# Correcting mean-field approximations for birth-death-movement processes

Ruth E. Baker*
*Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3PN, United Kingdom*

Matthew J. Simpson
*School of Mathematical Sciences, Queensland University of Technology, G.P.O. Box 2434, Brisbane, Queensland 4001, Australia*
(Received 26 May 2010; published 5 October 2010)

On the microscale, migration, proliferation and death are crucial in the development, homeostasis and repair of an organism; on the macroscale, such effects are important in the sustainability of a population in its environment. Dependent on the relative rates of migration, proliferation and death, spatial heterogeneity may arise within an initially uniform field; this leads to the formation of spatial correlations and can have a negative impact upon population growth. Usually, such effects are neglected in modeling studies and simple phenomenological descriptions, such as the logistic model, are used to model population growth. In this work we outline some methods for analyzing exclusion processes which include agent proliferation, death and motility in two and three spatial dimensions with spatially homogeneous initial conditions. The mean-field description for these types of processes is of logistic form; we show that, under certain parameter conditions, such systems may display large deviations from the mean field, and suggest computationally tractable methods to correct the logistic-type description.

DOI: 10.1103/PhysRevE.82.041905 PACS number(s): 87.10.Ed, 87.10.Hk, 87.10.Mn, 87.10.Rt

## I. INTRODUCTION

Cell motility, proliferation and death play critical roles in the development and repair of organisms [1]. In combination, these cell-level processes can ultimately determine the fate of cell populations. For example, a reduction in cell motility can lead to crowding effects and suppressed growth. This was demonstrated by Barrandon and co-workers [2] who showed that small colonies of cells undergo exponential growth, but that this growth becomes linear as the colony becomes larger and crowding effects become important (Fig. 1). Addition of epidermal growth factor (EGF) and transforming growth factor $\alpha$ (TGF-$\alpha$) increased the growth rates substantially, with colonies becoming 30–50 times larger. The enhanced growth of the population was attributed to the ability of these chemicals to increase the rate of cell migration. This demonstrates that the net growth of a cell population is determined by a complicated interaction between cell motility and cell proliferation owing to cell-cell contact and cell crowding effects. Similarly, several tumor cell lines display unregulated cell motility [3,4], and it has been suggested that altering tumor cell movement can disrupt epithelial homeostasis, leading to excessive growth of the population [5].

Several models have been proposed to study the combined effects of cell motility, proliferation and death, and we outline a subset of these here. The Eden model [6] is a lattice-based model that can be used to investigate how a single cell can proliferate to form a cluster of cells, and to determine conditions required for the formation of symmetric or asymmetric clusters. Zygourakis and co-workers [7,8] consider a lattice gas cellular automata model in three dimensions. Their model accounted for cell movements, directional changes and the pauses in movement associated with cell-cell collisions. They show that cell migration is generally sufficient to overcome the adverse effects of contact inhibition on growth, and that increased migration speed leads to enhanced tissue growth. Further, they demonstrate that population growth rates depend crucially on the initial seeding cell density and location. However, due to the complex nature of the model they were unable to derive any analytical results and relied on simulation results alone. Hernández-

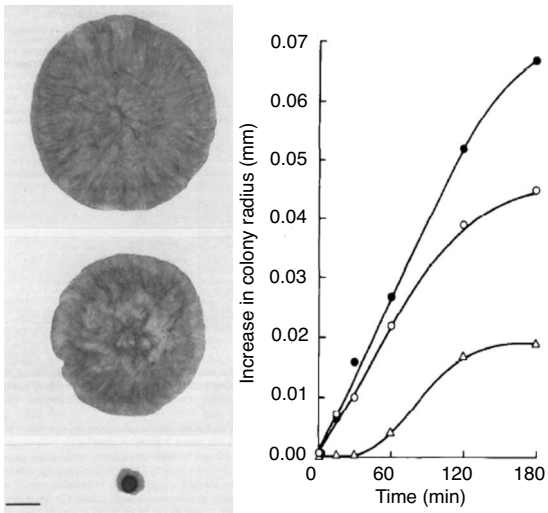


FIG. 1. Experimental data described by Barrandon and Green [2] illustrate the complicated relationship between cell motility and cell proliferation. Three colonies are shown after 24 days' incubation. The top colony (black circles) was supplied with TGF-$\alpha$, the middle colony (open circles) was supplied with EGF and the bottom colony (open triangles) was supplied with neither chemical. These chemicals upregulate cell motility and we see that the increase in motility leads to dramatic population growth over a period of 24 days. Reprinted from [2] with permission from Elsevier.

*ruth.baker@maths.ox.ac.uk

 ©2010 The American Physical Society

García and López [9] considered a "Brownian Bug" model in which both birth and death are first order events: they discuss the deviation from mean-field due to the development of spatial correlations and analyze the formation of spatial patterning. This previous work introduces a simple mechanism for cell-cell interactions for the birth and death processes, however, the cell migration mechanism does not include any cell-cell contact or contact inhibition of migration effects. Sander and co-workers consider a lattice-based exclusion process model to study cell migration and proliferation in the context of epidermal wound healing [10]. Their models replicate traveling-wave-like invasive fronts, the speed of these fronts matches the mean-field theory for small proliferation rates. Simpson and co-workers [11–14] have investigated a number of aspects of exclusion processes, including the connection between cell and population scales, the presence of multiple species, biased movement rates and different motility mechanisms and Codling and co-workers provide a thorough review of discrete motility mechanisms applied to cell biology problems in [16].

It has been well documented that the relative rates of migration, proliferation and death control whether spatial heterogeneity arises from an initially uniform field. This spatial heterogeneity is associated with the formation of spatial correlations which can retard population growth. These effects are typically neglected in modeling studies and simple phenomenological models, such as the logistic growth model or generalizations thereof, are used to represent population growth [17,18]. One important reason for the widespread use of these phenomenological models is that, to the best of our knowledge, there is no standard approach for incorporating the effects of correlations into models in a simple and computationally efficient manner. We seek to address this issue here, using an analytically tractable model of a birth-death-movement process to determine when these correlations effects become important, and demonstrate numerical techniques for dealing with them.

We address the issue of incorporating correlation effects by proposing a simple lattice-based exclusion process model that includes abstracted descriptions of cell motility, proliferation and death. The focus of this work is to study the growth of the population on a lattice that is initially populated such that the agents are distributed uniformly. We are motivated to study this problem as it is a standard experiment used to determine the growth rates of cells [18]. For example, Fig. 2 shows two images from an experiment used to measure the proliferation rate of a population of fibroblast cells *in vitro*. Initially, a small number of cells was placed uniformly into a two-dimensional culture medium [Fig. 2(a)]. Cells migrate and proliferate, and the number of cells present in the system increases with time [Fig. 2(b)]. The experimental growth curve is fitted to a logistic model and an estimate of the growth rate is obtained [Fig. 2(c)]. Similar experimental approaches have been used to understand how keratinocyte cells can be applied to burn wounds using aerosol delivery techniques [19]. The aerosol delivery results in an even distribution of cells onto the wound area, subsequently the cells migrate and proliferate to help restore normal tissue. In this work we model these kinds of experiments on two and three dimensional lattices and consider two distinct cases. *Case 1* involves simulating cell motility and proliferation only. This is appropriate for some cell biology systems where there is a complete absence of cell death [20,21]. *Case 2* involves simulating cell motility, proliferation and death which is appropriate for other cell biology systems where the cell death is thought to be important [22,23]. Our simulations and analysis show that there are important differences between these two cases. In both cases there can be large deviations from the mean-field approximation and we show how these mean-field approximations can be corrected. Several extensions of this work are also outlined.

### A. Outline of the problem

We consider an exclusion processes on a fixed, square lattice of dimension $d$. Each lattice site may be occupied by at most one agent and at any time the agent has a transition rate $P_m$ per unit time of moving to another lattice site, a proliferation rate $P_p$ per unit time of giving rise to another agent, and a death rate $P_d$ per unit time. Each of the processes of movement and proliferation may only take place if the target site is unoccupied. In each case we consider a

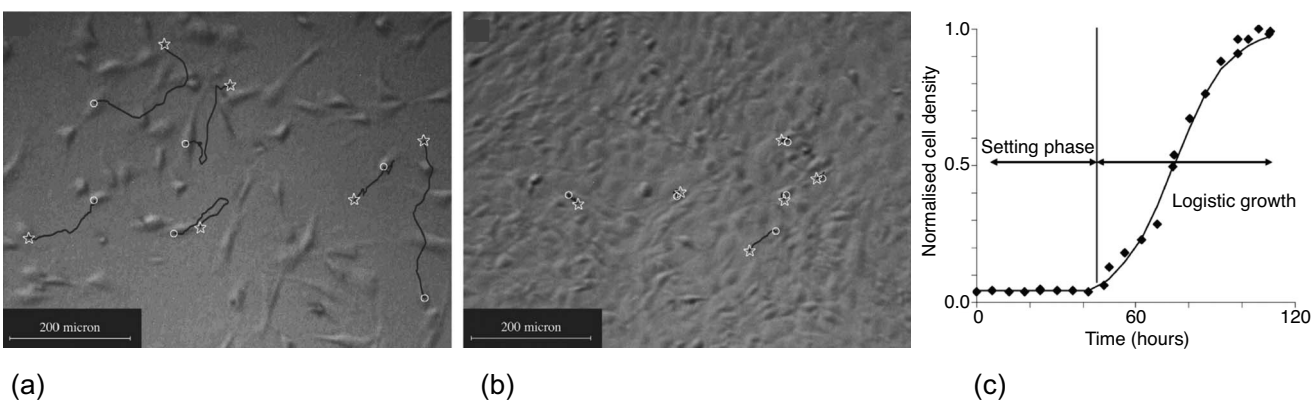


FIG. 2. Experiments performed by Tremel *et al.* [18] were used to estimate the proliferation rate of a population of fibroblast cells *in vitro*. Motile and proliferative fibroblast cells were placed uniformly in two-dimensional culture and the number of cells in the system was observed at regular intervals of time, for example, at (a) 54–60 and (b) 96–102 h after initial seeding. The growth in cell numbers was fitted to a logistic growth model (c). Reprinted from [18] with permission from Elsevier.

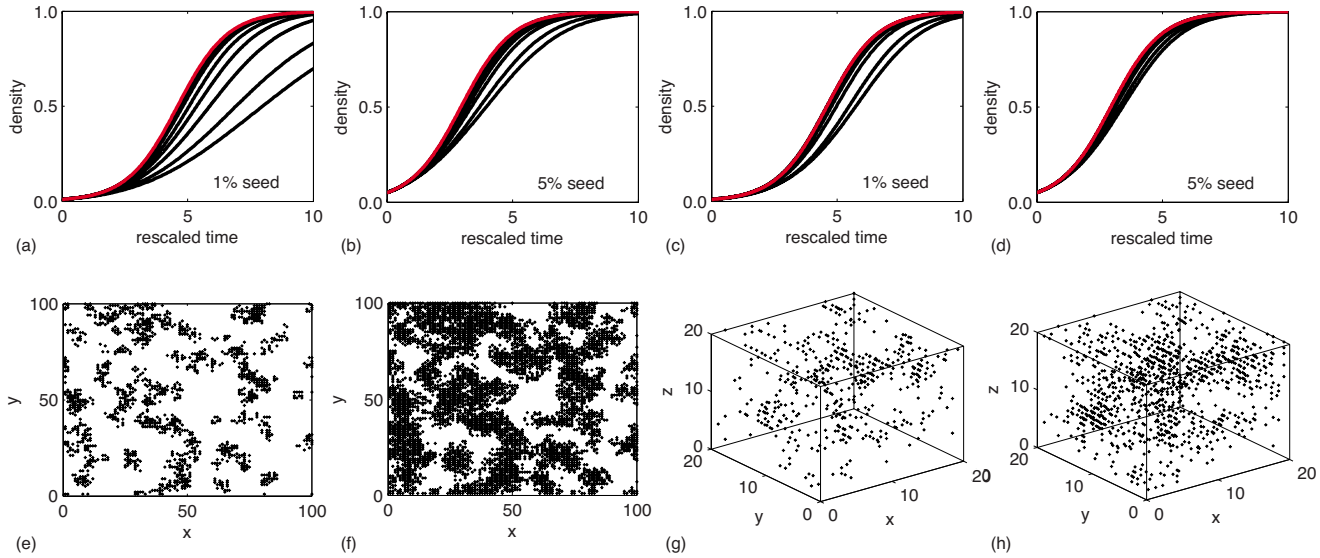


FIG. 3. (Color online) The influence of motility on proliferation patterns in two (left) and three (right) spatial dimensions for different initial conditions. The top plots show the evolution of density over time with (a) and (b) showing 2D results and (c) and (d) showing 3D results. The red (gray) lines show the solution of the logistic equation as a function of rescaled time, $\bar{t}=(P_p-P_d)t$, and the black lines the discrete results averaged over 40 realizations for $P_m=1.0$, $P_d=0.0$ and $P_p=0.005$, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0, again as a function of rescaled time. Agreement with the logistic equation becomes worse as $P_p$ is increased relative to $P_m$. The bottom plots show snapshots of the density at different time points for $P_p=0.5$ and illustrate the spatial correlations (clusters) that arise: (e) 2D simulation at rescaled time $t=4$; (f) 2D simulation at rescaled time $t=7$; (g) 3D simulation at rescaled time $t=2$; (g) 3D simulation at rescaled time $t=3$.

square lattice with fixed spacing $\Delta x$ in each direction, such that grid points are identified by their index, for example, $(i,j,k)$ where $i\in\{1,2,\ldots,N_x\}$ and $j\in\{1,2,\ldots,N_y\}$ and $k\in\{1,2,\ldots,N_z\}$. Periodic boundary conditions and uniform random initial conditions are imposed, and we denote the number of agents on the lattice at time $t$ by $Q(t)$. Throughout this report we suppose that both proliferation and movement can only take place to the $z=2d$ nearest neighbor sites, though all analysis is extendable to other cases.

### B. Aims for this work

As we show, the mean-field field prediction of this model, for initially uniformly seeded lattices, is that the population size grows logistically over time, with growth for different proliferation and death rates, $P_p$ and $P_d$, comparable on a time scale $\bar{t}=(P_p-P_d)t$. However, as Fig. 3 demonstrates, there may be a large deviation from the predicted mean-field growth rate as $P_p$ and $P_d$ are increased relative to $P_m$. Intuitively, we see that this comes about because correlations arise between the occupancy of lattice sites.

In order to tackle this problem, we outline a method for considering growth of the population with inclusion of the occupancy correlations between pairs of lattice sites. We begin by constructing expressions for evolution of the one- and two-point distribution functions describing the occupancy of lattice sites, where at each level there is dependence on higher-level distribution functions. We then outline an approximation that may be used in order to close our system at second order and compare our results with the mean-field and individual-based results. Finally, we demonstrate that our systems of equations may be simplified, rendering them efficient for numerical computation. We conclude our work with a short discussion our results and possible avenues for future research.

## II. CORRELATION ANALYSIS

In this section we outline the derivation of equations for the evolution of agent density, along with the occupancy correlations between lattice sites at all distances. For simplicity we consider only the following "reactions" between nearest neighbor lattice sites,

$$A+0\xrightarrow{P_p}A+A, \tag{1}$$

$$A+0\xrightarrow{P_m}0+A, \tag{2}$$

$$A\xrightarrow{P_d}0, \tag{3}$$

where the first "reaction" describes proliferation at rate $P_p$, the second movement at rate $P_m$, and the third death at rate $P_d$, all per unit time. All results are easily extendable to include different types of reaction, for example, where proliferation takes place to sites other than the nearest neighbor, or the proliferation or movement rates depend on local agent density [13]. Our aim is to accurately describe the evolution of the total number of agents over time. In order to realize this aim, we study the $k$-point distribution functions [24,25], $\rho^{(k)}$ $(k=1,2,\ldots)$, for agents on the lattice in order to gain information about the correlations between occupancy of different lattice sites—it is this which leads to deviation from the traditional logistic model. In essence the $\rho^{(k)}$ are multi-

variate probability distribution functions describing the occupancy of $k$-tuplets of lattice sites.

We define $\boldsymbol{l}$, $\boldsymbol{m}$, and $\boldsymbol{n}$ to be vectors that point to sites on the lattice, $\sigma_l \in \{0,A\}$ to be the lattice variable which describes the state of lattice site $\boldsymbol{l}$, and $z=2d$ to be the lattice constant (the number of "nearest neighbors" of each site).

For $k=1$ the distribution function does not depend on $\boldsymbol{l}$ and we may write

$$\rho^{(1)}(A_l) = c_A, \quad \rho^{(1)}(0_l) = c_0 = 1 - c_A, \tag{4}$$

where $c_i$ is the (total) density of agents of type $i$, in this case either agents of "type" A or empty sites, 0. In other words, $\rho^{(1)}(A_l)$ is the probability of finding an agent at site $\boldsymbol{l}$.

For $k=2$ the distribution function depends only on $|\boldsymbol{l}-\boldsymbol{m}|$ due to translational invariance of the lattice. We define the correlation functions [24,25] to be

$$F_{\lambda,\mu}(|\boldsymbol{l}-\boldsymbol{m}|) = \frac{\rho^{(2)}(\sigma_l,\sigma_m)}{\rho^{(1)}(\sigma_l)\rho^{(1)}(\sigma_m)} = \frac{\rho^{(2)}(\sigma_l,\sigma_m)}{c_\lambda c_\mu}, \tag{5}$$

for $|\boldsymbol{l}-\boldsymbol{m}| \neq 0$, and $F_{\lambda,\mu}(0)=0$. Note that, throughout, $\lambda$ denotes the state of site $\boldsymbol{l}$ and $\mu$ the state of site $\boldsymbol{m}$, and that $F_{\lambda,\mu}(|\boldsymbol{l}-\boldsymbol{m}|)$ implicitly depends on $t$.

As an asymptotic condition we have independence of the sites,

$$\lim_{|\boldsymbol{l}-\boldsymbol{m}|\to\infty} \rho^{(2)}(\sigma_l,\sigma_m) = \rho^{(1)}(\sigma_l)\rho^{(1)}(\sigma_m), \tag{6}$$

in other words that the occupancy of distant sites is uncorrelated. This gives

$$\lim_{|\boldsymbol{l}-\boldsymbol{m}|\to\infty} F_{\lambda,\mu}(|\boldsymbol{l}-\boldsymbol{m}|) = 1. \tag{7}$$

By summing over the lattice states we have

$$\sum_{\sigma_m} \rho^{(2)}(\sigma_l,\sigma_m) = \rho^{(1)}(\sigma_l), \tag{8}$$

i.e.,

$$\rho^{(2)}(A_l,A_m) + \rho^{(2)}(A_l,0_m) = \rho^{(1)}(A_l), \tag{9}$$

$$\rho^{(2)}(0_l,A_m) + \rho^{(2)}(0_l,0_m) = \rho^{(1)}(0_l). \tag{10}$$

Using this fact we see that the correlation functions are not independent,

$$c_A = c_A^2 F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|) + c_A(1-c_A)F_{A,0}(|\boldsymbol{l}-\boldsymbol{m}|), \tag{11}$$

and

$$(1-c_A) = c_A(1-c_A)F_{0,A}(|\boldsymbol{l}-\boldsymbol{m}|) + (1-c_A)^2 F_{0,0}(|\boldsymbol{l}-\boldsymbol{m}|), \tag{12}$$

give

$$F_{A,0}(|\boldsymbol{l}-\boldsymbol{m}|) = \frac{1-c_A F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)}{(1-c_A)}, \tag{13}$$

and

$$F_{0,0}(|\boldsymbol{l}-\boldsymbol{m}|) = \frac{1-2c_A+c_A^2 F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)}{(1-c_A)^2}. \tag{14}$$

In order to write down equations describing evolution of the two-point distributions/correlation functions, we require the three-point distribution functions, $\rho^{(3)}$, which satisfy a similar conservation rule,

$$\sum_{\sigma_n} \rho^{(3)}(\sigma_l,\sigma_m,\sigma_n) = \rho^{(2)}(\sigma_l,\sigma_m). \tag{15}$$

With one last piece of notation to define the nearest neighbors,

$$\alpha_{n,l} = \begin{cases} 1 & \text{if } \boldsymbol{n} \text{ and } \boldsymbol{l} \text{ are nearest neighbors,} \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

we are in a position to write down expressions for the one- and two-point distribution functions.

### A. One-point distribution functions

For the one-point distribution functions we have

$$\frac{d\rho^{(1)}(A_l)}{dt} = P_m \sum_n \frac{\alpha_{n,l}}{z}[\rho^{(2)}(0_l,A_n) - \rho^{(2)}(A_l,0_n)] + P_p \sum_n \frac{\alpha_{n,l}}{z}\rho^{(2)}(0_l,A_n) - P_d\rho^{(1)}(A_l),$$

which gives, upon simplification,

$$\frac{dc_A}{dt} = P_p c_A(1-c_A)F_{0,A}(1) - P_d c_A = P_p c_A[1 - F_{A,A}(1)c_A] - P_d c_A. \tag{17}$$

It is immediately obvious that, upon the assumption that neighboring sites are independent, and $F_{A,A}(1)\equiv 1$, the system reduces to the well-known logistic equation [17]. We rescale the logistic equation, using

$$\bar{c}_A = \left(\frac{P_p - P_d}{P_p}\right)c_A \Rightarrow \frac{d\bar{c}_A}{dt} = r\bar{c}_A(1-\bar{c}_A), \tag{18}$$

with solution

$$\bar{c}_A(t) = \frac{\bar{c}_A(0)e^{rt}}{1+\bar{c}_A(0)(e^{rt}-1)}, \tag{19}$$

where $r=(P_p-P_d)$ and $\bar{c}_A(0)=(P_p-P_d)/P_p c_A(0)$. Moreover, rescaling time using $\bar{t}=(P_p-P_d)t$ allows comparison of different parameter regimes,

$$\bar{c}_A(\bar{t}) = \frac{\bar{c}_A(0)e^{\bar{t}}}{1+\bar{c}_A(0)(e^{\bar{t}}-1)}. \tag{20}$$

Throughout this work we focus on the case where $r=(P_p-P_d)>0$ and we see net population growth, though all analysis is the same for $r\leq 0$ were we expect eventual population extinction.

We would like to point out that assuming that $F\equiv 1$ is a standard assumption made by us [12,14] and others [15]. While it is attractive to make this assumption to arrive at a simple mean-field model, this assumption is questionable, and without any analytical or computational tools to check

the validity of this assumption we have no way of assessing how good this assumption is. In comparison, here we have developed computational and analytical tools which allow us to relax this assumption so that we can quantify the effect of ignoring correlations.

Figure 3 compares the mean-field logistic approximation to the results of discrete simulations of the system, for $d=2,3$ where each lattice site is initially randomly occupied with probability 1% and 5%, without agent death. We see immediately that the logistic description becomes less accurate as $P_p$ is increased relative to $P_m$ and this can be understood intuitively by reasoning that agents disperse further between "birth" events when $P_p$ is small, allowing more of the agents space to proliferate. We also see that (i) the logistic description is less accurate with smaller initial seedings as correlations arise at lower population levels and (ii) the extent to which the logistic equation approximates the discrete simulation results increases as $d$, the number of spatial dimensions, increases [simulations in 1D (not shown) also corroborate this result].

### B. Two-point distribution functions

For the two-point distribution functions we have

$$\begin{aligned}\frac{d\rho^{(2)}(A_l,A_m)}{dt} &= P_m\sum_{n\neq l}\frac{\alpha_{n,m}}{z}[\rho^{(3)}(A_l,0_m,A_n)-\rho^{(3)}(A_l,A_m,0_n)]\\ &+P_m\sum_{n\neq m}\frac{\alpha_{n,l}}{z}[\rho^{(3)}(0_l,A_m,A_n)-\rho^{(3)}(A_l,A_m,0_n)]\\ &+P_p\left[\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(A_l,0_m,A_n)+\sum_{n\neq m}\frac{\alpha_{n,l}}{z}\right.\\ &\left.\times\rho^{(3)}(0_l,A_m,A_n)\right]+P_p\frac{\alpha_{l,m}}{z}[\rho^{(2)}(A_l,0_m)\\ &+\rho^{(2)}(0_l,A_m)]-2P_d\rho^{(2)}(A_l,A_m),\end{aligned}\tag{21}$$

where the first term on the last row describes the proliferation events taking place when $\boldsymbol{l}$ and $\boldsymbol{m}$ are nearest neighbors. The movement terms simplify using the conservation rule, for example,

$$\rho^{(3)}(A_l,0_m,A_n)+\rho^{(3)}(A_l,A_m,A_n)=\rho^{(2)}(A_l,A_n),\tag{22}$$

$$\rho^{(3)}(A_l,A_m,0_n)+\rho^{(3)}(A_l,A_m,A_n)=\rho^{(2)}(A_l,A_m),\tag{23}$$

while translational invariance of the lattice implies that terms centered around site $\boldsymbol{l}$ are equivalent to those centered around site $\boldsymbol{m}$. This gives

$$\begin{aligned}\frac{d\rho^{(2)}(A_l,A_m)}{dt} &= 2P_m\sum_{n\neq l}\frac{\alpha_{n,m}}{z}[\rho^{(2)}(A_l,A_n)-\rho^{(2)}(A_l,A_m)]\\ &-2P_d\rho^{(2)}(A_l,A_m)+P_p\frac{\alpha_{l,m}}{z}[\rho^{(2)}(A_l,0_m)\\ &+\rho^{(2)}(0_l,A_m)]+2P_p\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(A_l,0_m,A_n).\end{aligned}\tag{24}$$

We now rewrite these equations in terms of the correlation functions,

$$\rho^{(2)}(\sigma_l,\sigma_m)=c_\lambda c_\mu F_{\lambda,\mu}(|\boldsymbol{l}-\boldsymbol{m}|),\tag{25}$$

gives

$$\begin{aligned}\frac{d\rho^{(2)}(\sigma_l,\sigma_m)}{dt} &= c_\lambda c_\mu\frac{dF_{\lambda,\mu}}{dt}(|\boldsymbol{l}-\boldsymbol{m}|)\\ &+\left(c_\lambda\frac{dc_\mu}{dt}+c_\mu\frac{dc_\lambda}{dt}\right)F_{\lambda,\mu}(|\boldsymbol{m}-\boldsymbol{l}|),\end{aligned}\tag{26}$$

i.e., in our case

$$\begin{aligned}\frac{d\rho^{(2)}(A_l,A_m)}{dt} &= c_A^2\frac{dF_{A,A}}{dt}(|\boldsymbol{l}-\boldsymbol{m}|)+2c_A\frac{dc_A}{dt}F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)\\ &=c_A^2\frac{dF_{A,A}}{dt}(|\boldsymbol{l}-\boldsymbol{m}|)+2P_pc_A^2[1-c_AF_{A,A}(1)]\\ &\times F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)-2P_dc_A^2F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|).\end{aligned}\tag{27}$$

Therefore at each distance, $|\boldsymbol{l}-\boldsymbol{m}|$, the correlation functions evolve according to the equation

$$\begin{aligned}\frac{dF_{A,A}}{dt}(|\boldsymbol{l}-\boldsymbol{m}|) &= \frac{P_m}{d}\sum_{n\neq l}\alpha_{n,m}[F_{A,A}(|\boldsymbol{l}-\boldsymbol{n}|)-F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)]\\ &-2P_p[1-c_AF_{A,A}(1)]F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)\\ &+\frac{P_p}{dc_A}\alpha_{l,m}[1-c_AF_{A,A}(1)]+\frac{P_p}{d^2}\sum_{n\neq l}\alpha_{n,m}\\ &\times\rho^{(3)}(A_l,0_m,A_n).\end{aligned}\tag{28}$$

We note first that the only places in which we need to close the system are those where there are terms involving proliferation at rate $P_p$; the movement terms can all be reduced by considering the "conservation law." Second, we note that death does not explicitly feature in the evolution of the two-point distribution function as it is a first order "reaction." Finally, we recognize the first term, involving $P_m$, as a discrete lattice Laplacian [24,25],

$$\hat{F}_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)=\sum_{n\neq l}\alpha_{n,m}[F_{A,A}(|\boldsymbol{l}-\boldsymbol{n}|)-F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)].\tag{29}$$

### C. Closure approximation

The infinite chain of master equations cannot be closed to obtain a finite system of nonlinear ordinary differential equations (ODEs) without approximation, and it is the derivation of a suitable closure approximation that will concern the analysis here. As we have seen earlier, the only terms that require closure approximations are those that describe proliferation events and, as such, they are of the form

$$\rho^{(3)}(A_l,0_m,A_n),\tag{30}$$

where $\boldsymbol{m}$ and $\boldsymbol{n}$ are nearest neighbor sites. We now give a heuristic derivation of our choice of moment closure approximation, following [26]. We choose a symmetric, third

order closure approximation; testing of several, lower order, closure approximations found in the literature [27–29] gave less accurate approximations.

Supposing that sites $\boldsymbol{l}$ and $\boldsymbol{m}$ are independent we have

$$\rho^{(2)}(0_l,A_m) = \rho^{(1)}(0_l)\rho^{(1)}(A_m). \tag{31}$$

Using Bayes' Theorem we may write

$$\begin{aligned}\rho^{(3)}(A_l,0_m,A_n) &= \rho^{(3)}(0_m|A_l,A_n)\rho^{(2)}(A_l,A_n)\\ &= \rho^{(2)}(0_m|A_n)\rho^{(2)}(A_l,A_n)\\ &= \frac{\rho^{(2)}(0_m,A_n)}{\rho^{(1)}(A_n)}\rho^{(2)}(A_l,A_n).\end{aligned} \tag{32}$$

We then use Eq. (31) to write

$$\frac{\rho^{(2)}(0_l,A_m)}{\rho^{(1)}(0_l)\rho^{(1)}(A_m)} = 1. \tag{33}$$

Multiplying both sides of Eq. (32) by Eq. (33) gives the Kirkwood superposition approximation (KSA),

$$\rho^{(3)}(0_l,A_m,A_n) = \frac{\rho^{(2)}(0_l,A_n)\rho^{(2)}(0_l,A_m)\rho^{(2)}(A_m,A_n)}{\rho^{(1)}(0_l)\rho^{(1)}(A_m)\rho^{(1)}(A_n)}, \tag{34}$$

and hence

$$\begin{aligned}\rho^{(3)}(0_l,A_m,A_n) = c_A^2(1-c_A)F_{0,A}(|\boldsymbol{l}-\boldsymbol{m}|)F_{A,A}(|\boldsymbol{m}-\boldsymbol{n}|)\\ \times F_{0,A}(|\boldsymbol{n}-\boldsymbol{l}|).\end{aligned} \tag{35}$$

The KSA has been rigorously derived and shown to maximize entropy for some physical applications [26].

The system of correlation equations then becomes

$$\begin{aligned}\frac{dF_{A,A}}{dt}(|\boldsymbol{l}-\boldsymbol{m}|) &= \frac{P_m}{d}\hat{F}_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|) - 2P_p[1-c_AF_{A,A}(1)]\\ &\quad\times F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|) + \frac{P_p}{(1-c_A)}\alpha_{l,m}[1-c_AF_{A,A}(1)]\\ &\quad + \frac{P_p}{d(1-c_A)}[1-c_AF_{A,A}(1)]\\ &\quad\times[1-c_AF_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)]\Bigg[\hat{F}_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)\\ &\quad + \sum_{n\neq l}\alpha_{n,m}F_{A,A}(|\boldsymbol{l}-\boldsymbol{m}|)\Bigg].\end{aligned} \tag{36}$$

While these approximations were made under the assumption that, at least, sites $\boldsymbol{l}$ and $\boldsymbol{m}$ are independent, this is not generally the case. Particularly in this work as we are describing proliferation events in which agents deposit their progeny onto nearest neighbor sites. We will use numerical simulation, for a range of values of $P_p$ and $P_d$, to test the validity of this closure approximation, and we detail our results in the next sections.

## III. NUMERICAL METHODS

Here we outline the specific numerical methods used to solve the systems of closure equations and to generate the discrete results. In each case we use a regular, square lattice of size $N_x \times N_y = 100 \times 100$ in two dimensions (2D) and $N_x \times N_y \times N_z = 20 \times 20 \times 20$ in three dimensions (3D), giving approximately the same number of sites in each simulation. We use periodic boundary conditions, random initial conditions, and average over 40 identically prepared realizations of our discrete system to compare to our closure approximation. Without loss of generality, we take $P_m = 1.0$ throughout, varying both $P_p$ and $P_d$ relative to the motility rate. We present all results (unless otherwise stated) for a uniform initial seeding of 5%; extensive numerical simulations show that results are very similar for other initial seedings, with the KSA performing better (worse) for greater (smaller) initial seedings, as illustrated in Fig. 3. All results are plotted as a function of rescaled time, $\bar{t} = (P_p - P_d)t$, and rescaled density, $\bar{c}_A = (P_p - P_d)/P_p c_A$.

### A. Closure approximation

In order to compare the validity of our various closure approximations, we solve the systems of coupled ODEs arising from correlation analysis numerically in MATLAB using a standard fourth-order Runge-Kutta method with a constant time step of $\Delta t = 0.01$. In 2D/3D things are slightly complicated because of the requirement to include an equation for each distance $|\boldsymbol{l}-\boldsymbol{m}|$ on the grid—this becomes numerically expensive for large grids. In practice, we evaluate each distance up to the threshold $|\boldsymbol{l}-\boldsymbol{m}| = 5$ individually, and then move to a regular grid using a 2D/3D Laplacian with radial/spherical symmetry to approximate the lattice Laplacian terms up to a distance $\Delta x\sqrt{(N_x^2+N_y^2+N_z^2)}/2$ [24,25]. We will show, in later results, that truncating at $r=5$ gives an accurate approximation of the full system of equations. Remembering to scale the discrete Laplacian we have, in 2D, $2(\Delta x)^2 = (\Delta r)^2$ and

$$\nabla^2 f(r) = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} \tag{37}$$

$$\approx \frac{f(r_{i-1}) - 2f(r_i) + f(r_{i+1})}{(\Delta r)^2} + \frac{1}{r_i}\frac{f(r_{i+1}) - f(r_{i-1})}{2\Delta r}, \tag{38}$$

and in 3D, $3(\Delta x)^2 = (\Delta r)^2$ and

$$\begin{aligned}\nabla^2 f(r) = \frac{\partial^2 f}{\partial r^2} + \frac{2}{r}\frac{\partial f}{\partial r} &\approx \frac{f(r_{i-1}) - 2f(r_i) + f(r_{i+1})}{(\Delta r)^2}\\ &\quad + \frac{1}{r_i}\frac{f(r_{i+1}) - f(r_{i-1})}{\Delta r}.\end{aligned} \tag{39}$$

*Initial conditions.* A random initial seeding of the lattice ensures that all sites are uncorrelated and therefore that $F_{A,A}(r) = 1$, $\forall r$ at $t=0$.

*Boundary conditions.* By definition, $F_{A,A}(r) \to 1$ as $r \to \infty$ so we set $F_{A,A}(r_{\max}) \equiv 1$ (assuming that our domain is large enough to accommodate this approximation).

### B. Discrete simulations

We use a modified Gillespie approach [30] to update the discrete model. The system is updated at discrete time steps using the following algorithm:

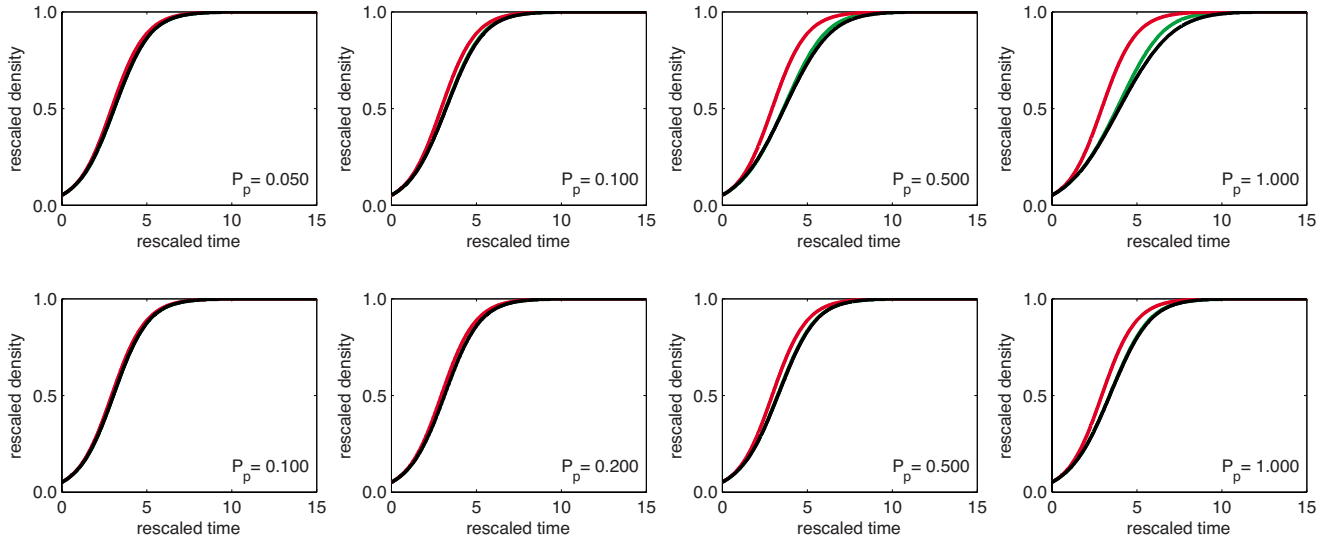


FIG. 4. (Color online) The influence of motility on proliferation patterns in 2D (top row) and 3D (bottom row). In each case, the red (dark gray) line shows the solution of the logistic equation, the black line the averaged discrete results and the green (light gray) line the results given by the inclusion of correlations. In most cases, the KSA results (green/light gray) are visually indistinguishable from the simulation results (black).

*Step 1.* Set $t=0$ and initialize the lattice by placing agents at the required lattice sites. Let $Q(t)$ be the number of agents.

*Step 2.* Calculate the total propensity function $a_0=(P_m+P_p+P_d)Q(t)$. Let $\tau=(1/a_0)\log(1/r_1^u)$ where $r_1^u$ is a uniform random number in the interval [0,1].

*Step 3.* Decide whether movement, proliferation or death occurs, according to the following: calculate $R=a_0 r_2^u$ where $r_2^u$ is a uniform random number in the interval [0,1].

(i) If $R\in[0,P_mQ(t))$ execute a movement event. Choose an agent at random and a target site for movement from the nearest neighbors (of which there are four in 2D and six in 3D). If the target site is empty then move the agent into that site. If not, abort the movement event.

(ii) If $R\in[P_mQ(t),(P_m+P_p)Q(t))$ execute a proliferation event. Choose an agent at random and a target site for proliferation from the nearest neighbors (of which there are four in 2D and six in 3D). If the target site is empty then let the agent proliferate by placing a new agent in the target site and let $Q(t)\mapsto Q(t)+1$, $a_0\mapsto a_0+(P_m+P_p+P_d)$. If not, abort the proliferation event.

(iii) If $R\in[(P_m+P_p)Q(t),(P_m+P_p+P_d)Q(t))$ execute a death event. Choose an agent at random and remove it from the system. Let $Q(t)\mapsto Q(t)-1$ and $a_0\mapsto a_0-(P_m+P_p+P_d)$.

*Step 4.* Update time by letting $t\mapsto t+\tau$.

*Step 5.* If $t\geq t_{final}$ exit, otherwise return to Step 2.

## IV. RESULTS

To get a good understanding of the effects of changing various parameter values and the number of spatial dimensions, we solved our systems numerically for a wide range of parameter values, with different initial seedings and in 2D/3D. We present a representative subset of these results in this section.

### A. Case 1: Without agent death

In the first instance we consider only birth and movement processes—though we have seen that on the mean-field level we can incorporate death by a simple rescaling, things are not so simple with correlations included, and hence we start by considering the simplest problem. Figure 4 shows results obtained with a variety of different values of $P_p$: we compare results from our discrete simulations, the mean-field (logistic) and KSA results.

The initial results are intuitive and simple to interpret. The KSA offers clear improvements upon the mean-field approximation in all cases. For almost all parameter regimes in 2D and 3D the KSA provides an accurate method for closing the systems of correlation equations. Although the results of the KSA are not always completely accurate in 2D they provide better estimates than the mean-field model. In particular, for $0.1<P_p/P_m<0.5$ the results are encouraging. Results in 3D are excellent, but suggest that we only need take correlations into account in a specific range of parameter space, where motility is highly restricted relative to proliferation.

### B. Case 2: Including agent death

We now consider the effect of agent death. In each case we take $P_p>P_d$ in order to ensure that the mean-field model predicts a positive growth rate, and a nonzero steady population density. Intuitively, we expect the inclusion of death to decrease correlations between neighboring lattice sites, thereby making the mean-field model more accurate. Results from simulations with a range of parameter values are shown in Fig. 5: we compare results from our discrete simulations with the mean-field and KSA results.

We see that the inclusion of agent death has somewhat counterintuitive effects on the accuracy of the KSA. One might expect death to decrease correlations, opening up gaps in crowded areas allowing agent movement, so that the mean-field approximation would fair better. However, this is not the case, as we see from Fig. 5 where the discrepancies between the simulation data and the mean-field model are much larger than comparable results in Fig. 4 without agent

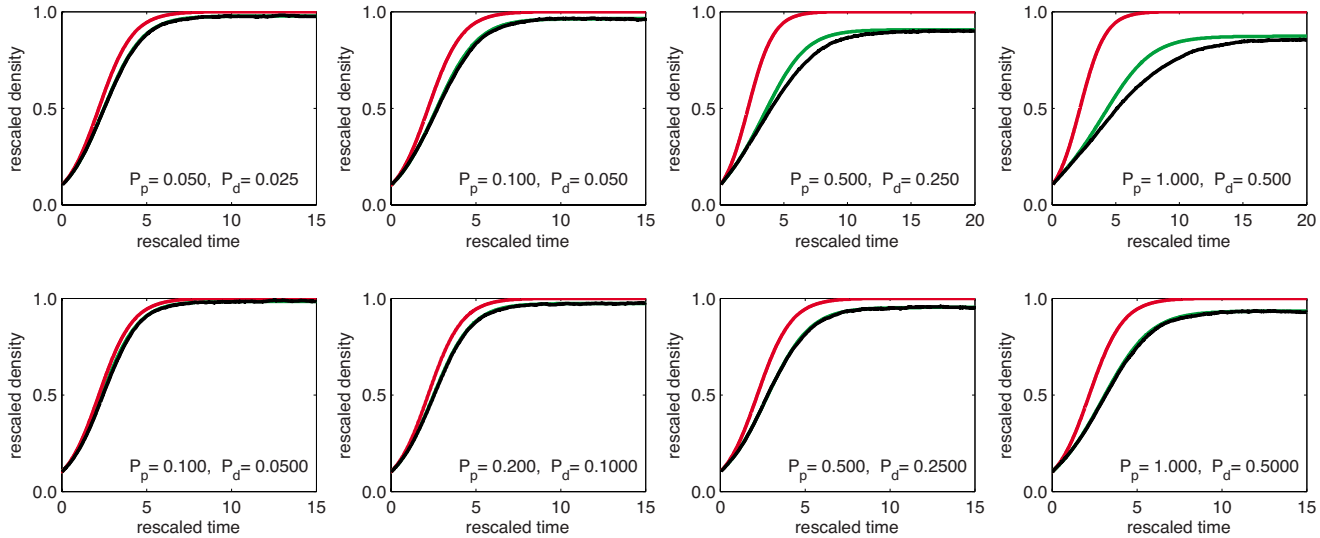


FIG. 5. (Color online) The influence of motility on proliferation patterns with the inclusion of agent death in 2D (top row) and 3D (bottom row). In each case, the red (dark gray) line shows the solution of the logistic equation, the black line the averaged discrete results and the green (light gray) line the results given by the inclusion of correlations.

death. While the KSA results in 2D are encouraging, particularly for proliferation rates up to approximately $P_p=0.5$, in 3D they are more so, with excellent agreement observed over all parameter regimes explored.

It should be noted that the most remarkable effect of including agent death is the deviation of the steady state population density from its predicted mean-field value: for example, we see up to a 10–20 % decrease in steady state population density when $P_d=P_p/2$. The KSA system correctly predicts this result; inclusion of the KSA is vital in order that these effects are properly captured. While one may have argued from initial results that inclusion of the correlations is not necessary in 3D, this deviation demonstrates why this is not the case. Intuitively we can understand this effect by reasoning that correlations impact only on proliferation rates and not on death rates because proliferation rates are aborted increasingly often as site occupancy becomes more correlated. So, by increasing the proliferation rate, and hence neighboring site correlations, we are effectively increasing the contribution of death relative to that of proliferation, leading to a suppression in steady state population numbers.

## V. ANALYZING THE EXTENT OF CORRELATION

The results of the previous section show that the KSA outperforms the mean-field equation at approximating agent density in both 2D and 3D. However, it can be time consuming to derive the exact equations to be solved (see the supplementary information [31] for more details) and also to develop code to solve them numerically. In order to simplify such systems, it is important to analyze the extent to which correlations are important in the model, and at what distance they become negligible. We carried out numerical simulations to investigate, and present our results in Fig. 6. In each case we show the correlation function at distances $r=1,2,\ldots,5$. The results in Fig. 6 also confirm that the inclusion of agent death increases the effects of correlations. We see that the maximum values of $F_{A,A}(1)$ is greater when we include agent death.

Overall, as one might expect, the correlations can be quite significant, especially between neighboring lattice sites. However, such effects quickly diminish with correlations rapidly tending to unity both in space and time. Moreover, as outlined in the previous section, it is now obvious that the inclusion of agent death increases intersite correlations, especially at $r=1,2$. In summary, our results suggest that, although correlations are at times quite large, we can truncate our system of correlation equations, without compromising accuracy, and it is to this we direct our attention in the following section.

### Reducing the system of correlation equations

Our analysis of the extent of correlations suggests that effects die off rapidly as the distance between lattice sites increases. Therefore we investigated the extent to which it is possible to truncate the system of correlation equations, solving for a range of distances less than $r_{\max}=\Delta x\sqrt{(N_x^2+N_y^2+N_z^2)}/2$ and using $F_{A,A}(r_{\max})=1$. Results for simulations with a range of parameter values and truncating the system of correlation equations at different distances are shown in Fig. 7.

Our results suggest that one can often truncate at relatively short distances without sacrificing accuracy. In 2D we can easily truncate at $r=3$ without changing the results appreciably. This involves considering seven equations in total (for $c_A$ and distances $r=1,\sqrt{2},2,\sqrt{5},\sqrt{8},3$). It is quick to solve such systems numerically, for example, using the standard Runge-Kutta fourth-order solver. As expected, results in 3D are even more encouraging, suggesting that $r=2$ is a suitable cutoff, and that even $r=1$ works well enough in some parameter regimes (see, for example, results with $P_p=1.0$, $P_d=0.5$, and $P_s=0.05$ in 3D). We conclude that truncation of the correlation functions on the nonuniform grid at

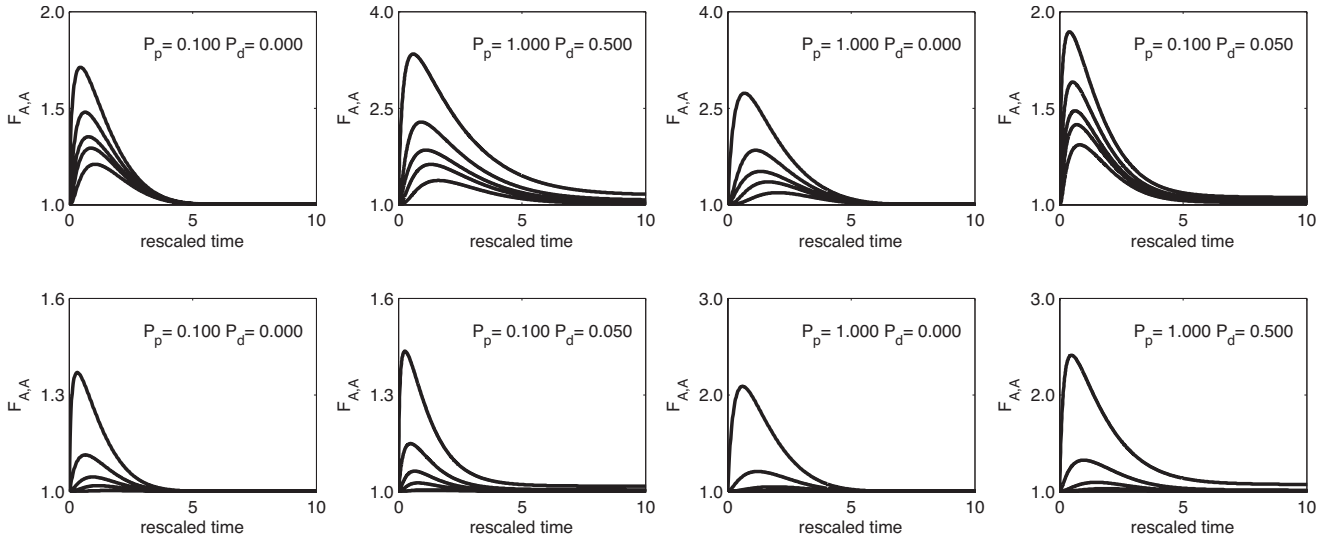


FIG. 6. Investigation of the correlation at different distances in 2D (top row) and 3D (bottom row). The results show $F_{A,A}(1), F_{A,A}(2), \ldots, F_{A,A}(5)$ (decaying as the distance increases).

$r \approx 3$ provides a good method for efficient computation of population sizes with both birth and death included; we detail the equations to be solved in these cases in the Supplementary Information [31].

## VI. DISCUSSION AND CONCLUSIONS

This work has been concerned with analyzing the extent to which spatial correlations affect the rates of population growth in a birth-death-movement exclusion process. Starting from a description of the probability of occupation of groups of lattice sites, we used a well-known closure approximation to reduce our system to one describing the total population numbers and pairwise correlations between lattice sites. Numerical results show that our approximation provides a vast improvement over the standard logistic approximation when birth rates become comparable to movement rates. Moreover, this approach can predict the decreases in steady state density that are observed when agent death is also included. Analysis of the pairwise correlations revealed that one may truncate the system of correlation equations at a distance $r=3$ lattice sites which leads to a computationally efficient and viable method of accurately predicting population growth rates in such systems.

Throughout this work we have made use of the KSA to truncate our system of $k$-point distribution functions at $k=2$. We note that, while a variety of different closure methods

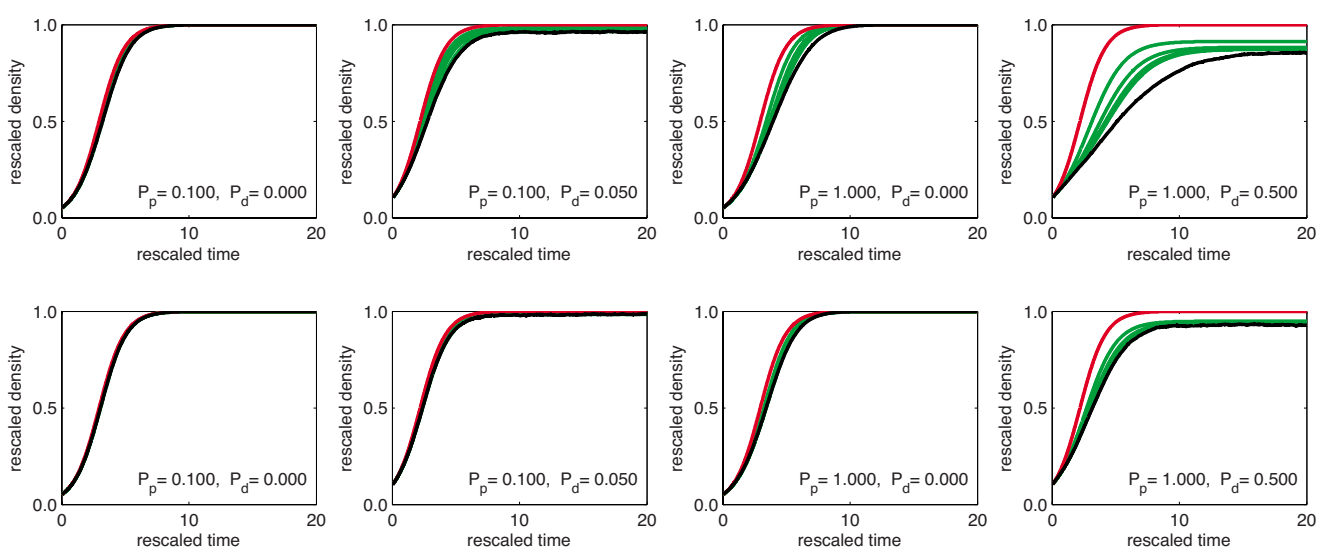


FIG. 7. (Color online) Investigating truncation of the correlation equations at different distances in 2D (top row) and 3D (bottom row). Mean field (red/dark gray) and simulation (black) results are presented together with a suite of solutions of the correlation equations (green/light gray). In each case, four solutions of the correlation equations are given with the most accurate representing the full system of correlations. The remaining three solutions of the correlation equations correspond to truncating the system at $r=1,2,3$ with the accuracy decreasing as $r$ decreasing.

can be found within the ecology literature (see, for example, [29] for a substantial list), the KSA provides the best approximation for our system (results not shown). We note also that, as a possible simplification of our system, we could solve our correlation equations on a regular spatial mesh, allowing one to avoid the time-consuming job of calculating the equations at each possible site separation. However, numerical results carried out to test this hypothesis were, in general, disappointing. Except for a mesh with spacing $\Delta r = 0.5$, corresponding reasonably well to the lattice distances 1, $\sqrt{2}$, and 2, a great deal of accuracy was lost. We conclude from our studies that the simple KSA gives an excellent method for closing the system of distribution equations, and that truncating the system at distance $r=3$ is sufficiently accurate for most purposes.

The inclusion of biased motility may simply be taken into account in our equations, as may other possible proliferation mechanisms, such as those where agents are deposited outside the set of nearest neighbor sites [14] (results not shown). Those mechanisms which involve more than two lattice sites for proliferation require higher order KSA approximations in order to truncate at $k=2$ and the analysis is an extension of that presented here. Similarly, the extension to multiple interacting or noninteracting cell populations is easy to include and for that reason we do not detail results here. Another obvious extension to this work is to consider systems which have nonhomogeneous initial conditions, or other types of boundary condition, as here the translational invariance of the lattice is lost and we must retain equations for the correlations at each individual lattice site. While it would be time consuming to write down such a system, our analysis for the homogeneous case suggests that some similar approximation may also be made in the nonhomogeneous case. Although this will still entail solving large systems of ODEs, we do now have the computing power to tackle such problems.

### A. Correcting the KSA

We considered a number of improvements to the KSA, in order that it may better enable us to predict population growth. First, we considered correcting the KSA in order that it satisfies the conservation laws described in this work [Eqs. (8) and (15)], as outlined in [24,25]. However, these corrections gave no noticeable improvements (results not shown) while adding to the algebraic complexity of the model. Second, we considered the suggestion of several authors [32–34], that the KSA can be corrected by incorporating a power series in the density, $c_A$, of the form

$$\rho^{(3)}(A_l,0_m,A_n) = c_A^2(1-c_A)F_{0,A}(|\boldsymbol{l}-\boldsymbol{m}|)F_{0,A}(|\boldsymbol{m}-\boldsymbol{n}|) \times F_{A,A}(|\boldsymbol{n}-\boldsymbol{l}|)\left[1+\sum_{i=1}^{\infty}\epsilon_i c_A^i\right]. \quad (40)$$

This suggests that we may close the system of two-point distribution equations more accurately if we are able to derive the coefficients of the expansion. We expect the $\epsilon_i$ to depend on $P_p$ and $P_d$ and we test this hypothesis numerically in 2D. We begin our investigation by truncating the power series and using the MATLAB function lsqnonlin to determine the $\epsilon_i$ for our simulation data: in each case we use time series data for agent density over a complete simulation, until the lattice is at least 99% full, and use the lsqnonlin function to determine the coefficients that give the best fit to this data. We calculate only a first order correction (i.e., finding only $\epsilon_1$) as the KSA is already reasonably accurate for most parameter regimes.

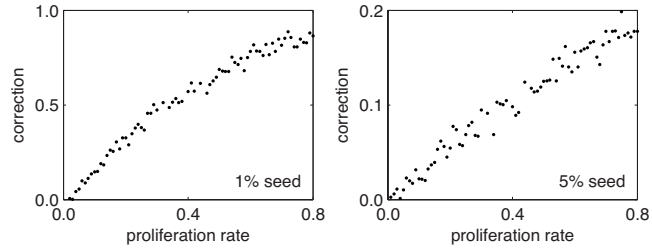


FIG. 8. Calculating the coefficient $\epsilon_1$ for the correction of the KSA with birth only and a 1% and 5% initial seedings. Note that a significantly larger correction is required for 1% seeding compared to 5% seeding.

Figure 8 shows the results of our analysis for birth-only simulations with both 1% and 5% initial seedings. We see that there is a clear relationship between the $\epsilon_i$, $P_p$ and the initial seeding and highlight inclusion of this correction term as important for future research in the area.

### B. Parameter estimation

The big question that arises from inclusion of correlation data is whether we can use such information when considering the inverse problem: given density and/or correlation data to what extent may parameters be recovered using optimization techniques? For example, how would predictions made by Tremel and co-workers [18] (Fig. 2) change if correlation information was included? To this end we considered fitting 2D simulation results to data using the MATLAB routine lsqnonlin, requiring the solver to minimize either

$$\sum_{i=0}^{N}[c_{A_i}^{data}-c_{A_i}^{sim}]^2, \quad (41)$$

or

$$\sum_{i=0}^{N}\{[c_{A_i}^{data}-c_{A_i}^{sim}]^2+[F_{A,A_i}^{data}(1)-F_{A,A_i}^{sim}(1)]^2\}, \quad (42)$$

depending on the amount of experimental data available: in principle, it should be possible to determine at least nearest neighbor correlations given snapshots of the population at each time point, though more research needs to be carried out to determine the extent to which this is possible. Without loss of generality, in each case we set $P_m=1.0$ and estimate $P_p$ and $P_d$. Note that we are able to do this as the equations may simply be rescaled, and that as a result we are only able to predict $P_p$ and $P_d$ relative to $P_m$.

Figure 9 shows the results of our analysis. The top row shows plots of fitting simulation data with $P_d=0.0$. Results are very promising: in particular, we see a large gain in the accuracy of our approximation compared estimation of the

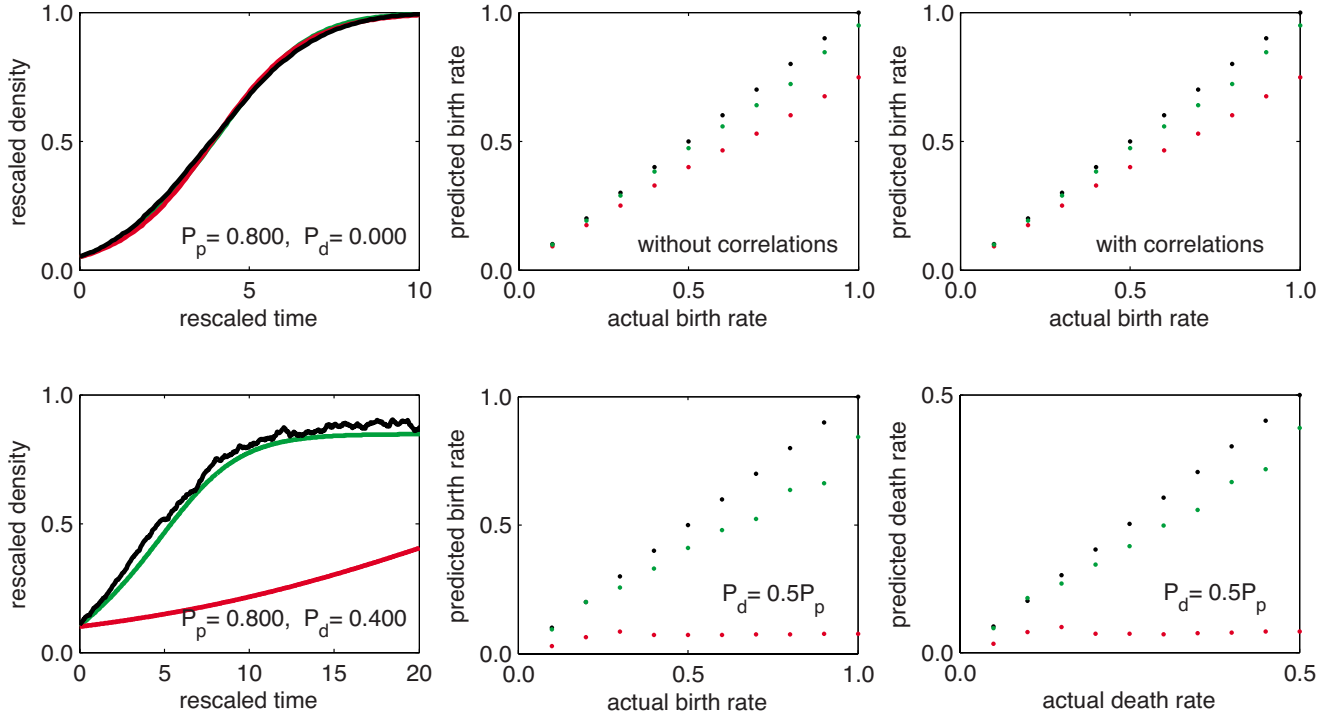


FIG. 9. (Color online) Estimation of parameter values with and without inclusion of the correlations. The top row shows the results of fitting with $P_d$=0.0. The left-hand figure shows the results of fitting density data only for $P_p$=0.8, $P_d$=0.0 with agent density shown in black, results from fitting with the inclusion of nearest neighbor correlations in green (light gray), and fitting the logistic curve in red (dark gray). The predicted values of $P_p$, $P_d$ are $P_p$=0.76, $P_d$=0.02 (correlations included) and $P_p$=0.60, $P_d$=0.02 (correlations not included). The center and right-hand figures compare the predicted birth rates given by fitting with the logistic curve (red/dark gray) and with the inclusion of nearest neighbor correlations (green/light gray). The center plot shows results from fitting density data only and the right-hand plot fitting with both density and correlation data. The bottom row shows results of fitting using both the density and nearest neighbor correlation data when death is nonzero. Details are as above, with the center plot showing the predicted birth rate and the right-hand plot the predicted death rate. The predicted values of $P_p$ and $P_d$ on the left-hand plot are $P_p$=0.64, $P_d$=0.33 (correlations included) and $P_p$=0.07, $P_d$=0.04 (correlations not included).

parameter using the logistic model, with accurate prediction of the parameter values even when $P_p \approx 1$. We also note that, without death, similar results are gained when fitting either solely to density data or to both density and nearest neighbor correlation data. With a nonzero death rate we see that the inclusion of correlations is vital for accurate prediction of the parameters (compare with the mean-field results). However, we note that the accuracy of estimation, even with correlation data included, decreases as $P_p$ and $P_d$ increase.

We note that one may be able to increase the accuracy of parameter estimation with careful consideration of the optimization function. For example, assuming $P_m$=1.0, one may estimate the birth rate using correlation data alone (as $P_d$ does not appear in the correlation equations). Then one may estimate the net birth rate ($r=P_m-P_d$) from density data, which then gives enough information to determine both $P_m$ and $P_d$. Other obvious methods include weighting the contributions from density and correlation data, dependent on whether one data source is more accurately measurable.

In summary, we conclude that including the effects of correlations can assist in determining the relative rates of movement, proliferation and death from experimental data. When death is negligible fitting to density-only data gives excellent results, however, when death is significant the data must be fitted using both density and correlation information.

### C. Outlook

The results presented here provide a simple, computationally efficient method for taking into account the effects of spatial correlations when considering populations undergoing birth-death-movement processes. With the enormous increases in computing power over the last decade we are now easily in a position to include correlations into general population models as standard. Moreover, our results show that the inclusion of correlations is vital in the recovery/estimation of biological parameters.

## ACKNOWLEDGMENTS

R.E.B. would like to thank the Research Councils U.K. and St. Hugh's College, Oxford for support.

[1] S. F. Gilbert, *Developmental Biology*, 8th ed. (Sinauer Associates Inc., Sunderland, MA, 2006).
[2] Y. Barrandon and H. Green, Cell **50**, 1131 (1987).
[3] L. Bao, M. Loda, P. A. Janmey, R. Stewart, B. Anand-Apte, and B. R. Zetter, Nat. Med. **2**, 1322 (1996).
[4] J. Liliental, S. Y. Moon, R. Lesche, R. Mamillapalli, D. Li, Y. Zheng, H. Sun, and H. Wu, Curr. Biol. **10**, 401 (2000).
[5] B. J. Lao and D. T. Kamei, J. Theor. Biol. **250**, 642 (2008).
[6] M. Eden, in *A Two-Dimensional Growth Process*, edited by J. Neyman, Proceedings of the Fourth Berkeley Sympomsium on Mathematical Statististics and Probability (University of California Press, Berkeley, CA, 1961), Vol. 4, pp. 223–239.
[7] Y. Lee, S. Kouvroukoglou, L. V. McIntire, and K. Zygourakis, Biophys. J. **69**, 1284 (1995).
[8] G. Cheng, B. B. Youssef, P. Markenscoff, and K. Zygourakis, Biophys. J. **90**, 713 (2006).
[9] E. Hernández-García and C. López, Phys. Rev. E **70**, 016216 (2004).
[10] T. Callaghan, E. Khain, L. M. Sander, and R. M. Ziff, J. Stat. Phys. **122**, 909 (2006).
[11] M. J. Simpson, A. Merrifield, K. A. Landman, and B. D. Hughes, Phys. Rev. E **76**, 021918 (2007).
[12] M. J. Simpson, K. A. Landman, and B. D. Hughes, Physica A **388**, 399 (2009).
[13] M. J. Simpson, K. A. Landman, B. D. Hughes, and A. E. Fernando, Physica A **389**, 1412 (2010).
[14] M. J. Simpson, K. A. Landman, and B. D. Hughes, Physica A **389**, 3779 (2010).
[15] C. Deroulers, M. Aubert, M. Badoual, and B. Grammaticos, Phys. Rev. E **79**, 031917 (2009).
[16] E. A. Codling, M. J. Plank, and S. Benhamou, J. R. Soc., Interface **5**, 813 (2008).
[17] J. D. Murray, *Mathematical Biology I: An Introduction*, 3rd ed. (Springer-Verlag, Berlin, 2003), Vol. I.
[18] A. Tremel, A. Cai, N. Tirtaatmadja, B. D. Hughes, G. W. Stevens, K. A. Landman, and A. J. O'Connor, Chem. Eng. Sci. **64**, 247 (2009).
[19] P. K. Denman, D. L. S. McElwain, D. G. Harkin, and Z. Upton, Bull. Math. Biol. **69**, 157 (2007).
[20] S. Gianino, J. R. Grider, J. Cresswell, H. Enomoto, and R. O. Heuckeroth, Development **130**, 2187 (2003).
[21] H. M. Young, A. J. Bergner, R. B. Anderson, H. Enomoto, J. Milbrandt, D. F. Newgreen, and P. M. Whitington, Dev. Biol. **270**, 455 (2004).
[22] A. S. Haider, J. Grabarek, B. Eng, P. Pedraza, N. R. Ferreri, E. A. Balazs, and Z. Darzynkiewicz, Cytometry, Part A **53A**, 1 (2003).
[23] D. C. Walker, J. Southgate, G. Hill, M. Holcombe, D. R. Hose, S. M. Wood, S. Mac Neil, and R. H. Smallwood, BioSystems **76**, 89 (2004).
[24] J. Mai, V. N. Kuzovkov, and W. von Niessen, J. Chem. Phys. **98**, 10017 (1993).
[25] J. Mai, V. N. Kuzovkov, and W. von Niessen, Physica A **203**, 298 (1994).
[26] A. Singer, J. Chem. Phys. **121**, 3657 (2004).
[27] U. Dieckmann and R. Law, Technical Report No. ir99040, 1999.
[28] J. A. N. Filipe and M. M. Maule, Math. Biosci. **183**, 15 (2003).
[29] D. J. Murrell, U. Dieckmann, and R. Law, J. Theor. Biol. **229**, 421 (2004).
[30] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[31] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevE.82.041905 for details of the correlation equations.
[32] G. H. A. Cole and A. Moreton, Mol. Phys. **13**, 501 (1967).
[33] G. H. A. Cole, Rep. Prog. Phys. **31**, 419 (1968).
[34] N. N. Bugaenko, A. N. Gorban, and I. V. Karlin, Theor. Math. Phys. **88**, 977 (1991).