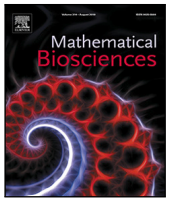
Mathematical Biosciences 365 (2023) 109084

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs



Original Research Article

Finding analytical approximations for discrete, stochastic, individual-based models of ecology

Linnéa Gyllingberg [a,*], David J.T. Sumpter [b], Åke Brännström [c,d]

[a] Department of Mathematics, Uppsala University, Uppsala, Sweden
[b] Department of Information Technology, Uppsala University, Uppsala, Sweden
[c] Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden
[d] Complexity Science and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Kunigami, Japan

ARTICLE INFO

Dataset link: https://github.com/linneagyllingberg/finding-analytical-approximations

Keywords:
Individual-based model
Site-based model
Spatial ecology
Approximation
Difference equations
Spatial correlations

ABSTRACT

Discrete time, spatially extended models play an important role in ecology, modelling population dynamics of species ranging from micro-organisms to birds. An important question is how 'bottom up', individual-based models can be approximated by 'top down' models of dynamics. Here, we study a class of spatially explicit individual-based models with contest competition: where species compete for space in local cells and then disperse to nearby cells. We start by describing simulations of the model, which exhibit large-scale discrete oscillations and characterize these oscillations by measuring spatial correlations. We then develop two new approximate descriptions of the resulting spatial population dynamics. The first is based on local interactions of the individuals and allows us to give a difference equation approximation of the system over small dispersal distances. The second approximates the long-range interactions of the individual-based model. These approximations capture demographic stochasticity from the individual-based model and show that dispersal stabilizes population dynamics. We calculate extinction probability for the individual-based model and show convergence between the local approximation and the non-spatial global approximation of the individual-based model as dispersal distance and population size simultaneously tend to infinity. Our results provide new approximate analytical descriptions of a complex bottom-up model and deepen understanding of spatial population dynamics.

1. Introduction

Spatial structure plays an important role in ecology [1] and numerous models have been developed and explored to advance understanding of spatial population dynamics [2–12]. The importance of using spatially explicit and discrete models was highlighted by Durrett and Levin in a seminal paper that elucidated differences in outcomes between plausible modelling assumptions: discrete individuals and explicit space were shown to be critically important for capturing salient features of ecological systems [13]. The importance of spatial structure has been manifested in many other ways as well. One well studied aspect is the effect of dispersal on population dynamics. For example, Gyllenberg et al. [14] and Hastings [15] show through simple models that dispersal between two metapopulations stabilizes population dynamics.

Among the different ways to model spatial population dynamics, spatially explicit individual-based models (arguably) stand out as the most realistic approach. These type of models are often referred to as 'bottom-up', since they explicitly represent individual actions and interactions. This contrasts with the more classical 'top-down' approach, where the population dynamics is modelled through partial or ordinary differential or difference equations. There are many benefits of individual-based models. One is the ease by which spatial structure can be represented. Another advantage, from a biological perspective, is that observations of individual behaviour and interactions can be included. However, individual-based models are rarely analytically tractable and often computationally demanding. Therefore, it is important to study the relationship between individual-based models and analytical, top-down models of populations dynamics.

Techniques that have been developed for this purpose include reaction–diffusion approximations [16–19], the method of moments [20–25] and pair approximations [26,27]. All these methods have advantages and drawbacks. For example, pair approximation methods work in epidemiological settings, but perform poorly when reproduction and dispersal take place on different spatial scales. Also, both

* Corresponding author.
E-mail address: linnea.gyllingberg@math.uu.se (L. Gyllingberg).

https://doi.org/10.1016/j.mbs.2023.109084
Received 17 July 2023; Received in revised form 15 September 2023; Accepted 23 September 2023
Available online 30 September 2023
0025-5564/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

pair approximations and the general method of moments fail to capture large scale patterns that can emerge through local interactions. Reaction–diffusion approximations are useful for deriving limit dynamics of individual-based models with continuous time and space [16], and there are rigorous results for many of these models. Recently, Patterson et al. [28] developed a new method demonstrating convergence to a mean-field limit for an ecologically motivated discrete-space, continuous-time stochastic individual-based model. These methods are essentially techniques for approximating interacting particle systems and spatial point processes, in which a continuous time approximation is possible.

Here, we are interested in a different class of individual-based models: what Berec refer to as Individual-based models or D-space, D-time IBMs [29]. These models are motivated by, for example, empirical studies of many insects [30,31], birds [32], microorganisms [33,34], and other species for which the population is distributed across resource sites and have a lifecycle involving discrete phases of: (1) competition for resource sites, (2) reproduction by one or more of these individuals and (3) dispersal of surviving offspring to new potential resource sites. In this case, competition and reproduction happen on the same timescale and, as a result, classical spatial methods like pair-approximations and reaction–diffusion approximations do not apply to models of these systems.

The simplest way to approximate spatial population dynamics in discrete-time discrete-space models is to neglect spatial structure altogether, which is sometimes referred to as a mean field approximation [35]. If the population is assumed to be well mixed, the approximation describes how the mean population changes over time through a set of differential or difference equations. A large class of discrete-time individual-based models, in which individuals from one or more species compete for resources in discrete sites but mix uniformly in space, have been studied exhaustively [36–40]. However, these studies do not extend to the spatial case when offspring are more likely to disperse to nearby sites and, in this setting, the task of deriving suitable approximations through first principles becomes more difficult. The discrete nature of time and space in these site-based models together with potentially long dispersal distances limit the applicability of existing spatial approximation techniques.

As a first effort to describe the spatial population dynamics of discrete-time, discrete-space site-based models, Brännström and Sumpter [41] introduced a new approximation method called the coupled map lattice approximation. The idea is to divide the lattice which individuals inhabit into nine or more sublattices and decouple the individual interactions within each sublattice from the dispersal between each sublattice. An advantage of their approach is the ability to disentangle the stochastic and deterministic parts of individual interactions and dispersal, but a downside is that it leads to a large system of difference equations which is not necessarily easier to analyse than the individual-based model itself. Brännström and Sumpter [41] attempted to reduce the model to a single analytical tractable equation, but in doing so they needed to make several simplifying assumptions that are not necessarily justified for the full system. We complement the previous work of Brännström and Sumpter [41] by studying and characterizing the spatial dynamics that can occur in discrete-time, discrete-space site-based models.

In this paper, we present an initial exploration of methods for approximating the dynamics of D-space, D-time IBMs, based on a selected spatial population model. The paper is organized as follows. We start in Section 2 by describing a simple discrete time, discrete space biologically motivated individual-based model with local dispersal and reproduction. The qualitative behaviour of the model is studied through bifurcation plots and the spatial statistics of the model is analysed, to gain insight into the spatial patterns of the individual-based model. From this, we continue in Section 3 where we derive three different approximations: first we derive a global approximation (also known as a mean field approximation), reducing the individual-based model to a one-dimensional discrete dynamical system. Based on the insight of the local cluster sizes and the spatial distribution of the individual-based model in Section 2.3, we derive a 'local correlation approximation' where the dispersal length is taken into account. The local correlation approximation is also given by a one-dimensional discrete dynamical system, but with the dispersal length included as a parameter in the model. We then derive a long-range dispersal approximation, in the form of a two-dimensional coupled map lattice. We continue to analyse these three models in Section 4, where we first present bifurcation plots for the three approximations, to compare the qualitative behaviour of the approximations with the individual-based model. We then carry out a stability analysis of the global approximation and local correlation approximation and then calculate a 'general extinction probability' of the individual-based model to explain parts of the discrepancies in stability between the global approximation and the individual-based model. Finally, we show that the local correlation approximation converges to the global approximation in the limit of large dispersal distances.

2. The individual-based model

2.1. Model description

We consider a model where individuals live on a $D \times D$ lattice with cyclic boundary conditions. There is thus a total of $n = D^2$ resource sites. We divide the life cycle of the individuals into a competition phase, a reproduction phase and a dispersal phase. Let X_{ij}^t be a random variable denoting the number of individuals at site (i, j) at time t. At the **competition phase** there may be more than one individual at a particular site. In order to capture competition for resources we define a survival function $\phi : \mathbb{N} \to \{0, 1\}$ that acts on X_{ij}^t giving a new site population $\phi(X_{ij}^t)$. The survival function maps the number of individuals at a given site to 0 if the site is over exploited or 1 if there is a survivor. This captures over-crowding: if there are too many individuals at a resource site, they cannot all reproduce.

After competition the next phase is **reproduction**. The surviving individual at a site (if there is one) produces R_{ij}^t offspring, where R_{ij}^t is a random variable with expectation $\mathrm{E}[R_{ij}^t] = r$ for all i, j and t. We consider discrete generations, so after the reproduction phase the parents die, resulting in $R_{ij}^t \cdot \phi(X_{ij}^t)$ individuals at each site (i, j). These offspring then **disperse** to a random site in a $(2s + 1) \times (2s + 1)$ Moore neighbourhood, with uniform probability, resulting in a new site count X_{ij}^{t+1}. The Moore neighbourhood of site (i, j) with range s, $N(i, j, s)$, is given by the set of sites (i', j') for which $|i - i'| \leq s$ and $|j - j'| \leq s$. Since the lattice has cyclic boundary conditions, individuals leaving the lattice in one end, end up on the opposite side, i.e. two sites (i, j) and (i', j') are identical if $i \pmod D \equiv i' \pmod D$ and $j \pmod D \equiv j' \pmod D$. The life cycle is complete and the process starts over again.

There are several ways to define the survival function and Brännström and Sumpter examine the population dynamics for different survival functions for the non-spatial case [37]. In our simulations of this individual-based model, however, we will define the survival function as

$$\phi(k) = \begin{cases} 1 & \text{if } k = 1 \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

which gives a classic scramble competition model, as first described by Nicholson [42]. In our simulations in the main article, we also assume R_{ij}^t to be equal to a constant r for all i, j and t, and thus not a random variable. In Supplementary material S3 we simulate the model for $R_{ij}^t \sim \mathrm{Bin}(r/q, q)$, for different values of q. Fig. 1 illustrates the individual-based model with scramble competition for $r = 4$ and $s = 2$. Individuals that are alone in a site survive, whereas in sites with more than one individual, all individuals die. The surviving individuals then produce $r = 4$ offspring each, that disperse to a randomly chosen site with uniform probability in a $(2s + 1) \times (2s + 1)$ Moore neighbourhood.

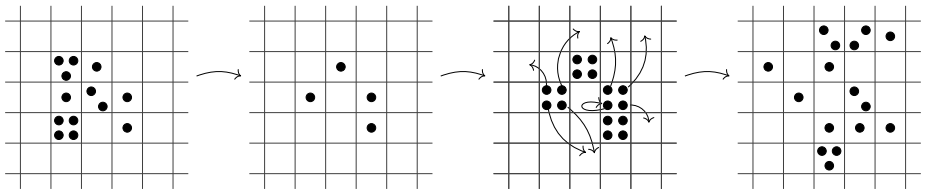


Fig. 1. An illustration of the individual-based model. Each box represents a site, and individuals are represented by black dots. After the competition phase, only the individuals that are alone in their boxes survive. They then reproduce $r = 4$ offspring that disperse in an $s = 2$ range (arrows for the dispersal illustrated for just two sites). We count the population in the third state.

2.2. Simulations

The model has a rich set of dynamics depending on the parameter values r and s. Fig. 2 shows how the spatial distribution of individuals on a 101×101 lattice evolves over time for four time steps for different values of s, when $r = 20$. When $s = 1$, there is no clear spatial clustering, and between generations the population is relatively stable. As s increases to 2 and 3 (Fig. 2(b)–(c)), we can distinguish spatial clusters. While the overall population size is still reasonably stable over time, there are local fluctuations. Patches of a size roughly $(2s+1)\times(2s+1)$, i.e. the same size as the dispersal range emerge creating a checkerboard effect. In Fig. 2(d) and (e), where $s = 5$ and $s = 10$ respectively, the global population is alternatingly large and small and fluctuates in the four consecutive snapshots. The last figure, Fig. 2(f), shows time evolution of the population for $s = 50$. Dispersal is now global, i.e. individuals disperse uniformly to any site on the lattice, since $s = (D-1)/2 = 50$ for $D = 101$. The figure displays chaotic oscillations between generations and there is no spatial clustering.

Further simulations are summarized in bifurcation plots shown in Fig. 3. Here we varied r, simulated the model for 5000 time steps, and the plot shows the outcome for the last 500 time steps. In Fig. 3(a), we see how very local dispersal ($s = 1$) typically gives rise to a stable population. This population reaches a maximum at around $r \approx 15$, and it dies out at $r = 23$. Fig. 3(b)–(c) shows a stable population for $s = 2$ and $s = 3$. Increasing s to 5, there is a period doubling when $r = 9$, and for $s = 10$, the period doubling bifurcation occurs already at $r = 8$, which we see in Fig. 3(d) and (e). When dispersal is global the population is stable for $r < e^2 (\approx 7.39)$, and undergoes a period doubling bifurcation after that, leading to an oscillating population size between generations when $8 < r < 14$. When $r > 15$ the population oscillates chaotically. In summary, for all values of s, the population is stable for small reproductive rates, but for higher values of r, the population goes from stable to oscillating to chaotic by increasing s.

In Supplementary material S3 we look at what happens to the dynamics if the reproductive rate is not constant, but instead the reproductive rate at each site and time step, R_{ij}^t, are i.i.d with $R_{ij}^t \sim \text{Bin}(r/q, q)$, by producing bifurcation plots for different values of s and q. We see that the overall behaviour is not affected by different values of q.

Increasing the lattice size to 201×201 and the reproductive rate to $r = 30$, the individual-based model produces interesting patterns, which is seen in Fig. 4. When $s = 10$, we see oscillating ring patterns, resembling reaction–diffusion patterns. As we increase s to 20, the population dynamics instead behave as travelling waves. When $s = 45$ there is a band formation jumping between the two halves of the lattice in each consecutive snapshot. In Fig. 4(c) the band formation is vertical, but the band formation can also be horizontal. Note that, although the particular vertical/horizontal band formation is most likely a result from the anisotropy produced by Moore neighbourhood for dispersal, we would also expect waves and bands to form at arbitrary angles in isotropic settings as they do in cellular automata models [43,44].

In the bifurcation diagram in Fig. 5(a), we fixed r to 30 and varied s, simulated the model for 5000 time steps, and the plot shows the outcome for the last 500 time steps. From this bifurcation diagram, we can conclude that the band formation seen in Fig. 4(c) gives rise to stable population dynamics. For $s = 10$, the population dynamics seem stable and for $s = 20$ the population is fluctuating. In the bifurcation diagram in Fig. 5(b), we instead fixed s to 45 and varied r, simulated the model for 5000 time steps, and the plot shows the outcome for the last 500 time steps. We see that the population is stable for $r < 8$, and exhibits a period doubling bifurcations after that. At $r = 22$, the population becomes stable again. We can thus see that long-range, but not global, dispersal has a stabilizing effect on the population dynamics for high reproductive rates.

In essence, the snapshots of the model together with the bifurcation plots, show that our model displays behaviour which are inherently discrete in time: the population density display chaotic behaviour for some values of r and s, and there are large-scale discrete oscillations in the spatial distribution of the population, rather than waves that move smoothly.

2.3. Spatial statistics

The snapshots of the individual-based model, together with the bifurcation diagrams, demonstrate that spatial structure strongly influence the population dynamics. In Section 3 we will make different assumptions on the spatial distribution of the individual-based model, in order to find analytical approximations of the model. Thus, understanding the spatial structure and scales of the individual-based model is valuable for the approximations done in 3. In the supplementary material S1, we use Moran's I (which measures spatial correlation) to show that spatial clustering changes with dispersal distance. This, in itself, does not give us any information of the scales of the spatial clusters. A better measure for detecting spatial pattern scales is the four-term local quadrat variance measure, referred to as 4TLQV. The idea behind 4TLQV is to look at a block of size $b \times b$ and then compare this block to three adjacent blocks. This is done by summing all the individuals in the first block, multiply by -3 and then sum all individuals of the other three blocks. Since any of the four blocks can be used for comparison, there are four possible ways to do this calculation, and the 4TLQV is calculated by taking the average of the squared values of these four possibilities (see the Figure S3 for an illustration of this). The peaks in 4TLQV as a function of b can be interpreted as the spatial scale of the pattern. We used the PySSaGE package for Python, to calculate the 4TQLV for 100 consecutive snapshots of 10 realizations of the individual-based model for various values s (i.e. 1000 snapshots for each parameter set) [45].

In Fig. 6 we see the average 4TQLV of 100 consecutive snapshots from 10 realizations (1000 snapshots in total) for different values of s when $r = 20$ as a function of the block size, b. We note that for $s = 1$ and $s = 2$, the peak in variance is already at $b = 1$, suggesting that there is no clear spatial structure. For $s = 3$, there is a peak around $b = 9$. As s increases, the value of the 4TLQV peak also increases. However, when $s = 50$ the peak is now at $b = 1$, suggesting there is no clear spatial structure.

In Fig. 7, the average 4TLQV variance is plotted for different values of s when $r = 30$ and the lattice size is 201×201. We see that when $s = 10$, there is a peak around $b = 30$. For $s = 20$, there are there is

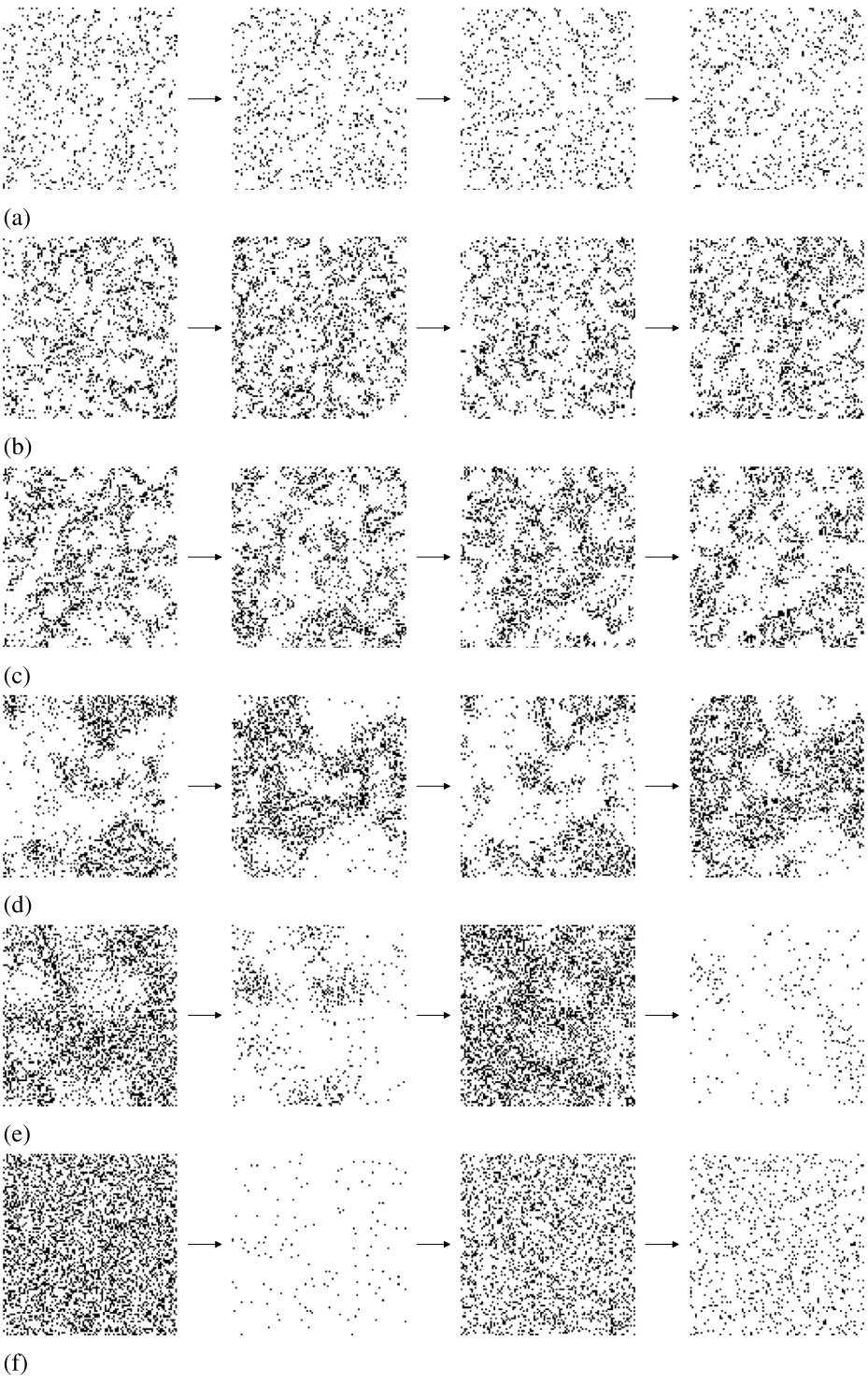


Fig. 2. Four sequential snapshots for the individual-based model on a 101 × 101 sized lattice when $r = 20$ and (a) $s = 1$ (very local dispersal), (b) $s = 2$, (c) $s = 3$, (d) $s = 5$, (e) $s = 10$ and (f) $s = 50$ (global dispersal). The snapshots show the population after competition, thus there are at most 1 individual in each site.

a peak around 60, and for $s = 45$, there is a peak at $b = 90$. What should also be noted, but is not shown here, is that for all parameter sets with clear spatial patterns, the variance of the 4TLQV increases as b increases, meaning that the spatial patterns sizes varies between time step. From these plots, we can conclude that if there is a spatial pattern, the size of the pattern is always larger than $2s+1$. This is not surprising, since the individuals disperse into a $(2s+1)\times(2s+1)$ blocks and thus compete with other individuals within that range, the spatial patterns should be at least of that size.

3. Approximation

As we saw in the previous section, our model, which is very simple to state, produces a rich variety of dynamics. The aim of this article is to look at various ways to approximate the simulation model using an appropriate dynamical system and then to use this approximation to understand why these dynamics arise. In this section we describe three such approximations.

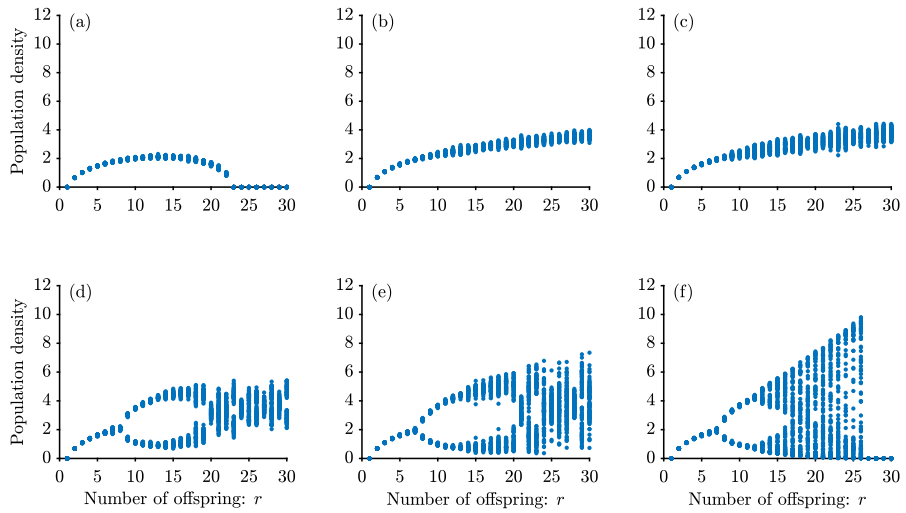


Fig. 3. Bifurcation diagram for the individual-based model, with $\phi(k)$ given as Eq. (1). In each subplot we vary r and simulate the model for 5000 time steps and plot last 500 time steps. The population density is given by the number of individuals divided by the number of resource sites, and is counted after reproduction. We give results for (a) $s = 1$ (very local dispersal) (b) $s = 2$, (c) $s = 3$, (d) $s = 5$, (e) $s = 10$ and (f) $s = 50$ (global dispersal).

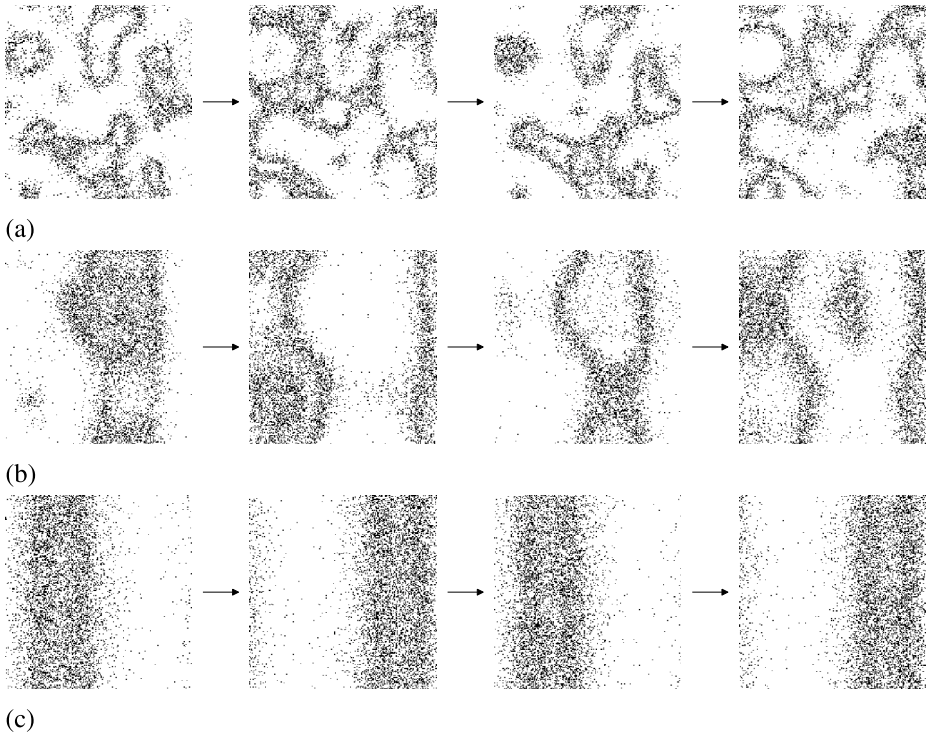


Fig. 4. Four sequential snapshots for the individual-based model on a 201 × 201 lattice when $r = 30$ and (a) $s = 10$, (b) $s = 20$, and (c) $s = 45$. The snapshots show the population after competition, thus there are at most 1 individual in each site.

3.1. Global approximation

We start with the model described in Section 2 with scramble competition described by Eq. (1). Let x_t be the density of the population at time t, where $t = 0, 1, 2, \ldots$ is the time after the competition phase, i.e. after the survival function $\phi(X_{ij}^t)$ has been applied to every resource site. The expected population density at time $t+1$ given the population density x_t at time t can be written as:

$$x_{t+1} = \mathrm{E}(X_{t+1}|X_t = x_t) = \frac{1}{D^2}\sum_{k=1}^{\infty}\sum_{i=1}^{D}\sum_{j=1}^{D} P(X_{ij}^t = k|X_t = x_t)\cdot\phi(k). \qquad (2)$$

Here $\phi(k)$ is the survival function and $P(X_{ij}^t = k|X_t = x_t)$ is the probability to find k individuals in a site (i, j) at time t, given that the population density is x_t. The question of approximating the population dynamics becomes a question of finding a good approximation p_k of $P(X_{ij}^t = k)$, which utilizes as much information as possible about the correlations between neighbouring sites. Thus p_k is an approximation of $P(X_{ij}^t = k)$, which is not dependent on a specific site (i, j). From this, we can reduce Eq. (2) to

$$x_{t+1} = \frac{1}{D^2}\sum_{k=1}^{\infty}\sum_{i=1}^{D}\sum_{j=1}^{D} p_k\cdot\phi(k) = \sum_{k=1}^{\infty} p_k\cdot\phi(k), \qquad (3)$$

i.e. p_k is site independent.

One possible way to define p_k, is to assume that the offspring disperse independently to any site on lattice with uniform probability.

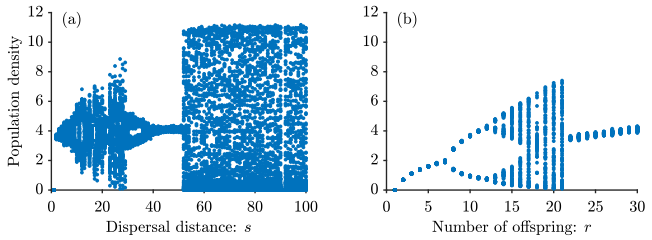


Fig. 5. (a) Bifurcation diagram for the individual-based model, with $\phi(k)$ given as Eq. (1), $r = 30$ and s as the bifurcation parameter. For each value of s, we simulate the model for 5000 time steps and plot last 500 time steps. In (b), $s = 45$ and r is varied, and for each value of r we simulate the model for 5000 time steps, and the plot shows the outcome for the last 500 time steps. The population density is given by the number of individuals divided by the number of resource sites, and is counted after reproduction.

When the lattice is infinitely large, X_{ij}^t is Poisson distributed with mean rx_t for all i and j, in which case

$$p_k = P(X_{ij}^t = k) = \frac{e^{-rx_t}(rx_t)^k}{k!}, \tag{4}$$

giving the following population dynamics

$$x_{t+1} = \sum_{k=1}^{\infty} \frac{e^{-rx_t}(rx_t)^k}{k!} \cdot \phi(k). \tag{5}$$

In this case, the entire population dynamics for the scramble model is

$$x_{t+1} = rx_t e^{-rx_t}. \tag{6}$$

This equation, sometimes called the mean field equation [35], we refer to as the global approximation equation of the individual-based model. This is the same equation as derived by Sumpter et al. [46] and Brännström and Sumpter [41] but with a change of variable $x_t' = rx_t$ to reflect the fact that we measure the population after reproduction. This global approximation works well for individual-based models with large dispersal, i.e. when s is close to $(D-1)/2$.

3.2. Local correlation approximation

From the simulations of the individual-based model, we saw that even though the scale of the spatial pattern is the same for very local dispersal ($s = 1$ and $s = 2$) and global dispersal, the population dynamics is vastly different: when dispersal is global, the population dynamics is chaotic, whereas the population dynamics is always stable for local dispersal. This suggests, that for local dispersal, ($s \ll D$), we need to take local correlations into account. To do this, we adopt an approach that looks at correlations which arise between dispersal and competition. Consider a particular lattice site, which we call the focal site. In order to better account for local correlations we look in the $(2s+1)\times(2s+1)$ Moore neighbourhood around the focal site. Let F be the total number of parents in the $(2s+1)\times(2s+1)$ Moore neighbourhood of the focal site (see Fig. 8).

For this approximation we assume – as we also did in our approximation of full independence between the sites – that the population is uniformly distributed at this stage with population density x_t. The spatial statistics in Section 2.3 together with the supplementary material S1, showed that this assumption is justified not just for global dispersal, but also for very local dispersal (i.e. when s is small). Now, if $s \ll D$, we can assume that each site in the neighbourhood of the focal site has an independent probability $p = x_t$ of containing an individual, meaning that $F \sim \text{Bin}((2s+1)^2, x_t)$.

In this new approximation, our aim is to calculate the number of offspring of all of the parents in the neighbourhood of the focal site that disperse to the focal site. We denote this as a random variable Y. The probability that a particular offspring of a particular parent within the $(2s+1)\times(2s+1)$ Moore neighbourhood ends up in the focal site is $1/(2s+1)^2$. We can thus denote the event that this offspring j lands on the focal site as a Bernoulli random variable, O_j with parameter $1/(2s+1)^2$. The dispersal of individuals is independent and thus Y is defined by the sum

$$Y = \sum_{j=1}^{F\cdot r} O_j \tag{7}$$

or, equivalently, $Y \sim \text{Bin}(F \cdot r, \frac{1}{(2s+1)^2})$.

Using the law of total probability we utilize Eq. (7) to write

$$p_k = P(Y = k) = \sum_{i=1}^{(2s+1)^2} P(Y = k|F = i)P(F = i). \tag{8}$$

The two terms in this convolution are, respectively,

$$P(Y = k|F = i) = \binom{ir}{k} \frac{1}{(2s+1)^{2k}} \left(1 - \frac{1}{(2s+1)^2}\right)^{ir-k}, \tag{9}$$

and

$$P(F = i) = \binom{(2s+1)^2}{i} x_t^i (1-x_t)^{(2s+1)^2-i}. \tag{10}$$

This gives the following explicit expression for p_k

$$p_k = \sum_{i=1}^{(2s+1)^2} \binom{(2s+1)^2}{i} x_t^i (1-x_t)^{(2s+1)^2-i} \binom{ir}{k} \times \frac{1}{(2s+1)^{2k}} \left(1 - \frac{1}{(2s+1)^2}\right)^{ir-k}. \tag{11}$$

With the approximation that all sites on the lattice experience the same dynamics as the focal site, the population dynamics are then given by

$$x_{t+1} = \sum_{k=1}^{\infty} p_k \cdot \phi(k). \tag{12}$$

For scramble competition, where $\phi(k)$ is defined as Eq. (1), the expression becomes

$$x_{t+1} = \sum_{i=1}^{(2s+1)^2} \binom{(2s+1)^2}{i} x_t^i (1-x_t)^{(2s+1)^2-i} ir \frac{1}{(2s+1)^2} \left(1 - \frac{1}{(2s+1)^2}\right)^{ir-1}. \tag{13}$$

By identifying the last factor as a derivative, the expression can be simplified to

$$\begin{aligned} x_{t+1} &= rx_t \left(1 - \frac{1}{(2s+1)^2}\right)^{r-1} \left(x_t \left(1 - \frac{1}{(2s+1)^2}\right)^{r} - x_t + 1\right)^{(2s+1)^2-1} \\ &= f(x_t). \end{aligned} \tag{14}$$

We now have an approximation of the individual-based model that, unlike the global approximation equation (Eq. (6)), takes in the spatial properties of the individual-based model, and has the dispersal distance, s, as a parameter. (See Supplementary material S4 for a detailed derivation.)

Now, if R_{ij}^t is not constant, but R_{ij}^t are i.i.d. random variables with $R_{ij}^t \sim \text{Bin}(r/q, q)$, the number of offspring the parents will produce will be given by

$$S_F = \sum_{i}^{F} R_i \tag{15}$$

where $R_i \sim \text{Bin}(r/q, q)$. Thus,

$$Y = \sum_{j=1}^{S_F} O_j = \sum_{j=1}^{\sum_i^F R_i} O_j. \tag{16}$$

Using probability generating functions (see Supplementary material S4), we can find a closed form of $p_1 = P(Y = 1)$, which will be given

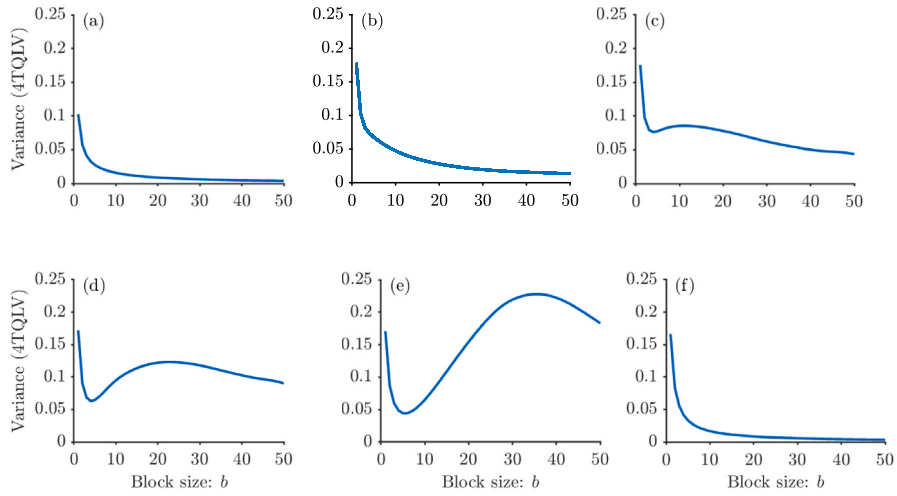


Fig. 6. The average 4TLQV variance against the block size b for 1000 snapshots from 10 realizations of the individual-based model simulated on a 101×101 grid when $r = 20$ and (a) $s = 1$, (b) $s = 2$, (c) $s = 3$, (d) $s = 5$, (e) $s = 10$, and (f) $s = 50$.

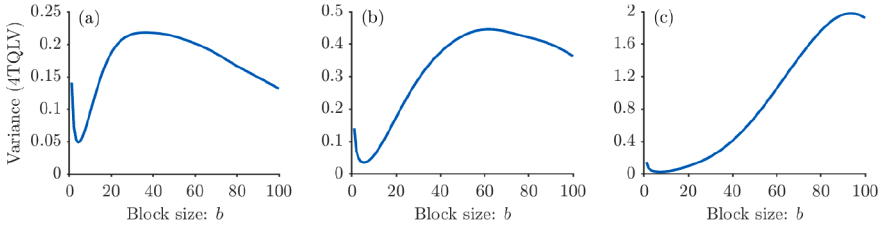


Fig. 7. The average 4TLQV variance against the block size b for 1000 snapshots from 10 realizations of the individual-based model simulated on a 201×201 grid when $r = 30$ (a) $s = 10$, (b) $s = 20$, and (c) $s = 45$.

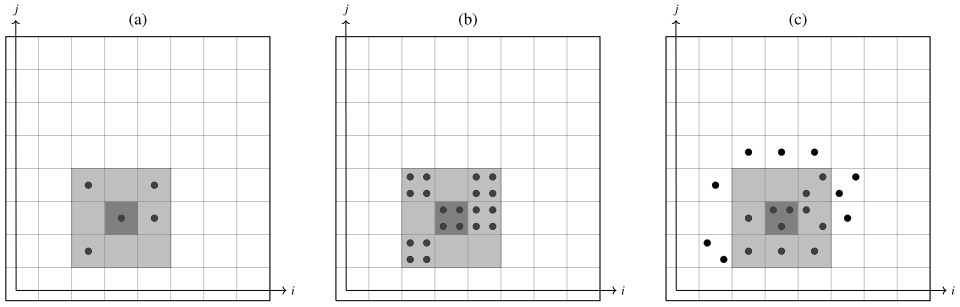


Fig. 8. An illustration of how p_k is calculated for the local correlation approximation when $s = 1$. The focal site is dark grey, while its neighbours within distance $s = 1$ are light grey. The figure shows how individuals are (a) distributed around the focal site initially, where the number of parents is $F = 5$, which in general is given by $F \sim \text{Bin}((2s+1)^2, x_t)$. In (b) the individuals have each produced $r = 4$ offspring, meaning there is $r \cdot F = 4 \cdot 5 = 20$ individuals within the neighbourhood of the focal site. In (c) the offspring have dispersed. The event that a particular offspring j will disperse into the focal site, O_j has probability $1/(2s+1)^2$. The number of offspring landing in the focal site is then given by $Y = \sum_{j=1}^{Fr} O_j = \sum_{j=1}^{20} O_j$, which in this example is equal to 3, ($Y = 3$).

by

$$p_1 = rx_t\left(\left(1-\frac{1}{(2s+1)^2}\right)q - q + 1\right)^{r/q-1} \left(x_t\left(\left(1-\frac{1}{(2s+1)^2}\right)q - q + 1\right)^{r/q} - x_t + 1\right)^{(2s+1)^2-1}. \quad (17)$$

Thus, when the reproductive rate, R_{ij}^t, is random with $R_{ij}^t \sim \text{Bin}(r/q, q)$, the population dynamics is given by

$$x_{t+1} = rx_t\left(\left(1-\frac{1}{(2s+1)^2}\right)q - q + 1\right)^{r/q-1} \left(x_t\left(\left(1-\frac{1}{(2s+1)^2}\right)q - q + 1\right)^{r/q} - x_t + 1\right)^{(2s+1)^2-1}. \quad (18)$$

In the rest of the paper, we will focus on analysing the case with deterministic r, but in Supplementary material S4, derive and analyse the local correlation approximation with random R.

3.3. Long-range dispersal approximation

The bifurcation diagram in Fig. 5 shows that for long-range dispersal, i.e., $s \approx D/2$, the population density stabilizes. The four consecutive snapshots of the individual-based model for $s = 45$ in Fig. 4(c) suggests that this happens when the population propagates as a band over the lattice, i.e., when the spatial pattern scale is around $D/2$, as seen in Fig. 7(c). Even though there is a clear global spatial structure in terms of the band formation, calculating the Moran's I within each band, suggest that the population is close to well mixed locally.

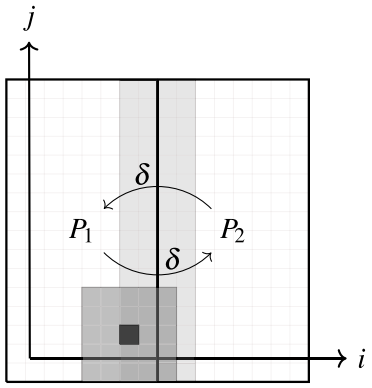


Fig. 9. Division of the lattice into two patches. The light grey area consists of the sites where dispersal to the other patch is possible, when the dispersal distance is $s = 2$. The dark grey area shows the sites reachable by dispersal from the black site when $s = 2$.

To capture this observation, we propose a long-range dispersal approximation, which divides the lattice into two patches, P_1 and P_2, with population size x_1 and x_2 respectively, with dispersal probability δ between the patches, as seen in Fig. 9. In this approximation, the population is assumed large and well mixed within each patch, allowing us to describe the population as a coupled version of the global approximation (Eq. (6)), i.e.

$$\begin{aligned} x_{1,t} &= (1-\delta)re^{-rx_{1,t}} + \delta re^{-rx_{2,t}} \\ x_{2,t} &= (1-\delta)re^{-rx_{2,t}} + \delta re^{-rx_{1,t}}. \end{aligned} \tag{19}$$

We now want to determine reasonable values of the dispersal probability δ as a function of the dispersal distance s, by reasoning from the individual-based model. For simplicity, we assume that an individual in P_1 enters P_2 only when it crosses the border in the middle. When crossing the other borders, it re-enters P_1. Since P_1 and P_2 are identical except for being laterally reversed, it suffices to determine the dispersal probability from P_1 to P_2 to find δ. To do this, we start by indexing the sites by (i, j), as in Fig. 9. In P_1 we have $1 \le i \le D/2$ and $1 \le j \le D$. The dispersal probability between P_1 and P_2 will be given by the probability that an individual k is situated at site (i, j) times the probability that it disperses to patch P_2, given that it is at site (i, j), summed over all sites in P_1:

$$\delta = \sum_{(i,j)\in P_1} P(k \text{ at site } (i,j) |\ k \text{ in } P_1) \cdot P(k \text{ disperse to } P_2 |\ k \text{ at site } (i,j)). \tag{20}$$

Assuming that within each patch the population is well mixed, meaning that individuals are uniformly distributed over all sites, we have that $P(k \text{ at site } (i,j) |\ k \text{ in } P_1)$ is given by $\frac{1}{D^2/2}$.

To determine $P(k \text{ disperse to } P_2 | k \text{ at site } (i,j))$, we want to find the number of sites in P_1 where P_2 is reachable by dispersal. This obviously depends on s. We have that

$$P(k \text{ disperse to } P_2 | \text{ at site } (i,j)) = \frac{(s+1)-i}{2s+1},$$

when $i < s$, and 0 when $i > s$.

Thus, we have

$$\begin{aligned} \delta &= \sum_{(i,j)\in P_1} P(k \text{ at site } (i,j) | k \text{ in } P_1) \cdot P(k \text{ disperse to } P_2 | k \text{ at site } (i,j)) \\ &= \sum_{j=1}^{D} \sum_{i=1}^{D/2} \frac{1}{D^2/2} P(k \text{ disperse to } P_2 | k \text{ at site } (i,j)) \\ &= D \sum_{i=1}^{D/4} \frac{1}{D^2/2} P(k \text{ disperse to } P_2 | k \text{ at site } (i,j)) = 2D \sum_{i=1}^{D/4} \frac{(s+1)-i}{2s+1} \\ &= \frac{s(s+1)}{D(2s+1)}. \end{aligned} \tag{21}$$

We now have three different approximations that describe the population dynamics for the individual-based model. For the rest of the paper we will analyse Eq. (14), (6) and (19) and compare it with the individual-based model described in Section 2.

4. Analysis

In Section 3 we have derived three different approximations of the individual-based model described in Section 2. In this section we discuss how these approximations are able to describe the properties of the individual-based model and the connections between the different approximations.

4.1. Bifurcations

As for the individual-based model in Fig. 3, we produced bifurcation plots for the local correlation approximation by iterating Eq. (14) for different values of r, which can be seen in Fig. 10. When dispersal is very local ($s = 1$), our approximation displays quantitatively the same behaviour as the individual-based model: the population is stable and at most around 2 (when $r \approx 15$). The only difference between the individual-based model and the local correlation approximation is that the population goes extinct already at $r = 23$ for the individual-based model, whereas it goes extinct for $r = 30$ for the approximation. In Fig. 10(b), we see that there is a period doubling bifurcation when $r = 10$, for $s = 2$, in contrast to the individual-based model, where the population is stable at least until $r = 30$ for $s = 2$. Increasing s to 3 (Fig. 10(c)), the approximation model undergoes a period-doubling route to chaos: the model exhibits period doubling first at $r = 8$ and then at $r = 16$, and becomes chaotic at $r = 26$. Increasing s further, the period doubling takes place earlier and earlier: in Fig. 10(d) we see that for $s = 5$, the first period doubling appears at $r = 8$ and the second at $r = 13$. For $s = 10$, the dynamics of the approximation is more or less the same as for $s = 50$. Quantitatively, the population density is higher for the local correlation approximation, than for the individual-based model, when $s > 2$. In summary, the local approximation seems to approach the global approximation quicker than the individual-based model approaches the well-mixed case. In Supplementary material S4 we show bifurcation plots for the local correlation approximation for stochastic reproductive rates, i.e. Eq. (18). We see in the plots that over all qualitative behaviour is not affected by the value of q, just as for the bifurcation plots of the individual-based model with stochastic reproductive rate.

For the long range dispersal approximation, we produced bifurcation plots with both s and r as the bifurcation parameter. In (a) and (b) we show the total population, $(x_{1,t} + x_{2,t})/2$, and in (c) and (d) we plot the difference in population, $(x_{1,t} - x_{2,t})/2$. In Fig. 11(a), we see that for s between 40 and 50, the approximation has both quantitatively and qualitatively the same behaviour as the individual-based model: the population is stable and has a size around 4. Just before that, for s around 33, there is bifurcation point, and for $s = 37$, the population becomes stable again. However, in contrast to the individual-based model, the long-range dispersal approximation exhibits a stable population already at $s = 15$ and up until $s = 70$. In (c) we see that, while the total population is chaotic for $s > 71$, the difference in population is stable for $s > 77$.

4.2. Stability analysis

One important property of the individual-based model, that is not captured in the global approximation equation, is that the population goes extinct for some values of r, both when s is very local ($s = 1$) and when s is global,as seen in Fig. 3(a) and (f). We want to find for which values of r this happens for the local correlation approximation. To do

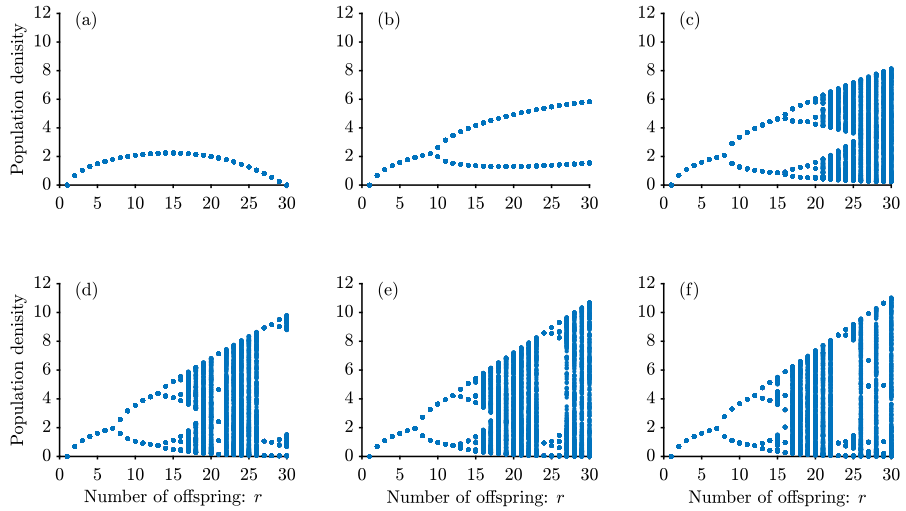


Fig. 10. Bifurcation diagram for the local correlation approximation, given in Eq. (14). In each subplot we vary r and iterate through x_t for 5000 time steps and plot last 500 time steps. We give results for (a) $s = 1$ (very local dispersal) (b) $s = 2$, (c) $s = 3$, (d) $s = 5$, (e) $s = 10$ and (f) $s = 50$ (close to global dispersal).

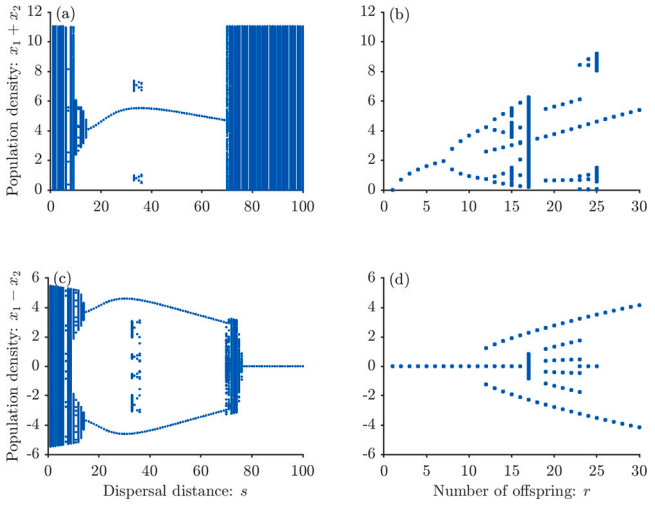


Fig. 11. Bifurcation diagram for the long-range dispersal approximation, given in Eq. (19). In (a) and (c) we fix r to 30, set $D = 201$, vary s, and iterate through $x_{1,t}$ and $x_{2,t}$ for 5000 time steps and plot last 500 time steps of 50 realizations of the total population, $(x_{1,t} + x_{2,t})/2$, (a) and the difference in population, $(x_{1,t} - x_{2,t})/2$, (c). In (b) and (d) we fix s to 45 and D to 201, vary r, and iterate through $x_{1,t}$ and $x_{2,t}$ for 5000 time steps and plot last 500 time steps of 50 realizations of the total population, $(x_{1,t} + x_{2,t})/2$, (b) and the difference in population, $(x_{1,t} - x_{2,t})/2$, (d).

so, we first observe that $x_* = 0$ is always a steady state for Eq. (14). We then check the stability for the steady state $x_* = 0$. Differentiating $f(x_*)$ we obtain:

$$f'(x_*) = r\left(\frac{(2s+1)^2-1}{(2s+1)^2}\right)^{r-1}\left(x_*\left(\left(\frac{(2s+1)^2-1}{(2s+1)^2}\right)^r - 1\right) + 1\right)^{(2s+1)^2-2} \times \left((2s+1)^2 x_*\left(\left(\frac{(2s+1)^2-1}{(2s+1)^2}\right)^r - 1\right) + 1\right) \tag{22}$$

When $x_* = 0$, we have

$$f'(0) = r\left(\frac{(2s+1)^2-1}{(2s+1)^2}\right)^{r-1} = r\left(1 - \frac{1}{(2s+1)^2}\right)^{r-1}. \tag{23}$$

The stability condition $|f'(x_*)| < 1$ for $x_* = 0$ yields

$$\left| r\left(1 - \frac{1}{(2s+1)^2}\right)^{r-1} \right| < 1, \tag{24}$$

which, for $s = 1$, is true for $1 < r < 30$. Thus, for our local correlation approximation, the population goes extinct when $r > 30$, which is seen in Fig. 10(a). This is a slightly larger value than what is seen in bifurcation diagram of the individual-based model (Fig. 3(a)), where, after $t = 5000$ time steps, extinction is observed already for $r = 23$. When s increases, the extinction value for r in the local correlation approximation increases as well. In Fig. 12(a), we see how the value of s affects $f'(x_*)$ when $x_* = 0$. For $s = 2$ e.g., r needs to be larger than 117 for the population to go extinct. For global dispersal, we have that

$$\lim_{s\to+\infty} f'(0) = \lim_{s\to+\infty} r\left(1 - \frac{1}{(2s+1)^2}\right)^{r-1} = r, \tag{25}$$

and thus the inequality (24) is only fulfilled for $r < 1$. This means that the local correlation approximation never result in extinction for global dispersal. In Fig. 12 below, we see a graphical illustration of this: when $s = 1$, $f'(x_*)$ crosses the line $y = 1$ when $r = 30$. As s increases, the slope of $f'(x_*)$ increases. For the global approximation equation, $f'(x_*) = r$, and will thus never cross the line $y = 1$. In Fig. 12(b) we see $f(x_t)$ as a function of x_t when $r = 20$ and for different values of s. In this cobweb diagram, we see that the population dynamics becomes unstable as s increases.

Now for our local correlation approximation with $R_{ij}^t \sim \text{Bin}(r/q, q)$, given by Eq. (18), $x_* = 0$ will also be a steady state, and $f'(x_*)|_{x^*=0}$ will be given by

$$f'(x_*)|_{x^*=0} = r\left(1 - \frac{q}{(2s+1)^2}\right)^{r/p-1}. \tag{26}$$

The stability of $x_* = 0$ of the local correlation approximation for stochastic reproductive rate will thus be affected by the value of q, and for smaller values, the population will become extinct for higher r values.

4.3. General extinction probability

The individual-based model clearly displays population extinction for some values of r when dispersal is global, as seen in the bifurcation diagram in Fig. 3(f). However, this extinction is not due to the stability of the stationary point $x_* = 0$, but due to the stochasticity of the individual-based model: when r is sufficiently large there is always a probability that all sites get overcrowded in one generation and the population goes extinct. We now want to find for which values of r this happens. To find an expression for this, we make use of the different life-cycle stages within one time step: the population before

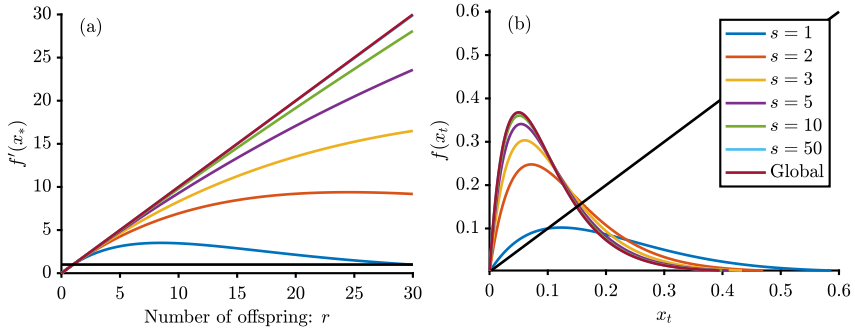


Fig. 12. (a) Plot of $f'(x_*)$, for $x_* = 0$ for various values of s (see legend) . The model denoted global approximation is $f(x_t) = rx_t e^{-rx_t}$, i.e. Eq. (6). The black line shows $f'(x_*) = 1$. (b) Plot of the approximation in Eq. (14) for various values of s (see legend) and for $r = 20$. The model denoted Global is $f(x_t) = rx_t e^{-rx_t}$, i.e. Eq. (6)). The black line shows $x_{t+1} = x_t$.

competition, the population after competition but before reproduction, and the population after reproduction. Consider the individual-based model with $n = D \times D$ resource sites. For the population to survive after competition, at least one resource site must contain exactly one individual (in case of scramble competition). Let x_t be the number of individuals before competition and reproduction at generation t, and $x_{t'}$ be the number of individuals after reproduction in the same time step. When dispersal is global, the probability that a particular resource site contains exactly one individual, $P(X = 1)$, will be a function of x_t, $p_1(x_t)$, since X is Poisson distributed with mean x_t/n. We can thus write

$$p_1(x_t) = \frac{x_t}{n} e^{-\frac{x_t}{n}}. \tag{27}$$

$p_1(x_t)$ will be maximized at $x_t = n$:

$$p_1(n) = \frac{n}{n} e^{-\frac{n}{n}} = e^{-1} = p_1^*. \tag{28}$$

Thus $p_1^* > p_1(x_t)\ \forall x_t$. When this is maximized, there will be $e^{-1} \cdot n$ individuals that survived after competition, meaning that, if the reproductive rate is constant, there will be

$$x_{t'} = nre^{-1} \tag{29}$$

individuals after reproduction. Thus, the probability for a particular resource site containing exactly one individual after reproduction

$$p_1(x_{t'}) = \frac{nre^{-1}}{n} e^{-\frac{nre^{-1}}{n}} = re^{-1}e^{-re^{-1}}. \tag{30}$$

Then the probability that none of the resource sites has exactly one individual is given by

$$p^*_{\text{extinct}} = (1 - p_1(x_{t'}))^n = (1 - re^{-1}e^{-re^{-1}})^n. \tag{31}$$

Since $p_1^* > p_1(x_t)\ \forall x_t$, $p^*_{\text{extinct}} < p_{\text{extinct}}(x_{t'})\ \forall x_t$. Thus, this is the minimum probability of going extinct for each generation. In Fig. 13 we see how this probability is affected for different values of D and r, and how this is connected to the actual simulations of the individual-based model (the black line and error bar). Recall that p_{extinct} is the minimum probability of going extinct *at each time step*, and thus when simulating over many generations (below for 5000 generations), the population will go extinct when p_{extinct} is quite low (≈ 0.1).

4.4. Convergence to the global approximation model

When deriving the local correlation approximation, we made the assumption that $s \ll D$. However Figs. 10 and 12(a) suggest that as s increases Eq. (14) comes increasingly close to the global approximation (Eq. (6)). Indeed, when $s = 50$ in Fig. 12(b), the curve defined by Eq. (14) is indistinguishable from the global approximation equation. These numerical results suggest that when s is large, the

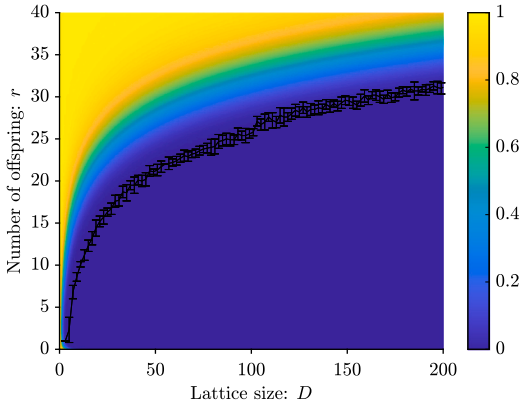


Fig. 13. Probability of going extinct in the next time step for different values of D and r. The heat map is produced from Eq. (31). The black line and the error bar show for each value of D, the mean and standard deviation of the lowest value of r for which the population was extinct after 5000 time steps for 10 simulation of each parameter set of the individual-based model.

local correlation approximation reproduces the dynamics of the global approximation, even though the $s \ll D$ assumption that was made for the approximation is not fulfilled. To rigorously determine whether this result holds, we need to check if our population dynamics model in Eq. (14) approaches the original global approximation as $s \to \infty$. This is equivalent to showing that the limit $\lim_{s\to+\infty} Y \to \text{Po}(rx_t)$, in distribution. Here Y is the number of offspring of all of the parents in the $(2s+1)\times(2s+1)$ Moore neighbourhood of a focal site that disperse to the focal site. Thus, $\lim_{s\to+\infty} Y \to \text{Po}(rx_t)$, means that when dispersal is global the number of offspring from the whole lattice that end up in a particular site, is Poisson distributed with mean rx_t. To show this limit, we can make use of the following theorem:

Theorem 1 (*Yannaros 1991*). *Let $\xi_1, \xi_2, \ldots$ be i.i.d. Bernoulli variables with $P(\xi_j = 1) = p$ and N a non-negative integer valued random variable which is independent of the ξ_j's. Let $S_N = \sum_{j=1}^{N} \xi_j$, then*

$$d(S_N, U_{pa}) \le \min\Big\{\frac{p}{2\sqrt{1-p}}, p\text{E}[1 - e^{-pN}]\Big\} + \min\Big\{p\text{E}[|N - a|], \frac{1}{2}\sqrt{p\frac{\text{E}[N-a]^2}{a}}\Big\}, \tag{32}$$

for any $a > 0$. Here d is the total variation distance and U_λ is the Poisson distribution with mean λ.

The total variation distance is a measure of the closeness of two probability distributions. Intuitively, the total variation distance between two probability distribution is the largest possible difference that

the two probability distributions can assign to the same event. Thus, a total variation distance equal to zero, indicates that the probability distributions are the same.

Proposition 1. *As $s \to \infty$, Y converges in distribution to a Poisson distribution with mean rx_t.*

Proof. By choosing a in the right way, we can use Theorem 1 to prove Proposition 1. First, recall that

$$Y = \sum_{j=1}^{F\cdot r} O_j. \tag{33}$$

Now since O_j are i.i.d. Bernoulli variables with $P(O_j = 1) = \frac{1}{(2s+1)^2}$, we can equate these with ξ_j, where $p = \frac{1}{(2s+1)^2}$. Likewise, we can equate $N = r \cdot F$, since it is an integer valued random variable which is independent on the O_j's. Thus Y is equal to S_N. We can now find an $a > 0$ such that $d(Y, U_{\frac{a}{(2s+1)^2}}) \to 0$ as $s \to \infty$.

Now choose a to be the expected value of $r \cdot F$, i.e. $a = \mathrm{E}[r \cdot F] = rx_t(2s+1)^2$. Since $p = \frac{1}{(2s+1)^2}$, $U_{pa} = U_{rx_t}$. We want to compute

$$\lim_{s\to+\infty} d(Y_{r\cdot F}, U_{rx_t}). \tag{34}$$

By applying Theorem 1 we know that

$$\begin{aligned}
&\lim_{s\to+\infty} d(Y_{r\cdot F}, U_{rx_t}) \\
&\leq \lim_{s\to+\infty} \Big(\min\Big\{ \frac{\frac{1}{(2s+1)^2}}{2\sqrt{1-(\frac{1}{(2s+1)^2})}}, \frac{1}{(2s+1)^2}\mathrm{E}[1-e^{-\frac{r\cdot F}{(2s+1)^2}}] \Big\} \\
&\quad + \min\Big\{ \frac{1}{(2s+1)^2}\mathrm{E}[|r\cdot F - a|], \frac{1}{2}\sqrt{p\frac{\mathrm{E}[r\cdot F-a]^2}{a}} \Big\} \Big).
\end{aligned} \tag{35}$$

Since $\frac{\frac{1}{(2s+1)^2}}{2\sqrt{1-(\frac{1}{(2s+1)^2})}} \to 0$ as $s \to \infty$, the first term, $\min\Big\{ \frac{\frac{1}{(2s+1)^2}}{2\sqrt{1-(\frac{1}{(2s+1)^2})}}, \frac{1}{(2s+1)^2}\mathrm{E}[1-e^{-\frac{r\cdot F}{(2s+1)^2}}] \Big\}$, will approach zero as $s \to \infty$.

For the second term, note that by choosing $a = E(r\cdot F) = rx_t(2s+1)^2$ we get that $\mathrm{E}(r\cdot F - a)^2 = \mathrm{E}[r\cdot F - \mathrm{E}[r\cdot F]] = \mathrm{Var}(r\cdot F) = (2s+1)^2 x_t(1-x_t)$. So the second term will be

$$\min\Big\{ \frac{1}{(2s+1)^2}\mathrm{E}[|r\cdot F - a|], \frac{1}{2}\sqrt{\frac{1}{(2s+1)^2}\frac{(2s+1)^2x_t(1-x_t)}{rx_t(2s+1)^2}} \Big\}.$$

Since $\frac{1}{2}\sqrt{\frac{1}{(2s+1)^2}\frac{(2s+1)^2x_t(1-x_t)}{rx_t(2s+1)^2}} = \frac{1}{2}\sqrt{\frac{(1-x_t)}{r(2s+1)^2}} \to 0$ as $s \to \infty$, we get that the second term goes to zero as $s \to \infty$. □

5. Discussion

With increasing computational capacity during the last 30 years, there has been a shift from 'top-down' models in ecology, describing the overall population structure, to 'bottom up' individual-based models capturing the local behaviour of the population. In light of this shift, a question arises as to whether it is possible to approximate individual-based models with a small number of analytically tractable equations, describing the overall population dynamics. For individual-based models in continuous time or space or in which evolution in space is 'smooth' [24,47–49], such as interacting particle systems and spatial point processes, there are a range of analytical approximations available [16,20,28]. For models which are, what Berec [29] refer to D-space, D-time, like the one we study here, analytical approaches have proved more limited.

The current work has started to look at ways into this problem, by studying a specific model which has both spatial patterning and large oscillations over a single time step. The spirit of our work here is perhaps similar to how Durrett and Levin first looked at problems for spatially discrete, but smoothly changing processes [13,50], which eventually led to the more rigorous treatments referenced above. But we would also emphasize our focus on qualitative methods to compare individual-based models with various approximations. We have used a standard method for qualitative comparison: studying the bifurcation plots of the models, as done by, e.g. [41,51,52]. We also produce a cobweb diagram for two of the models, as done by Anazawa [51], and look at how steady state stability depends on parameter values. These are a few of a number of comparison methods; studying time series outputs for different models as done by Nardini et al. [53], comparing mean and variance in population size between models [41] or computing the mean squared error between the time-series output of models [53], are other examples. The point of such approximations is not always to find a model for which we can prove convergence between approximation and individual-based model, but to gain a qualitative understanding of why an individual-based model produces the patterns that it does.

We have identified two novel qualitative approaches here. The first is a local correlation approximation, which led to Eq. (14). This provided a one-dimensional dynamical system, with two parameters – s for dispersal range and r for reproduction number – which allows us to the draw a bifurcation diagram illustrating how these parameters determine stability. In this way, we show that our individual-based model can be reasonably approximated for small dispersal distances (compare Figs. 3 (a) and 10 (a)). The second in the long-range approximation, consisting of two coupled maps, that captures some aspects of the simulation when s is large. Again, this allows us to draw a bifurcation diagram (Fig. 11).

Recently, interest has increased in the long-running discussion about the prevalence of chaos in natural populations [54–57]. After May's original paper [58] on chaotic dynamics in populations, empirical studies suggested that, in real ecological time series, chaos is infrequent [54,55,59]. Often, two explanations are given for this inconstancy: populations stabilize due to demographic stochasticity [60], or they are said to stabilize due to dispersal. With our local correlation approximation, we show that a deterministic difference equation does have this latter stabilizing effect on the population dynamics.

Interestingly, however, more recent empirical studies suggest that chaos is more prevalent in ecological systems than previously estimated [56,57]. One way which our model might help better understand when chaos will and won't be observed is to examine the scale on which observations are made. While our individual-based model is stable for small dispersal distances when measuring the whole population, the statistical measurements of the model show high variability on local level, suggesting that chaos could be present when measuring population sizes in more local areas. A similar argument can be made when comparing the total population and the difference in population of the long-range dispersal approximation: while the total population is chaotic for some parameter values, the difference in population is stable for the same values (see Fig. 11). It would be interesting then to look, in for example the 172 populations studied by Rogers et al. [56], at the relationship between the size of the area from which observations where collected (relative to the area over which the species disperse) and the degree to which the dynamics are chaotic.

The long-range approximation establishes a link between our model and two patch metapopulations discussed by [14,15]. Indeed, when dispersal is large in the individual-based model, the population dynamics can be well approximated with a two patch system, as seen in Figs. 5 and 11. Long-range dispersal has been shown to play an important role for seed dispersal and is key for understanding plant population dynamics and community composition [61].

The work presented here can be extended in several natural directions. First, our characterization of the spatial dynamics could be extended to other survival functions, like those examined by Brännström and Sumpter for the non-spatial case [37]. Second, our two approximation techniques can be considered also for situation with two sexes, or several populations. A better approximation should also result if correlations are tracked also between generation, as opposed to currently

within a single generation. The possibility to approximate spatially explicit individual-based models with analytically tractable dynamical systems does not only facilitate systematic exploration of parameter space by speeding up numerical investigations but can potentially reveal deeper insights about the role of space in population dynamics. Also, as we showed here, this approximation helped us establish an extinction probability curve, a result which could be extended further. We believe that the results presented here has contributed to this understanding and that future efforts will help to elucidate how and why spatial structure affects the dynamics of populations.

It is also worth pausing to think about the limits which our results point towards. Our starting question was the degree to which we could analyse the complex dynamics generated by a relatively simple individual-based model. The answer is that, through our two approximations and statistical measurements, we can get some insight into those dynamics, but this is not a complete mathematical explanation of the phenomena that arise, i.e. of the patterns we see in Figs. 2 and 4. In particular, neither of the approximations work particularly well for s between 3 and 10 in Fig. 2. This is in itself an interesting observation. Our model is far from being the most complex bottom-up model, yet there seems to be a limit to what we can currently say about it using mathematical approaches. Our approach goes beyond, for example, mean-field (well-mixed) equations, yet ultimately it still fails to capture the richness of the individual-based model. While we have little doubt that approximating stochastic individual-based models with analytical approaches is a useful contribution, it might also be that there are quite strong limits to what can readily be achieved with a mathematical analysis.

Declaration of competing interest

All authors declare that they have no conflicts of interest.

Data availability

The code for this article is available at https://github.com/linneagyllingberg/finding-analytical-approximations

Acknowledgements

We thank Tilo Wiklund, Erik Thörnblad and Svante Jansson for interesting and insightful discussions about the local correlation approximation. We thank the anonymous reviewers for helpful comments on earlier drafts.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.mbs.2023.109084.

References

[1] D.L. DeAngelis, S. Yurek, Spatially explicit modeling in ecology: a review, Ecosystems 20 (2) (2017) 284–300.
[2] S.W. Pacala, J. Silander Jr., Neighborhood models of plant population dynamics. I. Single-species models of annuals, Amer. Nat. 125 (3) (1985) 385–411.
[3] M.A. Nowak, R.M. May, Evolutionary games and spatial chaos, Nature 359 (6398) (1992) 826–829.
[4] G.B. Ermentrout, L. Edelstein-Keshet, Cellular automata approaches to biological modeling, J. Theoret. Biol. 160 (1) (1993) 97–133.
[5] M.C. Boerlijst, M.E. Lamers, P. Hogeweg, Evolutionary consequences of spiral waves in a host—parasitoid system, Proc. R. Soc. B 253 (1336) (1993) 15–18.
[6] S.W. Pacala, D. Tilman, Limiting similarity in mechanistic and spatial models of plant competition in heterogeneous environments, Amer. Nat. 143 (2) (1994) 222–257.
[7] I. Hanski, C.D. Thomas, Metapopulation dynamics and conservation: a spatially explicit model applied to butterflies, Biol. Cons. 68 (2) (1994) 167–180.
[8] J. Wu, S.A. Levin, A patch-based spatial modeling approach: conceptual framework and simulation scheme, Ecol. Model. 101 (2–3) (1997) 325–346.
[9] Y. Iwasa, M. Nakamaru, et al., Allelopathy of bacteria in a lattice population: competition between colicin-sensitive and colicin-producing strains, Evol. Ecol. 12 (7) (1998) 785–802.
[10] R. Durrett, Stochastic spatial models, SIAM Rev. 41 (4) (1999) 677–718.
[11] M.J. Keeling, H.B. Wilson, S.W. Pacala, Reinterpreting space, time lags, and functional responses in ecological models, Science 290 (5497) (2000) 1758–1761.
[12] D.L. DeAngelis, Individual-Based Models and Approaches in Ecology: Populations, Communities and Ecosystems, CRC Press, 2018.
[13] R. Durrett, S. Levin, The importance of being discrete (and spatial), Theor. Popul. Biol. 46 (3) (1994) 363–394.
[14] M. Gyllenberg, G. Söderbacka, S. Ericsson, Does migration stabilize local population dynamics? Analysis of a discrete metapopulation model, Math. Biosci. 118 (1) (1993) 25–49.
[15] A. Hastings, Complex interactions between dispersal and dynamics: lessons from coupled logistic equations, Ecology (1993) 1362–1372.
[16] K. Oelschläger, On the derivation of reaction-diffusion equations as limit dynamics of systems of moderately interacting stochastic processes, Probab. Theory Related Fields 82 (4) (1989) 565–586.
[17] A. Stevens, H.G. Othmer, Aggregation, blowup, and collapse: the ABC's of taxis in reinforced random walks, SIAM J. Appl. Math. 57 (4) (1997) 1044–1081.
[18] S.A. Isaacson, J. Ma, K. Spiliopoulos, How reaction-diffusion PDEs approximate the large-population limit of stochastic particle models, SIAM J. Appl. Math. 81 (6) (2021) 2622–2657.
[19] S.A. Isaacson, J. Ma, K. Spiliopoulos, Mean field limits of particle-based stochastic reaction-diffusion models, SIAM J. Math. Anal. 54 (1) (2022) 453–511.
[20] B. Bolker, S.W. Pacala, Using moment equations to understand stochastically driven spatial pattern formation in ecological systems, Theor. Popul. Biol. 52 (3) (1997) 179–197.
[21] U. Dieckmann, R. Law, J.A. Metz, The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, 2000.
[22] D.J. Murrell, U. Dieckmann, R. Law, On moment closures for population dynamics in continuous space, J. Theoret. Biol. 229 (3) (2004) 421–432.
[23] O. Ovaskainen, D. Finkelshtein, O. Kutoviy, S. Cornell, B. Bolker, Y. Kondratiev, A general mathematical framework for the analysis of spatiotemporal point processes, Theor. Ecol. 7 (1) (2014) 101–113.
[24] A. Surendran, M.J. Plank, M.J. Simpson, Spatial moment description of birth–death–movement processes incorporating the effects of crowding and obstacles, Bull. Math. Biol. 80 (11) (2018) 2828–2855.
[25] N. Bordj, N. El Saadi, Moment approximation of individual-based models. Application to the study of the spatial dynamics of phytoplankton populations, Appl. Math. Comput. 412 (2022) 126594.
[26] K. Sato, Pair approximation for lattice-based ecological models, in: The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, 2000, pp. 341–358.
[27] M. Van Baalen, Pair approximations for different spatial geometries, in: The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Vol. 742, Cambridge University Press Cambridge, UK, 2000, pp. 359–387.
[28] D.D. Patterson, S.A. Levin, C. Staver, J.D. Touboul, Probabilistic foundations of spatial mean-field models in ecology and applications, SIAM J. Appl. Dyn. Syst. 19 (4) (2020) 2682–2719.
[29] L. Berec, Techniques of spatially explicit individual-based models: construction, simulation, and mean-field analysis, Ecol. Model. 150 (1–2) (2002) 55–81.
[30] E.O. Wilson, et al., The Insect Societies, Cambridge, Massachusetts, USA, Harvard University Press [Distributed by ..., 1971.
[31] J.T. Costa, J.T. Costa, The Other Insect Societies, Harvard University Press, 2006.
[32] C.R. Brown, M.B. Brown, Coloniality in the Cliff Swallow: The Effect of Group Size on Social Behavior, University of Chicago Press, 1996.
[33] M.E. Hibbing, C. Fuqua, M.R. Parsek, S.B. Peterson, Bacterial competition: surviving and thriving in the microbial jungle, Nat. Rev. Microbiol. 8 (1) (2010) 15–25.
[34] L. Hall-Stoodley, J.W. Costerton, P. Stoodley, Bacterial biofilms: from the natural environment to infectious diseases, Nat. Rev. Microbiol. 2 (2) (2004) 95–108.
[35] A. Morozov, J.-C. Poggiale, From spatially explicit ecological models to mean-field dynamics: The state of the art and perspectives, Ecol. Complex. 10 (2012) 1–11.
[36] A. Johansson, D.J. Sumpter, From local interactions to population dynamics in site-based models of ecology, Theor. Popul. Biol. 64 (4) (2003) 497–517.
[37] Å. Brännström, D.J. Sumpter, The role of competition and clustering in population dynamics, Proc. R. Soc. Lond. Ser. B: Biol. Sci. 272 (1576) (2005) 2065–2072.
[38] M. Anazawa, Individual-based competition between species with spatial correlation and aggregation, Bull. Math. Biol. 76 (8) (2014) 1866–1891.
[39] T. Royama, Analytical Population Dynamics, Vol. 10, Springer Science & Business Media, 2012.
[40] M. Anazawa, Bottom-up derivation of discrete-time population models with the allee effect, Theor. Popul. Biol. 75 (1) (2009) 56–67.
[41] Å. Brännström, D.J. Sumpter, Coupled map lattice approximations for spatially explicit individual-based models of ecology, Bull. Math. Biol. 67 (4) (2005) 663–682.

[42] A.J. Nicholson, An outline of the dynamics of animal populations, Aust. J. Zool. 2 (1) (1954) 9–65.
[43] B. Schönfisch, Anisotropy in cellular automata, Biosystems 41 (1) (1997) 29–41.
[44] M. Markus, Dynamics of a cellular automaton with randomly distributed elements, in: Mathematical Population Dynamics, CRC Press, 2020, pp. 413–434.
[45] M. Rosenberg, PySSaGE, beta version, 2021, GitHub repository, GitHub, https://github.com/msrosenberg/pyssage.
[46] D. Sumpter, D. Broomhead, Relating individual behaviour to population dynamics, Proc. R. Soc. Lond. Ser. B: Biol. Sci. 268 (1470) (2001) 925–932.
[47] J. Touboul, Propagation of chaos in neural fields, Ann. Appl. Probab. 24 (3) (2014) 1298–1328.
[48] I. Omelyan, Y. Kozitsky, Spatially inhomogeneous population dynamics: beyond the mean field approximation, J. Phys. A 52 (30) (2019) 305601.
[49] P.J. Wallhead, A.P. Martin, M.A. Srokosz, Spatially implicit plankton population models: transient spatial variability, J. Theoret. Biol. 253 (3) (2008) 405–423.
[50] S.A. Levin, The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture, Ecology 73 (6) (1992) 1943–1967.
[51] M. Anazawa, Bottom-up derivation of population models for competition involving multiple resources, Theor. Popul. Biol. 81 (2) (2012) 158–167.
[52] W.T. Gibson, W.G. Wilson, Individual-based chaos: Extensions of the discrete logistic model, J. Theoret. Biol. 339 (2013) 84–92.
[53] J.T. Nardini, R.E. Baker, M.J. Simpson, K.B. Flores, Learning differential equation models from stochastic agent-based model simulations, J. R. Soc. Interface 18 (176) (2021) 20200987.
[54] M.P. Hassell, J.H. Lawton, R. May, Patterns of dynamical behaviour in single-species populations, J. Anim. Ecol. (1976) 471–486.
[55] R.M. Sibly, D. Barker, J. Hone, M. Pagel, On the stability of populations of mammals, birds, fish and insects, Ecol. Lett. 10 (10) (2007) 970–976.
[56] T.L. Rogers, B.J. Johnson, S.B. Munch, Chaos is not rare in natural ecosystems, Nat. Ecol. Evol. 6 (8) (2022) 1105–1111.
[57] S.B. Munch, T.L. Rogers, B.J. Johnson, U. Bhat, C.-H. Tsai, Rethinking the prevalence and relevance of chaos in ecology, Annu. Rev. Ecol. Evol. Syst. 53 (2022) 227–249.
[58] R.M. May, et al., Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos, Science 186 (4164) (1974) 645–647.
[59] S. Salvidio, Stability and annual return rates in amphibian populations, Amphibia-Reptilia 32 (1) (2011) 119–124.
[60] S. Jaggi, A. Joshi, Incorporating spatial variation in density enhances the stability of simple population dynamics models, J. Theoret. Biol. 209 (2) (2001) 249–255.
[61] D.J. Levey, J.J. Tewksbury, B.M. Bolker, Modelling long-distance seed dispersal in heterogeneous landscapes, J. Ecol. 96 (4) (2008) 599–608.