



DOCTOR OF SCIENCE (DSC)

Stochastic multiscale models of cell behaviour

Gavagnin, Enrico

Award date:
2020

Awarding institution:
University of Bath

[Link to publication](#)

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Stochastic multiscale models of cell behaviour

submitted by

Enrico Gavagnin

for the degree of *Doctor of Philosophy*

of the

University of Bath

Department of Mathematical Sciences

April 2020

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with the author and copyright of any previously published materials included may rest with third parties. A copy of this thesis has been supplied on condition that anyone who consults it understands that they must not copy it or use material from it except as licensed, permitted by law or with the consent of the author or other copyright owners, as applicable.

Declaration of any previous Submission of the Work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

.....

Enrico Gavagnin

Declaration of Authorship

I am the author of this thesis, and the work described therein was carried out by myself personally.

.....

Enrico Gavagnin

Acknowledgements

Like all important achievements, this PhD would not have been possible without the support of many people acting behind the scenes. The following lines are an attempt of giving them the credit and the gratitude that they deserve.

First, and foremost, I need to thank my two supervisors, Dr Kit Yates and Professor Tim Rogers. What they taught me goes well beyond the duty of a doctoral supervisor: from the value of letting your passion becoming your job, to the importance of preventing your passion from overcoming your life. Everything I have learnt from them has shaped an important part of who I am now. Thank you Kit and Tim, working with you has been as exciting as jumping down from a cliff into freezing water... and a true pleasure more than anything else.

A special acknowledgement goes to Professor Matthew Simpson for being a great host during my visit at the QUT, but also for all the inspiring conversations in the heat of Brisbane's summer - always in front of a flat white. I will treasure all his advices and those memories.

On the personal level, there are several people who I want to thank. Jonathan, as one of the very first people I met, who taught me so much about Britishness, biology and fresh pasta. If I have always felt so welcomed in this wet country, despite any cultural (or political..) barrier, is largely his merit. Claudio and Serena for being my alter-ego and my second-mum, respectively. Federico and his imperturbable politeness. Alice, for being the only one who truly understood my venetian origins. The whole Yates Group, Cameron and Jennifer, for supporting each others despite all possible difficulties we encountered. Yvonne, the Indian restaurant music and the Eastern Eye. The LittleItaly (aka office 2.15): Paolo, Umberto, Giacomo and the two Marchi, you all have been so kind with me during my last grumpy months. Thanks to Andrenino and all his cakes, Mirmmiu, Adwaye, Robbie, the duck fat and the long list of people who lived in Third Avenue at some point. The Scott and North road for not being so steep, at least not as much as Bathwick. Andreas and Laure, with whom I shared the adventure and the adversities of feeling European in Australia. All the *very* unconventional people with whom I have played at the open-mic of the Bell: Jim Blackman, Lawrie, Rutti, Luke DeSciscio and many others - you have been my weekly reminder that it is not all just a matter of science. Elisa, for leaving me probably the cheapest flat in Bath. I am very grateful to Alessandro and Serena, for all the time and dinners spent together and for always trying all my baking experiments without ever complaining. Oh and, obviously, to B.

A fundamental thanks goes to my family, my parents Rita and Gianni and my sister Elena. They are my unshakable roots and my saving sailors whenever I need them. The love and trust that bind us together define my values and my identity which I reflect in everything I do.

At last, it is imperative for me to thank my girlfriend Sara who has become a central part of my life in these years. You have been an endless source of love, motivation and strength for me, without which I wouldn't have been able to do anything of all this. All we have created together in the last two years is priceless for me and it represents the achievement of which I am most proud.

Abstract

Cells are the fundamental units of life and, as such, the most simple entities for which we can meaningfully talk about behaviour. All living creatures, including humans, can be considered the result of the behaviour of a group of cells and their interactions. Some cellular behaviours play more central roles than others. For example, cell migration and cell division constitute the main driving forces of many homeostatic and pathological processes in human body. Elucidating the role of a specific cell behaviour on the emergence of a biological phenomenon is a fundamental step for many medical interventions. Despite the great experimental advances of the past century, we are still lacking a full understanding of how complex biological processes, such as the development of a multicellular organism, arise from the interactions of significantly simpler entities - cells. In this context, multiscale mathematical modelling represents a powerful assistive tool which, by employing a combination of discrete and continuum approaches, provides a theoretical framework to bridge between the microscopic and macroscopic dynamics of the cell population.

The goal of this thesis is to elucidate some complex dynamics driven by cell migration and proliferation. Throughout the work we will make systematic use of multiscale modelling. We will introduce a series of stochastic and deterministic models describing cell migration and proliferation with increasing levels of realism. By exploiting the mutual benefits offered by these different paradigms, we will develop an equivalence framework which allows the connection of microscopic cellular properties with the dynamics of the total population. In particular, the first part of this study will focus on the interaction between the phenomena of directional persistence with excluding properties and on the spatial correlation that emerges from their interplay. In the second part, we will explore the role of stochastic cell proliferation in the context of an invading cellular front and its importance when interpreting experimental observations.

Contents

1	Introduction	1
1.1	Summary of original papers	2
1.2	A note on the thesis format	3
2	Stochastic and deterministic modelling of cell migration	4
2.1	Outline of the chapter	4
2.2	Conclusions	61
3	Modelling persistence of motion in a crowded environment: The diffusive limit of excluding velocity-jump processes	62
3.1	Outline of the article	62
3.2	Conclusions	82
4	Pair correlation functions for identifying spatial correlation in discrete domains	84
4.1	Outline of the article	84
4.2	Conclusions	99
5	The invasion speed of cell migration models with realistic cell cycle time distributions	116
5.1	Outline of the article	116
5.2	Conclusions	127
6	Synchronised oscillations in growing cell populations are explained by demographic noise	129
6.1	Outline of the article	129
6.6	Conclusions	149
7	Final Conclusions and Outlook	150
7.1	Future research	151

Chapter 1

Introduction

In many fields of biology, complex behaviours arise at the macro-scale as a result of many simple interactions at the micro-scale. Examples of these phenomena are ubiquitous in biological science, ranging from cell biology [Erban and Othmer, 2004, Baker et al., 2010, Yates et al., 2012] to animal behaviour [Couzin et al., 2002, Couzin, 2009, Yates et al., 2009, Berdahl et al., 2018]. The macroscopic level represents the level emergent phenomena manifest and, normally, the main object of interest is either to predict, or control, the occurrence of the event (e.g. tumour invasion [Sherratt and Chaplain, 2001], embryogenesis [Simpson et al., 2007, Mort et al., 2016], pest infestation [Yates et al., 2009]). The microscopic level describes the fundamental principles of the phenomenon and it often represents the level at which interventions can be made.

Among all the examples of complex behaviours observed in nature, those which take place at the cell level are of particular interest to me and have been the main object of the research during my PhD. There are several reasons why studying cell behaviour is important. The most evident are related to medical applications [McLean et al., 2005, Beltman et al., 2009]. Indeed a wide variety of human processes can be studied as collective cellular phenomena (morphogenesis [Keller, 2005, Mort et al., 2016], wound healing [Deng et al., 2006, Maini et al., 2004], tumour growth [Hanahan and Weinberg, 2000, Haass et al., 2014]).

However, there is another, often overlooked, motivation for studying these mechanisms: their intrinsic simplicity. Single cells, as the building blocks of life, are simpler than any other living organism (birds, fish, locusts, etc), but at the same time they represent the minimal living units for which we can meaningfully talk about behaviour. Therefore, multi-scale cellular behaviour represents one of the real frontiers between physics and biology, where active matter becomes a living system. As such, it is a field in which we can ask and answer fundamental questions about the basic principles underlying complex behaviours.

A crucial step in the understanding of complex cellular behaviour consists of linking the microscopic level of the individual cell with the macroscopic level of the whole population. This potentially allows us to explain the emergence of global events in terms of individual cell properties and, hence, to point out the main cellular mechanisms driving a particular biological phenomenon. In this context, mathematical modelling plays a central role for hypothesis formulation and data analysis. In particular, employing a multiscale modelling approach, which combines two, or more, levels of representation has become more and more popular in the past few decades [Othmer et al., 1988, Keller and Segel, 1971, Bodnar and Velazquez, 2005, Simpson et al., 2009].

The aim of my research is to improve the understanding of multi-scale phenomena

in cell populations with particular focus on the process of cell migration. During my studies I concentrated on two aspects of cell behaviour which together represent the main driving forces of cell invasion: cell motility and cell proliferation. Specifically, in the first part of my PhD I studied the role of cell directional persistence and its interplay with volume exclusion properties. Motivated by my findings in this, I also invested some time investigating methods for quantifying spatial correlation in discrete domains. In the second part of my PhD I studied the impact of accounting for realistic distributions of cell proliferation time in models of cell migration and population growth.

My work has led to a review book chapter [Gavagnin and Yates, 2018b] and four original papers (three published [Gavagnin and Yates, 2018a, Gavagnin et al., 2018, 2019] and one in preparation). The review chapter is presented in Chapter 2 and provides a broad overview of the state of the art of cellular multiscale modelling with particular focus on cell invasion. A substantial part of the review is devoted to explaining the derivation of macroscopic, deterministic representations of agent-based stochastic models of cell movement. In the remainder, a series of cell behaviours and cell-cell interactions are considered and some relevant modelling techniques are presented. Chapter 2 does not include original results, but it is used in this thesis as general introduction to the field. In rest of the thesis I present the results of my original research following the “Alternative Format” or “Thesis by Publications” style. In other words, each of the remaining chapters (Chapters 3, 4, 5 and 6) comprises one of my published or in preparation manuscripts and is accompanied by a short introduction and conclusion which help to place the corresponding chapter into the context of the thesis. In Chapter 7, I conclude by discussing my results from a broader perspective and suggesting some future work.

1.1. Summary of original papers

Chapter 3 - Modelling persistence of motion in a crowded environment: The diffusive limit of excluding velocity-jump processes

In this chapter we study the interplay of volume exclusion - the ability of cells to sense and avoid neighbours - and directional persistence - the tendency of cell to maintain the same direction of motion for short time scales. We develop a stochastic agent-based model embedded on a lattice which describes cell movement from which we derive an advection-diffusion PDE representing the total cell density at the macroscopic level. The results highlight some crucial differences and similarities between the microscopic (individual-based) and the macroscopic (advection-diffusion) modelling approaches. In particular, the macroscopic model is capable of capturing non-trivial phenomena, such as the spontaneous aggregation of cells.

Chapter 4 - Pair correlation functions for identifying spatial correlation in discrete domains

In this chapter we present a work about identifying and quantifying spatial correlation in multi-agent systems. We develop a set of statistical functions, specifically designed for discrete domains (such as lattices and networks), which determine whether the positions of individuals are positively or negatively correlated. These functions can then be used to detect the level of cell aggregation in agent-based models, such as those presented in Chapter 3.

Chapter 5 - The invasion speed of cell migration models with realistic cell cycle time distributions

In this chapter we focus on the realistic incorporation of the cell cycle into models of cell invasion, and the impact that this has on the speed of an invading wavefront. By using a multi-stage representation of the cell cycle, we extend previous models of cell invasion in order to account for a more realistic representation of the cell cycle. We demonstrate two analytical results obtained by adopting a mean field approximation, that elucidate the connection between the cell-cycle time distribution and the speed of the invasion.

Chapter 6 - Synchronised oscillations in growing cell populations are explained by demographic noise

In this chapter we further explore the importance of adopting realistic models of the cell cycle and its implications to growing population dynamics. We design multi-stage models representing two subpopulations of melanoma cells (depending on the phase of their cell cycle) in a proliferation assay. By comparing the results with the experimental data, the model explains the emergence of noise-induced oscillations on the relative proportion of the subpopulations which agrees with the experimental observations.

1.2. A note on the thesis format

Chapter 2 contains a review chapter published in the Elsevier Handbook of Statistics: Integrated Population in Biology. Chapters 3 and 4 contain two papers published in Physical Review E. Chapter 5 contains a paper published in the Journal of Theoretical Biology. Chapter 6 contains a preprint to be submitted for publication to Nature Communications. In all the published and submitted articles reported I am a lead author.

All the published material in this thesis (Chapter 2, 3, 4 and 5) is presented in the original format of the publication. The submitted manuscript (Chapter 6) is formatted to match the style of the thesis. The reader should consider the number in the footer as global page-number throughout the thesis.

Chapter 2

Stochastic and deterministic modelling of cell migration

The majority of this chapter comprises a review which I wrote for the Elsevier Handbook of Statistics: Integrated Population in Biology. The aim of this chapter is to give a general introduction to the field of multiscale cell modelling with particular emphasis on the application of cell invasion. In the review, we outline a series of stochastic and deterministic models which incorporate a broad range of cell behaviours and cell-cell interactions. However, the subject of cell motility plays a central role in the discussion and it is used as emblematic example to illustrate the main techniques for deriving deterministic representations from stochastic microscopic models.

In the context of the thesis, the role of this chapter is to provide the relevant background for the further papers. In particular, all the derivations in Section 2.1.1 of this chapter are relevant for Chapters 3 and 5. Section 3.1 contains an introduction to multi-stage models of the cell-cycle which is the main object of Chapters 5 and 6. Section 3.4 of this chapter can be considered to be an extended introduction to and a summary of Chapter 3. The remaining sections cover topics which have not been the direct object of my research and are, therefore, not necessary for the understanding of the following chapters. However, since they are closely related to the objects of the following papers, their role is to help to place the thesis in the wider context of the background literature.

2.1. Outline of the chapter

In Section 1, we give a general overview of the process of cell invasion, we highlight the need for mathematical models and we underline the strengths and weaknesses of typical modelling approaches. Section 2 is entirely dedicated to the problem of modelling cell movement. A series of deterministic diffusive models are derived from stochastic agent-based models based on random walks. The derivation is carried out in detail for one-dimensional discrete and continuum random walks, with and without volume excluding properties (Section 2.1). Extensions of these basic models to higher orders of approximation and higher spatial dimensions are discussed in Section 2.2 and 2.3, respectively. In Section 3 other model extensions are considered. Particular attention is given models of cell proliferation (Section 3.1), growing domains (Section 3.3) and directional persistence (Section 3.4). Some direct and indirect forms of cell-cell interactions are considered in Section 3.2. Finally, Section 4 contains a short concluding discussion.

Appendix B: Statement of Authorship

This declaration concerns the article entitled:									
Stochastic and Deterministic Modeling of Cell Migration									
Publication status (tick one)									
draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input checked="" type="checkbox"/>
Publication details (reference)	Book chapter: Integrated Population Biology and Modeling, Vol. 39. Ch. 2 Authors: Enrico Gavagnin and Christian A. Yates								
Candidate's contribution to the paper (detailed, and also given as a percentage).	All the simulations and the numerical computations have been performed by the author of the thesis (100%). All authors contributed equally to the presentation of the content (50%).								
Statement from Candidate	This paper reports on a review chapter which I wrote during the period of my Higher Degree by Research candidature.								
Signed							Date	27.4.2020	

Chapter 2

Stochastic and Deterministic Modeling of Cell Migration

Enrico Gavagnin¹ and Christian A. Yates

Department of Mathematical Sciences, University of Bath, Bath, United Kingdom

¹*Corresponding author: e-mail: e.gavagnin@bath.ac.uk*

ABSTRACT

Mathematical models are vital interpretive and predictive tools used to assist in the understanding of cell migration. There are typically two approaches to modeling cell migration: either microscale, discrete or macroscale, continuum. The discrete approach, using agent-based models (ABMs), is typically stochastic and accounts for properties at the cell-scale. Conversely, the continuum approach, in which cell density is often modeled as a system of deterministic partial differential equations (PDEs), provides a global description of the migration at the population level. Deterministic models have the advantage that they are generally more amenable to mathematical analysis. They can lead to significant insights for situations in which the system comprises a large number of cells, at which point simulating a stochastic ABM becomes computationally expensive. However, finding an appropriate continuum model to describe the collective behavior of a system of individual cells can be a difficult task. Deterministic models are often specified on a phenomenological basis, which reduces their predictive power. Stochastic ABMs have advantages over their deterministic continuum counterparts. In particular, ABMs can represent individual-level behaviors (such as cell proliferation and cell–cell interaction) appropriately and are amenable to direct parameterization using experimental data. It is essential, therefore, to establish direct connections between stochastic microscale behaviors and deterministic macroscale dynamics.

In this chapter we describe how, in some situations, these two distinct modeling approaches can be unified into a discrete-continuum equivalence framework. We carry out detailed examinations of a range of fundamental models of cell movement in one dimension. We then extend the discussion to more general models, which focus on incorporating other important factors that affect the migration of cells including cell proliferation and cell–cell interactions. We provide an overview of some of the more recent advances in this field and we point out some of the relevant questions that remain unanswered.

Keywords: Cell migration, Discrete-continuum equivalence, Agent-based models, Partial differential equation, Collective behavior

1 INTRODUCTION

The process of cell migration plays an essential role in several developmental and pathological mechanisms. For example, cell migration orchestrates morphogenesis throughout the development of the embryo (Gilbert, 2003; Keller, 2005) and plays a crucial role in wound-healing (Deng et al., 2006; Maini et al., 2004) and immune responses (Madri and Graesser, 2000). Migration also contributes to many pathological processes, including vascular disease (Raines, 2000) and cancer (Hanahan and Weinberg, 2000).

Many of these phenomena are highly complex and involve the collective behavior of a set of interacting cells. At the individual-cell-level, a variety of mechanisms can be involved, including cell–cell adhesion (Niessen, 2007; Trepap et al., 2009), attraction (Yamanaka and Kondo, 2014), and repulsion (Carmona-Fontaine et al., 2008). In other cases, cells can respond to a chemical gradient which regulates and guides their motility (*chemotaxis*) (Keynes and Cook, 1992; Ward et al., 2003). In addition, when the migration occurs over a sufficiently long time, cell proliferation and death can also play important roles in the process (Mort et al., 2016). An understanding of the impact that these (and other) individual-cell mechanisms have on global collective migration is important, since it can illuminate the origin of major diseases and suggest effective therapeutic approaches.

Over the past few decades, great progress has been made in understanding many aspects of cell migration (Friedl and Gilmour, 2009; Ridley et al., 2003). However, we still lack a proper understanding of many of its underlying mechanisms. In particular, there are aspects which remain impenetrable to experimental biologists (Staton et al., 2004). One of the major issues when studying cell migration is obtaining and interpreting experimental data. For example, a comprehensive controlled experiment (in vivo) or a realistic (in vitro) set up can be difficult or impossible to obtain. In other cases, it may be extremely complicated to investigate the microscopic origin of a macroscopic phenomenon, given the number of different actions that a single cell can perform (Dworkin and Kaiser, 1985; Tambe et al., 2011; Trepap et al., 2009; Westermann et al., 2003). In this context, mathematical modeling has become a necessary interpretive and predictive tool to assist in the understanding of such complex phenomena (Maini et al., 2004; Noble, 2002; Simpson et al., 2006).

Mathematical models have been employed as tools for validation of experimentally generated hypotheses. Moreover, due to the advances in computation of the last few decades, mathematical models are becoming more detailed and accurately parametrized, which makes them capable of generating experimentally testable hypotheses and exploring new questions that are still not approachable from an experimental prospective (Tomlin and Axelrod, 2007). For example, the use of stochastic mathematical models facilitates the investigation of biological systems in which randomness plays an important role and for which viewing only a single instance of the evolution

of a system can be inconclusive (Lee and Wolgemuth, 2011). In a mathematical modeling framework, the system can be repeatedly simulated using the same deterministic or randomized initial conditions. Equally, repeats with slightly altered initial conditions or parameters are useful for quantifying the sensitivity of systems in which a small fluctuation in the initial state can have a significant effect on the outcome of the experiment, or in understanding the sensitivity of the system to certain parameter choices, respectively. In addition, the outcome of a mathematical simulation can be examined easily since all the variables of the system are explicitly accessible (Flaherty et al., 2007).

There are typically two approaches to modeling cell migration, depending on the scale of interest. At the level of an individual cell, stochastic, discrete agent-based models (ABMs) are popular. Each cell is modeled as a single individual (agent) with its own rules of movement, proliferation, and death. Alternatively, to model collective cell migration at the population level, a deterministic, continuum partial differential equation (PDE) for the population density is normally used¹. These two modeling paradigms have complementary advantages and disadvantages which we summarize in Table 1.

TABLE 1 Summary of the Advantages and Disadvantages of Adopting a Discrete/Stochastic Approach, Such as an ABM, as Opposed to a Continuum/Deterministic Approach, Typically Represented by a PDE

	Advantages	Disadvantages
Discrete/ stochastic	Detailed structure and explicit implementation	High computational cost
	Direct connection to experimental data	Multiple simulations required
	Incorporation of randomness	Relatively inaccessible to mathematical analysis
Continuum/ deterministic	Fast to simulate	Lack of fine detailed structure
	Amenable to mathematical analysis	Difficult to link experimental data
	Suitable for systems of large numbers of cells	Ignore the effects of randomness

1. Although deterministic ABMs (Kurhekar and Deshpande, 2015) and stochastic continuum models (Dickinson and Tranquillo, 1993; Schienbein and Gruler, 1993) have also been considered, the most common pairing in the literature involves stochastic ABMs and population-based deterministic models. Therefore, these are the two modeling subtypes that we will consider in this review.

Generally speaking, ABMs are attractive because their macroscale behavior is completely self-induced, rather than being superimposed on a phenomenological basis (Ben-Jacob et al., 2000; Shapiro, 1988). This is particularly interesting when a complex structure emerges at the population-level (Othmer and Stevens, 1997; Thompson et al., 2011, 2012). In fact, although the agents represent the driving force of such structure, typically they are unaware, as individuals, of the macroscopic configuration of the system, since their behavior is dictated only by their local environment.

Another advantage of ABMs is that their formulation is generally more intuitive than continuum models. Each cell's actions can be easily incorporated in the model by implementing appropriate rules for the corresponding agent, which mimic specific known biological behaviors. The parameters regulating each of these rules can then be inferred directly from real observations which makes ABMs more easily relatable to experimental data.

Mort et al. (2016), for example, used a simple ABM to study the migration of mouse melanoblasts, the embryonic precursors of melanocytes, responsible for pigmentation (see Fig. 1). The authors studied the effects of a family of mutations in the gene for the receptor tyrosine kinase, Kit. These mutations affect the success of the colonization of the growing epidermis by melanoblasts, leading to unpigmented regions of hair and skin. Mort et al. (2016) employed an ABM to study the interplay between movement and proliferation of melanoblasts. Further they carried out a statistical analysis on single-cell

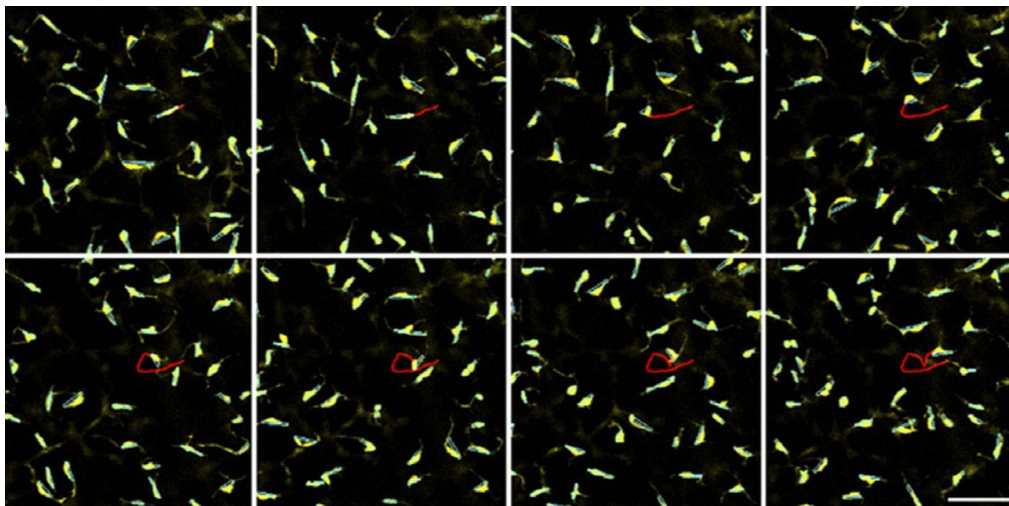


FIG. 1 A time-lapse sequence of melanoblasts migrating in ex vivo culture of E14.5 (14.5 days postfertilization) mouse skin. The Feret's diameter of each cell body is indicated in cyan. The path of a single migrating cell is indicated in red. Melanoblasts migrate along their Feret's diameter—the longest distance between any two points along a given cell boundary. *Reproduced from Mort, R.L., Ross, R.J.H., Hainey, K.J., Harrison, O.J., Keighren, M.A., Landini, G., Baker, R.E., Painter, K.J., Jackson, I.J., Yates, C.A., 2016. Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model. Nat. Commun. 7, 10288 with the permission of Nature Communications.*

trajectories in order to parametrize the model against experimental data. By simulating their ABM and comparing the results with experimental data, [Mort et al. \(2016\)](#) were able to show that belly-spot formation in Kit mutants is likely to be induced by a reduction in proliferation rate, rather than motility, a result which was contrary to the received wisdom in the experimental literature.

Deterministic models also have their advantages. For example, when the population size becomes large, a PDE model is generally desirable, since there exist a wide range of well-developed numerical tools for their rapid solution. This is in direct contrast to agent-based models which are typically coded up ad hoc and require multiple computationally intensive repeats in order to gather reliable ensemble statistics. Moreover, continuum models have the advantage that they are generally more amenable to mathematical analysis than ABMs, and this analysis can often lead to significant and general insights. For example, PDEs can be used to carry out stability analysis to determine the conditions which lead to pattern formation ([Anguige and Schmeiser, 2009](#)). In other scenarios, one can use traveling wave analysis to obtain expressions for the speed of the cell invasion in terms of the model parameters ([Murray, 2007](#)).

Depending on the biological questions of relevance and available experimental data, either deterministic-continuum or stochastic-discrete models (or some combination of both) may be appropriate. Ideally, the individual-level representation can be used to parametrize the model, and the continuum level-description should link the population-level results back to the parameters of interest. The problem of connecting the parameters of the individual-level model to those of a representative population-level description represents, therefore, a crucial step of the multiscale modeling process.

In this chapter, we provide an overview of a range of techniques which can be used to connect ABMs of cell migration to macroscopic PDEs for average cell density. We review a series of stochastic and deterministic models which are capable of reproducing some of the key features of the behavior of cells and describe how these two modeling regimes can be linked to form a multi-scale mathematical framework.

In [Section 2.1](#) we carry out a detailed derivation of deterministic models from ABMs which focus on cell movement. We present derivations in two different scenarios, depending on whether the spatial domain in which the ABM is defined is partitioned into a finite lattice (*the on-lattice case*) or not (*the off-lattice case*). In each of these situations, we first study the case in which cells move independently of others (*noninteracting cells*) and then we introduce the ability of cells to sense the occupancy of neighboring regions of space and to avoid overlapping with other cells (*interacting cells*). All the derivations in [Section 2.1](#) are carried out in one dimension and assuming a *mean-field* moment closure approximation. We discuss generalizations to higher dimensions and other closure approximations in [Sections 2.2](#) and [2.3](#), respectively.

The remaining part of this chapter is devoted to reviewing a series of biologically relevant features that can be incorporated in the models of cell migration. In each case we highlight the implications of introducing a given behavior both at the individual- and population-level. More precisely, in [Section 3.1](#) we present models which incorporate the ability of cells to reproduce by division. In [Section 3.2.1](#) we present models in which cells can interact indirectly through an external signal, e.g., *chemotaxis* or slime following. Some examples of direct forms of cell–cell interactions, such as adhesion–repulsion, pushing, and pulling, are discussed in [Section 3.2](#). In [Section 3.3](#) we review a series of recent papers which model cells migrating on a growing domain. Finally, we discuss how to derive a macroscopic limit for ABMs which are not based on simple random walks and which are capable of representing persistence of motion, in [Section 3.4](#). We conclude this chapter with a short discussion and final considerations in [Section 4](#).

2 CELL MOTILITY

In this section we present a detailed description of the basic models of cell motility, which represent the fundamental basis for the majority of the spatially extended representations of the remaining part of this chapter. We use the case of these models to illustrate standard approaches to deriving diffusive, deterministic, continuum representation from ABMs, based on occupancy master equations. We carry out the explicit derivations for one-dimensional versions of the models in [Section 2.1](#), and we discuss the generalization to higher dimensions in [Section 2.2](#). We consider two distinct cases, depending on whether the motility mechanism of the ABM is implemented on- or off-lattice. In each case, we focus on two variants of the model, both with and without crowding effects. We explain the standard techniques for deriving the corresponding deterministic descriptions at the population-level in each case. Note that, throughout this chapter, we adapt the notation of models taken from the literature for consistency.

2.1 Connecting Stochastic and Deterministic Models of Cell Movement

2.1.1 On-Lattice Models

Broadly speaking, on-lattice ABMs can be classified as cellular automata in which a set of agents occupy some or all sites of a lattice. In general, these agents have a number of state variables associated with them and a set of rules prescribing the evolution of their state and position. There exists a great variety of forms of ABMs. The appropriateness of each representation depends on the phenomenon that is being modeled. For example, cellular Potts models have been used by [Turner and Sherratt \(2002\)](#) and [Turner et al. \(2004\)](#) in the context of cancer modeling and by [Graner and Glazier \(1992\)](#) in the context of cell sorting via differential adhesion. [Othmer and Stevens \(1997\)](#)

modeled bacterial aggregation by using a position-jump process, and [Painter et al. \(1999\)](#) adopted a similar approach to study the interplay of chemotaxis and volume exclusion. One of the advantages of using these on-lattice models is that their formulation and analysis tend to be straightforward in comparison to their off-lattice counterparts. Moreover, the presence of the grid considerably reduces the computational cost of simulations when a large number of agents is involved. However, since in the majority of scenarios the assumption that cells move on a discrete grid is not appropriate, the implementation of cell behaviors is usually phenomenological.

Consider a one-dimensional domain, $[0, L]$, with periodic boundary conditions. An ABM on the domain $[0, L]$ comprises a set of $\mathcal{N} \in \mathbb{N}$ agents positioned in $[0, L]$ and a range of stochastic rules which govern their evolution in time. The ABMs that we consider throughout this section are all based on continuous-time *position-jump processes* ([Othmer and Stevens, 1997](#); [Othmer et al., 1988](#)). This means that the position of the agents undergoes series of sequential Markovian jumps, i.e., at any given time, the evolution of the position of each agent depends only on its current position. An alternative approach is to use *velocity-jump processes*, in which the Markov property applies to the velocity of the agents, instead of their position. We discuss this approach in [Section 3.4](#).

Discrete-time ABMs are popular in the literature ([Simpson et al., 2007, 2009](#); [Treloar et al., 2011, 2013](#)), although, for the examples presented here, we treat time as a continuous variable² ([Othmer and Stevens, 1997](#); [Othmer et al., 1988](#)). As time evolves, agents can attempt movement events which occur as a Poisson process with rate α . In other words, each agent attempts to move after an independent exponentially distributed waiting time with parameter α . When such attempts take place, we say that the agent has been selected to attempt a movement.

Consider the on-lattice scenario, in which the domain is partitioned in k intervals, each of length $\Delta = L/k$, whose centers are denoted by x_1, \dots, x_k , respectively (see [Fig. 2](#) for a schematic illustration). We denote the position of the n -th agent at time t as $c_n(t)$, hence $c_n(t) \in \{x_1, \dots, x_k\}$ for every $t \in \mathbb{R}^+$ and $n = 1, \dots, \mathcal{N}$.

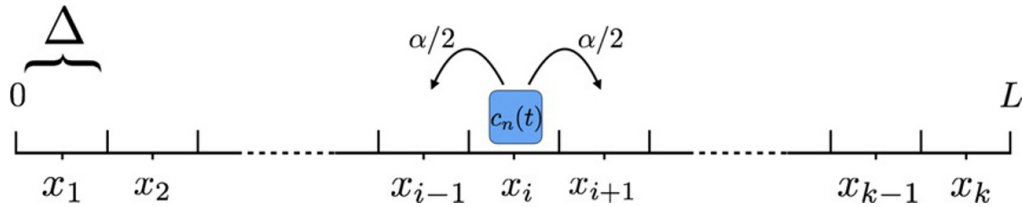


FIG. 2 Schematic of the on-lattice ABM in one dimension. The domain is partitioned in k intervals of length Δ . A single agent, n , is represented as a *blue square* and it occupies interval i , i.e., $c_n(t) = x_i$. The agent moves, with rate α , to one of the two adjacent sites.

2. In general, when the time discretization step is sufficiently small, the behavior of the discrete-time ABMs is similar to its continuous-time counterpart.

2.1.1.1 Noninteracting Cells Undergoing Unbiased Movement

In this section, we consider the basic case of noninteracting agents. In particular, the movement rates of each agent are independent of the other agents' positions and unbiased. If at time t an agent, n , is selected to move, it moves with equal probability, $1/2$, to either one of its adjacent sites, $c_n(t) \pm \Delta$. Equivalently, we can say that each agent moves to the right and to the left with rates $\mathcal{T}^\pm = \alpha/2$, respectively. Notice that multiple agents can occupy the same lattice site. We aim to obtain a deterministic representation of the mean evolution of the agent density at a given time t . Therefore, we assume the model is simulated until time t for a large number, M , of independent realizations, with each realization identically prepared. In Fig. 3A we consider 30 one-dimensional lattices with $L = 100$ and $\Delta = 1$. In each one-dimensional lattice we populated the 20 central sites, $i = 41, \dots, 60$, at random with 18 agents on average (with multiple agents per site possible). In Fig. 3C, E, and G we show three snapshots of these 30 identically prepared simulations of the ABM, as time evolves. We denote by $C^{(m)}(x_i, t)$ the number of agents which lie in the interval i at time t of the m -th repeat simulation, with $m = 1, \dots, M$. Namely

$$C^{(m)}(x_i, t) = |\{n | c_n(t) = x_i\}|. \quad (1)$$

If $C(x_i, t) = 0$, we say that the site i is *empty* and if $C(x_i, t) > 0$ we say it is *occupied*. We define the mean occupancy of site i at time t , averaged over the number of realizations, as

$$\bar{C}(x_i, t) = \frac{1}{M} \sum_{m=1}^M C^{(m)}(x_i, t). \quad (2)$$

By considering all the possible agents movements we can write down a conservation law for the average occupancy at time $t + \delta t$, where δt is a sufficiently small that the probability that two or more movements take place in the interval $[t, t + \delta t)$ is $o(\delta t)$. This reads

$$\bar{C}(x_i, t + \delta t) = \bar{C}(x_i, t) - \overbrace{\alpha \delta t \bar{C}(x_i, t)}^{\text{moving out of site } i} + \underbrace{\frac{\alpha \delta t}{2} [\bar{C}(x_{i-1}, t) + \bar{C}(x_{i+1}, t)]}_{\text{moving into site } i} + \mathcal{O}(\delta t^2). \quad (3)$$

The right-hand side of Eq. (3) comprises three parts: the first term, which accounts for the occupancy of site i at time t , and two terms which determine the expected loss of average occupancy due to agents moving out of site i and the gain due to agents moving into site i , respectively.

If we rearrange Eq. (3), divide through by δt and take the limit as $\delta t \rightarrow 0$, we obtain the Kolmogorov equation (or continuous-time master equation) of the process which reads

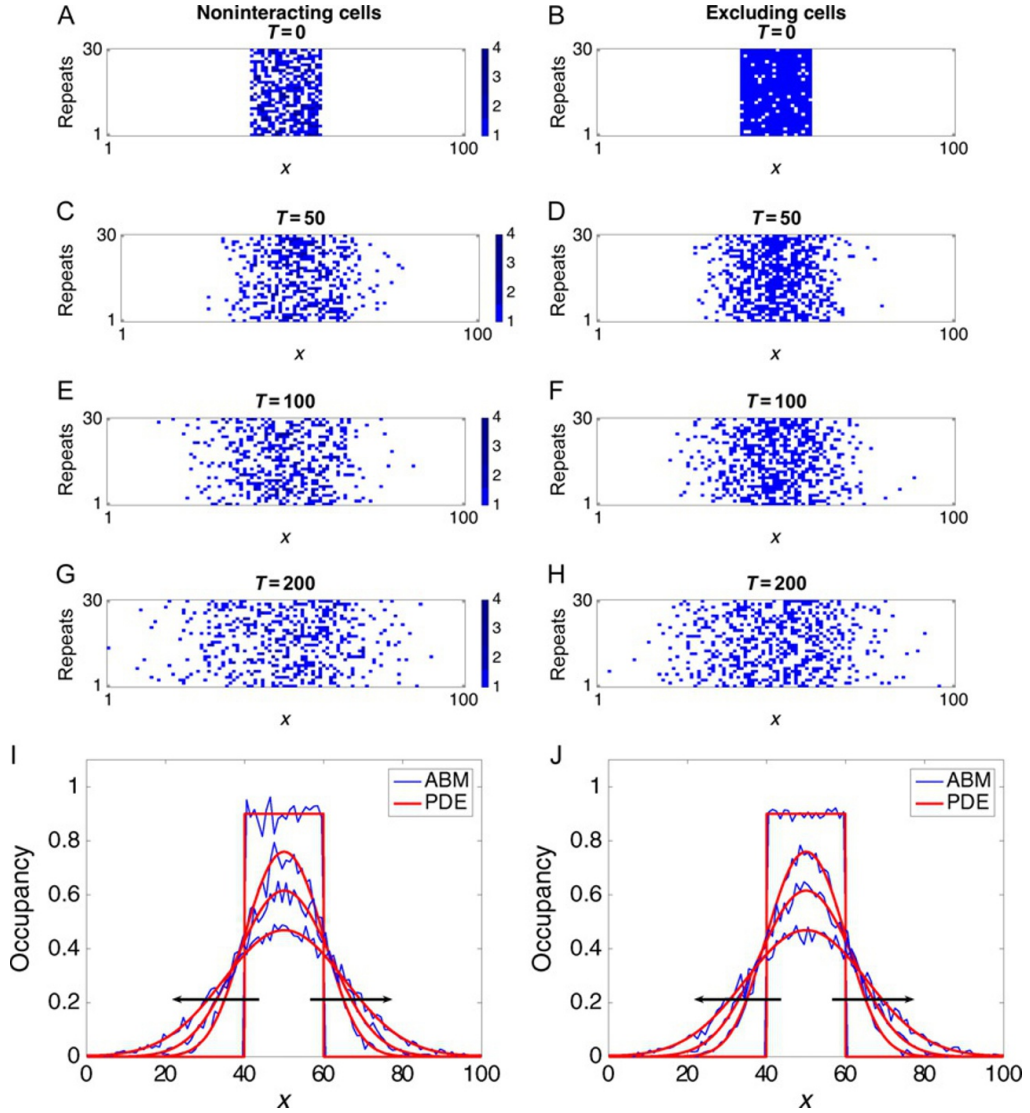


FIG. 3 Comparison between stochastic and deterministic models of cell motility in on a one-dimensional lattice. Panels (A)–(H) show four snapshots of 30 simulations of the ABM with non-interacting agents as described in [Section 2.1](#) (panels (A), (C), (E), and (G)) and with excluding agents (panels (B), (D), (F), and (H)). The motility rate is $\alpha = 1$ and the simulations are shown at times $T = 0, 50, 100$, and 200 . In each panel 30 independent repeats of the simulations are displayed on top each other. Empty sites are represented in *white* and occupied sites are represented in *blue*. For the noninteracting model, the occupied sites are *colored* with a graded intensity of *blue* corresponding to the number of agents which occupy the site, as indicated in the *color bar*. In panels (I) and (J) we show a comparison between the occupancy of the ABM (*blue line*), averaged over 500 repeats, and the numerical solution of corresponding deterministic PDE (*red line*). Panel (I) is for the noninteracting model and panel (J) is for the model with volume exclusion. In both cases, the profiles are displayed at the times $T = 0, 50, 100, 200$, with the direction of the *black arrows* indicating increasing time.

$$\frac{\partial \bar{C}(x_i, t)}{\partial t} = \frac{\alpha}{2} [\bar{C}(x_{i-1}, t) - 2\bar{C}(x_i, t) + \bar{C}(x_{i+1}, t)]. \quad (4)$$

Now we Taylor expand the terms $\bar{C}(x_{i\pm 1}, t)$ about the point x_i to the second order, i.e., we use the following approximations:

$$\bar{C}(x_{i+1}, t) = \bar{C}(x_i, t) + \Delta \frac{\partial \bar{C}(x_i, t)}{\partial x} + \frac{\Delta^2}{2} \frac{\partial^2 \bar{C}}{\partial x^2} + \mathcal{O}(\Delta^2), \quad (5a)$$

$$\bar{C}(x_{i-1}, t) = \bar{C}(x_i, t) - \Delta \frac{\partial \bar{C}(x_i, t)}{\partial x} + \frac{\Delta^2}{2} \frac{\partial^2 \bar{C}}{\partial x^2} + \mathcal{O}(\Delta^2). \quad (5b)$$

By substituting Eq. (5) into Eq. (4) and taking the limit as $\Delta \rightarrow 0$ while holding $\alpha\Delta^2$ constant, we recover the diffusion equation for the continuous approximation of the average occupancy function, \bar{C} :

$$\frac{\partial \bar{C}}{\partial t} = D \frac{\partial^2 \bar{C}}{\partial x^2}, \quad (6)$$

where D is the *diffusion coefficient* defined as

$$D = \lim_{\Delta \rightarrow 0} \frac{\alpha\Delta^2}{2}. \quad (7)$$

We refer to Fig. 3I for a comparison between the average occupancy of the ABM and the diffusion equation (6). The results confirm the good agreement between the two models.

The derivation of Eq. (6) from a simple random walk, as outlined above, is a well-known result (Codling et al., 2008; Deutsch and Dormann, 2007; Murray, 2007). Alternatively, we could have carried out the derivation for the probability density function, $P(x_i, t)$, of finding a single agent at position x_i at time t . This would result in a macroscopic PDE of the same form as Eq. (6). The equation for P can be obtained by dividing both sides of Eq. (6) by the total number of agents, \mathcal{N} .

The diffusion equation (6) is a classical PDE, also known as Fick's diffusion equation or the heat equation depending on the application. Extensive discussions on this type of equation can be found in Crank (1979) or Welty et al. (2009). The existence of an analytic solution of Eq. (6) depends on the imposed initial and boundary conditions. For example, if we assume an infinite domain and that all agents are initialized in the same position:

$$\bar{C}(x, 0) = \begin{cases} \mathcal{N} & x = x^*, \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

Eq. (6) admits the fundamental solution given by

$$\bar{C}(x, t) = \frac{\mathcal{N}}{\sqrt{4\pi Dt}} e^{-\frac{(x-x^*)^2}{4Dt}}. \quad (9)$$

In the basic model outlined above, agents can only move to their nearest-neighbor sites. A generalization of this model, in which agents have the ability to perform nonlocal jumps is studied by [Taylor et al. \(2015a\)](#). Specifically, in the model of [Taylor et al. \(2015a\)](#) agents are allowed to jump up to Q sites away from their current position. For $q = 1, \dots, Q$, each length- q jump occurs with rate $\mathcal{T}^{(\pm q)}$ in the right- and left-directions, respectively. When the motility is unbiased, i.e., $\mathcal{T}^{(q)} = \mathcal{T}^{(\pm q)}$ for every q , the authors show that the behavior of their agents in the continuum limit evolves according to the diffusion equation (6), but with diffusion coefficient given by

$$D^Q = \lim_{\Delta \rightarrow 0} \Delta^2 \sum_{i=1}^Q q^2 \mathcal{T}^{(-q)}. \quad (10)$$

This implies that, when the following condition is satisfied,

$$\alpha = \sum_{i=1}^Q q^2 \mathcal{T}^{(q)}, \quad (11)$$

the local ABM and the nonlocal ABM are described by the same macroscopic PDE. There are an infinite number of possible combinations of $\mathcal{T}^{(q)}$ s which satisfy condition (11). However, [Taylor et al. \(2015a\)](#) suggest that choosing

$$\mathcal{T}^{(q)} = \frac{\alpha}{q^2 Q}, \quad (12)$$

is particularly appropriate, since it preserves the well-known property of diffusive processes, that the mean-squared displacement scales linearly with time ([Codling et al., 2008](#); [Othmer et al., 1988](#)). A comparison of the simulations of the two ABMs confirms the accuracy with which the nonlocal models typically correspond to their local equivalent. The results also reveal a significant reduction in the average simulation time of the nonlocal ABM compared to the local ABM. This time-saving potential, given that condition (11) is satisfied, is highlighted by [Taylor et al. \(2015a\)](#) as an important application of the nonlocal models in order to reduce the computational cost of stochastic simulations.

However, it should be noted that, when dealing with steep gradients in agent numbers, the nonlocal model loses accuracy in comparison to the local model. This inaccuracy stems from the choice of the transitional rates (12) which match the terms of the local model only up to the second order of the expansion. In order to address this issue, the authors propose a spatially extended hybrid method in which regions containing steep gradients in agent numbers are dealt with using the local representation and regions in which gradients are more shallow using the nonlocal representation. This allows the acceleration of simulations afforded by the nonlocal representation, while maintaining the accuracy associated with the local method.

In the second part of the paper, [Taylor et al. \(2015a\)](#) derive a general class of boundary conditions for local and nonlocal ABMs, corresponding to classical boundary conditions in the deterministic, continuum, diffusive limit.

For a general nonlocal ABM, the authors study a class of first-order reactive boundary conditions known as Robin conditions which, for the left boundary, can be written as

$$D \frac{\partial C(0,t)}{\partial x} = C(0,t)B. \quad (13)$$

Here $B = 0$ corresponds to a purely reflective boundary and $B = \infty$ corresponds to a purely absorbing boundary. In order to obtain the corresponding stochastic boundary condition at the individual-level, [Taylor et al. \(2015a\)](#) allow agents which attempt length- q jumps that result in them hitting the boundary, to be absorbed (i.e., removed from the domain) with absorption probability $a_{q,Q}$. If no absorption occurs, the agents reach the boundary and are then reflected in the opposite direction for the remaining number of steps of the jump. By taking a diffusive limit from the corresponding occupancy master equation, the authors determine the expression of the absorption rates in terms of the parameter, B , of the Robin boundary condition. The absorption rates are given by (13):

$$a_{q,Q} = \Delta \frac{B}{D^Q} \left(1 + q^2 \sum_{i=q+1}^Q \frac{2}{i^2} \right) \text{ for } q = 1 \dots Q, \quad (14)$$

where Δ is the lattice step, Q is the maximum jump length, and D^Q is defined as in Eq. (10). In particular, when a local ABM is considered, the absorption rate, $a_{1,1}$, for agents at the site adjacent to the boundary is given by

$$a_{1,1} = \Delta \frac{B}{D}.$$

Finally, [Taylor et al. \(2015a\)](#) extend their study to the case in which Neumann boundary conditions are imposed at the population level. For the left boundary this condition is given as

$$\frac{\partial C(0,t)}{\partial x} = F, \quad (15)$$

where $F < 0$ represents the influx into the system at this boundary. The corresponding ABM implementation requires new agents to enter the domain. For a general nonlocal model with maximum jump length Q , these agents are positioned in the Q nearest sites to the boundary. By postulating a discrete master equation, which respects conservation of the total influx, F , and taking a diffusive limit as $\delta t, \Delta \rightarrow 0$, the authors derive an expression for the rate of introduction of agents into the k -th nearest site:

$$f_{k,Q} = -\frac{FD^Q}{Q\Delta} \left(\sum_{i=k}^Q \frac{2i-2k+1}{k^2} \right) \text{ for } k=1\dots Q, \quad (16)$$

which in the local case ($Q = 1$), becomes

$$f_{1,1} = -\frac{FD}{\Delta}.$$

2.1.1.2 Excluding Cells

We now incorporate volume exclusion in the model following the approach of [Simpson et al. \(2009\)](#). We modify the basic ABM of the previous section, by introducing a specific form of agent-agent interaction which prevents two or more agents from occupying the same lattice site. We initialize the domain by populating the central interval, with \mathcal{N} excluding agents, i.e., with the property that any given site can be occupied by at most one agent. In the example in [Fig. 3B](#) in each row we populated sites $i = 41, \dots, 60$ with 18 agents, on average, with a maximum of one agent per site. When an agent, n , is selected for a movement event, one of its two neighboring sites, $C_n(t) \pm \Delta$, is selected with equal probability, $1/2$. In order for the movement to be successful, however, the new selected site has to be empty. When the selected site is occupied, the movement is aborted and the selected agent remains in its position. This movement rule prevents agents from moving into occupied sites. Therefore, the excluding property of the initial condition is preserved as the time evolves. In particular, this implies that $C^{(m)}(x_i, t)$, defined in Eq. (1), takes only two values, 0 and 1 (see [Fig. 3D, F, and H](#) for snapshots of simulations of the ABM).

For a given a realization of the model, m , we make the usual moment closure approximation that the occupancies of different sites, i and j with $i \neq j$, are independent, i.e.:

$$\mathbb{P}[C^{(m)}(x_i, t) = 1, C^{(m)}(x_j, t) = 1] = \mathbb{P}[C^{(m)}(x_i, t) = 1] \mathbb{P}[C^{(m)}(x_j, t) = 1], \quad (17)$$

for every $t \in \mathbb{R}^+$. Eq. (17) is known as the *mean-field assumption*, and it provides a good approximation in many scenarios; however, it becomes invalid when spatial correlations play an important role in evolution, for example, when proliferation or attractive forces between agents are involved (see [Sections 3.1 and 3.2](#), respectively). We discuss the validity of this approximation and possible alternative approaches to the mean-field assumption in [Section 2.3](#).

By assuming the independence captured by Eq. (17), we can write down the transition rates from a given site x_i as

$$\begin{aligned} \mathcal{T}^+(x_i, t) &= \frac{\alpha \delta t}{2} \left(1 - C^{(m)}(x_{i+1}, t) \right), \\ \mathcal{T}^-(x_i, t) &= \frac{\alpha \delta t}{2} \left(1 - C^{(m)}(x_{i-1}, t) \right). \end{aligned}$$

Note that the terms $(1 - C^{(m)}(x_{i\pm 1}, t))$ keep into account the possibility of abortion of the movement due to volume exclusion. Hence the occupancy master equation at position x_i reads

$$C^{(m)}(x_i, t + \delta t) = C^{(m)}(x_i, t) - \overbrace{\frac{\alpha \delta t}{2} C^{(m)}(x_i, t) \left[(1 - C^{(m)}(x_{i-1}, t)) + (1 - C^{(m)}(x_{i+1}, t)) \right]}^{\text{moving out of } x_i} + \underbrace{\frac{\alpha \delta t}{2} (1 - C^{(m)}(x_i, t)) \left[C^{(m)}(x_{i-1}, t) + C^{(m)}(x_{i+1}, t) \right]}_{\text{moving into } x_i} + \mathcal{O}(\delta t^2).$$

If we divide by δt , rearrange and let $\delta t \rightarrow 0$, we obtain

$$\frac{\partial C^{(m)}(x_i, t)}{\partial t} = \frac{\alpha}{2} \left[C^{(m)}(x_{i-1}, t) - 2C^{(m)}(x_i, t) + C^{(m)}(x_{i+1}, t) \right], \quad (18)$$

for $m = 1, \dots, M$. By summing equations (18) over each repeat and dividing by the total number of realizations, M , we recover Eq. (4) for the average occupancy, $\bar{C}(x_i, t)$. Remarkably, we obtain exactly the same macroscopic representation as in the case of noninteracting agents, which is given by the canonical diffusion equation (6). In Fig. 3J we show a comparison between the average occupancy of the ABM with excluding agents and the corresponding diffusion equation at increasing times.

Notice that this inclusion of volume exclusion on a lattice does not affect the population-level dynamics. In other words, since the diffusion equation (6) arises as the limit equation of the standard random walks of noninteracting agents, the effect of the local interaction between the agent in the simple exclusion process disappears as we let $\Delta, \delta t \rightarrow 0$. Since agents are indistinguishable, the situation in which two agents occupy adjacent positions and block each other's movement is equivalent to the scenario in which the two neighboring agents swap their positions in the noninteracting random walk. If multispecies agents are considered, such equivalence no longer holds and the exclusion property leads to different continuous equations (Simpson et al., 2009).

Taylor et al. (2015b, 2016) study alternative approaches to implementing volume exclusion in compartment-based models at different spatial scales. They consider a *coarse-grained* representation of volume exclusion (as opposed to *fine-grained* representation described above, in which at most one agent can occupy a given site) in which S fine-grained sites are amalgamated together into a single coarse-grained site with capacity S and length $S\Delta$. This coarse-grained representation is referred to as a partially excluding ABM. See the schematic in Fig. 4 for an illustration. Agents can perform jumps between neighboring compartments with a rates proportional to $1/S^2$ and which scales linearly with the proportion of available space in the target compartment.

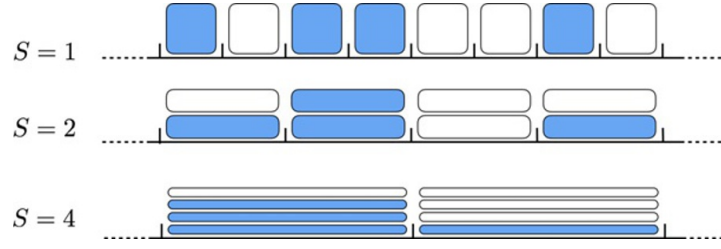


FIG. 4 Schematics of the partially excluding lattice with different values of the carrying capacity, $S = 1, 2, 4$. *Blue rectangle* represent agents and *white rectangles* represent unused spaces. For unit carrying capacity, $S = 1$, the fully excluding model is recovered, but as S increases the spatial resolution coarsens.

Taylor et al. (2015b) consider a uniform regular lattice. By comparing the fully excluding model, $S = 1$, and the partially excluding model, $S > 1$, the authors show that the mean and variance of the number of agents in each compartment is the same in both representations. In other words, it is possible to provide a consistent description of the effects of volume exclusion across different spatial scales with these coarse and fine representations. While the usage of a coarse-grained, partially excluding approach leads to a significant time saving in their example simulations, it also requires movement events to occur across a wide range of spatial and temporal scales (Walpole et al., 2013).

In a more recent work, Taylor et al. (2016) extend the study of these partially excluding models to nonuniform lattices in which the sites' carrying capacities can vary across the domain. In particular, they develop a set of hybrid methods which allow the interfacing of regions of partially excluding sites with regions of fully excluding sites. The advantage of these hybrid methods is that they allow for the study of complicated scenarios, in which the accuracy of the fully excluding model is required in some regions of space, but the partially excluding model can be exploited in other regions, allowing a considerable reduction of computational cost in comparison to the ubiquitous fully excluding model (Taylor et al., 2016).

Simpson et al. (2009) study other important extensions of the fundamental volume-excluding ABM outlined toward the start of this section. In particular, the authors incorporate the possibility of having multiple subpopulations of agents with different motility parameters and a deterministic directional bias.

The ABM of Simpson et al. (2009) is defined in discrete time and on a two-dimensional lattice with lattice step Δ . An exclusion property is implemented, requiring that a lattice site can be occupied by at most one agent at the time. Agents move according to a biased random walk with the bias modulated by a parameter $\phi \in [-1, 1]$.

In the first part of the paper, a single population of identical cells is considered. Simpson et al. (2009) derived an advection-diffusion continuum approximation of the model by writing down the occupancy master equation and letting the bias parameter scale with the spatial step $\phi \sim \mathcal{O}(\Delta)$.

The simulations of the ABM are averaged over the columns of the lattice and over many realizations. The resulting density profiles are compared with continuum descriptions showing good agreement with the corresponding PDE.

In the second part of the paper, the model is extended to describe the migration of U subpopulations or species of cells. The cells of each species move according to a biased random walk, with the motility rate $\alpha^{(u)}$ and bias intensity $\phi^{(u)}$, with $u = 1, \dots, U$, depending on the species. Using a similar derivation as in the one-species case, the authors derive a system of U advection-diffusion equations describing the evolution of the occupancy of the different species, $\bar{C}^{(u)}$. When the bias is turned off, the system reads

$$\frac{\partial \bar{C}^{(u)}}{\partial t} = D^{(u)} \frac{\partial}{\partial x} \left[\left(1 - \sum_{u=1}^U \bar{C}^{(u)} \right) \frac{\partial \bar{C}^{(u)}}{\partial x} + \bar{C}^{(u)} \frac{\partial}{\partial x} \sum_{u=1}^U \bar{C}^{(u)} \right], \quad u = 1, \dots, U \quad (19)$$

where

$$D^{(u)} = \lim_{\Delta \rightarrow 0} \frac{\alpha^{(u)} \Delta^2}{2}.$$

In this case, the exclusion property leads to nonlinear diffusivity for the individual species. However, in the case in which all species have the same motility parameters as each other, by summing all the equations, unsurprisingly, simple diffusion is recovered for the total population.

To compare the two levels of description, the authors consider a population of agents formed by two species initialized in adjacent regions with different initial densities. The results highlight a spontaneous aggregation in one of the species' density profiles in the continuous model. From this observation, [Simpson et al. \(2009\)](#) conclude that the single species densities do not obey any maximum principle, i.e., the monotonicity of the density profile is not preserved in time, although this is true for the total population.

2.1.2 Off-Lattice Models

In off-lattice ABMs the positions of the agents are represented in continuous space. This improvement adds realism to the model, since the movement of real cells is not constrained to a discrete grid. Moreover, the continuous framework introduces a larger variety of possible actions which cells can perform. For example, when modeling the migration of cells in two dimensions, the off-lattice framework allows both the distance and the direction of the movement to be a continuous variable, rather than being restricted to a discrete set of values, as in the on-lattice case. This extension is consistent with many biological observations of cell migration in which cells are not restricted to a lattice ([Plank and Sleeman, 2004](#); [Stokes and Lauffenburger, 1991](#)). However, it has the disadvantage that it makes the mathematical analysis more complicated and sometimes intractable.

2.1.2.1 Noninteracting Cells Undergoing Unbiased Movement

We consider an ABM on a one-dimensional domain, $[0, L]$, as in Fig. 5, with periodic boundary conditions. Each agent, n , is defined by its position at time t , $c_n(t) \in [0, L]$, and it occupies the interval $(c_n(t) - R, c_n(t) + R)$, where R is the analog of the cell's radius. The model is defined in continuous-time and agents perform unbiased jumps of fixed distance in one of the two directions. The rate and the distance of these jumps are denoted by α and d , respectively (see the schematics in Fig. 5 for an illustration of the model). A set of \mathcal{N} agents are initially located uniformly at random in the interval $[40, 60]$ (see Fig. 6A).

For now, we assume agents move independently of other agents' positions, as in Othmer et al. (1988). Consequently, a given point, $x \in [0, L]$, can be occupied by more than one agent simultaneously. In this case, obtaining a macroscopic description of the model can be achieved by employing a similar method to the corresponding on-lattice case (Section 2.1.1). We consider M identically prepared simulations of the model and aim to write down the master equation for the average occupancy of position x at time t , $\bar{C}(x, t)$, which is defined as

$$\bar{C}(x, t) = \frac{1}{M} \sum_{m=1}^M C^{(m)}(x, t), \quad (20)$$

where

$$C^m(x, t) = |\{n \mid c_n^m(t) = x\}|. \quad (21)$$

The master equation for $\bar{C}(x, t + \delta t)$ then reads

$$\bar{C}(x, t + \delta t) = \overbrace{\bar{C}(x, t) - \alpha \delta t \bar{C}(x, t)}^{\text{moving out of } x} + \underbrace{\frac{\alpha \delta t}{2} [\bar{C}(x - d, t) + \bar{C}(x + d, t)]}_{\text{moving into } x} + \mathcal{O}(\delta t^2), \quad (22)$$

where δt is chosen sufficiently small that at most one movement event can take place in $[t, t + \delta t)$. By Taylor expanding the terms $\bar{C}(x \pm d, t)$ to the second order about the point x and taking the limit $\delta t, d \rightarrow 0$, while keeping

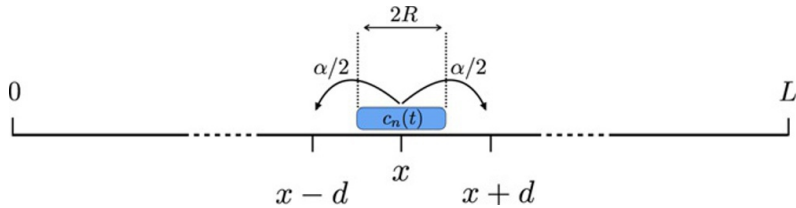


FIG. 5 Schematics of the off-lattice ABM in one dimension. A single agent, n , is represented by an blue interval of length R and whose center, $c_n(t)$, is defined as the position of the agent. The agent attempts to move with rate α a distance d in one of the two directions.

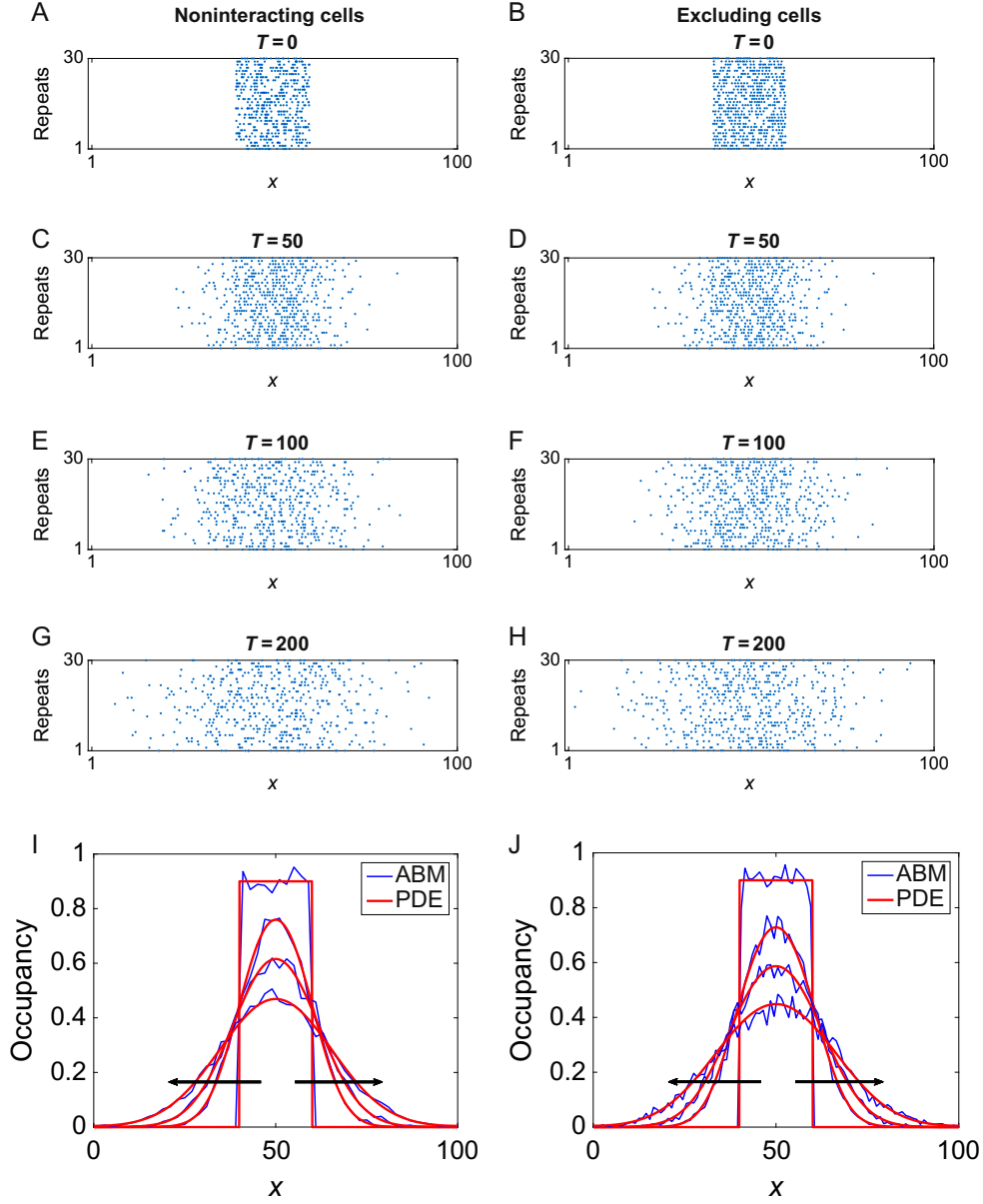


FIG. 6 Comparison between stochastic and deterministic models of cell motility on a one-dimensional off-lattice domain. Panels (A)–(H) show four snapshots of 30 simulations of the ABM with noninteracting agents as described in [Section 2.1](#) (panels (A), (C), (E), and (G)) and with excluding agents (panels (B), (D), (F), and (H)). The parameters of the model are $\alpha = 4$, $d = 0.5$, $R = 0.2$ and the simulations are shown at times $T = 0, 50, 100,$ and 200 . In each panel 30 independent repeats of the simulations are displayed on top each other. Agents are represented by a *blue ball* with radius R . In panels (I) and (J) we show a comparison between the occupancy of the ABM (*blue line*), averaged over 500 repeats, and the numerical solution of corresponding deterministic PDE (*red line*). Panel (I) is for the noninteracting model and panel (J) refers to the model with volume exclusion. In both cases, the profiles are displayed at the times $T = 0, 50, 100, 200$, with the direction of the *black arrows* indicating increasing time.

αd^2 a nonzero constant, we recover the diffusion equation (6), with the diffusion coefficient defined as

$$D = \lim_{d \rightarrow 0} \frac{\alpha d^2}{2}. \quad (23)$$

In other words, when agents are not interacting, the off-lattice framework of the ABM does not change the resulting population-level representation which was obtained for the on-lattice case. In Fig. 6I we show a comparison of the average agent density and the corresponding diffusion equation as time evolves. The results confirm a good agreement between the stochastic and the deterministic models.

2.1.2.2 Volume-Excluding Cells

Incorporating volume exclusion in an on-lattice model is a natural extension of the simple multiple occupancy model. However, in an off-lattice framework there are several ways that the effects of volume exclusion can be incorporated. In this section we follow the approach of Dyson et al. (2012), however, we highlight that other approaches can lead to slightly different results both at the individual- and population-levels.

Dyson et al. (2012) consider an exclusion mechanism in which an attempted move is aborted if it would lead to the overlap of two agents, i.e., the corresponding centers are closer than $2R$. A schematic in Fig. 7 shows an example in which an agent (blue) attempts to move in the rightwards direction and an examples of an agents (gray) which would obstruct the movement. In order to write down the occupancy master equation for the average number of agents, we need to compute the transition rates, $T^\pm(x, t)$. In particular, we need to determine the probabilities that an agent located at position x at time t , moves to the right- and left-directions, at time $t + \delta t$. By using the continuous form of the mean-field assumption (17), we can write

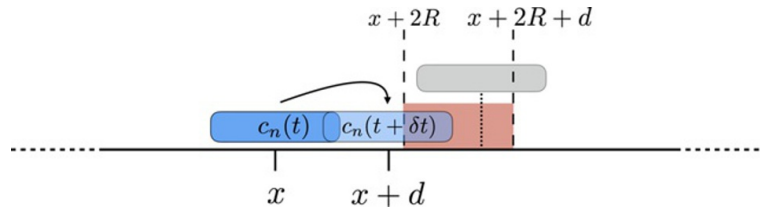


FIG. 7 Illustration of the exclusion property for the one-dimensional off-lattice model. The moving agent (*blue interval*) attempts to move in the right direction into position $x + d$ (*light blue interval*). In order for the movement to succeed, none of the other agents' centers can occupy the exclusion zone $[x + 2R, x + 2R + d)$ (highlighted in *pink*). An example of agent that would obstruct the movement is shown (*light gray interval*).

$$\mathcal{T}^+(x, t) = \frac{\alpha \delta t}{2} (1 - \mathcal{I}^+(x, t)), \quad (24a)$$

$$\mathcal{T}^-(x, t) = \frac{\alpha \delta t}{2} (1 - \mathcal{I}^-(x, t)), \quad (24b)$$

where $\mathcal{I}^+(x, t)$ and $\mathcal{I}^-(x, t)$ are the number agents in the exclusion zones $[x+2R, x+2R+d)$ and $[x-2R-d, x-2R)$, respectively, which impede the movement. See the interval highlighted in pink in Fig. 7 for an illustration right exclusion zone.

Hence the master equation for the occupancy, $C^{(m)}(x, t + \delta t)$, reads

$$\begin{aligned} C^{(m)}(x, t + \delta t) = & C^{(m)}(x, t) - \overbrace{\frac{\alpha \delta t}{2} C^{(m)}(x, t) [(1 - \mathcal{I}^-(x, t)) + (1 - \mathcal{I}^+(x, t))]}^{\text{moving out of } x} \\ & + \underbrace{\frac{\alpha \delta t}{2} (1 - C^{(m)}(x, t)) [\mathcal{I}^-(x, t) + \mathcal{I}^+(x, t)]}_{\text{moving into } x} + \mathcal{O}(\delta t^2), \end{aligned} \quad (25)$$

where $C^{(m)}$ is defined as in Eq. (21). In order to compute $\mathcal{I}^\pm(x, t)$, we reduce to the case in which there is *at most* one other agent obstructing the movement by assuming $d < 2R$. If we consider a number of agents, \mathcal{N} , sufficiently large and we use the continuous version of the mean-field approximation (17), we can write

$$\mathcal{I}^+(x, t) = \int_{2R}^{2R+d} C^{(m)}(x+y, t) dy, \quad (26a)$$

$$\mathcal{I}^-(x, t) = \int_{-2R-d}^{-2R} C^{(m)}(x+y, t) dy. \quad (26b)$$

Notice that, due to the assumption on d , the two integrals on the right-hand side of Eqs. (26) assume values in $\{0, 1\}$ and so the two transition probabilities defined in Eqs. (24) are meaningful. We can Taylor expand $\mathcal{I}^+(x, t)$ and $\mathcal{I}^-(x, t)$ to second order to obtain

$$\begin{aligned} \mathcal{I}^-(x, t) = & dC^{(m)}(x, t) + \frac{d}{2}(4R+d) \frac{\partial \bar{C}(x, t)}{\partial x} \\ & + \frac{d}{6}(12R^2 + 6Rd + d^2) \frac{\partial^2 C^{(m)}(x, t)}{\partial x^2} + \mathcal{O}[(2R+d)^4], \end{aligned} \quad (27a)$$

$$\begin{aligned} \mathcal{I}^+(x, t) = & dC^{(m)}(x, t) - \frac{d}{2}(4R+d) \frac{\partial \bar{C}(x, t)}{\partial x} \\ & + \frac{d}{6}(12R^2 + 6Rd + d^2) \frac{\partial^2 C^{(m)}(x, t)}{\partial x^2} + \mathcal{O}[(2R+d)^4]. \end{aligned} \quad (27b)$$

By substituting Eqs. (27) into Eq. (25) and taking the limit as $\delta t \rightarrow 0$, we obtain

$$\frac{\partial C^{(m)}}{\partial t} = \frac{\alpha d^2}{2} \frac{\partial}{\partial x} \left[\left(1 + (4R - d)C^{(m)} \right) \frac{\partial C^{(m)}}{\partial x} \right] + \mathcal{O}\left((R + d)^4\right), \quad (28)$$

for $m = 1, \dots, M$. We can now sum Eq. (28) over all values of m and divide by the total number of realizations, M , to obtain an equivalent equation for the average occupancy $\bar{C}(x, t,)$ using the assumption that all agents are identically uniformly distributed initially we find

$$\frac{\partial \bar{C}}{\partial t} = \frac{\alpha d^2}{2} \frac{\partial}{\partial x} \left[\left(1 + (4R - d)\bar{C} \right) \frac{\partial \bar{C}}{\partial x} \right] + \mathcal{O}\left((R + d)^4\right). \quad (29)$$

The dependence on the agents' size can be explained by noting that larger agents will collide more often. Dyson et al. (2012) identify that when d is large compared to R , the diffusion coefficient decreases to the point at which it may be negative, which might suggest the occurrence of cell aggregation. It is also important to notice that, if the agents are initialized on a lattice with step $\Delta = 2R$, and we choose $d = 2R$, the ABM is equivalent to the on-lattice ABM with excluding agents. However, for such choice of d , the derivation of Eq. (29) breaks down, which explains why simply setting $d = 2R$ in Eq. (29) does not recover the simple diffusion equation as might be expected. By taking the limit as $d \rightarrow 0$, while keeping αd^2 a nonzero constant, we arrive at a nonlinear diffusion equation:

$$\frac{\partial \bar{C}}{\partial t} = \frac{\partial}{\partial x} \left[D(\bar{C}) \frac{\partial \bar{C}}{\partial x} \right], \quad (30)$$

with

$$D(\bar{C}) = D [1 + 4R\bar{C}]$$

and D is defined as in Eq. (23).

Notice that the exclusion property results in a nonlinearity in the diffusion coefficient of the population-level equation. In particular, the term $1 + 4R\bar{C}$ in Eq. (30) leads to faster diffusion where the average occupancy is large and slower diffusion where the average occupancy decreases to zero, in which case the diffusion coefficient reaches its minimum value, D . In Fig. 6J we compare the average agent density with the numerical solution of Eq. (29) at four subsequent times. The results confirm the good agreement between the ABM and the corresponding continuum equation.

2.2 Higher Dimensions

Although the detailed derivations of the previous section are carried out for a one-dimensional interval domain, discrete-continuum equivalence frameworks can be extended to higher dimensions.

In fact, for models which do not account for crowding effects, either on- or off-lattice, the resulting macroscopic description can be obtained in a similar manner to the one-dimensional case (Codling et al., 2008; Deutsch and Dormann, 2007; Othmer et al., 1988). For the regular-square-lattice and the off-lattice models, the resulting deterministic description is the natural generalization of Eq. (6) which is given by

$$\frac{\partial \bar{C}}{\partial t} = D \nabla^2 \bar{C}, \quad (31)$$

where ∇^2 represents the Laplacian operator and the diffusion coefficient is given by

$$D = \lim_{\Delta \rightarrow 0} \frac{\alpha \Delta^2}{2r},$$

where r is the dimension. Eq. (31) is an isotropic PDE, i.e., it is not biased in any spatial direction. That such an isotropic equation can be derived from an ABM which is defined on a regular, anisotropic lattice is perhaps surprising. In other words, the intrinsic individual-level anisotropy of the ABM in the lattice directions vanishes in the diffusive macroscopic description (Codling et al., 2008; Deutsch and Dormann, 2007; Othmer et al., 1988). This same property has been demonstrated to not hold for models of cell movement which are based on velocity-jump processes (Gavagnin and Yates, 2018). We refer the reader to Section 3.4 for a more detailed discussion of such models.

Incorporating volume exclusion in higher dimensions does not lead to substantial changes for the on-lattice models in which case equation (31) still holds (Simpson et al., 2009). However, higher dimensions significantly increase the complexity of off-lattice ABMs with crowding effects. For example, Dyson and Baker (2014) study an extension of their previous one-dimensional model in two and three dimensions. In their ABMs, cells are represented as circular or spherical agents of radius R . To include volume exclusion, agents movements which would lead to an overlap of agents are aborted. In Fig. 8 we reproduce a schematic of the volume exclusion property for the ABM of Dyson and Baker (2014) in two dimensions. Notice that, to compute the probability of finding obstructing agents, it is necessary to integrate the occupancy function over the gray shaded region, A_i , of Fig. 8. This makes the computation complicated and intractable from a mathematical point of view. To overcome this problem, Dyson and Baker (2014) suggest a simplification of the calculation by extending the integral to the two blue regions, b . Clearly, if the distance of each jump, d_θ , is sufficiently small, this is a reasonable approximation which significantly simplifies the analytical calculation.

By adopting this simplification, Dyson and Baker (2014) carry out the derivation of diffusive PDEs for the average agent occupancy, leading to equations of the form of (29). In all cases, the nonlinear diffusion coefficients

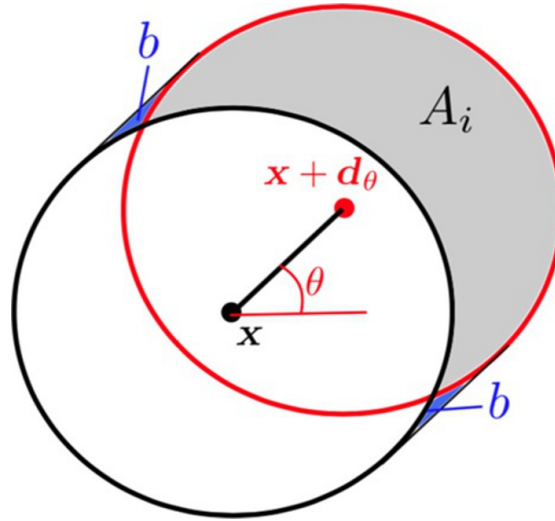


FIG. 8 Schematic representation of the volume exclusion property of the model of Dyson and Baker (2014) in two dimensions. A moving agent (black circle) is attempting a movement to a new position (red circle). The gray shaded region, A_i , is the area in which agent overlap must be avoided for a successful movement event. To simplify the algebra involved, the authors extend the integration region to include the blue shaded regions, b . Reproduced from Dyson, L., Baker, R.E., 2014. The importance of volume exclusion in modelling cellular migration. *J. Math. Biol.* 71(3), 691–711 with the permission of Journal of Mathematical Biology.

are increasing functions of the agents' radii, R , and decreasing functions of the jump distance, d .

Finally, the authors consider a system comprising multiple species of agents in which the size and the movement rates of the agents depend on the species to which they belong. The authors consider an example with two species, one comprising agents that are large and slow moving, and the other smaller but quicker. A system of PDEs for the densities of the two species is derived and the results show that the species with smaller agents is less affected by volume exclusion. Interestingly, the effect is reduced in the area where the two species coexist, since the majority of the area is occupied by smaller agents.

2.3 Higher Order Closure Approximations

In general, when agent–agent interactions are included in an ABM, as in the example of volume exclusion above, the positions of the agents are not independent. When this happens, the evolution of the average agent density depends on the distribution of agent pairs, which also depends on the distribution of agent triples, and so on. Therefore, if we aim to derive a deterministic representation of the model, we have to deal with an infinite system of unclosed equations for each of the spatial moments of the agents' distribution. In order to overcome this problem, a moment closure approximation is necessary in order to make the system amenable to mathematical analysis.

The mean-field closure, given by Eq. (17), represents the easiest form of moment closure which is typically used (Cheeseman et al., 2014; Dyson and Baker, 2014; Dyson et al., 2012; Simpson et al., 2009). Although such rudimentary approximation provides good results in a wide range of scenarios, when the spatial distribution of the agents is strongly correlated, the mean-field closure can lead to under- or overestimations of the total agent density, resulting in a poor agreement between the stochastic ABM and its deterministic representation.

Baker and Simpson (2010) investigate the role of spatial correlation in exclusion processes and its effect on the agreement between the averaged agent-based dynamics and the continuum approximation.

They consider an excluding ABM on two- and three-dimensional lattices. Agents can move to neighboring lattice sites, with rate α_m , and proliferate by placing daughter agents into neighboring lattice sites, with rate α_p . The authors derive a general set of master equations for the k -point distribution functions, $\rho^{(k)}$ (of which the 1-point distribution function is simply the density, the 2-point distribution function is the pairwise occupancy, etc.). This infinite hierarchy of master equations is unclosed: the differential equations for the $(k - 1)$ -point distribution functions depend on the k -point distribution functions. Therefore, a closure approximation is required in order to solve for the lower-order distribution functions. The authors compare the averaged density for the discrete model with the first-level moment closure (the *mean-field approximation*), in which neighboring sites are assumed independent (see Eq. 17). They also compare to a second-level moment closure (the Kirkwood superposition approximation), which takes into account pairwise spatial correlations. On the square lattice, the distance between two lattice sites increases irregularly ($\Delta, \sqrt{2}\Delta, 2\Delta, \sqrt{5}\Delta$) and the neighbors for each site, therefore, have to be calculated separately for each distance. This fact means that the number of ODEs for the correlation functions becomes intractable very quickly. The system of ODEs, therefore, needs to be truncated at a maximum distance r_{\max} , beyond which the sites are considered independent. Different values of r_{\max} are compared and the results suggest that, in two dimensions, the system can be truncated at $r_{\max} = 3$ without losing accuracy. For the three-dimensional case, the cut-off can be reduced to $r_{\max} = 2$. The inclusion of correlations in the model, even if closed at level two, provides a significant improvement in the approximation to the agent-based model.

Baker and Simpson (2010) also investigate the effects of motility, birth, and death events on spatial correlations. As the proliferation parameter increases with respect to the (fixed) motility parameter, spatial correlations play a more important role and the mean-field prediction appears to overestimate the growth of the population. This is due to the fact that cell motility cannot break up the correlations, caused by the appearance of new agents close to their parents, sufficiently quickly. This leads to cluster formation which, in

turn, reduces the number of successful proliferation events, slowing population growth. When agent death is included in the model, a counterintuitive effect appears. We may naively expect that deaths would decrease spatial correlation and so increase the agreement between the mean-field and the agent-based models. Instead the opposite happens. A high death rate has a similar effect to having a more sparsely populated initial seed. It provides opportunities for correlations to build up where previously sites occupancies were uncorrelated.

[Markham et al. \(2013a\)](#) continue the work of [Baker and Simpson \(2010\)](#) by deriving a deterministic, continuum analog of a spatially extended, on-lattice, agent-based model. As in the previous paper, they look at the influence of including spatial correlations in the continuum model and determine how this improves the agreement, in comparison with the mean-field model, which assumes absence of spatial correlations.

The ABM considered is the same as in [Baker and Simpson \(2010\)](#), i.e., a volume exclusion process on a lattice on which individuals can move, proliferate, and die with rates α_m , α_p , and α_d , respectively.

[Baker and Simpson \(2010\)](#) truncate the pairwise correlation functions at a maximum distance, r_{\max} , resulting in a system of ODEs. The aim of [Markham et al. \(2013a\)](#), however, is to derive a tractable deterministic PDE for the evolution of the pairwise spatial correlation function. Specifically, as the lattice step, Δ , is small relative to the size of the domain, one can Taylor expand the correlation functions up to second order in Δ and then move to radial coordinates. Finally, assuming that $\alpha\Delta^2$ remains constant as $\Delta \rightarrow 0$, one obtains a pair of PDEs.

A trivial analysis shows that, as the motility parameter, α_m , increases, the diffusion coefficient of the PDE for the correlation function increases. This agrees with the intuitive prediction that higher rates of movement break up clusters of agents more effectively. On the other hand, the reaction term of the PDE is negative and decreases as the proliferation rate, α_p , increases. This means that increasing proliferation leads to a decrease in correlations. This fact appears to contradict the results of [Baker and Simpson \(2010\)](#). Nevertheless, [Markham et al. \(2013a\)](#) provide an informal explanation for this phenomenon.

The results of spatially extended simulations show good agreement between the solution of the PDE and the discrete model. In particular, the PDE approximation behaves similarly to the ABM which, in the case of nonzero death rate, is significantly lower than the mean-field prediction. Moreover, an improvement in the agreement between the PDE and the ABM is obtained if the movement rate increases or if higher-dimensional domains are considered.

[Markham et al. \(2013a\)](#) find that there exists a region of parameter space for which the deterministic models (either with or without correlations) cannot replicate the ABM dynamics. This phenomenon corresponds to high values of

the death rate which leads the population of the discrete model to eventually go extinct.

Markham et al. (2013b) extend their previous spatial correlation model (Markham et al., 2013a) to an ABM with heterogeneous agents. They provide other examples in which spatial correlations play a crucial role in predicting the population-level behavior correctly. Agents are divided into U species and can move, proliferate, and die with rates α_m^I , α_p^I , and α_d^I , respectively, where $I \in \{0, \dots, U\}$ denotes the species of the selected agent. With the same idea as their previous work, Markham et al. (2013b) focus on the evolution of the pairwise correlation, closing the system at the level above. Firstly they derive a set of ODEs for the system in which the pairwise correlation is divided into *auto-correlation*, i.e., between agents of the same species, and *cross-correlation*, if the agents belong to different species. Subsequently they obtain a set of PDEs by Taylor-expanding and taking the limit as the lattice step goes to zero. A key assumption throughout the chapter is that the pairwise correlation depends only on the distance between the agents, i.e., it is translationally invariant and isotropic. For this reason, the agents are always assumed to be spread uniformly across the domain initially.

The results for two species show a good agreement between the PDE model and the ABM in comparison to the agreement between the mean-field approximation and the ABM. The authors assume agent death is negligible and consider two species of cells with the same proliferation rates but different movement rates. While the logistic dynamics of the mean-field model predicts that the densities of the two species will converge to the same steady state, in the ABM, the species with the greater movement rate reaches a higher density at the equilibrium. They also show, with specific values of proliferation rates, that the results can be the reversed, i.e., in the mean-field approximation, the density of one species at the equilibrium is greater than the other species' density, while the ABM predicts that, at equilibrium, the two species reach the same density. Remarkably, the PDE model incorporating agent-agent correlation shows a good agreement with the ABM in all these scenarios.

3 MODEL EXTENSIONS

The models, summarized in the previous section, represent the fundamental basis for the vast majority of spatially extended models of cell migration. However, during the process of migration, cells perform a variety of other actions and interactions whose role can dramatically impact upon the macroscopic behavior of the system (Carmona-Fontaine et al., 2008; Keynes and Cook, 1992; Niessen, 2007; Tambe et al., 2011; Treppe et al., 2009; Ward et al., 2003). Here we provide a brief overview of some of the most important extensions and modification of the standard ABMs and demonstrate their effects on the corresponding macroscopic models.

3.1 Cell Proliferation

In all of the models considered so far, the ability of cells to divide and produce daughter cells was not included. In reality, in certain circumstances (e.g., tumor invasion and wound healing) the role of cell proliferation is crucial for the dynamics of the system (Maini et al., 2004; Sherratt and Chaplain, 2001).

Simpson et al. (2007) propose an ABM which takes into account cells' motility and proliferation. The ABM is defined on a two-dimensional lattice with a simple exclusion property, meaning that each lattice site can be occupied by at most one agent at the time. The model advances in discrete time. At each time step, every agent can move and proliferate with probabilities P_m and P_p , respectively. Agents move according to a simple, unbiased random walk to one of the their four nearest sites (von Neumann neighbors). When a proliferation event occurs, the proliferating agent moves to one of its eight surrounding sites (Moore neighbors) and the new offspring is displaced to the site diametrically opposite. If either of the two sites is already occupied, the proliferation event is aborted. Both motility and proliferation events are limited by a carrying capacity $\kappa \in \{1, \dots, 8\}$. In particular, if an agent attempts to move or proliferate into a site with a number of surrounding neighbors greater than κ , the event is aborted.

Firstly, the authors simulate an invasion wave and investigate some of the features of the ABM. They relate and compare the behavior of their ABM with the traditional deterministic Fisher equation (Fisher, 1937):

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} + vC \left(1 - \frac{C}{K}\right), \quad (32)$$

where D is the diffusion coefficient, v is the growth rate, and K is the carrying capacity. Note that the average speed of invasion of the deterministic Fisher wave is known to be $v^* = 2\sqrt{Dv}$.

The results of the ABM simulations show that, consistent with the continuum equation, the wave speed of the invasion increases with motility parameter, P_m , and proliferation probability, P_p . However, the speed in the ABM is more sensitive to variation in proliferation than in motility, in contrast to the corresponding deterministic description which predicts equal sensitivity. As in the continuum setting, the speed of invasion is found to be independent of the carrying capacity.

To establish a connection between the ABM and the traditional continuum model given by Eq. (32), Simpson et al. (2007) compare the behavior of the two levels of description for scenarios in which either proliferation or motility are active, but not both. When agents can only proliferate, the total density evolves in a logistic manner and the authors identify a linear relationship between the parameters of the agent-based and population-level models. When agents can move but not proliferate, the results show that the diffusivity of the model is a decreasing function of the background density. Nevertheless,

a linear dependence is found between the motility probability of the ABM, P_m , and the diffusion coefficient at zero agent density.

Finally, the ABM is used to investigate the individual cells' trajectories within the invasion wave. The authors use synthetic data to obtain statistical information on the average direction of movement. The results compare the behavior of the most advanced cell in the invasion with a second cell close behind the wave front. Despite the cells moving according to an unbiased random walk, the crowding effects inhibit the motility into highly populated regions, inducing a bias toward the direction of the invasion in both of the cells considered. By using the individual agent trajectories, this bias can be quantified. Consistent with experimental data (Druckendbrod and Epstein, 2007), the observations on the ABM show a more evident bias for cells near the wave front in comparison to cells behind the wave front.

Cheeseman et al. (2014) continued the study of cell proliferation by modeling the spatial and temporal dynamics of different cellular lineages within an invasion wave. The authors defined an ABM model on a two-dimensional lattice with a volume exclusion property. The model is initialized with a population of cells located on one end of the domain (see Fig. 9A). Agents are labeled according either to their generation number or the lineage to which they belong. The results for the ABM simulation show a clear spatial organization in the distribution of the agent generation number. However, there is a large individual variability in the spatial distribution of a single agent's lineage tracing (see Fig. 9). In order to reproduce the dynamics of the agent generation number, a set of nonlinear diffusion equations are derived from the ABM and are studied in one dimension. In particular, by using the same approach as Simpson et al. (2009), Cheeseman et al. (2014) derive a system of conservation of mass equations describing the density profile for each generation $\eta_i(x, t)$. In order to investigate the lineage tracings, they develop a Generation-Dependent Galton-Watson (GDGW) process. The authors look at the Lorenz curves, which describe the proportional contributions of each of the initial seed cells' lineages to the final population (Lorenz, 1905).

The results from agent-based and PDE models show that the invasion wave is composed of spatially regular and predictable generations. As the invasion occurs, the older generations reach a steady state behind the traveling wave. This structure is more apparent when averaging over more realizations of the ABM model. Indeed, both the mean and the variance of the generation number increase linearly with the distance from the location of the initial group of cells. Conversely, the individual agent lineages exhibit a clear asymmetry in their contribution to the final population. A small group of cells (*superstars*) generate the majority of the total population. The GDGW model shows a good agreement with the ABM results. This fact is interesting, since the GDGW process ignores some of the spatial and temporal correlations between individual lineages. Nevertheless this phenomenon can be partially

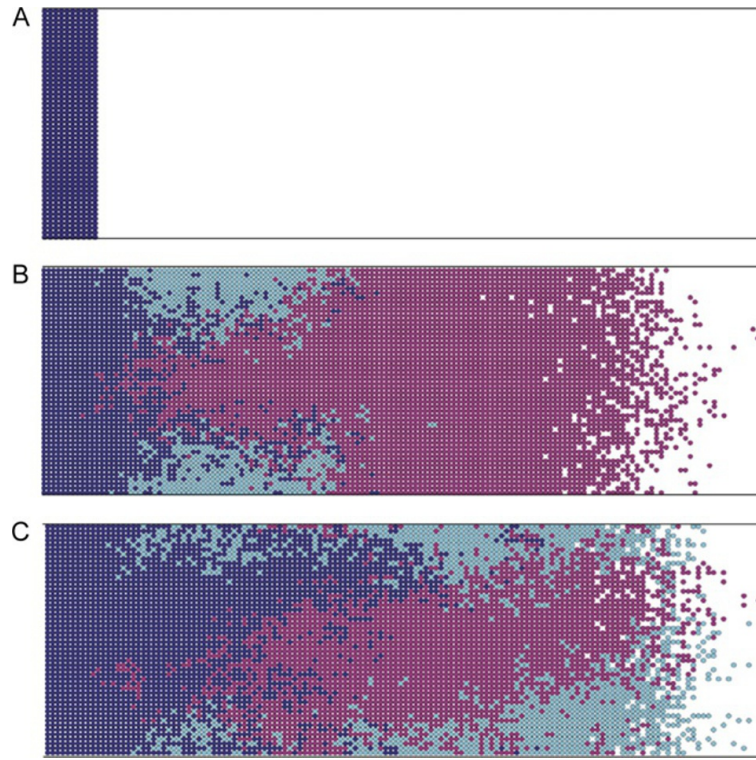


FIG. 9 Two simulations of the ABM of [Cheeseman et al. \(2014\)](#) for an invasion wave with agent lineage tracings. Panel (A) shows the initial condition for all simulations. Panels (B) and (C) show two realizations of the ABM in which the agents populate the domain through movement and proliferation. The largest and second largest single agent lineage tracings (pink and turquoise, respectively) and the 498 other agent lineage tracings (all collected together in blue) are illustrated. *Reproduced from Cheeseman, B.L., Newgreen, D.F., Landman, K.A., 2014. Spatial and temporal dynamics of cell generations within an invasion wave: a link to cell lineage tracing. J. Theor. Biol. 363, 344–356 with permission of Journal of Theoretical Biology.*

explained by noticing that the competition between agents affects the generation densities, η_i , which are used in the Galton–Watson process.

As highlighted in [Section 2](#), choosing between on- and off-lattice frameworks can lead to differing macroscopic descriptions for excluding ABMs incorporating motility only. The same can be said for models of cell proliferation. [Plank and Simpson \(2012, 2013\)](#) provide a pioneering comparison between on-lattice and off-lattice models of proliferating and migrating cells.

The authors develop a new off-lattice discrete-time ABM for migration and proliferation of cells in a two-dimensional domain. Agents are represented as incompressible circles of diameter Δ and the total number of agents at time t is denoted $\mathcal{N}(t)$. At each time step, $\mathcal{N}(t)$ agents are selected independently and are given the chance to move with probability P_m . A selected agent attempts to move a fixed distance, Δ , in a given direction, θ , which is chosen uniformly at random in $[0, 2\pi)$. To include crowding effects, the movement is aborted if any of the other agents lie within a distance Δ of the line segment

which connects the current location to its potential new location. Once all the motility events have been attempted, another $\mathcal{N}(t)$ agents are selected independently to attempt proliferation events with probability P_p . A proliferating agent attempts to divide into two daughter agents whose positions are chosen diametrically opposite each other at distance $\Delta/2$ from the original agent. The chosen proliferation event takes place only if it does not cause overlap between agents.

The authors derive a mean-field approximate ordinary differential equation for the spatially averaged agent density, $C_m(t)$. In doing this, they assume that the population of agents is homogeneous in the domain, which is known to be a poor assumption for large values of proliferation (Baker and Simpson, 2010; Markham et al., 2013b).

Plank and Simpson (2012) show a comparison between their off-lattice model and a standard on-lattice model (Simpson et al., 2010) for a spatially uniform initial condition. For small values of agent density, the absence of significant agent–agent interactions leads to similar behaviors for the on- and off-lattice models. However, the two types of approach show a substantial difference for high densities. The on-lattice model leads to a faster growth of the population density. The density eventually reaches the maximum value of unity when all the lattice sites are occupied. Conversely, it is impossible for agents in the off-lattice model to reach the theoretical maximum agent density. This is because the agents are not perfectly aligned, and at high densities become jammed in more realistic, yet irregular configurations. The appearance of a natural carrying capacity in the off-lattice model, as opposed to the artificial value induced by an on-lattice approach, suggests the off-lattice model is a more suitable representation for biological applications.

Plank and Simpson (2013) further investigate the behavior of their on-lattice and off-lattice models in scenarios in which the spatial variability of the cell density profile plays an important role. In particular, they initialize the domain by uniformly populating only the left-hand side region with a relatively low density. They then proceed to study the speed and the shape of the resulting invasion waves in the two models.

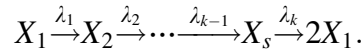
Their results highlight that the two models behave differently behind the invasion front. However, the models' behaviors are similar at the leading edge. These findings are consistent with the previous observation that the effects of crowding are more evident in the off-lattice model than in the on-lattice model (Plank and Simpson, 2012). Specifically, in the invaded region, where the agent density is higher, the population size of the off-lattice model approaches carrying capacity slower than the analogous on-lattice model. However, at the front of the invasion, the low value of agent density makes the two models almost indistinguishable. This implies that, in the long term, the speeds of the invasion fronts in the on-lattice and off-lattice models are the same.

Finally, Plank and Simpson (2013) carry out a least-squares parameter estimation in order to fit the two continuum representations to density profile

data simulated using the off-lattice ABM. The solutions for both the on-lattice and off-lattice PDEs are in good agreement with the simulated data and it is hard to distinguish between the two fitted curves. The authors present this as an example that highlights the difficulty of choosing the correct mathematical framework when modeling real experimental data.

Typically, cell proliferation is represented in ABMs as a Poisson process with a certain rate α_p , which means that each agent's interdivision times are exponentially distributed (Mort et al., 2016; Simpson et al., 2007; Treloar et al., 2013; Turner et al., 2009). One of the advantages of this approach is that, due to the memoryless property of the exponential distribution, the process can be efficiently simulated using the popular Gillespie algorithm (Gillespie, 1977).

However, assuming exponentially distributed cell interdivision times is not biologically realistic (Golubev, 2016). In particular, the monotonicity of the exponential distribution implies that the most likely time for a cell to divide is immediately after its own creation. Recently, Yates et al. (2017) have proposed a novel model for incorporating nonexponentially distributed cell cycle times, based on a multistage scheme. The authors divide the cell cycle into s stages. The waiting time for an agent to pass from the i -th phase to the $(i + 1)$ -th phase is exponentially distributed with rate λ_i . When an agent exits from the s -th stage, it divides into two daughter agents which are initialized in the first phase. This can be summarized by the following chain of reactions:



In general, the resulting probability distribution of the total cell cycle is a hypoexponential distribution. Note that this general implementation involves k independent parameters $(\lambda_1, \dots, \lambda_k)$, one for each stage transition and, if k is large, this may lead to issues of parameter identifiability. In order to reduce the number of free parameters while maintaining the advantage of using a multistage representation, Yates et al. (2017) focus on two cases: the case in which all the transition rates are identical, $\lambda_i = \lambda$ for $i = 1, \dots, k$, and the case in which all the transition rates are identical apart from one, $\lambda_i = \lambda$ for $i = 1, \dots, k - 1$. The corresponding distribution for the total cell cycle in these two cases are the Erlang distribution and exponentially modified Erlang distribution, respectively. The authors show these distributions to be both biologically plausible and computational feasible.

In order to investigate how the multistage representation of the cell cycle affects the total population growth, Yates et al. (2017) implemented two models (one spatial and one nonspatial). Firstly, they modified the model of Turner et al. (2009) in which the spatial position of the cells is considered unimportant. The authors investigate the stage distribution of agents at the steady state for large times. By writing down a system of ODEs describing the average proportion of agents at each stage j , M_j , for $j = 1, \dots, s$,

Yates et al. (2017) show that the number of agents in each stage is not proportional to the average length of that stage. In particular, for identical transition rates, as the total number of stages becomes large, the proportion of agents at the first stage approaches twice that of agents at the last stage.

The authors also modify the spatially extended ABM of Baker et al. (2010) to include their multistage proliferation scheme. The agents are initialized uniformly on a lattice with periodic boundary conditions and a standard volume exclusion property. Agents attempt movements with rate α_m . To facilitate the comparison with the traditional model with an exponentially distributed waiting time of rate α_p , the authors consider the case of identical transition rates $\lambda_i = s\alpha_p$ for $i = 1, \dots, k$. Hence the average waiting time required for progress through all the s stages is independent of the number of stages. In particular, the two models, with and without multistage scheme, have the same average attempted proliferation waiting time.

The simulations of the ABM highlight that, for low-density colonies, the multistage scheme leads to a slower population growth compared to the model with exponentially distributed proliferation. However, under the realistic assumption that agents that fail to proliferate due to crowding remain at the last stage (as opposed to returning back to the first stage of the cycle), a clear proliferating rim of (gray) cells can be seen with the bulk of cells being kept at stage s (see Fig. 10). Agents located in this rim, reattempt division after aborted events quicker than in the single-stage cell cycle model since they only have to wait a time α_p/k on average. Therefore, the effective average interdivision time for cells with a multistage cell cycle at the proliferating rim of a cluster decreases in comparison with cells to a single-stage cell cycle.

3.2 Cell Interactions

Cells can undergo a great variety of interactions in response to their environment and more specifically, their neighboring cells (Cai et al., 2006; Dworkin and Kaiser, 1985; Tambe et al., 2011; Trepat et al., 2009). Typically, these can be divided in two broad categories: *indirect* and *direct* (Othmer and Stevens, 1997).

Indirect interactions are when cells have the ability to detect the presence of an external signal and change their behavior according to its concentration. They can affect the movement speed or the turning rate (*kinesis*) (Cai et al., 2006), they may induce a directional bias (*taxis*) (Painter and Hillen, 2002), or they may involve a combination of these (Urban and Othmer, 2004). In general, the external signal can depend on the cells themselves. For example, it can be directly produced by a moving cell as it moves, forming a trail behind it, as it has been shown for Myxobacteria (Dworkin and Kaiser, 1985). Notice that this type of interaction does not involve direct cell–cell contact, communication is mediated only through the external signal. Direct interactions are based on the ability of cells to detect and interact with

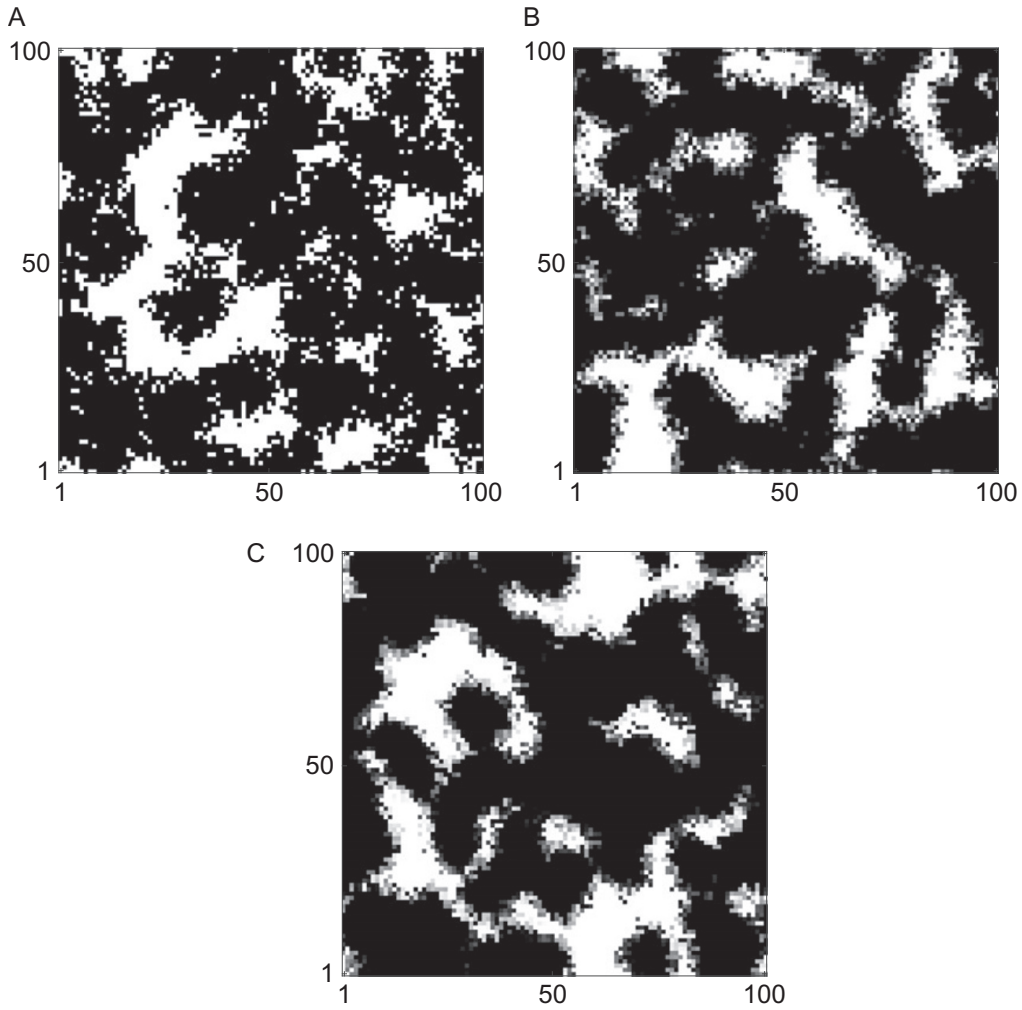


FIG. 10 Influence of multistage representation of the cell cycle on the spatial coverage of cells populating the domain at $t = 10$. Agents in different stages are represented by *different shades of gray*. Darker shading corresponds to later stages, and *white* indicates empty sites. Parameters are $\alpha_m = 1$, $\alpha_p = 1$. The initial condition consists of 100 agents uniformly distributed on the lattice. Panels (A), (B), and (C) correspond to the cases $s = 1, 10, 100$, respectively. Increasing s causes an increase in the total cell density in this scenario at time $t = 10$. *Reproduced from Yates, C.A., Ford, M.J., Mort, R.L., 2017. A multi-stage representation of cell proliferation as a Markov process. Bull. Math. Biol. 79 (12), 2905–2928 with permission of Bulletin of Mathematical Biology.*

surrounding cells. These include contact interactions such as volume exclusion (Abercrombie, 1979) and adhesion–repulsion (Tambe et al., 2011; Trepap et al., 2009).

The introduction of these interactions in models of cell migration is crucial, since they often constitute the driving forces that generate spatial structure at the population scale. In this section we focus our attention on a set of important cell interactions: the indirect response to a chemical signal, direct adhesion–repulsion forces between cells, and the ability of cells to push and pull each other.

3.2.1 Chemotaxis

Othmer and Stevens (1997) derive a general class of PDEs from a set of ABMs which incorporate a simple signal following mechanism. The aim of their paper is to determine whether stable bacterial aggregation can arise as a result of a simple chemotactic response.

The main ABM considered in the paper consists of a system of noninteracting agents which move on a one-dimensional lattice with unit step size, according to a continuous-time, nearest-neighbor random walk. As time evolves, agents produce an attracting signal, w , which sits on an embedded lattice of half the step size.

The authors distinguish three types of model depending on the information used to compute the transition rates. These are the *local* model, if the transitions depend only on the signal concentration at the current site, w_n ; the *barrier* model, in which cells only sense the value of the signal at the short-range, nearest-neighbor sites, $w_{n\pm 1/2}$; and the *gradient-based* model, in which the transition rates depend on the long-range, nearest-neighbor differences, $w_n - w_{n\pm 1}$. In the barrier and gradient-based cases, an additional distinction is made between normalized and unnormalized transition rates. For unnormalized rates, larger values of w lead to a faster total movement rate, whereas, for normalized rates, the average waiting time at each site does not depend on the signal concentration. In all cases, by using a limiting argument, the authors derive the corresponding continuous diffusive approximations for the average agent density.

Othmer and Stevens (1997) carry out a stability analysis of the solutions of the continuous barrier model with normalized transition rates. The analysis is repeated with three different models for the evolution of the signal: linear growth, exponential growth, and saturating growth. The results show that linear growth can only lead to uniform agent density profiles, even when starting from a single peak in the initial distribution of agents (*collapse*). On the other hand, when the signal grows exponentially, the system tends to converge to a single peaked distribution (*blow-up*). Finally, when a saturation in the production of w is assumed, stable aggregation can occur as a result of the interplay between the production of the signal and the short-range chemotactic response. The formation of such aggregates strongly depends on the decay rate of the signal and on the initial condition. In particular, a large decay rate or a small initial value of w can lead to aggregation or blow-up, whereas in the absence of signal decay, the agent density will eventually collapse to uniformity.

3.2.2 Adhesion–Repulsion

Anguige and Schmeiser (2009) investigate the emergence of aggregation through the mechanisms of cell–cell adhesion and volume exclusion. They predict, under their model, that the spontaneous aggregation of cells is not

possible if the initial cell density throughout the domain is too low, regardless of the intensity of the adhesion force.

To describe cell migration, [Anguige and Schmeiser \(2009\)](#) consider a discrete-space, continuous-time random walk model on the unit interval. Multiple agents can occupy the same compartment up to a carrying capacity, S , and they move at random with given rates to one of the two nearest-neighbor compartments. In order to include volume exclusion in the model, the transition rates in both directions are decreased linearly with the density of the target site. If a site is fully occupied, no transitions into that site can occur. To incorporate adhesive forces, the rate of moving in a particular direction is decreased linearly with the density of the adjacent site in the opposite direction, in proportion to an *adhesion parameter*, β .

By Taylor expanding and neglecting terms of third (or higher) order it is possible to derive the diffusive limit of the discrete model as the number of lattice sites goes to infinity. This macroscopic model is a nonlinear diffusion equation as Eq. (30) with

$$D(\bar{C}) = D \left[3\beta \left(\bar{C} - \frac{2}{3} \right)^2 + 1 - \frac{4}{3}\beta \right],$$

for which the homogenous density profile is the only steady state. [Anguige and Schmeiser \(2009\)](#) find that there is a critical value of the adhesion parameter $\beta_c = 0.75$. For the low-adhesion regime ($\beta < \beta_c$), pattern formation is not possible in either the discrete or continuum descriptions.

When the adhesion force is large ($\beta > \beta_c$), the results of the discrete model show complex behaviors such as pattern formation and spatial oscillations in density. Patterns in the discrete model are found to be transient and metastable. All clusters which arise eventually coalesce to a single stable aggregate, or to two aggregates separated by a single trough. For the macroscopic PDE representation, the authors identify an interval of unstable values of density for which the diffusion coefficient of the nonlinear PDE takes negative values making the continuum model ill-posed.

To overcome the issue of the ill-posedness in the PDE and to obtain a reasonable continuum description, [Anguige and Schmeiser \(2009\)](#) propose to include more terms from the Taylor expansion of their discrete model. These higher order corrections result in a fourth-order diffusion equation reminiscent of the Cahn–Hilliard equation ([Sun and Ward, 2000](#)). The presence of a viscosity term in the revised equation allows the authors to prove it has a well-posed initial value problem.

[Thompson et al. \(2012\)](#) continue the work of [Anguige and Schmeiser \(2009\)](#) on the modeling of cell–cell adhesion at multiple scales. The authors consider a modification of the ABM of [Anguige and Schmeiser \(2009\)](#) in order to study the interactions with a second species of cell. In this new model, agents are of two types, A and B , and there are four different

coefficients ($\beta_{A,A}$, $\beta_{B,B}$, $\beta_{A,B}$ and $\beta_{B,A}$) governing the intra- and interspecies adhesion forces, respectively. [Thompson et al. \(2012\)](#) show that the model is capable of reproducing three configurations (*complete sorting*, *engulfment*, and *mixing*), which have been observed previously both in experiments and continuous models ([Armstrong et al., 2006](#)). In [Fig. 11](#) we report an example of these three configurations depending on the choice of the adhesion parameters. Specifically, if interspecies adhesion is small and intraspecies adhesion is large, the system reaches a structured configuration (cell sorting) in which agents of the same species tend to cluster together (see left column of [Fig. 11](#)). If interspecies adhesion is small and one of the intraspecies adhesions is larger than the other, then engulfment of the species with the larger adhesion occurs (see middle column of [Fig. 11](#)). Finally, when interspecies adhesion is large and the intraspecies adhesion is small, then mixing occurs, i.e., agents of the two species are uniformly distributed across the domain (see right column of [Fig. 11](#)).

Recently, [Binny et al. \(2015\)](#) have studied a one-dimensional off-lattice ABM (later extended to two-dimensions ([Binny et al., 2016](#))) in which both the rate and the direction of the movement of each agent are influenced by the configuration of the neighboring agents. In their model an agent, n , moves with a rate, α_n , which is computed by

$$\alpha_n(t) = \max \left\{ 0, \tilde{\alpha} + \sum_{m \neq n} w(c_m(t) - c_n(t)) \right\},$$

where $\tilde{\alpha}$ is an intrinsic motility rate, independent of the other agents' positions, and $w(z)$ is a kernel function weighting the strength of interaction between agents positioned a distance z from each other. [Binny et al. \(2015\)](#) consider the case in which the kernel is Gaussian:

$$w(z) = \gamma_r \exp \left(-\frac{z^2}{\sigma_r^2} \right), \quad (33)$$

where γ_r and σ_r determine the intensity and the range of the interaction. Notice that if $\gamma_r > 0$, agents tend to move more often when they are surrounded by other close agents. Conversely, for $\gamma_r < 0$, the presence of close neighbors inhibits the motility. When a movement event takes place, the moving agent performs a jump of random length and direction. The length of jump is drawn at random from a Laplace distribution (i.e., short steps are more likely to be taken than long jumps), independent of the other agents' positions. For the direction of the movement, the authors incorporate a directional bias, $b_n(t)$, such that the presence of neighboring agents can affect the final direction. They defined the directional bias for an agent, n , as

$$b_n(t) = \sum_{m \neq n} v'(c_m(t) - c_n(t)),$$

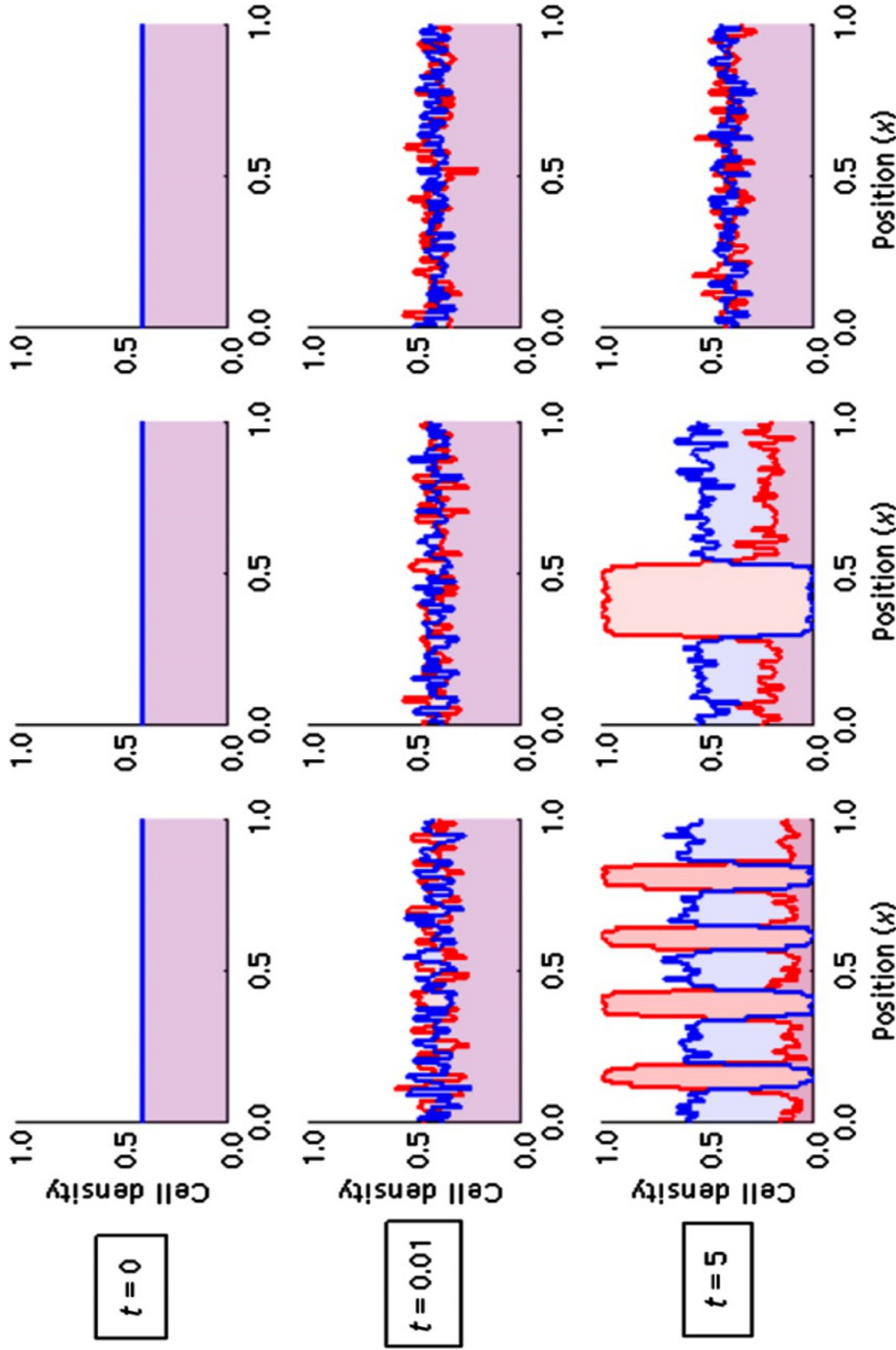


FIG. 11 Simulations of the two species ABM of Thompson et al. (2012) in one dimension. Blue lines denote the density of the type A agents and red lines the density of type B agents. The domain is initialized with homogeneously distributed agents of type A with density 0.4. Depending on the intra- and interadhesion forces, three qualitative configurations are possible: $\beta_{A,A} = 0.1$, $\beta_{B,B} = 0.9$, and $\beta_{A,B} = \beta_{B,A} = 0.1$, $\beta_{B,B} = 0.9$, and $\beta_{A,A} = 0.9$, and $\beta_{A,B} = \beta_{B,A} = 0.2$ lead to engulfment (middle) and $\beta_{A,A} = 0.1$, $\beta_{B,B} = 0.9$, and $\beta_{A,B} = \beta_{B,A} = 0.6$ produce a mixing configuration (right). Reproduced from Thompson, R.N., Yates, C.A., Baker, R.E., 2012. Modelling cell migration and adhesion during development. *Bull. Math. Biol.* 1–17 with permission of Bulletin of Mathematical Biology.

where $v(z)$ is a Gaussian kernel function of the same form of Eq. (33) with intensity and range given by γ_b and σ_b , respectively. If $\gamma_b > 0$, agents are biased to move away from highly concentrated regions, whereas if $\gamma_b < 0$, agents tend to move toward one another.

To study the resulting spatial structure from a deterministic prospective, Binny et al. (2015) develop a population-level model in terms of the first two spatial moments of the ABM. By using the Kirkwood superposition approximation (see Section 2.3 for a discussion about alternative approaches to the mean-field approximation), the authors derive a closed system of two ordinary differential equations (ODEs) describing the evolution of the average density of single agents, Z_1 and of pairs of agents separated by a distance ξ , $Z_2(\xi)$.

In order to test the accuracy of their deterministic representation against the numerically simulated individual-based model, the authors compare the pair correlation functions (PCFs) (Illian et al., 2008) of the ABM with the PCF predicted by the spatial moment model. The pair correlation function can be expressed in terms of the first two spatial moments as

$$PCF_{SM}(\xi) = \frac{Z(\xi)}{Z_1^2}. \quad (34)$$

Notice that $PCF_{SM} \equiv 1$ corresponds to the case of absence of spatial structure, whereas $PCF_{SM}(\xi) > 1$ denotes evidence of spatial correlation and $PCF_{SM}(\xi) < 1$ implies anticorrelation, at length ξ .

Their findings show that the moment model provides a good approximation of the spatial structure predicted by the ABM for a wide range of parameter choices. The moment model underestimates or overestimates the spatial structure only for large values of intensity γ_m and γ_b . Binny et al. (2015) suggest that this is due to higher order interactions which are not incorporated in the deterministic representation.

Binny et al. (2016) continue the study of agent-agent interactions by extending the previous model to two dimensions and including agent proliferation and death. In the new model, agents are allowed to proliferate with a rate which is dependent on their neighborhood and defined similarly to $\alpha_n(t)$ of Eq. (33). The rate of agent death is assumed independent of the location of other agents. Following Binny et al. (2015), the authors derive a deterministic model for the first two spatial moments of the ABM. After comparing the performance of four different types of moment closure, the authors select an asymmetric power-2 closure as the most appropriate (Murrell et al., 2004; Raghieb et al., 2011) which they adopt throughout the paper. The results show that the good agreement between the stochastic and deterministic models, found in Binny et al. (2015), is preserved when proliferation and death are incorporated in the model. Unsurprisingly, the presence of nonuniform spatial structure leads to both population growth and the average cell density

differing significantly from the predictions of a mean-field model that ignores agent–agent interactions.

Finally, [Binny et al. \(2015, 2016\)](#) investigate the role of neighbor-dependent movement and proliferation in the formation of spatial structure. In general, movement tends to break up spatial structure because new agents can move out of clusters generated by short-range proliferation ([Baker and Simpson, 2010](#)). When movement is biased away from neighboring agent, $\gamma_b < 0$, this effect appears more clearly, since agents undergo directed movement out of clusters. The same dispersal of spatial structure is found for the neighbor-dependent-inhibition of the motility rate, but it occurs in a less efficient way. The authors also highlight that the crowding-induced inhibition of proliferation is capable of counteracting the formation of clusters due to short-range proliferation.

3.2.3 *Pushing and Pulling*

[Yates et al. \(2015\)](#) study how cell–cell pushing affects cells’ dispersal and proliferation by considering on- and off-lattice ABMs. The authors begin by considering a simple ABM incorporating exclusion on a two-dimensional lattice, as in [Simpson et al. \(2007\)](#). They modify the basic model in order to incorporate the ability of cells to push their neighbors ([Ewald et al., 2008; Schmidt and Friedl, 2010](#)). Pushing is implemented according to four different mechanisms with variable degrees of complexity and realism.

Consider an agent at position (i, j) which attempts a movement into an occupied site to its right. In the most basic pushing mechanism, the authors allow the moving agent push the agent at position $(i + 1, j)$ into the site $(i + 2, j)$, with a given probability, H , with the pushing agent taking the pushed agents’ original position, $(i + 1, j)$. Pushing is successful only if the target site of the pushed agent, $(i + 2, j)$, is empty. Otherwise the entire movement is aborted. For the next two mechanisms, the assumption about the location of the target site of the pushed agent is relaxed. For one of these mechanisms the pushed agent is allowed to move into each of its empty neighboring sites with equal probability. If all neighboring sites of the pushed agent are occupied, then the whole event is aborted. In the other mechanism, a target site is chosen uniformly at random from among the pushed agent’s three nearest-neighbors, but if the chosen site is found to be occupied then the entire event is aborted. In the final case considered, the moving agent can attempt to push up to K other agents in a straight line in a chosen direction.

In all the cases considered, the authors derive diffusive PDE approximations for the total agent density. For the basic pushing mechanism, for example, the resulting nonlinear PDE takes the form of Eq. (30) with density-dependent coefficient given by

$$D(\bar{C}) = D(1 + 4H\bar{C}). \quad (35)$$

In most scenarios, the comparison between the ABM and the corresponding PDE is good. However, the more complicated pushing mechanisms, such as the linear pushing of multiple agents, introduce strong spatial correlation in the occupancies of adjacent sites. Since the derivation of the continuum models is based on a mean-field approximation, the presence of such spatial correlation leads to a divergence between the ABM and its deterministic counterpart.

Yates et al. (2015) also investigate the effect of introducing pushing in an off-lattice setup by modifying the one-dimensional model of Dyson et al. (2012) (see Section 2.1). In this model, a pushing event is attempted with probability H when a movement would lead to an overlap of agents, which are represented as interval of length $2R$. The moving agent can displace the pushed adjacent agent far enough for it to complete its movement, but only if this will not produce an overlap with a third agent. In this case the entire movement would be aborted. By using a similar approach to that of Dyson et al. (2012), a nonlinear diffusive PDE of the form (30) is derived with

$$D(\bar{C}) = D(1 + 2R(2 - H)\bar{C}), \quad (36)$$

and the numerical solution of the PDE shows good agreement with the averaged density of the ABM.

For both on- and off-lattice cases, Yates et al. (2015) find that the introduction of pushing behavior leads to density-dependent diffusion coefficients in the macroscopic descriptions. For the on-lattice models, pushing leads to a faster diffusion, due to the larger number of possible movements for each agent in comparison to the simple exclusion model. However, in the off-lattice scenario, increasing the tendency to push leads to a decreasing diffusivity. To explain this behavior, Yates et al. (2015) underline that when a pushing event occurs in the off-lattice model, the moving agent and the pushed agent end up being adjacent, even if they were not in contact prior to the movement. This aggregation effect leads to a slower dispersal of agents corresponding to a lower diffusion in comparison to the nonpushing case.

There is much evidence in the literature to suggest that cells can pull each other while they are undergoing collective migration (Bianco et al., 2007; Ghysen and Dambly-Chaudière, 2007). Chappelle and Yates (2018) explore the effect of allowing cell–cell pulling in both on- and off-lattice models of cell migration.

Firstly, the authors specify a simple on-lattice excluding ABM in which agents are capable of pulling each other. For example, if an agent is moving to the right and another agent is occupying the neighboring site to its left, the moving agent will pull the neighboring agent rightwards, with probability ω .

By writing down the occupancy master equation of the model and taking appropriate limits, the authors obtain a nonlinear diffusion equation for the average agent density, which takes the form of Eq. (30) with

$$D(\bar{C}) = D(1 + 3\omega\bar{C}^2).$$

Pulling produces a quadratic density dependence in the diffusion coefficient of the macroscopic PDE. A comparison with the corresponding model for agents which are able to *push* each other (see Eq. (35)) highlights that the density dependence of diffusivity due to pushing is larger than in the pulling case.

In order to study the effect of pulling in off-lattice models, Chappelle and Yates (2018) use a similar approach to that of Dyson et al. (2012) (see Section 2.1). The one-dimensional model in Dyson et al. (2012) is extended by introducing a *pulling distance* proportional to the moving distance, $l = kd$, such that, if an agent is chosen to move in a given direction and there is another agent whose center is within distance $l + 2R$ in the opposite direction, both agents move in the given direction. The corresponding PDE is of the form of Eq. (30) with

$$D(\bar{C}) = D[1 + 4R\bar{C}(1 - \omega k)],$$

which predicts that the effect of pulling is to decrease the effective diffusivity in the off-lattice model, leading to a slower dispersion at the macroscopic level. The apparent contradiction with the on-lattice counterpart is explained in the paper by using a similar argument to Dyson et al. (2012).

3.3 Growing Domains

Considering cell motility on a purely stationary domain may often be unrealistic for biological processes: cell growth, division, and movement itself will cause the size of the tissue to change dynamically in many biologically plausible situations (Rogulja and Irvine, 2005; Wolpert et al., 2015). Surprisingly, it is only in the last decade that attention has been given to investigating how this phenomenon affects models at multiple scales. Baker et al. (2010) represents a pioneering example of the incorporation of domain growth into models of cell migration.

For the motility scheme, Baker et al. (2010) follow the approach of Othmer and Stevens (1997). At the individual-cell-level they consider a continuous-time, discrete-space ABM. Agents are initialized on a one-dimensional lattice with zero-flux boundary conditions. Agents have the ability to sense the concentration of a signal profile, w , at their current lattice site and at their two nearest-neighbor sites. Agents can jump to immediately neighboring lattice sites. These jumps are regulated by two transition rates that are linear combinations of the signaling molecule concentration at the agent's current and immediately sites. The authors consider four types of transition rate comprising different linear combinations of these concentrations (*local*, *nonlocal*, *average*, and *difference*). If no signal profile exists then agents can either diffuse

randomly (jumping with constant rates independent of their position) or implement density-dependent transition rates. For each case, using a master equation, the authors derive population-level descriptions, which comprise an advection-diffusion PDE for the average agent density, $C(x, t)$, as a function of position, x , and time, t .

The inclusion of domain growth in the model is carried out in two phases. Firstly, [Baker et al. \(2010\)](#) consider the case of exponential growth. On the individual-level this implies that each lattice site divides at a constant rate per unit time. When the site divides, a new daughter site is added adjacent to the parent site. The agents in the parent site are divided between these two sites according to a symmetric distribution and the sites to the right of the daughter site (and their contents) are all shifted one site's width to the right. A PDE description of cell density is derived both from a phenomenological, continuum perspective, using a conservation of matter argument ([Crampin et al., 1999](#)), and from the master equation of the ABM. In order to derive the corresponding PDE from the master equation, the authors use a moment closure approximation on the number of agents in each lattice site. This approximation is based on the assumption that movement occurs on a faster time-scale than domain growth.

Secondly, the authors consider a more general type of domain growth which is density-dependent. As before, upon being chosen to undergo a growth event, a lattice site is divided into two daughter sites, but now with a density-dependent rate, $f(C(x, t))$. A continuous approximation is derived using a further moment closure assumption—that the mean of a nonlinear function of the agent density can be expressed as the same nonlinear function of the mean agent density. The authors prove that domain growth is linear for the case of linear density dependence on domain growth, while it must be evaluated for numerically in the general case.

The paper shows comparisons between the simulations of the individual-based models and the numerical solution of the PDEs. In addition, the predicted macroscopic domain length is compared with the average value from simulations. Results are displayed for the cases of the nongrowing domain, constant growth, linearly density-dependent growth and quadratically density-dependent growth. In each case, the results show a good agreement between the discrete model density and the continuous approximation. However, the predicted value of the domain length in the continuum approximation is underestimated in comparison to the true value from the stochastic model in the case of the nonlinear growth model.

The last part of the paper incorporates a signaling profile which agents can sense and respond to. Specifically, an exponentially decreasing concentration of signaling molecule, w , is placed in the domain and agents are assumed to interact with it, both via the local and nonlocal schemes. Results show good agreement between the simulation and the deterministic prediction for both linear and logistic domain growth. As a last application, [Baker et al. \(2010\)](#)

consider linearly density-dependent growth with both local and nonlocal sensing again demonstrating good agreement between the two modeling regimes.

Yates et al. (2012) generalize the existing ABMs of Baker et al. (2010) to a nonuniform lattice allowing the process of cell division on the underlying domain to be modeled in a more realistic manner. The domain is allowed to extend in a manner which can be made arbitrarily close to continuous, as opposed to the discrete increments by which the domain is extended in Baker et al. (2010).

The authors consider two distinct ways to discretize a one-dimensional domain in a nonuniform way: the *Voronoi partition method* and the *interval-centered method*. In the first case, the agent positions, $c(n, t)$ for $n = 1, \dots, \mathcal{N}$, are chosen first and the edges of the intervals are the bisectors of these points. In the second case, the edges of the intervals are specified first and the agents lie at the center of these intervals. As in Baker et al. (2010), no exclusion property is implemented and multiple agents can occupy the same lattice site.

In the first part of the paper the domain is fixed and agents move following a position-jump process on the irregular lattice. The transition rates of an agent in interval i are computed from the mean first passage times that a Brownian particle, initialized at x_i , takes to hit one of the neighboring particle positions (x_{i-1} or x_{i+1}). Using the master equation for agent densities, Yates et al. (2012) derive a macroscale description of the model in the diffusive limit. For the Voronoi partition, the density evolves according to the diffusion equation and the comparison between the simulations and the PDE shows a good agreement. In the case of the interval-centered partition, the agreement between the individual-level behavior and the population-level diffusion equation is poor since the transition rates are not inherently linked to the size of the intervals.

The remainder of the paper focuses on the introduction of domain growth to models of agent migration using the Voronoi partition. Yates et al. (2012) start by defining a deterministic scheme for domain growth. Every time an agent moves, each lattice interval grows an amount proportional to its length and proportional to the time step between movement events. When an interval reaches a threshold length, it splits into two daughter intervals. The boundaries of such daughter intervals are chosen in order to preserve the Voronoi property of the domain partition. Each agent is redistributed to a new interval with probability proportional to the overlap between the new intervals and the old intervals. The authors also consider an alternative growth mechanism in which, with a give rate, an interval is chosen at random to grow with probability proportionally to its length. When a *growth event* occurs, both the selected interval and an adjacent interval grow in order to preserve the Voronoi property. For both growth schemes, a population-level description can be obtained through the master equation for the densities. The resulting PDEs are of the same form as in Baker et al. (2010).

Finally, a series of comparisons between simulations and PDEs confirm the good agreement between the microscopic and macroscopic models for the Voronoi domain partition, but a poorer agreement for the interval-centered domain partition, as expected.

In addition to their investigations into cell–cell adhesion and volume exclusion on a static domain (see Section 3.2.2) Thompson et al. (2012) study the interplay of the same properties on a growing domain using a modification of the ABMs of Baker et al. (2010) and Yates et al. (2012). The authors incorporate a partially excluding property into the model of domain growth of Baker et al. (2010) by allowing the carrying capacity of each compartment to be proportional to its size (see Fig. 4 for a schematic illustration). When a growth event occurs, a compartment is chosen at random and its carrying capacity is increased by unity. Commensurately, its size it is also increased. When a compartment reaches a predefined threshold, it is split into two compartments each with carrying capacity set to half of the previous value. To compensate for the unequal compartment sizes generated by growth, the transition rates are amended as in Yates et al. (2012). A population-level description of the model is derived, but it is only valid for short time-scales, specifically until the first splitting event occurs.

The results demonstrate that domain growth decreases the chance of cell clustering occurring by cell–cell interactions. For high values of adhesion and with small growth rates, clusters can still appear over short time-sales. However, as the growth rate increases, the probability of clusters appearing decreases and all initial cell clusters are eventually destroyed.

Hywood et al. (2013) suggest a modified version of the model of Baker et al. (2010) to represent tissue growth. The authors initialize a one-dimensional lattice with a set of contiguous nonoverlapping agents (*tracers* agents). These agents are inactive (i.e., they do not perform any jumping movement between sites) and their role is only to mark the position of the site in which they are located as the underlying lattice grows. The authors focus on the case of exponential growth with constant rate, b , which is implemented similarly to Baker et al. (2010). If a marked site splits into two daughter sites, the corresponding tracer agent moves to the right daughter site (see Fig. 12 for an illustration). The usage of tracer agents follows from the previous work of Binder and Landman (2009) in which, contrastingly, domain growth is implemented in a deterministic manner.

By coupling the position of the tracers to a system of noninteracting random walkers and writing down the corresponding occupancy master equation for the tracers, Hywood et al. (2013) derive a formula for the infinitesimal mean and variance of the underlying stochastic process, $\mu(x, t)$ and $\sigma^2(x, t)$, respectively. These expressions can then be used as coefficients for an advection-diffusion PDE (Fokker–Planck equation (FPE)) describing the spatio-temporal evolution of the average occupancy of the tracer agents, $\bar{C}(x, t)$:

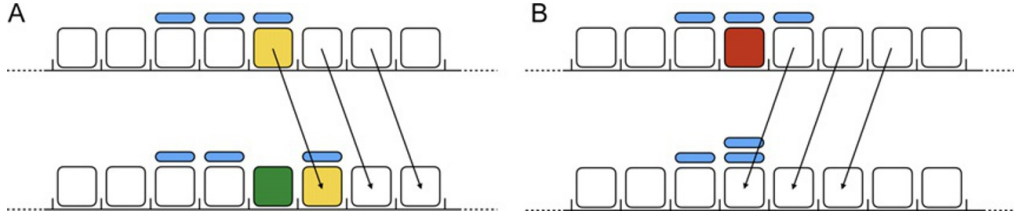


FIG. 12 Schematics of the ABM of Yates (2014). White squares represent empty lattice spaces and blue markers represent the tracer agents. Panel (A) shows an example of site division in which the dividing site (yellow) is also marked by a tracer agent. The tracer agent moves with the dividing agent to the right and a new site (green) is added in the vacant space. Panel (B) shows an example of site death. A marked site (red) is removed from the lattice, its tracer agent remains in its position and causes an overlap of two tracer agents.

$$\frac{\partial \bar{C}(x, t)}{\partial t} = \frac{\partial^2}{\partial x^2} \left[\frac{\sigma^2(x, t)}{2} \bar{C}(x, t) \right] - \frac{\partial}{\partial x} [\mu(x, t) \bar{C}(x, t)]. \quad (37)$$

Yates (2014) extends the work of Hywood et al. (2013) to more general scenarios in which the rate at which sites divide is time-dependent, $b(t)$. This extension allows the incorporation of a variety of biologically realistic mechanisms of domain growth. By using similar steps, Yates (2014) obtains a series of PDEs for the average occupancy of the tracer agents of the same form as Eq. (37). In particular, the author provides the expressions of $\mu(x, t)$ and $\sigma^2(x, t)$ for the case of linear growth, $\frac{dL}{dt} = r$, generalized logistic growth, $\frac{dL}{dt} = rL \left[1 - \left(\frac{L}{R} \right)^v \right]$, and Gompertzian growth, $\frac{dL}{dt} = rL \left(\ln \left(\frac{L}{R} \right) \right)$.

Finally, Yates (2014) studies the implications of implementing site death in the model which can be done by removing sites with a given rate, $d(t)$. When this occurs, all the site to the right of the removed site are moved to the left in order to fill the vacant space created by the site's removal. If a tracer agent occupied the removed site, it remains in its position. Notice that this new mechanism can lead to multiple tracer agents occupying the same site (see Fig. 12 for an illustration). The comparison between the ABM simulations and the corresponding PDE preserves a good agreement when site death is introduced, even for cases of net domain shrinkage, i.e., $b(t) < d(t)$.

The results of Yates (2014) highlight the danger of neglecting death events in the model, even when the net growth rate is positive, $b(t) - d(t) > 0$. In other words, although the mean growth rate can be estimated correctly by using purely growing dynamics (with growth rate given by $\tilde{b}(t) = b(t) - d(t)$), the second and higher moments of the process will be incorrect. For example, the higher value of variance leads to faster diffusivity of the agents when site death is explicitly incorporated in the model.

3.4 Persistence of Motion

A fundamental assumption of modeling cell movement with a position-jump process (such as those implemented in Section 2) is that the cells' positions are subject to a series of Markovian jumps in space and, therefore, the direction of motion between sequential jumps is totally uncorrelated. In reality, experimental observations suggest that many types of cell show a tendency to preserve their direction of motion for some time before reorienting (Berg and Brown, 1972; Gail and Boone, 1970; Hall, 1977; Wright et al., 2008), even when the movement remains unbiased in the long term. This behavior is known as *persistence of motion* or *of direction* (Patlak, 1953). The standard framework to model persistence of motion in an ABM based is to employ a *velocity-jump* process (Campos et al., 2010; Codling et al., 2008; Othmer and Hillen, 2000, 2002; Othmer et al., 1988) in which the variable which is performing Markovian jumps is the velocity rather than the position of the agents (Othmer et al., 1988).

A simple example of an velocity-jump process ABM in one dimension was analyzed by Goldstein (1951) and later by Kac (1974) and Othmer et al. (1988). The model can be formulated as follows. Consider an ABM in which agents have an assigned direction of motion, either right or left. As time evolves they can either move, with rate α , or change their direction, with rate λ . The two events are assumed to occur independently. When a movement event takes place, the moving agent performs a jump of distance, v , in its assigned direction. We can interpret $\pm v\alpha$ as the two possible values of the velocity of an agent, depending whether it is moving in the right- or left- direction, respectively. Notice that this formulation is valid for both on- and off-lattice models. However, in an on-lattice framework, the distance, v , has to be an integer multiple of the lattice step Δ .

Let $\bar{R}(x, t)$ and $\bar{L}(x, t)$ denote the average occupancy of agents in position x at time t with associated direction to the right and left, respectively. For a system of noninteracting agents, the corresponding occupancy master equations can be written as

$$\begin{cases} \bar{R}(x, t + \delta t) = \bar{R}(x, t) + \alpha \delta t [\bar{R}(x - v, t) - \bar{R}(x, t)] + \lambda \delta t (\bar{L} - \bar{R}), \\ \bar{L}(x, t + \delta t) = \bar{L}(x, t) + \alpha \delta t [\bar{L}(x + v, t) - \bar{L}(x, t)] + \lambda \delta t (\bar{R} - \bar{L}). \end{cases} \quad (38a,b)$$

By Taylor expanding the two terms $\bar{R}(x - v, t)$ and $\bar{L}(x + v, t)$ about x to first order and taking the limit $\delta t, v \rightarrow 0$ with $v\alpha$ kept constant, one obtains the system of advective PDEs given by

$$\begin{cases} \frac{\partial \bar{R}}{\partial t} = -V \frac{\partial \bar{R}}{\partial x} + \lambda (\bar{L} - \bar{R}), \\ \frac{\partial \bar{L}}{\partial t} = V \frac{\partial \bar{L}}{\partial x} + \lambda (\bar{R} - \bar{L}), \end{cases} \quad (39a,b)$$

where $V = \lim_{v \rightarrow 0} v\alpha$. By adding Eqs. (39) and differentiating with respect to t we can write

$$\frac{\partial^2(\bar{R} + \bar{L})}{\partial t^2} = V \frac{\partial^2(\bar{L} - \bar{R})}{\partial t \partial x}. \quad (40)$$

Similarly, by subtracting Eq. (39a) from Eq. (39b) and differentiating with respect to x , we obtain

$$\frac{\partial^2(\bar{L} - \bar{R})}{\partial x \partial t} = V \frac{\partial^2(\bar{R} + \bar{L})}{\partial x^2} - 2\lambda \frac{\partial(\bar{L} - \bar{R})}{\partial x}. \quad (41)$$

Finally, substituting Eq. (40) into Eq. (41) and recalling Eqs. (39) we can write

$$\frac{\partial^2 \bar{C}}{\partial t^2} + 2\lambda \frac{\partial \bar{C}}{\partial t} = V^2 \frac{\partial^2 \bar{C}}{\partial x^2}, \quad (42)$$

where $\bar{C} = \bar{R} + \bar{L}$ represents the total average occupancy. Eq. (42) is also known as a *telegraph equation* since it was originally derived to describe the propagation of signal waves traveling through a telegraph transmission wire (Goldstein, 1951). Othmer et al. (1988) were the first to obtain such equations from a system of noninteracting agents. Othmer and Hillen (2000) demonstrated that it is possible to recover the canonical diffusion equation as the parabolic limit of the telegraph equation by taking the limit as V and λ to infinity simultaneously, such that V^2/λ is constant. In other words, the two ABMs with and without persistence are governed by continuum models of the same form in the long term. This can be understood by noticing that the short-term directional bias due to the presence of persistence becomes less evident at the spatial scale which are much larger than the average distance moved by an agent before reorienting and temporal scales which are much larger than the average reorientation time (Codling et al., 2008; Othmer and Hillen, 2000, 2002).

Notice that the derivation of the Eq. (42) from Eqs. (39a) and (39b) is possible only for the case of noninteracting agents in one dimension. When the model is defined on a two-dimensional lattice, for example, we can still write down a system of four advective PDEs for the average occupancies of agents moving in the four direction of the lattice (Gavagnin and Yates, 2018). However, a closed formula for the total average occupancy is no longer obtainable.

A similar problem occurs when agent-agent interactions are added to the model, in which case deriving an analog of the telegraph equation and, consequently, its diffusive limit, is not possible. Treloar et al. (2011) have studied the implication of incorporating crowding effects in a model of persistence. At the stochastic-level, they consider a modification of the ABM of Othmer et al. (1988) on a one-dimensional lattice with three different volume exclusion properties each of increasing complexity. For each case, the authors derive a system of advective PDEs describing the evolution of the average occupancy of the two subpopulations of agents depending on their direction of movement.

The results of Treloar et al. (2011) show that the details of the crowding interactions lead to differences in the corresponding continuum models.

This highlights a substantial difference from the analogous persistence-free position-jump process (see Section 2.1), in which crowding effects do not change the corresponding macroscopic representation.

The continuum models derived by Treloar et al. (2011) do not admit a diffusive interpretation. A diffusive interpretation is desirable, since it allows a direct comparison with other diffusive models and establishes a direct connection with commonly used statistical tools of movement analysis. Recently, Gavagnin and Yates (2018) studied a generalization of the ABM of Treloar et al. (2011) in two dimensions for which is possible to obtain an accurate diffusive description at the population-level. The authors suggest the introduction of an additional parameter, $\varphi \in [0, 1]$, in the ABMs of Treloar et al. (2011) which modulates the intensity of the short-term directional bias and hence the intensity of the persistence.

By rescaling the parameter with the size of the lattice step, $\varphi \sim \mathcal{O}(\Delta)$, Gavagnin and Yates (2018) obtain a general set of advective-diffusive PDEs. A comparison of the column-averaged density shows a good agreement between the stochastic and the deterministic models for a wide range of parameters. However, when the agent jump length becomes large, the density profile of the ABM presents regular peaks in density which are not captured by the corresponding continuum models.

For strong values of persistence, Gavagnin and Yates (2018) find evidence of a spontaneous form of agent aggregation as result of the interplay of persistence and crowding effects in highly populated regions (see Fig. 13). Notice that such a form of agent aggregation is not possible in ABMs which do not incorporate persistence, since the overall behavior is governed by simple diffusion (Simpson et al., 2009). Their deterministic model is capable of providing a qualitatively matching description of such aggregation

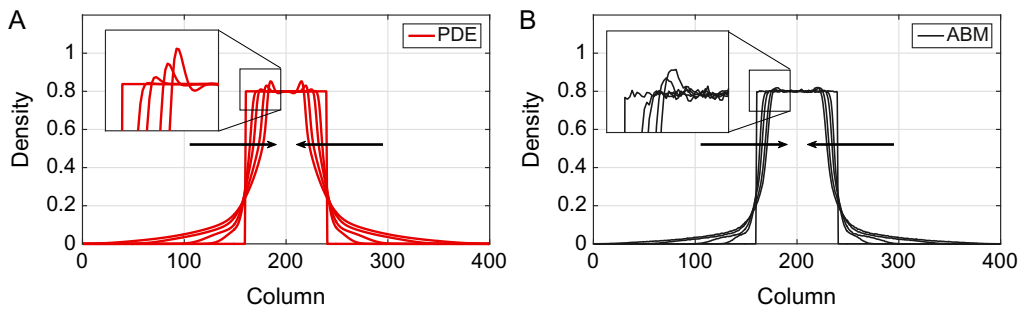


FIG. 13 Spontaneous aggregation induced by persistence and volume exclusion. The top panels show (A) the numerical solution of the PDE for the column-averaged total density and (B) column-averaged density of the ABM, averaged over $M = 100$ repeats. The profiles are shown at time $T = 0, 50, 100, 200, 300$ with the direction of the black arrows indicating increasing time. Reproduced from Gavagnin, E., Yates, C.A., 2018. Modeling persistence of motion in a crowded environment: the diffusive limit of excluding velocity-jump processes. *Phys. Rev. E* 97, 032416 with permission of Physical Review E.

phenomenon, even though the presence of short-range correlations affects the quality of the agreement (see Fig. 13).

Finally, the two-dimensional framework of Gavagnin and Yates (2018) highlights a positive anisotropy of the model in the axial directions of the lattice. This anisotropy appears as an intrinsic feature of the persistent model combined with the lattice environment. Off-lattice models should have the advantage that they are not afflicted by such anisotropy. However, the derivation of a corresponding macroscopic description becomes more complicated and sometimes intractable.

4 CONCLUSION

Cell migration is essential in a wide range of biological contexts, including many developmental and homeostatic mechanisms in the human body (Deng et al., 2006; Gilbert, 2003; Keller, 2005; Maini et al., 2004). As a result, a great deal of attention has been given to the study of such a complex phenomenon in the last few decades. However, when an experimental approach becomes difficult or even impossible, mathematical modeling provides an avenue through which investigations can continue. The power of modeling relies on its ability to test and verify experimental hypotheses, as well as to make predictions which can indicate appropriate experimental directions.

In general there is a dichotomy between the types of models used for the mathematical representation of cell migration. On one side of the divide are continuum models. These models are capable of representing the population-level characteristics of a group of cells and are often amenable to mathematical analysis. When required, such models are fast to simulate numerically and can often be linked explicitly to the model parameters. However, these models are less useful for capturing individual-level detail and are consequently more difficult to link directly to experimental data; abilities which are inherent to the discrete models, on the other side of the divide. Discrete models are often intuitive to formulate and are able to incorporate stochasticity in a natural manner. However, it is often difficult to link the results of such models directly to the model parameters in order to gain a population-level overview and their simulation can be computationally expensive. Building equivalence frameworks allows the exploitation of the two modeling paradigms' complementary strengths and the circumvention of their complementary weaknesses.

In the first part of this review we have outlined a class of methodologies designed to bridge the divide between the individual-level and population-level regimes. As fundamental examples, we have considered a series of simple ABMs representing cell migrating on a one-dimensional domain. We have recalled explicitly how the canonical diffusion equation (6) can be derived from these models when agents are not interacting with each other. The same equation is obtained for both on- and off-lattice systems. By recapitulating the computation for ABMs which incorporate crowding effects, we have underlined a

substantial difference between on-lattice ABMs and off-lattice ABMs. Although the on-lattice models for diffusion give the same PDE irrespective of whether volume exclusion is implemented or not, the same cannot be said in almost any other case, for example the off-lattice ABM that we considered.

In the second part of this chapter, we have provided a brief review of some of the important features which can influence cell behavior at the population level, either by suppressing or enhancing the collective movement. For each case, we have summarized the relevant studies devoted to incorporating such specific features into mathematical models both in the stochastic and deterministic regimes.

In conclusion, although significant progress has been made toward the understanding of cell migratory behaviors through mathematical modeling, many questions concerning the direct relationship between single-cell behaviors and collective invasion remain unclear. For example, there is still little understanding of the role of heterogeneity in cell behavior, such as leader–follower mechanisms (McLennan et al., 2012, 2015; Schumacher et al., 2017) or heterogeneous cell proliferation (Smadbeck and Stumpf, 2016). In this context, developing mathematical models capable of incorporating both these aspects represents an important avenue for future research.

REFERENCES

- Abercrombie, M., 1979. Contact inhibition and malignancy. *Nature* 281 (5729), 259.
- Anguige, K., Schmeiser, C., 2009. A one-dimensional model of cell diffusion and aggregation, incorporating volume filling and cell-to-cell adhesion. *J. Math. Biol.* 58 (3), 395–427.
- Armstrong, N.J., Painter, K.J., Sherratt, J.A., 2006. A continuum approach to modelling cell-cell adhesion. *J. Theor. Biol.* 243 (1), 98–113.
- Baker, R.E., Simpson, M.J., 2010. Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E* 82 (4), 041905.
- Baker, R.E., Yates, C.A., Erban, R., 2010. From microscopic to macroscopic descriptions of cell migration on growing domains. *Bull. Math. Biol.* 72 (3), 719–762.
- Ben-Jacob, E., Cohen, I., Levine, H., 2000. Cooperative self-organization of microorganisms. *Adv. Phys.* 49 (4), 395–554.
- Berg, H.C., Brown, D.A., 1972. Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* 239 (5374), 500–504.
- Bianco, A., Poukkula, M., Cliffe, A., Mathieu, J., Luque, C.M., Fulga, T.A., Rørth, P., 2007. Two distinct modes of guidance signalling during collective migration of border cells. *Nature* 448 (7151), 362–365.
- Binder, B.J., Landman, K.A., 2009. Exclusion processes on a growing domain. *J. Theor. Biol.* 259 (3), 541–551.
- Binny, R.N., Plank, M.J., James, A., 2015. Spatial moment dynamics for collective cell movement incorporating a neighbour-dependent directional bias. *J. R. Soc. Interface* 12 (106), 20150228.
- Binny, R.N., James, A., Plank, M.J., 2016. Collective cell behaviour with neighbour-dependent proliferation, death and directional bias. *Bull. Math. Biol.* 78 (11), 2277–2301.
- Cai, A.Q., Landman, K.A.A., Hughes, B.D., 2006. Modelling directional guidance and motility regulation in cell migration. *Bull. Math. Biol.* 68 (1), 25.

- Campos, D., Méndez, V., Llopis, I., 2010. Persistent random motion: uncovering cell migration dynamics. *J. Theor. Biol.* 267 (4), 526–534.
- Carmona-Fontaine, C., Matthews, H.K., Kuriyama, S., Moreno, M., Dunn, G.A., Parsons, M., Stern, C.D., Mayor, R., 2008. Contact inhibition of locomotion in vivo controls neural crest directional migration. *Nature* 456 (7224), 957.
- Chappelle, G., Yates, C.A., 2018. Pulling in models of cell migration. <http://arxiv.org/abs/1807.11852>.
- Cheeseman, B.L., Newgreen, D.F., Landman, K.A., 2014. Spatial and temporal dynamics of cell generations within an invasion wave: a link to cell lineage tracing. *J. Theor. Biol.* 363, 344–356.
- Codling, E.A., Plank, M.J., Benhamou, S., 2008. Random walk models in biology. *J. R. Soc. Interface* 5 (25), 813–834.
- Crampin, E.J., Gaffney, E.A., Maini, P.K., 1999. Reaction and diffusion on growing domains: scenarios for robust pattern formation. *Bull. Math. Biol.* 61 (6), 1093–1120.
- Crank, J., 1979. *The Mathematics of Diffusion*. Oxford University Press.
- Deng, M., Chen, W.L., Takatori, A., Peng, Z., Zhang, L., Mongan, M., Parthasarathy, R., Sartor, M., Miller, M., Yang, J., et al., 2006. A role for the mitogen-activated protein kinase kinase 1 in epithelial wound healing. *Mol. Biol. Cell* 17 (8), 3446–3455.
- Deutsch, A., Dormann, S., 2007. *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Applications, and Analysis*. Springer Science & Business Media.
- Dickinson, R.B., Tranquillo, R.T., 1993. A stochastic model for adhesion-mediated cell random motility and haptotaxis. *J. Math. Biol.* 31 (6), 563–600.
- Druckebrod, N.R., Epstein, M.L., 2007. Behavior of enteric neural crest-derived cells varies with respect to the migratory wavefront. *Dev. Dyn.* 236 (1), 84–92.
- Dworkin, M., Kaiser, D., 1985. Cell interactions in myxobacterial growth and development. *Science* 230 (4721), 18–24.
- Dyson, L., Baker, R.E., 2014. The importance of volume exclusion in modelling cellular migration. *J. Math. Biol.* 71 (3), 691–711.
- Dyson, L., Maini, P.K., Baker, R.E., 2012. Macroscopic limits of individual-based models for motile cell populations with volume exclusion. *Phys. Rev. E* 86 (3), 031903.
- Erban, R., Othmer, H.G., 2004. From individual to collective behavior in bacterial chemotaxis. *SIAM J. Appl. Math.* 65 (2), 361–391.
- Ewald, A.J., Brenot, A., Duong, M., Chan, B.S., Werb, Z., 2008. Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. *Dev. Cell* 14 (4), 570–581.
- Fisher, R.A., 1937. The wave of advance of advantageous genes. *Ann. Hum. Genet.* 7 (4), 355–369.
- Flaherty, B., McGarry, J.P., McHugh, P.E., 2007. Mathematical models of cell motility. *Cell Biochem. Biophys.* 49 (1), 14–28.
- Friedl, P., Gilmour, D., 2009. Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.* 10 (7), 445.
- Gail, M.H., Boone, C.W., 1970. The locomotion of mouse fibroblasts in tissue culture. *Biophys. J.* 10 (10), 980–993.
- Gavagnin, E., Yates, C.A., 2018. Modeling persistence of motion in a crowded environment: the diffusive limit of excluding velocity-jump processes. *Phys. Rev. E* 97, 032416.
- Ghyssen, A., Dambly-Chaudière, C., 2007. The lateral line microcosmos. *Genes Dev.* 21 (17), 2118–2130.
- Gilbert, S.F., 2003. The morphogenesis of evolutionary developmental biology. *Int. J. Dev. Biol.* 47 (7–8), 467.

- Gillespie, D., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81 (25), 2340–2361.
- Goldstein, S., 1951. On diffusion by discontinuous movements, and on the telegraph equation. *Q. J. Mech. Appl. Math.* 4 (2), 129–156.
- Golubev, A., 2016. Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression. *J. Theor. Biol.* 393, 203–217.
- Graner, F., Glazier, J.A., 1992. Simulation of biological cell sorting using a two-dimensional extended potts model. *Phys. Rev. Lett.* 69 (13), 2093–2096.
- Hall, R.L., 1977. Amoeboid movement as a correlated walk. *J. Math. Biol.* 4 (4), 327–335.
- Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. *Cell* 100 (1), 57–70.
- Hywood, J.D., Hackett-Jones, E.J., Landman, K.A., 2013. Modeling biological tissue growth: discrete to continuum representations. *Phys. Rev. E* 88 (3), 032704.
- Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons.
- Kac, M., 1974. A stochastic model related to the telegrapher's equation. *Rocky Mountain J. Math.* 4 (3), 497–509.
- Keller, R., 2005. Cell migration during gastrulation. *Curr. Opin. Cell Biol.* 17 (5), 533–541.
- Keynes, R.J., Cook, G.M.W., 1992. Repellent cues in axon guidance. *Curr. Opin. Neurobiol.* 2 (1), 55–59.
- Kurhekar, M.P., Deshpande, U.A., 2015. Agent based deterministic model of the adult subventricular neurogenesis. *Nano Comm. Netw.* 6 (3), 124–132.
- Lee, P., Wolgemuth, C.W., 2011. Crawling cells can close wounds without purse strings or signaling. *PLOS Comput. Biol.* 7 (3), e1002007.
- Lorenz, M.O., 1905. Methods of measuring the concentration of wealth. *Publ. Am. Stat. Assoc.* 9 (70), 209–219.
- Madri, J.A., Graesser, D., 2000. Cell migration in the immune system: the evolving inter-related roles of adhesion molecules and proteinases. *J. Immunol. Res.* 7 (2–4), 103–116.
- Maini, P.K., McElwain, D.L.S., Leavesley, D.I., 2004. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.* 10 (3–4), 475–482.
- Markham, D.C., Simpson, M.J., Baker, R.E., 2013a. Simplified method for including spatial correlations in mean-field approximations. *Phys. Rev. E* 87 (6), 062702.
- Markham, D.C., Simpson, M.J., Maini, P.K., Gaffney, E.A., Baker, R.E., 2013b. Incorporating spatial correlations into multispecies mean-field models. *Phys. Rev. E* 88 (5), 052713.
- McLennan, R., Dyson, L., Prather, K.W., Morrison, J.A., Baker, R.E., Maini, P.K., Kulesa, P.M., 2012. Multiscale mechanisms of cell migration during development: theory and experiment. *Development* 139 (16), 2935–2944.
- McLennan, R., Schumacher, L.J., Morrison, J.A., Teddy, J.M., Ridenour, D.A., Box, A.C., Semerad, C.L., Li, H., McDowell, W., Kay, D., et al., 2015. VEGF signals induce trailblazer cell identity that drives neural crest migration. *Dev. Biol.* 407 (1), 12–25.
- Mort, R.L., Ross, R.J.H., Hailey, K.J., Harrison, O.J., Keighren, M.A., Landini, G., Baker, R.E., Painter, K.J., Jackson, I.J., Yates, C.A., 2016. Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model. *Nat. Commun.* 7, 10288.
- Murray, J.D., 2007. *Mathematical Biology: I. An Introduction*. vol. 17. Springer Science & Business Media.
- Murrell, D.J., Dieckmann, U., Law, R., 2004. On moment closures for population dynamics in continuous space. *J. Theor. Biol.* 229 (3), 421–432.

- Niessen, C.M., 2007. Tight junctions/adherens junctions: basic structure and function. *J. Invest. Dermatol.* 127 (11), 2525–2532.
- Noble, D., 2002. Opinion: the rise of computational biology. *Nat. Rev. Mol. Cell Biol.* 3 (6), 459.
- Othmer, H.G., Hillen, T., 2000. The diffusion limit of transport equations derived from velocity-jump processes. *SIAM J. Appl. Math.* 61 (3), 751–775.
- Othmer, H.G., Hillen, T., 2002. The diffusion limit of transport equations II: chemotaxis equations. *SIAM J. Appl. Math.* 62 (4), 1222–1250.
- Othmer, H.G., Stevens, A., 1997. Aggregation, blowup, and collapse: the ABC's of taxis in reinforced random walks. *SIAM J. Appl. Math.* 57 (4), 1044–1081.
- Othmer, H.G., Dunbar, S.R., Alt, W., 1988. Models of dispersal in biological systems. *J. Math. Biol.* 26 (3), 263–298.
- Painter, K.J., Hillen, T., 2002. Volume-filling and quorum-sensing in models for chemosensitive movement. *Can. Appl. Math. Q.* 10 (4), 501–543.
- Painter, K.J., Maini, P.K., Othmer, H.G., 1999. Stripe formation in juvenile pomacanthus explained by a generalized turing mechanism with chemotaxis. *Proc. Natl. Acad. Sci. USA* 96 (10), 5549–5554.
- Patlak, C.S., 1953. Random walk with persistence and external bias. *Bull. Math. Biophys.* 15 (3), 311–338.
- Plank, M.J., Simpson, M.J., 2012. Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches. *J. R. Soc. Interface* 9 (76), 2983–2996.
- Plank, M.J., Simpson, M.J., 2013. Lattice-free models of cell invasion: discrete simulations and travelling waves. *Bull. Math. Biol.* 75 (11), 2150–2166.
- Plank, M.J., Sleeman, B.D., 2004. Lattice and non-lattice models of tumour angiogenesis. *Bull. Math. Biol.* 66 (6), 1785–1819.
- Raghib, M., Hill, N.A., Dieckmann, U., 2011. A multiscale maximum entropy moment closure for locally regulated space-time point process models of population dynamics. *J. Math. Biol.* 62 (5), 605–653.
- Raines, E.W., 2000. The extracellular matrix can regulate vascular cell migration, proliferation, and survival: relationships to vascular disease. *Int. J. Exp. Pathol.* 81 (3), 173–182.
- Ridley, A.J., Schwartz, M.A., Burridge, K., Firtel, R.A., Ginsberg, M.H., Borisy, G., Parsons, J.T., Horwitz, A.R., 2003. Cell migration: integrating signals from front to back. *Science* 302 (5651), 1704–1709.
- Rogulja, D., Irvine, K.D., 2005. Regulation of cell proliferation by a morphogen gradient. *Cell* 123 (3), 449–461.
- Schienenbein, M., Gruler, H., 1993. Langevin equation, Fokker-Planck equation and cell migration. *Bull. Math. Biol.* 55 (3), 585–608.
- Schmidt, S., Friedl, P., 2010. Interstitial cell migration: integrin-dependent and alternative adhesion mechanisms. *Cell Tissue Res.* 339 (1), 83.
- Schumacher, L.J., Maini, P.K., Baker, R.E., 2017. Semblance of heterogeneity in collective cell migration. *Cell Syst.* 5 (2), 119–127.
- Shapiro, J.A., 1988. Bacteria as multicellular organisms. *Sci. Am.* 258 (6), 82–89.
- Sherratt, J.A., Chaplain, M.A.J., 2001. A new mathematical model for avascular tumour growth. *J. Math. Biol.* 43 (4), 291–312.
- Simpson, M.J., Landman, K.A., Hughes, B.D., Newgreen, D.F., 2006. Looking inside an invasion wave of cells using continuum models: proliferation is the key. *J. Theor. Biol.* 243 (3), 343–360.
- Simpson, M.J., Merrifield, A., Landman, K.A., Hughes, B.D., 2007. Simulating invasion with cellular automata: connecting cell-scale and population-scale properties. *Phys. Rev. E* 76 (2), 021918.

- Simpson, M.J., Landman, K.A., Hughes, B.D., 2009. Multi-species simple exclusion processes. *Physica A* 388 (4), 399–406.
- Simpson, M.J., Landman, K.A., Hughes, B.D., 2010. Cell invasion with proliferation mechanisms motivated by time-lapse data. *Physica A* 389 (18), 3779–3790.
- Smadbeck, P., Stumpf, M.P.H., 2016. Coalescent models for developmental biology and the spatio-temporal dynamics of growing tissues. *J. R. Soc. Interface* 13 (117), 20160112.
- Staton, C.A., Stribbling, S.M., Tazzyman, S., Hughes, R., Brown, N.J., Lewis, C.E., 2004. Current methods for assaying angiogenesis in vitro and in vivo. *Int. J. Exp. Pathol.* 85 (5), 233–248.
- Stokes, C.L., Lauffenburger, D.A., 1991. Analysis of the roles of microvessel endothelial cell random motility and chemotaxis in angiogenesis. *J. Theor. Biol.* 152 (3), 377–403.
- Sun, X., Ward, M.J., 2000. Dynamics and coarsening of interfaces for the viscous Cahn–Hilliard equation in one spatial dimension. *Stud. Appl. Math.* 105 (3), 203–234.
- Tambe, D.T., Hardin, C.C., Angelini, T.E., Rajendran, K., Park, C.Y., Serra-Picamal, X., Zhou, E.H., Zaman, M.H., Butler, J.P., Weitz, D.A., et al., 2011. Collective cell guidance by cooperative intercellular forces. *Nat. mater.* 10 (6), 469.
- Taylor, P.R., Baker, R.E., Yates, C., 2015a. Deriving appropriate boundary conditions, and accelerating position-jump simulations, of diffusion using non-local jumping. *Phys. Biol.* 12, 016006.
- Taylor, P.R., Yates, C.A., Simpson, M.J., Baker, R.E., 2015b. Reconciling transport models across scales: the role of volume exclusion. *Phys. Rev. E* 92, 040701.
- Taylor, P.R., Baker, R.E., Simpson, M.J., Yates, C.A., 2016. Coupling volume-excluding compartment-based models of diffusion at different scales: Voronoi and pseudo-compartment approaches. *J. R. Soc. Interface* 13 (120), 20160336. ISSN 1742-5689.
- Thompson, A.G., Tailleur, J., Cates, M.E., Blythe, R.A., 2011. Lattice models of nonequilibrium bacterial dynamics. *J. Stat. Mech.* 2011 (02), P02029.
- Thompson, R.N., Yates, C.A., Baker, R.E., 2012. Modelling cell migration and adhesion during development. *Bull. Math. Biol.* 74 (12), 2793–2809.
- Tomlin, C.J., Axelrod, J.D., 2007. Biology by numbers: mathematical modelling in developmental biology. *Nat. Rev. Genet.* 8 (5), 331.
- Treloar, K.K., Simpson, M.J., McCue, S.W., 2011. Velocity-jump models with crowding effects. *Phys. Rev. E* 84 (6), 061920.
- Treloar, K.K., Simpson, M.J., McCue, S.W., 2013. Velocity-jump processes with proliferation. *J. Phys. A: Math. Theor.* 46 (1), 015003.
- Trepat, X., Wasserman, M.R., Angelini, T.E., Millet, E., Weitz, D.A., Butler, J.P., Fredberg, J.J., 2009. Physical forces during collective cell migration. *Nat. Phys.* 5 (6), 426.
- Turner, S., Sherratt, J.A., 2002. Intercellular adhesion and cancer invasion: a discrete simulation using the extended potts model. *J. Theor. Biol.* 216 (1), 85–100.
- Turner, S., Sherratt, J.A., Painter, K.J., Savill, N.J., 2004. From a discrete to a continuous model of biological cell movement. *Phys. Rev. E* 69 (2), 021910.
- Turner, C., Stinchcombe, A.R., Kohandel, M., Singh, S., Sivaloganathan, S., 2009. Characterization of brain cancer stem cells: a mathematical approach. *Cell Prolif.* 42 (4), 529–540.
- Walpole, J., Papin, J.A., Peirce, S.M., 2013. Multiscale computational models of complex biological systems. *Annu. Rev. Biomed. Eng.* 15, 137–154.
- Ward, M., McCann, C., DeWulf, M., Wu, J.Y., Rao, Y., 2003. Distinguishing between directional guidance and motility regulation in neuronal migration. *J. Neurosci.* 23 (12), 5170–5177.
- Welty, J.R., Wicks, C.E., Rorrer, G., Wilson, R.E., 2009. *Fundamentals of Momentum, Heat, and Mass Transfer*. John Wiley & Sons.

- Westermann, J., Söllner, S., Ehlers, E.M., Nohroudi, K., Blessenohl, M., Kalies, K., 2003. Analyzing the migration of labeled T cells in vivo: an essential approach with challenging features. *Lab. Invest.* 83 (4), 459.
- Wolpert, L., Tickle, C., Arias, A.M., 2015. *Principles of Development*. Oxford University Press, USA.
- Wright, A., Li, Y.H., Zhu, C., 2008. The differential effect of endothelial cell factors on in vitro motility of malignant and non-malignant cells. *Ann. Biomed. Eng.* 36 (6), 958–969.
- Yamanaka, H., Kondo, S., 2014. In vitro analysis suggests that difference in cell movement during direct interaction can generate various pigment patterns in vivo. *Proc. Natl. Acad. Sci.* 111 (5), 1867–1872.
- Yates, C.A., 2014. Discrete and continuous models for tissue growth and shrinkage. *J. Theor. Biol.* 350, 37–48.
- Yates, C.A., Baker, R.E., Erban, R., Maini, P.K., 2012. Going from microscopic to macroscopic on nonuniform growing domains. *Phys. Rev. E* 86 (2), 021921.
- Yates, C.A., Parker, A., Baker, R.E., 2015. Incorporating pushing in exclusion-process models of cell migration. *Phys. Rev. E* 91 (5), 052711.
- Yates, C.A., Ford, M.J., Mort, R.L., 2017. A multi-stage representation of cell proliferation as a Markov process. *Bull. Math. Biol.* 79 (12), 2905–2928.

2.2. Conclusions

This chapter provides an overview of the state of the art of mathematical modelling of cell behaviour. Some of the techniques illustrated here provide a fundamental mathematical basis which will be applied and extended throughout the thesis. For example, in Chapters 3 and 5 we employed the diffusive limit approach (which was explained in detail in Section 2.1 of the review) to derive macroscopic PDEs. In addition, the multi-stage model of cell proliferation introduced in Section 3.1 represents a key component of the analysis of the cell invasion study of Chapter 5. Overall, the dichotomy between microscopic, individual-level models and macroscopic, population-level descriptions represents the essence of the multiscale modelling approach and the leitmotif of this thesis.

In the wider context of the current literature, the role of this review is to bring together and summarise some of the main approaches of modelling cell behaviour. Many mathematical models described here are based on very few assumptions which makes this review applicable to broad areas of mathematical biology, beyond our primary interest for cell behaviour. For example, Jhawar et al. [2019] has recently referenced to this chapter in the general context of collective behaviour. Overall this chapter has received a good level of attention from the community of mathematical biology and it has been cited five times in the first year after being published.

It is important to underline that this review does not represent an exhaustive manual for new generations of mathematical biologists. Instead, its main goal is to give a general introduction to the topic of multiscale modelling and to highlight its potential applications to cell biology. Some alternative approaches could not be considered in this review, for reasons of space, even though they constitute an important component of the multiscale modelling. In particular, while the majority of the discussion of this chapter is devoted to the interface between microscale and macroscale, *i.e.* between ABMs and PDEs, the roles of mesoscopic models, such as stochastic differential equations (SDEs) or stochastic partial differential equations (SPDEs), are omitted. The idea behind mesoscopic modelling is to combine a continuum-deterministic model with some noise terms which modulate the intrinsic stochasticity of the system. These models are relevant for a wide range of applications which involve a relatively large number of cells, but in which stochasticity still plays an important role. In Chapter 6, we present an example in which employing a mesoscopic multiscale modelling approach turns out to be crucial for quantifying the role of demographic noise in a growing cell population. A detailed discussion of mesoscopic models, however, is beyond the scope of the review presented in this chapter. There are several more specific reviews on this topic which can be found in the literature and that can be adopted to accompany the reading of this chapter. Of particular relevance is the recent work of Jhawar et al. [2019], in which the authors explain how to derive mesoscopic models from ABMs of collective behaviour in a similar style to this chapter.

Chapter 3

Modelling persistence of motion in a crowded environment: The diffusive limit of excluding velocity-jump processes

This chapter comprises a paper which I published in Physical Review E [Gavagnin and Yates, 2018a]. In the paper we investigate the interplay between cell directional persistence and volume exclusion. Despite these two aspects of cell motion having been intensively studied in different contexts, a comprehensive theory for the phenomena arising from their interaction is still lacking. We define a series of ABMs based on a velocity-jump process which incorporate excluding interactions of increasing complexity. By taking a diffusive limit from the occupancy master equation of the ABMs, we obtain a set of PDEs describing the time evolution of the average cell density. We analyse and compare the two modelling regimes to highlight their versatility and their limitations.

3.1. Outline of the article

Section I of the paper contains an introduction to the problem of modelling directional persistence in the context of cell behaviour. In Section II we define the on-lattice ABM and four possible types of volume exclusion properties. The derivation of the PDE models is carried out in Section III. In Section IV we present the results of the comparison between ABMs and PDEs and we describe three phenomena of interest: the appearance of density spikes in the profile of the ABMs (Section IV-A), the intrinsic anisotropy of the lattice-based models (Section IV-B) and the emergence of spontaneous aggregation (Section IV-C). We conclude with a brief discussion in Section V. The appendices contain the complete set of master equations and PDEs which were omitted in the main text.

Appendix B: Statement of Authorship

This declaration concerns the article entitled:									
Modeling persistence of motion in a crowded environment: The diffusive limit of excluding velocity-jump processes									
Publication status (tick one)									
draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input checked="" type="checkbox"/>
Publication details (reference)	Journal: Physical Review E 97, 032416 Authors: Enrico Gavagnin, and Christian A. Yates								
Candidate's contribution to the paper (detailed, and also given as a percentage).	All the calculations have been performed by the author of the thesis (100%). All numerical computations and simulations have been performed by the author of the thesis (100%). All authors contributed equally to the presentation of the content (50%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed							Date	27.4.2020	

Modeling persistence of motion in a crowded environment: The diffusive limit of excluding velocity-jump processes

Enrico Gavagnin* and Christian A. Yates

Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom



(Received 24 October 2017; revised manuscript received 21 February 2018; published 26 March 2018)

Persistence of motion is the tendency of an object to maintain motion in a direction for short time scales without necessarily being biased in any direction in the long term. One of the most appropriate mathematical tools to study this behavior is an agent-based *velocity-jump process*. In the absence of agent-agent interaction, the mean-field continuum limit of the agent-based model (ABM) gives rise to the well known hyperbolic telegraph equation. When agent-agent interaction is included in the ABM, a strictly advective system of partial differential equations (PDEs) can be derived at the population level. However, no diffusive limit of the ABM has been obtained from such a model. Connecting the microscopic behavior of the ABM to a diffusive macroscopic description is desirable, since it allows the exploration of a wider range of scenarios and establishes a direct connection with commonly used statistical tools of movement analysis. In order to connect the ABM at the population level to a diffusive PDE at the population level, we consider a generalization of the agent-based velocity-jump process on a two-dimensional lattice with three forms of agent interaction. This generalization allows us to take a diffusive limit and obtain a faithful population-level description. We investigate the properties of the model at both the individual and population levels and we elucidate some of the models' key characteristic features. In particular, we show an intrinsic anisotropy inherent to the models and we find evidence of a spontaneous form of aggregation at both the micro- and macroscales.

DOI: [10.1103/PhysRevE.97.032416](https://doi.org/10.1103/PhysRevE.97.032416)

I. INTRODUCTION

Understanding the properties of cell movement is of fundamental interest in many biological contexts such as embryogenesis [1], epidermal wound healing [2], and tumor growth [3]. Mathematical models are now considered essential tools in cell biology for testing theoretical hypotheses, interpreting experimental data, and extracting biological parameters [4–7]. There are typically two approaches to modeling cell motion, either microscale discrete [4,8–13] or macroscale continuum [8,14–16]. The discrete approach, using agent-based models (ABMs), accounts for properties at the cell-scale, while the continuum approach, often presented as a system of partial differential equations (PDEs) or stochastic partial differential equations (SPDEs), gives a global description of the migration at the population level. Continuum models have the advantage that they are generally more amenable to mathematical analysis and can lead to significant insights for situations in which the system comprises a large number of agents, at which point simulating the ABM becomes computationally expensive. Nevertheless, finding the appropriate continuum model to describe the collective behavior of a system of moving agents can be a difficult task and continuum models are often specified on a phenomenological basis, which may reduce their predictive power. It is essential, therefore, to establish a connection between microscale properties, which can be inferred directly from experimental data, and macroscale dynamics [17–19].

Many analyses of cell migration are based on the hypothesis that the movement of a single cell can be described as a simple random walk on a lattice [1,5,10,18]. In many models, the behavior of a single cell is assumed to be independent of the other cells' positions, and multiple cells can occupy the same lattice site simultaneously [20,21]. In many applications, however, crowding effects play an important role that cannot be neglected [2,4]. Crowding is incorporated into such models via volume exclusion: each lattice site is allowed to be occupied by at most one cell [22–26]. A macroscopic continuum description of this type of model can be obtained by considering an average mass conservation law for each lattice site and taking an appropriate limit as the spatial and temporal discretization steps go to zero simultaneously [10].

One of the key aspects of a simple random walk is that the direction of motion undergoes a series of uncorrelated jumps in space. In reality, experimental observations indicate that many types of cell tend to preserve their direction of motion for a certain time before reorienting [21,27–29], even though the movement is globally unbiased. This tendency is normally called *persistence of motion* or *of direction* [30]. There is vast literature about modeling persistence at multiple scales for non-interacting agents [20,31–34], but it is only in recent years that there has been an increasing interest in studying the role of persistence of motion for systems of self-interacting agents [22–24,35–39] and persistence induced by crowding [40,41].

Typically, to incorporate persistence in the ABM, cell movement is represented as a correlated random walk (CRW), which is also known as a *velocity-jump process* [20,31]. In this model, the cell has an assigned direction of motion (left or right in one dimension) and it moves in this direction with constant

*Corresponding author: e.gavagnin@bath.ac.uk

velocity, v , until the assigned direction is changed, which occurs according to a Poisson process with a given rate, λ . Notice that the temporary preferential direction induces bias in the motion for short time scales, which represents persistence, but the resulting motion remains globally unbiased for longer time scales.

In the case of non-interacting agents, the macroscale behavior of the velocity-jump process in one dimension is well known to evolve according to the hyperbolic telegraph equation for the cell density $C(x, t)$ [31,42,43]:

$$\frac{\partial^2 C}{\partial t^2} + 2\lambda \frac{\partial C}{\partial t} = v^2 \frac{\partial^2 C}{\partial x^2}. \quad (1)$$

Notice that Eq. (1) was originally developed to describe the propagation of waves which travel and reflect through a telegraph transmission line [44]. The same type of equation can be derived from a system of non-interacting agents performing a velocity-jump process in one dimension. In particular, Othmer *et al.* [31] derived the telegraph equation for cells undergoing velocity-jump processes without interactions. Othmer and Hillen [32] demonstrated that it is possible to obtain a parabolic limit as v and λ tend to infinity simultaneously, such that v^2/λ remains constant. In this limit the canonical diffusion equation is recovered [32,33]. This is not a surprise, since the short-term correlation effects become less evident at large time scales and so the limit process is effectively equivalent to a simple random walk.

When direct agent-agent interactions are introduced, however, the derivation of an exact closed form PDE for the total mean agent density is not possible [20]. Recently, Treloar *et al.* derived a system of macroscopic advective equations from a velocity-jump process with three different forms of direct interaction [22,23]. Although their continuum model is successful in replicating the population-level behavior of the ABM for a limited range of model parameters, the first order approximation considered by Treloar *et al.* [22] enforces restrictions in the initial condition (which must be sufficiently smooth) and on the choice of the parameters.

The aim of our work is to ease these restrictions in order to provide a better connection between discrete and continuum models of volume excluding persistent agents. We consider a generalization of the ABM of Treloar *et al.* [22] in which we modulate the influence of persistence through an additional parameter φ . This allows us to take a *diffusive* limit if the new parameter φ scales with the lattice size. The resulting partial differential equation (PDE) description includes a non-linear diffusive term, which encapsulates the long-term diffusive behavior of cells, and an advective part, which is consistent with the findings of Treloar *et al.* [22]. Our new diffusive model represents an extension of the previous advective model and can be applied to study a wider range of scenarios. In particular, we can consider situations with a steep gradient in cell density, which have not previously been investigated. Moreover, a diffusive limit is appropriate for the study of the long-term behavior of the system, especially if we are interested in statistical tools which are related to the diffusion coefficient, such as the mean squared displacement and the mean dispersal distance [20].

In this paper we study the two-dimensional version of a model which incorporates persistence and volume exclusion.

We explain the derivation of the diffusive continuum description and finally we test the agreement between our discrete and continuum models using some illustrative examples. Our new diffusive PDEs correctly represent the population-level behavior of our ABMs, particularly in scenarios that could not be investigated with the previous advective models. Our investigation highlights some peculiar aspects of the excluding velocity-jump processes, which we discuss in the light of the new macroscopic description. In particular, a spontaneous form of aggregation, similar to that observed by Thompson *et al.* [35] and Sepúlveda and Soto [36], appears in both our agent-based and population-level models. We believe this is the first reported example in which such a phenomenon appears at both micro- and macroscales.

The paper is organized as follows. In Sec. II we define the ABM and introduce three forms of cell-cell interaction. In Sec. III, we derive the continuum diffusive description of the ABM from the occupancy master equations. Our numerical results on the comparison between the ABMs and the corresponding PDEs are shown in Sec. IV, together with our observations on some of the interesting model behaviors. We conclude with a short discussion of our results and possible avenues for future research in Sec. V.

II. THE AGENT-BASED MODEL

In this section we describe the basic ABM. The models presented in the following sections are all adaptations of this basic model. Cells are represented by agents on a square lattice of size $L_x \times L_y$ sites and lattice step Δ with periodic boundary conditions in the y direction and zero-flux boundary conditions in the x direction.¹ Each site of the lattice can be occupied by at most one cell, in which case we say the site is occupied; otherwise the site is said to be empty.

We assign to each agent a polarization in one of the four directions of the lattice. We denote such polarization with the corresponding initial capital letter: right (R), left (L), up (U), and down (D). Let $v \in \mathbb{N}^+$ be a positive integer which denotes the number of lattice sites that an agent can move during a single movement event. We can interpret this as a non-dimensional measure of the agent's velocity. Agents can move or reorient their polarization in continuous time. Both of these events occur at random as independent Poisson processes with rates P_m and P_r , respectively. The role of the polarization is to induce a temporary bias in the stochastic motion so that the polarized agent is more likely to move in the corresponding direction. Let $\varphi \in [0, 1]$ be a parameter which characterizes the intensity of the bias. Consider, for example, an R-polarized agent in two dimensions, located at site (i, j) . If the agent is chosen to move, one of the four sites $(i \pm v, j)$, $(i, j \pm v)$ is selected, at random, as a target site. The right-hand site $(i + v, j)$, corresponding to the R-polarization of that cell, is chosen with probability given by $\frac{1+\varphi}{4}$. In the opposite

¹We implement periodic boundary conditions in the vertical direction to avoid edge effects. In the horizontal direction we employ zero-flux boundary conditions, although in reality agents rarely, if ever, reach these boundaries. So other boundary conditions may be employed with little consequence.

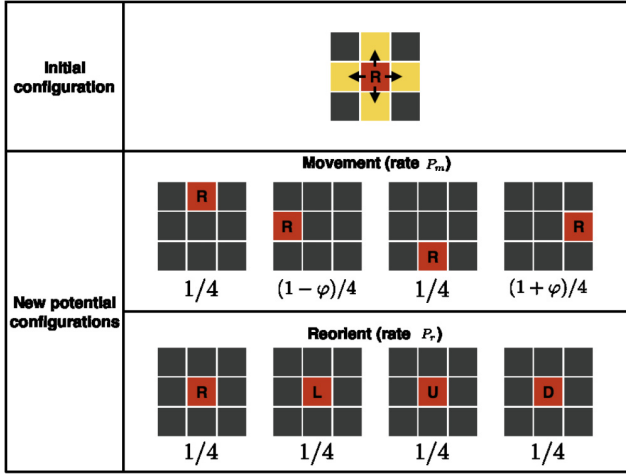


FIG. 1. Diagram of the motility mechanism of an R-polarized agent in the ABM with velocity $v = 1$. Red (dark grey) sites are occupied by the moving agent with the polarization denoted by the corresponding letter (R, right; L, left; U, up; D, down), black sites are empty, and yellow (light grey) sites highlight the von Neumann neighbor sites. The top panel shows the initial configuration and the neighbor sites reachable by the agents. The middle panel shows the four new potential configurations in the case that a movement occurs (with rate P_m) with the corresponding probabilities. The bottom panel shows the four new potential configurations in the case that a reorientation event occurs (with rate P_r), the agent remains in the same site, and its polarization is chosen uniformly at random.

direction to the polarization, the left-hand site $(i - v, j)$ is chosen with probability $\frac{1-\varphi}{4}$ and each of the sites in the vertical direction $(i, j \pm v)$ (orthogonal to the polarization direction) are chosen with probability $\frac{1}{4}$ (see the schematic in Fig. 1).² The transition probabilities for the other polarizations are obtained analogously. Finally, if a reorientation event occurs, with rate P_r , the agent changes its polarization uniformly at random to one of the four possible polarizations (including the possibility of maintaining its current polarization) (see Fig. 1).

Notice that if $\varphi = 0$, then the target site is chosen uniformly and the movement corresponds to a classic uncorrelated random walk (with non-local jumps for $v > 1$). On the contrary, if we let $\varphi = 1$ we achieve the strongest bias where agents cannot move in the opposite direction to their polarization.³

When the rates of reorientation and movement are chosen such that $P_r \ll P_m$ together with a large value of the parameter

²Alternative transition probabilities could be considered. For example, one could choose to reduce the movement probabilities in the direction orthogonal to the polarization (up and down in the example of an R-polarized agent) as well as in the direction opposite to the current polarization. A similar derivation can be applied, which will lead to a similar but slightly altered macroscopic model.

³Note that if the model is specified in one dimension and $\varphi = 1$, the model corresponds to the velocity-jump process described by Treloar *et al.* [22]. However, in the two-dimensional case the choice of maximum bias leads to a different model. In particular, the target site is still chosen at random (although not uniformly) between three of the nearest neighbors while in the work by Treloar *et al.* [22], upon an agent being selected to move, its target site is chosen deterministically.

φ , agents persist in their direction of motion. Figure 2 shows two trajectories of a single agent for parameters $P_m = 1$, $P_r = 0.05$, and $v = 1$ and for Fig. 2(a) $\varphi = 0$ (no persistence) and for Fig. 2(b) $\varphi = 0.8$ (strong persistence). Persistence of motion is clearly evident in the shape of the track in Fig. 2(b). The long-term behavior is unbiased, since none of the four directions is preferred in the long term. As there is only one agent in the domain, exclusion (specified in the next paragraph) does not play any role in the dynamics.

Once the target site is selected, the agent moves according to the exclusion property specified for the process. For consistency with Treloar *et al.* [22], we consider four different exclusion properties, one without agent interaction and three with a variety of interactions. Figure 3 shows two typical scenarios (for $v = 3$) in which an agent (red) at position i attempts to move to the target site at $i + 3$. In scenario A the target site is occupied by another agent (blue), while in scenario B the target site is empty and the site $i + 2$ is occupied. We use these two examples to explain the four exclusion properties as follows.

Type 0: Non-interacting agents. In this case the moving agent moves to the target site regardless of its occupancy. Such a process is not an exclusion process since arbitrarily many agents can occupy the same site. In both scenarios in Fig. 3 the moving agent moves to the target site. In scenario A the agent shares the site with the other agent (blue) and in scenario B it occupies the target site alone.

Type 1: Only move if target site is available. In this case the agent moves only when the target site is not occupied. In scenario A of Fig. 3 this exclusion property causes the entire movement to be aborted, whereas in scenario B the moving agent jumps over the blue agent occupying the site $i + 2$ in order to reach site $i + 3$.

Type 2: Farsighted agents. In this case the movement takes place only if the target site and all the intermediate sites are vacant. If at least one of the sites is occupied, the movement is aborted and the moving agent remains at the initial position. The assumption distinguishing this exclusion process is that agents know the occupancy of distant sites in order to decide whether to move to the target or not. If cells extend sensing filopodia, this behavior can be justified for short distances [45], but it becomes unrealistic for large values of v . In both scenarios of Fig. 3 the movement under this movement type is aborted since either the target or an intermediate site is occupied.

Type 3: Shortsighted agents. This is the most mathematically sophisticated and realistic form of interaction that we consider. With this scheme, the moving agent moves through the intermediate sites between its position and the target site and stops at the furthest site which it can reach without being blocked by any of the other agents. If no blocking occurs, the agent moves to the target site. In Fig. 3 we see that this exclusion property allows the agent to move in both scenarios; the distance depends on the position of the blocking agent (blue).

III. POPULATION-LEVEL MODEL

In this section we derive a family of PDEs which describe the behavior of the ABMs introduced in Sec. II at the population level. The general technique consists of writing down the

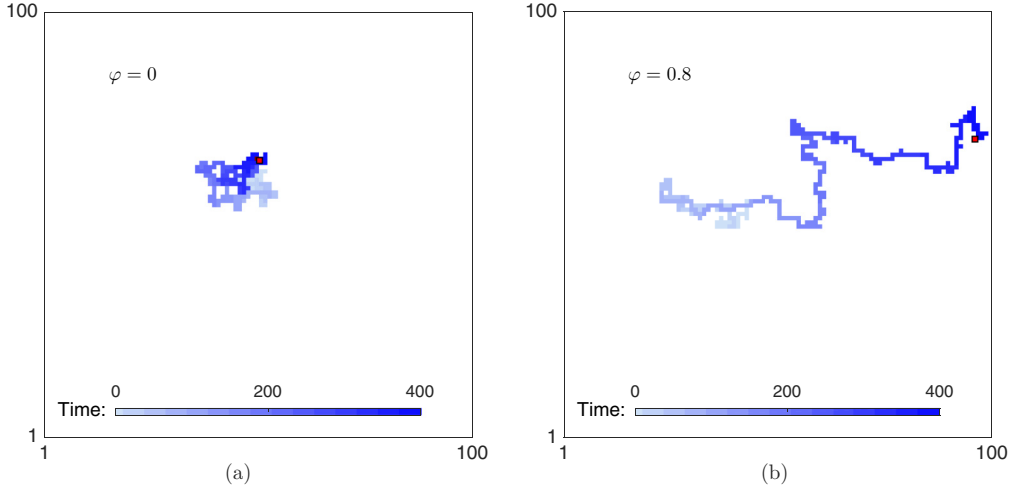


FIG. 2. Trajectories of a single agent moving according to the ABM scheme on a two-dimensional square lattice with $L_x = L_y = 100$ and $\Delta = 1$: (a) an example of a simple random walk ($\varphi = 0$) and (b) an example of a persistent random walk ($\varphi = 0.8$). The blue points represent the subsequent positions, from light ($t = 0$) to dark ($t = 400$) blue (grey), of a single agent initialized in the center of the domain, $(50, 50)$. The last position is highlighted in red (light grey with black border). In both panels the parameters are $P_m = 1$, $P_r = 0.05$, and $v = 1$.

continuous-time occupancy master equation for the average occupancy of a general site of the ABM, Taylor expanding, and finally taking a limit as the lattice step, Δ , and the time step, τ , go to zero jointly while holding Δ^2/τ constant (see Chap. 4 of [10]). The four types of agent interactions considered lead to different PDEs; therefore, in what follows, we consider those cases in different sections.

We denote by $C_{i,j}^t(m)$ the occupancy of the site (i, j) at time t for the m th simulation of the ABM; i.e., $C_{i,j}^t(m) = 1$ if

the site (i, j) at time t in the m th simulation is occupied and $C_{i,j}^t(m) = 0$ if it is empty. We denote with $\hat{C}_{i,j}^t$ the occupancy of the site (i, j) at time t averaged over all M realizations, i.e.,

$$\hat{C}_{i,j}^t = \frac{1}{M} \sum_{m=1}^M C_{i,j}^t(m). \quad (2)$$

In addition, we define occupancy variables for the four subpopulations of agents classified according to their polarization (right, left, up, or down) and we denote these occupancies with the capital corresponding to their first letter. For example, $R_{i,j}^t(m)$ represents the occupancy of right-polarized agents at the site (i, j) at time t for the m th simulation and $\hat{R}_{i,j}^t$ the corresponding occupancy averaged over the total number of realizations. The total occupancy can be obtained by adding together the occupancies of the four subpopulations as $\hat{C}_{i,j}^t = \hat{R}_{i,j}^t + \hat{L}_{i,j}^t + \hat{U}_{i,j}^t + \hat{D}_{i,j}^t$. From now on, we omit the hats in the averaged occupancies for simplicity. All the occupancies are defined for $t \in \mathbb{R}^+$ and $(i, j) \in \{1, \dots, L_x\} \times \{1, \dots, L_y\}$.

A. Type 0 interaction

For the case of non-interacting agents we can write down the occupancy master equations of the process, in sites away from the boundary,⁴ as

$$\begin{aligned} R_{i,j}^{t+\tau} = & R_{i,j}^t + \frac{\tau P_m}{4} \left[(1 + \varphi) R_{i-v,j}^t + (1 - \varphi) R_{i+v,j}^t \right. \\ & \left. + R_{i,j+v}^t + R_{i,j-v}^t - 4 R_{i,j}^t \right] \\ & + \frac{\tau P_r}{4} \left[L_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3 R_{i,j}^t \right] + \mathcal{O}(\tau^2), \end{aligned}$$

FIG. 3. Schematics of the four different exclusion properties considered in one dimension. The first row shows the initial configuration for the two scenarios considered. In both cases the moving agent (red) [dark grey] attempts to move to the target site (yellow T) three sites to the right ($v = 3$). The black sites are empty and the blue (light grey) sites are occupied by other agents. Each of the subsequent four rows represents the new configuration of cells under the exclusion property chosen.

⁴Note that, for sites on the boundary, the occupancy equations will be slightly different, but we employ zero-flux or periodic boundary conditions on the vertical and horizontal boundaries (respectively) in the continuum model to match our specification in the ABM.

$$\begin{aligned}
L_{i,j}^{t+\tau} &= L_{i,j}^t + \frac{\tau P_m}{4} [(1-\varphi)L_{i-v,j}^t + (1+\varphi)L_{i+v,j}^t \\
&\quad + L_{i,j+v}^t + L_{i,j-v}^t - 4L_{i,j}^t] \\
&\quad + \frac{\tau P_r}{4} [R_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3L_{i,j}^t] + \mathcal{O}(\tau^2), \\
U_{i,j}^{t+\tau} &= U_{i,j}^t + \frac{\tau P_m}{4} [U_{i+v,j}^t + U_{i-v,j}^t + (1-\varphi)U_{i,j+v}^t \\
&\quad + (1+\varphi)U_{i,j-v}^t - 4U_{i,j}^t] \\
&\quad + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + D_{i,j}^t - 3U_{i,j}^t] + \mathcal{O}(\tau^2), \\
D_{i,j}^{t+\tau} &= D_{i,j}^t + \frac{\tau P_m}{4} [D_{i+v,j}^t + D_{i-v,j}^t + (1+\varphi)D_{i,j+v}^t \\
&\quad + (1-\varphi)D_{i,j-v}^t - 4D_{i,j}^t] \\
&\quad + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + U_{i,j}^t - 3D_{i,j}^t] + \mathcal{O}(\tau^2), \quad (3)
\end{aligned}$$

where τ is sufficiently small such that the probability that more than one event occurs in the time interval $(t, t + \tau)$ is $\mathcal{O}(\tau^2)$. The terms in system (3) that contain P_m represent the transitions into and out of the site (i, j) due to the motility events, while the terms which contain P_r represent the transition from one polarization to another due to the reorienting events. In order to obtain a continuous approximation of the model, we Taylor expand terms such as $R_{i\pm v,j}^t$ and $R_{i,j\pm v}^t$ about position (i, j) as

$$\begin{aligned}
R_{i\pm v,j}^t &= R(x_{i\pm v}, y_j, t) \\
&= R(x_i, y_j, t) \pm v \Delta \frac{\partial R}{\partial x}(x_i, y_j, t) \\
&\quad + \frac{1}{2}(v \Delta)^2 \frac{\partial^2 R}{\partial x^2}(x_i, y_j, t) + \dots, \\
R_{i,j\pm v}^t &= R(x_i, y_{j\pm v}, t) \\
&= R(x_i, y_j, t) \pm v \Delta \frac{\partial R}{\partial y}(x_i, y_j, t) \\
&\quad + \frac{1}{2}(v \Delta)^2 \frac{\partial^2 R}{\partial y^2}(x_i, y_j, t) + \dots \quad (4)
\end{aligned}$$

In order to take the diffusive limit we assume that φ rescales with the spatial step of the lattice, i.e., $\varphi = \mathcal{O}(\Delta)$. This assumption, that the local bias tends to zero, is necessary in order to derive a finite advective term in the diffusive limit. A similar assumption is made in order to derive consistent continuum limits of models with global bias [5,20].

By substituting the truncated Taylor expansions [of which Eqs. (4) are an example] into Eqs. (3) and rearranging terms, we can take the diffusive limit as $\tau, \Delta \rightarrow 0$ in Eqs. (3) such that Δ^2/τ remains fixed. We obtain a system of PDEs for the continuous occupancy functions $R(x, y, t)$, $L(x, y, t)$, $U(x, y, t)$, and $D(x, y, t)$, where $t \in \mathbb{R}^+$ and $(x, y) \in [0, \Delta L_x] \times [0, \Delta L_y]$,

$$\begin{aligned}
\frac{\partial R}{\partial t} &= v^2 \mu \nabla^2 R - v v \frac{\partial R}{\partial x} + \frac{P_r}{4}(C - 4R), \\
\frac{\partial L}{\partial t} &= v^2 \mu \nabla^2 L + v v \frac{\partial L}{\partial x} + \frac{P_r}{4}(C - 4L),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial U}{\partial t} &= v^2 \mu \nabla^2 U - v v \frac{\partial U}{\partial y} + \frac{P_r}{4}(C - 4U), \\
\frac{\partial D}{\partial t} &= v^2 \mu \nabla^2 D + v v \frac{\partial D}{\partial y} + \frac{P_r}{4}(C - 4D), \quad (5)
\end{aligned}$$

where

$$\mu := \lim_{\Delta \rightarrow 0} \frac{\Delta^2 P_m}{4} \quad \text{and} \quad v := \lim_{\Delta \rightarrow 0} \frac{\varphi \Delta P_m}{2}. \quad (6)$$

The boundary conditions are chosen to be consistent with the ABM. In particular, for every $x \in [0, \Delta L_x]$, $y \in [0, \Delta L_y]$, and $t \in \mathbb{R}^+$ we impose

$$\begin{aligned}
R(x, 0, t) &= R(x, \Delta L_y, t), \quad \frac{\partial R}{\partial x}(0, y, t) = \frac{\partial R}{\partial x}(\Delta L_x, y, t) = 0, \\
L(x, 0, t) &= L(x, \Delta L_y, t), \quad \frac{\partial L}{\partial x}(0, y, t) = \frac{\partial L}{\partial x}(\Delta L_x, y, t) = 0, \\
U(x, 0, t) &= U(x, \Delta L_y, t), \quad \frac{\partial U}{\partial x}(0, y, t) = \frac{\partial U}{\partial x}(\Delta L_x, y, t) = 0, \\
D(x, 0, t) &= D(x, \Delta L_y, t), \quad \frac{\partial D}{\partial x}(0, y, t) = \frac{\partial D}{\partial x}(\Delta L_x, y, t) = 0. \quad (7)
\end{aligned}$$

The two limits in Eqs. (6) exist and are finite owing to the assumption on φ above. The right-hand sides of system (5) comprise three terms (in order from left to right): a diffusive term, an advective term, and a reactive term. The diffusive terms capture the long-term unbiased motion of the agents. In the case of the non-interacting agents, described above, the diffusion coefficient is independent of the agent density. The advective terms reflect the polarization of each subpopulation and, as such, they involve the first partial derivative of density in the direction of the polarization. The reactive terms represent the uniform changing of polarization.

We can write down the PDE for the total averaged density by adding the equations of system (5):

$$\frac{\partial C}{\partial t} = v^2 \mu \nabla^2 C + v v \frac{\partial}{\partial y}[(D - U)] + v v \frac{\partial}{\partial x}[(L - R)]. \quad (8)$$

Note that a closed form for the equations in terms of the total density is not possible unless $\varphi = 0$, in which case the total density evolves according to the canonical diffusion equation, consistently with [20,31].

Notice that advective-diffusive equations like the one of system (5) are an extensively studied class of PDEs which can be found in a wide range of applications. In particular, they are traditionally used to represent transport phenomena such as heat transfer [46,47], mass transfer [47], and virus propagation [48,49].

B. Type 1 interaction

For the first non-trivial type of interaction that we consider, the movement of the agents depends only on the occupancy of the target site. Specifically, movement is aborted if, and only if, the target site is occupied (see Fig. 3). The occupancy

master equation for the right-moving subpopulation reads

$$\begin{aligned}
 R_{i,j}^{t+\tau} = & R_{i,j}^t + \frac{\tau P_m}{4} (1 - C_{i,j}^t) [(1 + \varphi) R_{i-v,j}^t \\
 & + (1 - \varphi) R_{i+v,j}^t + R_{i,j+v}^t + R_{i,j-v}^t] \\
 & - \frac{\tau P_m}{4} R_{i,j}^t [(1 + \varphi)(1 - C_{i+v,j}^t) + (1 - \varphi)(1 - C_{i-v,j}^t) \\
 & + (1 - C_{i,j+v}^t) + (1 - C_{i,j-v}^t)] \\
 & + \frac{\tau P_r}{4} [L_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3R_{i,j}^t] + \mathcal{O}(\tau^2). \quad (9)
 \end{aligned}$$

The equations for the other three subpopulations are given in Appendix A. The main difference in comparison to the non-interacting type 0 models is the introduction of terms that reduce the probability of moving according to the density of the target site. For example, the term $(1 - C_{i,j}^t)$ determines the probability of success of a movement into the site (i, j) at time t . If at time t the site (i, j) is occupied in all the realizations of the ABM we have $C_{i,j}^t = 1$ so the corresponding probability of success is zero. Conversely, if the site (i, j) is empty in all the simulations, the probability of success is 1, since the movement is always allowed to take place. Notice that in writing down the occupancy master equation (9), we are making the mean-field assumption that the occupancies of neighboring sites are independent.

We can use the same steps as in Sec. III, for type 0 interactions, to obtain a system of diffusive PDEs for the density of the four different polarizations. The resulting equation for the right-moving subpopulation is given by

$$\begin{aligned}
 \frac{\partial R}{\partial t} = & v^2 \mu [R \nabla^2 C + (1 - C) \nabla^2 R] \\
 & - v \nu \frac{\partial}{\partial x} [R(1 - C)] + \frac{P_r}{4} (C - 4R), \quad (10)
 \end{aligned}$$

where μ and ν are as defined in Eqs. (6). See system (B1) of Appendix B for the full set of equations. We can see that the switching rates between subpopulations remain the same as in system (5), whereas the advective and the diffusive terms of Eq. (B1) depend linearly on the cell density. Specifically, both the advective and the original diffusive parts are scaled by a factor of $1 - C$, which takes into account the decrease in motility due to the volume exclusion. An additional diffusive term $\nabla^2 C$ appears, scaled by the density of each subpopulation. Notice that, by adding together the equations for the four subpopulations, we recover a normal diffusive term for the total cell density:

$$\begin{aligned}
 \frac{\partial C}{\partial t} = & v^2 \mu \nabla^2 C + v \nu \frac{\partial}{\partial y} [(D - U)(1 - C)] \\
 & + v \nu \frac{\partial}{\partial x} [(L - R)(1 - C)]. \quad (11)
 \end{aligned}$$

Nevertheless, as in the previous case (type 0), the advective terms make it impossible to close the PDE for the total density, $C(x, y, t)$, apart from in the trivial case, $\varphi = 0$.

C. Type 2 interaction

For the second type of non-trivial interaction, a chosen movement event takes place from the current site if, and only

if, the target site and all the intermediate sites are available (see Fig. 3). This leads to the following occupancy master equation for the right-moving subpopulation:

$$\begin{aligned}
 R_{i,j}^{t+\tau} = & R_{i,j}^t + \frac{\tau P_m}{4} \left[(1 + \varphi) R_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) \right. \\
 & + (1 - \varphi) R_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) \\
 & + R_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) + R_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \left. \right] \\
 & - \frac{\tau P_m}{4} R_{i,j}^t \left[(1 + \varphi) \prod_{s=1}^v (1 - C_{i+s,j}^t) \right. \\
 & + (1 - \varphi) \prod_{s=1}^v (1 - C_{i-s,j}^t) \\
 & + \prod_{s=1}^v (1 - C_{i,j+s}^t) + \prod_{s=1}^v (1 - C_{i,j-s}^t) \left. \right] \\
 & + \frac{\tau P_r}{4} [L_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3R_{i,j}^t] + \mathcal{O}(\tau^2). \quad (12)
 \end{aligned}$$

Again, we refer the reader to Appendix A [see system (A2)] for the other three occupancy master equations. Upon Taylor expansion and taking the appropriate limits, as before, we obtain

$$\begin{aligned}
 \frac{\partial R}{\partial t} = & v^2 \mu [R \nabla((1 - C)^{v-1} \nabla C) + (1 - C) \nabla((1 - C)^{v-1} \nabla R)] \\
 & - v \nu \frac{\partial}{\partial x} [R(1 - C)^v] + \frac{P_r}{4} (C - 4R), \quad (13)
 \end{aligned}$$

where μ and ν are defined in Eqs. (6). See system (B2) of Appendix B for the full set of equations. The main difference between Eqs. (B2) in comparison to the previous exclusion type, characterized by Eq. (B1), is that the rescaling factor, which accounts for the crowding effect, now depends on the v th power of the total density. Notice that for $v > 1$, by adding the equations for the four subpopulations, we obtain non-linear diffusion for the total cell density:

$$\begin{aligned}
 \frac{\partial C}{\partial t} = & v^2 \mu \nabla((1 - C)^{v-1} \nabla C) + v \nu \frac{\partial}{\partial y} [(D - U)(1 - C)^v] \\
 & + v \nu \frac{\partial}{\partial x} [(L - R)(1 - C)^v]. \quad (14)
 \end{aligned}$$

This suggests that the increase in aborted movements at the microscale in these type 2 interactions affects the long-term diffusive behavior.

D. Type 3 interaction

Finally, we consider the third (non-trivial) and most mathematically complex form of interaction. This consists of a focal agent moving to the furthest available site in its path towards the target site before (potentially) being blocked. The occupancy

master equation for the right subpopulation reads

$$\begin{aligned}
 R_{i,j}^{t+\tau} = & R_{i,j}^t + \frac{\tau P_m}{4} \left[(1+\varphi) R_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) (1-\varphi) R_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) + R_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) \right. \\
 & \left. + R_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \right] \\
 & - \frac{\tau P_m}{4} R_{i,j}^t \left[(1+\varphi) \prod_{s=1}^v (1 - C_{i+s,j}^t) + (1-\varphi) \prod_{s=1}^v (1 - C_{i-s,j}^t) + \prod_{s=1}^v (1 - C_{i,j+s}^t) + \prod_{s=1}^v (1 - C_{i,j-s}^t) \right] \\
 & + \frac{\tau P_m}{4} \left[(1+\varphi) C_{i+1,j}^t \sum_{k=1}^{v-1} R_{i-k,j}^t \prod_{s=0}^{k-1} (1 - C_{i-s,j}^t) + (1-\varphi) C_{i-1,j}^t \sum_{k=1}^{v-1} R_{i+k,j}^t \prod_{s=0}^{k-1} (1 - C_{i+s,j}^t) \right. \\
 & \left. + C_{i,j-1}^t \sum_{k=1}^{v-1} R_{i,j+k}^t \prod_{s=0}^{k-1} (1 - C_{i,j+s}^t) + C_{i,j+1}^t \sum_{k=1}^{v-1} R_{i,j-k}^t \prod_{s=0}^{k-1} (1 - C_{i,j-s}^t) \right] \\
 & - \frac{\tau P_m}{4} R_{i,j}^t \left[(1+\varphi) \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i+k,j}^t (1 - C_{i+s,j}^t) + (1-\varphi) \sum_{k=1}^v \prod_{s=1}^{k-1} C_{i-k,j}^t (1 - C_{i-s,j}^t) + \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j+k}^t (1 - C_{i,j+s}^t) \right. \\
 & \left. + \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j-k}^t (1 - C_{i,j-s}^t) \right] + \frac{\tau P_r}{4} [L_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3R_{i,j}^t] + \mathcal{O}(\tau^2). \quad (15)
 \end{aligned}$$

See system (A3) in Appendix A for the corresponding occupancy master equations for the other subpopulations. By Taylor expanding and taking the appropriate limits, as before, we obtain the continuum approximation given by

$$\frac{\partial R}{\partial t} = \mu \nabla \left[\sum_{k=1}^v (1-C)^{k-1} [(2k-1)(1-C) \nabla R - k(k-2) R \nabla C] \right] + v \frac{\partial}{\partial x} \left[\frac{(1-C)((1-C)^v - 1)}{C} R \right] + \frac{P_r}{4} (C - 4R), \quad (16)$$

where μ and v are as defined in system (6). The full set of equations is given in Appendix B. As with type 2 interactions, the polynomial rescaling factor due to the volume exclusion is of order v . Notice that the advective terms contain a factor C in the denominator. We choose to write the advective coefficients this way for notational convenience. Upon expansion of the numerator we see that it also contains a factor C , which cancels with the denominator, demonstrating that the coefficient, when simplified, is a polynomial rather than a quotient. The diffusive terms comprise a sum over $k = 1, \dots, v$, which reflects the possible movement events of length k . As for the previous case, by adding all the diffusive terms together for the four subpopulations, we obtain non-linear diffusion for the total population:

$$\begin{aligned}
 \frac{\partial C}{\partial t} = & \mu \nabla \left[\sum_{k=1}^v (1-C)^{k-1} [(2k-1) \nabla C + (1-k^2) C \nabla C] \right] \\
 & + v \frac{\partial}{\partial x} \left[\frac{(1-C)((1-C)^v - 1)}{C} (R - L) \right] \\
 & + v \frac{\partial}{\partial y} \left[\frac{(1-C)((1-C)^v - 1)}{C} (U - D) \right]. \quad (17)
 \end{aligned}$$

As expected, the three systems (B1)–(B3) for the interacting agent models (types 1, 2, and 3) are equivalent for $v = 1$. This is consistent with the ABMs, since the three forms of interaction

differ only when movements across multiple lattice sites are attempted, i.e., $v > 1$.

We should mention that the Taylor expansion in Eqs. (4) could be terminated at first order and, by following the same steps and taking the limit $\Delta \rightarrow 0$, $\tau \rightarrow 0$ as such that Δ/τ is constant, we would have obtained a family of equations similar to systems (5) and (B1)–(B3) without the contribution of the diffusive terms [22,23]. In this case the assumption on the parameter φ is no longer necessary. Othmer and Hillen [32,33] studied this type of system in detail in the case of non-interacting agents and for the particular case $\varphi = 1$ in one dimension. The model of Treloar *et al.* [22] represents a particular case of the one-dimensional version of the model defined in this paper with $\varphi = 1$. Treloar *et al.* [22] defined their model in terms of the *probabilities* of a single cell moving and reorienting in a given time step of length τ , which they denote P and λ , respectively. The reorienting rate, P_r , of our models corresponds to the limit $\Lambda = \lim_{\lambda, \tau \rightarrow 0} \lambda/\tau$ of the models of Treloar *et al.* [22]. Notice, in contrast to the suggestions of Treloar *et al.* [22], in our model there are no limitations on the rate P_r ; it can be chosen to be arbitrarily large. Treloar *et al.* considered a first order Taylor expansion that leads to a system of advective PDEs consistent with our continuum models. Apart from the special case $\lambda = 1/2$, for which a simple diffusion equation can be recovered, the nature of their model does not, in general, permit a diffusive limit to be taken. The introduction of the new parameter φ in our models allows us to consider a higher order Taylor expansion which results in the diffusive terms in Eqs. (5) and (B1)–(B3).

IV. RESULTS

In this section we compare the discrete simulations of the ABM with the continuous approximation. Then we investigate how the model behaves under particular choices of the parameters. We reveal three previously unobserved aspects of the model that appear when a high level of persistence is enforced: spike formation, anisotropy, and aggregation. Although such phenomena are interesting from a mathematical perspective, they represent potential obstacles for the application of such models to experimental data. We discuss the implications of such issues and future challenges in Sec. V.

All the ABMs and the corresponding PDEs are simulated on a two-dimensional domain. For the purpose of visualization, in most examples, we show the results for column-averaged cell density profiles (i.e., averaged over the y coordinates). For these examples, we define the total column-averaged density as

$$\bar{C}(x,t) = \frac{1}{L_y} \int_0^{L_y} C(x,y,t) dy, \quad (18)$$

and \bar{R} , \bar{L} , \bar{U} , and \bar{D} correspondingly. In these simulations we choose translationally invariant initial conditions in the vertical direction. In other words, $C(x,y,0) = C(x,0)$ for every $(x,y) \in [1, \Delta L_x] \times [1, \Delta L_y]$. Similarly, the polarized species, R , L , U , and D , are also initialized according to a translationally invariant condition. The periodic boundary conditions on the horizontal boundaries imply that translational invariance in the vertical direction is conserved as time evolves, namely,

$$C(x,y,t) = C(x,t), \quad (19)$$

for every $t \in \mathbb{R}^+$ and for every $(x,y) \in [1, \Delta L_x] \times [1, \Delta L_y]$ and similarly for the four subpopulations R , L , U , and D .

With this in mind, we can now derive the one-dimensional PDEs for the column-averaged densities from the corresponding two-dimensional equations. Formally, this is equivalent to dropping the dependence on y in all the density functions. As an example, we write down the averaged PDEs for the model without interaction (type 0 interactions). We omit the expressions for the other three cases, which can be obtained in a similar way. By column averaging system (5), we obtain the following system of equations:

$$\begin{aligned} \frac{\partial \bar{R}}{\partial t} &= v^2 \mu \frac{\partial^2 \bar{R}}{\partial x^2} - v \nu \frac{\partial \bar{R}}{\partial x} + \frac{P_r}{4} (\bar{C} - 4\bar{R}), \\ \frac{\partial \bar{L}}{\partial t} &= v^2 \mu \frac{\partial^2 \bar{L}}{\partial x^2} + v \nu \frac{\partial \bar{L}}{\partial x} + \frac{P_r}{4} (\bar{C} - 4\bar{L}), \\ \frac{\partial \bar{U}}{\partial t} &= v^2 \mu \frac{\partial^2 \bar{U}}{\partial x^2} + \frac{P_r}{4} (\bar{C} - 4\bar{U}), \\ \frac{\partial \bar{D}}{\partial t} &= v^2 \mu \frac{\partial^2 \bar{D}}{\partial x^2} + \frac{P_r}{4} (\bar{C} - 4\bar{D}), \end{aligned} \quad (20)$$

where μ and ν are defined as in Eqs. (6). The boundary conditions, for every $t \in \mathbb{R}^+$, are given by

$$\begin{aligned} \frac{\partial \bar{R}}{\partial x}(0,t) &= \frac{\partial \bar{R}}{\partial x}(\Delta L_x,t) = 0, & \frac{\partial \bar{L}}{\partial x}(0,t) &= \frac{\partial \bar{L}}{\partial x}(\Delta L_x,t) = 0, \\ \frac{\partial \bar{U}}{\partial x}(0,t) &= \frac{\partial \bar{U}}{\partial x}(\Delta L_x,t) = 0, & \frac{\partial \bar{D}}{\partial x}(0,t) &= \frac{\partial \bar{D}}{\partial x}(\Delta L_x,t) = 0. \end{aligned} \quad (21)$$

For Figs. 4–7 we use the same computational setup, which can be described as follows. The domain is a 400×400 lattice with $\Delta = 1$. We impose periodic boundary conditions on the horizontal boundaries and zero-flux boundary conditions on the vertical boundaries. The ABM is simulated using the Gillespie algorithm [50] for $M = 10$ identically prepared repeats. Few repeats are sufficient to compare the mean ABM behavior to the solutions of the PDEs for the mean occupancy because column averaging over 400 rows significantly reduces the noise in the ABM solutions. The numerical solutions of the PDEs are obtained through an implicit Euler method with spatial step $\delta x = 0.1$ and time step $\delta t = 0.1$, and using Picard iteration with tolerance $\epsilon = 10^{-3}$ to solve the non-linear equations.

Figure 4 shows the comparison between the total column-averaged density of the ABM and the PDE, for the four types of interactions at different times. The system is initialized such that the all sites with x coordinate between 161 and 240 are populated uniformly at random, with density $d = 0.5$. The polarization of the initial group of cells is chosen uniformly at random, so there is no bias towards any of the four polarizations. The PDE for the total column-averaged density is correspondingly initialized as constant $d = 0.5$ in the interval $[160, 240]$, with the density of the individual subpopulations also being constant at $d = 0.125$ in this region. The parameters for the model are $P_m = 1$, $P_r = 0.2$, $\varphi = 0.8$, and $v = 3$.

The agreement between the discrete and the continuous descriptions is generally very good for all four types of cell interactions, although we note that discrepancies are most noticeable for the type 3 interactions, for which our assumption of independence of site occupancy is least valid. We also tested our continuous approximation for different parameter values and found that the good agreement with the discrete model holds for a wide parameter range (results not shown). The good agreement is lost, however, as the value of persistence increases, i.e., $P_r \ll P_m$ and $\varphi \approx 1$ [see Fig. 9(c) for an example]. This disagreement is, in part, due to the significant spatial correlations induced by persistence in these models. Agents are highly likely to move to be adjacent to each other rather than aborting their movements (compare type 3 interactions to type 2 interactions, respectively). This tendency is ignored by our continuous approximations. We discuss this issue and potential improvements further in Sec. V.

The hyperbolic nature of previous continuum models of persistence of motion has meant that initial conditions with steep gradients have been difficult to investigate. Due to the diffusive nature of our continuous model, we are now able to examine initial conditions which have steep density gradients and still maintain a good agreement with the discrete model. In particular, this allows us to consider the initial condition described above with only a central region uniformly populated.

To compare the behavior of the different interaction mechanisms, we increase the initial total density to $d = 0.9$ and we display numerical solutions of the four PDEs (5) and (B1)–(B3) at time $T = 250$ (see Fig. 5). Non-interacting agents (type 0) lead to a faster agent spreading than any of the interacting types 1–3. Type 1 interactions cause a slightly slower spread of agents: although focal agents can jump

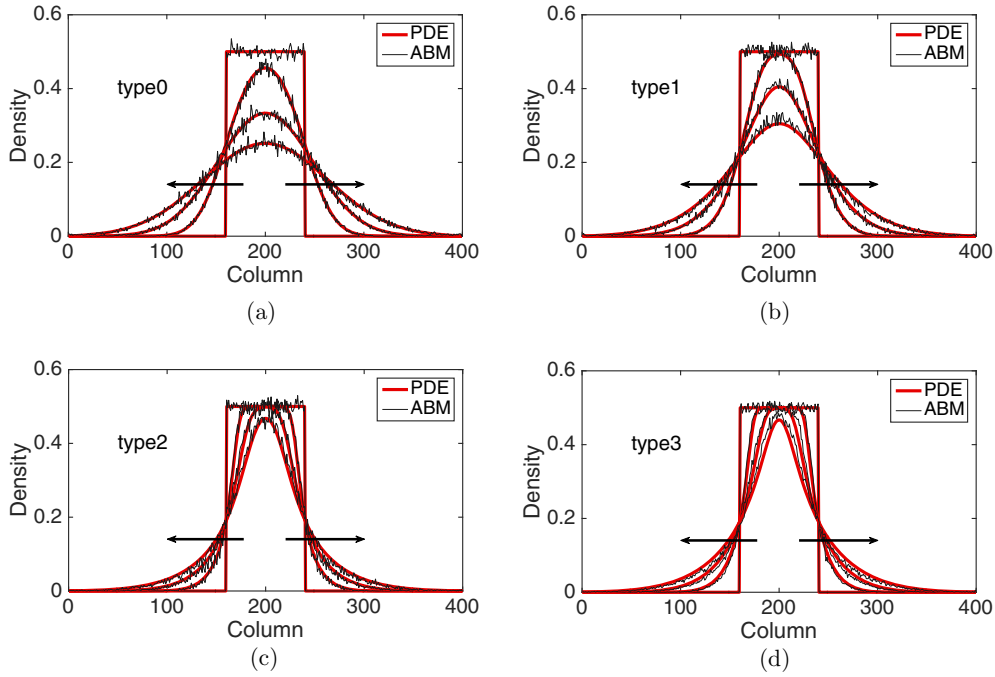


FIG. 4. Comparison between ABMs (thin black) and PDEs (thick red) for total column-averaged densities, $\bar{C}(x, t)$, for various forms of interaction. The solutions are displayed at the times $T = 0, 50, 150$, and 300 , with the direction of the black arrows indicating increasing time. The profiles are for (a) non-interacting agents [type 0, Eq. (5)], (b) the first type of non-trivial interaction [type 1, Eq. (B1)], (c) the second type of non-trivial interaction [type 2, Eq. (B2)], and (d) the third type of non-trivial interaction [type 3, Eq. (B3)]. All the ABMs are simulated using the Gillespie algorithm [50] averaged over $M = 10$ repeats.

over their neighbors, a small number of type 1 movement events are aborted if there is a cell in the target site. Type 2 interactions lead to the slowest spreading due to the high proportion of aborted movement events in which the focal agent stays stationary. Agents interacting through the type 3 mechanism spread slightly faster than agents undergoing type 2 interactions: although a significant proportion of movement events are aborted when agents are immediately adjacent to the focal agent, some small movement events are permitted towards a near neighbor which would otherwise be aborted under the implementation of type 2 interactions.

In Fig. 6 we compare the ABM and the continuum approximation for the right- and left-moving subpopulations in order

to evidence that the agreement between the models does not only hold at the population level. The density profiles of the up- and down-moving subpopulations are indistinguishable and we have omitted them for simplicity. However, the good agreement between the discrete and continuum models also holds for these cases (results not shown). We find good agreement, even for large values of the reorienting parameter, P_r , which has previously been thought not to be the case [22]. Note that the appearance of loss of total mass is due to the fact that we are only visualizing two of the four subpopulations in our two-dimensional model.

In the following sections we outline some of the inherent features of these models which have largely been overlooked, but which must be considered if the model is to be used in real applications.

A. Density spikes

In Fig. 7 we plot a zoomed-in density profile of the ABM and PDE for the second type of interacting agents (type 2). The parameters and the initial condition are as in Fig. 4 apart from $v = 5$ and $d = 0.8$. The new values of the parameters v and d are chosen in order to highlight the following phenomenon of the ABM. The results reveal a substantial difference between the behavior of the discrete model, in which regular spikes appear clearly in the density profile, and the behavior of the continuous model, whose density profile appears as a smooth function. In other words, the PDE provides correct information on the average number of agents in an interval of length v , but it fails to reveal the behavior of the ABM at smaller scales. Notice

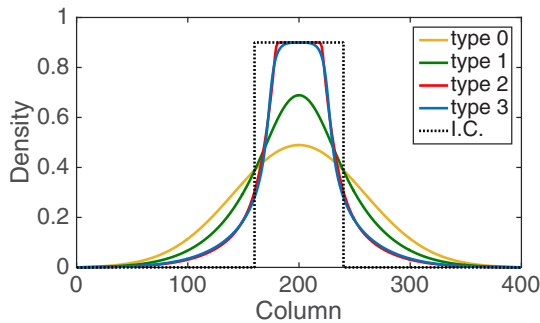


FIG. 5. Comparison of the numerical solutions of the column-averaged PDEs for the four types of interactions, types 0–3. The black dotted line illustrates the initial condition.

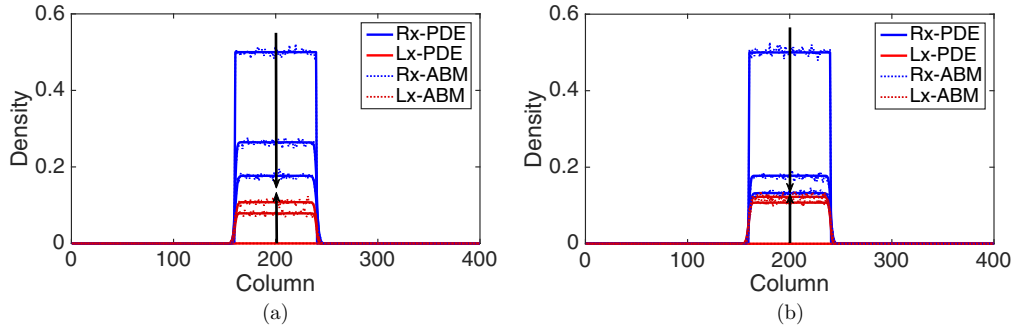


FIG. 6. Comparison between ABMs (dotted line) and PDEs (solid line) of right (blue) and left (red) subpopulations of non-interacting agents (type 0) for large values of reorienting rate P_r : (a) $P_r = 1$ and (b) $P_r = 2$. The simulations are initialized with the central region populated with density $d = 0.5$ by only right-polarized cells. The solutions are displayed at the times $T = 0, 1$, and 2 , with the direction of the black arrows indicating increasing time. The numerical solutions of the PDEs are obtained as described in the text.

that such discrepancy is not due to the stochasticity of the ABM but is a systematic feature of the model at the agent level. In order to gain an intuition of how the spikes appear in the ABM, notice that the agents which first leave the initial region (the region of the domain in which agents are initiated) by making a long jump of the maximum distance, $v = 5$, form an effective barrier for the following agents. Subsequent agents leaving the initial region accumulate behind this barrier. This mechanism repeats itself as the furthest agents jump again, producing a second effective barrier at distance $\Delta v = 5$. This mechanism produces a density profile characterized by multiple spikes at distances which are multiples of v . The amplitude of the spikes decreases with distance from the initial condition as some of the barrier agents or their successors change orientation or become less synchronized in their outward movements, leading to successively more porous barriers. The same behavior is also observed in the other form of complex interaction (type 3, results not shown). This departure from the PDE model becomes more evident as v increases.

B. Anisotropy

One of the key features of all of our models, for large value of persistence, is the appearance of (positive) anisotropy in the axial directions [35]. In Fig. 8 we show the two-dimensional density contour lines $C_{i,j}^T = 0.01$ of the ABMs (averaged over

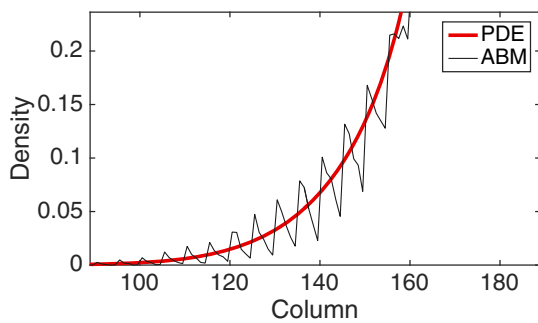


FIG. 7. A zoomed-in comparison between the column-averaged ABM (thin black) and the PDE (thick red) for $\bar{C}(x,t)$ for the second type of non-trivial interaction (type 2). Density spikes are clearly visible in the ABM, but not in the PDE.

$M = 1000$ realizations) evolve as time increases. All the ABM simulations are initialized by populating a central circular region of diameter $r = 40$ (i.e., the sites whose centers lie less than 40Δ away from the center of the domain). The contour density lines are recorded at times $T = 50, 100, 200$, and 300 . In all the examples we choose $P_m = 1$, $P_r = 0.01$, and $v = 1$. Notice that with this choice of parameters the three types of volume-excluding interactions (types 1–3) are equivalent. We repeat the simulations for non-interacting agents [Figs. 8(a) and 8(b)] and for interacting agents [Figs. 8(c) and 8(d)]. For each scenario we consider the non-persistent case, $\varphi = 0$ [Figs. 8(a) and 8(c)], and the strongly persistent case, $\varphi = 0.8$ [Figs. 8(b) and 8(d)]. When persistence is not included, the dynamics of the agents correspond to simple excluding walks. In this case the isotropy of the initial condition is known to be preserved as the system evolves [Figs. 8(a) and 8(c)] [10,15,20]. In particular, the density profiles conserve the circular shape of the initial region, meaning that there is no preferential direction of migration. When the persistence is switched on [Figs. 8(b) and 8(d)], the isotropy is lost. Cells spread faster in the four axial directions (red arrows), due to their polarizations, and this leads to density contour lines with a “diamond” shape.

It should be noticed that this phenomenon is not produced by the mechanism of jumping multiple lattice steps simultaneously; in fact we deliberately chose $v = 1$ to illustrate this. The anisotropy is, instead, an intrinsic feature of the persistent model combined with the lattice environment. Such anisotropic behavior has not been observed in previous studies, which focused on the one-dimensional scenario [22,23], because the higher-dimensional setup is a necessary condition for the anisotropy to appear [31].

C. Spontaneous aggregation

The last phenomenon that we highlight here is the emergence of short-term aggregation in density profiles driven by the interplay of persistence and volume exclusion. In Figs. 9(a) and 9(b) we show the total density of the column-averaged PDE and ABM, respectively, for the model with agent interaction (types 1–3) and parameters $P_m = 1$, $P_r = 0.01$, $\varphi = 0.9$, and $v = 1$. Such a choice of parameters leads to strong persistence. This is needed in order to make the aggregation phenomenon evident. Additionally, we increased the number of repeats,

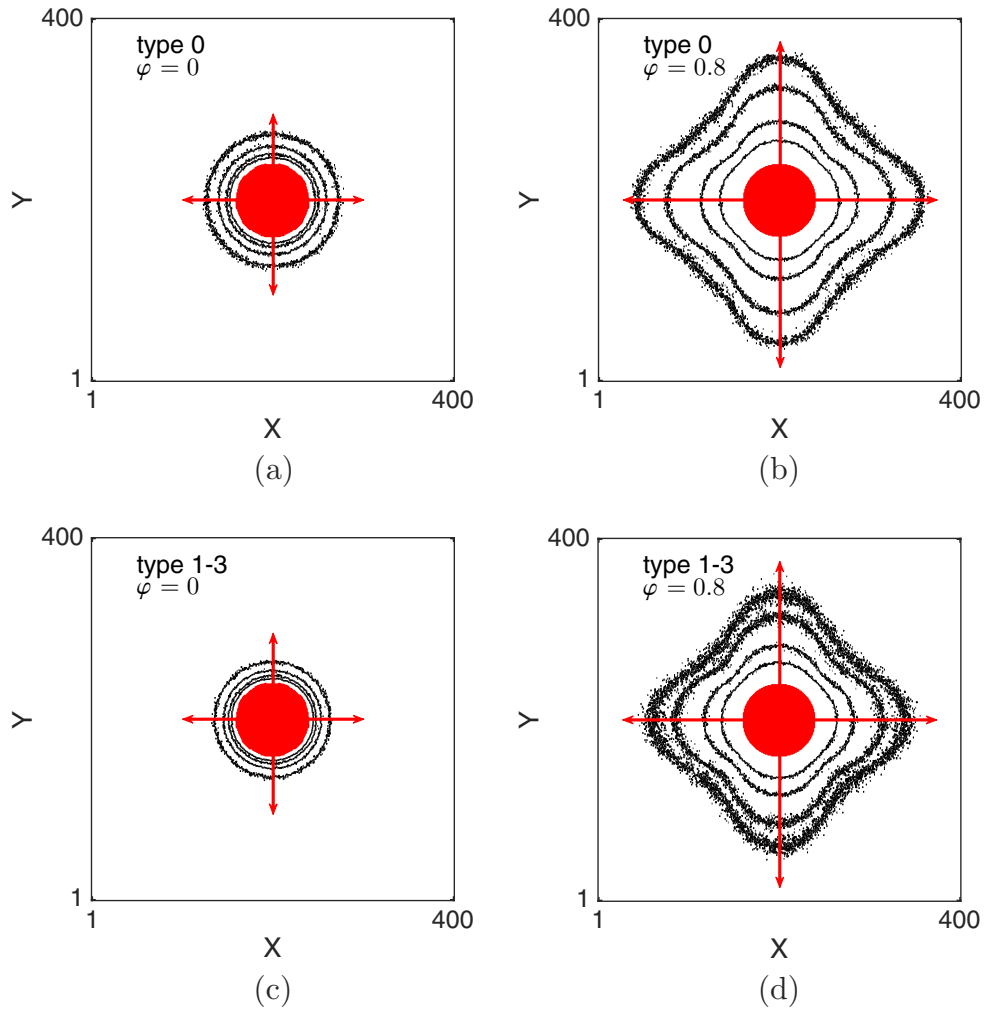


FIG. 8. Anisotropy of the ABM in two dimensions. The black contour lines represent values of equal total density, $C_{i,j}^T = 0.01$, at times $T = 50, 100, 200$, and 300 with time increasing in the direction of the red lines. Non-interacting agent profiles (type 0) (a) without persistence of motion, $\varphi = 0$, and (b) with strong persistence of motion, $\varphi = 0.8$. Agent profiles for the non-trivial agent interactions (types 1–3) (c) without persistence, $\varphi = 0$, and (d) with strong persistence, $\varphi = 0.8$.

to $M = 100$, to reduce the noise and better demonstrate the phenomenon in the ABM averaged profiles. The profiles are shown at times $T = 0, 50, 100, 200$, and 300 . The plots show a snapshot of the system in which total density towards the edge of the initially populated region increases above the initial value, producing a spontaneous non-monotonicity (traveling outwards from the center of the domain in either direction) in the density profile towards the edge of the initial interval. A closer look at the density for right- and left-polarized agents at time $T = 50$ [see Fig. 9(c)] reveals that the increment on the left-hand side of the initial condition is caused by an accumulation of right-polarized cells and vice versa for the other side. Although in the early stages ($T = 50$) the accumulation is visible only in the profiles of the differently polarized populations [see Fig. 9(c)], eventually ($T = 200$) the non-monotonicity appears at the total population level [see Figs. 9(a) and 9(b)].

In order to explain this phenomenon of *spontaneous aggregation* at the microscopic level, we consider a scenario with initially high density and high values of persistence ($P_r \ll 1$

and $\varphi \approx 1$). In Fig. 9(d) we display a portion of a single ABM simulation magnified in the region around the formation of the left peak. We partition the figure into three regions: *external*, which corresponds to the region outside the initially populated region and that is empty at time $t = 0$; *peripheral*, which represents the region containing the border of the initially populated region and where the aggregation takes place; and *internal*, which represents the center of the initially populated region. In the first phase of the simulation, the right-polarized cells in the peripheral region are blocked by the high density in the internal region and they are likely to remain in their position unless they reorient, which happens with low probability. Meanwhile, some of the other polarized cells, which occupy the peripheral region, spread into the external region on the left-hand side which creates a decrease in the total density of the peripheral region. This allows the right-polarized cells in the peripheral region to move rightwards, further into the peripheral region, to aggregate and hence to form a barrier for the cells in the internal region [see blue squares in the proximity of the dotted yellow line in Fig. 9(d), for example]. The agents in the internal

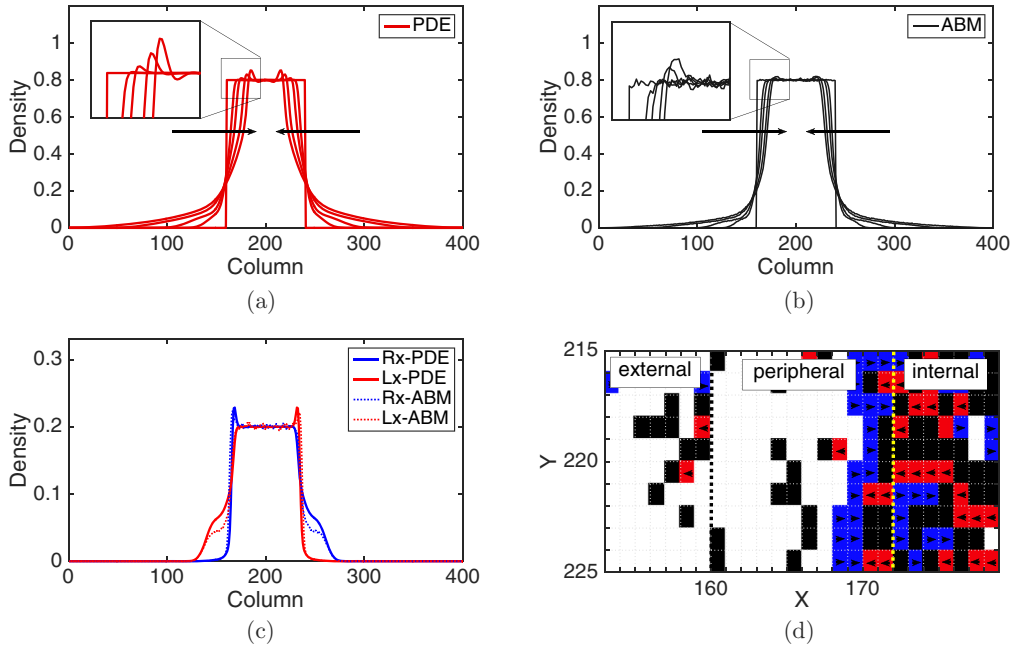


FIG. 9. Spontaneous aggregation induced by persistence and volume exclusion. (a) The numerical solution of the PDE for the column-averaged total density and (b) the column-averaged density of the ABM, averaged over $M = 100$ repeats, for the model with interacting agents (types 1–3 with $v = 1$). The profiles are shown at time $T = 0, 50, 100, 200$, and 300 with the direction of the black arrows indicating increasing time. (c) The solutions of the PDE (continuous lines) and ABM (dotted lines) are compared for the partial column-averaged densities of right-polarized agents (blue lines) and left-polarized agents (red lines) at time $T = 50$. (d) A zoomed-in snapshot of a single simulation of the ABM (types 1–3) is shown at time $T = 200$. White sites are empty, blue sites are occupied by a right-polarized agent, red sites are occupied by a left-polarized agent, and black sites are occupied by either an up- or down-polarized agent. The dotted black line represents the border of the initially populated region which divides the external region from the peripheral region, and the dotted yellow line distinguishes between the peripheral and the internal region.

region remain trapped by this obstruction. Notably, we can also see a weak form of aggregation on the internal sides of the two barriers in the PDE density profiles in Fig. 9(c). This is due to a similar mechanism that occurs in the internal region; the high density making the whole process slower and resulting in a weaker aggregate. The noisiness of the data for the ABM makes it difficult to see such a weak aggregation. As time evolves, more agents in the internal region escape the two barriers (and some of the cells forming the barriers reorient) and reach the external region. The two barriers slowly move towards the center and eventually coalesce.

The aggregation phenomenon appears in both the PDE and the ABM [Figs. 9(a) and 9(b)]. However, the high level of spatial correlation associated with the aggregation affects the quality of the agreement between the continuum and discrete models so the agreement is only qualitative and not quantitative.

Simpson *et al.* [5] observed a form of spontaneous non-monotonicity in the continuous description of their multi-species on-lattice ABM. The authors considered a population of two identical, but distinctly labeled, species of cells moving according to a simple excluding random walk. The two species are initially confined in two adjacent regions with different initial densities. As time evolves, a non-monotonicity appears in the continuum profile of the species with lower density as a consequence of the high density of the other species in the adjacent region. No clear evidence of such behavior is present in the corresponding ABM. Moreover, it should be noted that

any form of cell aggregation at the total population level is not possible in the model of Simpson *et al.* [5], since the overall behavior is governed by simple diffusion. Although the spontaneous formation of aggregates (jamming) has been described previously in models which incorporate persistence [24,35–39,51,52], to the best of our knowledge, non-monotonicity in average agent density has not been reported previously.

V. CONCLUSION

At the agent level, we have modeled persistence of motion for interacting agents through excluding velocity-jump processes. Traditionally, such processes are associated with systems of advective PDEs, which describe the model at a population scale. Our work continues the investigation of this type of model. First of all, we generalized the traditional velocity-jump model, which allowed us to derive a system of diffusive equations from the ABMs [Eqs. (5) and (B1)–(B3)]. Moreover, our observations reveal some unusual phenomena (density spikes, anisotropy, and aggregation) caused by the interplay of persistence and volume exclusion.

Despite the new diffusive PDEs correctly predicting the macroscopic behavior of the ABM for a wide range of parameters (including large values of reorienting rate), our findings highlight unrealistic behaviors for certain choices of parameters. Specifically, when the agent velocity is large and we implement one of the two more complex forms of

agent interaction (types 2 or 3), the density profile of the ABM presents regular peaks (spikes) in density which are not captured by the corresponding continuum models. We also give evidence of an inherent anisotropy that occurs when we implement persistence in an on-lattice context in two or more dimensions [35]. This phenomenon represents a problem when applying the model to experimental data, for which isotropy is usually a natural feature. One possibility for reducing the scale of this issue and still obtaining the macroscopic description would be to work on a hexagonal lattice. This would increase the number of preferential directions from four to six, making the anisotropy less evident although not completely removing it. Alternatively, one could allow cells to move in diagonal directions as in Ref. [35], which may also serve to mitigate, but not completely remove, anisotropy. Off-lattice models should have the advantage that they are not afflicted by anisotropy. However, the derivation of a corresponding macroscopic description becomes more complicated and sometimes intractable. Therefore, the problem of modeling persistence of motion in an on-lattice context at multiple scales, without incurring anisotropy in the lattice directions, remains an interesting challenge for future research.

Finally, the other main achievement of this work is that the new continuum approximation that we propose is capable of qualitatively reproducing the spontaneous aggregation driven by persistence and volume exclusion [24,35–39,51,52]. To our knowledge, this is the first time that such behavior in the ABM has been replicated at the macroscopic level. The non-intuitive consequence is that, in the case of strong persistence, the process of cell dispersion is initially slowed down by the aggregation phenomenon, which constrains some of the cells

in the internal region around the initial condition. Once the aggregates dissolve, the agents' dispersion is effectively faster than the normal diffusion. When such aggregation occurs, the agreement between the models at the two scales is qualitative rather than quantitative. More work might be done in order to recover a better agreement. In particular, this might be achieved by including a higher order of spatial correlations in the continuous model as done by Markham *et al.* [53,54]; however, this remains an open question.

ACKNOWLEDGMENT

The authors would like to thank the CMB/CNCB preprint club for constructive and helpful comments on a preprint of this paper.

APPENDIX A: OCCUPANCY MASTER EQUATIONS

In this section we report the occupancy master equations for the ABM described in Sec. II in the main text for the three types of non-trivial forms of agent interaction.

1. Type 0

The full set of occupancy master equations for agents moving according to type 0 interactions are reported in the main document [see system (3) in Sec. III of the main text].

2. Type 1

For type 1 agent interactions, which correspond to the case in which the movement is aborted if, and only if, the target site is occupied, the occupancy master equations read as follows:

$$\begin{aligned}
 R_{i,j}^{t+\tau} &= R_{i,j}^t + \frac{\tau P_m}{4} (1 - C_{i,j}^t) [(1 + \varphi) R_{i-v,j}^t + (1 - \varphi) R_{i+v,j}^t + R_{i,j+v}^t + R_{i,j-v}^t] \\
 &\quad - \frac{\tau P_m}{4} R_{i,j}^t [(1 + \varphi)(1 - C_{i+v,j}^t) + (1 - \varphi)(1 - C_{i-v,j}^t) + (1 - C_{i,j+v}^t) + (1 - C_{i,j-v}^t)] \\
 &\quad + \frac{\tau P_r}{4} [L_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3R_{i,j}^t] + \mathcal{O}(\tau^2), \\
 L_{i,j}^{t+\tau} &= L_{i,j}^t + \frac{\tau P_m}{4} (1 - C_{i,j}^t) [(1 - \varphi) L_{i-v,j}^t + (1 + \varphi) L_{i+v,j}^t + L_{i,j+v}^t + L_{i,j-v}^t] \\
 &\quad - \frac{\tau P_m}{4} L_{i,j}^t [(1 - \varphi)(1 - C_{i+v,j}^t) + (1 + \varphi)(1 - C_{i-v,j}^t) + (1 - C_{i,j+v}^t) + (1 - C_{i,j-v}^t)] \\
 &\quad + \frac{\tau P_r}{4} [R_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3L_{i,j}^t] + \mathcal{O}(\tau^2), \\
 U_{i,j}^{t+\tau} &= U_{i,j}^t + \frac{\tau P_m}{4} (1 - C_{i,j}^t) [(U_{i-v,j}^t + U_{i+v,j}^t + (1 - \varphi) U_{i,j+v}^t + (1 + \varphi) U_{i,j-v}^t] \\
 &\quad - \frac{\tau P_m}{4} U_{i,j}^t [(1 - C_{i+v,j}^t) + (1 - C_{i-v,j}^t) + (1 + \varphi)(1 - C_{i,j+v}^t) + (1 - \varphi)(1 - C_{i,j-v}^t)] \\
 &\quad + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + D_{i,j}^t - 3U_{i,j}^t] + \mathcal{O}(\tau^2), \\
 D_{i,j}^{t+\tau} &= D_{i,j}^t + \frac{\tau P_m}{4} (1 - C_{i,j}^t) [(D_{i-v,j}^t + D_{i+v,j}^t + (1 + \varphi) D_{i,j+v}^t + (1 - \varphi) D_{i,j-v}^t] \\
 &\quad - \frac{\tau P_m}{4} D_{i,j}^t [(1 - C_{i+v,j}^t) + (1 - C_{i-v,j}^t) + (1 - \varphi)(1 - C_{i,j+v}^t) + (1 + \varphi)(1 - C_{i,j-v}^t)] \\
 &\quad + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + U_{i,j}^t - 3D_{i,j}^t] + \mathcal{O}(\tau^2). \tag{A1}
 \end{aligned}$$

3. Type 2

Here we report the complete set of occupancy master equations for agents moving according to type 2 interactions. In this case agents move if, and only if, all the sites between the initial site and the target site are available. The equations are as follows:

$$\begin{aligned}
R_{i,j}^{t+\tau} &= R_{i,j}^t + \frac{\tau P_m}{4} \left[(1+\varphi) R_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + (1-\varphi) R_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) + R_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) \right. \\
&\quad \left. + R_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \right] - \frac{\tau P_m}{4} R_{i,j}^t \left[(1+\varphi) \prod_{s=1}^v (1 - C_{i+s,j}^t) + (1-\varphi) \prod_{s=1}^v (1 - C_{i-s,j}^t) + \prod_{s=1}^v (1 - C_{i,j+s}^t) \right. \\
&\quad \left. + \prod_{s=1}^v (1 - C_{i,j-s}^t) \right] + \frac{\tau P_r}{4} [L_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3R_{i,j}^t] + \mathcal{O}(\tau^2), \\
L_{i,j}^{t+\tau} &= L_{i,j}^t + \frac{\tau P_m}{4} \left[(1-\varphi) L_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + (1+\varphi) L_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) + L_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) \right. \\
&\quad \left. + L_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \right] - \frac{\tau P_m}{4} L_{i,j}^t \left[(1-\varphi) \prod_{s=1}^v (1 - C_{i+s,j}^t) + (1+\varphi) \prod_{s=1}^v (1 - C_{i-s,j}^t) + \prod_{s=1}^v (1 - C_{i,j+s}^t) \right. \\
&\quad \left. + \prod_{s=1}^v (1 - C_{i,j-s}^t) \right] + \frac{\tau P_r}{4} [R_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3L_{i,j}^t] + \mathcal{O}(\tau^2), \\
U_{i,j}^{t+\tau} &= U_{i,j}^t + \frac{\tau P_m}{4} \left[U_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + U_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) + (1-\varphi) U_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) \right. \\
&\quad \left. + (1+\varphi) U_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \right] - \frac{\tau P_m}{4} U_{i,j}^t \left[\prod_{s=1}^v (1 - C_{i+s,j}^t) \right. \\
&\quad \left. + \prod_{s=1}^v (1 - C_{i-s,j}^t) + (1+\varphi) \prod_{s=1}^v (1 - C_{i,j+s}^t) + (1-\varphi) \prod_{s=1}^v (1 - C_{i,j-s}^t) \right] \\
&\quad + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + D_{i,j}^t - 3U_{i,j}^t] + \mathcal{O}(\tau^2), \\
D_{i,j}^{t+\tau} &= D_{i,j}^t + \frac{\tau P_m}{4} \left[D_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + D_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) + (1+\varphi) D_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) \right. \\
&\quad \left. + (1-\varphi) D_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \right] - \frac{\tau P_m}{4} D_{i,j}^t \left[\prod_{s=1}^v (1 - C_{i+s,j}^t) \right. \\
&\quad \left. + \prod_{s=1}^v (1 - C_{i-s,j}^t) + (1-\varphi) \prod_{s=1}^v (1 - C_{i,j+s}^t) + (1+\varphi) \prod_{s=1}^v (1 - C_{i,j-s}^t) \right] \\
&\quad + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + U_{i,j}^t - 3D_{i,j}^t] + \mathcal{O}(\tau^2). \tag{A2}
\end{aligned}$$

4. Type 3

Finally, we report the occupancy master equations for agents moving according to type 3 interactions, in which the agents move to the furthest available site. These read

$$\begin{aligned}
R_{i,j}^{t+\tau} &= R_{i,j}^t + \frac{\tau P_m}{4} \left[(1+\varphi) R_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + (1-\varphi) R_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) \right. \\
&\quad \left. + R_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) + R_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \right] - \frac{\tau P_m}{4} R_{i,j}^t \left[(1+\varphi) \prod_{s=1}^v (1 - C_{i+s,j}^t) \right.
\end{aligned}$$

$$\begin{aligned}
& + (1 - \varphi) \prod_{s=1}^v (1 - C_{i-s,j}^t) + \prod_{s=1}^v (1 - C_{i,j+s}^t) + \prod_{s=1}^v (1 - C_{i,j-s}^t) \Big] \\
& + \frac{\tau P_m}{4} \left[(1 + \varphi) C_{i+1,j}^t \sum_{k=1}^{v-1} R_{i-k,j}^t \prod_{s=0}^{k-1} (1 - C_{i-s,j}^t) + (1 - \varphi) C_{i-1,j}^t \sum_{k=1}^{v-1} R_{i+k,j}^t \prod_{s=0}^{k-1} (1 - C_{i+s,j}^t) \right. \\
& + C_{i,j-1}^t \sum_{k=1}^{v-1} R_{i,j+k}^t \prod_{s=0}^{k-1} (1 - C_{i,j+s}^t) + C_{i,j+1}^t \sum_{k=1}^{v-1} R_{i,j-k}^t \prod_{s=0}^{k-1} (1 - C_{i,j-s}^t) \Big] \\
& - \frac{\tau P_m}{4} R_{i,j}^t \left[(1 + \varphi) \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i+k,j}^t (1 - C_{i+s,j}^t) + (1 - \varphi) \sum_{k=1}^v \prod_{s=1}^{k-1} C_{i-k,j}^t (1 - C_{i-s,j}^t) \right. \\
& + \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j+k}^t (1 - C_{i,j+s}^t) + \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j-k}^t (1 - C_{i,j-s}^t) \Big] \\
& + \frac{\tau P_r}{4} [L_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3R_{i,j}^t] + \mathcal{O}(\tau^2), \\
L_{i,j}^{t+\tau} = L_{i,j}^t & + \frac{\tau P_m}{4} \left[(1 - \varphi) L_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + (1 + \varphi) L_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) \right. \\
& + L_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) + L_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \Big] - \frac{\tau P_m}{4} L_{i,j}^t \left[(1 - \varphi) \prod_{s=1}^v (1 - C_{i+s,j}^t) \right. \\
& + \left((1 + \varphi) \prod_{s=1}^v (1 - C_{i-s,j}^t) + \prod_{s=1}^v (1 - C_{i,j+s}^t) + \prod_{s=1}^v (1 - C_{i,j-s}^t) \right) \\
& + \frac{\tau P_m}{4} \left[(1 - \varphi) C_{i+1,j}^t \sum_{k=1}^{v-1} L_{i-k,j}^t \prod_{s=0}^{k-1} (1 - C_{i-s,j}^t) + (1 + \varphi) C_{i-1,j}^t \sum_{k=1}^{v-1} L_{i+k,j}^t \prod_{s=0}^{k-1} (1 - C_{i+s,j}^t) \right. \\
& + C_{i,j-1}^t \sum_{k=1}^{v-1} L_{i,j+k}^t \prod_{s=0}^{k-1} (1 - C_{i,j+s}^t) + C_{i,j+1}^t \sum_{k=1}^{v-1} L_{i,j-k}^t \prod_{s=0}^{k-1} (1 - C_{i,j-s}^t) \Big] \\
& - \frac{\tau P_m}{4} L_{i,j}^t \left[(1 - \varphi) \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i+k,j}^t (1 - C_{i+s,j}^t) + (1 + \varphi) \sum_{k=1}^v \prod_{s=1}^{k-1} C_{i-k,j}^t (1 - C_{i-s,j}^t) \right. \\
& + \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j+k}^t (1 - C_{i,j+s}^t) + \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j-k}^t (1 - C_{i,j-s}^t) \Big] \\
& + \frac{\tau P_r}{4} [R_{i,j}^t + U_{i,j}^t + D_{i,j}^t - 3L_{i,j}^t] + \mathcal{O}(\tau^2), \\
U_{i,j}^{t+\tau} = U_{i,j}^t & + \frac{\tau P_m}{4} \left[U_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + U_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) + (1 - \varphi) U_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) \right. \\
& + (1 + \varphi) U_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \Big] - \frac{\tau P_m}{4} U_{i,j}^t \left[\prod_{s=1}^v (1 - C_{i+s,j}^t) \right. \\
& + \prod_{s=1}^v (1 - C_{i-s,j}^t) + (1 + \varphi) \prod_{s=1}^v (1 - C_{i,j+s}^t) + (1 - \varphi) \prod_{s=1}^v (1 - C_{i,j-s}^t) \Big] \\
& + \frac{\tau P_m}{4} \left[C_{i+1,j}^t \sum_{k=1}^{v-1} U_{i-k,j}^t \prod_{s=0}^{k-1} (1 - C_{i-s,j}^t) + C_{i-1,j}^t \sum_{k=1}^{v-1} U_{i+k,j}^t \prod_{s=0}^{k-1} (1 - C_{i+s,j}^t) \right. \\
& + (1 - \varphi) C_{i,j-1}^t \sum_{k=1}^{v-1} U_{i,j+k}^t \prod_{s=0}^{k-1} (1 - C_{i,j+s}^t) + (1 + \varphi) C_{i,j+1}^t \sum_{k=1}^{v-1} U_{i,j-k}^t \prod_{s=0}^{k-1} (1 - C_{i,j-s}^t) \Big]
\end{aligned}$$

$$\begin{aligned}
& -\frac{\tau P_m}{4} U_{i,j}^t \left[\sum_{k=2}^v \prod_{s=1}^{k-1} C_{i+k,j}^t (1 - C_{i+s,j}^t) + \sum_{k=1}^v \prod_{s=1}^{k-1} C_{i-k,j}^t (1 - C_{i-s,j}^t) \right. \\
& + (1 + \varphi) \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j+k}^t (1 - C_{i,j+s}^t) + (1 - \varphi) \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j-k}^t (1 - C_{i,j-s}^t) \left. \right] \\
& + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + D_{i,j}^t - 3U_{i,j}^t] + \mathcal{O}(\tau^2), \\
D_{i,j}^{t+\tau} = D_{i,j}^t & + \frac{\tau P_m}{4} \left[D_{i-v,j}^t \prod_{s=0}^{v-1} (1 - C_{i-s,j}^t) + D_{i+v,j}^t \prod_{s=0}^{v-1} (1 - C_{i+s,j}^t) + (1 + \varphi) D_{i,j+v}^t \prod_{s=0}^{v-1} (1 - C_{i,j+s}^t) \right. \\
& + (1 - \varphi) D_{i,j-v}^t \prod_{s=0}^{v-1} (1 - C_{i,j-s}^t) \left. \right] - \frac{\tau P_m}{4} D_{i,j}^t \left[\prod_{s=1}^v (1 - C_{i+s,j}^t) \right. \\
& + \prod_{s=1}^v (1 - C_{i-s,j}^t) + (1 - \varphi) \prod_{s=1}^v (1 - C_{i,j+s}^t) + (1 + \varphi) \prod_{s=1}^v (1 - C_{i,j-s}^t) \left. \right] \\
& + \frac{\tau P_m}{4} \left[C_{i+1,j}^t \sum_{k=1}^{v-1} D_{i-k,j}^t \prod_{s=0}^{k-1} (1 - C_{i-s,j}^t) + C_{i-1,j}^t \sum_{k=1}^{v-1} D_{i+k,j}^t \prod_{s=0}^{k-1} (1 - C_{i+s,j}^t) \right. \\
& + (1 + \varphi) C_{i,j-1}^t \sum_{k=1}^{v-1} D_{i,j+k}^t \prod_{s=0}^{k-1} (1 - C_{i,j+s}^t) + (1 - \varphi) C_{i,j+1}^t \sum_{k=1}^{v-1} D_{i,j-k}^t \prod_{s=0}^{k-1} (1 - C_{i,j-s}^t) \left. \right] \\
& - \frac{\tau P_m}{4} D_{i,j}^t \left[\sum_{k=2}^v \prod_{s=1}^{k-1} C_{i+k,j}^t (1 - C_{i+s,j}^t) + \sum_{k=1}^v \prod_{s=1}^{k-1} C_{i-k,j}^t (1 - C_{i-s,j}^t) \right. \\
& + (1 - \varphi) \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j+k}^t (1 - C_{i,j+s}^t) + (1 + \varphi) \sum_{k=2}^v \prod_{s=1}^{k-1} C_{i,j-k}^t (1 - C_{i,j-s}^t) \left. \right] \\
& + \frac{\tau P_r}{4} [R_{i,j}^t + L_{i,j}^t + U_{i,j}^t - 3D_{i,j}^t] + \mathcal{O}(\tau^2). \tag{A3}
\end{aligned}$$

APPENDIX B: COMPLETE DENSITY SYSTEMS

In this section we report the complete systems of PDEs for the four types of agent interaction.

1. Type 0

The full set of PDEs for the model with type 0 interactions is reported in the main text [see system (5) in Sec. III].

2. Type 1

For the type 1 form of interaction the system of diffusive PDEs for the four subpopulation reads as follows:

$$\begin{aligned}
\frac{\partial R}{\partial t} &= v^2 \mu [R \nabla^2 C + (1 - C) \nabla^2 R] - v \nu \frac{\partial}{\partial x} [R(1 - C)] + \frac{P_r}{4} (C - 4R), \\
\frac{\partial L}{\partial t} &= v^2 \mu [L \nabla^2 C + (1 - C) \nabla^2 L] + v \nu \frac{\partial}{\partial x} [L(1 - C)] + \frac{P_r}{4} (C - 4L), \\
\frac{\partial U}{\partial t} &= v^2 \mu [U \nabla^2 C + (1 - C) \nabla^2 U] - v \nu \frac{\partial}{\partial y} [U(1 - C)] + \frac{P_r}{4} (C - 4U), \\
\frac{\partial D}{\partial t} &= v^2 \mu [D \nabla^2 C + (1 - C) \nabla^2 D] + v \nu \frac{\partial}{\partial y} [D(1 - C)] + \frac{P_r}{4} (C - 4D), \tag{B1}
\end{aligned}$$

where μ and ν are defined as in Eqs. (6) and the boundary conditions are imposed as in Eqs. (21) of the main text.

3. Type 2

For the type 2 interaction, the complete set of PDEs is

$$\begin{aligned}\frac{\partial R}{\partial t} &= v^2 \mu [R \nabla((1-C)^{v-1} \nabla C) + (1-C) \nabla((1-C)^{v-1} \nabla R)] - v \nu \frac{\partial}{\partial x} [R(1-C)^v] + \frac{P_r}{4} (C-4R), \\ \frac{\partial L}{\partial t} &= v^2 \mu [L \nabla((1-C)^{v-1} \nabla C) + (1-C) \nabla((1-C)^{v-1} \nabla L)] + v \nu \frac{\partial}{\partial x} [L(1-C)^v] + \frac{P_r}{4} (C-4L), \\ \frac{\partial U}{\partial t} &= v^2 \mu [U \nabla((1-C)^{v-1} \nabla C) + (1-C) \nabla((1-C)^{v-1} \nabla U)] - v \nu \frac{\partial}{\partial y} [U(1-C)^v] + \frac{P_r}{4} (C-4U), \\ \frac{\partial D}{\partial t} &= v^2 \mu [D \nabla((1-C)^{v-1} \nabla C) + (1-C) \nabla((1-C)^{v-1} \nabla D)] + v \nu \frac{\partial}{\partial y} [D(1-C)^v] + \frac{P_r}{4} (C-4D),\end{aligned}\quad (\text{B2})$$

where μ and ν are defined as in Eqs. (6) and the boundary conditions are imposed as in Eqs. (21) of the main text.

4. Type 3

Finally, we report the complete diffusive system of the model with type 3 interaction. This reads

$$\begin{aligned}\frac{\partial R}{\partial t} &= \mu \nabla \left[\sum_{k=1}^v (1-C)^{k-1} [(2k-1)(1-C) \nabla R - k(k-2)R \nabla C] \right] + \nu \frac{\partial}{\partial x} \left[\frac{(1-C)((1-C)^v - 1)}{C} R \right] + \frac{P_r}{4} (C-4R), \\ \frac{\partial L}{\partial t} &= \mu \nabla \left[\sum_{k=1}^v (1-C)^{k-1} [(2k-1)(1-C) \nabla L - k(k-2)L \nabla C] \right] - \nu \frac{\partial}{\partial x} \left[\frac{(1-C)((1-C)^v - 1)}{C} L \right] + \frac{P_r}{4} (C-4L), \\ \frac{\partial U}{\partial t} &= \mu \nabla \left[\sum_{k=1}^v (1-C)^{k-1} [(2k-1)(1-C) \nabla U - k(k-2)U \nabla C] \right] + \nu \frac{\partial}{\partial y} \left[\frac{(1-C)((1-C)^v - 1)}{C} U \right] + \frac{P_r}{4} (C-4U), \\ \frac{\partial D}{\partial t} &= \mu \nabla \left[\sum_{k=1}^v (1-C)^{k-1} [(2k-1)(1-C) \nabla D - k(k-2)D \nabla C] \right] - \nu \frac{\partial}{\partial y} \left[\frac{(1-C)((1-C)^v - 1)}{C} D \right] + \frac{P_r}{4} (C-4D),\end{aligned}\quad (\text{B3})$$

where μ and ν are defined as in Eqs. (6) and the boundary conditions are imposed as in Eqs. (21) of the main text.

-
- [1] R. L. Mort, R. J. H. Ross, K. J. Hainey, O. J. Harrison, M. A. Keighren, G. Landini, R. E. Baker, K. J. Painter, I. J. Jackson, and C. A. Yates, Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model, *Nat. Commun.* **7**, 10288 (2016).
 - [2] P. K. Maini, D. L. S. McElwain, and D. I. Leavesley, Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells, *Tissue Eng.* **10**, 475 (2004).
 - [3] J. A. Sherratt and M. A. J. Chaplain, A new mathematical model for avascular tumour growth, *J. Math. Biol.* **43**, 291 (2001).
 - [4] M. J. Simpson, A. Merrifield, K. A. Landman, and B. D. Hughes, Simulating invasion with cellular automata: Connecting cell-scale and population-scale properties, *Phys. Rev. E* **76**, 021918 (2007).
 - [5] M. J. Simpson, K. A. Landman, and B. D. Hughes, Multi-species simple exclusion processes, *Phys. A* **388**, 399 (2009).
 - [6] S. T. Johnston, M. J. Simpson, D. L. S. McElwain, B. J. Binder, and J. V. Ross, Interpreting scratch assays using pair density dynamics and approximate Bayesian computation, *Opt. Biol.* **4**, 140097 (2014).
 - [7] R. J. H. Ross, R. E. Baker, A. Parker, M. J. Ford, R. L. Mort, and C. A. Yates, Using approximate Bayesian computation to quantify cell-cell adhesion parameters in a cell migratory process, *Syst. Biol. Appl.* **3**, 9 (2017).
 - [8] A. R. A. Anderson and M. A. J. Chaplain, Continuous and discrete mathematical models of tumor-induced angiogenesis, *Bull. Math. Biol.* **60**, 857 (1998).
 - [9] S. M. Peirce, E. J. Van Gieson, and T. C. Skalak, Multicellular simulation predicts microvascular patterning and in silico tissue assembly, *FASEB J.* **18**, 731 (2004).
 - [10] A. Deutsch and S. Dormann, *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Applications, and Analysis* (Springer Science & Business Media, New York, 2007).
 - [11] M. S. Alber, M. A. Kiskowski, J. A. Glazier, and Y. Jiang, in *Mathematical Systems Theory in Biology, Communications, Computation, and Finance*, edited by J. Rosenthal and D. S. Gilliam (Springer, New York, 2003), pp. 1–39.
 - [12] J. L. Segovia-Juarez, S. Ganguli, and D. Kirschner, Identifying control mechanisms of granuloma formation during M. tuberculosis infection using an agent-based model, *J. Theor. Biol.* **231**, 357 (2004).
 - [13] Z. Wang, J. D. Butner, R. Kerketta, V. Cristini, and T. S. Deisboeck, Simulating cancer growth with multiscale agent-based modeling, *Semin. Cancer Biol.* **30**, 70 (2015).
 - [14] M. J. Simpson, K. A. Landman, B. D. Hughes, and D. F. Newgreen, Looking inside an invasion wave of cells using continuum models: Proliferation is the key, *J. Theor. Biol.* **243**, 343 (2006).

- [15] J. D. Murray, *Mathematical Biology: I. An Introduction*, Vol. 17 (Springer Science & Business Media, New York, 2007).
- [16] S. M. Wise, J. S. Lowengrub, H. B. Frieboes, and V. Cristini, Three-dimensional multispecies nonlinear tumor growth—I: Model and numerical method, *J. Theor. Biol.* **253**, 524 (2008).
- [17] D. Noble, Opinion: The rise of computational biology, *Nat. Rev. Mol. Cell Biol.* **3**, 459 (2002).
- [18] R. E. Baker, C. A. Yates, and R. Erban, From microscopic to macroscopic descriptions of cell migration on growing domains, *Bull. Math. Biol.* **72**, 719 (2010).
- [19] C. A. Yates, R. E. Baker, R. Erban, and P. K. Maini, Going from microscopic to macroscopic on nonuniform growing domains, *Phys. Rev. E* **86**, 021921 (2012).
- [20] E. A. Codling, M. J. Plank, and S. Benhamou, Random walk models in biology, *J. R. Soc. Interface* **5**, 813 (2008).
- [21] H. C. Berg and D. A. Brown, Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking, *Nature (London)* **239**, 500 (1972).
- [22] K. K. Treloar, M. J. Simpson, and S. W. McCue, Velocity-jump models with crowding effects, *Phys. Rev. E* **84**, 061920 (2011).
- [23] K. K. Treloar, M. J. Simpson, and S. W. McCue, Velocity-jump processes with proliferation, *J. Phys. A: Math. Theor.* **46**, 015003 (2013).
- [24] A. B. Slowman, M. R. Evans, and R. A. Blythe, Jamming and Attraction of Interacting Run-and-Tumble Random Walkers, *Phys. Rev. Lett.* **116**, 218101 (2016).
- [25] H. G. Othmer and A. Stevens, Aggregation, blowup, and collapse: The ABC's of taxis in reinforced random walks, *SIAM J. Appl. Math.* **57**, 1044 (1997).
- [26] A. Stevens, A stochastic cellular automaton modeling gliding and aggregation of myxobacteria, *SIAM J. Appl. Math.* **61**, 172 (2000).
- [27] R. L. Hall, Amoeboid movement as a correlated walk, *J. Math. Biol.* **4**, 327 (1977).
- [28] M. H. Gail and C. W. Boone, The locomotion of mouse fibroblasts in tissue culture, *Biophys. J.* **10**, 980 (1970).
- [29] A. Wright, Y. H. Li, and C. Zhu, The differential effect of endothelial cell factors on in vitro motility of malignant and non-malignant cells, *Ann. Biomed. Eng.* **36**, 958 (2008).
- [30] C. S. Patlak, Random walk with persistence and external bias, *Bull. Math. Biophys.* **15**, 311 (1953).
- [31] H. G. Othmer, S. R. Dunbar, and W. Alt, Models of dispersal in biological systems, *J. Math. Biol.* **26**, 263 (1988).
- [32] H. G. Othmer and T. Hillen, The diffusion limit of transport equations derived from velocity-jump processes, *SIAM J. Appl. Math.* **61**, 751 (2000).
- [33] H. G. Othmer and T. Hillen, The diffusion limit of transport equations II: Chemotaxis equations, *SIAM J. Appl. Math.* **62**, 1222 (2002).
- [34] D. Campos, V. Méndez, and I. Llopis, Persistent random motion: Uncovering cell migration dynamics, *J. Theor. Biol.* **267**, 526 (2010).
- [35] A. G. Thompson, J. Tailleur, M. E. Cates, and R. A. Blythe, Lattice models of nonequilibrium bacterial dynamics, *J. Stat. Mech: Theory Exp.* (2011) P02029.
- [36] N. Sepúlveda and R. Soto, Coarsening and clustering in run-and-tumble dynamics with short-range exclusion, *Phys. Rev. E* **94**, 022603 (2016).
- [37] R. Soto and R. Golestanian, Run-and-tumble dynamics in a crowded environment: Persistent exclusion process for swimmers, *Phys. Rev. E* **89**, 012706 (2014).
- [38] F. Peruani, A. Deutsch, and M. Bär, Nonequilibrium clustering of self-propelled rods, *Phys. Rev. E* **74**, 030904 (2006).
- [39] G. S. Redner, M. F. Hagan, and A. Baskaran, Structure and Dynamics of a Phase-Separating Active Colloidal Fluid, *Phys. Rev. Lett.* **110**, 055701 (2013).
- [40] R. Grima, S. N. Yaliraki, and M. Barahona, Crowding-induced anisotropic transport modulates reaction kinetics in nanoscale porous media, *J. Phys. Chem. B* **114**, 5380 (2010).
- [41] S. Smith, C. Cianci, and R. Grima, Macromolecular crowding directs the motion of small molecules inside cells, *J. R. Soc. Interface* **14**, 20170047 (2017).
- [42] M. Kac, A stochastic model related to the telegrapher's equation, *Rocky Mt. J. Math.* **4**, 497 (1974).
- [43] S. Goldstein, On diffusion by discontinuous movements, and on the telegraph equation, *Q. J. Mech. Appl. Math.* **4**, 129 (1951).
- [44] G. Metzger, *Transmission Lines with Pulse Excitation* (Elsevier, New York, 2012).
- [45] K. Wolf, Y. I. Wu, Y. Liu, J. Geiger, E. Tam, C. Overall, M. S. Stack, and P. Friedl, Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion, *Nat. Cell Biol.* **9**, 893 (2007).
- [46] R. B. Bird, Transport phenomena, *Appl. Mech. Rev.* **55**, R1 (2002).
- [47] J. R. Welty, C. E. Wicks, G. Rorrer, and R. E. Wilson, *Fundamentals of Momentum, Heat, and Mass Transfer* (Wiley, New York, 2009).
- [48] U. S. Tim and S. Mostaghimi, Model for predicting virus movement through soils, *Groundwater* **29**, 251 (1991).
- [49] Y. Sim and C. V. Chrysikopoulos, Virus transport in unsaturated porous media, *Water Resour. Res.* **36**, 173 (2000).
- [50] D. Gillespie, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem* **81**, 2340 (1977).
- [51] D. Levis and L. Berthier, Clustering and heterogeneous dynamics in a kinetic Monte Carlo model of self-propelled hard disks, *Phys. Rev. E* **89**, 062301 (2014).
- [52] J. Bialké, H. Löwen, and T. Speck, Microscopic theory for the phase separation of self-propelled repulsive disks, *Europhys. Lett.* **103**, 30008 (2013).
- [53] D. C. Markham, M. J. Simpson, P. K. Maini, E. A. Gaffney, and R. E. Baker, Incorporating spatial correlations into multispecies mean-field models, *Phys. Rev. E* **88**, 052713 (2013).
- [54] D. C. Markham, M. J. Simpson, and R. E. Baker, Simplified method for including spatial correlations in mean-field approximations, *Phys. Rev. E* **87**, 062702 (2013).

3.2. Conclusions

This chapter reports my work on modelling cell persistence and its interplay with excluding interactions. The most important outcome of the paper presented is the derivation of an advection-diffusion PDE model for excluding velocity-jump processes. We highlighted some of the features of the microscopic model which are captured by the macroscopic representation and some which are not. In particular, we found that the PDE model is capable of qualitatively reproducing a spontaneous form of aggregation seen in the ABM.

Like all existing models, those presented in this chapter have some limitations. Some of these are highlighted and discussed in the paper, for example, the emergence of regular spikes in the density profiles and the intrinsic anisotropy of the lattice. Another, probably overlooked, aspect of our analysis is the scaling assumption on the persistence parameter φ , precisely $\varphi \sim \mathcal{O}(\Delta)$. Despite the fact that this step limit the applicability of the analysis, it represents a fundamental point for the entire paper and a crucial assumption for the derivation of the advection-diffusion PDE models of Section III. More details on this type of scaling assumption in the context of bias random walkers can be found in Simpson et al. [2009].

The results of this paper leave a set of unanswered questions about the possibility of modelling and analysing the emergence of the macroscopic behaviours in excluding velocity-jump processes. For example, in a recent work Zhang et al. [2019] have continued the study of the model presented in this paper by introducing global bias and multiple species populations. In Zhang et al. [2019] reorientation and motility events are coupled. Their work represents an important extension of our paper and an advancement in the understanding of these modelling techniques. We will further discuss the work of Zhang et al. [2019] in Section 4.2, in the context of emerging spatial correlation and pair correlation functions.

An important aspect of using mathematical modelling to study cell behaviour, such as directional persistence, is the choice of appropriate statistics to quantify the relevance of the behaviour in an experimental scenario. Several spatial statistics have been suggested to quantify directional persistence [Codling et al., 2008]; among these, the most relevant are Mean Dispersal Distance (MDD), Mean Square Displacement (MSD) and Tortuosity. The characteristic features of these statistics for non-interacting velocity-jump processes have been extensively studied for the last few decades [Othmer et al., 1988, Codling and Hill, 2005]. However, for models which account for particle interactions, as for example volume exclusion, similar studies have started to appear only in recent years and more work has to be done in this direction [Galanti et al., 2013, Bertrand et al., 2018]. As part of the study presented in this chapter, therefore, it would be interesting to investigate how the interplay between volume exclusion and directional persistence can influence these statistics. Some progress in this direction

has been made by Teomy and Metzler [2019b]. In this work the authors use a modified version of our persistence model to study the MSD in one, and two, dimensions for persistent excluding agents. Interestingly, the authors explore the behaviour of the model even in an *anti-persistent* regime, *i.e.* when agents are more likely to change their direction of motion at each step (corresponding to $\rho < 0$, in our model) [Teomy and Metzler, 2019a].

Finally, an important question raised in this chapter surrounds the role and the extent of the aggregation phenomenon observed in Section IV-C of the paper. In other words, whether it is possible to quantify and to predict the emergence of this form of aggregation from the parameters of the model. Motivated by this question, I focused part my research on designing appropriate methods for identifying and quantifying spatial correlation in multi-agent systems, which is the main topic of the next chapter.

Chapter 4

Pair correlation functions for identifying spatial correlation in discrete domains

This chapter is entirely dedicated to the problem of detecting and quantifying spatial correlation in discrete domains. It comprises a paper published in Physical Review E [Gavagnin et al., 2018] and the accompanying supplementary materials.

In a model of spatially interacting agents, the positions of the single individuals are often not independent of each other. Depending on the type of interactions and on the individual behaviour, the positions of the agents can be positively or negatively correlated. Studying the intensity of these correlations is typically an informative tool for the understanding of the model and for applications to experimental data. The spontaneous form of aggregation which we observe in Chapter 3 (Section IV-C) is a good example of a phenomenon which can be studied from the perspective of spatial correlations. With this in mind, we investigated statistical methods capable of quantifying the level of correlation in spatially discrete systems. In this paper, we design a series of novel pairwise correlation functions (PCFs), designed for multi-agent systems in discrete domains. Our functions improve previously existing PCFs and are extended to different metrics and higher dimensions.

4.1. Outline of the article

Section I of the paper provides a brief overview of the importance of studying spatial correlation and a summary of the relevant literature. In Section II of the paper we have a closer look at existing PCFs for discrete domains and we discuss their limitations. In Section III we introduce two new PCFs for square lattices accounting for different distance metrics and boundary conditions. We test the new PCFs by using examples of patterns taken from the biological scenarios and we present the results in Section IV. Extensions to other types of lattice and more general discrete domains are discussed in Sections V and VI, respectively. We conclude with a short discussion in Section VII. The accompanying supplementary material contains some derivations, omitted in the main text, and a summary of the normalisations of all the PCFs defined in the paper.

Appendix B: Statement of Authorship

This declaration concerns the article entitled:									
Pair correlation functions for identifying spatial correlation in discrete domains									
Publication status (tick one)									
draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input checked="" type="checkbox"/>
Publication details (reference)	Journal: Physical Review E 97, 062104 Authors: Enrico Gavagnin, Jennifer P. Owen and Christian A. Yates								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The first two authors contributed equally to the idea formulation, all the calculations and simulations. (50%). All authors contributed equally to the presentation of the content (33%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed							Date	27.4.2020	

Pair correlation functions for identifying spatial correlation in discrete domains

Enrico Gavagnin,^{*,†} Jennifer P. Owen,[‡] and Christian A. Yates

Centre for Mathematical Biology, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom



(Received 12 January 2018; published 4 June 2018)

Identifying and quantifying spatial correlation are important aspects of studying the collective behavior of multiagent systems. Pair correlation functions (PCFs) are powerful statistical tools that can provide qualitative and quantitative information about correlation between pairs of agents. Despite the numerous PCFs defined for off-lattice domains, only a few recent studies have considered a PCF for discrete domains. Our work extends the study of spatial correlation in discrete domains by defining a new set of PCFs using two natural and intuitive definitions of distance for a square lattice: the taxicab and uniform metric. We show how these PCFs improve upon previous attempts and compare between the quantitative data acquired. We also extend our definitions of the PCF to other types of regular tessellation that have not been studied before, including hexagonal, triangular, and cuboidal. Finally, we provide a comprehensive PCF for any tessellation and metric, allowing investigation of spatial correlation in irregular lattices for which recognizing correlation is less intuitive.

DOI: [10.1103/PhysRevE.97.062104](https://doi.org/10.1103/PhysRevE.97.062104)

I. INTRODUCTION

A system of agents is considered in a state of spatial correlation if, given any agent in the system, the likelihood that there are other agents at a certain, close distance, is either increased or decreased with respect to the situation in which the agents are distributed uniformly at random. Spatial correlation is a dominant feature of many biological and physical systems [1–12]. For example, in cell biology, spatial correlation can be seen in the form of patterns on animal fur or fish skin [1,3,13]. In a clinical setting, cell aggregation is a characteristic feature of melanoma and its identification is essential for early diagnosis and effective therapy [14,15]. Resource competition in ecology can lead to spatial correlation in the form of segregation, for example, in ant nest displacement in a competitive environment [16]. In epidemiology, spatial correlation can be observed in the occurrence of disease across different geographical regions [17].

The same spatial configuration can have different origins. For example, spatial aggregation in cell biology can be caused by a result of cell-to-cell adhesion [8], external signals, as in *chemotaxis* [5,6], or even slime following [7]. Alternatively, cells may form clusters during development due to a combination of a high proliferation rate and a low movement rate [9]. Given a system exhibiting spatial correlation, one may hypothesize an underlying mechanism responsible for these properties. These assumptions may form the basis of a mathematical model that can be simulated for the purpose of testing. Quantifying spatial correlation in both the simulation and observed experimental data can be a way to connect these studies and to validate or disprove such a theory. As a result, a great number of statistical tools have been developed in the past

decade to analyze and measure spatial correlation [12,18–23]. Among the most popular are pair correlation functions (PCFs) [11,12,20–22,24–28] and the fast Fourier transform (FFT) [29,30]. In this paper we focus our attention on the study of spatial correlation using PCFs.

Given a system of agents, a PCF determines whether pairs of agents are more or less likely to be found with a given separation than in the situation in which the agents are positioned uniformly at random in the domain. A PCF is considered effective if it fits two main criteria. First, the PCF distance metric should be well-defined, but most importantly be readily interpretable in the context of the system considered. This criteria is essential so that in the case of correlation (aggregation or segregation), the PCF can be used to obtain more details about the spatial configuration. For example, if the system exhibits aggregation, the PCF should be able to provide a measure for the average size of the clusters and their pairwise separation. Second, the PCF should be correctly calibrated. The PCF should be able to distinguish between three basic types of configurations: spatial randomness, aggregation, and segregation. For this, the PCF should be normalized correctly; i.e., the PCF should return the value unity at all pairwise distances (no correlation) when applied to a uniformly distributed set of agents. If the PCF is not normalized correctly, a spatially random set of agents may be incorrectly identified as a correlated system. This inconsistency makes PCF profiles hard to interpret.

Depending on the type of investigation, the mathematical framework can either be continuous (off-lattice) or discrete (on-lattice). The corresponding PCF has to be defined in accordance with the given framework. Despite the abundance of PCFs defined for off-lattice domains [11,26–28], only a few recent studies have defined a PCF for domains partitioned with a lattice [12,20–22]. On-lattice PCFs often assume exclusion properties, that is, that each lattice site in a domain can be occupied by at most one agent at any given time. This is consistent with typical on-lattice correlation studies, such as

*e.gavagnin@bath.ac.uk

[†]These authors contributed equally to this work.

[‡]j.owen@bath.ac.uk

those designed to quantify correlation in binary pixelated images, or to determine spatial correlation in exclusion processes simulated using a discrete domain.

Currently, there are two PCFs defined on-lattice. The first is a naive approach consisting of applying the classic off-lattice PCF to lattice-based systems. We refer to this from now on as the *annular* PCF. In the annular PCF, given some small positive δ , the number of agents at a distance m from a focal agent is defined as the number of agents whose centres lie in the annulus $(m - \delta, m]$, where distance is defined by the Euclidean metric. A limitation of this method is that, while the normalization is a good approximation for a continuous domain (see Sec. II for more details), it is poor in the case of a discrete domain, thus the PCF is not correctly calibrated. In more recent work, Binder and Simpson [12,21] defined a PCF specifically designed for a two-dimensional square on-lattice exclusion process, which we will refer to as the *rectilinear* PCF (see Sec. II for more details). While their approach correctly identifies the spatial correlation in many examples, due to an anisotropic definition of distance, spatial structures that are biased in either Cartesian directions can remain unidentified by this PCF. To summarize, to the best of our knowledge, a discrete isotropic PCF with correct normalization does not currently exist in the literature.

In this paper, we extend the study of pairwise spatial correlation for on-lattice exclusion processes which tackles the flaws of previous PCFs. We define isotropic PCFs for a square lattice on which distance is defined using two of the most natural and intuitive metrics for a discrete domain: the taxicab and uniform metric. We call these the *square taxicab* PCF and *square uniform* PCF, respectively, after the square lattice set up and metric type. We define them in both the nonperiodic and periodic boundary cases. Using synthetically generated data, we demonstrate that our PCF can correctly distinguish between spatial randomness, aggregation, and segregation. Furthermore, we show that it can also provide quantitative information about the structure of the system, such as approximate aggregate size or segregation distance both in the short and long scales. Moreover, we investigate how the choice of metric, uniform, or taxicab can affect this quantitative information. We demonstrate that our PCFs represent a significant improvement on previous on-lattice PCFs by showing that, first, our method is correctly calibrated (unlike the annular PCF) and, second, that it can identify anisotropic patterns of the type that are routinely missed by the rectilinear PCF. As a natural extension, we define PCFs for higher dimensions and other types of tessellations (cubic, triangular, and hexagonal) that have not been considered previously. We name these the *triangle* PCF, the *hexagon* PCF, the *cube taxicab* PCF, and the *cube uniform* PCF after the lattice set up and metric type.

Finally, we extend the concept of a PCF by introducing the *general* PCF. This PCF can be defined using any metric on any discrete domain type, with the caveat that it is more computationally expensive. We give an example of how we can use this PCF on a discrete irregular lattice (both tessellation and domain shape), where we define adjacent sites to be at unit distance from one another. We show how our PCF can identify aggregation and segregation on an irregular domain using some synthetic examples.

The paper is organized as follows. In Sec. II we discuss the successes and limitations of previous on-lattice PCFs. In

Sec. III we introduce our square taxicab and square uniform PCFs. We apply our square taxicab and square uniform PCFs to some relevant examples and make comparisons with previous on-lattice PCFs from the literature in Sec. IV. In Sec. V A we define the triangle PCF, the hexagon PCF, the cube taxicab PCF, and the cube uniform PCF. We extend our PCF to more generic and possibly irregular lattices by defining the general PCF in Sec. VI. For reference, in Sec. S.4 of the Supplemental Material, we supply a table summarizing all the formulae for the normalizations of our PCFs [31]. Finally, we conclude in Sec. VII by summarising the relevance of our results and discussing potential avenues for future work.

II. EXISTING ON-LATTICE PAIR CORRELATION FUNCTIONS

In this section we provide a summary of the only two existing PCFs defined for discrete domains: the annular PCF and the rectilinear PCF. For each, we describe their strengths and limitations.

First, consider a system of agents on a two-dimensional square lattice of size $L_x \times L_y$, with lattice step Δ , and with the exclusion property that, at any given time, each lattice site can be occupied by at most one agent. If N agents occupy the domain, then the occupancy of the lattice can be represented by a matrix M :

$$M_{xy} = \begin{cases} 0 & \text{if lattice site } (x, y) \text{ is vacant,} \\ 1 & \text{if lattice site } (x, y) \text{ is occupied,} \end{cases} \quad (1)$$

where

$$N = \sum_{x=1}^{L_x} \sum_{y=1}^{L_y} M_{xy} \leq L_x L_y. \quad (2)$$

Let ψ^M be the set of all agent pairs in the lattice defined by matrix M , i.e.,

$$\psi^M = \{(\mathbf{a}, \mathbf{b}) \in \mathbb{L} \times \mathbb{L} \mid \mathbf{a} = (x_a, y_a), \mathbf{b} = (x_b, y_b), \mathbf{a} \neq \mathbf{b}, M_{x_a, y_a} = M_{x_b, y_b} = 1\}, \quad (3)$$

where $\mathbb{L} = \{1, \dots, L_x\} \times \{1, \dots, L_y\}$ is the set of all sites in the lattice. With agents in configuration M , let us define the subset of agent pairs separated by distance m according to some (as yet unspecified) definition of distance, denoted by d , as

$$C_d(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi^M \mid \|\mathbf{a} - \mathbf{b}\|_d = m\}, \text{ for } m \in \mathcal{D}_d, \quad (4)$$

where \mathcal{D}_d is the set of possible distances under the metric d . We define the total number of pairs of agents for each value of distance $m \in \mathcal{D}_d$ as

$$c_d(m) = |C_d(m)|. \quad (5)$$

Similarly, we define the set of pairs of sites (regardless of their occupancy) which are separated by distance m according to the metric d as

$$S_d(m) = \{(\mathbf{a}, \mathbf{b}) \in \mathbb{L} \times \mathbb{L} \mid \|\mathbf{a} - \mathbf{b}\|_d = m\}, \text{ for } m \in \mathcal{D}_d, \quad (6)$$

hence the total number of pairs of sites at distance m is given by

$$s_d(m) = |S_d(m)|. \quad (7)$$

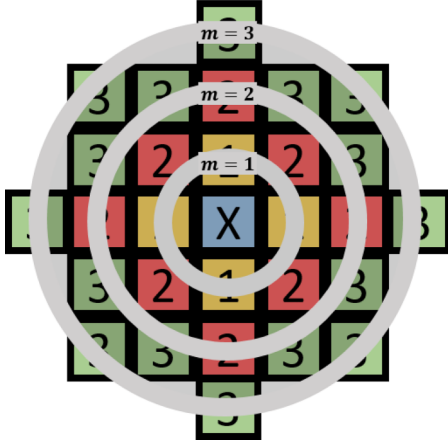


FIG. 1. Example agent pairs in the sets $C_A(m)$ with $m = 1, 2, 3$ and bandwidth $\delta = 1$. Concentric annuli, $(m - \delta, m]$, are superposed on top of the square lattice and the sites whose centres fall into each annulus are colored differently. Sites in yellow, red, and green (labeled 1, 2, and 3, respectively) are defined to be distance one, two, and three from the blue site, respectively, labeled X.

To produce a PCF, we aim to normalize the index $c_d(m)$ with the number of pairs we would expect at distance m if the system had no spatial correlation. That is, we consider the case in which the same number of agents are displaced uniformly at random on the lattice and compute the expected number of pairs at distance m . Let U be a random matrix such that $U_{xy} = 0$ for all sites $(x, y) \in \mathbb{L}$, apart from N sites chosen uniformly at random without replacement, for which $U_{xy} = 1$. Then we define

$$\bar{C}_d(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi^U \mid \|\mathbf{a} - \mathbf{b}\|_d = m\}, \quad \text{for } m \in \mathcal{D}_d. \quad (8)$$

Hence, for each value of $m \in \mathcal{D}_d$, the PCF at distance m is defined as

$$f_d(m) := \frac{c_d(m)}{\mathbb{E}[\bar{c}_d(m)]}, \quad (9)$$

where $\bar{c}_d(m) = |\bar{C}_d(m)|$ and \mathbb{E} represents the expectation operator.

A. The annular PCF

The annular PCF was originally designed for two-dimensional off-lattice systems [19] but can be extended to on-lattice systems with periodic boundary conditions (BC). For the annular PCF, the set C_A , where A denotes the annular metric, is defined as follows:

$$C_A(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi^M \mid \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \in (m - \delta, m]\}, \quad (10)$$

where $m \in \{\delta k \mid k \in \mathbb{N}^+\}$ and δ is a small real number which determines the bandwidth of the PCF. The schematics in Fig. 1 show a representation of some elements in the sets $C_A(1)$, $C_A(2)$, and $C_A(3)$ with $\delta = 1$.

The normalization factor is given by

$$\mathbb{E}[\bar{c}_A(m)] \approx \frac{N(N-1)(2\pi m\delta)}{L_x L_y}, \quad (11)$$

where $2\pi m\delta$ approximates the area of the m th annulus (assuming small δ) from any given agent. The annular PCF, f_A , follows from definition Eq. (9).

The normalization in Eq. (11) is a good approximation for a continuous domain with a small δ . However, when the agents are positioned on a lattice, this approach is no longer appropriate. The main issue is that the counts of agents in each annulus vary in an unpredictable manner with the distance, m , and the annular width, δ . For example, consider a square lattice with spacing Δ . The only possible distances two agents can be separated by are in the countable set

$$\begin{aligned} \mathcal{D}_A &= \{\Delta\sqrt{x^2 + y^2} \mid (x, y) \in \mathbb{N}^2 \setminus \{0, 0\}\} \\ &= \{\Delta, \sqrt{2}\Delta, 2\Delta, \sqrt{5}\Delta, \dots\}. \end{aligned} \quad (12)$$

Partitioning these distances into regularly spaced intervals, as it is required by the Euclidean distance metric, we can see that the number of agent pairs does not increase smoothly with the distance, m . Depending on the value of δ it may not even increase monotonically. However, the definition of the normalization factor Eq. (11) suggests that the expected number of pairs increases smoothly and monotonically with both m and δ . This disparity means the on-lattice annular PCF will not be properly normalized and will either be an over- or under- approximation, making results hard to interpret (see Fig. 7(b) as an example).

B. The Rectilinear PCF

In more recent work, Binder and Simpson [12, 21] define the Rectilinear PCF specifically for two-dimensional, on-lattice exclusion processes with nonperiodic BC. Their definition is easily extendible to periodic BC. They define two PCFs for the two Cartesian directions. In each case the distance is defined by the number of columns (or rows) separating two agents.

Thus, the set of pairs of agents separated by integer distance $m \in \mathbb{N}^+$ are defined in the x direction and y direction respectively as

$$C_{R_x}(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi^M \mid |x_a - x_b| = m\}, \quad (13a)$$

$$C_{R_y}(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi^M \mid |y_a - y_b| = m\}, \quad (13b)$$

where subscripts R_x, R_y refer to the metrics defined by the Rectilinear PCFs. The schematics in Fig. 2 represent examples of sites separated by distances $m = 0$, $m = 1$ and $m = 2$ for metrics R_x and R_y .

The counts are then normalized by the expected number of pairs of agents at distance m assuming N uniformly distributed agents:

$$\mathbb{E}[\bar{c}_{R_x}(m)] = \frac{N-1}{L_x L_y - 1} \frac{N}{L_x L_y} L_y^2 (L_x - m), \quad (14a)$$

$$\mathbb{E}[\bar{c}_{R_y}(m)] = \frac{N-1}{L_x L_y - 1} \frac{N}{L_x L_y} L_x^2 (L_y - m), \quad (14b)$$

respectively. For details of the derivation of these factors, see Ref. [12]. The final rectilinear PCF is defined as the arithmetic

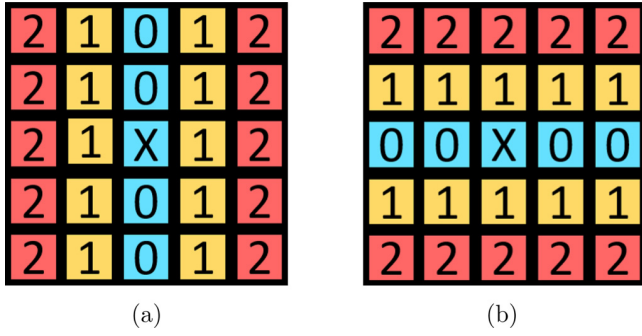


FIG. 2. Schematic of agent pairs using (a) the R_x metric, (b) the R_y metric. Sites in yellow and red (labelled 1, 2 respectively) are defined to be distance-one and -two neighbors, respectively, from the blue site labeled X.

average of the two orthogonal PCFs, i.e.,

$$f_R^M(m) = \frac{1}{2} [f_{R_x}(m) + f_{R_y}(m)],$$

where

$$f_{R_x}^M(m) = \frac{c_{R_x}}{\mathbb{E}[\bar{c}_{R_x}(m)]},$$

$$f_{R_y}^M(m) = \frac{c_{R_y}}{\mathbb{E}[\bar{c}_{R_y}(m)]}. \quad (15)$$

The rectilinear PCF correctly identifies spatial correlation in many examples and, unlike the annular PCF, is normalized correctly. However, one major issue that the rectilinear PCF suffers from is that, due to the inherent anisotropy of its definition, spatial structures that are biased in either Cartesian direction may be missed. For such patterns, the PCF given by the R_x and R_y metrics are approximately constant functions of distance, because the averaged row and column densities are constant along the axes, despite the fact that clustering can still be present. Examples of these spatial patterns include many biologically and chemically relevant cases, such as diagonal stripes and chessboard patterns [32,33] (see Fig. 8). We note that when the pattern structure is biased in only one Cartesian direction, the preaveraging rectilinear PCFs f_{R_x} and f_{R_y} will identify further information about the direction of the spatial pattern. Another limitation of the rectilinear PCF is that it applies only to regular square lattices and a generalisation to other forms of tessellations would be challenging.

III. THE SQUARE TAXICAB AND SQUARE UNIFORM PCFS

In this section we define two discrete PCFs for a square lattice: the square taxicab and square uniform PCF, using the taxicab and uniform metric, respectively, under both periodic and nonperiodic BC. Using the same notation as in Sec. II, we define the subsets of agent pairs separated by distance m under nonperiodic BC as

$$C_1^n(m) = \{(a, b) \in \psi^M \mid \|a - b\|_1 = m\}, \quad m \in \mathcal{D}_1^n, \quad (16a)$$

$$C_\infty^n(m) = \{(a, b) \in \psi^M \mid \|a - b\|_\infty = m\}, \quad m \in \mathcal{D}_\infty^n, \quad (16b)$$

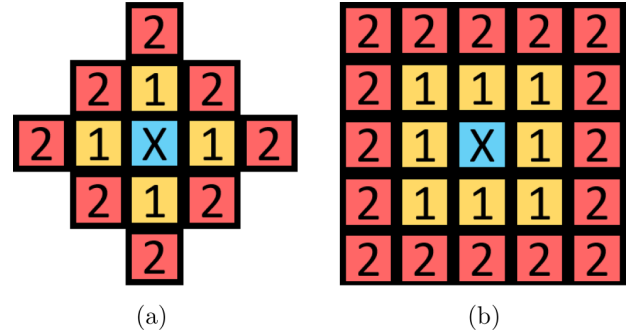


FIG. 3. Schematic of agent pairs using (a) the taxicab metric (b) the uniform metric. Sites in yellow and red, labeled with numbers 1 and 2, respectively, are defined to be distance-one and -two neighbors from the site marked in blue (labeled with X).

for the taxicab and uniform metric, respectively. Here, $\mathcal{D}_1^n = \mathcal{D}_\infty^n = \{1, 2, \dots, \min\{L_x, L_y\} - 1\}$ and the superscript n refers to the fact we are considering *nonperiodic* BC. Using the definitions of the uniform and taxicab metrics, we can express these sets as

$$C_1^n(m) = \{(a, b) \in \psi^M \mid |x_a - x_b| + |y_a - y_b| = m\}, \quad (17a)$$

$$C_\infty^n(m) = \{(a, b) \in \psi^M \mid \max\{|x_a - x_b|, |y_a - y_b|\} = m\}. \quad (17b)$$

Similarly, we define the subsets of agent pairs separated by distance m under periodic BC as

$$C_1^p(m) = \{(a, b) \in \psi^M \mid \min\{|x_a - x_b|, L_x - |x_a - x_b|\} + \min\{|y_a - y_b|, L_y - |y_a - y_b|\} = m\},$$

$$m \in \mathcal{D}_1^p, \quad (18a)$$

$$C_\infty^p(m) = \{(a, b) \in \psi^M \mid \max\{\min\{|x_a - x_b|, L_x - |x_a - x_b|\}, \min\{|y_a - y_b|, L_y - |y_a - y_b|\}\} = m\},$$

$$m \in \mathcal{D}_\infty^p, \quad (18b)$$

where $\mathcal{D}_1^p = \mathcal{D}_\infty^p = \{1, 2, \dots, \min\{\lfloor \frac{L_x}{2} \rfloor, \lfloor \frac{L_y}{2} \rfloor\}\}$. The corresponding definitions of S_1^n , S_∞^n , S_1^p , and S_∞^p can be obtained similarly by using Eq. (6). Here the superscript p refers to the fact we are considering *periodic* BC. Notice that we restrict the largest m to be $\min\{\lfloor \frac{L_x}{2} \rfloor, \lfloor \frac{L_y}{2} \rfloor\}$ to simplify the computation of the normalization factor (see Sec. III A). However, with some work this restriction could be relaxed. The schematics in Fig. 3 represent examples of sites separated by distance $m = 1$ and $m = 2$ using the taxicab (a) and uniform (b) metrics.

For the normalization factors, Binder and Simpson [12] use

$$\mathbb{E}[\bar{c}_d(m)] = \left(\frac{N}{L_x L_y} \right) \left(\frac{N-1}{L_x L_y - 1} \right) s_d(m), \quad (19)$$

where $s_d(m)$ is defined as in Eq. (7) and d refers to the metric used. In other words, the expected number of pairs of agents at distance m on a lattice with N uniformly distributed agents can

be written as the probability that two different sites at distance m are simultaneously occupied, multiplied by the total number of pairs of sites at distance m in the domain.

To complete the definitions of our square lattice PCFs we need to provide an expression for s_1 and s_∞ . We address this in the next two sections, distinguishing between the cases of periodic and nonperiodic BC.

A. Normalization of the square taxicab and square uniform PCF under periodic boundary conditions

As the derivation for the normalization is simple under periodic BC and more complicated under nonperiodic BC, we first consider the system with periodic BC and determine s_1^p , s_∞^p , where p denotes periodic BC. Let us define the number of sites separated by a distance m from any given reference site on a lattice as $t_1(m)$, $t_\infty(m)$ under the taxicab and uniform metric, respectively. These read

$$t_1(m) = 4m, \quad (20a)$$

$$t_\infty(m) = 8m. \quad (20b)$$

The proofs of Eqs. (20) are omitted, but they can be obtained easily by induction on m . Examples for $m = 1, 2$ can be seen in Fig. 3. Notice that for $m \leq \min \{\lfloor \frac{L_x}{2} \rfloor, \lfloor \frac{L_y}{2} \rfloor\}$, given any site on the lattice, the number of sites at distance m from this reference site in the case of periodic BC is exactly $t(m)$. Consider the lattice of size $L_x \times L_y$ with $L_x, L_y > 2$. If we multiply the total number of lattice sites by $t(m)$, we count each pair of sites separated by distance m exactly twice. Hence, we conclude that

$$s_{d=1,\infty}^p(m) = \frac{t(m)L_x L_y}{2}, \quad (21)$$

using the taxicab and uniform metrics. Substituting values for $t_1(m)$ and $t_\infty(m)$ from Eqs. (20) we deduce that

$$s_1^p(m) = 2mL_x L_y, \quad (22a)$$

$$s_\infty^p(m) = 4mL_x L_y. \quad (22b)$$

Therefore, by substituting Eqs. (22) into Eq. (19), the normalization factors under periodic BC are

$$\mathbb{E}[\bar{c}_1^p(m)] = \frac{2mN(N-1)}{L_x L_y - 1}, \quad (23a)$$

$$\mathbb{E}[\bar{c}_\infty^p(m)] = \frac{4mN(N-1)}{L_x L_y - 1}. \quad (23b)$$

B. Normalization of the square taxicab and square uniform PCF under nonperiodic BC

In this section we derive expressions for $s_1^n(m)$ and $s_\infty^n(m)$, where n denotes nonperiodic BC. Notice that, for all $m \in \mathcal{D}$, we have that $s^p(m) > s^n(m)$ since $s^p(m)$ includes pairs that cross the domain boundary, whereas $s^n(m)$ does not. Therefore, to find a formula for $s^n(m)$, it is enough to determine a formula for the remainders defined by

$$r_1(m) = s_1^p(m) - s_1^n(m), \quad (24a)$$

$$r_\infty(m) = s_\infty^p(m) - s_\infty^n(m). \quad (24b)$$

The remainders count the number of pairs of sites that cross a boundary under the periodic BC. For simplicity, throughout this section, we will only derive the normalization for the taxicab metric; however, the derivation for the uniform metric is similar and can be found in Supplemental Material Sec. S.1 for reference [31]. Let us define the set of pairs of sites separated by distance $m \in \mathcal{D}_1^n$ that cross the x boundary (horizontal axis) or y boundary (vertical axis), respectively, as

$$P_1^x(m) = \{(a, b) \in S_1^p(m) \mid |y_a - y_b| > L_y - |y_a - y_b|\}, \quad (25a)$$

$$P_1^y(m) = \{(a, b) \in S_1^p(m) \mid |x_a - x_b| > L_x - |x_a - x_b|\}. \quad (25b)$$

Of these pairs, let us consider those pairs that are at distance $k \in \{1, \dots, m\}$ rows or columns from each other, respectively. We define these subsets as

$$P_1^x(m, k) = \{(a, b) \in P_1^x(m) \mid L_y - |y_a - y_b| = k\}, \quad (26a)$$

$$P_1^y(m, k) = \{(a, b) \in P_1^y(m) \mid L_x - |x_a - x_b| = k\}. \quad (26b)$$

Notice that $P_1^x(m) = \bigcup_{k=1}^m P_1^x(m, k)$ and $P_1^y(m) = \bigcup_{k=1}^m P_1^y(m, k)$. Figures 4(a) and 4(b) give visualizations of pairs of sites within $P_1^x(m, m)$. Figure 4(c) gives examples of distances between pairs of sites in $P_1^x(m, k)$, for $k = 1, \dots, m$. By definition Eqs. (16a) and (18a) we have that

$$S_1^p(m) \setminus S_1^n(m) = P_1^x(m) \cup P_1^y(m). \quad (27)$$

Hence, by combining Eqs. (24a) and (7), we obtain

$$\begin{aligned} r_1(m) &= |P_1^x(m) \cup P_1^y(m)| \\ &= \sum_{k=1}^m |P_1^x(m, k)| + \sum_{k=1}^m |P_1^y(m, k)| - |P_1^x(m) \cap P_1^y(m)|. \end{aligned} \quad (28)$$

To conclude the computation we derive an expression for the two sums in Eq. (28) and the corresponding equation for the size of the intersection. By counting the contribution of each type of pair (see Fig. 4 for a visualization), one can write down the following expressions for the two sums in Eq. (28):

$$\sum_{k=1}^m |P_1^x(m, k)| = 2[L_x + 2L_x + \dots + L_x(m-1)] + L_x m, \quad (29a)$$

$$\sum_{k=1}^m |P_1^y(m, k)| = 2[L_y + 2L_y + \dots + L_y(m-1)] + L_y m. \quad (29b)$$

Hence,

$$\begin{aligned} \sum_{k=1}^m |P_1^x(m, k)| + \sum_{k=1}^m |P_1^y(m, k)| \\ = 2[L_x + 2L_x + \dots + L_x(m-1)] + L_x m \\ + 2[L_y + 2L_y + \dots + L_y(m-1)] + L_y m \end{aligned}$$

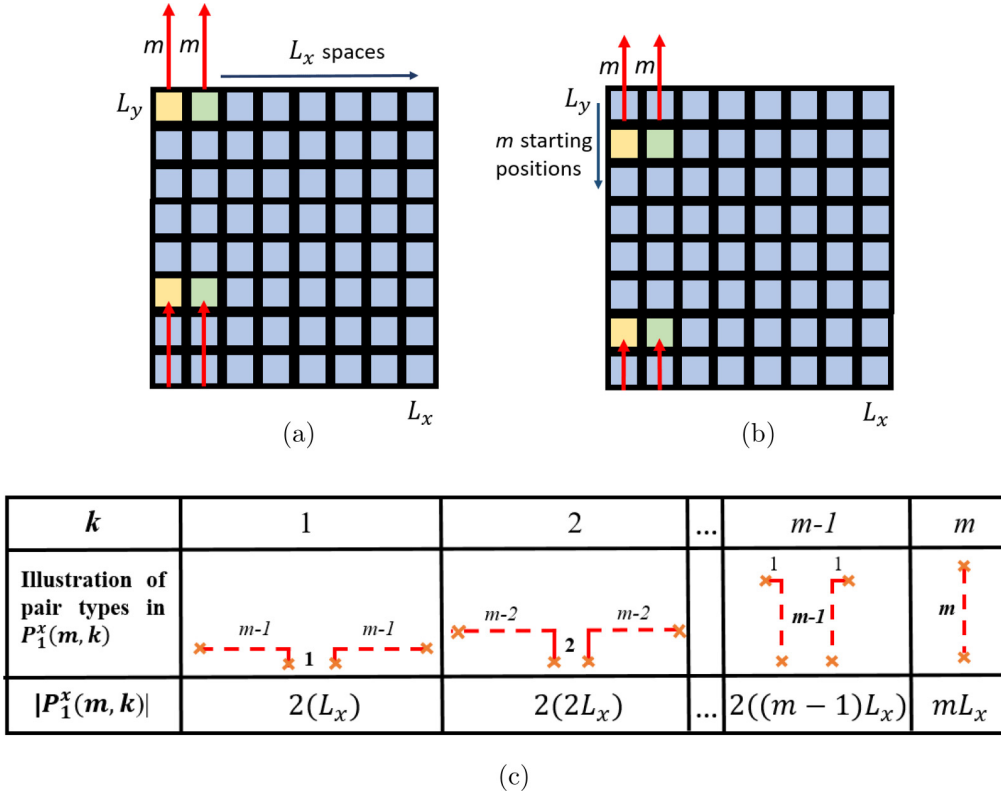


FIG. 4. A visualization of the pairs of sites in $P_1^x(m)$. Panels (a) and (b) show two different site pairs in $P_1^x(m)$. For each of the L_x columns, each site in the rows $\{L_y - m + 1, L_y - m + 2, \dots, L_y\}$ has a single corresponding site at distance m separated by m rows and 0 columns and reached by crossing the horizontal boundary. Therefore $|P_1^x(m, m)| = mL_x$. Panel (c) shows all the different types of pairs in $P_1^x(m, k)$ for $k = 1, \dots, m$ with the corresponding value of $|P_1^x(m, k)|$.

$$\begin{aligned}
 &= (L_x + L_y) \left(m + 2 \sum_{i=1}^{m-1} i \right) \\
 &= (L_x + L_y) \left[m + 2 \frac{(m-1)m}{2} \right] \\
 &= (L_x + L_y)m^2.
 \end{aligned} \tag{30}$$

We now focus on deriving an expression for the size of intersection, $|P_1^x(m) \cap P_1^y(m)|$, in Eq. (28). The set $P_1^x(m) \cap P_1^y(m)$ consists of pairs of sites separated by distance m that cross both the x and y boundaries simultaneously.

There are two regions of the domain where site pairs cross two boundaries. These are any two consecutive corners of the four corners of the domain. Examples of these regions and site pairs within these regions are visualized in Fig. 5. In Fig. 6 we give an illustrative example in which we count the number of these pairs for $m = 5$. All sites inside the boundaries of the domain colored in orange, purple, green, and yellow are distance $m = 5$ from other sites of the same color outside the boundaries of the domain. Notice that the yellow site in the corner at (L_x, L_y) is distance five from a total of four sites, reached by crossing the x and y boundaries, denoted by a 4 in the site. Similarly, the two green sites at $(L_x - 1, L_y)$ and $(L_x, L_y - 1)$ are distance m from three sites, reached by crossing the x and y boundary, denoted by a 3 in the two sites. $|P_1^x(5) \cap P_1^y(5)|$ is the sum of all the numbers in the colored

sites multiplied by two to account for the second corner region. Extrapolating, for any value of m , the number of pairs of sites

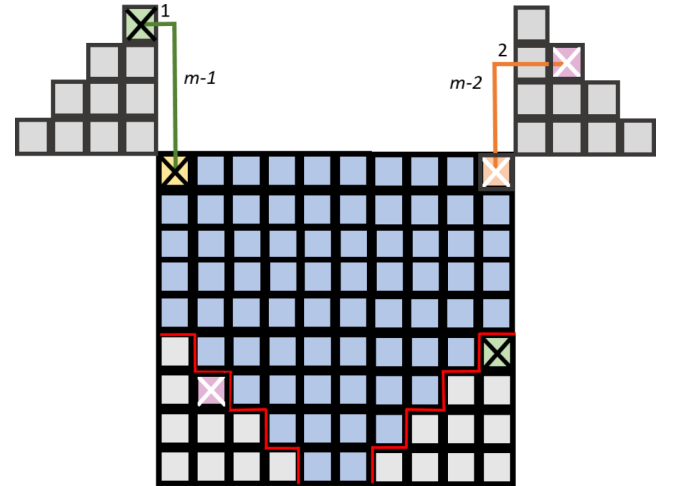


FIG. 5. Examples of pairs of sites separated by distance $m = 5$ that cross both the x and y boundaries, i.e., pairs of sites in $P_1^x(m) \cap P_1^y(m)$. The gray sites outside the domain correspond to the gray sites inside the domain in the diametrically opposite corner. As illustrations of site pairs at a distance m which cross both boundaries, the orange site containing a white cross is distance m from the pink site containing a white cross. Similarly, the yellow site containing a black cross is distance m from the green site containing a black cross.

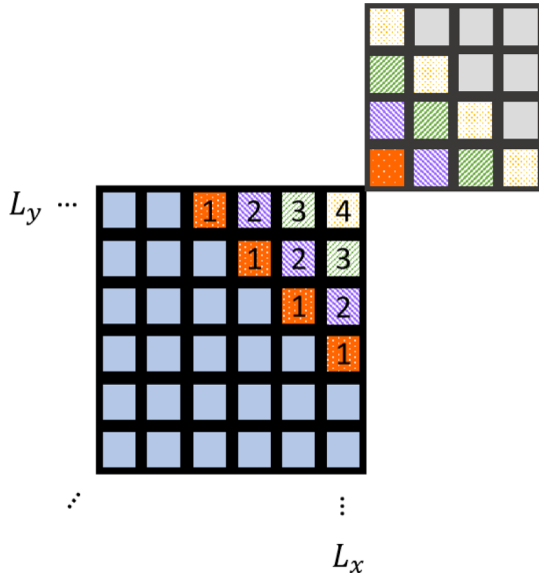


FIG. 6. Examples of pairs in $P_1^x(5) \cap P_1^y(5)$ on a zoomed-in corner of a larger domain. Sites with a given pattern (not plain) are distance $m = 5$ away from sites with the same pattern that can be reached by crossing the x and y boundaries. The number in each given site corresponds to the number of sites at distance $m = 5$, reached by crossing the x and y boundaries.

that cross the two boundaries is exactly

$$\begin{aligned} |P_1^x(m) \cap P_1^y(m)| &= 2[(m-1) + 2(m-2) \\ &\quad + 3(m-3) + \dots + m-1] \\ &= 2 \sum_{i=1}^{m-1} (m-i)i = \frac{m^3 - m}{3}. \end{aligned} \quad (31)$$

By substituting Eqs. (31) and (30) into Eq. (28) we gain an expression for the remainder $r_1(m)$. By rearranging Eq. (24a) we determine $s_1^n(m)$, which we then substitute into Eq. (19) to obtain the exact expression for the normalization in the nonperiodic case. This is given by

$$\begin{aligned} \mathbb{E}[\bar{c}_1^n(m)] &= \left(\frac{N}{L_x L_y} \right) \left(\frac{N-1}{L_x L_y - 1} \right) \\ &\quad \left[2m L_x L_y - (L_x + L_y)m^2 + \frac{m^3 - m}{3} \right]. \end{aligned} \quad (32)$$

A similar approach can be used to obtain the normalization factor for the uniform metric under nonperiodic BC:

$$\begin{aligned} \mathbb{E}[\bar{c}_\infty^n(m)] &= \left(\frac{N}{L_x L_y} \right) \left(\frac{N-1}{L_x L_y - 1} \right) \\ &\quad [4m L_x L_y - 3(L_x + L_y)m^2 + 2m^3]. \end{aligned} \quad (33)$$

For more details on the derivation of Eq. (33), see Supplemental Material Sec. S.1.

IV. RESULTS

In this section we use the square taxicab and square uniform PCFs defined in Sec. III to analyze the spatial correlation

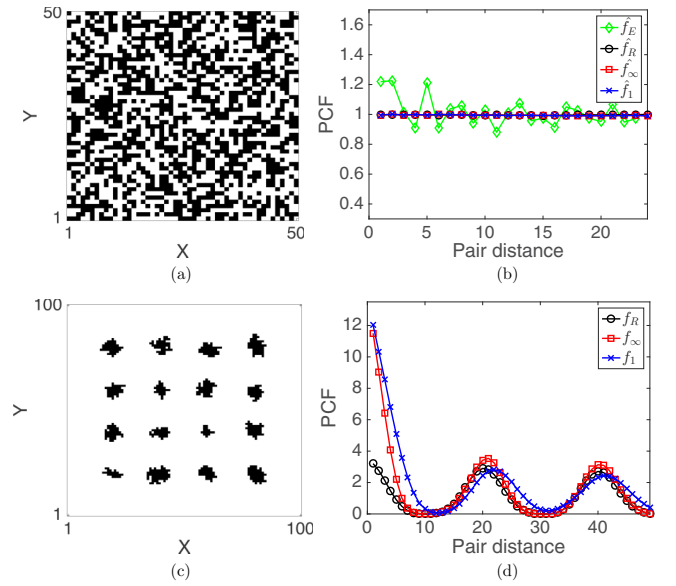


FIG. 7. Examples of spatial structure analysis. In panels (a) and (b) a case with no spatial correlation is considered. Panel (a) is an example visualization of an occupancy matrix uniformly populated with density 0.5. Occupied sites are colored in black and are white otherwise. Panel (b) shows the four PCFs each averaged over 50 uniformly populated matrices. Panels (c) and (d) refer to a discrete simulation of the agent-based model described in the text, at time $t = 10$. Panel (c) is a visualization of the occupancy matrix and in panel (d) our PCFs are compared with the Rectilinear PCF.

in some examples. We compare our results with previously suggested on-lattice PCFs.

We start by computing the PCFs for a system without any spatial correlation. We consider 50 independent occupancy matrices, U_i , $i = 1, \dots, 50$, populated uniformly at random with density 0.5 [see Fig. 7(a) for example]. For each realisation, U_i , we compute the corresponding PCF, $f_d^{U_i}$, and then we average the results over the 50 realisations which we denote \hat{f}_d . If the normalization is correct, $\hat{f}_d(m)$ should return the value unity for every pair distance, m , meaning that no spatial correlation is found. In Fig. 7(b) all four aforementioned averaged PCFs are plotted: \hat{f}_A , \hat{f}_R , \hat{f}_1 , and \hat{f}_∞ . The results show that both the averaged square uniform PCF, \hat{f}_1 , and square taxicab PCF, \hat{f}_∞ , correctly predict that there is no spatial correlation. The averaged rectilinear PCF, \hat{f}_R , also correctly predicts no spatial correlation. However, the averaged annular PCF, \hat{f}_A , has clear peaks, suggesting, incorrectly, the presence of spatial correlation. Since the results are averaged over multiple repeats, such a discrepancy can not be attributed to stochasticity, but is due to incorrect normalization as explained in Sec. II. Note that the annular PCF can still correctly identify spatial correlation in many examples; however, the incorrect normalization often makes the results hard to interpret. This is because the annular PCF makes it difficult to distinguish between genuine correlation and systematic error. For this reason, for the rest of this section, we omit the results of the annular PCF and continue to compare between our PCFs and the rectilinear PCF using nonperiodic BC.

Next we consider examples of strong spatial correlation. Figure 7(c) shows an example of aggregation driven by a

proliferation mechanism. The occupancy matrix is obtained by simulating an on-lattice agent-based model with periodic BC as described in Binder and Simpson [12], which we summarize as follows. The model is initialized with 16 agents, located at coordinates given by $\{(x, y) \mid x, y \in \{20, 40, 60, 80\}\}$ on a regular square lattice with $L_x = 100, L_y = 100$. Time is discretized with a time step $\tau = 1$ and the number of agents at time t is denoted by $n(t)$. At each time step the configuration at time $t + \tau$ is obtained from the configuration at time t , by repeating the following steps $n(t)$ times. (1) An agent is chosen uniformly at random from the $n(t)$ agents present at the end of the previous time-step; (2) one of its four von Neumann neighbors is selected at random with equal probability; (3) if the selected site is empty, then a new agent is placed in this site and $n(t + \tau) = n(t) + 1$; otherwise, the configuration is left unchanged.

Figure 7(c) shows a single realisation after 10 time steps and Fig. 7(d) shows the corresponding PCFs: f_R , f_1 , and f_∞ . The results indicate that all of the PCFs correctly identify aggregation. However, the quantitative information about aggregate sizes at different length scales provide by each PCF varies. For example, we see that all PCFs in Fig. 7(d) exhibit three peaks; at $m = 1$, $m \approx 20$, and $m \approx 40$. The different peaks and troughs of the PCF profiles have different qualitative meanings related to the correlation type. Due to the local approach of the square taxicab PCF and square uniform PCF, the first peak at $m = 1$ is three times higher than the peaks at larger values of distance. These differences in amplitude highlight the different peak origins. Specifically, the first and highest peak distinguishes the individual cluster aggregate and the later peaks indicate correlation between different clusters. In contrast, all three peaks in the rectilinear PCF are the same amplitude. Note that, in the case of aggregation, the average diameter of the aggregate corresponds to the first value of distance which achieves the minimum of the PCF. The rectilinear, square uniform, and square taxicab PCFs estimate the aggregate diameter to be 9, 9 and 11, respectively. Importantly, the PCFs capture the fact that this diameter depends on the metric used. In particular the distance between two sites measured using the uniform metric is always less than or equal to the taxicab distance. This phenomenon is seen more clearly in later examples.

We now consider a series of examples with spatial correlation constructed artificially to compare and evaluate the different PCFs. In Fig. 8 we compare our square uniform and square taxicab PCF with the rectilinear PCF for three different patterns with strong spatial correlation. All three examples (Fig. 8(a), diagonal stripes; Fig. 8(c), chessboard pattern; and Fig. 8(e), concentric circles) are chosen so that the column- and row-averaged densities are constant and hence the spatial structure is not recognised by the rectilinear PCF, as shown in Figs. 8(b), 8(d), and 8(f). This is in contrast to the approach of our PCFs (both square uniform and square taxicab), which successfully recognize the spatial structure in all three examples. In addition, these examples uncover other interesting differences between the taxicab and the uniform approaches. Consider the PCF for the case of diagonal stripes and the chessboard pattern [Figs. 8(b) and 8(d)]. Here the square uniform PCF quickly converges to unity (no spatial correlation) for large distance m , while in both cases, the square taxicab PCF still shows a strong oscillatory behavior for large distance m , suggesting spatial correlation. To give an intuitive

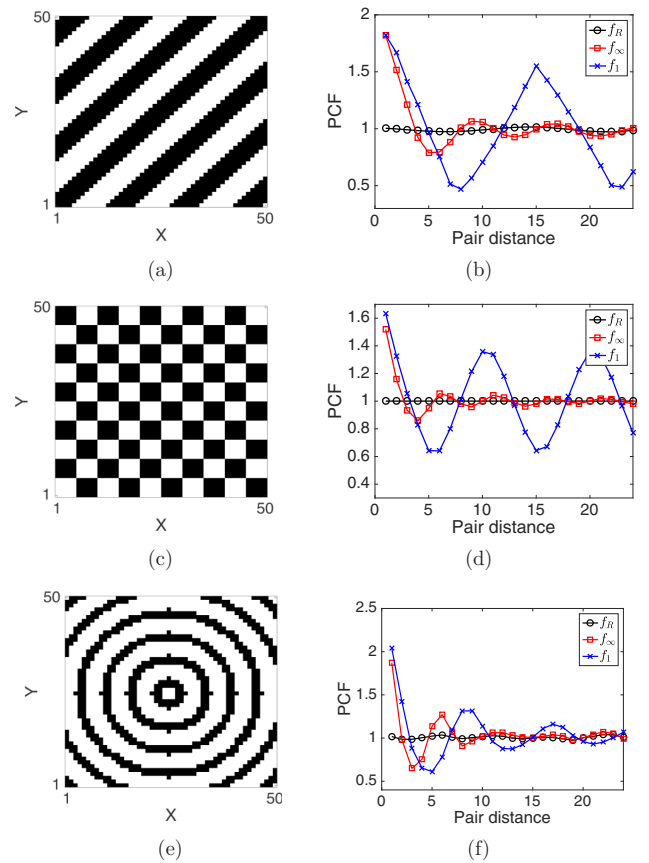


FIG. 8. Examples of pattern analysis. Panels (a), (c), and (e) visualize three constructed spatial patterns. Panels (b), (d), and (f) displays the corresponding square taxicab, square uniform and rectilinear PCF using the occupancy matrices for (a), (c), and (e), respectively.

explanation of this phenomenon, let us consider the shapes of balls of size m centered at a given site \mathbf{a} under the two metrics. These balls are defined as $B_m(\mathbf{a}) = \{\mathbf{b} \in \mathbb{L} \mid \|\mathbf{a} - \mathbf{b}\|_d \leq m\}$ with $d = 1, \infty$, respectively (see Fig. 3). The ball corresponding to the uniform metric [Fig. 3(b)] has a square shape with the sides aligned with the directions of the axis. This implies that when distance m becomes close to either L_x or L_y in size, the ball corresponding to the uniform metric of distance m contains most of the sites in the corresponding row or column at distance m . For large m , therefore, the uniform metric begins to work in a similar way to the Rectilinear PCF and thus fails to recognize anisotropic patterns biased in the Cartesian directions. The ball of the taxicab metric [see Fig. 3(a)], however, has a diamond shape. Consequently, the long-distance correlations appear clear even for patterns in which both the average column and row densities are constant, as in Fig. 8.

The examples in Fig. 8 were constructed specifically to underline the main differences between the three PCFs. Nevertheless, similar patterns also arise in many biologically and mathematically relevant applications [32–34]. We conclude this section by comparing the three PCF approaches applied to some real-world examples taken from the literature. In Fig. 9 we analyze three images representing examples of Turing patterns. A corresponding occupancy matrix is obtained by representing each pixel of the image as a value in a matrix

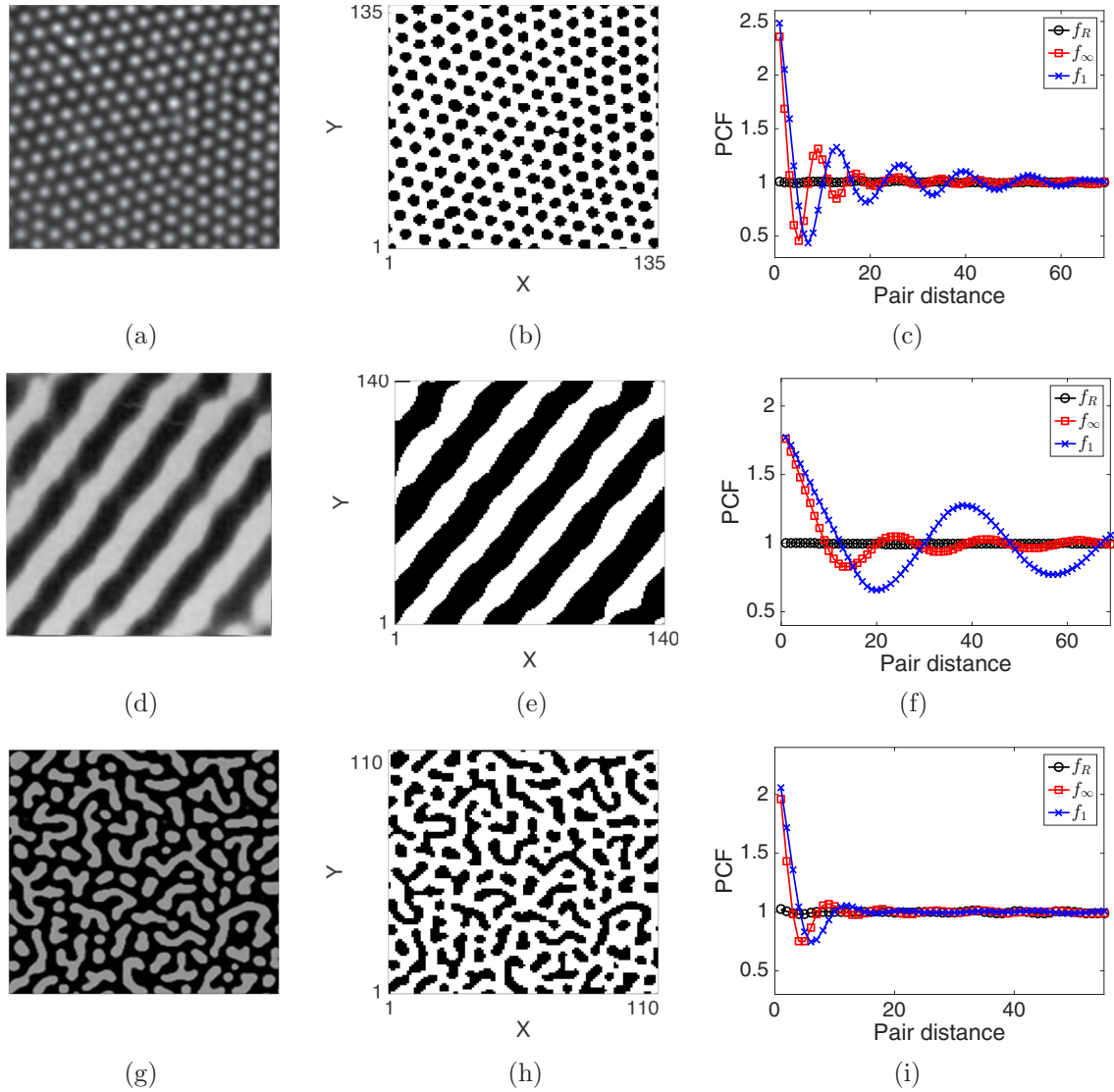


FIG. 9. Spatial analysis of Turing patterns. Panels (a), (d), and (g) show original images representing the results of a reaction-diffusion mechanism between two chemical substances, (a) is reprinted from Ref. [35], (d) from Ref. [32], and (g) from Ref. [36]. Panels (b), (e), and (h) visualize the occupancy matrices corresponding to the original images (described in text). In panels (c), (f), and (i) we compare the square taxicab, square uniform, and rectilinear PCFs for each of the examples.

which is 1 (i.e., occupied) if the three values of the RGB colorization of the pixel are above a certain threshold (80) and 0 otherwise. In all cases the column and row densities are almost constant, hence the spatial structure again remains largely undetected by the rectilinear PCF, while our square uniform and square taxicab PCF correctly identify the patterns. As already observed in the previous examples, we note that the estimated diameter of the aggregate, wavelength, and amplitude of the oscillations differ according to the metric used.

V. THE TRIANGLE, HEXAGON, AND CUBE PCFS

Despite the square lattice being the most popular set up for spatially discrete models [13,37–40], in some situations other types of tessellation, either regular or irregular, can be more suitable [13,17,41].

In the following subsections we extend our definition of the PCFs in Sec. III to more general types of tessellations. We define the triangle, hexagon, cube uniform, and cube taxicab PCFs under nonperiodic and periodic BC for triangular, hexagonal, and cuboidal tessellations, respectively. The following subsections represent qualitative discussions of the different cases. We refer the reader to the Supplemental Material Secs. S.2 and S.3 for the full details of the derivation of the PCF formulas [31].

A. Triangle and hexagon PCF

First, we define triangularly and hexagonally tessellated domains of size $L_x \times L_y$. These comprise an array of L_y rows of L_x regular triangles or hexagons, respectively. Examples for which $L_x = 6$ and $L_y = 3$ for each of the two cases are given in Figs. 10(a) and 10(b), respectively. Notice that, for a

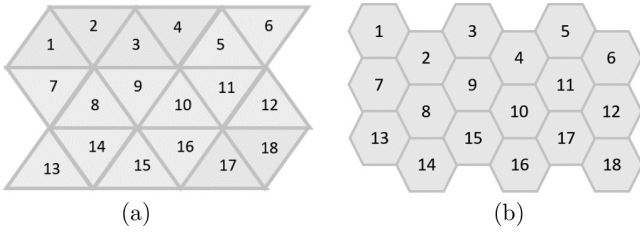


FIG. 10. Example domains for (a) triangular and (b) hexagonal tessellations in which $L_x = 6$ and $L_y = 3$.

periodic BC to be meaningful in these domain definitions, L_x must be even. Therefore, we enforce this as a condition in what follows.

In the context of triangular and hexagonal tessellations we focus our attention on the taxicab metric, both for simplicity and as the most natural metric on this domain type. Using the taxicab metric, the number of sites of distance m from any given reference site is given by

$$t_{\text{tri}}(m) = 3m, \quad (34a)$$

$$t_{\text{hex}}(m) = 6m. \quad (34b)$$

The proofs of Eqs. (34) are omitted, but they can be obtained easily by induction on m . Examples for $m = 1, 2, 3$ are visualized in Fig. 11. Using the same reasoning as in Sec. III under periodic BC:

$$s_{\text{tri}}^p(m) = 3m \frac{L_x L_y}{2}, \quad (35a)$$

$$s_{\text{hex}}^p(m) = 3m L_x L_y. \quad (35b)$$

Substituting Eqs. (35) into Eq. (19) we obtain the normalizations for the triangle and hexagon PCF, respectively, under periodic BCs, namely,

$$\mathbb{E}[\bar{c}_{\text{tri}}^p(m)] = \frac{3mN(N-1)}{2(L_x L_y - 1)}, \quad (36a)$$

$$\mathbb{E}[\bar{c}_{\text{hex}}^p(m)] = \frac{3mN(N-1)}{L_x L_y - 1}. \quad (36b)$$

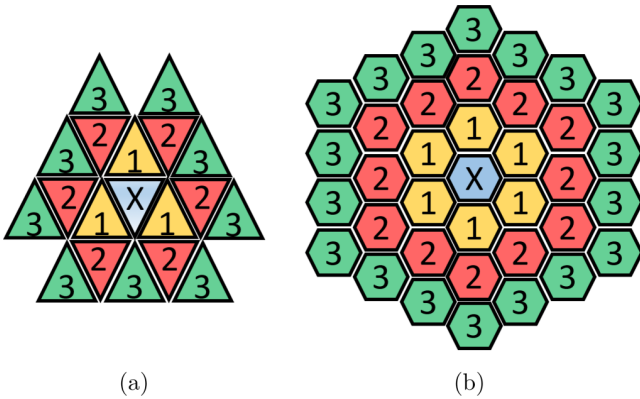


FIG. 11. Schematic of agent pairs with the (a) triangular tessellation and (b) hexagonal tessellation using the taxicab metric. Sites in yellow, red, and green, labeled 1, 2, and 3, respectively, are distance-one, -two, and -three neighbors from the blue site (labeled with X), respectively.

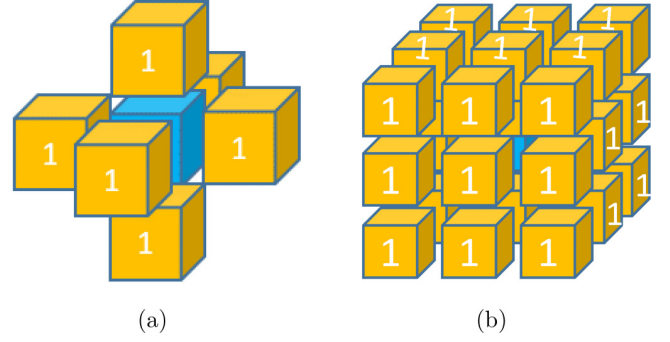


FIG. 12. Schematic of agent pairs using (a) the taxicab metric and (b) the uniform metric. Sites in yellow are defined to be distance-one neighbors from the site marked in blue.

From these expressions one can obtain the formulas of f_{tri} and f_{hex} under periodic BC by using the definition Eq. (9). The normalizations in the case of nonperiodic BC are given in the Supplemental Material Sec. S.2 [31].

B. Uniform cube and taxicab cube PCF

We define a three-dimensional $L_x \times L_y \times L_z$ cuboidal lattice with unit spacing. Using the taxicab and uniform metric, respectively, the number of sites of distance m from any given reference site is given by

$$t_{\text{cube}_1}(m) = 2(2m^2 + 1), \quad (37a)$$

$$t_{\text{cube}_\infty}(m) = 2(12m^2 + 1). \quad (37b)$$

The proofs of Eqs. (37) are omitted, but they can be obtained easily by induction on m . Examples of agent pairs for $m = 1$ are given in Figs. 12(a) and 12(b) for the taxicab and uniform metrics, respectively. Using the same reasoning as in Sec. III, under periodic BCs, the normalizations for taxicab and uniform cube PCFs, respectively, are as follows:

$$s_{\text{cube}_1}^p(m) = (2m^2 + 1)L_x L_y L_z, \quad (38a)$$

$$s_{\text{cube}_\infty}^p(m) = (12m^2 + 1)L_x L_y L_z. \quad (38b)$$

For simplicity we refer the reader to Sec. S.3 of the Supplemental Material for the normalization factors for the cases with nonperiodic BC [31].

VI. THE GENERAL PCF

In this section we provide a comprehensive method for generating a PCF for any tessellation type, BC, and metric but with the caveat of having a high computational cost.

This PCF is a valuable tool for irregular domain shapes and partitions although it can be used for any tessellation of any domain.

First, we consider a two-dimensional domain partitioned into Z regions (or sites) with arbitrary shapes and sizes, each labeled with a number from 1 to Z . Figure 13(a) shows an example of an irregularly shaped domain partitioned in $Z = 17$ regions. Given the domain, we choose a suitable metric. For the irregular lattice, which we consider in the following example, we consider the taxicab metric. This means that we define the

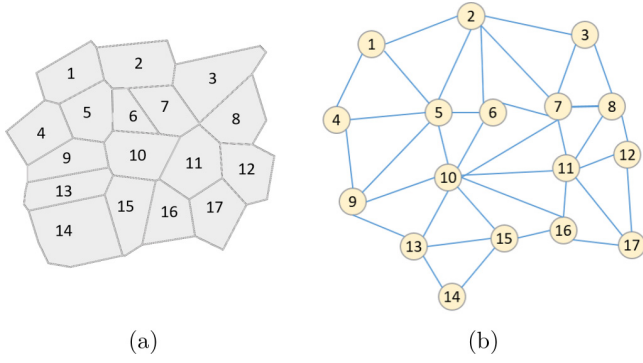


FIG. 13. An example of an irregular domain partition with its corresponding connectivity graph under the taxicab metric. Panel (a) shows A size 17 irregular lattice domain. Panel (b) shows corresponding connectivity graph for the tessellation in (a) under nonperiodic BC using the taxicab metric for distances.

distance, m , between two sites to be the minimum number of sites visited when starting at one site and moving consecutively through adjacent sites to the other. For example, in Fig. 13(a) the sites 4 and 7 are at distance three. Similarly, adjacent sites are defined to be at distance one.

Having chosen and defined a suitable metric, we may now represent the connections between lattice sites as an undirected connectivity graph $G(V, E)$, where each vertex represents a lattice site and each edge connects vertices whose corresponding sites are distance-one neighbors. Using the taxicab metric, edges connect vertices whose corresponding sites are adjacent. Figure 13(b) shows an example of such an association [applied to the irregular lattice in Fig. 13(a)] using the taxicab metric.

The corresponding adjacency matrix of graph G is a $Z \times Z$ matrix defined as follows:

$$A_{i,j}^G = \begin{cases} 1 & \text{for } (i, j) \in E, \\ 0 & \text{for } (i, j) \notin E. \end{cases} \quad (39)$$

We use properties of the adjacency matrix to determine the number of sites at a given distance. In particular, we can compute $(A^G)^m$ whose entries $(A^G)^m_{i,j}$ are the number of walks of length m from vertex i to vertex j . To compute the minimum walk between two sites (and hence the distance between them) we produce the *distance matrix* D^G . This is a $Z \times Z$ matrix defined as

$$D_{i,j}^G = \begin{cases} \min \{m \in \mathbb{N}^+ | (A^G)^m_{i,j} \neq 0\} & \text{for } i \neq j, \\ 0 & \text{for } i = j. \end{cases} \quad (40)$$

Notice that, each entry, $D_{i,j}^G$, denotes the distance between the vertices i and j on graph G and hence on the original lattice.

Given the distance matrix of the domain, D^G , and the set of the occupied sites $M \subseteq V$, the PCF of the system can be computed as follow. The number of pairs of agents at distance m for a general metric d is given by

$$c_d(m) = \frac{1}{2} |\{(i, j) \in M \times M | D_{i,j}^G = m\}|. \quad (41)$$

Similarly, we can express the number of pairs of sites at distance m as

$$s_d(m) = \frac{1}{2} |\{(i, j) \in V \times V | D_{i,j}^G = m\}|. \quad (42)$$

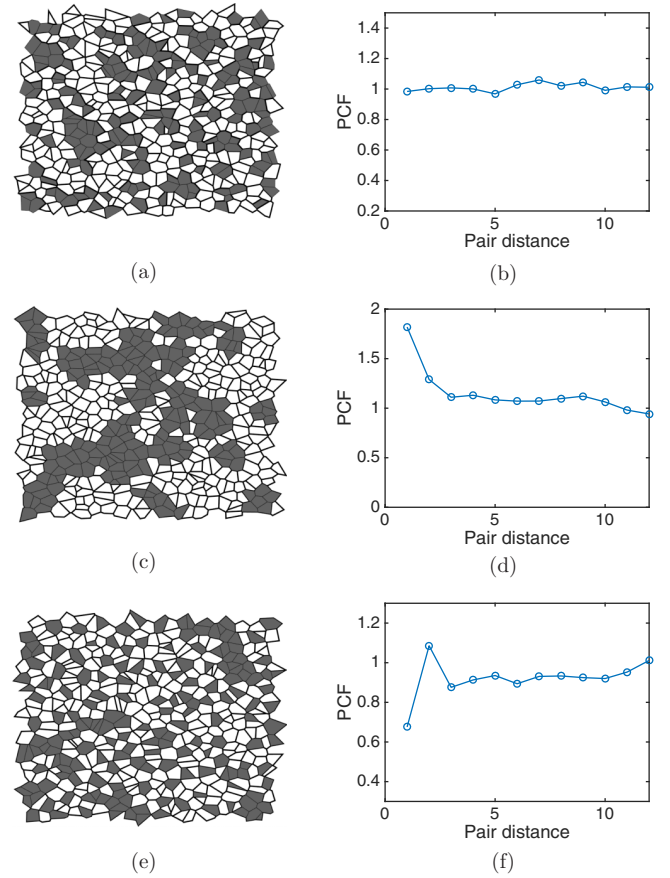


FIG. 14. Examples of spatial correlation analysis on an irregular domain. Panels (a), (c), and (e) show three examples of irregular lattices populated with density 0.4 with agents (gray sites). In panel (a), agents are displaced uniformly at random (no spatial correlation). In panel (c), agents are in a strong form of aggregation, while in panel (e) agents are displaced in a segregate manner. Panels (b), (d), and (f) are the corresponding general PCF evaluations for panels (a), (c), and (e), respectively.

To compute the normalization factor, denote the total number of agents as $N = |M|$, and hence using the same argument as in Sec. III we can write

$$\mathbb{E}[\bar{c}_d(m)] = \left(\frac{N}{Z}\right) \left(\frac{N-1}{Z-1}\right) s_d(m). \quad (43)$$

The general PCF is then defined by combining Eqs. (42) and (43) in Eq. (9). Notice that the computation of the normalization for the general PCF can be computationally expensive. This is because the computation of D^G involves calculating powers of matrices of size $Z \times Z$, where Z is often large. In particular, the cost of computing the normalization of the general PCF is $O(Z^3 m_{\max})$ in which m_{\max} is the maximum value m for which the PCF is computed. For this reason, the general PCF is better reserved for cases in which the expression for the normalization factor cannot be computed analytically, unlike in Secs. III, V A, and V B, although it can, of course, be used even if an analytical formula is available.

In Fig. 14 we apply the general PCF to three examples of agent-based systems on an irregular lattice. In all three examples, the irregular tessellation is the Voronoi partition

based on a set of randomly distributed points. The randomized points are obtained by starting with a square lattice, with lattice size Δ , and perturbing the coordinates of each point (x_i, y_i) to $(x_i + \delta_i^x, y_i + \delta_i^y)$ with each δ_i^x, δ_i^y chosen uniformly at random in the interval $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$.

In the first example, Fig. 14(a), N lattice sites selected uniformly at random to be occupied (gray sites). By eye, several larger clusters of occupied and unoccupied lattice sites are evident, indicating that there may be spatial correlation. Figure 14(b) shows the corresponding general PCF. The values are close to unity, correctly identifying that there is, indeed, no spatial correlation in the system. This highlights the importance of accurate quantitative methods for determining spatial correlation rather than a reliance on ad hoc judgements.

In the other two cases we test our PCF with examples of strong spatial correlation. In Fig. 14(c) we consider a system in an aggregated state. To generate such a configuration, we start with an empty domain, and select an empty site uniformly at random. We then place an agent in this site and all of its neighboring sites (if they are not already occupied). We repeat this process until we reach density 0.4. The process leads to a strong form of aggregation which the general PCF, in Fig. 14(d), correctly identifies. Finally, in Fig. 14(e) we consider a system in a segregated state. To generate such a configuration, we start with a fully populated domain, and repeatedly select at random an occupied site. We remove all agents occupying adjacent sites but leave the initially selected site occupied. The process ends when density 0.4 is reached. This mechanism generates a system which is unlikely to have adjacent sites occupied and more likely to have agents displaced at distance two. The corresponding PCF, shown in Fig. 14(f), correctly captures both of these features: the PCF has value 0.68 at pair distance one, which implies negative correlation at the shortest distance, and value 1.1 at pair distance two, highlighting the positive correlation at slightly larger distances.

Note that, since the normalization in the general PCF can give an exact value for the number of sites at certain pair distances, we used this method to check the normalization factors for all of our previously proposed PCFs. In all cases the results confirmed the analytical expressions given in Secs. III, V A, and V B and in the Supplemental Material [31].

VII. CONCLUSIONS

In this paper we have developed a set of tools to study spatial correlation on discrete domains. We derived two discrete

pair-correlation functions for an exclusion process on a square lattice: the square uniform and square taxicab PCF. We applied our PCFs to patterns observed in nature and to computational simulations and showed that our PCF can not only distinguish and quantify different types of correlation but also that it improves upon previous on-lattice PCFs. For example, we showed that our PCF was normalized correctly, unlike the annular PCF, and was able to identify and quantify anisotropic patterns such as the chessboard or the diagonal stripes that the rectilinear PCF [12] missed. Furthermore, we highlighted how different measures of distance, taxicab and uniform, can lead to different quantifications of spatial correlation.

We extended the calculation of appropriate PCFs to deal with exclusion processes on the other regular spatial tessellations in two dimensions as well as the cubic lattice in three dimensions. We derived the triangle PCF, hexagon PCF, cube taxicab PCF, and cube uniform PCF. These are the first PCFs defined specifically for these discrete lattice types. Finally, we derived a comprehensive PCF for any kind of discrete domain, BC, and metric, which we referred to as the general PCF. The method can be computationally expensive, however, it allows complete freedom in defining a suitable PCF for more complex cases, including those for which recognizing spatial correlation by eye becomes less intuitive.

All of our PCFs are designed for a single species of agents. However, in many applications the agents are divided into multiple species and it can be important to distinguish between different types of spatial correlations: either within agents of the same species (autocorrelation) or comparing the position of agents of different species (cross-correlation). Dini *et al.* [42] have recently investigated correlation in multiple species by using the Rectilinear PCF. We believe a similar approach to that of Dini *et al.* [42] can be applied to our isotropic PCFs to quantify heterogeneous correlations, yet it lies beyond the scope of this paper and, as such, we will tackle it in a future publication. The isotropic PCFs that we defined in this paper will be important in further studies and applications. In particular, our functions can improve previous studies (see Ref. [43], for example) which have used PCF as an efficient summary statistics to infer model parameters.

ACKNOWLEDGMENT

The authors would like to thank the CMB-CNCB preprint club for constructive and helpful comments on a preprint of this paper. J.P.O. acknowledges support from the SWBio DTP.

-
- [1] M. S. Steinberg, Adhesion in development: An historical overview, *Dev. Biol.* **180**, 377 (1996).
 - [2] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Scale-free correlations in starling flocks, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11865 (2010).
 - [3] J. E. F. Green, S. L. Waters, J. P. Whiteley, L. Edelstein-Keshet, K. M. Shakesheff, and H. M. Byrne, Nonlocal models for the formation of hepatocyte–stellate cell aggregates, *J. Theor. Biol.* **267**, 106 (2010).
 - [4] R. J. Thomas, A. Bennett, B. Thomson, and K. M. Shakesheff, Hepatic stellate cells on poly (dl-lactic acid) surfaces control the formation of 3D hepatocyte coculture aggregates *in vitro*, *eCells Mater.* **11**, 16 (2005).
 - [5] F. Chalub, Y. Dolak-Struss, P. Markowich, D. Oelz, C. Schmeiser, and A. Soreff, Model hierarchies for cell aggregation by chemotaxis, *Math. Models Methods Appl. Sci.* **16**, 1173 (2006).
 - [6] H. G. Othmer and A. Stevens, Aggregation, blowup, and collapse: The ABC's of taxis in reinforced random walks, *SIAM J. Appl. Math.* **57**, 1044 (1997).

- [7] A. Stevens, A stochastic cellular automaton modeling gliding and aggregation of myxobacteria, *SIAM J. Appl. Math.* **61**, 172 (2000).
- [8] H. Murakawa and H. Togashi, Continuous models for cell-cell adhesion, *J. Theor. Biol.* **374**, 1 (2015).
- [9] M. J. Simpson, B. J. Binder, P. Haridas, B. K. Wood, K. K. Treloar, D. McElwain, and R. E. Baker, Experimental and modeling investigation of monolayer development with clustering, *Bull. Math. Biol.* **75**, 871 (2013).
- [10] K. K. Treloar, M. J. Simpson, P. Haridas, K. J. Manton, D. I. Leavesley, D. S. McElwain, and R. E. Baker, Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies, *BMC Syst. Biol.* **7**, 137 (2013).
- [11] E. Hinde, F. Cardarelli, M. A. Digman, and E. Gratton, *In vivo* pair correlation analysis of EGFP intranuclear diffusion reveals DNA-dependent molecular flow, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16560 (2010).
- [12] B. J. Binder and M. J. Simpson, Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions, *Phys. Rev. E* **88**, 022705 (2013).
- [13] A. Deutsch and S. Dormann, *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Applications, and Analysis* (Springer Science & Business Media, Berlin, 2007).
- [14] R. J. Friedman, D. S. Rigel, and A. W. Kopf, Early detection of malignant melanoma: The role of physician examination and self-examination of the skin, *CA. Cancer J. Clin.* **35**, 130 (1985).
- [15] M. A. Weinstock, Early detection of melanoma, *JAMA* **284**, 886 (2000).
- [16] A. F. G. Bourke and N. R. Franks, *Social Evolution in Ants* (Princeton University Press, Princeton, NJ, 1995).
- [17] M. J. Keeling, The effects of local spatial structure on epidemiological invasions, *Proc. R. Soc., Ser. B, Biol. Sc., Lond.* **266**, 859 (1999).
- [18] P. J. Diggle, J. Besag, and T. J. Gleaves, Statistical analysis of spatial point patterns by means of distance methods, *Biometrics* **32**, 659 (1976).
- [19] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modeling of Spatial Point Patterns* (John Wiley & Sons, New York, 2008).
- [20] D. J. G. Agnew, J. E. F. Green, T. M. Brown, M. J. Simpson, and B. J. Binder, Distinguishing between mechanisms of cell aggregation using pair-correlation functions, *J. Theor. Biol.* **352**, 16 (2014).
- [21] B. J. Binder and M. J. Simpson, Spectral analysis of pair-correlation bandwidth: Application to cell biology images, *R. Soc. Open Sci.* **2**, 140494 (2015).
- [22] B. J. Binder, J. F. Sundstrom, J. M. Gardner, V. Jiranek, and S. G. Oliver, Quantifying two-dimensional filamentous and invasive growth spatial patterns in yeast colonies, *Plos Comput. Biol.* **11**, e1004070 (2015).
- [23] E. J. Hackett-Jones, K. J. Davies, B. J. Binder, and K. A. Landman, Generalized index for spatial data sets as a measure of complete spatial randomness, *Phys. Rev. E* **85**, 061908 (2012).
- [24] N. A. Bahcall and R. M. Soneira, The spatial correlation function of rich clusters of galaxies, *Astrophys. J.* **270**, 20 (1983).
- [25] A. Donev, S. Torquato, and F. H. Stillinger, Pair correlation function characteristics of nearly jammed disordered and ordered hard-sphere packings, *Phys. Rev. E* **71**, 011105 (2005).
- [26] W. R. Young, A. J. Roberts, and G. Stuhne, Reproductive pair correlations and the clustering of organisms, *Nature* **412**, 328 (2001).
- [27] R. N. Binny, A. James, and M. J. Plank, Collective cell behavior with neighbor-dependent proliferation, death, and directional bias, *Bull. Math. Biol.* **78**, 2277 (2016).
- [28] M. Raghib, N. A. Hill, and U. Dieckmann, A multiscale maximum entropy moment closure for locally regulated space-time point process models of population dynamics, *J. Math. Biol.* **62**, 605 (2011).
- [29] D. J. Bone, H. A. Bachor, and R. J. Sandeman, Fringe-pattern analysis using a 2D Fourier transform, *Appl. Optics* **25**, 1653 (1986).
- [30] M. Takeda, H. Ina, and S. Kobayashi, Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry, *J. Opt. Soc. Am.* **72**, 156 (1982).
- [31] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.97.062104> for the normalizations of the pair correlation functions that were omitted in the main text, as well as a summary of the normalization formulas. All the MATLAB codes which accompany the paper can be found as online supplementary material.
- [32] Q. I. Ouyang and H. L. Swinney, Transition from a uniform state to hexagonal and striped turing patterns, *Nature* **352**, 610 (1991).
- [33] S. Y. Bhide and S. Yashonath, Dependence of the self-diffusion coefficient on the sorbate concentration: A two-dimensional lattice gas model with and without confinement, *J. Chem. Phys.* **111**, 1658 (1999).
- [34] J. D. Murray, *Mathematical Biology: I. An Introduction*, Vol. 17 (Springer Science & Business Media, Berlin, 2007).
- [35] P. Ball, Forging patterns and making waves from biology to geology: A commentary on Turing (1952); The chemical basis of morphogenesis, *Phil. Trans. R. Soc. B* **370**, 20140218 (2015).
- [36] S. Kondo, An updated kernel-based turing model for studying the mechanisms of biological pattern formation, *J. Theor. Biol.* **414**, 120 (2017).
- [37] M. J. Simpson, K. A. Landman, and B. D. Hughes, Multi-species simple exclusion processes, *Phys. A* **388**, 399 (2009).
- [38] C. A. Yates, R. E. Baker, R. Erban, and P. K. Maini, Going from microscopic to macroscopic on nonuniform growing domains, *Phys. Rev. E* **86**, 021921 (2012).
- [39] R. J. H. Ross, C. A. Yates, and R. E. Baker, Inference of cell-cell interactions from population density characteristics and cell trajectories on static and growing domains, *Math. Biosci.* **264**, 108 (2015).
- [40] R. E. Baker, C. A. Yates, and R. Erban, From microscopic to macroscopic descriptions of cell migration on growing domains, *Bull. Math. Biol.* **72**, 719 (2010).
- [41] A. P. Browning, S. McCue, and M. J. Simpson, A Bayesian computational approach to explore the optimal duration of a cell proliferation assay, *Bull. Math. Biol.* **79**, 1888 (2017).
- [42] S. Dini, B. J. Binder, and J. E. F. Green, Understanding interactions between populations: Individual-based modeling and quantification using pair correlation functions, *J. Theor. Biol.* **439**, 50 (2018).
- [43] S. T. Johnston, M. J. Simpson, D. L. S. McElwain, B. J. Binder, and J. V. Ross, Interpreting scratch assays using pair density dynamics and approximate Bayesian computation, *Op. Bio.* **4**, 140097 (2014).

4.2. Conclusions

In this chapter, I presented my work on measuring spatial correlation in discrete domains. The aim of this paper is to improve the analysis of patterns in spatially discrete contexts and allow the study of spatial phenomena for on-lattice models.

In the context of this thesis, the results of this paper are inspired by and applicable to the investigation of the spontaneous aggregation phenomenon described in Section IV-C of Chapter 3. A first step in this direction has been taken recently by Zhang et al. [2019]. The authors adopted the Taxicab PCF, defined in this chapter to quantify the spatial correlation for an excluding, persistent ABM similar to the one introduced in Chapter 3 (see Figure 4-1). As the level of persistence increases, their findings confirm the emergence of positive correlation at short pair distance, which corresponds to cells aggregating. In their paper, Zhang et al. [2019] show that the Taxicab PCF can be used to distinguish persistent agent behaviours and they suggest it could also be applied to infer the level of persistence from experimentally derived data.

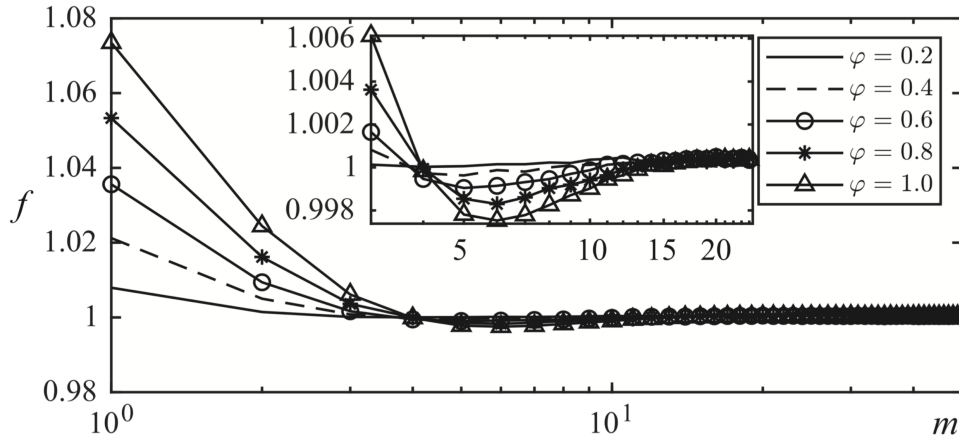


Figure 4-1: Comparison of the Taxicab PCF, f , of an excluding persistent ABM for increasing level of persistence $\varphi = 0.2, 0.4, 0.6, 0.8, 1$. The ABM is initialised uniformly on a 100×100 square lattice with density 0.5. The plots refer the PCF at time $t = 50$ averaged over 1000 independent simulations of the ABM. Reproduced with permission from Zhang et al. [2019].

In another study by Johnston and Crampin [2019] the Taxicab PCF has been adapted to discrete domains containing obstacles. This represents a natural, yet important, extension of our work, as it widens the applications of our discrete PCFs. For example, the formation of ganglia (clusters of glial cells) between nerve strands is highly constrained by the mesh-like structure of the nervous system. A spatial correlation analysis of glial cells, therefore, should take into account the influence of the underlying structured environment [Gabella, 1990, Rühl, 2005]. Johnston and Crampin [2019] use a similar counting argument to those presented in this paper and modify the normalisation factor of Eq. (32) to account for pairs of agents affected by the presence of the obstacles. The authors test their modified PCF with three obstacle patterns of different sizes, both in absence of spatial correlation and in clustering scenarios.

From a more general prospective, the PCFs designed in this chapter should be considered a set of promising tools for all those scenarios in which the level of spatial correlation has to be assessed. For instance, Cutright et al. [2020] have recently cited our paper in the context of microgel packing arrangement. When studying collective cell behaviour there are many situations in which spatial correlations naturally build-up. Among all forms of cell behaviour driving spatial correlation, cell division represents a central example and it will be the main object of study for the remaining part of this thesis.

Although we do not make direct use of PCFs in the remainder of the thesis, the applicability of the tools described in this chapter might lead to new avenues of research which we will discuss in Section 7.1, as part of future work.

Supplementary materials to accompany: “Pair correlation functions for identifying spatial correlation in discrete domains”

Enrico Gavagnin, Jennifer P. Owen and Christian A. Yates

*Department of Mathematical Sciences, University of Bath,
Claverton Down, Bath, BA2 7AY, UK*

This document contains the supplementary materials which accompany the paper of Gavagnin et al. 2018. In Section S.1 a full derivation of the normalisation factor for the Square Uniform PCF using non-periodic boundary conditions (BC) is provided. In Section S.2 we report the expressions of the normalisations for the Triangle PCF and Hexagon PCF, using non-periodic BC. Section S.3 contains the derivations of the normalisations for the Cube Taxicab PCF and the Cube Uniform PCF under non-periodic BC. This supplementary document concludes with a summary of the normalisation factors for all of our PCFs in Section S.4. We refer the reader to the main text for the definitions of all of our PCFs.

S.1. DERIVATION OF THE NORMALISATION FACTOR OF THE SQUARE UNIFORM PCF UNDER NON-PERIODIC BC

In this section we derive the normalisation factor for the Square Uniform PCF, defined in Section III in the main text, under non-periodic BC. This corresponds to computing $\mathbb{E}[\bar{c}_\infty^n(m)]$ of equation (9) of the main text. The approach that we use is similar to the one in Section III B. In particular, by using the equation (19) of the main text and the expression of $s_\infty^p(m)$ given by equation (22b) we only need to determine the expression of the reminder defined as

$$r_\infty(m) = s_\infty^p(m) - s_\infty^n(m). \quad (\text{S.1})$$

We define the set of pairs of sites separated by distance $m \in \mathcal{D}_\infty^n$ that cross the x boundary (horizontal axis) or y boundary (vertical axis), respectively, as

$$P_\infty^x(m) = \left\{ (\mathbf{a}, \mathbf{b}) \in S_\infty^p(m) \mid |y_a - y_b| > L_y - |y_a - y_b| \right\}, \quad (\text{S.2a})$$

$$P_\infty^y(m) = \left\{ (\mathbf{a}, \mathbf{b}) \in S_\infty^p(m) \mid |x_a - x_b| > L_x - |x_a - x_b| \right\}, \quad (\text{S.2b})$$

where $S_\infty^p(m)$ is defined as in equation (6) of the main text. Within these sets, P_∞^x and P_∞^y , we select those separated by $k \in \{1, \dots, m\}$ rows or columns, respectively. We define these subsets as:

$$P_\infty^x(m, k) = \{(\mathbf{a}, \mathbf{b}) \in P_\infty^x(m) \mid L_y - |y_a - y_b| = k\}, \quad (\text{S.3a})$$

$$P_\infty^y(m, k) = \{(\mathbf{a}, \mathbf{b}) \in P_\infty^y(m) \mid L_x - |x_a - x_b| = k\}. \quad (\text{S.3b})$$

Fig. S1 (a) and (b) show an example of pairs of sites within $P_\infty^x(m, m)$.

By following the same steps as in Section III B, we reduce to the expression for the remainder given by

$$\begin{aligned} r_\infty(m) &= |P_\infty^x(m) \cup P_\infty^y(m)| \\ &= \sum_{k=1}^m |P_\infty^x(m, k)| + \sum_{k=1}^m |P_\infty^y(m, k)| - |P_\infty^x(m) \cap P_\infty^y(m)|. \end{aligned} \quad (\text{S.4})$$

To conclude the computation we derive an expression for the two sums in equation (S.4) and the corresponding equation for the size of the intersection. By counting the contribution of each type of pair (see Fig. S1 (c) for a visualisation), one can write down the following expressions for the two sums in equation (S.21):

$$\sum_{k=1}^m |P_\infty^x(m, k)| = 2(L_x + 2L_x + \dots L_x(m-1)) + m(2m+1)L_x, \quad (\text{S.5a})$$

$$\sum_{k=1}^m |P_\infty^y(m, k)| = 2(L_y + 2L_y + \dots L_y(m-1)) + m(2m+1)L_y. \quad (\text{S.5b})$$

Hence

$$\begin{aligned} \sum_{k=1}^m |P_\infty^x(m, k)| + \sum_{k=1}^m |P_\infty^y(m, k)| &= 2(L_x + 2L_x + \dots L_x(m-1)) + m(2m+1)L_x \\ &\quad + 2(L_y + 2L_y + \dots L_y(m-1)) + m(2m+1)L_y \\ &= (L_x + L_y) \left(m(2m+1) + 2 \sum_{i=1}^{m-1} i \right) \\ &= (L_x + L_y) \left(m(2m+1) + 2 \frac{(m-1)m}{2} \right) \\ &= 3(L_x + L_y)m^2. \end{aligned} \quad (\text{S.6})$$

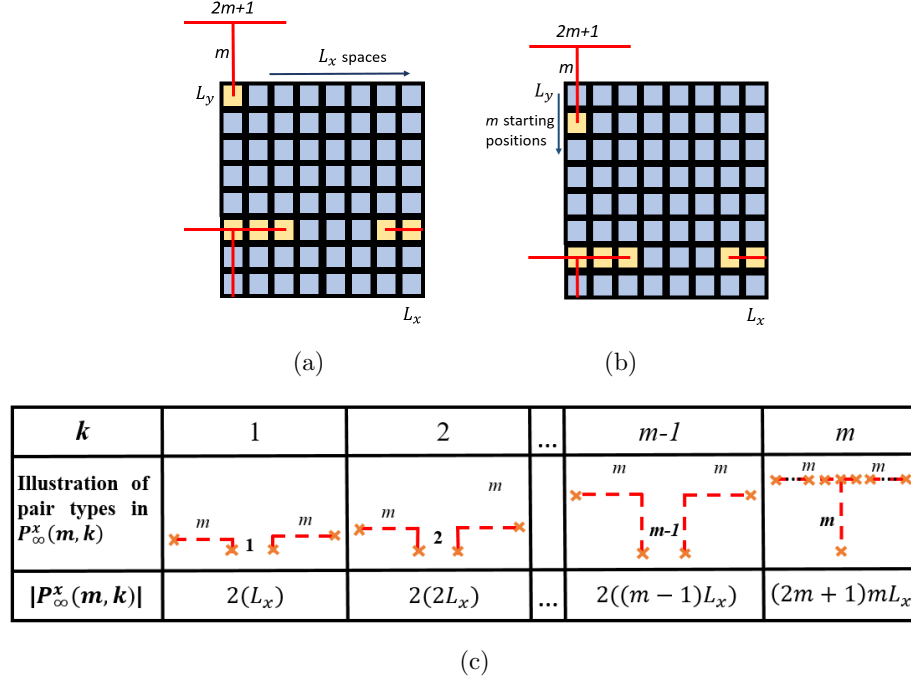


Fig. S1. A visualisation of examples of the pairs of sites in $P_\infty^x(m)$. Panels (a) and (b) show different site pairs in $P_\infty^x(m, m)$. Panel (c) shows all the different types of pairs in $P_\infty^x(m, k)$ for $k = 1, \dots, m$ with the corresponding value of $|P_\infty^x(m, k)|$. For example, for each of the L_x columns, each site in the rows $\{L_y - m + 1, L_y - m + 2, \dots, L_y\}$ has $2m + 1$ corresponding sites at distance m separated by m rows and reached by crossing the horizontal boundary. Therefore $|P_\infty^x(m, m)| = (2m + 1)mL_x$.

In order to calculate of the residue, $r_\infty(m)$, we need to compute the expression for the size of intersection, $|P_\infty^x(m) \cap P_\infty^y(m)|$, in equation (S.4). This counts the pairs of sites separated by distance m by crossing both the x and y boundaries. Fig. S2 shows an example of pairs of sites within $P_\infty^x(4) \cap P_\infty^y(4)$. The number in each site corresponds to the number of sites at distance $m = 4$, reached by crossing both the x and y boundaries. So $|P_\infty^x(4) \cap P_\infty^y(4)|$ is the sum of all the numbers in the coloured sites multiplied by two to account for the second corner region.

Extrapolating, for any value of m , the number of pairs of sites that cross the two bound-

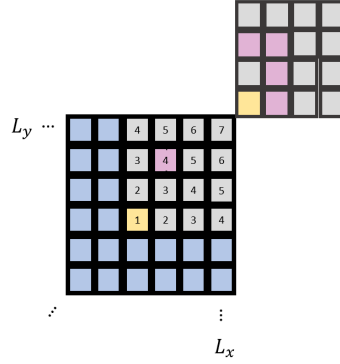


Fig. S2. Examples of pairs in $P_{\infty}^x(4) \cap P_{\infty}^y(4)$ on a zoomed-in corner of a larger domain. Sites in pink and yellow are distance $m = 4$ away from other sites in pink and yellow respectively. The number in each site correspond to the number of sites at distance $m = 4$, reached by crossing the x and y boundaries.

aries is exactly

$$\begin{aligned}
|P_\infty^x(m) \cap P_\infty^y(m)| &= 2 \sum_{i=1}^m \sum_{j=i}^{m+i-1} j \\
&= \sum_{i=1}^m [(m+i)(m+i-1) - i(i-1)] \\
&= \sum_{i=1}^m [m^2 + 2mi - m] \\
&= m^3 - m^2 + m^2(m+1) \\
&= 2m^3.
\end{aligned} \tag{S.7}$$

By substituting equations (S.7) and (S.6) into equation (S.4) we obtain the expression for the remainder $r_\infty(m)$. By rearranging equation (S.1) we determine $s_1^n(m)$, which we then substitute into equation (19) of the main text to obtain the exact expression for the normalisation in the non-periodic case. This is given by

$$\mathbb{E}[c_{\infty}^{U,n}(m)] = \left(\frac{N}{L_x L_y}\right) \left(\frac{N-1}{L_x L_y - 1}\right) \left(4m L_x L_y - 3(L_x + L_y)m^2 + 2m^3\right). \quad (\text{S.8})$$

S.2. NORMALISATION FACTOR FOR THE NON-PERIODIC TRIANGLE PCF AND HEXAGON PCF

Here we provide the expressions of the normalisation factors for the Triangle PCF and the Hexagon PCF defined in Section V, under non-periodic BC.

In both cases, the normalisation can be computed by using equation (19) of the main text. Hence, we only need to provide the expression of the terms $s_{tri}^n(m)$, for the Triangle PCF, and $s_{hex}^n(m)$, for the Hexagon PCF. These correspond to the exact numbers of site pairs at distance m , assuming non-periodic BC, in the two types of tessellations (see Section V of the main text for more details on the definition of the tessellated domain). We omit the derivations of the following expressions which can be obtained with similar steps to the square lattice cases. All the following expressions have been confirmed for a wide range of values of L_x , L_y and m by applying the General PCF (see Section VI of the main text).

Number of site pairs on a triangular tessellation with non-periodic BC

$$s_{tri}^n(1) = 3\frac{L_x L_y}{2} - \frac{L_x}{2} - L_y, \quad (\text{S.9a})$$

$$s_{tri}^n(2) = 3L_x L_y - 2L_x - 4L_y + 2, \quad (\text{S.9b})$$

$$\begin{aligned} s_{tri}^n(m) = & 3m\frac{L_x L_y}{2} - \frac{L_x m^2}{2} \quad (\text{for } m \geq 3) \\ & + L_y \left[2k_6(k_6 - 2k_3 + 1) + k_3(m - 6) - m^2 + m - 2 \right] \\ & + \frac{1}{3}(m - 1)(m^2 - 2m + 6) \\ & - \frac{1}{3}(k_7 + 1)(20k_7^2 + 37k_7 + 12) \\ & - (m - 7 - 4k_7)(k_7 + 1)(m + k_7 - 2), \end{aligned} \quad (\text{S.9c})$$

where $k_j = \left\lfloor \frac{m-j}{4} \right\rfloor$.

Number of site pairs on a hexagonal tessellation with non-periodic BC

$$s_{hex}^n(m) = 3L_x L_y - \frac{1}{4} (7m^2 + k) L_x - 2m^2 L_y + \frac{11m^3}{12} - \frac{(2-3k)m}{12}, \quad (\text{S.10})$$

where $k = m \pmod{2}$.

S.3. DERIVATION OF THE NORMALISATION FACTOR OF THE CUBE TAXI-CAB PCF AND CUBE UNIFORM PCF UNDER NON-PERIODIC BC

Consider a system of agents on a three-dimensional square lattice of size $L_x \times L_y \times L_z$, with lattice step Δ and with the exclusion property that, at any given time, each lattice site can be occupied by at most one agent. If N agents occupy the domain, then the occupancy of the lattice can be represented by a matrix M :

$$M_{xyz} = \begin{cases} 0 & \text{if } (x, y, z) \text{ is vacant,} \\ 1 & \text{if } (x, y, z) \text{ is occupied.} \end{cases} \quad (\text{S.11})$$

where

$$N = \sum_{x=1}^{L_x} \sum_{y=1}^{L_y} \sum_{z=1}^{L_z} M_{xyz} \leq L_x L_y L_z. \quad (\text{S.12})$$

Let ψ be the set of all agent pairs in the lattice i.e.

$$\psi = \{(\mathbf{a}, \mathbf{b}) \in \mathbb{L} \times \mathbb{L} \mid \mathbf{a} = (x_a, y_a, z_a), \mathbf{b} = (x_b, y_b, z_b), \mathbf{a} \neq \mathbf{b}, M_{x_a, y_a, z_a} = M_{x_b, y_b, z_b} = 1\}, \quad (\text{S.13})$$

where $\mathbb{L} = \{1, \dots, L_x\} \times \{1, \dots, L_y\} \times \{1, \dots, L_z\}$ is the set of all sites in the lattice.

In this Section we derive the normalisations for the Cube Taxicab PCF and the Cube Uniform PCF under non-periodic BC. Using the same notation as in Section II of the main text, we define the subsets of agent pairs separated by distance m under non-periodic BC

as

$$C_1^n(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi \mid \|\mathbf{a} - \mathbf{b}\|_1 = m\}, \quad m \in \mathcal{D}_1^n, \quad (\text{S.14a})$$

$$C_\infty^n(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi \mid \|\mathbf{a} - \mathbf{b}\|_\infty = m\}, \quad m \in \mathcal{D}_\infty^n, \quad (\text{S.14b})$$

for the taxicab and uniform metric respectively, where $\mathcal{D}_1^n = \mathcal{D}_\infty^n = \{1, 2, \dots, \max\{L_x, L_y, L_z\} - 1\}$. Using the definitions of the uniform and taxicab metrics, we can express these sets as:

$$C_1^n(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi \mid |x_a - x_b| + |y_a - y_b| + |z_a - z_b| = m\}, \quad (\text{S.15a})$$

$$C_\infty^n(m) = \{(\mathbf{a}, \mathbf{b}) \in \psi \mid \max\{|x_a - x_b|, |y_a - y_b|, |z_a - z_b|\} = m\}. \quad (\text{S.15b})$$

Similarly we define the subsets of agent pairs separated by distance m under periodic BC as

$$C_1^p(m) = \left\{ (\mathbf{a}, \mathbf{b}) \in \psi \mid \min\{|x_a - x_b|, L_x - |x_a - x_b|\} \right. \\ \left. + \min\{|y_a - y_b|, L_y - |y_a - y_b|\} \right. \\ \left. + \min\{|z_a - z_b|, L_z - |z_a - z_b|\} = m \right\}, \quad m \in \mathcal{D}_1^p, \quad (\text{S.16a})$$

$$C_\infty^p(m) = \left\{ (\mathbf{a}, \mathbf{b}) \in \psi \mid \max \left\{ \min\{|x_a - x_b|, L_x - |x_a - x_b|\}, \right. \right. \\ \min\{|y_a - y_b|, L_y - |y_a - y_b|\}, \\ \left. \left. \min\{|z_a - z_b|, L_z - |z_a - z_b|\} \right\} = m \right\}, \quad m \in \mathcal{D}_\infty^p, \quad (\text{S.16b})$$

where $\mathcal{D}_1^p = \mathcal{D}_\infty^p = \{1, 2, \dots, \min\left\{\left\lfloor \frac{L_x}{2} \right\rfloor, \left\lfloor \frac{L_y}{2} \right\rfloor, \left\lfloor \frac{L_z}{2} \right\rfloor\right\}\}$. The schematics in Fig. 12 of the main text represent examples of sites separated by distance $m = 1$ using the taxicab (a) and uniform (b) metrics in three dimensions. The formulae for the number of pairs of sites separated by distance m under periodic BC are given in Section V B of the main text. Here we compute the numbers of pairs of sites separated by distance m under non-periodic BC, *i.e.* $s_1^n(m)$ and $s_\infty^n(m)$, which we can then substitute into the equation (19) in the main text to obtain the normalisation factors of the corresponding PCFs.

Similarly to the two-dimensional cases, we focus on computing an expression the remainder given by

$$r_d(m) = s_d^p(m) - s_d^n(m), \quad (\text{S.17})$$

where $d = 1, \infty$. The expression of $s_d^n(m)$ will then follow by rearrangement.

Let us define the set of pairs of sites separated by distance $m \in \mathcal{D}_1^n$ that cross the y - z plane boundary (*i.e.* $x = 0$), x - z plane boundary (*i.e.* $y = 0$) and x - y plane boundary (*i.e.* $z = 0$), respectively as

$$P_1^{yz}(m) = \{(\mathbf{a}, \mathbf{b}) \in S_1^p(m) \mid |x_a - x_b| > L_x - |x_a - x_b|\}, \quad (\text{S.18a})$$

$$P_1^{xz}(m) = \{(\mathbf{a}, \mathbf{b}) \in S_1^p(m) \mid |y_a - y_b| > L_y - |y_a - y_b|\}, \quad (\text{S.18b})$$

$$P_1^{xy}(m) = \{(\mathbf{a}, \mathbf{b}) \in S_1^p(m) \mid |z_a - z_b| > L_z - |z_a - z_b|\}. \quad (\text{S.18c})$$

For $P_1^{yz}(m)$, $P_1^{xz}(m)$ and $P_1^{xy}(m)$ let us consider those pairs of sites separated by $k \in \{1, \dots, m\}$ sites in the corresponding orthogonal direction, *i.e.* x -, y - and z -direction, respectively. We define these subsets as:

$$P_1^{yz}(m, k) = \{(\mathbf{a}, \mathbf{b}) \in P_1^{yz}(m) \mid L_x - |x_a - x_b| = k\}, \quad (\text{S.19a})$$

$$P_1^{xz}(m, k) = \{(\mathbf{a}, \mathbf{b}) \in P_1^{xz}(m) \mid L_y - |y_a - y_b| = k\} \quad (\text{S.19b})$$

$$P_1^{xy}(m, k) = \{(\mathbf{a}, \mathbf{b}) \in P_1^{xy}(m) \mid L_z - |z_a - z_b| = k\}. \quad (\text{S.19c})$$

Notice that we have $P_d^{yz}(m) = \bigcup_{k=1}^m P_d^{yz}(m, k)$, $P_d^{xz}(m) = \bigcup_{k=1}^m P_d^{xz}(m, k)$ and $P_d^{xy}(m) = \bigcup_{k=1}^m P_d^{xy}(m, k)$. Figs. S.3 (a)-(b) provide a visualisation of pairs of sites within $P_1^{xy}(m, m)$. In Fig. S.3 (c) we show the possible types of pairs of sites in $P_1^x(m, k)$, for $k = 1, \dots, m$. By definitions (S.14) and (S.16) we have that

$$S_d^n(m) \setminus S_d^n(m) = P_d^{yz}(m) \cup P_d^{xz}(m) \cup P_d^{xy}(m), \quad (\text{S.20})$$

where $d = 1, \infty$. Hence, we obtain

$$\begin{aligned} r_d(m) &= |P_d^{yz}(m) \cup P_d^{xz}(m) \cup P_d^{xy}(m)| \\ &= \sum_{k=1}^m |P_d^{yz}(m, k)| + \sum_{k=1}^m |P_d^{xz}(m, k)| + \sum_{k=1}^m |P_d^{xy}(m, k)| \\ &\quad - |P_d^{yz}(m) \cap P_d^{xy}(m)| - |P_d^{xz}(m) \cap P_d^{yz}(m)| - |P_d^{xz}(m) \cap P_d^{xy}(m)| \\ &\quad + |P_d^{yz}(m) \cap P_d^{xz}(m) \cap P_d^{xy}(m)|, \end{aligned} \quad (\text{S.21})$$

for $d = 1, \infty$.

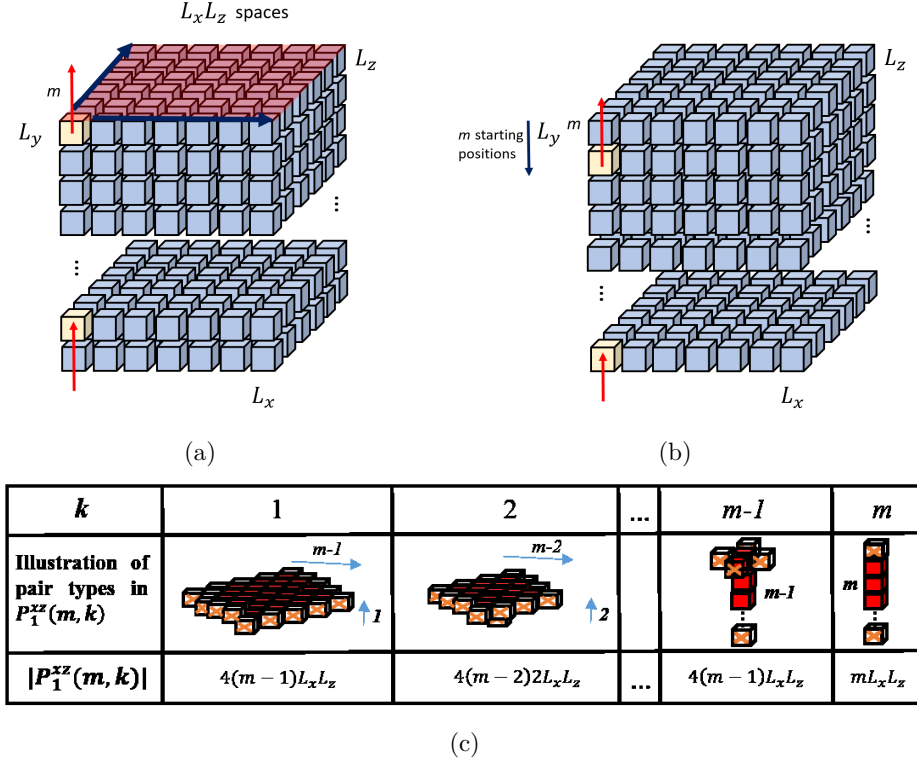


Fig. S.3. A visualisation of the pairs of sites in $P_1^{xz}(m)$. Panels (a) and (b) show two different site pairs in $P_1^{xz}(m, m)$. Panel (c) shows all the different types of pairs in $P_1^{xz}(m, k)$ for $k = 1, \dots, m$ with the corresponding value of $|P_1^{xz}(m, k)|$. For example, each site in the top m horizontal planes, *i.e.* $\{y = L_y - m + 1, y = L_y - m + 2, \dots, y = L_y\}$, has a single corresponding site at distance m separated by m horizontal planes and 0 vertical planes, which is reached by crossing the x - z boundary. Therefore $|P_1^{xz}(m, m)| = mL_xL_y$. The axis orientation is chosen to be consistent with the two-dimensional case.

By counting the contribution of each type of pair (see Fig. S.3 (c) for a visualisation), one can write down the following expressions for the first three sums in equation (S.21):

$$\sum_{k=1}^m |P_1^{ij}(m, k)| = L_i L_j \left(4(m-1) + 2(m-2) + 3(m-3) + \dots + m-1 + m \right), \quad (\text{S.22a})$$

$$\sum_{k=1}^m |P_\infty^{ij}(m, k)| = L_i L_j \left(8m(m-1) + (m-2) + (m-3) + \dots + 1 + (2m+1)^2 \right), \quad (\text{S.22b})$$

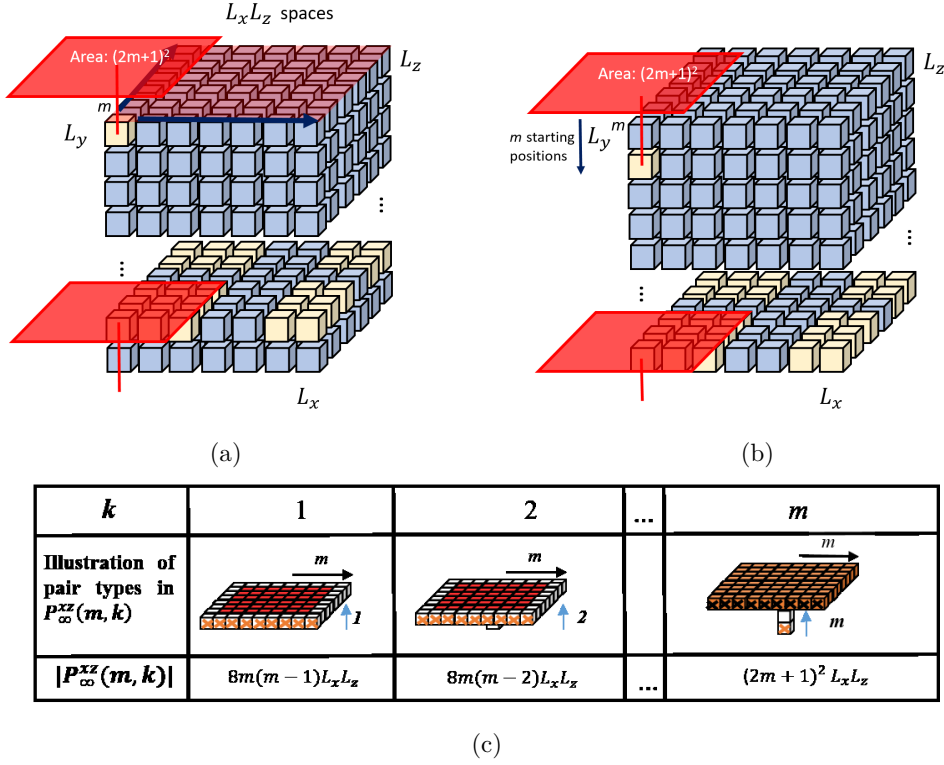


Fig. S4. A visualisation of the pairs of sites in $P_\infty^{xz}(m)$. Panels (a) and (b) show two different site pairs in $P_\infty^{xz}(m, m)$. Panel (c) shows all the different types of pairs in $P_\infty^{xz}(m, k)$ for $k = 1, \dots, m$ with the corresponding value of $|P_\infty^{xz}(m, k)|$. For example, each site in the top m horizontal planes, *i.e.* $\{y = L_y - m + 1, y = L_y - m + 2, \dots, y = L_y\}$, has $(2m + 1)^2$ sites at distance m separated by m horizontal planes which are reached by crossing the x - z boundary. Therefore $|P_\infty^{xz}(m, m)| = (2m + 1)^2 L_x L_z$.

where $ij = yz, xz, xy$. Hence for the case of taxicab metric we have

$$\begin{aligned}
& \sum_{k=1}^m |P_1^{yz}(m, k)| + \sum_{k=1}^m |P_1^{xz}(m, k)| + \sum_{k=1}^m |P_1^{xy}(m, k)| \\
&= \left(4 \sum_{i=1}^{m-1} i(m-i) \right) (L_x L_y + L_y L_z + L_z L_x) \\
&= \frac{2}{3} m(m-1)(m+1) (L_x L_y + L_y L_z + L_z L_x) \\
&= \frac{1}{3} (2m^3 + m) (L_x L_y + L_y L_z + L_z L_x); \quad (\text{S.23})
\end{aligned}$$

and similarly for the uniform metric:

$$\begin{aligned}
& \sum_{k=1}^m |P_{\infty}^{yz}(m, k)| + \sum_{k=1}^m |P_{\infty}^{xz}(m, k)| + \sum_{k=1}^m |P_{\infty}^{xy}(m, k)| \\
&= \left((2m+1)^2 + 8m \sum_{i=1}^{m-1} i \right) (L_x L_y + L_y L_z + L_z L_x) \\
&= m(8m^2 + 1)(L_x L_y + L_y L_z + L_z L_x). \tag{S.24}
\end{aligned}$$

We now focus on deriving an expression for the unknown expressions in (S.21), that is the size of intersections, $P_d^{xy}(m) \cap P_d^{yz}(m)$, $P_d^{xy} \cap P_d^{xz}(m)$ and $P_d^{xy} \cap P_d^{xz}(m)$. These sets consist of pairs of sites separated by distance m that cross *at least* two of the three y - z , x - z , x - y boundaries. Examples of pairs of sites in $P_1^{xz}(m) \cap P_1^{yz}(m)$ are illustrated in Fig. S 5 (a) and Fig. S 5 (b). We show that $P_1^{xz}(m) \cap P_1^{yz}(m)$ consists of all of the pairs of sites located in the top left and right edges (right demonstrated only) which cross the x - z and y - z boundaries of each of the L_z vertical planes making up the cube (see Fig. S 5 (b)). The sites that comprise pairs in $P_1^{xz}(m) \cap P_1^{yz}(m)$ can both be in the same x - y plane or they can be in neighbouring (up to $m-2$) planes away. The number of pairs in these sets can then be calculated by addition, enumerating possible pairs as in Fig. 6 in the main text for taxicab case and Fig. S 2 for the uniform case respectively to give the following. For any $s, t, u \in \{x, y, z\}$, such that $(s, t) \neq (t, u)$

$$\begin{aligned}
|P_1^{st}(m) \cap P_1^{tu}(m)| &= 2L_t \sum_{i=1}^m i^2 (m-i) \\
&= \frac{1}{6} (m-1)m^2(m+1)L_t, \tag{S.25}
\end{aligned}$$

$$\begin{aligned}
|P_{\infty}^{st}(m) \cap P_{\infty}^{tu}(m)| &= 2L_t \left((2m+1)m^3 + 2 \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} ij \right) \\
&= m^2(5m^2 + 1)L_t. \tag{S.26}
\end{aligned}$$

Finally, we compute $|P_d^{yz}(m) \cap P_d^{xz}(m) \cap P_d^{xy}(m)|$, the number of pairs of sites of distance m that cross the y - z , x - z and x - y plane boundaries. Notice that if $m < 3$ no pairs can be connected by crossing all three boundaries, which means that $P_1^{yz}(m) \cap P_1^{xz}(m) \cap P_1^{xy}(m)$ is the empty set. Fig. S 6 is an illustration of pairs of sites in $P_1^{yz}(4) \cap P_1^{xz}(4) \cap P_1^{xy}(4)$. These consist of the pairs of sites in four of the eight corners of the cube. In the case of the taxicab

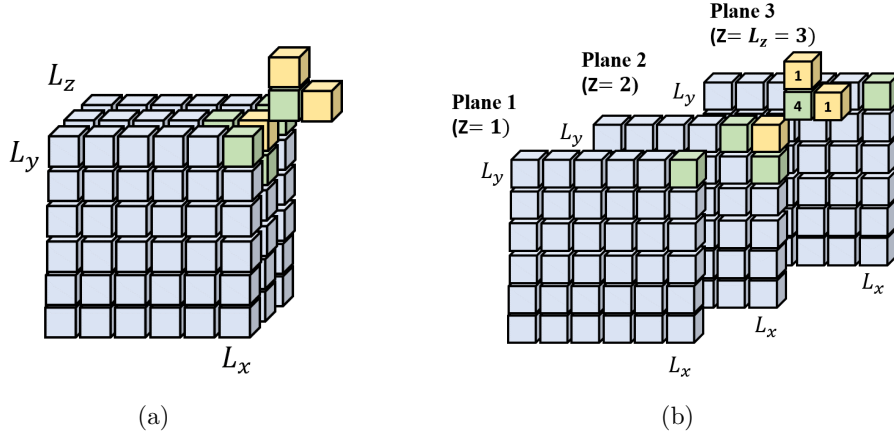


Fig. S5. Examples of pairs of sites separated by distance $m = 3$ that cross both the x - z and y - z plane boundaries, *i.e.* pairs of sites in $P_1^{xz}(3) \cap P_1^{yz}(3)$. Green and yellow sites inside of the cube domain are distance $m = 3$ from green and yellow sites outside of the domain respectively. Panel (a) illustrates an example three-dimensional cube that can be split into L_z vertical planes, panel (b) illustrates how the x - y plane $z = 2$ contributes sites that are distance m from other sites by crossing both the x - z and y - z boundaries.

metric, the number of pairs of sites in $P_1^{yz}(m) \cap P_1^{xz}(m) \cap P_1^{xy}(m)$ can be calculated by first considering the site at the very corner of the cube (marked in green in Fig. S6). There is exactly $Tr(1) = 1$ site at the very corner of the cube where $Tr(x)$ refers to the x th triangular number. This site is distance m from $Tr(m-2)$ sites (yellow sites outside of the domain) by crossing the y - z , x - z and x - y plane boundaries. Next we consider the nearest neighbouring sites to the very corner of the cube. There are $Tr(2) = 3$ of these sites (marked in yellow). Each of these sites are distance m from $Tr(m-3)$ sites by crossing the y - z , x - z and x - y plane boundaries (marked in yellow outside of the domain). We continue this method to count site pairs until we consider sites that separated by $m-3$ sites from the corner site (green site in Fig. S6). Hence we deduce that:

$$\begin{aligned}
 |P_1^{yz}(m) \cap P_1^{xz}(m) \cap P_1^{xy}(m)| &= 4 \sum_{i=1}^{m-2} Tr(i) Tr(m-i-1) \\
 &= 4 \sum_{i=1}^{m-2} \frac{i(i+1)}{2} \cdot \frac{(m-i-1)(m-i)}{2}, \\
 &= \frac{m^5}{30} - \frac{m^3}{6} + \frac{2m}{15}.
 \end{aligned} \tag{S.27}$$

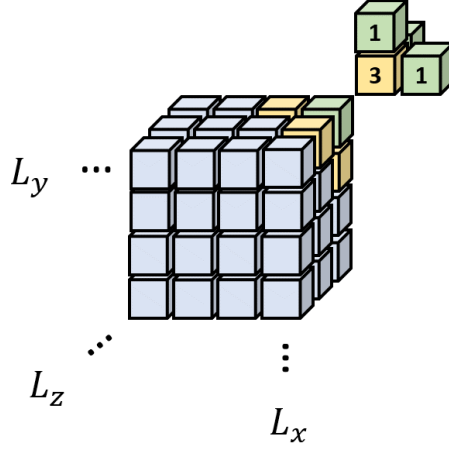


Fig. S6. Examples of pairs of sites separated by distance $m = 4$ that cross the x - y , y - z and x - z boundaries, *i.e.* pairs of sites in $P_1^{xz}(4) \cap P_1^{yz}(4) \cap P_1^{xy}(4)$. Green and yellow sites inside of the cube domain are distance $m = 3$ from green and yellow sites outside of the domain. The numbers correspond to the number of distance $m = 4$ neighbours each site has which lie in $P_1^{xz}(4) \cap P_1^{yz}(4) \cap P_1^{xy}(4)$.

With a similar argument, one can obtain the corresponding expression by using the uniform metric which reads

$$|P_\infty^{yz}(m) \cap P_\infty^{xz}(m) \cap P_\infty^{xy}(m)| = m^3(3m^2 + 1). \quad (\text{S.28})$$

The formulae (S.27) and (S.28) complete the formulae needed for the computation of the remainder of equation (S.21). This can be used together with Eqs. (38) in the main text to rearrange equation (S.17). The resulting expressions for the numbers of pairs at distance m under non-periodic BC in three dimensions, for the taxicab and uniform metrics, respectively, read

$$\begin{aligned} s_{cube_1}^n(m) = & (2m^2 + 1)L_x L_y L_z - \frac{1}{3}(2m^3 + m)(L_x L_y + L_y L_z + L_z L_x) \\ & + \frac{m^2(m^2 - 1)}{6}(L_x + L_y + L_z) - \frac{m^5}{30} + \frac{m^3}{6} - \frac{2m}{15}, \end{aligned} \quad (\text{S.29a})$$

and

$$s_{cube_\infty}^n(m) = (12m^2 + 1)L_x L_y L_z - m(8m^2 + 1)(L_x L_y + L_y L_z + L_z L_x) \\ + m^2(5m^2 + 1)(L_x + L_y + L_z) - m^3(3m^2 + 1). \quad (\text{S.29b})$$

The normalisations for the Cube Taxicab PCF and Cube Uniform PCF under non-periodic BC are obtained by substituting expressions (S.29) into equation (19) in the main text.

S.4. SUMMARY OF NORMALISATION FACTORS

In this section we summarise all of the normalisation factors which have been computed in the paper. For all cases considered, the normalisation, $\mathbb{E}[\bar{c}_d(m)]$, is obtained by equation (19) in the main text. In Table I we report only the counts of pairs of sites separated by distance m , $s_d(m)$, depending on the metric and the BC used.

PCF (d)	Periodic (s_d^p)	Non-periodic (s_d^n)
Square Taxicab (1)	$2mL_xL_y$	$2mL_xL_y - (L_x + L_y)m^2 + \frac{m^3 - m}{3}$
Square Uniform (∞)	$4mL_xL_y$	$4mL_xL_y - 3(L_x + L_y)m^2 + 2m^3$
Triangle (tri)	$\frac{3}{2}mL_xL_y$	<p>for $m = 1$: $3\frac{L_xL_y}{2} - \frac{L_x}{2} - L_y$, for $m = 2$: $3L_xL_y - 2L_x - 4L_y + 2$,</p> <p>for $m \geq 3$: $3m\frac{L_xL_y}{2} - \frac{m^2}{2}L_x + [2k_6(k_6 - 2k_3 + 1) + k_3(m - 6) - m^2 + m - 2]L_y$ $+ \frac{1}{3}(m - 1)(m^2 - 2m + 6) - \frac{1}{3}(k_7 + 1)(20k_7^2 + 37k_7 + 12)$ $-(m - 7 - 4k_7)(k_7 + 1)(m + k_7 - 2)$, where $k_j = \left\lfloor \frac{m-j}{4} \right\rfloor$</p>
Hexagon (hex)	$3mL_xL_y$	$3L_xL_y - \frac{1}{4}(7m^2 + k)L_x - 2m^2L_y + \frac{11m^3}{12} - \frac{(2-3k)m}{12}$ where $k = m \pmod{2}$
Cube Taxicab ($cube_1$)	$(2m^2 + 1)L_xL_yL_z$	$(2m^2 + 1)L_xL_yL_z - \frac{1}{3}(2m^3 + m)(L_xL_y + L_yL_z + L_zL_x)$ $+ \frac{m^2(m^2-1)}{6}(L_x + L_y + L_z) - \frac{m^5}{30} + \frac{m^3}{6} - \frac{2m}{15}$
Cube Uniform ($cube_\infty$)	$(12m^2 + 1)L_xL_yL_z$	$(12m^2 + 1)L_xL_yL_z - m(8m^2 + 1)(L_xL_y + L_yL_z + L_zL_x)$ $+ m^2(5m^2 + 1)(L_x + L_y + L_z) - m^3(3m^2 + 1)$

TABLE I. Summary of the number of sites pairs, $s_d(m)$, according to the metric and BC used.

Chapter 5

The invasion speed of cell migration models with realistic cell cycle time distributions

Within this chapter, we investigate the importance of incorporating realistic representations of cell proliferation in models of cellular invasion. The chapter contains a paper published in *Journal of Theoretical Biology* [Gavagnin et al., 2019] which concerns the impact of realistic cell-cycle time distributions on the speed of an invading cell population.

The majority of the literature connecting microscopic properties of cells with global invasion speeds is based on the assumption that cells divide after an exponentially distributed time. In this study, we analyse experimentally characterised behaviour of mouse fibroblasts in order to show that this assumption does not capture real cell division time distributions correctly. Using both stochastic and deterministic models based on a multi-stage representation of the cell cycle, we derive a series of analytical results which allow us to compute the invasion speed of a cell population with a general cell-cycle time distribution. Our results show that for a general distribution, the average cell-cycle time determines a lower bound for the speed; whereas there is not an upper bound and the invasion can be arbitrary fast, depending on the higher moments of the distribution. However, if we restrict the analysis to a class of biologically realistic distributions, known as hypo-exponential, we find a substantially different scenario: the average cell-cycle determines a bounded interval for the range of invasion speed while the contribution of the higher moments is limited.

5.1. Outline of the article

In Section 1, we introduce our work and we place the paper in the context of the existing literature. In Section 2, we follow a macroscopic approach based on an age-structured model to connect general distributions of the cell-cycle time with the corresponding invasion speeds. Section 3 is dedicated to the multi-stage modelling approach. We define an agent-based model using this modelling technique and, by deriving a set of reaction-diffusion partial differential equations for the average cell densities, we investigate the speed of the invading wavefront. In Section 4 we briefly discuss our findings.

Appendix B: Statement of Authorship

This declaration concerns the article entitled:									
The invasion speed of cell migration models with realistic cell cycle time distributions									
Publication status (tick one)									
draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input checked="" type="checkbox"/>
Publication details (reference)	Journal: Journal of Theoretical Biology, vol. 481, p. 91–99 Authors: Enrico Gavagnin, Matthew J. Ford, Richard L. Mort, Tim Rogers and Christian A. Yates								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The bulk of the calculations have been performed by the author of the thesis (80%). All the numerical computations and simulations have been performed by the author of the thesis (100%). E. Gavagnin, T. Rogers and C.A. Yates contributed equally to the presentation of the content (33%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed						Date	27.4.2020		



The invasion speed of cell migration models with realistic cell cycle time distributions

Enrico Gavagnin^{a,*}, Matthew J. Ford^b, Richard L. Mort^c, Tim Rogers^a, Christian A. Yates^a

^a Department of Mathematical Sciences University of Bath, Claverton Down, Bath, BA2 7AY, UK

^b Centre for Research in Reproduction and Development McGill University, Montréal, H3G 1Y6, Québec

^c Division of Biomedical and Life Sciences Faculty of Health and Medicine Lancaster University, Bailrigg, Lancaster LA1 4YG, UK

ARTICLE INFO

Article history:

Received 11 June 2018

Revised 7 September 2018

Accepted 10 September 2018

Available online 14 September 2018

Keywords:

Cell migration

Multi-stage model

Cell cycle time distribution

Invasion speed

Agent-based model

Travelling wave.

ABSTRACT

Cell proliferation is typically incorporated into stochastic mathematical models of cell migration by assuming that cell divisions occur after an exponentially distributed waiting time. Experimental observations, however, show that this assumption is often far from the real cell cycle time distribution (CCTD). Recent studies have suggested an alternative approach to modelling cell proliferation based on a multi-stage representation of the CCTD.

In this paper we investigate the connection between the CCTD and the speed of the collective invasion. We first state a result for a general CCTD, which allows the computation of the invasion speed using the Laplace transform of the CCTD. We use this to deduce the range of speeds for the general case. We then focus on the more realistic case of multi-stage models, using both a stochastic agent-based model and a set of reaction-diffusion equations for the cells' average density. By studying the corresponding travelling wave solutions, we obtain an analytical expression for the speed of invasion for a general N -stage model with identical transition rates, in which case the resulting cell cycle times are Erlang distributed. We show that, for a general N -stage model, the Erlang distribution and the exponential distribution lead to the minimum and maximum invasion speed, respectively. This result allows us to determine the range of possible invasion speeds in terms of the average proliferation time for any multi-stage model.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Cellular invasion is a process of fundamental importance in numerous morphogenetic and pathological mechanisms. Important examples of processes in which cell migration plays a crucial role include embryonic development (Gilbert, 2003; Keller, 2005), wound healing (Deng et al., 2006; Maini et al., 2004) and tumour invasion (Hanahan and Weinberg, 2000).

Understanding how the properties of the individual cells contribute to the formation and the propagation of the invasion wave is of fundamental importance. In fact, this can reveal the micro-scale mechanisms that are responsible for a given phenomenological aspect, and hence suggest effective therapeutic approaches to inhibit, or enhance, cell migration by interrupting the cell cycle (Gray-Schopfer et al., 2007; Haass and Gabrielli, 2017; Sadeghi et al., 1998).

Despite the large variety of actions and interactions which cells can undergo, there are at least two aspects of cells' behaviour

which are essential in order for the invasion to take place. These are cell motility and cell proliferation (Mort et al., 2016; Simpson et al., 2007). If one of these two aspects does not occur properly, the impact on the collective invasion is typically evident and it can affect the success of the colonisation. For example, Mort et al. (2016) show, using a joint experimental and a modelling approach, that the failure of colonisation of the mouse embryo by melanoblasts in Kit mutants is probably driven by reduced proliferation.

Extensive research has focused on the effect that cell motility and proliferation behaviours have on the speed of the invasion, c . The common approach makes use of simple mathematical models which typically take the form of a stochastic agent-based model (ABM) (Anderson and Chaplain, 1998; Deutsch and Dormann, 2007) or a deterministic partial differential equation (PDE) (Murray, 2007; Wise et al., 2008). By computing the invasion speed of the model, either analytically or numerically, it is possible to link the parameters which modulate the movement and proliferation with the speed of invasion.

Many studies have investigated this link in more general contexts, beginning with the seminal work of Fisher (1937) and

* Corresponding author.

E-mail address: e.gavagnin@bath.ac.uk (E. Gavagnin).

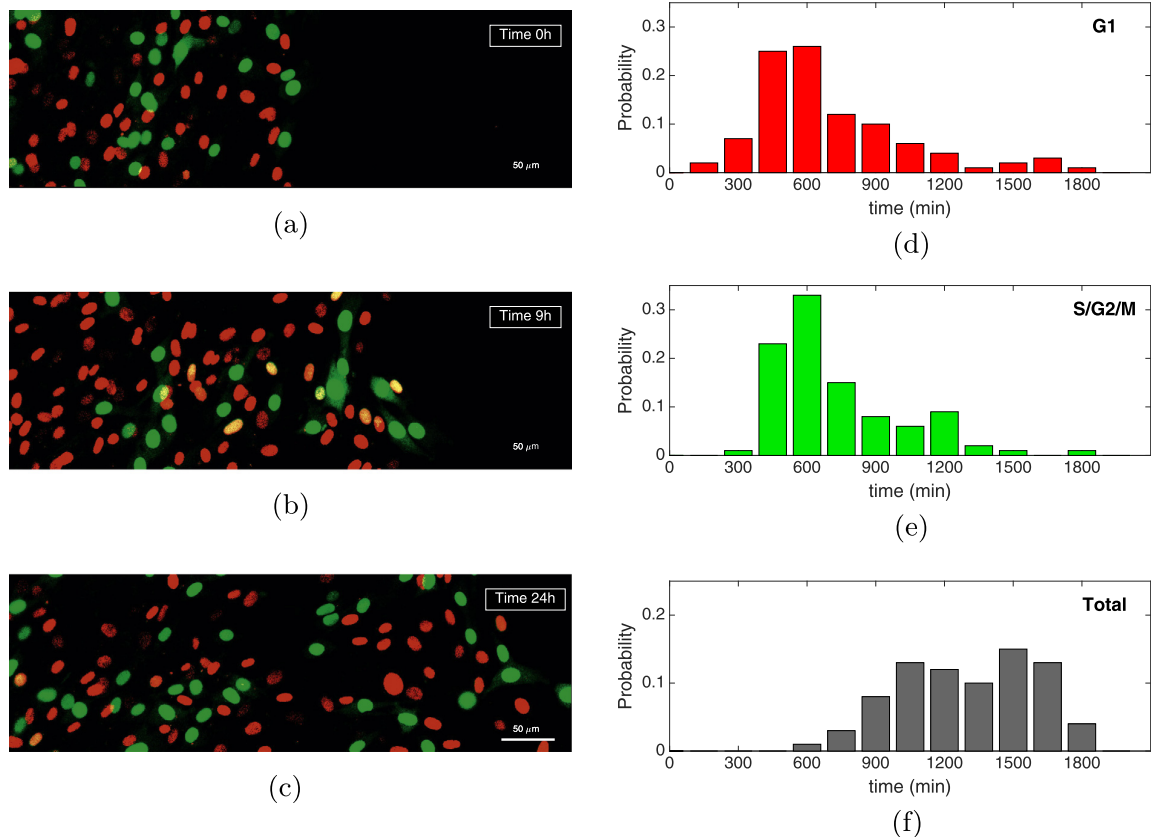


Fig. 1. Panels (a–c): Mouse NIH-3T3 fibroblasts with Fucci2a status migrating into open space (Mort et al., 2014). The Fucci2a system incorporates genetically encoded probes that highlight in red the nuclei of cells in the G1 phase and in green those of cells in one of the other phases (S/G2/M). Panels (d–f): experimental distributions of the time length of the G1 phase (panel (d)), S/G2/M phases (panel (e)) and total CCTD (panel (f)). Both the G1 and S/G2/M distributions show a clear non-monotonicity, which indicates that are not exponentially distributed. To capture both these non-monotonicities using a MSM for the CCTD, a minimum of four stages is required, two for each of the two phases.

Piscounov Kolmogorov and Petrovskii (1991), and more recently more complex models of populations with multiple types or stages (Elliott and Cornell, 2012; Neubert and Caswell, 2000). From these studies, it is well known that, when agents' motility is modelled as diffusion with diffusion coefficient D and proliferation occurs at rates λ , the invasion speed is proportional to the square root of the product of the rates, i.e. $c \propto \sqrt{D\lambda}$ (Fisher, 1937).

It is important to notice that the majority of the literature on the speed of invasion of travelling waves is based on the assumption that proliferation events occur as independent Poisson processes (Mort et al., 2016; Simpson et al., 2007). In the context of cell migration, this is equivalent to assuming that cells proliferate after an exponentially distributed random time. However, experimental observations show that the cell cycle time distribution (CCTD) is typically non-monotonic and differs substantially from an exponential distribution (see Fig. 1 (f) for an example) (Chao et al., 2018; Golubev, 2016; Yates et al., 2017).

There is a vast literature regarding the appropriate representation of the CCTD (Csikász-Nagy et al., 2006; Gérard and Goldbeter, 2009; Powathil et al., 2012). One class of representations, known as multi-stage models (MSMs), have gained particular attention in several recent studies (Chao et al., 2018; Vittadello et al., 2018; Yates et al., 2017). The main idea of MSMs is to partition the cell cycle into N sequential stages. As time evolves, each cell can transit from one stage, i , to the next one, $i + 1$, after an exponentially distributed waiting time with parameter λ_i . When a cell is found at the last stage, N , it can proliferate with rate λ_N , which leads the cell to split into two daughter cells, both initialised at the first

stage. The main motivation that makes MSMs mathematically appealing is the Markov property of the exponentials which simplifies both the analytical investigation of the model and its computational implementation. Moreover, MSMs lead to CCTDs that are hypoexponential and hypoexponential distributions have been shown to provide an excellent agreement with experimental data (Chao et al., 2018; Golubev, 2016; Yates et al., 2017).

Despite the fact that there is some evidence to suggest that the cell cycle comprises a series of uncoupled exponentially distributed phases (Chao et al., 2018), Yates et al. (2017) were at pains to point out that the stages in their MSM do not correspond to the phases in the cell cycle, but are tools which allowed them to fit the correct cell cycle distribution. Similarly, here, we are reticent to link the N stages of our model to N realistic steps in biological cell cycles. Especially since, when fitting to experimental data, different choices of N can give almost equally good agreement to cell cycle distribution data. In particular, the stages of the MSMs should not be confused with the biological phases of the cell cycle which, in general, are not exponentially distributed (see Fig. 1) (Chao et al., 2018).

Whilst previous studies have investigated MSMs extensively in the case of spatially uniform scenarios (Yates et al., 2017; Billy et al., 2014), there is still little understanding about the effect which MSMs have on invading waves of cells. In particular, it is not clear how, and to what extent, a multi-stage representation of the CCTD can impact on the speed of invasion.

The most recent progress on this was made by Vittadello et al. (2018). In their work, the authors derive an

analytical expression for the invasion speed of a 2-stage MSM in terms of the two rates of stage transition, λ_1 , λ_2 , and the diffusion coefficient of cells, D :

$$c = \sqrt{2D \left[-\lambda_1 - \lambda_2 + \sqrt{\lambda_1^2 + 6\lambda_1\lambda_2 + \lambda_2^2} \right]}. \quad (1)$$

The findings of Vittadello et al. (2018) provide useful insights in the qualitative effect of the MSMs. However a general expression for the invasion speed, as in Eq. (1), but for biologically realistic MSMs, which typically have ten or more exponentially distributed stages (Chao et al., 2018; Yates et al., 2017), is not feasible analytically. Hence, there are important questions about the quantitative effect of MSMs on the invasion speed which remain unanswered. In particular, the range of variability in speed for a general N -stage MSM has yet to be studied.

To investigate the effect of incorporating a general CCTD into the invasion models, we follow two distinct approaches. In the first part of the paper, we formulate a generalisation of the Fisher-KPP equation in which the cell population is structured by age. By studying the traveling wave solutions of the model, we derive an implicit equation for the speed of invasion in terms of the Laplace transform of the CCTD. We obtain an expression for the minimum wave speed under this model and show that for a general CCTD the invasion can be arbitrarily fast.

In the second part of the paper, we focus our attention on MSMs. We study a spatially extended ABM which is designed to mimic cell invasion on a regular two-dimensional lattice. For each agent, we implement a general N -stage MSM to simulate the stochastic waiting time before the agent attempts to divide into two daughters. Through a mean-field closure approximation on the average agent density, we derive a system of N reaction-diffusion PDEs which represents a generalisation of the model of Vittadello et al. (2018). By applying the front propagation method (Saarloos, 2003) to the system of PDEs, we reduce the computation of the invasion speed to an eigenvalue problem in terms of the rates of transition between consecutive stages, λ_i . We use this result to study the case of identical transition rates, that corresponds to modelling the CCTD as Erlang. In this case we provide the exact analytical expression for the speed. Finally, we formulate a result for the maximum and minimum speed for a general N -stage MSM.

The paper is organised as follows. In Section 2 we define the age-structured model and we derive the implicit equation for the invasion speed for general CCTD. In Section 3 we define two MSMs: a stochastic ABM and the corresponding mean-field approximation. In Section 3.1 we explain how to apply the front propagation method and we state the eigenvalue problem. We present our results on Erlang distributed cell cycle times and the general hypoexponential case in Section 3.2. We conclude in Section 4 with a brief discussion of this work and future challenges.

2. Age-Structured model

The Fisher-KPP equation implicitly assumes Markov dynamics for the individual cells making up the population, implying a cell cycle time with an exponential distribution (Fisher, 1937). One way to adapt the model to allow for an arbitrary cell cycle time distribution is through the addition of age-structure. Cells have an associated age, denoted by a , which takes values in the positive real numbers and increases as time evolves. Cell motility is modelled as diffusion, with diffusivity D , and they proliferate with an age-dependent rate, $h(a)$.

We can write down a simple linear PDE for the density of cells with age a and spatial location x at time t , $C(a, x, t)$, as follows

$$\begin{aligned} \frac{\partial}{\partial t} C(a, x, t) &= -\frac{\partial}{\partial a} C(a, x, t) + D \frac{\partial^2}{\partial x^2} C(a, x, t) - h(a)C(a, x, t) \\ C(0, x, t) &= 2 \int_0^\infty h(s)C(s, x, t) ds. \end{aligned} \quad (2)$$

The function $h(s)$ is the hazard rate, related to the probability density function $f(s)$ of the age at which cells divide (i.e. the CCTD) via

$$h(s) = \frac{f(s)}{\int_s^\infty f(a) da}, \quad f(s) = h(s) \exp \left(- \int_0^s h(a) da \right). \quad (3)$$

The boundary condition for $C(0, x, t)$ in Eq. (2) gives the density of newborn cells as twice the total rate of cell division. Note that we have neglected from our formulation in system (2) any non-linear terms arising from crowding effects, as these are not relevant to the speed of the front propagation. This model is a simple spatial adaptation of the McKendrick-Von Foerster equation for growing age-structured populations, and has been studied before (Al-Omari and Gourley, 2002; Gabriel et al., 2012; Webb and Webb, 1985; Billy et al., 2014).

As our first result, we show that the speed of propagation for the model (2) is determined by the Laplace transform of the CCTD, defined by

$$\mathcal{L}\{f\}(s) = \int_0^\infty e^{-sa} f(a) da. \quad (4)$$

Theorem 1 If $\lim_{s \rightarrow \infty} \mathcal{L}\{f\}(s) < 1/2$ then the PDE (2) admits travelling wave solutions with propagation speed $c > 2\sqrt{D\lambda}$, where $\lambda > 0$ is the unique solution to

$$\mathcal{L}\{f\}(\lambda) = 1/2. \quad (5)$$

Proof. The system (2) is separable, hence we seek solutions of the form $C(a, x, t) = v(a)w(x - ct)$, corresponding to a travelling wave with speed c and internal age structure given by v . Inserting into (2) and rearranging, we find

$$c \frac{w'}{w} + D \frac{w''}{w} = \frac{v'}{v} + h. \quad (6)$$

The left-hand side here is a function only of $x - ct$, whilst the right-hand side is a function only of a . We thus determine that both are equal to a constant, say $-\lambda$. The w equation becomes

$$\lambda w + cw' + Dw'' = 0, \quad (7)$$

which is well-known as the linearisation of the Fisher-KPP equation, admitting travelling wave solutions for all $c > 2\sqrt{D\lambda}$. The equation for v has solution

$$v(a) = v(0) \exp \left(-a\lambda - \int_0^a h(\alpha) d\alpha \right). \quad (8)$$

The boundary condition then gives us

$$1 = 2 \int_0^\infty h(a) \exp \left(-a\lambda - \int_0^a h(\alpha) d\alpha \right) da, \quad (9)$$

from which the definition of the hazard rate, Eq. (3), gives the result $1 = 2\mathcal{L}\{f\}(\lambda)$. Uniqueness of the solution (when one exists) follows from the monotonicity of the Laplace transform of a probability density. \square

We can use the previous result to investigate the range of speeds for an arbitrary CCTD with a given positive mean, $\bar{\mu} > 0$. By using Jensen's inequality we have that for any positive supported f with mean $\bar{\mu}$

$$\mathcal{L}\{f\}(\lambda) \leq e^{-\lambda\bar{\mu}} = \mathcal{L}\{\delta_{\bar{\mu}}\}(\lambda), \quad (10)$$

where $\delta_{\bar{\mu}}$ is the Dirac delta function concentrated at $\bar{\mu}$. From the monotonicity of the Laplace transform of a probability density, it follows that the minimum speed is obtained by using $f = \delta_{\bar{\mu}}$, which gives

$$c \geq 2\sqrt{\frac{D \ln 2}{\bar{\mu}}} \quad (11)$$

We now use Theorem 1 to show that there is no upper bound for the speed of invasion of a general CCTD with a given mean. Consider the set of probability density functions defined as

$$f_\varepsilon(x) = \frac{1}{2}(\delta_\varepsilon \bar{\mu} + \delta_{(2-\varepsilon)\bar{\mu}}), \quad (12)$$

where $\varepsilon \leq 1$. It follows immediately that each member of this set of functions have mean $\bar{\mu}$ and Laplace transform given by:

$$\mathcal{L}\{f_\varepsilon\}(\lambda) = \frac{1}{2}(e^{-\lambda\varepsilon\bar{\mu}} + e^{-\lambda(2-\varepsilon)\bar{\mu}}). \quad (13)$$

By substituting the expression (13) into Eq. (5) and rearranging, we obtain the implicit equation for λ given by

$$\lambda\varepsilon\bar{\mu} = -\ln(1 - e^{-2\lambda\bar{\mu}}). \quad (14)$$

The right-hand side of Eq. (14) is a strictly decreasing function of λ that converges to 0 as $\lambda \rightarrow \infty$. Therefore, we can always choose ε small enough so that the solution of Eq. (14) is arbitrarily large.

This demonstrates that, assuming that the CCTD is a general function with mean $\bar{\mu}$ and positive support, the range of possible invasion speeds is given by

$$c \in \left[2\sqrt{\frac{D \ln 2}{\bar{\mu}}}, \infty \right). \quad (15)$$

The result in Theorem 1 is important because it establishes the connection between a general CCTD and the corresponding invasion speed. However, for some particular classes of distributions, solving Eq. (5) analytically can be challenging and the method of this Section does not provide any deeper insights. In particular, this is true for hypoexponential distributions, which are of special interest in the context of cell proliferation. In remaining part of the paper we further explore this class of distributions using a MSM of cell migration.

3. Multi-Stage models

In this section we introduce the two MSMs that we will use throughout the remainder of this paper. Firstly, we define a discrete ABM, in which the multi-stage representation of the CCTD is implemented as a stochastic feature of each cell at the microscale. Secondly, we introduce a system of deterministic PDEs describing the average cell density in a macroscopic manner.

The ABM We consider a continuous-time ABM on a two-dimensional regular square lattice, with a given spacing denoted by Δ . Each cell is modelled as a single agent which moves and proliferates. Volume exclusion is incorporated by allowing at most one agent to occupy a given lattice site.

Agents are moved to a simple excluding random walk on the lattice. Each agent attempts a movement after an exponentially distributed waiting time with rate α . When this happens, a new position is chosen uniformly from one of the four nearest neighbouring sites and the movement takes place only if the selected site is empty. The event is aborted otherwise.

We implement cell proliferation using a MSM. We divide the cell cycle into N sequential stages. Agents at one of the first $N-1$ stages, $i = 1, \dots, N-1$, move to the next stage after an exponentially distributed waiting time of rate λ_i . Agents at the last stage, N , can attempt a proliferation event, after a further exponentially distributed waiting time of rate λ_N . In order to attempt a proliferation event, a target site is selected uniformly at random from one of the four nearest neighbouring sites. If such site is empty, a new first-stage agent is located on it, and the proliferating agent is returned to the first stage. If the target site is occupied, the proliferation event is aborted and the proliferating agent remains at the last stage¹.

We simulate the cell invasion by populating the first 10 columns of the lattice with agents at stages that are chosen uniformly at random. We impose zero flux boundary conditions on the x -direction and periodic boundary conditions on the y -direction. Agents are displaced uniformly at random in the vertical direction, so we can reduce the dimensionality of the problem by considering the average column density (Simpson et al., 2009).

The PDE model Here we define the continuous model for the average column density which will be the object of the wavespeed analysis.

We denote by $S_i(x, t)$ the density of i -stage agents in the column x at time t , averaged over multiple realisations of the ABM. Let $C(x, t)$ be the total density of column x at time t , i.e.

$$C(x, t) = \sum_{i=1}^N S_i(x, t). \quad (16)$$

By writing down the master equation of S_i , for $i = 1, \dots, N$ and taking the limit as $\Delta \rightarrow 0$, while keeping $\alpha\Delta^2$ constant, one can derive a system of reaction-diffusion PDEs for the column densities of the different stages:

$$\begin{cases} \frac{\partial S_1}{\partial t} = D \frac{\partial}{\partial x} \left[(1-C) \frac{\partial S_1}{\partial x} + S_1 \frac{\partial C}{\partial x} \right] + 2\lambda_N(1-C)S_N - \lambda_1 S_1 \\ \frac{\partial S_i}{\partial t} = D \frac{\partial}{\partial x} \left[(1-C) \frac{\partial S_i}{\partial x} + S_i \frac{\partial C}{\partial x} \right] + \lambda_{i-1} S_{i-1} - \lambda_i S_i \\ \text{for } i = 2, \dots, N-1 \\ \frac{\partial S_N}{\partial t} = D \frac{\partial}{\partial x} \left[(1-C) \frac{\partial S_N}{\partial x} + S_N \frac{\partial C}{\partial x} \right] + \lambda_{N-1} S_{N-1} - \lambda_N(1-C)S_N, \end{cases} \quad (17)$$

where $D = \lim_{\Delta \rightarrow 0} \frac{\alpha\Delta^2}{4}$. Notice that other types of tessellations than the regular square lattice are common in the literature (Deutsch and Dormann, 2007; Simpson et al., 2018). However, the model formulation and the corresponding mathematical analysis in these cases do not change substantially. For example, a detailed derivation for the three-stage model on an hexagonal lattice can be found in Simpson et al. (2018). Moreover, we should underline that the diffusivity of the cells in our model is independent of their stage which is not always true for real cells. We discuss this and other possible generalisations of our model in Section 4.

System (17) consists of a set of N reaction-diffusion PDEs with non-linearities in both the diffusion and the proliferation terms due to the effect of volume exclusion. Specifically, the term $(1-C)$ accounts for the reduction in rate due to volume exclusion. Notice that by summing all the equations in (17), we obtain

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} + \lambda_N(1-C)S_N. \quad (18)$$

In other words, although the diffusion terms in each of the Eq. (17) are non-linear, the motility at the population-level is simple diffusion (Simpson et al., 2009). Conversely, due to the dependence of Eq. (18) on S_N , it is not possible to obtain a closed PDE for the total agent density without further assumptions.

We conclude this section by showing a comparison of the two models in Fig. 2. In the example, we choose realistic values of motility rate, α , and proliferation rate, $\bar{\lambda}$, as in Treloar et al. (2013) and Haass et al. (2014), respectively. We consider an ABM with five stages with increasing rates (the

¹ Alternatively, we could choose to return the proliferating agent to the first stage every time an abortion occurs. This model has been studied in Yates et al. (2017) for homogeneously distributed agents. This modification does not substantially change our results. This is because our analysis of the speed of the wave front is based on low density regions, where abortion of events does not play an important role. For this reason, we decided to focus only on the stated version of the model.

stage-to-stage transition rates are chosen to facilitate the visualisation of the different density profiles). In panels (a), (b) and (c) three successive snapshots are shown and the formation of the travelling wave appears clearly. As previously observed by Vittadello et al. (2018), due to the presence of volume exclusion, the travelling wave solutions of the N subpopulations of cells are of two qualitatively different types. The density profile of the first $N - 1$ subpopulations have the form of moving pulses located at the front of the total wave with the amplitude which depends on the rate of the corresponding stage. The profile of the last stage subpopulation, instead, appears as a moving wavefront which dominates the density at the back of the total wave.

The numerical solutions of the PDEs agree well with the average behaviour of the ABM. Therefore, we focus our attention on the speed of the PDE model which we can investigate using an analytical approach (see Section 3.1).

It is important to note that the quantitative validity of our results on the PDE model will extend to the ABM only for the range of parameters which preserves the good agreement between the two models. In Fig. 3 we compare the total averaged column density profiles of the ABM and PDE for different parameters. The heat map shows the histogram distance error (HDE)² between ABM and the PDE model for different rates of movement and proliferation. When the rate of proliferation is large compared to the motility rate, the mean-field approximation loses its accuracy. This is a well known phenomenon which is caused by the presence of strong spatial correlations between occupied sites, induced by the proliferation (Middleton et al., 2014). Increasing the motility parameter tends to break up spatial correlations of neighbouring sites and, consequentially, to improve the accuracy of the mean-field approximation.

The results of Figs. 2 and 3 confirm that for realistic choice of parameters ($\alpha \approx 4$ and $\tilde{\lambda} \approx 0.02$ (Haass et al., 2014; Treloar et al., 2013)) the PDE model provides a good approximation of the ABM. This motivates us to focus our analysis on the continuum model. It is possible to derive more accurate descriptions in those cases where the agreement is lost using higher order moment closure schemes (see for example Baker and Simpson (2010); Markham et al. (2013)), but this is beyond the scope of this paper.

3.1. Wavespeed analysis

In this Section we apply the front propagation method of Saarloos (2003) to system (17) to study the speed of invasion of the PDE model.

The system of Eq. (17) has two equilibria, an unstable empty state, $S_i(x, t) \equiv 0$ for $i = 1, \dots, N$, and a stable occupied state, $S_i(x, t) \equiv 0$ for $i = 1, \dots, N - 1$ and $S_N(x, t) \equiv 1$. Firstly we linearise the system about the unstable steady state, giving

$$\begin{cases} \frac{\partial S_1}{\partial t} = D \frac{\partial^2 S_1}{\partial x^2} + 2\lambda_N S_N - \lambda_1 S_1 \\ \frac{\partial S_i}{\partial t} = D \frac{\partial^2 S_i}{\partial x^2} + \lambda_{i-1} S_{i-1} - \lambda_i S_i \end{cases} \quad \text{for } i = 2, \dots, N. \quad (19)$$

We substitute

$$S_i(x, t) \propto \exp(-i\omega(k)t + ikx),$$

into Eq. (19), where i is the imaginary unit, $\omega(k)$ is the dispersion angular frequency of the Fourier modes and k is the spatial wavenumber. Upon simplification, we obtain

$$\begin{cases} -i\omega(k) S_1 = -Dk^2 S_1 + 2\lambda_N S_N - \lambda_1 S_1 \\ -i\omega(k) S_i = -Dk^2 S_i + \lambda_{i-1} S_{i-1} - \lambda_i S_i \end{cases} \quad \text{for } i = 2, \dots, N.$$

² The HDE between two normalised histograms with values a_i and b_i at point i (i.e. $\sum a_i = \sum b_i = 1$) is defined as $HDE = \sum |a_i - b_i|/2$ (Cao and Petzold, 2006).

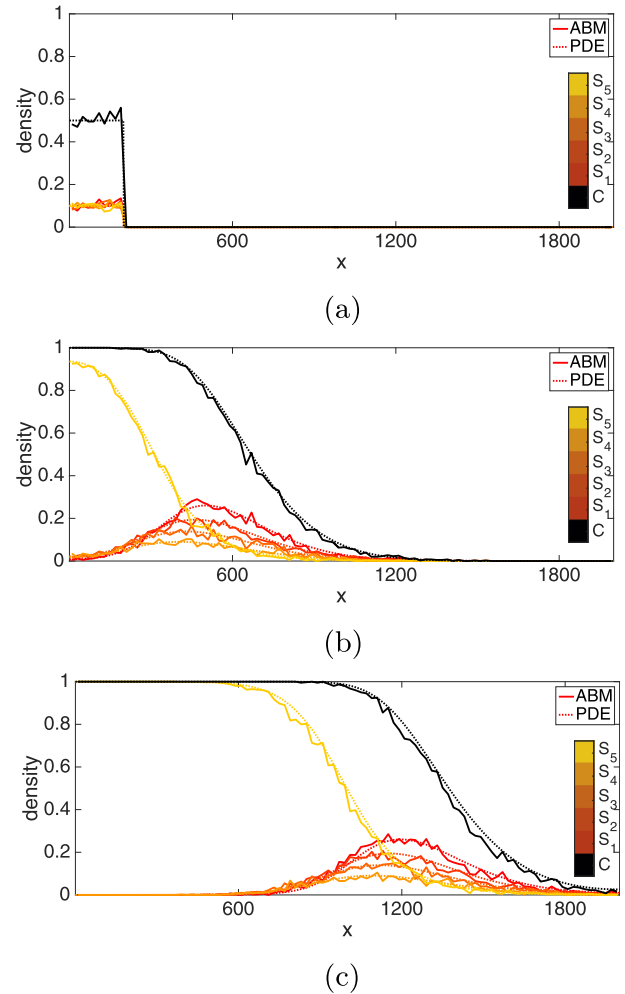


Fig. 2. Comparison between the average column density for the ABM (full lines) and the PDE model (dotted lines) with a five-stage MSM. The panels show three snapshots of the evolution of the two models at time 0 (a), 150 (b) and 300 (c). In all cases, the profiles for the five different subpopulations are shown in different gradations of orange and the total density is plotted in black. The ABM profiles are obtained by averaging over 20 identically prepared simulations on a 2000×400 lattice. The other parameters of the models are $\Delta = 20$, $\alpha = 4$, $\tilde{\lambda} = 0.0233$, $\lambda_1 = 0.15$, $\lambda_2 = 0.19$, $\lambda_3 = 0.25$, $\lambda_4 = 0.37$ and $\lambda_5 = 0.75$.

Following the front propagation method (Saarloos, 2003), the expression of the wave speed, c , is given by

$$c = \frac{\text{Im}[\omega(k^*)]}{\text{Im}[k^*]}, \quad (20)$$

where $k^* = iq$, with q real, and such that

$$\frac{d\omega}{dk}(k^*) = \frac{\text{Im}[\omega(k^*)]}{\text{Im}[k^*]}. \quad (21)$$

Notice that we can write down $i\omega(k)$ in the form

$$i\omega(k) = k^2 D - \rho, \quad (22)$$

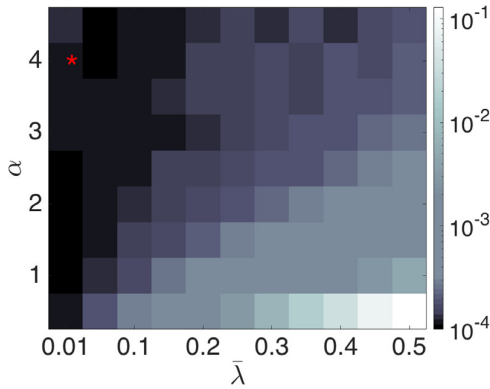


Fig. 3. Histogram distance error (HDE) between column density profiles of ABM and the mean-field model for different parameters of movement and proliferation. The colour of each square denotes the HDE between the total density of the PDE and the ABM (averaging over 1000 simulations) as described in the legend. Dark squares represent small HDE which denotes good agreement between the two models, whereas light squares represent large HDE which denote a loss of agreement. The models are simulated on a 2000×400 domain with $\Delta = 20$ and the HDE is computed at the time when half of the domain in the mean-field model is occupied. All the ABMs are simulated with $N = 5$ and with transition rates proportional to the one of Fig. 2. The red mark denotes the realistic parameter choice corresponding to Fig. 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where ρ is an eigenvalue of the matrix

$$\Lambda = \begin{bmatrix} -\lambda_1 & 0 & \dots & 0 & 2\lambda_N \\ \lambda_1 & -\lambda_2 & 0 & \dots & 0 \\ 0 & \lambda_2 & -\lambda_3 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & \lambda_{N-1} & -\lambda_N \end{bmatrix}. \quad (23)$$

From expression (22) it follows that

$$\frac{d\omega}{dk}(k^*) = 2qD, \quad (24a)$$

$$\frac{\text{Im}[\omega(k^*)]}{\text{Im}[k^*]} = \frac{q^2 D + \text{Re}[\rho]}{q}. \quad (24b)$$

By substituting Eqs. (24) into (21), we obtain $q^2 = \text{Re}[\rho]/D$. Note that Λ is Metzler, and hence by Perron-Frobenius its rightmost eigenvalue is real. Hence, from (20), the speed of the invasion is given by

$$c = 2\sqrt{D\rho}, \quad (25)$$

where ρ is the maximum real eigenvalue of Λ , defined in terms of the characteristic polynomial of the matrix Λ , $\mathcal{P}_\Lambda(x)$, as follows

$$\rho(\Lambda) = \max \{x \in \mathbb{R} \mid \mathcal{P}_\Lambda(x) = 0\}. \quad (26)$$

This shows that the problem of finding the speed of invasion of the PDE model is equivalent to computing the maximum real eigenvalue of the matrix Λ , $\rho(\Lambda)$.

3.2. Results

The characteristic polynomial of the matrix Λ can be computed directly from the matrix and it reads

$$\mathcal{P}_\Lambda(x) = \prod_{i=1}^N (\lambda_i + x) - 2 \prod_{i=1}^N \lambda_i. \quad (27)$$

In general, an analytical formula of the roots of the polynomial function $\mathcal{P}_\Lambda(x)$ is not available. In this section we first consider the case of $\lambda_i = \lambda$ for $i = 1, \dots, N$ for which the maximum eigenvalue $\rho(\Lambda)$ can be computed analytically. This corresponds to a special case of the general hypoexponential distribution, known as the Erlang distribution. We conclude by proving a theorem in which we state the range of speed variability for the general hypoexponential CCTD.

The Erlang distribution Consider the case $\lambda_i = \lambda$ for $i = 1, \dots, N$, which corresponds the Erlang CCTD. Under this assumption, we can write down the characteristic equation of the matrix Λ , using formula (27), as

$$(\lambda + x)^N = 2\lambda^N. \quad (28)$$

The eigenvalues of Λ are then given by the solutions of Eq. (28) which are $x_j = \lambda(\xi^j \sqrt[N]{2} - 1)$ for $j = 1, \dots, N$, where $\xi = \exp(2\pi i/N)$ is the primitive N -th root of unity. Hence, we obtain that

$$\rho(\Lambda) = \lambda(\sqrt[N]{2} - 1). \quad (29)$$

By substituting the expression (29) into Eq. (25) we obtain the formula for the speed of invasion for the model with Erlang distribution

$$c = 2\sqrt{D\lambda(\sqrt[N]{2} - 1)}. \quad (30)$$

Notice that for $N = 1$, which corresponds to exponential CCTD, we recover the well known expression of the speed for the Fisher-KPP equation, $2\sqrt{D\lambda}$.

The general case For the case of a general hypoexponential distribution, there is no analytical formula for the expression of the maximum real eigenvalue of the matrix Λ . However, we find that the Erlang case and the exponential case, for which we do have the analytical formula of the speed, correspond to the lower and upper bound (respectively) for the speed of travelling waves with hypoexponential CCTD and a given total proliferation rate, $\bar{\lambda}$. This result follows directly from the following theorem on the range of $\rho(\Lambda)$.

Theorem 2 Let $\rho(\Lambda)$ be defined by Eq. (26) as the maximum real eigenvalue of the matrix Λ . Then

$$\bar{\lambda}N(\sqrt[N]{2} - 1) \leq \rho(\Lambda) < \bar{\lambda}, \quad (31)$$

where $\bar{\lambda} = (\sum_{i=1}^N 1/\lambda_i)^{-1}$.

A proof of Theorem 2 can be found in the appendix. It is immediate to interpret the result of the Theorem 2 in terms of invasion speeds. In particular, by using Eq. (25), together with the two inequalities (31), we deduce that the speed of the invasion of the PDE model with diffusion coefficient D and a general N -stage representation of the CCTD with total growth rate given by $\bar{\lambda}$, lies in the interval

$$c \in \left[2\sqrt{D\bar{\lambda}N(\sqrt[N]{2} - 1)}, 2\sqrt{D\bar{\lambda}} \right). \quad (32)$$

We can generalise this result even further by taking the limit as $N \rightarrow \infty$ in the right-hand side of Eq. (32). Hence we obtain a general interval which holds for any multi-stage representation, regardless of the number of stages, which reads

$$c \in \left(2\sqrt{D\bar{\lambda} \ln 2}, 2\sqrt{D\bar{\lambda}} \right), \quad (33)$$

where we used $N(\sqrt[N]{2} - 1) = \ln 2 + \mathcal{O}(N^{-1})$.

Notice that the lower bound of the interval (33) is equivalent to the lower bound for the general CCTD, obtained in (15) of Section 2. This can be intuitively understood by observing that, as we let number of stages of an hypoexponential distribution go to infinity while keeping the total rate, $\bar{\lambda}$, fixed, the variance of the

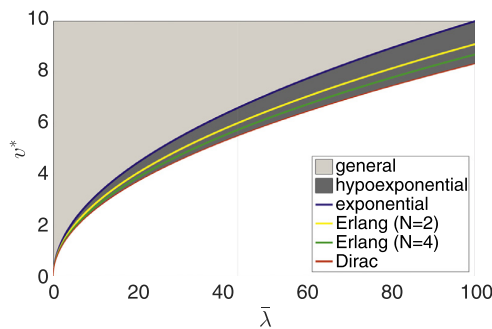


Fig. 4. Illustration of the range of invasion speeds for a given mean proliferation rate, $\bar{\lambda}$, and fixed diffusion coefficient, $D = 1$. The two coloured regions represent the range of speed for a general CCTD. The dark grey subregion highlights the range of speeds for hypoexponential CCTDs. The global minimum speed is obtained by using the Dirac distribution (red line). The exponential CCTD (blue line) is the hypoexponential distribution which leads to maximum speed. There is no upper bound for the general case. Two examples of Erlang CCTDs with two stages (yellow line) and four stages (green line) are also shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distribution tends to zero. Consequently, the distribution converges to a Dirac function concentrated in the mean, $\bar{\mu} = \bar{\lambda}^{-1}$, which we have proved in Section 2 to be the distribution corresponding to the minimum invasion speed. In Fig. 4 we summarise our findings about the range of invasion speed for different CCTD through a graphical representation.

4. Conclusion

In this work we investigated the quantitative effect of implementing a realistic CCTD into models of cell invasion. Firstly, we derived a general result from a generalised version of the Fisher-KPP equation. Then we investigated the case of MSMs by implementing a simple ABM of cells undergoing undirected migration and proliferation by division, in which the time between successive divisions is modelled using a multi-stage representation (i.e. the CCTD is hypoexponential). By studying a continuous version of the ABM, we connected the type of CCTD to the speed of the corresponding invasion.

The results indicate that, for a fixed mean division time, the minimum speed of invasion is obtained by the Dirac distribution, while there is no upper bound. In other words, the invasion can be, in general, infinitely fast. However, when we focus our attention to the case of MSMs, which are known to represent well the experimental CCTD, our analysis shows that the speed can vary in a bounded interval (see Fig. 4). More precisely, we show that the maximum invasion speed is reached by adopting an exponential CCTD, which leads to the classic Fisher-KPP model. On the other hand, the minimum speed is obtained by partitioning the CCTD into multiple exponential stages with identical rates, which corresponds to the case of Erlang CCTD. Finally, by considering the limiting case of infinitely many stages, we find that the infimum value of the speed for the class of hypoexponential CCTD coincides with the global minimum for a general CCTD.

The results indicate the invasion speed changes with the variance of the CCTD, i.e. decreasing the variance in the proliferation time distribution leads to slower invasion. We found that the maximum reduction in comparison to the classical formula for the Fisher-KPP model, is given by a multiplicative factor of $\sqrt{\ln 2} \approx 0.83$. Whilst interpreting this result in the context of experimental data is beyond the aim of this work, we want to stress that for number of stages $N \gg 1$, which is typically the case for experimentally observed distributions (Chao et al., 2018; Golubev, 2016; Yates et al., 2017), the speed converges to the lower bound of Eq.

(33) with order given by $\mathcal{O}(N^{-1})$. This suggests that, with the only information of the mean of the CCTD (equivalently, the total rate), including the factor $\sqrt{\ln 2}$ in the formula for the speed leads to a more accurate estimation than the classic expression of Fisher-KPP.

In Section 3 we used a discrete ABM, but it is important to notice that alternative modelling approaches might lead to different results. Although discrete space ABMs are widespread in the literature (Cheeseman et al., 2014; Deutsch and Dormann, 2007; Mort et al., 2016; Simpson et al., 2018, 2007; Vittadello et al., 2018), a considerable number of studies focus on lattice-free ABMs in which cells' positions are not constrained to a grid (Dyson and Baker, 2014; Dyson et al., 2012; Grima, 2008; Matsiaka et al., 2017; Middleton et al., 2014). Another alternative approach, known as compartment-based model, consists in allowing multiple cells occupying a single lattice site (Cianci et al., 2017; Taylor et al., 2016; 2015). In the context of our work, adopting alternative modelling regimes, such as lattice-free or compartment-based models, would lead to different nonlinear factors in system (17). Since the analysis of the wave speed is based on a linearisation of system (17), we believe that our results would still hold qualitatively. However, a rigorous comparison of these modelling approaches is beyond the scope of this paper.

An important question that remains unanswered is the role of motility heterogeneity within the cell cycle. Experimental studies have found that the motility of a cell can depend on its cell cycle phase (Vittadello et al., 2018). For example, during the mitotic phase, cells tend to reduce their movement (Mort et al., 2016). In order to investigate this phenomenon in the light of the invasion speed, we could modify our model to allow different diffusion coefficients, D_i for $i = 1, \dots, N$, for each stage in the system (17). Another aspect of the cell movement that can vary within the cell cycle is the directional persistence. Our models do not incorporate directional persistence of cells. However, it is possible to combine a MSM with existing models of directional persistence (Codling et al., 2008; Gavagnin and Yates, 2018). Unfortunately, the application of the front propagation method of Saarloos (2003) (see Section 3.1) to these models leads to difficulties and it may be necessary to study the problem using a different approach. We will investigate this in future research.

Acknowledgements

The authors would like to thank the Centre for Mathematical Biology and Centre for Networks and Collective behaviour of the University of Bath for constructive and helpful comments on a preprint of this paper. MJF was supported by a MRC funded scholarship. RLM was supported by core funding from the Medical Research Council. TR acknowledges the support of the Royal Society.

Appendix A. Proof of Theorem 2

Proof. Let $\mu_i = 1/\lambda_i$ for every $i = 1, \dots, N$. By writing the characteristic equation $P_\Lambda(x) = 0$ in terms of the parameters μ_i and upon rearranging, we obtain

$$\prod_{i=1}^N (\mu_i x + 1) = 2. \quad (\text{A.1})$$

We can write $\rho(\Lambda) = \rho(\mu_1, \dots, \mu_N) = \rho(\underline{\mu})$ as

$$\rho(\underline{\mu}) = \max \left\{ x \in \mathbb{R} \mid \prod_{i=1}^N (\mu_i x + 1) = 2 \right\}, \quad (\text{A.2})$$

for every $\underline{\mu} \in \{\mathbb{R}_{>0}\}^N$. It is easy to observe that $\rho(\underline{\mu})$ is a positive continuous function and we can extend the definition (A.2) to $\underline{\mu} \in \{\mathbb{R}_{\geq 0}\}^N \setminus \{(0, \dots, 0)\}$, by continuity.

Now fix $\bar{\lambda} = (\sum_{i=1}^N \mu_i)^{-1}$; without loss of generality we can take $\sum_{i=1}^N \mu_i = 1$, whence (31) becomes $N(\sqrt[N]{2} - 1) \leq \rho(\underline{\mu}) < 1$. The case of general $\bar{\lambda}$ follows by multiplying by a rescaling factor. Since ρ is a continuous function, we aim to find the stationary points of $\rho(\underline{\mu})$ in the N -dimensional simplex:

$$\mathcal{U}_N = \left\{ (\mu_1, \dots, \mu_N) \in (0, 1)^N \mid \sum_{i=1}^N \mu_i = 1 \right\}. \quad (\text{A.3})$$

We apply the Lagrange multipliers method. Hence we study the Lagrangian function given by

$$\mathcal{L}(\mu_1, \dots, \mu_N, \sigma) = \rho(\underline{\mu}) + \sigma \left(\sum_{i=1}^N \mu_i - 1 \right). \quad (\text{A.4})$$

Throughout we adopt the notation $\mathcal{L}_j = \frac{\partial \mathcal{L}}{\partial \mu_j}$ and $\rho_j = \frac{\partial \rho}{\partial \mu_j}$. By imposing $\mathcal{L}_j = 0$ we obtain

$$\rho_j = -\sigma, \quad (\text{A.5})$$

for all $j = 1, \dots, N$. We can now differentiate Eq. (A.1) respect to μ_j , which gives us

$$0 = \sum_{i=1}^N \prod_{k \neq i} (1 + \mu_k \rho) (\rho \delta_{i,j} + \mu_i \rho_j) \quad (\text{A.6})$$

where $\delta_{i,j}$ denotes the Kronecker delta. If we multiply and divide each term of the right-hand side of Eq. (A.6) by $(1 + \mu_i \rho)$, we obtain

$$\begin{aligned} 0 &= \sum_{i=1}^N \frac{\rho \delta_{i,j} + \mu_i \rho_j}{1 + \mu_i \rho} \\ &= \frac{\rho}{1 + \mu_j \rho} + \rho_j \sum_{i=1}^N \frac{\mu_i}{1 + \mu_i \rho}, \end{aligned} \quad (\text{A.7})$$

By combining Eqs. (A.5) and (A.7) we gain a condition on the coordinate μ_j of the stationary points, namely

$$\frac{\rho}{1 + \mu_j \rho} = \sigma \sum_{i=1}^N \frac{\mu_i}{1 + \mu_i \rho}. \quad (\text{A.8})$$

Notice that Eq. (A.8) holds for every $j = 1, \dots, N$ and the right hand side is independent of j , hence the only stationary point of $\rho(\underline{\mu})$ in the simplex \mathcal{U}_N is the given by the centre $\underline{\mu}_N^* = (1/N, \dots, 1/N)$.

To conclude we need study the value of $\rho(\underline{\mu})$ on the boundary of the simplex, defined as

$$\partial \mathcal{U}_N = \left\{ (\mu_1, \dots, \mu_N) \in [0, 1]^N \mid \sum_{i=1}^N \mu_i = 1 \text{ and } \min \mu_i = 0 \right\}. \quad (\text{A.9})$$

Let us consider the elements of $\partial \mathcal{U}_N$ with exactly n non-zero coordinates, with $n = 1, \dots, N-1$. Without loss of generality we can focus on the points of the form

$$(\mu_1, \dots, \mu_n, 0, \dots, 0) \in \partial \mathcal{U}_N, \quad (\text{A.10})$$

where $(\mu_1, \dots, \mu_n) \in \mathcal{U}_n$. Notice that the $\rho(\underline{\mu})$ is well defined in such points by continuity, as observed before. By repeating the Lagrange multiplier method in the sub-simplex \mathcal{U}_n , we find that the only stationary point of $\rho(\underline{\mu})$ of the form (A.10) is the one with $\mu_1 = \mu_2 = \dots = \mu_n$, i.e.:

$$\underline{\mu}_n^* = (\underbrace{1/n, \dots, 1/n}_n, 0, \dots, 0) \in \partial \mathcal{U}_N. \quad (\text{A.11})$$

This holds for every $n = 1, \dots, N-1$, so we can write all the stationary points of $\rho(\underline{\mu})$ in $\partial \mathcal{U}_N$ upon permutation of the coordinates in the form (A.11).

All the stationary points $\underline{\mu}_n^*$, for $n = 1, \dots, N$, correspond to an Erlang distribution for which we can compute the expression of ρ directly from the definition (A.2) as

$$\rho(\underline{\mu}_n^*) = n(\sqrt[n]{2} - 1), \quad (\text{A.12})$$

for $n = 1, \dots, N$. The right-hand side of Eq. (A.12) is a decreasing function of n . We deduce that the centre of the simplex, $\underline{\mu}_N^* \in \mathcal{U}_N$, corresponds to the global minimum, i.e. for all $\underline{\mu} \in \mathcal{U}_N$

$$\rho(\underline{\mu}) \geq \rho(\underline{\mu}_N^*) = N(\sqrt[N]{2} - 1). \quad (\text{A.13})$$

Finally, $\underline{\mu}_1^* \in \partial \mathcal{U}_N$ and all the points obtained by permuting its coordinates, correspond to supremum points, i.e. for all $\underline{\mu} \in \mathcal{U}_N$

$$\rho(\underline{\mu}) < \rho(\underline{\mu}_1^*) = 1. \quad (\text{A.14})$$

□

References

- Al-Omari, J., Gourley, S.A., 2002. Monotone travelling fronts in an age-structured reaction-diffusion model of a single species. *J. Math. Biol.* 45 (4), 294–312.
- Anderson, A.R.A., Chaplain, M.A.J., 1998. Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull. Math. Biol.* 60 (5), 857–899.
- Baker, R.E., Simpson, M.J., 2010. Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E* 82 (4), 041905.
- Billy, F., Clairambault, J., Fercoq, O., Gaubert, S., Lepoutre, T., Ouillon, T., Saito, S., 2014. Synchronisation and control of proliferation in cycling cell population models with age structure. *Math. Comput. Simulation* 96, 66–94.
- Cao, Y., Petzold, L., 2006. Accuracy limitations and the measurement of errors in the stochastic simulation of chemically reacting systems. *J. Comp. Phys.* 212 (1), 6–24.
- Chao, H.X., Fakhreddin, R.I., Shimerov, H.K., Kumar, R.J., Gupta, G.P., Purvis, J.E., 2018. Evidence that the cell cycle is a series of uncoupled, memoryless phases. *bioRxiv* 283614.
- Cheeseman, B.L., Newgreen, D.F., Landman, K.A., 2014. Spatial and temporal dynamics of cell generations within an invasion wave: a link to cell lineage tracing. *J. Theor. Biol.* 363, 344–356.
- Cianci, C., Smith, S., Grima, R., 2017. Capturing brownian dynamics with an on-lattice model of hard-sphere diffusion. *Phys. Rev. E* 95 (5), 052118.
- Codling, E.A., Plank, M.J., Benhamou, S., 2008. Random walk models in biology. *J. R. Soc. Interface* 5 (25), 813–834.
- Csikász-Nagy, A., Battogtokh, D., Chen, K.C., Novák, B., Tyson, J.J., 2006. Analysis of a generic model of eukaryotic cell-cycle regulation. *Biophys. J.* 90 (12), 4361–4379.
- Deng, M., Chen, W.L., Takatori, A., Peng, Z., Zhang, L., Mongan, M., Parthasarathy, R., Sartor, M., Miller, M., Yang, J., et al., 2006. A role for the mitogen-activated protein kinase kinase kinase 1 in epithelial wound healing. *Mol. Biol. Cell* 17 (8), 3446–3455.
- Deutsch, A., Dormann, S., 2007. Cellular automaton modeling of biological pattern formation: characterization, applications, and analysis. Springer Science & Business Media.
- Dyson, L., Baker, R.E., 2014. The importance of volume exclusion in modelling cellular migration. *J. Math. Biol.* 71, 691–711.
- Dyson, L., Maini, P.K., Baker, R.E., 2012. Macroscopic limits of individual-based models for motile cell populations with volume exclusion. *Phys. Rev. E* 86 (3), 031903.
- Elliott, E.C., Cornell, S.J., 2012. Dispersal polymorphism and the speed of biological invasions. *PLoS ONE* 7 (7), e40496.
- Fisher, R.A., 1937. The wave of advance of advantageous genes. *Ann. Hum. Genet.* 7 (4), 355–369.
- Gabriel, P., Garbett, S.P., Quaranta, V., Tyson, D.R., Webb, G.F., 2012. The contribution of age structure to cell population responses to targeted therapeutics. *J. Theor. Biol.* 311, 19–27.
- Gavagnin, E., Yates, C.A., 2018. Modeling persistence of motion in a crowded environment: the diffusive limit of excluding velocity-jump processes. *Phys. Rev. E* 97, 032416.
- Gérard, C., Goldbeter, A., 2009. Temporal self-organization of the cyclin/cdk network driving the mammalian cell cycle. *Proc. Natl. Acad. Sci.* 106 (51), 21643–21648.
- Gilbert, S.F., 2003. The morphogenesis of evolutionary developmental biology. *Int. J. Dev. Biol.* 47 (7–8), 467.
- Golubev, A., 2016. Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression. *J. Theor. Biol.* 393, 203–217.
- Gray-Schopfer, V., Wellbrock, C., Marais, R., 2007. Melanoma biology and new targeted therapy. *Nature* 445 (7130), 851.
- Grima, R., 2008. Multiscale modeling of biological pattern formation. *Curr. Top. Dev. Biol.* 81, 435–460.
- Haass, N.K., Beaumont, K.A., Hill, D.S., Anfosso, A., Mrass, P., Munoz, M.A., Kinjyo, I., Weninger, W., 2014. Real-time cell cycle imaging during melanoma growth, invasion, and drug response. *Pigment Cell Melanoma Res.* 27 (5), 764–776.
- Haass, N.K., Gabrielli, B., 2017. Cell cycle-tailored targeting of metastatic melanoma: challenges and opportunities. *Exp. Dermatol.* 26 (7), 649–655.
- Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. *Cell* 100 (1), 57–70.

- Keller, R., 2005. Cell migration during gastrulation. *Curr. Opin. Cell Biol.* 17 (5), 533–541.
- Maini, P.K., McElwain, D.L.S., Leavesley, D.I., 2004. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.* 10 (3–4), 475–482.
- Markham, D.C., Simpson, M.J., Maini, P.K., Gaffney, E.A., Baker, R.E., 2013. Incorporating spatial correlations into multispecies mean-field models. *Phys. Rev. E* 88 (5), 052713.
- Matsiaka, O.M., Penington, C.J., Baker, R.E., Simpson, M.J., 2017. Continuum approximations for lattice-free multi-species models of collective cell migration. *J. Theor. Biol.* 422, 1–11.
- Middleton, A.M., Fleck, C., Grima, R., 2014. A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion. *J. Theor. Biol.* 359, 220–232.
- Mort, R.L., Ford, M.J., Sakaue-Sawano, A., Lindstrom, N.O., Casadio, A., Douglas, A.T., Keighren, M.A., Hohenstein, P., Miyawaki, A., Jackson, I.J., 2014. Fucci2a: a bicistronic cell cycle reporter that allows cre mediated tissue specific expression in mice. *Cell Cycle* 13 (17), 2681–2696.
- Mort, R.L., Ross, R.J.H., Hainey, K.J., Harrison, O.J., Keighren, M.A., Landini, G., Baker, R.E., Painter, K.J., Jackson, I.J., Yates, C.A., 2016. Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model. *Nat. Commun.* 7, 10288.
- Murray, J.D., 2007. *Mathematical biology: I. an introduction*, volume 17. Springer Science & Business Media.
- Neubert, M.G., Caswell, H., 2000. Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. *Ecology* 81 (6), 1613–1628.
- Piscounov, N., Kolmogorov, A., Petrovskii, I., 1991. A Study of the Diffusion Equation with Increase in the Amount of Substance, and Its Application to a Biological Problem. In: Tikhomirov, I.V.M. (Ed.), *Selected works of AN Kolmogorov*. Springer, pp. 248–270.
- Powathil, G.G., Gordon, K.E., Hill, L.A., Chaplain, M.A.J., 2012. Modelling the effects of cell-cycle heterogeneity on the response of a solid tumour to chemotherapy: biological insights from a hybrid multiscale cellular automaton model. *J. Theor. Biol.* 308, 1–19.
- Saarloos, W.V., 2003. Front propagation into unstable states. *Phys. Rep.* 386 (2–6), 29–222.
- Sadeghi, H.M., Seitz, B., Hayashi, S., LaBree, L., McDonnell, P.J., 1998. In vitro effects of mitomycin-c on human keratocytes. *J. Refract. Surg.* 14 (5), 534–540.
- Simpson, M.J., Jin, W., Vittadello, S.T., Tambyah, T., Ryan, J., Gunasingh, G., Haass, N., McCue, S., 2018. Stochastic models of cell invasion with fluorescent cell cycle indicators. *Phys. A* 510, 375–386.
- Simpson, M.J., Landman, K.A., Hughes, B.D., 2009. Multi-species simple exclusion processes. *Phys. A* 388 (4), 399–406.
- Simpson, M.J., Merrifield, A., Landman, K.A., Hughes, B.D., 2007. Simulating invasion with cellular automata: connecting cell-scale and population-scale properties. *Phys. Rev. E* 76 (2), 021918.
- Taylor, P.R., Baker, R.E., Simpson, M.J., Yates, C.A., 2016. Coupling volume-excluding compartment-based models of diffusion at different scales: voronoi and pseudo-compartment approaches. *J. R. Soc. Interface* 13 (120). ISSN 1742–5689.
- Taylor, P.R., Yates, C.A., Simpson, M.J., Baker, R.E., 2015. Reconciling transport models across scales: the role of volume exclusion. *Phys. Rev. E* 92 (040701).
- Treloar, K.K., Simpson, M.J., Haridas, P., Manton, K.J., Leavesley, D.I., McElwain, D.S., Baker, R.E., 2013. Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies. *BMC Syst. Biol.* 7 (1), 137.
- Vittadello, S.T., Cue, S.W.M., Gunasingh, G., Haass, N.K., Simpson, M.J., 2018. Mathematical models for cell migration with real-time cell cycle dynamics. *Biophys. J.* 114 (5).
- Webb, G.F., Webb, G., 1985. *Theory of nonlinear age-dependent population dynamics*. CRC Press.
- Wise, S.M., Lowengrub, J.S., Frieboes, H.B., Cristini, V., 2008. Three-dimensional multispecies nonlinear tumor growth—i: model and numerical method. *J. Theor. Biol.* 253 (3), 524–543.
- Yates, C.A., Ford, M.J., Mort, R.L., 2017. A multi-stage representation of cell proliferation as a markov process. *Bull. Math. Biol.* 79 (12), 2905–2928.

5.2. Conclusions

The aim of this paper is to investigate the impact of accounting for realistic and general distributions of the cell-cycle time in models of cell invasion. By adopting a multiscale approach, we elucidate the connection between the distribution of cell-cycle time and the speed of invasion. One of the more significant findings to emerge from this study is that when the cell-cycle time distribution is chosen from a class of biologically realistic distributions, the asymptotic speed of the invasion lies in a tightly bounded interval. The results of this observation suggest that the characteristics of the cell-cycle time distribution might have a limited impact on the speed of the invasion.

Age-structured models, like those presented in this work, have received increasing attention in recent years. The results of this work, therefore, need to be considered in light of the current research on this topic. In the first year after its publication, this paper has been cited several times in relatively different contexts. Bobadilla et al. [2019], for example, employed an age-structured model to recapitulate the appearance of two phases in the growth rate of a proliferating cell population. The authors model the cell cycle as a resource-dependent threshold age, after which cells can proliferate at constant rate. The model of Bobadilla et al. [2019] belongs to the same class as the age-structured model studied in Section 2 of this paper. The authors cite our work as an example of invasion speed analysis, however the focus of their study is substantially different from ours, as they consider crowded scenarios in which the limited-resource condition leads to a logistic dynamic.

One aspect of this chapter which has received particular attention is related with possibility of extending the front propagation analysis to the case of phase-dependent motility. The reason why such an extension would be particularly important is because the current literature is unclear on whether cell movement is significantly affected by the cell-cycle phase or not, and recent experimental studies supporting both arguments have been reported [Mort et al., 2016, Haass et al., 2014, Vittadello et al., 2018, 2019a]. It is essential, therefore, to understand the extent to which a phase-dependent motility can affect the speed of a cellular invasion. The multi-stage model presented in this paper sets the basis of a new potential approach to answer this question. However the derivation of the formula of (25) for the invasion speed does not hold for phase-dependent diffusivity and other methods have to be explored. More specifically, by replacing the diffusive coefficient D in system (19) with D_i , for $i = 1, \dots, N$, and

following the steps of the paper, one can express $\omega(k)$ as the eigenvalue of the matrix

$$\Gamma = \begin{bmatrix} k^2 D_1 - \lambda_1 & 0 & \dots & 0 & 2\lambda_N \\ \lambda_1 & k^2 D_2 - \lambda_2 & 0 & \dots & 0 \\ 0 & \lambda_2 & k^2 D_3 - \lambda_3 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & & \lambda_{N-1} & k^2 D_N - \lambda_N \end{bmatrix}. \quad (5.1)$$

Notice that when $D_i = D$, for every $i = 1 \dots, N$, one recovers immediately equation (22) of the paper since we can decompose $\Gamma = k^2 D \mathbf{1} + \Lambda$, where $\mathbf{1}$ represents the identity matrix. Unfortunately, for a general choice of D_i the eigenstructure of Γ becomes very complicated. Vittadello et al. [2018] have studied in details the case of $N = 2$ and $N = 3$, for which they obtained analytical formulas of the invasion speed, but similar formulas for larger values of N are currently unknown.

A possible way to gain some insights on the case of phase-dependent motility is by focusing on particular combinations of the parameters D_i . For example, one can turn off the diffusivity of one, or multiple, stages by setting $D_j = 0$ for some $j \in \{0, \dots, N\}$, while keeping uniform diffusivity in the remaining stages. The combination of *mobile* and *immobile* stages is likely to simplify the eigenstructure of the matrix Γ and it might help making a first step toward the understanding of the more general case. In addition, the idea of having different phases within the cell-cycle for motility and proliferation events, is consistent with the highly-debated “go-or-grow” hypothesis [Hatzikirou et al., 2012, Vittadello et al., 2019a].

Finally, the topic tightly connected with this chapter, despite not being explicitly mentioned in paper, is the dynamical evolution of the cell-cycle phase distribution within a growing cell population. Whilst this aspect of the model has not been investigated in this paper, there is an evident connection between the cell-cycle time and the phase distribution observed in the population. Understanding this connection is crucial for a wide range of applications and it has been the object a series of recent studies [Pirjol et al., 2017, Jafarpour et al., 2018, Jafarpour, 2019]. Of particular relevance is the work of Jafarpour [2019] in which the author studies bacterial populations with Gaussian distributed cell-cycles and a cell-size regulation mechanisms. In the following chapter we continue to investigate the effects of including multi-stage descriptions of the cell cycle in models of growing cell populations. In particular, we present a case of study in which accounting for realistic cell-cycle distributions has important implications for the model and for the interpretation of the results.

Chapter 6

Synchronised oscillations in growing cell populations are explained by demographic noise

This chapter contains a draft manuscript to be submitted for publication in Nature Communications. The formulation of the project and part of the research of this chapter were carried out while visiting the group of Professor Matthew J. Simpson at the Queensland University of Technology in Brisbane, Australia. In this paper, we apply the multi-stage modelling approach described in Chapter 5 to study the relation between demographic noise and emerging oscillations in subpopulations of cells.

More precisely, we study a proliferation assay of melanoma cells to investigate the origin of inherent oscillations in the subpopulation of cells in the first phase of the cell cycle (G1). By deriving a series of deterministic and stochastic models based on a multi-stage representation of the cell-cycle time distribution, we explore the hypothesis that the observed fluctuations might be a transient phenomenon originated, and amplified, by the finite-size effect intrinsic in the population. We elucidate the transient and asymptotic phases of our models and we derive an analytical formula to quantify the effect of stochasticity in the appearance of the oscillations. The results suggest that the intrinsic demographic noise of the population is capable of explaining the oscillations in the proportion of G1 cells.

6.1. Outline of the article

As the manuscript reported is a preprint, it is formatted to match the style of the rest of the thesis. Section 6.2 contains the introduction of the manuscript with a short overview of the literature to contextualise our study. In Section 6.3 we present the results of our work and the comparison with the experimental data. In Section 6.4 we comment on our findings and we discuss their implications from a broad perspective. Section 6.5 represents the methods part of the manuscript where we report the details of the mathematical models used and all the analytical derivations. Section 6.6 provides a short conclusion to the chapter.

Appendix B: Statement of Authorship

This declaration concerns the article entitled:									
Synchronised oscillations in growing cell populations are explained by demographic noise									
Publication status (tick one)									
draft manuscript	<input checked="" type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input type="checkbox"/>
Publication details (reference)	In preparation for Nature Communications Authors: Enrico Gavagnin, Sean T. Vittadello, Gency Guanasingh, Nikolas K. Haass, Matthew J. Simpson, Tim Rogers and Christian A. Yates								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The bulk of the calculations have been performed by the author of the thesis (80%). All the numerical computations and simulations have been performed by the author of the thesis (100%). The data collection and analysis have been performed by the author of the thesis (100%). E. Gavagnin, T. Rogers and C.A. Yates contributed equally to the presentation of the content (33%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed							Date	27.4.2020	

Synchronised oscillations in growing cell populations are explained by demographic noise

Enrico Gavagnin^{*1}, Sean T. Vittadello², Gency Guanasingh³, Nikolas K. Haass³,
Matthew J. Simpson², Tim Rogers¹, and Christian A. Yates¹

¹*Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, UK*

²*School of Mathematical Sciences, Queensland University of Technology, Brisbane QLD 4001, Australia*

³*The University of Queensland, The University of Queensland Diamantina Institute, Translational Research Institute, Woolloongabba, Brisbane QLD 4102, Australia*

Abstract

We investigate the origin of recently discovered sustained oscillations in growing cell populations. Differential fluorescent staining during distinct phases of the cell cycle allows for the experimental observation of previously hidden inherent synchronisation in growing cell populations. We explore the hypothesis that the oscillations we observe are a long-lasting transient phenomenon originated, and amplified, by finite-size effects of demographic noise. We develop a mathematical theory of multi-stage cell growth and division which accurately reproduces the synchronised oscillations observed in proliferation assays of melanoma cells. By elucidating the transient and asymptotic phases of the dynamics, we derive an analytical formula to quantify the effect of demographic noise in the appearance of the oscillations. Our results give a simple and compelling explanation for the emergence of inherent synchronisation in growing cell populations, and our methods provide a framework for analysing the broader implications of this phenomenon.

Keywords: cellular population dynamics, cell-cycle time distribution, demographic noise, finite-size effects, multi-stage model

6.2. Introduction

Cell proliferation is crucial in a vast range of biological processes, from morphogenesis to tumour growth [Gilbert, 2000, Evan and Vousden, 2001]. Understanding and predicting the time evolution of a growing cell population is, therefore, of fundamental medical interest [Mort et al., 2016, Haass and Gabrielli, 2017].

A standard modelling approach consists of assuming that cell divisions are independent events with exponentially distributed waiting times. This gives rise to exponential growth of an unstructured population of cells [Murray, 2007]. This approach has been supported by classic experimental studies for large populations under favourable growth conditions [Monod, 1949, Laird, 1965]. However, when smaller populations are considered - for example a single progenitor cell - a simple model of exponential growth is incapable of explaining the experimental results and more sophisticated models are necessary [Baker and Simpson, 2010, Yates et al., 2017, Jafarpour, 2019, Pirjol et al., 2017, Lang et al., 2009, Kuritz et al., 2018]. Moreover, due to the technological advances of the past decade, we are now able to access accurate data revealing the structure of a dynamic cell population which is left completely unexplained by simple exponential models [Chao et al., 2018, Vittadello et al., 2018, Simpson et al., 2018].

The typical experimental protocol to study a growing cell population is a proliferation assay. This is an *in vitro* experimental procedure which monitors the number of cells over time, the number of cellular divisions or DNA synthesis [Riss et al., 2016]. In a recent work [Vittadello et al., 2019b], we considered a proliferation assay of melanoma cells labelled with FUCCI (Fluorescent Ubiquitous Cell Cycle Indicator [Sakaue-Sawano et al., 2008] - see Figure 6-1) which allowed us to track the number of cells in particular phases of the cell cycle over a timespan of 48h. We found that the number of cells in the first phase of the cell-cycle, gap 1 (G1), shows evident and unexpected fluctuations during the entire duration of the experiment. Understanding this phenomenon is crucial, as it might have a significant impact on the efficiency of cell-cycle-inhibiting drugs [Beaumont et al., 2016, Haass and Gabrielli, 2017] and on the reproducibility of these experiments. However, the biological mechanism causing these fluctuations and their long term behaviour, are yet to be fully understood.

In this study, we employed a multi-stage (MS) mathematical model for cell proliferation by representing the cell-cycle time (CCT) as a series of discrete stages. The waiting time distribution between consecutive stages is exponential, meaning that the CCT follows a more general class of distributions, known as hypo-exponential. This family of distributions has been shown to provide good agreement with the experimental cell-cycle time distribution data [Yates et al., 2017, Gavagnin et al., 2019, Simpson et al., 2018, Vittadello et al., 2018]. By deriving a deterministic representation of the population dynamics under the MS approach we reproduced the cell-cycle fluctuations observed in the experiments which suggests that MS models are a suitable framework for investigating the phenomenon of cell-cycle synchronisation. However, since the

parametrisation of the model was carried out individually for each experimental trajectory, our previous study did not explain the origin of such oscillatory phenomena nor their asymptotic behaviour.

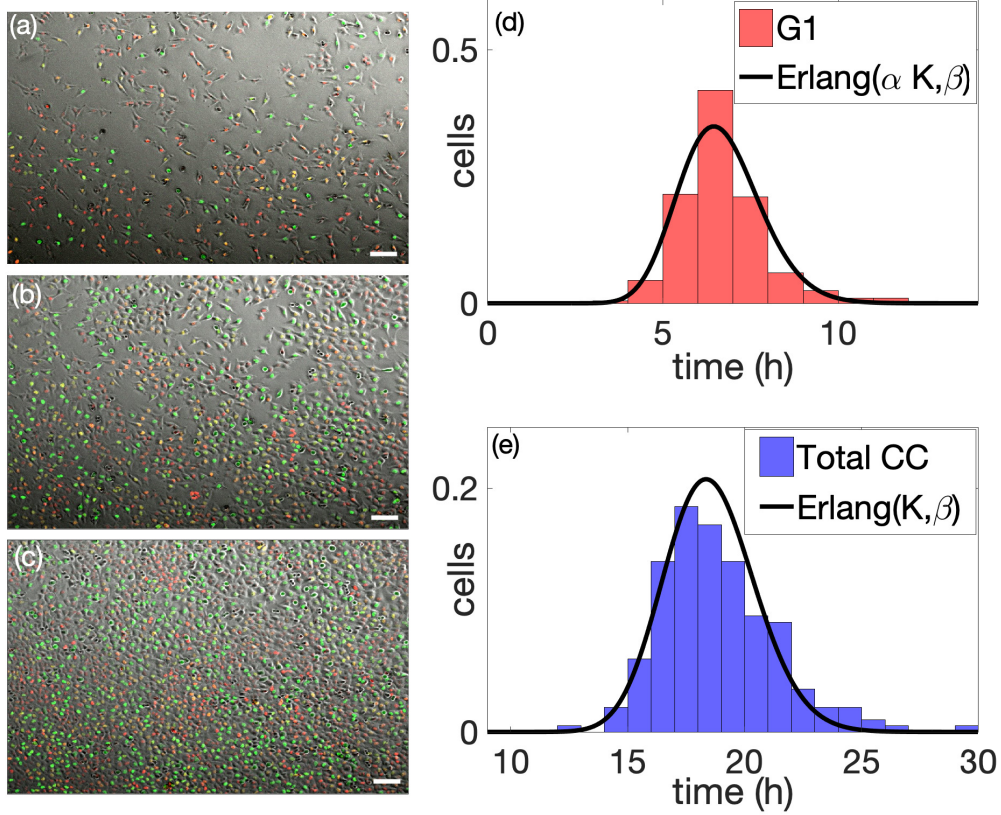


Figure 6-1: Proliferation assay with melanoma C8161 cells labelled with Fucci. Panels (a) to (c) show three snapshots of a microscopy image time series. Panel (a) is taken at the beginning of the recording, $t = 24h$, panel (b) is taken half-way through the recording, $t = 48h$, and at panel (c) at the end of the recording, $t = 72h$ (scale bar $100\mu m$). The cells with red nuclei are in the G1 (gap 1) phase of the cell cycle, the ones with yellow nuclei are in the eS (early-synthesis) phase and those with green nuclei are in one of the remaining consecutive phases S (synthesis), G2 (gap 2) or M (mitosis). Panel (d) shows the comparison between the distribution of the duration of the G1 phase obtained by tracking 200 randomly chosen cells (red histogram) and the Erlang distribution with the same mean and variance (black curve). Panel (e) shows the comparison between the distribution of the full cell cycle time of the same tracked cells (blue histogram) together with the corresponding Erlang distribution. See Section 6.5.4 for full details.

The fluctuations which appear in the MS model are a typical example of damping oscillations about an equilibrium [Harrison and Topiwala, 1974]. In other words, if the initial phase-distribution - the distribution of cells in each phase of the cell cycle - is sufficiently far from its invariant distribution, the model enters a *transient* phase characterised by oscillations of decaying amplitude, followed by an *asymptotic* phase in which the invariant distribution is reached and the total population grows exponentially. The presence of these two regimes, the transient-oscillatory regime and asymptotic-exponential regime, is a common feature of many structured growing population [Jafarpour, 2019, Jafarpour et al., 2018, Pirjol et al., 2017, Baker and Röst, 2019]. Is not surprising, therefore, that these two phases play distinct but critically

important roles in the dynamics of a growing cell populations.

It remains unclear, however, what causes the initial departure from the invariant distribution in the experiments which might have triggered the oscillatory behaviour. All the experiments were prepared following a standard procedure [Beaumont et al., 2016] which is considered to produce a cell population in its invariant state. However, since only a small proportion of the total population is recorded from the microscopy, there is a natural form of finite-size departure from the theoretical invariant distribution due to the relatively small size of the initial population.

In this paper, we explore the hypothesis that the oscillations reported by Vittadello et al. [2019b] can be quantitatively explained by the effects of the finite-size stochasticities alone. We first analyse the MS model with a particular focus on characterising the transient and asymptotic phases. By deriving a stochastic mesoscopic model, we study the effect of stochasticity in the system and obtain an analytical formula that can be used to quantify the amplitude of the fluctuations due to finite-size effects. Finally, we parametrise the MS model by fitting the G1 and total cell-cycle time distributions, obtained from single-cell tracking data, and compare our predictions with the time series obtained by the experiments.

Our central finding is that the fluctuations in the subpopulation of G1 cells in the proliferation assay are of the same magnitude of those induced by demographic noise alone, which suggests finite-size effects as the main origin of the synchronisation. In particular, our study predicts that the observed oscillations are a transient phenomenon for which we can predict the corresponding characteristic decay time.

6.3. Results

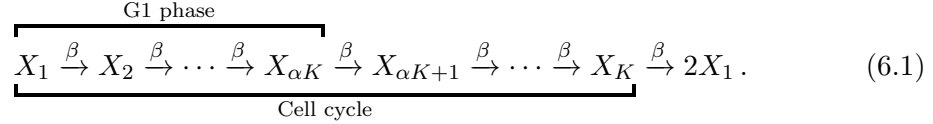
6.3.1 Multistage model recapitulates experimental observations

We adopt an agent-based model (ABM) for the growth and division of cells, following [Yates et al., 2017, Gavagnin et al., 2019, Vittadello et al., 2018]. In this formulation the cell-cycle is represented as a series of K stages through which a cell progresses before it divides. We choose the waiting time to progress from one stage to the next to be an exponential random variable with rate β , independent from all other events. When a cell passes through the final stage, it divides into two new daughter cells, both initialised at stage one. This is a highly simplified model of the cell cycle, however, it is sufficient for the purposes of this study since (as we will show later) it gives a good fit to experimentally observed distributions of cell cycle time, and is representative of a broad class of biologically realistic unimodal distributions.

The K stages of our model are grouped into sections corresponding to the known phases of the cell cycle. In particular, we say that a cell is in the G1 phase if it is in one of the first αK stages, where α is a constant to be determined by comparison with data. Expressed as a sum of exponential random variables, the duration of both the G1 and the entire cell cycle are Erlang distributed with parameters (K, β) and

$(\alpha K, \beta)$, respectively. Figure 6-1 (d) and (e) show the maximum likelihood fit of the model simultaneously to both the duration of the G1 phase and the total cell-cycle time for the melanoma cell line C8161. In this example we find parameters $K = 92, \alpha K = 33, \beta = 4.96h^{-1}$. The measured cell cycle time has an average of 18.5 hours with standard deviation around 2 hours (see Section 6.5.4).

To keep track of the growth of a population of cells, we define the state vector $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_K(t))$, where $X_k(t)$ denotes the number of cells in stage k at time t . Our model is then represented as a series of chemical reactions, namely



We write $N = \sum_{k=1}^K X_k$ for the total number of cells, $G = \sum_{k=1}^{\alpha K} X_k$ for the number of cells in G1 phase. As the population grows, the proportion of cells in each stage will eventually converge to a fixed value $X_i/N \xrightarrow{t \rightarrow \infty} u_i$, for $i = 1, \dots, N$, known as the *invariant stage distribution*. In Section 6.3.2 we prove this fact and derive an exact expression for the limit \mathbf{u} . On shorter time horizons, the behaviour of the system is less easy to predict.

To assess the strength of oscillations, in what follows we develop a mathematical theory for the behaviour of the proportion $Q = G/N$ of G1-phase cells. The first part of our analysis reveals long-lived damped oscillations in the expected value of Q in a growing population, while the second shows how this effect is initiated and sustained by demographic noise.

Our experimental data are 30 image time series taken from proliferation assays of C8161 melanoma cells expressing Fucci, as previously reported in Vittadello et al. [2019b] - see Figure (a), (b) and (c) 6-1 for three snapshots of the microscopy images. Each time series refers to a 48h time window following an incubation of 24h. In Figure 6-2 we present a comparison between an experimental time series (blue line) and the envelope of two standard deviations, Ω (light grey region) obtained from the MS model. Although the trajectory shows clear oscillations about the mean (about three complete cycles from time 24h to time 72h), 97% of the data points lie inside Ω . We repeat the comparison for all 30 time series from the experiments (reported in Figure 6-8(a)). The plot shows that the envelope Ω provides a good approximation for the amplitude of the fluctuations for most of the experimental trajectories.

6.3.2 Understanding the transient and asymptotic dynamics

In order to understand the interplay between the transient oscillatory dynamics and asymptotic exponential growth, we begin by writing down the equations governing the dynamics of the expected number of cells in each stage $\bar{\mathbf{x}} = \mathbb{E}[\mathbf{X}]$. Here “expected” should be interpreted to mean the average over many experiments with precisely the same initial condition — we will later see that the variability of the initial condition is

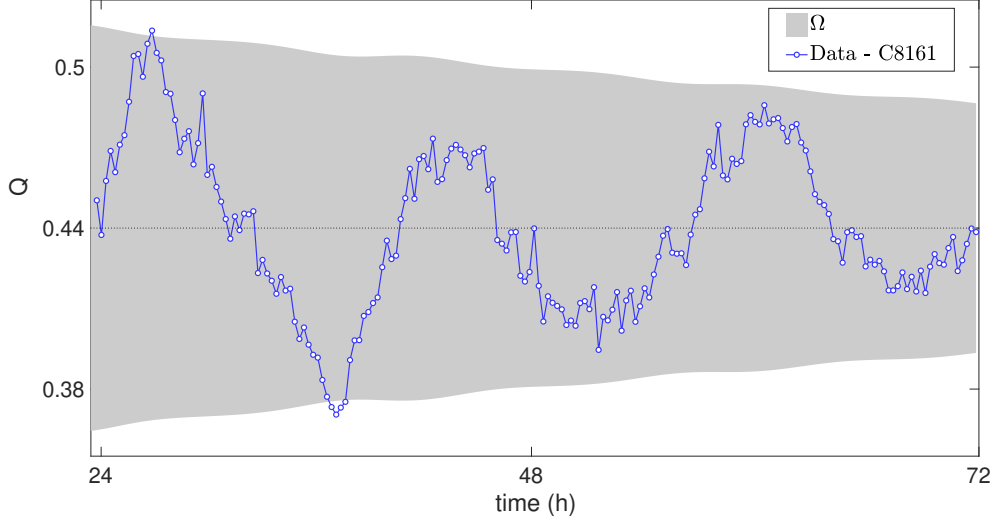


Figure 6-2: Comparison between experimental data and model prediction for the time evolution of $Q(t)$. One time series trajectory obtained from the experiments is plotted (blue line), together with the envelope of two standard deviation (light grey region) predicted using the MS model. The parameters of the MS models are obtained by fitting the distribution of the total cell-cycle time and G1 duration (see Section 6.5.4): $K = 92$, $\alpha K = 33$, $\beta = 4.96h^{-1}$ and $N_0 = 155$.

a strong contributor to the emergence of oscillations. From the model formulation we directly obtain

$$\frac{d\bar{\mathbf{x}}}{dt} = \beta \mathcal{S} \bar{\mathbf{x}}, \quad (6.2)$$

where β is the rate of progression through the model stages, and \mathcal{S} is the corresponding stoichiometry matrix. This matrix has non-zero entries $\mathcal{S}_{k,k} = -1$, $\mathcal{S}_{k,k+1} = 1$ for $k = 1, \dots, K-1$ describing progression between stages, and $\mathcal{S}_{K,K} = -1$, $\mathcal{S}_{1,K} = 2$ describing cell division.

For the purpose of the analysis, we assume $\beta = K$ throughout, so that the average cell-cycle time is normalised to 1. It is an easy exercise to see that the characteristic polynomial of the matrix \mathcal{S} is given by $\mathcal{P}(y) = (y+1)^K - 2$, from which the eigenvalues of \mathcal{S} are computed as $\lambda_k = \xi_k 2^{\frac{1}{K}} - 1$ for $k = 1, \dots, K$, where $\xi_k = e^{2\pi i k / K}$ is a K -th root of unity. By solving a series of recursive equations, one can write down the left- and right-eigenvectors associated with the k -th eigenvalue of \mathcal{S} , which we denote \mathbf{u}^k and \mathbf{v}^k , respectively. Specifically, we have

$$u_j^k = \frac{2\lambda_K}{(1+\lambda_k)^j}, \quad v_j^k = \frac{1}{K} \frac{(1+\lambda_k)^j}{2\lambda_K}. \quad (6.3)$$

We drop the index k whenever we refer to the eigenvalue with maximum real part and the corresponding eigenvectors, i.e. $\lambda = \lambda_K = 2^{\frac{1}{K}} - 1$, $\mathbf{u} = \mathbf{u}^K$ and $\mathbf{v} = \mathbf{v}^K$.

Notice that from system (6.2) we can write $\bar{\mathbf{x}}(t) = e^{Kt\mathcal{S}}\bar{\mathbf{x}}_0$, where $\bar{\mathbf{x}}_0$ denotes the initial number of cells per stage. In order to study the matrix exponential $e^{Kt\mathcal{S}}$, we first notice that we can write down the (i, j) element in terms of the eigenvalues and

eigenvectors of \mathcal{S} as

$$[e^{Kt\mathcal{S}}]_{i,j} = \sum_{k=1}^K u_i^k v_j^k e^{K\lambda_k t}. \quad (6.4)$$

Notice that as $t \rightarrow \infty$ the leading term of the righthand side in equations (6.4) is $u_i v_j e^{K\lambda t}$ and, hence $\bar{x}(t) \sim \mathbf{u} e^{K\lambda t}$ determines the long-time behaviour of system (6.2). We can use this fact to study the limiting behaviour of Q : we write $\bar{Q}(t) = \sum_{i=1}^{\alpha K} [e^{K\mathcal{S}t} \bar{x}_0]_i / \sum_{i=1}^K [e^{K\mathcal{S}t} \bar{x}_0]_i$ and by looking at the first two leading order terms of (6.4), we obtain $\lim_{t \rightarrow \infty} Q(t) = \sum_{i=1}^{\alpha K} \mathbf{u}_i = 2(1 - 2^{-\alpha}) = Q_*$. Notice that convergence to Q_* occurs with an exponential decay rate given by the spectral gap of the stoichiometry matrix, $\Re[\lambda_{K-1}] - \lambda_K$ (see Figure 6-3).

We now focus on the transient behaviour of the system (6.2). We substitute the expressions (6.3) into the formula (6.4) and, by exploiting a remarkable identity of the Mittag-Leffler function [Paris, 2002], we are able to transform the finite sum over eigenvalues on the right-hand side of (6.4) into an infinite sum over the cycles of the oscillatory solutions. Precisely, we write

$$[e^{Kt\mathcal{S}}]_{i,j} = \frac{1}{K} \sum_{k=1}^K (1 + \lambda_k)^{j-i} e^{K\lambda_k t}. \quad (6.5)$$

$$= \sum_{n=0}^{+\infty} \frac{1}{2\pi i} \oint 2^n (1+z)^{-1-Kn-(i-j)} e^{Kzt} dz \quad (6.6)$$

$$= \sum_{n=0}^{+\infty} \varphi_n(t, i, j), \quad (6.7)$$

where $\varphi_n(t, i, j) = \frac{(Kt)^{Kn+i-j}}{(Kn+i-j)!} 2^n e^{-Kt}$.

We can now use the expression (6.7) to approximate $e^{Kt\mathcal{S}}$ for short times, by truncating the sum over n to a finite index, \bar{n} . For example, let us consider an initial population of N_0 cells perfectly synchronised at the beginning of the cell cycle, *i.e.* $\bar{x}_0 = N_0 \mathbf{e}_1$. Then we define $G_{\bar{n}} = N_0 \sum_{k=1}^{\alpha K} \sum_{n=0}^{\bar{n}} \varphi(n, i, 1)$ and $N_{\bar{n}} = N_0 \sum_{k=1}^K \sum_{n=0}^{\bar{n}} \varphi(n, i, 1)$. In Figure 6-3 we plot $Q_{\bar{n}} = G_{\bar{n}}/N_{\bar{n}}$, for $\bar{n} = 0, 1$ and 3, together with Q obtained by solving system (6.2) numerically. The plot illustrates how each term of the sum (6.7) contributes one additional oscillation to the transient dynamics of the proportion of G1-phase cells. We now have a complete picture of how oscillations propagate on average in the growing population. It remains for us to show how they are created and sustained.

6.3.3 Finite-size effects trigger and amplify oscillations

There are two sources of randomness that are relevant to our model cell population growth: the choice of the initial state, and the timing of cell division events. In order to take into account the stochasticity in the initial population of cells we mimic the sampling procedure of the experiments. We assume that a sample of average size N_0 cells is drawn from the flask at time $t = 0$. Since the cell culture in the flask has

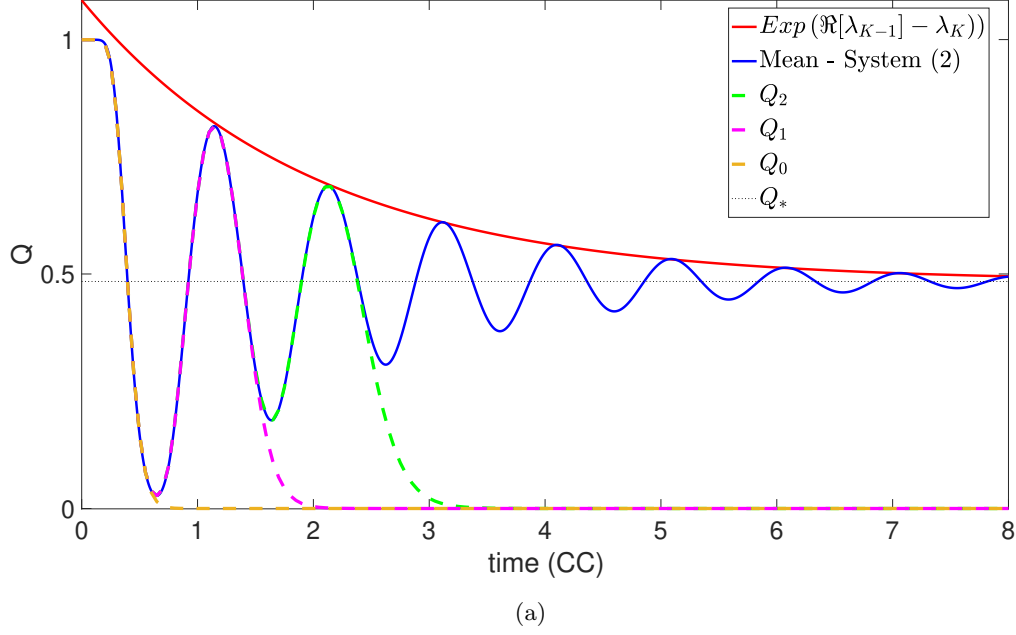


Figure 6-3: The transient oscillatory dynamics. The figure shows the plot of the ratio $Q(t)$ obtained numerically solving the deterministic system (6.2) (blue solid line) initialised with $\mathbf{X}(0) = N_0 \mathbf{e}_1$ and parameters $K = 40$, $N_0 = 100$ and $\alpha = 0.4$. The dashed lines represent the short-time approximation obtained by truncating expression (6.7) up to $\bar{n} = 0$ (yellow), $\bar{n} = 1$ (pink) and $\bar{n} = 2$ (green). The red solid line shows the exponential decay of the oscillations.

been maintained at a subconfluent cell density in fresh growth medium to prevent any synchronisation due to G1-arrest [Beaumont et al., 2016], we consider the population of the flask to have reached the invariant stage distribution \mathbf{u} . The random sampling of the number of cells in each stage is modelled by using K independent Poisson random variables, $X_i(0) \sim \text{Po}(u_i N_0)$, for $i = 1, \dots, K$, describing the number of cells sampled for each stage.

Next, we aim to quantify the effects of inherent stochasticity in the agent-based model. Performing a finite-size expansion of the master equation associated with the model [Morris and Rogers, 2014, Gardiner, 2009], we derive a system of stochastic differential equations for the density of cells relative to the initial population size. Let $\mathbf{x} = \mathbf{X}/N_0$, then for large but finite N_0 we obtain the Langevin equation

$$\frac{d\mathbf{x}}{dt} = K\mathcal{S}\mathbf{x} + \sqrt{\frac{K}{N_0}}\mathcal{S}\boldsymbol{\eta}(t), \quad (6.8)$$

where $\boldsymbol{\eta}(t)$ is a K -dimensional white noise vector with correlator $\mathbb{E}[\eta_i(t)\eta_j(t')] = x_i\delta_{ij}\delta(t-t')$. The first term on the right describes the average behaviour of the model, and is the same as in equation (6.2). The second term captures the stochastic contributions arising from the finiteness of the population.

To get more insight into the behaviour of this model, we first write down an Ornstein-Uhlenbeck (OU) model which approximates the behaviour of the Langevin equation (see Section 6.5.1). By studying the OU process, and in particular its corre-

lation matrix (see Section 6.5.2), we calculate the envelope of two standard deviations of Q , defined as

$$\Omega(t) = [Q_* - 2\sigma_Q(t), Q_* + 2\sigma_Q(t)], \quad (6.9)$$

where $\sigma_Q(t)$ denotes the standard deviation of $Q(t)$. All the details of the derivation are discussed in Section 6.5.

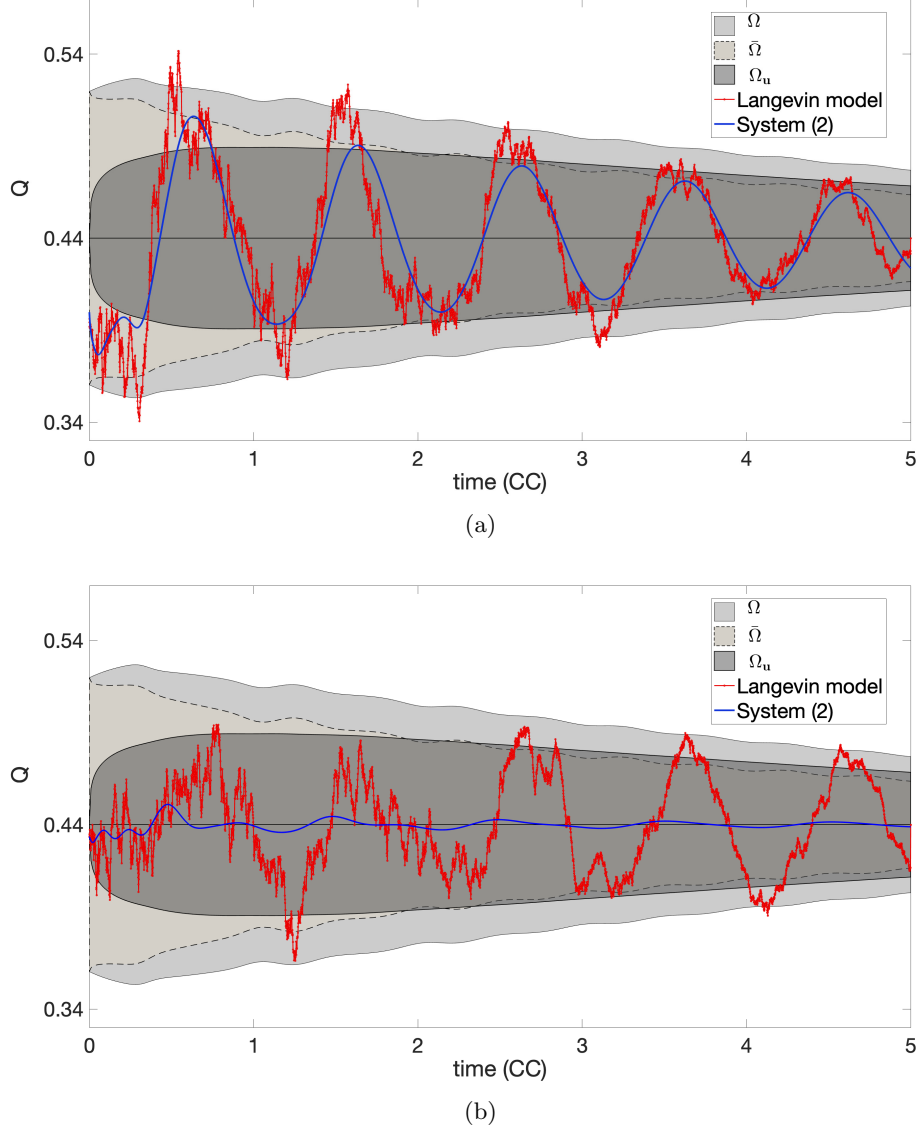


Figure 6-4: *Finite-size effects amplify the the oscillations of Q . The two panels show the overlay of the three envelopes Ω (light grey region with solid line), $\bar{\Omega}$ (medium grey region with dashed line) and $\Omega_{\mathbf{u}}$ (dark grey region with solid line) together with two trajectories of $Q(t)$ obtained by solving numerically (by using the Euler-Maruyama method with time step $\Delta t = 10^{-3}$) the Langevin model (red line) and the deterministic system (6.2) (blue line) with the same, random initial condition. The two panels show two independent realisations of the stochastic models.*

We denote with Ω , and $\bar{\Omega}$, the envelopes of the Langevin model and the system (6.2), respectively, with initial random sampling. We denote with $\Omega_{\mathbf{u}}$ the envelope of the Langevin model with deterministic initial condition $x_0 = \mathbf{u}$. In the two panels of Figure 6-4 we overlay Ω , $\bar{\Omega}$ and $\Omega_{\mathbf{u}}$, together with the numerical trajectories of Q , one

(red) obtained by solving the Langevin equation (6.8) and one (blue) by solving system (6.2), both initialised with the same random initial condition.

The results of Figure 6-4 show that in all three cases considered, accounting for the finite-size stochasticity can lead to a persistent departure of Q from the equilibrium value. In the two cases which account for the initial random sampling, Ω and $\bar{\Omega}$, the envelopes present an evident initial departure from the equilibrium which is sustained for several cell-cycle times and halving after approximately four periods. The inherent dynamical stochasticity of the Langevin model tends to amplify the departure from the equilibrium as evident in Ω . Interestingly, both these envelopes have slightly fluctuating edges. On the contrary, the envelope initialised at the invariant distribution, $\Omega_{\mathbf{u}}$ shows an initial, fast expansion of the envelope, followed by a phase of slower decay. Notice that $\Omega_{\mathbf{u}}$ lies well inside Ω for all the time interval considered which suggests that the effects of the initial random sampling play a role for the entire duration of the simulation. The numerical trajectories overlaid show good agreement with these findings. In particular, the solution of system (6.2) (blue line) lies well inside $\bar{\Omega}$ while the simulation of the Langevin equation (red line) shows a larger departure and it remains almost entirely inside the envelope Ω .

Notice that both trajectories considered in Figure 6-4(a) show clear oscillations about the origin with similar phase. The Langevin solution leads to an amplification of the amplitude of the oscillation appearing in system (6.2). Although this phenomenon is common, due to the stochasticity of the Langevin model, in some cases the oscillations appear only in the Langevin model and not in the deterministic model as shown in Figure 6-4(b).

In order to quantify the appearance of the oscillations, we look at the time autocorrelation function of $G(t)$, that we define as

$$A(t, t') = \rho [G(t), G(t')] , \quad (6.10)$$

where ρ denotes the correlation coefficient, defined in equation (6.20) and can be computed using the formula for the correlation matrix (see Section 6.5). Figure 6-5 shows the evolution of the autocorrelation function, $A(t, t')$ as function of t' , for $t = 0, 2$ and 4 , respectively. In each panel we plot $A(t, t')$ calculated analytically, using the correlation matrix, (black solid line) and the simulated value obtained by averaging 50 independent trajectories of the Langevin model (red dashed line). All three panels show a good agreement between the analytical formula and the simulated counterpart. Moreover, the results confirm the presence of strong fluctuations on the time autocorrelation of G with a period of exactly one cell cycle. As Q converges to the equilibrium, the amplitude of the oscillations decreases and the autocorrelation function $A(t, t')$ tends to unity.

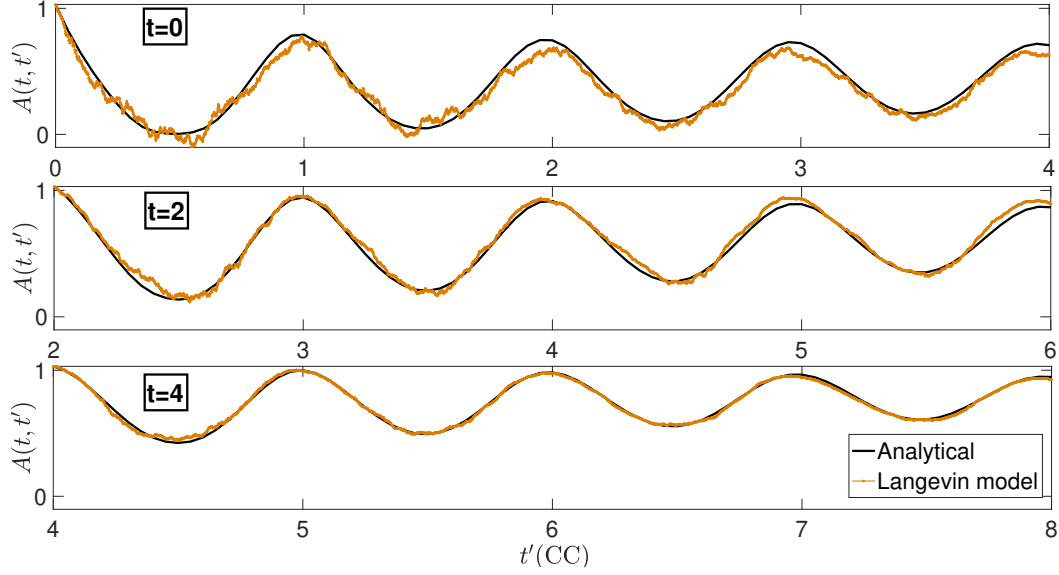


Figure 6-5: *The time-autocorrelation function. The three panels show the time-autocorrelation function, $A(t, t')$ at time $t = 0, 2$ and 4 , as indicated in the figure, obtained analytically (black line) and by averaging over 50 independent simulations of the Langevin model (red dashed line). The parameters of the model are the same as in Figure 6-2. Time is normalised with respect to the average cell-cycle time.*

6.4. Discussion

In this paper we studied the impact of demographic noise in a proliferation assay of melanoma cells and the extent to which finite-size effects can explain the synchronisation in the cell-cycle phase that we previously observed [Vittadello et al., 2019b]. We adopted a MS approach to model the cell-cycle time distribution and the G1 duration and we derived both a deterministic and a stochastic representation for the time evolution of Q . We found that the stochasticity in the initial sampling of cells leads to a departure from the invariant distribution which triggers a transient oscillatory phase. The presence of intrinsic stochasticity in the dynamics tends to amplify these oscillations and delay their exponential decay.

We characterised the transient and asymptotic phases of the MS model by deriving an analytical formula for the envelope of two standard deviations for the amplitude of the oscillations. Finally, by comparing our results with the experimental data from a proliferation assay of C8161 melanoma cells, we found that the amplitude of the experimentally observed fluctuation lies inside the envelope predicted by the model. Our findings suggest that finite-size stochasticity has a crucial role in the population dynamics of the proliferation assay and it can provide an explanation for observed synchronisation in the subpopulation of the cell cycle phases.

From an experimental point of view, our results bring attention to the importance of the sample size when performing experiments which involve small populations. In particular, any data interpretation should be carried with the role played by finite-size stochasticity in mind. In this paper we develop a general protocol that, in principle,

can be applied to other experiments in order to determine the extent of effects due to finiteness of the cell population.

From a theoretical point of view, our study provides a further understanding of the relation between cell-cycle distribution and global population dynamics. Whilst our analysis employed a MS mode, the results of our analysis are amenable to extension to more general type of cell-cycle distributions. In fact, for certain choices of the model parameters, the Erlang distribution adopted in this paper is an excellent approximation of a Gaussian distribution. In Section 6.5.5 we compute the relative entropy (Kullback-Leibler divergence) between an Erlang and Gaussian distribution to show that for increasing values of K , the two distributions converge. In principle, one could use this fact to study the applicability of our findings to Gaussian cell-cycle time and, hence, compare our results with other similar studies which rely on a Gaussian approach [Jafarpour, 2019, Pirjol et al., 2017].

The mathematical framework that we designed in this paper represents a first step towards understanding the fluctuating phenomenon observed in proliferating cell populations. In order to simplify the analysis, however, some biological details were omitted in the model and they will be the object of future studies. For instance, in all the models considered in this paper we do not take into account the spatial extent of the system and individual cells were not associated with a physical location. Accounting for cell motility and spatial correlations of mother-daughter cells is likely to play an important role in the context of synchronising subpopulations [Baker and Simpson, 2010]. In particular, we should expect cell motility to break the spatial correlation of cells leading to faster decay of the oscillations.

Another avenue of improvement of the models presented here is accounting for cell-cell interactions, such as cell adhesion or contact inhibition. Obviously, this would lead to a wide range of questions about the relation between these mechanisms and the appearance, or sustenance, of the oscillations. Despite their great biological interest, these questions are beyond the scope of this paper and we will investigate them in the future.

6.5. Methods

6.5.1 The OU approximation

We can simplify the Langevin model given by equation (6.8) by replacing the dependence on x in the correlator of $\boldsymbol{\eta}(t)$ with $\mathbb{E}[x] = K\mathbf{u}e^{K\lambda t}$. The resulting equation consists of the high-dimensional non-autonomous Ornstein-Uhlenbeck (OU) process

$$\frac{d\hat{x}}{dt} = K\mathcal{S}\hat{x} + K\sqrt{\frac{e^{K\lambda t}}{N_0}}\mathcal{S}\boldsymbol{\psi}(t), \quad (6.11)$$

where $\psi(t)$ is a K -dimensional white noise vector with correlator $\mathbb{E}[\eta_i(t)\eta_j(t')] = u_i\delta_{ij}\delta(t - t')$. We test the behaviour of the two models, the OU process given by equation (6.11), and the Langevin equation (6.8) in Figure 6-6. The results suggest that the OU process is an accurate approximation of the Langevin equation, in particular the presence of the oscillations is evident in both the modelling regimes (Figure 6-6(a)). In Figure 6-6(b), we compare the distributions of $Q(t)$ at times $t = 1, 3$ and 5 obtained by averaging over 1000 independent simulations, which show good agreement between the two models.

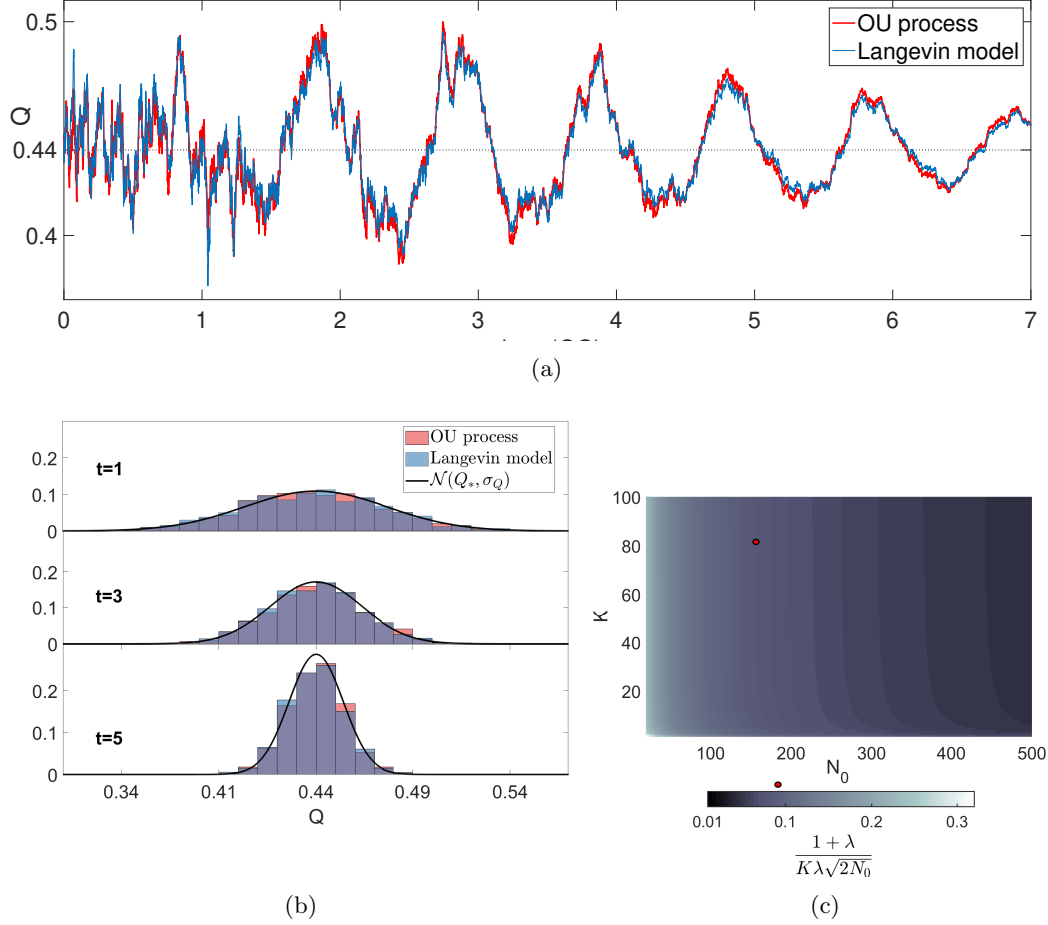


Figure 6-6: Comparison of OU process and the Langevin equation model. Panel (a) shows two evolutions of Q for the OU process (red) and the Langevin equation (blue). The two trajectories are realised using Euler-Maruyama method with time step $\Delta t = 10^{-3}$ and the same randomly generated numbers. The parameters are the same as Figure 6-2 and time is normalised with reference to the average cell-cycle time. In panel (b) we plot the distribution of Q at three time points ($t = 1$, $t = 3$ and $t = 5$). The two overlaid histograms represent the distributions of 1000 independent simulations of the OU process (red) and the Langevin Equation (blue). The black line represent the distribution $N(Q_*, \sigma_Q(t))$. All the parameters are the same as Figure 6-2. Panel (c) shows the asymptotic value of CV_G and CV_N for a range of parameters choices, N_0 and K . The red dot corresponds to the parameters inferred from the data in Section 6.5.4, indicating that Q will be Gaussian distributed for biologically realistic parameter values.

6.5.2 The correlation matrix

For a stochastic initial condition, x_0 , as described in Section 6.3.3, we can compute the correlation matrix at time $t = 0$, as

$$C_0 = \mathbb{E}^0 [x_i(0)x_j(0)] = \begin{cases} u_i u_j & \text{for } i \neq j \\ u_i^2 + \frac{u_i}{N_0} & \text{for } i = j \end{cases}. \quad (6.12)$$

We can rewrite this as $C_0 = \mathbf{u}\mathbf{u}^T + \frac{1}{N_0}M$, where $M = \text{Diag}(\mathbf{u})$.

We then focus on computing the correlation matrix $C(t, t') = \mathbb{E} [\hat{x}(t) \hat{x}^T(t')]$ for the OU process (6.11), as an approximation for the correlation matrix of the Langevin model. By applying general results for OU processes (See Section 4.5 of [Gardiner, 2009]) we have:

$$C(t, t') = e^{KtS} C_0 e^{Kt'S^T} + \frac{K}{N_0} \int_0^{\min(t, t')} e^{K(t-\tau)S} \mathcal{S} \left(M e^{K\lambda\tau} \right) \mathcal{S}^T e^{K(t'-\tau)S^T} d\tau. \quad (6.13)$$

We can use expression (6.5) and the fact that $e^{KtS} \mathbf{u}\mathbf{u}^T e^{Kt'S^T} = \mathbf{u}\mathbf{u}^T e^{K(t+t')\lambda}$, to write down the (i, j) element of (6.13) for $t < t'$ as

$$\begin{aligned} C_{i,j}(t, t') = & u_i u_j e^{K(t+t')\lambda} + \frac{1}{N_0 K^2} \sum_{k,l,m=1}^K \frac{1}{(1+\lambda_k)^{i-m}} \frac{1}{(1+\lambda_l)^{j-m}} \frac{2\lambda}{(1+\lambda)^m} e^{K(t\lambda_k + t'\lambda_l)} \\ & + \frac{1}{N_0 K} \sum_{k,l,m=1}^K \frac{\lambda_k}{(1+\lambda_k)^{i-m}} \frac{\lambda_l}{(1+\lambda_l)^{j-m}} \frac{2\lambda}{(1+\lambda)^m} \int_0^t e^{K[(t-\tau)\lambda_k + (t'-\tau)\lambda_l + \lambda\tau]} d\tau. \end{aligned} \quad (6.14)$$

Substituting the expressions (6.3) for \mathbf{u}^k and \mathbf{v}^k , using the formula

$$\sum_{m=1}^K \frac{(1+\lambda_k)^m (1+\lambda_l)^m}{(1+\lambda)^m} = \frac{(1+\lambda_k)(1+\lambda_l)}{(1+\lambda_k)(1+\lambda_l) - (1+\lambda)} \quad (6.15)$$

and upon rearranging terms, we obtain

$$\begin{aligned} C_{i,j}(t, t') = & \frac{4\lambda^2}{(1+\lambda)^{i+j}} e^{K(t+t')\lambda} + \frac{2\lambda}{N_0 K^2} \sum_{k,l=1}^K \frac{(1+\lambda_l)^{1-j}}{(1+\lambda_k)^{i-1}} \frac{1}{(\lambda_k + \lambda_l - \lambda)} \left[e^{K(t\lambda_k + t'\lambda_l)} \right. \\ & \left. - \frac{\lambda_k \lambda_l}{(1+\lambda_k)(1+\lambda_l) - (1+\lambda)} e^{K((t'-t)\lambda_l + t\lambda)} \right]. \end{aligned} \quad (6.16)$$

6.5.3 The envelope of two standard deviations of Q

We recall the definition of the envelope of two standard deviations of $Q(t)$ as $\Omega(t) = [Q_* - 2\sigma_Q(t), Q_* + 2\sigma_Q(t)]$, where $\sigma_Q(t)$ denotes the standard deviation of $Q(t)$. To compute σ_Q we employ the OU approximation (see Section 6.5.1). From a fixed initial condition, the solutions of (6.11) evolve as a Gaussian process with mean $\bar{x}(t)/N_0$. We

can write $G(t) \sim \mathcal{N}(\mu_G(t), \sigma_G(t))$ and $N(t) \sim \mathcal{N}(\mu_N(t), \sigma_N(t))$ where

$$\mu_G(t) = Q_* e^{K\lambda t}, \quad \sigma_G^2(t) = \sum_{i,j=1}^{\alpha K} C_{i,j}(t, t) - Q_*^2 e^{2K\lambda t}, \quad (6.17a)$$

$$\mu_N(t) = e^{K\lambda t}, \quad \sigma_N^2(t) = \sum_{i,j=1}^K C_{i,j}(t, t) - e^{2K\lambda t}. \quad (6.17b)$$

Notice that Q is defined as a ratio between two Gaussian distributions and, in general, this does not imply that $Q(t)$ is Gaussian. However, Hayya et al. [1975] showed that the ratio of two Gaussian can be well approximated as a Gaussian, under certain conditions on the *coefficient of variation* (CV) of the numerator and denominator. Precisely, provided that

$$CV_N = \frac{\sigma_N}{\mu_N} < 0.39 \quad \text{and} \quad CV_{G1} = \frac{\sigma_{G1}}{\mu_{G1}} > 0.005 \quad (6.18)$$

Hayya et al. [1975] demonstrate that Q is close to a Gaussian distribution. Moreover, we can approximate the variance of Q by Taylor expanding to the second order which leads to

$$\begin{aligned} \sigma_Q^2 &\approx \sigma_N^2 \frac{\mu_G^2}{\mu_N^4} + \frac{\sigma_G^2}{\mu_N^2} - 2\rho\mu_G \frac{\sigma_N\sigma_G}{\mu_N^3} \\ &= \frac{1}{\mu_N^2} [\sigma_N^2 Q_*^2 + \sigma_G^2 - 2\sigma_N\sigma_G Q_* \rho [G, N]] , \end{aligned} \quad (6.19)$$

where ρ denotes the correlation coefficient, defined as

$$\rho[Y_1, Y_2] = \frac{\mathbb{E}[Y_1 Y_2] - \mathbb{E}[Y_1] \mathbb{E}[Y_2]}{\sqrt{\text{Var}[Y_1] \text{Var}[Y_2]}}. \quad (6.20)$$

Notice that we can compute $\mathbb{E}[G(t)N(t)]$ in equation (6.20) in terms of the correlation matrix C as

$$\mathbb{E}[G(t)N(t)] = \sum_{i=1}^K \sum_{j=1}^{\alpha K} C_{i,j}(t, t).$$

We now need to check that the conditions (6.18) are satisfied for biologically relevant parameter choices. By studying the expressions (6.16) and (6.17), we obtain that $CV_N(0) = 1/\sqrt{N_0}$ and $CV_G(0) = 1/\sqrt{Q_* N_0}$ which satisfy the conditions (6.18) for $\alpha \in [0, 1]$ and $N_0 \in [10, 10^4]$. In order to check the validity of the conditions in the long-term, we look at the leading terms of the expression (6.16). We find that

$$\lim_{t \rightarrow +\infty} CV_N(t) = \lim_{t \rightarrow +\infty} CV_G(t) = \frac{1 + \lambda}{K\lambda\sqrt{2N_0}}. \quad (6.21)$$

In Figure. 6-6(c) we evaluated this expression for any $K \in [1, 100]$ and $N_0 \in [20, 500]$. Our findings show that the limit of CV_G and CV_N for $t \rightarrow +\infty$ lies in the interval $(0.01, 0.32)$ for the range of parameters considered which suggests that the conditions (6.18) are satisfied for biologically relevant choices of the parameters. Notice that the

plots in Figure 6-6(b) provide further confirmation of this by showing good agreement between the distribution of Q and the Gaussian distribution $\mathcal{N}(Q_*, \sigma_Q(t))$.

6.5.4 Parameters inference

To infer the parameters of the MS model, we simultaneously fit the distribution of the total cell-cycle time and of the G1 duration of 200 randomly selected cells.

Let \mathbf{H}_T and \mathbf{H}_{G1} denote the histogram representations of the pdf of the total cell-cycle time and the G1 duration, respectively, with a bin width of one hour. For example, $(\mathbf{H}_T)_i$ denotes the proportion of cells with a cell-cycle time in the interval $[ih, (i+1)h)$. We denote with $\mathbf{H}_{E(K,\beta)}$ the histogram obtained by discretising an Erlang distribution with parameters (K, β) with the same bin width, *i.e.* $(\mathbf{H}_{E(K,\beta)})_i = \frac{\beta^K}{(K-1)!} \int_i^{i+1} x^{K-1} e^{-\beta x} dx$.

For a given combination of parameters, (K, β, α) , one can consider the statistic

$$I(K, \beta, \alpha) = \|\mathbf{H}_T - \mathbf{H}_{E(K,\beta)}\|_1 + c \|\mathbf{H}_{G1} - \mathbf{H}_{E(\alpha K, \beta)}\|_1, \quad (6.22)$$

where $c > 0$ is a constant and $\|\cdot\|_1$ denotes the 1-norm. Notice that the constant c can be interpreted as a weight to give more ($c > 1$) or less ($c < 1$) priority at the fitting of the G1 distribution compared to the one of the total cell-cycle time distribution. For simplicity we choose $c = 1$, which corresponds to equal levels of priority for the two distribution fits.

To determine the parameter combination which provides the best simultaneous fit of the two distribution, we evaluated the function I in the parameter range $K \in [10, 150]$, $\beta \in [1, 10]$ and $\alpha \in [0, 1]$. We find that the combination $K^* = 92$, $\beta^* = 4.96$ and $\alpha^* = 33/92$ minimises the statistic I in the parameter region considered and, hence, we select these parameters for the MS model.

To infer the average population size at the moment of the initial sampling, N_0 , we first measure the average population size at the beginning of the recording, $N_{24} = 381.1$, averaged over the 30 experiments. Since the average population size grows exponentially at rate λ , we project back from time $t = 24h$ and we obtain the average sample size as $N_0 = N_{24} \exp(-24\lambda) \approx 155$.

6.5.5 The Kullback Leibler divergence between Erlang and Gaussian distribution

We compute the relative entropy (Kullback-Leibler divergence, D_{KL}) between an Erlang and a Gaussian distribution as a measure of the distance between the two distributions.

For two distributions, $p(x)$ and $q(x)$, the KL divergence is defined as:

$$D(p, q) = \int_{-\infty}^{\infty} p(x) \log \left[\frac{p(x)}{q(x)} \right] dx. \quad (6.23)$$

We set $p(x)$ to be the probability density function (pdf) of an $Erlang(K, \beta)$ and $q(x)$ to be the pdf of a Gaussian with same mean and variance, *i.e.* $\mathcal{N}\left(\frac{K}{\beta}, \frac{K}{\beta^2}\right)$. We then obtain

$$\begin{aligned}
D(K, \beta) &= D\left(p(K, \beta), q\left(\frac{K}{\beta}, \frac{K}{\beta^2}\right)\right) \\
&= \frac{\beta^K}{(K-1)!} \int_0^\infty x^{K-1} e^{-\beta x} \log \left[\frac{\beta^{K-1} \sqrt{2\pi K}}{(K-1)!} x^{K-1} e^{\frac{\beta^2}{2K} \left(x - \frac{K}{\beta}\right)^2 - \beta x} \right] dx \\
&= \log \left[\frac{\beta^{K-1} \sqrt{2\pi K}}{(K-1)!} \right] \frac{\beta^K}{(K-1)!} \int_0^\infty x^{K-1} e^{-\beta x} dx \\
&\quad + (K-1) \frac{\beta^K}{(K-1)!} \int_0^\infty x^{K-1} \log(x) e^{-\beta x} dx \\
&\quad + \frac{\beta^K}{(K-1)!} \int_0^\infty \left[-\beta x + \frac{\beta^2}{2K} \left(x - \frac{K}{\beta}\right)^2 \right] x^{K-1} e^{-\beta x} dx. \tag{6.24}
\end{aligned}$$

Notice that the first integral of Equation (6.24) is exactly the pdf of an Erlang, which simplifies to unity. The second and third integral in (6.24) require more work. By using integration by parts and upon simplification, we get to the final expression

$$D(K, \beta) = \log \left[\frac{\beta^{K-1} \sqrt{2\pi K}}{(K-1)!} \right] + (K-1) (\mathcal{H}_{K-1} - \log(\beta) - \gamma) - K + \frac{1}{2}, \tag{6.25}$$

where $\mathcal{H}_{K-1} = \sum_{i=1}^{K-1} \frac{1}{i}$ is the $(K-1)$ -th harmonic number and γ denotes the Euler-Mascheroni constant.

Using the expression (6.25) it is possible to show that $D(K, \beta)$ is a decreasing function of K and $D(K, \beta) \sim \mathcal{O}(K^{-1})$ for $K \rightarrow +\infty$. This is not a surprise, since by central limit theorem we know that the Erlang distribution converges to a Gaussian with same mean and variance. Since the CV of an $Erlang(K, \beta)$ is given by $K^{-\frac{1}{2}}$, we can rephrase by saying that the KL divergence scales proportionally to the square of the CV of the Erlang distribution.

Figure 6-7 shows the plot of $D(K, \beta)$ with $\beta = K$ for different values of K . In the overlaid panels the two distributions are compared for $K = 5, 10, 20, 30$ and 60 . The results highlight the good level of similarity between the Erlang and Gaussian distributions for large K - small values of the CV. For example, for $K > 25$, *i.e.* $CV < 0.2$, we have $D(K, K) < 0.02$ which corresponds to good agreement between the two distributions.

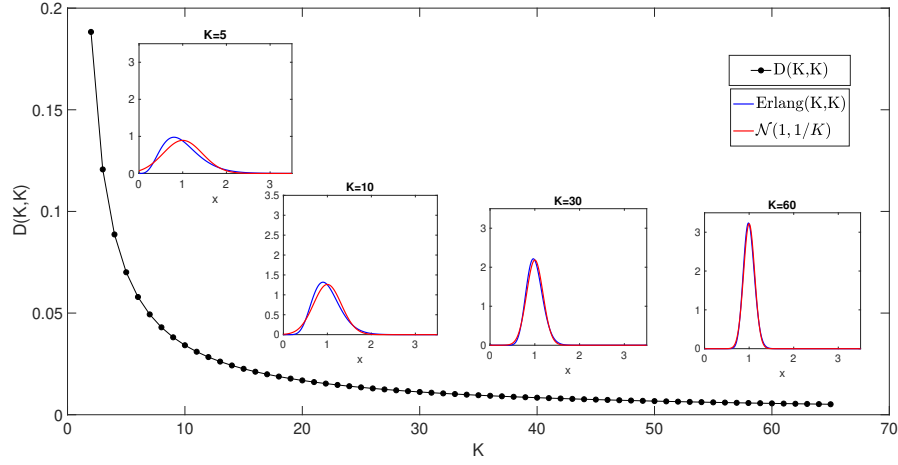


Figure 6-7: The Kullback Leibler (KL) divergence between an Erlang distribution and Gaussian distribution. The black dotted line in the main panel shows the KL divergence between an Erlang distribution of parameters (K, K) and a Gaussian distribution of parameters $(1, 1/K)$ as function of K . The four overlaid panels show the comparison of the two distributions, Erlang (blue) and Gaussian (Red), for $K = 5, 10, 30$ and 60 (from left to right).

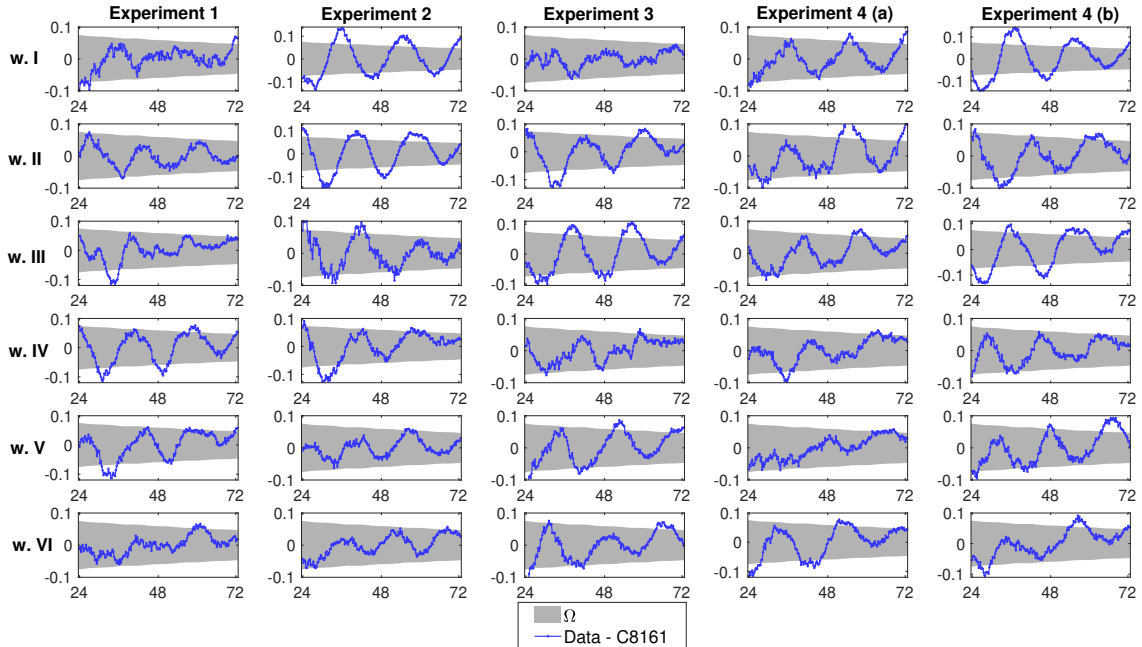


Figure 6-8: Comparison 30 time series obtained from the data (blue lines), together with the envelope of two standard deviations, Ω (light grey regions) predicted using the MS model. The parameters of the MS models are obtain by fitting the distribution of the total cell-cycle time and G1 duration (see Section 6.5.4): $K = 92, \alpha K = 33, \beta = 4.96h^{-1}$ and $N_0 = 155$.

6.6. Conclusions

In this chapter we studied the impact of stochasticity in a structured population of melanoma cells. We employed a multi-stage model to represent the cell cycle and by studying corresponding macroscopic and mesoscopic models we elucidated the role of finite-size effects in the emergence of stochastic fluctuations. The results of the comparison with the experimental data show that the demographic noise intrinsic in cell proliferation events is capable of recapitulating the fluctuations observed in the experiments.

In the context of the thesis, the focus of this chapter is on how to interpret biological observations which are naturally affected by intrinsic noise due to the stochasticity of cellular events. This study provides a good example of a model system in which accounting for stochastic effects at the microscopic level of a single cell plays a decisive role in the analysis of the experimental observations and the interpretation of macroscopic phenomena.

Chapter 7

Final Conclusions and Outlook

The main object of my PhD has been the study of complex phenomena arising in cellular populations. In this thesis I present a series of studies in which, by employing a multiscale modelling approach, I investigate the effect of certain microscopic properties of cell behaviour on the total population dynamics.

In Chapter 2 we set the context of the thesis by providing a general introduction to the topic of multiscale modelling paying particular attention to the process of cell migration. We reviewed the common mathematical framework which allows the derivation of a deterministic, macroscopic PDEs from a stochastic ABM of cellular behaviour. In Chapter 3 we presented our work on modelling the interaction of directional persistence and volume exclusion in a multiscale fashion. We derived diffusive macroscopic models from a set of ABMs which incorporate the tendency of cells to persist in their direction for short periods of time together with exclusion properties of increasing complexity. The results show good agreement between the two levels of description, but they also highlight some peculiar behaviours of such models including density spikes, anisotropy and a spontaneous form of aggregation. Inspired by these results, in Chapter 4 we studied the problem of detecting the presence of aggregation in a multi-agent system by quantifying the spatial pairwise correlation of the agents. We analysed a list of PCFs accounting for different distance metrics and boundary conditions which allow the measurement of pair-correlation in discrete domains such as two- and three-dimensional lattices and other more general examples.

In the second part of the thesis, we shifted the main focus of the study to cell proliferation. In Chapter 5 we investigated the role played by the cell-cycle time distribution in the context of an invasion front. In particular, we elucidated to what extent and in what way the distribution of the cell cycle can affect the speed of a cellular invasion. One of the most important results of this study is provided by Theorem 2, in which we determined analytically the range of speed for a class of biologically motivated distributions known as hypo-exponentials. Our findings suggest that, for realistic distributions, the range of invasion speed is primarily determined by the average cell-cycle time, whereas the higher moments of the distribution play a relatively small role and their contribution is limited. Contrastingly, in Chapter 6 we presented a study in which accounting for the cell-cycle time distribution is crucial and it can influence the analysis of experimental observations. We showed that the stochasticity intrinsic in the process of cell proliferation generates long-lasting damped oscillations in the proportion of cells in the G1-phase of the cell cycle. By parametrising the model using experimentally derived data from a proliferation assay of melanoma cells, we found that the oscillations

triggered by the intrinsic demographic noise are quantitatively consistent with those observed in the experiments. Our findings underline the importance of accounting for a realistic representation of the stochasticity due to cell proliferation events. We also described a mathematical protocol to quantify the emergence of noise-induced macroscopic phenomena in age-structured populations.

Overall, with this thesis I aim to improve the understanding of the connection between single cell behaviour and global population dynamics. The research presented in this thesis left several questions unanswered and raised more challenges for future studies. In the following section I highlight some avenues of further investigation which I consider of particular interest.

7.1. Future research

PCF as spatial summary statistics for directional persistence

One aspect of the research which is yet to be explored is whether the PFC defined in Chapter 4 is a suitable summary statistic to measure directional persistence in crowded environments. The typical approach to infer a set of parameters, Θ , for a general mathematical model $M(\Theta)$, consists of using a likelihood function which expresses the probability of observed data, D , under a choice of parameters, Θ . Nevertheless an analytical formula for the likelihood function for complex models is often hard to obtain or computationally prohibitive and alternative approaches are necessary. Approximate Bayesian computation (ABC) is an alternative probabilistic approach for estimating the posterior probability distribution that Θ are the model parameters which generate the data D . Depending on the nature of the biological process, the dimensionality of the data can make comparison between data sets extremely expensive or intractable. A way to reduce the dimension of the data, is to utilise a summary statistic of the data, $S(D)$, and an appropriate measure of distance between outcomes of the summary statistic, $d_S(D1, D2)$. Given a chosen summary statistic $S(D)$, the corresponding distance measure, d_S , a prior distribution π and a threshold ϵ , the ABC algorithm consists of repeating the following steps n times :

1. sample a set of parameters $\tilde{\Theta}$ from the prior distribution π ;
2. generate the data \tilde{D} using the model with parameters $\tilde{\Theta}$: $\tilde{D} \sim M(\tilde{\Theta})$;
3. compute the distance between the synthetic data and the experimental data according to the summary statistic, $\rho = d_S(\tilde{D}, D)$;
4. accept the value of $\tilde{\Theta}$ if $\rho < \epsilon$.

For large values of n , the set of accepted values of the algorithm is a sample of parameter values distributed approximately according to the desired posterior distribution. The choice of the summary statistic is not obvious, yet is crucial, since the effectiveness

of the algorithm to identify the model parameters depends on the chosen S [Ross et al., 2017]. One way to test if a certain summary statistic S is effective for estimating the model parameters from real data, is to employ the ABC algorithm on synthetic data, $\hat{D} \sim M(\hat{\Theta})$, generated for a chosen set of parameters $\hat{\Theta}$. If the resulting posterior distribution is largely focused around the correct parameters, $\hat{\Theta}$, the summary statistic is suitable.

Many spatial statistics have been suggested to measure directional persistence (see Codling et al. [2008] for a thorough review), the most popular being the mean squared displacement (MSD) [Okubo and Levin, 2013], the mean dispersal distance (MDD) [Bovet and Benhamou, 1988, Byers, 2001] and the path tortuosity [Benhamou, 2004]. These statistics have complementary strengths and limitations, as discussed by Codling et al. [2008]. However, since they all focus on single particle trajectories, their usage is typically limited to situations in which particle interactions are rare and in which excluding properties are often ignored [Codling and Hill, 2005].

PCFs have been shown to perform well as summary statistics in scenarios where spatial correlation is important [Johnston et al., 2014]. In the context of directional persistence, therefore, it is plausible that employing a PCF as summary statistic - or combining it with one of other existing functions [Harrison and Baker, 2017]), might improve the performance of an ABC algorithm in scenarios in which crowding plays an important role. Chapter 3 and 4 provide all the fundamental elements to investigate this hypothesis using a spatially discrete approach, which will be undertaken in future work.

Phase-dependent persistence and invasion speed

Another interesting aspect which requires further work is the connection between directional persistence and the cell-cycle. Despite these two aspects of cell behaviour having been extensively studied independently, it is only in recent years that there has been an increasing interest in studying the connections between properties of cell movement and cell-cycle phase [Vittadello et al., 2019a, Baker and Röst, 2019]. In particular, it remains unclear whether the intensity of directional persistence can change during the cell-cycle and what impact this phenomenon would have on an invading front.

In this thesis, we introduce two new modelling approaches to study persistence of motion and the cell-cycle progression, in Chapter 3 and Chapter 5, respectively. By considering an appropriate combination of the ABMs and of the corresponding deterministic representations, it is possible to account for phase-dependent persistence and to investigate the impact on the invasion speed. However, several challenges will arise in the analysis of these models. In particular the presence of stage-dependent diffusivities in the deterministic representation undermines most of the analytical wavespeed analysis carried out in Chapter 5 and more sophisticated techniques need to be employed.

The role of spatial dispersion in structured growing populations

An important avenue of research regarding cell proliferation is related to the last paper presented (Chapter 6). Throughout this study, the role of spatial effects and cell motility has been omitted completely for mathematical simplicity. In reality, cell division naturally induces positive correlation at short-distances by displacing daughter cells at the same cell-cycle phase, close to each other. This local synchronisation phenomenon, however, is counteracted by the effect of cell migration which tends to disperse synchronised cells through the domain. In other words, within this tradeoff between proliferation-induced correlation and motility-induced dispersion lies an important dynamic of the proliferation assay which should be taken into account when studying fluctuating oscillatory events, like those presented in Chapter 6.

In fact, the study reported in Chapter 6 represents a first step for the understanding of the emerging fluctuations. Clearly, the same mathematical framework could be adapted to account for spatial extent and cell motility. For example, agents could be allowed to perform a simple random walk on a two-dimensional lattice which would lead to diffusive terms in the macroscopic description of the same form as those presented in Chapter 5. However, most of the correlation analysis carried out in Section 6.5 would become considerably more complicated and probably intractable. A way to simplifying the analysis, would be by considering a finite set of Langevin equations, as equation (6.8), describing the dynamics of the cell-stage distribution in local regions of the domain. Cell motility can then be incorporated in the model as a series of reactions terms between Langevin equations of adjacent regions. By performing a similar analysis, as in Chapter 6, for each Langevin model one can study the emergence of local oscillations as function of cell motility and of local density. An alternative approach consists of systematically simulating the multi-stage ABM and employing a PCF, as in Chapter 4, to study the interplay of proliferation and dispersion in light of the emergence of local phase synchronisation.

In summary, this thesis represents an exploration of multiscale phenomena arising from cell behaviours and their interaction. We designed and analysed a series of microscopic and macroscopic models capturing some fundamental aspects of cell migration and cell proliferation. From a general perspective, the studies presented in this thesis demonstrate the power of a multiscale modelling approach which, by exploiting complementary strengths of different models allows a deeper understanding of complex phenomena.

Bibliography

- R.E. Baker and G. Röst. Global dynamics of a novel delayed logistic equation arising from cell biology. *arXiv:1901.07817*, 2019.
- R.E. Baker and M.J. Simpson. Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E*, 82(4):041905, 2010.
- R.E. Baker, C.A. Yates, and R. Erban. From microscopic to macroscopic descriptions of cell migration on growing domains. *Bull. Math. Biol.*, 72(3):719–762, 2010.
- K. A Beaumont, D.S. Hill, S.M. Daignault, Goldie Y.L.L., D.M. Sharp, B. Gabrielli, W. Weninger, and N.K. Haass. Cell cycle phase-specific drug resistance as an escape mechanism of melanoma cells. *J. Investig. Dermatol.*, 136(7):1479–1489, 2016.
- J.B. Beltman, A.F.M. Marée, and R.J. De Boer. Analysing immune cell migration. *Nat. Rev. Immunol.*, 9(11):789, 2009.
- S. Benhamou. How to reliably estimate the tortuosity of an animal’s path: straightness, sinuosity, or fractal dimension? *J. Theor. Biol.*, 229(2):209–220, 2004.
- A.M. Berdahl, A.B. Kao, A. Flack, P.A.H. Westley, E.A. Codling, I.D. Couzin, A.I. Dell, and D. Biro. Collective animal navigation and migratory culture: from theoretical models to empirical evidence. *Phil. Trans. R. Soc. B*, 373(1746):20170009, 2018.
- T. Bertrand, P. Illien, O. Bénichou, and R. Voituriez. Dynamics of run-and-tumble particles in dense single-file systems. *N. J. Phys.*, 20(11):113045, 2018.
- A.V.P. Bobadilla, T. Carraro, H.M. Byrne, P.K. Maini, and T. Alarcón. Age structure can account for delayed logistic proliferation of scratch assays. *Bull. Math. Biol.*, 81(7):2706–2724, 2019.
- M. Bodnar and J. Jose L. Velazquez. Derivation of macroscopic equations for individual cell-based models: a formal approach. *Math. Models Methods Appl. Sci.*, 28(15):1757–1779, 2005.
- P. Bovet and S. Benhamou. Spatial analysis of animal movements using a correlated random walk model. *J. Theor. Biol.*, 131(4):419–433, 1988.
- J.A. Byers. Correlated random walk equations of animal dispersal resolved by simulation. *Ecology*, 82(6):1680–1690, 2001.
- H.X. Chao, R.I. Fakhreddin, H.K. Shimerov, R.J. Kumar, G.P. Gupta, and J.E. Purvis. Evidence that the cell cycle is a series of uncoupled, memoryless phases. *bioRxiv*, page 283614, 2018.
- E.A. Codling and N.A. Hill. Calculating spatial statistics for velocity jump processes with experimentally observed reorientation parameters. *J. Math. Biol.*, 51(5):527–556, 2005.
- E.A. Codling, M.J. Plank, and S. Benhamou. Random walk models in biology. *J. R. Soc. Interface*, 5(25):813–834, 2008.
- I.D. Couzin. Collective cognition in animal groups. *T. in cogn. sci.*, 13(1):36–43, 2009.
- I.D. Couzin, J. Krause, R. James, G.D. Ruxton, and N.R. Franks. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.*, 218(1):1–11, 2002.
- C. Cutright, Z. Brotherton, L. Alexander, J. Harris, K. Shi, S. Khan, J. Genzer, and S. Menegatti. Packing density, homogeneity, and regularity: Quantitative correlations between topology and thermoresponsive morphology of pnipam-co-paa microgel coatings. *Appl. Surf. Sci.*, 508:145129, 2020.
- M. Deng, W.L. Chen, A. Takatori, Z. Peng, L. Zhang, M. Mongan, R. Parthasarathy, M. Sartor, M. Miller, J. Yang, et al. A role for the mitogen-activated protein kinase kinase 1 in epithelial wound healing. *Mol. Biol. Cell*, 17(8):3446–3455, 2006.

- R. Erban and H.G. Othmer. From individual to collective behavior in bacterial chemotaxis. *SIAM J. Appl. Math.*, 65(2):361–391, 2004.
- G.I. Evan and K.H. Vousden. Proliferation, cell cycle and apoptosis in cancer. *nature*, 411(6835):342, 2001.
- G. Gabella. On the plasticity of form and structure of enteric ganglia. *Auton. Neurosci.*, 30:S59–S66, 1990.
- M. Galanti, D. Fanelli, and F. Piazza. Persistent random walk with exclusion. *Eur. Phys. J. B*, 86(11):456, 2013.
- C. Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.
- E. Gavagnin and C.A. Yates. Modeling persistence of motion in a crowded environment: The diffusive limit of excluding velocity-jump processes. *Phys. Rev. E*, 97:032416, 2018a.
- E. Gavagnin and C.A. Yates. Chapter 2 - stochastic and deterministic modeling of cell migration. In Arni S.R. Srinivasa Rao and C.R. Rao, editors, *Integrated Population Biology and Modeling, Part A*, volume 39 of *Handbook of Statistics*, pages 37 – 91. Elsevier, 2018b.
- E. Gavagnin, J.P. Owen, and C.A. Yates. Pair correlation functions for identifying spatial correlation in discrete domains. *Phys. Rev. E*, 97:062104, 2018.
- E. Gavagnin, M.J. Ford, R.L. Mort, T. Rogers, and C.A. Yates. The invasion speed of cell migration models with realistic cell cycle time distributions. *J. Theor. Biol.*, 481:91–99, 2019.
- S.F. Gilbert. *Developmental Biology*. Sunderland, MA. Sinauer Associates, Inc, 2000.
- N.K. Haass and B. Gabrielli. Cell cycle-tailored targeting of metastatic melanoma: Challenges and opportunities. *Exp. Dermatol.*, 26(7):649–655, 2017.
- N.K. Haass, K.A. Beaumont, D.S. Hill, A. Anfosso, P. Mrass, M.A. Munoz, I. Kinjyo, and W. Weninger. Real-time cell cycle imaging during melanoma growth, invasion, and drug response. *Pigment Cell Melanoma Res.*, 27(5):764–776, 2014.
- D. Hanahan and R.A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.
- D.E.F. Harrison and H.H. Topiwala. Transient and oscillatory states of continuous culture. In *Adv. Bio. Eng. Vol. 3*, pages 167–219. Springer, 1974.
- J.U Harrison and R.E. Baker. An automatic adaptive method to combine summary statistics in approximate bayesian computation. *arXiv preprint arXiv:1703.02341*, 2017.
- H. Hatzikirou, D. Basanta, M. Simon, K. Schaller, and A. Deutsch. ‘go or grow’: the key to the emergence of invasion in tumour progression? *Math. Med. Biol.*, 29(1):49–65, 2012.
- J. Hayya, D. Armstrong, and N. Gressis. A note on the ratio of two normally distributed variables. *Manag. Sci.*, 21(11):1338–1341, 1975.
- F. Jafarpour. Cell size regulation induces sustained oscillations in the population growth rate. *Phys. Rev. Lett.*, 122(11):118101, 2019.
- F. Jafarpour, C.S. Wright, H. Gudjonson, J. Riebling, E. Dawson, K. Lo, A. Fiebig, S. Crosson, A.R. Dinner, and S. Iyer-Biswas. Bridging the timescales of single-cell and population dynamics. *Phys. Rev. X*, 8(2):021007, 2018.
- J. Jhavar, R.G Morris, and V. Guttal. Deriving mesoscopic models of collective behavior for finite populations. In *Handbook of Statistics*, volume 40, pages 551–594. Elsevier, 2019.
- S.T. Johnston and E.J. Crampin. Corrected pair correlation functions for environments with obstacles. *Phys. Rev. E*, 99(3):032124, 2019.

- S.T. Johnston, M.J. Simpson, D.L.S. McElwain, B.J. Binder, and J.V. Ross. Interpreting scratch assays using pair density dynamics and approximate bayesian computation. *Op. bio.*, 4(9):140097, 2014.
- E.F. Keller and L.A. Segel. Model for chemotaxis. *J. Theor. Biol.*, 30(2):225–234, 1971.
- R. Keller. Cell migration during gastrulation. *Curr. Opin. Cell Biol.*, 17(5):533–541, 2005.
- K. Kuritz, W. Halter, and F. Allgöwer. Passivity-based ensemble control for cell cycle synchronization. In *Emerging Applications of Control and Systems Theory*, pages 1–13. Springer, 2018.
- A.K Laird. Dynamics of tumour growth: comparison of growth rates and extrapolation of growth curve to one cell. *Br. J. Cancer*, 19(2):278, 1965.
- M. Lang, S. Waldherr, and F. Allgöwer. Amplitude distribution of stochastic oscillations in biochemical networks due to intrinsic noise. *PMC biophysics*, 2(1):10, 2009.
- P.K. Maini, D.L. S. McElwain, and D.I. Leavesley. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.*, 10(3-4):475–482, 2004.
- G.W. McLean, N.O. Carragher, E. Avizienyte, J. Evans, V.G. Brunton, and M.C. Frame. The role of focal-adhesion kinase in cancer—a new therapeutic opportunity. *Nat. Rev. Cancer*, 5(7):505, 2005.
- J. Monod. The growth of bacterial cultures. *Annu. Rev. Microbiol.*, 3(1):371–394, 1949.
- R.G. Morris and T. Rogers. Growth-induced breaking and unbreaking of ergodicity in fully-connected spin systems. *J. Phys. A-Math. Theor.*, 47(34):342003, 2014.
- R.L. Mort, R.J.H. Ross, K.J. Hainey, O.J. Harrison, M.A. Keighren, G. Landini, R.E. Baker, K.J. Painter, I.J. Jackson, and C.A. Yates. Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model. *Nat. Commun.*, 7, 2016.
- J.D. Murray. *Mathematical biology: I. An introduction*, volume 17. Springer Science & Business Media, 2007.
- A. Okubo and S.A. Levin. *Diffusion and ecological problems: modern perspectives*, volume 14. Springer Science & Business Media, 2013.
- H.G. Othmer, S.R. Dunbar, and W. Alt. Models of dispersal in biological systems. *J. Math. Biol.*, 26(3):263–298, 1988.
- R.B. Paris. Exponential asymptotics of the Mittag-Leffler function. *Proc. R. Soc., Ser. B, London*, 458(2028):3041–3052, 2002.
- D. Pirjol, F. Jafarpour, and S. Iyer-Biswas. Phenomenology of stochastic exponential growth. *Phys. Rev. E*, 95(6):062406, 2017.
- T.L. Riss, R.A. Moravec, A.L. Niles, S. Duellman, H.A. Benink, T.J. Worzella, and L. Minor. Cell viability assays. In *Assay Guidance Manual*. Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2016.
- R.J.H. Ross, R.E. Baker, A. Parker, M.J. Ford, R.L. Mort, and C.A. Yates. Using approximate bayesian computation to quantify cell–cell adhesion parameters in a cell migratory process. *npj Syst. Biol. Appl.*, 3, 2017.
- A. Rühl. Glial cells in the gut. *Neurogastroenterol. Motil.*, 17(6):777–790, 2005.
- A. Sakaue-Sawano, H. Kurokawa, T. Morimura, A. Hanyu, H. Hama, H. Osawa, S. Kashiwagi, K. Fukami, T. Miyata, H. Miyoshi, et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell*, 132(3):487–498, 2008.
- J.A. Sherratt and M.A.J. Chaplain. A new mathematical model for avascular tumour growth. *J. Math. Biol.*, 43(4):291–312, 2001.

- M.J. Simpson, D.C. Zhang, M. Mariani, K.A. Landman, and D.F. Newgreen. Cell proliferation drives neural crest cell invasion of the intestine. *Dev. Biol.*, 302(2):553–568, 2007. ISSN 0012-1606.
- M.J. Simpson, K.A. Landman, and B.D. Hughes. Multi-species simple exclusion processes. *Phys. A*, 388(4):399–406, 2009.
- M.J. Simpson, W. Jin, S.T. Vittadello, T.A. Tambyah, J.M. Ryan, G. Gunasingh, N.K. Haass, and S.W. McCue. Stochastic models of cell invasion with fluorescent cell cycle indicators. *Phys. A*, 510:375–386, 2018.
- E. Teomy and R. Metzler. Correlations and transport in exclusion processes with general finite memory. *J. Stat. Mech: Theory Exp.*, 2019(10):103211, 2019a.
- E. Teomy and R. Metzler. Transport in exclusion processes with one-step memory: density dependence and optimal acceleration. *J. Phys. A.-Math. Theor.*, 52(38):385001, 2019b.
- S.T. Vittadello, S.W. McCue, G. Gunasingh, N.K. Haass, and M.J. Simpson. Mathematical models for cell migration with real-time cell cycle dynamics. *Biophys. J.*, 114(5), 2018.
- S.T. Vittadello, S.W. McCue, G. Gunasingh, N.K. Haass, and M.J. Simpson. Examining go-or-grow using fluorescent cell-cycle indicators and cell cycle-inhibiting drugs. *bioRxiv*, page 797142, 2019a.
- S.T. Vittadello, S.W. McCue, G. Gunasingh, N.K. Haass, and M.J. Simpson. Mathematical models incorporating a multi-stage cell cycle replicate normally-hidden inherent synchronization in cell proliferation. *J. R. Soc. Interface*, 16(20190382), 2019b.
- C.A. Yates, R. Erban, C. Escudero, I.D. Couzin, J. Buhl, I.G. Kevrekidis, P.K. Maini, and D.J.T. Sumpter. Inherent noise can facilitate coherence in collective swarm motion. *Proc. Natl. Acad. Sci. USA*, 106(14):5464–5469, 2009.
- C.A. Yates, R.E. Baker, R. Erban, and P.K. Maini. Going from microscopic to macroscopic on nonuniform growing domains. *Phys. Rev. E*, 86(2):021921, 2012.
- C.A. Yates, M.J. Ford, and R.L. Mort. A multi-stage representation of cell proliferation as a markov process. *Bull. Math. Biol.*, 79(12):2905–2928, 2017.
- S. Zhang, A. Chong, and B.D. Hughes. Persistent exclusion processes: Inertia, drift, mixing, and correlation. *Phys. Rev. E*, 100(4):042415, 2019.