mathbf{x},\bar{\boldsymbol{\theta}},\sigma^2\right)\right\} \tag{A.4.9}$$

$$\begin{aligned}\mathrm{WAIC_A} &= -2\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(y_j \mid x_j,\boldsymbol{\theta}_i,\sigma^2\right)\right\}\\ &+2\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(y_j \mid x_j,\boldsymbol{\theta}_i,\sigma^2\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(y_j \mid x_j,\boldsymbol{\theta}_i,\sigma^2\right)\right\}\right]^2\right\}\end{aligned} \tag{A.4.10}$$

In the second method I incorporate the bootstrap data into the calculations:

$$\mathrm{DIC_B} = \frac{2\sum_{i=1}^{m}-2\log\left\{P\left(\mathbf{r}_i \mid \mathbf{q}_i,\boldsymbol{\theta}_i,\sigma^2\right)\right\}}{m}+2\log\left\{P\left(\mathbf{y} \mid \mathbf{x},\bar{\boldsymbol{\theta}},\sigma^2\right)\right\} \tag{A.4.11}$$

$$\begin{aligned}\mathrm{WAIC_B} &= -2\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(r_{i,j} \mid q_{i,j},\boldsymbol{\theta}_i,\sigma^2\right)\right\}\\ &+2\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(r_{i,j} \mid q_{i,j},\boldsymbol{\theta}_{i,i},\sigma^2\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(r_{i,j} \mid q_{i,j},\boldsymbol{\theta}_{i,i},\sigma^2\right)\right\}\right]^2\right\}\end{aligned} \tag{A.4.12}$$

A.4.4. Calculation of AICc and BIC

For each polynomial model considered, I also calculated two more basic model selection criteria, AICc and BIC, which do not account for parameter uncertainty and tend to select models that over-fit and under-fit the data respectively (Ripplinger and Sullivan 2008). This involved first finding the maximum likelihood parameters $\boldsymbol{\beta}^*$ for each model, given by

$$\boldsymbol{\beta}^* = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{A.4.13}$$

and then finding the maximum likelihood L^* by inserting these parameters into equation (A.4.3). AICc and BIC could then be calculated as

$$\mathrm{AIC_C} = -2\log L^* + 2k + \frac{2k\left(k+1\right)}{n-k-1} \tag{A.4.13}$$

$$\mathrm{BIC} = -2\log L^* + k\log n \tag{A.4.14}$$

where k is the number of model parameters.

A.4.5. Results & discussion

I found a generally strong correspondence between the standard DIC and WAIC values and the approximations from bootstrap sampling (Fig. A.4.2), with correlation coefficients in excess of 0.999 for all relationships, except that between the standard WAIC and WAIC$_\mathrm{B}$, which had a correlation coefficient of 0.98. The standard DIC and WAIC both select the eighth degree polynomial from the nine candidate models (Table A.4.1). DIC_A and WAIC_A are successful in selecting this same best model, and in exactly replicating the full model ranking observed for the standard DIC and WAIC, suggesting that these statistics are a valid approximation of the standard DIC and WAIC. DIC_B and WAIC_B were less successful, showing a preference for the ninth degree polynomial. Based on these results, I chose to use the WAIC_A approximation when comparing the models of cell movement (equation (2.13)).

While WAIC and DIC estimated from the optimisations on bootstrap samples of the data have been shown in this case to give a good approximation to the same statistics estimated from the true posterior, it should be noted that AICc and BIC, which are typically less reliable model comparison statistics, also showed good agreement with the standard DIC and WAIC (Table A.4.1). Both AICc and BIC selected the same best model as DIC and WAIC, and AICc successfully reproduces the full model ranking (there are some inconsistencies in the ranking by BIC). This suggests that, for this particular study, little accuracy in model selection was gained by calculating DIC and WAIC from the optimisations on bootstrap samples; a simpler analysis based on AICc or BIC would have been just as effective in selecting the correct model. Ultimately, further testing of the bootstrapping method is required in cases where AICc and BIC fail to give the right answer, so as to verify whether this method provides any improvement over these more basic comparison statistics.

Table A.4.1: Model comparison statistics for each polynomial model fitted to the radiocarbon data (Fig. A.4.1). The standard values of DIC and WAIC are calculated using the true posterior, while the alternative estimates are obtained through the bootstrapping technique. The best model based on each statistic is indicated by *.

Degree	Standard DIC	DIC_A	DIC_B	Standard WAIC	WAIC_A	WAIC_B	AICc	BIC
1	-927	-926	-936	-927	-925	-556	-143	-136
2	-1065	-1063	-1080	-1064	-1061	-850	-701	-689
3	-1409	-1408	-1425	-1408	-1406	-1106	-1403	-1388
4	-1407	-1406	-1427	-1407	-1404	-1112	-1402	-1383
5	-1412	-1410	-1438	-1411	-1404	-1125	-1408	-1386
6	-1411	-1407	-1441	-1409	-1393	-1124	-1407	-1381
7	-1448	-1445	-1481	-1447	-1439	-1184	-1448	-1417
8	-1457*	-1456*	-1493	-1457*	-1452*	-1213	-1457*	-1423*
9	-1456	-1454	-1495*	-1456	-1446	-1219*	-1455	-1418

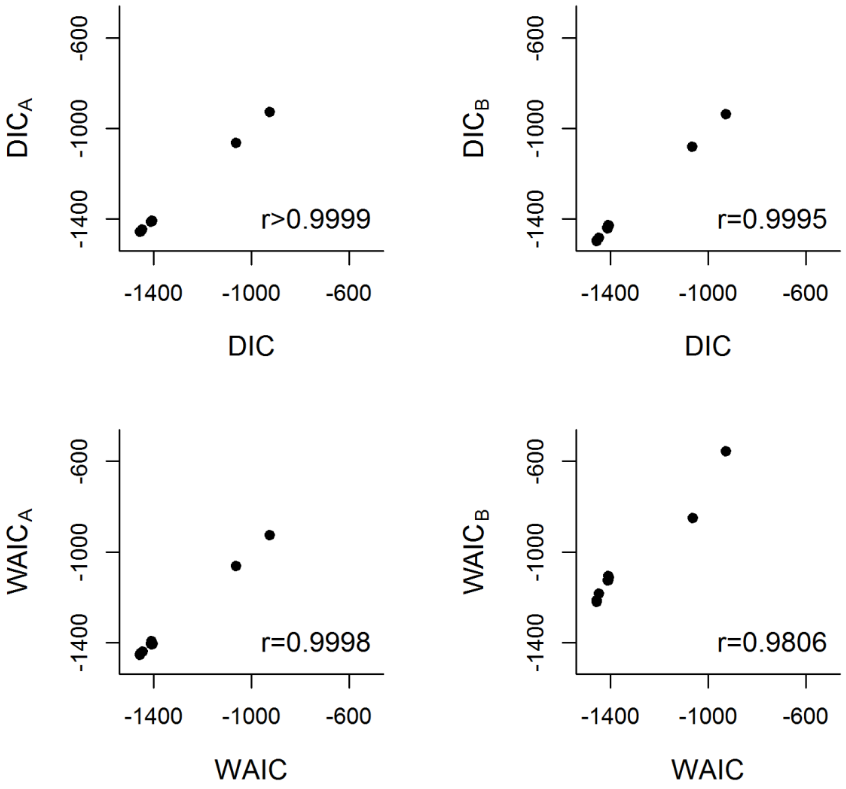

Figure A.4.2: Comparison of DIC and WAIC values calculated using bootstrapping and standard approaches. Plots of the DIC and WAIC approximations obtained for the nine polynomial models through bootstrapping against the standard DIC and WAIC values obtained for the models by direct sampling from the posterior. The value of Pearson's correlation coefficient is indicated for each comparison.

A.5. Calculation of standard errors for WAIC

WAIC was calculated for the cell movement models as:

$$\begin{aligned} \text{WAIC} &= -2\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\} \\ &\quad +2\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right\} \end{aligned} \qquad (\text{A.5.1})$$

To calculate the variance of the first term, I first obtained the variances of the mean likelihoods of each observation $\left(x_j, t_j\right)$ using:

$$\begin{aligned} \operatorname{var}\left(\frac{1}{m}\sum_{i=1}^{m}P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right) &= \frac{1}{m}\operatorname{var}\left(P\left(x_j \mid t_j,\boldsymbol{\theta}\right)\right) \\ &= \frac{1}{m}\left(\frac{1}{m}\sum_{i=1}^{m}\left[\left\{P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)-\frac{1}{m}\sum_{i=1}^{m}P\left(x_j \mid t_j,\boldsymbol{\theta}_i\right)\right\}^2\right]\right) \end{aligned} \qquad (\text{A.5.2})$$

The univariate delta method was then applied to get the variances of the log mean likelihoods as:

$$\operatorname{var}\left(\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right)=\left(\frac{1}{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)}\right)^2\operatorname{var}\left(\frac{1}{m}\sum_{i=1}^{m}P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right)\quad\text{(A.5.3)}$$

and the variance of the sum of the log mean likelihoods was obtained as:

$$\operatorname{var}\left(\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right)=\sum_{j=1}^{n}\operatorname{var}\left(\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right)\qquad\text{(A.5.4)}$$

The variance of the second term in the WAIC was obtained by first calculating the variance of the sample variance of the log likelihood of each observation $\left(x_j,t_j\right)$ as:

$$\operatorname{var}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right\}=\frac{\left(m-1\right)^2}{m^3}\mu_4-\frac{\left(m-1\right)\left(m-3\right)\mu_2^2}{m^3}\quad\text{(A.5.5)}$$

where μ_2 and μ_4 are the 2nd and 4th central moments of $\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}$, calculated by:

$$\begin{aligned}\mu_2&=\frac{1}{m}\sum_{i=1}^{m}\left(\left[\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}-\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)\\ \mu_4&=\frac{1}{m}\sum_{i=1}^{m}\left(\left[\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}-\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^4\right)\end{aligned}\qquad\text{(A.5.6)}$$

These variances of sample variances are then summed to get:

$$\begin{aligned}&\operatorname{var}\left[\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right\}\right]\\ &=\sum_{j=1}^{n}\operatorname{var}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right\}\end{aligned}\qquad\text{(A.5.7)}$$

The standard error of the full WAIC can be obtained as:

$$\begin{aligned}\operatorname{se}\left(\text{WAIC}\right)=2&\left(\operatorname{var}\left(\sum_{j=1}^{n}\log\left\{\frac{1}{m}\sum_{i=1}^{m}P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right)+\right.\\ &\left.\operatorname{var}\left[\sum_{j=1}^{n}\left\{\frac{1}{m}\left(\sum_{i=1}^{m}\left[\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right)-\left[\frac{1}{m}\sum_{i=1}^{m}\log\left\{P\left(x_j\,|\,t_j,\boldsymbol{\theta}_i\right)\right\}\right]^2\right\}\right]\right)^{\frac{1}{2}}\end{aligned}\qquad\text{(A.5.8)}$$

A.6. Supplementary tables

Table A.6.1: Selection of the degree of the polynomials used to describe the time-varying parameters for *Dictyostelium*. Values of the statistics are based on fits of the full model (equation (2.8)) with different polynomial degrees. Both AICc and BIC show a preference for a degree of three. Based on these results, I used a polynomial degree of three when fitting the remaining models to this dataset (see Table 2.1 in the main text).

Degree	**log$\tilde{L}^{*}$**	**AICc**	**BIC**
0	-44114.5	88249.0	88316.3
1	-43929.2	87886.5	87980.8
2	-43792.8	87621.7	87742.8
3	-43771.0	87586.1*	87734.1*
4	-43771.0	87594.1	87769.0

Table A.6.2: Selection of the degree of the polynomials used to describe the time-varying parameters for melanoma. Values of the statistics are based on fits of the full model (equation (2.8)) with different polynomial degrees. AICc shows a strong preference for a degree of one, while BIC (a comparison statistic known for its tendency to select models that are overly simple (Ripplinger and Sullivan 2008)) shows only a slight preference for a degree of zero (i.e. no time variance). Based on these results, I used a polynomial degree of one when fitting the remaining models to this dataset (see Table 2.1 in the main text).

Degree	**log$\tilde{L}^{*}$**	**AICc**	**BIC**
0	-2850.5	5717.3	5751.4*
1	-2838.5	5701.5*	5752.5
2	-2837.7	5708.5	5776.2

Table A.6.3: AICc and BIC based comparisons of the six models fitted to the *Dictyostelium* data. The model comparison statistics were calculated using the maximum weighted log-likelihood fits (Appendix A.2).

Model	**AICc**	**BIC**
Diffusion	88356.81	88383.75
Basic	87831.87	87932.83
Receptor Saturation	87587.29	87694.98*
Receptor Saturation & Interaction	87584.05*	87725.36
Receptor Saturation & Overcrowding	87589.00	87703.41
Full	87586.06	87734.09

Table A.6.4: AICc and BIC based comparisons of the six models fitted to the melanoma data. The model comparison statistics were calculated using the maximum weighted log-likelihood fits (Appendix A.2).

Model	AICc	BIC
Diffusion	6003.3	6011.9
Basic	5711.9	5741.8
Receptor Saturation	5701.1*	5735.3*
Receptor Saturation & Interaction	5702.1	5748.9
Receptor Saturation & Overcrowding	5703.2	5741.5
Full	5701.5	5752.5

Table A.6.5: Consequences of removing the time-variance in the parameters of the receptor saturation model fitted to the *Dictyostelium* dataset. The receptor saturation model was the best model for this dataset based on WAIC (Table 2.1). Removing variation in α gives poorer (higher) values of AICc and BIC, while removing variation in D_C improves BIC but gives a poorer AICc. Making γ constant improves BIC and has little effect on AICc.

Time-varying parameters	log$\tilde{L}^*$	AICc	BIC
α,γ,D_C	-43777.6	87587.3*	87695.0
γ,D_C	-43823.5	87673.0	87760.6
α,D_C	-43780.9	87588.0	87675.5*
α,γ	-43783.7	87593.4	87680.9
α	-43830.6	87681.3	87748.6
γ	-43853.0	87726.0	87793.4
D_C	-44094.8	88209.6	88276.9
none	-44120.2	88256.4	88310.3

Table A.6.6: Consequences of removing the time-variance in the parameters of the receptor saturation and overcrowding model fitted to the melanoma dataset. The receptor saturation and overcrowding model was the best model for this dataset based on WAIC; Table 2.1). Note that there is virtually no change in the maximum weighted log-likelihood provided that α is retained as a time-varying parameter. There is also no increase in either AICc or BIC unless both α and γ are removed as time-varying parameters, suggesting that these two parameters are able to compensate for one another to some degree.

Time-varying parameters	log$\tilde{L}^*$	AICc	BIC
α,γ,D_C	-2842.4	5703.2	5741.5
γ,D_C	-2843.5	5703.3	5737.4
α,D_C	-2842.4	5701.2	5735.3
α,γ	-2842.4	5701.2	5735.3
α	-2842.4	5699.1*	5729.0*
γ	-2843.9	5702.0	5731.9
D_C	-2849.1	5712.3	5742.2
none	-2851.7	5715.5	5741.1

A.7. Supplementary figures

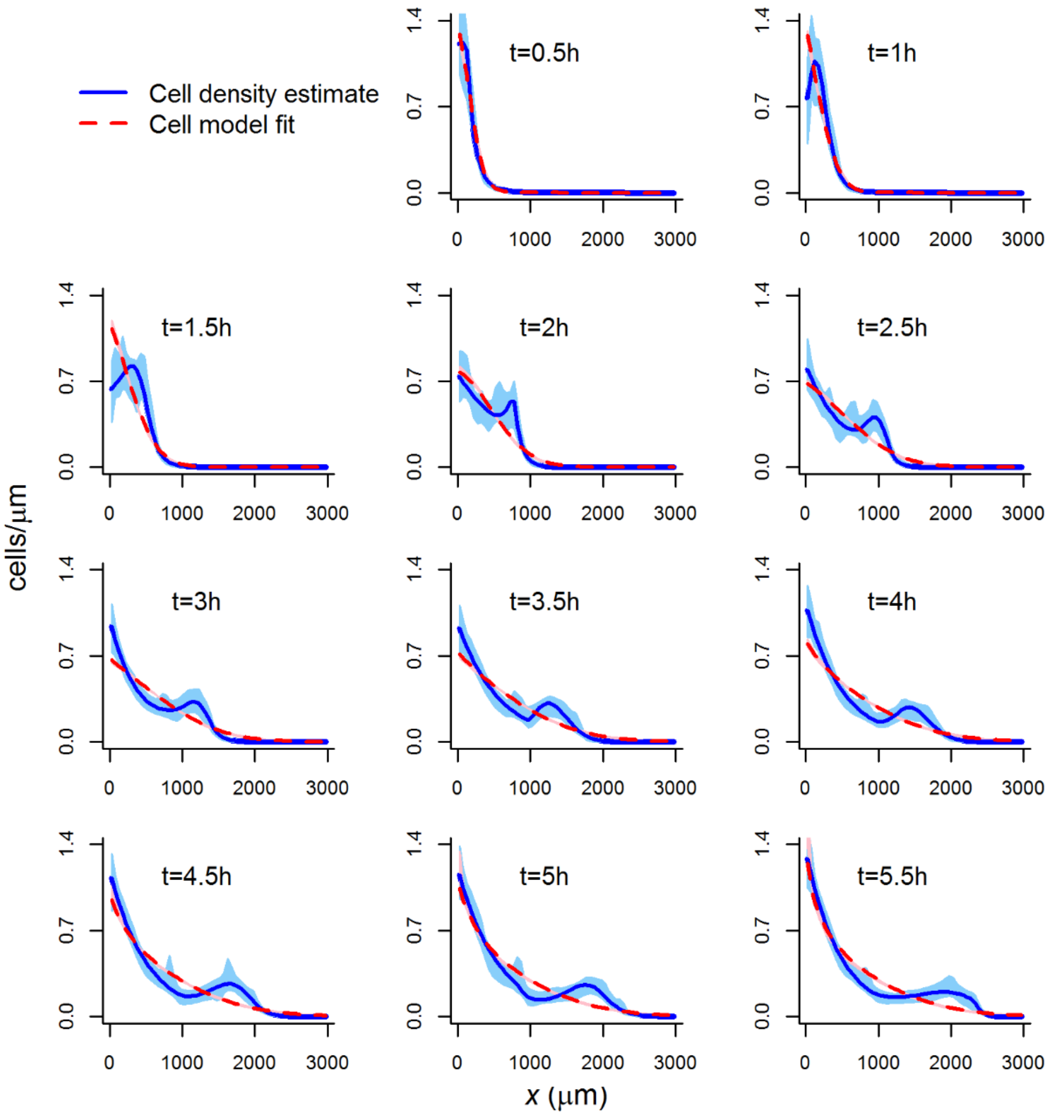

Figure A.7.1: Diffusion model fitted to the *Dictyostelium* data. Dashed red lines show *Dictyostelium* cell distributions at half-hour intervals produced by the diffusion model (equation (2.2)) using the optimised parameters from the bootstrap optimisation that gave the highest value of the weighted log-likelihood (equation (2.12)). Pink shaded areas show the 95 percentile interval for the modelled cell densities, based on 200 samples from the pseudo-posterior. Cell distributions obtained from the data using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) are shown by blue lines, with 95 percentile intervals obtained using 10,000 bootstrap samples of the data indicated by blue shaded areas.

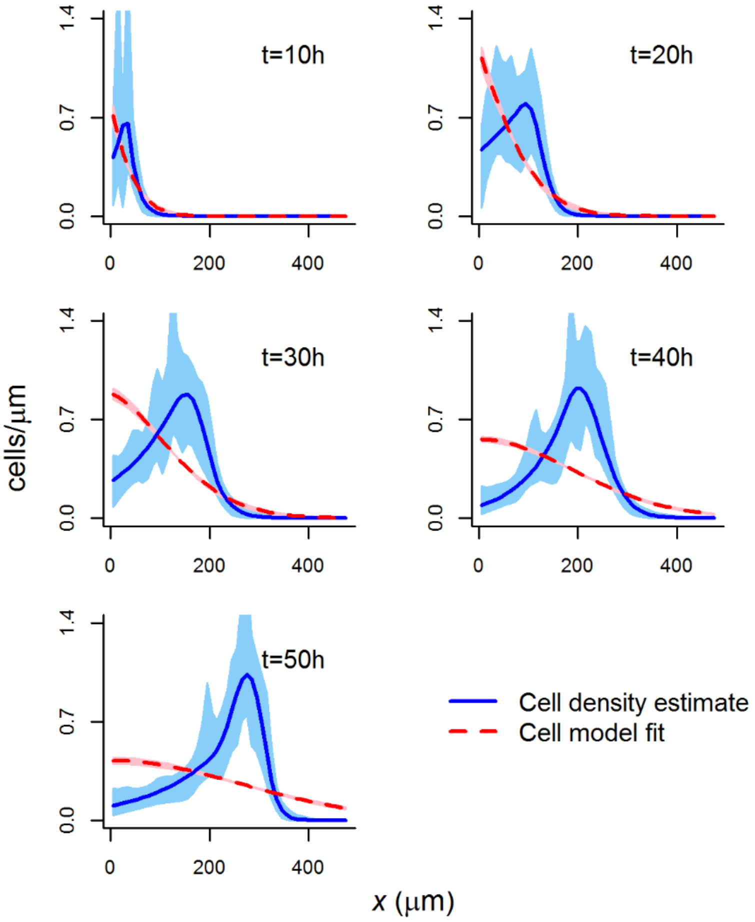

Figure A.7.2: Diffusion model fitted to the melanoma data. Dashed red lines show melanoma cell distributions at 10-hour intervals produced by the diffusion model (equation (2.2)) using the optimised parameters from the bootstrap optimisation that gave the highest value of the weighted log-likelihood (equation (2.12)). Pink shaded areas show the 95 percentile interval for the modelled cell densities, based on 200 samples from the pseudo-posterior. Cell distributions obtained from the data using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) are shown by blue lines, with 95 percentile intervals obtained using 10,000 bootstrap samples of the data indicated by blue shaded areas.

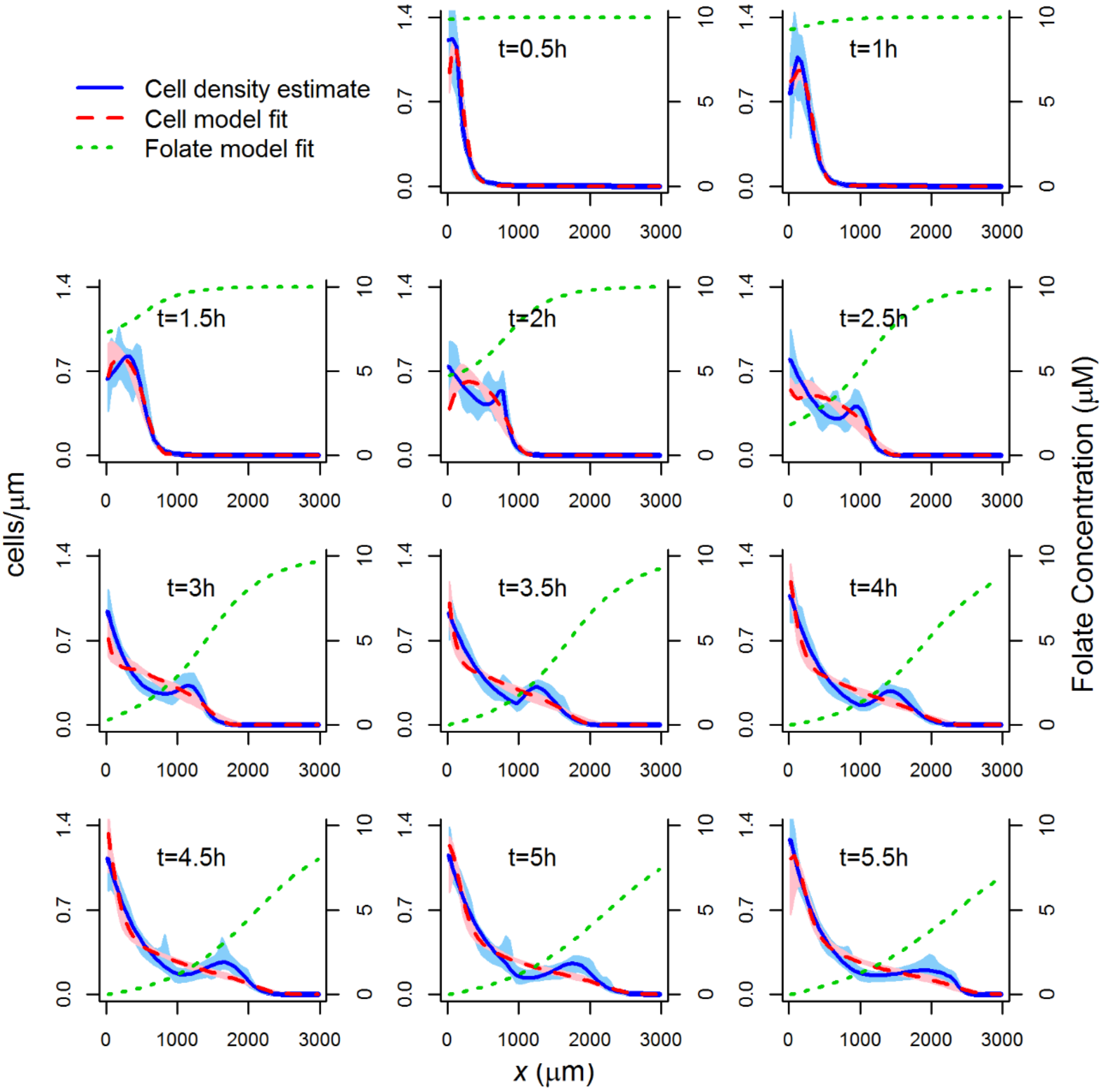

Figure A.7.3: Basic model fitted to the *Dictyostelium* data. Dashed red lines show *Dictyostelium* cell distributions at half-hour intervals produced by the basic model (equation (2.3)) using the optimised parameters from the bootstrap optimisation that gave the highest value of the weighted log-likelihood (equation (2.12)). Pink shaded areas show the 95 percentile interval for the modelled cell densities, based on 200 samples from the pseudo-posterior. The corresponding folate distributions predicted by this model are indicated by green dotted lines. Cell distributions obtained from the data using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) are shown by blue lines, with 95 percentile intervals obtained using 10,000 bootstrap samples of the data indicated by blue shaded areas.

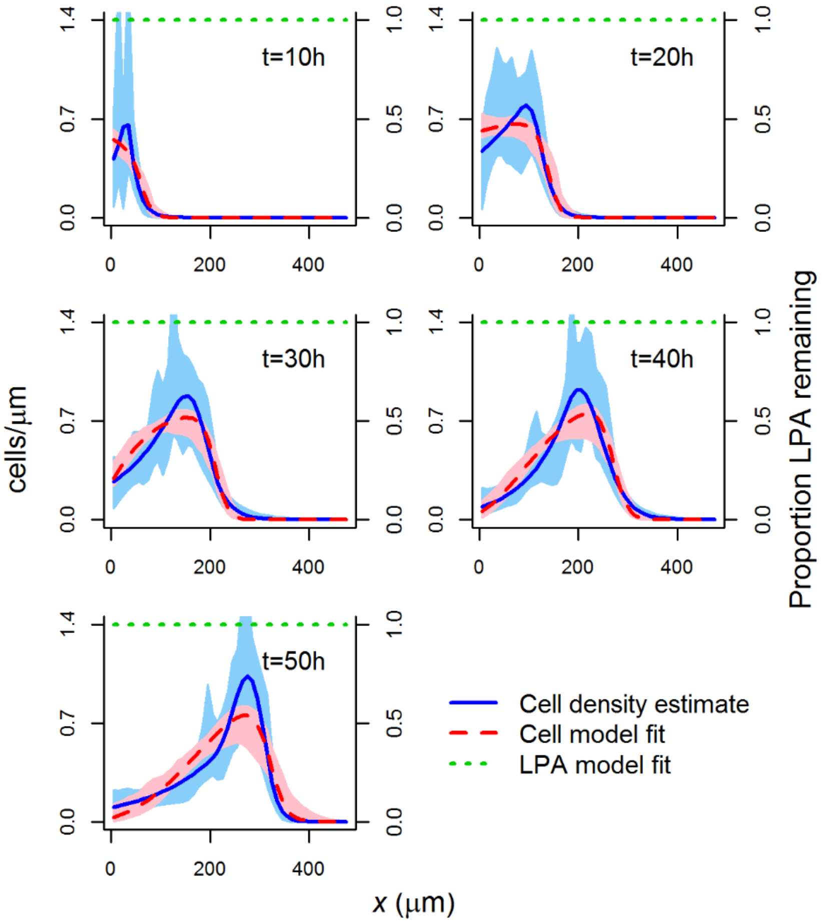

Figure A.7.4: Basic model fitted to the melanoma data. Dashed red lines show melanoma cell distributions at 10-hour intervals produced by the basic model (equation (2.3)) using the optimised parameters from the bootstrap optimisation that gave the highest value of the weighted log-likelihood (equation (2.12)). Pink shaded areas show the 95 percentile interval for the modelled cell densities, based on 200 samples from the pseudo-posterior. The corresponding LPA distributions predicted by this model are indicated by green dotted lines. Cell distributions obtained from the data using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) are shown by blue lines, with 95 percentile intervals obtained using 10,000 bootstrap samples of the data indicated by blue shaded areas.

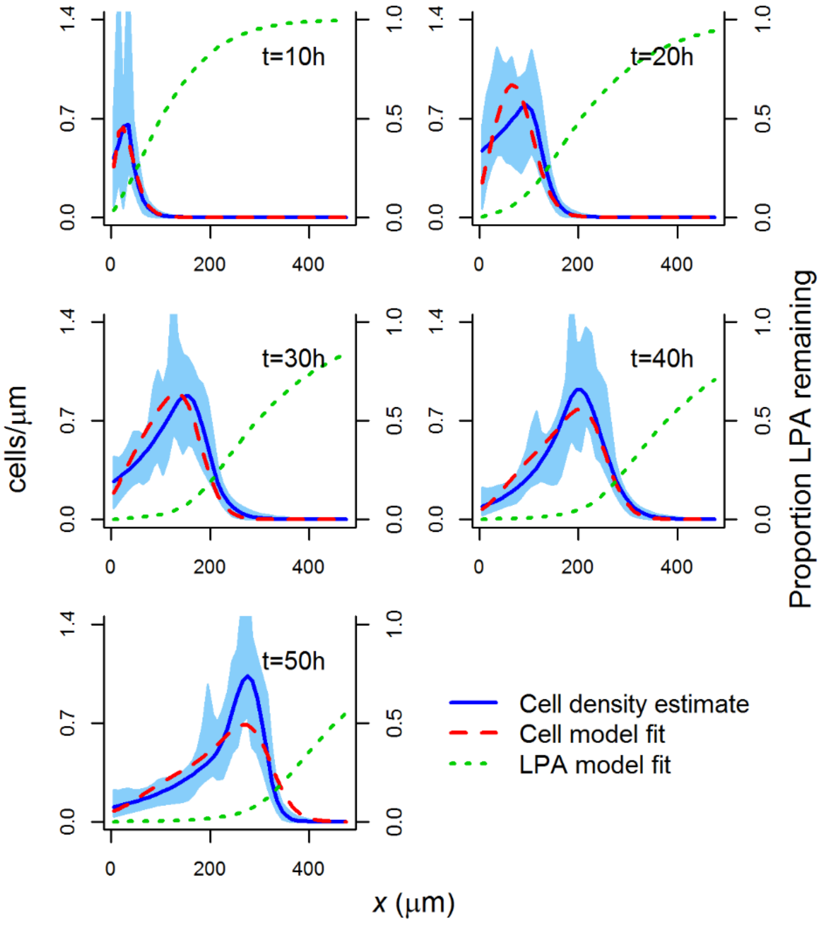

Figure A.7.5: Time invariant receptor saturation and overcrowding model fitted to the melanoma data. Dashed red lines show melanoma cell distributions at 10-hour intervals produced by the receptor saturation and overcrowding model (equation (2.3)) using the parameters optimised to give the maximum value of the weighted log-likelihood (see Appendix A.2). The corresponding LPA distributions predicted by this model are indicated by green dotted lines. Cell distributions obtained from the data using logspline density estimation (Kooperberg and Stone 1992, Stone et al. 1997, Kooperberg 2015) are shown by blue lines, with 95 percentile intervals obtained using 10,000 bootstrap samples of the data indicated by blue shaded areas.

A.8. Supplementary video descriptions

Supplementary videos are available online at: https://theses.gla.ac.uk/8942/

A.8.1. Supplementary video 2.1

This video is composed of microscopy images, captured every 90 seconds, of *Dictyostelium discoideum* cells moving under agarose. The first image was captured around an hour after the cells were introduced to a trough cut into the agarose, which is visible along the far left of the images. The agarose contained folate at an initially homogeneous concentration of 10µM. No folate was present in the trough area. The cells are observed to move to the right over time, leaving the trough and moving under the agarose. These images were collected by Tweedy et al. (2016)

A.8.1. Supplementary video 2.2

This video is composed of microscopy images, captured every 30 minutes, of human melanoma cells moving between two wells connected by a bridge in a direct visualisation chamber (Muinonen-Martin et al. 2010) that was homogeneously filled with 10% FBS (foetal bovine serum) The wells are visible to the far left and right of the images. The cells move from the left well to the right well over time. These images were collected by Muinonen-Martin et al. (2014).

Appendix B: Additional information for chapter 3

B.1. Representing the data in 1D

When fitting the cell movement models, I chose to discard the y-dimension, where, owing to the experimental set-up, there was nothing biologically interesting happening. Running the models in 1D space as opposed to 2D space allowed computational costs to be decreased by an order of magnitude. To check that there would be no significant misrepresentation of the data caused by this decision, I carried out two statistical tests. The first used Kolmogorov-Smirnov tests for each time point in each of the two datasets to confirm that the cell coordinates in *y* were not significantly different from samples from uniform distributions, indicating that there are no interesting features to be explained in this dimension (Table B.1.1). The second test was used to confirm that the *x* and *y* dimensions were independent by first calculating the mutual information for each time point for each dataset as:

$$\begin{aligned} I(x,y) &= H(x)+H(y)-H(x,y) \\ &= -\int p(x)\log p(x)dx - \int p(y)\log p(y)dy + \int p(x,y)\log p(x,y)dxdy \end{aligned} \tag{B.1.1}$$

where the probability density functions were obtained by kernel density estimation using the sm package in R (Bowman and Azzalini 2014). I then created 1,000 sample datasets for each time point in each dataset, under an assumption of independence of *x* and *y*, by carrying out slice sampling (e.g. section 24.5 of Murphy (2012)) on the marginal distributions $p(x)$ and $p(y)$ and randomly pairing the sampled *x* and *y* coordinates. The mutual information was then calculated for each of these sample datasets. I, thus, found that the mutual information values calculated from the original data were not significantly larger than would be expected if *x* and *y* were independent (Figs B.1.1-2).

Table B.1.1: Kolmogorov-Smirnov tests for uniformity of cell distributions in *y*. P-values from Kolmogorov-Smirnov tests used to check for significant deviations of the cell locations in *y* from samples from uniform distributions. In the 10μM folate dataset, two of the P-values were below the 0.05 significance level, but, following adjustment of the values for multiple testing (values shown in brackets; see Benjamini and Hochberg (1995) for calculation), I conclude that there is no evidence for significant deviation from a uniform distribution for either dataset.

Time Point	**P-value**	
	0μM folate	**10μM folate**
0	0.522	0.008 (0.102)
0.5	0.497	0.101 (0.304)
1.0	0.750	0.080 (0.321)
1.5	0.403	0.666 (0.799)
2.0	0.321	0.320 (0.480)
2.5	0.192	0.596 (0.795)
3.0	0.426	0.694 (0.757)
3.5	0.561	0.808 (0.807)
4.0		0.300 (0.599)
4.5		0.317 (0.543)
5.0		0.254 (0.609)
5.5		0.024 (0.146)

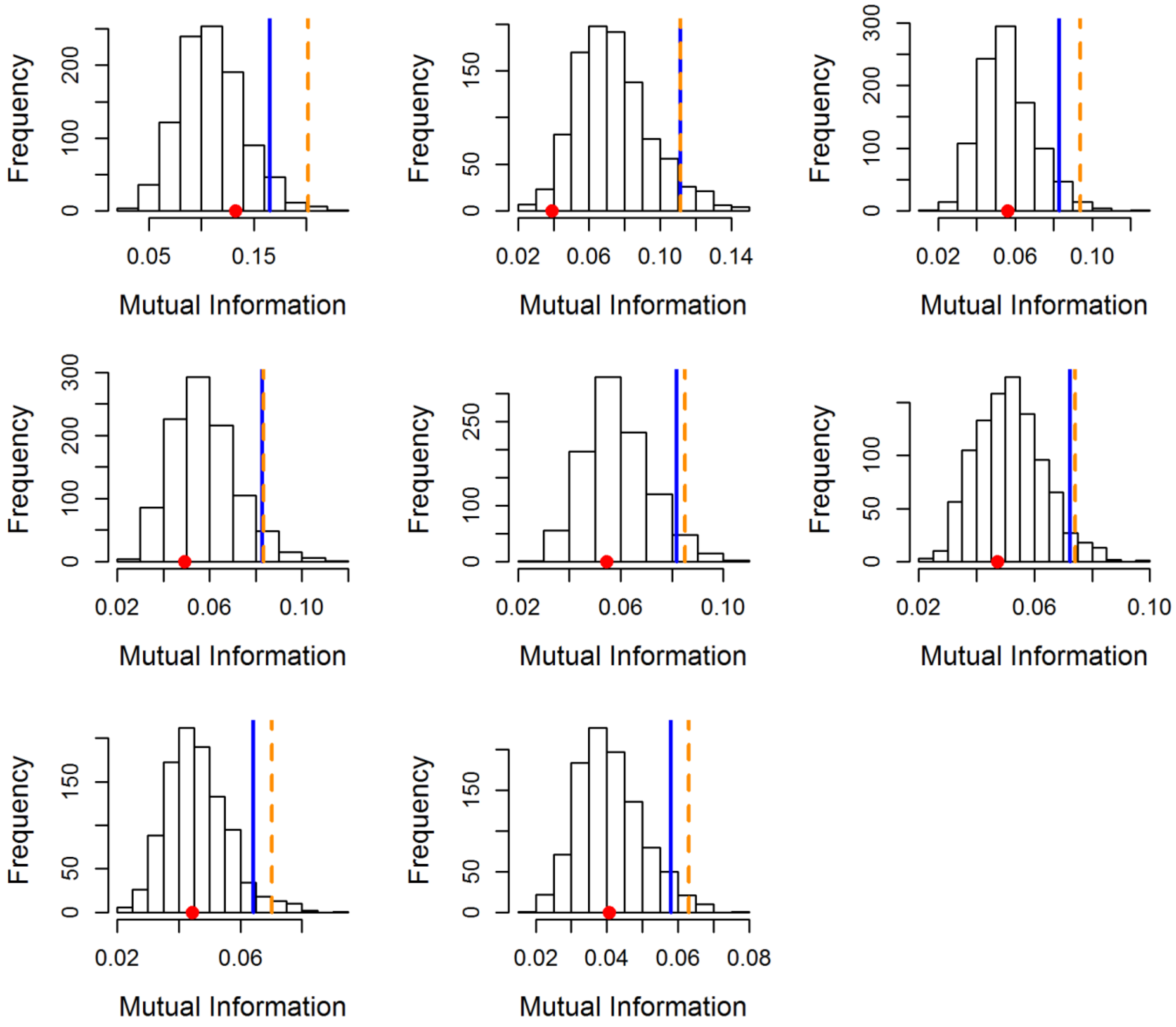

Figure B.1.1: Test of independence of the cell distributions in *x* and *y* for the 0µM folate data. Histograms of the mutual information between x and y for 1,000 sample datasets drawn assuming independence of x and y for each time point in the 0µM folate dataset. The mutual information values calculated from the real data are indicated by the red points. Solid blue lines show the mutual information beyond which the maximum 5% of the distribution is found. Dashed orange lines show the mutual information below which the red points must lie to indicate that x and y in the real data are independent, when multiple testing is controlled for (Benjamini and Hochberg 1995).

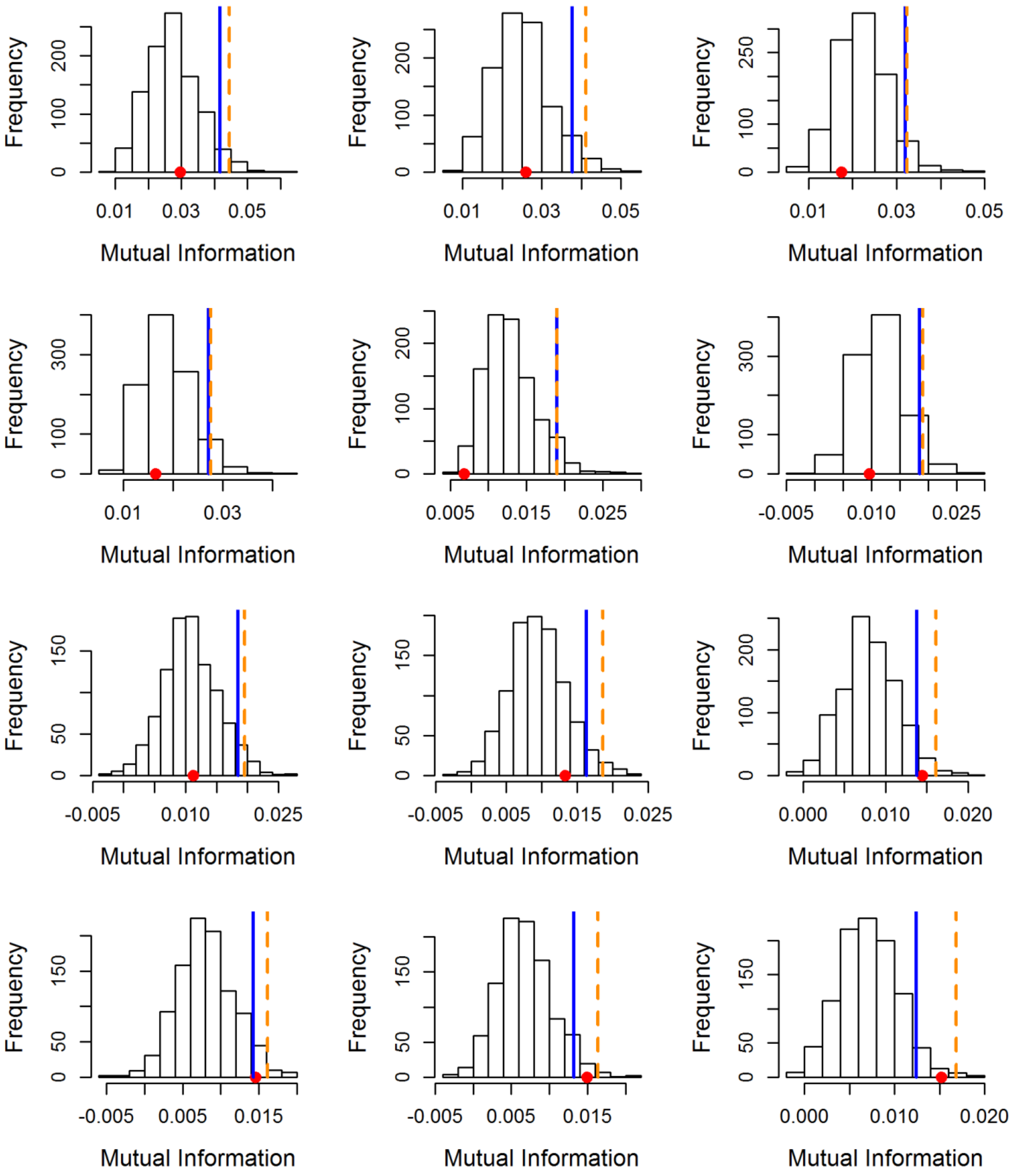

Figure B.1.2: Test of independence of the cell distributions in *x* and *y* for the 10μM folate data. Histograms of the mutual information between x and y for 1,000 sample datasets drawn assuming independence of x and y for each time point in the 10μM folate dataset. The mutual information values calculated from the real data are indicated by the red points. Solid blue lines show the mutual information beyond which the maximum 5% of the distribution is found. Dashed orange lines show the mutual information below which the red points must lie to indicate that x and y in the real data are independent, when multiple testing is controlled for (Benjamini and Hochberg 1995).

B.2. Prior distributions of model parameters

Details on all the priors applied to the parameters in the cell movement models are provided in Table B.2.1. Priors for two parameters, δ and ε, that respectively describe the steepness and position of the sigmoidal initial attractant distribution (equation (A.1.8)) are also included in Table B.2.1. As the initial attractant distribution was unobserved, these parameters were inferred during model fitting, with upper and lower bounds being introduced to prevent the distribution becoming unrealistic. I set the parameter bounds to δ_{min}=0.002, δ_{max}=1, ε_{min}=0 and ε_{max}=600. These bounds were selected in the same way as I selected those for the *Dictyostelium* dataset described in chapter 2 (see Appendix A.1.2), by comparing the initial cell distribution obtained from the data to initial attractant distributions obtained from a range of parameter values, and selecting the extremes that the distribution could realistically take (Fig. B.2.1). Increasing δ above δ_{max} has very little effect on the distribution, since the curve cannot get much steeper than it already is, making this a reasonable cut-off. A δ of less than δ_{min} either leads to the depleted attractant region extending too far beyond the initial distribution of the cells, into an area that should be at the undepleted maximum attractant value, or causes attractant to be too abundant in the trough region, where it is known that there was initially no attractant. An ε value greater than ε_{max} will similarly lead to the depleted area extending too far into the region where there are no cells, while a value lower than ε_{min} puts the inflection point into to the trough area, where attractant concentration should be low.

Table B.2.1: Prior distributions of model parameters.

Parameter	Prior	Notes
D_R	Beta(shape1=2, shape2=1.163), Rescaled to min=150μm^2/s, max=200μm^2/s	Folate diffusion coefficient; literature values are 192 and 194μm^2/s (Kalimuthu and John 2009, Ershad et al. 2013). This prior has a mode at 193μm^2/s
Kd	Gamma(shape=1.2, scale=0.08)	Dissociation constant; literature value of 0.016μM (De Wit et al. 1986) at which the mode of this gamma prior is positioned.
C_{max}	Gamma(shape=3.05, scale=50) Rescaled to have a minimum of 2.09	Maximum cell density; mode is at 50 times the maximum observed cell density, minimum is at the maximum observed cell density.
λ	Exponential(scale=4)	Describes decline in cell-cell attraction/repulsion as cell density increases
D_C intercept	Exponential(scale=50,000)	Cell diffusion coefficient. Prior is for the exponential of this parameter.
α intercept	Exponential(scale=50,000)	Advection in response to the gradient in folate/receptor saturation. Prior is for the exponential of this parameter.

γ intercept	Exponential(scale=800)	Folate depletion rate. Prior is for the exponential of this parameter.
η intercept	Normal(mean=0, sd=500,000)	Advection in response to the cell density gradient
α, γ & D_C time polynomial coefficients	Normal(mean=0, sd=20*0.5^(power of t))	standard deviations start at 20 for t^1 and progressively halve for each higher order of t
α & D_C space polynomial coefficients	Normal(mean=0, sd=20*0.5^(power of x))	standard deviations start at 20 for x^1 and progressively halve for each higher order of x
η time polynomial coefficients	Normal(mean=0, sd=500,000*0.5^(power of t))	standard deviations start at 500,000 for t^1 and progressively halve for each higher order of t
η space polynomial coefficients	Normal(mean=0, sd=500,000*0.5^(power of x))	standard deviations start at 500,000 for x^1 and progressively halve for each higher order of x
δ	Beta(shape1=1.5, shape2=1.5), Rescaled to min=0.002, max=1	Steepness of the sigmoid describing initial folate distribution (equation (A.1.8)). Boundaries of this distribution are as illustrated in Fig. B.2.1.
ε	Beta(shape1=1.5, shape2=1.5), Rescaled to min=0, max=600	Position of the inflection point of the sigmoid describing initial folate distribution (equation (A.1.8)). Boundaries of this distribution are as illustrated in Fig. B.2.1.

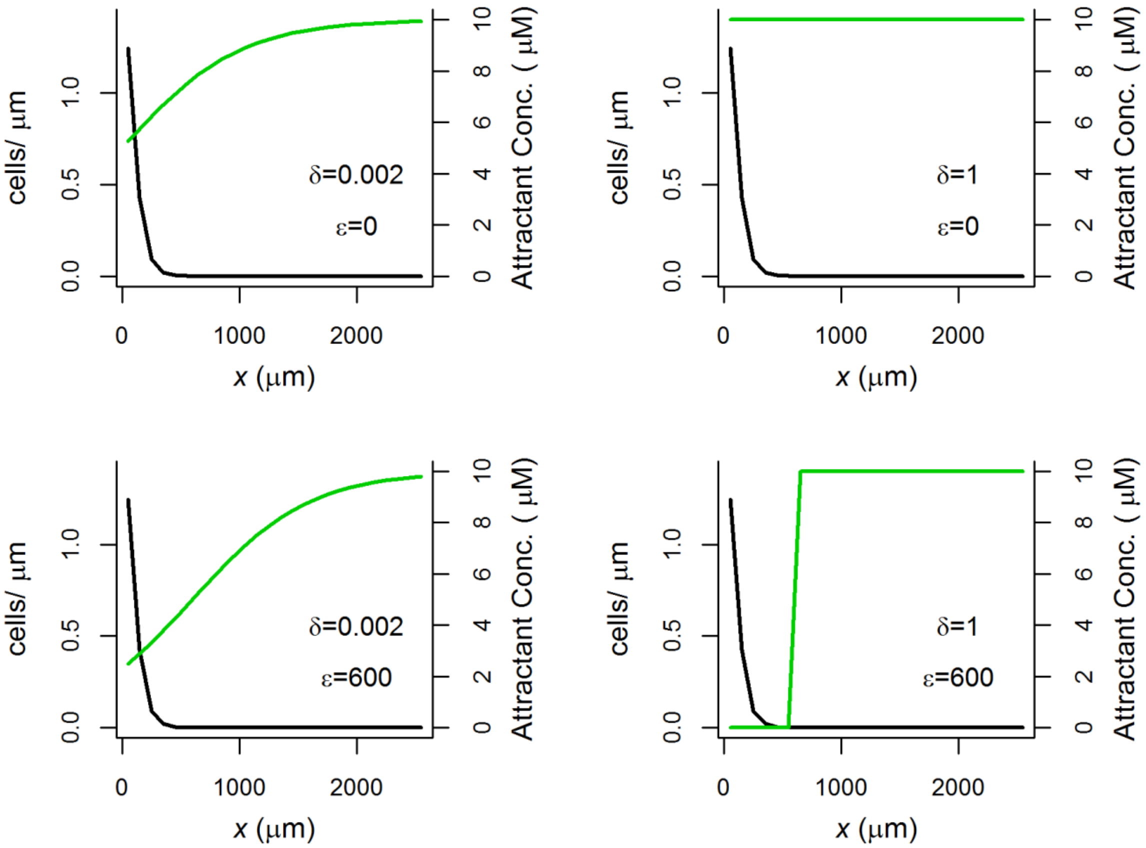

Figure B.2.1: Extremes that the initial folate distribution was permitted to take during model fitting to the 10μM folate dataset. Green lines show the initial attractant distributions calculated from equation (A.1.8) using each combination of the maximum and minimum values of the parameters δ and ε. The initial cell distribution (obtained by density estimation from the data) is included for reference (black line).

B.3. Cell numbers in the spatial region of interest over time

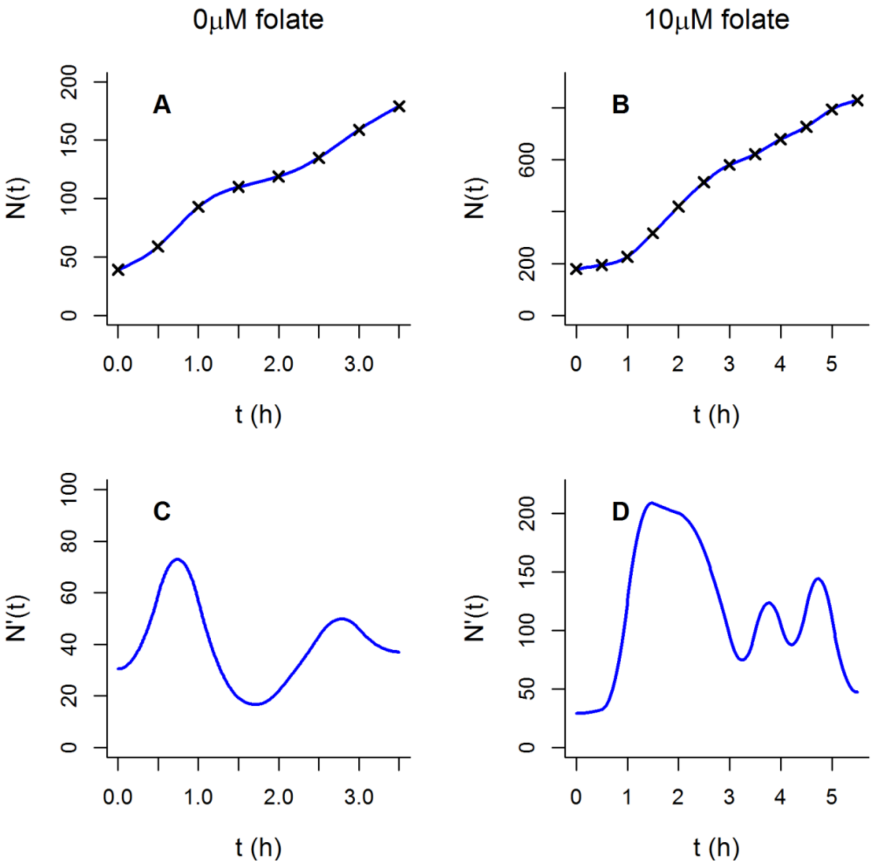

Figure B.3.1: Changes in the number of *Dictyostelium* cells in the region of interest for each dataset over time. A-B) Numbers of *Dictyostelium* cells observed in microscopy images at half-hourly intervals (black cross**es**), interpolated using a cubic spline $N(t)$ (blue line) for the datasets with 0μM folate (**A**) and 10μM folate (**B**). **C-D)** Derivatives of the curves in **A** and **B**.

B.4. Test of Bayesian inference method on simulated data

When carrying out inference for the models with advection coefficients (as discussed in section 3.4), high computational costs meant that achieving convergence of MCMC chains from hyperdispersed starting points was infeasible. I, therefore, first used repeated maximisations of the log-likelihood to obtain a good approximation of the MAP (maximum a posteriori parameter configuration). I then started two independent MCMC simulations of a minimum 80,000 MCMC steps from the MAP, and checked for convergence based on consistency of the WAIC scores obtained from two sections (the middle and end thirds of the MCMC chains, discarding the first third of steps as burn-in) of each MCMC run (giving 4 WAIC scores overall). Here, I provide a demonstration of the effectiveness of this approach on data simulated from a test model, the *N*-variable Goodwin model of biochemical oscillatory control (Goodwin 1965):

$$\begin{aligned}\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= \frac{a_1}{1+a_2\left(x_n(t)\right)^{\rho}} - \beta x_1(t)\\ \frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= k_1 x_1(t) - \beta x_2(t)\\ &\vdots\\ \frac{\mathrm{d}x_n(t)}{\mathrm{d}t} &= k_{N-1}x_{N-1}(t) - \beta x_N(t)\end{aligned} \tag{B.4.1}$$

This model can produce oscillating solutions that lead to highly multi-modal likelihood surfaces, so that, as for the cell movement models, MCMC chains used to infer the parameters of this model frequently become trapped on local optima (see Fig 8.3 of Girolami et al. (2010)). I set myself the same model selection problem described in Girolami et al. (2010), as outlined below.

Data were simulated from two versions of the model, using $N=3$ and $N=5$. The parameters from which the data were simulated were all drawn randomly from $\mathrm{Gamma}(2,1)$ distributions (with the exception of ρ, which was set to 10 throughout this analysis to ensure oscillating responses), and the models were numerically integrated using these parameters over a time period from $t=0$ to $t=60$, and initial conditions of zero for all variables. The values of the first two variables x_1 and x_2 were obtained at time intervals of 0.5, and Gaussian noise with variance 0.2 was added to these observations to create two datasets (Fig. B.4.1).

For a given parameter set, the probability of each data point was obtained from a Gaussian distribution with variance 0.2, centred on the model output for the variable (x_1 or x_2) to which that data point corresponds, at the time point at which the data point was obtained. The log-likelihood of the parameter set is then given by the sum of the log-probabilities over all data points. Note that when calculating the log-likelihood, I discarded the data points for which $t \le 20$, allowing the models to reach steady state.

I ran ten likelihood maximisations for each of the two models on each of the two datasets, drawing initial parameter values randomly from $\mathrm{Gamma}(2,1)$ distributions. For each model-dataset combination, I then identified the set of optimised parameters that gave the highest likelihood. These best parameter sets were used to initialise two MCMC chains of 20,000 iterations, using $\mathrm{Gamma}(2,1)$ priors for the parameters. WAIC values were then calculated from the middle and end thirds of each MCMC chain, and the two models were compared based on the mean of these four values for each dataset. This process of carrying out ten likelihood maximisations, running MCMC chains from the best optimised parameters and comparing WAIC values was repeated a further nine times, and the results are shown in Table B.4.1. It can be seen that, in every case, the mean WAIC is lower for the true model, suggesting that this approach to inference and model selection is generally accurate. In addition, based on the standard errors of the WAIC values (Table B.4.1), the inference and model selection approach I have developed is considerably more precise than model comparison using Bayes factors computed from standard MCMC sampling using an adaptive Metropolis algorithm (see Table 8.1 in Girolami et al. (2010)). My approach offers a similar level of precision to model selection using Bayes factors obtained from population MCMC with parallel tempering (see Table 8.2 in Girolami et al. (2010)). Note that this test of the inference method used both a relatively small number of initial optimisations (ten; the same number as I used in the main study) and short MCMC chains (20,000 iterations; I used a minimum of 80,000 in the main study). I expect the scheme to become even more accurate in identifying the correct model as the number of optimisations and the length of the MCMC chains are increased, as this increases the probability that good starting positions are obtained for the MCMC chains and that these MCMC chains reach convergence.

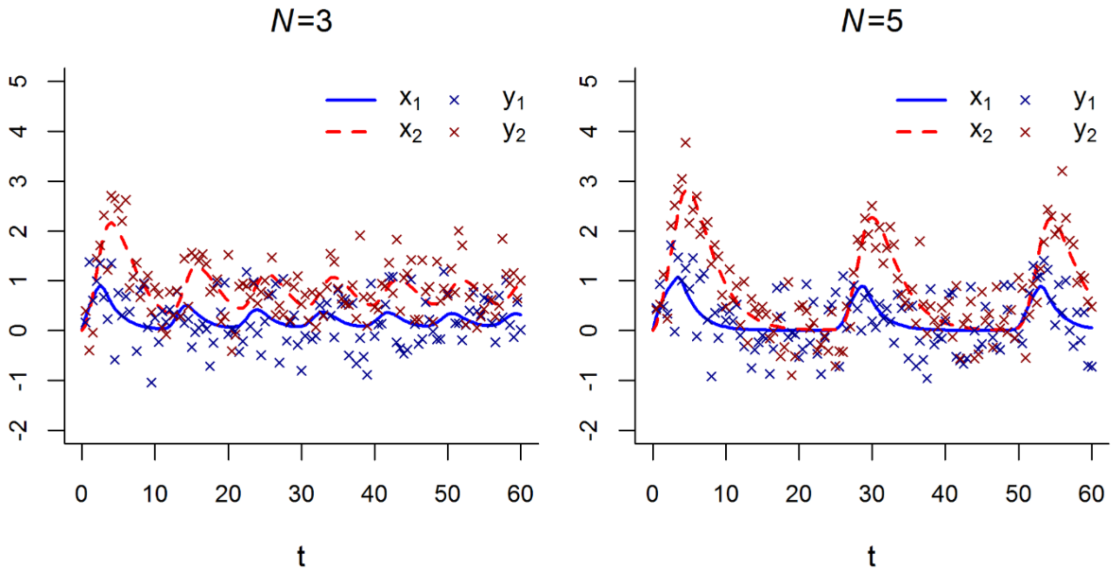

Figure B.4.1: Simulation of datasets from the Goodwin model of biochemical oscillatory control. Lines show the values of the first two variables x_1 and x_2 obtained by numerical integration of the Goodwin model (equation (B.4.1)) with $N=3$ and $N=5$, using parameters drawn from $\text{Gamma}(2,1)$ distributions. Crosses show data simulated by adding independent Gaussian noise with variance 0.2 to the model output at time intervals of 0.5.

Table B.4.1: Test of the inference method's ability to identify the correct model for the simulated datasets. Mean of the four WAIC values obtained from the middle and end thirds of the two MCMC chains run for each model-dataset combination during each of the ten replicates of the inference scheme. The model selected in each replicate is marked *.

Replicate	***N*=3 dataset**		***N*=5 dataset**	
	***N*=3 model**	***N*=5 model**	***N*=3 model**	***N*=5 model**
1	314.2 (se=10.14) *	322 (se=3.87)	627.2 (se=1.64)	296.3 (se=4.38) *
2	315.0 (se=8.68) *	317.8 (se=1.12)	624.7 (se=0.6)	619.9 (se=4.99) *
3	302.4 (se=0.89) *	315 (se=0.58)	625 (se=0.79)	622.8 (se=3.91) *
4	301.9 (se=0.84) *	316.8 (se=1.55)	625.9 (se=1.45)	297.4 (se=4.56) *
5	302.1 (se=0.97) *	317.1 (se=0.99)	827.2 (se=76.97)	294.4 (se=4.38) *
6	300.5 (se=0.83) *	337.3 (se=0.54)	642 (se=0.37)	296.3 (se=4.99) *
7	314.6 (se=7.57) *	330.2 (se=2.57)	624.5 (se=0.19)	620.3 (se=3.91) *
8	302.2 (se=0.78) *	328.5 (se=5.16)	643.1 (se=0.49)	552.3 (se=4.56) *
9	302.0 (se=0.54) *	316.4 (se=0.89)	842.9 (se=0.91)	297.6 (se=4.38) *
10	301.9 (se=2.91) *	331.9 (se=1.48)	625.4 (se=0.82)	621.3 (se=4.99) *

B.5. Selecting the degrees of the polynomials describing the dependencies of the model parameters on time and space

The WAIC values used to select the optimal degrees of the polynomials describing spatial and temporal dependencies in the diffusion coefficient of the cells (D_C; equation (3.5)) for the diffusion model fitted to the 0µM folate dataset are provided in Tables B.5.1-2. For both the standard and weighted likelihoods (equations (2.11-12)), I select a polynomial degree of 4 in time and 2 in space based on WAIC. While I have more confidence in WAIC as a model comparison statistic, due to its reduced reliance on asymptotics and relaxation of the assumption that all parameters are well-determined by the data (see Chapter 7 of Gelman et al. (2013)), I also calculated AICc (Akaike Information Criterion corrected for small sample sizes; Akaike (1974), Hurvich and Tsai (1989)) and BIC (Bayesian Information Criterion; Schwarz (1978)) values for each combination of polynomial degrees, using the highest likelihood point visited by the MCMC chains as an estimate of the maximum likelihood (Tables B.5.3-6). I find a close agreement between WAIC and AICc, increasing confidence in the WAIC results, though the agreement between WAIC and BIC is poorer (Fig. B.5.1).

Table B.5.1: WAIC-based selection (using the standard likelihood) of the degrees of the polynomials describing the spatio-temporal dependence of the diffusion coefficient of the diffusion model for the 0µM folate dataset. WAIC values are given for various combinations of degrees of the spatial and temporal polynomials (which are defined in equation (3.5)). Note that the minimum value was subtracted from all of the values to aid comparison. The optimal combination of degrees is marked *.

		Degree in time						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Degree in space	**0**	87.9	79	61.5	56.6	57.6	50.3	44.1
	1	48.2	41.7	37.2	33.9	33.3	36.7	41
	2	9.5	2.9	4.6	2.7	0*	1.9	5.9
	3	11	4.4	6.2	4.4	1.8	4	8
	4	13.5	6.6	8.3	6.3	4.4	6.7	10.4
	5	7.8	1	3	1.6	1.3	8.3	11.2
	6	87.9	79	61.5	56.6	57.6	50.3	44.1

Table B.5.2: WAIC-based selection (using the weighted likelihood) of the degrees of the polynomials describing the spatio-temporal dependence of the diffusion coefficient of the diffusion model for the 0µM folate dataset. WAIC values are given for various combinations of degrees of the spatial and temporal polynomials (which are defined in equation (3.5)). Note that the minimum value was subtracted from all of the values to aid comparison. The optimal combination of degrees is marked *.

		Degree in time						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Degree in space	**0**	95.4	99.1	63.4	58.8	67.6	65.1	52.2
	1	50.6	49.9	39.3	33.6	33.1	36.1	42.8
	2	10.4	7.2	7.3	2.4	0*	2.3	5.9
	3	12.2	8.9	8.8	3.5	1.9	4.3	8.7
	4	14.2	11.2	11.6	6.6	4.7	7.7	12.4
	5	12.3	9	9.4	4.8	4.3	9.8	18.9
	6	95.4	99.1	63.4	58.8	67.6	65.1	52.2

For both the standard and weighted likelihoods, BIC selects a lower polynomial degree in time than WAIC, but the same degree in space (Tables B.5.3-4). This reduction in the complexity of the preferred model when using BIC is expected, as this statistic is known for its tendency to select models that are overly simple (Ripplinger and Sullivan 2008). The shape of the time polynomial has been substantially simplified in the BIC selected model (Fig. B.5.2D) compared to the WAIC-selected model (Fig. B.5.2.B). The shape of the polynomial in space, however, is unchanged between the models selected by WAIC and BIC (Fig. B.5.2A,C).

Table B.5.3: BIC-based selection (using the standard likelihood) of the degrees of the polynomials describing the spatio-temporal dependence of the diffusion coefficient of the diffusion model for the 0µM folate dataset. BIC values are given for various combinations of degrees of the spatial and temporal polynomials (which are defined in equation (3.5)). Note that the minimum value was subtracted from all of the values to aid comparison. The optimal combination of degrees is marked *.

		Degree in time						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Degree in space	**0**	69.9	64.3	52.4	51.9	55.1	53.2	50.6
	1	35.4	33.4	33.5	34.5	37.5	44.4	51.3
	2	1.7	0*	6.5	9.2	4.7	17.8	24.7
	3	7.9	6	12.2	15.2	17	24.1	31
	4	14.6	12.8	19.1	21.9	23.9	30.7	37.8
	5	12.9	11.2	17.3	20	23.1	30	38.8
	6	87.9	79	61.5	56.6	57.6	50.3	44.1

Table B.5.4: BIC-based selection (using the weighted likelihood) of the degrees of the polynomials describing the spatio-temporal dependence of the diffusion coefficient of the diffusion model for the 0µM folate dataset. BIC values are given for various combinations of degrees of the spatial and temporal polynomials (which are defined in equation (3.5)). Note that the minimum value was subtracted from all of the values to aid comparison. The optimal combination of degrees is marked *.

		Degree in time						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Degree in space	**0**	74.7	81.4	50.7	50.3	56.9	57.3	54.6
	1	35.2	38.8	32.3	30.7	33.8	40.6	47
	2	0*	1.5	6.1	5.8	8	14.6	21.8
	3	6.3	7.7	12	11.7	13.9	20.6	27.7
	4	13.1	14.5	18.8	18.6	20.9	27.9	34.9
	5	14.6	16.1	20	20.2	23.4	30.5	38.8
	6	18.7	20.2	24.3	24.5	27.6	34.4	42.1

For the weighted likelihood, AICc selects the same model as WAIC, with a degree of 4 in time and 2 in space (Table B.5.6). For the standard likelihood, AICc shows a slight preference for an increased polynomial degree of 6 in space (Table B.5.5), but it should be noted that there is a similar level of support (difference in AICc of only 1.0) for the degree of 2 in space that was selected by WAIC. AICc is known to typically select models that are overly complex (Ripplinger and Sullivan 2008), so this slight disagreement between WAIC and AICc is to be expected. Using a polynomial degree of 6 in space results in a considerably more complex pattern in the space polynomial (Fig. B.5.2E), but the time polynomial is largely unchanged (Fig. B.5.2F).

Table B.5.5: AICc-based selection (using the standard likelihood) of the degrees of the polynomials describing the spatio-temporal dependence of the diffusion coefficient of the diffusion model for the 0µM folate dataset. AICc values are given for various combinations of degrees of the spatial and temporal polynomials (which are defined in equation (3.5)). Note that the minimum value was subtracted from all of the values to aid comparison. The optimal combination of degrees is marked *.

		Degree in time						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Degree in space	**0**	89.8	79.5	62.9	57.7	56.2	49.5	42.1
	1	50.6	43.9	39.2	35.6	33.8	36	38.1
	2	12.2	5.8	7.5	5.5	1	4.7	6.9
	3	13.6	7	8.5	6.8	3.9	6.3	8.4
	4	15.6	9.1	10.7	8.8	6.1	8.2	10.5
	5	9.2	2.8	4.1	2.1	0.5	2.8	6.9
	6	8.2	2	3.9	2.1	0*	1.9	5.6

Table B.5.6: AICc-based selection (using the weighted likelihood) of the degrees of the polynomials describing the spatio-temporal dependence of the diffusion coefficient of the diffusion model for the 0µM folate dataset. AICc values are given for various combinations of degrees of the spatial and temporal polynomials (which are defined in equation (3.5)). Note that the minimum value was subtracted from all of the values to aid comparison. The optimal combination of degrees is marked *.

		Degree in time						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Degree in space	**0**	95.1	97.1	61.7	56.5	58.4	54.1	46.7
	1	50.9	49.8	38.6	32.2	30.6	32.7	34.3
	2	11	7.8	7.6	2.6	0*	2	4.5
	3	12.5	9.2	8.8	3.7	1.3	3.3	5.6
	4	14.6	11.3	10.9	6	3.5	5.9	8.2
	5	11.4	8.2	7.4	2.9	1.3	3.8	7.4
	6	10.8	7.6	7	2.4	0.9	3	6

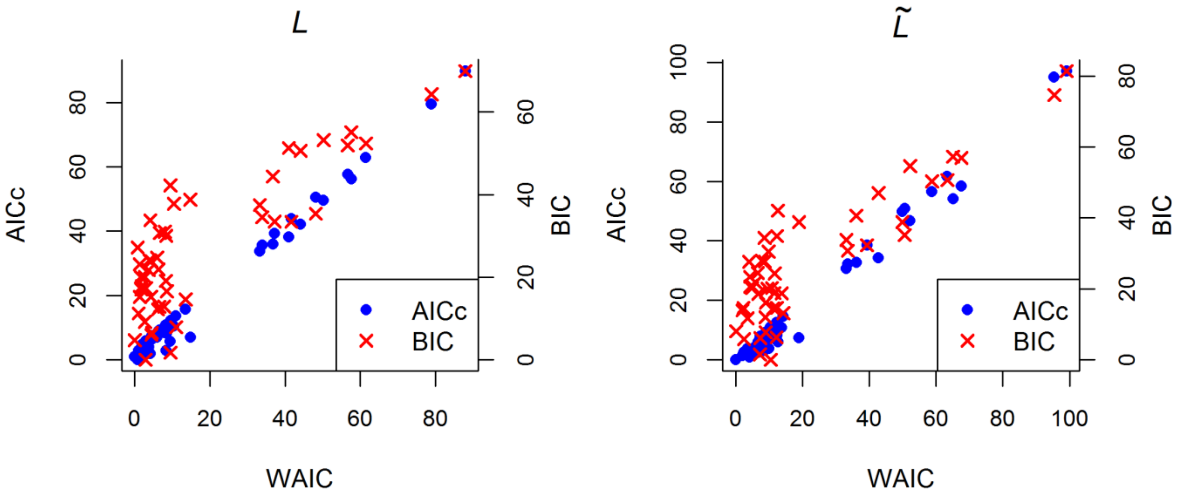

Figure B.5.1: Comparison of WAIC, AICc and BIC. Plots of AICc (blue points) and BIC (red crosses) against WAIC for versions of the diffusion model that used different combinations of degrees for the polynomials describing the temporal and spatial dependencies, fitted to the 0μM folate data using both the standard and weighted likelihoods (L and $\tilde{L}$, equations (2.11-12)).

When carrying out the local readjustment of the polynomial degrees for the 10μM folate dataset, I first identified polynomial coefficients where the posterior distribution was focussed around zero (Fig. B.5.3). Those parameters with a relatively high posterior density at zero were associated with the time polynomial, suggesting that the degree of this polynomial could be reduced. Using WAIC, I thus reduce the time polynomial degree from the value of 4 obtained from the 0μM folate dataset to a value of 3 for the standard likelihood and 2 for the weighted likelihood (Table B.5.7). An AICc comparison shows close agreement with the WAIC results. BIC is in agreement with WAIC and AICc for the weighted likelihood, but, predictably, selects a simpler time polynomial than WAIC and AICc for the standard likelihood (Table B.5.7).

Table B.5.7: Local readjustment of the temporal polynomial degree for the 10μM folate data. WAIC, AICc and BIC values for the diffusion model, with different degrees of the polynomial describing the dependence of D_C on time (equation (3.5)), fitted to the 10μM folate dataset using both the standard (equation (2.11)) and weighted (equation (2.12)) likelihoods, L and $\tilde{L}$. The degree of the polynomial describing dependence in space was fixed to 2, the value suggested from fits to the 0μM folate dataset (Tables B.5.1-2). For both L and $\tilde{L}$, the minimum value has been subtracted from each statistic to aid comparison. Standard errors (in brackets) for WAIC were calculated as described in Appendix A.5. Note that for all statistics and both L and $\tilde{L}$, the optimal degree in time (marked *) is lower than the value of 4 suggested by the WAIC results from the 0μM folate dataset (Tables B.5.1-2).

Degree in	**WAIC**		**AICc**		**BIC**	
Time	L	$\tilde{L}$	L	$\tilde{L}$	L	$\tilde{L}$
1	253.0 (0.1)	314.1 (0.08)	253.0	314.5	243.2	307.8
2	2.7 (0.1)	0 (0.09)*	3.2	0*	0*	0*
3	0 (0.1)*	1.4 (0.11)	0*	1.2	3.5	7.9
4	1.6 (0.1)	3.9 (0.12)	1.9	3.2	12.1	16.6

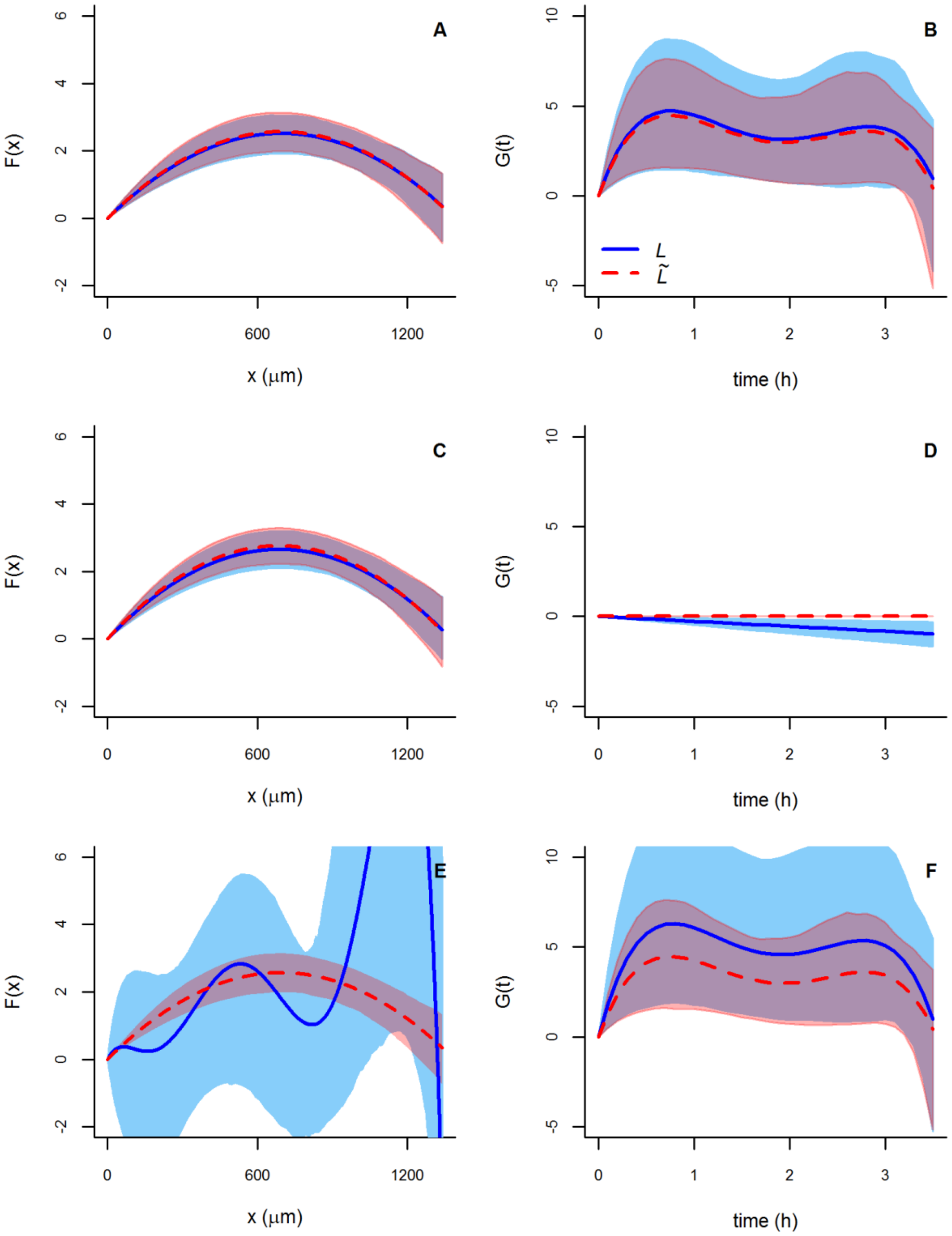

Figure B.5.2: Spatial and temporal dependencies of the cell diffusion coefficient D_C fitted to the 0μM folate data. Plots show the polynomials F(x) and G(t) (equation (3.5)) estimated from the data with the degrees selected by WAIC (**A-B**), BIC (**C-D**) and AICc (**E-F**) (see Tables B.5.1-6)). Polynomials obtained using both the standard and weighted likelihoods (L and $\tilde{L}$, equations (2.11-12)) are shown. 95 percentile intervals were obtained from 1,000 samples from the posterior distribution.

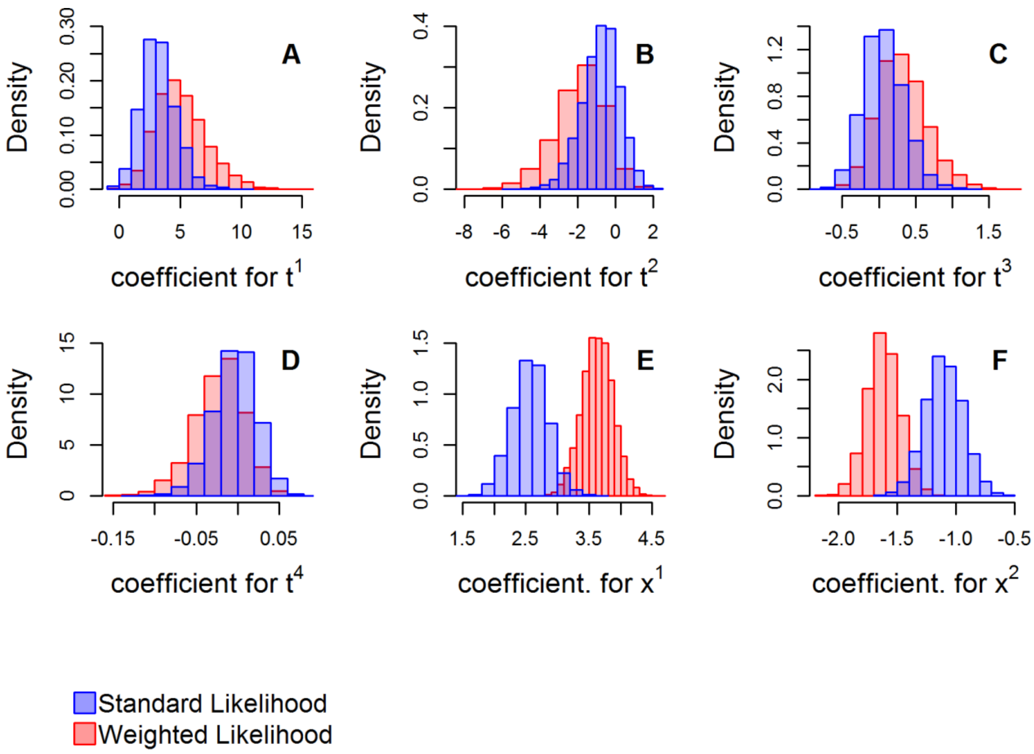

Figure B.5.3: Identification of polynomial coefficients with a posterior that is focussed around zero for the 10µM dataset. Posterior distributions for the coefficients of the polynomials describing spatial and temporal dependencies of the cell diffusion coefficient D_C (see equation (3.5)) from sampling from the posterior distribution of the diffusion model using the 10µM folate dataset and both the standard and weighted likelihoods (L and $\tilde{L}$). Here, I have used a polynomial degree of four in time and two in space (the degrees selected from fitting to the 0µM folate dataset (Tables B.5.1-2)). Note that, while zero has a very low posterior density for the coefficients of the spatial polynomial (plots E-F), it is well within the main bulk of the posterior distribution for three of the coefficients in the time polynomial (plots B-D), suggesting that a polynomial in time of degree four may be overly complex. I used this information to guide a local readjustment of the time polynomial (Table B.5.7).

B.6. Additional plots of dependence of cell behaviour on time and space

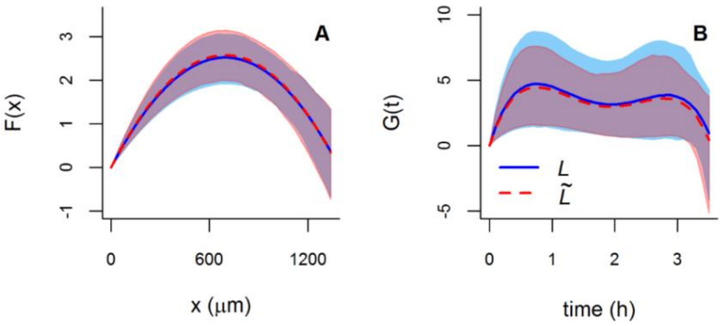

Figure B.6.1: Dependence of the diffusion coefficient D_C fitted to the 0μM folate dataset on space and time. Spatial (**A**) and temporal (**B**) dependencies of the cell diffusion coefficient D_C from fitting the diffusion model to the 0μM folate dataset, using both the standard and weighted likelihoods (L and $\tilde{L}$, equations (2.11-12)). Plots show the polynomials F(x) and G(t) (see equation (3.5)), which have degrees of two and four respectively (the degrees selected by WAIC (Tables B.5.1-2)). 95 percentile intervals were obtained from 1,000 samples from the posterior distribution.

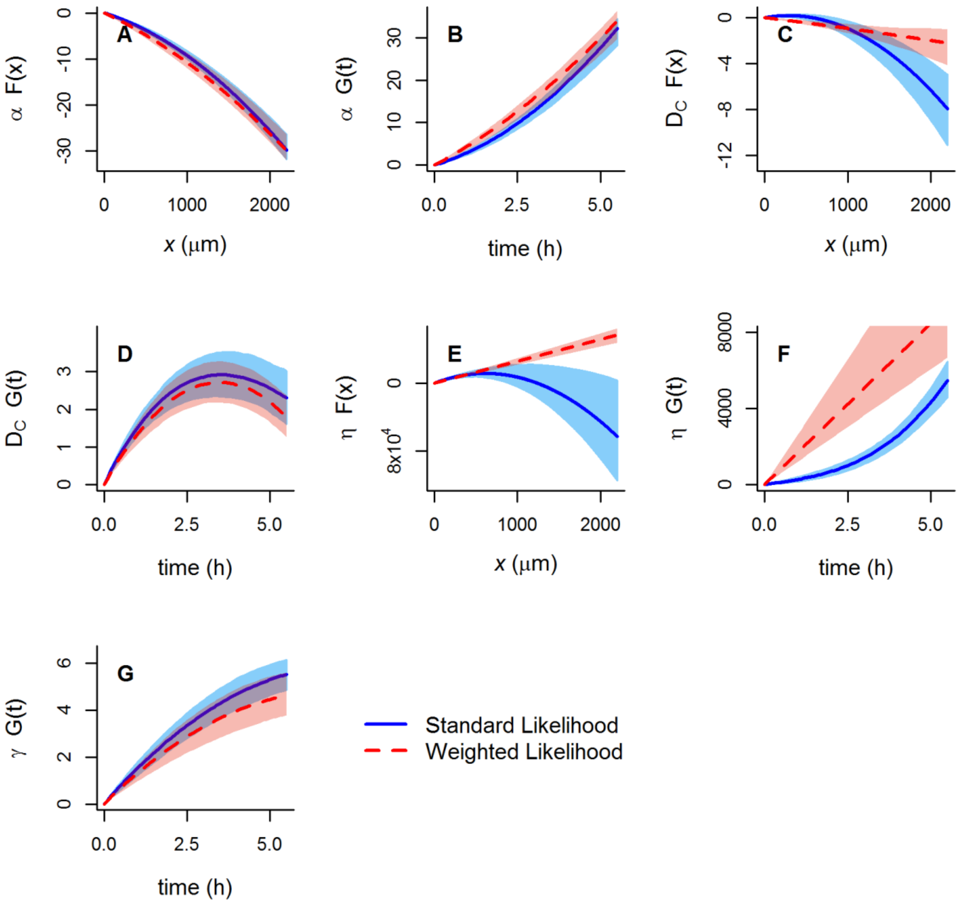

Figure B.6.2: Spatial and temporal dependencies of the parameters of the interaction model fitted to the 10μM folate dataset. The polynomials (see equations (3.4-5)) were estimated using both the standard and weighted likelihoods (equations (2.11-12)). 95 percentile intervals were obtained by sampling 1,000 parameter sets evenly from the latter two thirds of both MCMC chains obtained for this model. Lines show the mean values of the functions.

B.7. WAIC tables for comparison of full set of candidate models

Table B.7.1: WAIC values for models fitted to the 10µM folate dataset using the standard likelihood (equation (2.11)). Two values of the WAIC are given for each MCMC chain; one using samples from the middle third of the chain, and one using samples from the final third of the chain. The mean WAIC value for each model is taken as the mean of the 4 values calculated from the mid and end sections of the chains for that model. Standard errors are included in brackets and were calculated as outlined in Appendix A.5, with the exception of the standard error of the mean, which was obtained as the standard deviation of the four values for each model, divided by √4. Note that the minimum value was subtracted from all values to aid comparison. B=Basic, RS=Receptor saturation, I=Interaction, O=Overcrowding.

Model	**WAIC**				
	Chain1		**Chain2**		**Mean**
	Mid	**End**	**Mid**	**End**	
B	6.8 (0.29)	6.6 (0.29)	4.9 (0.3)	4.6 (0.29)	5.8 (0.58)
RS	17 (0.28)	16.6 (0.28)	14.1 (0.3)	12 (0.28)	14.9 (1.16)
I	1.6 (0.28)	0.7 (0.29)	0.5 (0.33)	3 (0.33)	1.5 (0.55)
O	5.4 (0.28)	4.6 (0.27)	4.3 (0.31)	5.5 (0.3)	5 (0.29)
RS+I	11.7 (0.31)	13.1 (0.32)	14.2 (0.29)	14.8 (0.27)	13.4 (0.69)
RS+O	14.2 (0.31)	13.2 (0.3)	14.3 (0.34)	13.5 (0.35)	13.8 (0.26)
I+O	2.7 (0.27)	6.6 (0.26)	0 (0.23)	4.5 (0.24)	3.4 (1.39)
RS+I+O	12 (0.3)	12.3 (0.33)	12.5 (0.3)	8.7 (0.35)	11.4 (0.9)

Table B.7.2: WAIC values for models fitted to the 10µM folate dataset using the weighted likelihood (equation (24)). Two values of the WAIC are given for each MCMC chain; one using samples from the middle third of the chain, and one using samples from the final third of the chain. The mean WAIC value for each model is taken as the mean of the 4 values calculated from the mid and end sections of the chains for that model. Standard errors are included in brackets and were calculated as outlined in Appendix A.5, with the exception of the standard error of the mean, which was obtained as the standard deviation of the four values for each model, divided by √4. Note that the minimum value was subtracted from all values to aid comparison. B=Basic, RS=Receptor saturation, I=Interaction, O=Overcrowding.

Model	**WAIC**				
	Chain1		**Chain2**		**Mean**
	Mid	**End**	**Mid**	**End**	
B	8.4 (0.36)	10.1 (0.36)	5.4 (0.35)	5.3 (0.35)	7.3 (1.18)
RS	18.6 (0.33)	17.4 (0.32)	18.3 (0.34)	19.5 (0.35)	18.5 (0.44)
I	5 (0.32)	5.9 (0.33)	0 (0.32)	0.5 (0.33)	2.8 (1.52)
O	5.3 (0.31)	6.5 (0.32)	6 (0.32)	7.3 (0.32)	6.3 (0.42)
RS+I	8.9 (0.32)	9.8 (0.38)	8.9 (0.33)	10.4 (0.32)	9.5 (0.37)
RS+O	13.4 (0.28)	16.1 (0.3)	15.5 (0.29)	12.5 (0.3)	14.3 (0.85)
I+O	4.3 (0.32)	7.8 (0.32)	5.8 (0.33)	5.3 (0.32)	5.8 (0.73)
RS+I+O	9.6 (0.34)	16.8 (0.37)	11.2 (0.34)	12 (0.35)	12.4 (1.55)

B.8. Diffusion model fit to 10µM folate data

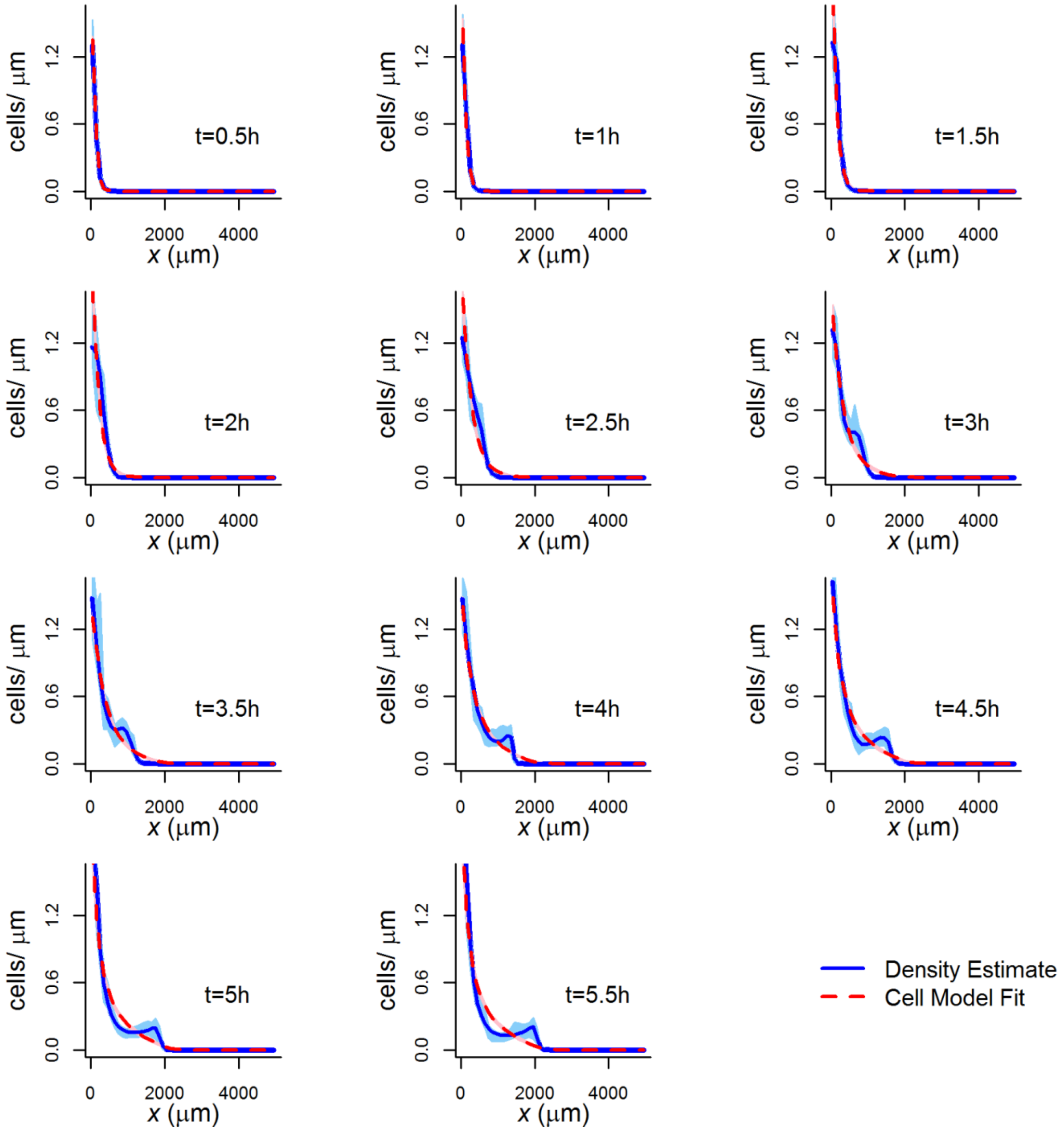

Figure B.8.1: Fit of the diffusion model to the 10µM folate data. Plots of the cell distributions at half-hourly intervals simulated (using the posterior mean parameters) from the diffusion model fitted to the 10µM folate data using the standard likelihood (equation (2.11), with polynomial degrees of three and two describing the temporal and spatial dependencies of the diffusion coefficient respectively. Direct density estimates from the data, obtained using logspline density estimation (Stone et al. 1997), are included for comparison. 95 percentile intervals for the density estimates (blue shaded areas) were obtained by non-parametric bootstrapping, using 10,000 samples of the data. 95 percentile intervals for the model (pink shaded areas) were obtained from 500 samples from the posterior distribution.

B.9. Residual Analysis

In standard residual analysis, the residuals are computed by taking the difference between the observed data and the predictions from the model, and then using them in standard diagnostic plots, to test e.g. independence or distributional assumptions. This is not immediately feasible in this study, because the model target is not directly observed. Time-varying cell locations are observed, while the model predicts time-varying spatial cell distributions. I therefore proceeded by using the time-varying cell locations to obtain an estimate of the MAP (maximum a posteriori parameter configuration) for the selected model (the interaction model, Table 3.1) as described in section 3.4, and using this to predict time-varying cell distributions from the model. I then obtained independent density estimates from the same time-varying cell locations using the logspline method described in Stone et al. (1997), and computed the difference between these probability densities and those predicted by the model. To obtain 95% confidence intervals around the density estimates, I repeated the density estimation procedure 10,000 times on 10,000 bootstrap replicates of the cell data. The results for the selected model are shown in Fig. B.9.1. I find that the differences between the model predictions and the density estimates from the data lie clearly within the relevant confidence regions, suggesting that there is no significant model mismatch.

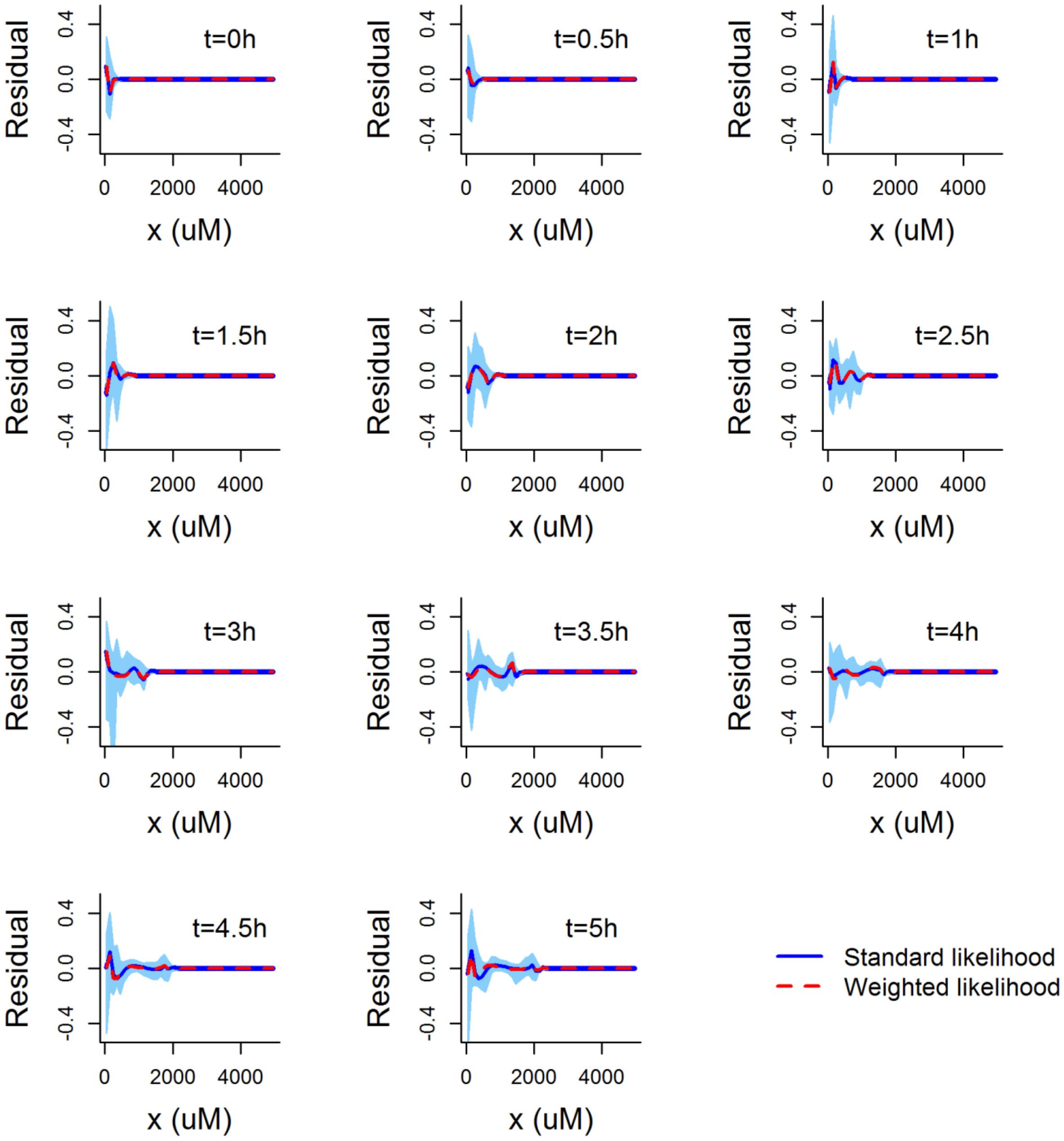

Figure B.9.1: Residual analysis. The lines show the difference between the cell distribution estimated from the 10µM folate data (using logspline density estimation (Stone et al. (1997)) and the cell distribution predicted from the MAP (maximum a posteriori parameter configuration) of the interaction model fitted using the standard likelihood and the weighted likelihood, plotted against x (the spatial coordinate) at different times *t*. The blue shaded areas show the 95% confidence regions obtained by logspline density estimation on bootstrap samples of the data, indicative of intrinsic estimation uncertainty. Note that the difference between model prediction and direct estimate from the data (the residual) is always included in the confidence intervals, suggesting that there is no significant model mismatch.

Appendix C: Additional information for chapter 4

C.1. Supplementary video descriptions

Supplementary videos are available online at: https://theses.gla.ac.uk/8942/

C.1.1. Supplementary video 4.1

This video is composed of 33 maps showing data on the distribution of wildebeest in the Serengeti ecosystem collected through aerial surveys over the period between August 1969 and August 1972. Each 25km^2 cell in a 56x46 grid is shown by its colour as belonging to one of five ordinal wildebeest abundance categories. The wildebeest density ranges that fall into each category are outlined in the scale bar.

C.1.2. Supplementary video 4.2

This video shows daily maps of wildebeest density obtained from the ordinal categorical wildebeest distribution data (see supplementary video 4.2) using the GAM-based spatio-temporal smoothing method outlined in section 4.3. The densities, in wildebeest/25km^2, are indicated by the colours of the grid cells as described in the scale bar. The two contours indicate the boundaries between abundance categories 0, 1 and 2 (which respectively contain 0, 1-25, and 26-250 wildebeest/25km2).

Appendix D: Additional information for chapter 5

D.1. Preparation of canopy cover data layer

I had access to three sources of data on canopy cover in the region. The first two were collected in 1962 and 1972 by Norton-Griffiths (1979) using aerial photography. These two datasets are grids of 10x10km cells, where the proportion of each cell that is tree cover is recorded (Fig. D.1A-B). Both of these grids have many missing values however (see white grid cells in Fig. D.1A-B). The third data source was the official Serengeti management map (Frankfurt Zoological Society and Harvey Maps 2010). This is a spatial polygon dataset, where each polygon is assigned to one of four ordinal categories, describing the proportion of canopy cover: 0-0.02, 0.02-0.2, 0.2-0.5 and 0.5-1.0. These categories were assigned based on the vegetation map produced by Reed et al. (2009), which used satellite images collected in 1999 and 2000. When using these data, I assumed that the proportion of canopy cover at a point in space in this map has a proportion of canopy cover equal to the mid-point of the ordinal category assigned to that point (Fig. D.1C).

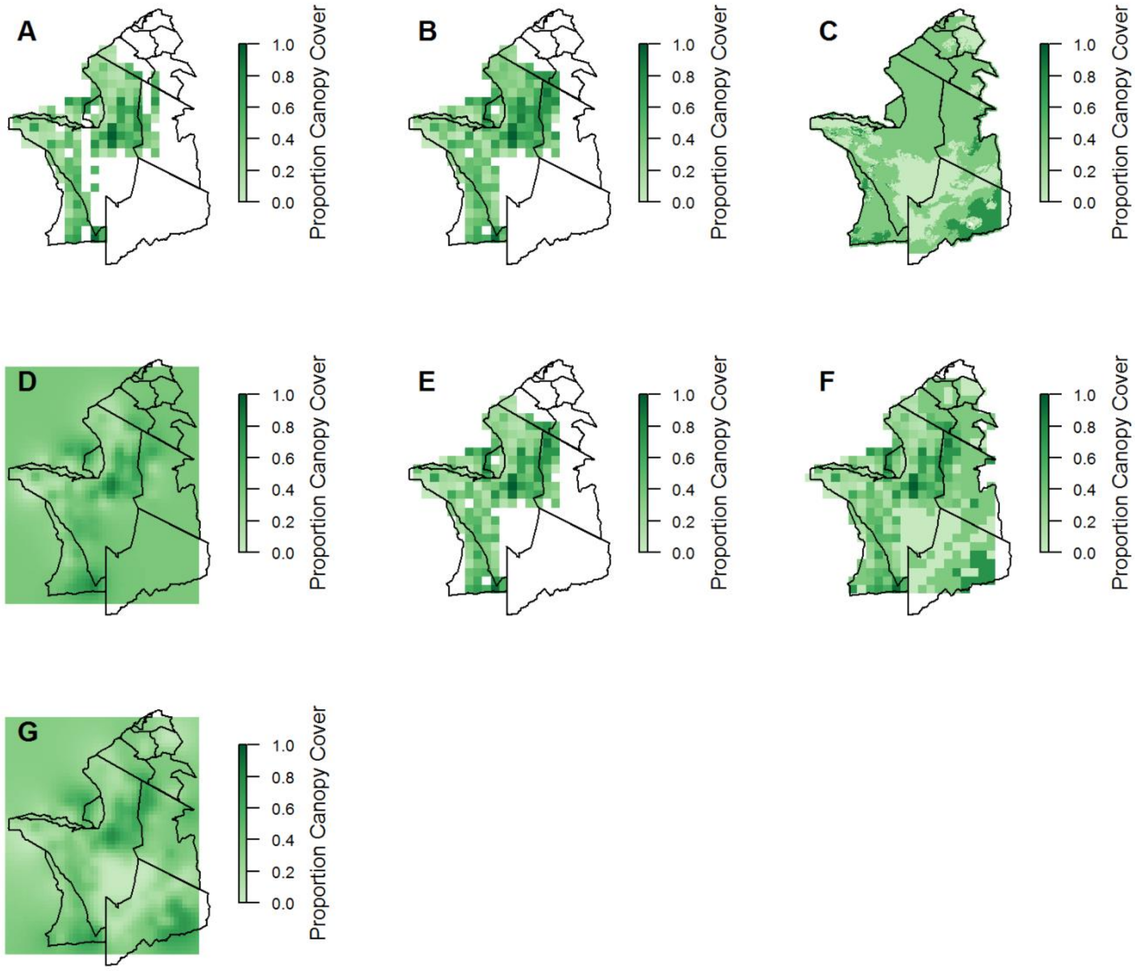

Figure D.1.1: Creating a map of canopy cover. A) 1972 canopy cover data. **B)** 1962 canopy cover data. **C)** 1999/2000 canopy cover data. **D)** 1972 data kriged onto the spatial grid used in our wildebeest movement models. **E)** 1972 data with missing values filled in with the 1962 data where available. **F)** 1972 data with missing values filled in using 1962 and 1999/2000 data where available. **G)** Data from panel **F** kriged onto the spatial grid used in the wildebeest movement models.

Since the wildebeest distribution data I am using in my analyses are from 1969-1972, it may be tempting to simply use the 1972 canopy data (Fig. D.1A), which overlap this time period, to explain the movement patterns observed in the data. However, as there are data missing, an interpolation method like kriging would be needed to fill in the gaps (Fig. D.1D), and this leads to results that are known to be inaccurate, such as the plains in the southwest of the region being given a proportion canopy cover of ~ 0.3, when this area was – as it is now – almost entirely treeless. Therefore, I instead incorporated information from all three datasets (Fig. D.1A-C) when creating a raster of canopy cover to be used in the analyses. Starting from the 1972 map, I filled in as many of the 10x10km grid cells with missing data as possible using the 1962 data (Fig. D.1E). The 1999/2000 data were then used to fill as many of the remaining missing data grid cells as possible (Fig. D.1F). In this way, I gave priority to the datasets that were collected at times that were closer to the period at which the wildebeest data were collected. Finally, I carried out ordinary kriging on this amalgamated dataset to get canopy cover values on the same grid being used in the wildebeest movement models (Fig. D.1G). Kriging was carried out using the autoKrige function from the automap package (Hiemstra et al. 2009) in R (R Core Team 2015), which tests a range of variogram models and selects the one giving the lowest residual sum of squares with the sample variogram.

D.2. Supplementary video descriptions

Supplementary videos are available online at: https://theses.gla.ac.uk/8942/

D.2.1. Supplementary Video 5.1

This video shows the changing abundances of green and dry grass estimated over the region of interest by numerical integration of the grass dynamics model described in section 5.4, alongside the wildebeest densities and monthly rainfall used as inputs to this model. The time series shown covers the period from January 1967, where I initialised the model with a grass abundance of zero, until August 1972. Note that the wildebeest data being analysed in this study were collected between August 1969 and August 1972, so that the grass was given more than 2.5 years to reach realistic levels of abundance.

In the video, rainfall is shown to change monthly, while green and dry grass abundance change daily. The wildebeest density in the region changes monthly for the initialisation period prior to August 1969, after which it begins to change daily. This is because daily wildebeest abundances for the period August 1969 to August 1972 were obtained from a GAM (Generalised Additive Model) fitted to the wildebeest distribution data, while for each month of the year in the period prior to August 1969, a wildebeest abundance map was obtained by averaging the daily estimates from the GAM for the same month in the three subsequent years. Note that I produced this video using the least complex of the GAMs fitted to the wildebeest distribution data (see Table 5.1).

The key mechanisms driving green and dry grass abundance can be observed in the video. Both green and dry grass become depleted in the presence of high wildebeest densities. If there is sufficient rainfall, the green grass recovers after the wildebeest have moved, and the dry grass recovers somewhat later when the new green grass starts to mature and dry. If there is little rainfall, depleted patches are unable to recover and any remaining areas of green grass disappear as they dry out. Since there is little rainfall in the south of the region during the dry season months, this leads to green grass periodically disappearing almost entirely from this area, while areas further north retain some rainfall and thus maintain a supply of green grass throughout. From the video, it can be seen that the patches with the most green grass are those that have previously been heavily depleted, but now have high rainfall, low wildebeest, and not enough standing dry grass to suppress new grass growth.

D.2.2. Supplementary Video 5.2

This video was developed from supplementary video 5.1 (a description is available in Appendix D.2.1) by removing the initialisation period, so that the video now covers the period from August 1969 to August 1972 (the time period of interest). In this video, I also set the grass abundances outside the area encompassing the range of the wildebeest migration to zero. This is because the levels of grass estimated outside the protected areas within which the wildebeest move are likely to have been unrealistically high (the grass dynamics model does not account for the effects of human-related activities in these areas), and also because any grass that is present in these unprotected areas is largely inaccessible to the wildebeest herds in any case. Assuming that the grass abundance is zero outside the wildebeest range prevents grass in these areas exercising an unrealistic draw on the animals within the wildebeest movement models.

D.2.3. Supplementary Video 5.3

Video showing comparisons of $\partial W/\partial t$ as estimated from the best PDE model fitted to the least complex GAM (Table 5.2; suggested to be the best PDE/GAM combination based on pAICc and pBIC (Fig.5.6)) using the associated optimised parameters (left plots) and $\partial W/\partial t$ as estimated directly from the least complex GAM by finite differencing (right plots) across the spatial region at all time points present in the original wildebeest data. The PDE used to produce the results in this video had time-varying fitted parameter values (see Fig. 5.9).

D.2.4. Supplementary Video 5.4

Video showing comparisons of $\partial W/\partial t$ as estimated from the best constant-parameter PDE model fitted to the least complex GAM (see tables in Appendix D.3) using the associated optimised parameters (left plots) and $\partial W/\partial t$ as estimated directly from the least complex GAM by finite differencing (right plots) across the spatial region at all time points present in the original wildebeest data. Note the much lower agreement between the PDE model and GAM in this video compared with supplementary video 5.3, where the parameters were time-varying, not constant.

D.3. Model comparison tables

Here, I provide a collection of tables containing AICc_{PDE} and BIC_{PDE} values equations (5.21-22) calculated for each PDE model of wildebeest movement fitted to each GAM complexity (indicated by the number of knots in the spatial marginals). The models are described based on which effects on wildebeest movement they contain: G=green grass abundance; IG=green grass intake; N=plant nitrogen concentration; W=wildebeest density; W_{max}=maximum tolerated wildebeest density. There are eight tables for each of $AICc_{PDE}$ and BIC_{PDE}, with these eight tables representing all possible combinations of constant versus time-varying parameters, local (equations (5.2-3)) versus non-local (equations (5.11-12)) versions of the models, and the small step size versus large step size schemes for the finite difference approximations of the partial derivatives from the GAMs (equation (5.20)). Values highlighted in yellow indicate the best model for a given GAM complexity for a given table, while values highlighted in blue represent the best model for a given GAM complexity over all tables. The $AICc_{PDE}$ and BIC_{PDE} values given in the tables have all had the lowest values of these statistics subtracted, so that the overall best model for each statistic has a value of zero, to aid comparison.

Table D.3.1: $AICc_{PDE}$ values for each constant parameter, local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	60241	60241	60238	60238	62128	62128	60349	60349	60239
8	199308	198929	199367	199071	199305	198895	199503	199065	199468
10	283046	283751	284780	285536	283045	283749	284875	285594	283749
12	283253	283568	283248	283453	283251	283454	283997	284114	283758
20	349213	349213	349193	349193	349211	349211	351676	351676	349211

Table D.3.2: BIC_{PDE} values for each constant parameter, local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	58736	58736	58724	58724	60614	60614	58835	58835	58725
8	197803	197424	197853	197557	197791	197381	197988	197551	197954
10	281542	282246	283266	284022	281531	282235	283361	284080	282235
12	281748	282063	281734	281939	281736	281939	282483	282600	282244
20	347708	347708	347679	347679	347697	347697	350162	350162	347697

Table D.3.3: AICc$_{PDE}$ values for each time-varying parameter, local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	34073	35459	36568	36568	43554	45385	38911	38803	41142
8	139350	123861	139995	130560	146461	131977	143846	128560	176894
10	243068	240699	251849	246864	254418	252218	253071	248867	258326
12	248166	233814	258011	242477	254899	240174	259535	242151	257785
20	293582	511387	283348	276893	284755	271333	306204	298635	279451

Table D.3.4: BIC$_{PDE}$ values for each time-varying parameter, local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	34064	35449	36250	36251	43236	45067	38593	38486	40824
8	139341	123851	139678	130242	146143	131659	143529	128242	176576
10	243059	240689	251532	246547	254100	251900	252753	248549	258009
12	248157	233805	257694	242159	254581	239856	259217	241833	257467
20	293573	511378	283030	276576	284437	271016	305886	298317	279134

Table D.3.5: AICc$_{PDE}$ values for each constant parameter, non-local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	61716	60649	61714	60744	61714	60647	62011	61284	61806
8	198995	197986	199076	198083	198993	197984	198993	198069	199323
10	283511	282591	284217	283102	283509	282589	283962	283312	283913
12	282060	281807	282223	281802	282058	281940	282337	282727	282738
20	339497	334942	339330	334750	339999	335713	350008	347229	339757

Table D.3.6: BIC$_{PDE}$ values for each constant parameter, non-local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	60220	59153	60209	59239	60209	59142	60506	59779	60301
8	197500	196491	197571	196578	197489	196479	197489	196564	197818
10	282015	281095	282712	281597	282004	281084	282458	281807	282409
12	280565	280312	280719	280298	280553	280435	280832	281222	281234
20	338002	333446	337825	333246	338494	334208	348504	345725	338252

Table D.3.7: AICc$_{PDE}$ values for each time-varying parameter, non-local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	19670	25479	22745	27079	22376	27897	35795	40614	31420
8	137585	150311	141010	152289	145819	159389	151856	172742	164766
10	232148	235540	237575	240167	242909	249256	243399	251486	249015
12	239348	237420	248042	245135	248131	246501	261881	256819	243907
20	263108	266710	273844	275106	276946	276116	288635	284638	278017

Table D.3.8: BIC$_{PDE}$ values for each time-varying parameter, non-local PDE model fitted to each GAM complexity using the small step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	19670	25479	22437	26771	22067	27588	35486	40306	31112
8	137585	150311	140702	151981	145511	159081	151548	172434	164458
10	232148	235540	237267	239859	242601	248948	243091	251177	248706
12	239348	237420	247734	244827	247823	246192	261573	256510	243599
20	263108	266710	273535	274798	276638	275807	288327	284329	277709

Table D.3.9: AICc$_{PDE}$ values for each constant parameter, local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	42034	41935	42102	42092	42070	41964	42387	42362	42032
8	130829	130545	130827	130543	130828	130545	130838	130551	130827
10	218440	218487	220389	220427	218438	218485	220446	220494	218485
12	210487	210430	210485	210427	210485	210428	210526	210479	210485
20	257839	257881	257834	257876	257837	257879	258269	258307	257891

Table D.3.10: BIC$_{PDE}$ values for each constant parameter, local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	40529	40431	40587	40578	40556	40449	40873	40848	40518
8	129324	129041	129313	129029	129313	129031	129324	129037	129313
10	216936	216983	218875	218913	216924	216971	218932	218980	216971
12	208982	208925	208971	208913	208971	208914	209012	208965	208971
20	256334	256376	256320	256362	256323	256365	256755	256793	256377

Table D.3.11: AICc$_{PDE}$ values for each time-varying parameter, local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	26374	26519	29257	28742	27183	27124	32764	33315	28258
8	112219	108297	114386	110543	113543	109631	119874	117977	114406
10	201384	201552	206743	208669	203649	203858	209866	213102	204514
12	191574	189735	199161	197967	193210	191422	201522	200363	193201
20	234911	235491	240919	239047	236214	236115	248151	249047	239251

Table D.3.12: BIC$_{PDE}$ values for each time-varying parameter, local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	26365	26509	28939	28425	26865	26806	32447	32997	27940
8	112209	108287	114068	110226	113225	109313	119556	117659	114088
10	201375	201542	206426	208352	203332	203541	209548	212784	204196
12	191564	189725	198843	197649	192892	191104	201205	200045	192883
20	234902	235482	240601	238729	235896	235798	247833	248730	238933

Table D.3.13: AICc$_{PDE}$ values for each constant parameter, non-local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	41734	40539	41732	40618	41732	40537	41994	40941	41783
8	130298	129490	130366	129578	130296	129488	130315	129574	130665
10	218833	218406	219884	219712	218831	218404	218925	219504	218945
12	208434	209126	208641	209201	208432	209209	208588	209570	209273
20	257649	257629	257775	257626	257830	257857	257660	257680	257902

Table D.3.14: BIC$_{PDE}$ values for each constant parameter, non-local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	40239	39044	40228	39113	40228	39032	40489	39436	40279
8	128803	127995	128862	128074	128792	127984	128810	128069	129160
10	217338	216910	218379	218207	217326	216899	217420	218000	217441
12	206938	207630	207136	207696	206927	207704	207083	208065	207768
20	256154	256133	256270	256121	256325	256352	256155	256175	256398

Table D.3.15: AICc$_{PDE}$ values for each time-varying parameter, non-local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	0	1367	4941	4824	2517	4186	17240	18229	6602
8	93111	90629	96523	94319	99632	96729	98810	101145	100529
10	187068	186794	193081	192868	191019	193337	197238	199634	194952
12	175073	173381	183704	181482	178957	178661	187977	188575	179237
20	220452	227835	229196	234668	227497	234277	234377	239869	232832

Table D.3.16: BIC$_{PDE}$ values for each time-varying parameter, non-local PDE model fitted to each GAM complexity using the large step size finite differencing scheme.

	Model								
Knots in Space	**I_G + N + W + W_{max}**	**G + N + W + W_{max}**	**I_G + N + W**	**G + N + W**	**I_G + W + W_{max}**	**G + W + W_{max}**	**I_G + N + W_{max}**	**G + N + W_{max}**	**N + W + W_{max}**
6	0	1367	4633	4515	2209	3877	16932	17921	6293
8	93111	90629	96215	94011	99324	96421	98501	100837	100220
10	187068	186794	192773	192560	190711	193029	196930	199326	194644
12	175073	173381	183396	181173	178649	178352	187669	188267	178928
20	220452	227835	228888	234360	227189	233969	234068	239561	232524