You may also be interested in:

Stochastic simulation tools and continuum models for describing two-dimensional collective cell spreading with universal growth functions
Wang Jin, Catherine J Penington, Scott W McCue et al.

Velocity-jump processes with proliferation
Katrina K Treloar, Matthew J Simpson and Scott W McCue

De-pinning of disordered bosonic chains
N Vogt, J H Cole and A Shnirman

Mechanisms of the formation of biological signaling profiles
Hamid Teimouri and Anatoly B Kolomeisky

Comparing a discrete and continuum model of the intestinal crypt
Philip J Murray, Alex Walter, Alexander G Fletcher et al.

Model of reversible vesicular transport with exclusion
Paul C Bressloff and Bhargav R Karamched

Extended Plefka expansion for stochastic dynamics
B Bravi, P Sollich and M Opper

Hydrodynamics of local excitations after an interaction quench in 1D cold atomic gases
Fabio Franchini, Manas Kulkarni and Andrea Trombettoni

Soliton splitting in quenched classical integrable systems
O Gamayun and M Semenyakin

New Journal of Physics

The open access journal at the forefront of physics

PAPER

CrossMark

A new and accurate continuum description of moving fronts

S T Johnston[1], R E Baker[2] and M J Simpson[1]

[1] Mathematical Sciences, Queensland University of Technology, Brisbane, Australia
[2] Mathematical Institute, University of Oxford, Oxford, United Kingdom

E-mail: matthew.simpson@qut.edu.au

Abstract

Processes that involve moving fronts of populations are prevalent in ecology and cell biology. A common approach to describe these processes is a lattice-based random walk model, which can include mechanisms such as crowding, birth, death, movement and agent–agent adhesion. However, these models are generally analytically intractable and it is computationally expensive to perform sufficiently many realisations of the model to obtain an estimate of average behaviour that is not dominated by random fluctuations. To avoid these issues, both mean-field (MF) and corrected mean-field (CMF) continuum descriptions of random walk models have been proposed. However, both continuum descriptions are inaccurate outside of limited parameter regimes, and CMF descriptions cannot be employed to describe moving fronts. Here we present an alternative description in terms of the dynamics of groups of contiguous occupied lattice sites and contiguous vacant lattice sites. Our description provides an accurate prediction of the average random walk behaviour in all parameter regimes. Critically, our description accurately predicts the persistence or extinction of the population in situations where previous continuum descriptions predict the opposite outcome. Furthermore, unlike traditional MF models, our approach provides information about the spatial clustering within the population and, subsequently, the moving front.

Introduction

Moving fronts feature ubiquitously throughout biological and ecological processes [1–12]. The introduction of non-native species can result in a catastrophic invasion wave if the introduced species out-competes native fauna [2, 13]. For example, the cane toad *Bufo marinus* was introduced to north-eastern Australia in 1935, and has subsequently invaded much of northern Australia due to a lack of natural predation [9, 14]. Similarly, malignant tumours spread through the invasion of previously healthy tissue, such as glioma cells moving throughout the brain to form glioblastoma [1, 3, 12, 15].

Lattice-based random walk models that include crowding, birth, death, movement and agent–agent adhesion are commonly used to describe processes that involve moving fronts [6, 16–21]. For example, these random walk models have been used to interpret *in vitro* cell biology experiments by considering the position of the leading edge of the cell front or the cell density profile [6, 19, 21]. Illien *et al* [18] consider random walks in the context of single-file diffusion models of active transport, that is, transport that requires energy due to an opposing force. However, the stochastic nature of random walks makes it problematic to efficiently examine the collective behaviour of a population, as a large number of realisations of the random walk must be performed to reduce the influence of stochastic fluctuations. Furthermore, it is difficult to determine meaningful population behaviour through analysis of the discrete process. There is, therefore, considerable interest in approaches that are both analytically tractable and avoid the computational expense of repeated simulations.

A common technique to analyse random walk processes is to consider a deterministic, continuum approximation of the discrete process [19, 20, 22–30]. The standard approximation, known as a mean-field (MF) approximation, results in a partial differential equation (PDE) [19, 20, 25]. The resulting PDE is amenable to analysis but only provides an accurate approximation of the discrete process in an extremely limited set of
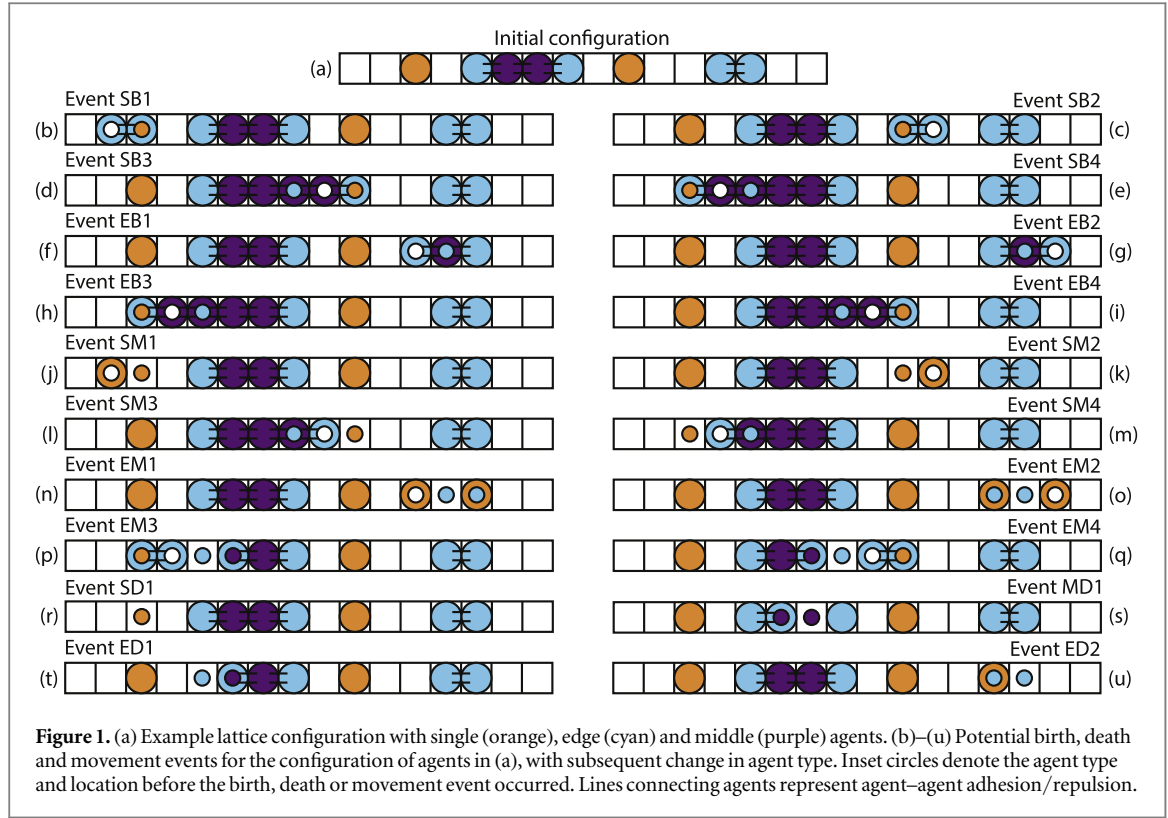
parameter regimes where spatial correlations are weak [22, 24, 31]. Hence, using these approximations to model moving fronts may provide an inaccurate estimate of the velocity of the moving front if the spatial correlations are important. This inaccuracy could have significant consequences if, for example, the front velocity is used to make decisions about implementing ecological control measures. To address the influence of spatial correlations, alternative approximation techniques have been proposed [22, 24, 28, 32, 33]. The corrected mean-field (CMF) approximation, which explicitly describes pairwise correlations, results in a system of ordinary differential equations (ODEs) that accurately approximate the discrete process for a wider range of parameter regimes, compared to the MF approximation [22, 24]. However, the CMF description is still invalid in parameter regimes where spatial correlations are sufficiently strong [28]. Furthermore, the CMF cannot be used to study moving fronts, as the governing equations cannot be evaluated anywhere that has zero agent density [22], such as areas where the population has yet to invade. The chain-and-gap approach (*C&G*), proposed by Johnston *et al* [28], considers the dynamics of groups of contiguous occupied and vacant sites, and results in a system of ODEs that provide an accurate approximation of the discrete process in all parameter regimes. These groups are termed *chains* and *gaps* for the contiguous occupied and vacant sites, respectively. Additionally, the *C&G* description provides information about the spatial clustering and patchiness present in the system. There is considerable interest in determining the influence of local spatial structure on the persistence of a species [34, 35]. However, the *C&G* description has previously only been applied to discrete processes that are, on average, spatially uniform [28]. As such, the *C&G* description is not currently suitable for describing processes that contain moving fronts.

Here the *C&G* description presented by Johnston *et al* [28] is extended to incorporate spatial variation so that the description can be applied to moving fronts. We interpret the discrete process in terms of chains and gaps, noting that the left-most site in each chain or gap can occur at any lattice site. The corresponding system of ODEs is derived and presented, and we demonstrate that the numerical solution of the ODE system provides an accurate approximation of the average discrete behaviour in all cases, even in parameter regimes where both the MF and CMF descriptions are inaccurate. This allows for the robust prediction of whether a population persists or becomes extinct, as well as reliable estimates of the velocity of the moving front. In addition, for the first time, the *C&G* description has been extended to include rates of birth, death and movement that are dependent on the length of the chain an agent belongs to. Furthermore, the *C&G* description includes explicit information about the spatial clustering present within a moving front.

Random walk model

We consider a one-dimensional lattice-based random walk model where each lattice site may be occupied by, at most, one agent [36]. The lattice is interpreted as a combination of groups of contiguous occupied and vacant sites [28]. Agents on the lattice undergo birth, death and movement events. These events occur at rates P_p^n, P_d^n and P_m^n per unit time, respectively, where $n \in [1, N]$ is the length of the chain an agent belongs to and N is the total number of sites. During a potential birth event, an agent randomly selects a nearest-neighbour site and attempts to place a daughter agent at that site. The birth event is successful provided that the target site is vacant [24]. During a death event, an agent is removed from the lattice [24]. During a potential movement event, an agent selects a nearest-neighbour site and attempts to move to that site, and is successful provided that the target site is vacant [24]. The target site selection is unbiased if both nearest-neighbour sites are vacant. If one nearest-neighbour site is occupied, the vacant nearest-neighbour site is selected with probability $(1 - \alpha)/2$, where $\alpha \in [-1, 1]$ [26]. The constant parameter α represents the strength of agent–agent adhesion/repulsion. Setting $\alpha = 0$ means that there is no agent–agent adhesion/repulsion, whereas setting $\alpha \neq 0$ simulates agent–agent adhesion ($\alpha > 0$) or repulsion ($\alpha < 0$) [26]. Note that α does not depend on the chain length.

Due to crowding, the success of birth and movement events depends on whether an agent has zero, one or two nearest-neighbour agents. These agents are referred to as *single*, *edge* or *middle* agents, respectively [28]. An example lattice configuration highlighting the different types of agents is presented in figure 1. The necessary information to obtain the average numbers of single, edge and middle agents at each site $i \in [1, N]$ is encoded within the average number of chains of length $n \in [1, N]$ that contain the site i. The average number of single, edge and middle agents at site i are denoted $N_i^S(t)$, $N_i^E(t)$ and $N_i^M(t)$, respectively. The average number of chains at time t where the left-most agent in the chain is at site $i \in [1, N]$ and the length is $n \in [1, N - i + 1]$ is denoted $C_i^n(t)$. Similarly, the average number of gaps at time t where the left-most vacant site in the gap is at site $i \in [1, N]$ and the length is $m \in [1, N - i + 1]$ is denoted $G_i^m(t)$. The spatially dependent restriction on the length is due to the choice of no-flux boundary conditions. Note that $N_i^S(t)$, $N_i^E(t)$, $N_i^M(t)$, $C_i^n(t)$ and $G_i^m(t)$ are all temporally dependent but, for convenience, we do not explicitly include this dependence. The average numbers of edge, middle and single

Figure 1. (a) Example lattice configuration with single (orange), edge (cyan) and middle (purple) agents. (b)–(u) Potential birth, death and movement events for the configuration of agents in (a), with subsequent change in agent type. Inset circles denote the agent type and location before the birth, death or movement event occurred. Lines connecting agents represent agent–agent adhesion/repulsion.

agents at site i are given by:

$$N_i^E = \sum_{j=2}^{N-i+1} C_i^j + \sum_{j=2}^{i-1} C_{i-j+1}^j,$$

$$N_i^M = \sum_{j=1}^{i-1} \sum_{k=3}^{N-j+1} C_j^k, \quad \text{and}$$

$$N_i^S = C_i^1.$$

Note that middle agents cannot exist at sites $i = 1$ and $i = N$ due to the choice of boundary conditions.

Similar to the approach of Johnston *et al* [28], we consider how birth, death and movement events change the location and lengths of the chains and gaps, rather than the occupancy of an individual lattice site. This approach avoids making an assumption about the probability that a particular site is occupied or vacant, as the sites either side of a chain or gap are necessarily vacant or occupied, respectively. There are twenty different types of events that change either the location of the left-most agent in a chain, the length of a chain, or both. An example of each event is presented in figure 1 for a particular configuration of agents. Events can have more than one potential outcome. For example, the potential outcomes of a single agent at site i undergoing a birth event can be classified into four groups. The daughter agent can be placed at site $i - 1$ or $i + 1$, and the gap that the daughter agent is placed in can be length one or greater. Here we detail each possible event and the subsequent change in configuration with respect to the number of chains and gaps. Events are referred to by a nomenclature describing the type of agent undergoing the event and the mechanism of the event itself, followed by a number highlighting the potential for multiple outcomes to arise from a specific event. For example, birth events for single agents are referred as SB events. Since there are four different types of SB events, we refer to these as SB1, SB2, SB3 and SB4 events. The details of each event are as follows:

Event SB1: A single agent at site i places a daughter agent at site $i - 1$, where the gap that includes site $i - 1$ is greater than length one. C_i^1 and G_j^{i-j} decrease, C_{i-1}^2 and G_j^{i-j-1} increase (figure 1(b)).

Event SB2: A single agent at site i places a daughter agent at site $i + 1$, where the gap that includes site $i + 1$ is greater than length one. C_i^1 and G_{i+1}^j decrease, C_i^2 and G_{i+2}^{j-1} increase (figure 1(c)).

Event SB3: A single agent at site i places a daughter agent at site $i - 1$, where the gap that includes site $i - 1$ is of length one. C_i^1, G_{i-1}^1 and C_j^{i-j-1} decrease, C_j^{i-j+1} increases (figure 1(d)).

Event SB4: A single agent at site i places a daughter agent at site $i + 1$, where the gap that includes site $i + 1$ is of length one. C_i^1, G_{i+1}^1 and C_{i+2}^j decrease, C_i^{j+2} increases (figure 1(e)).

Event EB1: An edge agent at site i places a daughter agent at site $i - 1$, where the gap that includes site $i - 1$ is greater than length one. C_i^j and G_k^{i-k} decrease, C_{i-1}^{j+1} and G_k^{i-k-1} increase, where $j \geqslant 2$ (figure 1(f)).

Event EB2: An edge agent at site i places a daughter agent at site $i + 1$, where the gap that includes site $i + 1$ is greater than length one. C_j^{i-j+1} and G_{i+1}^k decrease, C_j^{i-j+2} and G_{i+2}^{k-1} increase, where $j \geqslant 2$ (figure 1(g)).

Event EB3: An edge agent at site i places a daughter agent at site $i - 1$, where the gap that includes site $i - 1$ is of length one. C_i^j, G_{i-1}^1 and C_k^{i-k-1} decrease, C_k^{i+j-k} increases, where $j \geqslant 2$ (figure 1(h)).

Event EB4: An edge agent at site i places a daughter agent at site $i + 1$, where the gap that includes site $i + 1$ is of length one. C_j^{i-j+1}, G_{i+1}^1 and C_{i+2}^k decrease, $C_j^{i-j+k+2}$ increases, where $j \geqslant 2$ (figure 1(i)).

Event SM1: A single agent at site i moves to site $i - 1$, where the gap that includes site $i - 1$ is greater than length one. C_i^1, G_j^{i-j} and G_{i+1}^k decrease, C_{i-1}^1, G_j^{i-j-1} and G_i^{k+1} increase (figure 1(j)).

Event SM2: A single agent at site i moves to site $i + 1$, where the gap that includes site $i + 1$ is greater than length one. C_i^1, G_j^{i-j} and G_{i+1}^k decrease, C_{i+1}^1, G_j^{i-j+1} and G_{i+2}^{k-1} increase (figure 1(k)).

Event SM3: A single agent at site i moves to site $i - 1$, where the gap that includes site $i - 1$ is of length one. C_i^1, G_{i-1}^1, G_{i+1}^j and C_k^{i-k-1} decrease, C_k^{i-k} and G_i^{j+1} increase (figure 1(l)).

Event SM4: A single agent at site i moves to site $i + 1$, where the gap that includes site $i + 1$ is of length one. C_i^1, G_{i+1}^1, G_j^{i-j} and C_{i+2}^k decrease, C_{i+1}^{k+1} and G_j^{i-j+1} increase (figure 1(m)).

Event EM1: An edge agent at site i moves to site $i - 1$, where the gap that includes site $i - 1$ is greater than length one. C_i^j and G_k^{i-k} decrease, C_{i-1}^1, G_i^1, C_{i+1}^{j-1} and G_k^{i-k-1} increase, where $j \geqslant 2$ (figure 1(n)).

Event EM2: An edge agent at site i moves to site $i + 1$, where the gap that includes site $i + 1$ is greater than length one. C_j^{i-j+1} and G_{i+1}^k decrease, C_{i+1}^1, G_i^1, C_j^{i-j} and G_{i+2}^{k-1} increase, where $j \geqslant 2$ (figure 1(o)).

Event EM3: An edge agent at site i moves to site $i - 1$, where the gap that includes site $i - 1$ is of length one. C_i^j, G_{i-1}^1 and C_k^{i-k-1} decrease, C_{i+1}^{j-1}, G_i^1 and C_k^{i-k} increase, where $j \geqslant 2$ (figure 1(p)).

Event EM4: An edge agent at site i moves to site $i + 1$, where the gap that includes site $i + 1$ is of length one. C_j^{i-j+1}, G_{i+1}^1 and C_{i+2}^k decrease, C_j^{i-j}, G_i^1 and C_{i+1}^{k+1} increase, where $j \geqslant 2$ (figure 1(q)).

Event SD1: A single agent at site i dies. C_i^1, G_j^{i-j} and G_{i+1}^k decrease, $G_j^{i+k-j+1}$ increases (figure 1(r)).

Event MD1: A middle agent at site i dies. C_j^k decreases, C_j^{i-j}, $C_{i+1}^{j+k-i-1}$ and G_i^1 increase, where $j \geqslant 3$ (figure 1(s)).

Event ED1: An edge agent at site i dies, where site $i + 1$ is occupied and site $i - 1$ is vacant. C_i^j and G_k^{i-k} decrease, C_{i+1}^{j-1} and G_k^{i-k+1} increase, where $j \geqslant 2$ (figure 1(t)).

Event ED2: An edge agent at site i dies, where site $i - 1$ is occupied and site $i + 1$ is vacant. C_j^{i-j+1} and G_{i+1}^k decrease, C_j^{i-j} and G_i^{k+1} increase, where $j \geqslant 2$ (figure 1(u)).

To obtain expressions for the time rate of change of C_i^n and G_i^m, we consider the rate at which each event occurs at site i, and all possible results of each event. Birth events are always successful for single agents and, as such, occur at rate $P_p^1 C_i^1$. Similarly, movement events for single agents are always successful and are not influenced by agent–agent adhesion/repulsion. Therefore single agent movement events occur at rate $P_m^1 C_i^1$. Birth events for edge agents are, on average, unsuccessful half the time due to crowding. Hence, birth events occur at rates $P_p^n C_i^n / 2$ for $n \in [2, N]$. Movement events for edge agents are influenced by both crowding and agent–agent adhesion/repulsion, and occur at rates $P_m^n C_i^n (1 - \alpha)/2$ for $n \in [2, N]$. Neither birth or movement events can occur for middle agents. For all agent types, death events are not influenced by crowding and occur at rates $P_d^n C_i^n$ for $n \in [1, N]$. Note that all rates are chain-length dependent. Practical examples that could be described using chain-length dependent rates arise in a variety of situations [28, 37]. For example, Hedayati *et al* [37] demonstrate that nanoparticle-mediated heating causes cytotoxicity in prostate cancer cells, provided the volume of cells is above a threshold value. Furthermore, the cytotoxicity increases with the cell number. Our framework would be suitable for modelling this process, as we are able to impose death rates that are zero below a threshold length, and an increasing function with respect to length otherwise.

While the rates at which events occur for each mechanism and agent type combination are known, multiple events can occur for a mechanism and agent type combination. For example, there are four types of birth events for single agents and, as such, we require the proportion of single birth events that are SB1, SB2, SB3 and SB4

events. For a single birth event to be an SB1 event, the agent at site i must place a daughter agent at site $i - 1$ and the gap that contains site $i - 1$ must be of length two or greater. Note that the gap cannot include site i, as the selected agent occupies that site. The proportion of single birth events where the agent at site i selects a target site at $i - 1$ is $1/2$. The proportion of single birth events where the agent selects a target site that is part of a gap of length two or greater is $\sum_{j=1}^{i-2} G_j^{i-j} / \sum_{j=1}^{i-1} G_j^{i-j}$, that is, the number of gaps including site $i - 1$ that are of length two or greater divided by the total number of gaps including site $i - 1$. Hence the rate at which SB1 events occur at site i, and subsequently decrease C_i^1 and increase C_{i-1}^2, is $(1/2) P_p^1 C_i^1 \sum_{j=1}^{i-2} G_j^{i-j} / \sum_{j=1}^{i-1} G_j^{i-j}$. SB1 events also decrease G_j^{i-j} and increase G_j^{i-j-1}, where $j \in [1, i - 2]$. The proportion of SB1 events that change G_j^{i-j} and G_j^{i-j-1} for a specific $j \in [1, i - 2]$ is $(1/2) P_p^1 C_i^1 G_j^{i-j} / \sum_{k=1}^{i-1} G_k^{i-k}$. Therefore, we can determine the expected rate of change for all chains and gaps affected by an SB1 event at site i. Following a similar process for all events we obtain transition rates for C_i^n and G_i^m, where $i \in [1, N]$, $n \in [1, N - i + 1]$ and $m \in [1, N - i + 1]$. The resulting system of ODEs is presented in appendix A.1 for C_i^n and appendix A.2 for G_i^m.

Traditional MF descriptions

Traditional MF descriptions of lattice-based random walk models containing crowding, birth, death, movement and agent–agent adhesion do not have the flexibility to describe processes where the rate of birth, death and/or movement is arbitrarily chain-length dependent. However, with certain simplifying assumptions, MF descriptions of the discrete process can be derived [20, 38]. Here we examine two special cases of the discrete process, where continuum MF descriptions have been presented previously.

Special case 1: One simplifying assumption is that the birth, death and movement rates are independent of chain length. Hence we define $P_p = P_p^1 = P_p^2 = \ldots = P_p^N$, $P_d = P_d^1 = P_d^2 = \ldots = P_d^N$, and $P_m = P_m^1 = P_m^2 = \ldots = P_m^N$. There is, therefore, no positive or negative benefit associated with an agent being near other agents. Furthermore, there is no agent–agent adhesion/repulsion, and hence $\alpha = 0$. The MF description for this case takes the form of a reaction-diffusion equation, known as the Fisher–Kolmogorov model [20, 38–40],

$$\frac{\partial \rho(x, t)}{\partial t} = D \frac{\partial^2 \rho(x, t)}{\partial x^2} + (\lambda - K) \rho(x, t)(1 - \rho(x, t)), \tag{1}$$

where

$$D = \lim_{\Delta \to 0} \frac{P_m \Delta^2}{2}, \; \lambda = P_p, \; K = P_d,$$

and $\rho(x, t)$ is the agent density [20, 38].

Special case 2: An alternative simplifying assumption is that birth, death and movement rates depend on whether an agent is isolated or not. An isolated agent has zero nearest-neighbours [38], corresponding to a chain of length one. All grouped agents, that is, agents with at least one nearest-neighbour, undergo birth, death and movement events at the same rate. Grouped agents correspond to agents that are part of a chain of length two or greater. Hence we define $P_p^G = P_p^2 = \ldots = P_p^N$, $P_d^G = P_d^2 = \ldots = P_d^N$, and $P_m^G = P_m^2 = \ldots = P_m^N$. This assumption introduces a potential positive or negative benefit associated with an agent being adjacent to other agents. For example, if $P_d^1 > P_d^G$ then isolated agents are more likely to die, compared to other agents, and hence there is a positive benefit associated with being adjacent to other agents. This allows for significant flexibility in modelling a variety of competitive or co-operative processes [38]. Again, there is no agent–agent adhesion/ repulsion, and $\alpha = 0$. The MF description for this case is a reaction-diffusion equation with a nonlinear diffusivity function and Allee kinetics [38],

$$\frac{\partial \rho(x, t)}{\partial t} = \frac{\partial}{\partial x}\left(F(\rho(x, t)) \frac{\partial \rho(x, t)}{\partial x}\right) + R(\rho(x, t)), \tag{2}$$

where

$$F(\rho(x, t)) = D_I(1 - 4\rho(x, t) + 3\rho(x, t)^2) + D_G(4\rho(x, t) - 3\rho(x, t)^2),$$
$$R(\rho(x, t)) = (\lambda_I - \lambda_G - K_I + K_G)\rho(x, t)(1 - \rho(x, t))^2$$
$$+ \lambda_G \rho(x, t)(1 - \rho(x, t)) - K_G \rho(x, t),$$

and

$$D_I = \lim_{\Delta \to 0} \frac{P_m^1 \Delta^2}{2}, \; \lambda_I = P_p^1, \; K_I = P_d^1,$$
$$D_G = \lim_{\Delta \to 0} \frac{P_m^G \Delta^2}{2}, \; \lambda_G = P_p^G, \; K_G = P_d^G.$$

Figure 2. (a)–(c) Twenty identically prepared realisations of the one-dimensional discrete model at (a) $t = 0$, (b) $t = 100$, and (c) $t = 200$, as indicated. (d) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and *C&G* description (cyan, solid). Note that the cyan and red curves are visually indistinguishable at this scale. All results are obtained using $P_m^i = 1$, $P_p^i = 0.005$, $P_d^i = 0$ for $i \in [1, 200]$, $N = 200$, $M = 10^3$ and $\Delta = 1$. These parameter choices correspond to special case 1. Profiles are presented at $t = 100$ and $t = 200$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

While CMF descriptions have been presented for certain lattice-based random walk models, these descriptions are unsuitable for studying problems containing moving fronts as the system of governing ODEs is singular at zero agent density [22, 24, 26].
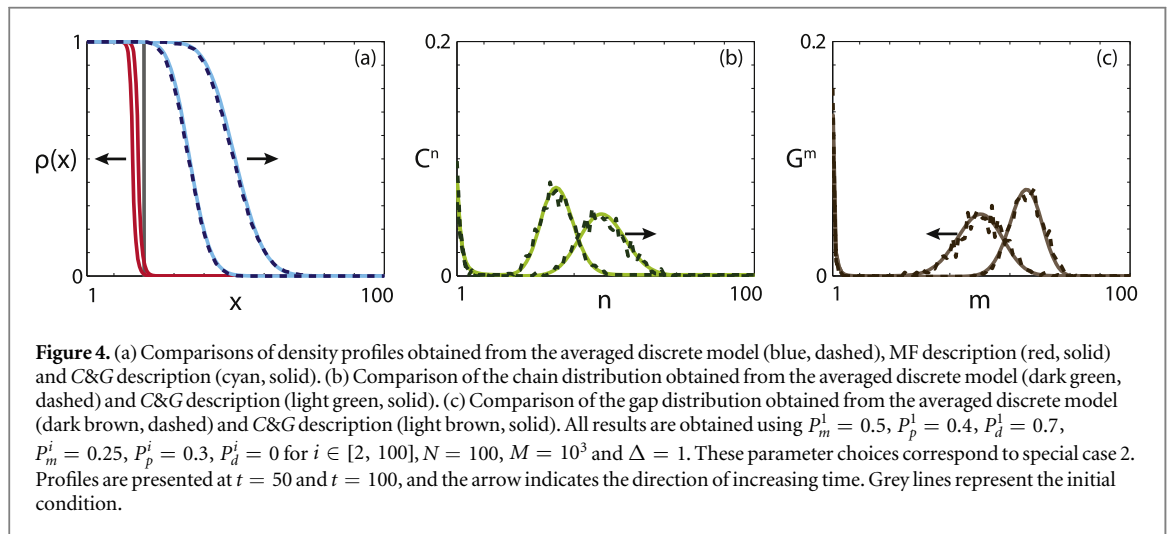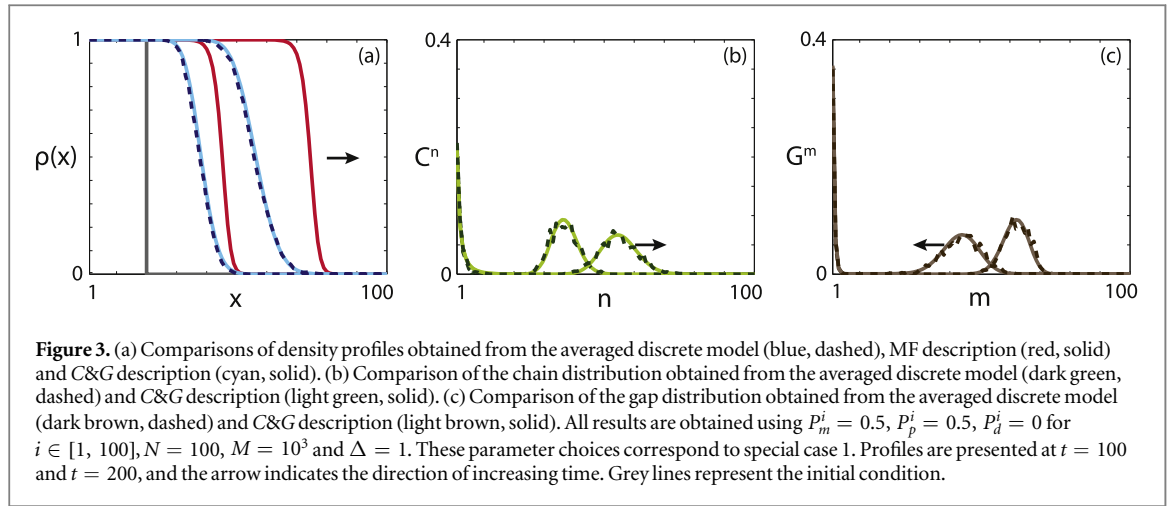
Results

The solution of the MF model leads to a prediction of the average agent density profile as a function of position and time, whereas the *C&G* description provides the number of chains and gaps of all possible lengths as a function of location and time. Hence to compare the MF descriptions with the *C&G* description, it is necessary to reconstruct the agent density at each location, ρ_i, from C_i^n and G_i^m. The agent density at site i is the sum of all possible chains that include site i, namely,

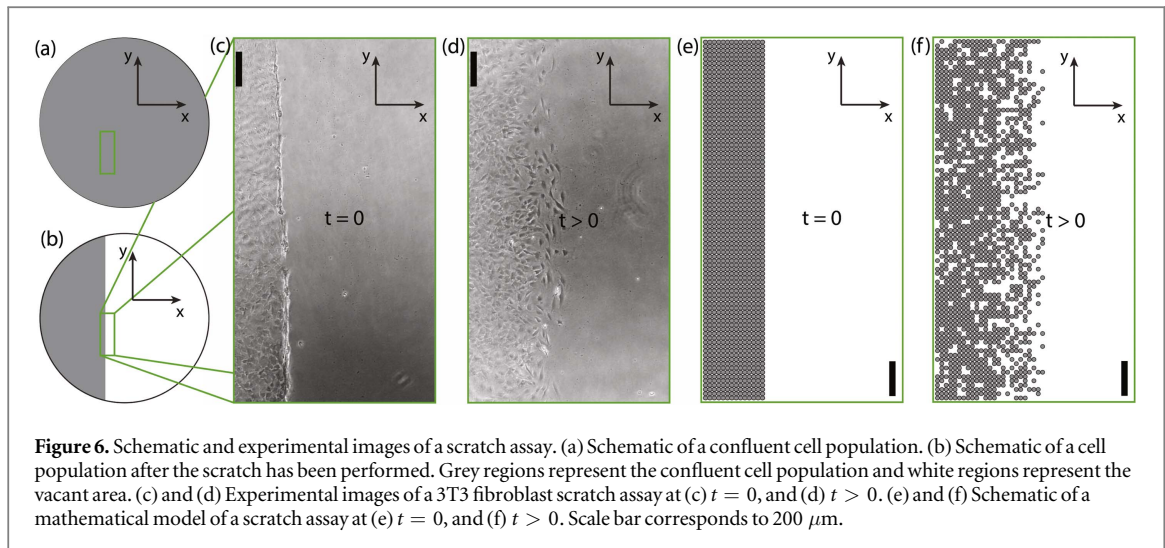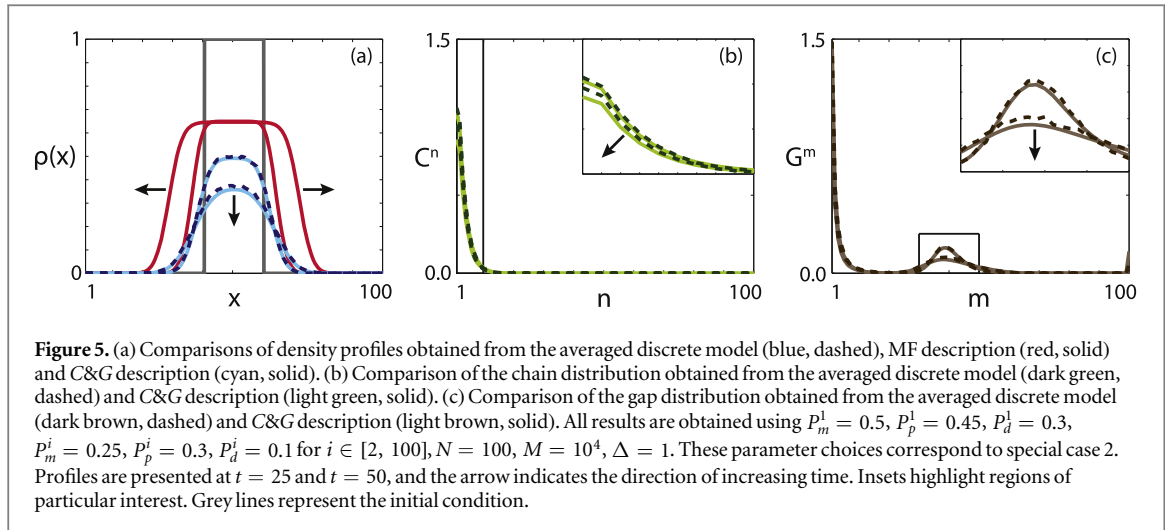$$\rho_i(t) = \sum_{j=1}^{i} \sum_{k=i-j+1}^{N-j+1} C_j^k.$$ (3)

An example of the output from the one-dimensional discrete model, illustrating twenty identically prepared realisations, is presented in figures 2(a)–(c) at $t = 0$, $t = 100$ and $t = 200$, respectively. Identically prepared realisations refer to simulations of the discrete model performed with the same initial condition, parameter regime and boundary conditions. Initially, the domain is fully occupied for $61 \leqslant x \leqslant 140$, and vacant otherwise. As time increases, the population spreads into the initially vacant region. Note that figures 2(a)–(c) each show twenty one-dimensional simulations, rather than one two-dimensional simulation. To obtain the average behaviour of the agent population, M identically prepared realisations of the discrete model are performed and the binary lattice occupancy at each site i, $\hat{\rho}_i$, is calculated for each realisation. Note that the discrete model is simulated with the Gillespie algorithm [41]. The binary lattice occupancy is then averaged, giving $\rho_i = (1/M)\sum_M \hat{\rho}_i$. The averaged density profile from the discrete model is presented in figure 2(d), with the numerical solutions to both equation (1) and the *C&G* governing equations superimposed, at $t = 100$ and $t = 200$. Both continuum descriptions match the averaged discrete model predictions extremely well. Details of the numerical techniques used to solve the PDEs and systems of ODEs are presented in appendix B.

The parameter regime considered in figure 2 has $P_m/P_p \gg 1$ and $P_d = 0$, and, as such, we expect the MF description to approximate the average discrete behaviour well since this combination of parameters avoids the formation of significant agent clustering [20, 22]. We now consider a parameter regime where P_m/P_p is $\mathcal{O}(1)$. In such parameter regimes, the spatial correlations are significant and, subsequently, the MF description does not provide a valid approximation of the average discrete behaviour [22, 28]. To highlight this, a comparison between the average discrete behaviour and the numerical solution of equation (1) is presented in figure 3(a). For these results, the domain is initially fully occupied for $x \leqslant 20$, and is vacant otherwise. The MF description predicts a front that is significantly sharper, and has a higher front speed, compared to the discrete model. In contrast, the numerical solution of the *C&G* description predicts the average discrete behaviour well, matching both the shape and position of the averaged discrete data. Furthermore, the distribution of chain and gap lengths, C^n and G^m, matches the observed average distribution of chains and gaps in the discrete model, as shown in figures 3(b) and (c). As such, the *C&G* description provides an accurate estimate of the front shape and speed, as well as a valid prediction of the clustering of occupied and vacant sites in the system.

Figure 3. (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and *C&G* description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and *C&G* description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and *C&G* description (light brown, solid). All results are obtained using $P_m^i = 0.5$, $P_p^i = 0.5$, $P_d^i = 0$ for $i \in [1, 100]$, $N = 100$, $M = 10^3$ and $\Delta = 1$. These parameter choices correspond to special case 1. Profiles are presented at $t = 100$ and $t = 200$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

Figure 4. (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and *C&G* description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and *C&G* description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and *C&G* description (light brown, solid). All results are obtained using $P_m^1 = 0.5$, $P_p^1 = 0.4$, $P_d^1 = 0.7$, $P_m^i = 0.25$, $P_p^i = 0.3$, $P_d^i = 0$ for $i \in [2, 100]$, $N = 100$, $M = 10^3$ and $\Delta = 1$. These parameter choices correspond to special case 2. Profiles are presented at $t = 50$ and $t = 100$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

If the birth, death and movement rates depend on whether an agent has zero or at least one nearest-neighbour agent then the MF description of the discrete model is equation (2) [38]. A comparison of the average discrete behaviour, the numerical solution of equation (2) and the numerical solution of the *C&G* description in an appropriate parameter regime is presented in figure 4(a). Interestingly, the MF description predicts that the agent population moves in the negative *x* direction, and would subsequently become extinct. In contrast, both the discrete model and the *C&G* description suggest that the agent population persists, and spreads in the positive *x* direction. Again, the *C&G* description matches the average discrete behaviour extremely well. The results in figure 4(a) highlight the need for an accurate approximation. If the naive approach of implementing a MF approximation to describe the spread of an invasive species is taken, it might be recommended that no culling measures are required to curtail the spread of the species. Obviously, such a recommendation is incorrect if the aim is to halt the invasion of the alien species. The clustering present in the system is highlighted in figures 4(b) and (c), for chains and gaps, respectively. Intuitively, as time increases and the population spreads, the average chain length increases and the average gap length decreases. The *C&G* description predicts both the average chain and gap distributions in the discrete model well.

Introducing a non-zero death rate for chains of length two or greater reduces the carrying capacity in the MF description of the discrete model [38]. To determine whether this reduction is an accurate reflection of the average discrete behaviour, we present a comparison of the average discrete behaviour, and the numerical solutions of both equation (2) and the *C&G* governing equations in figure 5(a). Both the discrete model and the *C&G* description predict that the peak agent density near $x = 50$ decreases between $t = 25$ and $t = 50$, whereas the MF description predicts that the peak agent density at this location is approximately constant, at $\rho = 0.646$ [38], after $t = 25$. Interestingly, both the discrete model and the *C&G* description predict that the population eventually goes extinct. In contrast, the MF description predicts that the population persists and spreads

Figure 5. (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and *C&G* description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and *C&G* description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and *C&G* description (light brown, solid). All results are obtained using $P_m^1 = 0.5$, $P_p^1 = 0.45$, $P_d^1 = 0.3$, $P_m^i = 0.25$, $P_p^i = 0.3$, $P_d^i = 0.1$ for $i \in [2, 100]$, $N = 100$, $M = 10^4$, $\Delta = 1$. These parameter choices correspond to special case 2. Profiles are presented at $t = 25$ and $t = 50$, and the arrow indicates the direction of increasing time. Insets highlight regions of particular interest. Grey lines represent the initial condition.

Figure 6. Schematic and experimental images of a scratch assay. (a) Schematic of a confluent cell population. (b) Schematic of a cell population after the scratch has been performed. Grey regions represent the confluent cell population and white regions represent the vacant area. (c) and (d) Experimental images of a 3T3 fibroblast scratch assay at (c) $t = 0$, and (d) $t > 0$. (e) and (f) Schematic of a mathematical model of a scratch assay at (e) $t = 0$, and (f) $t > 0$. Scale bar corresponds to 200 μm.
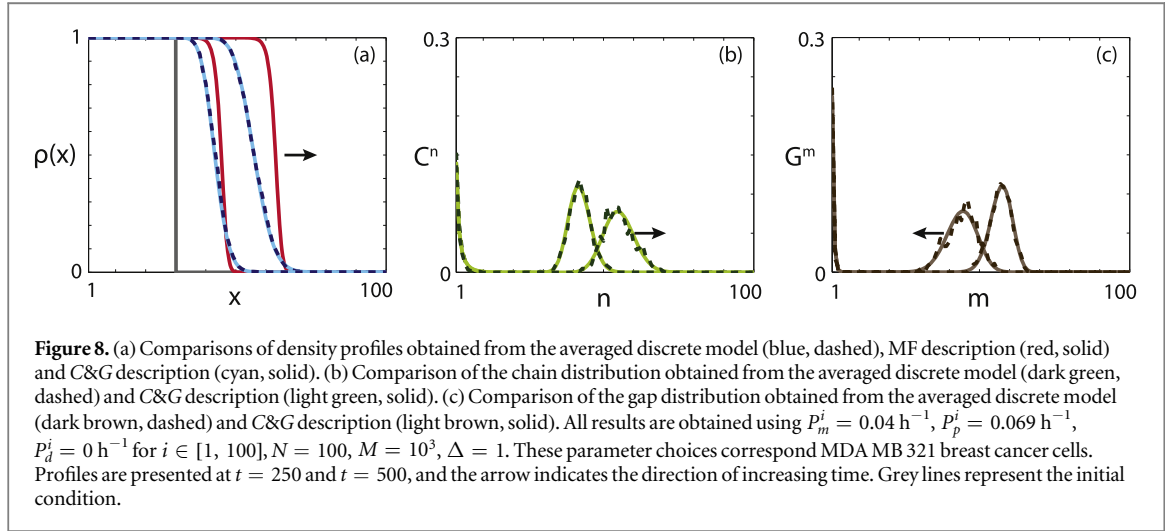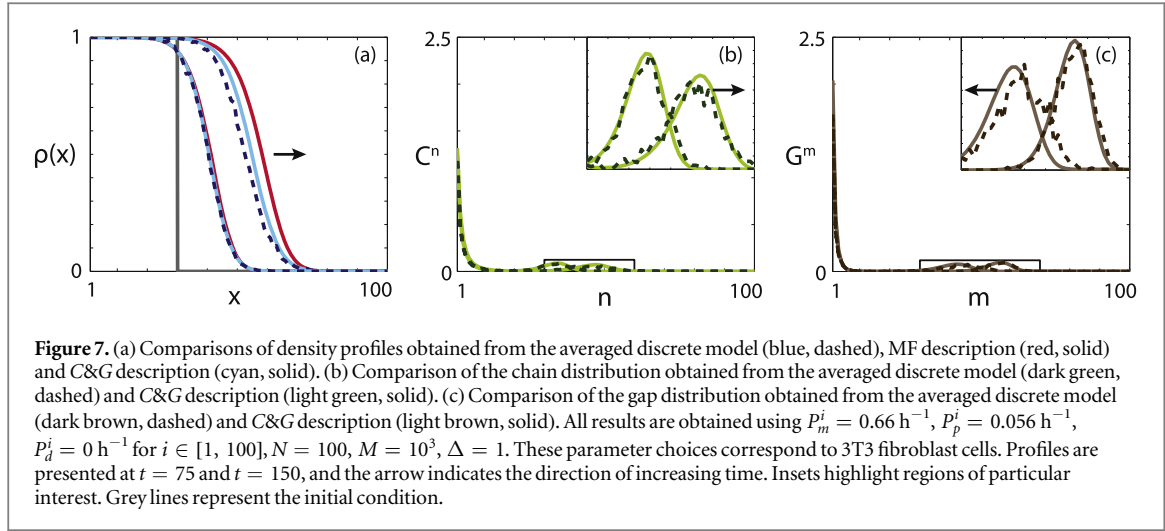
throughout the domain. Again, these results highlight the importance of implementing an accurate approximation to obtain meaningful conclusions, as well as the robust nature of the *C&G* description.

Note that all results presented here have been performed with $\alpha = 0$, and hence no agent–agent adhesion/repulsion. Simulations performed with $\alpha \neq 0$ (not presented) confirm that the *C&G* description accurately predicts the average discrete behaviour in all cases, even with strong agent–agent adhesion/repulsion.

Experimental case study

To highlight the insight provided about spatial clustering by the *C&G* description, we consider a case study motivated by a scratch assay. Scratch assays are widely used to observe the collective behaviour of a cell population in response to a model wound [42]. We present a schematic representation and experimental images of a scratch assay in figure 6. In a scratch assay, a cell population is placed on a dish and allowed to grow to confluence. A portion of the cell population is then removed, and the remaining cells spread, through a combination of migration and proliferation, into the newly vacant space [42]. A schematic representation of the confluent population before and after the scratch is performed is presented in figures 6(a) and (b), with a typical experimental field of view highlighted in green. To mimic the geometry of this experiment, we consider $1 \leqslant x \leqslant 100$ and initially set $\rho(x) = 1$ for $x \leqslant 30$ and $\rho(x) = 0$, otherwise. Note that scratch assays are two-dimensional processes, as observed in the experimental images of a scratch assay for a 3T3 fibroblast population in figures 6(c) and (d), and the corresponding schematics for this experiment in figures 6(e) and (f). Full experimental details are given in [6].

Figure 7. (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and *C&G* description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and *C&G* description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and *C&G* description (light brown, solid). All results are obtained using $P_m^i = 0.66\ \text{h}^{-1}$, $P_p^i = 0.056\ \text{h}^{-1}$, $P_d^i = 0\ \text{h}^{-1}$ for $i \in [1, 100]$, $N = 100$, $M = 10^3$, $\Delta = 1$. These parameter choices correspond to 3T3 fibroblast cells. Profiles are presented at $t = 75$ and $t = 150$, and the arrow indicates the direction of increasing time. Insets highlight regions of particular interest. Grey lines represent the initial condition.

Figure 8. (a) Comparisons of density profiles obtained from the averaged discrete model (blue, dashed), MF description (red, solid) and *C&G* description (cyan, solid). (b) Comparison of the chain distribution obtained from the averaged discrete model (dark green, dashed) and *C&G* description (light green, solid). (c) Comparison of the gap distribution obtained from the averaged discrete model (dark brown, dashed) and *C&G* description (light brown, solid). All results are obtained using $P_m^i = 0.04\ \text{h}^{-1}$, $P_p^i = 0.069\ \text{h}^{-1}$, $P_d^i = 0\ \text{h}^{-1}$ for $i \in [1, 100]$, $N = 100$, $M = 10^3$, $\Delta = 1$. These parameter choices correspond MDA MB 321 breast cancer cells. Profiles are presented at $t = 250$ and $t = 500$, and the arrow indicates the direction of increasing time. Grey lines represent the initial condition.

As the experiment is approximately spatially uniform in one direction, we can approximate the scratch assay with a one-dimensional model [43]. We consider two scratch assays performed with two different cell lines where parameter estimates for the cell motility and cell proliferation rates have been presented previously: 3T3 fibroblast cells (3T3 cells) and MDA MB 231 breast cancer cells (231 cells) [44]. The investigation performed by Simpson *et al* [44] resulted in parameter estimates of $P_p^i = 0.056\ \text{h}^{-1}$, $P_m^i = 0.66$ h^{-1} and $P_d^i = 0\ \text{h}^{-1}$ for all i for 3T3 cells, and $P_p^i = 0.069\ \text{h}^{-1}$, $P_m^i = 0.04\ \text{h}^{-1}$, $P_d^i = 0\ \text{h}^{-1}$ for all i for 231 cells. Note that the ratio P_p^i/P_m^i is approximately one order of magnitude higher for the breast cancer cells compared to the fibroblasts, which implies that the spatial correlations between breast cancer cells will be more significant [22].

For the numerical solution corresponding to the 3T3 cell population, presented in figure 7, both the MF description and the *C&G* description approximate the average discrete behaviour reasonably well. However, the *C&G* description provides additional information regarding the clustering present within the migrating cell population. The chain distribution, presented in figure 7(b), suggests that the 3T3 population does not form significant clusters, as the majority of the chains are short length.

In contrast, for the numerical solution corresponding to the 231 cell population, presented in figure 8, the *C&G* description accurately approximates the average discrete behaviour, while the MF description does not. This is result is intuitive as we observe that there is significantly more clustering present in the system, compared to the numerical solution corresponding to the 3T3 cell population. That is, the chain distribution in figure 8(b) contains significantly fewer chains of short length, compared to the chain distribution in figure 7(b). For example, at the times shown, C^1 is approximately 25 times higher in the 3T3 cell population, compared to the 231 cell population. Critically, the *C&G* description accurately predicts the experimental observation that 231 cells form clusters while 3T3 cells do not [44].

Table 1. Time in seconds taken to perform: (i) a single realisation of the discrete model; (ii) 1000 realisations of the discrete model, and (iii) a numerical solution of the *C&G* system of equations for the parameter values in figures 2–7. All solutions are obtained using a single 3.0 GHz Intel i7-3540 M desktop processor. Note that the computation time for 1000 realisations is lower than performing 1000 repeats of a single realisation due to the time associated with initial set-up.

	Figure 2	Figure 3	Figure 4	Figure 5	Figure 6	Figure 7
Discrete model (single realisation)	3.42	1.13	0.22	0.11	0.37	0.27
Discrete model (1000 realisations)	2838.1	954.4	134.1	39.1	313.6	177.9
C&G description	601.3	52.3	25.2	34.7	27.1	27.6

Specifically, during monolayer formation, 3T3 cells form an approximately spatially uniform monolayer while 231 cells develop into clusters [44].

Discussion and conclusions

Processes that involve moving fronts are common in cell biology and ecology [1, 2, 4–7, 9–12], and lattice-based random walks are widely employed to describe these processes [6, 16, 18–21, 45]. Due to the stochastic nature of random walks, it can be computationally intractable to perform sufficiently many realisations of a random walk model to obtain average behaviour that is not dominated by fluctuations. Furthermore, it is difficult to extract meaningful information about population behaviour through analysis of the random walk. The standard approach to overcome these issues is to derive a MF description of the random walk [19, 20, 25]. However, this approach relies on the assumption that any spatial correlations within the random walk are weak [22, 24]. CMF descriptions that account for the spatial correlations have been proposed [22, 24, 46]. Unfortunately, these descriptions are not applicable to problems involving moving fronts as the ODEs governing the CMF description are singular in regions where the density of agents is zero [22, 24].

Here we develop and present an accurate continuum description for moving fronts associated with lattice-based random walks that contain crowding, birth, death, movement and agent–agent adhesion. We consider processes that are spatially variable, and include birth, death and movement rates that are chain-length dependent. Our *C&G* description provides predictions that match the average behaviour of the discrete model well in all parameter regimes. In contrast, the MF description is less flexible in terms of the birth, death and movement rates and only provides a valid approximation of the average discrete behaviour in extremely limited parameter regimes. Furthermore, for all cases considered in this work, the *C&G* description requires less computation time than performing 1000 realisations of the discrete model. A comparison between the time taken to perform a single realisation of the discrete model, 1000 realisations of the discrete model, and to obtain the numerical solution to the *C&G* system of equations is presented in table 1.

For the special case where the rates of birth, death and movement are independent of the chain length, the MF description correctly predicts the persistence of the population but inaccurately predicts the front velocity. For the special case where the rates of birth, death and movement are different depending on whether the agents are part of a chain of length one, or are part of a chain of length two or greater, the MF description predicts persistence when the population becomes extinct, and predicts extinction when the population persists. In both these cases the *C&G* description accurately predicts the front velocity, and the persistence or extinction of the population. Furthermore, the *C&G* description provides information about the spatial clustering of both occupied and unoccupied sites, and the clustering predictions approximate the clustering observed in the discrete model accurately.

The work presented here could be extended in several ways. The influence of local spatial structure on the persistence of species is a key question in ecology [34, 35, 45, 47]. The *C&G* description provides an explicit estimate of the spatial clustering of both agents and unoccupied space. Therefore, it would be instructive to apply the *C&G* description to ecological processes to obtain insight into the clustering present in the system for parameter regimes where the agent population becomes extinct. Another approach would be to investigate a truncated system of governing equations, where there is a maximum chain or gap length. If there is prior knowledge about the long-time density of the system, an assumption could be made that chains or gaps above a threshold length could be neglected, and hence the system of equations could be truncated. This truncation would reduce the computational cost associated with obtaining a numerical solution to the governing equations. It would be insightful to examine the trade-off between the reduction in computational cost and the decreased accuracy caused by the truncation. Alternatively, the *C&G* description presented here could be calibrated to experimental data from the cell biology literature. For example, lattice-based random walks have been calibrated to *in vitro* cell biology

experiments to obtain estimates of cell diffusivity and cell proliferation rates [6, 21]. However, the calibration of stochastic models to experimental data is computationally expensive [6, 48]. As the *C&G* description accurately approximates the average random walk behaviour in all parameter regimes, it would be instructive to determine whether similar cell diffusivity and cell proliferation rates could be obtained from calibration of the deterministic *C&G* description to experimental data, and to quantify the reduction in computation time to obtain the parameter estimates. However, we leave these extensions for future consideration.

Acknowledgments

Appendix A. Governing equations

A.1. Chains
Here we present the time rate of change for each chain length and location, obtained by considering the potential outcomes of each type of birth, death and movement event.

$$
\begin{aligned}
\frac{dC_1^1}{dt} &= \frac{1}{2}\left[-P_p^1 C_1^1 - \sum_{i=1}^{N-2} P_p^1 Y(C_3^i, G_2^1, C_1^1)\right] + \left[\sum_{i=2}^{N}\{P_d^i C_1^1\} - P_d^1 C_1^1\right] \\
&\quad + \frac{P_m^1}{2}[-C_1^1 + C_2^1 - Y(C_3^1, G_2^1, C_1^1)] \\
&\quad + \frac{1-\alpha}{2}\left[P_m^2 C_1^2 + \sum_{i=2}^{N-1}\{P_m^i C_2^i\} - \sum_{i=2}^{N-2} P_m^i Y(C_3^i, G_2^1, C_1^1)\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{dC_2^1}{dt} &= \frac{1}{2}\left[-2P_p^1 C_2^1 - \sum_{i=1}^{N-3} P_p^i Y(C_4^i, G_3^1, C_2^1)\right] + \left[\sum_{i=2}^{N-1}\{P_d^i C_2^i\} + P_d^2 C_1^2 - P_d^1 C_2^1\right] \\
&\quad + \frac{P_m^1}{2}[C_1^1 - Y(C_1^1, G_2^1) - 2C_2^1 + C_3^1 - Y(C_3^1, G_2^1) - Y(C_4^1, G_3^1, C_2^1)] \\
&\quad + \frac{1-\alpha}{2}\left[P_m^2 C_2^2 + \sum_{i=2}^{N-2}\{P_m^i C_3^i - P_m^i Y(C_3^i, G_2^1)\} - \sum_{i=2}^{N-3} P_m^i Y(C_4^i, G_3^1, C_2^1)\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{dC_k^1}{dt} &= \frac{1}{2}\left[-2P_p^1 C_k^1 - \sum_{i=1}^{N-k-1} P_p^i Y(C_{k+2}^i, G_{k+1}^1, C_k^1) - \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^1)\right] \\
&\quad + \left[\sum_{i=2}^{k+1}\{P_d^i C_k^i\} + \sum_{i=2}^{k}\{P_d^i C_{k-i+1}^i\} - P_d^1 C_k^1\right] \\
&\quad + \frac{P_m^1}{2}[-2C_k^1 + C_{k-1}^1 - Y(C_{k-1}^1, G_1^1) + C_{k+1}^1 - Y(C_{k+1}^1, G_k^1) \\
&\quad - Y(C_{k-2}^1, G_{k-1}^1, C_k^1) - Y(C_{k+2}^1, G_{k+1}^1, C_k^1)] \\
&\quad + \frac{1-\alpha}{2}\left[P_m^2 C_k^2 + P_m^2 C_{k-1}^2 + \sum_{i=2}^{k-1}\{P_m^i C_{k-i}^i - P_m^i Y(C_{k-i}^i, G_k^1)\}\right. \\
&\quad + \sum_{i=2}^{N-k}\{P_m^i C_{k+1}^i - P_m^i Y(C_{k+1}^i, G_k^1)\} - \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^1) \\
&\quad \left. - \sum_{i=2}^{N-k-1} P_m^i Y(C_{k+2}^i, G_{k+1}^1, C_k^1)\right], \quad k = 3, \ldots, N-2,
\end{aligned}
$$

$$\frac{dC_{N-1}^1}{dt} = \frac{1}{2}\left[-2P_p^1 C_{N-1}^1 - \sum_{i=1}^{N-3} P_p^i Y(C_{N-i-2}^i, G_{N-2}^1, C_{N-1}^1) \right]$$

$$+ \left[P_d^2 C_{N-1}^2 + \sum_{i=2}^{N-1} \{P_d^i C_{N-i}^i\} - P_d^1 C_{N-1}^1 \right]$$

$$+ \frac{P_m^1}{2}[C_N^1 - Y(C_N^1, G_{N-1}^1) - 2C_{N-1}^1 + C_{N-2}^1 - Y(C_{N-2}^1, G_{N-1}^1)$$

$$- Y(C_{N-3}^1, G_{N-2}^1, C_{N-1}^1)]$$

$$+ \frac{1-\alpha}{2}\left[P_m^2 C_{N-2}^2 + \sum_{i=2}^{N-2} P_m^i \{C_{N-i-1}^i - Y(C_{N-i-1}^i, G_{N-1}^1)\} \right]$$

$$- \sum_{i=2}^{N-3} P_m^i Y(C_{N-i-2}^i, G_{N-2}^1, C_{N-1}^1) \Bigg],$$

$$\frac{dC_N^1}{dt} = \frac{1}{2}\left[-P_p^1 C_N^1 - \sum_{i=1}^{N-2} P_p^i Y(C_{N-i-1}^i, G_{N-1}^1, C_N^1) \right] + \left[\sum_{i=2}^{N} P_d^i \{C_{N-i+1}^i\} - P_d^1 C_N^1 \right]$$

$$+ \frac{P_m^1}{2}[-C_N^1 + C_{N-1}^1 - Y(C_{N-2}^1, G_{N-1}^1, C_N^1)]$$

$$+ \frac{1-\alpha}{2}\left[P_m^2 C_{N-1}^2 + \sum_{i=2}^{N-1} P_m^i \{C_{N-i}^i\} - \sum_{i=2}^{N-2} P_m^i Y(C_{N-i-1}^i, G_{N-1}^1, C_N^1) \right],$$

$$\frac{dC_1^L}{dt} = \frac{1}{2}\left[P_p^{L-1} C_1^{L-1} + P_p^{L-1} C_2^{L-1} - P_p^{L-1} Y(C_1^{L-1}, G_L^1) - P_p^L C_1^L \right.$$

$$- \sum_{i=1}^{N-L-1} P_p^i Y(C_{L+2}^i, G_{L+1}^1, C_1^L) + \sum_{i=1}^{L-2} P_p^i Y(C_1^1, G_{i+1}^1, C_{i+2}^{L-i-1})$$

$$\left. + \sum_{i=1}^{L-2} P_p^i Y(C_{L-i+1}^i, G_{L-i}^1, C_1^{L-i-1}) \right]$$

$$+ \left[\sum_{i=L+1}^{N} P_d^i \{C_1^i\} - P_d^L L C_1^L \right] + \frac{P_m^1}{2}[Y(C_{L+1}^1, G_L^1, C_1^{L-1}) - Y(C_{L+2}^1, G_{L+1}^1, C_1^L)]$$

$$+ \frac{1-\alpha}{2}\left[P_m^{L+1} C_1^{L+1} - C_1^L + \sum_{i=2}^{N-L} P_m^i Y(C_{L+1}^i, G_L^1, C_1^{L-1}) \right.$$

$$\left. - \sum_{i=2}^{N-L-1} P_m^i Y(C_{L+2}^i, G_{L+1}^1, C_1^L) \right], \quad L = 2, \ldots, N-2,$$

$$\frac{dC_2^L}{dt} = \frac{1}{2}\left[P_p^{L-1} C_2^{L-1} + P_p^{L-1} C_3^{L-1} - 2P_p^L C_2^L - P_p^{L-1} Y(C_2^{L-1}, G_{L+1}^1) \right.$$

$$- P_p^{L-1} Y(C_3^{L-1}, G_2^1) + \sum_{i=1}^{L-2} P_p^i Y(C_2^i, G_{i+2}^1, C_{i+3}^{L-i-1})$$

$$\left. + \sum_{i=1}^{L-2} P_p^i Y(C_{L-i+2}^i, G_{L-i+1}^1, C_2^{L-i-1}) - \sum_{i=1}^{N-4} P_p^i Y(C_{L+3}^i, G_{L+2}^1, C_2^L) \right]$$

$$+ \left[\sum_{i=L+1}^{N-1} P_d^i \{C_2^i\} + P_d^{L+1} C_1^{L+1} - P_d^L L C_2^L \right]$$

$$+ \frac{P_m^1}{2}[Y(C_1^1, G_2^1, C_3^{L-1}) + Y(C_{L+2}^1, G_{L+1}^1, C_2^{L-1}) - Y(C_{L+3}^1, G_{L+2}^1, C_2^L)]$$

$$+ \frac{1-\alpha}{2}\left[P_m^{L+1} C_2^{L+1} - 2P_m^L C_2^L + \sum_{i=2}^{N-L-1} P_m^i Y(C_{L+2}^i, G_{L+1}^1, C_2^{L-1}) \right.$$

$$\left. - \sum_{i=2}^{N-L-2} P_m^i Y(C_{L+3}^i, G_{L+2}^1, C_2^L) \right], \quad L = 2, \ldots, N-3,$$

$$\frac{\mathrm{d}C_k^L}{\mathrm{d}t} = \frac{1}{2}\Bigg[P_p^{L-1}C_k^{L-1} + P_p^{L-1}C_{k+1}^{L-1} - 2P_p^L C_k^L - P_p^{L-1}Y(C_k^{L-1}, G_{L+k-1}^1)$$

$$- P_p^{L-1}Y(C_{k+1}^{L-1}, G_k^1) + \sum_{i=1}^{L-2} P_p^i Y(C_k^i, G_{k+i}^1, C_{k+i+1}^{L-i-1})$$

$$+ \sum_{i=1}^{L-2} P_p^i Y(C_{L+k-i}^i, G_{L+k-i-1}^1, C_k^{L-i-1}) - \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^L)$$

$$- \sum_{i=1}^{N-L-k} P_p^i Y(C_{L+k+1}^i, G_{L+k}^1, C_k^L) \Bigg]$$

$$+ \Bigg[\sum_{i=L+1}^{N-k+1} P_d^i\{C_k^i\} + \sum_{i=L+1}^{L+k-1} P_d^i\{C_{L+k-i}^i\} - P_d^L L C_k^L \Bigg] + \frac{P_m^1}{2}[Y(C_{k-1}^1, G_k^1, C_{k+1}^{L-1})$$

$$+ Y(C_{L+k}^1, G_{L+k-1}^1, C_k^{L-1}) - Y(C_{k-2}^1, G_{k-1}^1, C_k^L) - Y(C_{L+k+1}^1, G_{L+k}^1, C_k^L)]$$

$$+ \frac{1-\alpha}{2}\Bigg[P_m^{L+1}C_k^{L+1} + P_m^{L+1}C_{k-1}^{L+1} - 2P_m^L C_k^L + \sum_{i=2}^{k-1} P_m^i Y(C_{k-i}^i, G_k^1, C_{k+1}^{L-1})$$

$$+ \sum_{i=2}^{N-L-k+1} P_m^i Y(C_{L+k}^i, G_{L+k-1}^1, C_k^{L-1}) - \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^1, C_k^L)$$

$$- \sum_{i=2}^{N-L-k} P_m^i Y(C_{L+k+1}^i, G_{L+k}^1, C_k^L) \Bigg],$$

$$L = 2, \ldots, N-4, \quad k = 3, \ldots, N-L-1,$$

$$\frac{\mathrm{d}C_{N-L}^L}{\mathrm{d}t} = \frac{1}{2}\Bigg[P_p^{L-1}C_{N-L}^{L-1} + P_p^{L-1}C_{N-L+1}^{L-1} - 2P_p^L C_{N-L}^L - P_p^{L-1}Y(C_{N-L}^{L-1}, G_{N-1}^1)$$

$$- P_p^{L-1}Y(C_{N-L+1}^{L-1}, G_{N-L}^1) + \sum_{i=1}^{L-2} P_p^i Y(C_{N-L}^i, G_{N-L+i}^1, C_{N-L+i+1}^{L-i-1})$$

$$+ \sum_{i=1}^{L-2} P_p^i Y(C_{N-i}^i, G_{N-i-1}^1, C_{N-L}^{L-i-1}) - \sum_{i=1}^{N-L-2} P_p^i Y(C_{N-L-i-1}^i, G_{N-L-1}^1, C_{N-L}^L) \Bigg]$$

$$+ \Bigg[P_d^{L+1}C_{N-L}^{L+1} + \sum_{i=L+1}^{N-1} P_d^i\{C_{N-i}^i\} - P_d^L L C_{N-L}^L \Bigg]$$

$$+ \frac{P_m^1}{2}[Y(C_N^1, G_{N-1}^1, C_{N-L}^{L-1}) + Y(C_{N-L-1}^1, G_{N-L}^1, C_{N-L+1}^{L-1})$$

$$- Y(C_{N-L-2}^1, G_{N-L-1}^1, C_{N-L}^L)]$$

$$+ \frac{1-\alpha}{2}\Bigg[P_m^{L+1}C_{N-L-1}^{L+1} - 2P_m^L C_{N-L}^L + \sum_{i=2}^{N-L-1} P_m^i Y(C_{N-L-i}^i, G_{N-L}^1, C_{N-L+1}^{L-1})$$

$$- \sum_{i=2}^{N-L-2} P_m^i Y(C_{N-L-i-1}^i, G_{N-L-1}^1, C_{N-L}^L) \Bigg], \quad L = 2, \ldots, N-3,$$

$$\frac{\mathrm{d}C_{N-L+1}^L}{\mathrm{d}t} = \frac{1}{2}\Bigg[P_p^{L-1}C_{N-L+1}^{L-1} + P_p^{L-1}C_{N-L+2}^{L-1} - P_p^L C_{N-L+1}^L$$

$$- P_p^{L-1}Y(C_{N-L+2}^{L-1}, G_{N-L+1}^1) - \sum_{i=1}^{N-L-1} P_p^i Y(C_{N-L-i}^i, G_{N-L}^1, C_{N-L+1}^L)$$

$$+ \sum_{i=1}^{L-2} P_p^i Y(C_{N-L+1}^i, G_{N-L+i+1}^1, C_{N-L+i+2}^{L-i-1})$$

$$+ \sum_{i=1}^{L-2} P_p^i Y(C_{N-i+1}^i, G_{N-i}^1, C_{N-L+1}^{L-i-1}) \Bigg] + \Bigg[\sum_{i=L+1}^{N} P_d^i\{C_{N-i+1}^i\} - P_d^L L C_{N-L+1}^L \Bigg]$$

$$+ \frac{P_m^1}{2}[Y(C_{N-L}^1, G_{N-L+1}^1, C_{N-L+2}^{L-1}) - Y(C_{N-L-1}^1, G_{N-L}^1, C_{N-L}^L)]$$

$$+ \frac{1-\alpha}{2}\Bigg[P_m^{L+1}C_{N-L}^{L+1} - P_m^L C_{N-L+1}^L$$

$$+ \sum_{i=2}^{N-L} P_m^i Y(C_{N-L-i+1}^i, G_{N-L+1}^1, C_{N-L+2}^{L-1})$$

$$- \sum_{i=2}^{N-L-1} P_m^i Y(C_{N-L-i}^i, G_{N-L}^1, C_{N-L}^L) \Bigg], \quad L = 2, \ldots, N-2,$$

$$
\begin{aligned}
\frac{\mathrm{d}C_2^{N-2}}{\mathrm{d}t} = \frac{1}{2}\Bigg[& P_p^{N-3}C_2^{N-3} + P_p^{N-3}C_3^{N-3} - 2P_p^{N-2}C_2^{N-2} \\
& - P_p^{N-3}Y(C_2^{N-3}, G_{N-1}^1) - P_p^{N-3}Y(C_3^{N-3}, G_2^1) \\
& + \sum_{i=1}^{N-4} P_p^i Y(C_2^i, G_{i+2}^1, C_3^{N-i-3}) + \sum_{i=1}^{N-4} P_p^i Y(C_{N-i}^i, G_{N-i-1}^1, C_2^{N-i-3}) \Bigg] \\
& + [P_d^{N-1}C_2^{N-1} + P_d^{N-1}C_1^{N-1} - P_d^{N-2}(N-2)C_2^{N-2}] \\
& + \frac{P_m^1}{2}[Y(C_1^1, G_2^1, C_3^{N-3}) + Y(C_N^1, G_{N-1}^1, C_2^{N-3})] + \frac{1-\alpha}{2}[-2P_m^{N-2}C_2^{N-2}],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_1^{N-1}}{\mathrm{d}t} = \frac{1}{2}\Bigg[& P_p^{N-2}C_1^{N-2} + P_p^{N-2}C_2^{N-2} - P_p^{N-1}C_1^{N-1} + \sum_{i=1}^{N-3} P_p^i Y(C_1^i, G_{i+1}^1, C_{i+2}^{N-i-2}) \\
& + \sum_{i=1}^{N-3} P_p^i Y(C_{N-i}^i, G_{N-i-1}^1, C_1^{N-i-2}) \Bigg] + [P_d^N C_1^N - P_d^{N-1}(N-1)C_1^{N-1}] \\
& + \frac{P_m^1}{2}[Y(C_N^1, G_{N-1}^1, C_1^{N-2})] + \frac{1-\alpha}{2}[-P_m^{N-1}C_1^{N-1}],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_2^{N-1}}{\mathrm{d}t} = \frac{1}{2}\Bigg[& P_p^{N-2}C_2^{N-2} + P_p^{N-2}C_3^{N-2} - P_p^{N-1}C_2^{N-1} - P_p^{N-2}Y(C_3^{N-2}, G_2^1) \\
& + \sum_{i=1}^{N-3} P_p^i Y(C_2^i, G_{i+2}^1, C_3^{N-i-2}) + \sum_{i=1}^{N-3} P_p^i Y(C_{N-i+1}^i, G_{N-i}^1, C_2^{N-i-2}) \Bigg] \\
& + [P_d^N C_1^N - P_d^{N-1}(N-1)C_2^{N-1}] + \frac{P_m^1}{2}[Y(C_1^1, G_2^1, C_3^{N-2})] \\
& + \frac{1-\alpha}{2}[-P_m^{N-1}C_2^{N-1}],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}C_1^N}{\mathrm{d}t} = \frac{1}{2}\Bigg[& P_p^{N-1}C_1^{N-1} + P_p^{N-1}C_2^{N-1} + \sum_{i=1}^{N-2} P_p^i Y(C_1^i, G_{i+1}^1, C_{i+2}^{N-i-1}) \\
& + \sum_{i=1}^{N-2} P_p^i Y(C_{N-i+1}^i, G_{N-i}^1, C_1^{N-i-1}) \Bigg] + P_d^N[-NC_1^N].
\end{aligned}
$$

A.2. Gaps
Here we present the time rate of change for each gap length and location, obtained by considering the potential outcomes of each type of birth, death and movement event.

$$
\begin{aligned}
\frac{\mathrm{d}G_1^1}{\mathrm{d}t} = \frac{1}{2}\Bigg[& \sum_{i=1}^{N-2} P_p^i Y(C_3^i, G_1^2) - \sum_{i=1}^{N-1} P_p^i Y(C_2^i, G_1^1) \Bigg] + \Bigg[\sum_{i=2}^{N} P_d^i \{C_1^i\} - \sum_{i=1}^{N-1} Y(C_2^i, G_1^1) \Bigg] \\
& + \frac{P_m^1}{2}[C_1^1 - 2C_2^1 + Y(C_3^1, G_1^2)] + \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{N-1} P_m^i \{C_2^i\} + \sum_{i=2}^{N-2} P_m^i Y(C_3^i, G_1^2) \Bigg],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}G_2^1}{\mathrm{d}t} = \frac{1}{2}\Bigg[& \sum_{i=1}^{N-3} P_p^i Y(C_4^i, G_2^2) - \sum_{i=1}^{N-2} P_p^i \{Y(C_3^i, G_2^1)\} - P_p^1 Y(C_1^1, G_2^1) \Bigg] \\
& + \Bigg[\sum_{i=3}^{N} P_d^i \{C_1^i\} - P_d^1 Y(C_1^1, G_2^1) - \sum_{i=1}^{N-1} P_d^i Y(C_3^i, G_2^1) \Bigg] \\
& + \frac{P_m^1}{2}[-Y(C_1^1, G_2^1) - 2Y(C_3^1, G_2^1) + Y(C_4^1, G_2^2)] \\
& + \frac{(1-\alpha)}{2}\Bigg[P_m^2 C_1^2 + \sum_{i=2}^{N-1} P_m^i \{C_2^i\} - \sum_{i=2}^{N-2} P_m^i \{Y(C_3^i, G_2^1)\} + \sum_{i=2}^{N-3} P_m^i Y(C_4^i, G_2^2) \Bigg],
\end{aligned}
$$

$$\frac{\mathrm{d}G_k^1}{\mathrm{d}t} = \frac{1}{2}\left[\sum_{i=1}^{N-k-1} P_p^i Y(C_{k+2}^i, G_k^2)\right.$$

$$\left. + \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^2) - \sum_{i=1}^{N-k} P_p^i Y(C_{k+1}^i, G_k^1) - \sum_{i=1}^{k-1} P_p^i Y(C_{k-i}^i, G_k^1)\right]$$

$$+ \left[\sum_{i=1}^{k-1}\sum_{j=k-i+2}^{N-i+1} P_d^j \{C_i^j\} - \sum_{i=1}^{k-1} P_d^i Y(C_{k-i}^i, G_k^1) - \sum_{i=1}^{N-k} P_d^i Y(C_{k+1}^i, G_k^1)\right]$$

$$+ \frac{P_m^1}{2}[-2Y(C_{k-1}^1, G_k^1) - 2Y(C_{k+1}^1, G_k^1) + Y(C_{k-2}^1, G_{k-1}^2) + Y(C_{k+2}^1, G_k^2)]$$

$$+ \frac{1-\alpha}{2}\left[\sum_{i=2}^{k} P_m^i \{C_{k-i+1}^i\} + \sum_{i=2}^{N-k+1} P_m^i \{C_k^i\} - \sum_{i=2}^{k-1} P_m^i \{Y(C_{k-i}^i, G_k^1)\}\right.$$

$$\left. - \sum_{i=2}^{N-k} P_m^i \{Y(C_{k+1}^i, G_k^1)\} + \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^2) + \sum_{i=2}^{N-k-1} P_m^i Y(C_{k+2}^i, G_k^2)\right],$$

$$k = 3, \ldots, N-2,$$

$$\frac{\mathrm{d}G_{N-1}^1}{\mathrm{d}t} = \frac{1}{2}\left[\sum_{i=1}^{N-3} P_p^i \{Y(C_{N-i-2}^i, G_{N-2}^2)\} - P_p^1 Y(C_N^1, G_1^{N-1}) - \sum_{i=1}^{k-1} P_p^i Y(C_{N-i-1}^i, G_{N-1}^1)\right]$$

$$+ \left[\sum_{i=1}^{N-2} P_d^i \{C_{N-i+1,i}\} - \sum_{i=1}^{N-2} P_d^i \{Y(C_{N-i-1}^i, G_{N-1}^1)\} - P_d^1 Y(C_N^1, G_k^1)\right]$$

$$+ \frac{P_m^1}{2}[-2Y(C_{N-2}^1, G_{N-1}^1) - Y(C_N^1, G_{N-1}^1) + Y(C_{N-3}^1, G_{N-2}^2)]$$

$$+ \frac{1-\alpha}{2}\left[\sum_{i=2}^{N-1} P_m^i \{C_{N-i}^i\} + P_m^2 C_{N-1}^2 - \sum_{i=2}^{N-2} P_m^i \{Y(C_{N-1-i}^i, G_{N-1}^1)\}\right.$$

$$\left. + \sum_{i=2}^{N-3} P_m^i Y(C_{N-i-2}^i, G_{N-2}^2)\right],$$

$$\frac{\mathrm{d}G_N^1}{\mathrm{d}t} = \frac{1}{2}\left[\sum_{i=1}^{N-2} P_p^i Y(C_{N-i-1}^i, G_{N-1}^2) - \sum_{i=1}^{N-1} P_p^i Y(C_{N-i}^i, G_N^1)\right]$$

$$+ \left[\sum_{i=2}^{N} P_d^i C_{N-i+1}^i - \sum_{i=1}^{N-1} P_d^i Y(C_{N-i}^i, G_N^1)\right]$$

$$+ \frac{P_m^1}{2}[C_N^1 - 2C_{N-1}^1 + Y(C_{N-2}^1, G_{N-1}^2)]$$

$$+ \frac{1-\alpha}{2}\left[-\sum_{i=2}^{N-1} P_m^i \{C_{N-i}^i\} + \sum_{i=2}^{N-2} P_m^i Y(C_{N-i-1}^i, G_{N-1}^2)\right],$$

$$\frac{\mathrm{d}G_1^L}{\mathrm{d}t} = \frac{1}{2}\left[\sum_{i=1}^{N-L-1} P_p^i Y(C_{L+2}^i, G_1^{L+1}) - \sum_{i=1}^{N-L} P_p^i Y(C_{L+1}^i, G_1^L)\right] + \left[P_d^1 Y(C_1^1, G_2^{L-1})\right.$$

$$+ \sum_{i=2}^{N-L+1} P_d^i Y(C_L^i, G_1^{L-1}) + \sum_{i=1}^{L-2} P_d^1 Y(C_{i+1}^1, G_1^i, G_{i+2}^{L-i-1}) - \sum_{i=1}^{N-L} P_d^i Y(C_{L+1}^i, G_1^L)\right]$$

$$+ \frac{P_m^1}{2}[Y(C_L^1, G_1^{L-1}) - 2Y(C_{L+1}^1, G_1^L) + Y(C_{L+2}^1, G_1^{L+1})]$$

$$+ \frac{1-\alpha}{2}\left[-\sum_{i=2}^{N-L} P_m^i \{Y(C_{L+1}^i, G_1^L)\} + \sum_{i=2}^{N-L-1} P_m^i Y(C_{L+2}^i, G_1^{L+1})\right],$$

$$L = 2, \ldots, N-2,$$

$$
\begin{aligned}
\frac{\mathrm{d}G_2^L}{\mathrm{d}t} =& \frac{1}{2}\Bigg[\sum_{i=1}^{N-L-2} P_p^i Y(C_{L+3}^i, G_2^{L+1}) - \sum_{i=1}^{N-L-1} P_p^i \{Y(C_{L+2}^i, G_2^L)\} - P_p^1 Y(C_1^1, G_2^L) \Bigg] \\
&+ \Bigg[P_d^2 Y(C_1^2, G_3^{L-1}) + \sum_{i=2}^{N-L} P_d^i Y(C_{L+1}^i, G_2^{L-1}) + \sum_{i=1}^{L-2} P_d^1 \{Y(C_{i+2}^1, G_2^i, G_{i+3}^{L-i-1})\} \\
&- P_d^1 Y(C_1^1, G_2^L) - \sum_{i=1}^{N-L-1} P_d^i Y(C_{L+2}^i, G_2^L) \Bigg] + \frac{P_m^1}{2}[Y(C_2^1, G_3^{L-1}) + Y(C_{L+1}^1, G_2^{L-1}) \\
&- Y(C_1^1, G_2^L) - 2Y(C_{L+2}^1, G_2^L) + Y(C_{L+3}^1, G_2^{L+1})] \\
&+ \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{N-L-1} P_m^i Y(C_{L+2}^i, G_2^L) + \sum_{i=2}^{N-L-2} P_m^i Y(C_{L+3}^i, G_2^{L+1}) \Bigg], \\
& L = 2,\ldots,N-3,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}G_k^L}{\mathrm{d}t} =& \frac{1}{2}\Bigg[\sum_{i=1}^{N-L-k} P_p^i Y(C_{L+k+1}^i, G_k^{L+1}) + \sum_{i=1}^{k-2} P_p^i Y(C_{k-i-1}^i, G_{k-1}^{L+1}) \\
&- \sum_{i=1}^{N-L-k+1} P_p^i Y(C_{L+k}^i, G_k^L) - \sum_{i=1}^{k-1} P_p^i Y(C_{k-i}^i, G_k^L) \Bigg] \\
&- \Bigg[\sum_{i=2}^{k} P_d^i Y(C_{k-i+1}^i, G_{k+1}^{L-1}) + \sum_{i=2}^{N-L-k+2} P_d^i Y(C_{L+k-1}^i, G_k^{L-1}) \\
&+ \sum_{i=1}^{L-2} P_d^1 Y(C_{k+i}^1, G_k^i, G_{k+i+1}^{L-i-1}) - \sum_{i=1}^{k-1} P_d^i Y(C_{k-i}^i, G_k^L) - \sum_{i=1}^{N-L-k+1} P_d^i Y(C_{L+k}^i, G_k^L) \Bigg] \\
&+ \frac{P_m^1}{2}[Y(C_k^1, G_{k+1}^{L-1}) + Y(C_{L+k-1}^1, G_k^{L-1}) - 2Y(C_{k-1}^1, G_k^L) - 2Y(C_{L+k}^1, G_k^L) \\
&+ Y(C_{k-2}^1, G_{k-1}^{L+1}) + Y(C_{L+k+1}^1, G_k^{L+1})] \\
&+ \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{k-1} P_m^i Y(C_{k-i}^i, G_k^L) - \sum_{i=2}^{N-L-k+1} P_m^i Y(C_{L+k}^i, G_k^L) \\
&+ \sum_{i=2}^{k-2} P_m^i Y(C_{k-i-1}^i, G_{k-1}^{L+1}) + \sum_{i=2}^{N-L-k} P_m^i Y(C_{L+k+1}^i, G_k^{L+1}) \Bigg], \\
& L = 2,\ldots,N-4, \quad k = 3,\ldots,N-L-1,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}G_{N-L}^L}{\mathrm{d}t} =& \frac{1}{2}\Bigg[\sum_{i=1}^{N-L-2} P_p^i \{Y(C_{N-L-i-1}^i, G_{N-L-1}^{L+1})\} - P_p^1 Y(C_N^1, G_{N-L}^L) \\
&- \sum_{i=1}^{N-L-1} P_p^i Y(C_{N-L-i}^i, G_{N-L}^L) \Bigg] + \Bigg[\sum_{i=2}^{N-L} P_d^i \{Y(C_{N-L-i+1}^i, G_{N-L+1}^{L-1})\} \\
&+ P_d^2 Y(C_{N-1}^2, G_{N-L}^{L-1}) + \sum_{i=1}^{L-2} P_d^1 Y(C_{N-L+i}^1, G_{N-L}^i, G_{N-L+i+1}^{L-i-1}) \\
&- \sum_{i=1}^{N-L-1} P_d^i \{Y(C_{N-L-i}^i, G_k^L)\} - P_d^1 Y(C_N^1, G_{N-L}^L) \Bigg] \\
&+ \frac{P_m^1}{2}[Y(C_{N-L}^1, G_{N-L+1}^{L-1}) + Y(C_{N-1}^1, G_{N-L}^{L-1}) - 2Y(C_{N-L-1}^1, G_{N-L}^L) \\
&- Y(C_N^1, G_{N-L}^L) + Y(C_{N-L-2}^1, G_{N-L-1}^{L+1})] \\
&+ \frac{1-\alpha}{2}\Bigg[-\sum_{i=2}^{N-L-1} P_m^i Y(C_{N-L-i}^i, G_{N-L}^L + \sum_{i=2}^{N-L-2} P_m^i Y(C_{N-L-i-1}^i, G_{N-L-1}^{L+1}) \Bigg], \\
& L = 2,\ldots,N-3,
\end{aligned}
$$

$$\frac{\mathrm{d}G_{N-L+1}^{L}}{\mathrm{d}t} = \frac{1}{2}\left[\sum_{i=1}^{N-L-1} P_p^i Y(C_{N-L-i}^i, G_{N-L}^{L+1}) - \sum_{i=1}^{N-L} P_p^i Y(C_{N-L-i+1}^i, G_{N-L+1}^L)\right]$$

$$+\left[P_d^1 Y(C_N^1, G_{N-L+1}^{L-1}) + \sum_{i=1}^{N-L+1} P_d^i Y(C_{N-L-i+2}^i, G_{N-L+2}^{L-1})\right.$$

$$\left.-\sum_{i=1}^{N-L} P_d^i Y(C_{N-L-i+1}^i, G_{N-L+1}^L) + \sum_{i=1}^{L-2} P_d^1 Y(C_{N-L+i+1}^1, G_{N-L+1}^i, G_{N-L+i+2}^{L-i-1})\right]$$

$$+\frac{P_m^1}{2}[Y(C_{N-L+1}^1, G_{N-L+2}^{L-1}) - 2Y(C_{N-L}^1, G_{N-L+1}^L) + Y(C_{N-L-1}^1, G_{N-L}^{L+1})]$$

$$+\frac{1-\alpha}{2}\left[-\sum_{i=2}^{N-L} P_m^i Y(C_{N-L-i+1}^i, G_{N-L+1}^L) + \sum_{i=2}^{N-L-1} P_m^i Y(C_{N-L-i}^i, G_{N-L}^{L+1})\right],$$

$$L = 2,\ldots,N-2,$$

$$\frac{\mathrm{d}G_2^{N-2}}{\mathrm{d}t} = \frac{1}{2}\left[-P_p^1 Y(C_N^1, G_2^{N-2}) - P_p^1 Y(C_1^1, G_2^{N-2})\right]$$

$$+[P_d^2 Y(C_1^2, G_3^{N-3}) + P_d^2 Y(C_{N-1}^2, G_2^{N-3})$$

$$+\sum_{i=1}^{N-4} P_d^1\{Y(C_{i+2}^1, G_2^i, G_{i+3}^{N-i-3})\} - P_d^1 Y(C_1^1, G_2^{N-2}) - P_d^1 Y(C_N^1, G_2^{N-2})\bigg]$$

$$+\frac{P_m^1}{2}[Y(C_2^1, G_3^{N-3}) + Y(C_{N-1}^1, G_2^{N-3}) - Y(C_1^1, G_2^{N-2}) - Y(C_N^1, G_2^{N-2})],$$

$$\frac{\mathrm{d}G_1^{N-1}}{\mathrm{d}t} = \frac{1}{2}\left[-P_p^1 Y(C_N^1, G_1^{N-1})\right] + [P_d^1 Y(C_1^1, G_2^{N-2}) + P_d^2 Y(C_{N-1}^2, G_1^{N-2})$$

$$+\sum_{i=1}^{N-3} P_d^1\{Y(C_{i+1}^1, G_1^i, G_{i+2}^{N-i-2})\} - P_d^1 Y(C_N^1, G_1^{N-1})\bigg]$$

$$+\frac{P_m^1}{2}[Y(C_{N-1}^1, G_1^{N-2}) - Y(C_N^1, G_1^{N-1})],$$

$$\frac{\mathrm{d}G_2^{N-1}}{\mathrm{d}t} = \frac{1}{2}\left[-P_p^1 Y(C_1^1, G_2^{N-1})\right] + [P_d^1 Y(C_N^1, G_2^{N-2}) + P_d^2 Y(C_1^2, G_3^{N-2})$$

$$-P_d^1 Y(C_1^1, G_2^{N-1}) + \sum_{i=1}^{N-3} P_d^1 Y(C_{i+2}^1, G_2^i, G_{i+3}^{N-i-2})\bigg]$$

$$+\frac{P_m^1}{2}[Y(C_2^1, G_3^{N-2}) - Y(C_1^1, G_2^{N-1})],$$

$$\frac{\mathrm{d}G_1^N}{\mathrm{d}t} = P_d^1\left[Y(C_1^1, G_2^{N-1}) + Y(C_N^1, G_1^{N-1}) + \sum_{i=1}^{N-2} Y(C_{i+1}^1, G_1^i, G_{i+2}^{L-i-1})\right].$$

The *Y* function represents the number of configurations that contain the specified chains and gaps within the parentheses, and is defined as

$$Y(C_j^i, G_l^k) = C_j^i P(G_l^k|C_j^i),$$

where $P(G_l^k|C_j^i)$ represents the probability that a gap of length *k* at site *l* exists, given that there is a chain of length *i* at site *j*, and can be calculated as

$$P(G_l^k|C_j^i) = G_l^k / A_G(C_j^i),$$

where $A_G(C_j^i)$ are all the possible gaps that can exist on either side of the chain of length *i* at site *j*. If we are interested in gaps on the positive *x* side of the chain, we require $l = i + j$ and $k \leqslant N - i - j + 1$. If we are interested in gaps on the negative *x* side of the chain, we require $l \leqslant j - 1$, $k = j - k$. Similarly,

$$Y(C_j^i, G_l^k, G_n^m) = C_j^i P(G_l^k|C_j^i) P(G_n^m|C_j^i, G_l^k),$$

where $P(G_n^m|C_j^i, G_l^k)$ is the probability a gap of length *m* at site *n* exists, given that there is a chain of length *i* at site *j* and a gap of length *k* at site *l*. Note that the gaps at site *l* and site *n* must necessarily be on opposing sides of the chain at site *j* and hence $P(G_n^m|C_j^i, G_l^k)$ is calculated in the same manner as $P(G_l^k|C_j^i)$.

Appendix B. Numerical techniques

B.1. Ordinary differential equations

The system of ODEs describing the dynamics of the chains and gaps, presented in appendices A.1 and A.2, are solved using an adaptive Runge–Kutta method with a strict truncation error control of 10^{-6} [49]. All results presented are found to be insensitive to a reduction in the strict truncation error control.

B.2. Partial differential equations

The two MF descriptions of the discrete process, equations (1) and (2), are discretised onto a spatially uniform finite difference grid with grid spacing δx. The spatial derivative terms are approximated using a central finite difference approximation. We approximate the temporal derivative using the backward Euler method with constant time step δt, and the resulting system of nonlinear algebraic equations is solved using Picard iteration with absolute convergence tolerance ϵ. Finally, the system of tridiagonal algebraic equations is solved using the Thomas algorithm [49]. In all cases, $\delta x = 0.1$, $\delta t = 0.01$ and $\epsilon = 10^{-6}$.

References

[1] Alvord E C Jr and Shaw C M 1991 *Neoplasms Affecting the Nervous System of the Elderly* (Oxford: Oxford University Press)
[2] Clavero M and García-Berthou E 2005 *Trends Ecol. Evol.* **20** 110
[3] Gabriely G, Wurdinger T, Kesari S, Esau C C, Burchard J, Linsley P S and Krichevsky A M 2008 *Mol. Cell. Biol.* **28** 5369–80
[4] Gerlee P and Nelander S 2012 *PLoS Comput. Biol.* **8** e1002556
[5] Gerlee P and Nelander S 2016 *Math. Biosci.* **276** 75–81
[6] Johnston S T, Simpson M J and McElwain D L S 2014 *J. R. Soc. Interface* **11** 20140325
[7] Martin P 1997 *Science* **276** 75–81
[8] Nnetu K D, Knorr M, Käs J and Zink M 2012 *New J. Phys.* **14** 115012
[9] Phillips B L, Brown G P, Greenlees M, Webb J K and Shine R 2007 *Austral Ecol.* **32** 169–76
[10] Poujade M, Grasland-Mongrain E, Hertzog A, Jouanneau J, Chavrier P, Ladoux B, Buguin A and Silberzan P 2007 *Proc. Natl Acad. Sci.* **104** 15988–93
[11] Sherratt J A, Lewis M A and Fowler A C 1995 *Proc. Natl Acad. Sci.* **92** 2524–8
[12] Swanson K R, Bridge C, Murray J and Alvord E C Jr 2003 *J. Neurol. Sci.* **216** 1–10
[13] Townsend C R 1996 *Biol. Conservation* **78** 13–22
[14] Mungomery R 1935 *Cane Growers Q. Bull.* **3** 21–7
[15] Fort J and Sole R V 2013 *New J. Phys.* **15** 055001
[16] Deroulers C, Aubert M, Badoual M and Grammaticos B 2009 *Phys. Rev.* E **79** 031917
[17] Fort J and Pujol T 2007 *New J. Phys.* **9** 234
[18] Illien P, Bénichou O, Mejía-Monasterio C, Oshanin G and Voituriez R 2013 *Phys. Rev. Lett.* **111** 038102
[19] Khain E, Katakowski M, Charteris N, Jiang F and Chopp M 2012 *Phys. Rev.* E **86** 011904
[20] Simpson M J, Landman K A and Hughes B D 2010 *Physica* A **389** 3779–90
[21] Simpson M J, Treloar K K, Binder B J, Haridas P, Manton K J, Leavesley D I, McElwain D L S and Baker R E 2013 *J. R. Soc. Interface* **10** 20130007
[22] Simpson M J and Baker R E 2011 *Phys. Rev.* E **83** 051922
[23] Anguige K and Schmeiser C 2009 *J. Math. Biol.* **58** 395–427
[24] Baker R E and Simpson M J 2010 *Phys. Rev.* E **82** 041905
[25] Fernando A E, Landman K A and Simpson M J 2010 *Phys. Rev.* E **81** 011903
[26] Johnston S T, Simpson M J and Baker R E 2012 *Phys. Rev.* E **85** 051922
[27] Johnston S T, Simpson M J and Baker R E 2015 *J. Theor. Biol.* **370** 81–92
[28] Johnston S T, Baker R E and Simpson M J 2016 *Phys. Rev.* E **93** 042413
[29] Middleton A M, Fleck C and Grima R 2014 *J. Theor. Biol.* **359** 220–32
[30] Turner S, Sherratt J A, Painter K J and Savill N J 2004 *Phys. Rev.* E **69** 021910
[31] Grima R 2008 *Curr. Top. Dev. Biol.* **81** 435–60
[32] Sharkey K J, Fernandez C, Morgan K L, Peeler E, Thrush M, Turnbull J F and Bowers R G 2006 *J. Math. Biol.* **53** 61–85
[33] Wilkinson R R and Sharkey K J 2014 *Phys. Rev.* E **89** 022808
[34] Law R, Purves D W, Murrell D J and Dieckmann U 2001 *Br. Ecol. Soc.* **14** 21–44
[35] Murrell D J 2010 *Ecology* **91** 1605–16
[36] Chowdhury D, Schadschneider A and Nishinari K 2005 *Phys. Life Rev.* **2** 318–52
[37] Hedayati M *et al* 2013 *Nanomedicine* **8** 29–41
[38] Johnston S T, Baker R E, McElwain D L S and Simpson M J 2017 *Sci. Rep.* **7** 42134
[39] Fisher R A 1937 *Ann. Eugenics* **7** 355–69
[40] Kolmogorov A N, Petrovsky I G and Piskunov N S 1937 *Mosc. Univ. Math. Bull.* **1** 1–25
[41] Gillespie D T 1977 *J. Phys. Chem.* **81** 2340–61
[42] Liang C C, Park A Y and Guan J L 2007 *Nat. Protocols* **2** 329–33
[43] Johnston S T, Shah E T, Chopin L K, McElwain D L S and Simpson M J 2015 *BMC Syst. Biol.* **9** 38
[44] Simpson M J, Binder B J, Haridas P, Wood B K, Treloar K K, McElwain D L S and Baker R E 2013 *Bull. Math. Biol.* **75** 871–89
[45] Hiebeler D E, Houle J, Drummond F, Bilodeau P and Merckens J 2016 *J. Theor. Biol.* **407** 212–24
[46] Markham D C, Simpson M J and Baker R E 2013 *Phys. Rev.* E **87** 062702
[47] Brown R G, James A F, Pitchford J W and Plank M J 2012 *J. Theor. Biol.* **310** 231–8
[48] Johnston S T, Ross J V, Binder B J, McElwain D L S, Haridas P and Simpson M J 2016 *J. Theor. Biol.* **400** 19–31
[49] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 2007 *Numerical Recipes 3rd Edition: The Art of Scientific Computing* (Cambridge: Cambridge University Press)