From microscopic to macroscopic descriptions of cell migration on growing domains

Ruth E. Baker[1]⋆, **Christian A. Yates**[1], **Radek Erban**[2]

[1] Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK
[2] Oxford Centre for Collaborative Applied Mathematics, Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK

Abstract   Cell migration and growth are essential components of the development of multicellular organisms. The role of various cues in directing cell migration is widespread, in particular, the role of signals in the environment in the control of cell motility and directional guidance. In many cases, especially in developmental biology, growth of the domain also plays a large role in the distribution of cells and, in some cases, cell or signal distribution may actually drive domain growth. There is an almost ubiquitous use of partial differential equations (PDEs) for modelling the time evolution of cellular density and environmental cues. In the last twenty years, a lot of attention has been devoted to connecting macroscopic PDEs with more detailed microscopic models of cellular motility, including models of directional sensing and signal transduction pathways. However, domain growth is largely omitted in the literature. In this paper, individual-based models describing cell movement and domain growth are studied, and correspondence with a macroscopic-level PDE describing the evolution of cell density is demonstrated. The individual-based models are formulated in terms of random walkers on a lattice. Domain growth provides an extra mathematical challenge by making the lattice size variable over time. A reaction-diffusion master equation formalism is generalised to the case of growing lattices and used in the derivation of the macroscopic PDEs.

1 Introduction

Cell migration is fundamental to the development and maintenance of multicellular organisms; during formation of the nervous system and vasculature, for tissue repair, wound healing and functioning of the immune system, to name but a few examples. Unregulated migration can lead to issues such as vascular disease, chronic inflammatory diseases, and tumour formation and metastasis, making the management of cell migration one of the most important topics in biology.

The role of environmental cues in directional guidance and motility regulation of these migrating cells is widespread. Chemical factors, such as secreted molecules, are often linked to the control of cell migration (chemotaxis and chemokinesis). Examples include: neutrophil migration to the sites of infection, crucial to the body's intrinsic immune defence system, being controlled by gradients of factors such as interleukin 8 [31]; angiogenesis, the growth of new blood vessels from pre-existing ones, being driven, during tumour formation, by gradients in angiogenic factors [45]; neuronal migration being controlled by factors such as the secreted protein Slit [46] and aggregation of the

---

⋆ Corresponding author. E-mail: ruth.baker@maths.ox.ac.uk.

cellular slime mould *Dictyostelium discoideum* in response to the chemical cAMP [47]. The motion of cells up gradients in adhesive sites or substrate-bound chemoattractants is known as haptotaxis, and plays a role, for example, in axon guidance [22]. In addition to these and other similar factors, the density and composition of the surrounding environment plays a role in the ability of a cell to migrate.

Growth is essential in the development of all organisms. Embryos change in size many-fold during morphogenesis and, on top of this, enormous amounts of cell migration and re-arrangement must take place in order to correctly specify the vertebrate body plan. For example, outgrowth of the vertebrate limb involves maintenance of a progress zone filled with rapidly proliferating cells by the apical ectodermal ridge. Members of the FGF family are involved in induction of the limb bud and Sonic hedgehog is expressed in the polarising region during cell specification [48].

In other cases, distribution of certain gene products in a developing organ may control growth of that organ. This may be seen in *Drosophila melanogaster* wing disc formation, where the graded distribution of morphogens such as Decapentaplegic control the rate of cell proliferation, and hence of disc growth [42]. Restriction of proliferation and movement of Decapentaplegic by Ultrabithorax in the *Drosophila melanogaster* haltere (small wing used for balance) disc results in a much smaller appendage [13]. Therefore the integration of domain growth into models is increasingly important if we truly wish to further our insight into the biological mechanisms underlying development.

1.1 Models for cell migration

Macroscopic mathematical models of cell migration are often formulated in terms of the density of cells, $c(\boldsymbol{x}, t)$, at time t and position $\boldsymbol{x}$. Let us assume that cells alter their movement according to one (chemical) environmental cue, $s(\boldsymbol{x}, t)$. Then it is often postulated that $c(\boldsymbol{x}, t)$ evolves according to the classical chemotaxis (Keller-Segel) equation [28,32,41]

$$\frac{\partial c}{\partial t} = \nabla \cdot (D\nabla c - \chi c \nabla s), \tag{1}$$

where D is the diffusion constant and χ is the chemotactic sensitivity that can depend on c, s, $\boldsymbol{x}$ and t. The minus sign on the second term indicates that the chemical is a chemoattractant, and so cells move up gradients in chemical concentration. The classical chemotaxis equation (1) has been used in the macroscopic modelling of a wide range of systems, including the bacterium *Escherichia coli* [7,29] or the cellular slime mould *Dictyostelium discoideum* [27,35], although their microscopic (individual-based) behaviour differs significantly.

Some cells, including certain types of bacteria, use the "run and tumble" strategy: a cell runs in a chosen direction, then stops for a brief while, reorients itself (tumbling state) and continues running in a new direction [4,5]. The time spent travelling in the new direction is chosen according to information that the cell aquires about the signal along its trajectory. On the other hand, *Dictyostelium discoideum* has receptors in the cell membrane and is large enough to detect typical differences in signal concentration over its body length [40]. In either case, one can argue that the behaviour of individual cells should determine the collective properties of cellular populations; that is the coefficients D and χ of the classical chemotaxis equations should be derivable from the individual-level models.

The mathematical derivation of the diffusion coefficient, D, is relatively simple for many microscopic models [5,33], although it is more challenging when cells are described as objects of a finite size, as in the case of the cellular Potts model [1,44]. For some models D depends on the concentration, c [37], or it is even replaced by a more general (anisotropic) diffusion tensor, obtained by the diffusion approximation of the transport equation that describes the underlying random walk model [24,34]. A similar analysis can be applied to random walks biased according to the gradient of an external signal to derive the expression for the chemotactic sensitivity, χ [8, 34]. The derivation of χ becomes more challenging, however, when the details of the intracellular biochemical processes are taken into account. For example, flagellated bacteria, such as *Escherichia coli*, detect chemical signals by receptors on their surface and process this information through a signal transduction network [3,43]. The bacterium "decides" where to go according to the output

of this network (a particular protein, CheY-P, communicates with the flagellar motors [9]). An important question is how the chemotactic sensitivity, χ, depends on the kinetic parameters of the signal transduction network and parameters of models of flagellar motors and receptors. This question has been answered partially for a simplified model of signal transduction in [17, 18] where the formula for chemotactic sensitivity, in terms of excitation and adaptation times of the signal transduction model, is given. Its derivation is based on analysis of the transport equation describing velocity jump processes with internal variables.

This mathematical formalism is useful even for larger cells like *Dictyostelium discoideum*, for example, for explaining the so called back-of-the-wave paradox [14]. However, the derivation of macroscopic equations is more challenging because the intracellular models have to take into account the finite size of the cell [30, 40]. A macroscopic description is derived in [19] but only for a simplified model of intracellular dynamics that is similar to the eukaryotic sensing model [30]. It is worth noting that the resulting macroscopic equations are not given in the form of the classical chemotaxis equation (1), but as a system of moment equations. In fact, for some individual-based models, it might be too ambitious to expect that the macroscopic description can be written in terms of a relatively simple equation, like equation (1)—see, for example, the discussion in [16].

1.2 Models for domain growth

Considering conservation of matter arguments, the classical chemotaxis equation (1) modifies to the following PDE for the density of individuals, $c(\boldsymbol{x}, t)$, at time t and position $\boldsymbol{x}$ on the growing domain $\Omega(t)$ [10–12]:

$$\frac{\partial c}{\partial t} + \nabla \cdot (\boldsymbol{u}c) = \nabla \cdot (D\nabla c - \chi c \nabla s), \qquad (\boldsymbol{x}, t) \in \Omega(t) \times [0, \infty), \qquad (2)$$

where $\boldsymbol{u} = \mathrm{d}\boldsymbol{x}/\mathrm{d}t$ is the velocity field due to domain growth. The derivation of this equation (and a more complex version of it) is given in Section 3.1 and Appendix B.1. Considering a stochastic individual-based model of cell motility on a growing domain, the fundamental questions stay the same. Does the cell density evolve according to equation (2), or is the macroscopic model more complicated? How do the coefficients of equation (2) depend on the parameters of the individual-level model?

1.3 Outline

In this paper we will focus on lattice-based random walk models of cellular migration. In the case of non-growing domains, our approach is equivalent to the work of Othmer and Stevens [36]. They use a continuous-time, discrete space random walk approach to investigate a number of ways that cells may sense their environment, including local, barrier and average methods. They derive several types of PDEs in the limit that the jump size tends to zero, and show that a range of behaviours, such as aggregation, blow-up or collapse, may arise from these models depending on the dynamics. In Section 2, we summarize the results obtained for the fixed domain case. We also introduce the reaction-diffusion master equation (RDME) formalism used for the analysis of random walks in this paper. Although lattice-based models are often used for getting insights into cellular motility and its effects on the collective behaviour of cells [26, 36, 37, 39], domain growth has not been included in these studies. In this paper, we aim to fill this hole in the literature. In Section 3, we study the simplest possible case—domain growth at a constant rate. We start by summarizing the macroscopic theory in Section 3.1. Then, in Section 3.2, we introduce random walks on a growing lattice and the corresponding set of RDMEs suitable for analysis of this model. Working in a one-dimensional setting, we show that the macroscopic behaviour is described by a simplified form of equation (2). In Section 4 a more general form of domain growth is studied. We couple the models from Section 2 (directional sensing in fixed domains) and Section 3 (domain growth modelling) and show that the macroscopic behaviour of the combined model is given by (2). We provide explicit formulae showing how the coefficients $\boldsymbol{u}$, χ and D depend on the parameters of

the individual-based model. Finally, in Section 5 we apply our results for domain growth to models where cell motility depends on environmental signals, highlighting the successes and short-comings of our approach for different growth and motility functions, and in Section 6 we conclude with a short discussion.

2 Modelling cell migration in fixed domains

In all models explored in this paper, we consider a population of motile cells contained in a one-dimensional domain (we may assume periodic boundary conditions, or some similar approximation for the other dimensions, which we are then able to neglect from our studies). In this section we assume that the cells may not enter or leave the domain and that (unless explicitly stated) there is no cell proliferation or death on the time scale of interest. It is relatively straightforward to include these processes, and we refer interested readers to Appendix A. Movement of the cells is regulated by signals from the external environment which may bias both the rate at which cells move and the direction in which they move. We approach the problem on an individual cell level: that is, considering each cell as a discrete entity that moves according to probabilistic rules. In this section, we study cell motility in a fixed (time-independent) domain $[0, L_0]$ which we divide into k boxes of length $\Delta x = L_0/k$. The number of cells in box i is denoted by C_i, such that the evolution of cell density over time can be represented by the vector $\boldsymbol{C}(t) = [C_1(t), \ldots, C_k(t)]$.

To simplify matters we assume that cells are restricted to move on a grid with grid points situated at the mid-points of each box. In this manner box i is defined as the interval $[(i-1)\Delta x, i\Delta x]$ and a cell in box i sits at $x_i = (2i-1)\Delta x/2$. We denote the transition rates per unit time of a cell moving left or right from box i as $T_i^{\pm}$ where each $T_i^{\pm}$ may be dependent on an external signal, $s(x)$, and time, t, i.e. $T_i^{\pm} \equiv T_i^{\pm}(s, t)$.

In order to construct a RDME describing the evolution of $\boldsymbol{C}$ we let $P(\boldsymbol{n}, t)$ be the joint probability that $C_i = n_i$ at time t for $i = 1, 2, \ldots, k$ where $\boldsymbol{n} = [n_1, \ldots, n_k]$. We define the operators $J_i^+ : \mathbb{R}^k \to \mathbb{R}^k$, for $i = 1, 2, \ldots, k-1$, and $J_i^- : \mathbb{R}^k \to \mathbb{R}^k$, for $i = 2, \ldots, k$, by

$$J_i^+ : [n_1, \ldots, n_i, \ldots, n_k] \to [n_1, \ldots, n_{i-2}, n_{i-1}, n_i + 1, n_{i+1} - 1, n_{i+2}, \ldots, n_k], \tag{3}$$

$$J_i^- : [n_1, \ldots, n_i, \ldots, n_k] \to [n_1, \ldots, n_{i-2}, n_{i-1} - 1, n_i + 1, n_{i+1}, n_{i+2}, \ldots, n_k]. \tag{4}$$

Put simply, J_i^+ adds a cell to box i and removes a cell from box $(i+1)$, whilst J_i^- adds a cell to box i and removes a cell from box $(i-1)$. Then the time evolution of $P(\boldsymbol{n}, t)$ is described by the following RDME:

$$\frac{\partial P(\boldsymbol{n}, t)}{\partial t} = \sum_{i=1}^{k-1} T_i^+ \left\{ (n_i + 1)P(J_i^+ \boldsymbol{n}, t) - n_i P(\boldsymbol{n}, t) \right\}$$

$$+ \sum_{i=2}^{k} T_i^- \left\{ (n_i + 1)P(J_i^- \boldsymbol{n}, t) - n_i P(\boldsymbol{n}, t) \right\}. \tag{5}$$

The first term on the right-hand side represents all the possible right-hand movements, so that $T_i^+(n_i + 1)P(J_i^+ \boldsymbol{n}, t)\Delta t$ represents the chance of a cell moving right from box i to box $(i+1)$ in time Δt multiplied by the probability of the system being in state $J_i^+ \boldsymbol{n}$. Similarly, the second term on the right-hand side describes the possible left-hand movements. We define

$$\boldsymbol{M}(t) = [M_1(t), \ldots, M_k(t)] = \sum_{\boldsymbol{n}} \boldsymbol{n} P(\boldsymbol{n}, t) \equiv \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \ldots \sum_{n_k=0}^{\infty} \boldsymbol{n} P(\boldsymbol{n}, t), \tag{6}$$

that is, $M_i(t)$ is the average number of cells in the ith box at time t. Multiplying (5) by n_i, summing over $\boldsymbol{n}$ and using the fact that the $T_i^{\pm}$ are independent of $\boldsymbol{n}$ [15], it can be shown that $\boldsymbol{M}(t)$ evolves according to the system of equations

$$\frac{\partial M_1}{\partial t} = T_2^- M_2 - T_1^+ M_1, \tag{7}$$
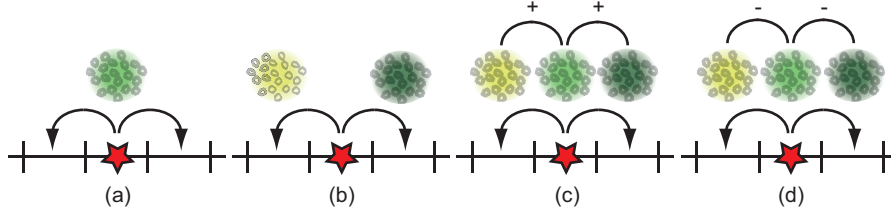
Fig. 1 Illustration of the different sensing mechanisms: (a) local sensing; (b) non-local sensing; (c) average sensing; (d) difference sensing. See the text for more details of the individual movement rules associated with each mechanism.

$$\frac{\partial M_i}{\partial t} = T_{i-1}^+ M_{i-1} - (T_i^+ + T_i^-)M_i + T_{i+1}^- M_{i+1}, \tag{8}$$

$$\frac{\partial M_k}{\partial t} = T_{k-1}^+ M_{k-1} - T_k^- M_k, \tag{9}$$

where equation (8) holds for $i = 2, \ldots, k-1$. Equation (8) is the master equation for a random walk on a lattice studied in [26,36,37,39]. Notice that we have derived equation (8) from the RDME under the assumption that the transition rates, $T_i^\pm$, are independent of n_i. If the $T_i^\pm$ do depend on the n_i then the RDME is not equivalent to equation (8) and this complicates the analysis, as we shall see later in this paper (Appendix A and later sections about growing domains).

2.1 Types of cell sensing

There are various scenarios for how a cell can sense its environment [6,30,40]. Following [36], we suppose that cells have the ability to sense the concentration of a signalling molecule at their current grid point and at grid points either side, and that the transition rates, $T_i^\pm$, of moving left or right from grid point i are dictated by the signalling molecule density at these points:

$$T_i^\pm = d_l s_i + d_n s_{i\pm 1}, \tag{10}$$

where s_i represents signalling molecule density at grid point i and d_l and d_n are constants (dependent on Δx) representing local and non-local contributions to the transition rates. Four cases are illustrated in Figure 1: local; non-local; average; difference.

Local sensing. Cells sense the concentration of signalling molecule only at their current site. The rate of moving right from the current site, T_i^+, is the same as that of moving left, T_i^-, so that $T_i^+ = d_l s_i = T_i^-$.

Non-local sensing. Cells sense the signalling molecule over a wider region: transition rates are purely dependent on the concentration of signalling molecule at neighbouring sites. $T_i^- = d_n s_{i-1}$ and $T_i^+ = d_n s_{i+1}$.

Average sensing. Cells sense signalling molecule concentration at both current and neighbouring sites and move according to an averaged reading: $T_i^\pm = d_l s_i + d_n s_{i\pm 1}$.

Difference sensing. Cells sense signalling molecule concentration at both current and neighbouring sites and move according to a differenced reading: $T_i^\pm = d_0 + d_l s_i - d_n s_{i\pm 1}$, where d_0 is a base transition rate required to ensure positivity.

2.2 Local sensing

Supposing that cells have the ability to locally sense their environment, which does not change over time, the transition probabilities assume the form $T_i^{\pm} = d_l s(x_i)$ and (7)-(9) becomes

$$\frac{\partial M_1}{\partial t} = d_l s(x_2) M_2 - d_l s(x_1) M_1, \tag{11}$$

$$\frac{\partial M_i}{\partial t} = d_l s(x_{i-1}) M_{i-1} - 2 d_l s(x_i) M_i + d_l s(x_{i+1}) M_{i+1}, \tag{12}$$

$$\frac{\partial M_k}{\partial t} = d_l s(x_{k-1}) M_{k-1} - d_l s(x_k) M_k, \tag{13}$$

where equation (12) holds for $i = 2, \ldots, k-1$. The steady state mean particle density can be found by solving a tri-diagonal matrix equation subject to the constraint that $\sum_{i=1}^{k} M_i = M_0$, where M_0 is the total number of cells contained in the domain. In order to draw correspondence with a population-level description of cell movement, we expand terms such as $M_{i\pm1}$ about M_i:

$$M_{i\pm1} = M(x_{i\pm1}) = M(x_i) \pm (\Delta x) \frac{\partial M}{\partial x}(x_i) + \frac{1}{2}(\Delta x)^2 \frac{\partial^2 M}{\partial x^2}(x_i) + \ldots, \tag{14}$$

where Δx is the distance between spatial grid points. Substituting into equations (11)-(13) gives

$$\frac{\partial M}{\partial t}(x_i, t) = d_l (\Delta x)^2 \frac{\partial^2 (sM)}{\partial x^2}(x_i, t) + o((\Delta x)^2), \tag{15}$$

which, upon letting $\Delta x \to 0$ such that $\lim_{\Delta x \to 0} d_l (\Delta x)^2 = D_l$, gives the corresponding PDE description for cell density, $c(x, t)$:

$$\frac{\partial c}{\partial t} = D_l \frac{\partial}{\partial x} \left[s \frac{\partial c}{\partial x} + c \frac{\partial s}{\partial x} \right], \qquad (x, t) \in (0, L_0) \times [0, \infty), \tag{16}$$

with zero flux boundary conditions, $\partial(sc)/\partial x = 0$ for $x = 0, L_0$. The right-hand side of equation (16) may be viewed as a combination of a spatially dependent diffusion term and a spatially varying advection term; so long as s varies over the domain, purely local measurements induce a directional response and the steady state is no longer spatially homogeneous:

$$c_{st}(x) = \frac{A}{s(x)}, \tag{17}$$

where A is a constant determined by the initial conditions. Figure 2 shows the results of numerical simulation of a local sensing mechanism using an exponentially distributed external signal, $s(x) = d \exp(-ax)$. As expected, we see excellent agreement between the stochastic average and continuum solutions.
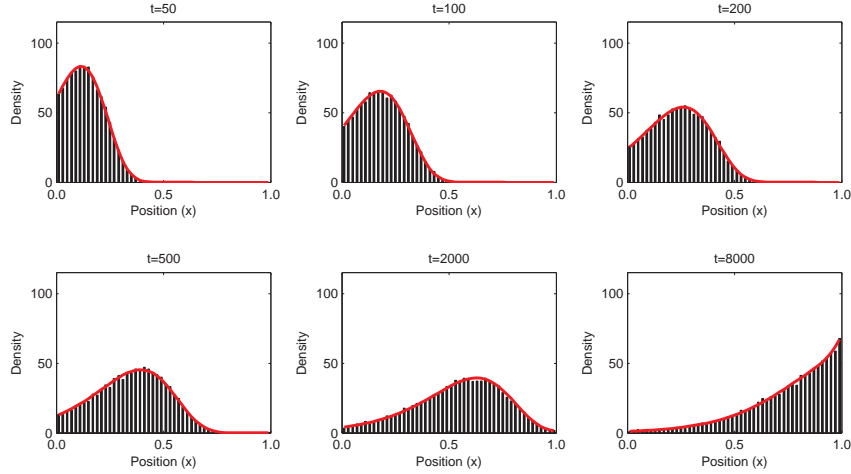
2.3 Non-local sensing

Suppose now that cells sense their environment in a non-local manner: the transition rates assume the form $T_i^{\pm} = d_n s(x_{i\pm1})$. Substituting into equations (7)-(9) and taking the limit as $\Delta x \to 0$, such that $\lim_{\Delta x \to 0} d_n (\Delta x)^2 = D_n$, gives

$$\frac{\partial c}{\partial t} = D_n \frac{\partial}{\partial x} \left[s \frac{\partial c}{\partial x} - c \frac{\partial s}{\partial x} \right], \qquad (x, t) \in (0, L_0) \times [0, \infty), \tag{18}$$

with zero flux boundary conditions. As with local sensing, the right-hand side of equation (18) may be viewed as a combination of a spatially dependent diffusion term and some kind of spatially varying advection term, but we note that in this case the taxis occurs up gradients in signalling molecule concentration. The steady state takes the form

$$c_{st}(x) = B s(x), \tag{19}$$

where B is determined by the initial conditions.

Fig. 2 Cells moving according to a local sensing mechanism. The histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines show the result of numerical simulation of equation (16) for $s(x) = \exp(-ax)$ using the finite difference approximation outlined in Appendix C.1. All 1000 particles are released in box 1 ($x = 0.01$) and the boundaries are assumed to be reflecting. Parameters are as follows: $k = 50$, $d_l = 1.0$, $a = 4.0$, $L_0 = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

2.4 Average sensing

Equations for the stochastic means and macroscopic PDEs can be derived from equations (7)-(9), as in the previous sections. For average sensing the limiting PDE is

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left[\{D_l s + D_n s\}\frac{\partial c}{\partial x} + c\frac{\partial}{\partial x}\{D_l s - D_n s\}\right]. \tag{20}$$

Importantly, we note that in this case if $d_l = d_n$ the advection term is lost for average sensing and we simply have spatially varying diffusion:

$$\frac{\partial c}{\partial t} = D\frac{\partial}{\partial x}\left[s\frac{\partial c}{\partial x}\right], \tag{21}$$

where $D = D_l + D_n$. Zero flux boundary conditions result in a homogeneous steady state.

The biological implication of this is that a purely motility regulating effect, and hence a homogeneous steady state density in the presence of a signalling molecule, can only be achieved when cells base their movements on evenly weighted measurements of the signal concentration.

2.5 Other considerations

Firstly, we note that signal transduction may amplify the signal, making the shape of the transition rate profile different from that of the signalling molecule concentration profile. This could be represented by letting the transition rates take the following form $T_i^{\pm} = d_l f(s_i) + d_n f(s_{i\pm1})$ [36]. Such considerations lead to a trivial extension of the results presented in this work—therefore we do not consider amplification any further here.

Secondly, we note that we may renormalise the transition rates. As suggested by Othmer and Stevens [36], the transition rates can be normalised so that their sum is one, point-wise. Further, we can separate the effects of "knowing when to jump" and "where to jump". Suppose that we

wish the mean waiting time to leave a site not to depend on $s(x)$. Then it must be constant across the lattice and the transition rates redefined as

$$N_i^\pm = \frac{T_i^\pm}{T_i^+ + T_i^-}, \tag{22}$$

where the $N_i^\pm$ are the normalised transition rates. This may introduce longer-range dependencies into the transition rates. For example, the average case would give

$$N_i^\pm = \frac{d_l s_i + d_n s_{i\pm1}}{d_n s_{i-1} + 2d_l s_i + d_n s_{i+1}}, \tag{23}$$

and the difference case

$$N_i^\pm = \frac{d_0 + d_n s_{i\pm1} - d_l s_i}{d_n s_{i-1} + 2d_0 - 2d_l s_i + d_n s_{i+1}}. \tag{24}$$

These formulations have the added advantage that they tend to stop transition rates becoming large and hence cells from moving too quickly. However, for simplicity we will not investigate this case here.

It is also possible to include the effects of cell flux and decay, density-dependent transition probabilities and stochastic modelling of the signal. The results are a fairly straightforward extension of those presented here, and we refer the readers to Appendix A for more details.

3 The inclusion of constant domain growth

We now consider the case in which the domain grows uniformly over time at constant rate r. We first derive a population-level description for the system using a PDE approach and then move to the individual-level description. Throughout this section we consider constant transition rates for cell movement of the form $T_i^\pm = d$, i.e. diffusion but no taxis.

3.1 Population-level behaviour with constant domain growth

In deriving the equation for cell density, $c(x,t)$, at time t and position x on the growing domain $\Omega(t)$ we follow Crampin [10] by considering conservation of matter arguments in a generalised approach for a single chemical species. We consider the net rate of production, the reaction kinetics and flux of the species in and through an elemental volume, $v(t)$, which moves with the flow caused by domain growth. Conservation arguments imply

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{v(t)} c(x,t)\,\mathrm{d}x = \int_{\partial v(t)} -j(x,t) \cdot \mathrm{d}s + \int_{v(t)} R(c(x,t))\,\mathrm{d}x, \tag{25}$$

where the time varying element, $v(t)$, is arbitrary. The flux, $j(x,t)$, is through the closed surface $\partial v(t)$ and $R(c(x,t))$ represents any proliferation/decay taking place.

Using the Divergence Theorem we can re-write equation (25) as

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{v(t)} c\,\mathrm{d}x = \int_{v(t)} [-\nabla \cdot j + R(c)]\,\mathrm{d}x. \tag{26}$$

We apply the Reynolds Transport Theorem to the left-hand side:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{v(t)} c\,\mathrm{d}x = \int_{v(t)} \left[\frac{\partial c}{\partial t} + \nabla \cdot (uc)\right]\mathrm{d}x, \tag{27}$$

where $u = \mathrm{d}x/\mathrm{d}t$ is the velocity field due to domain growth, to arrive at

$$\int_{v(t)} \left[\frac{\partial c}{\partial t} + \nabla \cdot (uc)\right]\mathrm{d}x = \int_{v(t)} [-\nabla \cdot j + R(c)]\,\mathrm{d}x. \tag{28}$$

Applying the usual argument relating to the arbitrary choice of $v(t)$ implies that equation (28) holds everywhere on the growing domain. Therefore

$$\frac{\partial c}{\partial t} + \nabla \cdot (uc) = -\nabla \cdot j + R(c), \qquad (x,t) \in \Omega(t) \times [0,\infty). \tag{29}$$

As shown in Appendix B.2, cell conservation arguments imply that in order to be consistent with earlier simulations zero flux boundary conditions must be implemented on $\partial\Omega(t)$. The two new terms introduced due to domain growth are $u \cdot \nabla c$, which represents transport of chemical around the domain, and $c\nabla \cdot u$, which represents dilution due to local volume increase. We note, also, that all the analysis carried out above is applicable to systems of higher spatial dimension.

Here we will consider the domain, $x \in [0, L(t)]$, with cell movements due to a diffusive flux and no taxis, $j = -D\partial c/\partial x$, and no cell proliferation or decay, $R(c) = 0$. This gives

$$\frac{\partial c}{\partial t} + \frac{\partial(uc)}{\partial x} = D\frac{\partial^2 c}{\partial x^2}, \qquad (x,t) \in (0, L(t)) \times [0,\infty), \tag{30}$$

and $u(x,t) = \mathrm{d}x/\mathrm{d}t$.

For the case of constant growth at rate r and an initial domain of length unity, we have $L(t) = e^{rt}$ and $u = rx$ so that

$$\frac{\partial c}{\partial t} + rx\frac{\partial c}{\partial x} = D\frac{\partial^2 c}{\partial x^2} - rc, \qquad (x,t) \in (0, e^{rt}) \times [0,\infty), \tag{31}$$

with $\partial c/\partial x = 0$ for $x = 0$ and $x = e^{rt}$.

3.2 Individual-level behaviour with constant domain growth

We return to considering the RDME approach to describe individual-level behaviour with the aim of comparison with equation (31). In the first instance, we consider simple diffusion with domain growth such that each box divides at rate r per unit time. Cells in the dividing box are split between the two new boxes according to some discrete probability distribution, π. All boxes to the right of the dividing box have their index increased by one to "make room" for the newly created box.

In order to write down a RDME describing growth we must introduce further operators: we define "inverse" growth operators $G_i : \mathbb{R}^k \to \mathbb{R}^{k-1}$, $i = 1, 2, \ldots, k-1$, by

$$G_i : [n_1, \ldots, n_{i-1}, n_i, n_{i+1}, \ldots, n_k] \to [n_1, \ldots, n_{i-1}, n_i + n_{i+1}, \ldots, n_k], \tag{32}$$

so that G_i joins boxes i and $(i+1)$. Ignoring diffusion for the moment, the RDME can be written

$$\frac{\partial P^k(\boldsymbol{n},t)}{\partial t} = r\sum_{i=1}^{k-1} \pi(n_i, n_{i+1}|n_i + n_{i+1}) \times P^{k-1}(G_i\boldsymbol{n},t) - r\sum_{i=1}^{k} P^k(\boldsymbol{n},t), \tag{33}$$

where the superscript on the joint probability density function explicitly indicates the number of boxes at time t. $\pi(n_i, n_{i+1}|n_i + n_{i+1})$ indicates the probability that n_i cells ends up in box i and n_{i+1} in box $(i+1)$ after splitting of the ith box (which before splitting contained $n_i + n_{i+1}$ cells). r now corresponds to the (constant) rate of box splitting.

Symmetry Assumption (S): We make the simple assumption that the probability density function π is symmetric about its mid-point, so that

$$\pi(n_i, n_{i+1}|n_i + n_{i+1}) = \pi(n_{i+1}, n_i|n_i + n_{i+1}). \tag{34}$$

This assumption states that if there are n cells in a splitting box, the division of cells into the new boxes is just as likely to be a and $n - a$ right and left, respectively, as it is $n - a$ and a, for each $a = 0, 1, \ldots, n$. In mathematical terms, we anticipate that this assumption will ensure no artificial drift is introduced into the model due to box splitting.

We note also that π satisfies the normalization condition:

$$\sum_{n=0}^{n_i+n_{i+1}} \pi(n, n_i + n_{i+1} - n | n_i + n_{i+1}) = 1. \tag{35}$$

In this work we will consider two types of distribution: firstly, a binomial-type splitting approach where

$$\pi(n_i, n_{i+1} | n_i + n_{i+1}) = \frac{1}{2^{(n_i+n_{i+1})}} \binom{n_i + n_{i+1}}{n_i}, \tag{36}$$

and, secondly, an even-splitting approach:

$$\pi(n_i, n_{i+1} | n_i + n_{i+1}) = \begin{cases} 1 & \text{for} \quad n_i = \lfloor (n_i + n_{i+1})/2 \rfloor, \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

Other types of splitting, such as uniform splitting of the form $\pi(n_i, n_{i+1} | n_i + n_{i+1}) = (n_i + n_{i+1} + 1)^{-1}$, may also be used, but we do not discuss this further here. The main reason being that so long as diffusion/migration is the dominant process we do not expect the specific box-splitting algorithm to have much affect on the model results.

We define the stochastic means in the same way as before:

$$\boldsymbol{M}^k(t) = [M_1^k(t), \dots, M_k^k(t)] = \sum_{\boldsymbol{n}} \boldsymbol{n} P^k(\boldsymbol{n}, t), \tag{38}$$

to give, away from the end boxes,

$$\frac{\partial M_j^k}{\partial t} = r \sum_{i=1}^{k-1} \sum_{\boldsymbol{n}} n_j \pi(n_i, n_{i+1} | n_i + n_{i+1}) P^{k-1}(G_i \boldsymbol{n}, t) - r \sum_{i=1}^{k} M_j^k(t). \tag{39}$$

For $i \neq j, (j-1)$ we may evaluate terms of the form

$$\sum_{\boldsymbol{n}} n_j \pi(n_i, n_{i+1} | n_i + n_{i+1}) P^{k-1}(G_i \boldsymbol{n}, t), \tag{40}$$

by combining indices in the sum, so that the above becomes

$$\sum_{G_i \boldsymbol{n}} \left[\sum_{\bar{n}_i=0}^{n_i+n_{i+1}} \pi(\bar{n}_i, n_i + n_{i+1} - \bar{n}_i | n_i + n_{i+1}) \right] n_j P^{k-1}(G_i \boldsymbol{n}, t). \tag{41}$$

However, the term in square brackets is simply the sum over all the possible ways of splitting $n_i + n_{i+1}$ into two parts and therefore equals unity. This leaves us with

$$\sum_{G_i \boldsymbol{n}} n_j P^{k-1}(G_i \boldsymbol{n}, t) = \begin{cases} M_{j-1}^{k-1} & \text{for } i < j - 1, \\ \\ M_j^{k-1} & \text{for } i > j. \end{cases} \tag{42}$$

We approach terms involving $i = j, (j-1)$ in a similar manner and utilise assumption **(S)**. By swapping indices, we can consider

$$\sum_{\boldsymbol{n}} \pi(n_j, n_{j+1} | n_j + n_{j+1}) n_j P^{k-1}(G_j \boldsymbol{n}, t), \tag{43}$$

as being identical to

$$\sum_{\boldsymbol{n}} \pi(n_j, n_{j+1} | n_j + n_{j+1}) n_{j+1} P^{k-1}(G_j \boldsymbol{n}, t). \tag{44}$$

Adding the two together, and evaluating the result gives

$$\sum_{\boldsymbol{n}} \pi(n_j, n_{j+1}|n_j + n_{j+1})(n_j + n_{j+1})P^{k-1}(G_j\boldsymbol{n}, t)$$

$$= \sum_{G_j\boldsymbol{n}} \left[\sum_{\bar{n}_j=0}^{n_j+n_{j+1}} \pi(\bar{n}_j, n_j + n_{j+1} - \bar{n}_j|n_j + n_{j+1})\right] (n_j + n_{j+1})P^{k-1}(G_j\boldsymbol{n}, t),$$

$$= M_j^{k-1}, \tag{45}$$

as the term in square brackets is again equal to unity. Therefore

$$\sum_{\boldsymbol{n}} \pi(n_j, n_{j+1}|n_j + n_{j+1})n_j P^{k-1}(G_j\boldsymbol{n}, t) = \frac{1}{2}M_j^{k-1}, \tag{46}$$

and an almost identical argument can be used to show that

$$\sum_{\boldsymbol{n}} \pi(n_{j-1}, n_j|n_{j-1} + n_j)n_j P^{k-1}(G_{j-1}\boldsymbol{n}, t) = \frac{1}{2}M_{j-1}^{k-1}. \tag{47}$$

Combining these results gives

$$\frac{\partial M_j^k}{\partial t} = r(j - 3/2)M_{j-1}^{k-1} + r(k - j - 1/2)M_j^{k-1} - rkM_j^k. \tag{48}$$

In order to derive correspondence with a PDE, as in earlier sections of this work, we make a heuristic approximation to relate terms involving $(k - 1)$ boxes with those involving k boxes:

$$\frac{(k - 1)}{k}M_j^{k-1} \approx M_j^k. \tag{49}$$

That is, supposing that diffusion occurs on a faster timescale than growth, the cell density tends to "dilute" in the natural way.

We then use the same Taylor series expansions about M_j^{k-1} to give

$$\frac{(k - 1)}{k}\frac{\partial M^{k-1}}{\partial t}(x_j, t) = r(j - 3/2)\left[M^{k-1}(x_j, t) - \Delta x\frac{\partial M^{k-1}}{\partial x}(x_j, t) + \ldots\right]$$

$$+ r(k - j - 1/2)M^{k-1}(x_j, t) - r(k - 1)M^{k-1}(x_j, t),$$

$$= -r\left[x_j\frac{\partial M^{k-1}}{\partial x}(x_j, t) - M^{k-1}(x_j, t)\right] + o(\Delta x). \tag{50}$$

Taking the limit as $\Delta x = 1/k(0) \to 0$, relabelling and bringing diffusion back in gives the familiar PDE

$$\frac{\partial M^k}{\partial t} + rx\frac{\partial M^k}{\partial x} = D\frac{\partial^2 M^k}{\partial x^2} - rM^k, \tag{51}$$

where the superscript $k = k(t)$ indicates the domain length at time t.

3.2.1 Domain length. $P_k(t)$, the probability of having k boxes at time t, satisfies the master equation

$$\frac{\mathrm{d}P_k}{\mathrm{d}t} = (k - 1)rP_{k-1} - krP_k, \quad \text{for} \quad k = k_0, k_0 + 1, \ldots, \tag{52}$$

with initial condition

$$P_k(0) = \begin{cases} 1 & k = k_0, \\ 0 & k \neq k_0, \end{cases} \tag{53}$$

where k_0 is the initial number of boxes.

This infinite set of ODEs may be solved inductively to give $P_k(t) \equiv 0$ for $k < k_0$ and

$$P_k(t) = \binom{k-1}{k_0-1} \left(1 - e^{-rt}\right)^{k-k_0} e^{-k_0 rt}, \tag{54}$$

for $k \geq k_0$.

The mean number of boxes is then given by

$$K(t) = \sum_{k=k_0}^{\infty} k P_k(t) = \sum_{k=k_0}^{\infty} k \binom{k-1}{k_0-1} \left(1 - e^{-rt}\right)^{k-k_0} e^{-k_0 rt} = k_0 e^{rt}, \tag{55}$$

i.e. the mean domain length grows exponentially with time.

In summary, our analysis shows that the cell density approximately satisfies the PDE

$$\frac{\partial c}{\partial t} + rx\frac{\partial c}{\partial x} = D\frac{\partial^2 c}{\partial x^2} - rc, \qquad (x,t) \in (0, e^{rt}) \times [0, \infty), \tag{56}$$

assuming that $L_0 = k_0 \Delta x = 1$. This is identical to the PDE formulation of the problem—see equation (31)—and our analysis holds for any symmetric π satisfying equation (34).

3.3 Numerical results

Figure 3 shows the results of numerical simulation of the system for constant domain growth at rate $r = 0.001$ under Symmetry Assumption **(S)** using the binomial box-splitting approach of equation (36). The first four plots show bar charts of the average of 20 stochastic realisations of the system along with numerical solution of the corresponding PDE using the method outlined in Appendix C. We see good agreement between the two. The fifth plot shows the dynamics of the system over a longer time period with cell density indicated by the shading tone and the positions of 10 initially equally spaced boxes over time indicated by coloured lines. Throughout this section, if one of the tracked boxes splits we follow the left-hand most box for subsequent times. The sixth plot compares the evolution domain length for each of the stochastic realisations with the average length and that predicted by the corresponding population-level equation. We see excellent agreement between the mean and the continuum equivalent. In Appendix D we justify the moment closure approximation, equation (49), made in deriving our PDE formulation by calculating the accuracy of our assumption over the 20 realisations presented in Figure 3: the approximation is satisfied to a good degree throughout.

We note here, that as the domains grow in each of the stochastic realisations, some domains will be shorter and some longer than that predicted by the population-level model. This is a natural consequence of the individual-level formulation and results in parts of the histograms of Figure 3 extending beyond the reach of the deterministic result. This phenomenon is seen to some extent throughout the work presented here.

## 4 Including more general types of domain growth

We now turn our attention to more general forms of domain growth, with the hope that we may find a similar correspondence between the PDE and RDME formulation for the problem. In line with previous simulations we consider simple transitions probabilities, $T_i^{\pm} = d$, and cell conservation on the domain. The local domain growth will be taken to be $f(x, t, c(x, t))$ *i.e.* some function of x, t and cell density.

Derivation of the corresponding continuum equation is given in detail in Appendix B.1 and results in a PDE description of the form

$$\frac{\partial c}{\partial t} + \frac{\partial}{\partial x}\left(c\int_0^x f(\bar{x}, t, c(\bar{x}, t))\mathrm{d}\bar{x}\right) = D\frac{\partial^2 c}{\partial x^2}, \quad (x,t) \in (0, L(t)) \times [0, \infty), \tag{57}$$

with zero flux boundary conditions at $x = 0$ and $x = L(t)$ (see Appendix B.2 for justification of the choice of boundary conditions).

Note that in Section 3 we studied $f(x, t, c(x, t)) = r$. In this case equation (57) is equivalent to equation (31) and equation (127) implies $\mathrm{d}L(t)/\mathrm{d}t = rL$, *i.e.* $L(t) = e^{rt}$.
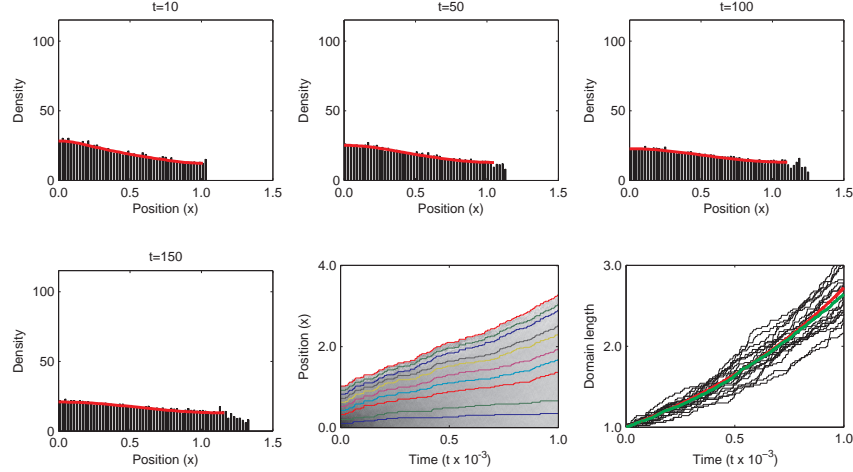
Fig. 3 Cells diffusing at a constant rate with constant growth. This histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines show the result of numerical simulation of the corresponding PDE using the finite difference algorithm outlined in Appendix C.2. The boundaries are assumed to be reflecting and the initial cell distribution consists of 1000 cells approximately distributed as $c(x,0) = 1000e^{-x}/\sum_{i=1}^{k} e^{-x_i}$. The fifth plot shows the results of one stochastic realisation over a longer time frame. The coloured lines indicate the trajectories of 10 boxes and the shading the average cell density (black high, white low). The sixth plot shows the evolution of domain length over time for each of the 20 realisations (black lines), the average stochastic length (green line) and the length predicted by the corresponding continuum system (red line). Parameters are as follows: $k_0 = 50$, $d = 1.0$, $r = 0.001$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

4.1 Individual-level behaviour for general domain growth

We work with exactly the same set up as before, only this time we assume that the box splitting rate may be dependent upon cell density. We neglect the possibility that the splitting rate may also be time- and/or space-dependent, as these may easily be explored using the arguments laid out for constant domain growth, and we present some examples in Section 5.

We suppose that the rate of box splitting is some function, f, of the number of cells in that box. As before, we neglect diffusion from our calculations, and write down the RDME:

$$\frac{\partial P^k(\boldsymbol{n},t)}{\partial t} = \sum_{i=1}^{k-1} f(n_i + n_{i+1})\pi(n_i, n_{i+1}|n_i + n_{i+1})P^{k-1}(G_i\boldsymbol{n},t) - \sum_{i=1}^{k} f(n_i)P^k(\boldsymbol{n},t), \quad (58)$$

where, once again, the superscript on the joint probability density function explicitly indicates the number of boxes at time t and π is the probability density function describing the distribution of cells upon box splitting.

We define the stochastic means by equation (38) to give, away from the end boxes,

$$\frac{\partial M_j^k}{\partial t} = \sum_{i=1}^{k-1} \sum_{\boldsymbol{n}} n_j f(n_i + n_{i+1})\pi(n_i, n_{i+1}|n_i + n_{i+1})P^{k-1}(G_i\boldsymbol{n},t)$$
$$- \sum_{i=1}^{k} \sum_{\boldsymbol{n}} n_j f(n_i)P^k(\boldsymbol{n},t). \quad (59)$$

As before, for $i \neq j, (j-1)$ we may evaluate terms of the form

$$\sum_{\boldsymbol{n}} \pi(n_i, n_{i+1}|n_i + n_{i+1})n_j P^{k-1}(G_i\boldsymbol{n},t), \quad (60)$$

by combining indices in the sum to give

$$\sum_{G_i \boldsymbol{n}} n_j f(n_i + n_{i+1}) P^{k-1}(G_i \boldsymbol{n}, t) = \begin{cases} \langle n_{j-1}^{k-1} f(n_i^{k-1}) \rangle & \text{for } i < j-1, \\[2ex] \langle n_j^{k-1} f(n_i^{k-1}) \rangle & \text{for } i > j, \end{cases} \tag{61}$$

where $\langle \cdot \rangle$ is the moment defined in the usual way.

We approach terms involving $i = j, (j-1)$ as before. Making, once again, the simple assumption that the probability density function π satisfies Symmetry Assumption (**S**) allows us to write

$$\sum_{G_j \boldsymbol{n}} n_j f(n_j + n_{j+1}) \pi(n_j, n_{j+1} | n_j + n_{j+1}) P^{k-1}(G_j \boldsymbol{n}, t) = \frac{1}{2} \langle n_j^{k-1} f(n_j^{k-1}) \rangle, \tag{62}$$

and

$$\sum_{G_{j-1} \boldsymbol{n}} n_{j-1} f(n_{j-1} + n_j) \pi(n_{j-1}, n_j | n_{j-1} + n_j) P^{k-1}(G_{j-1} \boldsymbol{n}, t)$$
$$= \frac{1}{2} \langle n_{j-1}^{k-1} f(n_{j-1}^{k-1}) \rangle. \tag{63}$$

Combining these results gives

$$\frac{\partial M_j^k}{\partial t} = \sum_{i=1}^{j-2} \langle n_{j-1}^{k-1} f(n_i^{k-1}) \rangle + \frac{1}{2} \langle n_{j-1}^{k-1} f(n_{j-1}^{k-1}) \rangle + \frac{1}{2} \langle n_j^{k-1} f(n_j^{k-1}) \rangle$$
$$+ \sum_{i=j+1}^{k-1} \langle n_j^{k-1} f(n_i^{k-1}) \rangle - \sum_{i=1}^{k} \langle n_j^k f(n_i^k) \rangle. \tag{64}$$

In order to derive correspondence with a PDE, we now need to make some moment closure approximations (the validity of which are discussed in Appendix D). Firstly, we assume

$$\langle n_j^k f(n_i^k) \rangle \approx \langle n_j^k \rangle f(\langle n_i^k \rangle). \tag{65}$$

We also make the assumption that

$$\sum_{i=1}^{k} \langle n_j^k \rangle f(\langle n_i^k \rangle) \approx \sum_{i=1}^{k-1} \langle n_j^{k-1} \rangle f(\langle n_i^{k-1} \rangle). \tag{66}$$

This gives, using M_j^k to denote $\langle n_j^k \rangle$,

$$\frac{(k-1)}{k} \frac{\partial M_j^{k-1}}{\partial t} = \sum_{i=1}^{j-1} M_{j-1}^{k-1} f(M_i^{k-1}) + \sum_{i=j}^{k-1} M_j^{k-1} f(M_i^{k-1}) - \sum_{i=1}^{k-1} M_j^{k-1} f(M_i^{k-1})$$
$$- \frac{1}{2} M_{j-1}^{k-1} f(M_{j-1}^{k-1}) - \frac{1}{2} M_j^{k-1} f(M_j^{k-1}). \tag{67}$$

We then use the Taylor series expansions, as before, for the means, M_{j-1}^{k-1}, in order to make the change to a continuous space coordinate, x. The resulting equation is

$$\frac{(k-1)}{k} \frac{\partial M^{k-1}}{\partial t}(x_j, t) = -\frac{\partial M^{k-1}}{\partial x}(x_j, t) \sum_{i=1}^{j} \Delta x f(M^{k-1}(x_i, t))$$
$$- M^{k-1}(x_j, t) f(M^{k-1}(x_j, t)) + o(\Delta x). \tag{68}$$

Taking the limit as $\Delta x \to 0$, recognising the summation term as an integral and bringing diffusion back in gives the PDE

$$\frac{\partial M^{k-1}}{\partial t} + M^{k-1} f(M^{k-1}) + \left(\int_0^x f(M^{k-1}(\bar{x}, t)) d\bar{x} \right) \frac{\partial M^{k-1}}{\partial x} = D \frac{\partial^2 M^{k-1}}{\partial x^2}, \tag{69}$$

where the superscript $k-1$ indicates the domain length.

4.1.1 Domain length.   In this case we are unable to explicitly calculate an expression for the domain length. We may write down the following expression for $P_k(t)$, the probability of having k boxes at time t:

$$\frac{\mathrm{d}P_k}{\mathrm{d}t} = \sum_{j=1}^{k-1} \langle f(n_j + n_{j+1}) \rangle - \sum_{j=1}^{k} \langle f(n_j) \rangle, \quad \text{for} \quad k = k_0, k_0 + 1, \ldots, \tag{70}$$

with initial condition given, as before, by equation (53). The formulation comes from summing the master equation for domain growth over $\boldsymbol{n}$ where

$$P_k(t) = \sum_{\boldsymbol{n}} P^k(\boldsymbol{n}, t). \tag{71}$$

That there is dependence on the number of cells in each box at time t prevents us from writing down an explicit expression for $P_k(t)$.

The mean number of boxes is given by (assuming we may swap integral and sum)

$$\frac{\mathrm{d}K(t)}{\mathrm{d}t} = \sum_{k=k_0}^{\infty} k \frac{\mathrm{d}P_k(t)}{\mathrm{d}t} = k_0 P_{k_0} + \sum_{k=k_0+1}^{\infty} \left[\sum_{j=1}^{k-1} \langle f(n_j + n_{j+1}) \rangle - \sum_{j=1}^{k} \langle f(n_j) \rangle \right], \tag{72}$$

which is not closed for $K(t)$. This area will be the subject of future investigation by us.

4.2 Numerical results

Figure 4 shows the results of numerical simulation of the system for domain growth of the form $f(c) = rc$ where r is a constant. As before, the first four plots show histograms of the average of 20 stochastic realisations along with numerical solution of the corresponding PDE. Good agreement between the two is seen. The fifth plot shows the domain growth over a longer time period with the trajectories of 10 initially evenly spaced boxes shown and the sixth compares the evolution of domain length for each realisation.

Finally, Figure 5 shows the results of numerical simulation of the system for $f = rc^2$, where r is a constant, in the same format as those presented in Figure 4. Once again, excellent agreement between stochastic and deterministic methods is demonstrated, although we see a noticeable decrease in the agreement between the average stochastic domain length and that predicted by the continuum model as the domain size gets large. We anticipate that this is a result of the gradual breakdown of our moment closure approximations combined with stochastic effects as the number of cells in each box becomes small.

In Appendix D we present statistics gathered from our stochastic realisations that justify the moment closure approximations made in equations (65) and (66). Figure 14 demonstrates the validity of equation (66) for both linear and quadratic growth whilst Figure 15 (16) shows the validity of equation (65) for linear (quadratic) growth. In all cases we see the approximations are clearly valid over the time scales presented.

4.2.1 Domain length for linear dependence.   In the case in which the dependence of domain growth upon density is linear in $c(x, t)$ we can explicitly find the domain length. Consider equation (127) in this case:

$$\frac{\mathrm{d}L(t)}{\mathrm{d}t} = r \int_0^{L(t)} c(x, t) \mathrm{d}x. \tag{73}$$

Assuming total cell numbers remain conserved across the growing domain, we have

$$\frac{\mathrm{d}L(t)}{\mathrm{d}t} = rC, \tag{74}$$
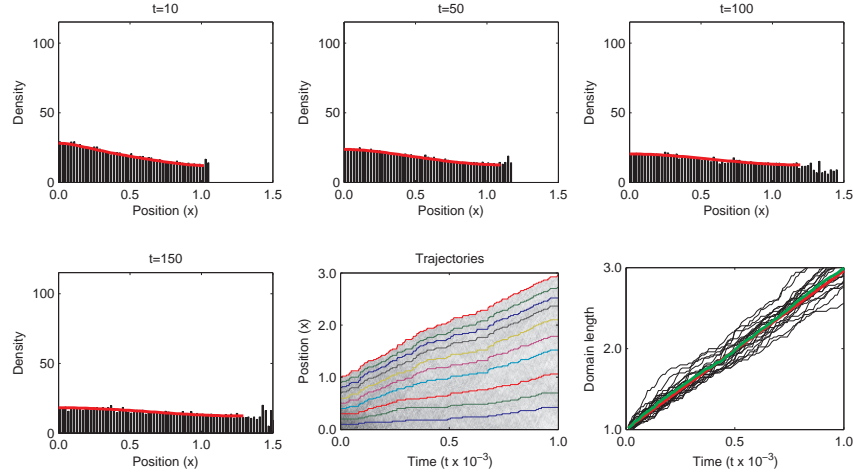
Fig. 4 Cells diffusing at a constant rate with density-dependent growth of the form $f = rc$, where r is a constant. This histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines show the result of numerical simulation of the corresponding PDE using the NAG routine D03PE as outlined in Appendix C.3. The boundaries are assumed to be reflecting and the initial cell distribution consists of 1000 cells approximately distributed as $c(x,0) = 1000e^{-x}/\sum_{i=1}^{k} e^{-x_i}$. The fifth plot shows the results from one stochastic realisation over a longer time frame. The coloured lines indicate the trajectories of 10 boxes and the shading the average cell density (black high, white low). The sixth plot shows the evolution of domain length over time for each of the 20 realisations (black lines), the average stochastic length (green line) and the length predicted by the corresponding continuum system (red line). Parameters are as follows: $k_0 = 50$, $d = 1.0$, $r = 0.0001$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

where C is a constant representing the total cell number. Hence $L(t) = L_0 + rCt$. This can be validated using the numerical simulation shown in Figure 4 with $L_0 = 1$:

$$L(t) = 1 + rt \int_0^1 c(x,0)\mathrm{d}x = 1 + \frac{1000rt}{\sum_{i=1}^{k} e^{-x_i}} \int_0^1 e^{-x}\mathrm{d}x \approx 1 + 0.00195757\,t, \qquad (75)$$

for the parameters used in the figure. This gives a domain length of $L(1000) \approx 2.9576$, consistent with the numerical solution to within four decimal places. We also note the good agreement between $L(t)$ and the average stochastic domain length in Figure 4.

We note, however, that this does not imply that $u(x,t) = \mathrm{d}x/\mathrm{d}t$ is constant. The flow still changes across the domain as we have
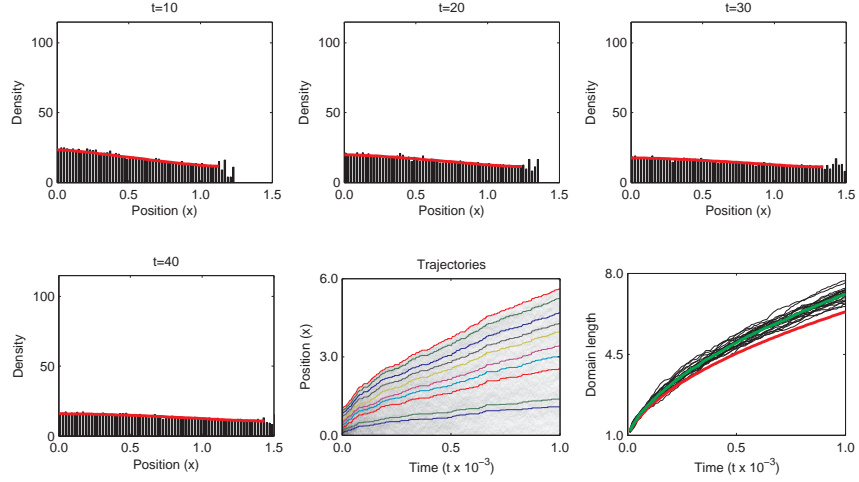
$$u(x,t) = \int_0^x c(\bar{x},t)\mathrm{d}\bar{x}, \qquad (76)$$

for an arbitrary point x in the domain.

5 Domain growth with the inclusion of regulated cell migration

In this section we combine the results of previous sections, to show the results of regulated cell migration on growing domains. Throughout, we use the "even-splitting" method, as given by equation (37).

Fig. 5 Cells diffusing at a constant rate with density-dependent growth of the form $f = rc^2$, where r is a constant. Other details of the plots are given in the caption of Figure 4. Parameters are as follows: $k_0 = 50$, $d = 1.0$, $r = 0.00005$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

5.1 Linear domain growth with non-local sensing

Our first example considers non-local sensing with (isotropic) linear growth, such that

$$L(t) = L_0 \alpha(t) \quad \text{with} \quad \alpha(t) = 1 + \rho t, \tag{77}$$

where, as before, L_0 is the initial domain length (which we will take to be unity) and ρ the rate of growth. We note that this type of domain growth corresponds to having $f(x, t, c(x, t)) = \rho/(1 + \rho t)$ in equation (127). In this case, the PDE description of domain growth, given by equation (30) of Section 3.1, has

$$u(x, t) = \frac{\rho x}{(1 + \rho t)}. \tag{78}$$

There are two choices we could make for the rate of domain growth in the stochastic formulation. Firstly, we could make the simple change $r \mapsto \rho(1 + \rho t)^{-1}$ in the formulation used for constant splitting, equation (33). This would result in the same formulation as before, and the relationship to the equivalent PDE would arise in exactly the same manner. More interesting though is to consider the stochastic formulation of linear growth in which

$$r(t) = \frac{\psi}{\text{number of boxes at time } t}, \tag{79}$$

for some constant $\psi > 0$ which will be determined subsequently. This leads to the RDME (excluding movements based on non-local measurements):

$$\frac{\partial P^k(\boldsymbol{n}, t)}{\partial t} = \frac{\psi}{k-1} \sum_{i=1}^{k-1} \pi(n_i, n_{i+1} | n_i + n_{i+1}) P^{k-1}(G_i \boldsymbol{n}, t) - \frac{\psi}{k} \sum_{i=1}^{k} P^k(\boldsymbol{n}, t), \tag{80}$$

where π and G_i, $i = 1, \ldots, n$ are defined earlier, and a system of mean equations of the form

$$\frac{\partial M_j^k}{\partial t} = \frac{\psi}{k-1} \left[(j - 3/2) M_{j-1}^{k-1} + (k - j - 1/2) M_j^{k-1} \right] - \psi M_j^k. \tag{81}$$

In order to derive correspondence with a PDE, as in earlier sections of this work, it is useful to consider the mean number of boxes. As before, we let $P_k(t)$ denote the probability of having k boxes at time t and write down the master equation satisfied by the P_k:

$$\frac{dP_k}{dt} = \psi P_{k-1} - \psi P_k, \quad \text{for} \quad k = k_0, k_0 + 1, \ldots, \tag{82}$$

with initial condition given by equation (53).

This infinite set of ODEs may be solved inductively to give $P_k(t) \equiv 0$ for $k < k_0$ and

$$P_k(t) = \frac{1}{(k - k_0)!}(\psi t)^{k-k_0} e^{-\psi t}, \tag{83}$$

for $k \geq k_0$.

The mean number of boxes is then given by

$$K(t) = \sum_{k=k_0}^{\infty} k P_k(t) = \sum_{k=k_0}^{\infty} \frac{k}{(k - k_0)!}(\psi t)^{k-k_0} e^{-\psi t}. \tag{84}$$

Relabelling indices, it is trivial to see that

$$K(t) = k_0 + \psi t, \tag{85}$$

i.e. that the mean domain length increases linearly with time. Comparing with the deterministic equation, we see that we must take $\psi = k_0 \rho$ in order that the mean stochastic growth rate corresponds to the deterministic growth rate.

We substitute the expression for the mean number of boxes into equation (81) and require the approximate relation $M_j^{k-1} \approx M_j^k$, in order to see correspondence with the deterministic equation:

$$\frac{\partial c}{\partial t} + \frac{\rho x}{1 + \rho t}\frac{\partial c}{\partial x} = D_n \frac{\partial}{\partial x}\left(s(x)\frac{\partial c}{\partial x} - c\frac{\partial s(x)}{\partial x}\right) - \frac{\rho}{1 + \rho t}c, \quad (x,t) \in (0, L_0(1 + \rho t)) \times [0, \infty). \tag{86}$$

Results from numerical simulation of the system with $s(x) = d\exp(-ax)$ are shown in Figure 6. As expected, there is good correlation between the average stochastic density and the PDE solution. We also note the close agreement between average stochastic domain length and domain length given by the deterministic formulation.

5.2 Logistic domain growth with local sensing

We next consider local sensing with (isotropic) logistic growth of the form

$$L(t) = L_0 \alpha(t) \quad \text{with} \quad \alpha(t) = \frac{\xi e^{\rho t}}{\xi + (e^{\rho t} - 1)}, \tag{87}$$

where L_0 is the initial domain length, ρ represents the rate of growth and ξ the ratio between final and initial domain sizes. This can be seen more easily by noting

$$\frac{d\alpha}{dt} = \rho \alpha \left(1 - \frac{\alpha}{\xi}\right), \tag{88}$$

so that the PDE description of domain growth, given by equation (30) of Section 3.1 has $u(x,t) = \rho(1 - \alpha(t)/\xi)$. The RDME takes the same form as equation (33) but with

$$r(t) = \rho\left(1 - \frac{\text{number of boxes at time } t}{\xi \times \text{number of boxes at time } 0}\right), \tag{89}$$
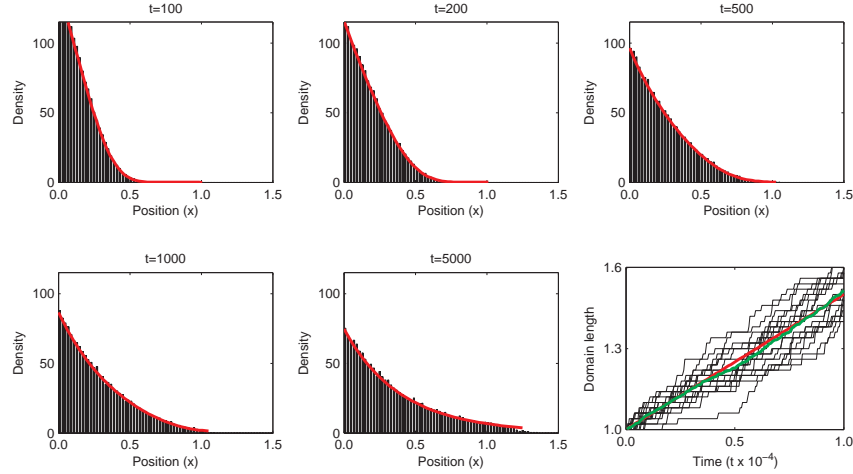
Fig. 6 Cells moving according to a non-local sensing mechanism with linear growth. The histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines the result of numerical simulation of the corresponding PDE using the finite difference algorithm outlined in Appendix C.2. The boundaries are assumed to be reflecting and all cells are initially in box 1 ($x = 0.01$). The sixth plot shows the evolution of domain length over time for each of the 20 realisations (black lines), the average stochastic length (green line) and the length predicted by the corresponding continuum system (red line). Parameters are as follows: $k_0 = 50$, $d = 1.0$, $\rho = 0.00005$, $a = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

whilst the number of boxes is less than ξ times the number of boxes at time 0, and $r(t) = 0$ otherwise. Excluding movements based on local measurements we have

$$\frac{\partial P^k(\boldsymbol{n}, t)}{\partial t} = \rho \left(1 - \frac{k-1}{\xi k_0} \right) \sum_{i=1}^{k-1} \pi(n_i, n_{i+1} | n_i + n_{i+1}) P^{k-1}(G_i \boldsymbol{n}, t) \tag{90}$$

$$- \rho \left(1 - \frac{k}{\xi k_0} \right) \sum_{i=1}^{k} P^k(\boldsymbol{n}, t),$$

for $k = k_0, \ldots, k_f$, where π and G_i, $i = 1, \ldots, n$ are defined as earlier and $k_f = \xi k_0$, the final number of boxes. The result is a system of mean equations of the form

$$\frac{\partial M_j^k}{\partial t} = \rho \left(1 - \frac{k-1}{\xi k_0} \right) \left[(j - 3/2) M_{j-1}^{k-1} + (k - j - 1/2) M_j^{k-1} \right] - \rho k \left(1 - \frac{k}{\xi k_0} \right) M_j^k. \tag{91}$$

We consider finding the mean domain length, again by considering a master equation for the number of boxes at time t:

$$\frac{\mathrm{d}P_k}{\mathrm{d}t} = \rho(k-1) \left(1 - \frac{k-1}{k_f} \right) P_{k-1} - \rho k \left(1 - \frac{k}{k_f} \right) P_k, \quad \text{for} \quad k = k_0, k_0 + 1, \ldots, k_f. \tag{92}$$

Initial conditions are as before, k_0 boxes at time $t = 0$. This infinite set of ODEs may be solved inductively[1] to give $P_k(t) \equiv 0$ for $k > k_f$, $k < k_0$ and

$$P_k(t) = \prod_{i=k_0}^{k-1} N_i \times \sum_{j=k_0}^{k} \frac{e^{-\rho N_j t}}{\prod_{l=k_0, j \neq l}^{k} (N_j - N_l)}, \tag{94}$$

---

[1] We note that this provides a general solution for a system of ODES of the form

$$\frac{\mathrm{d}P_k}{\mathrm{d}t} = N_{k-1} P_{k-1} - N_k P_k. \tag{93}$$

for $k_0 \le k \le k_f$, where $N_j = j(1 - j/k_f)$.

In this case, the equation for the mean number of boxes over time is not easy to evaluate and in order to gain further insight, we consider deriving an expression for $K(t)$ directly from the master equation:

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \sum_{k=0}^{\infty} k \frac{\mathrm{d}P_k}{\mathrm{d}t}, \tag{95}$$

$$= \sum_{k=1}^{\infty} \rho k(k-1)\left(1 - \frac{k-1}{k_f}\right) P_{k-1} - \sum_{k=0}^{\infty} \rho k^2 \left(1 - \frac{k}{k_f}\right) P_k, \tag{96}$$

$$= \rho \left(\langle k \rangle - \frac{\langle k^2 \rangle}{k_f} \right), \tag{97}$$

where $\langle k^n \rangle$ denotes the nth moment of k and $\langle k \rangle = K$. This result is to be expected: since the master equation is non-linear in k, we cannot derive a closed form for the mean number of boxes. However, correspondence with the deterministic case can be drawn by assuming, as we have several times in this work, that $\langle k^2 \rangle \approx \langle k \rangle^2$ in order to write

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \rho K \left(1 - \frac{K}{k_f}\right) \quad \implies \quad K(t) = \frac{k_0 e^{\rho t}}{1 + \frac{1}{\xi}(e^{\rho t} - 1)}, \tag{98}$$

c.f. the deterministic case.

Returning to the original RDME, we can now derive an approximate relation between M_j^{k-1} and M_j^k (not shown) in order to draw correspondence with the anticipated PDE:

$$\frac{\partial c}{\partial t} + \rho x \left(1 - \frac{\alpha(t)}{\xi}\right) \frac{\partial c}{\partial x} = D_n \frac{\partial}{\partial x}\left(s\frac{\partial c}{\partial x} + c\frac{\partial s}{\partial x}\right) - \rho\left(1 - \frac{\alpha(t)}{\xi}\right) c, \tag{99}$$

for $(x, t) \in (0, L_0\alpha(t)) \times (0, \infty)$ and $\alpha(t)$ as given by equation (87).

Results from numerical simulation of the system with $s(x) = d\exp(-ax)$ are shown in Figure 7. As expected, there is good correlation between the average stochastic density and the PDE solution. The results are particularly good as the upper bound on the number of boxes means that, in this case, all realisations attain the maximum number of boxes, and no more. Subsequent simulations will illustrate the issues that may arise if this is not the case.

5.3 Density-dependent domain growth with local and non-local sensing

We now turn to our final application—considering density-dependent domain growth of the form $f(x, t, c(x, t)) = rc$ with both local (Figure 8) and non-local (Figure 9) sensing and a signal of the form $s(x) = d\exp(-ax)$. In this case, little further analysis can be carried out to guide our explorations and we present only numerical results based on the results presented in Section 4.

It is here that we highlight the possible pitfalls of our approximations. For local sensing and a domain growing at a rate linearly proportional to cell density, we see that the domain grows more quickly at the right-hand edge which is precisely the region in which cells build up. Numerical simulation of our PDE approximation appears to over estimate the number of cells at the right-hand boundary for the simple reason that in some simulations the domain length exceeds that predicted by the PDE formulation and this results in lower average cell densities in the region. For non-local sensing, the build up of cells occurs at the left-hand edge of the domain, so growth occurs most rapidly in a region where all cells are contained and, as a result, the PDE formulation agrees better with the stochastic realisations.
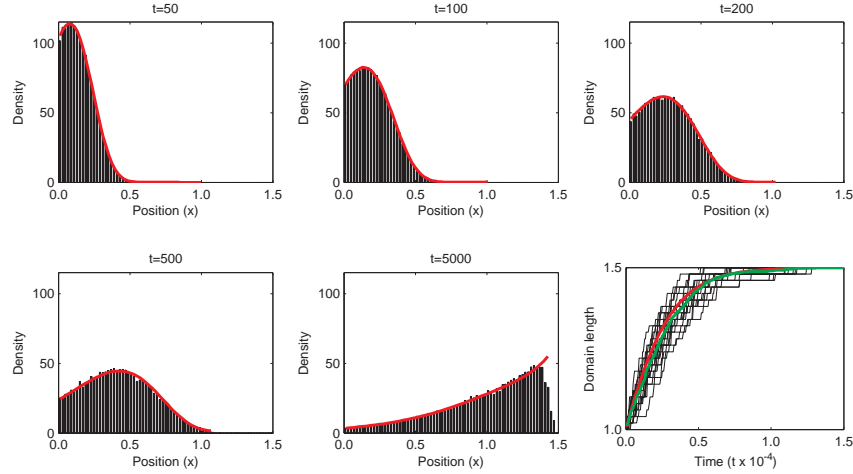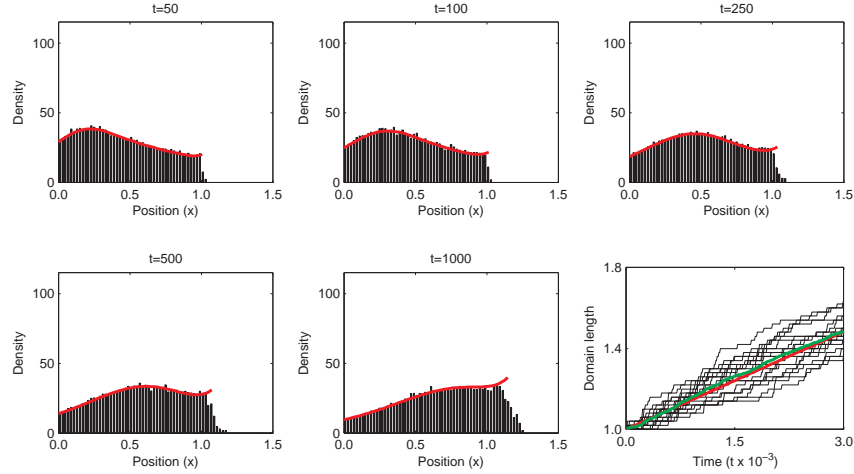
Fig. 7 Cells moving according to a local sensing mechanism with logistic growth. The histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines the result of numerical simulation of the corresponding PDE using the finite difference algorithm outlined in Appendix C.2. The boundaries are assumed to be reflecting and all cells are initially in box 1 ($x = 0.01$). The sixth plot shows the evolution of domain length over time for each of the 20 realisations (black lines), the average stochastic length (green line) and the length predicted by the corresponding continuum system (red line). Parameters are as follows: $k_0 = 50$, $d = 1.0$, $\rho = 0.0005$, $\xi = 1.5$, $a = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

6 Discussion

We began this work by outlining a stochastic framework for considering cell motility: by dividing the domain under consideration into boxes, we used a continuous-time RDME to describe the evolution of probability density. In Section 2 equations describing evolution of the stochastic means were derived and linked to a PDE describing cell density by using a Taylor series expansion to convert from discrete to continuous space. We applied the framework to consider three types of cell sensing of an external signal: local; non-local; average. In each case we found a correspondence between the microscopic, cell level description of movement and a macroscopic PDE description of cell density. Numerical simulations confirmed the validity of our results. In Appendix A, we present some extensions to the model that allow more general biological systems to be modelled: cell proliferation/flux/decay, sensing of cell density and stochastic modelling of the signal density.

We next considered the incorporation of domain growth: first, in Section 3, the simple case in which each box splits at the same, constant rate as every other and then, in Section 4, the case in which box splitting is dependent upon the cell density in that box. In each case we were able to draw correspondence with a PDE describing the dynamics of the system as the domain changes size. Numerical simulations validate our results for the range of parameter values considered. Both average density and domain lengths for the individual-level models closely followed those predicted by the corresponding population-level model for all types of movement and growth considered here. Deviations arise in the quadratic growth simulation (Figure 5) as the domain grows significantly larger than its original size and we point this phenomenon out for as a significant avenue for future research. For a constant box splitting rate we were able to derive an expression for the mean number of boxes over time, and this matched the domain length predicted by the PDE formulation. For density-dependent splitting we have, as yet, been unable to derive a general expression—this is another area for future research.

Finally, in Section 5, we applied our results to consider local and non-local sensing on domains growing isotropically (logistically and linearly, respectively) and anisotropically, with dependence
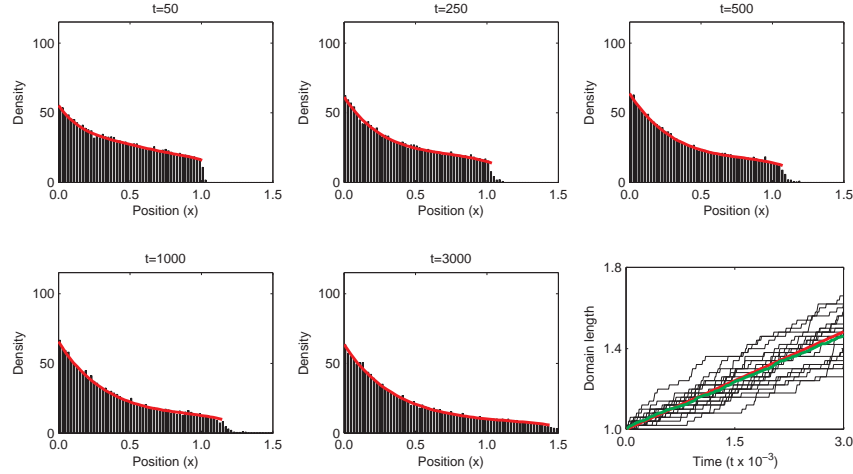
Fig. 8 Cells moving according to a local sensing mechanism with density-dependent growth of the form $f = rc$, where r is a constant. This histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines the result of numerical simulation of the corresponding PDE using the NAG routine D03PE as outlined in Appendix C.3. The boundaries are assumed to be reflecting and the initial cell distribution consists of 1000 cells approximately distributed as $c(x,0) = 1000e^x / \sum_{i=1}^{k} e^{x_i}$. The sixth plot shows the evolution of domain length over time for each of the 20 realisations (black lines), the average stochastic length (green line) and the length predicted by the corresponding continuum system (red line). Parameters are as follows: $k_0 = 50$, $d = 1.0$, $r = 0.000005$, $a = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

on local cell density. In most cases, agreement with the PDE formulation is excellent and demonstrates the equivalence of the stochastic and deterministic approaches described here.

Symmetry Assumption (S). Throughout our explorations of growing domains we have made use of Symmetry Assumption **(S)**: that the probability density function describing division of cells between the original and new boxes is symmetric about its mid-point. We justify this approximation biologically by assuming that there is no bias in the splitting of cells between original and new boxes. Mathematically this prevents any artificial drift from being introduced into the model as a result of box splitting. Having said this, Symmetry Assumption **(S)** is crucial only if the diffusion is on a comparable (or slower) time scale than growth. However, in many applications, we have fast diffusion compared to growth (see later in Discussion). Any asymmetry in splitting would be compensated by the diffusive flux (from boxes with more cells) before the next splitting occurs. Thus, Symmetry Assumption **(S)** could be relaxed by a suitable closure approximation and the same equations would be obtained.

Moment closure approximations. Our other main assumptions have come in the form of the moment closure approximations that enable us to draw parallels between the RDME and PDE formulations of the system. Our first approximation, equation (49) can be thought of as representing the "dilution" effects that take place as the domain grows—we expect the smoothing effects of diffusion to re-distribute cells between each box splitting event. Our subsequent moment closure approximation, equation (66), has a less intuitive explanation but relies on a similar "smoothing" argument.

We have not, thus far, carried out a thorough analytical investigation into the accuracy of our assumptions but we present numerical estimations drawn from our stochastic realisations in Appendix D. We show, that for the models (and specific parameter values) considered here, the approximations are valid to within a very reasonable tolerance and so lend support to our model. Our studies also indicate that these approximations are most accurate when growth occurs on

Fig. 9 Cells moving according to a non-local sensing mechanism with density-dependent growth of the form $f = rc$, where r is a constant. This histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines the result of numerical simulation of the corresponding PDE using the NAG routine D03PE as outlined in Appendix C.3. The boundaries are assumed to be reflecting and the initial cell distribution consists of 1000 cells approximately distributed as $c(x,0) = 1000e^x / \sum_{i=1}^{k} e^{x_i}$. The sixth plot shows the evolution of domain length over time for each of the 20 realisations (black lines), the average stochastic length (green line) and the length predicted by the corresponding continuum system (red line). Parameters are as follows: $k_0 = 50$, $d = 1.0$, $r = 0.000005$, $a = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

a long timescale compared with those of cell movements, and when cell numbers are reasonably large. Just how far we can push these approximations remains to be tested in more depth: in the following paragraph we present an example where our approximations fall down.

An example where our moment closure methods are not sufficient. We consider the situation in which domain growth (box splitting) is regulated by the presence of a signalling molecule produced in the underlying tissue according to the birth-death process

$$\emptyset \xrightarrow{k_p} S, \qquad S \xrightarrow{k_d} \emptyset, \tag{100}$$

i.e. S is produced at rate k_p and linearly degraded at rate k_d. We assume that the box splitting rate per unit time is a non-linear (in this case, saturating) function of S:

$$\text{splitting rate of box } i = v(S_i) = \frac{rS_i}{K + S_i}, \tag{101}$$

where r and K are positive parameters and S_i is the concentration of signalling molecule in box i. Further, we assume that S is bound to the tissue (*i.e.* may not diffuse) so that we have the reactions given by equation (100) taking place independently in each box. If we start the system with $S^* = k_p/k_d$ molecules of S in each box (*i.e.* at steady state) then the mean-field description of the system would suggest that the number of boxes evolves like

$$\frac{\mathrm{d}k}{\mathrm{d}t} = v(S^*)k \quad \implies \quad k(t) = k(0)e^{v(S^*)t}. \tag{102}$$

However, the extent to which the individual-level model mimics this growth rate depends on the nonlinearities of the growth function and the ability of the system to buffer against fluctuations in S in the absence of diffusion. We demonstrate this in Figure 10. In the left-hand plot, K is much smaller than S^* and the effects of variation in signalling molecule concentration are buffered to a
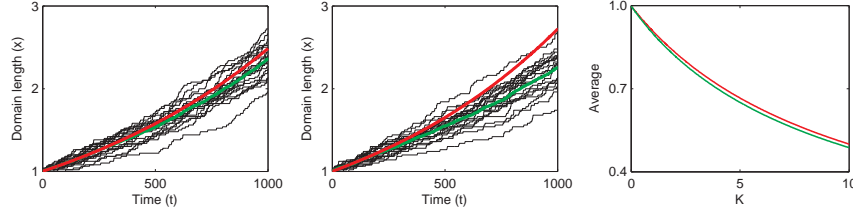
Fig. 10 The effects of stochastic regulation of domain growth. The red line shows the domain length estimated by equation (102) whilst the green lines show the average stochastic length over the 20 realisations (each shown in black). In the left-hand plot $r = 0.001$ and $K = 1.0$ and in the centre plot $r = 0.001$ and $K = 10.0$. In all plots $k_p = 0.01$ and $k_d = 0.001$ giving a steady state for S of 10 molecules. The right-hand plot shows a comparison between the mean-field approximation of the growth rate (red line) and that predicted from the stationary distribution of S (green line). In both cases the growth rate is per box and we take $r = 1.0$ for clarity.

large extent. However, in the centre plot K is of the same order of magnitude as S^* and we see that fluctuations in S cause the stochastic domain length to grow, in general, more slowly than the mean-field description permits. Mathematically our argument can be illustrated by noting that, in the individual-level case, the stationary distribution for each S_i is of Poisson form with mean k_p/k_d. In this case the mean growth rate is given by

$$\left\langle \sum_{i=1}^{k} \frac{rS_i}{K + S_i} \right\rangle = k \left\langle \frac{rS}{K + S} \right\rangle = \sum_{s=0}^{\infty} \left(\frac{rS}{K + S} \right) \frac{1}{S!} \left(\frac{k_p}{k_d} \right)^s e^{-\frac{k_p}{k_d}} \neq kv(\langle S \rangle), \qquad (103)$$

where $\langle S \rangle$ is the average number of S molecules. The right-hand plot of Figure 10 illustrates our results, plotting the mean growth rate estimated from calculation of the stationary distribution (equation (103)) alongside results from the mean-field calculation (equation (102)).

From this study we see that, for certain parameter regimes, the mean-field description over predicts growth of the system and the individual-level model must be used to generate accurate predictions of the domain length. This simple example highlights the importance of individual-level models.

Parameter values. In the applications which we have in mind, a typical diffusion constant is of the order $D = 10^{-8} - 10^{-7}$ cm^2 sec^{-1} [25] and the domain length is of order $L_0 = 500\,\mu$m-1 mm [42]. Using the nondimensionalisation $\hat{x} = x/L_0$ and $\tau = t/t_0$ where $t_0 = 1$ min, $L_0 = 1$ mm, we obtain the dimensionless diffusion coefficient $D_d = Dt_0/L_0^2 \approx 6 \times 10^{-5} - 6 \times 10^{-4}$. Using $\Delta x = 0.02$ (the value used in our illustrative computations) we obtain $d = D_d/\Delta_x^2 \approx 1$ which is the value of d used in Figure 3. Thus our dimensionless results can be expressed in units [cm] and [minutes]. In particular, the results in our illustrative figures show the time evolution over several hours, and the domain growth rate $r = 0.001$ used throughout the paper is equal to [1 micrometer per minute] which is also in a realistic range. In fact, parameters from [42] suggest an even smaller rate for growth. Based on *Drosophila melanogaster* wing disc expansion from approximately 40 to 50,000 cells over 4 days [42], and assuming cells are approximately cubic with side length 10 μm, gives a growth rate in the region of [0.05 micrometers per minute]: an even more extreme separation of growth and diffusion time scales than those studied here. As discussed, this could have implications for the relaxation of Symmetry Assumption **(S)**. On the other hand, for larger values of r we expect the accuracy of our closure approximations to decrease as r increases relative to the diffusion rate, d, and this will be the subject of further investigation by us. However, results from numerical simulations (not shown) indicate that the rate of domain growth in the constant splitting case (Section 3) maybe increased by at least an order of magnitude from that shown in Figure 3 without a noticeable decrease in accuracy.

6.1 A general framework for spatio-temporal dynamics on growing domains

The most general description of the spatio-temporal evolution of populations, such as those described here, needs to be able to include migration (both random diffusion and controlled/directed migration), cell proliferation and decay, and the interactions of different cell types with each other and chemical signals in the environment. The results presented in this work provide a general framework for modelling such populations on growing domains using RDME's and how to connect these models to more well-known systems of PDEs of the form of equation (29). As such, our results provide a framework for considering more general population interactions, for example, the effects of quorum sensing or volume filling studied in the deterministic sense by Painter and Hillen [38]. Further, these results can be applied to investigate specific biological systems in which the interactions between chemical species and growth of the domain are tightly coupled [42].

6.2 Open problems

There are numerous avenues for future research. Firstly, the possibility of applying moment closure methods to cases in which transition or box splitting rates are a function of cell density should be explored to further test the validity of assumptions made in order to derive a limiting PDE describing cell density. Secondly, the application of these methods to two and three spatial dimensions—especially the exploration of domain growth—is non-trivial and should be investigated in order for such models to be applied to biological systems where the geometry of the system cannot easily be reduced to one dimension. Thirdly, cell proliferation is often one of the driving forces of domain expansion. We have attempted to include this phenomenon in our models by including density-dependent domain growth but this problem deserves a more thorough consideration as it is of vital importance in many aspects of embryonic development. Finally, the goal of this research is to provide a platform for furthering our insight into the mechanisms underlying cell migration and domain growth in biology. Proper model building, analysis and parametrisation, followed by hypothesis-driven testing, should be carried out in order to fulfill this goal and also to suggest ways in which the models may be improved.

A Extensions of the model

We outline three fairly trivial extensions to the framework that, when combined, enable a wider set of scenarios to be described using the RDME approach.

A.1 Cell flux and decay

We consider spatially dependent cell influx and decay by assuming that a cell in box i decays at rate p_i^d and that p_i^p is the rate of influx of cells into box i. The latter could be adjusted also to reflect cell proliferation. Neglecting diffusion, the RDME becomes

$$\frac{\partial P(\boldsymbol{n}, t)}{\partial t} = \sum_{i=1}^{k} p_i^d \left\{ (n_i + 1) P(A_i^c \boldsymbol{n}, t) - n_i P(\boldsymbol{n}, t) \right\} + \sum_{i=1}^{k} p_i^p(t) \left\{ P(A_i^a \boldsymbol{n}, t) - P(\boldsymbol{n}, t) \right\}, \quad (104)$$
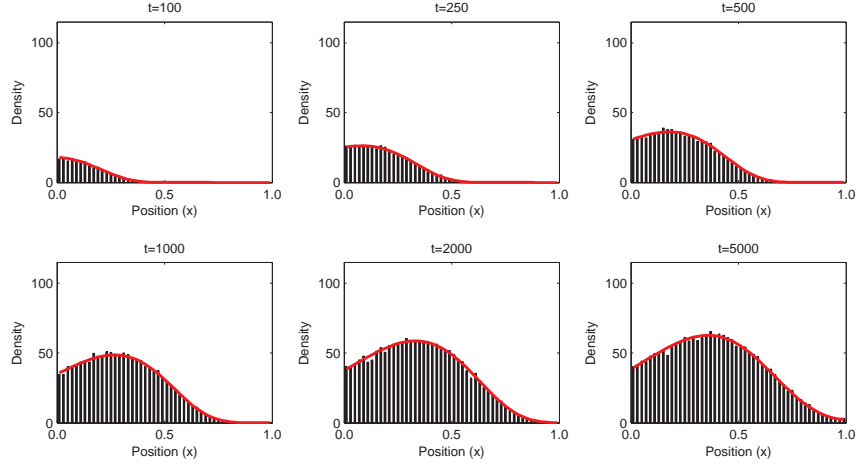
Fig. 11 Cells moving according to a local sensing mechanism with cell flux and decay. The histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines show the result of numerical simulation of the corresponding PDE using the finite difference approximation outlined in Appendix C.1. The local sensing mechanism is the same as for Figure 2, $p_1^p = p_c$, $p_i^p = 0$ for $i = 2, \ldots, k$ and $p_i^d = \lambda_c$, $i = 1, \ldots, k$. Initially the domain is empty and the boundaries are assumed to be reflecting. Parameters are as follows: $k = 50$, $d_l = 1.0$, $a = 2.0$, $p_c = 2.0$, $\lambda_c = 0.001$, $L_0 = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

where the creation and annihilation operators $A_i^c, A_i^a : \mathbb{R}^k \to \mathbb{R}^k$, are defined by

$$A_i^c : [n_1, \ldots, n_i, \ldots, n_k] \to [n_1, \ldots, n_i + 1, \ldots, n_k],$$
$$A_i^a : [n_1, \ldots, n_i, \ldots, n_k] \to [n_1, \ldots, n_i - 1, \ldots, n_k],$$

for $i = 1, \ldots, k$. Equation (104) implies

$$\frac{\partial M_i}{\partial t} = p_i^p - p_i^d M_i. \tag{105}$$

The corresponding PDE is easily seen to be

$$\frac{\partial c}{\partial t} = p^p(x) - p^d(x)c, \qquad (x, t) \in (0, L) \times [0, \infty), \tag{106}$$

with corresponding boundary conditions.

Figure 11 shows the results of simulating cell movement according to local measurement rules with cell flux and decay included. There is an influx of cells into box 1 at rate p_c and all cells decay at rate λ_c. By $t = 5000$ a quasi-steady state is approximately reached with cell densities in the stochastic realisation approximately those suggested by numerical solution of the corresponding PDE.

A.2 Density-dependent diffusion

In the case of density-dependent diffusion, we assume that the transition rates depend on cell density at the current and/or neighbouring grid points so that

$$T_i^\pm = d_l f(n_i) + d_n g(n_{i\pm 1}) \quad \text{or} \quad T_i^\pm = df(n_i)g(n_{i\pm 1}). \tag{107}$$

Specific forms for density-dependence were investigated by Painter and Hillen [37] in the context of volume filling and quorum sensing—recognising the fact that cells have a finite size and an ability

to sense density, both locally and non-locally. We take a more general approach here, that may be compared to their results by sensible choice of f and g.

Then the RDME can still be written in the form (5), noting that the addition of density-dependence introduces non-linearities into the calculation of the stochastic means. In the first case we have

$$\frac{\partial \langle n_i \rangle}{\partial t} = \langle [d_l f(n_{i-1}) + d_n g(n_i)] \rangle + \langle [d_l f(n_{i+1}) + d_n g(n_i)] \rangle \tag{108}$$
$$- \langle [d_l f(n_i) + d_n g(n_{i-1})] \rangle - \langle [d_l f(n_i) + d_n g(n_{i+1})] \rangle,$$

and in the second

$$\frac{\partial \langle n_i \rangle}{\partial t} = \langle df(n_{i-1})g(n_i) \rangle + \langle df(n_{i+1})g(n_i) \rangle \tag{109}$$
$$- \langle df(n_i)g(n_{i-1}) \rangle - \langle df(n_i)g(n_{i+1}) \rangle,$$

for $i = 2, \ldots, k - 1$, where $\langle f(\boldsymbol{n}) \rangle = \sum_{\boldsymbol{n}} f(\boldsymbol{n})P(\boldsymbol{n}, t)$. Similar equations hold for for $i = 1, k$. We cannot derive a closed system of equations for the first moments (means), $\langle n_j \rangle$, as we have dependencies on higher order moments. However, in the limit of large numbers of cells and where the correction terms are small, we can still expect the stochastic means to approximately satisfy,

$$\frac{\partial M_i}{\partial t} \approx [d_l f(M_{i-1}) + d_n g(M_i)] + [d_l f(M_{i+1}) + d_n g(M_i)] \tag{110}$$
$$- [d_l f(M_i) + d_n g(M_{i-1})] - [d_l f(M_i) + d_n g(M_{i+1})],$$

and

$$\frac{\partial M_i}{\partial t} \approx df(M_{i-1})g(M_i) + df(M_{i+1})g(M_i) \tag{111}$$
$$- df(M_i)g(M_{i-1}) - df(M_i)g(M_{i+1}),$$

away from the boundaries $i = 1$ and $i = k$.

Expanding terms of the form $f(M_{i\pm 1})$, correct to 2nd order in Δx, gives

$$f(M_{i\pm 1}) = f \left(M(x_i) \pm (\Delta x) \frac{\partial M}{\partial x}(x_i) + \frac{1}{2}(\Delta x)^2 \frac{\partial^2 M}{\partial x^2}(x_i) + o\left((\Delta x)^2\right) \right),$$
$$= f(M(x_i)) \pm (\Delta x) \frac{\partial M}{\partial x}(x_i) f'(M(x_i))$$
$$+ \frac{1}{2}(\Delta x)^2 \left[\frac{\partial^2 M}{\partial x^2}(x_i) f'(M(x_i)) + \left(\frac{\partial M}{\partial x}(x_i) \right)^2 f''(M(x_i)) \right]$$
$$+ o\left((\Delta x)^2\right). \tag{112}$$

Substituting into equations (110) and (111) and taking the limit as $\Delta x \to 0$ such that $\lim_{\Delta x \to 0} d_l (\Delta x)^2 = D_l$, $\lim_{\Delta x \to 0} d_n (\Delta x)^2 = D_n$ and $\lim_{\Delta x \to 0} d(\Delta x)^2 = D$ results in the limiting PDEs

$$\frac{\partial c}{\partial t} = \frac{\partial^2}{\partial x^2} \left\{ [D_l f(c) - D_n g(c)] c \right\}, \qquad (x, t) \in (0, 1) \times [0, \infty), \tag{113}$$

and

$$\frac{\partial c}{\partial t} = Dg(c) \frac{\partial^2}{\partial x^2} [cf(c)] - Df(c) \frac{\partial^2}{\partial x^2} [cg(c)], \qquad (x, t) \in (0, 1) \times [0, \infty), \tag{114}$$

respectively. Zero flux boundary conditions hold at $x = 0$ and $x = 1$.

Figure 12 shows the results of simulating density-dependent diffusion with local sensing and the diffusion rate a monotonic decreasing function of cell density at the current site. We see that the average of 20 stochastic realisations agrees well with numerical solution of the corresponding PDE. We note that as all cells are initially situated in box 1, progression towards the homogeneous quasi-steady state (given by setting the right-hand side of equation (113) to zero) is initially slow.
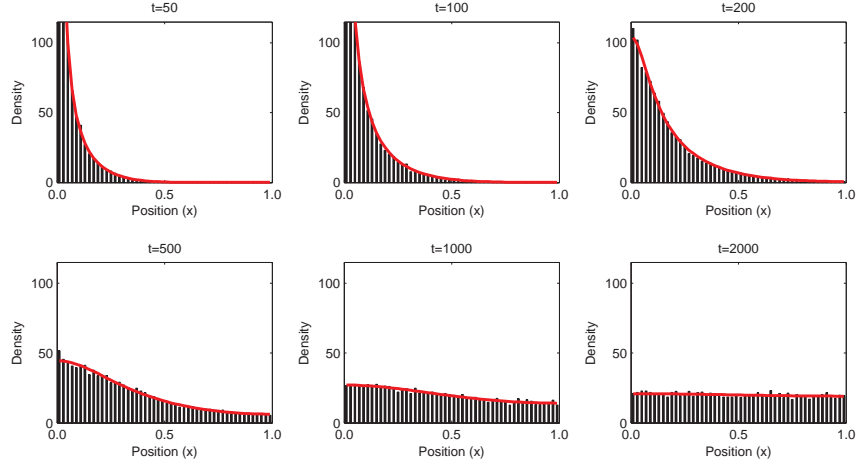
Fig. 12 Cells moving according to a density-dependent, local sensing mechanism. The histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines show the result of numerical simulation of equation (113) using the finite difference approximation outlined in Appendix C.1. The transition rates are given by $T_i^{\pm} = d_l A/(A + n_i)$, all cells are initially in box 1 and the boundaries are assumed to be reflecting. Parameters are as follows: $k = 50$, $d_l = 1.0$, $A = 50.0$, $L_0 = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

A.3 Stochastic modelling of the environment

The idea can be extended to include cases in which the cells control the signal density themselves. As a simple example we consider a system consisting of a signalling molecule moving with transition rates $T_{s,i}^{\pm} = d_s$ and cells whose movements are controlled according to signalling molecule concentration in a local manner, $T_{c,i}^{\pm} = d_c f(s_i)$ where s_i is the number of signalling molecules in box i and f is a function describing signal amplification. We assume (i) a production of cells in the left-hand most box at rate p_c, (ii) that cells produce the signalling molecule at rate p_s and (iii) that cells and signalling molecules decay linearly at rates λ_c and λ_s, respectively. Zero flux conditions are implemented on both boundaries.

We let $P(\boldsymbol{n}, \boldsymbol{s}, t)$ be the joint probability that $C_i = n_i$ and $S_i = s_i$ for $i = 1, 2, \ldots, k$ where $\boldsymbol{n} = [n_1, n_2, \ldots, n_k]$ and $\boldsymbol{s} = [s_1, s_2, \ldots, s_k]$. The RDME can be written

$$\frac{\partial}{\partial t} P(\boldsymbol{n}, \boldsymbol{s}, t) = \sum_{i=1}^{k-1} d_s \left\{ (s_i + 1) P(\boldsymbol{n}, J_i^+ \boldsymbol{s}, t) - s_i P(\boldsymbol{n}, \boldsymbol{s}, t) \right\} \tag{115}$$

$$+ \sum_{i=2}^{k} d_s \left\{ (s_i + 1) P(\boldsymbol{n}, J_i^- \boldsymbol{s}, t) - s_i P(\boldsymbol{n}, \boldsymbol{s}, t) \right\}$$

$$+ \sum_{i=1}^{k} p_s \left\{ n_i P(\boldsymbol{n}, A_i^a \boldsymbol{s}, t) - n_i P(\boldsymbol{n}, \boldsymbol{s}, t) \right\}$$

$$+ \sum_{i=1}^{k} \lambda_s \left\{ (s_i + 1) P(\boldsymbol{n}, A_i^c \boldsymbol{s}, t) - s_i P(\boldsymbol{n}, \boldsymbol{s}, t) \right\}$$

$$+ \sum_{i=1}^{k-1} d_c f(s_i) \left\{ (n_i + 1) P(J_i^+ \boldsymbol{n}, \boldsymbol{s}, t) - n_i P(\boldsymbol{n}, \boldsymbol{s}, t) \right\}$$

$$+ \sum_{i=2}^{k} d_c f(s_i) \left\{ (n_i + 1) P(J_i^- \boldsymbol{n}, \boldsymbol{s}, t) - n_i P(\boldsymbol{n}, \boldsymbol{s}, t) \right\}$$

$$+ \sum_{i=1}^{k} \lambda_c \left\{ (n_i + 1)P(A_i^c \boldsymbol{n}, \boldsymbol{s}, t) - n_i P(\boldsymbol{n}, \boldsymbol{s}, t) \right\}$$

$$+ p_c \left\{ P(A_1^c(\boldsymbol{n}, \boldsymbol{s}, t) - P(\boldsymbol{n}, \boldsymbol{s}, t) \right\}. \tag{116}$$

The stochastic means for signalling molecule number evolve according to

$$\frac{\partial \langle s_1 \rangle}{\partial t} = d_s \left(\langle s_2 \rangle - \langle s_1 \rangle \right) + p_s \langle n_1 \rangle - \lambda_s \langle s_1 \rangle, \tag{117}$$

$$\frac{\partial \langle s_i \rangle}{\partial t} = d_s \left(\langle s_{i+1} \rangle - 2\langle s_i \rangle + \langle s_{i-1} \rangle \right) + p_s \langle n_i \rangle - \lambda_s \langle s_i \rangle, \tag{118}$$

$$\frac{\partial \langle s_k \rangle}{\partial t} = d_s \left(\langle s_{k-1} \rangle - \langle s_k \rangle \right) + p_s \langle n_k \rangle - \lambda_s \langle s_k \rangle, \tag{119}$$

where equation (118) holds for $i = 2, \ldots, k - 1$. Those for the cell number evolve according to

$$\frac{\partial \langle n_1 \rangle}{\partial t} = d_c \langle n_2 f(s_2) \rangle - d_c \langle n_1 f(s_1) \rangle + p_c - \lambda_c \langle n_1 \rangle, \tag{120}$$

$$\frac{\partial \langle n_i \rangle}{\partial t} = d_c \langle n_{i-1} f(s_{i-1}) \rangle - 2d_c \langle n_i f(s_i) \rangle + d_c \langle n_{i+1} f(s_{i+1}) \rangle - \lambda_c \langle n_i \rangle, \tag{121}$$

$$\frac{\partial \langle n_k \rangle}{\partial t} = d_c \langle n_{k-1} f(s_{k-1}) \rangle - d_c \langle n_k f(s_k) \rangle - \lambda_c \langle n_k \rangle, \tag{122}$$

where, similarly, equation (121) holds for $i = 2, \ldots, k - 1$. We assume $\langle n_i f(s_j) \rangle = \langle n_i \rangle f(\langle s_j \rangle)$ in order to draw correspondence with the following system of PDEs:

$$\frac{\partial s}{\partial t} = D_s \frac{\partial^2 s}{\partial x^2} + p_s n - \lambda_s s, \tag{123}$$

$$\frac{\partial c}{\partial t} = D_c \frac{\partial}{\partial x} \left[f(s) \frac{\partial c}{\partial x} + c \frac{\partial f(s)}{\partial x} \right] - \lambda_c c, \tag{124}$$

for $(x, t) \in (0, 1) \times [0, \infty)$. s and c represent signalling molecule and cell density, respectively, and D_s, D_c are defined in the same manner as before. Boundary conditions are specified by the requirement for consistency with equations (120), (122) and result in a flux of signalling molecule at rate p_c into the domain at the left-hand boundary and zero flux conditions at the right-hand boundary. The same approach yields similar results for non-local cell measurements *etc.*

Figure 13 shows the results of numerical simulation the system described in this section. Once again we see good agreement between the stochastic realisation and numerical solution of the PDEs describing signalling molecule concentration and cell density.

B Population-level results

In this section we outline the derivation of some of the PDEs used to describe domain growth on a population-level. Whilst the results are not necessarily new, we provide them for completeness.

B.1 Population-level behaviour for general domain growth

We follow the derivation of Section 3.1 as far as equation (30). This leaves us requiring expressions for the length of the domain and the flow due to growth. In order to derive an equation for these we consider dividing the domain up into N intervals of length Δx_i for $i = 1, \ldots, N$. For a time interval of length $\Delta t \ll 1$ we may write

$$L(t + \Delta t) = \sum_{i=1}^{N} \left[1 + f(x_i, t, c(x_i, t)) \Delta t \right] \Delta x_i,$$

$$= L(t) + \sum_{i=1}^{N} f(x_i, t, c(x_i, t)) \Delta t \Delta x_i. \tag{125}$$
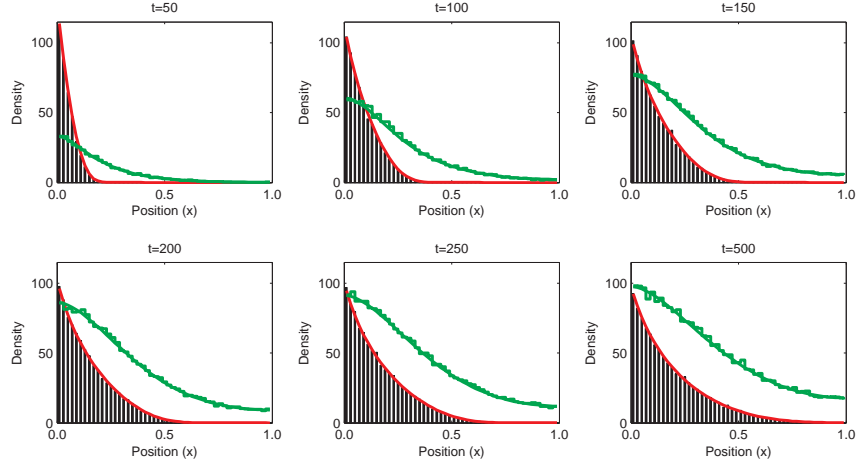
Fig. 13 Cells moving according to a local sensing mechanism with signalling molecule dynamics included. The histograms show the average of 20 stochastic realisations of the system using the Gillespie algorithm [21] (see Appendix C.4 for more details) and the solid lines show the result of numerical simulation of equations (123)-(124) using the finite difference approximation outlined in Appendix C.1. The cell transition probabilities are give by $T_i^{\pm} = d_c s_i$, the domain is initially empty, there is a flux of cells into box 1 at rate p_c and the boundaries are assumed to be reflecting. Parameters are as follows: $k = 50$, $d_s = 5.0$, $d_c = 0.01$, $p_c = 10.0$, $p_s = 0.05$, $\lambda_c = 0.01$ $\lambda_s = 0.02$, $L_0 = 1.0$, $\Delta t = 0.01$ and $\Delta x = 0.02$.

Re-arranging and taking the limit as $\Delta t \to 0$ we have

$$\frac{\mathrm{d}L(t)}{\mathrm{d}t} = \sum_{i=1}^{N} f(x_i, t, c(x_i, t)) \Delta x_i. \tag{126}$$

Taking the limit as $\Delta x_i \to 0$ for $i = 1, \ldots, N$ gives an integro-differential equation for $L(t)$:

$$\frac{\mathrm{d}L(t)}{\mathrm{d}t} = \int_0^{L(t)} f(x, t, c(x, t)) \mathrm{d}x, \tag{127}$$

which, upon integrating, gives

$$L(t) = \int_0^t \int_0^{L(\bar{t})} f(x, \bar{t}, c(x, \bar{t})) \mathrm{d}x \mathrm{d}\bar{t}. \tag{128}$$

This method may also be used to describe the flow due to domain growth, $u(x, t, c(x, t))$, at any point in the domain:

$$u(x, t, c(x, t)) = \frac{\mathrm{d}x}{\mathrm{d}t} = \int_0^x f(\bar{x}, t, c(\bar{x}, t)) \mathrm{d}\bar{x}. \tag{129}$$

Substituting the expression for flow due to growth, equation (129), into equation (30) gives the PDE we are looking for:

$$\frac{\partial c}{\partial t} + \frac{\partial}{\partial x} \left(c \int_0^x f(\bar{x}, t, c(\bar{x}, t) \mathrm{d}\bar{x} \right) = D \frac{\partial^2 c}{\partial x^2}, \quad (x, t) \in (0, L(t)) \times [0, \infty). \tag{130}$$

We close the system by specifying zero flux boundary conditions at $x = 0$ and $x = L(t)$ (see Appendix B.2 for justification of the choice boundary conditions).

B.2 Choice of boundary conditions in the domain growth case.

As derived in Section 3, the PDE for a species diffusing on a growing domain is given by

$$\frac{\partial c}{\partial t} + \frac{\partial (uc)}{\partial x} = D\frac{\partial^2 c}{\partial x^2}, \quad \text{for} \quad (x,t) \in (0, L(t)) \times [0, \infty), \tag{131}$$

where u is the flow due to domain growth, D is the diffusion constant and $L(t)$ is specified by equation (127). We make a change of variables, $\xi = x/L(t)$ and $\tau = t$, in order to normalise the domain to $\xi \in [0,1]$. In this case we have

$$\frac{\mathrm{d}L(\tau)}{\mathrm{d}\tau} = \int_0^1 L(\tau)f(c(\xi,\tau))\mathrm{d}\xi, \tag{132}$$

Substituting into the transformed PDE (equation (131)),

$$-\frac{\xi}{L(\tau)}\frac{\mathrm{d}L}{\mathrm{d}\tau}\frac{\partial c}{\partial \xi} + \frac{\partial c}{\partial \tau} + f(c(\xi L(\tau),\tau))c(\xi L(\tau),\tau)$$
$$+ \left(\int_0^\xi L(\tau)f(c(\bar{\xi},\tau))d\bar{\xi}\right)\frac{1}{L(\tau)}\frac{\partial c}{\partial \xi} = \frac{D}{L(\tau)^2}\frac{\partial^2 c}{\partial \xi^2}, \tag{133}$$

gives

$$\frac{\partial c}{\partial \tau} - \left(\xi\int_0^1 f(c(\xi,\tau))\mathrm{d}\xi\right)\frac{\partial c}{\partial \xi} + f(c(\xi L(\tau),\tau))c(\xi L(\tau),\tau)$$
$$+ \left(\int_0^\xi f(c(\bar{\xi},\tau))d\bar{\xi}\right)\frac{\partial c}{\partial \xi} = \frac{D}{L(\tau)^2}\frac{\partial^2 c}{\partial \xi^2}, \tag{134}$$

which can be simplified to

$$\frac{\partial c}{\partial \tau} = \left(\xi\int_0^L f(c(\xi,\tau))d\xi - \int_0^\xi f(c(\bar{\xi},\tau))d\bar{\xi}\right)\frac{\partial c}{\partial \xi}$$
$$-f(c(\xi L(\tau),\tau))c(\xi L(\tau),\tau) + \frac{D}{L(\tau)^2}\frac{\partial^2 c}{\partial \xi^2}. \tag{135}$$

Our boundary conditions for domain growth are derived using a cell conservation argument, so that total cell number in the domain stays constant (assuming no proliferation or decay):

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^{L(t)} c(x,t)\mathrm{d}x = 0. \tag{136}$$

Differentiating under the integral gives

$$\int_0^{L(t)} \frac{\partial c}{\partial t}\mathrm{d}x + \frac{\mathrm{d}L}{\mathrm{d}t}c(L(t),t) = 0. \tag{137}$$

We replace $\partial c/\partial t$ using the evolution equation to give

$$\int_0^{L(t)} \frac{\partial}{\partial x}\left(D\frac{\partial c}{\partial x} - \left\{\int_0^x f(c(\bar{x},t))d\bar{x}\right\}c(x,t)\right)\mathrm{d}x + \frac{\mathrm{d}L}{\mathrm{d}t}c(L(t),t) = 0,$$
$$\Rightarrow \left[D\frac{\partial c}{\partial x} - \left\{\int_0^x f(c(\bar{x},t))d\bar{x}\right\}c(x,t)\right]_0^{L(t)} + \frac{\mathrm{d}L}{\mathrm{d}t}c(L(t),t) = 0. \tag{138}$$

Noting that the second two terms cancel, we are left with

$$\left[\frac{\partial c}{\partial x}\right]_0^{L(t)} = 0, \tag{139}$$

which are zero flux boundary conditions on the growing domain. When we change coordinates the corresponding boundary conditions on the transformed domain become

$$\left[\frac{1}{L(\tau)}\frac{\partial c}{\partial \xi}\right]_0^1 = 0 \Rightarrow \left[\frac{\partial c}{\partial \xi}\right]_0^1 = 0, \tag{140}$$

zero flux boundary conditions once again.

C Numerical simulations

We outline the main numerical methods used in solving the systems of equations presented in this work. In all cases simulations were conducted using either FORTRAN 90 or MATLAB and the results displayed using MATLAB. Unless otherwise stated, zero flux boundary conditions are used, and the parameter values are stated in the caption of the figure.

C.1 Finite difference approximations for the PDE on a static domain.

The spatial discretisation is carried out using a centered finite difference scheme of the following form:

$$\frac{\partial}{\partial x}\left(f\frac{\partial g}{\partial x}\right) \approx \frac{1}{\Delta x}\left(\left[f\frac{\partial g}{\partial x}\right]_{i+\frac{1}{2}} - \left[f\frac{\partial g}{\partial x}\right]_{i-\frac{1}{2}}\right), \tag{141}$$

$$\approx \frac{f_{i+\frac{1}{2}}}{\Delta x}\left(\frac{g_{i+1}-g_i}{\Delta x}\right) - \frac{f_{i-\frac{1}{2}}}{\Delta x}\left(\frac{g_i - g_{i-1}}{\Delta x}\right). \tag{142}$$

Averaging the functions evaluated at mid-points,

$$f_{i\pm\frac{1}{2}} \approx \frac{f_i + f_{i\pm1}}{2}, \tag{143}$$

gives

$$\frac{\partial}{\partial x}\left(f\frac{\partial g}{\partial x}\right) \approx \frac{(f_{i-1}+f_i)g_{i-1}}{2(\Delta x)^2} - \frac{(f_{i-1}+2f_i+f_{i+1})g_i}{2(\Delta x)^2} + \frac{(f_i+f_{i+1})g_{i+1}}{2(\Delta x)^2}. \tag{144}$$

The temporal discretisation is of the form

$$\frac{\partial f}{\partial t} \approx \frac{f^{j+1} - f^j}{\Delta t}. \tag{145}$$

C.2 Numerical solution of the PDE for constant domain growth.

For the constant growth case, we have $L(t) = L(0)e^{rt}$ which results in the transformed equation

$$\frac{\partial c}{\partial t} = \frac{D}{L(t)^2}\frac{\partial^2 c}{\partial x^2} - rc, \qquad (x,t) \in (0,1) \times [0,\infty). \tag{146}$$

This equation can be solved using finite differences in the usual way, paying attention to the changing effective diffusion constant, $D/L(t)^2$.

C.3 Numerical solution of the PDE for density-dependent domain growth.

In this case we use a Lagrangian formulation, making a transformation of coordinates of the form [2, 10–12]

$$x = \Gamma(X, \tau) \quad \text{and} \quad t = \tau, \tag{147}$$

where the advection due to domain growth is defined by the strain rate, $\sigma = u_x$ satisfying

$$\sigma = f(c(x,t)) = \frac{\Gamma_{X\tau}}{\Gamma_X}, \tag{148}$$

and the subscripts denote partial derivatives. This leads to the system of PDEs:

$$c_\tau = \frac{D}{\Gamma_X} \left(\frac{c_X}{\Gamma_X} \right)_X - \sigma c, \tag{149}$$

$$\Gamma_{X\tau} = \sigma \Gamma_X, \tag{150}$$

where $\Gamma(X,0) = X$, $\Gamma(0,t) = 0$ and the usual zero flux boundary conditions apply.

We transform the system to a first order one and employ the NAG library routine D03PE and the NAG MATLAB toolbox in order to solve the system numerically.

C.4 Gillespie Algorithm.

The stochastic simulations were carried out using the Gillespie algorithm [21]. No approximate methods for speeding up the simulations were used as the simulation times were not significant and the original algorithm avoids issues such as species numbers becoming negative. One could also use more efficient exact methods equivalent to the Gillespie algorithm, like the Next Reaction Method [20] or the Next Subvolume Method [23].

D Justification of the moment closure approximations

Firstly, we attempt to justify the moment closure approximations made in equations (49) and (66) using the results of our stochastic simulations. For each of the 20 realisations we output the cell numbers in each box for $t = 1, 2, \ldots, 200$. For each domain length arising in the 20 realisations we calculate the average number of cells in each box of the domain at that length. Since the domains grow at approximately the same rates, this gives estimates of M_j^k for $j = 1, \ldots, k$. In Figure 14 we demonstrate the accuracy of our moment closure approximations by plotting

$$\text{Relative error in } j^{\text{th}} \text{ box} = \left(M_j^k - \frac{(k-1)}{k} M_j^{k-1} \right) \Big/ M_j^k, \tag{151}$$

for constant growth (the difference between the right-hand and left-hand sides of equation (49)) and

$$\text{Relative error in } j^{\text{th}} \text{ box} = \left(\sum_{i=1}^{k} \langle n_j^k \rangle f(\langle n_i^k \rangle) - \sum_{i=1}^{k-1} \langle n_j^{k-1} \rangle f(\langle n_i^{k-1} \rangle) \right) \Big/ \sum_{i=1}^{k} \langle n_j^k \rangle f(\langle n_i^k \rangle), \tag{152}$$

for linear and quadratic growth (the difference between the right-hand and left-hand sides of equation (66)).

The graphs demonstrate that, in general, the moment closure approximations are accurate for the types of domain growth shown here. Similarly good agreement can be found for the other types considered in this work. As expected, the approximations generally become less accurate as the domain size increases and our studies indicate that this is because the variance in domain length increases over time.
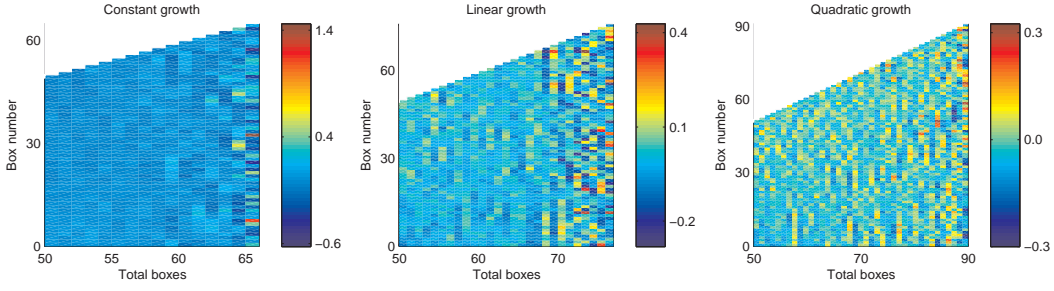
Fig. 14 Justification of the moment closure approximations made for constant (see Section 3.2, equation (31) and Figure 3), linear (see Section 4.1, equation (57) and Figure 4) and quadratic (see Section 4.1, equation (57) and Figure 5) domain growth. In each case, the relative error between the left-hand and right-hand sides of the moment closure equations, equation (49) for constant growth and equation (66) for linear and quadratic growth, are plotted for all values of domain length. We note that total boxes corresponds to the index k whilst box number to the index j of our moment equations.
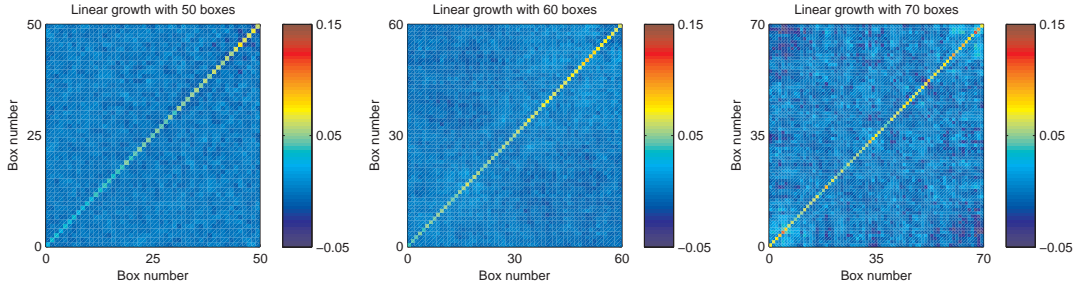
Fig. 15 Justification of the first moment closure approximation for linear domain growth (see Section 4.1, equation (57) and Figure 4). We plot the relative error between the left-hand and right-hand sides of the moment closure equation (65) for three different domain lengths. The graphs are symmetric about $i = j$ by definition. We note that box number corresponds to the index k of our moment equations.

Secondly, we attempt to justify the moment closure approximations made in equation (65) for linear and quadratic growth by plotting

$$\left(\langle n_j^k f(n_i^k) \rangle - \langle n_j^k \rangle f(\langle n_i^k \rangle) \right) / \langle n_j^k f(n_i^k) \rangle . \tag{153}$$

As before, we use the output of 20 realisations with density distributions recorded for $t = 1, 2, 3, \ldots, 200$. We used all data at time points with the required number of boxes in order to calculate estimates of the right-hand and left-hand sides of equations (65). Figure 15 shows the results of our investigation for linear growth and Figure 16 for quadratic growth. In each case the approximation is extremely accurate throughout the field, with a slight increase when $i = j$ (in particular for small j).

References

1. M. Alber, N. Chen, P. M. Lushnikov, and S. A. Newman, *Continuous macroscopic limit of a discrete stochastic model for interaction of living cells*, Phys. Rev. Lett. **99** (2007), no. 16, 168102–4.
2. R. E. Baker and P. K. Maini, *A mechanism for morphogen-controlled domain growth*, J. Math. Biol. **54** (2007), no. 5, 597–622.
3. N. Barkai and S. Leibler, *Robustness in simple biochemical networks*, Nature **387** (1997), 913–917.
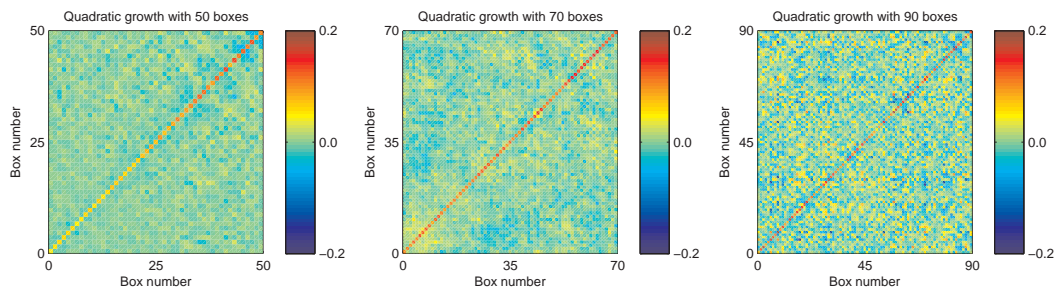
Fig. 16 Justification of the first moment closure approximation for quadratic domain growth (see Section 4.1, equation (57) and Figure 5). We plot the relative error between the left-hand and right-hand sides of the moment closure equation (65) for three different domain lengths. The graphs are symmetric about $i = j$ by definition. We note that box number corresponds to the index k of our moment equations.

4. H. C. Berg, *How bacteria swim*, Sci. Amer. **233** (1975), 36–44.

5. ———, *Random Walks in Biology*, Princeton University Press, 1983.

6. H. C. Berg and E. M. Purcell, *Physics of chemoreception*, Biophys. J. **20** (1977), no. 2, 193–219.

7. M. P. Brenner, L. S. Levitov, and E. O. Budrene, *Physical mechanisms for chemotactic pattern formation by bacteria*, Biophys. J. **74** (1998), no. 4, 1677–1693.

8. F. Chalub, P. Markowich, B. Perthame, and C. Schmeiser, *Kinetic models for chemotaxis and their drift-diffusion limits*, Monatshefte für Mathematik **142** (2004), no. 1-2, 123–141.

9. P. Cluzel, M. Surette, and S. Leibler, *An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells*, Science **287** (2000), 1652–1655.

10. E. J. Crampin, E. A. Gaffney, and P. K. Maini, *Reaction and diffusion on growing domains: scenarios for robust pattern formation*, Bull. Math. Biol. **61** (1999), no. 6, 1093–1120.

11. E. J. Crampin, W. W. Hackborn, and P. K. Maini, *Pattern formation in reaction-diffusion models with nonuniform domain growth*, Bull. Math. Biol. **64** (2002), no. 4, 747–769.

12. E. J. Crampin and P. K. Maini, *Reaction-diffusion models for biological pattern formation*, Meth. App. Anal. **8** (2001), no. 3, 415–428.

13. M. A. Crickmore and R. S. Mann, *Hox control of organ size by regulation of morphogen production and mobility*, Science **313** (2006), no. 5783, 63–68.

14. Y. Dolak and C. Schmeiser, *Kinetic models for chemotaxis: Hydrodynamic limits and spatio-temporal mechanisms*, J. Math. Biol. **51** (2005), no. 6, 595–615.

15. R. Erban, S. J. Chapman, and P. K. Maini, *A practical guide to stochastic simulations of reaction-diffusion processes*, 34 pages, available as http://arxiv.org/abs/0704.1908, 2007.

16. R. Erban, I. Kevrekidis, and H. G. Othmer, *An equation-free computational approach for extracting population-level behavior from individual-based models of biological dispersal*, Physica D **215** (2006), no. 1, 1–24.

17. R. Erban and H. G. Othmer, *From individual to collective behaviour in bacterial chemotaxis*, SIAM J. Appl. Math. **65** (2004), no. 2, 361–391.

18. ———, *From signal transduction to spatial pattern formation in* E. coli*: A paradigm for multi-scale modeling in biology*, Multiscale Model. Sim. **3** (2005), no. 2, 362–394.

19. ———, *Taxis equations for amoeboid cells*, J. Math. Biol. **54** (2007), no. 6, 847–885.

20. M. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A **104** (2000), 1876–1889.

21. D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. **81** (1977), no. 25, 2340–2361.

22. T. M. Gomez and P. C. Letourneau, *Filopodia initiate choices made by sensory neuron growth cones at laminin/fibronectin borders* in vitro, J. Neurosci. **14** (1994), no. 10, 5959–5972.

23. J. Hattne, D. Fange, and J. Elf, *Stochastic reaction-diffusion simulation with MesoRD*, Bioinformatics **21** (2005), no. 12, 2923–2924.

24. T. Hillen and H. G. Othmer, *The diffusion limit of transport equations derived from velocity-jump processes*, SIAM J. Appl. Math. **61** (2000), 751–775.

25. T. Höfer, J. A. Sherratt, and P. K. Maini, *Cellular pattern formation during dictyostelium aggregation*, Physica D **85** (1995), no. 3, 425–444.

26. D. Horstmann, K. J. Painter, and H. G. Othmer, *Aggregation under local reinforcement: From lattice to continuum*, Eur. J. Appl. Math. **15** (2004), 545–576.

27. E. Keller and L. Segel, *Initiation of slime mold aggregation viewed as an instability*, J. Theor. Biol. **26** (1970), no. 3, 399–415.

28. _____, *Model for chemotaxis*, J. Theor. Biol. **30** (1971), 225–234.

29. _____, *Traveling bands of chemotactic bacteria: A theoretical analysis*, J. Theor. Biol. **30** (1971), 235–248.

30. H. Levine, D. Kessler, and W. Rappel, *Directional sensing in eukaryotic chemotaxis: A balanced inactivation model*, Proc. Natl. Acad. Sci. USA **103** (2006), no. 26, 9761–9766.

31. F. Lin, C. M.-C. Nguyen, S.-J. Wang, W. Saadi, S. P. Gross, and N. L. Jeon, *Effective neutrophil chemotaxis is strongly influenced by mean IL-8 concentration*, Biochem. Bioph. Res. Co. **319** (2004), no. 2, 576–581.

32. J. D. Murray, *Mathematical Biology*, Springer Verlag, 2002.

33. H. G. Othmer, S. Dunbar, and W Alt, *Models of dispersal in biological systems*, J. Math. Biol. **26** (1988), 263–298.

34. H. G. Othmer and T. Hillen, *The diffusion limit of transport equations II: chemotaxis equations*, SIAM J. Appl. Math. **62** (2002), no. 4, 1222–1250.

35. H. G. Othmer and P. Schaap, *Oscillatory cAMP signaling in the development of Dictyostelium discoideum*, Comm. Theor. Biol. **5** (1998), 175–282.

36. H. G. Othmer and A. Stevens, *Aggregation, blowup, and collapse: The ABC's of taxis in reinforced random walks*, SIAM J. Appl. Math. **57** (1997), no. 4, 1044–1081.

37. K. J. Painter and T. Hillen, *Volume-filling and quorum-sensing in models for chemosensitive movement*, Canadian Appl. Math. Quarterly **10** (2002), no. 4, 501–543.

38. _____, *Volume filling and quorum-sensing in models for chemosensitive movement*, Canadian Appl. Math. Quarterly **10** (2002), no. 4, 501–543.

39. K. J. Painter, D. Horstmann, and H. G. Othmer, *Localization in lattice and continuum models of reinforced random walks*, Appl. Math. Lett. **16** (2003), no. 3, 375–381.

40. C. Parent and P. Devreotes, *A cell's sense of direction*, Science **284** (1999), 765–770.

41. C. Patlak, *Random walk with persistence and external bias*, Bull. Math. Biophys. **15** (1963), no. 3, 311–338.

42. D. Rogulja and K. Irvine, *Regulation of cell proliferation by a morphogen gradient*, Cell **123** (2005), no. 3, 449–461.

43. P. Spiro, J Parkinson, and H. G. Othmer, *A model of excitation and adaptation in bacterial chemotaxis*, Proc. Natl. Acad. Sci. USA **94** (1997), 7263–7268.

44. S. Turner, J. A. Sherratt, and K. J. Painter, *From a discrete to continuous model of biological cell movement*, Phys. Rev. E **69** (2004), 021910–1–021910–10.

45. T. Veikkola, M. Karkkainen, L. Claesson-Welsh, and K. Alitalo, *Regulation of angiogenesis via vascular endothelial growth factor receptors*, Cancer Res. **60** (2000), no. 2, 203–212.

46. M. Ward, C. McCann, M. DeWulf, J. Y. Wu, and Y. Rao, *Distinguishing between directional guidance and motility regulation in neuronal migration*, J. Neurosci. **23** (2003), no. 12, 5170–5177.

47. D. Wessels, J. D. Murray, and D. Soll, *Behavior of Dictyostelium amoebae is regulated primarily by the temporal dynamic of the natural cAMP wave*, Cell Motil. Cytoskel. **23** (1992), no. 2, 145–156.

48. L. Wolpert, R. Beddington, T. Jessell, P. Lawrence, E. Meyerowitz, and J. Smith, *Principles of Development*, 3rd ed., Oxford University Press, 2006.