

## Research



**Cite this article:** Johnston ST, Simpson MJ, Crampin EJ. 2020 Predicting population extinction in lattice-based birth–death–movement models. *Proc. R. Soc. A* **476**: 20200089.  
<http://dx.doi.org/10.1098/rspa.2020.0089>

Received: 12 February 2020

Accepted: 6 May 2020

**Subject Areas:**

applied mathematics, mathematical modelling

**Keywords:**

random walk, continuum limit, extinction, diffusion, mathematical modelling, population dynamics

**Author for correspondence:**

Stuart T. Johnston

e-mail: [stuart.johnston@unimelb.edu.au](mailto:stuart.johnston@unimelb.edu.au)

# Predicting population extinction in lattice-based birth–death–movement models

Stuart T. Johnston<sup>1,2</sup>, Matthew J. Simpson<sup>3</sup> and Edmund J. Crampin<sup>1,2,4</sup>

<sup>1</sup>Systems Biology Laboratory, School of Mathematics and Statistics, and Department of Biomedical Engineering, <sup>2</sup>ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, Melbourne School of Engineering, and <sup>3</sup>School of Medicine, Faculty of Medicine Dentistry and Health Sciences, University of Melbourne, Parkville, Victoria, Australia

<sup>4</sup>School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

STJ, 0000-0003-4532-5573; MJS, 0000-0001-6254-313X

The question of whether a population will persist or go extinct is of key interest throughout ecology and biology. Various mathematical techniques allow us to generate knowledge regarding individual behaviour, which can be analysed to obtain predictions about the ultimate survival or extinction of the population. A common model employed to describe population dynamics is the lattice-based random walk model with crowding (exclusion). This model can incorporate behaviour such as birth, death and movement, while including natural phenomena such as finite size effects. Performing sufficiently many realizations of the random walk model to extract representative population behaviour is computationally intensive. Therefore, continuum approximations of random walk models are routinely employed. However, standard continuum approximations are notoriously incapable of making accurate predictions about population extinction. Here, we develop a new continuum approximation, the *state-space diffusion approximation*, which explicitly accounts for population extinction. Predictions from our approximation faithfully capture the behaviour in the random walk model, and provides additional information compared to standard approximations.

We examine the influence of the number of lattice sites and initial number of individuals on the long-term population behaviour, and demonstrate the reduction in computation time between the random walk model and our approximation.

## 1. Introduction

Determining whether the behaviour of individuals in a population results in the extinction or survival of the population is a key question in cell biology, population biology and ecology [1–9]. For example, tumour cells can undergo migration from the original tumour site to form metastases [10]. However, this relies on the successful colonization of a new site by a small number of tumour cells. As such, the question of whether the new colony of cells will survive or go extinct is paramount [11]. Similarly, in the field of drug development, determining whether a novel treatment results in the eradication of a target cell population is crucial [12]. In an ecological context, predicting whether the introduction of a threatened species to a particular area will be successful is critical for selecting suitable sites [3,8].

Mathematical models provide a framework for incorporating knowledge about population behaviour such that the evolution of the population can be evaluated [1,4,7,13–20]. The most common mathematical models used to investigate population dynamics are *continuum* models, such as ordinary differential equation (ODE) or partial differential equation (PDE) models. These approaches aim to predict the temporal and spatio-temporal distribution of the average density of individuals, respectively [15,20–22]. While such models are employed widely, they are ill-suited for describing extinction of finite populations, as the population is not restricted to integer states, that is, whole numbers of individuals within the population. *Discrete* models enforce the number of individuals to be an integer, as each individual in the model is simulated explicitly [23]. Further, discrete models have the benefit that biological mechanisms, such as birth, death, and movement, can be directly imposed at the level of individuals [23–26].

A popular choice of discrete model for describing population dynamics is the lattice-based random walk [14,23–29]. Individuals in the model reside on a pre-defined lattice, and undergo processes such as birth, death and movement in a stochastic manner according to biologically inspired rules [14,23–26]. The lattice provides a natural method for incorporating population limits and finite size effects, known as an exclusion process, by imposing the restriction that each lattice site contains, at most, one individual [30,31]. These lattice-based random walks have been successfully used to investigate processes in cell biology and ecology, among others [14,27,32,33].

While the simplicity and flexibility of the lattice-based random walk is appealing, there are limitations to this modelling framework. In particular, the stochastic nature of the random walk means that a considerable number of simulations of the random walk process must be performed to obtain representative average behaviour of the population. This can be extremely computationally demanding, particularly if the random walk model is employed to perform parameter inference [33,34]. Furthermore, the random walk model is typically analytically intractable, and the amount of analytic insight is limited. To address this, continuum approximations of random walk processes have been developed, where time is treated as a continuous variable [23,24,26,35]. Provided a judicious choice of approximation is made, typically depending on the rates of birth, death and movement, these approximations accurately describe the average behaviour in the random walk [24,26,35]. However, as noted previously, standard continuum models are ill-suited for describing population extinction. Therefore, the development of a continuum approximation that avoids the computational burden of repeated simulations of a stochastic random walk process, while faithfully capturing the behaviour in the random walk and accurately predicting population extinction would be beneficial.

Here we introduce a new PDE approximation of the underlying birth–death–movement lattice-based random walk, which we refer to as the state-space diffusion approximation (SSDA). The SSDA involves two major components. First, the underlying random walk is approximated

via a state-space representation. The approximate state space of this random walk corresponds to the number of sites occupied by individuals. Second, the state-space representation is approximated via the diffusion method, which is also known as the Fokker–Planck approximation [36,37]. In contrast to other PDE approximations of lattice-based random walks, the dependent variables in the PDE are time and *state space*, rather than time and physical space. As such, the solution to the PDE can be interpreted as the distribution of the number of occupied sites at a particular time. Here, we consider two mechanisms for birth and death [14,25]:

- (i) Logistic growth, where the intrinsic rates of birth and death of an individual are independent of the neighbouring lattice sites, and;
- (ii) Allee growth, where the intrinsic rates of birth and death depend on whether an individual has zero occupied neighbouring sites, or at least one occupied neighbour site.

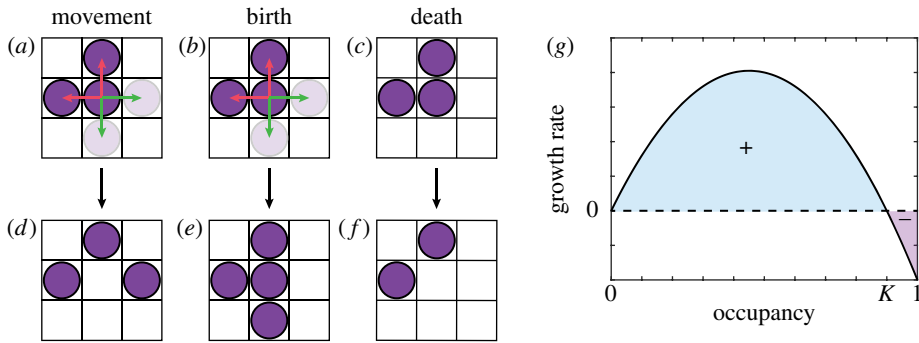
The logistic growth model is widely used to describe populations where low-density growth is close to exponential but the population eventually tends to a finite carrying capacity density due to competition for a shared resource [38]. Note that growth refers to the net birth/death rate of the population, whereas intrinsic rates of birth/death refers to the birth/death of an individual. The logistic growth model has been used to describe the dynamics of populations of tumour cells [15,20], spruce budworms [38] and jellyfish polyps [39], among others. The logistic model is relatively simple, but still captures key features of population behaviour. The Allee growth model is slightly more complicated, and describes population dynamics where growth is either inhibited or negative below a threshold population density [7,17,40–42]. This reduced growth rate can represent mechanisms where a threshold number of individuals are needed to maintain a population, such as for effective predator avoidance or for a viable number of breeding pairs [40,41]. Reduced or negative growth rates have been observed in populations of cancer cells [7,17], as well as in animal populations, such as the gypsy moth [43]. The widespread relevance of both the logistic and Allee growth models ensure that insight gleaned from analysis of population extinction in such models will be instructive in a range of research areas.

For both the logistic and Allee models, we demonstrate that the SSDA faithfully predicts the behaviour of the random walk process, provided that the standard mean-field assumption is satisfied. Further, as the SSDA explicitly accounts for the possibility that an individual random walk will undergo extinction, we demonstrate that the SSDA provides a more accurate approximation of the average random walk behaviour compared to standard mean-field ODEs. The SSDA provides an estimate of both the rate of extinction and the state-space distribution, and we show that these estimates are consistent with the underlying random walk process. Finally, we highlight that the time taken to obtain the solution to the SSDA is orders of magnitude lower than the time taken to perform sufficiently many realizations of the random walk.

## 2. Model

### (a) Logistic model

Consider an  $n$ -dimensional lattice-based birth–death–movement random walk with exclusion on a lattice with  $N$  sites [24,30]. While, in theory,  $n$  is arbitrary as continuum approximations of the random walk under spatially uniform initial conditions are independent of  $n$ , the most relevant  $n$  for biological and ecological processes are  $n = 1, 2, 3$ . Here, we present results corresponding to a square lattice with  $n = 2$ . Individuals in the random walk undergo birth, death and movement events at rates per unit time  $P_b$ ,  $P_d$  and  $P_m$ , respectively [24,44] (figure 1a–f). During a birth event, we select an individual at random, and select one of  $2n$  nearest-neighbour sites in the von Neumann neighbourhood of the selected individual. The individual attempts to place a daughter individual at the target site and is successful provided that the target site is unoccupied; otherwise the event is aborted. Similarly, during a movement event, we select an individual at random, as well as a nearest-neighbour site. The individual attempts to move to the target site, and the



**Figure 1.** Mechanisms in the logistic random walk model. (a) Potential movement events for the individual in the centre of the lattice. (b) Potential birth events for the individual in the centre of the lattice. Green arrows indicate directions that would result in a successful event, red arrows indicate directions that would result in an unsuccessful event. (c) Potential death events for the individual in the centre of the lattice. (d)–(f) Lattice configuration after the corresponding event in (a)–(c). (g) Typical growth rate curve as a function of the occupancy of the lattice. Cyan regions correspond to positive growth and purple regions correspond to negative growth.  $K = (P_b - P_d)/P_b$  is the carrying capacity. (Online Version in colour.)

event is only successful if the target site is unoccupied. During a death event, an individual is selected at random and removed from the lattice. Initially,  $N_0$  sites are selected at random, and an individual is placed at each selected site. We select this method of determining whether an event is successful to be consistent with previous investigations into continuum approximations of lattice-based birth–death–movement processes [24,44]. For spatially uniform initial conditions, periodic boundary conditions and under the mean-field assumption, where the probability that a particular lattice site is occupied is independent of all other sites, the evolution of the average proportion of occupied sites  $0 \leq \bar{S} \leq 1$  is given by [24,44]

$$\frac{d\bar{S}}{dt} = P_b \bar{S}(1 - \bar{S}) - P_d \bar{S}, \quad (2.1)$$

as highlighted in figure 1g.

Here we instead consider the probability,  $f(S, t)$ , that the system has  $S$  occupied sites at time  $t$ . The system evolves in terms of the state variable  $S$  during a single time step of duration  $\tau$ , denoted  $\delta f(S, t) = f(S, t + \tau) - f(S, t)$ , according to

$$\begin{aligned} \delta f(S, t) = & P_b \left( \frac{S-1}{N} \right) \left( 1 - \frac{S-2}{N-1} \right) f(S-1, t) + P_d \left( \frac{S+1}{N} \right) f(S+1, t) \\ & - P_b \frac{S}{N} \left( 1 - \frac{S-1}{N-1} \right) f(S, t) - \frac{1}{N} P_d S f(S, t). \end{aligned} \quad (2.2)$$

The first term corresponds to the transition from state  $S-1$  to state  $S$  due to an individual undergoing a birth event. This occurs proportional to the birth rate,  $P_b$ , the proportion of the lattice that is occupied by individuals,  $(S-1)/N$ , the probability that a nearest-neighbour site is unoccupied,  $1 - (S-2)/(N-1)$ , and the probability that the system is in state  $S-1$ ,  $f(S-1, t)$ . We note that this term involves the standard mean-field assumption that the probability that a site is occupied by an individual is independent of whether a nearest-neighbour site is occupied [24]. The second term corresponds to the transition from state  $S+1$  to state  $S$  due to an individual undergoing a death event. This occurs proportional to the death rate,  $P_d$ , the proportion of the lattice that is occupied by individuals,  $(S+1)/N$ , and the probability that the system is in state  $S+1$ ,  $f(S+1, t)$ . The third and fourth terms represent events that result in the system moving from state  $S$  to states  $S+1$  and  $S-1$ , corresponding to birth and death events, respectively. Note that the rate of movement,  $P_m$ , does not appear in equation (2.2), as movement events do not change the number of individuals in the system. Instead, the ratio of  $P_m/P_b$  and  $P_m/P_d$  influences

whether the assumption of independence between the occupancy of nearest-neighbour sites is appropriate [24].

The system has two boundary cases that occur at  $S = 1$ , where

$$\delta f(1, t) = \frac{2}{N} P_d f(2, t) - \frac{1}{N} (P_b + P_d) f(1, t),$$

implying that probability mass is lost to the system at  $S = 1$  corresponding to  $P_d f(1, t)/N$ ; and  $S = N$ , where

$$\delta f(N, t) = P_b \left( \frac{N-1}{N} \right) \left( 1 - \frac{N-2}{N-1} \right) f(N-1, t) - P_d f(N, t),$$

and hence no probability mass passes through  $S = N$ .

Introducing a change of variables  $S = Ns$ , where  $1/N \leq s \leq 1$  is the proportion of occupied sites [45,46], the system becomes

$$\begin{aligned} \delta f(s, t) = & P_b \frac{N}{N-1} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) f \left( s - \frac{1}{N}, t \right) + P_d \left( s + \frac{1}{N} \right) f \left( s + \frac{1}{N}, t \right) \\ & - P_b \frac{N}{N-1} s(1-s) f(s, t) - P_d s f(s, t). \end{aligned}$$

Expanding  $f(s, t)$  in a Taylor series, neglecting terms  $O(N^{-3})$ , and dividing by the time step  $\tau$ , we obtain

$$\begin{aligned} \frac{\delta f(s, t)}{\tau} = & \frac{1}{2N^2\tau} \left[ P_b \frac{N}{N-1} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) + P_d \left( s + \frac{1}{N} \right) \right] \frac{\partial^2 f(s, t)}{\partial s^2} \\ & - \frac{1}{N\tau} \left[ P_b \frac{N}{N-1} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) - P_d \left( s + \frac{1}{N} \right) \right] \frac{\partial f(s, t)}{\partial s} \\ & + \frac{1}{N\tau} \left[ P_b \frac{N}{N-1} \left( 2s - 1 - \frac{1}{N} \right) + P_d \right] f(s, t) + O(N^{-3}), \end{aligned}$$

or in conservative form,

$$\begin{aligned} \frac{\delta f(s, t)}{\tau} = & \frac{1}{N\tau} \frac{\partial}{\partial s} \left( \frac{1}{2N} \left[ P_b \frac{N}{N-1} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) + P_d \left( s + \frac{1}{N} \right) \right] \frac{\partial f(s, t)}{\partial s} \right. \\ & - \left. \left[ P_b \frac{N}{N-1} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) - P_d \left( s + \frac{1}{N} \right) \right] \right. \\ & \left. + \frac{1}{2N} \left\{ P_b \frac{N}{N-1} \left( 1 - 2s + \frac{2}{N} \right) + P_d \right\} f(s, t) \right). \end{aligned}$$

Taking the limit  $N \rightarrow \infty$  and  $\tau \rightarrow 0$  jointly such that  $N\tau$  is constant [47], and noting that, in practice,  $N$  and  $\tau$  are finite, we obtain

$$\begin{aligned} \frac{\partial f(s, t)}{\partial t} = & \frac{\partial}{\partial s} \left( \frac{1}{2N} \left[ P_b \frac{N}{N-1} \left( s - s^2 + \frac{2s}{N} \right) + P_d \left( s + \frac{1}{N} \right) \right] \frac{\partial f(s, t)}{\partial s} \right. \\ & - \left. \left[ P_b \frac{N}{N-1} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) - P_d \left( s + \frac{1}{N} \right) \right] \right. \\ & \left. + \frac{1}{2N} \left\{ P_b \frac{N}{N-1} \left( 1 - 2s + \frac{2}{N} \right) + P_d \right\} f(s, t) \right). \end{aligned}$$

Simplifying, the PDE gives

$$\frac{\partial f(s, t)}{\partial t} = \frac{\partial}{\partial s} \left( \frac{1}{2N} \left[ a(s) + b(s) \right] \frac{\partial f(s, t)}{\partial s} + \left[ b(s) - a(s) - \frac{1}{2N} \{ a'(s) + b'(s) \} \right] f(s, t) \right), \quad (2.3)$$

where

$$a(s) = P_b \frac{N}{N-1} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right), \quad (2.4)$$

and

$$b(s) = P_d \left( s + \frac{1}{N} \right), \quad (2.5)$$

noting that terms  $O(N^{-3})$  are neglected. The boundary conditions are given by

$$\frac{1}{2N} [a(s) + b(s)] \frac{\partial f(s, t)}{\partial s} + \left[ b(s) - a(s) - \frac{1}{2N} \{a'(s) + b'(s)\} \right] f(s, t) = \frac{P_d}{N} f(s, t) \text{ at } s = \frac{1}{N},$$

and

$$\frac{1}{2N} [a(s) + b(s)] \frac{\partial f(s, t)}{\partial s} + \left[ b(s) - a(s) - \frac{1}{2N} \{a'(s) + b'(s)\} \right] f(s, t) = 0 \text{ at } s = 1.$$

We note that the approach employed here is similar to obtaining diffusion approximations of infectious disease processes, where the size of the state space renders solution impractical [36,45, 46,48]. However, to the best of our knowledge, this type of approach has not been employed to analyse and approximate lattice-based random walks with exclusion.

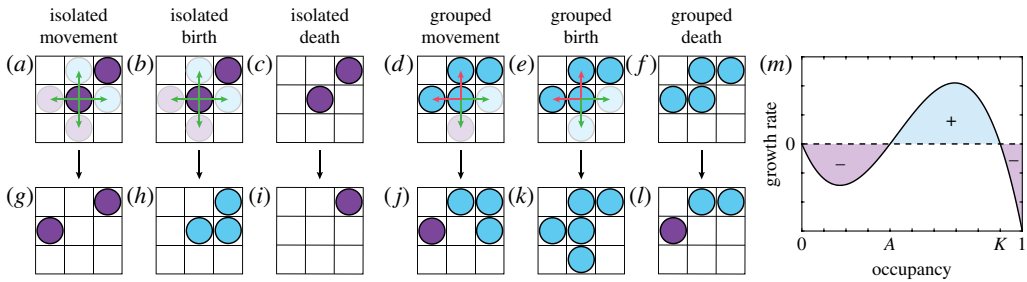
Consistent with the hyperbolic scaling imposed in the SSDA, we expect advection to provide the dominant contribution to transport of the probability distribution. That is, we expect the initial condition to evolve primarily according to the advective velocity  $b(s) - a(s)$ . While this does neglect the contribution of diffusive transport, this approximation is appropriate provided that  $N$  is large, as the diffusive contribution is inversely proportional to  $N$  (see equation (2.3)). As such, we expect the mode of the probability distribution,  $s_{\text{mo}}$ , to evolve from the initial location of the peak of the initial condition to the stable steady state of

$$\frac{ds_{\text{mo}}}{dt} = a(s_{\text{mo}}) - b(s_{\text{mo}}).$$

Note that this is similar to the standard mean-field ODE, but that this describes the *mode* of the probability distribution, rather than the *mean* of the probability distribution.

## (b) Allee model

We now consider a two-dimensional birth–death–movement random walk with exclusion on a lattice with  $N$  sites [30]. Here the birth and death rates depend on whether an individual agent has zero nearest-neighbour agents or at least one nearest-neighbour agent, referred to as isolated and grouped individuals, respectively [14,25] (figure 2*a–l*). These rates are  $P_b^i, P_b^g, P_d^i$  and  $P_d^g$  for birth and death of isolated and grouped individuals, respectively. This corresponds to either a competitive or co-operative process, depending on whether it is beneficial to be part of a group of individuals or to be isolated [25]. It has been shown that in one dimension these birth and death mechanisms give rise to an Allee effect [25] and, in two dimensions, these mechanisms give rise to a *per capita* growth rate that is qualitatively similar to the Allee effect [14]. In certain parameter regimes, growth rates can be obtained that are qualitatively similar to the weak, strong and hyper Allee effects [14]. The weak Allee effect refers to a *per capita* growth rate where growth is inhibited, but remains positive, at low individual densities [17,41]. The strong Allee effect refers to the case where growth is negative below a threshold individual density, and positive otherwise [41,49]. The hyper Allee effect refers to per capita growth rates that may have additional steady states [14]. As the strong Allee effect is perhaps the more interesting model with respect to population extinction, as fluctuations in the number of individuals that take the population below the threshold density can result in extinction [49], we focus on parameter suites that correspond to the strong Allee effect, noting that it is straightforward to analyse other parameter regimes. In figure 2*m*, we present the growth curve as a function of the proportion of occupied lattice sites. Under the standard mean-field assumption for a spatially uniform initial condition and periodic boundary conditions, the evolution of the average proportion of occupied sites,  $0 \leq \bar{s} \leq 1$ , for the



**Figure 2.** Mechanisms in the Allee random walk model. (a)–(c) Potential movement, birth and death events, respectively, for the isolated individual in the centre of the lattice. (d)–(f) Potential movement, birth and death events, respectively, for the grouped individual in the centre of the lattice. Green arrows indicate directions that would result in a successful event, red arrows indicate directions that would result in an unsuccessful event. (g)–(l) Lattice configuration after the corresponding event in (a)–(f). Isolated individuals are shown in purple, grouped individuals are shown in cyan. (m) Typical growth rate curve as a function of the occupancy of the lattice. Cyan regions correspond to positive growth and purple regions correspond to negative growth.  $A$  is the threshold parameter and  $K$  is the carrying capacity. (Online Version in colour.)

strong Allee effect is [25,41]

$$\frac{d\bar{S}}{dt} = r\bar{S}(K - \bar{S})(C(\bar{S}) - C(A)), \quad (2.6)$$

where  $r > 0$  is the net birth rate,  $0 \leq K \leq 1$  is the carrying capacity,  $0 < A < K$  is the threshold parameter, and  $C(\bar{S})$  satisfies  $C(\bar{S}) < C(A)$  for  $0 \leq \bar{S} < A$ , and  $C(\bar{S}) > C(A)$  for  $A < \bar{S} \leq K$ . The most common form of the strong Allee effect [25,41] has  $C(\bar{S}) = \bar{S}$ , though we note that the qualitative dynamics do not change provided that the above conditions for  $C(\bar{S})$  are satisfied.

For the strong Allee effect, it is known that the standard mean-field ODE is inaccurate near the unstable steady state, where the growth rate transitions from negative to positive [14]. As above, we consider how the system evolves in terms of the state variable, the number of occupied sites, during a single time step:

$$\begin{aligned} \delta f(S, t) = & P_b^g \frac{S-1}{N} \left(1 - \frac{S-2}{N-1}\right) \left(1 - \left[1 - \frac{S-2}{N-2}\right] \left[1 - \frac{S-2}{N-3}\right] \left[1 - \frac{S-2}{N-4}\right]\right) f(S-1, t) \\ & - P_b^g \frac{S}{N} \left(1 - \frac{S-1}{N-1}\right) \left(1 - \left[1 - \frac{S-1}{N-2}\right] \left[1 - \frac{S-1}{N-3}\right] \left[1 - \frac{S-1}{N-4}\right]\right) f(S, t) \\ & + P_b^i \frac{S-1}{N} \left(1 - \frac{S-2}{N-1}\right) \left(1 - \frac{S-2}{N-2}\right) \left(1 - \frac{S-2}{N-3}\right) \left(1 - \frac{S-2}{N-4}\right) f(S-1, t) \\ & - P_b^i \frac{S}{N} \left(1 - \frac{S-1}{N-1}\right) \left(1 - \frac{S-1}{N-2}\right) \left(1 - \frac{S-1}{N-3}\right) \left(1 - \frac{S-1}{N-4}\right) f(S, t) \\ & + P_d^g \frac{S+1}{N} \left(1 - \left[1 - \frac{S}{N-1}\right] \left[1 - \frac{S}{N-2}\right] \left[1 - \frac{S}{N-3}\right] \left[1 - \frac{S}{N-4}\right]\right) f(S+1, t) \\ & - P_d^g \frac{S}{N} \left(1 - \left[1 - \frac{S-1}{N-1}\right] \left[1 - \frac{S-1}{N-2}\right] \left[1 - \frac{S-1}{N-3}\right] \left[1 - \frac{S-1}{N-4}\right]\right) f(S, t) \\ & + P_d^i \frac{S+1}{N} \left(1 - \frac{S}{N-1}\right) \left(1 - \frac{S}{N-2}\right) \left(1 - \frac{S}{N-3}\right) \left(1 - \frac{S}{N-4}\right) f(S+1, t) \\ & - P_d^i \frac{S}{N} \left(1 - \frac{S-1}{N-1}\right) \left(1 - \frac{S-1}{N-2}\right) \left(1 - \frac{S-1}{N-3}\right) \left(1 - \frac{S-1}{N-4}\right) f(S, t). \end{aligned}$$

Note that if  $P_b^g = P_b^i$  and  $P_d^g = P_d^i$ , the above system reduces to the logistic model. Following the same limit approach as previously, we obtain the following PDE describing the distribution of the

state of the random walk process

$$\frac{\partial f(s, t)}{\partial t} = \frac{\partial}{\partial s} \left( \frac{1}{2N} [a(s) + b(s)] \frac{\partial f(s, t)}{\partial s} + \left[ b(s) - a(s) - \frac{1}{2N} \{a'(s) + b'(s)\} \right] f(s, t) \right), \quad (2.7)$$

where

$$\begin{aligned} a(s) = & P_b^g \frac{N^4}{(N-1)(N-2)(N-3)(N-4)} \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) \\ & \times \left( \frac{(N-2)(N-3)(N-4)}{N^3} - [1-s] \left[ 1 - s - \frac{1}{N} \right] \left[ 1 - s - \frac{2}{N} \right] \right) \\ & + P_b^i \frac{N^4}{(N-1)(N-2)(N-3)(N-4)} \\ & \times \left( s - \frac{1}{N} \right) \left( 1 - s + \frac{1}{N} \right) (1-s) \left( 1 - s - \frac{1}{N} \right) \left( 1 - s - \frac{2}{N} \right), \end{aligned} \quad (2.8)$$

and

$$\begin{aligned} b(s) = & P_d^g \frac{N^4}{(N-1)(N-2)(N-3)(N-4)} \left( s + \frac{1}{N} \right) \\ & \times \left( \frac{(N-1)(N-2)(N-3)(N-4)}{N^4} - \left[ 1 - s - \frac{1}{N} \right] \left[ 1 - s - \frac{2}{N} \right] \left[ 1 - s - \frac{3}{N} \right] \left[ 1 - s - \frac{4}{N} \right] \right) \\ & + P_d^i \frac{N^4}{(N-1)(N-2)(N-3)(N-4)} \\ & \times \left( s + \frac{1}{N} \right) \left( 1 - s - \frac{1}{N} \right) \left( 1 - s - \frac{2}{N} \right) \left( 1 - s - \frac{3}{N} \right) \left( 1 - s - \frac{4}{N} \right). \end{aligned} \quad (2.9)$$

The boundary conditions are given by

$$\frac{1}{2N} [a(s) + b(s)] \frac{\partial f(s, t)}{\partial s} + \left[ b(s) - a(s) - \frac{1}{2N} \{a'(s) + b'(s)\} \right] f(s, t) = P_d^i f(s, t) \text{ at } s = \frac{1}{N}$$

and

$$\frac{1}{2N} [a(s) + b(s)] \frac{\partial f(s, t)}{\partial s} + \left[ b(s) - a(s) - \frac{1}{2N} \{a'(s) + b'(s)\} \right] f(s, t) = 0 \text{ at } s = 1.$$

Note that the boundary condition at  $s = 1/N$  only depends on  $P_d^i$  as it is impossible to have any grouped agents at this  $s$  value, which corresponds to a single occupied lattice site. Similar to the logistic model, the standard approach to examine how the system evolves is to use a mean-field approximation, which results in the following ODE [14,25]

$$\frac{d\bar{S}}{dt} = P_b^g \bar{S}(1 - \bar{S})(1 - (1 - \bar{S})^3) + P_b^i \bar{S}(1 - \bar{S})^4 - P_d^g \bar{S}(1 - (1 - \bar{S})^4) - P_d^i \bar{S}(1 - \bar{S})^4. \quad (2.10)$$

Note that equation (2.10) is equivalent to equation (2.6) in parameter regimes that correspond to the strong Allee effect. In these parameter regimes,  $C(\bar{S})$  is a cubic function such that  $C(\bar{S}) < C(A)$  for  $0 \leq \bar{S} < A$  and  $C(\bar{S}) > C(A)$  for  $A < \bar{S} \leq K$ .

In general, the form of the PDE arising from a lattice-based birth–death–movement random walk process is

$$\frac{\partial f(s, t)}{\partial t} = \frac{\partial}{\partial s} \left( \frac{1}{2N} [a(s) + b(s)] \frac{\partial f(s, t)}{\partial s} + \left[ b(s) - a(s) - \frac{1}{2N} \{a'(s) + b'(s)\} \right] f(s, t) \right),$$

where  $a(s)$  are the coefficients of the  $f(S - 1, t)$  terms and  $b(s)$  are the coefficients of the  $f(S + 1, t)$  terms, cast with respect of the scaled variable  $s$  rather than  $S$ . That is, the coefficients represent the probability of transitioning from  $S - 1$  to  $S$  and from  $S + 1$  to  $S$ , respectively.



### 3. Results

We first highlight the stochastic nature of the random walk, and the resulting need for an approximation technique that provides the distribution of the number of occupied sites. In figure 3, we present simulation results for both logistic and Allee models where, for two identically prepared simulations, two completely different lattice configurations arise at  $t = T_{\text{end}}$ . In one realization, the population persists (figure 3*b,f*), while in another realization, the population undergoes extinction (figure 3*c,g*). This information is not captured by the corresponding mean-field ODE, presented in figure 3*d,h*, which only provides information regarding the average occupancy. However, as detailed in §2, the SSDA explicitly accounts for population extinction and describes the distribution of the proportion of occupied sites. We refer to the latter throughout as the occupancy distribution.

We next verify that the predictions obtained from the SSDA are consistent with the average random walk behaviour and the mean-field ODE for parameter regimes where it is known that the mean-field approximation is appropriate, and where we expect the rate of population extinction to be small. Specifically, these parameter regimes correspond to regimes where  $P_m/P_b \gg 1$  and  $P_m/P_d \gg 1$  [24]. Details of the numerical techniques employed to obtain solutions to the SSDA PDE are provided in appendix A. In figure 4*a,d*, we demonstrate that the evolution of the average occupancy obtained from the solution to the SSDA matches the average occupancy obtained from both the random walk and the mean-field ODE for both the logistic and Allee models. Results are obtained from  $10^4$  identically prepared realizations of the random walk. The average occupancy obtained from the SSDA is calculated via

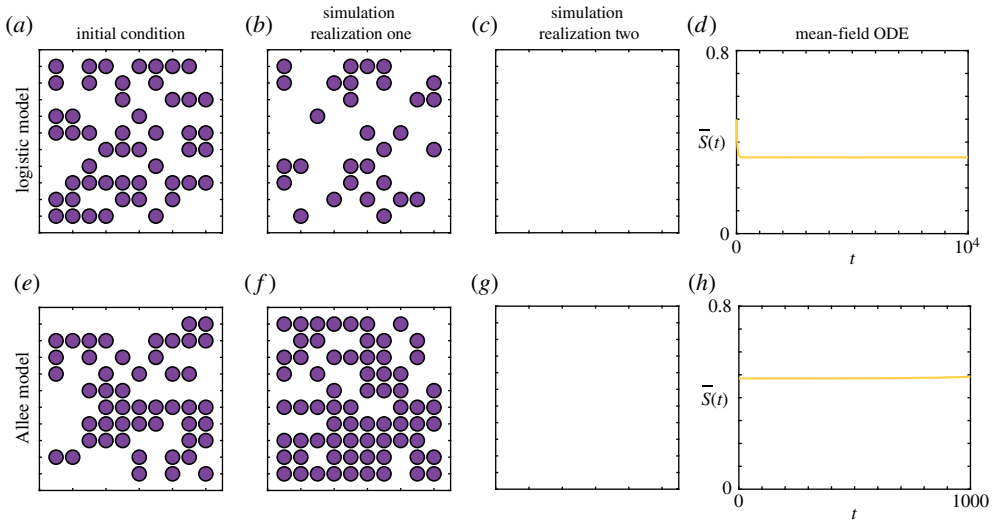
$$\bar{S}(t) = \sum_i s_i f(s_i, t) \Delta s,$$

where  $s_i$  are the values of  $s$  where a grid point is defined in the numerical solution, and  $\Delta s$  is the space between grid points (appendix A). The SSDA provides additional information, compared to the mean-field ODE, with respect to the occupancy distribution, presented in figure 4*b,e*, and the probability that the population has undergone extinction by time  $t$ ,  $P_E(t)$ , presented in figure 4*c,f*. The extinction probability is calculated via

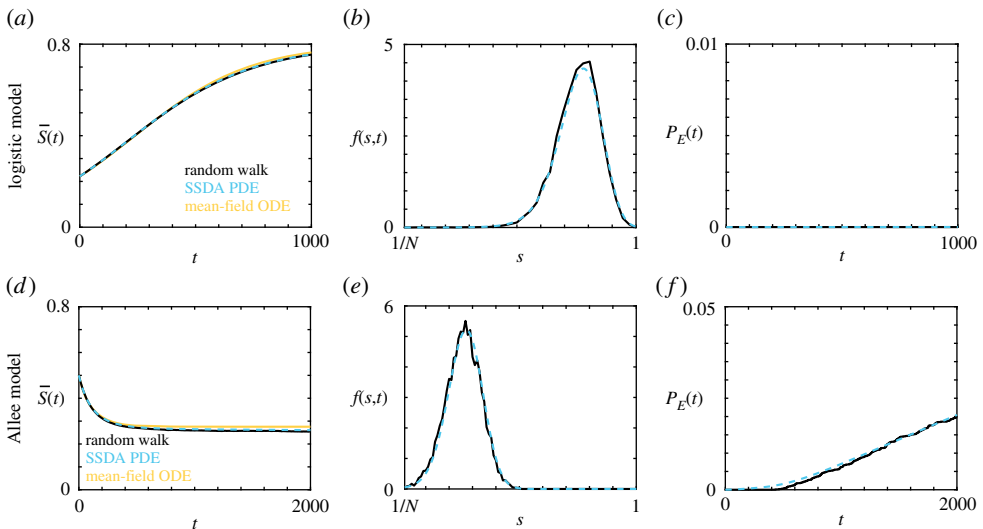
$$P_E(t) = \sum_i f(s_i, 0) \Delta s - \sum_i f(s_i, t) \Delta s,$$

that is, the difference in probability mass between the initial occupancy distribution and the occupancy distribution at time  $t$ . Note that throughout this work  $t$  refers to the number of time steps of arbitrary length in the underlying random walk, and does not refer to any particular physical time scale. For both the occupancy distribution and the probability of population extinction, the predictions obtained from the SSDA match the average random walk behaviour.

We now consider two parameter regimes for the logistic model where the possible extinction of the population has a significant impact, calculate the evolution of the average occupancy, the occupancy distribution and the probability of extinction, and present the results in figure 5. Note that for all cases we choose parameters such that  $P_m/P_b \gg 1$  and  $P_m/P_d \gg 1$  to satisfy the mean-field approximation [24]. In all cases, results obtained from the SSDA match the average behaviour of the random walk well. In the first regime, presented in figure 5*a-c*, the birth rate is twice the death rate, and hence the positive steady state in equation (2.1) occurs at  $\bar{S} = 1/2$ . Interestingly, despite the steady state existing well away from  $\bar{S} = 0$ , there is a non-zero chance that the population undergoes extinction. As such, the mean-field ODE overestimates the average occupancy in the random walk as time increases, as this approximation does not account for population extinction. At early time, it is unlikely that the population has undergone extinction, so the mean-field ODE provides a reasonable approximation. By contrast, the SSDA accounts for population extinction, and hence accurately reflects the average occupancy for all  $t$  considered. The non-zero extinction chance can be observed in the occupancy distribution (figure 5*b*). The distribution is centred around  $s = 1/2$  but the distribution is broad relative to the range of

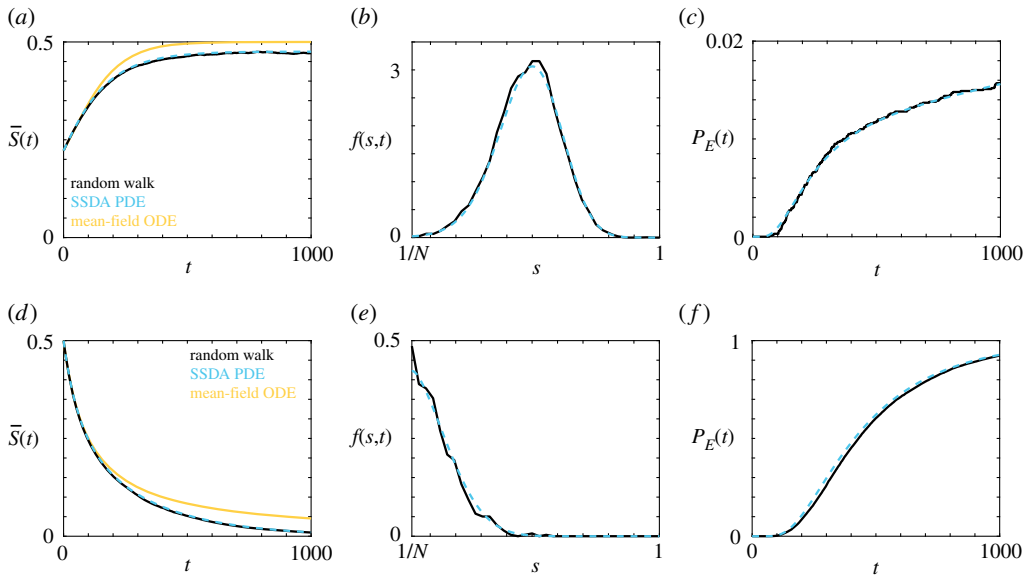


**Figure 3.** Differences in the random walk obtained from identically-prepared initial conditions for (a)–(d) logistic growth model and the (e)–(h) Allee model. (a),(e) The initial condition in the random walk model. (b),(c),(f),(g) The final result of the random walk model for two realizations. (d),(h) The evolution of average occupancy obtained from the mean-field ODE. Parameters used are (a)–(d)  $P_m = 1$ ,  $P_b = 0.075$ ,  $P_d = 0.05$ ,  $T_{\text{end}} = 10^4$ ,  $N = 100$ ,  $N_0 = 50$  and (e)–(h)  $P_m = 1$ ,  $P_b^i = 0.01$ ,  $P_b^g = 0.01$ ,  $P_d^i = 0.04$ ,  $P_d^g = 0.0025$ ,  $T_{\text{end}} = 1000$ ,  $N = 100$ ,  $N_0 = 49$ . (Online Version in colour.)



**Figure 4.** Comparison between the average behaviour of the lattice-based random walk (black), the SSDA PDE (cyan, dashed) and the mean-field ODE (orange) for the (a)–(c) logistic growth model and the (d)–(f) Allee model. (a),(d) The evolution of the average occupancy. (b),(e) The occupancy distribution probability density function. (c),(f) The evolution of the extinction probability. Parameters used are (a)–(c)  $P_m = 1$ ,  $P_b = 0.005$ ,  $P_d = 0.001$ ,  $T_{\text{end}} = 1000$ ,  $N = 36$ ,  $N_0 = 8$  and (d)–(f)  $P_m = 1$ ,  $P_b^i = 0.03$ ,  $P_b^g = 0.01$ ,  $P_d^i = 0.02$ ,  $P_d^g = 0.01$ ,  $T_{\text{end}} = 2000$ ,  $N = 100$ ,  $N_0 = 50$ . Results are obtained from  $10^4$  identically prepared realizations of the random walk. (Online Version in colour.)

possible average occupancy values. As such, the flux through  $s = 1/N$  is non-negligible. In the second parameter regime, presented in figure 5d–f, there is no positive steady-state solution to equation (2.1), and we expect the population to tend toward extinction. Both the mean-field ODE

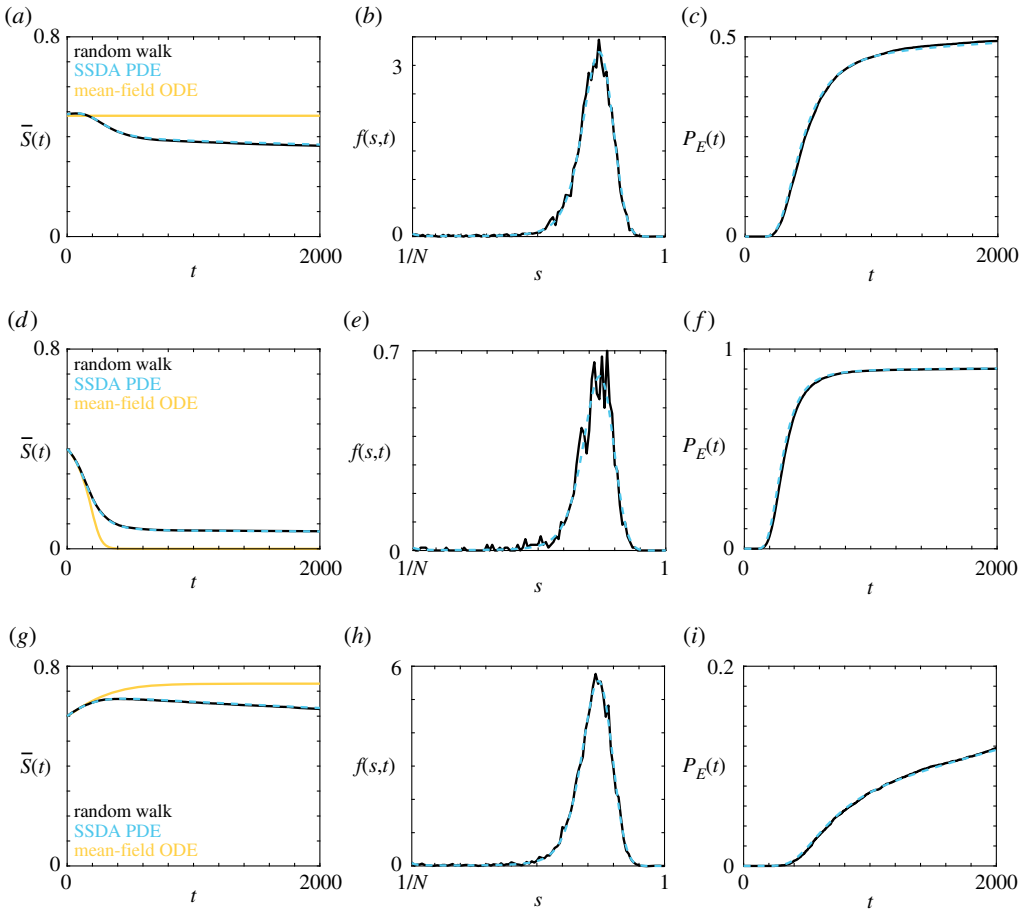


**Figure 5.** Comparison between the average behaviour of the lattice-based random walk (black), the SSDA PDE (cyan, dashed) and the mean-field ODE (orange) for the logistic growth model. (a,d) The evolution of the average occupancy. (b),(e) The occupancy distribution probability density function. (c),(f) The evolution of the extinction probability. Parameters used are  $P_m = 1, P_b = 0.02, P_d = 0.01, T_{\text{end}} = 1000, N = 36, N_0 = 8$  and (d)–(f)  $P_m = 1, P_b = 0.02, P_d = 0.02, T_{\text{end}} = 1000, N = 36, N_0 = 18$ . Results are obtained from  $10^4$  identically prepared realizations of the random walk. (Online Version in colour.)

and the SSDA predict that the average occupancy will tend to zero; however, the mean-field ODE significantly overestimates the average occupancy compared to the SSDA. For example, at  $t = 1000$ , the mean-field ODE predicts an average occupancy that is approximately five times larger than that obtained from both the random walk and the SSDA. Further, it is unclear how the average occupancy is related to the probability that the population has undergone extinction, and hence the additional information provided by the SSDA is instructive in this regard. By  $t = 1000$ , there is approximately a 92% chance that an individual population will have gone extinct and, for the non-extinct populations, the probability density of the occupancy distribution is monotonically decreasing with respect to the number of occupied sites. This indicates that for non-extinct populations, it is likely that the population is nearly extinct. This is in contrast to the results in figure 5b, where non-extinct populations are not close to extinction.

Next, we consider three cases of the Allee model. In each case, we maintain the same rates of birth and death; only the number of initially occupied sites changes. The choice of birth and death rates is for illustrative purposes only; we could select other parameter values provided that the parameters result in a strong Allee effect. The three different numbers of initially occupied sites correspond to initial average occupancies that are at  $\bar{S}(0) = A$ , figure 6a–c), below  $\bar{S}(0) < A$ , figure 6d–f) and above  $\bar{S}(0) > A$ , figure 6g–i) the unstable steady state. As such, the mean-field ODE, equation (2.6), predicts that the evolution of the average occupancy should remain on the unstable steady state, tend towards extinction and tend towards the carrying capacity, respectively.

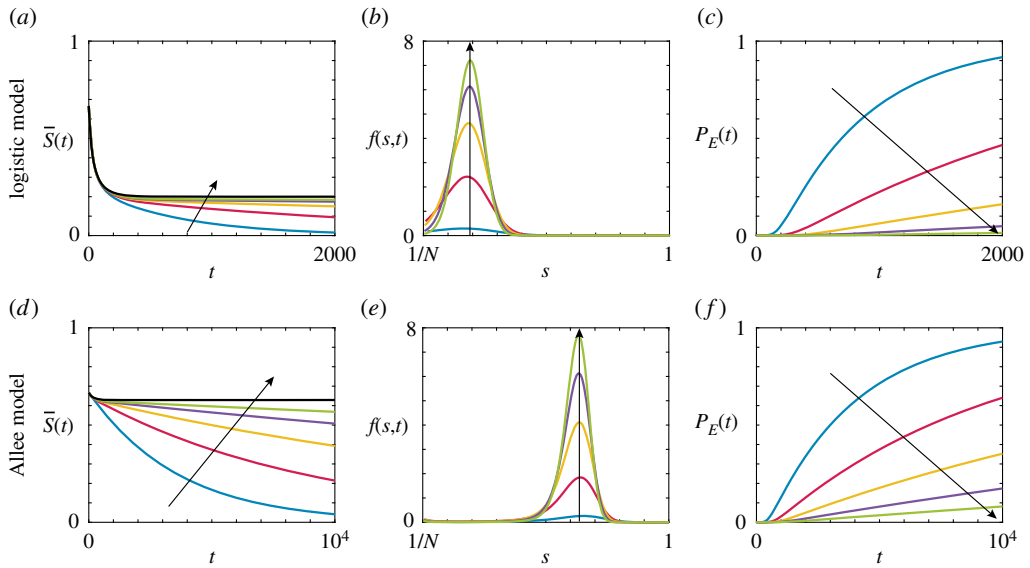
For the initial occupancy corresponding to  $\bar{S}(0) = A$ , we observe that the average occupancy obtained from both the SSDA and the random walk tends below  $\bar{S} = A$  as time increases, but does not rapidly undergo extinction. This result is perhaps somewhat unintuitive based solely on the evolution of the average occupancy, but is clearer when considered in context of both the extinction probability and the occupancy distribution. Interestingly, the occupancy distribution is still centred around the carrying capacity  $\bar{S} = K$ , despite the average occupancy being below the



**Figure 6.** Comparison between the average behaviour of the lattice-based random walk (black), the SSDA PDE (cyan, dashed) and the mean-field ODE (orange) for the Allee model. (a,d,g) The evolution of the average occupancy. (b,e,h) The occupancy distribution probability density function. (c,f,i) The evolution of the extinction probability. Parameters used are  $P_m = 1$ ,  $P_b^i = 0.01$ ,  $P_b^j = 0.01$ ,  $P_d^i = 0.04$ ,  $P_d^j = 0.0025$ ,  $T_{\text{end}} = 2000$ ,  $N = 100$  and (a)–(c)  $N_0 = 49$ , (d)–(f)  $N_0 = 40$ , (g)–(i)  $N_0 = 60$ . Results are obtained from  $10^4$  identically prepared realizations of the random walk. (Online Version in colour.)

unstable steady state. Combined with the probability of extinction at  $t = 2000$  of approximately 50%, this implies that in an individual realization of a random walk one of two possibilities occur: either the population will become extinct or the population will tend to the carrying capacity. Due to this dichotomy, reporting the average occupancy is not an informative measure of average population behaviour. Instead, both the occupancy distribution and the extinction probability are required to understand the population behaviour.

Similar results are observed for the initial average occupancies corresponding to  $\bar{S}(0) < A$  and  $\bar{S}(0) > A$ . The average occupancy in the case where  $\bar{S}(0) < A$  appears to tend to a finite positive value in the random walk, and the corresponding SSDA. By contrast, the mean-field ODE predicts that the average occupancy will be near-zero by  $t = 400$ . Again, this discrepancy can be understood by considering the occupancy distribution, and the extinction probability. Here, a single population will become extinct approximately 90% of the time, but will otherwise tend to the carrying capacity. This is again observed for the case where  $\bar{S}(0) > A$ , albeit with a much lower probability of extinction due to the higher initial occupancy. We observe that the occupancy distribution is extremely similar for all three initial occupancies, to a scaling factor. This suggests that provided a population does not go extinct, we expect the population to exist

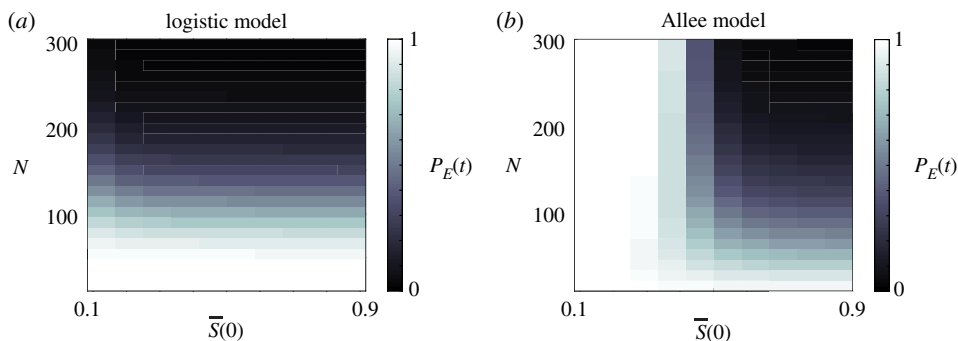


**Figure 7.** SSDA predictions for the (a,d) the evolution of the average occupancy, (b,e) the occupancy distribution probability density function, and (c,f) the evolution of the extinction probability for (a)–(c) the logistic model and (d)–(f) the Allee model for a range of  $N$  values. Arrow indicates direction of increasing  $N$ . Parameters used are (a)–(c)  $P_m = 1, P_b = 0.05, P_d = 0.04, T_{\text{end}} = 2000$  and (d)–(f)  $P_m = 1, P_b^i = 0.02, P_b^o = 0.015, P_d^i = 0.04, P_d^o = 0.005, T_{\text{end}} = 10^4$ . Coloured lines indicate different population sizes: (blue)  $N = 60, N_0 = 40$ , (red)  $N = 120, N_0 = 80$ , (orange)  $N = 180, N_0 = 120$ , (purple)  $N = 240, N_0 = 180$ , (green)  $N = 300, N_0 = 240$ . The grey line represents the mean-field ODE solution, that is,  $N \rightarrow \infty$ . (Online Version in colour.)

near the carrying capacity. Note that the additional noise in figure 6(e) is due to the higher probability of extinction. Further, the probability that a population goes extinct depends on the initial occupancy. However, this dependence is less strict than in the standard Allee model, equation (2.6), where the population will tend to extinction with certainty if  $\bar{S}(0) < A$  and will tend to the carrying capacity with certainty if  $\bar{S}(0) > A$ .

As the difference between the SSDA and the standard mean-field approximation arises from whether the population size is finite or infinite, it is instructive to examine how different population sizes impact the average occupancy, occupancy distribution and the probability of extinction. We consider both the logistic model and the Allee model for a range of population sizes, with a consistent proportion of the lattice initially occupied. The parameter regimes considered have a stable steady state that is below the initial average occupancy but occurs at positive average occupancy. As such, we expect that the average occupancy should tend towards this positive value, aside from extinction events associated with finite population size. For the results obtained from the logistic model, presented in figure 7a–c, we observe that the size of the population has a dramatic effect. For the smallest population ( $N = 60$ ), we observe that the average occupancy is close to zero by  $t = 2000$  and that the extinction probability is approaching one. Increasing the population size, we observe that the average occupancy increases, corresponding to a decrease in extinction probability. Further, we see that the spread in the occupancy distribution decreases with an increase in population size. This corresponds to a smaller proportion of this distribution occurring near the boundary corresponding to extinction. The population size has a similar effect in the Allee model, as observed in the results presented in figure 7d–f. As noted in the results presented in figure 6, provided that the population does not undergo extinction, the occupancy distribution is centred around the stable steady state.

Finally, we evaluate the probability that a population will undergo extinction by a particular time for a suite of population sizes and initial average occupancies for both the logistic and Allee models, and present the results in figure 8. For the logistic model, we observe that increasing the



**Figure 8.** SSDA predictions for the probability of population extinction by (a)  $T_{\text{end}} = 2000$  and (b)  $T_{\text{end}} = 10^4$  for (a) the logistic model and (b) the Allee model. Parameters used are (a)  $P_m = 1, P_b = 0.05, P_d = 0.04$  and (b)  $P_m = 1, P_b^i = 0.02, P_b^g = 0.015, P_d^i = 0.04, P_d^g = 0.005$ . (Online Version in colour.)

**Table 1.** Comparison between the time taken, in seconds, to perform  $10^4$  realizations of the random walk process and the time taken to obtain the numerical solution to the SSDA PDE. All simulations were performed on a 3.1 GHz Intel i7 Processor (7920HQ) in Matlab R2018b.

result	figure 4(a)–(c)	figure 4(d)–(f)	figure 5(a)–(c)	figure 5(d)–(f)	figure 6(a)–(c)	figure 6(d)–(f)	figure 6(g)–(i)
random walk	$1.17 \times 10^3$ s	$7.09 \times 10^2$ s	$7.09 \times 10^2$ s	$1.93 \times 10^2$ s	$3.43 \times 10^3$ s	$8.87 \times 10^2$ s	$5.27 \times 10^3$ s
SSDA PDE	$1.03 \times 10^0$ s	$2.41 \times 10^0$ s	$1.94 \times 10^0$ s	$1.93 \times 10^0$ s	$1.79 \times 10^0$ s	$1.73 \times 10^0$ s	$1.78 \times 10^0$ s

population size corresponds to a decrease in the extinction probability. Interestingly, the initial occupancy is less important than the number of sites. This is consistent with the observation that the occupancy distribution will approach a quasi-steady-state distribution on a faster timescale than extinction [42]. For the Allee model, the initial average occupancy is important due to the presence of the Allee threshold. If the initial average occupancy is below this threshold density, the population will tend to extinction. However, if the population size is sufficiently small, fluctuations in average occupancy can result in the population size crossing this threshold, and hence a portion of populations will tend to the carrying capacity. Similarly, if the initial average occupancy is above the Allee threshold density, the majority of populations will tend to the carrying capacity. The exact size of this majority depends on the distance between the initial average occupancy and the Allee threshold density, as well as the population size. This can be observed in figure 8b, where increasing both the initial average occupancy and the population size results in a decrease in the probability of extinction.

In table 1, we present a comparison of the computation time required to perform sufficient realizations of the random walk process to obtain representative average behaviour, and the time required to obtain a numerical solution to the SSDA PDE, for the results presented in figures 4–6. For all cases, we observe that it is significantly faster, up to three orders of magnitude, to obtain predictions from the SSDA rather than the random walk.

## 4. Discussion and conclusion

Determining whether a population will undergo extinction is relevant throughout biology and ecology [1–3,6–9]. Population extinction may be beneficial, such as in the case of treating a malignant tumour cell population [10], or detrimental, such as in the case of endangered animal populations [3,8]. However, in both cases, being able to accurately predict whether population extinct will occur is invaluable. Lattice-based random walks with finite size effects are widely used to model the dynamics of populations of individuals [14,27,32,33]. Performing

sufficiently many realizations of such random walks to obtain representative average behaviour is computationally intensive, which has motivated the use of continuum approximations of random walks [24]. However, previous continuum approximations are incapable of predicting population extinction.

The new continuum approximation of a lattice-based birth–death–movement random walk presented here is capable of describing and predicting population extinction. This approximation involves a state-space representation of the lattice-based random walk, obtained via a mean-field assumption, followed by a continuum scaling of both the state space and time. Previous continuum approximations of lattice-based random walks have focused on the average occupancy of the lattice, which is ill-suited for analysing extinction [24,26]. By contrast, the SSDA explicitly accounts for the possibility that an individual random walk will undergo extinction via the distribution of the number of lattice sites occupied by individuals.

We have demonstrated that the SSDA faithfully captures the behaviour in the corresponding random walk for two common models of population dynamics, the logistic growth model and the Allee growth model [14,24,25]. Specifically, we have shown that the evolution of average occupancy obtained from the SSDA matches the average occupancy obtained from the corresponding random walk. This is in contrast with the standard mean-field ODE, which provides inaccurate predictions of the evolution of average occupancy. Further, the SSDA provides further relevant information about the average population behaviour, namely, the occupancy distribution and the probability that the population has undergone extinction. Neither of these measures can be obtained from the standard mean-field ODE. We have demonstrated that this additional information accurately reflects the behaviour in the corresponding random walk. Finally, we have highlighted the benefit of using the SSDA rather than the random walk by comparing the computational time required to obtain representative population behaviour.

Significant research effort has been directed towards continuum approximations of lattice-based random walks [14,23,24,26,35,44]. In particular, the vast majority of previous work has investigated the validity of the standard mean-field assumption for a suite of biologically inspired mechanisms [24,26,35,44]. It is well known that the mean-field assumption is satisfied provided that the positions of individuals in the population are not correlated [24,50]. That is, the probability that a particular lattice site is occupied does not depend on the occupancy of other lattice sites. However, certain mechanisms in the random walk induce short-range correlation [33]. For example, the birth mechanism in the random walk considered here results in two individuals occupying neighbouring lattice sites. Hence, without sufficiently many movement events, compared to birth events, the mean-field assumption will be invalid due to this short range correlation. To address this issue, corrected mean-field models have been proposed [24,26,35,44,51]. Corrected mean-field models consider the evolution of pairs of lattice sites, and hence can account for correlation between lattice sites. These models, and similar approaches, extend the parameter regimes under which the random walk can be accurately approximated [24,26,35,44,51]. However, all of these models have focused on the issue of correlation in the random walk, and how to select an appropriate continuum approximation, based on the strength of correlation between lattice sites in the underlying random walk.

In contrast, the continuum approximation considered here addresses the issue of population extinction. In standard continuum approximations, the evolution of the occupancy of the lattice is described via a continuum equation [24,26,44,51]. As such, the average occupancy will never be exactly zero, and will eventually approach the stable steady state in the continuum approximation equation. However, in the corresponding random walk, there is a non-zero probability, monotonically increasing in time, that the population has undergone extinction. As we have demonstrated in this work, this extinction probability significantly impacts the evolution of the average occupancy, even if the mean-field assumption is valid. Therefore, to accurately predict the average occupancy, the probability that the population undergoes extinction must be considered. We expect that a technique that incorporates *both* the probability of population extinction and the correlation in lattice site occupancy will be necessary to accurately

predict population behaviour for parameter suites where the mean-field assumption is not valid. However, we leave this investigation for future work.

The work presented can here can be extended in several directions. While we have only considered a population of individuals that is a single species, a natural extension would be to consider more than one species [51]. Increasing the number of species allows for the description of more detailed individual behaviour, such as predation. However, the size of the state space increases exponentially with the number of species; an  $n$ -species model would have a state space of size  $O(N^n)$ . This quickly becomes intractable to analyse via the state space representation, necessitating the use of an appropriate continuum approximation [50]. The combination of the SSDA and a multispecies model would allow for the analysis of competition between species, and which species undergo extinction or survive. Alternatively, the PDE governing the population behaviour, equation (2.3), could be investigated analytically via separation of variables or an asymptotic expansion to obtain insight into the late-time population behaviour. Another potential future direction would be to develop further mechanisms in the underlying random walk model, informed by experimental data. This approach would not only develop new partial differential equations, but may also provide insight into experimental mechanisms that were hitherto unclear [14]. Finally, we could develop an extended version of the SSDA that preserves information about the location of individuals in the random walk in order to describe invasion-type processes, such as the spread of invasive species [25,41]. However, such an extension is beyond the scope of this work.

**Data accessibility.** The code used to generate the results presented here can be found on Github at <https://github.com/DrStuartJohnston/state-space-diffusion-approximation>

**Authors' contributions.** S.T.J. conceived the study, performed the analysis and numerical experiments and also wrote the manuscript. S.T.J., M.J.S. and E.J.C. designed the analysis and numerical experiments and edited the manuscript. All authors gave final approval for publication.

**Competing interests.** We declare we have no competing interests.

**Funding.** This work was in part supported by the Australian Research Council Centre of Excellence in Convergent Bio-Nano Science and Technology (project no. CE140100036) and the Australian Research Council (grant nos. DP170100474, DE200100988).

**Acknowledgements.** The authors thank the two anonymous referees for their helpful feedback.

## Appendix A.

### (a) Numerical solution of the SSDA PDE

To obtain numerical solutions of the SSDA PDE, we first apply finite difference techniques to the spatial derivatives in the governing equation, using a central difference approximation [52]. For all results, we choose a node spacing of  $\Delta s = 10^{-5}$ . The node spacing is selected such that the numerical solution was insensitive to further reduction in the node spacing. For the temporal derivative, we implement the backwards Euler method [52] with time step  $\Delta t = T_{\text{End}}/1000$ . The resulting tridiagonal system of algebraic equations is solved using the Thomas algorithm [52].

### (b) Numerical solution of the mean-field ODE

To obtain numerical solutions of the mean-field ODE, we use Matlab's inbuilt *ode45* routine, which implements an adaptive time step Runge–Kutta method [53].

## References

1. Anderson RM, Anderson B, May RM. 1992 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
2. Axelrod R, Axelrod DE, Pienta KJ. 2006 Evolution of cooperation among tumor cells. *Proc. Natl Acad. Sci. USA* **103**, 13 474–13 479. (doi:10.1073/pnas.0606053103)



3. Berger J. 1990 Persistence of different-sized populations: an empirical assessment of rapid extinctions in bighorn sheep. *Conserv. Biol.* **4**, 91–98. (doi:10.1111/j.1523-1739.1990.tb00271.x)
4. Brightwell G, House T, Luczak M. 2018 Extinction times in the subcritical stochastic SIS logistic epidemic. *J. Math. Biol.* **77**, 455–493. (doi:10.1007/s00285-018-1210-5)
5. Cantrell RS, Cosner C. 1998 On the effects of spatial heterogeneity on the persistence of interacting species. *J. Math. Biol.* **37**, 103–145. (doi:10.1007/s002850050122)
6. Keeling MJ, Ross JV. 2007 On methods for studying stochastic disease dynamics. *J. R. Soc. Interface* **5**, 171–181. (doi:10.1098/rsif.2007.1106)
7. Neufeld Z, von Witt W, Lakatos D, Wang J, Hegedus B, Czirok A. 2017 The role of Allee effect in modelling post resection recurrence of glioblastoma. *PLoS Comput. Biol.* **13**, e1005818. (doi:10.1371/journal.pcbi.1005818)
8. Saltz D, Rubenstein DI. 1995 Population dynamics of a reintroduced Asiatic wild ass (*Equus hemionus*) herd. *Ecol. Appl.* **5**, 327–335. (doi:10.2307/1942025)
9. Thomas CD *et al.* 2004 Extinction risk from climate change. *Nature* **427**, 145–148. (doi:10.1038/nature02121)
10. Coghlin C, Murray GI. 2010 Current and emerging concepts in tumour metastasis. *J. Pathol.* **222**, 1–15. (doi:10.1002/path.2727)
11. Korolev KS, Xavier JB, Gore J. 2014 Turning ecology and evolution against cancer. *Nat. Rev. Cancer* **14**, 371–380. (doi:10.1038/nrc3712)
12. Schmidt P, Kopecky C, Hombach A, Zigrino P, Mauch C, Abken H. 2011 Eradication of melanomas by targeted elimination of a minor subset of tumor cells. *Proc. Natl Acad. Sci. USA* **108**, 2474–2479. (doi:10.1073/pnas.1009069108)
13. Chaplain MAJ, Lorenzi T, Macfarlane FR. 2020 Bridging the gap between individual-based and continuum models of growing cell populations. *J. Math. Biol.* **80**, 343–371. (doi:10.1007/s00285-019-01391-y)
14. Fadai NT, Johnston ST, Simpson MJ. 2019 Unpacking the Allee effect: determining individual-level mechanisms that drive global population dynamics. *bioRxiv*, p. 774000.
15. Gerlee P. 2013 The model muddle: in search of tumor growth laws. *Cancer Res.* **73**, 2407–2411. (doi:10.1158/0008-5472.CAN-12-4355)
16. Gourley SA, So JWH. 2003 Extinction and wavefront propagation in a reaction-diffusion model of a structured population with distributed maturation delay. *Proc. R. Soc. Edinb. A: Math.* **133**, 527–548. (doi:10.1017/S0308210500002523)
17. Johnson KE, Howard G, Mo W, Strasser MK, Lima EA, Huang S, Brock A. 2019 Cancer cell population growth kinetics at low densities deviate from the exponential growth model and suggest an Allee effect. *PLoS Biol.* **17**, e3000399. (doi:10.1371/journal.pbio.3000399)
18. Othmer HG, Hillen T. 2000 The diffusion limit of transport equations derived from velocity-jump processes. *SIAM J. Appl. Math.* **61**, 751–775. (doi:10.1137/S0036139999358167)
19. Painter KJ, Maini PK, Othmer HG. 2000 A chemotactic model for the advance and retreat of the primitive streak in avian development. *Bull. Math. Biol.* **62**, 501–525. (doi:10.1006/bulm.1999.0166)
20. Sarapata EA, de Pillis LG. 2014 A comparison and catalog of intrinsic tumor growth models. *Bull. Math. Biol.* **76**, 2010–2024. (doi:10.1007/s11538-014-9986-y)
21. Gatenby RA, Gawlinski ET. 1996 A reaction-diffusion model of cancer invasion. *Cancer Res.* **56**, 5745–5753.
22. Gatenby RA, Gawlinski ET, Gmitro AF, Kaylor B, Gillies RJ. 2006 Acid-mediated tumor invasion: a multidisciplinary study. *Cancer Res.* **66**, 5216–5223. (doi:10.1158/0008-5472.CAN-05-4193)
23. Codling EA, Plank MJ, Benhamou S. 2008 Random walk models in biology. *J. R. Soc. Interface* **5**, 813–834. (doi:10.1098/rsif.2008.0014)
24. Baker RE, Simpson MJ. 2010 Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E* **82**, 041905. (doi:10.1103/PhysRevE.82.041905)
25. Johnston ST, Baker RE, McElwain DLS, Simpson MJ. 2017 Co-operation, competition and crowding: a discrete framework linking Allee kinetics, nonlinear diffusion, shocks and sharp-fronted travelling waves. *Sci. Rep.* **7**, 42134. (doi:10.1038/srep42134)
26. Simpson MJ, Baker RE. 2011 Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena. *Phys. Rev. E* **83**, 051922. (doi:10.1103/PhysRevE.83.051922)
27. Bubba F, Lorenzi T, Macfarlane FR. 2019 From a discrete model of chemotaxis with volume-filling to a generalised Patlak-Keller-Segel model. *arXiv* p. 1912.07488.

28. El-Hachem M, McCue SW, Jin W, Du Y, Simpson MJ. 2019 Revisiting the Fisher–Kolmogorov–Petrovsky–Piskunov equation to interpret the spreading–extinction dichotomy. *Proc. R. Soc. A* **475**, 20190378. (doi:10.1098/rspa.2019.0378)
29. Field TR, Tough RJA. 2010 Coupled dynamics of populations supported by discrete sites and their continuum limit. *Proc. R. Soc. A* **466**, 2561–2586. (doi:10.1098/rspa.2010.0049)
30. Chowdhury D, Schadschneider A, Nishinari K. 2005 Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms. *Phys. Life Rev.* **2**, 318–352. (doi:10.1016/j.plrev.2005.09.001)
31. Embacher P, Dirr N, Zimmer J, Reina C. 2018 Computing diffusivities from particle models out of equilibrium. *Proc. R. Soc. A* **474**, 20170694. (doi:10.1098/rspa.2017.0694)
32. Johnston ST, Simpson MJ, McElwain DLS. 2014 How much information can be obtained from tracking the position of the leading edge in a scratch assay. *J. R. Soc. Interface* **11**, 20140325. (doi:10.1098/rsif.2014.0325)
33. Johnston ST, Simpson MJ, McElwain DLS, Binder BJ, Ross JV. 2014 Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. *Open Biol.* **4**, 140097. (doi:10.1098/rsob.140097)
34. Johnston ST, Ross JV, Binder BJ, McElwain DLS, Haridas P, Simpson MJ. 2016 Quantifying the effect of experimental design choices for in vitro scratch assays. *J. Theor. Biol.* **400**, 19–31. (doi:10.1016/j.jtbi.2016.04.012)
35. Johnston ST, Baker RE, Simpson MJ. 2016 Filling the gaps: a robust description of adhesive birth-death-movement processes. *Phys. Rev. E* **93**, 042413. (doi:10.1103/PhysRevE.93.042413)
36. Doering CR, Sargsyan KV, Sander LM. 2005 Extinction times for birth-death processes: exact results, continuum asymptotics, and the failure of the Fokker–Planck approximation. *Multiscale Model. Simul.* **3**, 283–299. (doi:10.1137/030602800)
37. Gardiner CW. 1985 *Handbook of stochastic methods*, vol. 3. Berlin, Germany: Springer.
38. Murray JD. 2003 *Mathematical biology I: an introduction*, vol. I. Berlin, Germany: Springer.
39. Melica V, Invernizzi S, Caristi G. 2014 Logistic density-dependent growth of an *Aurelia aurita* polyps population. *Ecol. Modell.* **291**, 1–5. (doi:10.1016/j.ecolmodel.2014.07.009)
40. Courchamp F, Clutton-Brock T, Grenfell B. 1999 Inverse density dependence and the Allee effect. *Trends Ecol. Evol.* **14**, 405–410. (doi:10.1016/S0169-5347(99)01683-3)
41. Taylor CM, Hastings A. 2005 Allee effects in biological invasions. *Ecol. Lett.* **8**, 895–908. (doi:10.1111/j.1461-0248.2005.00787.x)
42. Ovaskainen O. 2001 The quasistationary distribution of the stochastic logistic model. *J. Appl. Probab.* **38**, 898–907. (doi:10.1017/S0021900200019112)
43. Johnson DM, Liebhold AM, Tobin PC, Bjørnstad ON. 2006 Allee effects and pulsed invasion by the gypsy moth. *Nature* **444**, 361–363. (doi:10.1038/nature05242)
44. Markham DC, Simpson MJ, Baker RE. 2013 Simplified method for including spatial correlations in mean-field approximations. *Phys. Rev. E* **87**, 062702. (doi:10.1103/PhysRevE.87.062702)
45. Chalub FACC, Souza MO. 2011 The SIR epidemic model from a PDE point of view. *Math. Comput. Modell.* **53**, 1568–1574. (doi:10.1016/j.mcm.2010.05.036)
46. Chalub FACC, Souza MO. 2014 Discrete and continuous SIS epidemic models: a unifying approach. *Ecol. Complex.* **18**, 83–95. (doi:10.1016/j.ecocom.2014.01.006)
47. Chalub FACC, Souza MO. 2009 From discrete to continuous evolution models: a unifying approach to drift-diffusion and replicator dynamics. *Theor. Popul. Biol.* **76**, 268–277. (doi:10.1016/j.tpb.2009.08.006)
48. Rebuli NP, Bean NG, Ross JV. 2017 Hybrid Markov chain models of S–I–R disease dynamics. *J. Math. Biol.* **75**, 521–541. (doi:10.1007/s00285-016-1085-2)
49. Dennis B. 2002 Allee effects in stochastic populations. *Oikos* **96**, 389–401. (doi:10.1034/j.1600-0706.2002.960301.x)
50. Sharkey KJ. 2008 Deterministic epidemiological models at the individual level. *J. Math. Biol.* **57**, 311–331. (doi:10.1007/s00285-008-0161-7)
51. Johnston ST, Simpson MJ, Baker RE. 2015 Modelling the movement of interacting cell populations: a moment dynamics approach. *J. Theor. Biol.* **370**, 81–92. (doi:10.1016/j.jtbi.2015.01.025)
52. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 2007 *Numerical recipes 3rd edition: the art of scientific computing*. Cambridge, UK: Cambridge University Press.
53. Shampine LF, Reichelt MW. 1997 The MATLAB ODE suite. *SIAM J. Sci. Comput.* **18**, 1–22. (doi:10.1137/S1064827594276424)