Distinguishing between mean-field, moment dynamics and stochastic descriptions of birth–death–movement processes

Matthew J. Simpson\textsuperscript{a,b,}\textasteriskcentered, Jesse A. Sharp\textsuperscript{a}, Ruth E. Baker\textsuperscript{c}

\textsuperscript{a} Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001, Australia
\textsuperscript{b} Tissue Repair and Regeneration Program, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
\textsuperscript{c} Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK

HIGHLIGHTS

- Discrete model of birth–death–movement processes is analyzed.
- Mean-field and moment dynamics descriptions are derived.
- The mean-field and moment dynamics descriptions are not always valid.
- Standard calibration procedures are misleading.
- Propose an indirect measure of patchiness to indicate when each description is valid.

ABSTRACT

Mathematical descriptions of birth–death–movement processes are often calibrated to measurements from cell biology experiments to quantify tissue growth rates. Here we describe and analyze a discrete model of a birth–death–movement process applied to a typical two-dimensional cell biology experiment. We present three different descriptions of the system: (i) a standard mean-field description which neglects correlation effects and clustering; (ii) a moment dynamics description which approximately incorporates correlation and clustering effects; and (iii) averaged data from repeated discrete simulations which directly incorporates correlation and clustering effects. Comparing these three descriptions indicates that the mean-field and moment dynamics approaches are valid only for certain parameter regimes, and that both these descriptions fail to make accurate predictions of the system for sufficiently fast birth and death rates where the effects of spatial correlations and clustering are sufficiently strong. Without any method to distinguish between the parameter regimes where these three descriptions are valid, it is possible that either the mean-field or moment dynamics model could be calibrated to experimental data under inappropriate conditions, leading to errors in parameter estimation. In this work we demonstrate that a simple measurement of agent clustering and correlation, based on coordination number data, provides an indirect measure of agent correlation and clustering effects, and can therefore be used to make a distinction between the validity of the different descriptions of the birth–death–movement process.

© 2013 Elsevier B.V. All rights reserved.
1. Introduction

Mathematical models describing birth–death–movement processes are often used to interpret cell biology experiments such as the growth to confluence experiments shown in Fig. 1(a) [1–3]. In these experiments a population of cells is initially distributed, approximately uniformly, at low density on a two-dimensional substrate. The individual cells undergo motility, birth and death events which leads to changes in the density of the cells on the substrate. Such experiments play an important role in informing our understanding of wound healing and tissue engineering [1–3]. Typical approaches to modeling this kind of experiment involve applying standard continuum descriptions, such as the logistic model, without necessarily examining the underlying assumptions [3,4].

In this work we consider a discrete, lattice-based model of a birth–death–movement process and apply this model to replicate a growth to confluence experiment [5]. We show that the standard mean-field description of this discrete model gives accurate predictions for relatively slow agent proliferation and death rates [6]. A more sophisticated description, based on a moment dynamics approach which accounts for the density of agents and density of pairs of agents [7–15], provides an accurate description of averaged data from the discrete model for moderate proliferation and death rates [16,17]. For rapid proliferation and death rates both the mean-field and moment dynamics descriptions fail to predict the averaged discrete behavior and we must rely on using repeated, computationally intensive and time consuming, discrete simulations.

Since we have several potential mathematical descriptions of the same birth–death–movement process, two of which become inaccurate for sufficiently large birth and death rates, it is relevant for us to develop an understanding of the different parameter regimes where each description is valid. Without such information it is conceivable that a particular model, such as the standard mean-field logistic model, could be calibrated to match experimental data without any explicit consideration of whether that description is appropriate [3,16]. Such an oversight could lead to incorrect parameter estimation as an inaccurate model is calibrated to the observed data [3,16]. We will demonstrate this problem explicitly in Section 5.

In summary, we present a through parameter investigation of a birth–death–movement model that is applied to replicate a growth to confluence experiment from the cell biology literature [1–3]. We show that both the standard mean-field description and a more sophisticated moment dynamics description of the system can fail to produce accurate predictions of the averaged discrete data depending on the movement, birth and death rates in the discrete model, and the degree to which the distribution of agents is spatially correlated. Using simulation data we show that it is possible to distinguish between the applicability of different descriptions of the system using a relatively straightforward estimation of the agent coordination number [18] which provides a measure of the degree to which the distribution of agents is spatially correlated.

2. Discrete model

We consider a discrete model of biological cell motility, proliferation and death processes which has been described previously [16]. In brief, the discrete model consists of a two-dimensional square lattice with lattice spacing ∆, in which each lattice compartment can be occupied by, at most, a single agent. Each agent has a transition rate $P_m$ per unit time describing the motion of agents to a nearest neighbor site, a proliferation rate $P_p$ per unit time describing the production of new agents and a death rate $P_d$ per unit time. We suppose that motility events are unbiased so that an agent at $(x, y)$ attempts to step to $(x ± ∆, y)$ or $(x, y ± ∆)$ such that each target site is chosen with equal probability. Proliferation events are also unbiased so that a proliferative agent at $(x, y)$ attempts to deposit a daughter agent at $(x ± ∆, y)$ or $(x, y ± ∆)$ with each target site chosen with equal probability. A dying agent at site $(x, y)$ is simply removed from the system. The model is an exclusion process [19] since potential motility and proliferation events can only take place if the target site is vacant [20]. To be consistent with the experimental images in Fig. 1(a), we consider the initial distribution of agents to be spatially uniform [3] and we denote the number of agents on the lattice at time $t$ by $Q(t)$. Discrete simulations of this process are performed using the Gillespie algorithm with periodic boundary conditions [16,21].

Typical snapshots of the discrete process are shown in Fig. 1(b)–(d) where each simulation was randomly initiated by occupying each site with probability 5%. Snapshots are presented at later times $T_1$ and $T_2$, with $0 ≤ T_1 ≤ T_2$, and we see that the growing population exhibits different spatial patterns depending on the parameters in the simulation. The simulation in Fig. 1(b) illustrates a situation where the distribution of agents remains uniformly distributed with time whereas the simulations in Fig. 1(c)–(d) illustrate significant pattern development that is associated with agent clustering. Both uniform and clustered population growth processes have been observed previously in the experimental literature [3] and it is relevant for us to develop mathematical descriptions of these processes in order to interpret such experimental observations.

3. Continuum models

We use $k$-point distribution functions [3], $ρ^{(k)} (k = 1, 2, 3, \ldots)$, to describe the evolution of the average properties of the discrete model. The $ρ^{(k)}$ functions are multivariate probability distribution functions describing the occupancy of $k$-tuplets of sites. We use $l, m$ and $n$ to denote various lattice sites, and $c_l \in \{0, A\}$ to be the lattice variable describing the state of site $l$. With $k = 1$ we have

$$ρ^{(1)}(l) = c_l, \quad ρ^{(1)}(0) = 1 − c_l,$$

where $c_l$ is the density of agents at site $l$. Since we consider a translationally invariant system, $c_l$ represents the density of agents at any site $l$ and we now drop the subscript for notational convenience. For our initial condition the distribution
Fig. 1. (a) Snapshots of a growth to confluence experiment showing an initial low density distribution of 3T3 fibroblast cells at \( t = 0 \). Snapshots are shown at later times, \( t_1 = 12 \) h and \( t_2 = 60 \) h \[2\], and we see that the density of cells increases with time as the cells undergo individual migration and proliferation events. The scale bar in each experimental image corresponds to 200 \( \mu \)m and individual trajectories of particular cells are superimposed on the experimental images. See Ref. \[2\] for a complete description of the experiments. All images in (a) are reproduced with the kind permission of Elsevier.

(b)-(d) Snapshots from the discrete model described in Section 2 on a \( 100 \times 100 \) lattice randomly populated at \( t = 0 \) so that each site is occupied with probability 5%. All simulation snapshots are shown at \( T_1 = 10 \) and \( T_2 = 20 \) where \( T \) is a nondimensional time variable, given by \( T = t (p_p - p_d) \). All simulations correspond to \( p_m = 1 \) and the three sets of results illustrate the effect of varying the birth and death rates. Snapshots in row (b) correspond to \( p_p = 0.01 \) and \( p_d/p_p = 0.7 \), snapshots in row (c) correspond to \( p_p = 0.2 \) and \( p_d/p_p = 0.7 \), while snapshots in row (d) correspond to \( p_p = 0.5 \) and \( p_d/p_p = 0.7 \).

Function with \( k = 2 \) depends only on the distance between two lattice sites, \( r = |l - m| \geq \Delta \). We make use of a correlation function \[7,8\], written as

\[
F(|l - m|) = \frac{\rho^{(2)}(A_l, A_m)}{\rho^{(1)}(A_l)\rho^{(1)}(A_m)},
\]

to measure the correlation between occupancies of different sites since setting \( F(|l - m|) \equiv 1 \) implies that the occupancy of sites \( l \) and \( m \) are independent whereas if \( F(|l - m|) \neq 1 \) the occupancy status of sites \( l \) and \( m \) are correlated. The occupancy status of distant sites will be essentially uncorrelated so that we have \( F(|l - m|) \to 1 \) as \( |l - m| \to \infty \) \[17\].

We now develop expressions for the time rate of change of \( \rho^{(1)} \) and \( \rho^{(2)} \). For \( \rho^{(1)} \) we have

\[
\frac{d\rho^{(1)}(A_l)}{dt} = p_m \sum_n \alpha_{n,l} \left[ \rho^{(2)}(0_l, A_n) - \rho^{(2)}(A_l, 0_n) \right] + p_p \sum_n \alpha_{n,l} \left[ \rho^{(2)}(0_l, A_n) - p_d \rho^{(1)}(A_l) \right],
\]

where \( \alpha_{n,l} = 1 \) if \( l \) and \( n \) are nearest neighbor sites and \( \alpha_{n,l} = 0 \) otherwise. The summation in Eq. (3) is considered to be over \( n \), which represents all sites. Each term on the right of Eq. (3) corresponds to a potential motility, proliferation or death event that would alter the occupancy of site \( l \). Ref. \[16\] shows that Eq. (3) simplifies to

\[
\frac{dc}{dt} = P_p c (1 - F(\Delta)c) - P_d c.
\]
If we make the standard mean-field assumption that the occupancies of sites are independent \[6,22–25\] and \( F(\Delta) \equiv 1 \), Eq. (4) simplifies to,

\[
\frac{dc_{mf}}{dt} = P_p c_{mf}(1 - c_{mf}) - P_d c_{mf},
\]

where we have introduced the subscript notation \( c_{mf}(t) \) to emphasize that we have invoked the usual mean-field approximation. Eq. (5) can be re-written as a logistic model by introducing a nondimensional time variable, \( T = t(P_p - P_d) \), which gives

\[
\frac{dc_{mf}}{dT} = c_{mf}(1 - Kc_{mf}),
\]

where \( K = P_p/(P_p - P_d) \), which has the solution

\[
c_{mf}(T) = \frac{c_{mf}(0)}{Kc_{mf}(0)} \left[1 - e^{-T}\right] + e^{-T}.
\]

Invoking an independence assumption, such as we have described here, to arrive at a mean-field differential equation description of a particular discrete process is a standard approach to derive an approximate continuum description of a stochastic process \[6,22–25\]. Unfortunately, as we will show, the solution of such mean-field models does not always give an accurate prediction of the average behavior of the corresponding discrete system since the occupancies of lattice sites are not always independent \[22,25\].

Instead of invoking the mean-field assumption we can develop a moment dynamics description of the system by solving Eq. (4) without invoking the assumption that \( F(\Delta) \equiv 1 \). To do this we consider the time evolution of the 2-point distribution functions. Here, we have

\[
\frac{d\rho^{(2)}(A_l, A_m)}{dt} = P_m \sum_{n \neq m} \alpha_{n,m} \left[ \rho^{(3)}(A_l, A_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) \right] + P_p \sum_{n \neq m} \alpha_{n,m} \left[ \rho^{(3)}(0_l, A_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) \right] + P_p \alpha_{l,m} \left[ \rho^{(2)}(A_l, A_m) + \rho^{(2)}(0_l, A_m) \right] - 2P_d \rho^{(2)}(A_l, A_m).
\]

Each term on the right of Eq. (8) corresponds to a potential motility, proliferation or death event that would alter the probability of finding a pair of agents at sites \( l \) and \( m \). Eq. (8) can be combined with Eq. (2) and simplified to show how the correlation functions at each distance evolve in time, namely,

\[
\frac{dF}{dt}(|l - m|) = \frac{P_m}{2} \sum_{n \neq m} \alpha_{n,m} \left[ F(|l - n|) - F(|l - m|) \right] - 2P_p \left[1 - cF(\Delta)\right] F(|l - m|)
\]

\[
+ \frac{P_p}{2c} \alpha_{l,m} \left[1 - cF(\Delta)\right] + \frac{P_p}{2c^2} \sum_{n \neq m} \alpha_{n,m} \rho^{(3)}(A_l, A_m, A_n).
\]

To close Eq. (9) we need to express the remaining \( \rho^{(3)} \) terms as \( \rho^{(2)} \) terms and we do this by using the Kirkwood superposition approximation (ksa) \[26–28\]:

\[
\rho^{(3)}(A_l, A_m, A_n) = \frac{\rho^{(2)}(A_l, A_m) \rho^{(2)}(A_l, A_n) \rho^{(2)}(A_m, A_n)}{\rho^{(1)}(A_l) \rho^{(1)}(A_m) \rho^{(1)}(A_n)}.
\]

The ksa allows us to write

\[
\frac{dF}{dt}(|l - m|) = \frac{P_m}{2} \tilde{F}(|l - m|) - 2P_p \left[1 - cF(\Delta)\right] F(|l - m|) + \frac{P_p}{2c} \alpha_{l,m} \left[1 - cF(\Delta)\right] \left[ \tilde{F}(|l - m|) + \sum_{n \neq m} \alpha_{n,m} F(|l - m|) \right],
\]

where \( \tilde{F}(|l - m|) \) is the discrete lattice Laplacian

\[
\tilde{F}(|l - m|) = \sum_{n \neq m} \alpha_{n,m} \left[ F(|l - n|) - F(|l - m|) \right].
\]
others in the literature. In particular, we refer the reader to Table 1 in Murrell, Dieckmann and Law [13] for a comprehensive review. In this study the authors show that the ksa closure has several desirable properties such as certain symmetry and positivity conditions, and that it recovers the correct behavior at large distances [13]. Although it is well-known that other closure relationships have been derived and implemented elsewhere [9,10,12,13,28], here we choose to focus on the ksa since this is perhaps the most commonly used closure approximation. Furthermore, as we reported previously in Ref. [16], we have compared the performance of the ksa and other closure approximations, such as the power-1 and power-2 closures, elsewhere. In that previous study we found that the ksa gave the best results for this discrete model motivated by cell biology.

To make predictions using the moment dynamics model we solve Eq. (4) for \( c(t) \) and Eq. (11) for \( F(r) \) simultaneously [16,29]. Since we are interested in comparing the predictions of the moment dynamics model with other descriptions we will refer to the solution of Eq. (4) with Eq. (11) as \( \text{c}_{\text{mf}}(t) \). To obtain this solution we consider a truncated system by solving Eq. (11) for \( F(r) \) considering all appropriate increasing lattice distances \( \Delta \leq r \leq r_{\text{max}} \) with \( F(r) = 1 \) for \( r > r_{\text{max}} \). Since we are working on a two-dimensional square lattice we consider the appropriate lattice distances \( r = \Delta, \sqrt{2}\Delta, 2\Delta, \sqrt{5}\Delta, \sqrt{8}\Delta, 3\Delta, \ldots \). We solve this system of nonlinear ordinary differential equations using a fourth order Runge–Kutta method with constant time steps of duration \( \delta t \) [31]. For all results reported here we have \( r_{\text{max}} = 5\Delta \) and we tested that this truncation was appropriate by resolving all problems with \( r_{\text{max}} = 10\Delta \) to check that the results were independent of the truncation (not shown).

Although we will demonstrate, in Section 4, that the moment dynamics model is readily solved numerically, one of the difficulties of the moment dynamics model is that the governing system of nonlinear ordinary differential equations is not straightforward to interpret and it is difficult to obtain any analytical insight into the relationships between the dependent variables \( c, F(\Delta) \), \( F(\sqrt{2}\Delta) \), \( F(2\Delta) \), \( F(\sqrt{5}\Delta) \), \( \ldots \) and the parameters \( P_m, P_d \) and \( P_t \). Very recently we have made some progress in this area by showing that the moment dynamics model can be re-written in as an ordinary differential equation for \( c \) that is coupled to an ordinary differential equation for \( F(\Delta) \) and a nonlinear partial differential equation for \( F(r) \), \( r > \Delta \) [17]. Written in this way it is possible for us to obtain some insight into the relationships between the parameters and the dependent variables as discussed in Ref. [17]. However, since we are dealing with a coupled system of nonlinear differential equations there appears to be no opportunity to solve the model analytically.

4. Comparing average discrete simulation data, mean-field and moment dynamics descriptions

To generate averaged simulation data we consider simulations using the same lattice and initial condition reported in Fig. 1. If the number of agents present at time \( t \) in the \( m \)th identically prepared realization is \( Q_m(t) \), we estimate the average occupancy as

\[
\langle c_s(t) \rangle = \frac{1}{ML^2} \sum_{s=1}^{M} Q_s(t)
\]

(13)

where \( L \) is the size of the lattice, \( M \) is the total number of identically prepared realizations and the subscript \( s \) is used to emphasize that \( \langle c_s(t) \rangle \) corresponds to the discrete simulations. Results showing the evolution of \( \langle c_s(T) \rangle \) are given in Fig. 2, where \( T = t(P_p - P_d) \). These averaged density profiles correspond to \( P_m = 1 \), \( M = 50 \) and various \( P_p \) and \( P_d \). Results in Fig. 2(a)-(c) show the evolution of the system for \( P_d/P_p = 0.5, 0.6 \) and \( 0.7 \), respectively. For each value of \( P_d/P_p \), we consider six different proliferation rates \( P_p = 0.01, 0.1, 0.2, 0.3, 0.4 \) and \( 0.5 \). These parameter values are chosen so that we have \( P_p > 0 \) and \( P_d > P_p \), which ensures that we never observe the population becoming extinct on the timescale of the discrete simulations, which is consistent with observations in cell biology experiments [1–3]. The \( \langle c_s(T) \rangle \) profiles in Fig. 2 indicates that the density of agents increases as we reach an approximately steady state, and we note that in each case considered the temporal change in \( \langle c_s(T) \rangle \) appears to be negligible after \( T \approx 20 \).

We would like to point out that, upon first glance, the averaged density profiles in Fig. 2 could be interpreted to suggest that larger proliferation rates lead to a slower increase in the observed averaged agent density profile. We would like to point out that this impression is caused by our re-scaling of time, \( T = t(P_p - P_d) \), and we note that when the results are plotted as a function of the dimensional variable, \( t \), we clearly see that larger proliferation rates lead to a faster increase in the observed averaged agent density profile as expected. Unfortunately, plotting the results as a function of dimensional time, \( t \), means that it is difficult to compare results for different choices of parameters since the timescale required to reach steady state depends on the parameters. Since we are interested in considering a wide range of parameter values, working with the dimensional variable \( t \) is inconvenient and we prefer to present our results in terms of nondimensional time, \( T \).

The averaged discrete data in Fig. 2 are compared with Eq. (7), which is the standard mean-field representation of this process where the occupancies of lattice sites are assumed to be independent. Comparing solutions of the mean-field model and the averaged discrete data indicates that the mean-field model accurately predicts the evolution of the system for sufficiently small proliferation and death rates since we see that the mean-field prediction and the averaged discrete data are indistinguishable in the first column of Fig. 2 where the proliferation and death rates are sufficiently small relative to \( P_p \). As \( P_p \) increases, for each \( P_d/P_p \) ratio considered, the deviation between the averaged discrete data and Eq. (7) increases. For \( P_p = 0.5 \) and \( P_d/P_p = 0.5, 0.6 \) and \( 0.7 \), the mean-field model over predicts the agent density by \( 11\%, 26\% \) and \( 97\% \) at \( T = 20 \), respectively. Such large deviations between the mean-field description and the averaged simulation data has motivated the development of more sophisticated descriptions that approximately relax the independence assumptions.
A suite of averaged discrete results, \(c(T)\) (blue), are compared with the solution of the mean-field model, \(c_{mf}(T)\) (red), and the solution of the moment dynamics model, \(c_{md}(T)\) (green). Results in rows (a)-(c) correspond to \(P_p/P_d = 0.5, 0.6\) and \(0.7\), respectively. Results in each row are given for different birth and death rates with \(P_p = 0.01, 0.1, 0.2, 0.3, 0.4\) and \(0.5\), as indicated. All results correspond to \(P_0 = 1\) and are plotted as a function of nondimensional time, \(T = t(P_p - P_d)\). Averaged discrete data (blue) correspond to simulations on a \(100 \times 100\) lattice with periodic boundary conditions. Each discrete simulation was initiated with \(5\%\) of lattice sites randomly occupied and averaged data was obtained by generating \(M = 50\) identically prepared realizations of the same stochastic process. The solution of the mean-field model is given by Eq. (7) whereas the solution of the moment dynamics model corresponds to a numerical solution of Eq. (4) and Eq. (11). The numerical solution corresponds to \(st = 0.1\). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Averaged discrete data in Fig. 2 are also compared with the solution of the moment dynamics model, Eqs. (4) and (11). Comparing the results in Fig. 2 indicates that the moment dynamics description is indistinguishable from the averaged discrete data and the solution of the mean-field model provided that \(P_p\) and \(P_d\) are sufficiently small relative to \(P_m\). This confirms that the moment dynamics model and the mean-field model can produce equivalent results for the special case where \(F(r) \equiv 1\) and the occupancies of lattice sites are independent. For larger proliferation and death rates we observe a transition where the mean-field model fails to match the averaged discrete data whereas the moment dynamics model provides a good match. This is particularly clear in the second and third columns of Fig. 2 where \(P_p = 0.1\) and \(P_p = 0.2\), respectively. For this parameter regime the mean-field model systematically overpredicts the averaged discrete data whereas the moment dynamics model provides an excellent match to the averaged simulation data. As the rates of proliferation and death increase further we observe a consistent trend where the moment dynamics model always outperforms the mean-field model. However, for sufficiently large proliferation and death rates we also observe a significant deviation between the averaged discrete data and the solution of the moment dynamics model. For \(P_p = 0.5\) and \(P_d/P_p = 0.5, 0.6\) and \(0.7\), we see that the moment dynamics model over predicts the agent density by \(1\%\), \(3\%\) and \(20\%\) at \(T = 20\), respectively. The deviation between the averaged discrete data and the moment dynamics model can be attributed to the failure of the ksa closure scheme since it is known that Eq. (10) is an approximation that can fail to hold under certain circumstances [10–12,28].

5. Parameter fitting to density data

In this section we would like to highlight the problems that can arise when calibrating a mathematical model of a birth–death–movement processes to averaged density data. It is important to be aware of these issues since standard approaches to interpreting these kinds of cell biology experiments can lead to errors. For example, Fig. 1 in Tremel et al. [2] shows a cell density curve, similar to the averaged discrete data in Fig. 2. To describe these experimental observations, Tremel et al. [2] calibrated a mean-field logistic model to estimate the proliferation rate. Here we aim to show that this standard kind of calibration procedure should be performed with care.

Results in Fig. 3 show averaged density data from the discrete model for three particular choices of parameters in Fig. 2 namely: (i) \(P_p = 0.100, P_d = 0.050\); (ii) \(P_p = 0.300, P_d = 0.180\); and (iii) \(P_p = 0.500, P_d = 0.350\). For each of these three data sets we naively calibrate both the mean-field and the moment dynamics model to the observed density data using the Levenberg–Marquardt [32,33] nonlinear least-squares algorithm to find a least-squares estimate of \(P_p\) and \(P_d\) that best matches the density data. Results in Fig. 3(a)–(c) show the solution of the mean-field model with the least-squares estimates of \(P_p\) and \(P_d\). Similarly, results in Fig. 3(d)–(f) show the solution of the moment dynamics model with the least-squares estimates of \(P_p\) and \(P_d\). For all three sets of parameters we see that both the mean-field and the moment dynamics descriptions can be calibrated to produce an excellent match to the observed density data. Unfortunately, the apparently excellent match between the data and the models lead to least-squares parameter estimates which are incorrect.
The data in Table 1 summarizes the least-squares estimates of $P_p$ and $P_d$ for the density data in Fig. 3. Here we see two main results. Firstly, we observe that the moment dynamics model leads to parameter estimates that are, in all cases considered, closer to the real parameter values than the corresponding least-squares estimates obtained for the mean-field model. Secondly, the least-squares estimates of the model parameters are, in all cases considered, quite different from the real parameters in the discrete model. Again, we would like to emphasize that the standard procedure of fitting models of birth–death–movement processes to observed density data, as demonstrated here, can lead to major errors in parameter estimation [2,3]. To help alleviate this difficulty we now consider an additional measure that could be made using snapshots from discrete simulations or by analyzing images from an experiment [34] to provide an insight into how we can sensibly distinguish between the validity of mean-field and moment dynamics descriptions of such a birth–death–movement process.

### 6. Measuring the degree of population patchiness and clustering

To provide a simple but insightful measure of the agent correlation and clustering observed in snapshots from the discrete model we propose to use a measure of the agent coordination number [18]. We define the agent coordination number, $K$, using a Moore neighborhood, to be the total number of the eight closest sites that are occupied, giving $0 \leq K \leq 8$. One of the key attractions of using $K$ data is that for an uncorrelated uniformly occupied lattice, at density $C \in [0, 1]$, $K$ is binomially distributed:

$$P(K = k) = \binom{8}{k} C^k (1 - C)^{8-k},$$

with mean $\mu = 8C$ and variance $\sigma^2 = 8C (1 - C)$. This simple mathematical expression for the expected distribution of $K$ means that it is straightforward to compare simulation data, or appropriately discretized images from an experiment [30], with the theoretical result. We note that $K$ can be defined for other neighborhoods on the square lattice, such as the von Neumann neighborhood [18], for which the appropriate analog of Eq. (14) is $P(K = k) = \binom{4}{k} C^k (1 - C)^{4-k}$, for $0 \leq K \leq 4$.

To demonstrate the utility of the coordination number data we took the discrete data used to produce the average density data in Fig. 2 and we measured the distribution of $K$ at $T = 20$. The measurements were made using just $M = 10$ identically prepared realizations, and the mean values are plotted in Fig. 4, where the error bars correspond to one standard deviation from the mean. We also superimpose the shape of the theoretical distribution, given by Eq. (14), using $C = \langle c_s(T) \rangle$ at $T = 20$, onto these distributions. The data in Fig. 4 indicates that estimates of $K$ reveal insightful information that could be used to make a distinction between the validity of the mathematical models. Results in the first column, for sufficiently small proliferation and death rates, indicate that the observed distribution of $K$ is very well approximated by Eq. (14) confirming that the distribution of agents in the discrete simulations is very close to a random distribution without any spatial correlations. This observation is consistent with the results in Fig. 2 where we saw that these parameter combinations correspond to a regime where the traditional mean-field model accurately captures the average behavior of the system.

Results in the second through to the fifth columns of Fig. 4 illustrate that the distribution of $K$ is no longer accurately described by Eq. (14). In particular we see that the observed distribution of agents with high coordination number, $K \geq 5$, is greater than that predicted by Eq. (14) whereas the observed distribution of agents with low coordination number, $K \leq 3$, is less than that predicted by Eq. (14). These trends are consistent with the snapshots in Fig. 1 where we observe that the distribution of agents becomes increasingly patchy and clustered as the proliferation and death rates increase. This patchiness is reflected in the distribution of $K$. The trend within each column of Fig. 4 indicates that the match between Eq. (14) and the observed distribution of $K$ decreases as the $P_d/P_p$ ratio increases. This is also consistent with the trends in Fig. 2 where we observed that as the match between the averaged discrete data and the solution of the mean-field model became increasingly poor the $P_d/P_p$ ratio increased.

A visual comparison of the data in Fig. 4 indicates that as we further increase the proliferation and/or death rate, the deviation between the observed distribution of $K$ and Eq. (14) becomes increasingly large. This trend suggests that a measurement of this deviation could be used to indicate a transition where the moment dynamics model fails to predict the

### Table 1

<table>
<thead>
<tr>
<th>Actual $P_p$</th>
<th>Actual $P_d$</th>
<th>Mean-field Least squares $P_p$ (Relative error %)</th>
<th>Mean-field Least squares $P_d$ (Relative error %)</th>
<th>Moment dynamics Least squares $P_p$ md (Relative error %)</th>
<th>Moment dynamics Least squares $P_d$ md (Relative error %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.05</td>
<td>0.084 (16%)</td>
<td>0.044 (12%)</td>
<td>0.096 (4%)</td>
<td>0.048 (4%)</td>
</tr>
<tr>
<td>0.30</td>
<td>0.18</td>
<td>0.197 (34%)</td>
<td>0.130 (28%)</td>
<td>0.269 (10%)</td>
<td>0.164 (9%)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.35</td>
<td>0.259 (48%)</td>
<td>0.219 (37%)</td>
<td>0.409 (18%)</td>
<td>0.298 (15%)</td>
</tr>
</tbody>
</table>
Real parameters $P_p = 0.100, P_d = 0.050$

Least squares $P_p = 0.084, P_d = 0.044$

Least squares $P_p = 0.096, P_d = 0.045$

Least squares $P_p = 0.197, P_d = 0.130$

Least squares $P_p = 0.259, P_d = 0.219$

Least squares $P_p = 0.300, P_d = 0.180$

Least squares $P_p = 0.269, P_d = 0.164$

Least squares $P_p = 0.409, P_d = 0.230$

System, such as the results in Fig. 2, for $P_p \geq 0.2$ where we begin to see significant differences between the moment dynamics model and the averaged discrete simulation data.

In summary, we observe that the deviation between the expected distribution for $K$ when the population is distributed uniformly, Eq. (14), and the observed distribution of $K$ from a snapshot produced by this kind of simulation is a convenient way to obtain an estimate of the degree of patchiness and clustering in the population. This estimate can therefore provide us with a method to distinguish between the suitability of the three different models described here. We now quantify the
Fig. 5. Results in (a) illustrate how coordination number data from Fig. 4 deviate from the expected mean-field distribution, Eq. (14). Three plots of $E_K$, given by Eq. (15) in (a) are for $P_d/P_p = 0.5, 0.6$ and $0.7$, (circles, diamonds and squares, respectively). Results in (b) illustrate how the mean-field density in Fig. 2, $c_{mf}(20)$, deviates from the simulated density, $\langle c_s(20) \rangle$. Three plots of $E_c$, given by Eq. (16), in (b) are for $P_d/P_p = 0.5, 0.6$ and $0.7$, (circles, diamonds and squares, respectively). Results in (c) illustrate how the moment dynamics density in Fig. 4, $c_{md}(20)$, deviates from the simulated density, $\langle c_s(20) \rangle$. Three plots of $E_c$, given by Eq. (16), in (c) are for $P_d/P_p = 0.5, 0.6$ and $0.7$, (circles, diamonds and squares, respectively). The solid horizontal lines indicate potential threshold values for the mean-field model while the horizontal dashed lines indicate potential threshold values for the moment dynamics model.

deviation between the observed and theoretical distributions of $K$ data by considering

$$
E_K = \sum_{k=1}^{8} |\mathbb{P}(K = k) - \langle \mathbb{P}_s(K = k) \rangle|,
$$

where $\mathbb{P}(K = k)$ is given by Eq. (14) and $\langle \mathbb{P}_s(K = k) \rangle$ is an estimate of the probability that an agent observed in a simulation has coordination number $k$. Using the simulation data in Fig. 4 we plot $E_K$ as a function of $P_p$ for $P_d/P_p = 0.5, 0.6$ and $0.7$ in Fig. 5(a). The results in Fig. 5(a) confirm that $E_K$ increases with $P_p$, and with the ratio $P_d/P_p$.

Results in Fig. 5(a) are analogous to the qualitative observations we made about the data in Fig. 2: the deviation between the averaged simulation data and the predictions of the mean-field and moment dynamics models increases with $P_p$ and with the ratio $P_d/P_p$. To quantify the deviation between the averaged simulation data and the solutions of the mean-field and moment dynamics models we now consider the following quantity,

$$
E_c = |\langle c(T) \rangle - c(T)|,
$$

where $c(T)$ corresponds to the mean-field or moment dynamics density at time $T$. Results showing $E_c$ for the mean-field density at $T = 20$ from the averaged data in Fig. 2 are given in Fig. 5(b) confirm that we observe the same qualitative trend in the $E_c$ data as we have in the $E_K$ data. This indicates that if we can estimate $E_K$ from discrete snapshots of a birth–death–movement process, and we have an estimate of an acceptable threshold value of $E_K$, we could make an informed decision,
a priori, about the suitability of the traditional mean-field model for interpreting a particular observation of such a birth–death–movement process. For example, for a density threshold of $E_c \leq 0.02$, as indicated by the solid horizontal line in Fig. 5(b), the corresponding coordination number threshold, illustrated by the solid horizontal line in Fig. 5(a) indicates that snapshots with $E_c \leq 0.12$ could be modeled using the mean-field description.

Similarly, data in Fig. 5(c) showing $E_c$ for the moment dynamics density at $T = 20$ from the averaged data in Fig. 5 confirm that we have the same qualitative trend in $E_c$ for the moment dynamics model as we observe in the $E_c$ data. This data indicates that having an estimate of $E_c$ would allow us to make an informed decision about the suitability of the moment dynamics model for a particular application. As an example, for a density threshold of $E_c \leq 0.02$, as indicated by the dashed horizontal line in Fig. 5(c), the corresponding coordination number threshold, illustrated by the dashed horizontal line in Fig. 5(a), indicates that snapshots with $E_c \leq 0.48$ could be modeled using the moment dynamics description.

7. Discussion and conclusion

In this work we have presented a discrete birth–death–movement model that can be applied to experiments from the cell biology literature to describe the growth of a population of cells from an initially sparse, low density distribution. We generate averaged data from the discrete model that describes how the average density of the population increases with time to reach some steady value. Snapshots of the discrete process indicate that the growth process can either remain relatively uniform with time or can exhibit spatial structure, clustering and patchiness within the population. The formation of clustering depends on the birth and death rates relative to the motility rate. To model the averaged density of the discrete process we derive two continuum descriptions: the standard mean-field description which neglects spatial correlation effects and a moment dynamics description which approximately incorporates correlation effects. Comparing the averaged discrete data and the solutions of the two continuum models confirms that both descriptions match the averaged data for relatively slow birth and death rates. The standard mean-field model fails to match the averaged discrete data for moderate birth and death rates whereas the moment dynamics model provides an accurate description of the discrete process for moderate birth and death rates. However, both the standard mean-field and moment dynamics models fail to match the averaged data when the birth and death rates are sufficiently high. Given that standard applications of these models to experimental data are made without detailed knowledge of the birth and death rates [2] it is possible that either the mean-field or moment dynamics descriptions could be used in an inappropriate parameter regime leading to errors in parameter estimation [3]. To overcome this limitation additional data from the birth–death–movement process needs to be considered so that we can make a quantitative assessment to distinguish between the regimes where different mathematical descriptions are valid.

We have presented a straightforward method for assessing the degree of clustering and patchiness within the birth–death–movement process based on an estimate of $K$, the agent coordination number [18]. Using this approach we show that a measurement of the deviation between the expected mean-field distribution of $K$ and the observed distribution of $K$ can be made very simply, and that this deviation can be used to distinguish between different parameter regimes where the mean-field and moment dynamics models are either valid or invalid. Although there are many other candidate measures of the degree of clustering, or conversely the degree of mixing, such as those reviewed by Ref. [35], the coordination number is convenient since we have a very simple expression for the expected mean-field result. This means that we can assess the degree of clustering by comparing a simple mathematical expression with some observed quantity. Other options for quantifying the degree of clustering, such as measuring the distribution of cluster size [36], do not lend themselves such simple analytic expressions and could only be used by comparing two quantities that are derived from simulation data.

There are several options for extending the work described here with an obvious extension to three-dimensional birth–death–movement processes which are relevant to interpreting cell biology experiments performed in three-dimensional gels rather than those on two-dimensional substrates. While our work has focused on the ksa closure, we note that there are several other possible closure schemes and the development of new and improved closure approximations is an active area of research [28]. Therefore, another extension of this work would be to repeat our analysis using alternative closure schemes. A further extension is to consider applying the two-dimensional birth–death–movement process with inhomogeneous initial conditions to represent cell biology experiments which describe moving fronts of cells [29,30]. Developing techniques that can reliably distinguish between mean-field and moment dynamics descriptions of such spatially-variable processes is a challenge that we will address in the future.

Acknowledgments

This research is supported by the Australian Research Council Discovery Project DP120100551 and the 2011 International Exchange Scheme funded by the Royal Society. We acknowledge the Australian Mathematical Sciences Institute who provided a summer vacation scholarship to Mr. Sharp and we appreciated the review comments from Dr. Michael Raghb and one other anonymous referee.

Appendix. Simplification of the moment dynamics equations

In this appendix we outline some key results that allow us to simplify Eq. (8). The first step in simplifying this equation is to make use of conservation statements which allow us to simplify some of the $\rho^{(3)}$ terms. An example of this kind of
conservation statement can be written as
\[\rho(3)(A_l, A_m, A_n) + \rho(3)(0, A_m, A_n) = \rho(2)(A_m, A_n).\] (17)

Using these kinds of conservation statements allow us to re-write the \(\rho(3)\) terms proportional to \(P_m\) on the right of Eq. (8) using equivalent \(\rho(2)\) terms as we have outlined previously [16, 17, 25, 29]. Using these conservation statements allows us to simplify the \(\rho(3)\) terms proportional to \(P_m\) on the right of Eq. (8) into the first term on the right of Eq. (11).

Many of the \(\rho(2)\) terms in Eq. (8) can be simplified by making use of Eq. (2). For example, the term representing the rate of change of the \(\rho(2)\) term on the left of Eq. (8) can be re-written as
\[
\frac{d}{dt} \rho(2)(A_l, A_m) = \frac{d}{dt} \left( F(|l - m|) \rho(1)(A_l) \rho(1)(A_m) \right)
\]
\[
\begin{align*}
&= \frac{d}{dt} \left( F(|l - m|) c^2 \right) \\
&= c^2 \frac{dF(|l - m|)}{dt} + F(|l - m|) \frac{dc}{dt}.
\end{align*}
\] (18)

Substituting Eq. (4) into Eq. (18) gives us an expression for \(\frac{dF(|l - m|)}{dt}\) in terms of \(c\), and \(F(|l - m|)\) for \(|l - m| \geq \Delta\). When this expression is combined with the simplified form of Eq. (8) we arrive at Eq. (9). Following this procedure carefully we see that all terms proportional to \(P_m\) cancel leaving the final expression for \(\frac{d\rho(1)(A_l)}{dt}\) appearing to be independent of \(P_m\). We note, however, that the solution of Eq. (4) is explicitly dependent on \(P_m\) which means that the solution for \(F(|l - m|)\) is implicitly dependent on \(P_m\).

References