Follow the leader: Collective cell migration in cancer and developmental biology

William Duncan Martinson
Balliol College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2022

In loving memory of Marj Martinson (1929 – 2021), a wonderful Nana.

Acknowledgements

Firstly, I would like to thank my supervisors Professors Philip Maini and Helen Byrne. You have both been wonderful mentors, and I am especially grateful that you encouraged me to continue pursuing post-graduate research by transferring to the DPhil program here. Without your support and guidance, this thesis would certainly not exist. It is also thanks to you that I gained the confidence and skills to be a mathematical biologist. Thank you.

I would also like to thank the Keasbey Memorial Foundation and the Oxford Mathematical Institute, who funded my studies at Oxford these past few years.

I am grateful to those individuals who have helped contribute to this work presented in this thesis. In particular, I want to thank Hirokazu Ninomiya, who suggested looking for self-similar solutions for the work in Chapter 3 and provided insightful comments; Paul Kulesa, Rebecca McLennan, and Mary McKinney, whose experimental evidence on neural crest cell migration inspired and guided the development of the agent-based model in Chapters 4 and 5; Ruth Baker, who provided valuable insights into different techniques that could be used to model neural crest cell migration; and Lance Davidson, who helped inspire several *in silico* tests of the neural crest cell agent-based model and supplied information about fibronectin and cell remodelling of the extracellular matrix.

I am eternally grateful to my officemate Roisin Stephens: you are the closest friend I have in Oxford, and I want to thank you for putting up with my many distractions over the course of this DPhil and for providing some of the most memorable moments I have here (specifically the carrots incident). I would also like to thank the other members of the Wolfson Centre for Mathematical Biology for being almost like a second family. You all are a part of the best graduate program to which I could ever belong.

I am indebted to Mimi and Paul Harris, who took me into their home at the start of the Covid-19 pandemic in March 2020 and treated me like a member of their family for three whole months. You are the greatest aunts and uncles in the world, and I am grateful everyday that you were always close by to visit during my studies. Thank you so much to you and to Elizabeth, Mary, Eleanor, Joseph, James, and Peter.

Last, but certainly not least, I would like to thank my own family (Monica, Marty, Andrew, Peter, Anne). I know that having a 24-25 year old living at home is not usually a fun experience, but I enjoyed being with you all again (and playing fetch with Duke and Dutchess) during the pandemic. Thank you for supporting me since the day I was born, and thank you for the amazing experiences we have had together.

Abstract

Collective cell migration is an essential phenomenon in biology, occurring in diverse processes such as neural crest stem cell migration and angiogenesis (the formation of new blood vessels from pre-existing ones). One particular mode of collective cell migration involves two distinct phenotypes: leader cells, which migrate towards external biophysical signals, and follower cells, which move in response to leader-created cues. Mechanisms underpinning leader-follower migration remain poorly characterised in general, partly because they occur over multiple spatial and temporal scales. Mathematical modelling can address such issues, as it can be used to tractably evaluate different biological hypotheses, connect individual cell interactions to the behaviour of collectives, and guide the design of *in vivo* experiments. Open questions remain, however, regarding the validity of certain mathematical models and the conditions for which they accurately capture data from biological experiments.

In this thesis, we address different questions relating to leader-follower migration within the contexts of angiogenesis and neural crest cell migration. By comparing agent-based model simulations for angiogenesis to solutions from both a classical phenomenological framework and an alternative coarse-grained model, we identify parameter regimes for which the phenomenological framework accurately describes results of the other two systems. In doing so, we uncover implicit physical assumptions that underpin the continuum models. We also collaborate with experimental colleagues to develop a new agent-based model for cranial neural crest cell migration that incorporates cell remodelling of, and movement through, a naïve extracellular matrix. We perform a global sensitivity analysis on the model and rank the mechanisms that most dominate collective cell behaviour. Further tests of this model suggest that migration is most efficient when leader and follower cell phenotypes are not identical. We conclude by providing, to our knowledge, the first detailed investigation of mechanisms that ensure robust neural crest cell migration.

Statement of Originality

Some of the work presented in Chapter 2 has previously appeared in the following journal article, of which I am the first author: **Martinson W.D.**, Byrne H.M., Maini P.K. (2020). "Evaluating snail-trail frameworks for leader-follower behavior with agent-based modeling." Physical Review E 102 (6), 062417. The text and analysis presented in that chapter are wholly my own work. I have clearly indicated the sources of the software libraries I have used to generate numerical solutions in the text.

Some of the work presented in Chapter 3 has previously appeared in the following journal article, of which I am the first author: **Martinson W.D.**, Ninomiya H., Byrne H.M., Maini P.K. (2021). "Comparative analysis of continuum angiogenesis models." Journal of Mathematical Biology 82 (21), online. DOI: 10.1007/s00285-021-01570-w. We thank Dr. Hirokazu Ninomiya (Meiji University, Japan) for suggesting that self-similar solutions exist for the leading order angiogenesis models presented in that chapter. The specific equations and derivations in that chapter, however, are wholly my own work.

The development of the neural crest cell agent-based model, described in Chapters 4 and 5, was carried out in collaboration with my supervisors Prof. Philip Maini and Prof. Helen Byrne, in addition to Prof. Ruth Baker (University of Oxford), Dr. Paul Kulesa of Stowers Institute for Medical Research (Kansas City, MO, USA), and Dr. Lance Davidson of the University of Pittsburgh (USA). The text and analysis in those chapters are wholly my own work. I have clearly indicated the sources of any software libraries that I have used within the text.

List of Abbreviations

ABM	agent-based model
ODE	ordinary differential equation
PDE	partial differential equation
P–ABM	Pillay *et al.* agent-based model developed in Pillay *et al.* (2017)
ST–PDE	snail-trail PDE model
P–PDE	Pillay *et al.* PDE model derived in Pillay *et al.* (2017)
TC	endothelial tip cell
EC	endothelial stalk cell
TAF	tumour angiogenic factor
VEGF	vascular endothelial growth factor
NCC	neural crest cell
ECM	extracellular matrix
FN	fibronectin
NT	neural tube
CIL	contact inhibition of locomotion
LHS	latin hypercube sample/sampling
PRCC	partial rank correlation coefficient
eFAST	extended Fourier amplitude sensitivity test/testing
VE	volume exclusion
CG	contact guidance

Contents

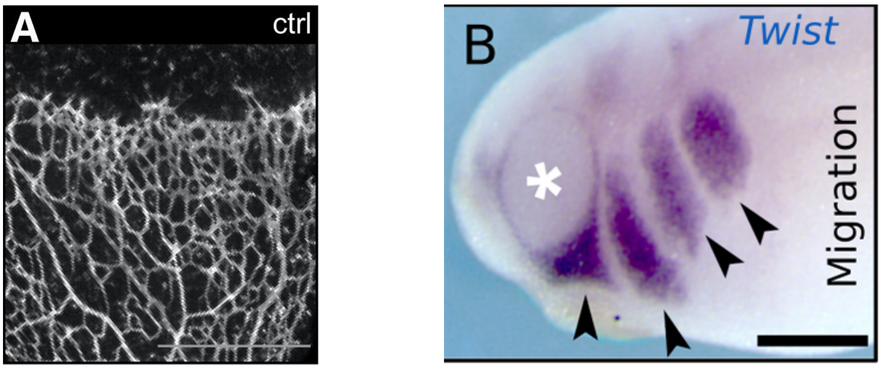
Chapter 1

Introduction

Collective cell migration – the coordinated movement of multiple cells – constitutes an essential phenomenon in biology. It contributes to tissue morphogenesis during development, for instance via the migration of neural crest stem cells (Giniunaite et al., 2019a), and helps to establish homeostasis in mature organisms through processes such as sprouting angiogenesis, in which new blood vessels are created from pre-existing ones (Duran et al., 2017). Collective cell migration represents a paradigm of emergent behaviour, in the sense that groups of cells adopt robust, intricate behaviours that are not observed in individuals and which result from their biophysical interactions (Shellard and Mayor, 2020).

In this thesis, we focus on a specific type of migration in which one can distinguish between two cell phenotypes. The first subtype, "leader" cells, exhibits unique morphological and transcriptomic features that enable efficient perception of – and migration towards – external biological, chemical, and/or physical cues. The second, "follower" cells, recognises and moves in response to signals created by leaders. We differentiate leader-follower migration from other types of collective behaviour such as sheet migration, where cells more tightly adhere to each other and move as a two-dimensional structure, and border cell migration, where a group of cells detaches from a cluster and migrates as a single unit (Rørth, 2009; Horne-Badovinac and Bilder, 2005).

Leader-follower migration admits a diverse array of migratory patterns. Thin cell chains arise during sprouting angiogenesis when leader cells, known as "tip cells" (TCs) in this context, move up spatial gradients of soluble tumour angiogenic factors (TAFs). Follower



(a) Angiogenesis network (b) Neural crest cell streams

Figure 1.1: Photographs of (a) vascular networks created by angiogenesis, and (b) neural crest cell streams in *Xenopus* embryos. In (a), the image shows grey-stained retinal vessels of mice pups five days past birth after being treated with DMSO (scale bar: 50 μm). In (b), *in situ* hybridisation stains for cells expressing Twist, a protein marker for NCCs (scale bar: 300 μm, white asterisk: eye/optic vesicle, black arrowheads: NCC streams). Photo (a) is cropped and reprinted from Figure 4 of Cell Metabolism, 19, Sandra Schoors, Katrien De Bock, Anna Rita Cantelmo, Maria Georgiadou, Bart Ghesquire, Sandra Cauwenberghs, Anna Kuchnio, Brian W. Wong, Annelies Quaegebeur, Jermaine Goveia, Francesco Bifari, Xingwu Wang, Raquel Blanco, Bieke Tembuyser, Ivo Cornelissen, Ann Bouché, Stefan Vinckier, Santiago Diaz-Moralli, Holger Gerhardt, Sucheta Telang, Marta Cascante, Jason Chesney, Mieke Dewerchin, Peter Carmeliet, Partial and Transient Reduction of Glycolysis by PFKFB3 Blockade Reduces Pathological Angiogenesis, 37-48, Copyright (2014), with permission from Elsevier. Figure (b) cropped from Figure 1 in PLoS Computational Biology, 15, András Szabó, Eric Theveneau, Melissa Turan, Roberto Mayor, Neural crest streaming as an emergent property of tissue interactions during morphogenesis, e1007002, Copyright (2019), reprinted under terms of the Creative Commons Attribution License (CC BY 4.0).

cells, known as endothelial stalk cells (ECs) in this context, follow the path of TCs by proliferating and tightly adhering to cell membranes. Both TCs and ECs may branch to create new sprouts; the combined actions of sprout formation and extension establish the new capillary network (Carmeliet and Jain, 2011). Alternatively, leader-follower migration may generate wider, ribbon-like streams of cells. This is the case during development, when certain stem cells known as neural crest cells (NCCs) migrate from the neural tube to different parts of the embryo (Szabó and Mayor, 2018; Kulesa et al., 2010, 2004). As in angiogenesis, transcriptomic analysis of NCCs suggests a dichotomy between leading cells at the front and trailing cells closer to the neural tube (Morrison et al., 2017; McLennan et al., 2015a). The streams consist of dense, loosely connected NCCs but are distinct, with cell-free regions bordering the migratory pathways (Szabó et al., 2019; Trainor and Tam, 1995; Lumsden et al., 1991). Figure 1.1 presents examples of patterns that arise both in angiogenesis and cranial NCC migration.

Given the importance of leader-follower migration to embryonic development and tissue homeostasis, it is unsurprising that disruptions to this process can have serious medical consequences. The over-stimulation of angiogenesis during tumour growth, for example, supplies oxygen to hypoxic tumours, promotes further cancer cell proliferation, and facilitates metastasis (Flegg et al., 2020; Carmeliet and Jain, 2011). Furthermore, the vessels created during tumour-induced angiogenesis consist of tortuous, disordered structures that hamper efficient drug delivery (De Palma et al., 2017; Tong et al., 2004). Similarly, since NCCs can differentiate into a variety of different cell types such as melanocytes, bone and cartilage cells, and ganglia of the peripheral nervous system (Tang and Bronner, 2020), failure of NCC migration can lead to the acquisition of developmental defects known as neurocristopathies, which include Meckel-Gruber syndrome and Von Hippel-Lindau syndrome (Vega-Lopez et al., 2018).

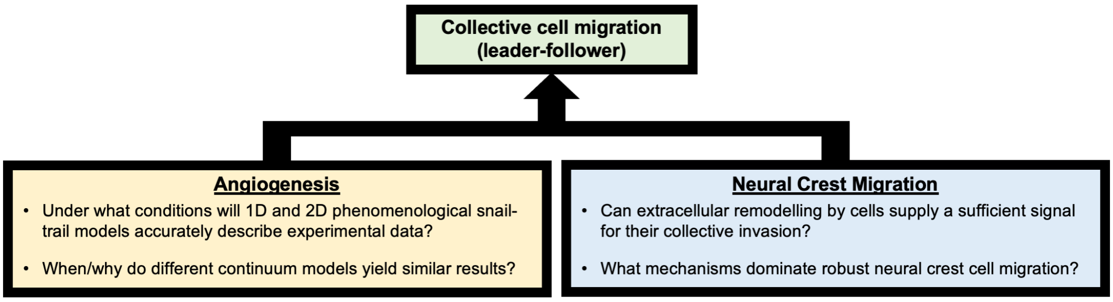
Insights into the mechanisms that drive collective cell migration may help researchers develop new therapeutic strategies to combat these diseases. The information obtained from studying collective cell migration in one biological context may also improve our understanding of its role in another: previous research, for instance, suggests a strong relationship between melanoma metastasis and NCC migration. Indeed, the cancerous cells adopt a phenotype very similar to migratory NCCs upon explantation to the neural tube of a developing embryo, and the two cell lines express many of the same genes (Kulesa et al., 2021, 2013, 2006; Bailey et al., 2012).

Complete understanding of how different physical, chemical, and biological signals influence collective cell migration remains elusive, however, due to the range of temporal and spatial scales on which these processes act. Endothelial cells involved in sprouting angiogenesis, for instance, can travel distances on the order of millimetres over a period of weeks (Nowak-Sliwinska et al., 2018). Similarly, NCCs migrate several millimetres from the neural tube over a relatively short time period of days (McKinney et al., 2020; Hutchins et al., 2018; Theveneau and Mayor, 2012; LeDouarin and Kalcheim, 1999). Cells may even sense cues at distances far from their centres of mass by extending multiple filopodia in their microenvironment (Fischer et al., 2019; Szabó and Mayor, 2018; Teddy and Kulesa, 2004).

Mathematical modelling of collective cell migration can help to resolve these multiscale dynamics, as it provides a framework within which to investigate multiple biological hypotheses and offers a tractable means of connecting individual cell mechanisms (e.g., chemotaxis, haptotaxis, cell-cell contact) to changes in collective behaviours (Kulesa et al., 2021; Flegg et al., 2020; Giniunaite et al., 2019b; Byrne, 2010). In doing so, these theoretical approaches can generate ideas for new *in vivo* experiments, increase understanding of the mechanisms driving collective cell migration, and inspire new strategies to regulate it.

Open questions remain, however, regarding the applicability of these abstract descriptions to the study of collective cell migration. Many mathematical models that describe cells in terms of continuous quantities, for instance, rely on techniques such as mean-field approximations to derive differential equations for their dynamics. While this approach yields closed-form systems, it requires that no spatial correlations exist between individuals (Davies et al., 2014; Markham et al., 2013; Simpson and Baker, 2011). This assumption may not be valid when groups of cells migrate together, however, and the parameters for when such mean-field models break down are not well characterised. Theoretical descriptions that instead represent cells as discrete entities are not immune from limitations either: because of their stochastic nature, they tend to be computationally intensive and less amenable to analysis (An et al., 2009). The resolution of such issues may establish conditions for which the mathematical models will accurately describe the underlying biology and thereby make useful predictions.

This thesis aims to address different questions relating to the phenomenon of leader-follower migration and its mathematical modelling, using angiogenesis and NCC migration as paradigms. Our work on angiogenesis focusses on issues regarding the applicability of different modelling frameworks to collective cell migration, and identifies conditions and regions of parameter space for which such descriptions are, in some sense, "equivalent" to each other. Our investigations of NCC migration, by contrast, use mathematical models to determine which biological mechanisms likely dominate that process and its robust pattern



- Under what conditions will 1D and 2D phenomenological snail-trail models accurately describe experimental data?
- When/why do different continuum models yield similar results?

- Can extracellular remodelling by cells supply a sufficient signal for their collective invasion?
- What mechanisms dominate robust neural crest cell migration?

Figure 1.2: A schematic detailing the questions we investigate in this thesis, which relate to the mathematical modelling of leader-follower migration in the contexts of angiogenesis and neural crest cell migration. Questions regarding models for angiogenesis investigate the frameworks themselves, while those related to neural crest migration apply mathematical modelling to gain a better biological understanding of the mechanisms underpinning that process.

forming behaviour. As part of this study, we apply global sensitivity analysis techniques (detailed in §1.3) to determine how model outputs are affected by perturbations to ABM parameter values. Figure 1.2 presents a schematic of the questions we address in angiogenesis and NCC migration, and how they relate to the broader context of leader-follower migration. In the remaining sections of this chapter, we detail the biology behind tumour-induced angiogenesis and NCC migration, survey and critique mathematical models that have been developed to describe them, and summarise several techniques for global sensitivity analysis. We conclude by outlining the content of the remaining thesis chapters.

1.1 Tumour-induced angiogenesis

1.1.1 Biological background

In this thesis, we use the term “angiogenesis” to specifically refer to sprouting angiogenesis – the formation of new vessels due to TC and EC migration from existing vasculature. This process differs from other modes of angiogenesis such as intusseptive angiogenesis, in which new vessels arise from vessel splitting and no bulk cell movement takes place (Uccelli et al., 2019; Viallard and Larrivée, 2017; Risau, 1997).

During cancer, angiogenesis is stimulated by hypoxic conditions within the interior of solid tumours. Cancer cells initially exist in an avascular environment, where the diffusion of nutrients, such as oxygen, largely determines the extent of their proliferation (Folkman and

Hochberg, 1973). Continuous tumour growth increases the metabolic stress on cells within its core: when the mass reaches a size that exceeds the diffusion limit of oxygen, these interior cells begin to starve. This metabolic stress leads to the accumulation of activated hypoxia-inducible transcription factors (HIFs); these proteins upregulate the expression of several genes that enable cells to adapt their phenotype to one that is better suited for hypoxic conditions (Samanta and Semenza, 2018; Pugh and Ratcliffe, 2003; Maxwell et al., 1997). One of these upregulated genes encodes a protein known as vascular endothelial growth factor (VEGF), which is a major regulator of angiogenesis (Fong, 2008). Hypoxic cells secrete multiple isoforms of VEGF into their microenvironment, some of which are soluble and diffuse freely before binding to membrane receptors on the endothelial cells of existing vessels (Uccelli et al., 2019; Duran et al., 2017).

The binding of VEGF triggers a signalling cascade within endothelial cells that is regulated by the molecules Notch and Delta-like ligand 4 (Delta). This cascade serves to partition endothelial cells lining the pre-existing vessel into a motile tip cell phenotype (corresponding to high levels of Delta/low levels of Notch) or a more proliferative endothelial stalk cell phenotype (corresponding to low Delta/high Notch) (Vega et al., 2020; Blanco and Gerhardt, 2013; Bentley et al., 2008; Hellström et al., 2007). TCs extend several filopodia, which they use to sense VEGF and other molecules present in their microenvironment. They migrate up increasing gradients of soluble VEGF via a process known as chemotaxis (Flegg et al., 2020; Betz et al., 2016; Potente et al., 2011; Gerhardt et al., 2003; Bowersox and Sorgente, 1982). ECs, meanwhile, proliferate along the path of TCs in order to extend the sprout and establish the new vessel lumen. In certain cases, microenvironmental cues may stimulate ECs more distant from the growing sprout to adopt a TC phenotype, thus adding more branches to the network. Alternatively, new sprouts can form through a dynamic re-shuffling of TCs and ECs (Stepanova et al., 2021; Ubezio et al., 2016; Arima et al., 2011; Jakobsson et al., 2010). Closed loops in the network form through a process known as anastomosis, in which a TC fuses with another TC or EC (Moreira-Soares et al., 2018; Nesmith et al., 2017; Lenard et al., 2013). Other cells, known as pericytes, help to regulate signalling pathways that determine TC/EC phenotypes, while smooth muscle cells assist ECs in supporting and maintaining the new vessel (Kang et al., 2019; Betz et al., 2016). As TCs and ECs grow closer to the tumour, branching becomes more frequent; this "brush-border" effect serves to increase the vascularity of the tumour and supply additional nutrients to the cancerous mass (Byrne and Chaplain, 1995; Muthukkaruppan et al., 1982).

Aside from VEGF, other molecular regulators of angiogenesis exist to guide the morphogenesis of the new capillary network. TAFs such as angiopoietin-2 and fibroblast growth factor, for instance, act alongside VEGF to guide TCs during migration (Duran et al., 2017; Carmeliet and Jain, 2011; Folkman and Klagsbrun, 1987). Interactions between migrating cells and the extracellular matrix (ECM) may also significantly contribute to the angiogenic process. TCs remodel the ECM during their migration by secreting both matrix metalloproteases, which degrade ECM proteins, in addition to fibronectin, a constituent ECM protein that can serve as a substrate for cells during haptotaxis – the migration of cells up increasing adhesive gradients (Flegg et al., 2020; De Palma et al., 2017; Duran et al., 2017; Paweletz and Knierim, 1989). Recent *in vitro* studies also suggest that fibronectin binds VEGF in acidic environments, and may affect cell phenotypes by triggering signalling pathways that involve such molecules; this hypothesis has yet to be evaluated *in vivo*, however (Usuelli et al., 2021).

1.1.2 Mathematical modelling of angiogenesis

Many mathematical models for angiogenesis now exist to provide insight into how signals such as chemotaxis and haptotaxis can guide cell migration (Flegg et al., 2020; Vilanova et al., 2017; Scianna et al., 2013; Mantzaris et al., 2004). These representations may be roughly classified as discrete or continuum models. The latter frameworks approximate quantities of interest using continuous variables and describe their dynamics with differential equations. These models most commonly describe macroscopic (tissue-level) observations, such as the total cell density, and may not always resolve individual cells. By contrast, discrete approaches describe the behaviour of single "agents" (e.g., cells or molecules), and, as such, tend to represent phenomena that occur at a microscopic (cellular) scale (Metzcar et al., 2019; Van Liedekerke et al., 2015; An et al., 2009). The continuum and discrete frameworks are not mutually exclusive: they may be combined to create so-called “hybrid models” that are often used to study interactions at the microscopic scale (Rejniak and Anderson, 2011).

Early mathematical descriptions for angiogenesis largely belong to a macroscopic continuum framework, known as the “snail-trail”, that tracks average TC and EC densities over space and time. Snail-trail models aim to reproduce results from assays in which angiogenesis is triggered by implanting a pellet of VEGF in the cornea of rabbits (Nowak-Sliwinska et al., 2018; Stahl et al., 2010; Auerbach et al., 2003). As vessels grow towards the pellet, investigators record pictures of the new networks and measure the density of vessels. Balding and McElwain (1985) developed the snail-trail model by adapting an existing framework for slime mould migration (Edelstein, 1982), which exhibits branching networks similar to those observed in corneal assays. Snail-trail models rely on phenomenological arguments for their construction – that is, their dynamics are inferred from tissue-level observations. Their defining feature is that the vessel density increases at a rate proportional to the flux of TCs; this modelling assumption is motivated by the experimental observation that vessels are created along the paths that TCs travel. In other words, if $\mathbf{J}$ denotes the net TC flux vector in two-dimensional (2D) planar geometry, then the rate of increase in vascular density, $\rho(\mathbf{x}, t)$, is

$$\frac{\partial \rho}{\partial t} = \|\mathbf{J}\|_2 = \sqrt{\mathbf{J} \cdot \mathbf{J}} = \sqrt{(J_x)^2 + (J_y)^2}, \tag{1.1}$$

where $\|\cdot\|_2$ denotes the Euclidean norm and J_x and J_y denote the x- and y-components of the net TC flux, respectively. The norm in Eq. (1.1) ensures that the vessel density remains non-negative for all simulation times (Jackson and Zheng, 2010).

Byrne and Chaplain (1995) extended the basic snail-trail framework by considering the impact of random motility on vessel formation in addition to mechanisms such as tip-to-sprout anastomosis and TAF-dependent branching. Other investigators contributed to the snail-trail framework by accounting for additional phenomena such as tip-to-tip anastomosis, tip cell death, and haptotaxis (Orme and Chaplain, 1997; Pettet et al., 1996b). More recently, Connor et al. (2015) applied snail-trail models to determine how multiple

TAFs affect TC guidance during migration. These models helped identify potential clinical targets for anti-angiogenic therapies by predicting how they would prevent or control vessel growth towards tumours.

Although snail-trail models represent a simple description of angiogenesis and have increased the understanding of the mechanisms that dominate its progression, they suffer from several disadvantages. Since snail-trail models track spatially averaged quantities, they cannot resolve individual vessels. Additionally, it is difficult to relate the parameters of a snail-trail framework to experimental results obtained at a microscopic scale because of their phenomenological construction. Discrete and hybrid models for angiogenesis have been created to address such shortcomings and, in recent years, have become increasingly prevalent (Metzcar et al., 2019; Van Liedekerke et al., 2015). Stokes and Lauffenburger (1991) considered one of the first discrete models for vessel growth by proposing a stochastic partial differential equation (PDE) for TC movement. Variations of this approach continue to be applied in hybrid models of angiogenesis, in part because they can produce branching networks similar to those observed *in vivo*. One such framework describes TC dynamics via a stochastic PDE, but models EC densities with a continuous phase-field approach due to the relatively large number of ECs that arise during angiogenesis (Vilanova et al., 2013; Travasso et al., 2011). In recent years, this modelling framework has been extended to consider the effects of haptotaxis, interstitial fluid flow within the ECM, and chemotaxis on network formation (Vilanova et al., 2014, 2018).

Alternatively, hybrid frameworks may represent the dynamics of both TCs and ECs via stochastic PDEs, while using differential equations to describe the spatiotemporal evolution of molecules such as VEGF (Bonilla et al., 2014). These frameworks have also been used to investigate the roles of chemotaxis and haptotaxis in angiogenesis. They may even be analytically tractable: Bonilla et al. (2016; 2020), for instance, derived closed-form expressions for one- and two-dimensional solitary TC travelling wave solutions from such models. A third subset of hybrid frameworks consists of physically motivated, force-based descriptions of cell dynamics; these have been used to investigate TC and EC migration in three spatial dimensions (Perfahl et al., 2017).

While agents in the above frameworks can occupy any spatial location, discrete on-lattice agent-based models (ABMs) for angiogenesis restrict cells to a finite set of locations within the domain. Anderson and Chaplain (1998) derived one of the first on-lattice ABMs for angiogenesis – a single cell occupancy cellular automaton – from a continuous PDE for cell dynamics. Cellular automata continue to be widely used to study tumour vascularisation, as they are usually simpler to implement than other discrete models (Pillay et al., 2017; Jackson and Zheng, 2010; Qutub and Popel, 2009). Cellular Potts models extend the on-lattice approach by allowing an individual cell to occupy multiple lattice sites (Vega et al., 2020; Boas et al., 2018; van Oers et al., 2014; Daub and Merks, 2013; Scianna et al., 2011; Bauer et al., 2007). Since cellular Potts models can capture fluctuations in cell morphologies (similar to certain off-lattice models) and can be used to investigate interactions between cell protrusions and the ECM, this framework has become a popular on-lattice approach for modelling angiogenesis. Lattice-based models need not resolve individual cells, however: for example, compartment-based frameworks adopt a coarser, mesoscale approach by allowing multiple cells to occupy a single lattice site. To our knowledge, such descriptions of angiogenesis have only been considered by Spill et al. (2015), where they were used to derive macroscopic snail-trail models.

Specialised software exists to streamline the simulation of on- and off-lattice models for angiogenesis. These libraries, which include Ya||a, Microvessel Chaste, PhysiCell, and CompuCell3D (to name but four), contain templates for one or several of the discrete and hybrid frameworks approaches listed above, and thus allow investigators to more quickly implement them in a computational setting (Germann et al., 2019; Ghaffarizadeh et al., 2018; Grogan et al., 2017; Swat et al., 2009). Furthermore, many of these libraries parallelise simulations, allowing modellers to more efficiently generate solutions.

Despite the advantages and wide use of ABMs, understanding remains limited about how different parameter values and agent interaction rules affect their solutions over large spatial and temporal scales (An et al., 2009). Osborne et al. (2017), for instance, showed that model-specific assumptions underpinning several types of discrete descriptions for cell sorting (including cellular automata, cellular Potts models, overlapping spheres models, Voronoi tessellations, and vertex models) could affect the biological conclusions drawn from them. Although such studies represent a much needed first step for evaluating discrete approaches and their underlying rules, it is challenging to gain rigorous insight into the general, long-time behaviour of ABMs unless there is a common mathematical framework within which to compare them.

Macroscopic PDEs represent a possible candidate for this framework. They can be derived by coarse-graining discrete ABMs to continuous PDEs and can describe how the average distribution of agents in a discrete model evolves over time and space (Chaplain et al., 2020; Motsch and Peurichard, 2018; Buttenschön et al., 2018; Bonilla et al., 2016; Dyson and Baker, 2015; Markham et al., 2013; Baker et al., 2010; Othmer et al., 1988). Since continuum models are often easier to analyse and simulate than discrete ones, they can be used to (indirectly) determine how the microscopic rules of a discrete model generate complex collective dynamics. For instance, Pillay et al. (2017) derived a coarse-grained system of PDEs for angiogenesis from an idealised on-lattice ABM, thus linking continuum parameters to those of the discrete model. The same authors later extended their methodology to determine how volume exclusion between cells affects their macroscopic dynamics (Pillay et al., 2018). Mean-field models derived from discrete-to-continuum approaches, however, tend to be nonlinear, with the degree of nonlinearity depending on the rules of the underlying discrete model (Penington et al., 2014; Bruna and Chapman, 2012; Penington et al., 2011; Simpson et al., 2009). To our knowledge, little attention has focussed on the systematic comparison of nonlinear mean-field models derived from discrete-to-continuum approaches. We shall address issues related to the validity of mean-field approximations and the similarity between continuum models in Chapters 2 and 3 of this thesis, where we will use numerical methods and asymptotic analysis to compare solutions from a discrete ABM, the classical snail-trail model, and a coarse-grained model for angiogenesis.

1.2 Neural crest cell migration

1.2.1 Biological background

The neural crest is an ideal system for studying long-distance collective cell migration, as over the course of hours cells colonise domains that grow to about 1000 μm in length. While neural crest cell migration occurs in multiple vertebrate species, how it is accomplished varies between species and even between different cell populations within an individual (Shellard and Mayor, 2020; McLennan et al., 2017; Richardson et al., 2016; Szabó et al., 2016). NCCs originate from the outer ectoderm germ layer of developing embryos, and form when the neural plate folds on itself following gastrulation to create the neural tube (Szabó and Mayor, 2018; Bronner and LeDouarin, 2012; Kulesa et al., 2010; Selleck and Bronner-Fraser, 1995). Cells initially exhibit an epithelial-like phenotype: they remain largely stationary, and tightly adhere to each other using membrane proteins such as E-cadherin. As the embryo continues to develop, however, a combination of external and internal signals trigger an epithelial-to-mesenchymal transition in NCCs that causes them to downregulate their strong adhesive properties, develop apoptotic resistance, and acquire an invasive phenotype (Piacentino et al., 2020; Barriga et al., 2018; Scarpa and Mayor, 2016; Kerosuo and Bronner-Fraser, 2012; Theveneau and Mayor, 2012). The cells migrate from the dorsal neural tube to a variety of tissues: some cranial NCCs, for instance, colonise a target corridor located along the second branchial arch, also known as the second pharyngeal arch in humans, and subsequently differentiate into cells that ultimately form the mandible, neck cartilage, and bones of the middle ear (Frisdal and Trainor, 2014).

Multiple biophysical cues guide NCCs along their migratory corridors; interestingly, some of these cues also act on TCs and ECs during angiogenesis. Chemotaxis is known to play a role in NCC migration, as migrating streams of NCCs can be steered by known chemoattractants such as VEGF, which is produced by cells belonging to the mesoderm and ectoderm germ layers (McLennan et al., 2010). NCCs may also locally secrete chemical signals to promote their migration: in *Xenopus* embryos, for instance, cells chemotactically attract neighbours via secretion of complement 3a (C3a) and stromal-derived factor 1 (Sdf1) into the microenvironment; in the chick cranial neural crest, NCC secretion of angiopoietin-2 helps to maintain stream migration (McKinney et al., 2016; Carmona-Fontaine et al., 2011; Theveneau et al., 2010). NCCs may also sense molecules that inhibit, rather than promote, cell migration; such inhibitors include DAN and Sema3A, which are produced by mesoderm cells (Bajanca et al., 2019; McLennan et al., 2017).

Contact-mediated mechanisms between NCCs influence their migration as well. Contact inhibition of locomotion (CIL) represents the most studied of these interactions: during CIL, cells halt when their protrusions touch, then move away from each other along the direction of contact. This process is regulated at the molecular level via signalling pathways that include N-cadherin, Wnt, epherins, and platelet-derived growth factor (Piacentino et al., 2020; Roycroft et al., 2018; Bahm et al., 2017; Stramer and Mayor, 2016; Carmona-Fontaine et al., 2008). The pathways polarise NCCs through expression of GTPases Rac1 and RhoA, which subsequently create the forces necessary for cell propulsion (Shellard et al., 2018). Although Abercrombie and Heaysman (1954) first characterised CIL in chick heart fibroblasts, the investigation of its role in NCC migration began with studies by Erickson (1985). Subsequent *in vitro* and *in vivo* experiments support the conclusion that certain populations of NCCs, including those of *Xenopus* embryos and the trunk of chick embryos, exhibit CIL (Shellard and Mayor, 2019; Li et al., 2019; Roycroft et al., 2018; Scarpa et al., 2015; Theveneau and Mayor, 2012; Carmona-Fontaine et al., 2008). It remains unclear, however, whether NCCs universally interact via this process: studies of chick cranial NCCs, for instance, have reached differing conclusions as to whether they interact via CIL (Genuth et al., 2018; Richardson et al., 2016; Kulesa et al., 2004).

More recent studies that investigate mechanisms for NCC migration have focussed on interactions between NCCs and their ECM (Yamada and Sixt, 2019; Vega and Schwarzbauer, 2016; Mittal et al., 2010). Experimental evidence suggests an intricate relationship between the ECM and NCCs, in which changes in one affect the other (D'Alessandro et al., 2021). NCCs, for instance, secrete a variety of molecules (such as matrix metalloproteases) which are known to degrade and remodel the ECM (Leonard and Taneyhill, 2020). At the same time, changes to the ECM, such as its stiffening prior to NCC migration, have been associated with the onset of epithelial-to-mesenchymal transition in NCCs and their later differentiation (Barriga et al., 2018; Chan et al., 2017). Fibronectin, in particular, constitutes a critical component for NCC migration: knocking out fibronectin expression in mice significantly reduces the number of cardiac NCCs that successfully migrate along target corridors (Wang and Astrof, 2016). Furthermore, NCCs preferentially orient fibronectin fibres in the direction of their migration and move along dense regions of fibronectin *in vitro* (Newgreen, 1989; Rovasio et al., 1983). These observations suggest that two ECM-related mechanisms, known to occur in individual cell migration, also act during the collective migration of NCCs: (1) haptotaxis, which guides cells towards increasing adhesive gradients; and (2) contact guidance, which aligns cells along ECM fibres (Piacentino et al., 2020; Walma and Yamada, 2020; Autenrieth et al., 2016; Haeger et al., 2015; Provenzano et al., 2008; Théry et al., 2006; Teixeira et al., 2003).

1.2.2 Mathematical modelling of neural crest cell migration

Neural crest cell migration, like angiogenesis, presents a particularly interesting system to study with mathematical models due to the plethora of signals thought to affect migration. Models used to study NCCs include discrete, hybrid, and continuous approaches, and they generally aim to determine which mechanisms dominate migration (Giniunaite et al., 2019a; Schumacher et al., 2016). For example, Simpson et al. (2007) applied a coupled system of Fisher–Kolmogorov–Petrovsky–Piskunov PDEs to study NCC colonisation of the chick gut. Through model analysis and subsequent biological experiments, they confirmed that NCCs at the leading edge of streams largely drove successful invasion via their migration and proliferation, whereas NCCs further behind the front did not move towards unexplored areas and were more quiescent. Landman et al. (2011) developed cellular automata to confirm that cell proliferation could dominate collective NCC migration. Cheeseman et al. (2014) extended these cellular automata to investigate the lineages of cells within successfully invading streams and found that most were descendants of a relatively small number of NCCs; these results led them to hypothesise that resource competition was a major factor in promoting migration and proliferation.

While these proliferation-driven models can explain the migration of enteric NCCs, cell division may not be the dominant driver of migration in all NCC populations. Division of chick cranial NCCs typically occurs near the end of migration, for example, and a relatively small fraction of cells proliferate during migration (Ridenour et al., 2014). Hence other mechanisms, in addition to proliferation, are likely to influence the migration of other populations of NCCs *in vivo*. Below, we describe two modelling frameworks that have been used to investigate such alternative mechanisms.

Cell-induced gradient models

Cell-induced gradient models investigate the role of chemotaxis in NCC migration. In this framework, leading NCCs consume chemoattractants such as VEGF, thereby creating a chemotactic gradient to guide their movement. Consequently, cells can migrate even when the chemoattractant distribution is initially homogeneous (i.e., when no gradient exists prior to migration). Such conditions can occur *in vivo*: for example, the VEGF distribution prior to the migration of chick cranial NCCs does not favour a particular direction (McLennan et al., 2010).

McLennan et al. (2012) constructed a hybrid ABM to study whether a cell-induced gradient of VEGF could lead to the formation of chick cranial NCC streams. They discovered that, within the context of their model, a homogeneous population of cells could not produce long-distance, cohesive migrating streams such as those observed *in vivo*. Instead, stream breaks – cell-free regions within the migratory pathway – tended to develop. This occurred because cells at the front of streams tended to deplete nearly the entire concentration of chemoattractant in their neighbourhood before migrating away. As a result, trailing NCCs sensed a weaker chemotactic gradient and moved more randomly than cells at the front. Since cells tend to travel shorter distances when their migration is not biased in a particular direction, this would explain why cells at the back of streams were left behind (McKinney et al., 2020; Giniunaite et al., 2019b; McLennan et al., 2017). The authors used their ABM to explore other ways by which cohesive streams could form during chick cranial neural crest cell migration. They determined that a heterogeneous population, consisting of "leader" cells that consume and migrate up VEGF gradients, and “follower” cells, which respond by touch to (unspecified) short range directional cues produced by the leader cells, reduced the number of stream breaks that occur. Furthermore, the resulting *in silico* streams successfully colonised the migratory corridor. Subsequent morphologic and transcriptomic analysis of *in vivo* chick cranial NCCs showed that while there are, indeed, distinct leader and follower phenotypes at the front and back of the stream, respectively, cells within the middle of streams appeared to express phenotypes along a continuous spectrum between the two (McLennan et al., 2020; Schumacher, 2019; Morrison et al., 2017; McLennan et al., 2015a). Extensions to this hybrid framework have examined the synergistic roles of other biological mechanisms on stream cohesiveness and collective migration. These include rules for phenotype switching (Schumacher, 2019), VEGF-mediated induction of leader/follower phenotypes (McLennan et al., 2015b), non-uniform domain growth (McKinney et al., 2020), and inhibition of NCC movement by molecules such as DAN (McLennan et al., 2017).

Although cell-induced gradient models can produce migrating NCC streams that resemble those observed *in vivo*, these results come with certain caveats. Increases in the VEGF consumption rate of leader NCCs can generate stream breaks, suggesting that cell-induced gradients do not provide a robust mechanism for stream formation. Although phenotype switching and movement inhibitors such as DAN can reduce the probability that stream breaks form, the question of robustness remains (Giniunaite et al., 2019b; Schumacher, 2019; McLennan et al., 2017). The cell-induced gradient framework also tends to require a confined domain or an initial pre-patterned “stripe” of chemoattractant in order to generate streams. While these simplifications ensure that NCCs cannot move lateral to the target corridor, they have little experimental justification in chick cranial NCC migration: the cells can, in principle, move laterally to other parts of the embryo and there is no definitive *in vivo* evidence of chemoattractant pre-patterning at this time (Szabó et al., 2019; McLennan et al., 2017, 2010). In *Xenopus* embryos, by contrast, the ECM protein versican may prevent lateral NCC migration by acting as a physical barrier for cells (Szabó et al., 2016).

Contact inhibition of locomotion models

Contact-based mechanisms for NCC migration, and in particular CIL, have also been investigated through mathematical modelling. Carmona-Fontaine et al. (2011) constructed a CIL-based model for *Xenopus* NCCs in order to determine if this mechanism could cause collective migration. The authors found that cells moving only via CIL could colonise their target corridor, but also exhibited more lateral movement than observed *in vivo* and *in vitro*. However, if NCCs are also attracted to each other by locally secreting C3a, then cohesive groups akin to those observed experimentally could form. Woods et al. (2014) developed a more detailed version of this ABM, and found that increasing the attractive force between cells generated more persistent migratory patterns. Szabó et al. (2016) further extended the framework to consider the effects of confinement on cell migration. In combination with experimental studies, they determined that increased confinement of cells, mediated through the ECM protein versican, could enhance collective cell migration along the target corridor. Other ABMs that rely on CIL suggest that the rate at which cells enter the domain is also an important parameter for determining successful migration (Khataee et al., 2021).

Painter et al. (2015) considered integro-differential equations to study CIL in (potentially heterogeneous) groups of NCCs. Unlike traditional PDEs, the framework developed by these authors can model cell interactions at distances far from their centres of mass through integral terms. Numerical results showed that extending the radius of non-local interactions, or increasing the strength of cell-cell repulsion, concurrently increased the distance that cells travel. However, the authors were unable to derive an analytic relationship between the migration speed and model parameters.

Merchant et al. (2018) investigated how the proteins Rac1 and RhoA affect NCC migration by developing an ABM in which each cell membrane was represented as a set of vertices that move according to an overdamped Newton's law. Simulations of their ABM in a confined, rectangular domain showed that collective cell migration could arise through CIL and co-attraction alone – i.e., cells did not require chemotaxis to persistently migrate along their target corridor. A similar conclusion has been found using other ABM frameworks

(Camley et al., 2016).

More recent CIL-based ABMs seek to determine how cell interactions with other tissues, such as the mesoderm, create emergent collective behaviours. For instance, a cellular Potts model created by Szabó et al. (2019) examines whether the "chase-and run" mechanism described in Theveneau et al. (2013) can lead to stream formation. In this model, placode cells, which surround the neural tube, secrete Sdf1 in order to attract NCCs. However, CIL also acts to repel NCCs and the placode cells whenever they come into contact. The model shows that the resulting balance of attractive and repulsive cues, along with other environmental cues such as C3a, can generate discrete ribbon-like patterns that resemble *in vivo* NCC streams. An off-lattice ABM from Colombi et al. (2020) further examines the chase-and-run framework in an open geometry. Although this model does not generate streaming patterns of NCCs, it does reproduce several *in vitro* observations of NCC and placode cell behaviours. Furthermore, it generates hypotheses that experimentalists can test in future studies.

Despite the insights provided by CIL-based mathematical models for NCC migration, this mechanism may not account for all aspects of the biological phenomenon. As we noted above, it remains unclear whether chick cranial NCCs even exhibit CIL. Instead, cells may repel each other, but not necessarily along the direction of contact (Schumacher et al., 2016). For the rest of this thesis, we will refer to such general methods of cell-cell repulsion as "volume exclusion" within the context of neural crest cell migration. On-lattice ABMs that incorporate this type of volume exclusion suggest that the combination of these signals with other cues, such as chemotaxis, can also account for migratory patterns observed during NCC migration (Wynn et al., 2013). Other studies using integro-differential equations have further suggested that continuous phenotype switching and chemotaxis alone can support invasive travelling wave solutions (Lorenzi and Painter, 2021).

ECM guidance cues in neural crest migration

Unlike cell-induced gradient mechanisms and CIL, ECM-related guidance cues tend to be under-represented in mathematical models of NCC migration. Such mechanisms present a particularly interesting avenue for exploration, as preliminary mathematical models demonstrate that they can generate complex patterns of migration. Painter (2009), for example, constructed a kinetic model for mesenchymal cell migration based on the Boltzmann equation. The results of this mesoscopic model suggest that, through contact guidance, cells can respond to ECM fibre orientations and thereby migrate collectively in anisotropic patterns.

Outside the context of the neural crest, models that incorporate ECM-related guidance cues suggest an important role for this structure in determining collective cell migration. Dallon et al. (1999), for instance, developed an off-lattice hybrid ABM for fibroblast migration in which cells could respond to and remodel collagen fibres of the ECM. They found that contact guidance promotes cell alignment, and that the initial fibre orientation is an important determinant of collective behaviour. Subsequent extensions of the model reveal that chemotaxis and increased cell influx into the domain can also enhance group alignment (McDougall et al., 2006; Dallon et al., 2000). A more recent ABM from Wershof et al. (2019), which extends an off-lattice framework model for cell alignment (Vicsek et al., 1995) by including ECM interactions, suggests that contacts between cells promotes ECM anisotropy and, consequently, further aligns cells to each other. Other ABMs for fibroblast migration suggest that ECM remodelling and contact guidance enhance the cohesiveness of migrating cell clusters (Groh and Wagner, 2011; Groh and Louis, 2010). In cancer biology, both on- and off-lattice frameworks for metastasis show that different strengths of ECM interactions can create a variety of tissue-level behaviours (Pramanik et al., 2021; Suveges et al., 2021; Azimzade et al., 2019).

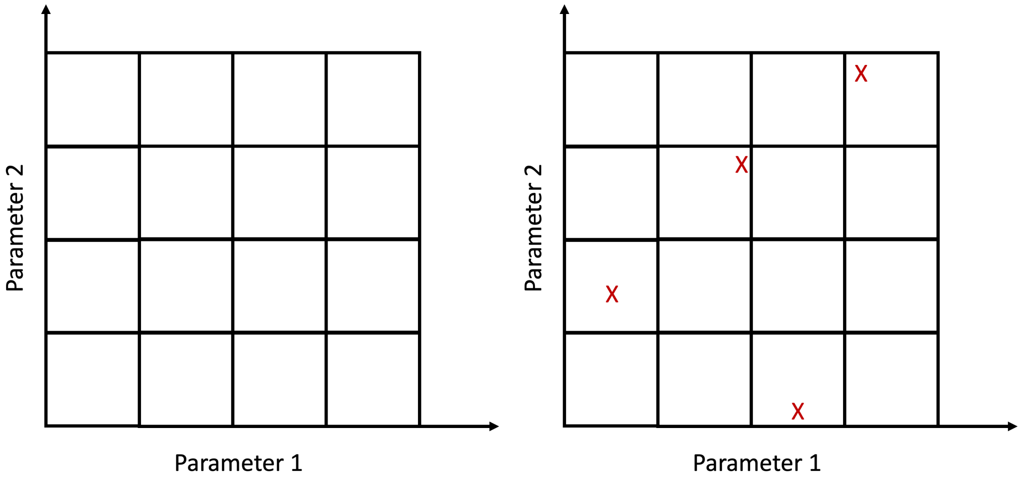
In Chapters 4 and 5 of this thesis, we will investigate the role of NCC-ECM interactions in robust collective cell migration by developing a new ABM for this process.

1.3 Global sensitivity analysis techniques

In spite of many mathematical models investigating the roles of haptotaxis, contact guidance, chemotaxis, and cell-cell repulsion in collective cell migration, it remains largely unclear how (or whether) these mechanisms act synergistically to promote collective behaviour. This presents a challenging task, not least because different mechanisms can lead to similar macroscopic outcomes. Global sensitivity analysis offers a potential solution to this problem, as it systematically quantifies how model outputs are affected by parameter perturbations. Since model parameters are connected to the biological mechanisms underpinning the modelling framework, this analysis can be used to infer how different signalling cues affect particular macroscopic behaviours, such as the distance that cells travel. Global sensitivity analysis approaches vary multiple parameter values at a time, and hence explore larger regions of input space than local sensitivity analysis methods, such as one-at-a-time approaches, which only vary a single parameter. Additionally, they better characterise complicated, nonlinear relationships than local approaches (Hamis et al., 2021; Saltelli et al., 2019; Saltelli and Annoni, 2010). In this thesis, we apply Latin hypercube sampling (LHS) and extended Fourier Amplitude Sensitivity Testing (eFAST) to analyse the roles of mechanisms for NCC migration (Qian and Mahdi, 2020; Renardy et al., 2019; Marino et al., 2008; Helton and Davis, 2003; Saltelli et al., 1999). We provide more detail on LHS and eFAST below.

1.3.1 Latin hypercube sampling and partial rank correlation coefficient analysis

LHS algorithms generate parameter regimes that minimise correlations between samples in parameter space, in order to better characterise their relationships with summary statistics. In this way, they differ from traditional Monte Carlo algorithms, which potentially generate highly clustered samples in parameter space. The LHS algorithm subdivides each axis of the parameter space into N intervals, where N is equal to the number of samples. Parameter regimes are then sampled uniformly at random within each resulting hypercube, but with the rule that no two samples, after having been projected onto a given axis, may exist in the same subinterval. The LHS algorithm is most easily visualised when the parameter space is created by two parameters, as illustrated in Figure 1.3: in this case, the algorithm partitions the parameter space into N^2 squares. Samples are then chosen such that there is one unique parameter set for every row and column of squares (Qian and Mahdi, 2020; Read et al., 2012; Marino et al., 2008; Helton and Davis, 2003; McKay et al., 2000).



(a) Division of parameter space (b) Sampling of parameter space

Figure 1.3: A diagram showing LHS for two parameters and $N = 4$ random samples. In LHS, the parameter space (a) is partitioned into equally sized squares, so that it consists of N rows and N columns. The parameter regimes (b) are then sampled uniformly within each square, but in such a way that only one parameter regime (red "X"s) belongs to each row and column of the partitioned space.

One advantage of LHS compared to traditional Monte Carlo sampling is that the former can generate statistics with lower variances, but equal mean values, compared to metrics obtained using the latter. Although this means that statistics generated with LHS converge more quickly to their expected values than those obtained using traditional sampling techniques, this effect becomes less pronounced when the parameter space under investigation is very large (Helton and Davis, 2003). This makes it difficult to determine a strict rule for the appropriate number of samples to use for LHS, since the sample size depends on the parameter space under investigation. There are *post hoc* techniques for determining if the number of samples used to generate a dataset is sufficient, however. Top-down correlation coefficients, for instance, determine whether rankings of parameter sensitivity for statistics produced by LHS are likely to significantly change when more samples are added; the sample size is chosen as the smallest value that ensures the rankings are likely to remain unchanged when the sample size increases (Marino et al., 2008; Helton and Davis, 2003). We describe top-down correlation coefficients in more detail in Appendix C.

Although LHS generates parameter regimes that can be simulated with a mathematical model, it alone cannot analyse parameter sensitivities. For that purpose, Spearman partial rank correlation coefficients (PRCCs) are computed on the summary statistics generated by LHS (Qian and Mahdi, 2020; Renardy et al., 2019; Marino et al., 2008; Helton and Davis, 2003). These PRCCs measure the extent to which the data are monotonically increasing or decreasing with respect to a given parameter value. Their magnitudes can be used to order parameter values by their impact on the summary statistic, with larger absolute values of PRCCs from the model indicating stronger effects. One caveat to the LHS/PRCC approach is that only monotonic behaviours can be characterised, i.e., PRCCs cannot capture more complicated nonlinearities such as sinusoidal behaviour (Qian and Mahdi, 2020; Marino et al., 2008; Helton and Davis, 2003). For this reason, LHS/PRCC analysis is generally used as a first step for analysing the sensitivity of summary statistics to model parameters. For more detailed relationships, eFAST is a preferable tool to use.

1.3.2 eFAST: extended Fourier amplitude sensitivity testing

In contrast to LHS, eFAST allows one to uncover nonlinear, non-monotonic relationships in data using Sobol indices. Sobol indices indicate the extent to which changes in the value of a given parameter affect the variance of an output statistic (Qian and Mahdi, 2020; Marino et al., 2008; Sobol, 2001; Saltelli et al., 1999). The eFAST method computes both first-order, S_1, and total-order, S_{tot}, indices. A first-order index measures the fraction of variance generated by changing only the parameter of interest, while the total-order index measures the fraction generated by its interactions with all other perturbed parameters (since these interactions include those used to compute the first-order index, $S_{\text{tot}} \geq S_1$). Since S_{tot} may count the same interactions more than once, the following inequalities hold:

$$\sum_{k=1}^{d} S_1^{(k)} \leq 1, \qquad \sum_{k=1}^{d} S_{\text{tot}}^{(k)} \geq 1, \tag{1.2}$$

where d is the total number of parameters, $S_1^{(k)}$ is the first-order Sobol index for parameter k, and $S_{\text{tot}}^{(k)}$ is the total-order Sobol index for parameter k. Equality holds for both expressions in Eq. (1.2) when the mathematical model is additive with respect to the statistic of interest, i.e. when the output data are not affected by high-order interactions between parameters (this also means that $S_1 \approx S_{\text{tot}}$ when the model is additive). The magnitude of the Sobol indices determines the sensitivity of the output statistics to particular parameters of interest: higher values indicate that the statistic is more sensitive to changes in that parameter. Large discrepancies between first-order and total-order indices may also be informative, as they can indicate when a statistic is sensitive to interactions between different model parameters.

The eFAST algorithm calculates Sobol indices in the following manner: first, the algorithm samples parameter values from periodic distributions that ensure the values lie within user-defined ranges; the frequencies of such periodic distributions (and their higher-order harmonics) are unique for each parameter. After computing model outputs and summary statistics for each parameter regime, the algorithm next performs a Fourier transform on the variance of each statistic; this allows investigators to quantify how much of the overall variance is attributable to the frequencies (and higher-order harmonics) of a given parameter value through Fourier coefficients. The algorithm finally calculates Sobol indices by measuring the fraction of the variance that is generated by the Fourier coefficients uniquely associated to each parameter.

One can also compute PRCCs from outputs generated by the parameter regimes sampled during eFAST; this allows for a more straightforward comparison of conclusions obtained from eFAST and LHS/PRCC analysis (Marino et al., 2008).

One drawback to using eFAST is that many parameter regimes must be sampled to calculate the Sobol indices: if N is the number of values to test per parameter and d is the number of ABM parameters, then a minimum of $(N \cdot d)$ parameter regimes are required. Often, though, researchers are interested in determining whether the Sobol indices from eFAST are statistically significant from the variation caused by a null parameter: in such cases, the investigator tests an additional dummy parameter and then conducts N_{r} runs of the eFAST sampling algorithm. As a result, the total number of parameter regimes to simulate in eFAST is equal to $N \cdot (d+1) \cdot N_{\text{r}}$ (Marino et al., 2008; Saltelli et al., 1999). Furthermore, if the model in question is a stochastic ABM, multiple realisations should be run for each parameter regime in order to ensure the summary statistic distribution is properly characterised. This can make eFAST computationally infeasible for certain models, in particular discrete ABMs. For example, the eFAST data shown in Chapter 4 required approximately one month to compute on a workstation consisting of four cores (Intel(R) Core(TM) i5-7500T CPU @ 2.70 GHz).

1.4 Thesis outline

In this chapter, we reviewed the biology and mathematical modelling of two phenomena that exhibit leader-follower collective cell migration: tumour-induced angiogenesis and chick cranial NCC migration. In the remaining chapters of this thesis, we aim to address the following issues regarding collective cell migration and the models that describe it.

In Chapter 2, we compare the classical snail-trail continuum model for angiogenesis with a minimal discrete model for the same process, in order to determine the conditions for which one-dimensional (1D) and 2D versions of the phenomenological model are likely to accurately capture the underlying cell dynamics. By numerically solving both models, we find that the snail-trail model consistently underestimates the follower cell density because it does not account for cell movement in directions other than that of the migrating front. Using a mean-field approach, we derive a new corrective term for the classical framework that is valid for scenarios in which leader cell movement is dominated by chemotaxis. This analysis thus reveals an implicit physical assumption that determines when the snail-trail framework accurately captures discrete results, and identifies metrics (such as the average cell displacement) that can indicate when the mean-field model will likely break down.

In Chapter 3, we apply techniques from asymptotic analysis to investigate the relationship between the corrected snail-trail model and a more nonlinear, coarse-grained system for angiogenesis developed in Pillay et al. (2017). We demonstrate that solutions to both continuum frameworks are identical at leading order, provided cells do not branch frequently and chemotaxis dominates leader cell migration. This analysis explains why numerical solutions to both models are indistinguishable for certain parameter regimes and also supplies a reduced model for angiogenesis, which describes the same dynamics but is easier to analyse and simulate. We construct self-similar asymptotic solutions to the leading order system, which are valid for early times. This work demonstrates how a comparative analysis of theoretical frameworks can identify simpler models of collective cell migration, and how there can be conditions for which phenomenological models can be equivalent to more complicated frameworks obtained from systematic discrete-to-continuum derivations.

In Chapter 4, we develop a new ABM for NCC migration that investigates whether a cell-reinforced migratory cue provides a sufficient signal for collective migration. In this description, leader NCCs remodel an initially punctate and immature ECM into a fibrous structure for follower cells, and follower cells further adjust the orientation of ECM fibres. At the same time, cells use the distribution of nearby punctate and fibrous ECM molecules to determine their direction of migration. With this model, we confirm that such ECM interactions, along with volume exclusion (cell-cell repulsion), haptotaxis, and contact guidance, produce migrating streams of NCCs that resemble those found *in vivo*. We then perform LHS/PRCC and eFAST analysis to provide one of the first comprehensive investigations of how key individual mechanisms affect collective cell migration.

In Chapter 5, we perform *in silico* tests of the NCC ABM to determine what mechanisms most contribute to "normal" migration. While many tests support the conclusions obtained from the aforementioned global sensitivity analysis, they also unexpectedly indicate that cell migration is most efficient when leader cells specialise in guiding the stream and remodelling the ECM. We test this conjecture by introducing a new mechanism to the ABM that steers leader cells along the target corridor. While this new signal reduces lateral migration, there are extreme cases for which stream breaks occur. We systematically identify parameter combinations that prevent leader-follower separation; these tests reveal that contact guidance is the key mechanism that controls this behaviour and, consequently, likely governs robustness in this system. This work constitutes, to our knowledge, the first detailed study of combinations of mechanisms that can avoid the onset of stream breaks during NCC migration. Finally, in Chapter 6 we discuss how the results add to the current understanding of leader-follower migration, summarise possible connections between angiogenesis and NCC migration, and highlight avenues for future work.

Chapter 2

Evaluating snail-trail frameworks for leader-follower behaviour with agent-based modelling

Most of the work in this chapter has previously appeared in the following journal article: Martinson WD, Byrne HM, Maini PK (2020) Evaluating snail-trail frameworks for leader-follower behavior with agent-based modeling. Phys Rev E 102(6): 062417.

2.1 Overview

Continuum snail-trail models represent a simple and analytically tractable approach for describing tip cell (TC) and stalk cell (EC) dynamics during angiogenesis (and, more generally, leader-follower dynamics in other processes that involve branching morphogenesis). Many investigations that utilise these frameworks, however, tend to phenomenologically formulate snail-trail systems in one spatial dimension (1D). While this assumption permits the derivation of equations whose solutions are in good qualitative agreement with experiments, its validity has yet to be assessed in a more theoretical setting. Consequently, the conditions for which 1D snail-trail systems accurately describe the inherently multidimensional biological process of angiogenesis are largely unclear. Such research, which is the focus of this chapter, will help identify biological settings for which snail-trail models are appropriate for describing results from *in vitro* and/or *in vivo* angiogenesis assays.

In this chapter, we address such open questions by comparing solutions of a modified 2D snail-trail model to ensemble average distributions from a 2D agent-based model (ABM) for angiogenesis originally developed in Pillay et al. (2017). The latter model, which we take to represent the biological “ground truth”, incorporates the same mechanisms as the snail-trail framework (including tip cell random motion, chemotaxis, branching, and anastomosis). We modify the 2D snail-trail model by introducing a new scaling factor in the stalk cell evolution equation. This factor, which we derive using a coarse-graining approach, accounts for stalk cell proliferation in directions other than that of the moving tip cell solution front; its functional form has not appeared in previous versions of the continuum framework. We demonstrate that the modified phenomenological snail-trail model is a good approximation to the ABM results, provided chemotaxis dominates tip cell migration. We then use the corrected 2D snail-trail framework to derive the 1D version of the model, and use numerical simulations to determine when the 1D model is able to accurately approximate the column averaged results of the 2D system. We find that 1D models are poor approximations to column averaged results if there is sufficiently large spatial variation in the TAF distribution and/or cell densities. This work thereby identifies assumptions that are implicitly invoked when spatially reduced models are applied to study angiogenesis.

2.1.1 Chapter outline

In §2.2, we present a 2D version of the classical snail-trail model for tip cell and stalk cell dynamics during angiogenesis, which has been modified by a novel multiplicative factor in the stalk cell rate equation. We introduce this factor in order to correct for an underestimation of stalk cell proliferation in directions other than that of the moving solution front. In §2.3 we derive a formula for the corrective factor, which suggests that snail-trail models are valid approximations of tip and stalk cell dynamics when cell movement is dominated by chemotaxis. We confirm the accuracy of this formula in §2.4-§2.5 by comparing numerical solutions of the 2D snail-trail model with ensemble averages of the ABM from Pillay et al. (2017), henceforth denoted as the P–ABM. Finally, in §2.6 we discuss when the 2D snail-trail model can be reduced to a 1D system and make explicit conditions for which it is appropriate to use 1D snail-trail models to describe experimental data.

2.2 Snail-trail model development in 2D

The 2D snail-trail model tracks the densities of TCs and ECs within the unit square $(x, y) \in [0, 1] \times [0, 1]$ and time interval $t \in [t_0, \infty)$, where $t_0 \geq 0$ (all quantities have been non-dimensionalised). We use continuous functions $n(x, y, t)$ and $e(x, y, t)$ to describe the mean-field TC and EC densities, respectively, and write $c(x, y, t)$ to denote the concentration of a generic tumour angiogenic factor (TAF). In general, the TAF dynamics are governed by a differential equation that may include diffusion and/or uptake by TCs and ECs (Tweedy et al., 2020; Connor et al., 2015; Pettet et al., 1996b). However, in this chapter our aim is to determine appropriate equations for describing cell movement and proliferation. For this reason, we simplify the model by assuming that the TAF field is at steady state, such that it is independent of t (i.e., $c = c(x, y)$).

The basis for modelling the tip cell density is the continuity equation

$$\underbrace{\frac{\partial n}{\partial t}}_{\text{tip cell rate of change}} + \underbrace{\nabla \cdot \mathbf{J}}_{\text{net movement of tip cells}} = \underbrace{F(n, e, c),}_{\text{tip cell production/elimination}} \tag{2.1}$$

where $\mathbf{J}$ is the net tip cell flux vector and the function F accounts for tip cell production and elimination. As is common within the modelling community (Chaplain et al., 2020; Perfahl et al., 2017; Bonilla et al., 2017; Connor et al., 2015; Bentley et al., 2009; Orme and Chaplain, 1997; Pettet et al., 1996a; Byrne and Chaplain, 1995), we view TC movement as a biased random walk towards increasing TAF gradients. Thus the net TC flux is given by

$$\mathbf{J} = \underbrace{-D\nabla n}_{\text{random motion}} + \underbrace{\chi(c) n \nabla c,}_{\text{chemotaxis}} \tag{2.2}$$

where ∇ is the gradient operator, D is a constant and positive random motion coefficient, and $\chi(c)$ is a positive function that measures the chemotactic sensitivity of cells to the TAF. We assume for simplicity that $\chi(c) = \chi$ is a positive constant, although one could also use other functions that, for instance, account for cell desensitisation to TAF (Hillen and Painter, 2009).

We represent the creation of new vessel branches by increases in the TC density. Observations of network morphogenesis in several angiogenesis assays indicate that such "sprout branching" depends on the TAF concentration (Uccelli et al., 2019; Gordon-Weeks et al., 2017; Gerhardt et al., 2003; Ruhrberg et al., 2002), thus for simplicity we assume that the TC proliferation rate is proportional to the TAF concentration. The particular form of the branching term originates from Byrne and Chaplain (1995), who used it to explain *in vivo* observations such as the brush-border effect, which is described in more detail in Chapter 1 (Muthukkaruppan et al., 1982).

We suppose tip-to-tip and tip-to-sprout anastomosis dominate TC elimination and assume that these processes occur at rates proportional to the product of their respective cell densities. In all, then, the net rate of change of TCs is given by

$$F(n,e,c) = \underbrace{\lambda nc}_{\text{branching}} - \underbrace{\beta_e ne}_{\text{tip-to-sprout anastomosis}} - \underbrace{\beta_n n^2,}_{\text{tip-to-tip anastomosis}} \tag{2.3}$$

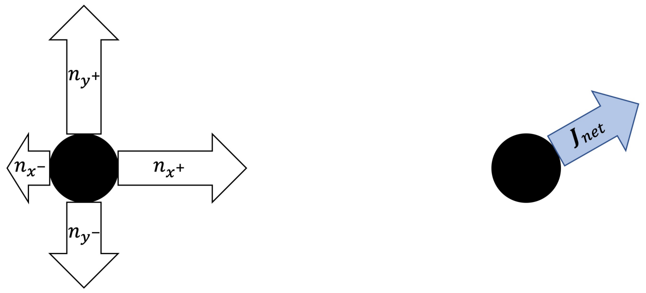
where λ, β_e, and β_n are positive constant parameters related to the branching rate, the tip-to-sprout anastomosis rate, and the tip-to-tip anastomosis rate, respectively.

In accordance with the snail-trail framework, we consider the rate of blood vessel production to be proportional to the magnitude of net tip cell flux. However, since we are interested in stalk cell, rather than vessel, density, we divide this rate by h, a constant that represents the length of a typical stalk cell. For reasons that will be made clear shortly, we also introduce a corrective factor $\kappa(x,y)$ in this rate equation, which is new to the 2D snail-trail framework. In §2.3, we derive an approximate value for this expression, and find that it is inversely proportional to $\chi\|\nabla c\|_2$, where $\|\cdot\|_2$ denotes the Euclidean norm. The full non-dimensional system is therefore given by

$$\frac{\partial n}{\partial t} = D\nabla^2 n - \chi\nabla\cdot(n\nabla c) + \lambda nc - \beta_e ne - \beta_n n^2, \tag{2.4}$$

$$\frac{\partial e}{\partial t} = \frac{\kappa(x,y)}{h}\|\chi n\nabla c - D\nabla n\|_2 = \frac{\mu}{\chi\|\nabla c\|_2}\|\chi n\nabla c - D\nabla n\|_2, \tag{2.5}$$

where $\mu = 4Dh^{-2}$. We remark that the above 2D snail-trail model is consistent with those already in the literature: if we set the fraction $\kappa(x,y)/h$ to be equal to one, for instance, then we recover a version of the classical 2D snail-trail model which appears in Jackson and Zheng (2010). We will discuss the physical interpretation of $\kappa(x,y)$ in later sections of this



(a) Tip cell flux in cardinal directions

(b) Resultant net flux vector $\mathbf{J}_{\text{net}}$

Figure 2.1: The net tip cell flux vector $\mathbf{J}_{\text{net}}$ underestimates the total stalk cell proliferation rate. From (a) a given point (black circle), one may measure the total flux of tip cells moving in any direction; each quantity should be considered when calculating the change in stalk cell density. The vector $\mathbf{J}_{\text{net}}$ (b), however, only considers the flow of cells in the direction of the tip cell wave front: as a result, its magnitude is always less than or equal to the total tip cell mass that moves within a given time period.

chapter. We close Eq.s (2.4)-(2.5) by imposing no-flux boundary conditions

$$(\chi n \nabla c - D \nabla n) \cdot \mathbf{a} = 0, \tag{2.6}$$

where $\mathbf{a}$ is the outward unit normal vector of the domain, and initial conditions

$$n(x, y, t_0) = G(x, y), \quad e(x, y, t_0) = H(x, y), \tag{2.7}$$

where $G(x, y)$ and $H(x, y)$ are non-negative functions. Boundary conditions are not required for the EC rate equation because it is a first order ordinary differential equation (ODE) with respect to time t. We also remark that, since the TAF concentration is independent of time, we prescribe the functional form of $c(x, y)$ in simulations.

The multiplicative factor $\kappa(x, y)$ is included in Eq. (2.5) because the net tip cell flux term, which is used to calculate the stalk cell density rate of change, neglects vessel growth in directions other than that of the invading front. To see this, we imagine decomposing the TC density into four populations: n_{x^+}, n_{x^-}, n_{y^+}, and n_{y^-}, which correspond to the density of TCs moving in the positive x-direction, the negative x-direction, and so on. Figure 2.1(a) presents a sketch of these quantities for a given time period. New vessels proliferate along the paths that these TCs traverse, hence the total change in vessel density is proportional to the sum of these quantities. The TC flux, however, does not consider this value: instead, it records the vessel density produced in a net direction defined by $(n_{x^+} - n_{x^-})$ and $(n_{y^+} - n_{y^-})$, as shown in Figure 2.1(b). Consequently, the magnitude of the net TC flux almost always underestimates the density of stalk cells produced, unless a corrective factor is included.

Although this scaling factor is new to the 2D snail-trail modelling framework, previous researchers have identified, and attempted to correct for, EC underestimation in 1D versions of snail-trail models (Pillay et al., 2017; Connor et al., 2015). In those studies, however, the authors consider the corrective factor to be constant, and use *ad hoc* methods to determine its value. By contrast, we present a general argument for determining an approximate functional form for κ in the next section, and find that it varies with respect to space. We demonstrate in §2.5 that constant-valued corrective factors do not always result in the 2D snail-trail model accurately approximating ensemble average ABM results, which suggests that the approach taken in this chapter improves upon those of previous researchers.

We note that our motivation for including a corrective factor originates from usage of the net TC flux, but does not account for further complications that could arise from branching and anastomosis. As a result, our derivation in §2.3 neglects these other processes (i.e., we focus on TC movement and EC proliferation).

2.3 Derivation of a functional form for the snail-trail multiplicative factor

We have argued that the corrective factor, κ, accounts for vessel production in directions other than that of the migrating front. Thus, if $\mathbf{J}_{\text{net}}$ measures the average net tip cell flux within a time period Δt and spatial interval h, then $\kappa \|\mathbf{J}_{\text{net}}\|_2$ is equal to the (total) vessel density produced (note that we have absorbed any dependence of h and Δt into κ). In this section, we derive a closed-form expression for κ by translating this statement to a discrete setting and coarse-graining the resulting expressions to find a relationship in terms of the continuum parameters.

We consider a general on-lattice framework to represent a discrete version of the snail-trail model without branching or anastomosis (note that this means multiple cells may occupy a single lattice site). The ABM is assumed to have N lattice sites that are equally spaced with step size h, where h represents the length of a typical cell. We assume for simplicity that a tip cell length is equivalent to that of an stalk cell – at the end of this section, we discuss how the formula for $\kappa(x, y)$ changes when we relax this assumption. As in the previous section, we assume that the TAF concentration $c(x, y)$ is prescribed and independent of time t.

In this ABM, we describe TC movement as a biased random walk towards increasing concentrations of TAF. Within a time step Δt, TCs are selected to move with constant probability P_m. Once chosen to move, the TC selects a direction according to the probability of moving left, P_{x^-}, right, P_{x^+}, up, P_{y^+}, and down, P_{y^-}. These probabilities are defined at $(x_i, y_j) \in (0, 1)^2$ as

$$P_{x^\pm} = \frac{1 \pm g_x(x_i, y_j)}{4}, \qquad P_{y^\pm} = \frac{1 \pm g_y(x_i, y_j)}{4}, \tag{2.8}$$

where for $0 < i, j < (N - 1)$,

$$\begin{aligned} g_x(x_i, y_j) &:= k\big(c(x_i + h, y_j) - c(x_i - h, y_j)\big), \\ g_y(x_i, y_j) &:= k\big(c(x_i, y_j + h) - c(x_i, y_j - h)\big). \end{aligned} \tag{2.9}$$

It can be readily verified that the four probabilities in Eq. (2.8) sum to 1. The parameter value k does not vary with respect to space and is chosen to ensure that $|g_x(x, y)| \leq 1$ and $|g_y(x, y)| \leq 1$ (consequently, no probability can become negative).

We incorporate a discrete version of the snail-trail assumption by letting ECs proliferate in the space vacated by a TC when it moves from a lattice site. In other words, any new vessel production within a time step Δt is equal to the number of times TCs move, since ECs proliferate to occupy the resulting empty space. Hence the rate at which ECs are produced is based on both the value of P_m, the probability that a tip cell moves, and the number of tip cells currently at the lattice site.

We now define the discrete analogue of the net TC flux magnitude, $\|\mathbf{J}_{\text{net}}\|_2$, as the expected net number of times TCs move from lattice site (x_i, y_j) (we use expected values because of the stochastic nature of the discrete model). This means that κ records the total number of ECs produced per unit of net TC movement.

In order to calculate the expected net number of times a TC leaves the lattice site (x_i, y_j), we construct random variables X_R, X_L, Y_U, and Y_D that measure the total number of rightward/leftward/upward/downward moves, respectively, from this location. Since TCs can only move in these four directions, these random variables can be reasonably modelled with a multinomial distribution. We also define net random variables $X_{\text{net}} = X_R - X_L$ and $Y_{\text{net}} = Y_U - Y_D$, which measure the net number of rightward/upward jumps, respectively, made from the lattice site. By the argument above, the components of the discrete TC flux vector are given by $J_x := \mathbb{E}[X_{\text{net}}]$ and $J_y := \mathbb{E}[Y_{\text{net}}]$, while its magnitude is $\sqrt{(J_x)^2 + (J_y)^2}$. We remark that X_{net} and Y_{net} will always underestimate the total stalk cell production rate when tip cell movement is purely diffusive, as in this case both expected values are equal to zero.

Since $\kappa \|\mathbf{J}_{\text{net}}\|_2$ is the total number of ECs produced (or, equivalently, total times TCs move) from lattice site (x_i, y_j) within the time step Δt and spatial interval h, it follows that when the value of $\|\mathbf{J}_{\text{net}}\|_2$ is normalised to 1, κ becomes equal to the total number of times TCs move. The expected values of X_R, X_L, Y_U, and Y_D can be calculated in a straightforward manner: by using the definition of expectation for discrete random variables and the multinomial theorem, we calculate

$$\mathbb{E}[X_R] = \kappa P_{x^+}, \qquad \mathbb{E}[X_L] = \kappa P_{x^-}, \qquad \mathbb{E}[Y_U] = \kappa P_{y^+}, \qquad \mathbb{E}[Y_D] = \kappa P_{y^-}, \tag{2.10}$$

when $\|\mathbf{J}_{\text{net}}\|_2$ is normalised to 1, so that

$$\begin{aligned} \mathbb{E}[X_{\text{net}}] &= \mathbb{E}[X_R] - \mathbb{E}[X_L] = \kappa(P_{x^+} - P_{x^-}), \\ \mathbb{E}[Y_{\text{net}}] &= \mathbb{E}[Y_U] - \mathbb{E}[Y_D] = \kappa(P_{y^+} - P_{y^-}). \end{aligned} \tag{2.11}$$

Since the above equations apply when $\|\mathbf{J}_{\text{net}}\|_2 = 1$, we deduce from Eq. (2.8) that

$$\|\mathbf{J}_{\text{net}}\|_2 = \sqrt{(J_x)^2 + (J_y)^2} = \sqrt{\mathbb{E}[X_{\text{net}}]^2 + \mathbb{E}[Y_{\text{net}}]^2} = \frac{\kappa}{2}\sqrt{(g_x)^2 + (g_y)^2} = 1, \tag{2.12}$$

with g_x and g_y defined by Eq. (2.9). Hence

$$\kappa = \frac{2}{\sqrt{(g_x)^2 + (g_y)^2}} > 0. \tag{2.13}$$

We may apply Taylor's theorem to simplify Eq. (2.13): assuming that the average cell length is sufficiently small relative to the domain length such that $0 < h \ll 1$, we may write g_x as

$$g_x(x,y) = k(c(x+h,y) - c(x-h,y)) = k\Big(2h\frac{\partial c}{\partial x} + O(h^3)\Big) \approx 2kh\frac{\partial c}{\partial x}(x,y). \tag{2.14}$$

A similar formula holds for g_y. Substitution into Eq. (2.13) then yields

$$\kappa(x,y) \approx \frac{1}{kh\|\nabla c(x,y)\|_2}, \tag{2.15}$$

so that $\kappa(x,y)$ is inversely proportional to the magnitude of the TAF gradient. We remark that κ is a function of space because the TAF gradient may depend on the variables x and y. Although we do not consider cases in which the TAF field is time-dependent, it is straightforward to extend the above analysis to such cases.

We may further transform Eq. (2.15) so that it is in terms of the continuum parameters D and χ: we do this by exploiting the relationships between discrete and continuum parameters that were originally identified in Pillay et al. (2017) and Simpson et al. (2009). Namely, if P_m denotes the probability that a TC moves within a time step Δt, then we may write

$$D = \lim_{h\to 0,\Delta t\to 0} \frac{P_m h^2}{4\Delta t}, \quad \chi = \lim_{h\to 0,\Delta t\to 0} \frac{P_m k h^2}{\Delta t}, \tag{2.16}$$

assuming that the above limits exist, are non-zero, and are finite. Substituting these expressions into Eq. (2.15) yields

$$\kappa(x,y) = \frac{4D}{h\chi\|\nabla c(x,y)\|_2} = \frac{\mu h}{\chi\|\nabla c(x,y)\|_2}, \tag{2.17}$$

where we have defined $\mu := P_m/\Delta t = 4Dh^{-2}$.

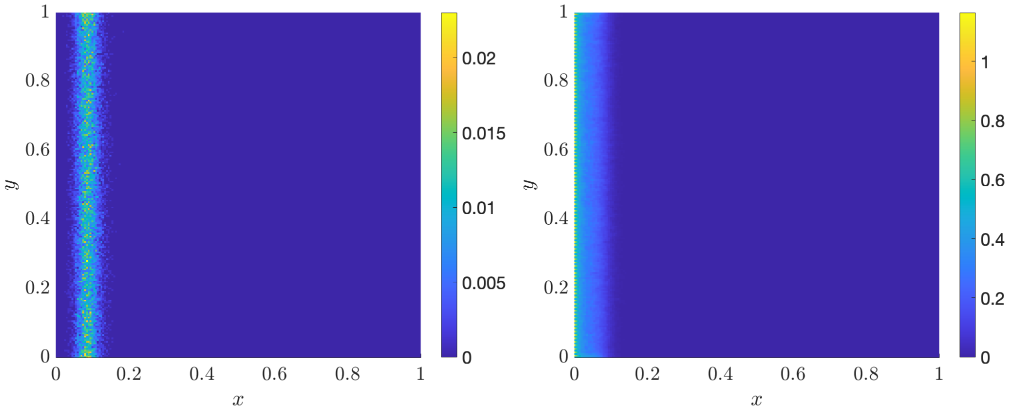
Equation (2.17) predicts that $\kappa \to \infty$ when $\chi\|\nabla c\|_2 \to 0$. Since both χ and ∇c relate to chemotactic mechanisms in the snail-trail framework, we conclude that the formula for the corrective factor breaks down when chemotaxis does not dominate TC movement. This occurs because the corrective factor attempts to correct for a near-zero vessel production rate: when there is no TAF gradient, TCs move randomly and isotropically such that $\|\mathbf{J}_{\text{net}}\|_2 \approx 0$. If we ignore the corrective factor in Eq. (2.5), then we anticipate that no new vessels will be created. In the discrete framework considered above, however, TCs that move randomly still produce ECs when they leave a lattice site – thus the corrective factor becomes large in order to account for this non-zero number of ECs.

We therefore view Eq. (2.17) as a "leading-order" approximation that is accurate when χ and $\|\nabla c\|_2$ are sufficiently large, so that chemotaxis dominates TC movement.

We conclude this section by discussing how the analysis is affected when TCs and ECs do not have the same length h. We account for different cell lengths by defining the discrete lattice step size as equal to a typical TC length (h_{TC}), while the EC length is denoted by the separate quantity h_{EC}. Since we have only employed the ABM lattice length to derive Eq. (2.17), the expression for $\kappa(x, y)$ is unchanged (the value of h in that equation, however, should be written as h_{TC} for clarity). The EC rate of change, however, does depend on h_{EC} in this scenario. To see this, we consider γ stalk cells to be equivalent to a single TC, such that $h_{\mathrm{TC}} = \gamma h_{\mathrm{EC}}$. Since $\kappa(x, y) = \mu h_{\mathrm{TC}}/(\chi \|\nabla c\|_2)$, this means that the ratio $\kappa(x, y)/h_{\mathrm{EC}}$ that appears in Eq. (2.5) reduces to $\mu\gamma/(\chi\|\nabla c\|_2)$. This expression is nearly identical to the one obtained when TC and EC lengths are equal – the only difference is that it is now multiplied by γ. Hence the EC rate of change is altered by a factor of γ when TC and EC lengths differ. This result makes sense because changing the EC length should only affect the density of ECs (not vessels) located along TC paths.

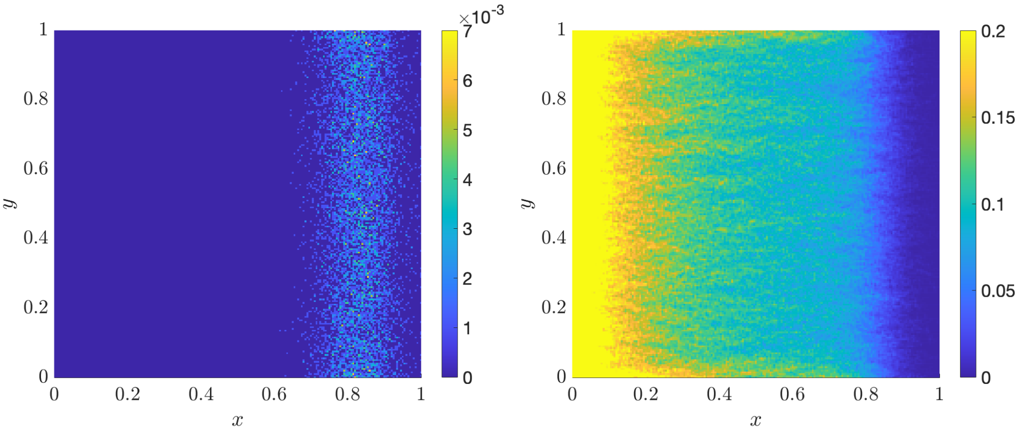
2.4 A multiplicative factor is necessary in the 2D snail-trail stalk cell equation

In this section, we confirm that the 2D snail-trail model underestimates the "true" stalk cell density when the corrective factor κ is not included in the framework. Specifically, we compare numerical solutions of the 2D snail-trail model, with $\kappa(x, y)/h = 1$, to those of a 2D ABM for angiogenesis developed by Pillay et al. (2017). We use this particular ABM (henceforth denoted as the P–ABM) because it is a discrete, cell-based representation of the verbal description that the classical snail-trail formulation purports to model. In other words, TC and EC movement in this ABM follow the same rules as those for the discrete model presented in §2.3. The P–ABM is a more detailed description of angiogenesis,



(a) Tip cell distribution, $t = 0.2$

(b) Stalk cell distribution, $t = 0.2$



(c) Tip cell distribution, $t = 2$

(d) Stalk cell distribution, $t = 2$

Figure 2.2: Heat maps of (a, c) the ensemble average P–ABM tip cell distributions and (b, d) its corresponding stalk cell distributions, which are computed over one thousand realisations and are presented for times (a, b) $t = 0.2$ and (c, d) $t = 2$. Colour bars indicate the average number of cells at each lattice point. §A.1 lists the P–ABM parameter values and initial conditions.

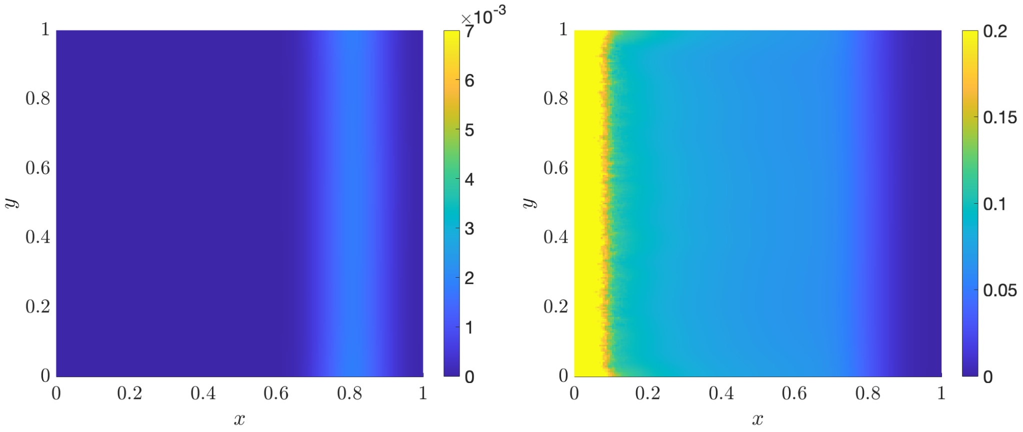
however, because it directly models phenomena such as anastomosis and TC branching that were ignored in the previous section. Full details on the P–ABM are found in §A.1.

In Figure 2.2, we present snapshots of ensemble average P–ABM TC and EC distributions generated in response to the TAF field $c(x, y) = x$, using initial conditions and parameter values outlined in §A.1. Although we observe noise in the TC distributions for the two time points presented in Figures 2.2(a) and 2.2(c), both densities are approximately constant in the y-direction for this choice of initial conditions and TAF field. Comparison of Figures 2.2(a) and 2.2(c) also reveals that TCs in the P–ABM travel to the right and spread out over larger spatial areas as time increases, and that the maximum number of TCs within a single lattice site decreases with time.

Figures 2.2(b) and 2.2(d) present the corresponding ensemble average EC distributions for this TAF field, initial conditions, and parameter regime. We observe that the EC distribution also travels up the TAF gradient towards the right hand boundary. Furthermore, the average number of ECs found within a single lattice site appears to decrease along the x-direction. Unlike the TC distribution, however, the EC results may not be constant in the y-direction: at sites near the top and bottom edges of the lattice, for instance, the EC density appears to be greater than at other points in the domain interior. This is likely an edge effect due to the boundary conditions of the P–ABM: TCs that attempt to exit the domain (either on the left boundary or the upper and lower boundaries) are reflected back to the original lattice site, but this action still produces an EC. Hence the total number of ECs will be greater near the boundaries than within the domain interior.

Since such edge effects are likely to lead to differences between the P–ABM and 2D snail-trail model solutions, we opt to reduce their influence by initializing the 2D snail-trail model with P–ABM data at time $t_0 = 0.2$, rather than $t_0 = 0$, to avoid simulating the snail-trail PDE too close to the left-hand boundary. Hence solutions to the 2D snail-trail model were calculated for the time interval $t \in [0.2, 2]$, rather than $t \in [0, 2]$.

We compare these P–ABM distributions to results from the 2D snail-trail model given by Eq.s (2.4)-(2.7), using expressions derived in Pillay et al. (2017) to determine the values of D, χ, λ, and β_n from the P–ABM parameters. Although the parameter β_e cannot be determined *a priori*, we opt to use an *ad hoc* approach to determine its value by minimising the sum of squared residuals between 2D snail-trail and P–ABM solutions (see §A.2 for details on the numerical methods). Since we assume in this section that no corrective factor is required, we also fix $\kappa(x,y)/h = 1$. We present the corresponding continuum parameter values in Table 2.1.



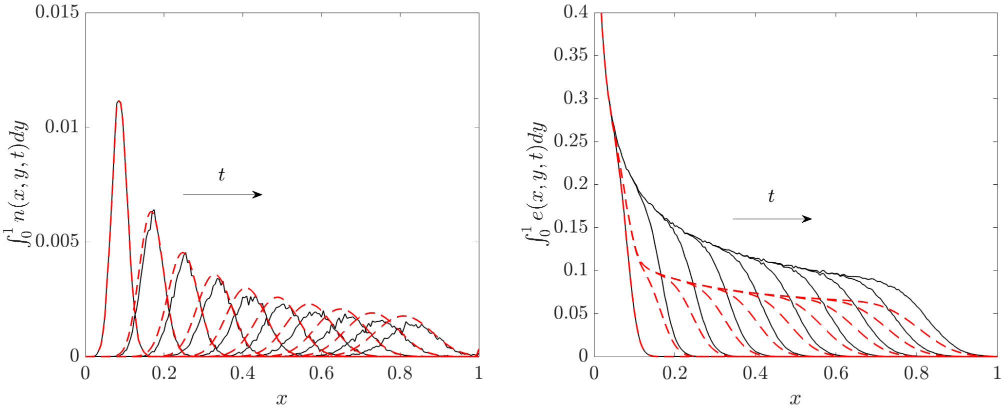
(a) Snail-trail PDE, tip cell density (b) Snail-trail PDE, stalk cell density

Figure 2.3: Heat maps of (a) the tip cell, $n(x,y,t)$, and (b) the stalk cell, $e(x,y,t)$, solutions for the 2D snail-trail model, which are subject to the TAF field $c(x,y) = x$, at $t = 2$. Colour bars indicate the cell density. Continuum parameters are listed in Table 2.1. Note that we have fixed $\kappa(x,y)/h = 1$ for this simulation. The parameter β_e was fitted to P–ABM results using numerical methods described in §A.2 ($\beta_e = 4.65 \times 10^{-12}$, 95% CI: $[0, 0.123]$). The snail-trail model was initialised at $t = 0.2$ using the average P–ABM distribution (see Figure 2.2).

D	χ	λ	β_n	h	β_e
10^{-3}	0.4	0.16	160	200^{-1}	(see figure caption)

Table 2.1: Snail-trail model parameter values used to generate Figures 2.3-2.5. These are determined from the P–ABM parameters (see §A.1), using relationships derived in Pillay et al. (2017).

Figure 2.3 shows the resulting 2D snail-trail TC and EC solutions at time $t = 2$. We observe in Figure 2.3(a) that the TC density appears to be approximately uniform in the y-direction, and resembles that of the P–ABM. However, differences between the two solutions are larger than one would expect from visual inspection. We quantify such discrepancies by computing the sum of squared differences between the P–ABM and snail-trail densities



(a) Tip cell densities

(b) Stalk cell densities

Figure 2.4: Solutions for (a) the tip cell, $n(x, y, t)$, and (b) the stalk cell, $e(x, y, t)$, densities of the P–ABM and 2D snail-trail model at $t = 0.2$, 0.4, ..., 2, which have been column averaged in the y-direction. We have fixed $\kappa(x, y)/h = 1$ in the snail-trail model. Key: P–ABM distribution (solid black lines); column averaged 2D snail-trail solution (red dashed lines). Initial conditions and parameter values: as in Figures 2.2-2.3.

across all lattice sites at the final time point, and then taking its square root (for the rest of this chapter, we will refer to this metric as the root sum of squared residuals). For the data above, this value is approximately equal to 0.10, and is about 69% of the corresponding norm for the average P–ABM distribution at this time point (the latter value is found by squaring the P–ABM density at every lattice site, summing the results, and then taking the square root of the resulting number – for the rest of this chapter, we will refer to this as the root sum of the squared average P–ABM density). These large relative differences suggest that the snail-trail model does not accurately describe average P–ABM distributions.

Figure 2.3(b) presents the 2D snail-trail EC solution at time $t = 2$. Although we observe that it decreases along the x-direction, just as the P–ABM EC distribution, comparison of this graph with Figure 2.2(d) reveals clear discrepancies between the two solutions. For instance, in Figure 2.3(b) the snail-trail EC density appears to be constant in the y-direction. At $t = 2$, the root sum of squared residuals between the P–ABM and snail-trail EC solutions across all lattice sites is 8.2, which is about 26% of the root sum of the squared average P–ABM stalk cell density. These results indicate that the snail-trail model does not accurately capture the ensemble average EC density when we ignore the corrective factor.

To better visualise the extent of differences between the discrete and continuum solutions, we numerically integrate the 2D results of each model in the y-direction. The resulting column averaged data are presented in Figure 2.4. We observe in Figure 2.4(a) that the snail-trail TC solutions appear to have the same bell curve shape as P–ABM distributions, and that their leading edges travel with roughly identical speeds (they are within 3% of each other). However, the continuum TC solution tends to overestimate the discrete distribution at all time points shown. We conclude that while the snail-trail model with $\kappa/h = 1$ captures some aspects of the P–ABM TC distribution, it is not entirely accurate.

Figure 2.4(b) presents the column averaged discrete and continuum EC densities. We observe that the snail-trail solution underestimates P–ABM results for all of the time points shown. This underestimation persists even when we exclude possible edge effects and column average the 2D results over a smaller interval given by $y \in [0.05, 0.95]$ (data not shown). Such results indicate that edge effects do not wholly cause the large differences between the snail-trail and P–ABM EC solutions – instead, they arise from within the domain interior.

These differences also do not originate from branching and anastomosis. We confirm this by setting the parameters of those processes equal to zero in both models (i.e., $\lambda = \beta_e = \beta_n = 0$). The resulting column averaged snail-trail EC solution is still always less than that of the P–ABM for this parameter regime (data not shown). In fact, its density is approximately half the value of the P–ABM distribution.

These results demonstrate that when the corrective factor is not included in the continuum framework, the 2D snail-trail model underestimates the total discrete EC density and thus does not accurately describe its results. We conclude that the EC rate equation should include a multiplicative factor κ/h, not necessarily equal to one, in order to accurately capture the P–ABM EC distribution. Furthermore, the underestimation of the P–ABM EC density occurs even when branching and anastomosis are neglected, which is consistent with the arguments from §2.2 and §2.3.

Case	χ	$c(x,y)$	$\|\nabla c\|_2$	$\kappa(x,y)/h$
1	0.4	x	1	2
2	0.4	$1-(x-\frac{1}{2})^2-(y-\frac{1}{2})^2$	$2\sqrt{(x-\frac{1}{2})^2+(y-\frac{1}{2})^2}$	$\frac{1}{\sqrt{(x-\frac{1}{2})^2+(y-\frac{1}{2})^2}}$
3	0.4	xy	$\sqrt{y^2+x^2}$	$2(y^2+x^2)^{-1/2}$
4	0.04	x	1	20

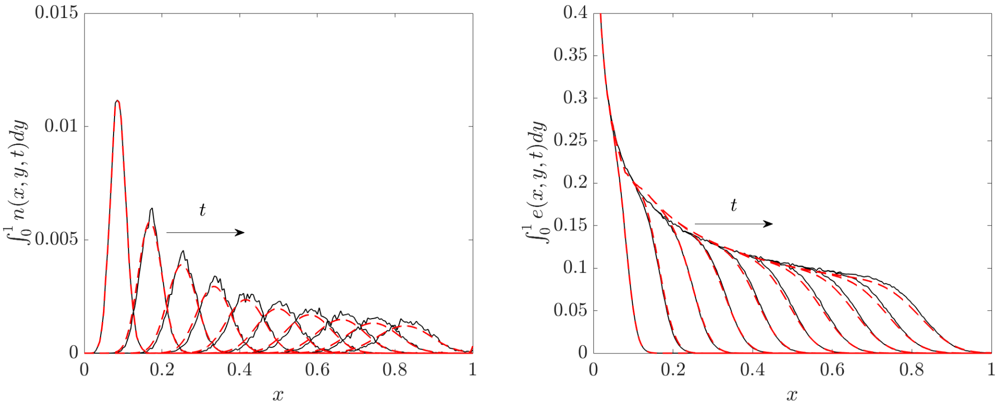
Table 2.2: TAF fields, gradient magnitudes, and expressions for κ used to generate the figures presented in §2.5-§2.6. In all cases, $D = 10^{-3}$, $\chi = 0.4$, and $h = 200^{-1}$. Except for Case 1, the other parameter values used in the snail-trail model were $\lambda = \beta_e = \beta_n = 0$. Results for Cases 3 and 4 are not shown here.

2.5 Validation of the snail-trail corrective factor by numerical simulation

These results and earlier analysis suggest that including the multiplicative factor $\kappa(x,y)$, given by Eq. (2.17), will cause the 2D snail-trail stalk cell evolution equation to more accurately describe ensemble average P–ABM distributions, provided TC movement is dominated by chemotaxis. In this section, we test this hypothesis by examining continuum solutions for different TAF fields. The scenarios that we consider are presented in Table 2.2.

We first examined the case from §2.4 for which $c(x,y) = x$. For this scenario, we use Eq. (2.17) to conclude that $\kappa(x,y)/h = 2$. Although the resulting 2D snail-trail solution profiles (data not shown) resemble those for the case in which $\kappa(x,y)/h = 1$, further investigation reveals that, for both TC and EC solutions, the snail-trail model with $\kappa(x,y)/h = 2$ more accurately estimates the average P–ABM distribution.

To see this, we present the column averaged TC and EC solutions in Figure 2.5. We observe in Figure 2.5(a) that the continuum TC density is almost indistinguishable from the corresponding P–ABM distribution (the maximum pointwise difference between the two solutions is at most 30% of the maximum TC height for the time points shown, and the total TC mass of the continuum and discrete models are within 4% of each other). Comparison of this graph with Figure 2.4(a) further reveals that the snail-trail model with $\kappa/h = 2$ overestimates the P–ABM solution to a lesser extent than when $\kappa/h = 1$. Figure 2.5(b) presents



(a) Tip cell densities

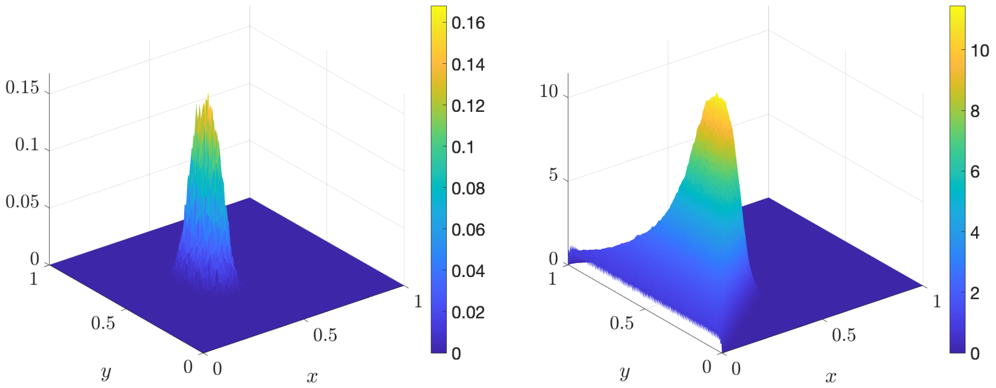
(b) Stalk cell densities

Figure 2.5: Solutions of (a) the tip cell, $n(x, y, t)$, and (b) the stalk cell, $e(x, y, t)$, densities given by the P–ABM and 2D snail-trail model, column averaged in the y-direction, at $t = 0.2$, 0.4, ..., 2, with $D = 10^{-3}$, $\chi = 0.4$, $\beta_n = 160$, $\lambda = 0.16$, and $\kappa(x, y)/h = 2$. The parameter β_e was fitted to the P–ABM results using the numerical methods described in §A.2 ($\beta_e = 4.77$, 95% CI: $[4.73, 4.82]$). Key: P–ABM distribution (solid black lines); 2D snail-trail solution (red dashed lines). The PDE was initialised using P–ABM data at $t = 0.2$ and simulated on the interval $t \in [0.2, 2]$.

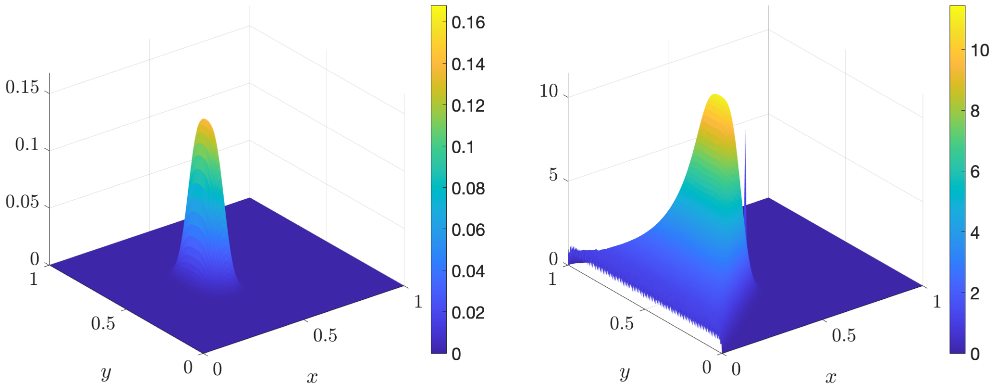
the column averaged EC distributions. We again observe that the continuum and discrete solutions are almost indistinguishable from each other. The maximum pointwise difference between the continuum and discrete solutions is approximately 1% of the maximum EC height for each time point shown, for instance. These observations indicate that including a corrective term given by our formula for $\kappa(x, y)$ improves the agreement between solutions of the 2D snail-trail model and the P–ABM.

In the aforementioned example, the chemotactic gradient varies in only one spatial dimension. However, these results can be extended to 2D chemotactic gradients: we have verified this for the TAF fields $c(x, y) = 1 - (x - 0.5)^2 - (y - 0.5)^2$ and $c(x, y) = xy$. For such simulations, we neglect branching and anastomosis because we have determined that neither process will affect the value of the multiplicative factor (in other words, for the remaining results presented in this section, $\lambda = \beta_e = \beta_n = 0$).

Figures 2.6-2.7 present results for the TAF field $c(x, y) = 1 - (x - 0.5)^2 - (y - 0.5)^2$ (Case 2 in Table 2.2). The 2D P–ABM and snail-trail results at $t = 2$ are presented in Figure 2.6. We observe in Figure 2.6(a) that TCs in the P–ABM travel from the left-hand boundary

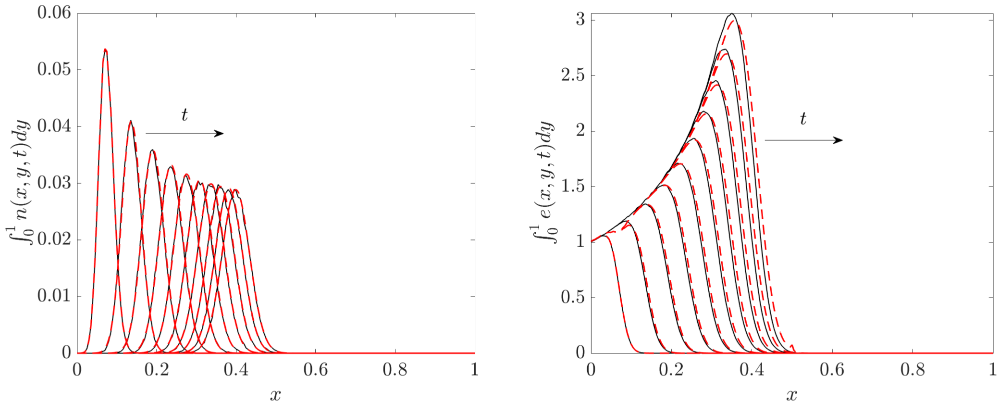


(a) 2D P–ABM distribution, tip cells (b) 2D P–ABM distribution, stalk cells



(c) 2D snail-trail solution, tip cells (d) 2D snail-trail solution, stalk cells

Figure 2.6: Results for (a) the P–ABM tip cell distribution, (b) the P–ABM stalk cell distribution, (c) the snail-trail model tip cell density, and (d) the snail-trail model stalk cell density at $t = 2$, which are subject to the TAF field $c(x, y) = 1 - (x - 0.5)^2 - (y - 0.5)^2$. See Table 2.2 for parameter values used to simulate the continuum model. The P–ABM used the same initial condition described in §A.1, and its solution at $t = 0.2$ was used to initialise the snail-trail PDE.



(a) Tip cell densities

(b) Stalk cell densities

Figure 2.7: Results for (a) the tip cell, $n(x, y, t)$, and (b) the stalk cell, $e(x, y, t)$, densities from the P–ABM and 2D snail-trail model, which are subject to the TAF field $c(x, y) = 1 - (x - 0.5)^2 - (y - 0.5)^2$ and are column averaged in the y-direction, at times $t = 0.2$, 0.4, ..., 2. Initial conditions and parameter values: as in Figure 2.6. Key: P–ABM distribution (solid black lines), snail-trail solution (dashed red lines).

and cluster near the point $(x, y) = (0.5, 0.5)$, which is the location of the maximum TAF concentration. This behaviour is expected, since the TAF gradient is directed towards the domain centre – thus chemotaxis will cause TCs to move towards this location. We also see in Figure 2.6(b) that the EC density becomes very large near the domain centre. This is because the magnitude of the TAF gradient becomes small near $(x, y) = (0.5, 0.5)$: TCs in the P–ABM thus become more likely to move via random motion in this area which, in turn, produces many ECs since we neglect anastomosis.

Figures 2.6(c) and 2.6(d) present the corresponding snail-trail solutions for this TAF field. We observe that the continuum TC density in Figure 2.6(c) resembles that of the P–ABM, although there are some differences due to noise. The root sum of squared residuals between the tip cell solutions, for example, is within 11% of the root sum of the squared average P–ABM tip cell distribution for this time point. Figure 2.6(d) reveals that the snail-trail EC solution also resembles that of the P–ABM (the root sum of squared residuals between the two solutions is also within 11% of the corresponding root sum of the squared average P–ABM stalk distribution). The most evident discrepancy with the P–ABM EC results occurs near the point $(x, y) = (0.5, 0.5)$, where a "spike" of stalk cells arises because the value of $\kappa(x, y)$ becomes large there. In the numerical scheme, we have approximated the blow-up of $\kappa(x, y)$ when $\chi\|\nabla c\|_2 \to 0$ by setting $\|\nabla c\|_2$ to be a positive value below machine precision ($< 10^{-16}$), in order to ensure that the numerical solution converges to a finite solution throughout the domain.

We observe this overestimation of the P–ABM stalk cell density even when both sets of 2D solutions are column averaged in the y-direction (results are presented in Figure 2.7). We also find that the maximum difference between the column averaged tip cell solutions is less than 5% of the maximum tip cell density height for the time points shown in Figure 2.7(a). Furthermore, the total tip cell masses computed by the discrete and continuum models are within 1% of each other. Both of these results indicate that the snail-trail model accurately estimates the P–ABM tip cell solution. However, there is greater discrepancy between the column averaged stalk densities in Figure 2.7(b). Although the total stalk mass given by the discrete and continuum models are very similar to each other (with approximately 1% difference), the maximum pointwise difference between the two solutions is on the order of 15% of the maximum EC height for the time points shown here. Such (relatively small) differences arise because the snail-trail model overestimates the stalk cell density rate of change when the cell densities approach $x = 0.5$. This extra stalk production also explains why the column averaged snail-trail EC solution appears to travel a farther distance than that of the P–ABM.

We also computed solutions for the TAF field $c(x, y) = xy$, which corresponds to Case 3 of Table 2.2. The conclusions for this case are similar to those for Case 2, hence we do not present the 2D and column averaged solutions here.

We have also conducted another experiment that confirms the conclusion that the discrepancies between the ABM and 2D snail-trail results increase when the value of χ, the chemotactic sensitivity of TCs, decreases. Specifically, we simulated both models for a case in which the cells invade a linear TAF field ($c(x, y) = x$), but have a value of χ that is an order of magnitude smaller than its value in Figure 2.5 (this corresponds to Case 4 of Table 2.2). While we indeed find that solutions of the 2D snail-trail model fail to accurately capture ensemble average P–ABM distributions, we do not present these results here.

The numerical results support the hypothesis that Eq. (2.17), which defines the multiplicative factor $\kappa(x, y)$, causes the snail-trail model solutions to approximate P–ABM results well, except in regions where chemotaxis does not dominate TC migration.

We conclude this section by testing whether it is possible to fix $\kappa(x, y)$ to be constant and still obtain similar agreement between the discrete and continuum models. This experiment is inspired by the *ad hoc* methods of previous investigators, who employed nonlinear least square approaches to estimate a fixed corrective EC factor in 1D snail-trail models (Pillay et al., 2018, 2017; Connor et al., 2015). We accordingly solve the 2D snail-trail PDE with a constant value of κ, which is selected to minimise the squared difference between the continuum and discrete solutions, for the first three TAF fields listed in Table 2.2 (we describe the numerical methods in §A.2). When the magnitude of the TAF gradient is constant, as in the case when $c(x, y) = x$, we find that the optimal value of κ is within 1% of the value predicted by Eq. (2.17) (data not shown). For more general TAF fields whose gradients vary in space, however, we find that the discrepancies between the snail-trail and P–ABM solutions are greater when κ is taken to be constant than in cases where its value is determined by Eq. (2.17) (this is the case for the TAF field $c(x, y) = 1-(x-0.5)^2-(y-0.5)^2$, although we do not present the results here). Thus for general TAF fields whose gradients vary in space, the optimal multiplicative factor is not necessarily constant. This contrasts with the assumptions made in previous approaches for estimating the corrective factor.

2.6 Derivation of the 1D snail-trail model

While we have shown that the 2D snail-trail model presented in §2.2 can accurately estimate results from the P–ABM, its relationship with other 1D snail-trail systems remains unclear. Such 1D models ostensibly describe TC and EC dynamics that have been column averaged in the direction perpendicular to that of the moving TC front (Connor et al., 2015; Byrne and Chaplain, 1995), but it is unknown what assumptions are required to ensure that this interpretation is appropriate. In order to address these open questions, we formally reduce the 2D snail-trail system to one spatial variable by deriving the dynamics that govern its column averaged solutions and comparing them to existing 1D snail-trail models. In doing so, we elucidate assumptions that are required in order for 1D snail-trail models to accurately describe experimental results during angiogenesis.

To do this, we define the new column averaged dependent variables

$$N(x,t) = \int_0^1 n(x,y,t)dy, \quad E(x,t) = \int_0^1 e(x,y,t)dy, \quad C(x) = \int_0^1 c(x,y)dy. \tag{2.18}$$

Integrating Eq.s (2.4)-(2.5) with respect to y and applying the Neumann boundary conditions yields

$$N_t = \int_0^1 Dn_{xx} - \chi(nc_x)_x dy + \lambda \int_0^1 ncdy - \beta_e \int_0^1 nedy - \beta_n \int_0^1 n^2 dy, \tag{2.19}$$

$$E_t = \frac{1}{h} \int_0^1 \kappa(x,y) \| D\nabla n - \chi n \nabla c \|_2 dy, \tag{2.20}$$

where subscript notation has been used to indicate partial derivatives taken with respect to x and t. For general TAF fields and TC/EC cell solutions, the above system cannot be reduced further. However, under the following assumptions we can simplify both equations to one spatial variable. We first make a mean-field approximation in the tip cell rate equation so that

$$\begin{gathered}
\int_0^1 ncdy \approx \Big[\int_0^1 ndy\Big] \times \Big[\int_0^1 cdy\Big] = NC, \\
\int_0^1 nedy \approx \Big[\int_0^1 ndy\Big] \times \Big[\int_0^1 edy\Big] = NE, \\
\int_0^1 n^2 dy \approx \Big[\int_0^1 ndy\Big] \times \Big[\int_0^1 ndy\Big] = N^2, \\
\int_0^1 nc_x dy \approx \Big[\int_0^1 ndy\Big] \times \Big[\int_0^1 c_x dy\Big] = \Big[\int_0^1 ndy\Big] \times \Big[\frac{\partial}{\partial x}\int_0^1 cdy\Big] = NC_x.
\end{gathered} \tag{2.21}$$

In making the mean-field approximations, and expressing rates as products of cell densities and/or TAF concentrations, we are assuming that these quantities are independent of each other (Penington et al., 2011; Simpson and Baker, 2011). When there are spatial correlations between the cell densities, as is the case for the tip and stalk cell densities (since stalk cells are produced along tip cell paths), we expect the mean-field approximation to become invalid. For the simulations presented above the cell densities are small enough that this breakdown is not apparent, however it is more pronounced in situations with frequent sprout branching and tip-sprout anastomosis (Pillay et al., 2017). By applying the Leibniz integral rule and substituting in the above mean-field approximations, we find that the first integral in Eq. (2.19) reduces to $DN_{xx} - \chi(NC_x)_x$.

In order to reduce the stalk cell evolution equation to a form that resembles that of the classical 1D snail-trail model, we again assume a mean-field approximation is valid and that the y-component $(Dn_y - \chi nc_y)$ of the net tip cell flux is negligible; the former assumption permits us to consider the integrals of $\kappa(x, y)$ and the tip cell flux separately, while the latter simplifies the magnitude of tip cell flux to an absolute value function. Specifically, these assumptions lead to the equation

$$E_t \approx \frac{1}{h}\int_0^1 \kappa(x,y)dy \times \int_0^1 \Big|Dn_x - \chi nc_x\Big| dy. \tag{2.22}$$

If the TAF field varies in the y-direction, then the first integral may diverge: for example it is undefined for the TAF field $c(x, y) = 1 - (x - 0.5)^2 - (y - 0.5)^2$. We therefore define a new multiplicative factor $\widetilde{\kappa}(x)$ that applies only to the column averaged dynamics. We use a similar formula to Eq. (2.17): if the column averaged TAF field is given by $C(x)$, then we write the 1D version of the corrective factor as

$$\widetilde{\kappa}(x) \approx \frac{4D}{h\chi|C_x|}. \tag{2.23}$$

For the 2D TAF fields listed in Table 2.2, we have found that using Eq. (2.23) allows the 1D snail-trail model to accurately approximate column averaged 2D continuum results in regions where the TAF gradient is non-zero (data not shown).

If we also assume that the sign of $(Dn_x - \chi nc_x)$ does not change in the domain, then we may take the absolute value sign out of the second integral of Eq. (2.22). This assumption is motivated by experimental observations that TCs are not likely to change direction as they migrate towards TAF sources (Connor et al., 2015), and by our previous simplification that the chemotactic gradient is fixed with respect to time t. We therefore anticipate that the net TC flux will not change direction (i.e., not change sign), as TCs migrate to the TAF source at the right-hand boundary. Since we have already assumed that a mean-field approximation is valid, we may use the Leibniz integration rule to simplify the second integral to $|DN_x - \chi NC_x|$.

Under the above assumptions, the 1D version of Eq.s (2.4)-(2.6) is

$$\frac{\partial N}{\partial t} = D\frac{\partial^2 N}{\partial x^2} - \chi\frac{\partial}{\partial x}\Big(N\frac{\partial C}{\partial x}\Big) + \lambda NC - \beta_e NE - \beta_n N^2, \tag{2.24}$$

$$\frac{\partial E}{\partial t} = \frac{\widetilde{\kappa}(x)}{h}\Big|D\frac{\partial N}{\partial x} - \chi N\frac{\partial C}{\partial x}\Big|, \tag{2.25}$$

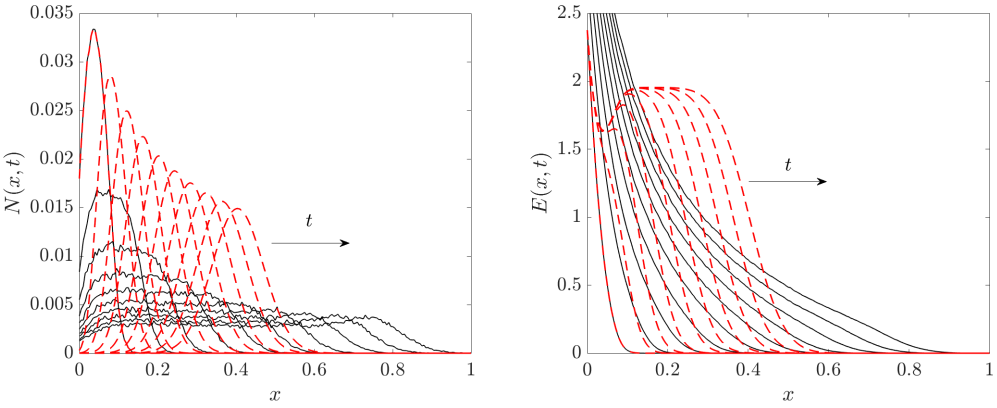
$$D\frac{\partial N}{\partial x} - \chi N\frac{\partial C}{\partial x} = 0 \text{ for } x \in \{0, 1\}, \tag{2.26}$$

where $\widetilde{\kappa}(x)$ is defined by Eq. (2.23).

When $c(x, y) = x$, we find that $\widetilde{\kappa}(x) = 2$ and the 1D snail-trail model is equivalent to the one used in Pillay et al. (2017). Additionally, the snail-trail model from Byrne and Chaplain (1995) may be recovered if one sets $\beta_n = 0$ and $\widetilde{\kappa}(x)/h = 1$, includes an additional term in Eq. (2.24) to account for TC branching from vessels, and introduces a term in Eq. (2.25) to account for EC elimination due to vessel regression. We conclude that the 1D snail-trail model is consistent with existing spatially reduced frameworks for angiogenesis, provided that the assumptions listed above are valid.

It is evident from this derivation that many conditions must be satisfied to ensure that the 1D snail-trail system captures the column averaged 2D continuum dynamics. We have found, however, that there are certain TAF fields, such as $c(x, y) = C(x) = x$, for which these conditions are all approximately satisfied, and for which the 1D snail-trail solutions accurately capture column averaged P–ABM results (data not shown).

When TAF fields exhibit marked variation in 2D, many of the above assumptions are likely to break down. In such cases, the 1D snail-trail model is unlikely to accurately approximate the column averaged dynamics of the P–ABM or 2D continuum model. When $c(x, y) = xy$, for instance, we violate the assumptions that c_x is independent of y (since $\nabla c = (y \quad x)^{\mathrm{T}}$ for this case) and that the y-component of net tip cell flux is negligible. Figure 2.8 presents column averaged P–ABM results for this TAF field, along with 1D snail-trail model solutions generated using the column-averaged 1D field $C(x) = 0.5x$. We observe that the snail-trail TC and EC densities are poor approximations to the P–ABM results for the time points shown. For example, the P–ABM TC and EC solution fronts appear to travel with faster speeds than the 1D snail-trail solutions. This is due to the difference



(a) Tip cell densities

(b) Stalk cell densities

Figure 2.8: Solutions for (a) the tip cell, $N(x, t)$, and (b) the stalk cell, $E(x, t)$, densities given by the 1D snail-trail model and the P–ABM at $t = 0.2$, 0.4, ..., 2. The P–ABM is subject to the TAF field $c(x, y) = xy$, and its results have been column averaged over the interval $y \in [0, 1]$; the 1D snail-trail model is subject to the column averaged TAF field ($C(x) = 0.5x$), such that $\widetilde{\kappa}(x) = 4$. Key: column averaged P–ABM distribution (solid black lines), 1D snail-trail solution (dashed red lines). Parameter values: as in Table 2.2. The P–ABM uses the same initial condition outlined in §A.1, while the 1D snail-trail model was initialised with the column averaged P–ABM results at $t = 0.2$.

between the 2D TAF field $c(x, y) = xy$ and its column averaged counterpart $C(x) = 0.5x$: since TCs in both frameworks move faster in larger TAF gradients, and the x-component of the 2D TAF gradient is equal to y, TCs in the 2D P–ABM move more quickly near the top domain boundary. We confirm this by visualising the P–ABM TC solutions in 2D (data not shown). The column averaged TAF field ($C(x) = 0.5x$), however, ignores any variation of the TAF distribution with respect to y, hence the 1D snail-trail model ignores these differences in TC speeds. We conclude that when the difference between the 2D TAF field and its column averaged counterpart is sufficiently large, the 1D snail-trail model will be a poor approximation to column averaged P–ABM results. In such cases, the 2D version of the snail-trail model is more appropriate to use, as it can capture the spatial variations in these quantities. Alternatively, changing the orientation of the reference frame may help establish the validity of the assumptions underlying the 1D snail-trail framework. This is the case, for example, for the 2D TAF field $c(x, y) = 0.5(x + y)$: by making the transformation $u = 0.5(x + y)$, $v = 0.5(x - y)$, the TAF field becomes $c(u, v) = u$, which only varies with respect to one variable.

2.7 Discussion and conclusion

In this chapter, we evaluated the validity of the classical snail-trail framework for angiogenesis (both its 1D and 2D version) by comparing its numerical solutions to those of an ABM developed in Pillay et al. (2017). The ABM incorporates the same mechanisms as the continuum model and is assumed to represent the underlying "ground truth". In particular, we formulated a 2D version of the snail-trail model for a domain with a generic chemoattractant at steady state, and modified its stalk cell evolution equation by introducing a novel multiplicative factor that has not previously appeared in the literature. By reasoning that this factor accounts for the subtleties of multi-directional cell movement, we derived a closed-form expression for this factor with a coarse-graining approach. Numerical solutions of the modified 2D snail-trail model are indeed accurate approximations of ensemble ABM distributions when chemotaxis dominates tip cell migration. This last condition represents a crucial, but heretofore implicit, assumption that must be satisfied for the snail-trail modelling framework to accurately describe leader-follower migration during angiogenesis.

After determining the conditions for which solutions to the 2D snail-trail model represent ABM data well, we then reduced the continuum framework to its 1D version by column averaging the model, making several mean-field approximations, and assuming that the 2D cell/TAF densities are closely represented by their column averaged fields. The numerical results indicate that when such assumptions are violated, the 1D snail-trail model (unlike its 2D counterpart) cannot accurately describe the discrete model solutions. The 1D snail-trail model is consistent with other existing continuum descriptions of angiogenesis (Pillay et al., 2017; Connor et al., 2015; Byrne and Chaplain, 1995). Hence the results in this chapter provide insight into when those frameworks can accurately capture cell dynamics in angiogenesis.

The results also imply that metrics that indicate when chemotaxis dominates cell migration – for example, the average cell displacement – could be used to infer when snail-trail models would provide accurate descriptions of angiogenesis. In future work, we aim to explore the ability of such metrics to predict how well such continuum frameworks describe experimental data (potentially including those from *in vitro* or *in vivo* assays).

There are several possible extensions that can be made to the current snail-trail framework to improve its applicability to experimental data. We assumed in this chapter, for instance, that all sources of cell motion (besides chemotaxis) could be modelled as an isotropic random motion term in the PDE system and P–ABM. Although this is a common modelling choice, there are certain cell lines whose motion cannot be described in this fashion (Wu et al., 2014). Thus, other models for cell migration, such as stochastic differential equations that incorporate anisotropy, should be considered and evaluated within the experimental context of angiogenesis. Furthermore, we analysed TC and EC dynamics for ideal situations where the TAF concentration was independent of time. In practice, it is also likely that the TAF field exhibits more complicated dynamics than those investigated in Table 2.2: for example, cells may degrade the chemoattractant to create a cell-induced gradient (Tweedy et al., 2020). It may therefore be necessary to extend the snail-trail modelling framework to consider the impact of such dynamics, for example by adding in a term into the TAF rate equation that is proportional to both the TC density and the TAF concentration (i.e., a term of the form $-\beta_c nc$, where β_c is a positive constant).

Since our focus in this chapter was to evaluate the accuracy of snail-trail models within the domain interior, we did not resolve the "edge effects" discovered in §2.4-§2.5 that contributed to differences between the P–ABM and 2D snail-trail solutions. Understanding how to correct for such boundary conditions will be required if snail-trail models are used to investigate experimental data, however. We aim to address this issue in future work.

While the corrective factor derived in this chapter is new to the 2D snail-trail framework, a term nearly identical to Eq. (2.17) appears in a hybrid ABM for angiogenesis developed in Vilanova et al. (2014) and Travasso et al. (2011). Although the latter framework does not use the net TC flux to calculate capillary densities, the similarity between the two expressions does not arise by chance: indeed, we show in §A.3 that one can re-derive Eq. (2.17) by employing the same logic that Vilanova et al. (2014) use to relate the discrete and continuum parts of their hybrid model. This equivalence results from a shared physical assumption underlying both modelling frameworks: that new vessels proliferate along the path of TCs. Hence this work may indicate that a core physical assumption of trail-following behaviour translates into a mathematical term that is conserved across distinct models.

This last observation raises the question of whether comparative analysis of other models for angiogenesis could similarly reveal the biological assumptions upon which they are based, along with conditions under which they accurately estimate experimental data. As a particular example, we could consider the corrected 1D snail-trail model and a coarse-grained PDE for angiogenesis (the P–PDE) that was derived from the P–ABM in Pillay et al. (2017). Although the P–PDE contains more nonlinear terms than the snail-trail model and its equation governing the EC dynamics is different, the authors identified parameter regimes for which the P–PDE generated indistinguishable solutions from those of a 1D snail-trail model that is equivalent to Eq.s (2.24)-(2.26). In Chapter 3, we aim to explain this observation by analysing the corrected snail-trail model and P–PDE, thereby determining more general regions of parameter space for which solutions to the two continuum models are indistinguishable. This may also provide insight into the biological conditions for which it is appropriate to use phenomenological models to describe experimental results from angiogenesis.

Chapter 3

Comparative analysis of continuum angiogenesis models

Much of the work in this chapter has previously appeared in the following journal article: Martinson WD, Ninomiya H, Byrne HM, Maini PK (2021) "Comparative analysis of continuum angiogenesis models" Journal of Mathematical Biology, online (DOI: 10.1007/s00285-021-01570-w).

3.1 Overview

In Chapter 2, we modified the snail-trail model, a classical continuum framework for angiogenesis, by reasoning that the system fails to account for leader tip cell (TC) migration in directions other than that of the leading solution front. We then compared its solutions to ensemble average distributions from an ABM developed by Pillay et al. (2017). This model, which we called the P–ABM, tracks the evolution of TCs that chemotactically migrate up gradients of a generic tumour angiogenic factor (TAF), while trailing stalk cells (ECs) proliferate along the path of the TCs. We demonstrated that a 1D version of the corrected snail-trail model can accurately estimate ensemble average discrete tip and stalk cell distributions. We did not determine, however, whether the modified phenomenological framework can also capture the dynamics of other continuum models for angiogenesis. We aim to resolve this question in the current chapter by comparing snail-trail model solutions to those from a particular coarse-grained PDE system derived in Pillay et al. (2017), which describes the mean-field dynamics of the P–ABM cell densities over time and space.

Although the full system of coarse-grained equations is formulated in 2D, it may be reduced to one spatial variable by column averaging the dependent variables, making a mean-field approximation, and non-dimensionalising the equations with appropriate length and time scales. The resulting 1D coarse-grained system, which we refer to as the P–PDE, can be written in Cartesian coordinates as follows: we denote by $N(x,t)$ and $E(x,t)$ the column averaged TC and EC densities, respectively, at location $x \in [0,L]$ and time $t \in [t_0,\infty)$, with $t_0 \geq 0$, and by $C(x,t)$ the column averaged TAF concentration. The system is closed by imposing no-flux boundary conditions on the TCs (boundary conditions are not required to close the EC rate equation because it is a first-order ODE with respect to t – see below). We assume that the source of TAF is located at $x = L > 0$, that there is no TAF at the original vessel where $x = 0$, and that the decay of TAF and its uptake by cells are negligible. Under these assumptions, the coarse-grained model is given by

$$\frac{\partial N}{\partial t} = \underbrace{\left(D\frac{\partial^2 N}{\partial x^2} - \chi\frac{\partial}{\partial x}\left(N\frac{\partial C}{\partial x}\right)\right)}_{\text{random motion + chemotaxis}}(1 - \underbrace{a_n N - a_e E}_{\text{anastomosis}}) + \underbrace{\lambda N C}_{\text{branching}} \underbrace{- \mu a_n N^2 - \mu a_e N E}_{\text{anastomosis}}, \tag{3.1}$$

$$\frac{\partial E}{\partial t} = \underbrace{\mu N}_{\text{production from tip cell movement}} \underbrace{+ a_n \mu N^2 + a_n N\left(D\frac{\partial^2 N}{\partial x^2} - \chi\frac{\partial}{\partial x}\left(N\frac{\partial C}{\partial x}\right)\right)}_{\text{production from tip-to-tip anastomosis}}, \tag{3.2}$$

$$\gamma\frac{\partial C}{\partial t} = \underbrace{\frac{\partial^2 C}{\partial x^2}}_{\text{diffusion of TAF}}, \tag{3.3}$$

$$\begin{gathered} D\frac{\partial N}{\partial x} - \chi N\frac{\partial C}{\partial x} = 0 \quad \text{for} \quad x \in \{0, L\} \\ C(0,t) = 0, \quad C(L,t) = 1, \\ N(x,t_0) = G(x), \quad E(x,t_0) = H(x), \end{gathered} \tag{3.4}$$

where $D > 0$ represents the TC random motility coefficient, $\chi > 0$ the TC chemotactic sensitivity to TAF, $\lambda > 0$ the TC branching rate, and $\mu > 0$ the EC proliferation rate. The non-negative functions $G(x)$ and $H(x)$ represent the initial TC and EC densities, respectively. The discrete-to-continuum derivation of the P–PDE demonstrates that the values of D, χ, μ, and λ may be written in terms of the P–ABM parameters (Pillay et al., 2017). In Eq. (3.1), the parameter $0 \leq a_e \leq 1$ is related to the prevalence of tip-to-sprout anastomosis, and in practice is fit to P–ABM data using a nonlinear least squares method. The parameter a_n, by contrast, is a model choice parameter: it either takes the value zero or one, depending on whether tip-to-tip anastomosis is included in the microscopic model. For the remainder of this chapter, we fix $a_n = 1$.

In Eq. (3.1), the expressions representing tip cell movement are multiplied by the nonlinear term $(1-a_nN-a_eE)$, which derives from the volume exclusion rules incorporated in the P–ABM and models the effect of cell elimination due to tip-to-tip and tip-to-sprout anastomosis. This term, however, may cause the P–PDE to become ill-posed if $a_nN+a_eE > 1$ (Pillay et al., 2017, 2018). This is because the tip cell rate equation becomes similar to a negative diffusion model – which yields ill-posed solutions that do not depend continuously on the initial data – when the latter condition holds. Thus, P–PDE solutions may become ill-posed when the TC and EC densities are sufficiently large. Although in practice multiple signalling cues act to prevent the emergence of such high cell density regions, for instance by creating a "salt-and-pepper" pattern of tip cells as they emerge from a pre-existing vessel (Duran et al., 2017; Blanco and Gerhardt, 2013; Bentley et al., 2008), these subcellular processes are outside the scope of the coarse-grained model. To ensure well-posedness of the PDE solutions, we opt to simulate the system from a time $t_0 > 0$ by which enough TCs have been sufficiently eliminated by anastomosis so that $(1 - a_nN - a_eE) \geq 0$. Guided by previous numerical implementations from Pillay et al. (2017), we choose $t_0 = 0.2$ and use column averaged P–ABM results as initial conditions (see §B.5 for details on the numerical methods).

Equation (3.3) describes the TAF dynamics. Byrne and Chaplain (1995) determined that for typical parameter values $0 < \gamma \ll 1$, hence a quasi-steady state approximation can be made for $C(x,t)$ by setting $\gamma = 0$. The resulting leading order solution, subject to the boundary conditions given in Eq. (3.4), is $C(x,t) \approx C(x) = \nu x$, where $\nu = L^{-1}$. We assume for the remainder of this chapter that the TAF concentration is given by this expression.

We compare solutions of the P–PDE to those of the 1D corrected snail-trail model (ST–PDE) from Chapter 2. If $N(x,t)$, $E(x,t)$, $C(x,t)$, D, χ, a_n, a_e, μ, and λ have the same physical interpretations as in the P–PDE, then the non-dimensional version of this

Snail-Trail Model (ST–PDE)
$N_t = DN_{xx} - \chi(NC_x)_x + \lambda NC - a_e\mu NE - a_n\mu N^2,$ $E_t = \kappa(x)\Big\vert DN_x - \chi NC_x\Big\vert = \frac{\mu}{\chi\nu}\Big\vert DN_x - \chi NC_x\Big\vert$
Pillay PDE Model (P–PDE)
$N_t = (DN_{xx} - \chi(NC_x)_x)(1 - a_nN - a_eE) + \lambda NC - a_e\mu NE - a_n\mu N^2,$ $E_t = \mu N + a_n\mu N^2 + a_nN(DN_{xx} - \chi(NC_x)_x)$

Table 3.1: The two continuum angiogenesis models investigated in this chapter, with subscript notation denoting partial derivatives with respect to x or t. Key: $N(x,t)$ (tip cell density), $E(x,t)$ (stalk cell density), $C(x,t)$ (TAF concentration, governed by Eq. (3.3), whose leading order solution is $C(x,t) = x/L = \nu x$ under the assumption that $0 < \gamma \ll 1$). The biological significance of each parameter is described in the text. All models are closed with the initial and boundary conditions given by Eq. (3.4). The function $\kappa(x)$ is defined in Eq. (3.7).

framework may be written in Cartesian coordinates as

$$\frac{\partial N}{\partial t} = \underbrace{D\frac{\partial^2 N}{\partial x^2} - \chi\frac{\partial}{\partial x}\Big(N\frac{\partial C}{\partial x}\Big)}_{\text{random motion + chemotaxis}} + \underbrace{\lambda NC}_{\text{branching}} - \underbrace{a_n\mu N^2 - a_e\mu NE,}_{\text{anastomosis}} \tag{3.5}$$

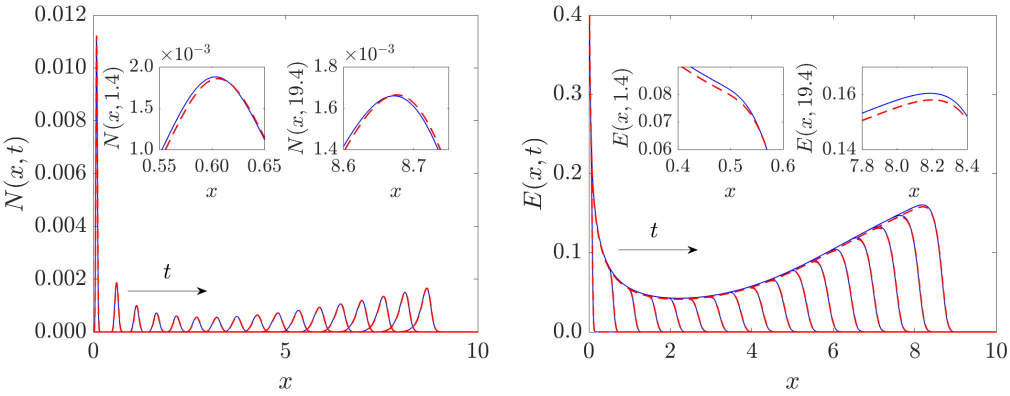
$$\frac{\partial E}{\partial t} = \underbrace{\kappa(x)\Big\vert D\frac{\partial N}{\partial x} - \chi N\frac{\partial C}{\partial x}\Big\vert}_{\text{production from net tip cell flux}}. \tag{3.6}$$

The initial conditions, boundary conditions, and TAF dynamics for this system are also given by Eq.s (3.3)-(3.4). We again make a quasi-steady state approximation for the TAF dynamics to conclude that the TAF concentration in the snail-trail model is given by $C(x,t) \approx C(x) = \nu x$. As we showed in the previous chapter, the formula for $\kappa(x)$ is given in terms of the continuum parameters by

$$\kappa(x) = \frac{\mu}{\chi\Big\vert\frac{\partial C}{\partial x}\Big\vert} = \frac{\mu}{\chi\nu}. \tag{3.7}$$

We note that, because of our quasi steady-state approximation for the TAF concentration, the value of $\kappa(x)$ is constant. Additionally, we remark that the equation for $\kappa(x)$ takes into account the factor of h^{-1} that appeared in the snail-trail models of the previous chapter (where h represents a typical cell length). We compare the equations for the ST–PDE and P–PDE in Table 3.1.

Although the P–PDE has a different stalk cell evolution equation than the ST–PDE and contains additional nonlinear terms, there are certain parameter regimes for which numerical



(a) TC density

(b) EC density

Figure 3.1: Numerical solutions of the ST–PDE and P–PDE (Table 3.1) showing (a) the TC density, $N(x,t)$, and (b) the EC density, $E(x,t)$, at times $t = 0.2$, 1.4, …, 19.4 within the domain $x \in [0,10]$. The insets in both figures show zoomed-in views of the results at (left inset) $t = 1.4$ and (right inset) $t = 19.4$. Key: P–PDE solutions (solid blue lines), ST–PDE solutions (dashed red lines). Parameter values: $D = 5 \times 10^{-4}$, $\chi = 0.425$, $\mu = 150$, $\lambda = 0.1$, $a_e = 0.05$, $a_n = 1$. The PDEs were initialised at $t = 0.2$ with column averaged data from the P–ABM. We refer to §B.5 for details on the numerical methods used to compute solutions.

solutions to the two continuum models are indistinguishable. We observe this behaviour in Figures 3.1(a)-(b) which, respectively, present TC and EC solutions for a random set of parameter values. At each time point shown, the largest pointwise difference between TC solutions is less than 5% of the maximum TC height. Similarly, the largest pointwise difference between EC solutions is less than 3% of the maximum EC density for each time point shown. Results such as these raise the question of whether solutions to the ST–PDE and P–PDE agree for larger regions of parameter space. Addressing this issue is important, as it would establish conditions under which relatively simple, phenomenologically-derived systems such as the ST–PDE may accurately capture the results of more nonlinear models that are rigorously derived from first principles, such as the P–PDE.

3.1.1 Chapter outline

In this chapter, we employ techniques from asymptotic analysis to study when the solutions to the ST–PDE and P–PDE will have relatively small pointwise differences. In §3.2, we derive the leading order ST–PDE and P–PDE dynamics for parameter regimes that correspond to chemotaxis-dominated TC movement and small branching rates. We find that, in such regions of parameter space, both PDE models reduce to the same system at leading order within the domain interior. Numerical simulation confirms that pointwise differences between both solutions are relatively small under the conditions outlined above. In §3.3, we construct asymptotic self-similar solutions to the leading order system that are valid for early times. Numerical simulation demonstrates that such analytic expressions are in good agreement with ST–PDE and P–PDE results. The approach taken in this chapter may be used to compare other nonlinear coarse-grained and/or phenomenological models for a given biophysical process, provided a perturbation parameter can be identified.

3.2 Both models have identical leading order dynamics

The procedure we use to reduce the ST–PDE and P–PDE to their dominant, leading order dynamics is presented below. For clarity, the details are shown only for the ST–PDE: the analysis for the P–PDE follows *mutatis mutandis*, and is presented in §B.1. We focus the analysis only on the domain interior, since most TCs have moved into this region by the initial time point $t_0 = 0.2$. Our focus on the domain interior also follows from the context in which both modelling frameworks apply: the P–PDE and ST–PDE each seek to capture TC and EC migration towards a TAF source, but not their subsequent infiltration into the TAF-producing region. This means that both models cease to be valid biological descriptions when the majority of TCs approach the right-hand boundary at $x = L$.

We make the following assumptions:

(A1) All parameter values in the ST–PDE are equal to those in the P–PDE.

(A2) The initial and boundary conditions for the ST–PDE and P–PDE are identical.

After substituting Eq. (3.7) into the ST–PDE stalk cell rate equation, we recast our dependent and independent variables to eliminate as many parameters as possible and simplify the analysis. We opt to use the transformations

$$u = \frac{a_n \mu}{\lambda} N, \qquad w = \frac{a_e \mu}{\lambda} E, \qquad c = C, \qquad \tau = \lambda t, \qquad X = x \sqrt{\frac{\lambda}{D}}, \tag{3.8}$$

so that the time scale of interest is λ^{-1} and the length scale is $\sqrt{D/\lambda}$. Since the TAF concentration is bounded between zero and one, and because the above transformations do not alter this range, we neglect to write the explicit functional form of $c(X, \tau)$. Instead, we denote it by the single variable c, with the understanding that it takes values in the range $[0, 1]$. These changes yield the following system of PDEs for $u(X, \tau)$ and $w(X, \tau)$:

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial X^2} - \frac{1}{\epsilon}\frac{\partial u}{\partial X} + u(c - u - w), \tag{3.9}$$

$$\frac{\partial w}{\partial \tau} = \beta \left| \epsilon \frac{\partial u}{\partial X} - u \right|. \tag{3.10}$$

We also transform the boundary and initial conditions using Eq. (3.8), to obtain

$$\epsilon \frac{\partial u}{\partial X} - u = 0 \text{ for } X \in \left\{0,\ L\sqrt{\frac{\lambda}{D}}\right\}, \tag{3.11}$$

$$u(X, \tau_0) = \frac{a_n \mu}{\lambda} G(X\sqrt{D/\lambda}) = g(X),\ w(X, \tau_0) = \frac{a_e \mu}{\lambda} H(X\sqrt{D/\lambda}) = h(X). \tag{3.12}$$

In Eq.s (3.9)-(3.12) we have introduced the dimensionless parameter groupings

$$\epsilon := \frac{\sqrt{D\lambda}}{\chi \nu}, \qquad \alpha := \frac{\lambda}{\mu}, \qquad \beta := \frac{a_e \mu}{a_n \lambda} = \frac{a_e}{a_n \alpha}. \tag{3.13}$$

Since we assumed that the TAF gradient is constant, the values of $\kappa(x)$ and, consequently, β are fixed in space and time. In more general scenarios where the TAF gradient varies in space, however, we note that β would depend on the value of X. Under the following assumption, we may further reduce the ST–PDE using standard perturbation methods (Verhulst, 2005; Bender and Orszag, 1999; Hinch, 1991):

(A3) Tip cell movement is dominated by chemotaxis and the branching rate is sufficiently small so that $0 < \epsilon \ll 1$ and $0 < \alpha = \epsilon^2 \Psi \ll 1$, with $\Psi \sim O(1)$.

Assumption (A3) is motivated by research that suggests chemotaxis dominates TC movement towards TAF sources (Duran et al., 2017; Betz et al., 2016; Carmeliet and Jain, 2011; Bowersox and Sorgente, 1982). We may write α in terms of ϵ because of Assumption (A1): the P–PDE model parameters directly relate to those of the P–ABM (Pillay et al., 2017; Simpson et al., 2009). In particular, if k and h represent P–ABM parameters related to the response of tip cells to TAF and the spatial step size of the lattice, respectively, then the continuum parameters D and χ are given as

$$D = \frac{\mu h^2}{4}, \qquad \chi = \mu k h^2. \tag{3.14}$$

Substituting these expressions into the formula for ϵ, we find that $\epsilon = \sqrt{\alpha}/(2kh\nu) = \sqrt{\alpha}\Psi^{-1/2}$, where we define $\Psi := (2kh\nu)^2$ as an $O(1)$ parameter.

For the remainder of this section, we consider the ratio a_e/α to be bounded in the limit as $\alpha \to 0$, so that $\beta \sim O(1)$. This is done to more clearly present the analysis: in §B.2 we derive the leading order ST–PDE and P–PDE dynamics for different limiting behaviours of β.

With $0 < \epsilon \ll 1$, we seek asymptotic solutions for $u(X,\tau)$ and $w(X,\tau)$ of the form

$$u(X,\tau) \sim u_0(X,\tau) + \epsilon u_1(X,\tau), \qquad w(X,\tau) \sim w_0(X,\tau) + \epsilon w_1(X,\tau). \tag{3.15}$$

We substitute these expressions into Eq. (3.10) and equate coefficients of $O(\epsilon^0)$. This leads to the following PDE for the dynamics of $w_0(X,\tau)$:

$$\frac{\partial w_0}{\partial \tau} = \beta |u_0|. \tag{3.16}$$

Ideally, we would identify the leading order TC dynamics with the same procedure. This naïve approach fails, however, because the Eq. (3.9) contains a term of $O(\epsilon^{-1})$: in particular, this yields a constant leading order TC solution that clearly does not match the observations of the ST–PDE and P–PDE TC solutions in Figure 3.1(a). To eliminate the $O(\epsilon^{-1})$ term, we change variables again by setting

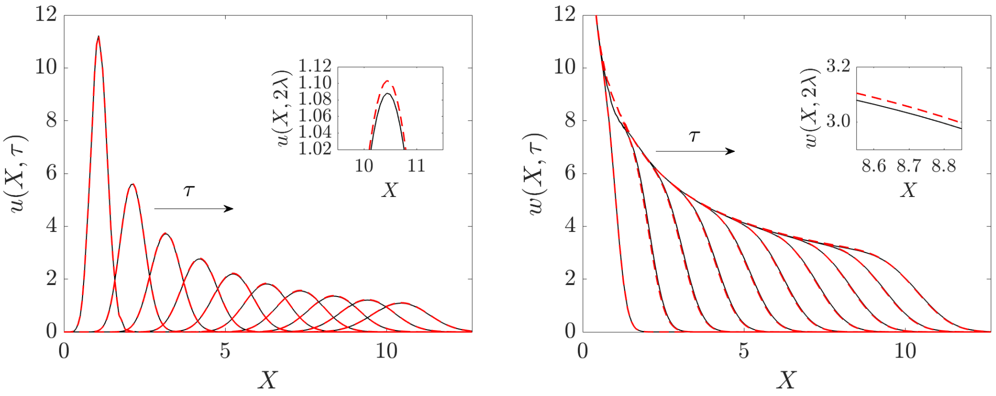
$$y = X - \frac{\tau + X_0}{\epsilon}, \qquad U = u, \qquad W = w, \qquad \widetilde{C} = c, \tag{3.17}$$

where X_0 is an arbitrary constant that ensures the location of the right-hand boundary, given in Eq. (3.19), is strictly positive over the time interval of interest. The above transformations do not change the range of the TAF concentration, which remains bounded between zero and one. We therefore use the variable $\widetilde{C}$ to denote the TAF distribution, with the understanding that the value of $\widetilde{C}$ can vary between zero and one within the domain. Substitution of the new variables into Eq. (3.9) yields

$$\frac{\partial U}{\partial \tau} = \frac{\partial^2 U}{\partial y^2} + U(\widetilde{C} - U - W), \tag{3.18}$$

with boundary conditions

$$\epsilon \frac{\partial U}{\partial y} - U = 0 \text{ for } y \in \left\{ -\frac{\tau + X_0}{\epsilon},\ \frac{\frac{\lambda}{\chi \nu} L - \tau - X_0}{\epsilon} \right\}. \tag{3.19}$$



(a) TC density

(b) EC density

Figure 3.2: Numerical solution of the full ST–PDE and leading order dynamics (Table 3.1 and Eq. (3.16), Eq. (3.21), and Eq. (3.22), respectively), which show how (a) the TC density, $u(X,\tau)$, and (b) the EC density, $w(X,\tau)$, evolve over times $\tau = 0.2\lambda$, 0.4λ, $\ldots$, 2λ, with $\lambda = 0.16$. The insets in both (a) and (b) show a zoomed-in view of the results at time $\tau = 2\lambda$. Key: ST–PDE solution (solid black lines), leading order solutions (dashed red lines). Parameter values: $D = 10^{-3}$, $\chi = 0.4$, $a_n = 1$, $a_e = 0.0391$, $\mu = 160$ (this corresponds to $\epsilon = 10^{-3/2}$, $\alpha = 10^{-3}$, and $\beta = 39.1$). The PDEs were simulated on $\tau \in [0.2\lambda, 2\lambda]$, $X \in [0, \sqrt{\frac{\lambda}{D}}]$. P–ABM solutions at $\tau = 0.2\lambda$, column averaged in the y-direction, were used as initial conditions. We refer to §B.5 for details on the numerical methods used to compute solutions.

We seek a regular perturbation series solution for U of the form

$$U(y,\tau) \sim U_0(y,\tau) + \epsilon U_1(y,\tau) + O(\epsilon^2). \tag{3.20}$$

At leading order, we recover the following PDE describing the dynamics of U_0:

$$\frac{\partial U_0}{\partial \tau} = \frac{\partial^2 U_0}{\partial y^2} + U_0(\widetilde{C} - U_0 - W_0), \tag{3.21}$$

where the dynamics of W_0 are described by Eq. (3.16) (after a suitable change of variables). The boundary conditions for the leading order solution, found in the limit as $\epsilon \to 0^+$, are

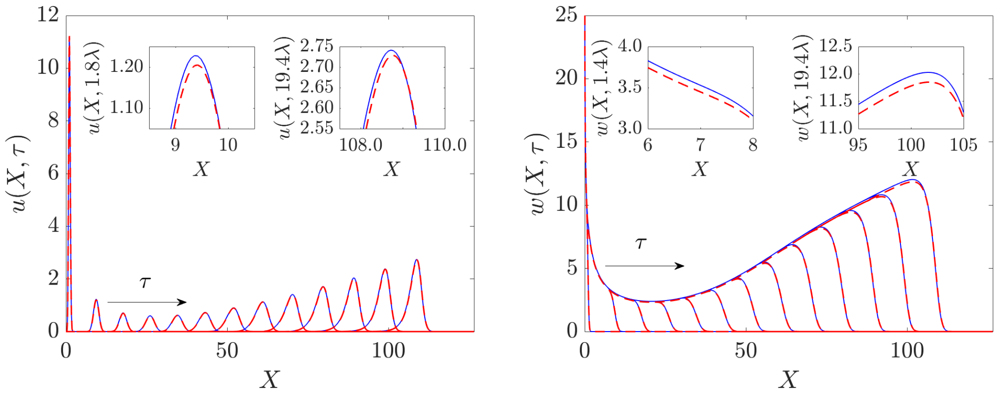
$$\lim_{y \to \pm\infty} U_0(y,\tau) = 0, \tag{3.22}$$

which are homogeneous Dirichlet boundary conditions. We show in §B.3 that solutions to Eq. (3.21) with these boundary conditions are non-negative, provided their initial conditions are also non-negative. Consequently, we may ignore the absolute value sign in Eq. (3.16).

Figure 3.2 presents the numerical solution to the leading order system in the independent variables X and τ, along with numerical results from the original ST–PDE. At each time point shown in Figure 3.2(a), the largest pointwise difference between the two TC solutions is less than 2% of the maximum tip cell density. Similarly, we observe in Figure 3.2(b) that the leading order EC solution is almost indistinguishable from that of the original ST–PDE for this set of parameter values. Indeed, at each time point shown the maximum pointwise difference between the two solutions is less than 2% of the maximum stalk cell density. These results support the earlier conjecture that the leading order dynamics of the ST–PDE within the domain interior are accurately described by Eq.s (3.16) and (3.21) when $0 < \epsilon \ll 1$ and $0 < \alpha \ll 1$.

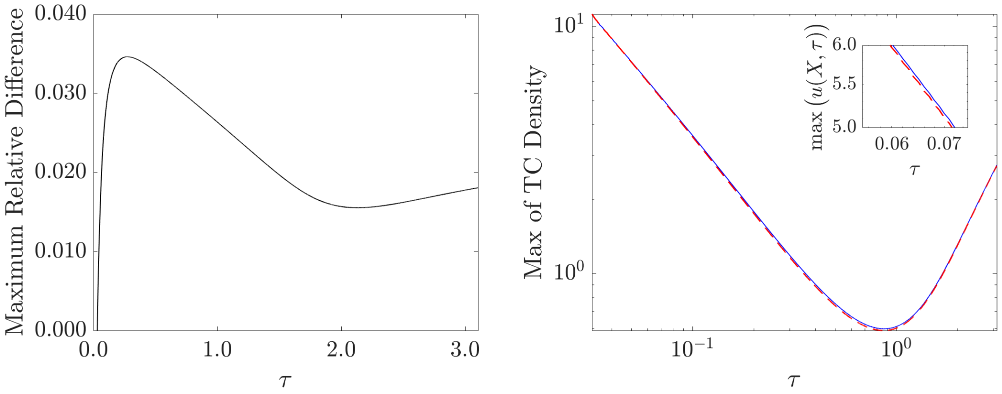
Crucially, we show in §B.1 that the procedure outlined above also reduces the P–PDE system to Eq.s (3.16) and (3.21). This equivalence between the leading order ST–PDE and P–PDE dynamics holds for other limiting behaviours of β (i.e., for cases in which β becomes infinitely large or small in the limit as $\epsilon \to 0$). Such derivations may be found in §B.2. In biological terms, this suggests that the ST–PDE and P–PDE are identical at leading order when chemotaxis dominates tip cell movement and the branching rate is sufficiently small; meanwhile the rate at which tip-to-sprout anastomosis occurs does not affect the correspondence between the two sets of solutions. Since the higher-order terms in the tip and stalk cell asymptotic series will be small when $0 < \epsilon \ll 1$ and $0 < \alpha \ll 1$, this also provides a reason for why solutions to the full ST–PDE and P–PDE models are indistinguishable within the domain interior for parameter regimes that correspond to such biological scenarios (such as those in Figure 3.1).

We have validated these conclusions with numerical results: Figure 3.3 presents numerical solutions of the full ST–PDE and P–PDE models, using a parameter regime for which $0 < \epsilon \ll 1$, $0 < \alpha \ll 1$, and $\beta \sim O(1)$ (values are listed in the figure caption; see §B.5 for details of the numerical methods). We observe in Figure 3.3(a) that at each time point, the TC solution profiles resemble bell curves. We also see that the TC density height increases gradually as the solution travels to the right. This occurs because the TC proliferation rate is proportional to the (linear) TAF concentration in both the ST–PDE and P–PDE: consequently, the TC density increases more quickly as it travels closer to the right hand boundary. It also appears in this graph that the ST–PDE and P–PDE TC solutions travel



(a) TC density

(b) EC density



(c) Maximum relative difference between TC solutions

(d) TC density height vs. time

Figure 3.3: Numerical solutions of the ST–PDE and P–PDE systems (Table 3.1) showing how (a) the tip cell density, $u(X,\tau)$, and (b) the stalk cell density, $w(X,\tau)$, evolve over times $\tau = 0.2\lambda, 1.8\lambda, ..., 19.4\lambda$, where $\lambda = 0.16$. The insets in (a) and (b) show zoomed-in views of the solutions at (left insets) $\tau = 1.8\lambda$ and (right insets) $\tau = 19.4\lambda$. The graph in (c) shows how the maximum difference between TC solutions of the ST–PDE and P–PDE, relative to their heights, varies with time. The log-log plot (d) shows how the TC height for both PDE systems evolves over time. The inset in that graph shows a zoomed-in view for small values of τ. Key: ST–PDE solution (dashed red line), P–PDE solution (solid blue line). Parameter values: $D = 10^{-3}$, $\chi = 0.4$, $a_e = 0.0391$, $a_n = 1$, $\mu = 160$ (this corresponds to $\epsilon = 10^{-3/2}$, $\alpha = 10^{-3}$, $\beta = 39.1$). We generated these graphs by simulating Eq.s (3.1)-(3.6) for $t \in [0.2, 20]$, $x \in [0, 10]$, and then transforming variables according to Eq. (3.8). P–ABM solutions at $t = 0.2$, column averaged in the y-direction, were used as initial conditions for both systems. We refer to §B.5 for details on the numerical methods used to compute solutions and the maximum relative difference between TC results.

with similar speeds. In fact, further investigation shows that the P–PDE speed is within 0.5% of the ST–PDE wave speed, and that both solutions gradually accelerate over time (see §B.5 for details on the numerical methods used to arrive at this conclusion).

The ST–PDE tip cell solutions appear to be indistinguishable from those of the P–PDE. We investigated this further by computing the maximum pointwise difference between the solutions and normalising that value by the largest TC height at each time point (see §B.5 for details). The resulting relative differences are presented in Figure 3.3(c). We observe that these values lie within the approximate range $[0.015, 0.035]$ (or 1.5%–3.5% of the maximum TC height) for the time points shown.

Figure 3.3(b) presents the ST–PDE and P–PDE EC densities for this parameter regime. We see that the P–PDE solution is greater than that of the ST–PDE, even though the solutions to both models otherwise appear to have similar profiles at each time point shown. Nevertheless, the size of the differences between the ST–PDE and P–PDE solution is small: we calculated the maximum relative pointwise difference between EC solutions using a similar method to the one described above (results not shown, see §B.5 for details). In this case, the relative difference between the two sets of stalk cell solutions lies within the range $[0.01, 0.02]$ (1%–2% of the maximum EC density) for the time interval considered here.

We observe in the TC density plots that the heights of both solutions initially decrease at a relatively rapid rate. We create a log-log plot of the TC density height against time in Figure 3.3(d) to further investigate this behaviour. Both graphs initially resemble straight lines that decrease with time. To identify the rate at which this occurs, we use least squares regression to compute a line of best fit to the ST–PDE and P–PDE curves for data points with $\tau < 0.4$, and find that a line with slope -1.02 approximates these data points with less than 2% error. This analysis suggests that the TC height is approximately proportional to τ^{-1} for early time points. We explain this observation in §3.3, where we derive self-similar solutions to the leading order ST–PDE and P–PDE dynamics that are initially proportional to τ^{-1}. For larger times, however, we see that the TC height increases – this is caused by the branching of new sprouts, which in turn leads to larger densities of TCs.

We next confirm that decreasing the chemotactic sensitivity of TCs (or increasing their random motion), increases the magnitude of differences between ST–PDE and P–PDE solutions, as both scenarios increase the value of ϵ. Specifically, we simulate two cases: (1) where the TC random motility coefficient, D, increases from 10^{-3} to 10^{-1}, and (2) where the chemotactic sensitivity of TCs, χ, decreases from 0.4 to 0.04. In both scenarios, ϵ increases by an order of magnitude (data not shown). In both situations, we can visually distinguish between the TC and EC solutions from both models. Quantification of the maximum relative differences of TC and EC densities confirms that there are larger discrepancies between the ST–PDE and P–PDE solutions when compared with the results in Figure 3.3. For the case in which $D = 10^{-1}$, we find that the maximum relative difference between TC and EC solutions lies within the range $[0.15, 0.3]$ (or 15%-30% of the maximum solution height) after an initial transient. When $\chi = 0.04$, the relative differences lie within the range of $[0.7, 0.85]$ (or 70%-85% of the maximum TC/EC height) for both cell densities.

We also verify that larger TC branching rates increase the difference between ST–PDE and P–PDE solutions by changing the value of λ from 10^{-3} to 0.0625 (data not shown; this corresponds to increasing both ϵ and α in the rescaled equations by an order of magnitude). In this case, the differences between ST–PDE and P–PDE TC/EC densities becomes noticeably large only over long time periods. The two sets of solutions are indistinguishable from each other at early time points, however. This apparent discrepancy from the analysis likely arises because the branching rate is not just proportional to λ but also to the TAF concentration. Since we have assumed the latter quantity is linear in space, the overall branching rate will remain small close to the left hand boundary, even for larger values of λ. It follows that, for early times, solutions near the left hand boundary are less likely to be affected by changes in λ, while over longer time periods differences between the ST–PDE and P–PDE will be larger because TCs will have invaded TAF-rich regions.

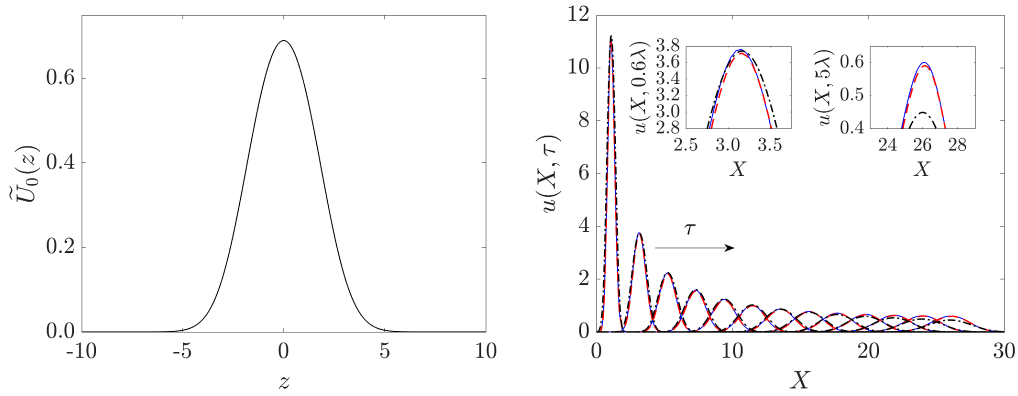
Finally, we confirmed that increasing the rate of tip-to-sprout anastomosis does not affect the degree of similarity between the ST–PDE and P–PDE solutions by changing the value of a_e to its maximum value of one (data not shown; this corresponds to increasing the parameter β to 1000). It is difficult to visually distinguish between the TC solutions generated by both models – in fact, further analysis reveals that the maximum difference between the TC densities is within 3% of their heights. The EC densities also appear indistinguishable: the relative difference between the ST–PDE and P–PDE solutions lies within the range $[0.015, 0.05]$ (or 1.5%–5% of the maximum EC height) for most time points (there is also an initial transient behaviour where the relative difference increases to a maximum of about 10%, but it quickly decays).

These numerical results largely corroborate the conclusions from the asymptotic analysis above. Namely, we find that for regimes in which one cannot necessarily reduce the ST–PDE and P–PDE models to the leading order system given by Eq.s (3.16) and (3.21) (which correspond to large values of $\epsilon = \sqrt{D\lambda}/(\chi\nu)$ and/or $\alpha = \lambda/\mu$), numerical solutions to the ST–PDE and P–PDE models are no longer indistinguishable. For parameter values in which the assumptions underpinning the analysis are valid, however, we find that the numerical solutions have relatively small pointwise differences. Such parameter regimes correspond to biological scenarios in which chemotaxis dominates tip cell movement and vessel branching is rare.

3.3 Self-similar leading order tip cell solutions can exist for early time periods

We now construct asymptotic solutions for the leading order system derived in §3.2. While the leading order model is a system of coupled PDEs, we show in §B.4 that, in cases where $0 < \tau \ll 1$, the two equations decouple. For such times, we can therefore restrict our attention to the leading order tip cell dynamics given by Eq. (3.21). Although the differential equation does not have a general closed-form solution to our knowledge, the numerical results from Figure 3.3(d) demonstrate that the height of the solution decreases at a rate approximately proportional to τ^{-1} for early times. This motivates us to search for self-similar solutions in cases where $0 < \tau \ll 1$.

We present the derivation of such solutions in §B.4 and list the main results of that analysis here. In the case where $\beta \sim O(1)$, the leading order TC density, $U_0(y, \tau)$, may be written in terms of the self-similar expression $\tau^{-1}\widetilde{U}_0(z)$, where $z = \tau^{-1/2}(y - \sigma(\tau)\tau)$ and $\sigma(\tau)$ is the wave speed at time τ. After substituting these expressions into Eq. (3.21) and



(a) Self-similar solution

(b) TC density

Figure 3.4: Numerical solution of the boundary value problem given by Eq. (3.23) in (a) the self-similar variable z, and (b) the independent variable X for $\tau = 0.2\lambda$, 0.6λ, ..., 5λ, where $\lambda = 0.16$. We have also plotted numerical ST–PDE and P–PDE results in (b), which are the same as those presented in Figure 3.3(a). To obtain the graph in (b), we rescaled the self-similar solution so that its maximum value at $\tau = 0.2\lambda$ is equal to that of the ST–PDE and P–PDE models; its height at future time values is then taken to be proportional to τ^{-1}. The insets in (b) show a zoomed-in view of the results at (left inset) $\tau = 0.6\lambda$ and (right inset) $\tau = 5\lambda$. Key for (b): P–PDE solutions (solid blue lines), ST–PDE solutions (dashed red lines), solution to Eq. (3.23) (dashed-dot black line). We refer to §B.5 for details on the numerical methods and to Table 3.2 for parameters of the self-similar solution.

noting that terms proportional to positive powers of τ become negligible as $\tau \to 0$, we find that in this limit the ODE describing the dynamics of $\widetilde{U}_0(z)$ is

$$\frac{d^2\widetilde{U}_0}{dz^2} + \frac{z}{2}\frac{d\widetilde{U}_0}{dz} + \widetilde{U}_0 = \widetilde{U}_0^2, \quad \widetilde{U}_0(\pm\infty) = 0, \tag{3.23}$$

which describes the TC density shape. The terms in Eq. (3.23) correspond to those from tip-to-tip anastomosis and TC migration. By contrast, the terms in Eq. (3.21) that arise from TC branching and tip-to-sprout anastomosis become proportional to τ following the self-similar transformation, hence they vanish as $\tau \to 0$. The self-similar solution therefore is valid in cases where branching and tip-to-sprout anastomosis are negligible. For longer time periods, however, those terms dominate the behaviour of solutions and it becomes impossible to reduce the dynamics for $\widetilde{U}_0$ to one independent variable. We conclude that the self-similar solution breaks down as time increases.

While it is not analytically known whether a non-trivial, closed-form solution exists for the boundary value problem given by Eq. (3.23) (i.e., which is not $\widetilde{U}_0(z) \equiv 0$), we can compute a non-trivial numerical solution. The resulting graph is presented in Figure 3.4(a). The solution resembles a bell curve, which is consistent with observations of numerical TC solutions from the previous section. We compare this function to the ST–PDE and P–PDE solutions by transforming the self-similar solution into a function that depends on the untransformed independent variable X: if we redefine the wave speed as $\phi(\tau) := \epsilon^{-1} + \sigma(\tau)$, then this transformation is given by

$$X = z\sqrt{\tau} + \phi(\tau)\tau + \frac{X_0}{\epsilon}, \tag{3.24}$$

which follows from the relationships listed above and in Eq. (3.17). The transformed self-similar solution is plotted in Figure 3.4(b) over several time points for a parameter regime in which $\epsilon = 10^{-3/2}$, $\alpha = 10^{-3}$, $\beta = 39.1$, and $\lambda = 0.16$; the numerical solutions to the ST–PDE and P–PDE are also plotted for comparison (see §B.5 for details on the numerical methods used to estimate $\phi(\tau)$ and X_0).

We observe in Figure 3.4(b) that the self-similar solution is indistinguishable from both the ST–PDE and P–PDE solutions for small time points (the largest pointwise difference between the self-similar solution and the other models is roughly 10% of the tip cell density height). At larger time points, however, the self-similar solution underestimates the ST–PDE and P–PDE results and the differences between the solutions are larger. This observation agrees with the earlier argument that the self-similar solution is invalid as time increases. The results shown in Figure 3.4 thus suggest that the leading order system does admit a similarity solution for early times, and that its profile is well described by Eq. (3.23).

At this point, we have only presented self-similar solutions in the case where $\beta \sim O(1)$. However, similarity solutions also exist for different limiting behaviours of β (this can be shown using a similar procedure to the one presented in §B.4). When $\beta \ll 1$, for instance,

ϵ	X_0	$\phi(\tau)$
$10^{-3/2}$	1.7×10^{-3}	32.45

Table 3.2: Dimensionless quantities used to generate the self-similar solutions presented in Figure 3.4(b). We refer to §B.5 for details on how such values are computed.

the self-similar solution is identical to the one described above. This makes sense, since terms representing tip-to-sprout anastomosis are negligible for this parameter regime – this result therefore is consistent with the conclusion that Eq. (3.23) describes the leading order solution when terms for sprout branching and tip-to-sprout anastomosis are negligible compared to those for tip-to-tip anastomosis.

When $\beta \gg 1$, however, the self-similar solution is different to the one described by Eq. (3.23). This results from the fact that, for this parameter regime, terms corresponding to tip-to-tip anastomosis are negligible in the leading order TC dynamics. A similar procedure to the one presented in §B.4 demonstrates that as $\tau \to 0$, the ODE for the self-similar solution is now given by

$$\frac{d^2\widetilde{U}_0}{dz^2} + \frac{z}{2}\frac{d\widetilde{U}_0}{dz} - A\widetilde{U}_0 = 0, \quad \widetilde{U}_0(\pm\infty) = 0, \tag{3.25}$$

where $z = \tau^{-1/2}(y - \sigma(\tau)\tau)$ is the same similarity variable as before and A is a constant such that $U_0(y,\tau) = \tau^A\widetilde{U}_0(z)$. The terms in this equation represent TC movement only: terms describing tip-to-sprout anastomosis and branching vanish as $\tau \to 0$. Thus, this similarity solution is applicable in cases where terms that represent branching, tip-to-tip anastomosis, and tip-to-sprout anastomosis are negligible compared to those for TC migration. We uniquely determine A by solving the ODE, satisfying the two boundary conditions, and enforcing the constraint that the solution must be non-negative. This sets $A = -1/2$, and the solution to Eq. (3.25) becomes

$$\widetilde{U}_0(z) = C_1 e^{-z^2/4}, \tag{3.26}$$

where C_1 is a non-negative constant. The solution is a Gaussian function and thus has a bell curve profile. For parameter regimes in which we expect this self-similar solution to be valid, we have confirmed that it has relatively small pointwise differences with ST–PDE and P–PDE numerical solutions for early time periods (data not shown).

3.4 Discussion and conclusion

In this chapter, we analysed two continuum frameworks for angiogenesis: the snail-trail and Pillay PDE models. The Pillay PDE is a nonlinear mean-field model that describes the coarse-grained behaviour of a discrete, rule-based ABM (Pillay et al., 2017). By contrast, the snail-trail model is a simpler system for angiogenesis that is constructed from phenomenological arguments (Byrne and Chaplain, 1995; Balding and McElwain, 1985; Orme and Chaplain, 1997; Flegg et al., 2020; Connor et al., 2015). Despite such differences, asymptotic solutions to the two models are identical at leading order within the (spatial) domain interior under parameter regimes in which chemotaxis dominates tip cell movement and sprout branching is rare. Hence in such biological situations we anticipate that solutions to the ST–PDE will be in good agreement with those of the P–PDE. We confirmed with numerical simulation that parameter regimes for which the asymptotic analysis is valid generate indistinguishable solutions, and that regimes for which it does not hold yield markedly different TC and EC densities.

In §3.3 we demonstrated that the leading order dynamics of the ST–PDE and P–PDE can admit self-similar solutions for early times. Pointwise differences in numerical results between the similarity solution, ST–PDE, and P–PDE solutions are relatively small. This indicates that higher order terms in the asymptotic expansions are indeed negligible under parameter regimes in which TC movement is dominated by chemotaxis and branching is rare.

There are several possible extensions to the work in this chapter. We may, for instance, apply a similar asymptotic analysis to study the leading order solutions of other continuum angiogenesis models. As a particular example, we cite a coarse-grained system from Bonilla et al. (2016), which was derived from a discrete model in which a coupled system of stochastic differential equations are used to update cell and vessel locations (Bonilla et al., 2014). Although the PDEs governing the mean-field system dynamics differ from those of the ST–PDE and P–PDE, their numerical solutions resemble those investigated in this chapter. The TC density of the Bonilla et al. PDE model, for instance, travels at constant speed and resembles a Gaussian function, while the EC density increases at a rate linearly proportional to that of the TC density (Bonilla et al., 2020, 2016). It would be interesting in future work to analyse the leading order solutions for other such continuum angiogenesis models, as this may provide insight into when different theoretical frameworks yield similar results and when simple phenomenological models accurately describe more complicated systems obtained from rigorous discrete-to-continuum derivations. This analysis would be contingent, however, on whether the other continuum models also admit a small perturbation parameter such as the one we identified when sprout branching is rare and chemotaxis dominates TC migration.

Additionally, the rules of the discrete model underpinning the P–PDE can greatly affect the terms within the coarse-grained framework – different rules for anastomosis and/or volume exclusion at the cell scale, for instance, have been found to generate nonlinear diffusion operators at the macroscale (Pillay et al., 2018). In the future, we may use perturbation methods to study the leading order behaviour of these other continuum models to determine how different microscopic rules may (or may not) affect the macroscopic evolution of tip and stalk cell densities over time and space.

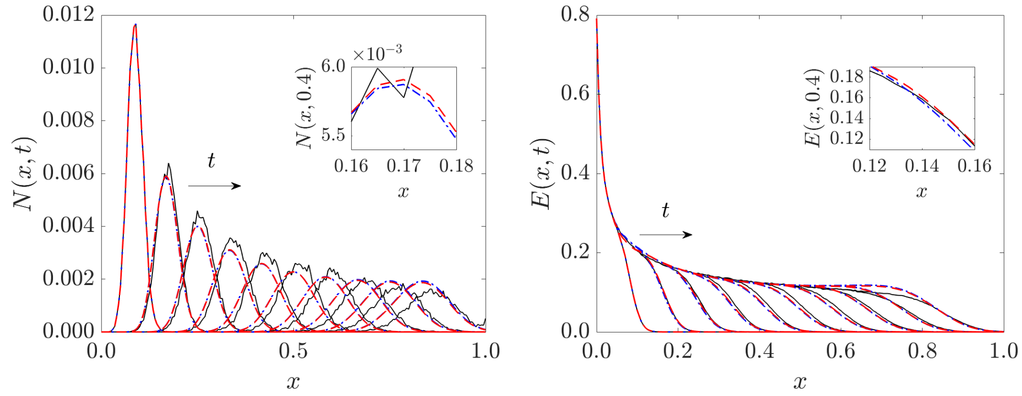
One limitation of this work is that we restricted our search of asymptotic solutions only to those within the domain interior, reasoning that they were most relevant to the numerical results given the initial conditions and underlying assumptions of each modelling framework. Boundary layers may exist, however: indeed, numerical results (not shown) reveal that stalk cell solutions of the ST–PDE and P–PDE are visibly different near the left-hand boundary. Thus, one extension of this work is to identify the leading order dynamics of P–PDE and ST–PDE solutions within these boundary layers and subsequently match them to the interior solutions from §3.2.

Another extension is to consider time-dependent TAF dynamics: although we used a quasi steady-state approximation in this chapter to simplify the asymptotic analysis, TAF dynamics may be more complicated in reality. TCs may create cell-induced gradients by consuming TAFs, for instance, which would generate time-dependent and spatially inhomogeneous TAF fields not considered in this chapter (Tweedy et al., 2020). In future work, we will investigate these more complicated dynamics and determine their impact on TC and EC trajectories. If we can identify a perturbation parameter, then we could use a similar approach to the one outlined in this chapter to analyse asymptotic solutions for both models. Alternatively, we could perform a large-scale survey of model outcomes from parameter space to identify how these more complicated dynamics affect tip cell and stalk cell solutions.

Although we identified parameter regimes for which solutions to two particular continuum models are indistinguishable, one open question we did not address is whether such solutions also capture the results of discrete models (which can more closely represent the underlying biology). Previous work (Pillay et al., 2017, 2018; Markham et al., 2013; Simpson and Baker, 2011) suggests that coarse-grained models poorly describe the results of their underlying discrete frameworks when the spatial correlations between agents cease to be negligible. In the context of angiogenesis, this may correspond to cases in which many cells are located in a given spatial region (in the P–ABM, for instance, the breakdown of the mean-field approximation originates from volume exclusion rules which are imposed for anastomosis).

Figure 3.5 presents column averaged P–ABM distributions along with ST–PDE and P–PDE solutions for a parameter regime in which the branching rate of TCs is relatively large (and the mean-field assumption is likely to be invalid). The two continuum models exhibit relatively small pointwise differences from each other (the greatest discrepancy between the TC densities at each time point, for example, is less than 4% of their heights). Discrepancies between the P–ABM and continuum models are larger, however, as the ST–PDE and P–PDE both overestimate the TC density along the trailing edge of the wave front. Such results support the conclusion that continuum models do not completely describe discrete ABM results, especially for situations in which the mean-field approximation may be invalid.

Further inspection of Figure 3.5, however, shows that the continuum models do not completely fail at capturing some aspects of the discrete solutions, even if their pointwise differences are relatively large. The shape of continuum TC solutions, for example, resembles that of the discrete results even though their heights are different. In addition, we observe that the speed at which the leading edge of TC and EC solutions travel in the P–ABM, P–PDE, and ST–PDE are almost identical (they are within 3.5% of each other). We conclude that the continuum models can accurately describe some aspects of the underlying discrete solutions, such as their invasion speeds and shapes, even though they may yield incorrect



(a) TC density (b) EC density

Figure 3.5: Numerical solutions of (a) the TC density, $N(x,t)$, and (b) the EC density, $E(x,t)$, from the ST–PDE, P–PDE, and P–ABM at times $t = 0.2, 0.4, \ldots, 2$. One thousand realisations of the P–ABM, with branching parameter $P_p = 5 \times 10^{-2}$, were column averaged in the y-direction to generate the discrete model results. Column averaged P–ABM solutions at $t = 0.2$ were used to initialise the two PDE models, which were simulated on the interval $x \in [0,1]$, $t \in [0.2, 2]$. The insets in both panels show a zoomed-in view of the solutions at time $t = 0.4$ (we remark that, in the inset of (a), the ABM solution exceeds the panel range for larger values of x). Key: column averaged P–ABM distributions (solid black lines), ST–PDE solution (dashed red lines), P–PDE solution (dashed-dot blue lines). PDE parameter values: $D = 10^{-3}$, $\chi = 0.4$, $a_n = 1$, $\mu = 160$. The PDE parameters λ and a_e minimise the squared difference between the PDE solutions and the column averaged ABM data for the time points shown (see §B.5 for details). For the ST–PDE, we have $\lambda = 0.99$ (95% CI: [0.985, 0.992]), $a_e = 0.0363$ (95% CI: [0.0362, 0.0364]). For the P–PDE, we have $\lambda = 1.115$ (95% CI: [1.111, 1.119]), $a_e = 0.0424$ (95% CI: [0.0423, 0.0425]).

values for the cell densities.

This raises the question of how one should quantify discrepancies between discrete and continuum results. For example, “good agreement” between discrete and continuum solutions could be characterised by sufficiently small pointwise differences, or by small differences in their invasion speeds. Clarification on this issue may provide insight into metrics (such as the average cell displacement or the average speed of the travelling sprouts) that could indicate both the extent of angiogenic progression and when it would be appropriate to apply simple continuum models to study and analyse *in silico* (and eventually *in vivo*) angiogenesis data.

Chapter 4

Neural crest cell interactions with the extracellular matrix dominate their collective migration

4.1 Overview

In Chapters 2 and 3, we investigated questions related to the mathematical modelling of leader-follower migration by analytically comparing different theoretical frameworks for angiogenesis. While such work demonstrates a degree of equivalence between such abstract models and suggests metrics to infer when they will accurately characterise cell migration, it does not address how specific biological mechanisms contribute to the macroscopic behaviour of leaders and followers. These signals – which include chemotaxis, haptotaxis, contact guidance, and cell-cell repulsion (see Figure 4.1) – all act to guide the collective migration of cells. Previous modelling studies demonstrate that not all of these signals are required for cell invasion: as we noted in Chapter 1, both cell-induced gradient models (which rely mainly on chemotaxis) and CIL-based models (which combine cell-cell interactions and chemotaxis) can produce collectively migrating groups of NCCs (Giniunaite et al., 2019a; Szabó et al., 2019). Combinations of haptotaxis, cell-cell repulsion, and chemotaxis can drive sprout elongation in angiogenesis as well (Stepanova et al., 2021; Flegg et al., 2020). In the context of neural crest cell migration, however, it is less well known how cell interactions with the ECM can affect their collective migration.

Indeed, recent observations from our experimental collaborators (Paul Kulesa and Mary

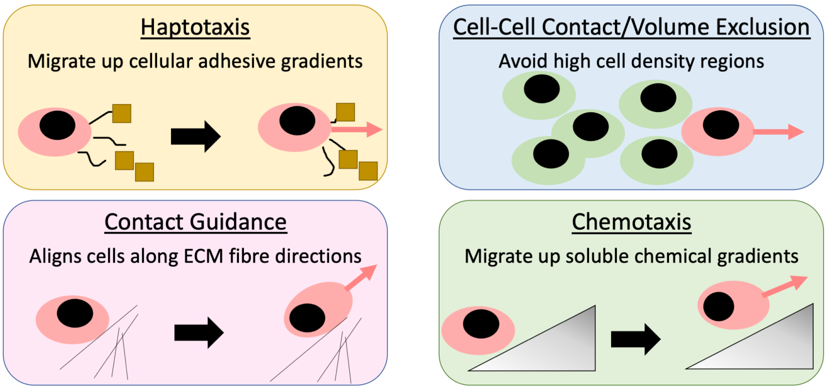


Figure 4.1: A cartoon depicting four mechanisms thought to drive neural crest cell migration. Haptotaxis (top left box) causes cells (pink ellipses with black circles) to migrate up increasing adhesive gradients of insoluble glycoproteins (yellow squares). Contact guidance (bottom left box), on the other hand, aligns cells along the direction of ECM fibres. Previous models of NCC migration have also considered the role of cell-cell repulsion (top right box, also sometimes referred to as volume exclusion in the literature), which guides NCCs toward regions with lower cell densities, and chemotaxis (bottom right panel), which attracts cells up increasing gradients of a soluble chemoattractant (grey triangle). The agent-based model introduced in this chapter does not consider the role of chemotaxis in guiding NCC migration, but does include the other three mechanisms.

McKinney, Stowers Institute for Medical Research, Kansas City, MO, USA) suggest a strong correlation between the deposition/remodelling of the fibronectin (FN) matrix and successful NCC migration. Their data indicate that, prior to NCC migration, the FN matrix within the mesodermal tissue consists of discrete, punctate structures – in other words, it does not resemble the cross-linked, fibrillar FN matrix of mature organisms (Walma and Yamada, 2020; Frantz et al., 2010). As leader NCCs move through the ECM, however, the matrix acquires a fibrous structure along cell paths, with more FN molecules concentrated near the tips of NCC filopodia within the stream front. We refer to Figure 4.2 for an example of the *in vivo* FN matrix during migration, and how its structure differs based on the presence or absence of NCCs. In combination with the fact that FN fibril assembly is mediated by integrin proteins, which are located on cell membranes (Kechagia et al., 2019; Schwarzbauer and DeSimone, 2011; Singh et al., 2010), these observations indicate that NCCs may contribute to and remodel the FN matrix by secreting the glycoprotein and assembling it into fibres.

Such experimental evidence motivates us to construct a new agent-based model (ABM)

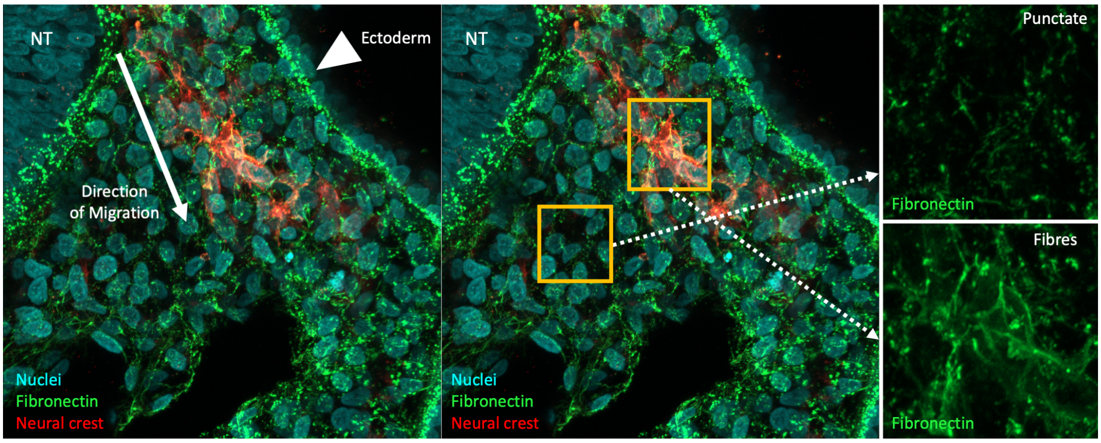


Figure 4.2: Neural crest cells secrete and assemble fibronectin along the migratory pathway. During migration in the vertebrate head (left panel), HNK1-labeled (red) neural crest cells (NCCs) exit the dorsal neural tube and migrate just underneath the ectoderm and through DAPI-labeled mesoderm (blue) and extracellular matrix glycoproteins such as fibronectin (green) to peripheral targets. In contrast to a mature and structured microenvironment, the embryonic pattern of fibronectin is largely punctate and unassembled in the mesoderm in advance of the migrating NCCs (bottom yellow inset in middle panel and top right panel). After lead NCCs pass through (upper yellow inset in middle panel and bottom right panel), however, the fibronectin matrix exhibits a more fibrous, connected structure, strongly suggesting that migrating NCCs secrete and assemble fibronectin that follower cells utilise for guidance. Images of these transverse tissue sections through the chick head on day 2 of development were collected using confocal microscopy, courtesy of Mary McKinney and Paul Kulesa (unpublished, with permission).

for chick cranial NCC migration that explicitly incorporates ECM-mediated signals. The modelling framework specifically considers a cell-reinforced migratory cue, in which a leader cell phenotype remodels an initially punctate and immature FN matrix to create a signal for trailing cells that exhibit a more inert follower phenotype. To evaluate whether such signals between cells and the ECM can lead to collective migration, we apply tools from global sensitivity analysis (see §1.3) to identify which of the migratory mechanisms in the ABM – haptotaxis, contact guidance, and cell-cell repulsion – most dominate the behaviour of NCCs.

We note that, because we are most interested in evaluating whether cell-reinforced signals can promote collective invasion, the ABM ignores other processes that occur during NCC migration. We therefore ignore the effects of cell proliferation, phenotype switching, domain growth, and chemotaxis in the modelling framework. Additionally, we neglect the interactions that NCCs have with mesoderm cells when they migrate. Mesoderm cells, which are located along the NCC migratory pathway prior to migration, are responsible for generating the initial *in vivo* FN matrix. Previous research has found that NCCs can migrate past mesoderm cells, and that the *in vivo* dynamics between the two cell types appeared to be coupled: NCCs, for example, can affect the proliferation of mesoderm cells (McKinney et al., 2020). Since the focus of this chapter is to evaluate whether FN remodelling can produce a strong enough signal for trailing NCCs to enable collective migration, we ignore the effect of mesoderm cells in the ABM. We note, however, that the ABM may be extended in the future to consider their influence.

4.1.1 Chapter outline

The rest of this chapter is organised as follows: in §4.2, we describe the NCC ABM and its underlying rules. After presenting example ABM simulations in §4.3, we detail several summary statistics for collective migration in §4.4. Finally, in §4.5 we apply Latin hypercube sampling (LHS) and extended Fourier amplitude sensitivity testing (eFAST) to determine how the summary statistics change in response to ABM parameter perturbations. In doing so, we infer which mechanisms most dominate collective migration and identify how they affect the macroscopic behaviour of NCCs.

4.2 Development of an ABM for NCC migration

The ABM is a 2D discrete off-lattice model in which individuals are assumed to be spherically symmetric with a user-specified constant cell body radius R_{cell}, as shown in Figure 4.3(a). Such frameworks are known as "overlapping spheres" models in the literature because the spheres can intersect, with the area of overlap between two cells being proportional to the repulsive force between them. The cells are treated as point particles and their movement is determined by balancing the forces that act on their centres of mass (Mathias et al., 2020; Osborne et al., 2017). We consider two cell phenotypes in the ABM – leaders and followers – each of which move through an initially punctate FN matrix. For each cell i, we update its centre position ($\mathbf{x}_i$) through force-based dynamics. These forces include frictional cues (which are assumed to be proportional to the cell velocity) and those from cell-ECM mechanisms such as haptotaxis and contact guidance ($\mathbf{F}_{\text{ECM}}$). To separate cells in the ABM, we include another force arising from cell-cell repulsion ($\mathbf{F}_{\text{VE}}$), which we call "volume exclusion" in the context of this ABM. We note, however, that this phrase does not imply that cells cannot overlap in the ABM; indeed, if the NCC-ECM force is sufficiently large, then two cell centres may become arbitrarily close. We instead use the term "volume exclusion" to describe a force that arises when a cell centre falls within the range of filopodial protrusions originating from a neighbouring cell; the resulting repulsive force is oriented in a direction that is likely to be (but not always) biased away from the vector connecting the cell centres. This is consistent with terminology introduced by Schumacher et al. (2016), who used it to distinguish such modes of cell-cell repulsion from contact inhibition of locomotion (see §4.2.2 for more details). We remark that the cell-cell force can potentially be extended to account for interactions arising from cell-cell adhesion, although we ignore this mechanism for the remainder of this thesis. Assuming that the mass, m, and acceleration of each cell are negligible compared to its velocity and friction coefficient, η, trajectories can be simplified to an overdamped Newton's law:

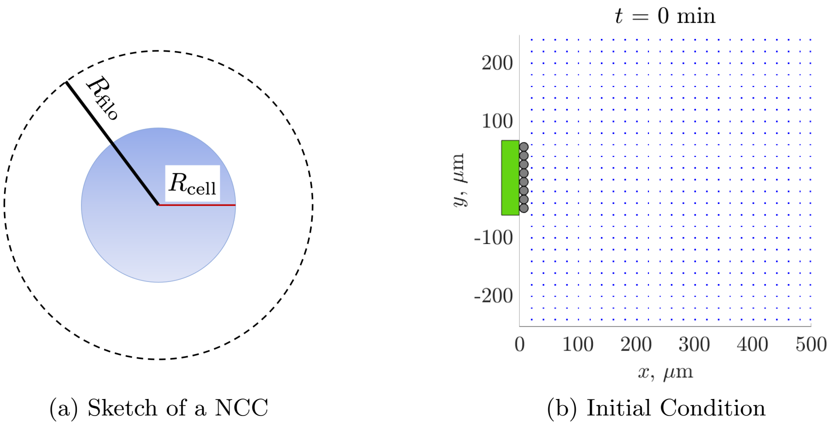
$$m\frac{\mathrm{d}^2\mathbf{x}_i}{\mathrm{d}t^2} = -\eta\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} + \mathbf{F}_{\text{ECM}} + \mathbf{F}_{\text{VE}} \qquad \Rightarrow \qquad \mathbf{v}_i := \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} \approx \mathbf{F}_{\text{ECM}} + \mathbf{F}_{\text{VE}}. \tag{4.1}$$

The assumption that inertial effects are negligible compared to frictional forces is reasonable within the context of cell biology, as it is frequently noted that the *in vivo* environments within which cells are situated are highly viscous (Van Liedekerke et al., 2015; Odell et al., 1981). We also remark that we have abused notation in the second expression of Eq. (4.1) by subsuming a factor of η^{-1} (where η is proportional to the frictional force) into the two "force" terms. We implement the ABM using the `C++` library PhysiCell (version 1.7.1) because this software specialises in efficiently running overlapping spheres models and has fewer dependencies compared to other libraries such as Chaste (Ghaffarizadeh et al., 2018; Pitt-Francis et al., 2009). In practice, we solve Eq. (4.1) with a second-order Adams–Bashforth method to determine cell trajectories. Aside from the initial and boundary conditions for the model (which we describe below), most of this section is dedicated to explaining how we determine the forces $\mathbf{F}_{\text{ECM}}$ and $\mathbf{F}_{\text{VE}}$ in Eq. (4.1).

We idealise NCC migration as a planar 2D process because of previous observations in the chick head that suggest the 3D environment in which cells migrate is shallow, i.e. approximately three to four cell diameters in depth (Kulesa and Fraser, 1998). Unlike existing frameworks for NCC migration, which consider domains that confine cells to a migratory corridor of up to 150 μm in height, cells in this ABM invade an open 500 μm × 500 μm domain where the left-hand boundary represents the neural tube. Since the FN matrix prior to NCC entry does not appear to favour a particular migration direction (Ostrovsky et al., 1983; Duband and Thiery, 1982; Newgreen and Thiery, 1980), we model the initial FN matrix as a punctate isotropic square lattice with equal spacing λ_{FN} between puncta, as in Figure 4.3(b). We also assume that FN does not diffuse or decay as the timescale of NCC invasion is shorter than that of these processes (McKinney et al., 2020; Kulesa et al., 2004). Furthermore, experimental research indicates that at most 10% of NCCs undergo mitosis during the first 12 hours of migration (Ridenour et al., 2014), which suggests cell division does not greatly contribute to invasion. We therefore ignore cell proliferation in the ABM.

Each NCC in the ABM has a fixed cell radius, R_{cell}, that determines its surface area and volume. As NCCs can extend filopodia from their cell bodies, we also associate a second radius $R_{\text{filo}} > R_{\text{cell}}$ to agents that denotes the maximum length that a cell can extend a filopodium. This approach models cells with variable filopodia lengths, as NCCs can sense any FN between R_{filo} and R_{cell}. We remark, however, that puncta within a distance R_{cell} are considered to be "covered" by the NCC and not able to be sensed. This assumption was introduced after observing individual ABM simulations, as we found that cells otherwise tended to "stick" to FN puncta and not migrate further. Hence in practice the assumption favours longer distance NCC migration because cells become more likely to travel toward distant FN puncta.

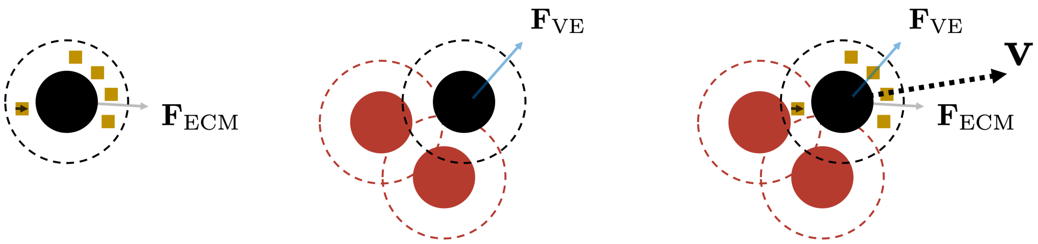
We separate the NCC population into two phenotypes: (1) leader cells, which can secrete new FN puncta, and (2) follower cells, which cannot secrete new FN puncta. We remark that this constitutes the only distinguishing feature between leaders and followers



(a) Sketch of a NCC

(b) Initial Condition

Figure 4.3: (a) An NCC agent (blue circle) with its maximum interaction area (dashed black circle), and (b) the ABM initial condition. The cell body radius, R_{cell}, is indicated as a red line in (a), while the maximum filopodial distance, R_{filo}, is denoted by the solid black line. Agents can only sense neighbouring cells and/or fibronectin puncta within the dashed circle. The parameters used to generate the configuration in (b) are $\lambda_{\text{FN}} = 20$ μm, $l_{\text{entr}} = 120\mu$m, and $R_{\text{cell}} = 7.5$ μm. Leader NCCs are denoted by black circles, while fibronectin puncta are represented as blue squares. The neural tube entrance strip, along which new cells enter the domain, is denoted by a green rectangle. We note that new follower NCCs will be placed at the locations currently marked by the black circles, provided they do not overlap with another cell, when they enter the domain.



(a) NCC, FN puncta, and $\mathbf{F}_{\text{ECM}}$

(b) NCCs and $\mathbf{F}_{\text{VE}}$

(c) NCCs, FN puncta, and $\mathbf{v}$

Figure 4.4: Cartoon of a NCC (black circle), fibronectin puncta (yellow squares), and neighbouring cells (red circles) showing (a) the force from NCC-ECM interactions, $\mathbf{F}_{\text{ECM}}$ (grey arrow), (b) the force from cell-cell repulsion/volume exclusion, $\mathbf{F}_{\text{VE}}$ (blue arrow), and (c) the resultant cell velocity $\mathbf{v}$ computed from Eq. (4.1) (black dotted arrow). In each panel, the dashed circles represent the maximum length of cell filopodia.

in the ABM. Leader cells secrete new FN puncta at their cell centres, according to time intervals that are drawn from a memoryless exponential distribution with user-defined mean T_{ave}, provided they also sense another FN puncta. We assume FN secretion depends on FN sensing because of *in vivo* observations that only leading NCCs appear to deposit FN puncta and that secretion occurs when cell protrusions are in contact with FN molecules in the microenvironment (Paul Kulesa, *pers.comm.*).

We initialise cells by placing a column of non-intersecting leaders a distance R_{cell} away from a subset of the left-hand boundary (with user-defined distance l_{entr} centred along the line $y = 0$), as shown in Figure 4.3(b). This "entrance strip" is intended to model the emergence of NCCs from a discrete subset of the neural tube. To guarantee that each NCC senses fibronectin at the beginning of simulations, the first column of FN puncta is placed a distance $R_{\text{cell}} + 0.5R_{\text{filo}}$ to the right of the leader cell centres. At every subsequent time step, new follower cells are introduced in the domain at the same positions where the leader cells were initialised. However, this action is aborted if the new agent were to intersect with an existing migrating NCC. This rule incorporates volume exclusion into the boundary condition and, in practice, prevents the continuous addition of cells. Cells already within the domain experience reflective conditions at all boundaries, which in practice is implemented as if they "bounce back" from the boundary. This ensures that no cell can exit the domain, and is a discrete analogue of a no-flux boundary condition.

As Eq. (4.1) indicates, the two forces that determine the cell velocity, $\mathbf{v}$, arise from cell-cell repulsion/volume exclusion ($\mathbf{F}_{\text{VE}}$) or NCC interactions with the FN matrix ($\mathbf{F}_{\text{ECM}}$); these are depicted as cartoons in Figure 4.4. In the following subsections, we describe in more detail how we compute these forces using the current cell and FN distributions. The process is summarised in Figure 4.5.

4.2.1 Computing the force from NCC-ECM interactions

In the ABM, the direction of the force from cell-ECM interactions, $\hat{\mathbf{F}}_{\text{ECM}}$, is determined by a linear combination of two vectors. The first, which corresponds to haptotaxis, biases NCC migration up increasing gradients of FN puncta and is denoted by $\hat{\mathbf{u}}_{\text{hap}}$. It models

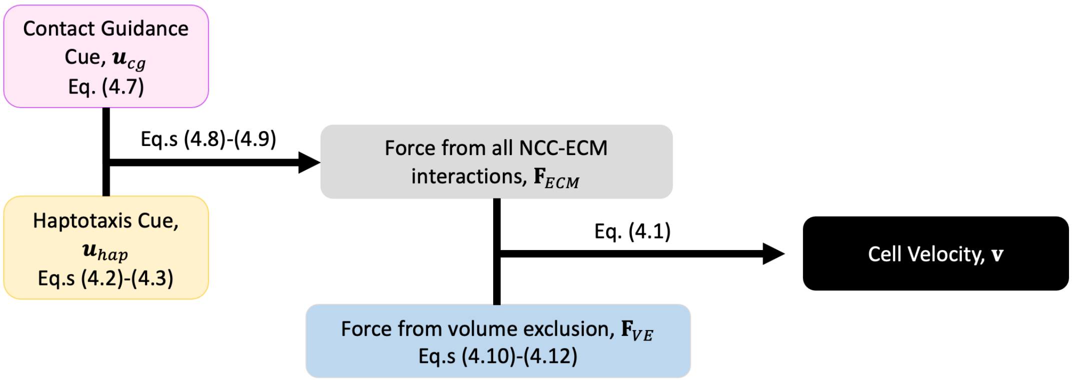


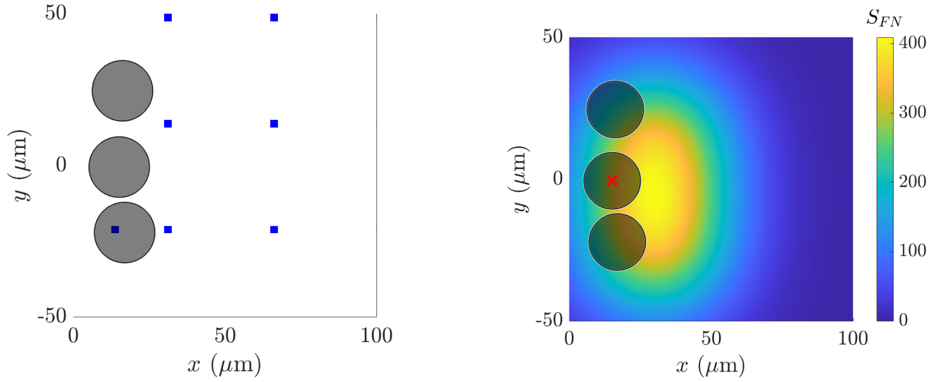
Figure 4.5: Flowchart depicting the relationships between the contact guidance cue ($\mathbf{u}_{\mathrm{cg}}$), haptotaxis cue ($\mathbf{u}_{\mathrm{hap}}$), NCC-ECM force ($\mathbf{F}_{\mathrm{ECM}}$), cell-cell repulsion/volume exclusion force ($\mathbf{F}_{\mathrm{VE}}$), and cell velocity ($\mathbf{v}$). In this flowchart, arrowheads represent a combination of two cues, which are calculated according to the listed equations. We refer to the text for a discussion of how each cue is determined from the current NCC/FN configurations.

the observation that NCCs preferentially migrate towards FN-dense environments (Rovasio et al., 1983). The second component, $\hat{\mathbf{u}}_{\mathrm{cg}}$, corresponds to contact guidance. We now discuss in more detail how we compute these individual signals.

Determining a migrational cue from haptotaxis

We specify the haptotaxis cue by sampling from a continuous probability distribution that is biased towards increasing densities of FN. This sampling approach allows us to model *in vivo* events in which NCCs sense multiple FN puncta within a given time interval; furthermore, it also allows us to naturally incorporate noise in how a given cell senses these signals. Since the FN distribution is discrete in the model, however, we must relate individual FN puncta to a continuous density; we do this via a similar procedure to one given in Binny et al. (2015, 2016) and Browning et al. (2018). In brief, we model FN puncta as the sum of continuous Gaussian functions, then use the gradient of the resulting function to parametrise a probability distribution for the haptotaxis cue. We describe the method in more detail below.

At time t, we record the locations, $\mathbf{q}_j$, of all FN puncta located within an annulus of inner radius R_{cell} and outer radius of R_{filo} from the cell centre $\mathbf{r}$ that are not covered by other NCCs. We then construct a “fibronectin surface”, $S_{\mathrm{FN}}(\mathbf{r})$, which is proportional to



(a) NCC and FN Puncta Locations (b) Colour plot of $S_{\text{FN}}(\mathbf{r})$

Figure 4.6: (a) Given a configuration of NCCs (black shaded circles) and FN puncta (blue squares), we compute (b) the FN surface $S_{\text{FN}}(\mathbf{r})$ using Eq. (4.2) for the middle cell (labelled with a red "X"). We note that FN puncta that are covered by cells, such as the lower left FN punctum in (a), are not included in the calculation of $S_{\text{FN}}(\mathbf{r})$. The parameter values used to calculate $S_{\text{FN}}(\mathbf{r})$ were $R_{\text{filo}} = 30$ μm, $R_{\text{cell}} = 10$ μm, and $\gamma_{\text{FN}} = 300$ (arbitrary units).

the FN density, by summing Gaussian functions at each FN location using the formula

$$S_{\text{FN}}(\mathbf{r}) = \sum_{j \in N_{\text{FN}}(t)} \gamma_{\text{FN}} \times \exp\left(-\frac{\|\mathbf{r}-\mathbf{q}_j\|^2}{2(R_{\text{filo}}-R_{\text{cell}})^2}\right), \tag{4.2}$$

where $N_{\text{FN}}(t)$ is the total number of uncovered FN puncta sensed by the cell at time t, γ_{FN} is a user-defined parameter that specifies the height of the Gaussian function, and $\|\cdot\|$ denotes the Euclidean norm. We note that the width of the Gaussian function in Eq. (4.2) is proportional to the annular radius, $R_{\text{filo}} - R_{\text{cell}}$. This ensures that the function $S_{\text{FN}}(\mathbf{r})$ is continuous and that its gradient is not negligible at the cell centre $\mathbf{r}$, so that cells are able to sense puncta even when they are located at distances close to R_{filo}. We remark that all FN puncta are otherwise considered to be identical in Eq. (4.2). For a visual representation of $S_{\text{FN}}(\mathbf{r})$ for a given configuration of NCCs and FN puncta, we refer to Figure 4.6.

We use $S_{\text{FN}}(\mathbf{r})$ to parameterise a von Mises distribution from which we sample $\hat{\mathbf{u}}_{\text{hap}}$ (Johnston and Painter, 2019; Sra, 2012). The von Mises distribution resembles a "wrapped around" normal distribution and models the probability of selecting a direction $\theta \in [-\pi, \pi)$ with mean μ and "concentration parameter" κ, the latter of which is analogous to the multiplicative inverse of the standard deviation (Mardia and Jupp, 2000). We set μ and κ

to be the direction and magnitude, respectively, of the FN surface gradient at $\mathbf{r}$. Hence

$$\mu = \operatorname{Arg}\Big(\nabla S_{\mathrm{FN}}(\mathbf{r})\Big), \qquad \kappa = \|\nabla S_{\mathrm{FN}}(\mathbf{r})\|, \qquad p(\theta|\mu,\kappa) = Ce^{\kappa\cos(\theta-\mu)}, \tag{4.3}$$

where $p(\cdot)$ is the probability density function (pdf) of the von Mises distribution and C is a constant. It is periodic with period 2π and is equivalent to a uniform distribution on $[-\pi,\pi)$ when $\kappa = 0$. It also behaves like a Dirac δ-distribution as $\kappa \to \infty$, in the sense that its pdf height increases without bound and its width becomes infinitesimally small in this limit. Thus, the haptotaxis cue is likely to be aligned with the gradient of the FN density when it is steep; if the gradient is small in magnitude, however, then the cell chooses a directional cue uniformly at random.

Once the direction $\hat{\theta}$ is sampled, the unit directional cue $\hat{\mathbf{u}}_{\mathrm{hap}}$ is given by

$$\hat{\mathbf{u}}_{\mathrm{hap}} = \begin{pmatrix}\cos(\hat{\theta})\\ \sin(\hat{\theta})\end{pmatrix}. \tag{4.4}$$

NCC remodelling of fibronectin informs the contact guidance cue

In order for NCCs to move via contact guidance, they need to sense the orientation of fibrous FN. Since FN fibres do not initially exist in the model, however, we must add a mechanism by which NCCs assemble them from immature FN puncta. We thus distinguish between two types of FN in our modelling framework: puncta and fibres. Fibres are nearly identical to FN puncta, but are equipped with an additional angle ϕ that represents their orientation. This distinction is only important in the calculation of the contact guidance cue: when computing the haptotaxis cue (described above), fibres are treated in the same way as FN puncta.

We note that this simplistic approach for modelling FN fibres cannot investigate how fibre lengths, or their crosslinks, may impact cell migration. Nor do we consider the existence of bipolar fibres, which admit two orientations (a forward and reverse direction) along which cells migrate. Such omissions differ from the biological reality of FN matrices: dense, fibrillar matrices can impede invasion, for instance, and FN fibres can admit cell migration in forward and reverse directions (Hillen, 2006; Perumpanani and Byrne, 1999; Yurchenco et al., 1994). We do not seek to create a highly detailed model of fibrous FN matrices, however – rather, we aim to develop a model that determines whether mechanisms of ECM remodelling and contact guidance play important roles during collective migration. Since this simplistic approach can answer the latter question, we do not consider other methods of modelling FN fibres in this chapter.

We model the conversion of FN puncta into fibres via the following procedure: when a punctum is passed over by a NCC – that is, when a FN punctum falls within the cell radius of a leader or follower cell – we equip the FN object with an orientation ϕ equal to the angle θ that the NCC velocity vector makes with the x-axis. This new FN fibre orientation is not fixed for all future times; in fact, it can be further altered when other NCCs (both leader and follower cells) pass over it. The dynamics that cells use to update the FN orientation are based on a differential equation from Dallon et al. (1999; 2000) that maximises the scalar product between the FN orientation and cell velocity:

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \zeta \sin(\theta - \phi), \tag{4.5}$$

where the parameter ζ controls how quickly the FN orientation changes. Under the assumptions that the angular difference $(\theta - \phi)$ is small and θ is constant, we may Taylor expand the right-hand side of Eq. (4.5), neglect higher order terms, and solve the resulting ODE to relate ζ to T_{half}, the time it takes to halve the angular distance:

$$\zeta \approx \frac{\ln(2)}{T_{\mathrm{half}}}. \tag{4.6}$$

The parameter T_{half} is specified by the user prior to the simulation, and lets us investigate how the time scale of FN fibre reorientation by NCCs affects invasion. Figure 4.7 presents a sketch of how the orientations of FN puncta are altered by a NCC.

If a NCC senses FN fibres with angles ϕ_i, then we assume the directional cue from contact guidance, $\mathbf{u}_{\mathrm{cg}}$, is the average fibre orientation:

$$\mathbf{u}_{\mathrm{cg}} = \frac{1}{N_{\mathrm{FN}}(t)} \sum_{i=1}^{N_{\mathrm{FN}}(t)} \begin{pmatrix} \cos(\phi_i) \\ \sin(\phi_i) \end{pmatrix}, \tag{4.7}$$

where $N_{\mathrm{FN}}(t)$ denotes the number of FN fibres that are sensed by the cell at time t within the annular region of inner radius R_{cell} and outer radius R_{filo}. We can normalise this vector to obtain the unit directional cue $\hat{\mathbf{u}}_{\mathrm{cg}}$. We remark that this contact guidance cue is

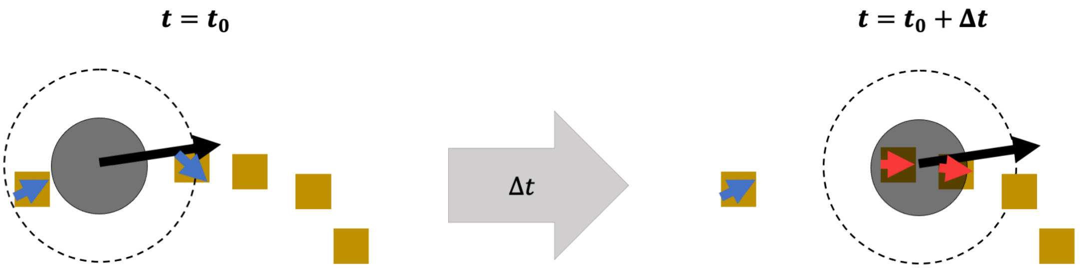


Figure 4.7: A cartoon depiction of a NCC (black circle) at time t_0 (left), moving with a given velocity (black arrow) towards a collection of FN puncta with orientations (yellow squares with blue arrowheads). Over a time period of Δt (right), the NCC passes over FN and remodels their orientations (red arrowheads), either by using Eq. (4.5) – if they already have orientations – or by equipping FN puncta with orientations equal to that of the cell velocity. Only FN covered by the NCC can change their orientations. In both panels, the dashed circle around the NCC represents the maximum length of its filopodial extensions.

unidirectional, and as such will bias cell movement. We could relax this assumption and follow ideas from Hillen and Painter (2013) to extend the ABM by considering bidirectional contact guidance cues, in which cells can also travel backwards along the fibre. We expect, however, that backwards movement along fibres will be restricted in practice because of the forces arising from cell-cell repulsion and haptotaxis. Hence we do not further consider bidirectional contact guidance cues in this chapter. We also note that the contact guidance signal does not directly incorporate stochastic noise; this assumes that cells "read" their microenvironment geometry more accurately than they read differences in the concentrations of various glycoproteins. This assumption follows from *in vitro* studies of migrating mesenchymal cells, which determined that cell polarity is highly dependent on the ECM geometry and that cells can sense and adapt to new fibre orientations with a high degree of accuracy (Nuhn et al., 2018; Théry et al., 2006).

Combining the haptotactic and contact guidance cues

As discussed at the beginning of this section, the direction of the total force from all NCC-ECM interactions, $\hat{\mathbf{F}}_{\mathrm{ECM}}$, is modelled as a linear combination of the haptotaxis and contact guidance cues. To compute this value, we introduce a user-defined parameter value $\rho \in [0, 1]$, which represents the bias of $\hat{\mathbf{F}}_{\mathrm{ECM}}$ towards $\hat{\mathbf{u}}_{\mathrm{hap}}$, the unit haptotactic cue, and take $\hat{\mathbf{F}}_{\mathrm{ECM}}$

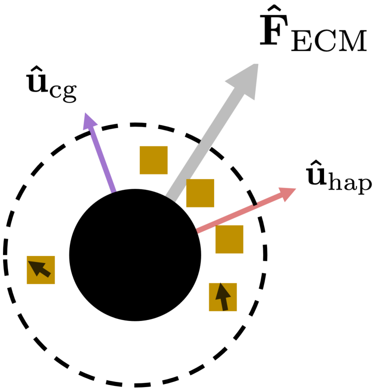


Figure 4.8: A cartoon depicting a NCC (black circle), its unit contact guidance cue, $\hat{\mathbf{u}}_{\text{cg}}$ (purple arrow), its unit haptotaxis cue, $\hat{\mathbf{u}}_{\text{hap}}$ (red arrow), and the direction of its resultant NCC-ECM interaction force, $\hat{\mathbf{F}}_{\text{ECM}}$ (grey arrow), for a particular configuration of FN puncta (yellow squares) and fibres (yellow squares with black arrows). The haptotaxis direction is biased towards increasing densities of FN, while the contact guidance cue is the average FN fibre orientation. The NCC-ECM interaction force is a linear combination of the haptotaxis and contact guidance cues.

to be the sum of $\hat{\mathbf{u}}_{\text{hap}}$ and $\hat{\mathbf{u}}_{\text{cg}}$ weighted by $0 \leq \rho \leq 1$:

$$\hat{\mathbf{F}}_{\text{ECM}} = \frac{\rho \hat{\mathbf{u}}_{\text{hap}} + (1-\rho)\hat{\mathbf{u}}_{\text{cg}}}{\|\rho \hat{\mathbf{u}}_{\text{hap}} + (1-\rho)\hat{\mathbf{u}}_{\text{cg}}\|}, \tag{4.8}$$

where $\|\cdot\|$ denotes the Euclidean norm. If $\rho = 1$ in Eq. (4.8), then $\hat{\mathbf{F}}_{\text{ECM}} = \hat{\mathbf{u}}_{\text{hap}}$, while if $\rho = 0$ then $\hat{\mathbf{F}}_{\text{ECM}} = \hat{\mathbf{u}}_{\text{cg}}$. Hence by adjusting the value of ρ we can control whether haptotaxis or contact guidance dominates the direction of the NCC-ECM force. Figure 4.8 provides a visual example of $\hat{\mathbf{F}}_{\text{ECM}}$, $\hat{\mathbf{u}}_{\text{cg}}$, and $\hat{\mathbf{u}}_{\text{hap}}$ for a particular configuration of cells and FN puncta.

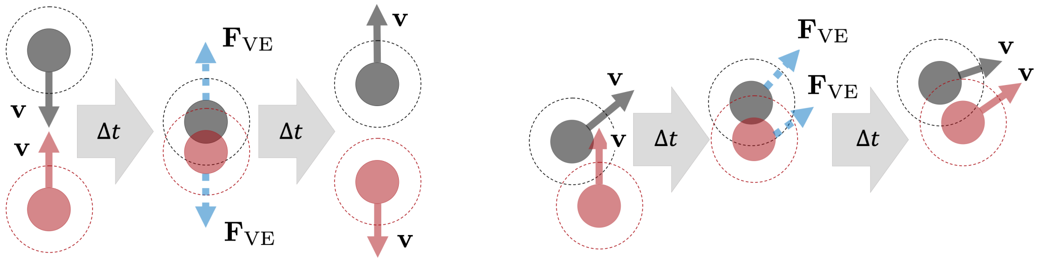
We apply physics-based approaches to determine the magnitude of the NCC-ECM force, F_{ECM}. Specifically, we imagine that elastic springs "pull" a NCC to all FN puncta and fibres that it senses and calculate F_{ECM} as the magnitude of the resulting net force (note that fibres are treated in the same way as FN puncta in this calculation). We assume that the force of each spring decreases quadratically with respect to distance and is non-zero only within a sphere centred at $\mathbf{r}$ with radius R_{filo}. The quadratic function ensures that cells are more attracted to close FN puncta and fibres, and are less affected by more distant species. This formulation is inspired by implementations of cell-cell adhesion and repulsion in other overlapping spheres models, and we remark that other potentials could be used to model cell-ECM attraction, including Leonard-Jones, Hertz, and Morse potentials; the quadratic function is preferred here because of its relative simplicity (Mathias et al., 2020). Given the above discussion, then, F_{ECM} is given by

$$F_{\text{ECM}} = \left\| \sum_{j \in N_{\text{FN}}(t)} \left(1 - \frac{\|\mathbf{d}_j\|}{R_{\text{filo}}}\right)^2 \hat{\mathbf{d}}_j \right\|, \tag{4.9}$$

where $N_{\text{FN}}(t)$ is the number of uncovered FN puncta sensed at time t by the cell, $\mathbf{d}_j$ is the displacement vector from the cell centre, $\mathbf{r}$, to the FN puncta and fibre locations, $\mathbf{q}_j$, and $\|\cdot\|$ denotes the Euclidean norm. The total force from all NCC-ECM interactions is given by $\mathbf{F}_{\text{ECM}} = F_{\text{ECM}} \hat{\mathbf{F}}_{\text{ECM}}$.

4.2.2 Computing the volume exclusion force

The volume exclusion force separates cells and prevents their intersection. Specifically, it incorporates repulsive forces that arise from cell-cell contact. The force can also be extended to consider interactions due to cell-cell adhesion, although these effects are not explicitly considered here for simplicity. We highlight again that the phrase “volume exclusion” does not imply that cells can never intersect but instead was chosen to be consistent with the terminology introduced by Schumacher et al. (2016) to denote a particular mode of cell-cell repulsion that is distinct from CIL, as described below. Cell separation can arise through a variety of mechanisms, as indicated in Figure 4.9: NCCs may interact via CIL, for instance, in which case they repel each other along the direction of contact upon their collision, as observed in *Xenopus* embryos (Theveneau and Mayor, 2012). Alternatively, cells may use other mechanisms to separate from each other: chick cranial NCCs, for instance, do not necessarily interact via CIL because they have been seen migrating in directions besides the one along the point of contact (Genuth et al., 2018; Kulesa et al., 2004). Since our model specifically aims to model observations in chick cranial NCCs, we model cell-cell repulsion in a way that prevents close cells from overlapping but also allows them to select more random directions (possibly towards each other) upon contact when they are further apart. We determine this direction similarly to the haptotaxis signal: specifically, we record the locations, $\mathbf{p}_j$, of all neighbouring NCCs within a distance R_{filo} of a cell centred at location



(a) Contact inhibition of locomotion

(b) Volume exclusion

Figure 4.9: A cartoon depicting the differences between (a) CIL and (b) volume exclusion. In CIL (a), when cells (red and black circles) begin to intersect, they re-orient their velocities (black and red arrows) away from the point of contact. Volume exclusion (b), however, admits more general directions after travel: for instance, NCC velocities could be aligned with each other, or even be unchanged, after contact. In each panel, the dashed circle around each cell depicts its maximum filopodial length, while the cell-cell force at the time of intersection, $\mathbf{F}_{\mathrm{VE}}$, is a light blue dashed arrow.

$\mathbf{r}$. We then represent neighbours as a continuous "cell density" $S_{\mathrm{cell}}(\mathbf{r})$ using Gaussian functions:

$$S_{\mathrm{cell}}(\mathbf{r}) = \sum_{j \in N_{\mathrm{cell}}(t)} \gamma_{\mathrm{cell}} \times \exp\left(- \frac{\|\mathbf{r} - \mathbf{p}_j\|^2}{2R_{\mathrm{cell}}^2} \right), \tag{4.10}$$

where $N_{\mathrm{cell}}(t)$ represents the total number of neighbouring cells located in a ball of radius R_{filo} with centre $\mathbf{r}$ at time t and γ_{cell} is a user-defined parameter that specifies the height of a single Gaussian function. The width of the Gaussian function is proportional to the cell radius R_{cell}, which ensures that the gradient of $S_{\mathrm{cell}}(\mathbf{r})$ is steepest when two cells are about to overlap. This construction thus increases the probability that a cell moves away from its nearest neighbour (in a direction that likely prevents overlapping) when they become close.

The gradient of S_{cell} at $\mathbf{r}$ is then used to parametrise a von Mises distribution for the volume exclusion force direction, $\hat{\mathbf{F}}_{\mathrm{VE}}$. Since cells tend to move towards lower cell density regions, we set the mean of this distribution to be the opposite direction of the gradient, while its width is inversely proportional to its magnitude. The resulting probability density function is given by

$$\mu_{\mathrm{VE}} = \mathrm{Arg}\Big(- \nabla S_{\mathrm{cell}}(\mathbf{r}) \Big), \quad \kappa_{\mathrm{VE}} = \|-\nabla S_{\mathrm{cell}}(\mathbf{r})\|, \quad p(\theta | \mu_{\mathrm{VE}}, \kappa_{\mathrm{VE}}) = C e^{\kappa_{\mathrm{VE}} \cos(\theta - \mu_{\mathrm{VE}})}, \tag{4.11}$$

where $p(\cdot)$ is the probability density function of the von Mises distribution and C is a constant. When the magnitude of the gradient is large, the force is likely oriented towards sparse regions. By contrast, the force direction is more likely to be random when the cell density gradient magnitude is small.

We compute the magnitude of $\mathbf{F}_{\mathrm{VE}}$ in a similar manner to that of the force arising from NCC-ECM interactions. In particular, we imagine that a given NCC is repelled by its neighbours, and the net sum of these repulsive forces determines the magnitude of the volume exclusion force. We again model the strength of these forces using a decreasing quadratic function whose support is the ball of radius R_{filo} centred at $\mathbf{r}$. This follows previous implementations of cell-cell repulsion in other overlapping spheres models, although we note that other functions such as Morse, Hertz, and Leonard–Jones potentials may alternatively be used (Mathias et al., 2020). The magnitude of $\mathbf{F}_{\mathrm{VE}}$ for cell i is given by

$$F_{\mathrm{VE}} = \left\| \sum_{j \in N_{\mathrm{cell}}(t)} \sqrt{c_i c_j} \left(1 - \frac{\|\mathbf{d}_j\|}{R_{\mathrm{filo}}} \right)^2 \hat{\mathbf{d}}_j \right\|, \tag{4.12}$$

where $\|\cdot\|$ is the Euclidean norm, $N_{\mathrm{cell}}(t)$ denotes the number of neighbouring cells at time t located within a sphere of radius R_{filo} centred at $\mathbf{r}$, c_i is the (user-defined) strength of volume exclusion in the NCC of interest, c_j is the volume exclusion strength of its neighbour, and $\mathbf{d}_j$ is the displacement vector from $\mathbf{p}_j$ (the neighbouring NCC centre) to $\mathbf{r}$. We remark that, in practice, c_i and c_j control the size of F_{VE} and they allow us to investigate how the relative dominance of volume exclusion and ECM-related forces affect NCC migration (we note that there are no equivalent parameters for the NCC-ECM force strength in Eq. (4.9); they are set equal to one for simplicity). Additionally, we note that, for the remainder of this chapter, both leader and follower cells have identical values of c_i. Hence we can simplify Eq. (4.12) by rewriting $\sqrt{c_i c_j} = c_i$. We will consider scenarios in which the values of c_i for leader cells differ from those of follower cells in Chapter 5. The total weighted directional cue due to volume exclusion is given by $\mathbf{F}_{\mathrm{VE}} = F_{\mathrm{VE}} \hat{\mathbf{F}}_{\mathrm{VE}}$.

4.2.3 Combining directional cues to find a resultant NCC velocity

Once both $\mathbf{F}_{\mathrm{ECM}}$ and $\mathbf{F}_{\mathrm{VE}}$ are known, we use Eq. (4.1) to compute the resulting cell velocity. However, since the sum of these forces may generate unrealistically large speeds, we add an additional rule that controls the maximum speed at which cells travel. Our experimental collaborators have observed evidence of a chemokinesis-like mechanism in which NCCs moving through FN-rich areas travel faster than NCCs migrating in areas devoid of FN (data not shown). We coarsely model this result by stipulating two maximum speeds at which cells may move: the first, $s_{\text{max}}^{\text{FN}}$, represents the maximum cell speed when a NCC senses at least one FN punctum. If the NCC does not sense any FN, then the maximum speed is given by $s_{\text{max}}^{\text{off}} < s_{\text{max}}^{\text{FN}}$, a user-defined parameter that represents the maximum cell speed in FN-free areas.

This method of accounting for cell speeds is deliberately simplistic. In reality, a more complicated non-monotonic relationship exists between cell speeds and the ECM density. Specifically, while it is true that cells travel faster when the ECM is sparse, very dense FN matrices can also impede cell migration (Zaman et al., 2005; Perumpanani and Byrne, 1999). It is possible to adapt the existing modelling framework to account for this behaviour – for instance, one could use a bell-shaped function to compute the maximum cell speed instead of the two-speed model discussed above. In this chapter, however, we are most interested in investigating how different mechanisms of migration (such as haptotaxis and contact guidance) affect collective behaviour, rather than providing a complete description of *in vivo* NCC migration. Since this two-speed model allows us to sufficiently answer the former question, we shall not further consider those more complicated scenarios.

4.3 Typical ABM results

We list all of the user-specified ABM parameters in Table 4.1, along with default values that we use to compute simulations. Most values derive from the literature: as in previous overlapping spheres models for NCC migration, we take R_{cell} to be between 7.5 and 12.5 μm to reflect *in vivo* sizes of NCCs (McLennan et al., 2020; Szabó et al., 2019). Studies in *Xenopus* embryos also suggest that the time scale at which cells remodel fibronectin is on the order of 30 to 60 min (Rozario et al., 2009; Davidson et al., 2008). Thus, we select values of T_{ave} and T_{half} that lie within this range. Since NCCs emerge from sections of the neural tube that are typically between 100–200 μm in length, we select a value for l_{entr}

Parameter	**Description (units)**	**Value**
Δt	time step (min)	0.1
R_{cell}	NCC radius (μm)	10
R_{filo}	maximum filopodial extension length (μm)	50
λ_{FN}	initial spacing between FN puncta (μm)	35
l_{entr}	height of neural tube corridor (μm)	150
T_{ave}	average time interval before leader FN secretion (min)	30
T_{half}	approx. half-life of FN re-orientation by NCCs (min)	60
ρ	bias of ECM signal to haptotaxis direction (a.u.)	0.5
$s_{\text{max}}^{\text{FN}}$	max cell speed when NCCs sense FN (μm/min)	0.75
$s_{\text{max}}^{\text{off}}$	max cell speed when NCCs do not sense FN (μm/min)	0.05
c_i	relative strength of cell-cell repulsion (a.u.)	1.25
γ_{FN}	height of Gaussian function for $S_{\text{FN}}(\mathbf{r})$ (a.u.)	300
γ_{cell}	height of Gaussian function for $S_{\text{cell}}(\mathbf{r})$ (a.u.)	100

Table 4.1: Table of user-specified ABM parameters, along with the default values used to compute the plots in §4.3 (a.u. = arbitrary units).

within this range (McLennan et al., 2012).

We infer the maximum speed at which individual NCCs travel through experimental measurements of cells within migrating streams. Time lapse imaging of NCCs over time periods of two hours reveals, for instance, that cells tend to travel at speeds ranging from 0.3 to 1.25 μm/min (Ridenour et al., 2014). Since these measurements indicate that most NCCs travel at lower speeds *in vivo*, we set $s_{\text{max}}^{\text{FN}} = 0.75$ μm/min. We also set $s_{\text{max}}^{\text{off}} = 0.05$ μm/min to be roughly an order of magnitude smaller than $s_{\text{max}}^{\text{FN}}$, so as to better distinguish cells that sense FN from those that do not. Finally, while most experiments record NCC filopodia on the order of 30 μm in length, values greater than 50 μm have also been used in other modelling frameworks (Colombi et al., 2020; McLennan et al., 2012; McLennan and Kulesa, 2010). We select $R_{\text{filo}} = 50$ μm to represent values near the larger end of this range.

Other values are, unfortunately, nearly impossible to parameterise from existing data. To our knowledge, for instance, there are no detailed measurements of the FN distribution prior to NCC migration. Instead, we set λ_{FN} to be roughly half of the NCC filopodial radius in order to ensure that NCCs can sense multiple FN puncta as they invade. The parameters γ_{cell}, γ_{FN}, and Δt are chosen based on their ability to yield migratory patterns that closely resemble those of *in vivo* NCCs, and are selected based on results of multiple ABM simulations. Finally, we randomly sample ρ and c_i from ranges for which contact guidance, haptotaxis, and volume exclusion all act on cells. Based on preliminary simulations of the ABM, we determined that qualitative behaviours of solutions did not appear to be significantly affected when ρ is varied within the range $[0, 0.25]$ (i.e., when contact guidance strongly dominates NCC-ECM interactions). We later confirmed that summary statistics, such as the distance that cells travel and the the average nearest neighbour distance, do not change very much when ρ is varied within this region (data not shown). As a result, we only consider values within the range $\rho \in [0.25, 1]$. Additionally, we have used a positive but small lower bound for the range of the parameter c_i (the range is $[10^{-5}, 5]$) to ensure that cell-cell repulsion can always occur.

We remark that, although many of the values in Table 4.1 cannot be readily inferred from existing data sets, the global sensitivity analysis of the ABM in the next sections will investigate how dependent its results are upon the choice of parameter values.

Figure 4.10 presents snapshots of an ABM simulation that uses the parameter values listed in Table 4.1. We observe that the leader cells (black circles) migrate towards the right boundary, in the direction of the target corridor (which we assume is a rectangle of height l_{entr} and width 500 μm, as depicted in Figure 4.12(b)). Furthermore, follower cells (red circles) are able to sense the paths of the leader cells. Over the course of 12 hours, NCCs travel in a unified stream that resembles those observed *in vivo.* This simulation implies that NCC-ECM interactions and volume exclusion can be sufficient for leaders and followers to migrate collectively, since the ABM includes only these mechanisms.

We also observe in Figure 4.10 that the leader cells appear to maintain their positions at the front of the stream over time. Indeed, we have observed that it is rare for follower cells to overtake leader cells, even though they can do so in principle. Such a result is likely due to the fact that cells at the front of streams (whether they exhibit a “leader” or “follower” phenotype) have fewer individuals directly in their path. Consequently, they are

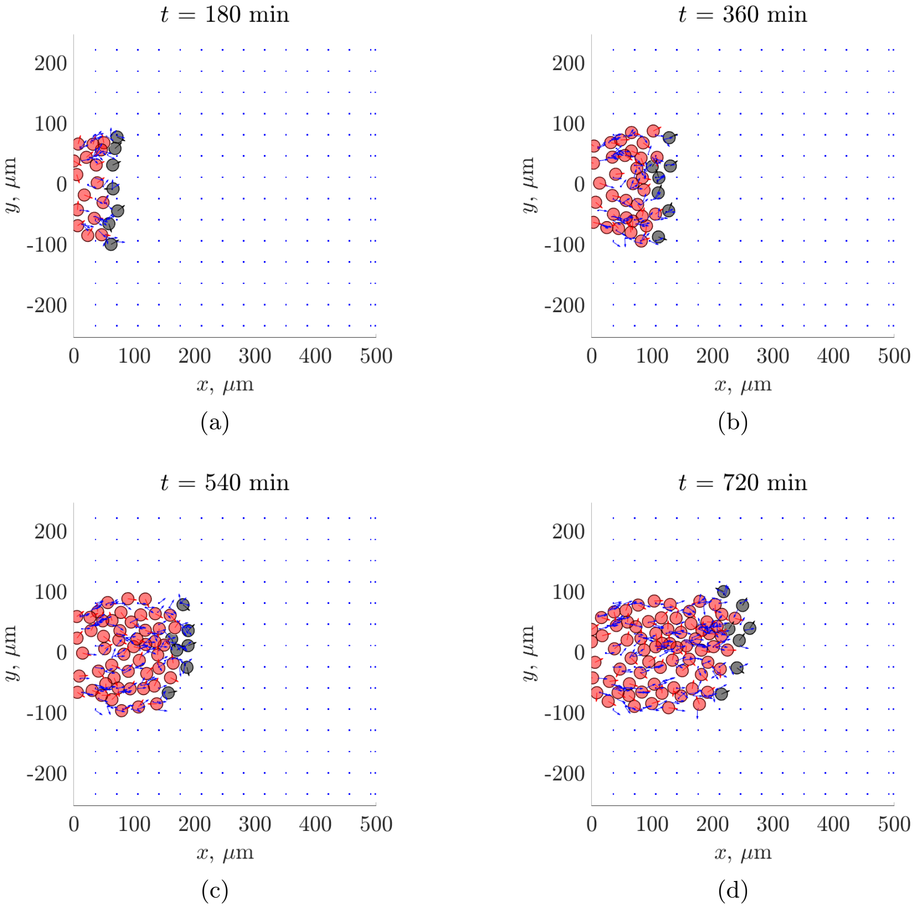
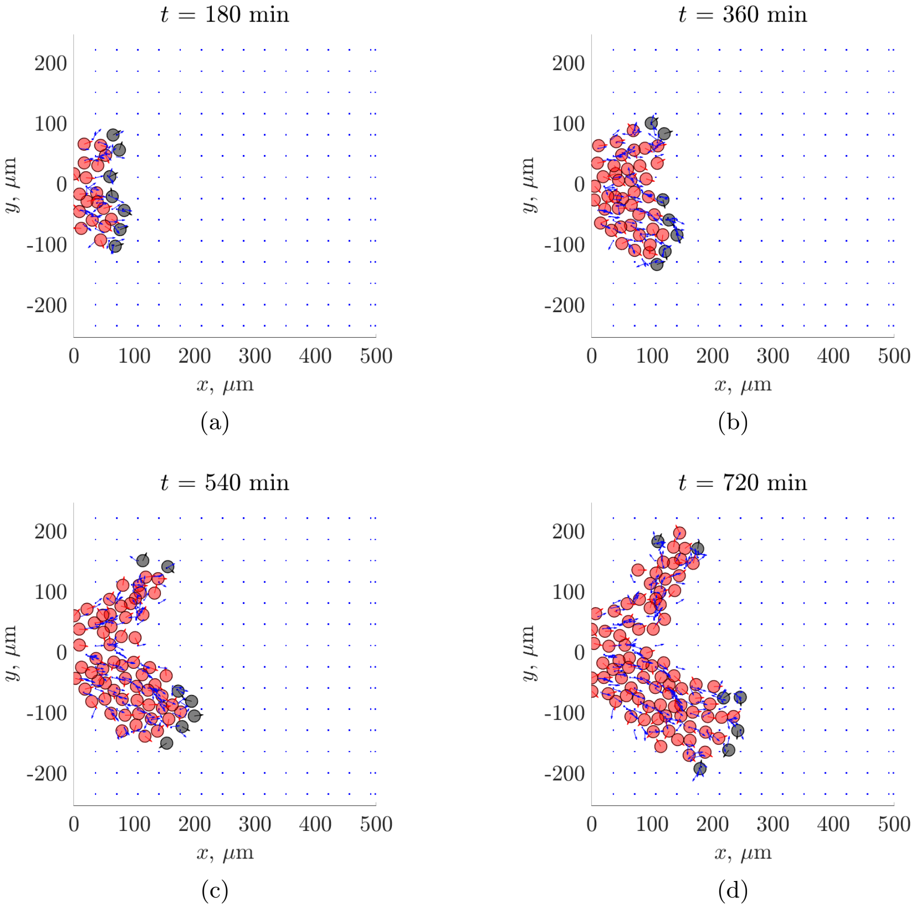


Figure 4.10: Snapshots showing the cell locations of leader cells (black circles), follower cells (red circles), and FN puncta (blue squares) at times (a) $t = 180$ min, (b) $t = 360$ min, (c) $t = 440$ min, and (d) $t = 720$ min. The parameter values for the simulation are given in Table 4.1. Red and black arrows represent the direction of current cell velocities, while blue arrows indicate FN fibre orientations.

more likely to sense uncovered FN puncta, and to travel faster, without being “blocked” by neighbours. This suggests that the ability of leader cells to maintain their position at the front of streams depends on the initial conditions we have used, i.e., that this result depends on our assumption that the leader (secretory) cells first emerge from the neural tube at time $t = 0$ min. We stress, however, that this choice of initial conditions is motivated by *in vivo* observations from our biological collaborators (Paul Kulesa, *pers. comm.*), and is why we do not consider other initial conditions in this chapter.

Cells do not always migrate along the correct target corridor in the ABM, however. Figure 4.11 presents the results of another ABM realisation that uses the parameter values

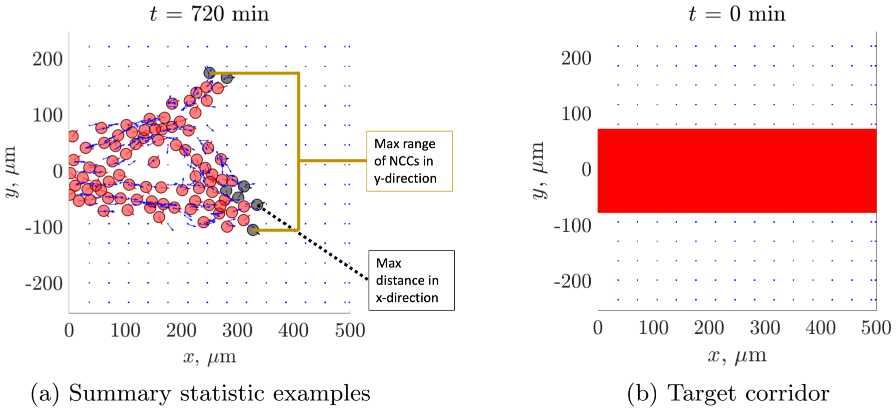


(a) (b)

(c) (d)

Figure 4.11: Snapshots showing the cell locations of leader cells (black circles), follower cells (red circles), and FN puncta (blue squares) at times (a) $t = 180$ min, (b) $t = 360$ min, (c) $t = 440$ min, and (d) $t = 720$ min. The parameter values for the simulation are the same as in Figure 4.10. Red and black arrows represent the direction of current cell velocities, while blue arrows indicate FN fibre orientations.

listed in Table 4.1. At early time points (e.g., at $t = 180$ min), we observe that NCCs appear to migrate as a single collective stream, just as in the previous simulation. Over time, however, the stream splits into two branches that invade regions perpendicular (or lateral) to the target corridor (which is assumed to be located horizontally from where NCCs enter the domain). This branching phenomenon likely results from NCCs at the front sensing and migrating towards FN puncta and fibres that are located lateral to the target corridor. As these cells convert puncta into fibres and secrete new FN, they create a continually reinforced signal that stimulates trailing cells to follow them along the errant paths.



(a) Summary statistic examples (b) Target corridor

Figure 4.12: Given (a) a configuration of NCCs (red and black circles) and FN puncta (blue squares/arrows), we compute summary statistics such as the maximum distance NCCs travel in the x-direction (dashed black line) and the range of NCCs in the y-direction (solid yellow lines); the latter statistic provides a rough indication of how far NCCs may have travelled laterally from (b) the target corridor (red rectangle), which we assume to be of height l_{entr} and width 500 μm. For the configuration of NCCs and FN puncta shown in (a), the summary statistics are as follows: Maximum x-distance of NCCs: 335.0 μm, range of NCCs: 279.2 μm, average nearest distance between NCCs: 20.75 μm, order parameter for follower cells: 0.75. Note that distances are measured from cell centres.

4.4 Constructing summary statistics for collective migration

The stochastic nature of the ABM motivates us to develop summary statistics to characterise NCC migration, as the previous figures indicate that different patterns are likely to be generated over multiple realisations. We describe these metrics in more detail below; two are depicted in Figure 4.12(a).

4.4.1 Maximum invasion distance along the target corridor direction

Since the NCC target corridor is assumed to be perpendicular to the neural tube (the left-hand boundary), we measure successful invasion by recording the maximum distance that any NCC travels in the x-direction. We also compute the average distances that leaders and followers travel, the distributions of NCCs along the x-direction, and their average speed. However, we find that these latter statistics tend to produce similar conclusions as the first metric mentioned here (data not shown). Thus, we do not discuss these other statistics further.

4.4.2 Maximum lateral range of NCCs

We measure the difference between the maximum and minimum position of all NCCs in the y-direction in order to infer the extent of lateral migration from the target corridor (whose height is assumed to coincide with that of the neural tube). We note, however, that while this range is correlated with the lateral migration distance, it cannot determine the number of NCCs that are outside the target corridor. Furthermore, since multiple stream branches may form during the simulation, this statistic does not necessarily measure the width of NCC streams.

To gain further insight into the direction that NCCs travel, we have also recorded the average FN fibre orientation. Our motivation for measuring this value arises from the ABM rule that aligns fibres to NCC velocities. We find, however, that this statistic tends to be less informative (it often yields values close to zero radians, as cells move up or down equally likely – this is the case for Figures 4.10 and 4.11, as the average FN fibre orientation in either graph is no more than 0.12 radians in magnitude). Thus we do not discuss this statistic further.

4.4.3 Average nearest distance to another NCC

The next summary statistic we compute is the average distance from a NCC to its nearest neighbour. By recording this value, which has also been used to characterise NCC migration in *in vitro* and *in vivo* settings (McKinney et al., 2020), we aim to infer the cell density within a migrating collective. We have also recorded a more direct statistic for cell density: the average number of neighbouring cells that are found within a ball of 50 μm from a NCC. We found, however, that this statistic is correlated with the nearest neighbour distance (data not shown). Furthermore, the latter statistic may have a deceptively low value depending on the number of streams that form. For instance, in a case where there are two branched streams, some cells will have fewer neighbours because of the gap between branches; this will decrease the statistic value, even though the cell density within each individual stream may be large (this is the case in Figure 4.11). For these reasons, we only report the results of the nearest neighbour distance.

4.4.4 Order parameter for cell alignment

The final summary statistic we consider is known as the order parameter Φ, and measures the degree of cell alignment. It is inspired by a similar metric from Vicsek et al. (1995), and is computed as the magnitude of the average unit velocity vector:

$$\Phi = \left\| \frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} \right\|, \tag{4.13}$$

where $\|\cdot\|$ denotes the Euclidean norm and N the number of cells at the time point of interest. Note that $\Phi \in [0, 1]$: it is approximately zero when cell velocities are uncorrelated and random, but equal to unity when all cells move in the same direction. Thus, the order parameter value indicates the extent of alignment between NCCs: higher values suggest that a large number of cells share the same behaviour and migrate collectively.

4.5 Sensitivity analysis reveals dominant mechanisms affecting NCC behaviours

Having identified summary statistics for collective migration, we now aim to determine their sensitivity to perturbations in ABM parameter values using Latin hypercube sampling (LHS) and extended Fourier Amplitude Sensitivity Testing (eFAST). The benefits of this analysis are twofold: firstly, it can determine regions of parameter space for which collective migration likely occurs. Additionally, since certain ABM parameters directly relate to mechanisms of cell migration (such as haptotaxis, contact guidance, and volume exclusion), this analysis can be used to infer how such processes contribute to certain macroscopic behaviours, such as the total distance cells migrate.

4.5.1 LHS/PRCC results

As discussed in §1.3, LHS can be combined with the calculation of partial rank correlation coefficients (PRCCs) to identify monotonic relationships between ABM summary statistics and parameter values. We aim to analyse parameters related to the underlying migratory mechanisms of the ABM: these include those related to NCC-ECM interactions, (λ_{FN}, ρ, T_{ave}, T_{half}) and volume exclusion (c_i). Additionally, we consider how other parameters (such

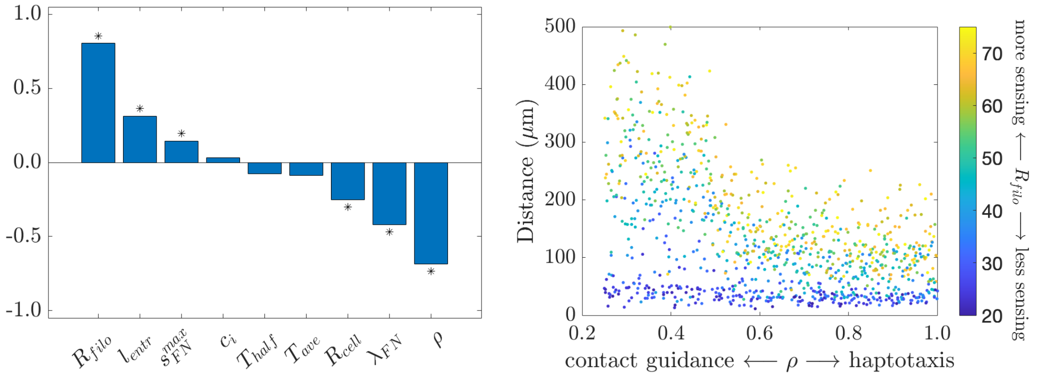
Parameter	Range	Units	Parameter	Range	Units
R_{filo}	$[20, 75]$	μm	l_{entr}	$[30, 500]$	μm
λ_{FN}	$[20, 75]$	μm	R_{cell}	$[7.5, 17.5]$	μm
ρ	$[0.25, 1]$	a.u.	$s_{\max}^{\text{FN}}$	$[0.5, 1.25]$	μm/min
T_{ave}	$[15, 90]$	min	c_i	$[10^{-5}, 5]$	a.u.
T_{half}	$[2, 90]$	min			

Table 4.2: List of ABM parameters and ranges of values investigated using LHS (a.u. = arbitrary units). For a discussion of how these parameters relate to the ABM rules, we refer to §4.2. All other parameters are fixed at the values from Table 4.1.

as the filopodial radius, cell size, maximum cell speed and neural tube height) affect NCC invasion. The full list of parameter values investigated using LHS, including their ranges, is presented in Table 4.2. The ranges are chosen based on values from the literature that we detailed in §4.3. Unlisted ABM parameters are taken to be constant, and are set equal to the default values given in Table 4.1.

We generated one thousand unique ABM parameter sets from LHS using the `lhsdesign` function of the Statistics and Machine Learning Toolbox in Matlab (version 2021a). This number of samples covers as much of the parameter space as possible, while simultaneously ensuring that the analysis completes within a reasonable time frame. We adopt an approach from Marino et al. (2008) to confirm that $N = 1000$ samples are sufficient for the LHS/PRCC analysis. This method uses a metric known as the top-down correlation coefficient (Iman and Conover, 1987) to determine the extent to which PRCCs are ordered consistently by their effect on output statistics when the number of samples increases. Results from this analysis (presented in Appendix C.2) indicate that as few as 300 samples are needed to consistently rank the sensitivity of parameter values for all summary statistics listed above, hence one thousand regimes is sufficient for the analysis presented here.

Since the ABM is stochastic, we must compute multiple ABM realisations for each parameter regime in order to properly capture the average summary statistic values. We simulate twenty ABM replicates for each parameter regime, as taking more realisations causes the LHS/PRCC analysis to become computationally intractable. For example, the



(a) PRCCs (b) Scatter Plot for ρ, coloured by R_{filo}

Figure 4.13: PRCCs (a) for the maximum distance of NCCs in the x-direction at $t = 720$ min, along with (b) a scatter plot of the statistic versus ρ, with points coloured by the maximum filopodial length R_{filo}. Black asterisks in (a) indicate whether the PRCC is significantly different from zero at the 99% level (following a *post hoc* Bonferroni correction). The p-values were computed using a two-sided t-test. For a description of ABM parameter values, we refer to the text and Table 4.2.

data presented below took approximately three weeks to compute on a single workstation consisting of four cores (Intel(R) Core(TM) i5-7500T CPU @ 2.70 GHz).

In this analysis, we only compute summary statistics at $t = 720$ min (or, equivalently, 12 hours): this is done to ensure we capture the long-term emergent behaviour of the model, rather than the transient dynamics. Additionally, we choose this time point because NCC migration occurs on time scales of roughly this magnitude (McKinney et al., 2020; Szabó et al., 2019, 2016).

Figure 4.13(a) presents PRCCs that indicate how the parameters in Table 4.2 affect the maximum distance that NCCs travel in the x-direction. The largest PRCC belongs to R_{filo}, the maximum filopodial length, and indicates a monotonically increasing relationship exists between the statistic and the parameter value. Indeed, the scatter plot for this parameter (not shown) is nearly linear. The initial FN spacing (λ_{FN}) and cell radius (R_{cell}) also have statistically significant negative PRCCs, which suggests that increasing the initial distance between FN puncta, or the cell size, reduces the invasion distance. All three of the above parameters relate to the ability of NCCs to sense FN: if NCCs have longer filopodia, then they are more likely to sense FN puncta and other cells. Additionally, since FN puncta that are covered by cells cannot be sensed in this ABM, we would expect that increasing the cell body radius would decrease the distance that NCCs travel. Figure 4.13(b) presents a scatter plot of the statistic versus the parameter ρ, which is coloured by the value of R_{filo}. We observe that cells travel distances of about 50 μm or less when the filopodial radius is small, and that altering the parameter ρ does not seem to affect the distance for this subset of data. For cells with larger filopodia, however, we see in the graph that NCCs can travel nearly the entire length of the domain. This scatter plot confirms that NCCs must sense FN in order to successfully colonise the target corridor; migration is also more likely to occur when the initial FN distribution is denser and/or the cell size is small (as both changes increase the probability that cells sense FN).

The parameter ρ, which determines whether contact guidance or haptotaxis dominate NCC-ECM interactions, supplies the second largest PRCC by magnitude. Its value is negative, suggesting that NCCs travel shorter distances when haptotaxis dominates NCC-ECM interactions. This likely occurs because haptotactic cues, unlike those of contact guidance, are sampled from probabilistic distributions. Since this means that the directions of these cues likely change at every time point, cells whose NCC-ECM interactions are dominated by haptotaxis tend to travel less persistently (and, consequently, invade smaller distances over time) than cells whose interactions are dominated by contact guidance. Figure 4.13(b) presents a scatter plot for the maximum invasion distance versus ρ, with data points coloured by the filopodial length R_{filo}. We observe two trends in the data: for cells with relatively small filopodia (that are unlikely to sense FN), the maximum distance travelled is roughly constant with respect to ρ and has a relatively small value. Regimes for which cells sense multiple FN, however, exhibit a monotonically decreasing relationship. This plot thus confirms the correlation suggested by the PRCC and supports the conclusion that NCC-ECM interactions largely determine successful NCC collective migration.

The other parameters that yield statistically significant PRCCs in Figure 4.13(a) include the maximum cell speed, $s_{\text{FN}}^{\max}$, and the neural tube height, l_{entr}. However, scatter plots of the statistic versus each of these parameters (not shown) do not show clear evidence of correlations. The parameter for the volume exclusion force strength (c_i) does not yield a significant PRCC either, and further investigation of its scatter plot (not shown) does not reveal a clear correlation with the maximum invasion distance.

Analysis of the statistic for lateral migration – the range of NCCs in the y-direction – provides similar results those listed above (data not shown). For that case, the largest PRCC corresponds to l_{entr}, the height of the neural tube section from which cells emerge, reflecting the fact that longer neural tube sections lead to a wider range of NCCs along the y-direction. The scatter plot for this parameter (not shown) confirms a nearly linear relationship with the statistic. Comparison of the PRCCs with Figure 4.13(a) also reveals that nearly all of the parameters that yield statistically significant PRCCs for one metric are also statistically significant for the other (the only difference is that the range of NCCs in the y-direction has a significant positive PRCC with respect to c_i, the parameter related to the strength of volume exclusion forces). Scatter plots for parameters other than l_{entr} (not shown), however, do not indicate that such correlations exist. Thus, we cannot conclude from the LHS/PRCC analysis which mechanisms most contribute to lateral NCC migration.

The two statistics investigated above are calculated for all cells, rather than just leader or follower cells. Consequently, they alone cannot be used to determine the extent to which follower cells travel along the paths of leader cells. We can use the order parameter Φ to answer the latter question, since order parameters with values close to unity indicate that cells are aligned in the same direction (which is indicative of collective migration). The PRCCs for this statistic (not shown) are relatively small, which signals that follower cell alignment is more robust to parameter perturbations than other metrics. The largest PRCC by magnitude corresponds to ρ, and its negative value suggests that biasing NCC-ECM forces to haptotaxis reduces follower cell alignment. This again appears to result from the fact that NCCs stochastically sample their haptotactic cues, but not their contact guidance cues: we thus would expect cells to exhibit similar velocity directions if their movement is dominated by contact guidance as opposed to haptotaxis. The corresponding scatter plot of Φ versus ρ (not shown) also indicates that a weak, monotonically decreasing relationship exists.

While the volume exclusion strength parameter, c_i, and cell radius, R_{cell}, also generate significant PRCCs for this statistic, their corresponding scatter plots (not shown) demonstrate very weak relationships with the metric at best. We therefore cannot say with certainty that mechanisms related to these parameters greatly affect the collective alignment of follower cells.

By contrast, these two parameters strongly affect the separation between NCCs (both leader and follower cells). The two largest PRCCs by magnitude for the average nearest neighbour distance (not shown), for example, belong to R_{cell} and c_i. Their coefficients are both positive, which suggests that increasing the size of the volume exclusion force (or, alternatively, making cells larger) will increase the distance between NCCs, as we intuitively expect. The scatter plot for the parameter R_{cell} (not shown) confirms that a strong linear relationship exists between the cell radius and the separation distance. The scatter plot for c_i (not shown) reveals a relatively disperse but monotonically increasing relationship. Since c_i controls the volume exclusion force strength, and because cells are more likely to interact via volume exclusion when they become larger, these results indicate that the volume exclusion force separates NCCs, as we expect.

The other three parameters that produce significant PRCCs for the average nearest neighbour distance are T_{ave}, the average time interval for FN secretion by leader cells, ρ, and R_{filo}. Only T_{ave}, however, exhibits a scatter plot with a monotonically increasing relationship (data not shown, the points are dispersed and have a slight linearly increasing trend). This correlation implies that greater FN secretion is associated with more NCC clustering.

The conclusions from the LHS/PRCC analysis may be summarised as follows: (1) successful invasion is most sensitive to whether NCCs can sense FN. This implies that collective migration is more heavily dependent on NCC-ECM interactions than forces related to volume exclusion. (2) Cells travel further distances when contact guidance dominates NCC-ECM interactions. Furthermore, follower cells that move via contact guidance tend to be more aligned with each other, which indicates that this mechanism likely promotes collective migration. (3) Haptotaxis primarily clusters cells together, as parameters that increase the dominance of this mechanism (such as ρ and T_{ave}) are associated with smaller distances between NCCs. Upregulation of haptotaxis is also associated with more uncorrelated follower cell velocities. (4) Although parameters related to volume exclusion do not yield statistically significant relationships with the distance that NCCs travel (both along and lateral to the target corridor), they are unsurprisingly associated with increased separation between NCCs.

4.5.2 eFAST results

While the LHS/PRCC analysis provides some insight into how the different mechanisms incorporated into the ABM produce complicated macroscopic behaviour, it does not represent the only approach to global sensitivity analysis. Methods such as eFAST, for example, offer a different way to characterise sensitivities. Moreover, eFAST supplies additional information that LHS/PRCC analysis cannot capture: for instance, eFAST can characterise nonlinear relationships that are also non-monotonic. Furthermore, it uses a variance-based metric (the Sobol index) to describe model sensitivities, rather than a correlation coefficient: as a result, sensitive parameters found by LHS/PRCC and eFAST may differ (Marino et al., 2008). This motivates us to conduct another global sensitivity analysis of the ABM using eFAST.

Due to the relatively large computational requirements for eFAST – the algorithm must test a minimum of at least 65 regimes per parameter, for example – we cannot investigate all nine parameters from Table 4.2 within a reasonable time frame. We can, however, perform eFAST on a subset of parameters that are of particular interest. We focus on three: R_{filo}, the maximum filopodial radius; ρ, the parameter controlling whether the NCC-ECM force is biased toward haptotaxis or contact guidance; and c_i, the parameter that controls the volume exclusion force strength. We choose to analyse the filopodial radius, R_{filo}, and bias towards contact guidance cues, ρ, because the LHS/PRCC analysis indicates that they are very important to determining the success of migration. We also investigate c_i because it is most directly related to the mechanism of volume exclusion, and because the LHS/PRCC analysis unexpectedly indicates that it does not significantly affect the distance that cells migrate.

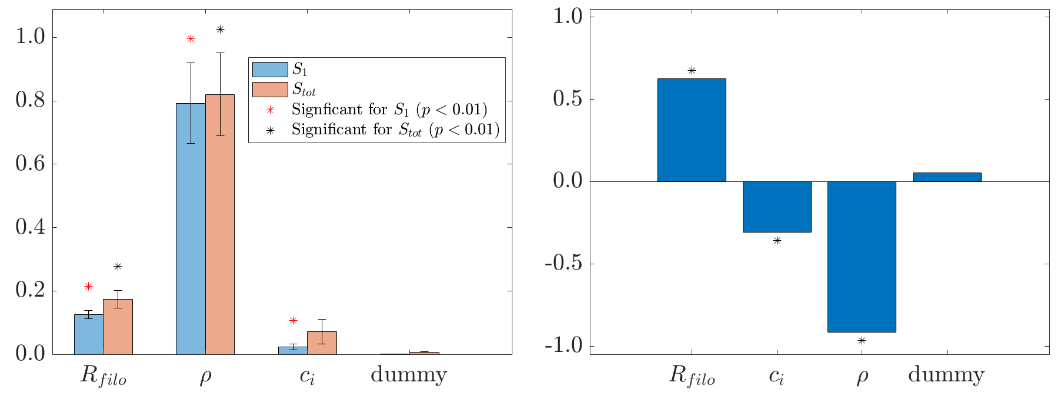
Parameter	**Range of values**	**Units**
R_{filo}	$[30, 75]$	μm
ρ	$[0.25, 1]$	a.u.
c_i	$[0, 5]$	a.u.

Table 4.3: List of ABM parameters and ranges investigated using eFAST (a.u. = arbitrary units). For a discussion of how these parameters relate to the ABM rules, we refer to §4.2. All other parameters are fixed at the values from Table 4.1.

Table 4.3 presents the parameters and their ranges of values to be investigated by eFAST. We remark that the intervals for R_{filo} and c_i are slightly different from the LHS/PRCC analysis: we change the range of c_i to include zero values so that we consider cases where volume exclusion does not occur in ABM simulations, and we adjust that of R_{filo} to ensure that it is greater or approximately equal to λ_{FN}, such that cells are likely to sense FN (since the LHS/PRCC analysis indicates that cells fail to invade when $R_{\text{filo}} \ll \lambda_{\text{FN}}$). Any parameters not investigated with eFAST are set to the default values found in Table 4.1.

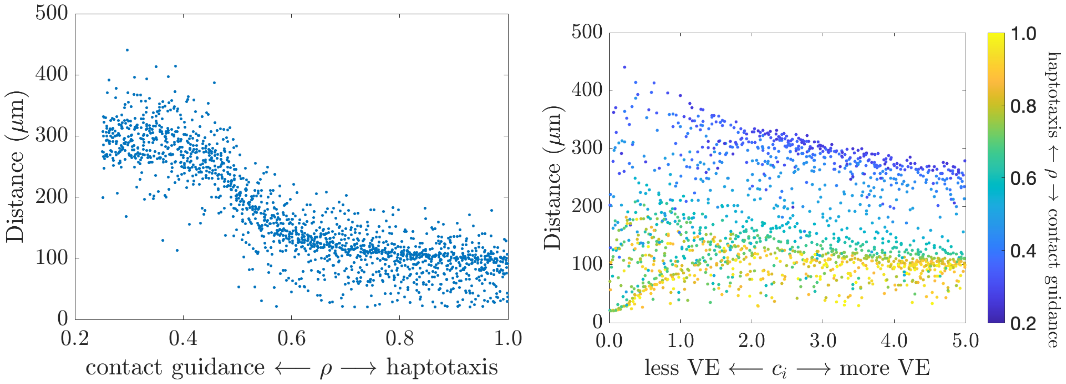
We carry out parameter sampling and Sobol index computation using the SALib library in Python (version 3.9.1), which provides unit-tested and open source functions for eFAST (Herman and Usher, 2017). We sample $N = 131$ values – slightly more than double the minimum amount – for each of the three parameters of interest plus a dummy parameter, and repeat this stochastic sampling algorithm three times in order to evaluate the statistical significance of Sobol indices. We thus analyse a total of 1,572 parameter regimes. For each set of values, we again average the summary statistics over twenty ABM replicates to accurately determine the average values of summary statistics within a reasonable time frame. We also compute PRCCs for each parameter using the `partialcorr` function in Matlab (version 2021a). As noted in Chapter 1, the data presented below required approximately one month of computation time on a single workstation consisting of four cores (Intel(R) Core(TM) i5-7500T CPU @ 2.70 GHz).

Figure 4.14 presents Sobol indices, PRCC values, and scatter plots for the maximum distance NCCs travel in the x-direction. We observe in Figure 4.14(a) that the parameter ρ produces the largest first- and total-order Sobol indices, which implies that the statistic



(a) Sobol Indices

(b) PRCCs



(c) Scatter Plot for ρ

(d) Scatter Plot for c_i, coloured by ρ

Figure 4.14: Sobol indices (a), PRCCs (b), and scatter plots (c)-(d) for the maximum distance that NCCs have travelled in the x-direction at 720 min. Error bars in (a) represent two standard deviations of the Sobol index. Asterisks in (a) denote whether S_1 (red) or S_{tot} (black) is significantly different from the respective dummy parameter index at the 99% significance level. Asterisks in (b) indicate if the PRCC is significantly different from zero at the 99% significance level (following the application of a *post hoc* Bonferroni correction). The p-values for the Sobol indices and PRCCs were both calculated using a two-tailed t-test.

is most sensitive to this parameter. Its PRCC, shown in Figure 4.14(b), indicates that the parameter exhibits a significantly negative correlation with the statistic, which we confirm by its corresponding scatter plot in Figure 4.14(c). These results are consistent with the conclusion from LHS/PRCC analysis that NCC-ECM interactions dominated by contact guidance are very important for successful invasion.

The parameter R_{filo} supplies the next largest first- and total-order Sobol indices, which are both significantly different from those of the dummy parameter. Its PRCC suggests a positive correlation with the statistic, which we confirm by its scatter plot (not shown). These results again agree with the conclusion from the LHS/PRCC analysis that cells travel larger distances when they sense more FN.

The parameter c_i has a small but significant first-order Sobol index, which indicates that the metric is also sensitive to changes of this parameter value (albeit to a lesser degree than for R_{filo} and ρ). The PRCC for c_i indicates that it exhibits a negative correlation with the maximum invasion distance. Its scatter plot, however, suggests a more complicated relationship: Figure 4.14(d) presents this graph, which is coloured by the value of ρ. For points with $\rho < 0.5$, the statistic appears to monotonically decrease with respect to c_i; for points with $\rho > 0.8$, however, the distance increases with c_i before reaching a plateau. These results imply that volume exclusion can increase the distance NCCs travel, so long as their NCC-ECM interactions are dominated by haptotaxis. When contact guidance governs such forces, however, then stronger volume exclusion cues tend to decrease the distance that NCCs migrate along the target corridor. We remark that this conclusion differs from that obtained using LHS/PRCC analysis, which found that invasion distance was not sensitive to volume exclusion-related parameters. Thus, eFAST has supplied information that cannot be obtained by the former approach.

We next analyse the range of NCCs in the y-direction in order to investigate how ABM parameter perturbations affect the lateral migration of cells. Although we do not present the results here, we find that ρ again produces the largest Sobol indices, followed by the filopodial radius. The parameter c_i, by contrast, does not generate statistically significant values. PRCCs for these parameters (not shown) confirm this ordering of parameters: ρ exhibits the largest PRCC by magnitude, and its strong monotonically decreasing relationship can be seen in the corresponding scatter plot (data not shown). The PRCCs for R_{filo} and c_i, on the other hand, suggest weaker relationships. Subsequent analysis of their scatter plots (not shown) suggests that there is very little correlation between these parameters and the range of NCCs in the y-direction: aside from two clusters of data points (which correspond to $\rho > 0.7$ and $\rho < 0.5$, respectively), it is difficult to discern a clear trend for the statistic with respect to c_i. Overall, then, these results suggest that NCCs tend to move more laterally when NCC-ECM interactions are dominated by contact guidance, but that parameters related to volume exclusion and FN sensing do not affect this macroscopic behaviour much (if at all).

Unlike the results for the above statistics, we did not find any statistically significant Sobol indices for Φ, the order parameter that measures the extent of follower cell alignment. The PRCCs for this statistic (not shown), however, indicate that both ρ and c_i have significant negative correlations with the order parameter. We confirm from their scatter plots (not shown) that the order parameter does strongly, and monotonically, decrease with the value of ρ. The scatter plot for c_i also exhibits a monotonically decreasing relationship (data not shown). There are also two clusters of data points that respectively relate to whether NCC-ECM interactions are dominated by haptotaxis ($\rho > 0.7$) or contact guidance ($\rho < 0.5$). These results thus suggest that haptotaxis and volume exclusion are associated with less cell alignment. Conversely, increasing the dominance of contact guidance in NCC-ECM interactions promotes the collective alignment of follower cells to each other.

The parameters ρ and c_i also affect the average nearest distance between NCCs. Plots of the Sobol indices (not shown) reveal that these two parameters yield significant first-order Sobol indices for this statistic (with the volume exclusion parameter generating the largest of these values). The bar chart also reveals that the only significant Sobol index for R_{filo} is its total-order value, which implies that this parameter only affects the distances between cells through its higher order interactions with ρ and c_i. This makes sense, given that cells need to sense FN or neighbours in order for volume exclusion and NCC-ECM forces to act on them.

We obtain similar conclusions from plots of the PRCCs for this statistic (data not shown). The coefficients suggest an identical relationship to the Sobol indices: the largest values by magnitude correspond to c_i and ρ, for instance, while the parameter R_{filo} does not suggest that a statistically significant relationship exists. Scatter plots for ρ and c_i (not shown) confirm the relationships suggested by the PRCCs: we observe that the statistic monotonically decreases with ρ, and that it remains constant for roughly $\rho \geq 0.5$. The scatter plot for c_i also increases monotonically. These results suggest that the main mechanisms that act to separate cells are volume exclusion and contact guidance. Haptotaxis, on the other hand, appears to generate denser migrating collectives.

To summarise the eFAST results, we have found the following: (1) the invasion distance of NCCs is most sensitive to whether haptotaxis or contact guidance dominates NCC-ECM interactions, with contact guidance favouring longer streams. Forces dominated by this latter mechanism are also associated with greater follower cell alignment and increased lateral migration away from the target corridor. (2) Interactions dominated by haptotaxis, on the other hand, are associated with more random follower cell velocities and more compact clusters of NCCs. (3) As one might expect, larger volume exclusion forces increase the separation distance between NCCs. Parameters related to this mechanism have mixed effects on the distance that NCCs travel, however: for cells with haptotaxis-dominated NCC-ECM interactions, they are associated with greater distances travelled. Conversely, larger volume exclusion forces tend to hinder invasion for cells whose ECM interactions are dominated by contact guidance.

We expect these conclusions to hold even when we add noise to the contact guidance cue, since the latter mechanism largely determines how persistently cells migrate. Although we have not yet tested this conjecture, we plan to do so in future work.

4.6 Discussion and conclusion

In this chapter, we introduced a novel off-lattice ABM for NCC collective migration in which leading cells deposit and remodel an initially punctate FN matrix to create a reinforced signal for follower cells. To identify model parameter values that most affect macroscopic observations of collective NCC migration, we conducted a global sensitivity analysis of ABM summary statistics using tools such as LHS/PRCCs and eFAST. In doing so, we determined how the mechanisms of haptotaxis, contact guidance, and volume exclusion each contribute to collective migration. Some results are intuitively obvious, and help to verify the code: we find that volume exclusion-related parameters, for instance, increase the distance between cells, and that increasing the dominance of haptotaxis in NCC-ECM interactions attracts cells to each other (which makes sense, given that leader cells secrete FN within migrating streams). Furthermore, cells become more aligned to each other, and travel larger distances, when NCC-ECM interactions are dominated more by contact guidance (this makes sense because these cues are not sampled from a probability distribution, and thus lead to more persistent cell migration).

The global sensitivity analysis also yields counter-intuitive insights, however: the distance cells migrate is not sensitive to parameters that dictate the time scale at which NCCs reorient FN fibres, which is surprising considering that contact guidance is a key determinant of NCC migration (and requires fibres). This implies that the initial establishment of fibres by cells and/or the pre-existing FN distribution are more important to determining the success of invasion. Furthermore, the eFAST results suggest a context-dependent role for volume exclusion: it can hinder NCC invasion when NCC-ECM interactions are dominated by contact guidance, but enhance it when those forces are influenced more by haptotaxis. Thus, the analysis presented above sheds light on the black box that connects individual mechanisms to more complicated macroscopic behaviours.

This work adds to a rich history of mathematical models that investigate how the ECM affects collective cell migration in other diverse contexts such as (to name but three) cancer metastasis, angiogenesis, and somitogenesis (Stepanova et al., 2021; Suveges et al., 2021; Azimzade et al., 2019; van Oers et al., 2014; Martins et al., 2009; Baker et al., 2008; Olsen et al., 1997). Unlike previous models, however, the ECM in our modelling framework initially exhibits an immature, discrete, and punctate structure that is subsequently assembled into fibres by migrating NCCs (and is motivated by observations from our experimental collaborators). Despite this difference, certain conclusions we obtain from the ABM are similar to those found in other modelling frameworks. We found, for instance, that contact guidance and ECM remodelling are sufficient mechanisms to support collective migration (Painter, 2009). The establishment of oriented ECM fibres is also correlated with successful cell invasion (Suveges et al., 2021; Azimzade et al., 2019). Even if the initial FN distribution favours isotropic migration, cells in the ABM can ultimately move anisotropically by remodelling the ECM – this result is consistent with observations from off-lattice models for fibroblast migration (Park et al., 2019). The ABM also predicts that FN fibres and contact guidance can create a "memory" for follower cells that allows them to follow leader cell paths (D'Alessandro et al., 2021; Wershof et al., 2019).

The results indicate that volume exclusion (and, by extension, interactions due to cell-cell contacts) plays a more complicated role in NCC migration than most CIL-based models suggest. Namely, we find that cell-cell contacts may actually hinder cell invasion of the target corridor when the interactions between NCCs and the ECM are dominated by contact guidance rather than haptotaxis. The global sensitivity analysis suggests that, compared to contact guidance, volume exclusion does not play a dominant role in cell migration. Although this conclusion contrasts with those obtained from CIL-based models (Khataee et al., 2021; Szabó et al., 2016; Woods et al., 2014; Carmona-Fontaine et al., 2011), we note that such discrepancies may result from the fact that CIL is not directly implemented in the ABM: while cells do repel each other at close distances, more distant cells may actually move towards each other upon contact. Nevertheless, the complicated role for volume exclusion suggested by the ABM is an interesting avenue to explore in future studies. One could investigate, for instance, whether CIL produces different conclusions than those obtained from using the definition of volume exclusion; this could, in turn, inform *in vivo* experiments that aim to distinguish between the two types of cell-cell interaction.

Our goal in constructing this ABM is not to completely describe NCC migration but, instead, to evaluate how certain migratory mechanisms affect the macroscopic behaviours of cells. Indeed, the ABM does not capture all aspects of *in vivo* NCC migration: for example, NCCs in the ABM exhibit excessive migration lateral to the target corridor compared to real migrating streams. Furthermore, *in vivo* experiments demonstrate that NCCs typically migrate distances on the order of 1000 μm or more (McKinney et al., 2020); this is more than twice the distance that agents typically travel in the ABM.

These differences are valuable from a modelling perspective, however, as they provide insight into other mechanisms that may act on NCCs during their migration. For instance, we could add a new guidance cue in the ABM to direct NCCs along the appropriate target corridor; this could represent previously ignored mechanisms in chick cranial NCCs such as chemotaxis up VEGF gradients (McLennan et al., 2012). Alternatively, we could prevent lateral movement from the target corridor by adding an internal boundary that inhibits NCCs from migrating into such areas. Although such a boundary is not known to occur far away from the neural tube in the chick head, molecules such as versican do act to prevent lateral NCC migration in *Xenopus* embryos (Szabó et al., 2016). It would be interesting whether similar confinement occurs for chick cranial NCCs *in vivo*, and whether such a mechanism would promote collective migration in the context of the ABM. Other processes such as domain growth could also advect cells, allowing them to invade farther distances (McKinney et al., 2020). Additionally, previous *in silico* studies have also suggested that phenotype switching between leaders and followers may cause cells to invade more of the target corridor (Schumacher, 2019; McLennan et al., 2015a,b). In future studies, we plan to introduce these previously ignored mechanisms into the ABM framework in order to determine what combination of cues leads to robust collective migration.

Chapter 5

Agent-based modelling of neural crest cell migration provides insights into mechanisms during robust stream migration

5.1 Overview

In Chapter 4, we developed a new agent-based model (ABM) for neural crest cell (NCC) migration in which cell trajectories are determined by forces that arise from cell-cell repulsion (which we termed "volume exclusion") and interactions with the extracellular matrix (ECM). We applied techniques from global sensitivity analysis to determine the extent to which certain features of collective behaviour, such as the distance that cells travel during a fixed time period, change in response to perturbations in ABM parameter values. While this analysis provides some insight into the roles of different mechanisms incorporated into the ABM framework, it does not identify what combination of them are required for "normal" collective migration to occur. Such issues can be resolved by perturbing individual parameters for cell migration: by upregulating or removing mechanisms from the ABM framework, we can infer which ones are necessary for collective migration and which confer robustness to the process. Additionally, these tests can also guide the design of future biological experiments.

Such questions motivate the work of this chapter. In consultation with our experimental collaborators (Paul Kulesa and Mary McKinney, Stower's Institute for Medical Research, USA; Lance Davidson, University of Pittsburgh), we use the ABM to perform *in silico* experiments in which we perturb individual processes related to the fibronectin distribution, the ability of NCCs to remodel fibronectin, and the strength of volume exclusion forces acting on NCCs. The results strengthen the understanding of mechanisms underpinning collective migration, and go beyond the conclusions provided by the global sensitivity analysis. They suggest, for instance, that NCCs migrate more efficiently when leader and follower cells exhibit specialised roles (in which leader cells steer the collective by creating trails of remodelled fibronectin, while follower cells sense and move along the fibronectin trails). This observation, along with evidence that NCCs in the ABM migrate more laterally than *in vivo*, motivates us to introduce a new mechanism into the ABM that steers leader cells along a defined target corridor. We find that this "guiding signal", which could be chemotactic in nature (McLennan et al., 2010), reduces lateral migration and increases the distance that cells travel along the target corridor, however in extreme cases it can cause leader and follower cells to separate. We then provide, to our knowledge, the first detailed investigation of different mechanisms that can prevent stream breaks in models for NCC migration by identifying parameter combinations for which leader cells do not separate from follower cells.

5.1.1 Chapter outline

The rest of this chapter is organised as follows: in §5.2, we perturb individual mechanisms underpinning NCC migration in the ABM, using several summary statistics to determine how macroscopic behaviours change from those of a baseline parameter regime. In §5.3, we introduce into the modelling framework a new force, which we term a "guiding force" (possibly arising from chemotaxis or mechanical-based mechanisms), that steers NCCs along a defined target corridor. Although there are extreme cases for which the guiding force causes leader cells to separate from follower cells, in §5.4 we identify different combinations of parameter perturbations that do not give rise to such gaps between leader and follower cells. The combinations that successfully avoid leader-follower separation are then used to infer which mechanisms are likely to confer robustness to NCC migration.

5.2 ABM simulations explore the roles of different mechanisms in collective NCC migration

In Chapter 4, we found that several summary statistics for NCC collective migration were sensitive to changes in ABM parameters related to volume exclusion, NCC-ECM interactions, and the pre-existing fibronectin distribution. In this section, we perturb these values individually, to determine how each mechanism alters macroscopic outputs from a baseline state. In doing so, we infer what combination of mechanisms may be required for collective NCC migration to occur *in vivo*. We remark that all of the computational experiments performed in this section generate *in silico* results, and therefore are limited to the context of the ABM (although we prioritise tests that could most readily be validated by our experimental collaborators in the near future). We also only consider forces on NCCs that arise from volume exclusion, contact guidance, and haptotaxis. We will discuss in later sections how forces associated with other mechanisms – such as chemotaxis or mechanical-based cues – may alter the summary statistics.

We begin by defining a set of baseline ABM parameter values that are chosen to align closely with *in vivo* observations of NCCs. We infer these values from observations found in the literature, and from the results of the global sensitivity analysis in Chapter 4. We remark that the baseline parameter regime used here is slightly different from that used to generate example ABM simulations in that chapter. This is because we here estimate values that are closer to observations of *in vivo* and/or *in vitro* cell migration (see the discussion below). For simplicity, we will occasionally refer to this baseline parameter regime in graphs as the “wild-type” (WT) state. We stress, however, that this terminology does not imply that the baseline parameter regime represents the underlying biology: instead, it represents a benchmark against which we may compare the effects of different parameter perturbations.

Most parameter values are readily obtained from the literature: we set $R_{\text{filo}} = 27.5$ μm, for instance, since NCCs usually extend filopodia 20–30 μm from their centres of mass (Colombi et al., 2020; McLennan and Kulesa, 2010). We also choose the NCC radius to be $R_{\text{cell}} = 7.5$ μm; this follows previous studies that fix cell diameters to lie within the range

15–25 μm (McLennan et al., 2020; Szabó et al., 2019). We set $l_{\text{entr}} = 120$ μm because *in vivo* measurements of the neural tube indicate that NCCs emerge from regions of approximately this length (Szabó et al., 2019; McLennan et al., 2012). We remark that, because the number of leader cells is determined by the value of l_{entr}, this fixes the number of leader cells (to eight individuals) in all simulations. This is a reasonable assumption, however, given *in vivo* observations that there is a relatively small number of NCCs at the front of streams, that cells near the front tend to express genes related to a highly motile, invasive "trailblazer" phenotype, and that this "trailblazer" identity is not as highly expressed in the majority of cells further back in the stream (Morrison et al., 2017; McLennan et al., 2015a). While, to our knowledge, no studies detail how quickly chick cranial NCCs can remodel fibronectin, this process has been studied in *Xenopus* embryos and occurs on the time scale of 30 min to one hour (Rozario et al., 2009; Davidson et al., 2008). We therefore fix $T_{\text{ave}} = T_{\text{half}} = 30$ min. We note that it is reasonable to use this order of magnitude estimate for fibronectin remodelling, since the global sensitivity analysis in the previous chapter suggests that neither parameter significantly affects any summary statistic relating to the distance that NCCs travel.

The global sensitivity analysis also informs the selection of other parameters such as $s_{\text{max}}^{\text{FN}}$, the maximum speed at which individual NCCs travel when they sense fibronectin. This is necessary because *in vivo* experiments do not measure the rate at which individuals travel: instead, they measure the speed of cell groups, which typically lies within the range 0.3 to 0.7 μm/min (Ridenour et al., 2014). Using the results from Chapter 4, however, we can make a scatter plot that relates $s_{\text{max}}^{\text{FN}}$ to the average speed of all NCCs (data not shown). The graph indicates that cells which individually travel at a maximum speed of 0.8 μm/min yield group velocities within the *in vivo* range. The maximum speed at which NCCs travel when they do not sense fibronectin, $s_{\text{max}}^{\text{off}}$, is assumed to be an order of magnitude smaller (0.05 μm/min) than $s_{\text{max}}^{\text{FN}}$, reflecting the observation that NCCs do not migrate over large distances when fibronectin is absent (Rovasio et al., 1983).

We fixed parameter values related to the time step and heights of the Gaussian functions for the haptotaxis and volume exclusion cues (Δt, γ_{FN}, and γ_{cell}, respectively) to have the

Parameter	**Description (Units)**	**Value**
Δt	time step (min)	0.1
R_{cell}	NCC radius (μm)	7.5
R_{filo}	maximum NCC filopodial extension length (μm)	27.5
λ_{FN}	spacing between initial fibronectin puncta (μm)	20
l_{entr}	height of the neural tube corridor from which NCCs enter (μm)	120
T_{ave}	average time interval before leader fibronectin secretion (min)	30
T_{half}	approx. half-life of fibronectin re-orientation by NCCs (min)	30
ρ	bias of NCC-ECM interactions to haptotaxis (d.u.)	0.5
$s_{\text{max}}^{\text{FN}}$	max cell speed when NCCs sense fibronectin (μm/min)	0.8
$s_{\text{max}}^{\text{off}}$	max cell speed when NCCs do not sense fibronectin (μm/min)	0.05
c_i	relative strength of cell-cell repulsion (d.u.)	0.5
γ_{FN}	height of Gaussian function for $S_{\text{FN}}(\mathbf{r})$ (d.u.)	300
γ_{cell}	height of Gaussian function for $S_{\text{cell}}(\mathbf{r})$ (d.u.)	100

Table 5.1: List of baseline parameter values used in this chapter (d.u. = dimensionless units). See the main text for a discussion of how we selected these values. We refer to Chapter 4 for details about how each parameter affects cell trajectories in the ABM.

same values as in Chapter 4. This was done to be consistent with the values used in the global sensitivity analysis; furthermore, we know from the previous chapter that such values generate NCC streams that resemble those observed *in vivo*. The parameters λ_{FN}, ρ, and c_i, which correspond to the initial fibronectin lattice spacing, the proportion of the NCC-ECM interaction force towards the haptotaxis signal, and the strength of volume exclusion forces, respectively, cannot be identified from the literature either. We opt to use values for which NCCs sense at least one fibronectin punctum at any time point ($\lambda_{\text{FN}} = 20$ μm), give equal weight to NCC-ECM interactions arising from haptotaxis and contact guidance ($\rho = 0.5$), and separate from each other at an approximate distance of a cell diameter apart ($c_i = 0.5$). We remark that several of the computational experiments we perform below involve perturbations to these three indeterminate parameters, in order to better understand their influence on cell trajectories.

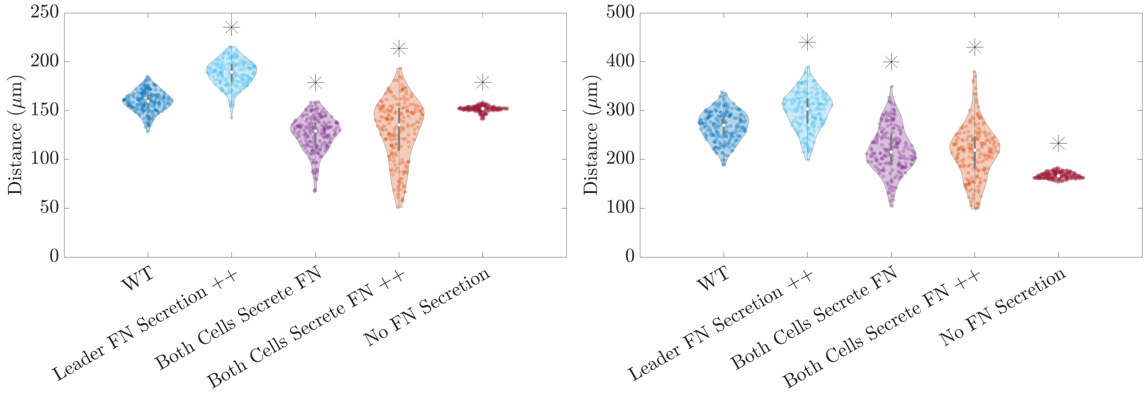
The complete list of baseline parameter values may be found in Table 5.1.

For each parameter regime, we characterise how successfully NCCs colonise the target corridor (which is assumed to be aligned along the x-axis) by calculating the maximum distance that cells travel in the x-direction. Since we found in Chapter 4 that NCCs may migrate laterally from the target corridor, we also quantify lateral dispersion by computing the range of NCCs in the y-direction (computed as the difference between the maximum and minimum y-coordinates of all NCC locations). We will henceforth refer to both statistics as the invasion distance and lateral dispersion of NCCs, respectively.

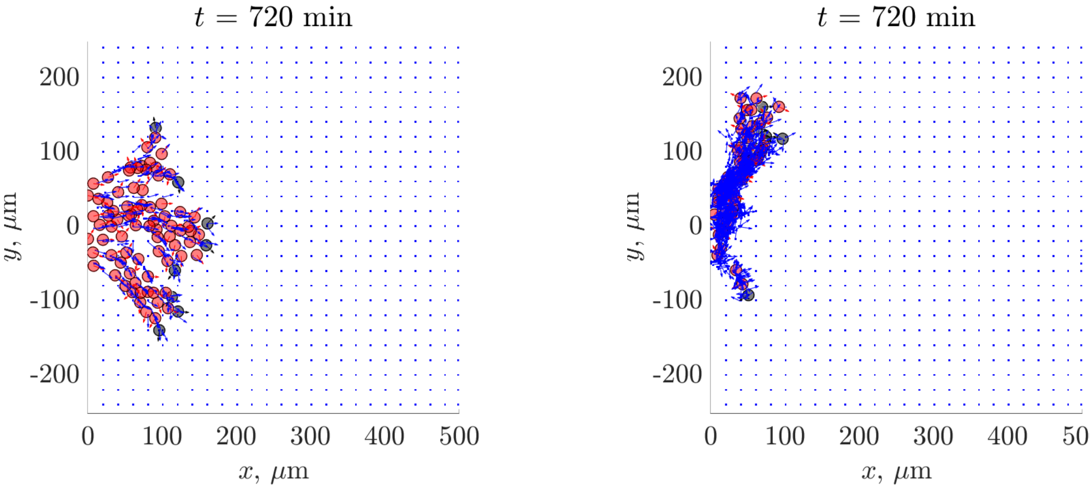
As these summary statistics will vary stochastically over multiple ABM realisations, we use their averages to describe each parameter regime. We employ consistency analysis (Hamis et al., 2021; Alden et al., 2013; Read et al., 2012; Vargha and Delaney, 2000) to determine the number of realisations that are required to ensure that these average values are not greatly affected by noise. More complete description of consistency analysis and its implementation for the modelling framework are given in §D.1. In brief, we find that using 200 realisations per parameter regime ensures, to a statistically significant degree, that differences in outputs obtained from distinct parameter regimes are not wholly generated by inherent ABM stochasticity.

Collective migration is most efficient when leader and follower phenotypes are not identical The first *in silico* experiments perturb the ability of NCCs to secrete new fibronectin (allowing us to anticipate what may happen when we overexpress or knock out fibronectin secretion in one or both cell types). We note, however, that leader and follower cells can still assemble ECM fibres from existing puncta in these experiments, even when they cannot secrete new fibronectin. Figures 5.1(a) and (b) present the resulting distributions for the invasion distance and lateral dispersion of NCCs, respectively, at $t = 720$ min. We observe that the median values of both statistics are significantly larger than those of the baseline case only when leader NCCs secrete fibronectin more rapidly. By contrast, if neither cell type secretes fibronectin, then NCCs travel shorter distances in the x- and y-directions compared to the baseline parameter regime; collective migration still occurs, however, because NCCs sense and move towards existing fibronectin.

We also see that if both leader and follower cells produce fibronectin (or over-secrete



(a) Violin plots, invasion distance (b) Violin plots, lateral dispersion



(c) Baseline parameter regime (d) Both cells secrete fibronectin at baseline rate

Figure 5.1: Violin plots of (a) the invasion distance and (b) the lateral dispersion of NCCs at $t = 720$ min, for experiments that alter the secretion rate of fibronectin by leader and/or follower NCCs. Example realisations of (c) the baseline parameter regime and (d) an experiment in which leader and follower cells both secrete fibronectin at the baseline rate are also shown. Abbreviations: WT (baseline parameter regime), Leader FN Secretion ++ (set $T_{\text{ave}} = 10$ min), Both Cells Secrete FN (leader and follower cells both secrete new fibronectin puncta on average every $T_{\text{ave}} = 30$ min), Both Cells Secrete FN ++ (leader and follower cells both secrete new fibronectin puncta on average every $T_{\text{ave}} = 10$ min), No FN Secretion (neither cell type secretes new fibronectin). We remark that, in contrast to the graphs of §5.3-§5.4, no leader cell experiences a guiding force directing them along the correct target corridor. Black asterisks in (a)-(b) indicate whether the distributions are significantly different from that of the baseline regime (see §D.2 for details). In (c)-(d), black circles denote leader NCCs, red circles follower NCCs, and blue squares fibronectin puncta and/or fibres. Arrows emanating from these objects denote either the current cell velocity direction (black or red) or fibre orientation (blue).

it), then NCCs migrate shorter distances in both the x- and y-directions to a statistically significant degree (see §D.2 for details on the numerical methods). These results imply that NCC migration is most successful when follower cells secrete negligible amounts of fibronectin. Since the only distinction between leader and follower cells in the ABM is that the former population can secrete fibronectin, this suggests that cells migrate smaller distances when leader and follower cell phenotypes are identical.

The observation that NCCs travel shorter distances when both cell types secrete fibronectin may appear counter-intuitive, especially when we consider that overexpressing fibronectin in only leader NCCs promotes migration. When both cell types secrete fibronectin, however, more fibronectin is produced within NCC streams; such high glycoprotein densities create NCC-ECM forces that attract cells towards previously visited areas. While these forces do not entirely prevent cells from migrating (since volume exclusion continues to push cells outward), they reduce the distance NCCs travel over time. By contrast, fibronectin is less dense within travelling streams when only leader cells can secrete it. Furthermore, leader cells near the leading edge of travelling streams will likely not sense newly secreted fibronectin (since they initially “cover” the new puncta and fibres). Consequently, leader NCCs will be more likely to migrate toward unexplored regions if they are the only secretory cell type. To confirm this hypothesis, we investigate a case in which leader NCCs secrete fibronectin at a much greater rate than in the baseline parameter regime (we do this by setting $T_{\text{ave}} = 1$ min, results not shown). For this case, cells indeed travel further distances (almost always reaching the right boundary of the domain), but they do not form wide, ribbon-like streams: instead, cells tend to move as several single-file chains of cells.

Overall, these results suggest a complex, nonlinear relationship between fibronectin secretion and the distance that cells travel towards target sites: while increased fibronectin secretion by leader cells is positively correlated with the invasion distance, too much fibronectin secretion can prevent the formation of wide streams; cells also tend to travel shorter distances when follower cells can secrete new fibronectin puncta. Potential negative correlations between fibronectin secretion and the invasion distance may be further compounded *in vivo*, given that cells travel at smaller speeds in dense ECM (Perumpanani and

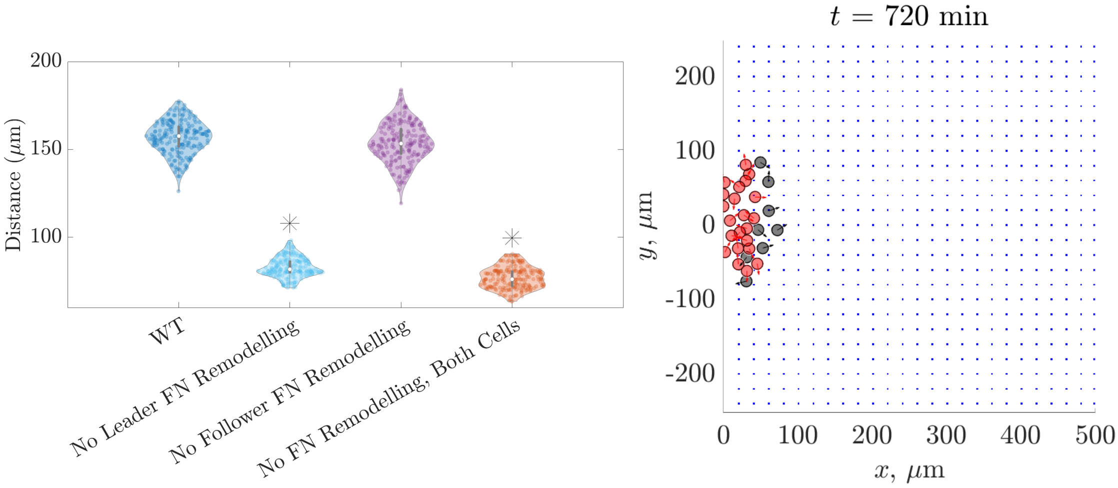
ABM Perturbation	**Leaders**	**Followers**	**Both cells**	**Figure**
No fibronectin remodelling (no secretion, no fibre assembly)	↓↓	—	↓↓	5.2
Fibronectin secretion + ($T_{\text{ave}} = 30$ min)	WT	n/a	↓	5.1
Fibronectin secretion ++ ($T_{\text{ave}} = 10$ min)	↑	n/a	↓	5.1
Fibronectin secretion – (no secretion)	n/a	WT	↓	5.1

Table 5.2: Summary of how the maximum distance that cells travel along the target corridor is changed by perturbations to fibronectin matrix remodelling in different cell types, and references to the figures in which these relationships appear. Unless otherwise noted, cells that cannot secrete fibronectin may still assemble fibres from the existing lattice. Arrows indicate whether the invasion distance is significantly less than or greater than that of the baseline parameter regime; a "–" indicates the change is not significantly different from the baseline case. In the table, "n/a" denotes *in silico* experiments that were not performed, while "WT" refers to statistics obtained using the baseline parameter regime.

Byrne, 1999) – while this effect is not directly incorporated in the ABM, we expect it to further reduce the distance that cells migrate *in vivo* when they over-secrete fibronectin.

Figures 5.1(c) and (d) present example realisations of the baseline parameter set and the case in which both cells secrete fibronectin. Together, they provide further insight into how the above perturbations affect the cell streams. We observe in the former graph that leader cells tend to split off in multiple directions, creating multiple branches along which follower cells travel. The cells are relatively dense, with no obvious separation between leader and follower cells. Comparison of the two graphs reveals that the cells travel shorter distances in Figure 5.1(d), and that there is more fibronectin in the domain. Additionally, cells are co-located with fibronectin that has already been passed over by cells. Overall, these snapshots corroborate the hypothesis that, at least in the context of this ABM, NCCs travel smaller distances along the target corridor when follower cells secrete fibronectin because of NCC-ECM interactions that direct them towards high fibronectin densities.

Successful collective migration depends on fibronectin remodelling by leader cells Although the above experiments show the importance of fibronectin in migration,

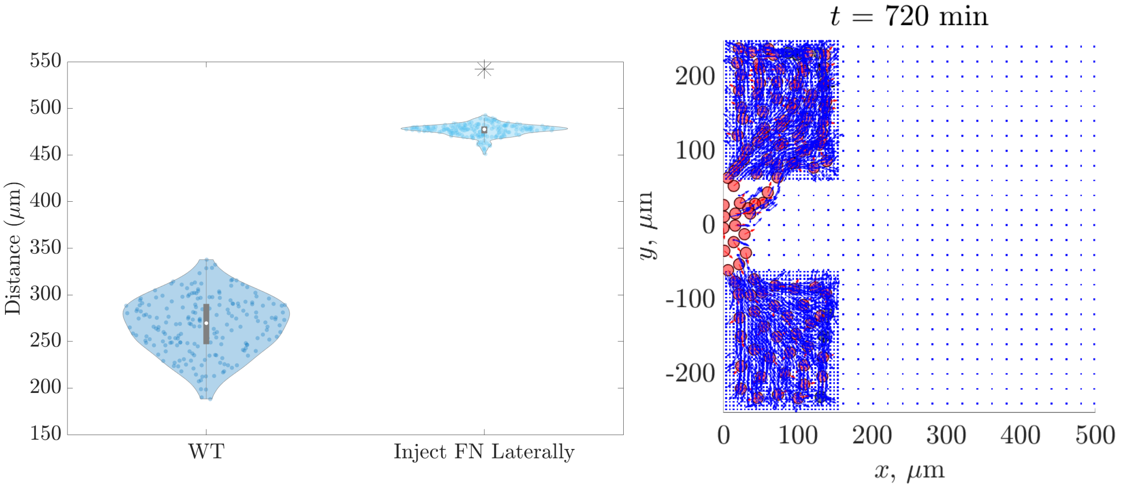


(a) Violin plots, invasion distance

(b) No fibronectin remodelling by either cell type, $t = 720$ min

Figure 5.2: (a) Violin plots for the NCC invasion distance at $t = 720$ min, for *in silico* experiments in which NCCs cannot alter the fibronectin matrix. Abbreviations: WT (baseline parameter regime), No Leader FN Remodelling (leader cells cannot assemble fibronectin fibres or secrete puncta), No Follower FN Remodelling (follower cells cannot assemble fibronectin fibres or secrete puncta), No FN Remodelling, Both Cells (neither cell type can assemble fibronectin fibres or secrete puncta). (b) A realisation of the ABM at $t = 720$ min for a parameter regime in which neither leader nor follower cells can alter the initial fibronectin matrix. See §D.2 for the interpretation of black asterisks in (a). We refer to the caption of Figure 5.1 for a description of the objects (circles, squares, etc.) found in (b).

it remains to be determined if one particular cell type dominates the establishment of collective behaviour. To answer this question, we conduct *in silico* experiments in which we prevent leader and/or follower cells from altering the fibronectin matrix (by eliminating fibronectin secretion and fibre assembly). Figure 5.2(a) presents the resulting distributions for the invasion distance after conducting such manipulations. We observe that this statistic decreases significantly when leader cells (or both cell types) cannot remodel the matrix. If follower cells cannot alter the fibronectin matrix, however, then there are no statistically significant differences from the baseline case. In fact, this is true for the distributions of other summary statistics (such as the lateral dispersion, the average nearest distance between NCCs, and the percentage of cells outside the target corridor, although the data are not shown here). Observation of individual ABM realisations for the baseline parameter regime (not shown) suggests that follower NCCs do not greatly alter the orientation of fibronectin



(a) Violin plots, lateral dispersion (b) Inject fibronectin lateral to corridor

Figure 5.3: (a) Violin plots of the NCC lateral dispersion at $t = 720$ min for the baseline parameter regime and for an *in silico* experiment in which punctate fibronectin lattices with spacing $\lambda_{\mathrm{FN}} = 5$ μm are placed adjacent to the target corridor. (b) An example realisation of the aforementioned experiment at $t = 720$ min. See §D.2 for the interpretation of black asterisks in (a). Abbreviations: WT (baseline parameter regime). We refer to the caption of Figure 5.1 for a description of the objects (circles, squares, etc.) found in (b).

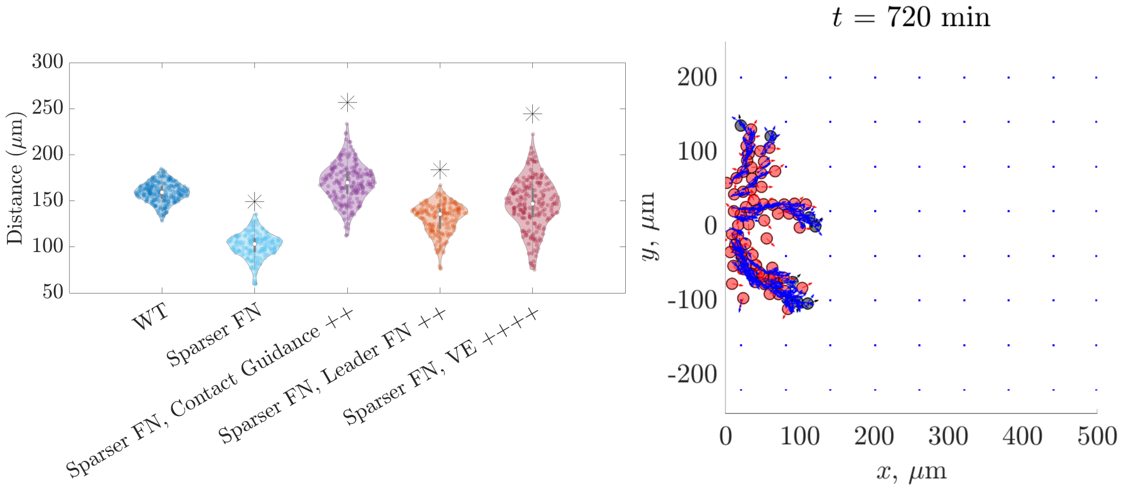
fibres over time. Since ECM fibres become aligned towards NCC velocities when they are "passed over" by cells, this result suggests that cells are largely aligned with each other during collective migration.

These results demonstrate that, within the ABM, fibronectin remodelling by leader NCCs plays a key role in determining the success of collective migration. By contrast, there are no discernible differences in collective behaviour when follower cells cannot remodel the fibronectin matrix. Such conclusions admit the possibility that leader and follower cells serve different but specialised roles, with leader NCCs guiding streams via fibronectin remodelling; follower cells, for their part, may not need to remodel the fibronectin matrix to produce collectively migrating streams. Instead, they may simply follow the trails left by leader cells, using forces that arise from contact guidance, haptotaxis, and volume exclusion.

Table 5.2 presents a summary of the relationships between the invasion distance and the different experimental perturbations of fibronectin remodelling in leader and follower NCCs from Figures 5.1-5.2.

High density regions of fibronectin can control where cells migrate Although the experiments above demonstrate how fibronectin remodelling by NCCs helps determine their collective behaviour, it remains to be established whether external sources of fibronectin can also control this process. This motivates the next *in silico* experiment in which we perturb the initial fibronectin distribution by placing rectangular lattices of fibronectin puncta (with spacing $\lambda_{\text{FN}} = 5$ μm) above and below the neural tube corridor from which NCCs emerge. This represents an experiment in which fibronectin is injected in regions lateral to the target corridor. For simplicity, we take these rectangles to be 150 μm wide and extend them laterally to the top/bottom boundaries of the domain. Figure 5.3(a) presents the resulting distributions for the lateral dispersion of NCCs at $t = 720$ min. For the "injection" experiment, we observe that this statistic is close to 500 μm – the entire domain height – for almost all 200 realisations. Furthermore, the invasion distance (distributions not shown) does not exceed 150 μm for any realisation. These results imply that cells travel to, and remain in, the regions of injected fibronectin. We confirm this in Figure 5.3(b), which presents a typical snapshot of the experiment at $t = 720$ min. The plot reveals that nearly all of the NCCs are within the denser fibronectin areas, even though some cells can sense fibronectin outside these regions. The results of this experiment highlight the importance of NCC-ECM interactions and the initial fibronectin distribution in migration within the context of the ABM: namely, high density regions of fibronectin attract cells regardless of whether NCCs secreted it.

Cells migrate shorter distances in more sparse fibronectin matrices We further investigate the importance of the initial fibronectin distribution by increasing the spacing between pre-existing fibronectin puncta (this is done by increasing the parameter λ_{FN} from 20 μm to 60 μm). Figure 5.4(a) presents the resulting distributions for the invasion distance at $t = 720$ min. We observe that cells travel significantly shorter distances on average in a more sparse fibronectin lattice when compared to the baseline case (although collective migration still occurs). Distributions for other summary statistics, such as the lateral dispersion of NCCs and the percentage of cells outside the target corridor (data not shown),



(a) Violin plots, invasion distance

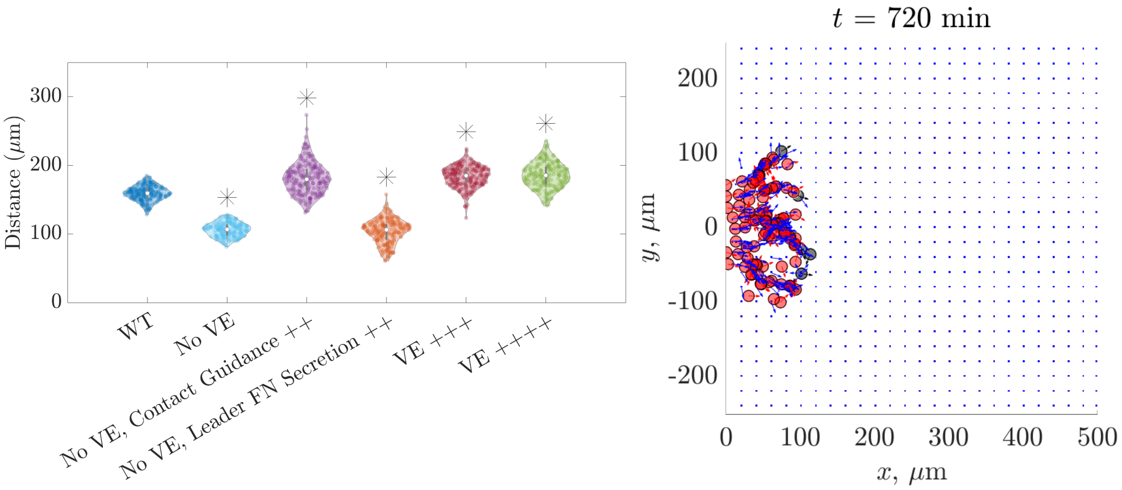
(b) More sparse fibronectin lattice with leader fibronectin secretion ++

Figure 5.4: (a) Violin plots for the NCC invasion distance at $t = 720$ min, computed for *in silico* experiments in which cells migrate through a more sparse fibronectin lattice ($\lambda_{\text{FN}} = 60$ μm) than in the baseline parameter regime. (b) A realisation in which NCCs migrate through a more sparse fibronectin lattice, but leader cells secrete fibronectin more often ($\lambda_{\text{FN}} = 60$ μm, $T_{\text{ave}} = 10$ min). Abbreviations: WT (baseline parameter regime), Sparser FN Lattice, CG ++ ($\lambda_{\text{FN}} = 60$ μm, $\rho = 0.33$), Sparser FN Lattice, Leader FN ++ ($\lambda_{\text{FN}} = 60$ μm, $T_{\text{ave}} = 10$ min), Sparser FN Lattice, VE ++++ ($\lambda_{\text{FN}} = 60$ μm, $c_i = 5$). See §D.2 for the interpretation of black asterisks in (a). We refer to the caption of Figure 5.1 for a description of the objects (circles, squares, etc.) found in (b).

indicate that cells travel shorter distances in directions lateral to the target corridor as well. We conclude that if NCCs are less likely to sense fibronectin, then they travel smaller distances (both towards, and lateral from, the target corridor) – at least within the context of the ABM.

These observations motivate the next *in silico* experiments, also shown in Figure 5.4(a), which investigate whether NCC can travel distances as far as the baseline case – even when the fibronectin lattice is more sparse – by altering the fibronectin secretion rate of leader NCCs, the bias of the NCC-ECM force towards contact guidance, and the strength of volume exclusion forces between cells. All three sets of perturbations increase the distance cells travel, albeit to different degrees. Upregulating contact guidance, for instance, causes the average distance cells travel to exceed that of the baseline case. Although increasing the volume exclusion strength, or the rate of leader cell fibronectin secretion, can also cause NCCs to travel farther distances in sparser fibronectin lattices, the average values of these metrics are smaller than those of the baseline case. Figure 5.4(b) presents a snapshot of the experiment in which cells secrete fibronectin at a higher rate but travel through a sparser lattice. We observe that several branches have formed by this time point, which are each led by at least one leader cell. Streams do not seem to favour a particular direction, as these branches migrate either horizontally or vertically. We have found that the distributions of other summary statistics, such as the lateral migration of cells (data not shown), exhibit similar trends as those obtained from the invasion distance.

Volume exclusion forces are not required for collective migration to occur The ABM can also be used to anticipate how simplifications to the theoretical framework will affect the invasion distance. As an example, we investigate whether volume exclusion is required for cells to successfully colonise the target corridor by perturbing c_i, the ABM parameter related to the strength of volume exclusion forces. Figure 5.5(a) presents the resulting distributions for the invasion distance at $t = 720$ min. The graph indicates that cells migrate along the target corridor even when there is no cell-cell repulsion, although they travel significantly shorter distances than in the baseline scenario. We confirm this in



(a) Violin plots, invasion distance

(b) No volume exclusion

Figure 5.5: (a) Violin plots for the NCC invasion distance at $t = 720$ min, computed for *in silico* experiments that alter the strength of volume exclusion forces. Abbreviations: WT (baseline parameter regime); No VE (No volume exclusion, set $c_i = 0$); No VE, Contact Guidance ++ (set $c_i = 0$, $\rho = 0.33$ in both cell types); No VE, Leader FN Secretion ++ (set $c_i = 0$, $T_{\text{ave}} = 10$ min); VE +++ (set $c_i = 2$ for both cell types); VE ++++ (set $c_i = 5$ for both cell types). (b) A typical realisation at $t = 720$ min for an experiment in which there is no volume exclusion (cell-cell repulsion). See §D.2 for the interpretation of black asterisks in (a). We refer to the caption of Figure 5.1 for a description of the objects (circles, squares, etc.) found in (b).

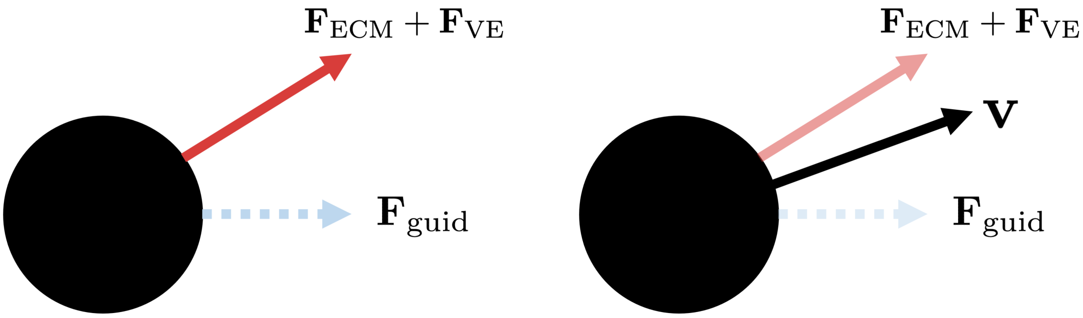
Figure 5.5(b), which presents a single ABM realisation of this experiment. We observe that both leader and follower cells infiltrate the domain (reaching a maximum distance of just over 100 μm), that many cells overlap (agreeing with our expectation), and that multiple finger-like branches led by one or two leader cells appear to form.

By contrast, if the NCCs experience stronger volume exclusion forces than in the baseline scenario, then Figure 5.5(a) indicates that they migrate significantly further distances. Similar trends are observed for the lateral dispersion of cells and the percentage of cells outside the target corridor (data not shown). This suggests that, at least within the context of the ABM, adding complexity to the theoretical framework by including volume exclusion forces can increase the distance that NCCs migrate (both along, and lateral to, the target corridor), but it is not required for collective migration to occur.

We now perform two experiments to assess whether perturbations to NCC-ECM interactions can cause cells to travel as far as in the baseline scenario, even when they experience no volume exclusion. For the first test, we increase the rate at which leader NCCs secrete fibronectin; for the second, we increase the persistence of cell movement along ECM fibres by decreasing ρ, the parameter that determines what proportion of the NCC-ECM force is directed towards the (stochastically sampled) haptotaxis cue. The resulting distributions in Figure 5.5(a) indicate that increasing the rate at which NCCs secrete fibronectin does not greatly increase the distance cells migrate, while upregulating contact guidance causes NCCs to travel further than in the baseline case. Similar trends are observed for the lateral dispersion of cells and the percentage of cells outside the target corridor (data not shown). We conclude that we do not need to introduce additional complexities into the ABM to ensure that NCCs migrate as far as in the baseline scenario – it is sufficient to simply adjust parameters related to cell persistence/contact guidance. This result may be useful in guiding future simplifications for the ABM framework.

5.3 Extending the ABM to account for directed leader cell migration

In the previous section, the only forces acting on the NCCs arise from volume exclusion and/or NCC-ECM interactions. The summary statistics reveal, however, that cells in the ABM exhibit excessive lateral migration: in the baseline parameter regime, for instance, the average lateral dispersion of NCCs is almost twice as wide as the neural tube corridor from which they migrate, whereas *in vivo* streams typically maintain a constant width from the neural tube (Szabó et al., 2016; Kulesa et al., 2004; Kulesa and Fraser, 1998). Such discrepancies with experimental data motivate us to extend the ABM by introducing a new mechanism that steers leader and/or follower NCCs along their target corridor (see Figure 5.6). This mechanism may represent processes previously neglected in the ABM, such as chemotaxis (McLennan et al., 2012, 2010) or mechanical-based cues (McKinney et al., 2020; Shellard and Mayor, 2019); as a coarse first approximation, we assume that its force, $\mathbf{F}_{\text{guid}}$ (henceforth generically referred to as the "guiding force"), simply directs NCCs along the target corridor. Since we take the target corridor to be perpendicular to the neural tube,



(a) $\mathbf{F}_{\mathrm{guid}}$, $\mathbf{F}_{\mathrm{ECM}}$, and $\mathbf{F}_{\mathrm{VE}}$ (b) Resultant velocity $\mathbf{v}$

Figure 5.6: (a) The guiding force $\mathbf{F}_{\mathrm{guid}}$ (light blue dashed arrow) and the net force from the volume exclusion and NCC-ECM forces (red solid arrow). (b) The total cell velocity ($\mathbf{v}$, black solid arrow) is obtained through a linear combination of these vectors, as indicated in Eq.s (5.3) and (5.4).

which is along the y-axis, the guiding force is directed along the x-axis such that

$$\hat{\mathbf{F}}_{\mathrm{guid}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{5.1}$$

where $\hat{\mathbf{F}}_{\mathrm{guid}}$ denotes the unit guiding force vector.

We use Newton's second law to model the dynamics of a given cell i in the ABM. The forces that act on the cell include a frictional force (which is assumed to be proportional to the cell velocity), a force from all cell-ECM interactions, $\mathbf{F}_{\mathrm{ECM}}$, a force from volume exclusion, $\mathbf{F}_{\mathrm{VE}}$, and the guiding force, $\mathbf{F}_{\mathrm{guid}}$, such that the cell dynamics are given by

$$m\frac{\mathrm{d}^2\mathbf{x}_i}{\mathrm{d}t^2} = -\eta\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} + \mathbf{F}_{\mathrm{ECM}} + \mathbf{F}_{\mathrm{VE}} + \mathbf{F}_{\mathrm{guid}}, \tag{5.2}$$

where m denotes the cell mass and η the cell coefficient of friction. As in Chapter 4, we assume that the mass and acceleration of a given cell i are much smaller than the frictional coefficient, velocity, and external forces acting on the cell. We may thus ignore the left-hand side of Eq. (5.2) and thus simplify Newton's law to the leading order overdamped equation

$$\mathbf{v}_i := \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} = (\mathbf{F}_{\mathrm{ECM}} + \mathbf{F}_{\mathrm{VE}}) + \mathbf{F}_{\mathrm{guid}} = \mathbf{F}_{\mathrm{orig}} + \mathbf{F}_{\mathrm{guid}}, \tag{5.3}$$

where $\mathbf{v}_i$ denotes the velocity of cell i. In Eq. (5.3), we have abused notation by subsuming a factor of η^{-1} into the forces. Although this means that the quantities on the right-hand side no longer have the same units as in Eq. (5.2) – and now represent components of the cell velocity – we will continue to refer to them as "forces". Additionally, in the equation above we have grouped $\mathbf{F}_{\text{ECM}}$ and $\mathbf{F}_{\text{VE}}$ into a net term $\mathbf{F}_{\text{orig}}$, which represents all of the forces from the "original" version of the NCC ABM presented in Chapter 4.

Since we cannot determine the magnitude of the guiding force *a priori*, in practice we generalise Eq. (5.3) in order to examine how the magnitude of the guiding force, relative to $\mathbf{F}_{\text{orig}}$, alters cell trajectories. We do this by assuming that the guiding force only alters the direction, but not the speed, with which cells migrate. We introduce a new user-specified parameter $z \in [0, 1]$, which measures the degree to which the cell velocity is aligned along the direction of the guiding force. The unit cell velocity, $\hat{\mathbf{v}}_i$, is then a linear combination of the unit forces $\hat{\mathbf{F}}_{\text{orig}}$ and $\hat{\mathbf{F}}_{\text{guid}}$, weighted by z:

$$\hat{\mathbf{v}}_i = \frac{(1-z)\hat{\mathbf{F}}_{\text{orig}} + z\hat{\mathbf{F}}_{\text{guid}}}{\|(1-z)\hat{\mathbf{F}}_{\text{orig}} + z\hat{\mathbf{F}}_{\text{guid}}\|}, \tag{5.4}$$

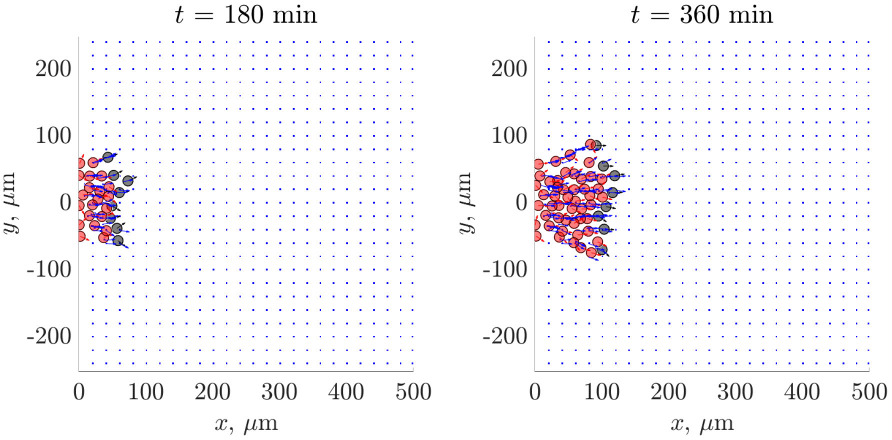
where $\|\cdot\|$ denotes the Euclidean norm. The speed of cell i, $\|\mathbf{v}_i\|$, is given by the norm of $\mathbf{F}_{\text{orig}}$, even for cases in which $z > 0$. Thus $\mathbf{v}_i = \|\mathbf{v}_i\|\hat{\mathbf{v}}_i = \|\mathbf{F}_{\text{orig}}\|\hat{\mathbf{v}}_i$ – note that this is the same speed that the cells have using the "original" version of the NCC ABM presented in earlier sections of this chapter and in Chapter 4.

When $z = 0$, cells move identically to those from the ABM in Chapter 4. When $z = 1$, however, cells migrate along the x-axis. Thus, the weight $z \in [0, 1]$ controls the extent to which the guiding force affects cell trajectories. By adjusting the value of z, we can determine how different "strengths" of the guiding force affect key features of collective migration, such as the invasion distance and lateral dispersion of cells. This is the focus of the experiments in the remainder of this section. We consider cases for which both leader and follower cells are affected by the guiding force, in addition to those for which only leader cells respond to changes in the value of z – this second set of tests aims to determine whether leader cells can guide the entire migrating collective, even if follower cells can otherwise move laterally from the target corridor.

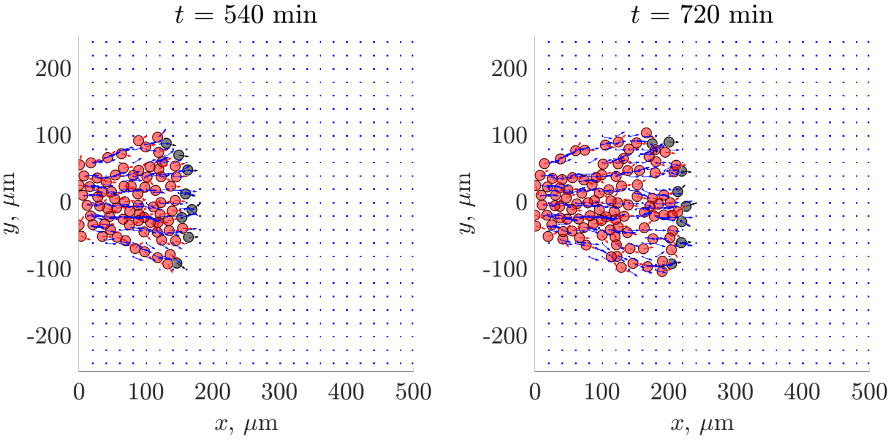
We stress again that this implementation of a guiding force is deliberately simplistic. The unidirectional nature of the guiding force, along with the user-defined weight z, allows us to gain preliminary insight into how other mechanisms, such as chemotaxis or mechanicallybased forces, might impact collective cell migration. If we were to model these forces in more detail, then the guiding force must also necessarily change to match *in vivo* observations of the specific mechanism under investigation. For example, we could make the value of the weight z dependent on the concentration of substrates that constitute the ECM; this would represent the fact that cells must exert traction forces on the ECM in order to migrate. Similarly, if the guiding force were to only correspond to chemotaxis, then we would make its direction dependent on the gradient of chemoattractant that the cell senses within a neighbourhood around its cell body. For the purposes of this chapter, however, the guiding force presented in Eq.s (5.3)-(5.4) is sufficient for further differentiating leader cells from follower cells and provides a means to characterise mechanisms that allow for robust migration. We therefore do not consider other functional forms for this force in the remainder of the chapter.

Figure 5.7 presents snapshots of an ABM simulation in which the guiding force acts on leader and follower cells (with $z = 0.5$). We observe that cell movement is predominantly horizontal, and that distinct cell branches do not form within the time period shown. The guiding force also reduces migration lateral to the target corridor: the lateral dispersion of NCCs at $t = 720$ min, for instance, is equal to 207 μm – approximately 50 μm (3.33 cell diameters) less than its median value when there is no guiding force (see Figure 5.1(b)). As we would expect, NCCs are also spaced roughly one cell diameter apart from each other due to the volume exclusion cue (the average nearest distance between cells at $t = 720$ min is 14 μm or, equivalently, just under one cell diameter). Taken together, these results indicate that adding the guiding force causes NCCs to migrate along the target corridor in a more focussed manner than for the baseline parameter regime.

When only leader cells sense the guiding force, however, there are extreme cases for which the leader cells separate from follower cells. Figure 5.8 presents snapshots of an ABM simulation in which the guiding force, with weight $z = 0.75$, acts only on leader cells. We see that, although leader and follower cells appear to migrate collectively for early times (e.g., at $t = 180$ min), leader cells eventually split away from follower cells. Cells also exhibit more chain-like migration: leader cells quickly branch off into two groups that are



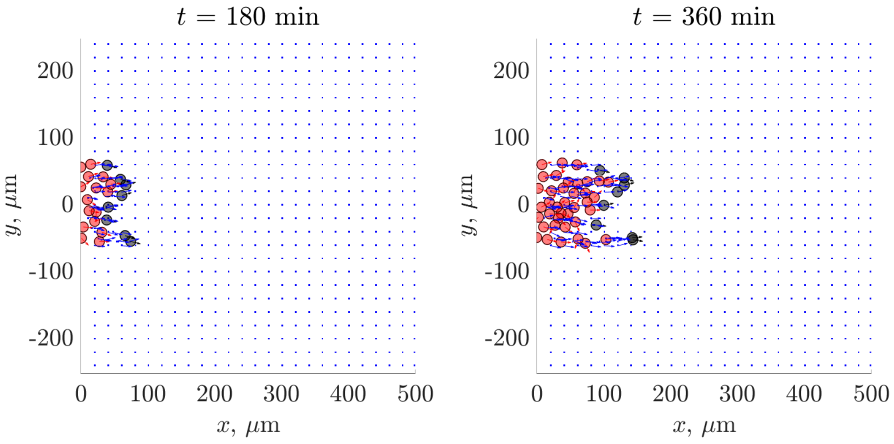
(a) $t = 180$ min (b) $t = 360$ min



(c) $t = 540$ min (d) $t = 720$ min

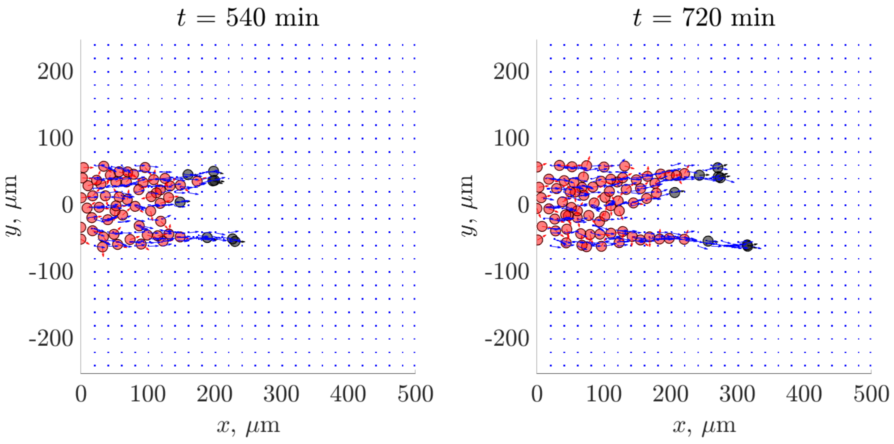
Figure 5.7: Snapshots of a typical ABM simulation in which leader and follower NCCs experience a guiding force, with $z = 0.5$, at times (a) $t = 180$, (b) $t = 360$, (c) $t = 540$, and (d) $t = 720$ minutes. Black circles denote leader cells, red circles denote follower cells, and blue squares mark the locations of fibronectin puncta and/or fibres. Arrows denote either the NCC velocity direction (red or black) or the fibronectin fibre orientation (blue).

one to two cells in width. The lateral migration of cells, however, remains less than that of the baseline case (the lateral dispersion of cells at $t = 720$ min, for instance, is 122 μm – only 2 μm larger than the height of the neural tube from which NCCs emerge). Follower cells migrate along the paths of leader cells, but the separation between the two cell types sometimes exceeds the distance of a cell diameter. These results suggest that if only leader cells are sensitive to the guiding force, then they may become separated from follower cells. This is unsurprising, given that other hybrid ABMs for neural crest cell migration have found that stream breaks can occur if leader cells exhibit more directed movement than



(a) $t = 180$ min

(b) $t = 360$ min



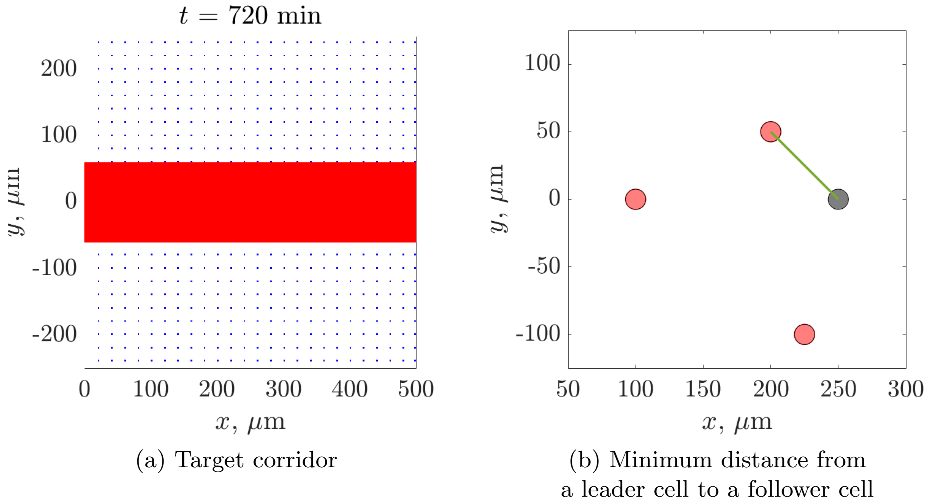
(c) $t = 540$ min

(d) $t = 720$ min

Figure 5.8: Snapshots of a typical ABM simulation for which only leader NCCs experience a guiding force (with $z = 0.75$) at times (a) $t = 180$, (b) $t = 360$, (c) $t = 540$, and (d) $t = 720$ minutes. Black circles denote leader cells, red circles denote follower cells, and blue squares mark the locations of fibronectin puncta and/or fibres. Arrows denote either the NCC velocity direction (red or black) or the fibronectin fibre orientation (blue).

follower cells (Giniunaite et al., 2019b). Unlike previous studies, however, our aim for the remainder of this chapter is to characterise when stream breaks occur in the ABM and to systematically identify combinations of mechanisms that avoid their onset.

To confirm that gaps between leader and follower cells persist over multiple ABM realisations, we compute summary statistics for the invasion distance, the lateral dispersion of cells, the number of cells outside the target corridor (which is defined to be the absolute number of cells whose centres lie outside a rectangle of width 500 μm and height l_{entr} centred at $y = 0$, as depicted in Figure 5.9(a)), and the size of the gaps between leader



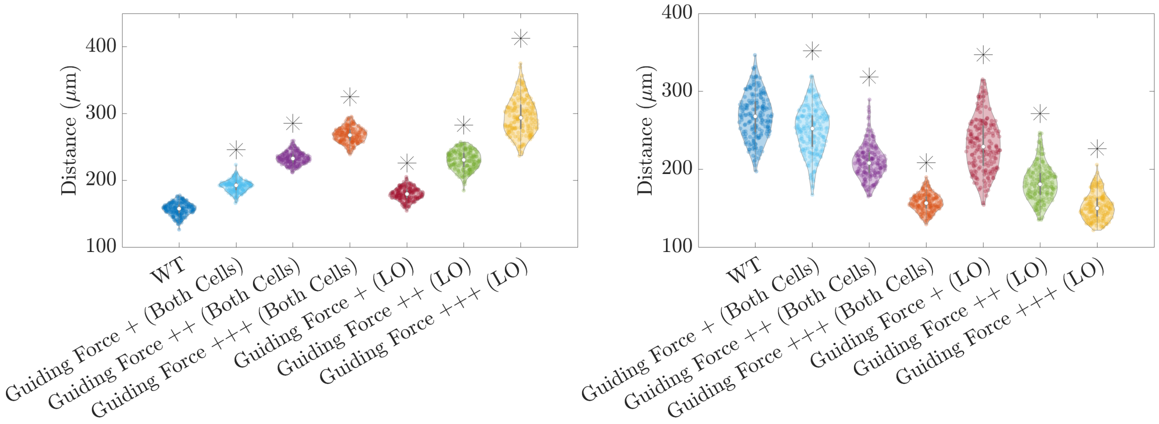
(a) Target corridor

(b) Minimum distance from a leader cell to a follower cell

Figure 5.9: (a) The target corridor (red rectangle), and (b) the nearest distance (solid green line) from a leader NCC (black circle) to any follower cell (light red circle). The height of the target corridor coincides with that of the neural tube from which cells emerge.

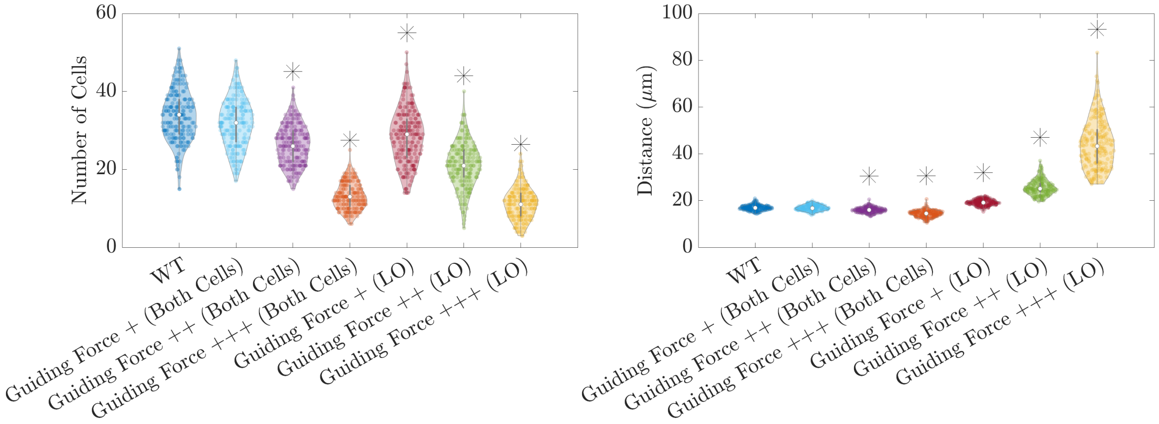
and follower cells. We approximate the size of gaps between leader and follower cells by calculating the average minimum distance from a leader cell to any follower, as depicted in Figure 5.9(b) for a configuration of NCCs with a single leader cell. While this is a crude metric for stream breaks, we have found that it can adequately determine cases for which leader cells are separated from follower cells (see below).

In Figure 5.10 we present the distributions of the four summary statistics for experiments in which we perturb the "strength" of the guiding force (by changing the value of z). We observe in Figure 5.10(a) that increasing the guiding force strength causes the cells to migrate longer distances along the target corridor, that this relationship appears to be linear with respect to z, and that it does not seem to matter whether leader cells – or both cell types – are affected by the guiding force (both sets of simulations yield similar distributions for each value of z considered). By contrast, in Figure 5.10(b) we observe that the lateral cell dispersion decreases as the guiding force strength increases. For smaller values of z, experiments for which only leader cells are affected by the guiding force generally yield smaller ranges than those for which it acts on both cell types. This suggests that cells exhibit less lateral movement when only leader cells sense a guiding force. This conjecture is further corroborated by the distributions of the number of cells outside the target corridor.



(a) Invasion distance

(b) Lateral dispersion of NCCs



(c) Number of cells outside the target corridor

(d) Average nearest distance from leader cells to follower cells

Figure 5.10: Violin plots for (a) the NCC invasion distance, (b) the lateral dispersion of NCCs, (c) the number of NCCs outside the target corridor, and (d) the average minimum distance from leader cells to a follower cell at $t = 720$ min, computed for *in silico* experiments that increase the strength of the guiding force by increasing the parameter value z. Abbreviations: WT (baseline parameter regime with $z = 0$), Guiding Force + ($z = 0.25$), Guiding Force ++ ($z = 0.5$), Guiding Force +++ ($z = 0.75$), LO (only leader NCCs affected). See §D.2 for the interpretation of the black asterisks.

The results, presented in Figure 5.10(c), mirror those for the previous metric: increasing the guiding force strength decreases the number of cells that lie outside the target corridor and, for each value of z, scenarios in which the cue only acts on leader cells exhibit smaller values than when both cell types are affected. This seemingly counter-intuitive observation may result from increased separation between the two cell types when the guiding force acts only on leader cells: since there are fewer cells, the follower cells are less likely to experience volume exclusion forces that propel them laterally from the target corridor.

Figure 5.10(d) confirms that the average minimum distance from leader cells to follower cells increases with the guiding force strength for scenarios in which only leader cells are sensitive to it (the average minimum distance when $z = 0.75$, for instance, is more than 30 μm – or two cell diameters – long). When the guiding force acts on both cell types, however, the increase in this distance is less pronounced (in fact, it appears to decrease as z increases). These results indicate that leader cells will separate from follower cells when the guiding force dominates their migration.

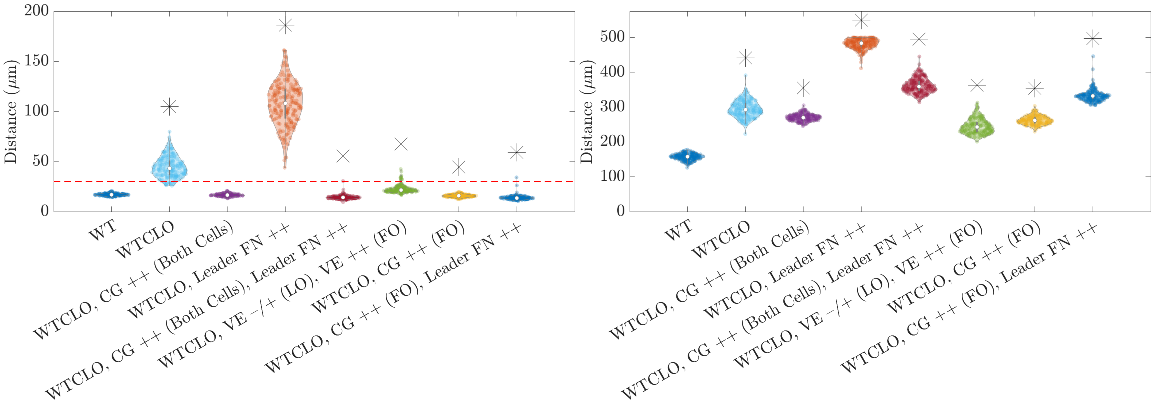
We also computed the four summary statistics for experiments in which the guiding force acts on leader cells invading a more sparse fibronectin lattice (with spacing $\lambda_{\mathrm{FN}} = 60$ μm between puncta), and for cases in which NCCs do not alter the initial fibronectin matrix. Our interpretation of the results (data not shown) is similar to those presented above: namely, increasing the guiding force strength increases the mean distance that NCCs travel in the direction of the target corridor but reduces the number of cells that travel laterally. When the guiding force dominates leader cell trajectories, however, the average minimum distance between leader cells and follower cells becomes greater than two cell diameters – indicating that leader cells separate from follower cells. Only when NCCs are unable to remodel the initial fibronectin matrix do the findings differ from the baseline case: we find that increasing the guiding force strength for both cell types tends to increase, rather than decrease, the number of cells outside the target corridor. The absolute number of cells that lie outside the target corridor, however, is smaller than that of the baseline case, and we conclude that this relationship is largely an artefact of the smaller number of NCCs that migrate when compared to the baseline case.

5.4 Contact guidance largely determines the robustness of collective migration in the ABM

In this section, we aim to identify mechanisms that prevent leader-follower separation in the ABM. In so doing so, we infer those mechanisms that most confer robustness to NCC migration. The sensitivity analysis from Chapter 4 and results in §5.2 suggest several potential candidate mechanisms. Given that ABM summary statistics tend to be most sensitive to ρ, the proportion of the NCC-ECM force that is directed towards the haptotaxis or contact guidance cues, we perform *in silico* experiments in which cells upregulate the latter mechanism. The results in §5.2 also show that increased fibronectin secretion by NCCs promotes cell clustering; consequently, we focus on cases in which leader and/or follower cells secrete fibronectin more often than in the baseline case. Finally, we investigate whether altering the strength of volume exclusion forces can reduce the distances between the two cell types. These "rescue" experiments all apply to the worst-case scenario from the previous section (in which the guiding force acts only on leader cells, with $z = 0.75$).

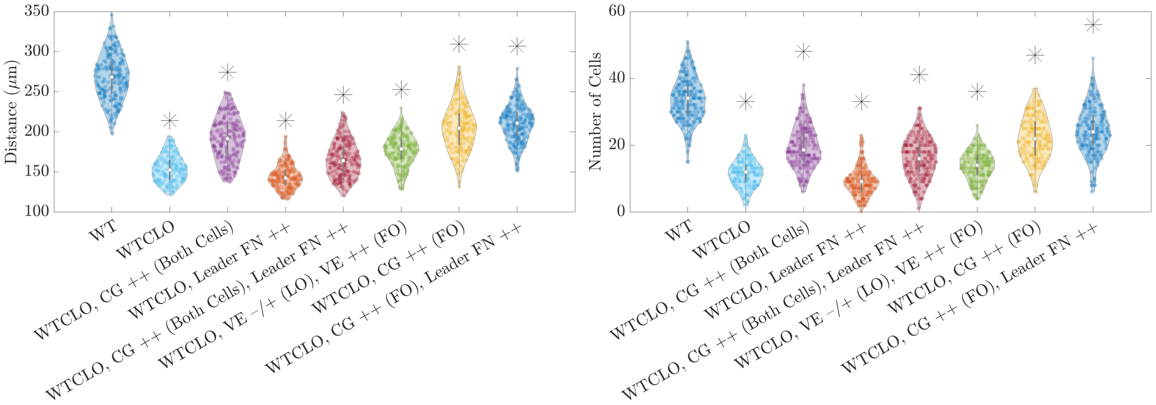
We define a "successful" rescue experiment as one for which the average minimum distance from leader cells to follower cells is less than two cell diameters (30 μm). We choose this threshold because, in the ABM, our choice of filopodial length means that cells cannot interact with each other beyond this distance; furthermore, values above this threshold yield noticeable empty spaces between the cells. Figure 5.11(a) presents distributions of the average minimum distance from leader cells to follower cells for several rescue experiments. The only scenario in this plot that fails to prevent leader-follower separation is where leader cells secrete fibronectin more rapidly than they do in the baseline case. We have also found that increasing, or reducing, the strength of volume exclusion forces in both cell types does not cause the distance between cell types to fall below 30 μm (data not shown).

Tables 5.3 and D.2 summarise the results from all of the parameter combinations that we tested. There is a clear trend in the former table: all but four of the successful rescue experiments upregulate contact guidance (in either follower cells or both cell types). In the four experiments that do not involve upregulation of contact guidance, follower NCCs



(a) Average nearest distance,
leader cells to follower cells

(b) Invasion distance



(c) Lateral dispersion of NCCs

(d) Number of cells outside the
target corridor

Figure 5.11: Violin plots for (a) the average nearest distance from leader cells to follower cells, (b) the invasion distance, (c) the lateral dispersion, and (d) the number of NCCs outside the target corridor at $t = 720$ min for different rescue experiments. Abbreviations: WT (baseline parameter regime), WTCLO (baseline parameters with a guiding force of weight $z = 0.75$ acting only on leader cells), Leader FN ++ ($T_{\text{ave}} = 10$ min), CG ++ ($\rho = 0.33$), VE ++ ($c_i = 2$), VE −/+ ($c_i = 0.05$), FO (change only affects follower cells), LO (change only affects leader cells). The dashed red line in (a) denotes a distance equivalent to two cell diameters (30 μm). See §D.2 for the interpretation of the black asterisks.

	CG ++ in both cell types	**CG ++ in follower cells**	**Weaker VE force in leader cells**
Contact guidance (CG) ++ in both cell types ($\rho = 0.33$)	✓	n/a	✓
Contact guidance ++ in follower cells ($\rho = 0.33$)	n/a	✓	✓
Leader fibronectin secretion ++ ($T_{\text{ave}} = 10$ min)	✓	✓	X
Follower fibronectin secretion + ($T_{\text{ave}} = 90$ min)	✓	✓	X
Weaker volume exclusion (VE) force in both cell types ($c_i = 0.05$)	✓	✓	n/a
Stronger volume exclusion force in both cell types ($c_i = 2$)	✓	✓	n/a
Weaker volume exclusion force in leader cells ($c_i = 0.05$)	✓	✓	✓
Stronger volume exclusion force in leader cells ($c_i = 2$)	✓	✓	n/a
Weaker volume exclusion force in follower cells ($c_i = 0.05$)	✓	✓	X
Stronger volume exclusion force in follower cells ($c_i = 2$)	✓	✓	✓
Less noisy haptotaxis cue in both cell types ($\gamma_{\text{FN}} = 3000$)	✓	✓	✓
Less noisy haptotaxis cue in follower cells ($\gamma_{\text{FN}} = 3000$)	✓	✓	✓

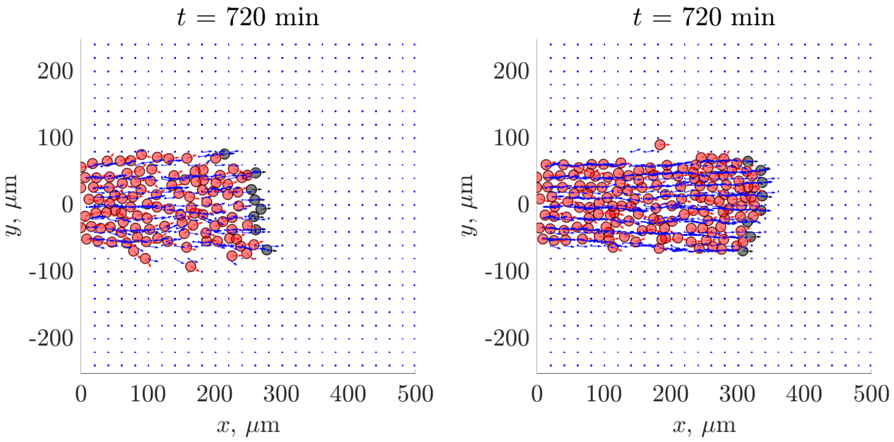
Table 5.3: Results of ABM parameter combinations that intend to prevent leader-follower separation. Green boxes with a "✓" denote combinations for which the average minimum distance from leader cells to follower cells is less than two cell diameters (30 μm), orange boxes with a "X" denote experiments that fail to reduce the distance below this threshold, and boxes with "n/a" denote tests that have not been performed. Unless otherwise noted in the vertical labels, ABM parameter values are the same as in Table 5.1. For each test, a guiding force with weight $z = 0.75$ also acts only on leader cells. See Table D.2 for the full table of experiments. Abbreviations: CG = contact guidance, VE = volume exclusion.

experience stronger volume exclusion forces than leader cells. These results suggest that contact guidance is a key determinant of robustness in NCC migration, because adjusting parameter values related to this mechanism nearly always avoids leader-follower separation from occurring. Separation can also be avoided if follower cells have greater strengths of cell-cell repulsion than leader cells, however this mechanism appears to be less robust than contact guidance because it comprises a smaller subset of successful rescue experiments.

We can also attain successful rescue experiments by changing the behaviour of only the follower cells. This result is interesting, because it is consistent with the earlier assertion that NCCs may specialise, adopting certain roles according to their location. Specifically, these data are consistent with a scenario in which leader NCCs steer the stream by upregulating mechanisms related to ECM remodelling and the sensing of external cues (such as those arising from chemotaxis), while follower cells upregulate mechanisms such as contact guidance to migrate more persistently along leader-created trails.

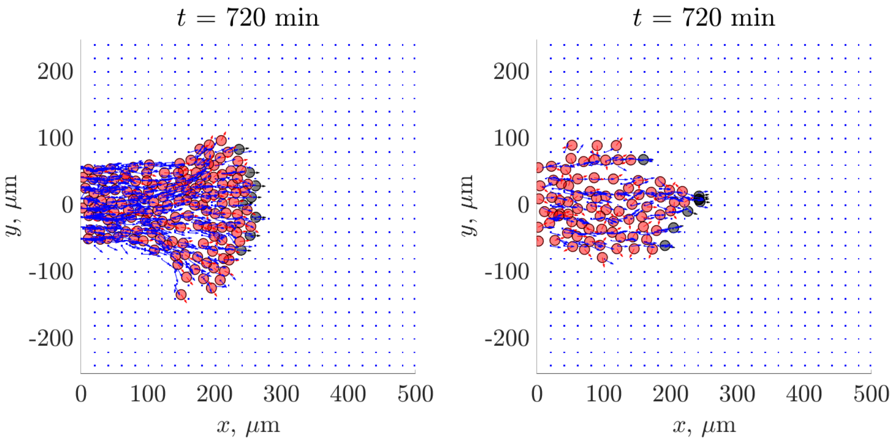
Figure 5.11(b) shows the corresponding distributions for the invasion distance statistic. Values of this metric lie within the range 250–400 μm for most experiments, with the largest distances corresponding to scenarios in which leader cells secrete fibronectin more rapidly than in the baseline case. By contrast, the smallest invasion distances correspond to experiments in which leader and follower cells exert different strengths of cell-cell repulsion (volume exclusion) forces. These results imply that NCCs travel further distances when leader cells secrete fibronectin, and that increasing the strength of volume exclusion (cell-cell repulsion) forces may not always promote efficient colonisation of the target corridor.

Figure 5.11(c) presents distributions for the lateral dispersion of NCCs. Most of the experiments yield smaller values than those of the baseline parameter regime, except those that involve fibronectin secretion by follower cells (not shown). This suggests that fibronectin secretion by follower cells increases lateral migration. We confirm these observations by comparing distributions for the number of cells outside the target corridor (Figure 5.11(d)). Again, all of the rescue experiments yield smaller values than those of the baseline parameter regime, except those that admit follower fibronectin secretion (not shown). We shall explain this result in the ensuing paragraphs.



(a) Contact guidance ++
in both cell types

(b) Contact guidance ++
in both cell types,
leader fibronectin secretion ++



(c) Contact guidance ++
in both cell types,
follower fibronectin secretion +

(d) Weaker volume exclusion forces in
leader cells,
stronger volume exclusion forces in
follower cells

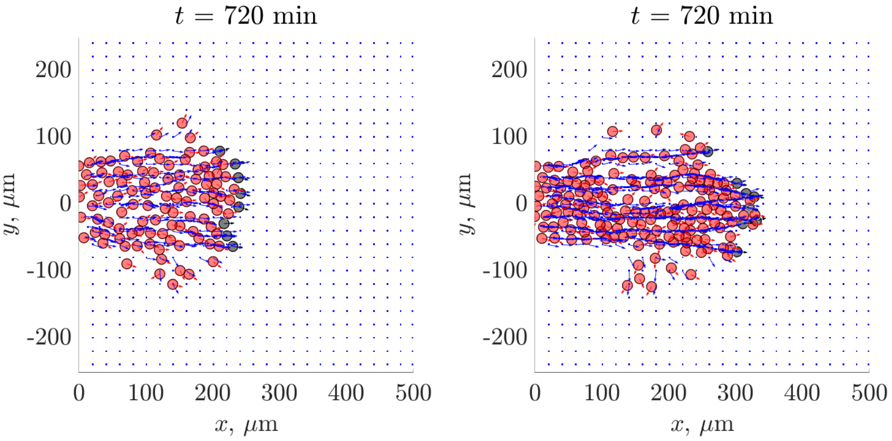
Figure 5.12: Typical snapshots of NCC and fibronectin distributions at $t = 720$ min for several experiments listed in Table 5.3. The graph in (a) corresponds to the case in which both cell types upregulate contact guidance (which is done by setting $\rho = 0.33$), that in (b) to a scenario for which both cell types upregulate contact guidance and leader cells secrete fibronectin more often ($\rho = 0.33$, $T_{\text{ave}} = 10$ min in leader cells), (c) to an experiment where both cell types upregulate contact guidance ($\rho = 0.33$) and follower cells secrete fibronectin with an average time interval of 90 min, and (d) to an experiment where follower cells experience stronger volume exclusion forces (by setting $c_i = 2$ in all follower cells), and leader cells experience weaker volume exclusion forces (by setting $c_i = 0.05$ in all leader cells) compared to the baseline parameter regime. Circles correspond to either leader cells (black) or follower cells (red). Blue squares denote fibronectin puncta and/or fibres. Arrows denote either the current cell velocity (red or black) or fibronectin fibre orientation (blue).

Figure 5.12 presents typical snapshots of the rescue experiments discussed above, so that we may see how these perturbations affect the patterns in which NCCs migrate. Figure 5.12(a) shows that increasing the bias of cells towards contact guidance creates a single ribbon-like stream of cells with approximately constant height and cell-cell separation. Figure 5.12(b) reveals that we obtain a qualitatively similar stream (albeit one that extends further into the domain) if we increase the rate at which leader cells secrete fibronectin. Cells likely travel greater distances in this case because they can travel at faster speeds when they sense fibronectin and move more persistently when more fibronectin fibres are created – both processes occur more often when leader cells secrete fibronectin more rapidly.

Although we have already noted that the lateral migration of NCCs increases when follower cells secrete fibronectin, it is unclear from the summary statistics whether multiple branches of NCCs form. Figure 5.12(c) indicates that this is likely not the case: in the plot, follower cells "fan out" over long distances, while cells closer to the neural tube travel in a narrower rectangular pattern. The snapshot suggests that volume exclusion is the main cause for this excessive lateral migration, since more follower cells are located near the front of the stream. Analysis of leader and follower speeds (not shown) corroborates this hypothesis: we find that, on average, leader cells travel with slower speeds in the horizontal direction than follower cells for this experiment. This would lead to a "pile up" of follower cells immediately behind leader cells; consequently, the follower cells experience stronger volume exclusion forces that propel them laterally from the target corridor.

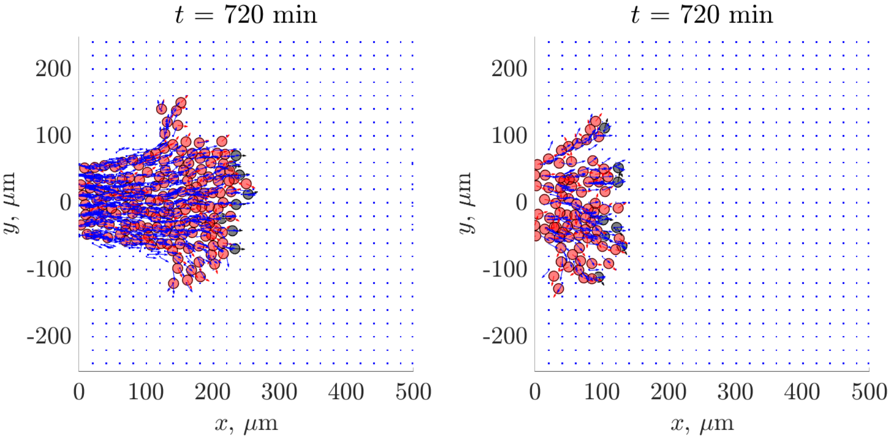
Figure 5.12(d) presents a simulation for which follower cells experience stronger volume exclusion forces than leader cells. Two branches of NCCs appear to form: one is led by a single leader cell and appears to be no more than two cells in width, while the other is larger and led by multiple leader cells. Within each branch, the distance between individual cells is roughly constant (except near the front where two leader cells overlap). This simulation demonstrates that different strengths of cell-cell repulsion between leader and follower cells can create stream-like patterns of migration, but also that additional mechanisms may be required to avoid the emergence of multiple branches and/or single-file cell chains.

Figure 5.13 presents snapshots of additional rescue experiments. Figure 5.13(a) shows



(a) Contact guidance ++
in follower cells

(b) Contact guidance ++
in follower cells,
leader fibronectin secretion ++



(c) Contact guidance ++
in follower cells,
follower fibronectin secretion +

(d) Weaker volume exclusion
force in leader cells

Figure 5.13: Typical snapshots of NCC and fibronectin distributions at $t = 720$ min for additional experiments listed in Table 5.3. The graph in (a) corresponds to a case in which follower cells upregulate contact guidance (by setting $\rho = 0.33$ in follower cells), (b) to a scenario in which follower cells upregulate contact guidance and leader cells secrete fibronectin more rapidly (by setting $\rho = 0.33$ in follower cells and $T_{\text{ave}} = 10$ min in leader cells), (c) to an experiment in which follower cells upregulate contact guidance (by setting $\rho = 0.33$ in follower cells) and secrete fibronectin with an average time interval of 90 min, and (d) to an experiment in which leader cells experience weaker volume exclusion forces (by setting $c_i = 0.05$ in leader cells). Circles correspond to either leader cells (black) or follower cells (red). Blue squares denote fibronectin puncta and/or fibres. Arrows denote either the current cell velocity (red or black) or fibronectin fibre orientation (blue).

an experiment for which follower cell migration is dominated by contact guidance. Most cells appear to travel along the paths of leader cells, however several follower cells also lie outside the target corridor. Indeed, the distribution plots from Figure 5.11 confirm that more cells lie outside the target corridor when only follower cells upregulate contact guidance than when both cell types upregulate that mechanism. This likely occurs via a process similar to that which arises when follower cells secrete fibronectin. If only follower cells upregulate contact guidance, then we expect them to migrate more persistently than leader cells. Consequently, follower cells move faster along the target corridor on average, which would lead to greater cell densities near the front. This, in turn, would cause follower cells to experience stronger volume exclusion forces that propel them to low cell density regions (located adjacent to the target corridor). A similar phenomenon appears to occur in Figure 5.13(b), which presents an experiment for which follower cells upregulate contact guidance and leader cells secrete fibronectin more rapidly: multiple follower cells lie outside the target corridor, although there generally appears to be a single stream.

When follower cells secrete fibronectin and upregulate contact guidance, Figure 5.13(c) reveals that NCCs “fan out” as they migrate along the target corridor. In the plot, a branch led by a single follower cell also forms perpendicular to the target corridor. The lateral migration of cells again appears to result from differences in the speeds with which leader and follower cells move (data not shown) – follower cells travel faster than leader cells, leading to a “pile-up” near the front that causes follower cells to experience stronger volume exclusion forces. This, in turn, propels follower cells lateral to the target corridor.

Finally, Figure 5.13(d) presents a snapshot for an experiment in which leader cells experience weaker volume exclusion forces than in the baseline parameter regime. We observe that multiple branches of NCCs tend to form in this experiment, and that multiple follower cells lie outside the target corridor. Additionally, cells travel smaller distances than in the other three plots of this figure, and there are several single-file streams that appear to be led by only a single leader cell. These results demonstrate how different strengths of cell-cell repulsion may create chain-like patterns of NCCs, rather than ribbon-like streams.

We verify that the success of rescue experiments is not affected by the fibronectin distribution by performing the same tests for cases in which cells migrate through a sparser fibronectin lattice (with $\lambda_{\mathrm{FN}} = 60\ \mu\mathrm{m}$). The resulting distributions of summary statistics (not shown) reveal that all of the successful ABM perturbations from Table 5.3 also prevent leader-follower separation for this new lattice. Additionally, we find that leader cells do not separate from follower cells when both cell types experience weaker volume exclusion forces. Most of the conclusions from the baseline case also remain unchanged: cells travel farthest when only leader cells secrete fibronectin, and all rescue experiments (except for those that exhibit fibronectin secretion by follower cells) reduce the number of NCCs outside the target corridor. Unlike in the baseline case, however, removing volume exclusion from the ABM causes NCCs to fail to colonise the target corridor (data not shown). This implies that when the fibronectin matrix is sparse, cells must push each other forward so that they may sense fibronectin ahead of the stream and successfully migrate.

Finally, we consider how these rescue experiments perform in situations where no cell can remodel fibronectin (either by secreting it or assembling it into fibres). The resulting distributions of summary statistics (not shown) reveals that none of the successful rescue experiments from Table 5.3 can prevent leader-follower separation if the cells are unable to alter the initial fibronectin matrix. Since cells still sense pre-existing FN puncta in these tests but do not sense any fibres (because fibres are not assembled), the only forces that act on cells arise from haptotaxis and volume exclusion. As contact guidance is not present in these simulations, and because cells fail to migrate, this strengthens our hypothesis that contact guidance plays an important role in robust collective cell migration.

5.5 Discussion and conclusion

In this chapter, we used the ABM from Chapter 4 to perform what is, to our knowledge, the first detailed study of mechanisms that promote robust chick cranial neural crest cell migration. By perturbing model parameters, we inferred how individual mechanisms for cell migration control macroscopic outcomes such as the furthest distance that cells travel. Based on these *in silico* experiments, we anticipate that, in the absence of chemotaxis and mechanical-based signals, knocking out fibronectin secretion in NCCs will reduce the distance they migrate, but will not entirely prevent their collective migration because cells will still sense and move along the pre-existing matrix. While increasing the fibronectin secretion rate causes NCCs to travel further distances along the target corridor, in extreme cases they can form chain-like patterns that are not observed *in vivo*, due to the effects of haptotaxis and contact guidance. When we prevent leader cells from remodelling the fibronectin matrix, NCCs fail to colonise the target corridor. By contrast, there are no significant changes from the baseline scenario if follower cells cannot remodel the ECM. We thus predict that follower cells do not need to remodel the ECM for successful NCC migration to occur. Although to our knowledge it is not presently possible to conduct some of these experiments *in vivo*, the results nevertheless demonstrate the importance of leader cells in establishing collective migration.

Alterations to the pre-existing fibronectin matrix also affect the success of neural crest cell migration in the ABM. We find, for instance, that cells preferentially travel towards, and remain within, high density regions of fibronectin – even when these regions are artificially placed adjacent to the target corridor along which cells normally migrate. We could test this prediction *in vivo*, by injecting exogenous fibronectin lateral to the target corridor and subsequently observing where NCCs migrate. We also found that making the pre-existing fibronectin matrix sparser in the ABM reduces the distances that NCCs travel. This prediction could also be tested *in vivo*: for instance, one could knock down fibronectin secretion within mesoderm and/or ectoderm cells to alter the fibronectin matrix before neural crest cells migrate from the neural tube.

The results also indicate that cell-cell repulsion (which we termed “volume exclusion”) may not be necessary for successful migration to occur in the ABM when the pre-existing fibronectin matrix is sufficiently dense, although they increase the distances that cells migrate (both along and lateral to the target corridor). While the implementation of cell-cell repulsion in this ABM is a mathematical abstraction that ensures cells do not intersect, in *Xenopus* embryos the molecular basis for cell-cell repulsion has been investigated (Shellard and Mayor, 2019; Roycroft et al., 2018; Bahm et al., 2017; Stramer and Mayor, 2016). Such studies may identify molecular targets for controlling cell-cell repulsion, which may help us verify some of the above hypothesis in the future.

Interestingly, we find that if there is no distinction between leader and follower cells, then cells travel shorter distances on average. These results suggest that NCCs migrate most persistently when leading and follower cells specialise in different roles, with leader cells guiding the NCC stream via fibronectin remodelling and follower cells upregulating mechanisms that allow them to better sense where leader cells have migrated. We expect many of these insights to hold even when other mechanisms (such as chemotaxis) are included in the ABM framework, although the magnitudes of such effects may become smaller.

These conjectures of phenotype specialisation, along with observations that NCCs migrate more laterally in the ABM than *in vivo*, motivated us to introduce a new guiding force that steers leader NCCs along the target corridor. We find that this guiding force, which represents the effects of chemotactic and/or mechanical-based cues in simple terms, reduces the lateral migration of all cells and increases the distance that NCCs migrate. In extreme cases, however, leader cells can separate from follower cells when the former cell type experiences strong guiding forces. Subsequent experiments in §5.4 demonstrate that leader-follower separation can be averted when cells upregulate contact guidance, or when leader and follower cells experience different strengths of volume exclusion forces. The resulting patterns of NCCs largely resemble the ribbon-like streams of *in vivo* experiments (Kulesa et al., 2004; Kulesa and Fraser, 1998). We further discovered that only follower cells need to be affected in the aforementioned way to prevent the onset of stream breaks, although such perturbations also increase the number of cells outside the target corridor. There are several mechanisms that could reduce the extent of such lateral movement: NCCs at the front of streams, for instance, could upregulate proteins that allow them to travel at faster speeds than follower cells (although this process must be tightly regulated to prevent leader-follower separation). Alternatively, an internal boundary could prevent follower NCCs from migrating laterally to the target corridor. It would be interesting to evaluate this hypothesis in chick cranial NCC migration, since molecules such as versican are already known to give rise to internal boundaries for *Xenopus* NCCs (Szabó et al., 2016).

Although we have used several summary statistics to characterise NCC migration, these represent very simple quantities such as the maximum distance travelled. Consequently, these metrics cannot supply very detailed information about the types of patterns that NCCs form (e.g., ribbon-like streams or single-file chains). Although simplistic metrics are appropriate here, given the aims of this chapter, it would be interesting to investigate other statistics that characterise NCC migration. For instance, spatial statistics such as pair correlation functions and Ripley's K-function can identify locations of cell clusters and quantify the length scales of different patterns (Bull et al., 2020; Lagache et al., 2013; Ripley, 1981). Alternatively, techniques from topological data analysis, such as persistent homology, can provide insights into the topological features of migrating NCC streams, including their connectedness, loops, and holes (Bhaskar et al., 2021; Stolz et al., 2017; Otter et al., 2015; Lum et al., 2013). Machine learning approaches may even identify new summary statistics for NCC migration: semi-supervised pipelines already exist to extract time-dependent features in gene-regulatory networks, for example (Wrede and Hellander, 2019). While the identification and evaluation of new summary statistics for NCC migration is outside the scope of this thesis, it represents an interesting avenue for future work.

This NCC ABM is deliberately simplistic so that it can connect cell-level perturbations to changes in macroscopic behaviour. As such, it does not completely describe NCC migration: it neglects, for instance, a more detailed representation of "guiding forces" arising from mechanisms such as chemotaxis (Tweedy et al., 2020; McLennan et al., 2012), phenotype switching between leader and follower cells (Schumacher, 2019), and domain growth (McKinney et al., 2020). Despite these deficiencies, we can use the experiments in this chapter to infer how such mechanisms will affect NCC migration. We expect domain growth, for instance, to extend the distance between existing fibronectin puncta (if no new ones are created) and make its distribution more sparse. Based on the *in silico* experiments in which cells migrate through a more sparse fibronectin environment, we anticipate that cells will travel smaller distances unless they upregulate mechanisms related to contact guidance and/or volume exclusion. Alterations to cell phenotypes may not be necessary, though, given that domain growth may also advect cells and cause them to travel larger distances.

If cells can switch phenotypes, then we expect more fibronectin secretion to occur near the front of migrating streams (since there will be more leader cells there). Based on observations from the *in silico* experiments in which leader cells secrete fibronectin more rapidly, we hypothesise that phenotype switching will increase the distance NCCs migrate – however, if NCCs switch phenotypes too quickly, then we expect cells will not travel as far because there will be practically no distinction between their phenotypes, leading to very large fibronectin densities within regions already traversed by cells. These conjectures demonstrate how a simple ABM framework can generate valuable insights into the possible interactions between cells, and provide testable hypotheses that can be evaluated *in vivo*.

Chapter 6

Discussion and future work

6.1 Thesis summary

We sought in this thesis to improve the understanding of mechanisms underpinning collective leader-follower cell migration and to resolve open questions regarding the mathematical models used to investigate it. To achieve these goals, we developed, simulated, and analysed models that describe two biological phenomena in which leader-follower migration occurs: angiogenesis and neural crest cell migration. Angiogenesis has historically attracted much interest from mathematical modellers, and a plethora of frameworks now exists to describe this process (Flegg et al., 2020; Scianna et al., 2013; Mantzaris et al., 2004). Consequently, the questions we addressed in the first part of the thesis (Chapters 2 and 3) were theoretical in nature, and focussed on determining conditions under which certain frameworks can be appropriately applied to study cell movement in angiogenesis. By contrast, the second part of the thesis (Chapters 4 and 5) focussed on biological issues concerning neural crest cell migration, and clarified the mechanisms that underpin this process. Specifically, we used mathematical modelling to investigate whether interactions between cells and their extracellular matrix (ECM) could support collective migration, and identified mechanisms that likely confer robustness to the system. As we make clear below, this thesis has extended the knowledge of leader-follower migration in angiogenesis and neural crest cell migration, and has raised interesting questions that could be addressed in future work.

In Chapter 2, we determined the conditions under which a classical “snail-trail” partial differential equation (PDE) model for angiogenesis could accurately describe column averaged results of an agent-based model (ABM) from Pillay et al. (2017), which incorporates a discrete version of the snail-trail assumption. We found that numerical solutions of the 2D snail-trail model consistently underestimated the ensemble average follower (stalk) cell distribution of the discrete model. This discrepancy arises because the phenomenological framework does not account for cell movement in directions other than that of the migrating front. We thus uncovered a key physical assumption that is implicitly invoked (without justification) when snail-trail models are used to describe angiogenesis and is, in certain circumstances, erroneous. Using a mean-field approach, we derived a corrective factor for the snail-trail stalk cell evolution equation that is valid when leader cell movement is dominated by chemotaxis. After confirming that this corrective factor causes the solutions of the 2D PDE model to accurately approximate those of the ABM, we then formally reduced the snail-trail model to 1D. In doing so, we determined what assumptions must be satisfied for spatially-reduced models to accurately describe this multidimensional process.

Although previous studies of snail-trail models have investigated qualitative differences between results from 2D and 1D systems for angiogenesis (Pillay et al., 2017; Orme and Chaplain, 1997), our work provided a more rigorous derivation of 1D snail-trail models, made explicit what assumptions must be satisfied for them to accurately capture column averaged discrete angiogenesis data, and identified regions of parameter space in which continuous theoretical frameworks of leader-follower migration, such as the snail-trail model, are valid descriptions of its underlying biological processes (here represented by the ABM). Furthermore, the results of this work suggested metrics, such as the average cell displacement, that may be used in the future to predict when such theoretical frameworks may be invalid and/or provide a poor approximation of the underlying biology.

After identifying the corrected version of the snail-trail model, in Chapter 3 we examined whether solutions from this phenomenological framework could accurately represent those from models derived using detailed discrete-to-continuum approaches. Specifically, we applied techniques from perturbation theory to investigate the relationship between the corrected snail-trail model and a fully nonlinear coarse-grained system for angiogenesis that was originally developed in Pillay et al. (2017). Although this analysis was non-standard, due to the complexity of the scaling of terms, we demonstrated that the solutions to both continuum models are identical at leading order if cells do not branch frequently and chemotaxis dominates leader cell migration. The shared leading order system explained why numerical solutions to both models are indistinguishable and supplied a reduced model for angiogenesis, which describes the same dynamics as both frameworks but is simpler to analyse and solve numerically. We constructed self-similar asymptotic solutions to the reduced model that are valid for early times. This work demonstrated how models constructed from phenomenological arguments can, in some sense, be "equivalent" to those derived from a coarse-graining approach and how one can exploit such similarities to identify simpler theoretical frameworks for leader-follower migration.

In contrast to the above chapters, which investigate theoretically motivated questions within the context of angiogenesis, the research into neural crest cell migration (Chapters 4 and 5) focussed on how mathematical models can provide insight into biological processes underpinning leader-follower migration. In Chapter 4, we developed a novel agent-based model for neural crest cell migration to investigate how cell interactions with the ECM influence collective cell behaviour. We incorporated a cell-reinforced migratory cue in which "leader" cells remodel an initially punctate and immature ECM, creating paths for trailing cells to follow. Within the ABM, the ECM influences cell trajectories in several ways: cells can nonlocally sense ECM molecules and fibres, move via haptotaxis up adhesive gradients, and align via contact guidance towards the orientation of fibres. We confirmed that such interactions, along with volume exclusion (cell-cell repulsion), could generate migrating streams of neural crest cells that resemble those observed *in vivo.* We then applied techniques from global sensitivity analysis to examine how changing certain parameter values alters summary statistics related to migration, such as the distance that cells travel and the average nearest neighbour distance. The analysis demonstrated that these metrics are most sensitive to parameters related to contact guidance, which suggests that this mechanism most influences collective cell migration – at least within the context of the ABM. Furthermore, this work highlighted the crucial role that cell-ECM interactions can play in establishing collective cell migration. This finding differs from those of other theoretical descriptions for neural crest cell migration (Khataee et al., 2021; Colombi et al., 2020; Painter et al., 2015; Woods et al., 2014; Wynn et al., 2013), which suggest that cell-cell interactions such as volume exclusion may be more important for general collective cell migration (i.e., with migratory patterns not restricted to ribbon-like streams) to occur. In our model, while volume exclusion forces can increase the distance that cells travel, cell-ECM interactions must be dominated by contact guidance rather than by haptotaxis for general long-distance migration to occur (otherwise, cells collectively travel shorter distances). This difference in conclusions may result from the fact that we incorporated a slightly different mechanism for volume exclusion than that of other models: in our ABM, for instance, distant cells do not necessarily repel each other upon contact.

In Chapter 5, we used the ABM to perform *in silico* experiments that investigate the roles of individual mechanisms for collective cell migration. These experiments largely supported the conclusions of the global sensitivity analysis but, unexpectedly, indicated that cell migration is most efficient when leader and follower cells exhibit unique phenotypes. In such a scenario, leader cells specialise in guiding the stream and remodelling the ECM. We tested this conjecture by adding a new mechanism to the modelling framework that steered leader cells along the correct target corridor. We found that, in extreme cases, leader cells could separate from follower cells and fail to establish ribbon-like coherent streams. Although other theoretical frameworks have also generated undesired breaks between leader and follower cells (Bernardi et al., 2021; Giniunaite et al., 2019b), we used the ABM to provide (to our knowledge) the first comprehensive study of biological mechanisms that can prevent leader-follower separation by systematically identifying different combinations of parameter values for which the distance between the two cell phenotypes is below a certain threshold. Nearly all successful combinations involved cells upregulating processes related to contact guidance, which further supports the importance of this mechanism in ensuring robustness of collective migration, at least within the context of the ABM.

6.2 Future work and perspectives

While this thesis has addressed many questions regarding collective cell migration in leader-follower systems and the validity of its theoretical frameworks, there are several limitations to the work. One major weakness is that we have relied on *in silico* data to evaluate different modelling frameworks and assess the roles of mechanisms involved in leader-follower migration. For instance, in Chapter 2 we used ensemble average distributions from an on-lattice ABM to determine whether 2D and 1D versions of the classical snail-trail model accurately describe tip and stalk cell densities over time and space. We did not, however, compare these solutions to densities obtained from *in vivo* datasets such as corneal angiogenesis assays (Nowak-Sliwinska et al., 2018). In spite of this, we note that such comparisons have already been performed using 1D versions of snail-trail models that have had their stalk cell evolution equations corrected using *ad hoc* approaches (Connor et al., 2015). In particular, that work demonstrated that 1D snail-trail models are accurate approximations of tip and stalk cell densities in corneal angiogenesis assays. Since we showed in Chapter 2 that the corrected snail-trail framework is equivalent to previous 1D models, and improved upon *ad hoc* approaches for estimating the corrective factor, we therefore expect that our 1D and 2D versions of the snail-trail model will also accurately approximate such *in vivo* data.

The lack of biological data also limits the impact of conclusions obtained from the neural crest cell ABM that we developed in this thesis. In Chapter 5, for example, we used *in silico* results from the framework to predict the outcomes of several biological experiments, but did not verify these hypotheses *in vivo* due to a lack of available lab space caused by the COVID-19 pandemic. We plan to address this in the future by conducting such experiments in consultation with our collaborators (Paul Kulesa and Mary McKinney, Stowers Institute for Medical Research, USA; Lance Davidson, University of Pittsburgh). Such investigations will greatly improve understanding of the mechanisms underpinning neural crest cell migration, regardless of their agreement with the *in silico* results: if an experiment generates results that are contrary to the conclusions from Chapter 5, for instance, then this suggests that the ABM framework is not equipped to accurately describe the process of interest, at least in its current state. This, in turn, may help guide improvements that can be made to the ABM by, for example, identifying mechanisms that must be modelled in greater detail.

As a particular example of how ABM results can inform the design of more detailed models, we highlight how the experiments in Chapters 4 and 5 indicate that summary statistics for collective migration are highly sensitive to parameter values related to contact guidance and cell-ECM interactions. While we can conclude from this that such mechanisms may play an important role in guiding cell migration, the ECM description in this ABM is overly simplistic to address more detailed questions. To do this, we require more comprehensive models of the ECM. These could, for example, explicitly incorporate a fibrous ECM geometry with crosslinks to investigate how such structures may impede cell migration and whether cells can overcome such obstacles via matrix metalloprotease secretion; we can take inspiration from previous studies in constructing such descriptions (Kato et al., 2021; Sfakianakis et al., 2020; Azimzade et al., 2019; Painter and Hillen, 2018; Painter, 2009; Hillen, 2006). Alternatively, we could develop a model that examines how the mechanical forces that cells exert on the ECM affect fibre alignment. While such interactions were not considered in this thesis, several models have attempted to account for them in other biological settings (Palsson, 2008; Dallon and Othmer, 2004; Palsson and Othmer, 2000). We may also describe fibre assembly in more detail, given that in the current version of the ABM, cells instantaneously create fibres directed towards their velocity. We could, for instance, represent fibre densities as a continuous variable, describe their evolution via differential equations, and use such values to weight the contact guidance cue when computing cell velocities (i.e., adjust Eq. (4.7) so that cells align more to denser fibres). Wershof et al. (2019) have considered a discrete version of this approach, which we could extend by considering continuous ECM deposition by cells and a wider array of fibre orientations. In so doing, we could investigate how the time scale of fibre assembly affects collective migration.

A further question we could address in future work is how model-specific implementations of cell-cell repulsion in a given ABM alter the long-term behaviour of migrating cells. Unlike models for neural crest cell migration based on contact inhibition of locomotion (CIL), the ABM used here suggests that the influence of cell-cell repulsion on the distance that cells travel along the target corridor depends on the mechanism(s) that dominate cell-ECM interactions. Since our approach to modelling volume exclusion differs from other studies (Khataee et al., 2021; Colombi et al., 2020; Szabó et al., 2016; Woods et al., 2014), it remains to be determined whether CIL also plays a context dependent role during neural crest cell (NCC) migration. We could address this future work by adjusting the rules of the NCC ABM to explicitly incorporate a CIL mechanism, rather than the volume exclusion-based one used in this thesis. By conducting a similar global sensitivity analysis to the one described in Chapter 4, we could establish whether CIL dominates collective migration to the same degree as mechanisms related to NCC-ECM interactions. Comparison of the two hybrid models may also guide the design of experiments that aim to determine the extent to which volume exclusion and/or CIL act on NCCs during their migration.

Another question raised by this thesis concerns the appropriate metrics for characterising collective cell migration. As noted in Chapter 3, theoretical frameworks may generate solutions that capture some features of biological data (for instance, by accurately describing how quickly cells travel, but underestimating cell densities in most of the spatial domain). Furthermore, in Chapters 4 and 5 we used simple metrics to evaluate the collective behaviour of cells, such as the maximum distance that they travel and the average nearest neighbour distance, however these could not distinguish between cases in which cells form patterns that resemble single-file chains or wider, ribbon-like streams. As discussed in Chapter 5, a variety of approaches could be used to measure collective behaviours in cells, including spatial statistics (Bull et al., 2020; Lagache et al., 2013; Ripley, 1981), topological data analysis (Bhaskar et al., 2021; Vipond et al., 2021; Stolz et al., 2017; Otter et al., 2015; Lum et al., 2013), semi-supervised learning methods (Wershof et al., 2019; Wrede and Hellander, 2019). It will be interesting to determine in future work which of these metrics are most effective at describing certain features of collective cell migration.

As a final example of future work, we discuss the possibility of modelling leader-follower migration by integro-differential equations (Bernardi et al., 2021; Colombi et al., 2021; Loy and Preziosi, 2020; Painter et al., 2015). The work in the second part of this thesis highlighted the importance of (nonlocal) contact guidance in establishing robust collective cell migration. While the neural crest ABM captures this process at the microscopic level, it is not clear what mathematical framework is best suited to investigate this phenomenon at the continuous, macroscopic scale. One possible approach may be to use integro-differential equations, which describe interactions that occur at distances far from cell centres of mass and (in certain cases) exhibit desirable stability properties when compared with more traditional PDEs. For example, while solutions to the classical Patlak–Keller–Segel system for chemotaxis can blow up in finite time, a similar system of integro-differential equations does not suffer from this issue (Chen et al., 2020; Eckardt et al., 2020). Despite these advantages, the analysis of integro-differential equations is less well developed than that of PDEs. Additionally, except for relatively simple examples, the derivation of the macroscopic equations from first principles is not well established despite its apparent natural connection to discrete models with nonlocal rules (Buttenschön et al., 2018; Carrillo et al., 2010; Morale et al., 2005). It would be interesting in future work to systematically derive closed-form system of integro-differential equations from discrete models such as the neural crest ABM considered here, as this would couple the microscopic dynamics to those at the macroscale. Furthermore, a continuous model may yield analytically tractable solutions that would be difficult to obtain from the stochastic ABM.

6.3 Conclusion and final remarks

This thesis examined different models for leader-follower cell migration using angiogenesis and neural crest cell migration as paradigms. By comparing different continuum mathematical models for angiogenesis using numerical simulation and techniques from asymptotic analysis, we identified conditions under which they generate similar solutions and, hence, when they may be appropriate descriptions of ensemble average discrete tip and stalk cell trajectories. In developing a novel agent-based model for neural crest cell migration and performing a variety of different *in silico* experiments (including a global sensitivity analysis), we highlighted the importance of ECM-related interactions on establishing robust collective cell migration. This work has also demonstrated how theoretical frameworks can be used to test biological hypotheses regarding the mechanisms by which collective cell migration arises, and how they can stimulate new ideas for *in vivo* (or possibly *in vitro*) experiments.

We conclude by commenting on the similarities between angiogenesis and neural crest cell migration and their possible significance to the mathematical modelling of collective cell migration. Although angiogenesis and neural crest cell migration occur under different biological conditions, they are affected by many of the same molecules and mechanisms: VEGF and angiopoietin-2, for instance, are known to guide the movement of tip cells and neural crest cells (Duran et al., 2017; McKinney et al., 2016; McLennan et al., 2010). Additionally, cells integrate many sources of information to determine their trajectories; in both angiogenesis and neural crest cell migration these signals arise from chemotaxis, volume exclusion, haptotaxis, and contact guidance. The extracellular matrix also plays an important role in both phenomena; it supplies a domain for cellular substrates and can physically prevent cells from moving in certain directions (Leonard and Taneyhill, 2020; Yamada and Sixt, 2019; De Palma et al., 2017). Furthermore, collective migration in angiogenesis and neural crest cell migration appears to depend on the same sets of ECM glycoproteins, such as fibronectin (Wang and Astrof, 2016; Stenzel et al., 2011), even though the exact mechanisms by which these molecules alter cell trajectories are not yet completely understood.

The biological similarities between angiogenesis and neural crest cell migration invite questions about whether their respective theoretical frameworks of study also resemble each other. Resolving such issues would be interesting, as an affirmative answer may point to the existence of a general model for collective cell migration that could accurately describe the spatial and temporal evolution of leader and follower cell densities in both phenomena. In such a framework, different biological phenomena would be recognised by distinct sets of parameter values, but would otherwise move according to the same underlying mechanisms. Addressing this question is outside the scope of this thesis, as it requires a deeper understanding of the mechanisms underlying each process and more detailed models of how such interactions individually alter cell trajectories. Nevertheless, it represents an intriguing long-term goal. In deriving a general system for collective cell migration, we would demonstrate a profound knowledge of the main mechanisms regulating this phenomenon and could apply it to further our understanding of the processes within which it occurs.

Appendix A

Chapter 2 Appendix

A.1 Review of the P–ABM

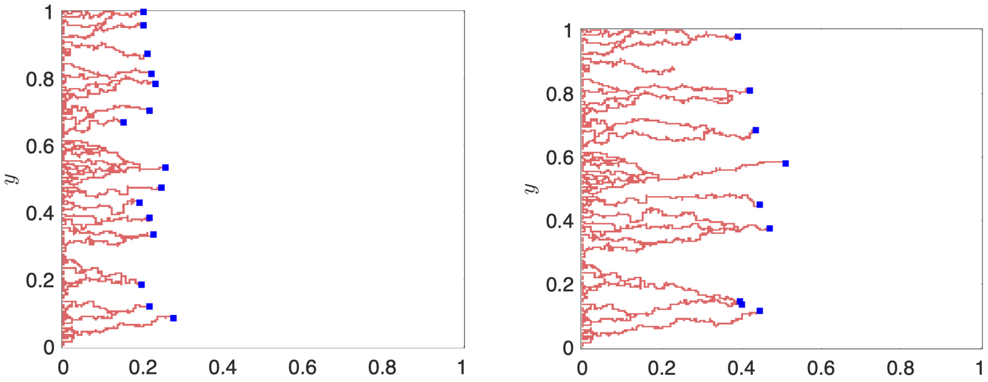
In this section, we review the rules for the discrete model originally described in Pillay et al. (2017) for the reader's reference. For a more detailed discussion of the P–ABM and its implementation, we refer to the original article.

The domain of the P–ABM is the unit square $(x, y) \in [0, 1]^2$. The lattice is equally spaced with N points in both the x and y directions with a spatial step size h, such that $(N - 1)h = 1$. We denote lattice sites with indices i, j, where $0 \leq i, j \leq (N - 1)$ and $(x_i, y_j) = (ih, jh)$. We assume that initially there is a blood vessel located at $x = 0$ and that there is a TAF field at steady state, such that it may be represented as a continuous function $c(x, y)$.

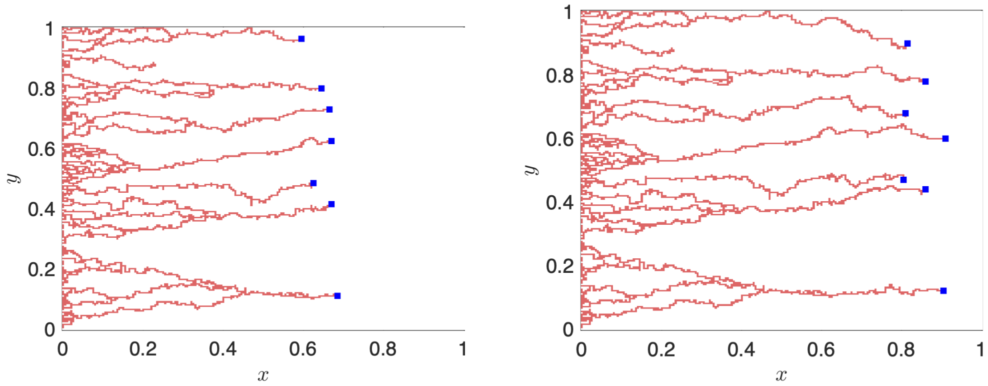
We distinguish between two cell types: leader tip cells (TCs) and endothelial stalk cells (ECs). TCs can move and proliferate to create new sprout branches, whereas ECs can only arise from TC movement and tip-to-tip anastomosis. At each time step, TCs, chosen uniformly at random with replacement, move with probability P_m. When a TC moves, its direction is chosen based on the probabilities listed in Eq.s (2.8)-(2.9) of §2.3.

Every time a TC moves from location (x_i, y_j), an EC is created at the lattice site. No-flux boundary conditions are applied along the domain, such that if a TC attempts to leave the lattice, then it is reflected back to its original position. We note that, in this event, an EC is still placed at the lattice site.

If a TC moves into a site occupied by one or more TCs, then tip-to-tip anastomosis occurs



(a) $t = 0.5$ (b) $t = 1$



(c) $t = 1.5$ (d) $t = 2$

Figure A.1: Network given by a single realisation of the P–ABM, subject to the TAF field $c(x, y) = x$, at (a) $t = 0.5$, (b) $t = 1$, (c) $t = 1.5$, and (d) $t = 2$. Tip cells are represented as dark blue squares, while stalk cells are light red squares. Initial and boundary conditions: see text of §A.1. Parameter values: see Table A.1.

with probability 1. In this event, all TCs at the lattice site are converted into a single EC (i.e., TC + A·TC → EC, where A $\geq$ 1). If a TC instead encounters an EC (tip-to-sprout anastomosis), then the TC is annihilated. Tip-to-tip anastomosis takes precedence over tip-to-sprout anastomosis. Following Pillay et al. (2017), TCs cannot anastomose with an EC from the same sprout, provided the TC has been at that site within 10 prior movement steps. This constraint prevents TC extinction via self-loop formation.

Branching is implemented in the following manner: TCs proliferate with probability P_b, which is assumed to be proportional to the TAF concentration at (x_i, y_j):

$$P_b = P_p c(x_i, y_j), \tag{A.1}$$

$c(x,y)$	N	h	k	P_m	P_p	Δt	t_{final}
(see figure caption)	201	200^{-1}	100	1	10^{-3} or 0	1/160	2

Table A.1: TAF field and parameter values used in P–ABM simulations. All values are dimensionless.

where P_p is a non-negative constant chosen with the restriction that $P_b \leq 1$.

In the P–ABM, TC branching only occurs in the y-direction. Thus, if a TC at (x_i, y_j) divides, then it is removed from the lattice site and two daughter TCs are created at $(x_i, y_{j\pm1})$. No ECs are created when a TC branches. When a TC divides, its daughter cells also cannot branch during that time step. In practice, a TC may deposit a daughter cell in a site that is already occupied by another cell, which (combined with our self-loop restriction) means that more than one cell may simultaneously occupy a lattice site.

At time $t = 0$, the domain is devoid of ECs while TCs are placed at alternating sites along the left edge of the lattice, so tip cells are located initially at $(x, y) = (0, 2h)$, $(0, 4h)$, ... $(0, (N-1)h)$. This initial condition is inspired by earlier investigations which showed that intercellular signalling creates a "salt and pepper" patterning of TCs and ECs, in which TCs are separated by no more than 2 stalk cells before they migrate towards the TAF source (Bentley et al., 2008).

For a review of the algorithm used to generate realisations of the P–ABM, we refer to §A.2.

Figure A.1 shows a network of TCs and ECs created by a typical realisation of the P–ABM using the initial conditions, parameter values, and boundary conditions listed above. We observe multiple anastomosis events (especially near the left boundary) and few branching events in this network. Figure 2.2 in §2.4 shows the average tip and stalk cell distributions generated by the P–ABM at $t = 0.2$ and $t = 2$, using the same parameter values and TAF field.

A.2 Numerical methods

The algorithm used to generate solutions of the P–ABM is given in Pillay et al. (2017) and reproduced here for reference:

Choose Δt, the time step, and t_{final}, the time at which to terminate the P–ABM solution. Set $t = 0$.

While $t < t_{\text{final}}$ and TCs have not reached $i = N - 1$ (the right edge of the lattice):

1. Choose N_{TC} tip cells at random with replacement, where N_{TC} is the current number of tip cells.

   Loop 1: For 1 to N_{TC}

   (a) Choose a random number $r_1 \in [0, 1]$.

   (b) If $r_1 \leq P_m$, the TC moves:

      i. Choose a random number $r_2 \in [0, 1]$.

      ii. The TC moves according to the probabilities $P_{x^\pm}, P_{y^\pm}$ in Eq.s (2.8)-(2.9) of §2.3, and an EC is left behind.

      iii. If tip-to-tip anastomosis can occur and the target site is occupied by one (or more) TCs, then all TCs at the lattice site are removed and an EC is placed there.

      iv. Else if tip-to-sprout anastomosis can occur, the target site is occupied by an EC, and the EC does not belong to the same sprout as the TC, then the TC is removed.

      v. Otherwise, the TC remains at the target site.

   End Loop 1

2. Choose N_{TC} tip cells at random with replacement, where N_{TC} is the number of tip cells after completing Loop 1.

   Loop 2: For 1 to N_{TC}:

   (a) Choose a random number $r_3 \in [0, 1]$.

(b) If branching can occur and $r_3 \leq P_b$, then the TC is removed and two TCs are placed at adjacent lattice sites in the y-direction.

End Loop 2

3. Set $t = t + \Delta t$.

End While loop

The 2D snail-trail system given by Eq.s (2.4)-(2.6) was solved using the finite element method, implemented with the FENICS library in PYTHON (Alnaes et al., 2015). The PDEs were solved on the unit square for $t \in [0.2, 2]$ using 200×200 second-degree continuous Lagrange finite elements. A Crank-Nicolson scheme with a time step of 0.001 was used to approximate the time derivative. The SCIPY library's `NearestNDInterpolator` function was used to set up a suitable initial condition on the mesh from the P–ABM solution at $t = 0.2$. Column-averaged snail-trail solutions in the y-dimension were obtained using the `trapz` function in Matlab (version 2019b), which implements a numerical trapezoidal integration scheme.

Nonlinear least squares parameter fitting for β_e and κ was conducted using the `least_squares` function in the SciPy library optimization package, which implements a trust-region-reflexive algorithm (Coleman and Li, 1996). 95% confidence intervals were computed using the `nlparci` function in Matlab (version 2019b).

A.3 An alternative derivation of the snail-trail corrective factor

In this appendix, we demonstrate that it is possible to derive Eq. (2.17), our expression for $\kappa(x, y)$, using the procedure given in Travasso et al. (2011) and Vilanova et al. (2014). In those articles, the modelling framework is not a snail-trail system given by Eq.s (2.4)-(2.5) but instead a hybrid ABM that utilises a discrete model for TC movement and a continuous model for capillary density. Despite such differences, the hybrid ABM contains a non-dimensional value that couples the discrete and continuum models and resembles Eq. (2.17). By deriving our formula for $\kappa(x, y)$ using the logic outlined in Vilanova et al.

(2014) and Travasso et al. (2011), we aim to show that the two factors arise from a common physical mechanism.

In the above articles, the coupling factor is calculated as “the ratio of the [vessel] material produced within the tip cell to the [area] of material swept as the cell migrates” (Vilanova et al., 2014). In other words, it is given as

$$\text{factor} = \frac{\text{area of vessel produced within time } \Delta t}{\text{area swept out by tip cells within time } \Delta t} = \frac{A_1}{A_2}. \tag{A.2}$$

We may physically interpret Eq. (A.2) as the TC “smearing” vessel density along the tunnel it creates within a given time step. In other words, the coupling factor arises as a result of a snail-trail assumption that new vessels proliferate along the path of TCs (this is similar to the assumption used to derive Eq. (2.17) for $\kappa(x, y)$).

We now re-express Eq. (A.2) using the discrete model rules from §2.3. We recall that new vessel production in the discrete model is equal to the number of times that TCs move from a lattice site. Since TCs move with probability P_m in time Δt and the area of a TC is equal to h^2, the total new vessel area produced by a TC, on average, is equal to $A_1 = h^2 P_m$.

The area per lattice length that a TC sweeps out when it moves is also equal to $h^2 P_m$. In order to calculate the total area swept out by the cell, however, this quantity must be multiplied by the average distance that the TC travels. Following our logic from §2.3, this latter value can be calculated as the expected net distance that a TC travels, which is given by

$$\sqrt{\mathbb{E}[X_{net}]^2 + \mathbb{E}[Y_{net}]^2} = \sqrt{(P_{x^+} - P_{x^-})^2 + (P_{y^+} - P_{y^-})^2} = \frac{1}{2}\sqrt{(g_x)^2 + (g_y)^2},$$

where $P_{x^\pm}$, $P_{y^\pm}$, g_x, and g_y are the same values as in Eq.s (2.8)-(2.9). By applying Taylor's theorem and neglecting terms of $O(h^3)$ or smaller, we find that the expected distance that a single TC travels is approximately $kh\|\nabla c\|_2$, hence the total area swept out by the tip cell within time Δt is approximately $A_2 = h^3 k P_m \|\nabla c\|_2$.

From here, it is straightforward to derive Eq. (2.17) by substituting the above expressions for A_1 and A_2 into Eq. (A.2) and defining $\mu := P_m/\Delta t$ and $\chi := P_m k h^2/\Delta t$.

Appendix B

Chapter 3 Appendix

B.1 Derivation of leading order P–PDE dynamics

Here we derive the leading order dynamics of the P–PDE, using the same procedure and assumptions listed in §3.2.

We recast the independent and dependent variables of Eq.s (3.1)-(3.2) in terms of the transformations given by Eq. (3.8), so that $N(x,t)$, $E(x,t)$, and $C(x,t)$ map to the functions $u(X,\tau)$, $w(X,\tau)$, and $c(X,\tau)$, respectively. This leads to the system

$$\frac{\partial u}{\partial \tau} = \Big(\frac{\partial^2 u}{\partial X^2} - \frac{1}{\epsilon}\frac{\partial u}{\partial X}\Big)\Big(1 - \alpha(u+w)\Big) + u(c-u-w), \tag{B.1}$$

$$\frac{\partial w}{\partial \tau} = \beta\Big[u + \alpha u^2 + \alpha^2 u\Big(\frac{\partial^2 u}{\partial X^2} - \frac{1}{\epsilon}\frac{\partial u}{\partial X}\Big)\Big], \tag{B.2}$$

which is subject to the initial and boundary conditions given by Eq.s (3.11)-(3.12). The parameter values ϵ, α, and β are defined as in Eq. (3.13). We again invoke assumption (A3) of §3.2 and let $0 < \epsilon \ll 1$, $0 < \alpha = \epsilon^2\Psi \ll 1$, where Ψ is a constant of $O(1)$. This allows us to expand the solutions to the P–PDE as a perturbation series. We will assume for the remainder of this section that $\beta \sim O(1)$ (we will identify the leading order dynamics for different limiting behaviours of β in §B.2).

Using the same procedure as outlined in §3.2, we find that the dynamics of the leading order stalk cell solution, $w_0(X,\tau)$, are given by

$$\frac{\partial w_0}{\partial \tau} = \beta u_0. \tag{B.3}$$

We identify the leading order TC dynamics by transforming the independent and dependent variables in Eq. (B.1) as follows:

$$y = X - \frac{\tau + X_0}{\epsilon} \qquad U = u, \qquad W = w, \qquad \widetilde{C} = c, \tag{B.4}$$

where X_0 is an arbitrary constant, so that $u(X, \tau)$, $w(X, \tau)$, and $c(X, \tau)$ map to functions $U(y, \tau)$, $W(y, \tau)$, and $\widetilde{C}(y, \tau)$. This leads to the equation

$$\frac{\partial U}{\partial \tau} = \frac{\partial^2 U}{\partial y^2} - \epsilon \Psi \Big(\epsilon \frac{\partial^2 U}{\partial y^2} - \frac{\partial U}{\partial y} \Big) \Big(U + W \Big) + U(\widetilde{C} - U - W), \tag{B.5}$$

We substitute the perturbation series

$$U(y, \tau) \sim U_0(y, \tau) + \epsilon U_1(y, \tau) + O(\epsilon^2), \tag{B.6}$$

and identify the leading order dynamics by equating terms of $O(\epsilon^0)$. This leads to the leading order PDE for the TC dynamics:

$$\frac{\partial U_0}{\partial \tau} = \frac{\partial^2 U_0}{\partial y^2} + U_0(\widetilde{C} - U_0 - W_0), \tag{B.7}$$

where the dynamics for the leading order EC solution, $W_0(y, \tau)$, are given by Eq. (B.3) after a suitable transformation of variables. Crucially, the resulting system is equivalent to that of the ST–PDE in §3.2 (Eq. (3.16) and Eq.s (3.21)-(3.22)), since we show in §B.3 that the leading order tip cell solutions to the ST–PDE are non-negative. We have confirmed that numerical solutions to the above system are in good agreement with results from the full P–PDE when assumptions (A1)-(A3) are satisfied (data not shown).

B.2 Leading order dynamics for different limiting behaviours of β

In §3.2 and §B.1, we assumed $\beta \sim O(1)$ when deriving the leading order ST–PDE and P–PDE dynamics. We now relax this condition and consider how the leading order ST–PDE dynamics change for cases where $\beta \gg 1$ and $\beta \ll 1$ (the analysis for the P–PDE follows naturally, and is omitted for clarity, since it yields identical leading order equations). Table B.1 summarises all of the possible leading order systems for the ST–PDE and P–PDE when $0 < \epsilon \ll 1$ and $0 < \alpha \ll 1$.

	Leading Order Stalk Cell Dynamics	**Leading Order Tip Cell Dynamics** $\left(y = X - \frac{\tau + X_0}{\epsilon},\ X_0 \in \mathbb{R}\right)$
$\beta \sim O(1)$	$\frac{\partial w_0}{\partial \tau} = \beta u_0$	$\frac{\partial U_0}{\partial \tau} = \frac{\partial^2 U_0}{\partial y^2} + U_0(\widetilde{C} - U_0 - W_0)$
$\beta \gg 1$ $(\beta = \widetilde{\beta}\epsilon^{-m})$	$\frac{\partial w_0}{\partial \tau} = \widetilde{\beta}\bar{u}_0$	$\frac{\partial \bar{U}_0}{\partial \tau} = \frac{\partial^2 \bar{U}_0}{\partial y^2} + \bar{U}_0(\widetilde{C} - W_0)$
$\beta \ll 1$ $(\beta = \widehat{\beta}\epsilon^{m})$	$\frac{\partial \bar{w}_0}{\partial \tau} = \widehat{\beta} u_0$	$\frac{\partial U_0}{\partial \tau} = \frac{\partial^2 U_0}{\partial y^2} + U_0(\widetilde{C} - U_0)$

Table B.1: Leading order dynamics of the ST–PDE and P–PDE for different positive values of β. In all cases, $u(X, \tau)$ and $U(y, \tau)$ represent the tip cell density, $w(X, \tau)$ and $W(y, \tau)$ the stalk cell density, and $c(X, \tau)$ and $\widetilde{C}(y, \tau)$ the TAF concentration. We refer to the main text of §3.2 and §B.1-§B.2 for the derivation of these systems.

When $\beta \gg 1$, we can rewrite it as a power of ϵ using the equation $\beta = \widetilde{\beta}\epsilon^{-m}$, where $m > 0$ and $\widetilde{\beta}$ is a constant of $O(1)$. To derive the leading order dynamics, we rescale the dependent variable u so that $u = \epsilon^m \bar{u}$ and $u(X, \tau) \mapsto \bar{u}(X, \tau)$. Substitution into Eq.s (3.9)-(3.10) yields

$$\frac{\partial \bar{u}}{\partial \tau} = \frac{\partial^2 \bar{u}}{\partial X^2} - \frac{1}{\epsilon}\frac{\partial \bar{u}}{\partial X} + \bar{u}(c - \epsilon^m \bar{u} - w), \tag{B.8}$$

$$\frac{\partial w}{\partial \tau} = \widetilde{\beta}\left|\epsilon\frac{\partial \bar{u}}{\partial X} - \bar{u}\right|. \tag{B.9}$$

From this point, the procedure for deriving the outer and inner systems is the same as in §3.2. We find that the dynamics governing the leading order stalk cell density, $w_0(X, \tau)$, are given by

$$\frac{\partial w_0}{\partial \tau} = \widetilde{\beta}|\bar{u}_0|. \tag{B.10}$$

Similarly, the leading order tip cell dynamics, $\bar{U}_0(y, \tau)$, may be identified using the same transformations presented in Eq. (3.17). This yields

$$\frac{\partial \bar{U}_0}{\partial \tau} = \frac{\partial^2 \bar{U}_0}{\partial y^2} + \bar{U}_0(\widetilde{C} - W_0), \tag{B.11}$$

where the dynamics of $W_0(y, \tau)$ are given by Eq. (B.10) after a suitable transformation of variables. The boundary conditions are given at leading order by

$$\lim_{y\to\pm\infty} \bar{U}_0(y,\tau) = 0. \tag{B.12}$$

A similar proof to the one in §B.3 shows that the leading order tip cell solution is non-negative, provided its initial condition is non-negative. Therefore we may drop the absolute value sign in Eq. (B.10). Crucially, the resulting leading order system is identical to the one derived for the P–PDE.

For the case in which $\beta \ll 1$, we rewrite the parameter in terms of a power of ϵ, such that $\beta = \widehat{\beta}\epsilon^m$, where $m > 0$ and $\widehat{\beta}$ is a constant of $O(1)$. We rescale the dependent variable w in the ST–PDE, such that $w = \epsilon^m \bar{w}$ and $w(X,\tau) \mapsto \bar{w}(X,\tau)$. Substitution into Eq.s (3.9)-(3.10) leads to the system

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial X^2} - \frac{1}{\epsilon}\frac{\partial u}{\partial X} + u(c - u - \epsilon^m \bar{w}), \tag{B.13}$$

$$\frac{\partial \bar{w}}{\partial \tau} = \widehat{\beta}\left|\epsilon\frac{\partial u}{\partial X} - u\right|. \tag{B.14}$$

We deduce that the dynamics of the leading order stalk cell solution, $\bar{w}_0(X,\tau)$, are given by

$$\frac{\partial \bar{w}_0}{\partial \tau} = \widehat{\beta}|u_0|. \tag{B.15}$$

To eliminate the ϵ^{-1} term in Eq. (B.13), we change variables using the same transformations given in Eq. (3.17) and expand solutions as a perturbation series. This allows us to identify the leading order TC dynamics as

$$\frac{\partial U_0}{\partial \tau} = \frac{\partial^2 U_0}{\partial y^2} + U_0(\widetilde{C} - U_0), \tag{B.16}$$

which are subject to the homogeneous Dirichlet boundary conditions

$$\lim_{y\to\pm\infty} U_0(y,\tau) = 0. \tag{B.17}$$

If the initial condition for the leading order tip cell solutions is non-negative, then a similar proof to the one outlined in §B.3 shows that their solutions are non-negative at all other times. We may thus drop the absolute value sign in the leading order EC rate equation. The resulting leading order system is identical to that of the P–PDE.

B.3 Leading order ST–PDE solutions are non-negative

We show in this appendix that Eq. (3.21), which describes the leading order ST–PDE TC dynamics, admits non-negative solutions when its initial condition is also non-negative. For this proof, we assume solutions are at least twice differentiable.

For sake of contradiction, assume that there exists a point $y^* \in \mathbb{R}$ such that at time $\tau_1 > 0$, $U_0(y^*, \tau_1) < 0$. Since the PDE solution is twice differentiable, it is continuous, so there must have existed some time $0 \leq \tau^* < \tau_1$ such that $U_0(y^*, \tau^*) = 0$ and $\frac{\partial U_0}{\partial \tau}(y^*, \tau^*) < 0$. We also conclude that $U_0(y, \tau^*) \geq 0$ for all other values of y, since y^* is the first point at which the solution becomes negative.

Using Taylor's theorem, the second derivative in Eq. (3.21) may be written as

$$\frac{\partial^2 U_0}{\partial y^2}(y^*, \tau^*) = \lim_{\Delta y \to 0} \frac{U_0(y^* + \Delta y, \tau^*) - 2U_0(y^*, \tau^*) + U_0(y^* - \Delta y, \tau^*)}{\Delta y^2}, \tag{B.18}$$

which is well-defined because $U_0(y, \tau)$ is twice differentiable. Since $U_0(y^*, \tau^*) = 0$ and $U_0(y^* \pm \Delta y, \tau^*) \geq 0$, the second derivative must be non-negative. Since the nonlinear reaction term of Eq. (3.21) is equal to zero when $U_0 = 0$, this means that $\partial U_0/\partial \tau \geq 0$ at (y^*, τ^*), which is a contradiction. We conclude that any twice differentiable solution to Eq. (3.21) with smooth, non-negative initial data is also non-negative for $\tau \geq 0$.

B.4 Derivation of self-similar tip cell solutions

In this appendix, we demonstrate that the solution to Eq. (3.21), which describes the leading order TC dynamics, is self-similar in the limit as $\tau \to 0$. We focus on the case for which $\beta \sim O(1)$; a similar procedure can be used to derive self-similar solutions for a wider range of values for β, using the leading order dynamics derived in §B.2.

In addition to the assumptions made in §3.2, we simplify our analysis by considering the following statements to be true:

(B1) The TC dynamics are given by Eq. (3.21) with boundary conditions (3.22).

(B2) The TAF concentration $\widetilde{C}(y, \tau)$ is bounded between zero and one.

(B3) The initial leading order TC and EC densities are non-negative and bounded.

(B4) The wave speed of the leading order tip cell solution, $\sigma(\tau)$, may depend on time but is bounded such that $\sigma(\tau)\tau$ and $\sigma'(\tau)\tau \to 0$ as $\tau \to 0$. This assumption is inspired by our numerical results, which suggest that solutions to the ST–PDE and P–PDE travel short distances in small time periods but have wave speeds that vary over time (see Figure B.1).

We seek self-similar solutions by substituting into Eq. (3.21) the expression

$$U_0(y,\tau) = \tau^A \widetilde{U}_0\big(\tau^B(y - \sigma(\tau)\tau)\big) = \tau^A \widetilde{U}_0(z), \tag{B.19}$$

where A and B are constants to be determined, $\sigma(\tau)$ is the tip cell wave speed at time τ, and $z = \tau^B(y - \sigma(\tau)\tau)$ is the self-similar variable. We recast the equation in terms of the independent variables z and τ. Additionally, we use the variable $\widetilde{c}$ to denote the TAF concentration, with the understanding that its value varies between zero and unity within the domain. Similarly, we denote the leading order stalk cell density as $\widetilde{W_0}$, and note that its value also depends on z and τ. These substitutions lead to the equation

$$A\widetilde{U}_0 + \left(Bz - \tau^{1+B}\Big(\sigma(\tau) + \tau\frac{d\sigma}{d\tau}\Big)\right)\frac{d\widetilde{U}_0}{dz} = \tau^{1+2B}\frac{d^2\widetilde{U}_0}{dz^2} + \Big(\tau\widetilde{c} - \tau\widetilde{W}_0 - \tau^{1+A}\widetilde{U}_0\Big)\widetilde{U}_0. \tag{B.20}$$

The values of A and B are chosen to eliminate as many powers of τ in Eq. (B.20) as possible, while simultaneously ensuring that no term blows up to infinity as $\tau \to 0$. This uniquely sets $A = -1$ and $B = -1/2$. However, the resulting equation does not describe self-similar solutions in general because there remain terms proportional to τ. We conclude that self-similar solutions do not exist for sufficiently large values time periods, as those terms will begin to dominate the dynamics. However, in the limit as $\tau \to 0$ one can recover an ODE for $\widetilde{U}$ as a function of z. This is because the TAF concentration, wave speed, and stalk cell densities are all bounded and are all multiplied by positive powers of τ in Eq. (B.20). Consequently, they vanish as $\tau \to 0$. Such simplifications lead to the ODE (3.23).

B.5 Numerical methods

We construct initial conditions for the ST–PDE and P–PDE by column averaging one thousand P–ABM realisations in the y-direction at $t = 0.2$, using ABM parameter values listed in Pillay et al. (2017). We initialise the PDEs at time $t = 0.2$, rather than $t = 0$, as previous investigations have shown that the P–PDE can become ill-posed if it is initialised at earlier times (Pillay et al., 2017, 2018).

The 1D PDE models were solved using the method of lines (Schiesser, 1991) with central finite difference formulae for spatial derivatives and a spatial step size $\Delta x = 1/200$. The systems were solved with the `ode15s` function in Matlab (version 2019b), which implements adaptive time-stepping for stiff ODE problems (Shampine and Reichelt, 1997).

In Figure 3.5, we fitted the parameters λ and a_e to P–ABM data using a nonlinear least squares method that minimises the error generated by the PDE solutions in the least squares sense. We computed such values on the time interval $t \in [0.2, 2]$ using the `lsqnonlin` function in MATLAB (version 2019b), which implements a trust-region-reflective algorithm (Coleman and Li, 1996).

In the figures presented in §3.2-§3.3, we transformed solutions of the ST–PDE and P–PDE from functions of the form $N(x, t)$ and $E(x, t)$ to $u(X, \tau)$ and $w(X, \tau)$, respectively, so as to facilitate comparison with our asymptotic analysis. This is done using the formulae listed in Eq. (3.8).

The maximum relative difference between solutions, which appears in Figure 3.3(c) and several supplementary figures (data not shown), was computed in the following manner: if $v_{\text{ST}}(X, \tau)$ denotes the (tip or stalk cell) ST–PDE solution at location X and time τ, and $v_{\text{P}}(X, \tau)$ depicts the corresponding P–PDE solution, then we calculate their maximum relative difference at time τ, $\zeta(\tau)$, as

$$\zeta(\tau) = \frac{\|v_{\text{ST}}(X, \tau) - v_{\text{P}}(X, \tau)\|_\infty}{\max\{\|v_{\text{ST}}(X, \tau)\|_\infty, \quad \|v_{\text{P}}(X, \tau)\|_\infty\}}, \tag{B.21}$$

where $\|\cdot\|_\infty = \sup_{X \in \Omega} |\cdot|$ denotes the supremum norm (the maximum value of the function over the spatial domain, Ω). We note that the relative difference between EC solutions is computed on only a subset of the domain (from $X = 0.05\sqrt{\lambda/D}$) because the initial condition tends to have a large EC density near the left-hand boundary that causes the relative difference to become artificially small.

When we compute the relative difference between the self-similar leading order tip cell

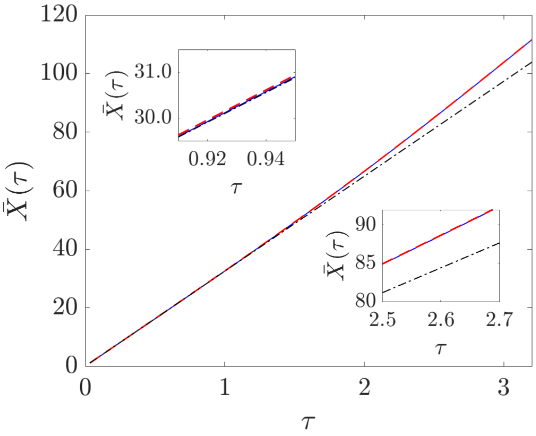


Figure B.1: Mean TC location, $\bar{X}(\tau)$, versus τ for numerical solutions for the ST–PDE and P–PDE (Table 3.1). These results were generated using the same initial condition and parameter regime that produced Figure 3.3. We have also plotted the line that minimises the sum of squared residuals for data points within the range $[0.2\lambda, 5\lambda]$, where $\lambda = 0.16$. The formula for this line of best fit is $\bar{X}(\tau) = 32.45\tau + 0.06$. The top left inset shows a zoomed-in view of the graph for small values of τ, while the bottom right inset shows a zoomed-in view of the results for longer time points. Key: ST–PDE wave front (dashed red line), P–PDE wave front (solid blue line), line of best fit to data in the interval $\tau \in [0.2\lambda, 5\lambda]$ (dashed-dot black line).

solution, $u_{\mathrm{SS}}(X,\tau)$, and those of the ST–PDE and P–PDE solutions we use a similar formula to Eq. (B.21):

$$\zeta(\tau) = \frac{\max\{\|u_{\mathrm{SS}}(X,\tau) - u_{\mathrm{ST}}(X,\tau)\|_\infty, \quad \|u_{\mathrm{SS}}(X,\tau) - u_{\mathrm{P}}(X,\tau)\|_\infty\}}{\max\{\|u_{\mathrm{ST}}(X,\tau)\|_\infty, \quad \|u_{\mathrm{P}}(X,\tau)\|_\infty, \quad \|u_{\mathrm{SS}}(X,\tau)\|_\infty\}}. \tag{B.22}$$

The solution to Eq. (3.23), presented in Figure 3.23(a), was computed numerically using the MATLAB library Chebfun, which applies spectral methods and Newton's method to solve nonlinear boundary value problems (Driscoll et al., 2014). To approximate the infinite domain, the solution was computed on the interval $z \in [-100, 100]$, which is two orders of magnitude larger than the region of interest, using the initial guess $\widetilde{U_0}(z) = \exp(-z^2)$.

The quantities X_0 and $\phi(\tau)$, which are used to transform the self-similar solution into a function of the independent variables X and τ in Figure 3.23(b), are calculated in the following manner. The wave speed $\phi(\tau)$ is taken to be the derivative of $\bar{X}(\tau)$, the location of the TC wave front at time τ. We define the wave front location as the average spatial location of TCs using the formula

$$\bar{X}(\tau) := \frac{\int_{-\infty}^{\infty} X u(X,\tau) dX}{\int_{-\infty}^{\infty} u(X,\tau) dX}. \tag{B.23}$$

Since we have found that $\bar{X}(\tau)$ is approximately linear, we calculate the wave speed as the slope of the line that minimises the sum of squared residuals with the graph of $\bar{X}(\tau)$ versus τ. The value of X_0/ϵ is similarly taken to be the y-intercept of this line.

Figure B.1 shows a plot of $\bar{X}(\tau)$ for the same parameter regime and initial conditions that generated Figure 3.3. We observe that the ST–PDE and P–PDE tip cell solutions yield roughly identical wave speeds for the time points considered here, and that they gradually accelerate over time. For early times, the two solutions appear to travel with constant speed: we confirmed this by plotting a straight line that minimises the sum of squared residuals with data points in the range $\tau \in [0.2\lambda, 5\lambda]$, where $\lambda = 0.16$. We observe that this line of best fit is a good approximation to the numerical results for sufficiently small values of τ (for $\tau < 0.8$, the R^2-value is greater than 0.99). We conclude that for $\tau \approx 0$, the wave speed $\phi(\tau)$ is approximately constant; its value is given in Table 3.2 of §3.3.

Appendix C

Chapter 4 Appendix

C.1 Algorithm for the NCC ABM

NCC ABM Algorithm

```
Create the initial distributions of FN puncta and NCCs

Initialise edge agents (these are located at the same initial position of leader cells, but
are stationary, do not interact with NCCs, and are only used to determine when new
follower cells enter the domain – see Step 6 below)

For all leader cells:
    1. Sample a time interval $r$ from an exponential distribution with user-defined
       mean $T_{\text{ave}}$

Set $t = 0$.

while $t < 720$ min:

    1. For all leader and follower cells, compute their new velocities:
        (a) List all NCCs and FN puncta/fibres within a radius $R_{\text{filo}}$ from the
            current cell.
            (i) For FN within a distance between $R_{\text{cell}}$ and $R_{\text{filo}}$:
                – Compute the FN density, $S_{\text{FN}}$, using Eq. (4.2).
                – Record that the cell senses FN and set the maximum cell speed to
                  be $s_{\text{max}}^{\text{FN}}$.
                – Sample $\hat{\mathbf{u}}_{\text{hap}}$ from the distribution given in Eq. (4.3).
                – If the NCC sensed any FN agents with orientations, use Eq. (4.7)
                  to compute $\hat{\mathbf{u}}_{\text{cg}}$. Otherwise, set it equal to $\mathbf{0}$.
                – Compute $\hat{\mathbf{F}}_{\text{ECM}}$ using Eq. (4.8).
                – Calculate $F_{\text{ECM}}$ using Eq. (4.9).
            (ii) For all neighbouring leader and follower cells within a distance $R_{\text{filo}}$:
                – Compute the cell density, $S_{\text{cell}}$, using Eq. (4.10).
                – Sample $\hat{\mathbf{F}}_{\text{VE}}$ from the distribution defined by Eq. (4.11).
                – Compute the magnitude of $\mathbf{F}_{\text{VE}}$ using Eq. (4.12).
```

```
        (iii) If the cell did not sense any FN, then set its maximum speed to be
              $s_{\max}^{\text{off}}$. If it also did not sense any neighbouring cells, do not change
              its velocity direction from the previous time step and proceed to (v).
        (iv) Compute the new NCC velocity using Eq. (4.1).
        (v) Set the norm of the velocity vector to be the maximum cell speed if
            it exceeds that value.

  2. Record the new cell velocity.

  3. Apply a second-order Adams–Bashforth scheme to update cell positions.

  4. Secrete new FN puncta in the domain:
     (a) For all NCCs that can secrete FN:
         (i) If the time since the last FN secretion by the NCCs has exceeded $r$,
             AND the NCC has sensed at least one FN puncta:
             – Create a new FN punctum located at the NCC centre
             – Sample a new time interval $r$ from an exponential distribution with
               mean $T_{\text{ave}}$

  5. Apply reflective boundary conditions to cells that have exited the domain.

  6. Add new follower cells to the domain:
     (a) For all edge agents:
         (i) If the edge agent centre is not covered by another NCC, then add a
             new follower cell at that location.

  7. Remodel the FN matrix:
     (a) For all FN puncta/fibres covered by a NCC:
         (i) If the FN agent does not currently have an orientation, then set it
             to be equal to the direction of the NCC velocity.
         (ii) Otherwise, use Euler's method to numerically simulate the ODE
              (4.5), with $\zeta = \ln(2)/T_{\text{half}}$.

  8. Set $t = t + \Delta t$.

end while loop
```

C.2 Determining the optimal LHS sample size

This section presents the results of a procedure originally presented in Marino et al. (2008), which determines the optimum minimal number of samples to use in LHS/PRCC analysis. This approach employs a metric known as the top-down correlation coefficient (TDCC), which quantifies the extent to which the PRCCs are ranked in a similar order when the number of LHS samples is increased (Iman and Conover, 1987). TDCCs are computed as the correlation coefficient of the Savage scores, which are values that weight PRCCs according to their ordering. For each set of PRCCs considered, the Savage score of the h^{th} largest PRCC, SS_h, is given by

$$SS_h = \sum_{j=h}^{k} j^{-1}, \tag{C.1}$$

where k is the number of parameters that are varied in the LHS/PRCC analysis. We remark that the largest Savage scores correspond to the highest-valued PRCCs. Since PRCCs may be ranked in different orders when the number of LHS samples increases, the Savage scores for a given parameter value may vary across datasets.

For a given matrix of PRCCs, we compute the Savage score of each entry and then calculate the TDCC as the Pearson correlation coefficient of the Savage scores matrix. By determining the statistical significance of the resulting TDCCs (Marino et al., 2008), we may infer the minimum number of samples to take for the LHS/PRCC analysis. We remark that, since the magnitude of PRCCs, not their raw ordering, determines the sensitivity of output statistics to parameter values, we compute TDCCs and the Savage scores using the absolute value of the PRCC matrix.

We computed TDCCs using the LHS/PRCC dataset given in Chapter 4. Specifically, we computed PRCCs for various sample sizes ($N = 100$, $N = 200$, $N = 300$, $N = 400$, $N = 500$, and $N = 1000$ samples) by randomly selecting subsets of the LHS dataset. We opted to randomly sample parameter regimes without replacement from the existing dataset, rather than use new parameter regimes, in order to limit the computational cost of the algorithm. To ensure that the results are not sensitive to the particular parameter regimes selected, we averaged all PRCCs over 30 replicates before calculating the Savage scores. Following the procedure outlined in Marino et al. (2008), we then calculated TDCCs on pairs of datasets ($N = 100$ and $N = 200$, $N = 200$ and $N = 300$, etc.). Table C.2 presents the resulting TDCCs for the summary statistics considered in Chapter 4. We observe that, for each summary statistic, many values of the TDCC are identical – this results from our averaging of PRCCs prior to computing the Savage scores. We also observe that the smallest sample size for which each metric has had a statistically significant TDCC (with

N	**Max Distance in x-Direction**	**Range in y-Direction**
100–200	0.9595	1.0000*
200–300	0.9595	1.0000*
300–400	1.0000*	1.0000*
400–500	1.0000*	1.0000*
500–1000	1.0000*	1.0000*

N	**Nearest Neighbour Distance**	**Follower Order Parameter**
100–200	0.8380	0.8135
200–300	1.0000*	0.9820*
300–400	0.8380	0.9820*
400–500	0.8380	0.8199
500–1000	1.0000*	0.8199

Table C.2: Top-down correlation coefficients (TDCCs) for the four summary statistics investigated in Chapter 4, which have been computed from the LHS dataset using subsets of $N = 100$, $N = 200$, $N = 300$, $N = 400$, $N = 500$, or $N = 1000$ parameter regimes. Asterisks indicate that the p-value for the particular TDCC of interest is less than 0.01.

$p < 0.01$) occurs for $N = 300$ samples. Thus, we need only sample 300 parameter regimes for the LHS/PRCC analysis. In order to better resolve relationships and trends, however, we opt to use the largest value of parameter regimes ($N = 1000$) for which the LHS/PRCC algorithm terminates in a reasonable time frame.

Appendix D

Chapter 5 Appendix

D.1 Consistency analysis results

Consistency analysis determines the minimum number of samples needed to reduce the degree to which stochastic effects (also known as aleatory uncertainty) affect the values of summary statistics obtained from a given ABM (Hamis et al., 2021). One technique for this analysis involves a non-parametric statistical test that relies on the Vargha–Delaney A-measure (Marino et al., 2008; Vargha and Delaney, 2000). This metric quantifies the extent of differences between two samples: specifically, it estimates the probability that a given measurement in Sample 1 is larger than that in Sample 2. If b_i denotes a measurement from Sample 1 (with $1 \leq i \leq n$) and c_j a measurement from Sample 2 (with $1 \leq j \leq m$), then the A-measure, $\hat{A}$, is given by

$$\hat{A} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} H(b_i - c_j), \tag{D.1}$$

where $H(x)$ represents the Heaviside step function, which is defined to be

$$H(x) = \begin{cases} 1, & x > 0, \\ 0.5, & x = 0, \\ 0, & x < 0. \end{cases} \tag{D.2}$$

In practice, we use the `ranksum` function in Matlab (version 2021a) to compute A-measures, following an implementation detailed in Hamis et al. (2021). It is clear from Eq.s (D.1)-(D.2) that the A-measure is equal to 0.5 when the two samples are identical, and that its value is bounded between zero and unity. Values of this metric that are sufficiently far from 0.5 indicate that there are differences between the two samples (Read et al., 2012). Since we are not interested in the ordering of samples, but rather whether the two samples are different from each other in magnitude, we rescale the A-measure using the equation

$$\hat{\mathbf{A}} = 0.5 + \left|0.5 - \hat{A}\right|, \tag{D.3}$$

where $\hat{\mathbf{A}} \in [0.5, 1]$ represents the scaled A-measure and $\hat{A} \in [0, 1]$ its raw estimate. In their original paper, Vargha and Delaney (2000) outline threshold values for determining statistical significance: scaled A-measures below 0.56 indicate "small" differences, while those above 0.71 describe significantly distinct samples.

To employ the scaled A-measure in consistency analysis, we carry out the following procedure originally outlined in Alden et al. (2013) and Read et al. (2012):

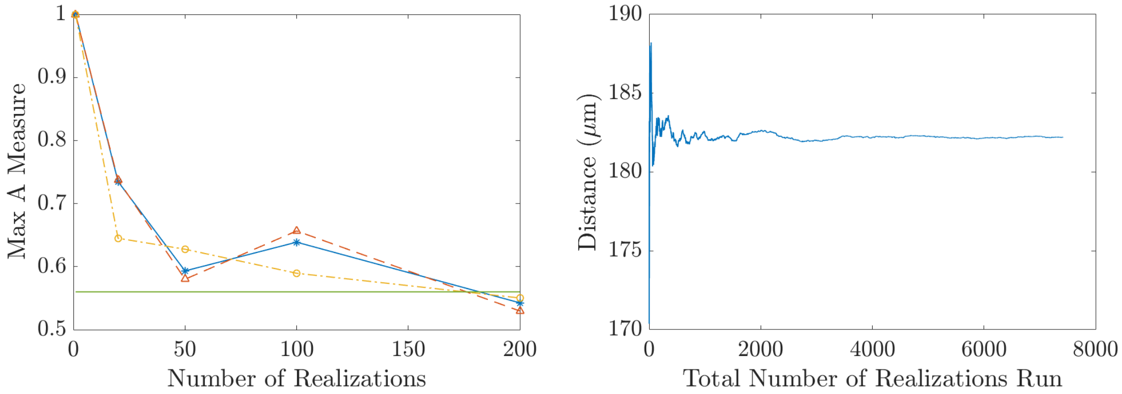
I. Define a finite set $\mathcal{C}$ that lists the number of realisations per parameter regime. As an example, let $\mathcal{C} = \{1, 5, 20, 50, 100, 200\}$ replicates of the ABM.

II. For each element $k \in \mathcal{C}$, compute twenty sets of summary statistics that are averaged over k ABM realisations. Following the above example, we will therefore obtain twenty sets of summary statistics averaged over one realisation, twenty sets averaged over five realisations, etc.

III. For each entry in $\mathcal{C}$, use Eq.s (D.1)-(D.3) to compare the magnitude of the stochastic differences between the first set of statistics and the other nineteen. Record the maximum value of the nineteen scaled A-measures.

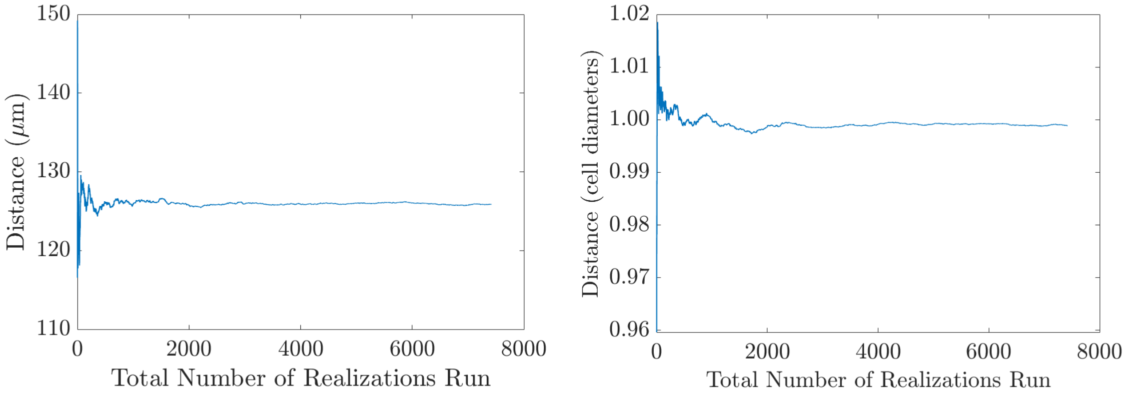
IV. Define the total number of realisations per parameter regime to be the smallest element of $\mathcal{C}$ for which the maximum scaled A-measure is below the 0.56 threshold for all summary statistics of interest.

Figure D.1(a) presents the Vargha–Delaney A-Test results for three summary statistics used to characterise several aspects of collective behaviour in our NCC ABM (the invasion distance, the lateral dispersion of NCCs, and the nearest neighbour distance). We observe that, for all three summary statistics shown, the maximum A-measure does not fall below the 0.56 significance threshold until 200 realisations are taken. We conclude from



(a) Maximum A-statistic value

(b) Cumulative mean, invasion distance



(c) Cumulative mean, lateral dispersion

(d) Cumulative mean,
average nearest distance between NCCs

Figure D.1: (a) Maximum A-measure values of the NCC ABM from Chapters 4 and 5, computed using Algorithm I.-IV. with $\mathcal{C} = \{1, 5, 20, 50, 100, 200\}$ realisations. The solid green horizontal line in (a) denotes the threshold below which there are no significant differences between the samples. The statistic for the invasion distance is represented as a blue line with asterisks, that of the lateral dispersion of NCCs is a dashed orange line with triangles, and that of the average nearest neighbour distance is denoted by a dashed-dot yellow line with open circles. The other plots present the cumulative means for (b) the invasion distance, (c) the lateral dispersion of NCCs, and (d) the average nearest neighbour distance, which are shown in increments of one realisation.

these results that we should use 200 realisations per parameter regime in order to ensure that inherent ABM stochasticity is not responsible for statistically significant differences in outputs obtained from distinct parameter regimes.

An alternative method for consistency analysis identifies the optimum number of ABM realisations per parameter regime by plotting the cumulative mean of summary statistics. The number of replicates is then chosen to be a value above which stochastic fluctuations about the mean are sufficiently small – for instance, below 5% of the mean value (Renardy et al., 2019). Figures D.1(b)-(d) present plots of the cumulative mean for the three NCC ABM summary statistics considered above. In all three plots, we observe that for significantly large numbers of realisations (over four thousand, for instance), there are few fluctuations from the mean value of all three statistics of interest. Although the graphs are noisier when fewer ABM realisations are used, for 200 realisations or more the amplitudes of such fluctuations are close to (i.e., within 3% of) their final respective mean values.

D.2 Numerical methods

The NCC ABM is simulated using the same algorithm as outlined in Chapter 4 and Appendix C. The model is simulated with the `C++` library PhysiCell (version 1.7.1), which efficiently simulates force-based overlapping spheres models (Ghaffarizadeh et al., 2018).

Violin plots were generated in Matlab (version 2021a), using a suite of open-source methods programmed by Bechtold (2016). Violin plots are designed to yield the same information as box plots, but also use kernel density estimation to calculate the approximate distribution of the summary statistic (Hintze and Nelson, 1998). In order to evaluate whether these distributions are significantly different from the baseline parameter regime, we employ two statistical tests: the Mann–Whitney U test and the two-sample Kolmogorov–Smirnov test. The Mann-Whitney U test is similar to the A-measure from §D.1, and determines whether two distributions are significantly different by measuring the probability that a given sample from one distribution is less than that of the other (Mann and Whitney, 1947). The two-sample Kolmogorov–Smirnov test, on the other hand, computes the largest absolute difference between the empirical distribution functions of each sample to determine whether they are significantly different (Massey, 1951). We calculate Mann–Whitney U-tests (and their p-values) using the `ranksum` function in Matlab, and the two-sample Kolmogorov–Smirnov tests (and p-values) using its `kstest2` function. We set the statistical significance threshold to be at the 1% level, so that tests with $p < 0.01$ are considered to be significant. In the violin plots, distributions that have both significant Mann–Whitney U tests and significant two-sample Kolmogorov–Smirnov tests are labelled with black asterisks.

We also note that, since we are comparing multiple samples, we must correct p-values before assessing statistical significance so as to reduce the false discovery rate (Goeman and Solari, 2014). We apply a Bonferroni correction – a conservative *post hoc* approach – for adjusting the statistical significance to account for multiple comparisons. Bonferroni corrections multiply raw p-values by the number of statistical tests performed. In Figure 5.1, for instance, we multiply the raw p-values of each test by the number of pairwise tests per statistic (three) and by number of statistics collected (we recorded eleven metrics, although only four are presented in this chapter). Thus, each p-value is multiplied by 33 before we determine the statistical significance of the Mann–Whitney U test and the two-sample Kolmogorov–Smirnov test.

D.3 Supplementary tables

Abbreviation	**Meaning**
CG ++, Both Cells	Both cells upregulate contact guidance ($\rho = 0.33$)
CG ++, FO	Followers upregulate contact guidance ($\rho = 0.33$)
Leader FN ++	Leaders upregulate fibronectin secretion ($T_{\text{ave}} = 10$ min)
Follower FN +	Followers can secrete fibronectin ($T_{\text{ave, follower}} = 90$ min)
VE −/+, Both Cells	Reduce volume exclusion strength in both cells ($c_i = 0.05$)
VE ++, Both Cells	Increase volume exclusion strength in both cells ($c_i = 2$)
VE −/+, LO	Reduce volume exclusion strength in leaders only ($c_i = 0.05$)
VE ++, LO	Increase volume exclusion strength in leaders only ($c_i = 2$)
VE −/+, FO	Reduce volume exclusion strength in followers only ($c_i = 0.05$)
VE ++, FO	Increase volume exclusion strength in followers only ($c_i = 2$)
Hap. Bias ++	Both cells sample less noisy haptotaxis cues ($\gamma_{\text{FN}} = 3000$)
Hap. Bias ++, FO	Followers sample less noisy haptotaxis cues ($\gamma_{\text{FN}} = 3000$)

Table D.1: List of abbreviations used in Table D.2 that identify what changes to the baseline parameter regime have been made.

	CG ++, Both Cells	**CG ++, FO**	**Leader FN ++**	**Follower FN +**	**VE −/+, Both Cells**	**VE ++, Both Cells**	**VE −/+, LO**	**VE ++, LO**	**VE −/+, FO**	**VE ++, FO**	**Hap. Bias ++**	**Hap. Bias ++, FO**
CG ++, Both Cells	✓	n/a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CG ++, FO	n/a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Leader FN ++	✓	✓	X	X	X	X	X	X	X	X	X	X
Follower FN +	✓	✓	X	X	X	X	X	X	X	X	X	X
VE −/+, Both Cells	✓	✓	X	X	X	n/a	n/a	n/a	n/a	n/a	X	X
VE ++, Both Cells	✓	✓	X	X	n/a	X	n/a	n/a	n/a	n/a	X	X
VE −/+, LO	✓	✓	X	X	n/a	n/a	✓	n/a	X	✓	✓	✓
VE ++, LO	✓	✓	X	X	n/a	n/a	n/a	X	X	X	X	X
VE −/+, FO	✓	✓	X	X	n/a	n/a	X	X	X	n/a	X	X
VE ++, FO	✓	✓	X	X	n/a	n/a	✓	X	n/a	X	X	X
Hap. Bias ++	✓	✓	X	X	X	X	✓	X	X	X	X	n/a
Hap. Bias ++, FO	✓	✓	X	X	X	X	✓	X	X	X	n/a	X

Table D.2: Results of ABM parameter combinations that intend to prevent leader-follower separation. Green boxes with a "✓" denote tests for which the average minimum distance from leader cells to follower cells is less than two cell diameters (30 μm), orange boxes with a "X" denote experiments for which the minimum distance is above this threshold, and boxes with "n/a" denote tests that have not been performed. Unless otherwise noted, ABM parameter values are the same as in Table 5.1. For each test, a guiding force with weight $z = 0.75$ also acts only on leader cells. Abbreviations: see Table D.1.

References

Abercrombie M, Heaysman JE (1954) Observations on the social behaviour of cells in tissue culture: II. Monolayering of fibroblasts. Exp Cell Res 6(2):293–306, DOI 10.1016/0014-4827(54)90176-7

Alden K, Read M, Timmis J, et al. (2013) Spartan: A comprehensive tool for understanding uncertainty in simulations of biological systems. PLoS Comput Biol 9(2):e1002916, DOI 10.1371/journal.pcbi.1002916

Alnaes MS, Blechta J, Hake J, et al. (2015) The FEniCS Project Version 1.5. Archive of Numerical Software 3(100):9–23, DOI 10.11588/ans.2015.100.20553

An G, Mi Q, Dutta-Moscato J, et al. (2009) Agent-based models in translational systems biology. Wiley Interdiscip Rev Syst Biol Med 1(2):159–171, DOI 10.1002/wsbm.45

Anderson ARA, Chaplain MAJ (1998) Continuous and discrete mathematical models of tumor-induced angiogenesis. Bull Math Biol 60(5):857–900, DOI 10.1006/BULM.1998.0042

Arima S, Nishiyama K, Ko T, et al. (2011) Angiogenic morphogenesis driven by dynamic and heterogeneous collective endothelial cell movement. Development 138(21):4763–4776, DOI 10.1242/DEV.068023

Auerbach R, Lewis R, Shinners B, et al. (2003) Angiogenesis assays: A critical overview. Clin Chem 49(1):32–40, DOI 10.1373/49.1.32

Autenrieth TJ, Frank SC, Greiner AM, et al. (2016) Actomyosin contractility and RhoGTPases affect cell-polarity and directional migration during haptotaxis. Integr Biol 8(10):1067–1078, DOI 10.1039/c6ib00152a

Azimzade Y, Saberi AA, Sahimi M (2019) Regulation of migration of chemotactic tumor cells by the spatial distribution of collagen fiber orientation. Phys Rev E 99(6):062414, DOI 10.1103/PhysRevE.99.062414

Bahm I, Barriga EH, Frolov A, et al. (2017) PDGF controls contact inhibition of locomotion by regulating N-cadherin during neural crest migration. J Cell Sci 130(14):e1.2–e1.2, DOI 10.1242/jcs.207860

Bailey CM, Morrison JA, Kulesa PM (2012) Melanoma revives an embryonic migration program to promote plasticity and invasion. Pigment Cell Melanoma Res 25(5):573–583, DOI 10.1111/j.1755-148X.2012.01025.x

Bajanca F, Gouignard N, Colle C, et al. (2019) In vivo topology converts competition for cell-matrix adhesion into directional migration. Nat Comm 10:1518, DOI 10.1038/s41467-019-09548-5

Baker RE, Gaffney EA, Maini PK (2008) Partial differential equations for self-organization in cellular and developmental biology. Nonlinearity 21(11):R251, DOI 10.1088/0951-7715/21/11/R05

Baker RE, Yates CA, Erban R (2010) From microscopic to macroscopic descriptions of cell migration on growing domains. Bull Math Biol 72(3):719–762, DOI 10.1007/s11538-009-9467-x

Balding D, McElwain DLS (1985) A mathematical model of tumour-induced capillary growth. J Theor Biol 114(1):53–73, DOI 10.1016/S0022-5193(85)80255-1

Barriga EH, Franze K, Charras G, et al. (2018) Tissue stiffening coordinates morphogenesis by triggering collective cell migration in vivo. Nature 554(7693):523–527, DOI 10.1038/nature25742

Bauer AL, Jackson TL, Jiang Y (2007) A cell-based model exhibiting branching and anastomosis during tumor-induced angiogenesis. Biophys J 92(9):3105–21, DOI 10.1529/biophysj.106.101501

Bechtold B (2016) Violin Plots for Matlab, Github Project. DOI 10.5281/zenodo.4559847

Bender CM, Orszag SA (1999) Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory. Springer Science + Business Media, New York

Bentley K, Gerhardt H, Bates PA (2008) Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation. J Theor Biol 250(1):25–36, DOI 10.1016/J.JTBI.2007.09.015

Bentley K, Mariggi G, Gerhardt H, et al. (2009) Tipping the balance: Robustness of tip cell selection, migration and fusion in angiogenesis. PLoS Comput Biol 5(10):1000549, DOI 10.1371/journal.pcbi.1000549

Bernardi S, Eftimie R, Painter KJ (2021) Leadership through influence: What mechanisms allow leaders to steer a swarm? Bull Math Biol 83(6):1–33, DOI 10.1007/S11538-021-00901-8/FIGURES/12

Betz C, Lenard A, Belting HG, et al. (2016) Cell behaviors and dynamics during angiogenesis. Development 143(13):2249–2260, DOI 10.1242/dev.135616

Bhaskar D, Zhang WY, Wong IY (2021) Topological data analysis of collective and individual epithelial cells using persistent homology of loops. Soft Matter 17(17):4653, DOI 10.1039/d1sm00072a

Binny RN, Plank MJ, James A (2015) Spatial moment dynamics for collective cell movement incorporating a neighbour-dependent directional bias. J Roy Soc Interface 12(106):20150228, DOI 10.1098/rsif.2015.0228

Binny RN, Haridas P, James A, et al. (2016) Spatial structure arising from neighbourdependent bias in collective cell movement. PeerJ 2016(2):e1689, DOI 10.7717/peerj.1689

Blanco R, Gerhardt H (2013) VEGF and Notch in tip and stalk cell selection. Cold Spring Harb Perspect Biol 3(1):a006569, DOI 10.1101/cshperspect.a006569

Boas SEM, Jiang Y, Merks RMH, et al. (2018) Cellular Potts model: Applications to vasculogenesis and angiogenesis. In: Louis P-Y, Nardi FR (eds) Probabilistic Cellular Automata, Springer International Publishing, New York, chap 18, pp 279–310, DOI 10.1007/978-3-319-65558-1_18

Bonilla LL, Capasso V, Alvaro M, et al. (2014) Hybrid modeling of tumor-induced angiogenesis. Phys Rev E 90:062716, DOI 10.1103/PhysRevE.90.062716

Bonilla LL, Carretero M, Terragni F (2016) Solitonlike attractor for blood vessel tip density in angiogenesis. Phys Rev E 94:062415, DOI 10.1103/PhysRevE.94.062415

Bonilla LL, Capasso V, Alvaro M, et al. (2017) On the mathematical modelling of tumor-induced angiogenesis. Math Biosci Eng 14(1):45–66, DOI 10.3934/mbe.2017004

Bonilla LL, Carretero M, Terragni F (2020) Two dimensional soliton in tumor induced angiogenesis. J Stat Mech 8:83402, DOI 10.1088/1742-5468/aba598

Bowersox JC, Sorgente N (1982) Chemotaxis of aortic endothelial cells in response to fibronectin. Cancer Res 42(7):2547–2551

Bronner ME, LeDouarin NM (2012) Development and evolution of the neural crest: An overview. Dev Biol 366(1):2–9, DOI 10.1016/j.ydbio.2011.12.042

Browning AP, McCue SW, Binny RN, et al. (2018) Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. J Theor Biol 437:251–260, DOI 10.1016/j.jtbi.2017.10.032

Bruna M, Chapman SJ (2012) Diffusion of multiple species with excluded-volume effects. J Chem Phys 137(20):204116, DOI 10.1063/1.4767058

Bull JA, Macklin PS, Quaiser T, et al. (2020) Combining multiple spatial statistics enhances the description of immune cell localisation within tumours. Sci Rep 10(1):1–12, DOI 10.1038/s41598-020-75180-9

Buttenschön A, Hillen T, Gerisch A, et al. (2018) A space-jump derivation for non-local models of cell-cell adhesion and non-local chemotaxis. J Math Biol 76(1-2):429–456, DOI 10.1007/s00285-017-1144-3

Byrne HM (2010) Dissecting cancer through mathematics: From the cell to the animal model. Nat Rev Cancer 10:221–230, DOI 10.1038/nrc2808

Byrne HM, Chaplain MAJ (1995) Mathematical models for tumour angiogenesis: Numerical simulations and nonlinear wave solutions. Bull Math Biol 57(3):461–486, DOI 10.1007/BF02460635

Camley BA, Zimmermann J, Levine H, et al. (2016) Emergent collective chemotaxis without single-cell gradient sensing. Phys Rev Lett 116(9):098101, DOI 10.1103/PhysRevLett.116.098101

Carmeliet P, Jain RK (2011) Molecular mechanisms and clinical applications of angiogenesis. Nature 473(7347):298–307, DOI 10.1038/nature10144

Carmona-Fontaine C, Matthews HK, Kuriyama S, et al. (2008) Contact inhibition of locomotion in vivo controls neural crest directional migration. Nature 456(7224):957–961, DOI 10.1038/nature07441

Carmona-Fontaine C, Theveneau E, Tzekou A, et al. (2011) Complement fragment C3a controls mutual cell attraction during collective cell migration. Dev Cell 21(6):1026–1037, DOI 10.1016/j.devcel.2011.10.012

Carrillo JA, Fornasier M, Toscani G, et al. (2010) Particle, kinetic, and hydrodynamic models of swarming. In: Modeling and Simulation in Science, Engineering and Technology, vol 51, Springer Basel, pp 297–336, DOI 10.1007/978-0-8176-4946-3_12

Chan CJ, Heisenberg CP, Hiiragi T (2017) Coordination of morphogenesis and cell-fate specification in development. Curr Biol 27(18):R1024–R1035, DOI 10.1016/j.cub.2017.07.010

Chaplain MAJ, Lorenzi T, Macfarlane FR (2020) Bridging the gap between individual-based and continuum models of growing cell populations. J Math Biol 80(1-2):343–371, DOI 10.1007/s00285-019-01391-y

Cheeseman BL, Zhang D, Binder BJ, et al. (2014) Cell lineage tracing in the developing enteric nervous system: Superstars revealed by experiment and simulation. J Roy Soc Interface 11(93):20130815, DOI 10.1098/RSIF.2013.0815

Chen L, Painter K, Surulescu C, et al. (2020) Mathematical models for cell migration: A non-local perspective. Phil Trans R Soc B 375:20190379, DOI 10.1098/rstb.2019.0379

Coleman TF, Li Y (1996) A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. SIAM J Optimiz 6(4):1040–1058, DOI 10.1137/S1052623494240456

Colombi A, Scianna M, Painter KJ, et al. (2020) Modelling chase-and-run migration in heterogeneous populations. J Math Biol 80(1-2):423–456, DOI 10.1007/s00285-019-01421-9

Colombi A, Falletta S, Scianna M, et al. (2021) An integro-differential non-local model for cell migration and its efficient numerical solution. Math Comput Simul 180:179–204, DOI 10.1016/j.matcom.2020.08.020

Connor AJ, Nowak RP, Lorenzon E, et al. (2015) An integrated approach to quantitative modelling in angiogenesis research. J Roy Soc Interface 12(110):20150546, DOI 10.1098/rsif.2015.0546

D'Alessandro J, Barbier-Chebbah A, Cellerin V, et al. (2021) Cell migration guided by long-lived spatial memory. Nat Comm 12(1):1–10, DOI 10.1038/s41467-021-24249-8

Dallon J, Sherratt J, Maini PK, et al. (2000) Biological implications of a discrete mathematical model for collagen deposition and alignment in dermal wound repair. Math Med Biol 17(4):379–393, DOI 10.1093/imammb/17.4.379

Dallon JC, Othmer HG (2004) How cellular movement determines the collective force generated by the Dictyostelium discoideum slug. J Theor Biol 231(2):203–222, DOI 10.1016/J.JTBI.2004.06.015

Dallon JC, Sherratt JA, Maini PK (1999) Mathematical modelling of extracellular matrix dynamics using discrete cells: fiber orientation and tissue regeneration. J Theor Biol 199(4):449–471, DOI 10.1006/jtbi.1999.0971

Daub JT, Merks RMH (2013) A cell-based model of extracellular-matrix-guided endothelial cell migration during angiogenesis. Bull Math Biol 75:1377–1399, DOI 10.1007/s11538-013-9826-5

Davidson LA, Dzamba BD, Keller R, et al. (2008) Live imaging of cell protrusive activity, and extracellular matrix assembly and remodeling during morphogenesis in the frog, Xenopus laevis. Dev Dyn 237(10):2684–2692, DOI 10.1002/dvdy.21600

Davies KJ, Green JEF, Bean NG, et al. (2014) On the derivation of approximations to cellular automata models and the assumption of independence. Math Biosci 253(1):63–71, DOI 10.1016/j.mbs.2014.04.004

De Palma M, Biziato D, Petrova TV (2017) Microenvironmental regulation of tumour angiogenesis. Nat Rev Cancer 17(8):457–474, DOI 10.1038/nrc.2017.51

Driscoll TA, Hale N, Trefethen LN (eds) (2014) Chebfun Guide. Pafnuty Publications, Oxford

Duband JL, Thiery JP (1982) Distribution of fibronectin in the early phase of avian cephalic neural crest cell migration. Dev Biol 93(2):308–323, DOI 10.1016/0012-1606(82)90120-8

Duran CL, Howell DW, Dave JM, et al. (2017) Molecular regulation of sprouting angiogenesis. Compr Physiol 8(1):153–235, DOI 10.1002/cphy.c160048

Dyson L, Baker RE (2015) The importance of volume exclusion in modelling cellular migration. J Math Biol 71(3):691–711, DOI 10.1007/s00285-014-0829-0

Eckardt M, Painter KJ, Surulescu C, et al. (2020) Nonlocal and local models for taxis in cell migration: A rigorous limit procedure. J Math Biol 81(6-7):1251–1298, DOI 10.1007/s00285-020-01536-4

Edelstein L (1982) The propagation of fungal colonies: A model for tissue growth. J Theor Biol 98:679–701

Erickson CA (1985) Control of neural crest cell dispersion in the trunk of the avian embryo. Dev Biol 111(1):138–157, DOI 10.1016/0012-1606(85)90442-7

Fischer RS, Lam PY, Huttenlocher A, et al. (2019) Filopodia and focal adhesions: An integrated system driving branching morphogenesis in neuronal pathfinding and angiogenesis. Dev Biol 451(1):86–95, DOI 10.1016/J.YDBIO.2018.08.015

Flegg JA, Menon SN, Byrne HM, et al. (2020) A current perspective on wound healing and tumour-induced angiogenesis. Bull Math Biol 82(2):1–22, DOI 10.1007/s11538-020-00696-0

Folkman J, Hochberg M (1973) Self-regulation of growth in three dimensions. J Exp Med 138(4):745, DOI 10.1084/JEM.138.4.745

Folkman J, Klagsbrun M (1987) Angiogenic factors. Science 235(4787):442–447

Fong GH (2008) Mechanisms of adaptive angiogenesis to tissue hypoxia. Angiogenesis 11(2):121–140, DOI 10.1007/S10456-008-9107-3

Frantz C, Stewart KM, Weaver VM (2010) The extracellular matrix at a glance. J Cell Sci 123(24):4195–4200, DOI 10.1242/jcs.023820

Frisdal A, Trainor PA (2014) Development and evolution of the pharyngeal apparatus. Wiley Interdiscip Rev Dev Biol 3(6):403–418, DOI 10.1002/WDEV.147

Genuth MA, Allen CDC, Mikawa T, et al. (2018) Chick cranial neural crest cells use progressive polarity refinement, not contact inhibition of locomotion, to guide their migration. Dev Biol 444:S252–S261, DOI 10.1016/j.ydbio.2018.02.016

Gerhardt H, Golding M, Fruttiger M, et al. (2003) VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. J Cell Biol 161(6):1163–77, DOI 10.1083/jcb.200302047

Germann P, Marin-Riera M, Sharpe J (2019) yalla: GPU-Powered Spheroid Models for Mesenchyme and Epithelium. Cell Syst 8:261–266, DOI 10.1016/j.cels.2019.02.007

Ghaffarizadeh A, Heiland R, Friedman SH, et al. (2018) PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLoS Comput Biol 14(2):e1005991, DOI 10.1371/JOURNAL.PCBI.1005991

Giniunaite R, Baker RE, Kulesa PM, et al. (2019a) Modelling collective cell migration: neural crest as a model paradigm. J Math Biol pp 1–24, DOI 10.1007/s00285-019-01436-2

Giniunaite R, McLennan R, McKinney MC, et al. (2019b) An interdisciplinary approach to investigate collective cell migration in neural crest. Dev Dyn pp 1–11, DOI 10.1002/dvdy.124

Goeman JJ, Solari A (2014) Multiple hypothesis testing in genomics. Stat Med 33(11):1946–1978, DOI 10.1002/SIM.6082

Gordon-Weeks AN, Lim SY, Yuzhalin AE, et al. (2017) Neutrophils promote hepatic metastasis growth through fibroblast growth factor 2dependent angiogenesis in mice. Hepatology 65(6):1920–1935, DOI 10.1002/hep.29088

Grogan JA, Connor AJ, Markelc B, et al. (2017) Microvessel chaste: An open library for spatial modeling of vascularized tissues. Biophys J 112(9):1767–1772, DOI 10.1016/j.bpj.2017.03.036

Groh A, Louis AK (2010) Stochastic modelling of biased cell migration and collagen matrix modification. J Math Biol 61(5):617–647, DOI 10.1007/s00285-009-0314-3

Groh A, Wagner M (2011) Biased three-dimensional cell migration and collagen matrix modification. Math Biosci 231(2):105–119, DOI 10.1016/j.mbs.2011.02.007

Haeger A, Wolf K, Zegers MM, et al. (2015) Collective cell migration: Guidance principles and hierarchies. Trends Cell Biol 25(9):556–566, DOI 10.1016/j.tcb.2015.06.003

Hamis S, Stratiev S, Powathil GG (2021) Uncertainty and sensitivity analyses methods for agent-based mathematical models: An introductory review. In: Gerstman BS (ed) The Physics of Cancer, World Scientific Publishing Co. Pte. Ltd., pp 1–37, DOI 10.1142/9789811223495_0001

Hellström M, Phng LK, Hofmann JJ, et al. (2007) Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. Nature 445(7129):776–780, DOI 10.1038/nature05571

Helton JC, Davis FJ (2003) Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 81(1):23–69, DOI 10.1016/S0951-8320(03)00058-9

Herman J, Usher W (2017) SALib: An open-source Python library for Sensitivity Analysis. J Open Source Softw 2(9):97, DOI 10.21105/joss.00097

Hillen T (2006) M5 mesoscopic and macroscopic models for mesenchymal motion. J Math Biol 53(4):585–616, DOI 10.1007/s00285-006-0017-y

Hillen T, Painter KJ (2009) A user's guide to PDE models for chemotaxis. J Math Biol 58(1-2):183–217, DOI 10.1007/s00285-008-0201-3

Hillen T, Painter KJ (2013) Transport and anisotropic diffusion models for movement in oriented habitats. Lecture Notes in Mathematics 2071:177–222, DOI 10.1007/978-3-642-35497-7_7

Hinch EJ (1991) Perturbation Methods. Cambridge University Press, Cambridge

Hintze JL, Nelson RD (1998) Violin plots: A box plot-density trace synergism. Am Stat 52(2):181–184

Horne-Badovinac S, Bilder D (2005) Mass transit: Epithelial morphogenesis in the Drosophila egg chamber. Dev Dyn 232(3):559–574, DOI 10.1002/dvdy.20286

Hutchins EJ, Kunttas E, Piacentino ML, et al. (2018) Migration and diversification of the vagal neural crest. Dev Biol 444:S98–S109, DOI 10.1016/J.YDBIO.2018.07.004

Iman RL, Conover WJ (1987) A measure of top-down correlation. Technometrics 29(3):351–357

Jackson T, Zheng X (2010) A cell-based model of endothelial cell migration, proliferation and maturation during corneal angiogenesis. Bull Math Biol 72:830–868, DOI 10.1007/s11538-009-9471-1

Jakobsson L, Franco CA, Bentley K, et al. (2010) Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. Nat Cell Biol 12(10):943–953, DOI 10.1038/ncb2103

Johnston ST, Painter KJ (2019) The impact of short- and long-range perception on population movements. J Theor Biol 460:227–242, DOI 10.1016/j.jtbi.2018.10.031

Kang TY, Bocci F, Jolly MK, et al. (2019) Pericytes enable effective angiogenesis in the presence of proinflammatory signals. Proc Natl Acad Sci USA 116(47):23551–23561, DOI 10.1073/PNAS.1913373116

Kato T, Jenkins RP, Derzsi S, et al. (2021) Interplay of adherens junctions and matrix proteolysis determines the invasive pattern and growth of squamous cell carcinoma. bioRxiv p 2021.11.24.469674, DOI 10.1101/2021.11.24.469674

Kechagia JZ, Ivaska J, Roca-Cusachs P (2019) Integrins as biomechanical sensors of the microenvironment. Nat Rev Mol Cell Biol 20(8):457–473, DOI 10.1038/s41580-019-0134-2

Kerosuo L, Bronner-Fraser M (2012) What is bad in cancer is good in the embryo: Importance of EMT in neural crest development. Semin Cell Dev Biol 23(3):320–332, DOI 10.1016/J.SEMCDB.2012.03.010

Khataee H, Czirok A, Neufeld Z (2021) Contact inhibition of locomotion generates collective cell migration without chemoattractants in an open domain. Phys Rev E 104(1):014405, DOI 10.1103/PhysRevE.104.014405

Kulesa P, Ellies DL, Trainor PA (2004) Comparative analysis of neural crest cell death, migration, and function during vertebrate embryogenesis. Dev Dyn 229(1):14–29, DOI 10.1002/dvdy.10485

Kulesa PM, Fraser SE (1998) Neural crest cell dynamics revealed by time-lapse video microscopy of whole embryo chick explant cultures. Dev Biol 204(2):327–344, DOI 10.1006/dbio.1998.9082

Kulesa PM, Kasemeier-Kulesa JC, Teddy JM, et al. (2006) Reprogramming metastatic melanoma cells to assume a neural crest cell-like pnenotype in an embryonic microenvironment. Proc Natl Acad Sci USA 103(10):3752–3757, DOI 10.1073/pnas.0506977103

Kulesa PM, Bailey CM, Kasemeier-Kulesa JC, et al. (2010) Cranial neural crest migration: New rules for an old road. Dev Biol 344(2):543–554, DOI 10.1016/j.ydbio.2010.04.010

Kulesa PM, Morrison JA, Bailey CM (2013) The neural crest and cancer: A developmental spin on melanoma. Cells Tissues Organs 198(1):12–21, DOI 10.1159/000348418

Kulesa PM, Kasemeier-Kulesa JC, Morrison JA, et al. (2021) Modelling cell invasion: A review of what JD Murray and the embryo can teach us. Bull Math Biol 83(4):1–13, DOI 10.1007/s11538-021-00859-7

Lagache T, Meas-Yedid V, Olivo-Marin JC (2013) A statistical analysis of spatial colocalization using Ripley's K function. In: IEEE International Symposium on Biomedical Imaging, pp 896–901, DOI 10.1109/ISBI.2013.6556620

Landman KA, Fernando AE, Zhang D, et al. (2011) Building stable chains with motile agents: Insights into the morphology of enteric neural crest cell migration. J Theor Biol 276(1):250–268, DOI 10.1016/J.JTBI.2011.01.043

LeDouarin NM, Kalcheim C (1999) The Neural Crest, 2nd edn. Cambridge University Press, Cambridge

Lenard A, Ellertsdottir E, Herwig L, et al. (2013) In vivo analysis reveals a highly stereotypic morphogenetic pathway of vascular anastomosis. Dev Cell 25(5):492–506, DOI 10.1016/J.DEVCEL.2013.05.010

Leonard CE, Taneyhill LA (2020) The road best traveled: Neural crest migration upon the extracellular matrix. Semin Cell Dev Biol 100:177–185, DOI 10.1016/j.semcdb.2019.10.013

Li Y, Vieceli FM, Gonzalez WG, et al. (2019) In vivo quantitative imaging provides insights into trunk neural crest migration. Cell Rep 26(6):1489–1500.e3, DOI 10.1016/J.CELREP.2019.01.039

Lorenzi T, Painter KJ (2021) Trade-offs between chemotaxis and proliferation shape the phenotypic structuring of invading waves. Int J Non Linear Mech in press

Loy N, Preziosi L (2020) Kinetic models with non-local sensing determining cell polarization and speed according to independent cues. J Math Biol 80(1-2):373–421, DOI 10.1007/s00285-019-01411-x

Lum PY, Singh G, Lehman A, et al. (2013) Extracting insights from the shape of complex data using topology. Sci Rep 3(1):1–8, DOI 10.1038/srep01236

Lumsden A, Sprawson N, Graham A (1991) Segmental origin and migration of neural crest cells in the hindbrain region of the chick embryo. Development 113(4):1281–1291, DOI 10.1242/DEV.113.4.1281

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 18(1):50–60, DOI 10.1214/AOMS/1177730491

Mantzaris NV, Webb S, Othmer HG (2004) Mathematical modeling of tumor-induced angiogenesis. J Math Biol 49(2):111–187, DOI 10.1007/s00285-003-0262-2

Mardia K, Jupp P (2000) Directional Statistics. John Wiley and Sons, Ltd, Chichester

Marino S, Hogue IB, Ray CJ, et al. (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. J Theor Biol 254(1):178–196, DOI 10.1016/j.jtbi.2008.04.011

Markham DC, Simpson MJ, Baker RE (2013) Simplified method for including spatial correlations in mean-field approximations. Phys Rev E 87(6):62702, DOI 10.1103/PhysRevE.87.062702

Martins GG, Rifes P, Amândio R, et al. (2009) Dynamic 3D cell rearrangements guided by a fibronectin matrix underlie somitogenesis. PLoS ONE 4(10):e7429, DOI 10.1371/JOURNAL.PONE.0007429

Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 46(253):68–78

Mathias S, Coulier A, Bouchnita A, et al. (2020) Impact of force function formulations on the numerical simulation of centre-based models. Bull Math Biol 82(10):132, DOI 10.1007/s11538-020-00810-2

Maxwell PH, Dachs GU, Gleadle JM, et al. (1997) Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth. Proc Natl Acad Sci USA 94(15):8104–8109, DOI 10.1073/PNAS.94.15.8104

McDougall S, Dallon J, Sherratt J, et al. (2006) Fibroblast migration and collagen deposition during dermal wound healing: Mathematical modelling and clinical implications. Philos Trans Royal Soc A 364(1843):1385–1405, DOI 10.1098/rsta.2006.1773

McKay MD, Beckman RJ, Conover WJ, et al. (2000) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 42(1):55–61, DOI 10.1080/00401706.2000.10485979

McKinney MC, McLennan R, Kulesa PM (2016) Angiopoietin 2 signaling plays a critical role in neural crest cell migration. BMC Biol 14(1):111, DOI 10.1186/s12915-016-0323-9

McKinney MC, McLennan R, Giniunaite R, et al. (2020) Visualizing mesoderm and neural crest cell dynamics during chick head morphogenesis. Dev Biol 461(2):184–196, DOI 10.1016/j.ydbio.2020.02.010

McLennan R, Kulesa PM (2010) Neuropilin-1 interacts with the second branchial arch microenvironment to mediate chick neural crest cell dynamics. Dev Dyn 239(6):1664–1673, DOI 10.1002/dvdy.22303

McLennan R, Teddy JM, Kasemeier-Kulesa JC, et al. (2010) Vascular endothelial growth factor (VEGF) regulates cranial neural crest migration in vivo. Dev Biol 339(1):114–125, DOI 10.1016/j.ydbio.2009.12.022

McLennan R, Dyson L, Prather KW, et al. (2012) Multiscale mechanisms of cell migration during development: Theory and experiment. Development 139(16):2935–2944, DOI 10.1242/dev.081471

McLennan R, Schumacher LJ, Morrison JA, et al. (2015a) Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front. Development 142(11):2014–2025, DOI 10.1242/dev.117507

McLennan R, Schumacher LJ, Morrison JA, et al. (2015b) VEGF signals induce trailblazer cell identity that drives neural crest migration. Dev Biol 407(1):12–25, DOI 10.1016/j.ydbio.2015.08.011

McLennan R, Bailey CM, Schumacher LJ, et al. (2017) DAN (NBL1) promotes collective neural crest migration by restraining uncontrolled invasion. J Cell Biol 216(10):3339–3354, DOI 10.1083/jcb.201612169

McLennan R, McKinney MC, Teddy JM, et al. (2020) Neural crest cells bulldoze through the microenvironment using Aquaporin 1 to stabilize filopodia. Development 147(1), DOI 10.1242/dev.185231

Merchant B, Edelstein-Keshet L, Feng JJ (2018) A Rho-GTPase based model explains spontaneous collective migration of neural crest cell clusters. Dev Biol 444:S262–S273, DOI 10.1016/j.ydbio.2018.01.013

Metzcar J, Wang Y, Heiland R, et al. (2019) A review of cell-based computational modeling in cancer biology. JCO Clin Cancer Inform 2(3):1–13, DOI 10.1200/cci.18.00069

Mittal A, Pulina M, Hou SY, et al. (2010) Fibronectin and integrin alpha 5 play essential roles in the development of the cardiac neural crest. Mechanisms of Development 127(9-12):472–484, DOI 10.1016/j.mod.2010.08.005

Morale D, Capasso V, Oelschläger K (2005) An interacting particle system modelling aggregation behavior: From individuals to populations. J Math Biol 50(1):49–66, DOI 10.1007/s00285-004-0279-1

Moreira-Soares M, Coimbra R, Rebelo L, et al. (2018) Angiogenic factors produced by hypoxic cells are a leading driver of anastomoses in sprouting angiogenesis - a computational study. Sci Rep 8(1):1–12, DOI 10.1038/s41598-018-27034-8

Morrison JA, McLennan R, Wolfe LA, et al. (2017) Single-cell transcriptome analysis of avian neural crest migration reveals signatures of invasion and molecular transitions. eLife 6:e28415, DOI 10.7554/eLife.28415

Motsch S, Peurichard D (2018) From short-range repulsion to Hele-Shaw problem in a model of tumor growth. J Math Biol 76(1-2):205–234, DOI 10.1007/s00285-017-1143-4

Muthukkaruppan VR, Kubai L, Auerbach R (1982) Tumor-induced neovascularization in the mouse eye. J Natl Cancer Inst 69(3):699–708, DOI 10.1093/jnci/69.3.699

Nesmith JE, Chappell JC, Cluceru JG, et al. (2017) Blood vessel anastomosis is spatially regulated by Flt1 during angiogenesis. Development 144(5):889–896, DOI 10.1242/DEV.145672

Newgreen D, Thiery JP (1980) Fibronectin in early avian embryos: Synthesis and distribution along the migration pathways of neural crest cells. Cell Tissue Res 211(2):269–291, DOI 10.1007/BF00236449

Newgreen DF (1989) Physical influences on neural crest cell migration in avian embryos: Contact guidance and spatial restriction. Dev Biol 131(1):136–148, DOI 10.1016/S0012-1606(89)80045-4

Nowak-Sliwinska P, Alitalo K, Allen E, et al. (2018) Consensus guidelines for the use and interpretation of angiogenesis assays. Angiogenesis 21(3):425–532, DOI 10.1007/s10456-018-9613-x

Nuhn JA, Perez AM, Schneider IC (2018) Contact guidance diversity in rotationally aligned collagen matrices. Acta Biomater 66:248–257, DOI 10.1016/J.ACTBIO.2017.11.039

Odell GM, Oster G, Alberch P, et al. (1981) The mechanical basis of morphogenesis: I. Epithelial folding and invagination. Dev Biol 85(2):446–462, DOI 10.1016/0012-1606(81)90276-1

van Oers RF, Rens EG, LaValley DJ, et al. (2014) Mechanical cell-matrix feedback explains pairwise and collective endothelial cell behavior in vitro. PLoS Comput Biol 10(8):1003774, DOI 10.1371/journal.pcbi.1003774

Olsen L, Sherratt JA, Maini PK, et al. (1997) A mathematical model for the capillary endothelial cell-extracellular matrix interactions in wound-healing angiogenesis. Mathematical Medicine and Biology 14(4):261–281, DOI 10.1093/imammb/14.4.261

Orme ME, Chaplain MAJ (1997) Two-dimensional models of tumour angiogenesis and anti-angiogenesis strategies. IMA J Math Med Biol 14:189–205

Osborne JM, Fletcher AG, Pitt-Francis JM, et al. (2017) Comparing individual-based approaches to modelling the self-organization of multicellular tissues. PLoS Comput Biol 13(2):e1005387, DOI 10.1371/journal.pcbi.1005387

Ostrovsky D, Cheney CM, Seitz AW, et al. (1983) Fibronectin distribution during somitogenesis in the chick embryo. Cell Differ 13(3):217–223, DOI 10.1016/0045-6039(83)90092-1

Othmer HG, Dunbar SR, Alt W (1988) Models of dispersal in biological systems. J Math Biol 26(3):263–298, DOI 10.1007/BF00277392

Otter N, Porter MA, Tillmann U, et al. (2015) A roadmap for the computation of persistent homology. EPJ Data Science 6(1), DOI 10.1140/epjds/s13688-017-0109-5

Painter KJ (2009) Modelling cell migration strategies in the extracellular matrix. J Math Biol 58:511–543, DOI 10.1007/s00285-008-0217-8

Painter KJ, Hillen T (2018) From random walks to fully anisotropic diffusion models for cell and animal movement. In: Modeling and Simulation in Science, Engineering and Technology, Springer Basel, pp 103–141, DOI 10.1007/978-3-319-96842-1_5

Painter KJ, Bloomfield JM, Sherratt JA, et al. (2015) A nonlocal model for contact attraction and repulsion in heterogeneous cell populations. Bull Math Biol 77(6):1132–1165, DOI 10.1007/s11538-015-0080-x

Palsson E (2008) A 3-D model used to explore how cell adhesion and stiffness affect cell sorting and movement in multicellular systems. J Theor Biol 254(1):1–13, DOI 10.1016/J.JTBI.2008.05.004

Palsson E, Othmer HG (2000) A model for individual and collective cell movement in Dictyostelium discoideum. Proc Natl Acad Sci USA 97(19):10448–10453, DOI 10.1073/PNAS.97.19.10448

Park D, Wershof E, Boeing S, et al. (2019) Extracellular matrix anisotropy is determined by TFAP2C-dependent regulation of cell collisions. Nat Mat DOI 10.1038/s41563-019-0504-3

Paweletz N, Knierim M (1989) Tumor-related angiogenesis. Crit Rev Oncol Hematol 9(3):197–242, DOI 10.1016/S1040-8428(89)80002-2

Penington CJ, Hughes BD, Landman KA (2011) Building macroscale models from microscale probabilistic models: A general probabilistic approach for nonlinear diffusion and multispecies phenomena. Phys Rev E 84:41120, DOI 10.1103/PhysRevE.84.041120

Penington CJ, Hughes BD, Landman KA (2014) Interacting motile agents: Taking a mean-field approach beyond monomers and nearest-neighbor steps. Phys Rev E 89:32714, DOI 10.1103/PhysRevE.89.032714

Perfahl H, Hughes BD, Alarcón T, et al. (2017) 3D hybrid modelling of vascular network formation. J Theor Biol 414:254–268, DOI 10.1016/J.JTBI.2016.11.013

Perumpanani AJ, Byrne HM (1999) Extracellular matrix concentration exerts selection pressure on invasive cells. Eur J Cancer 35(8):1274–1280, DOI 10.1016/S0959-8049(99)00125-2

Pettet G, Chaplain MAJ, McElwain DLS, et al. (1996a) On the role of angiogenesis in wound healing. Proc Roy Soc B 263(1376):1487–1493, DOI 10.1098/rspb.1996.0217

Pettet GJ, Byrne HM, McElwain DLS, et al. (1996b) A model of wound-healing angiogenesis in soft tissue. Math Biosci 136(1):35–63, DOI 10.1016/0025-5564(96)00044-2

Piacentino ML, Li Y, Bronner ME (2020) Epithelial-to-mesenchymal transition and different migration strategies as viewed from the neural crest. Curr Opin Cell Biol 66:43–50, DOI 10.1016/j.ceb.2020.05.001

Pillay S, Byrne HM, Maini PK (2017) Modeling angiogenesis: A discrete to continuum description. Phys Rev E 95(1):1–12, DOI 10.1103/PhysRevE.95.012410

Pillay S, Byrne HM, Maini PK (2018) The impact of exclusion processes on angiogenesis models. J Math Biol 77(6-7):1721–1759, DOI 10.1007/s00285-018-1214-1

Pitt-Francis J, Pathmanathan P, Bernabeu MO, et al. (2009) Chaste: A test-driven approach to software development for biological modelling. Comput Phys Commun 180(12):2452–2471, DOI 10.1016/j.cpc.2009.07.019

Potente M, Gerhardt H, Carmeliet P (2011) Basic and therapeutic aspects of angiogenesis. Cell 146(6):873–87, DOI 10.1016/j.cell.2011.08.039

Pramanik D, Jolly M, Bhat R (2021) Matrix adhesion and remodeling diversifies modes of cancer invasion across spatial scales. J Theor Biol 524:110733, DOI 10.1016/j.jtbi.2021.110733

Provenzano PP, Inman DR, Eliceiri KW, et al. (2008) Contact guidance mediated three-dimensional cell migration is regulated by Rho/ROCK-dependent matrix reorganization. Biophys J 95(11):5374–5384, DOI 10.1529/BIOPHYSJ.108.133116

Pugh CW, Ratcliffe PJ (2003) Regulation of angiogenesis by hypoxia: Role of the HIF system. Nat Med 9(6):677–684, DOI 10.1038/nm0603-677

Qian G, Mahdi A (2020) Sensitivity analysis methods in the biomedical sciences. Math Biosci 323:108306, DOI 10.1016/j.mbs.2020.108306

Qutub AA, Popel AS (2009) Elongation, proliferation and migration differentiate endothelial cell phenotypes and determine capillary sprouting. BMC Syst Biol 3(1):13, DOI 10.1186/1752-0509-3-13

Read M, Andrews PS, Timmis J, et al. (2012) Techniques for grounding agent-based simulations in the real domain: a case study in experimental autoimmune encephalomyelitis. Math Comput Model Dyn Syst 18(1):67–86, DOI 10.1080/13873954.2011.601419

Rejniak KA, Anderson ARA (2011) Hybrid models of tumor growth. Wiley Interdiscip Rev Syst Biol Med 3(1):115–125, DOI 10.1002/wsbm.102

Renardy M, Hult C, Evans S, et al. (2019) Global sensitivity analysis of biological multiscale models. Curr Opin Biomed Eng 11:109–116, DOI 10.1016/j.cobme.2019.09.012

Richardson J, Gauert A, Briones Montecinos L, et al. (2016) Leader cells define directionality of trunk, but not cranial, neural crest cell migration. Cell Rep 15(9):2076–2088, DOI 10.1016/j.celrep.2016.04.067

Ridenour DA, Mclennan R, Teddy JM, et al. (2014) The neural crest cell cycle is related to phases of migration in the head. Development 141(5):1095–1103, DOI 10.1242/dev.098855

Ripley BD (1981) Spatial Statistics. Wiley, Chichester

Risau W (1997) Mechanisms of angiogenesis. Nature 386(6626):671–674, DOI 10.1038/386671a0

Rørth P (2009) Collective cell migration. Annu Rev Cell Dev Biol 25:407–429, DOI 10.1146/annurev.cellbio.042308.113231

Rovasio RA, Delouvee A, Yamada KM, et al. (1983) Neural crest cell migration: Requirements for exogenous fibronectin and high cell density. J Cell Biol 96(2):462–473, DOI 10.1083/jcb.96.2.462

Roycroft A, Szabó A, Bahm I, et al. (2018) Redistribution of adhesive forces through Src/FAK drives contact inhibition of locomotion in neural crest. Dev Cell 45(5):565–579.e3, DOI 10.1016/j.devcel.2018.05.003

Rozario T, Dzamba B, Weber GF, et al. (2009) The physical state of fibronectin matrix differentially regulates morphogenetic movements in vivo. Dev Biol 327(2):386–398, DOI 10.1016/j.ydbio.2008.12.025

Ruhrberg C, Gerhardt H, Golding M, et al. (2002) Spatially restricted patterning cues provided by heparin-binding VEGF-A control blood vessel branching morphogenesis. Genes Dev 16(20):2684–2698, DOI 10.1101/gad.242002

Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis. Environ Model Softw 25(12):1508–1517, DOI 10.1016/j.envsoft.2010.04.012

Saltelli A, Tarantola S, Chan PS (1999) A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41(1):39–56, DOI 10.1080/00401706.1999.10485594

Saltelli A, Aleksankina K, Becker W, et al. (2019) Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environ Model Softw 114:29–39, DOI 10.1016/j.envsoft.2019.01.012

Samanta D, Semenza GL (2018) Metabolic adaptation of cancer and immune cells mediated by hypoxia-inducible factors. Biochim Biophys Acta Rev Cancer 1870(1):15–22, DOI 10.1016/J.BBCAN.2018.07.002

Scarpa E, Mayor R (2016) Collective cell migration in development. J Cell Biol 212(2):143–155, DOI 10.1083/jcb.201508047

Scarpa E, Szabó A, Bibonne A, et al. (2015) Cadherin switch during EMT in neural crest cells leads to contact inhibition of locomotion via repolarization of forces. Dev Cell 34(4):421–434, DOI 10.1016/j.devcel.2015.06.012

Schiesser WE (1991) The Numerical Method of Lines: Integration of Partial Differential Equations. Academic Press, San Diego

Schumacher LJ (2019) Neural crest migration with continuous cell states. J Theor Biol 481:84–90, DOI 10.1016/j.jtbi.2019.01.029

Schumacher LJ, Kulesa PM, McLennan R, et al. (2016) Multidisciplinary approaches to understanding collective cell migration in developmental biology. Open Biol 6:160056, DOI 10.1098/rsob.160056

Schwarzbauer JE, DeSimone DW (2011) Fibronectins, their fibrillogenesis, and in vivo functions. Cold Spring Harb Perspect Biol 3(7):1–19, DOI 10.1101/cshperspect.a005041

Scianna M, Munaron L, Preziosi L (2011) A multiscale hybrid approach for vasculogenesis and related potential blocking therapies. Prog Biophys Mol Biol 106(2):450–462, DOI 10.1016/j.pbiomolbio.2011.01.004

Scianna M, Bell CG, Preziosi L (2013) A review of mathematical models for the formation of vascular networks. J Theor Biol 333:174–209, DOI 10.1016/J.JTBI.2013.04.037

Selleck M, Bronner-Fraser M (1995) Origins of the avian neural crest: The role of neural plate-epidermal interactions. Development 121(2):525–538, DOI 10.1242/DEV.121.2.525

Sfakianakis N, Madzvamuse A, Chaplain MA (2020) A hybrid multiscale model for cancer invasion of the extracellular matrix. Multiscale Model Sim 18(2):824–850, DOI 10.1137/18M1189026

Shampine LF, Reichelt MW (1997) The MATLAB ODE suite. SIAM J Sci Comput 18(1):1–22, DOI 10.1137/S1064827594276424

Shellard A, Mayor R (2019) Integrating chemical and mechanical signals in neural crest cell migration. Curr Opin Genet Dev 57:16–24, DOI 10.1016/j.gde.2019.06.004

Shellard A, Mayor R (2020) Rules of collective migration: From the wildebeest to the neural crest. Phil Trans R Soc B 375:20190387, DOI 10.1098/rstb.2019.0387

Shellard A, Szabó A, Trepat X, et al. (2018) Supracellular contraction at the rear of neural crest cell groups drives collective chemotaxis. Science 362(6412):339–343, DOI 10.1126/science.aau3301

Simpson MJ, Baker RE (2011) Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena. Phys Rev E 83:51922, DOI 10.1103/PhysRevE.83.051922

Simpson MJ, Zhang DC, Mariani M, et al. (2007) Cell proliferation drives neural crest cell invasion of the intestine. Dev Biol 302(2):553–568, DOI 10.1016/J.YDBIO.2006.10.017

Simpson MJ, Landman KA, Hughes BD (2009) Multi-species simple exclusion processes. Physica A 388:399–406, DOI 10.1016/j.physa.2008.10.038

Singh P, Carraher C, Schwarzbauer JE (2010) Assembly of fibronectin extracellular matrix. Annu Rev Cell Dev Biol 26:397–419, DOI 10.1146/annurev-cellbio-100109-104020

Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 55(1-3):271–280, DOI 10.1016/S0378-4754(00)00270-6

Spill F, Guerrero P, Alarcon T, et al. (2015) Mesoscopic and continuum modelling of angiogenesis. J Math Biol 70(3):485–532, DOI 10.1007/s00285-014-0771-1

Sra S (2012) A short note on parameter approximation for von Mises-Fisher distributions: And a fast implementation of Is(x). Comput Stat 27(1):177–190, DOI 10.1007/s00180-011-0232-x

Stahl A, Connor KM, Sapieha P, et al. (2010) The mouse retina as an angiogenesis model. Investig Ophthalmol Vis Sci 51(6):2813–2826, DOI 10.1167/iovs.10-5176

Stenzel D, Lundkvist A, Sauvaget D, et al. (2011) Integrin-dependent and -independent functions of astrocytic fibronectin in retinal angiogenesis. Development 138(20):4451–4463, DOI 10.1242/DEV.071381

Stepanova D, Byrne HM, Maini PK, et al. (2021) A multiscale model of complex endothelial cell dynamics in early angiogenesis. PLoS Comput Biol 17(1):e1008055, DOI 10.1371/JOURNAL.PCBI.1008055

Stokes CL, Lauffenburger DA, Williams SK (1991) Migration of individual microvessel endothelial cells: Stochastic model and parameter measurement. J Cell Sci 99(2):419–430

Stolz BJ, Harrington HA, Porter MA (2017) Persistent homology of time-dependent functional networks constructed from coupled time series. Chaos 27(4):47410, DOI 10.1063/1.4978997

Stramer B, Mayor R (2016) Mechanisms and in vivo functions of contact inhibition of locomotion. Nat Rev Mol Cell Biol 18(1):43–55, DOI 10.1038/nrm.2016.118

Suveges S, Chamseddine I, Rejniak KA, et al. (2021) Collective cell migration in a fibrous environment: A hybrid multiscale modelling approach. Front Appl Math Stat 7:34, DOI 10.3389/FAMS.2021.680029

Swat MH, Hester SD, Balter A, et al. (2009) Multicell simulations of development and disease using the CompuCell3D simulation environment. In: Maly IV (ed) Systems Biology, Humana Press, New York, chap 13, pp 361–428

Szabó A, Mayor R (2018) Mechanisms of neural crest migration. Annu Rev Genet 52(1):43–63, DOI 10.1146/annurev-genet-120417-031559

Szabó A, Melchionda M, Nastasi G, et al. (2016) In vivo confinement promotes collective migration of neural crest cells. J Cell Biol 213(5):543–555, DOI 10.1083/jcb.201602083

Szabó A, Theveneau E, Turan M, et al. (2019) Neural crest streaming as an emergent property of tissue interactions during morphogenesis. PLoS Comput Biol 15(4):e1007002, DOI 10.1371/journal.pcbi.1007002

Tang W, Bronner ME (2020) Neural crest lineage analysis: From past to future trajectory. Development 147(20):dev193193, DOI 10.1242/dev.193193

Teddy JM, Kulesa PM (2004) In vivo evidence for short- and long-range cell communication in cranial neural crest cells. Development 131(24):6141–6151, DOI 10.1242/dev.01534

Teixeira AI, Abrams GA, Bertics PJ, et al. (2003) Epithelial contact guidance on well-defined micro- and nanostructured substrates. J Cell Sci 116(10):1881–1892, DOI 10.1242/JCS.00383

Théry M, Racine V, Piel M, et al. (2006) Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. Proc Natl Acad Sci USA 103(52):19771–19776, DOI 10.1073/pnas.0609267103

Theveneau E, Mayor R (2012) Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. Dev Biol 366(1):34–54, DOI 10.1016/j.ydbio.2011.12.041

Theveneau E, Marchant L, Kuriyama S, et al. (2010) Collective chemotaxis requires contact-dependent cell polarity. Dev Cell 19(1):39–53, DOI 10.1016/j.devcel.2010.06.012

Theveneau E, Steventon B, Scarpa E, et al. (2013) Chase-and-run between adjacent cell populations promotes directional collective migration. Nat Cell Biol 15(7):763–772, DOI 10.1038/ncb2772

Tong RT, Boucher Y, Kozin SV, et al. (2004) Vascular normalization by vascular endothelial growth factor receptor 2 blockade induces a pressure gradient across the vasculature and improves drug penetration in tumors. Cancer Res 64(11):3731–3736, DOI 10.1158/0008-5472.CAN-04-0074

Trainor P, Tam P (1995) Cranial paraxial mesoderm and neural crest cells of the mouse embryo: co-distribution in the craniofacial mesenchyme but distinct segregation in branchial arches. Development 121(8):2569–2582, DOI 10.1242/DEV.121.8.2569

Travasso RDM, Poiré EC, Castro M, et al. (2011) Tumor angiogenesis and vascular patterning: A mathematical model. PLoS ONE 6(5):e19989, DOI 10.1371/journal.pone.0019989

Tweedy L, Thomason PA, Paschke PI, et al. (2020) Seeing around corners: Cells solve mazes and respond at a distance using attractant breakdown. Science 369(6507), DOI 10.1126/science.aay9792

Ubezio B, Blanco RA, Geudens I, et al. (2016) Synchronization of endothelial Dll4-Notch dynamics switch blood vessels from branching to expansion. eLife 5(APRIL2016), DOI 10.7554/ELIFE.12167

Uccelli A, Wolff T, Valente P, et al. (2019) Vascular endothelial growth factor biology for regenerative angiogenesis. Swiss Med Wkly 149(0304):w20011, DOI 10.4414/smw.2019.20011

Usuelli M, Meyer T, Mezzenga R, et al. (2021) VEGF and VEGFR2 bind to similar pH-sensitive sites on fibronectin, exposed by heparin-mediated conformational changes. J Biol Chem 296:100584, DOI 10.1016/J.JBC.2021.100584

Van Liedekerke P, Palm MM, Jagiella N, et al. (2015) Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. Comput Part Mech 2(4):401–444, DOI 10.1007/s40571-015-0082-3

Vargha A, Delaney HD (2000) A critique and improvement of the CL common language effect size statistics of McGraw and Wong. J Educ Behav Stat 25(2):101–132, DOI 10.3102/10769986025002101

Vega ME, Schwarzbauer JE (2016) Collaboration of fibronectin matrix with other extracellular signals in morphogenesis and differentiation. Curr Opin Cell Biol 42:1–6, DOI 10.1016/j.ceb.2016.03.014

Vega R, Carretero M, Travasso RD, et al. (2020) Notch signaling and taxis mechanims regulate early stage angiogenesis: A mathematical and computational model. PLoS Comput Biol 16(1):e1006919, DOI 10.1371/journal.pcbi.1006919

Vega-Lopez GA, Cerrizuela S, Tribulo C, et al. (2018) Neurocristopathies: New insights 150 years after the neural crest discovery. Dev Biol 444:S110–S143, DOI 10.1016/J.YDBIO.2018.05.013

Verhulst F (2005) Methods and Applications of Singular Perturbations: Boundary layers and Multiple Timescale Dynamics. Springer, New York

Viallard C, Larrivée B (2017) Tumor angiogenesis and vascular normalization: Alternative therapeutic targets. Angiogenesis 20(4):409–426, DOI 10.1007/s10456-017-9562-9

Vicsek T, Czirok A, Ben-Jacob E, et al. (1995) Novel type of phase transition in a system of self-driven particles. Phys Rev Lett 75(6):1226–1229, DOI 10.1103/PhysRevLett.75.1226

Vilanova G, Colominas I, Gomez H (2013) Capillary networks in tumor angiogenesis: From discrete endothelial cells to phase-field averaged descriptions via isogeometric analysis. Int J Numer Method Biomed Eng 29(10):1015–1037, DOI 10.1002/cnm.2552

Vilanova G, Colominas I, Gomez H (2014) Coupling of discrete random walks and continuous modeling for three-dimensional tumor-induced angiogenesis. Comput Mech 53(3):449–464, DOI 10.1007/s00466-013-0958-0

Vilanova G, Colominas I, Gomez H (2017) Computational modeling of tumor-induced angiogenesis. Arch Comput Methods Eng 24(4):1071–1102, DOI 10.1007/s11831-016-9199-7

Vilanova G, Burés M, Colominas I, et al. (2018) Computational modelling suggests complex interactions between interstitial flow and tumour angiogenesis. J Roy Soc Interface 15:20180415, DOI 10.1098/rsif.2018.0415

Vipond O, Bull JA, Macklin PS, et al. (2021) Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors. Proc Natl Acad Sci USA 118(41):e2102166118, DOI 10.1073/pnas.2102166118

Walma DAC, Yamada KM (2020) The extracellular matrix in development. Development 147(10):dev175596, DOI 10.1242/dev.175596

Wang X, Astrof S (2016) Neural crest cell-autonomous roles of fibronectin in cardiovascular development. Development 143(1):88–100, DOI 10.1242/dev.125286

Wershof E, Park D, Jenkins RP, et al. (2019) Matrix feedback enables diverse higher-order patterning of the extracellular matrix. PLoS Comput Biol 15(10):e1007251, DOI 10.1371/journal.pcbi.1007251

Woods ML, Carmona-Fontaine C, Barnes CP, et al. (2014) Directional collective cell migration emerges as a property of cell interactions. PLoS ONE 9(9):e104969, DOI 10.1371/journal.pone.0104969

Wrede F, Hellander A (2019) Smart computational exploration of stochastic gene regulatory network models using human-in-the-loop semi-supervised learning. Bioinformatics 35(24):5199–5206, DOI 10.1093/bioinformatics/btz420

Wu PH, Giri A, Sun SX, et al. (2014) Three-dimensional cell migration does not follow a random walk. Proc Natl Acad Sci USA 111(11):3949–3954, DOI 10.1073/pnas.1318967111

Wynn ML, Rupp P, Trainor PA, et al. (2013) Follow-the-leader cell migration requires biased cell-cell contact and local microenvironmental signals. Phys Biol 10(3):35003–35016, DOI 10.1088/1478-3975/10/3/035003

Yamada KM, Sixt M (2019) Mechanisms of 3D cell migration. Nat Rev Mol Cell Biol 20(12):738–752, DOI 10.1038/s41580-019-0172-9

Yurchenco P, Birk D, Mechan RP (1994) Extracellular Matrix Assembly. Academic Press, San Diego

Zaman MH, Kamm RD, Matsudaira P, et al. (2005) Computational model for cell migration in three-dimensional matrices. Biophys J 89(2):1389–1397, DOI 10.1529/biophysj.105.060723