bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Extended logistic growth model for heterogeneous populations

Wang Jin[1], Scott W McCue[1], *Matthew J Simpson[1]

[1] *School of Mathematical Sciences, Queensland University of Technology (QUT) Brisbane, Queensland, Australia.*

Abstract

Cell proliferation is the most important cellular-level mechanism responsible for regulating cell population dynamics in living tissues. Modern experimental procedures show that the proliferation rates of individual cells can vary significantly within the same cell line. However, in the mathematical biology literature, cell proliferation is typically modelled using a classical logistic equation which neglects variations in the proliferation rate. In this work, we consider a discrete mathematical model of cell migration and cell proliferation, modulated by volume exclusion (crowding) effects, with variable rates of proliferation across the total population. We refer to this variability as *heterogeneity*. Constructing the continuum limit of the discrete model leads to a generalisation of the classical logistic growth model. Comparing numerical solutions of the model to averaged data from discrete simulations shows that the new model captures the key features of the discrete process. Applying the extended logistic model to simulate a proliferation assay using rates from recent experimental literature shows that neglecting the role of heterogeneity can, at times, lead to misleading results.

Key words: Cell proliferation, Heterogeneity, Population dynamics, Logistic growth.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Introduction

Cell proliferation is essential for regulating the dynamics of cell populations, and plays a vital role in collective cell spreading, cancer progression and tissue regeneration (Eladdadi and Isaacson, 2008; Evan and Vousden, 2001; Haridas et al., 2017; Pavlath et al., 1998). While it is clear that cells from different cell lines proliferate at different rates (Hayflick, 1965), recent experimental methods indicate that heterogeneity in cell proliferation arises even within the same cell line (Bajar et al., 2016; Guan et al., 2014; Sakaue-Sawano et al., 2008).

Many different types of experiments are used to quantify cell proliferation (An et al., 2001; Azzarone and Macieira-Coelho, 1982; Haass et al., 2014; Hayflick, 1965; Jin et al., 2017; Kaneoka et al., 1983; Willaime et al., 2013). The complexity of these experiments varies from simple *in vitro* proliferation assays in which the net expansion of a population of cells is observed and measured, such as the experiment shown in Figure 1, to more sophisticated experiments that use fluorescent cell cycle indicators to measure the duration of different phases of the cell cycle for individual cells (Haass et al., 2014; Sakaue-Sawano et al., 2008; Vittadello et al., 2018). A standard measure of cell proliferation is the doubling time, which is a measure of the duration of time required for a population of cells, at low density, to double (Hayflick, 1965; Jin et al., 2016). The doubling time quantifies cell proliferation from the perspective of the entire population, and any kind of variability amongst individual cells in the population is neglected. Modern experimental approaches, such as individual-level fluorescent cell cycle indicators and micro collagen gel arrays, allow us to quantify variations in the cell cycle of individual cells (Guan et al., 2014; Haass et al., 2014). This individual-level data shows that proliferation rates

* Corresponding author
Email address: `matthew.simpson@qut.edu.au`, *Telephone* `+ 61 7 31385241`, *Fax* `+ 61 7 3138 2310` (Wang Jin[1], Scott W McCue[1], *Matthew J Simpson[1]).

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of individual cells can vary significantly within the same cell line (Guan et al., 2014; Haass et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

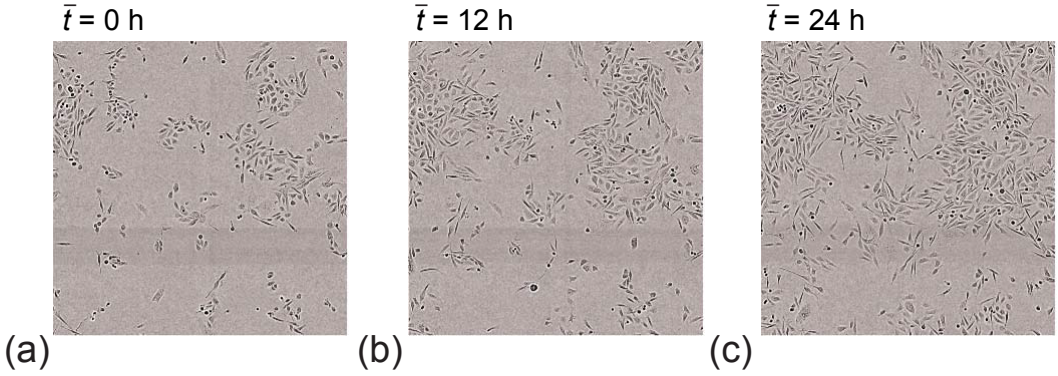

Fig. 1. ***In vitro* cell proliferation assay.** Population of PC-3 prostate cancer cells in a square field of view, of side length 1440 μm. Images correspond to (a) $\bar{t} = 0$ h, (b) 12 h, and (c) 24 h (Browning et al., 2018). Reproduced from Browning et al. (2018) with permission.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mathematical models are often used to mimic cell biology experiments, and to quantify rates of cell proliferation (Cai et al., 2007; Jin et al., 2017; Nardini et al., 2016). One approach is to apply an individual-level, agent-based model (Frascoli et al, 2014). In this kind of model, agents represent individual cells, and these agents migrate and proliferate according to certain rules thought to be relevant to the application of interest (Treloar et al., 2014). Although agent-based models offer the capability to investigate individual-level details, most of these models adopt a conventional assumption that the rate of proliferation of individual cells in the population is taken to be a constant. This assumption, however, may not be applicable to real situations where the proliferation rate of individual cells in the population varies significantly.

The most commonly-used continuum model of cell proliferation is the classical logistic growth model (Cai et al., 2007; Jin et al., 2016; Maini et al., 2004; Sengers et al., 2007; Sheardown and Cheng, 1996; Sherratt and Murray 1990; Vo et al., 2015; Warne et al. 2017). Although the classical logistic growth model is widely used to estimate the growth rate for populations of cells, there is an increasing awareness in the mathematical biology literature that cell populations do not always grow logistically (e.g. Gerlee, 2013; Powell et al., 2017; Sarapata and de Pillis 2014; Sewalt et al. 2016; West et al. 2001; Neufeld et al. 2017), and generalisations of the logistic growth model have been proposed (Jolicoeur and Pontier, 1989; Tsoularis and Wallace, 2002). Other types of models, where population growth is explicitly coupled to external factors, such as light availability (Pozzobon and Perré, 2018) and interactions with other populations (García-Algarra et al., 2014), have also been developed for specific biological applications. However, a limitation in each of these modelling frameworks is that the cell proliferation rate is treated as a constant, which amounts to neglecting heterogeneity.

In this work we consider a discrete modelling framework in which we deliberately introduce heterogeneity in the rates of migration and proliferation. The

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

continuum limit description of the discrete model leads to a complicated system of reaction-diffusion equations which simplifies to a generalisation of the classical logistic growth model when we apply the model to situations where there are, on average, no spatial variations in the agent density, such as the experimental image in Figure 1. We apply the extended logistic model to simulate a proliferation assay using a distribution of heterogeneous proliferation rates that are estimated from the cell biology literature.

We show that neglecting the role of heterogeneity can produce misleading results when population dynamics are interpreted in a standard way by simply calibrating the solution of the classical logistic growth model to match the data. Comparing the results from the logistic growth model and the extended model illustrates that the logistic growth model does not perform well in some cases. As we show, in these cases where the standard approach fails to capture the growth dynamics of heterogeneous populations, the new extended model performs very well. Unlike the classic logistic growth model, the extended model does not have an exact solution. With this in mind, we provide analytical insight into the role of heterogeneity by constructing approximate perturbation solutions in the limit of small variation in the proliferation rate.

Throughout this study we use a combination of dimensional and dimensionless parameters. Mathematical models, both discrete and continuum, are first presented using dimensional variables and dimensional parameters. All dimensional quantities are indicated with an overbar. Later, when we apply the mathematical models to experimental data, and when we present some analysis of the mathematical models, we always work with dimensionless variables and parameters for which the overbar notation is dropped.

2 Discrete model

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We use a conceptually straightforward exclusion process on a hexagonal lattice to simulate cell migration and cell proliferation (Jin et al., 2017). We apply this model to simulate *in vitro* cell proliferation assays, such as the experimental image in Figure 1. For this kind of proliferation assay, a standard way to estimate the proliferation rate is to count individual cells, and to use this information to construct time evolution of the average density profile (Jin et al., 2016). With this averaged density profile, we could calibrate the solution of the classical logistic equation to provide an estimate of the proliferation rate (Tremel et al. 2009). However, this standard approach neglects any variation in the proliferation rate. Therefore, here we use our model and take a different approach.

In our model, agents represent individual cells, and these agents are placed uniformly, at a specified initial density, on a hexagonal lattice. We use a lattice of size $I \times J$ lattice sites, with lattice spacing $\bar{\Delta}$. Here, $\bar{\Delta}$ can be thought of as a typical cell diameter, such as $\bar{\Delta} = 20\ \mu$m. To simulate crowding effects, each lattice site can be occupied by, at most, one agent. Each lattice site is indexed, $\mathbf{s} = (i, j)$, where $i, j \in \mathbb{Z}^+$, and each site is associated with a unique Cartesian coordinate. The total population of agents is composed of $N \geq 1$, potentially distinct, subpopulations. Agents in each subpopulation are characterised by a potentially distinct migration probability per time step, $P_m^{(n)} \in [0, 1]$ for $n = 1, 2, \ldots, N$. Furthermore, agents in each subpopulation are characterised by a potentially distinct proliferation probability per time step, $P_p^{(n)} \in [0, 1]$ for $n = 1, 2, \ldots, N$. The total number of agents at time t is $M(\bar{t})$.

Cell migration and proliferation are modelled using a random sequential random update method (Chowdhury et al., 2005). To advance the discrete model from time $\bar{t}$ to time $\bar{t} + \bar{\tau}$, $M(\bar{t})$ agents are selected at random, one at a time, with replacement. The selected agent attempts to move to one of the six nearest neighbour sites with probability $P_m^{(n)}$. The attempted migration event will be successful if the randomly chosen nearest neighbour target site is

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

vacant. After $M(\bar{t})$ potential migration events have been attempted, another $M(\bar{t})$ agents are selected at random, one at a time, with replacement. The selected agent will attempt to place a daughter agent on one of the six nearest neighbour sites with probability $P_p^{(n)}$. The attempted proliferation will only be successful if the randomly chosen nearest neighbour site is vacant. In the event that the potential proliferation event is successful, we make the simplest assumption that the daughter agent belongs to the same subpopulation as the mother agent. Once these potential motility and proliferation events have been attempted, we update $M(\bar{t} + \bar{\tau})$.

Using a random sequential update algorithm means that during each time step some agents may attempt to migrate or proliferate multiple times, whereas other agents may not attempt to migrate or proliferate at all. However, when we simulate over a large number of time steps, on average each agent in the population will attempt to migrate and proliferate once per time step. The random sequential update algorithm is conceptually straightforward, easy to implement, and is known to provide a good approximation to the dynamics of these kinds of simulations where we consider populations of motile and proliferative agents (Treloar et al., 2014).

3 Continuum limit description

The continuum limit description of the discrete model can be derived using standard averaging arguments and a mean field approximation. These kinds of arguments and approximations are widely invoked throughout the mathematical biology literature where continuum limit descriptions are derived from underlying stochastic models (e.g. Callaghan et al., 2006; Deroulers et al. 2009; Dyson and Baker, 2015; Plank and Simpson, 2012). The mean field assumption involves treating the occupancy status of lattice sites as being independent. While this assumption is questionable for any particular single realisation of

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the stochastic model, it turns out to be remarkably accurate when we consider an ensemble of stochastic simulations, as we will show later. Furthermore, this approximation is known to be accurate for a range of problems involving interactions between multiple subpopulations of agents (Simpson et al, 2009), and for other problems involving single populations of motile and proliferative agents (Simpson et al. 2010).

We denote the probability of finding an agent from subpopulation n at site $\mathbf{s} = (i, j)$ as $C_{\mathbf{s}}^{(n)} \in [0, 1]$, for $n = 1, 2, \ldots, N$. This probability can be thought of as corresponding to averaging the occupancy of site $\mathbf{s}$ over many identically prepared realisations of the stochastic model. Therefore, the probability of site $\mathbf{s}$ being vacant is $1 - \sum_{n=1}^{N} C_{\mathbf{s}}^{(n)}$. In the discrete model, migration events can act to either increase or decrease the occupancy of site $\mathbf{s}$, whereas proliferation events can only act to increase the occupancy of site $\mathbf{s}$. Accounting for these possibilities, the change in average occupancy at site $\mathbf{s}$ for agents from subpopulation n, from time $\bar{t}$ to time $\bar{t} + \bar{\tau}$ can be written as,

$$\begin{aligned}
\delta C_{\mathbf{s}}^{(n)} = & \overbrace{\frac{P_m^{(n)}}{6}\left(1 - \sum_{n=1}^{N} C_{\mathbf{s}}^{(n)}\right) \sum_{\mathbf{s}' \in \mathcal{N}\{\mathbf{s}\}} C_{\mathbf{s}'}^{(n)}}^{\text{increase in occupancy due to migration into site } \mathbf{s}} \\
& - \overbrace{\frac{P_m^{(n)}}{6} C_{\mathbf{s}}^{(n)} \sum_{\mathbf{s}' \in \mathcal{N}\{\mathbf{s}\}} \left(1 - \sum_{n=1}^{N} C_{\mathbf{s}'}^{(n)}\right)}^{\text{decrease in occupancy due to migration out of site } \mathbf{s}} \\
& + \overbrace{\frac{P_p^{(n)}}{6}\left(1 - \sum_{n=1}^{N} C_{\mathbf{s}}^{(n)}\right) \sum_{\mathbf{s}' \in \mathcal{N}\{\mathbf{s}\}} C_{\mathbf{s}'}^{(n)}}^{\text{increase in occupancy due to proliferation into site } \mathbf{s}} \quad , \qquad (1)
\end{aligned}$$

where $\mathcal{N}\{\mathbf{s}\}$ denotes the set of six nearest-neighbour sites around site $\mathbf{s}$. In Equation (1) we implicitly make the standard assumption that the average occupancy of each lattice site is independent. This is the mean field assumption. We expand each term in Equation (1) about site $\mathbf{s}$ using Taylor series, and neglect terms of $\mathcal{O}(\bar{\Delta}^3)$. This process leads to the cancellation of many terms. Here, we omit showing these intermediate steps as they have been outlined

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for similar models in our previous work (Jin et al., 2016; Simpson et al., 2009; Simpson et al., 2010). Dividing both sides of the resulting expression by $\bar{\tau}$, and taking the limit as $\bar{\Delta} \to 0$ and $\bar{\tau} \to 0$ jointly, with $\bar{\Delta}^2/\bar{\tau}$ held constant we obtain,

$$\frac{\partial C_n(\bar{x},\bar{y},\bar{t})}{\partial \bar{t}} = \overbrace{\left(\frac{P_m^{(n)}\bar{\Delta}^2}{4\bar{\tau}}\right)\nabla\cdot\Big[\left(1-S(\bar{x},\bar{y},\bar{t})\right)\nabla C_n(\bar{x},\bar{y},\bar{t}) + C_n(\bar{x},\bar{y},\bar{t})\nabla S(\bar{x},\bar{y},\bar{t})\Big]}^{\text{unbiased motility mechanism with exclusion}}$$
$$+ \overbrace{\left(\frac{P_p^{(n)}}{\bar{\tau}}\right) C_n(\bar{x},\bar{y},\bar{t})\Big[1-S(\bar{x},\bar{y},\bar{t})\Big]}^{\text{unbiased proliferation mechanism with exclusion}}, \tag{2}$$

for $n = 1, 2, \ldots, N$. $C_n(\bar{x},\bar{y},\bar{t})$ is the density of nth subpopulation, and $\bar{x}$ and $\bar{y}$ are the horizontal and vertical coordinates, respectively. The total population density is given by $S(\bar{x},\bar{y},\bar{t}) = \sum_{n=1}^{N} C_n(\bar{x},\bar{y},\bar{t})$.

In cell proliferation experiments, cells are placed uniformly on a two-dimensional substrate (Browning et al. 2018; Jin et al., 2017). Therefore, this kind of initialisation means that there are, on average, no spatial gradients in cell density provided we view the experiment at a sufficiently large spatial scale, such as in Figure 1. Under these conditions Equation (2) simplifies to

$$\frac{\mathrm{d}C_n(\bar{t})}{\mathrm{d}\bar{t}} = \bar{r}_n C_n(\bar{t})\Big[1 - S(\bar{t})\Big], \tag{3}$$

where $\bar{r}_n = P_p^{(n)}/\bar{\tau}$, for $n = 1, 2, 3, \ldots, N$. The total cell density $S(\bar{t})$ then can be obtained by summing over the governing equations of N subpopulations to give,

$$\frac{\mathrm{d}S(\bar{t})}{\mathrm{d}\bar{t}} = \left(\sum_{n=1}^{N} \bar{r}_n C_n(\bar{t})\right)\Big[1 - S(\bar{t})\Big]. \tag{4}$$

In this work we always deal with initial conditions without any spatial gradients, which corresponds to the experimental images shown in Figure 1. This implies that we are working with a system of ordinary differential equations

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(ODEs) instead of a system of partial differential equations. If we were to consider a different initial condition, such as a scratch assay or a barrier assay where there experiments are intentionally initialised with some spatial gradients present, then we would have to work with Equation (2) instead of Equation (3).

Without loss of generality, when we apply Equation (3) we adopt the convention that $\bar{r}_1 \geq \bar{r}_2 \geq \bar{r}_3 \geq \ldots \geq \bar{r}_N$, so that $\bar{r}_1$ is the proliferation rate of the fastest-proliferating subpopulation, $\bar{r}_2$ is the proliferation rate of the second fastest-proliferating subpopulation, and so on. We note that in the special case where we consider all the proliferation rates to be equal, $\bar{r}_1 = \bar{r}_2 = \ldots = \bar{r}_N$, we are dealing with a homogeneous population with a constant proliferation rate, $\bar{\lambda}$. The continuum limit description simplifies to,

$$\frac{\mathrm{d}C(\bar{t})}{\mathrm{d}\bar{t}} = \bar{\lambda}C(\bar{t})\Big[1 - C(\bar{t})\Big], \tag{5}$$

which is the classical logistic growth model (Murray, 2002), whose solution is given by,

$$C(\bar{t}) = \frac{C(0)}{\Big[1 - C(0)\Big]\mathrm{e}^{-\bar{\lambda}\bar{t}} + C(0)}. \tag{6}$$

This exact solution is a sigmoid curve that monotonically increases from $C(0)$, and approaches unity as $\bar{t} \to \infty$, provided that $C(0) < 1$. Since our system of ODEs, given by Equation (3), simplifies to the classical logistic model when all the proliferation rates are identical, we refer to Equation (3) as the *extended logistic growth model*.

To simplify our work we nondimensionalise time using the fastest proliferation rate, $t = \bar{r}_1\bar{t}$. Therefore Equation (3) becomes

$$\begin{aligned} \frac{\mathrm{d}C_1(t)}{\mathrm{d}t} &= C_1(t)\Big[1 - S(t)\Big], \\ \frac{\mathrm{d}C_n(t)}{\mathrm{d}t} &= r_n C_n(t)\Big[1 - S(t)\Big], \end{aligned} \tag{7}$$

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $r_n = \bar{r}_n/r_1$, for $n = 2, 3, \ldots, N$. Therefore, we now have a system of ODEs with non-dimensional proliferation rates of: unity, $r_2, r_3, \ldots, r_N$, with $1 \geq r_2 \geq r_3 \geq \ldots \geq r_N$. This non-dimensionalisation allows us to compare the solutions of the model for different systems that are characterised by very different proliferation rates. In the non-dimensional format, Equation (4) becomes

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = \left(C_1(t) + \sum_{n=2}^{N} r_n C_n(t) \right) \Big[1 - S(t) \Big]. \tag{8}$$

To be consistent, if we non-dimensionalise Equation (6) with $t = \bar{\lambda}\bar{t}$, we obtain

$$C(t) = \frac{C(0)}{\Big[1 - C(0)\Big]\mathrm{e}^{-t} + C(0)}. \tag{9}$$

Unlike the classical logistic model, Equation (7) does not have an exact solution. Therefore, we present numerical solutions that are obtained using a backward Euler approximation. In all cases we use a constant time step of $\delta t = 0.01$, and Picard iteration with convergence tolerance $\epsilon = 1 \times 10^{-5}$. These choices of δt and ϵ are sufficient to produce grid-independent numerical solutions of the model.

4 Results

4.1 Continuum-discrete match

All discrete results are presented in a non-dimensional format, on a lattice with unit lattice spacing and with time steps of unit duration, $\Delta = \tau = 1$. Note that Δ and τ can be re-scaled to correspond to any particular choices of dimensional $\bar{\Delta}$ and $\bar{\tau}$. This means that we can re-scale any of these dimensionless simulations to correspond to a population of cells with arbitrary cell

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

diameter, and arbitrary characteristic proliferation rate. Since we are focusing on the role of heterogeneity in cell proliferation, in all simulations we set $\Delta = P_m^{(n)} = 1$, for $n = 1, 2, \ldots, N$, and $\tau = P_p^{(1)} = 0.01$. In addition, we choose $I = 101$ and $J = 117$, so that the size of the simulation domain is 100×100. Periodic boundary conditions are applied to all simulations.

To explore the role of heterogeneity in population growth, we first consider simulations involving up to three subpopulations: subpopulation 1 has the fastest proliferation rate; subpopulation 2 has an intermediate proliferation rate; and subpopulation 3 has the slowest proliferation rate. We first perform three different types of discrete simulations initialised with different combinations of these three subpopulations. Each simulation is initialised so that the total number of agents occupies just 10% of the total number of lattice sites. In the first simulation we consider a homogeneous population that is composed entirely of agents from subpopulation 1, $N = 1$. The second simulation involves a heterogeneous population that is composed of equal proportions of agents from subpopulations 1 and 2, $N = 2$. The third simulation involves a heterogeneous population that is composed of equal proportions of agents from subpopulations 1, 2 and 3, $N = 3$. Snapshots from the discrete simulations are shown in Figure 2. A qualitative comparison of these snapshots shows that the growth dynamics are very different in the homogeneous and heterogeneous populations. Both the dynamics of the overall total population, and the dynamics of the various subpopulations depends on the details of the heterogeneity present in the system.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

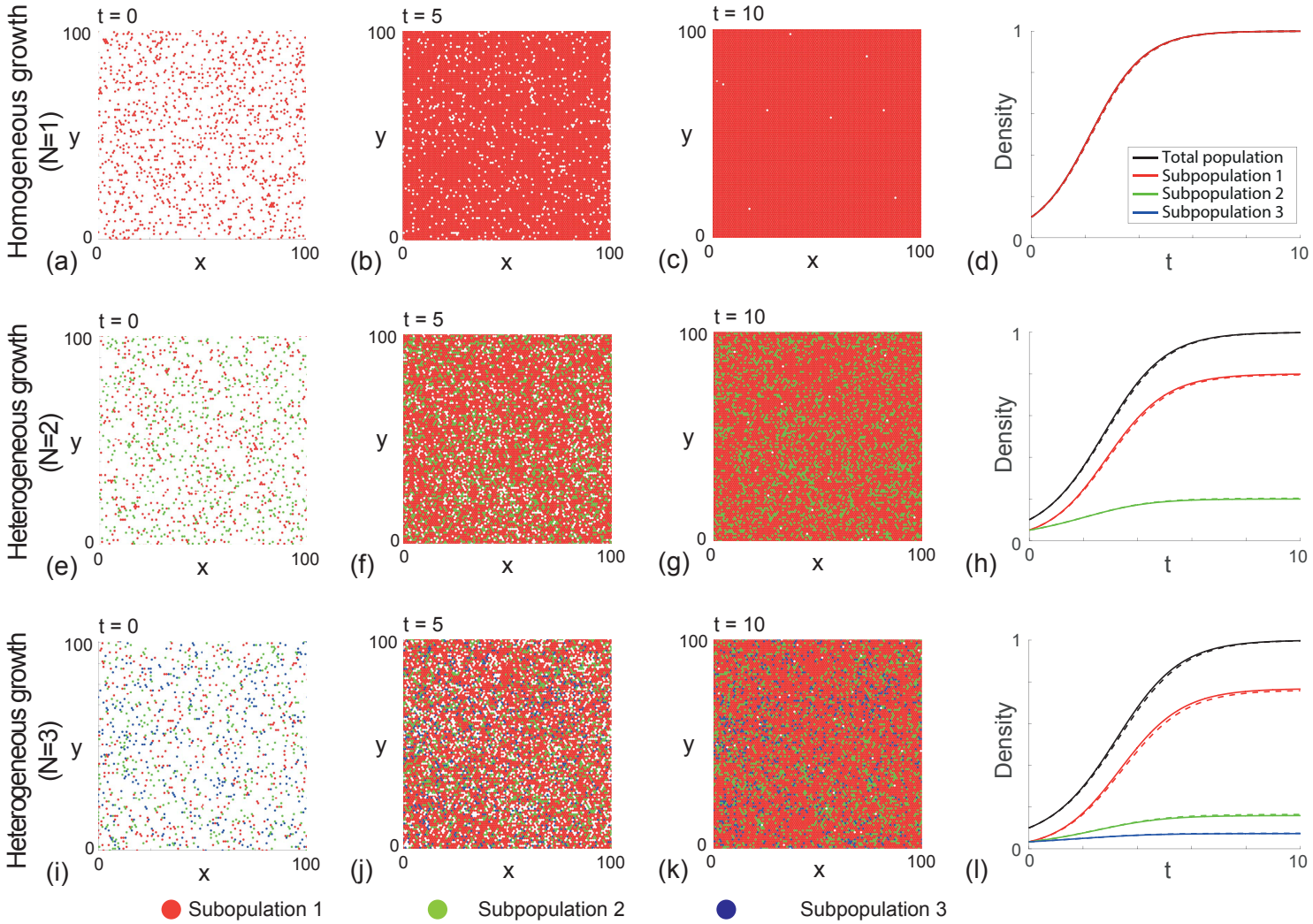

Fig. 2. **Discrete and continuum simulations of heterogeneous proliferation.** Snapshots are shown at $t = 0, 5$, and 10 for: (a)-(c) a homogeneous population, $N = 1$; (e)-(g) a heterogeneous population, $N = 2$; and (i)-(k) a heterogeneous population, $N = 3$. In each case, 10% of lattice sites are initially randomly occupied, with equal proportions of the various subpopulations. (d), (h), (l) show the corresponding continuum-discrete matches. The solid lines are solutions of the extended logistic model, dashed lines are averaged discrete results, obtained by considering 50 identically prepared realisations. In all plots red represents the fastest-proliferating subpopulation (subpopulation 1); green represents the intermediate subpopulation (subpopulation 2); and blue represents the slowest-proliferating subpopulation (subpopulation 3). All simulations correspond to $P_p^{(1)} = 0.01$, $P_p^{(2)} = 0.005$, $P_p^{(3)} = 0.0025$, $r_2 = 0.5$ and $r_3 = 0.25$.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To quantify the population growth, we plot the time evolution of the total averaged agent density, and superimpose the corresponding solution of Equation (7). For the homogeneous population in Figure 2(d), the continuum limit simplifies to the classical logistic growth model, while for the heterogeneous populations in Figure 2(h) and (l), the extended logistic model applies. Overall we see that the quality of the match between the solution of the continuum model and averaged data from the discrete simulations is excellent. Therefore, this comparison indicates that the continuum limit description is a useful and accurate mathematical tool that can be used to study the proliferation of heterogeneous populations without relying on repeated stochastic simulations.

4.2 Comparison of the classical logistic growth model and the extended logistic model

As mentioned in the Introduction, although the classical logistic growth model is widely used when interpreting data from cell biology experiments (Cai et al., 2007; Jin et al., 2016; Maini et al., 2004; Sengers et al., 2007; Sheardown and Cheng, 1996; Sherratt and Murray 1990; Vo et al., 2015; Warne et al. 2017), this standard approach neglects any heterogeneity in cell proliferation rate. To provide insight into how well the classical logistic growth model is able to predict and describe the growth of heterogeneous populations, we now calibrate the solution of the classical logistic growth model in an attempt to match the solution of the extended logistic model which explicitly accounts for heterogeneous growth.

We consider two different initial conditions for a population that is composed of three different subpopulations, $N = 3$. Again, we refer to these subpopulations as subpopulations 1, 2, and 3. For both initial conditions we consider, we distribute the total population uniformly across the domain so that the initial total density is 10% of the carrying capacity density. In the first case we

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

choose the initial condition so that the total population is initially composed of 75% of agents from subpopulation 1, 20% of agents from subpopulation 2, and 5% of agents from subpopulation 3, as shown in Figure 3(a). In the second case we choose the initial condition so that the total population is initially composed of 5% of agents from subpopulation 1, 20% of agents from subpopulation 2, and 75% of agents from subpopulation 3, as shown in Figure 3(b). These choices of initial condition mean that the first case is composed of a small proportion of relatively quiescent agents (di Fagagna et al., 2003), and the second case corresponds to a population that contains a small proportion of rapidly proliferating agents, such as is thought to be relevant to cancer progression (Davis et al., 2017). The solution of the extended logistic growth model for these two scenarios of heterogeneous growth are shown in Figure 3(b) and (d), respectively.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

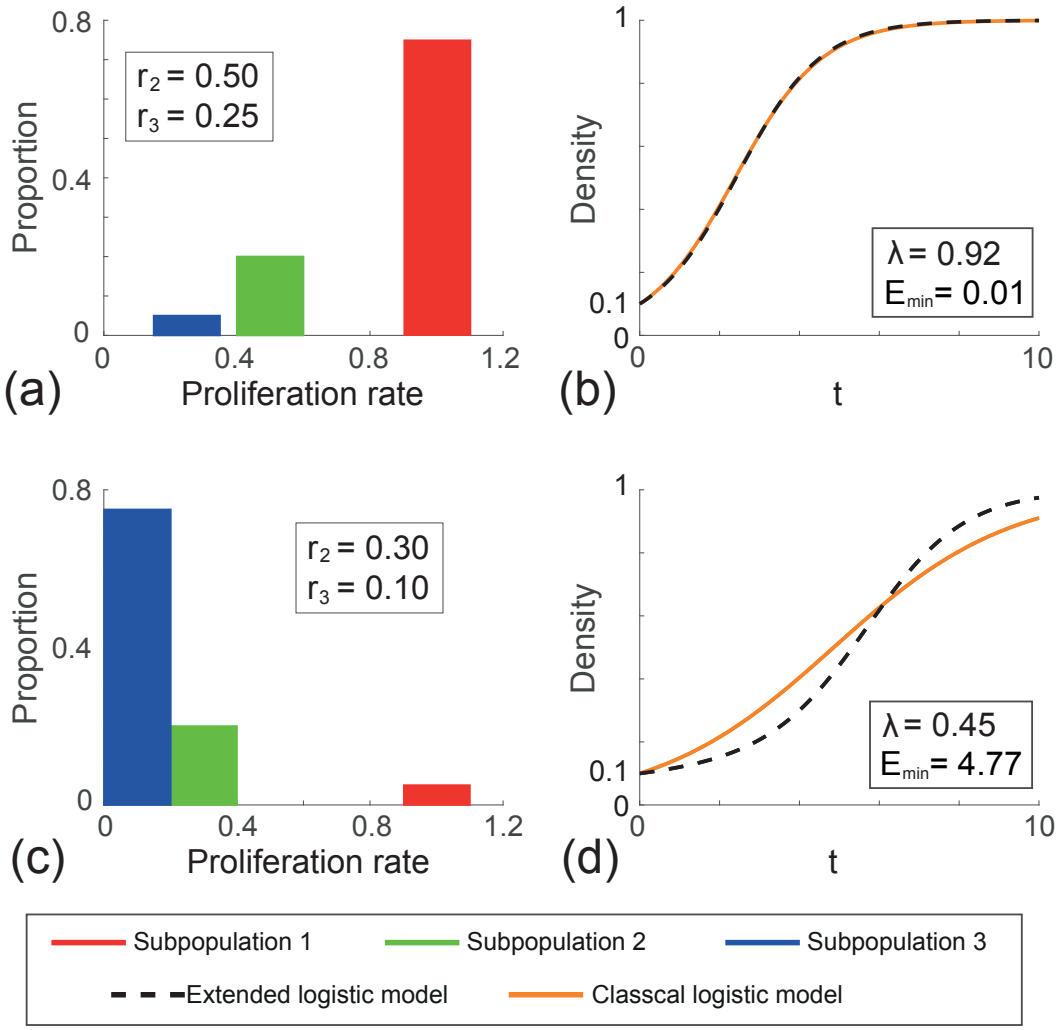

Fig. 3. **Comparison of the classical logistic growth model to the extended model for heterogenous growth.** (a) and (c) Initial distribution of proliferation rate for the two heterogenous populations. (a) corresponds to a population composed of a small proportion of relatively quiescent cells, and (c) corresponds to a population containing a small proportion of rapidly proliferating cells. In all cases, red represents the fastest-proliferating subpopulation (subpopulation 1); green represents the intermediate subpopulation (subpopulation 2); and blue represents the slowest-proliferating subpopulation (subpopulation 3). In each case, the time evolution of the total density from the extended model, together with the best-fit classical logistic growth curve, is plotted in (b) and (d).

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Again, we recall that standard approaches to interpreting cell proliferation assays is to calibrate the solution of the classical logistic growth model to match the experimental data, and to provide an estimate of the proliferation rate (Cai et al., 2007; Jin et al., 2016; Maini et al., 2004; Sengers et al., 2007; Sheardown and Cheng, 1996; Sherratt and Murray 1990; Vo et al., 2015; Warne et al. 2017). This standard approach implicitly neglects the role of heterogeneity, so it is of interest for us to take the population growth curves for these heterogeneous populations in Figure 3(b) and (d), and to calibrate the solution of the classical logistic growth model to match this data. We calibrate the solution of the classical logistic growth model to the density data in Figure 3(b) and (d) using MATLAB's *lsqcurvefit* function, and show the best match in Figure 3(b) and (d). This calibration provides an estimate of λ that is associated with the best fit of the standard model to the total density data. We are also interested in understanding the quality of match between the best fit solution of the classical logistic growth model and the heterogeneous density data. To quantify the quality of match we use a least-squares measure,

$$E = \sum_{l=1}^{1000} \left[C(t_l) - S(t_l)\right]^2 , \tag{10}$$

where $C(t)$ is the best fit solution of the classical logistic growth model, and $S(t)$ is the total density associated with the extended logistic growth model. Here we measure E in the interval $0 \leq t \leq 10$, by evaluating both $C(t)$ and $S(t)$ at 1000 equally-spaced time points, t_l for $t_l = 0, 0.01, 0.02, \ldots, 10$. We denote the minimum least–squares error as $E_{\min}$.

A simple visual comparison of the best fit classical logistic growth model and the solution of the extended model in Figure 3(b) shows that the standard approach of neglecting heterogeneity leads to an excellent match. However, results in Figure 3(d) indicate that the best fit classical logistic model matches the heterogeneous growth curve poorly. This qualitative assessment of the quality of match is confirmed quantitatively by our estimates of E, as re-

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ported in Figure 3. Overall, we see that the standard approach of neglecting heterogeneity can sometimes lead to reasonable outcomes, whereas in other cases the neglect of heterogeneity is unsatisfactory. We anticipate that the initial distribution of proliferation rates across the initial subpopulations plays an important role in determining the suitability of this standard approach. We now investigate this question further by applying our extended model to some data from the literature

5 Case study

We will now compare the performance of the classical logistic growth model and the extended logistic growth model by simulating proliferation assays in which the distributions of heterogeneous proliferation rates are taken from recent experimental measurements. In particular, we work with data from two human melanoma cell lines (Haass et al., 2014): (i) the 1205Lu cell line, which we refer to as cell line 1, and (ii) the WM983C cell line, which we refer to as cell line 2. The data we use to characterise the distribution of proliferation rates comes directly from Haass et al. (2014) where they use a specialised fluorescent technique to characterise the cell cycle of individual melanoma cells. Data from Haass et al. (2014) reports the duration of time spent in the S/G2/M phase of the cell cycle for groups of at least 20 individual cells from multiple melanoma cell lines. Since the S/G2/M cell cycle phase is closely related to the process of cell division, we treat the heterogeneity in these measurements as being representative of the heterogeneity present in the entire cell cycle.

Haass' data reports the duration of time that at least 20 individual cells spend in the S/G2/M cell cycle phase (Haass et al., 2014). We group these individual measurements of duration into three subgroups. We choose the subgroups so that each column in the histogram of this data has approximately the same width, as shown in Figure 4(a)-(b). Since Haass' data is reported in terms of a

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

duration of time spent in the cell cycle, we convert these durations into rates by dividing $\log_e(2)$ by the reported durations. Here, the $\log_e(2)$ term comes from making the simple assumption that cells are growing exponentially. The details of the raw experimental data are given in the Supplementary Material document. Presenting the dimensional rates in Figure 4(c)-(d) indicates that the distribution of proliferation rate in cell line 1 is approximately symmetric, whereas the distribution of rates for cell line 2 is positively skewed.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

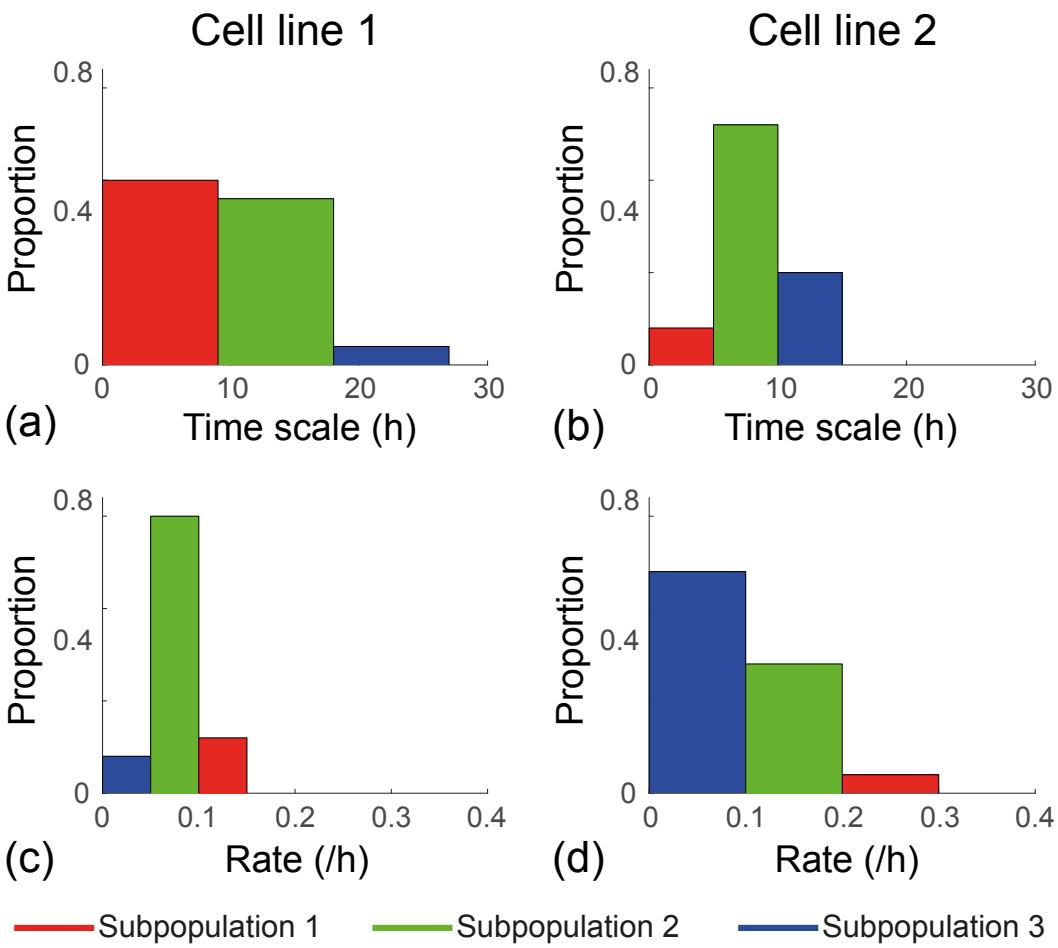

Fig. 4. **Experimental data from Haass et al. (2014).** (a)-(b) Histograms showing the distribution of the time scale associated with the S/G2/M cell cycle for cell lines 1 and 2, respectively. (c)-(d) Histograms showing the distribution of rates associates with the S/G2/M cell cycle for cell lines 1 and 2, respectively. The time scales are converted to rates by dividing $\log_e(2)$ by the time scale. For both cell lines: red indicates the fastest-proliferating subpopulation (subpopulation 1); green indicates the intermediate subpopulation (subpopulation 2); and blue indicates the slowest-proliferating subpopulation (subpopulation 3).

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To apply our models using the heterogeneous proliferation rate data in Figure 4, we non-dimensionalise the proliferation rate data by dividing each rate by the fastest-proliferation rate for each cell line. This data, presented as histograms in Figure 5(a)-(e), shows the distribution of non-dimensional proliferation rates for both cell lines. We now use these histograms to specify both the initial proliferation rates, and the initial distribution of the three subpopulations in the discrete model and the corresponding extended logistic continuum model.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

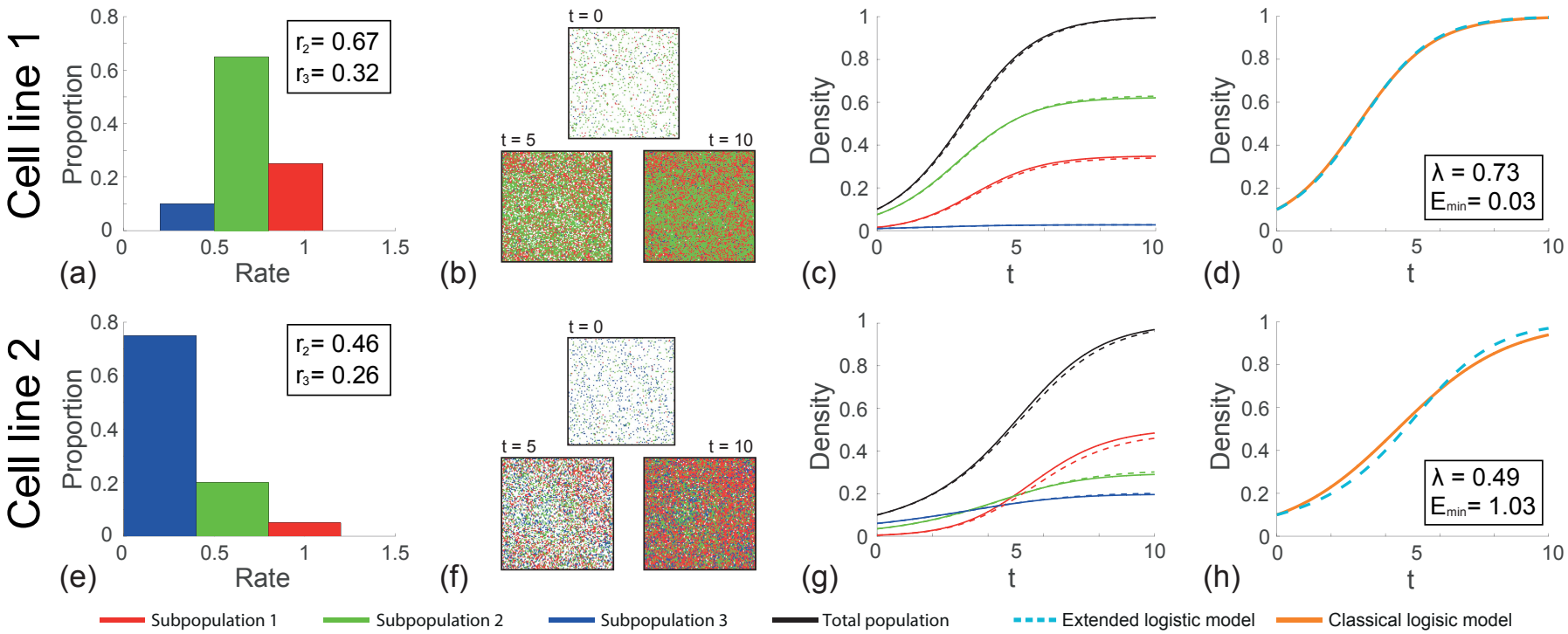

Fig. 5. **Comparison of the classical logistic growth model to the extended model for experimental cell lines 1 and 2.** (a) and (e) Initial distribution of proliferation rate for cell lines 1 and 2, respectively. (b) and (f) Snapshots of simulations at $t = 0, 5$, and 10 for both experimental cell lines. (c) and (g) The continuum-discrete match of the time evolution of the densities for experimental cell lines 1 and 2, respectively. Solid lines represent the numerical solutions of the extended model, and the dashed lines represent the averaged simulation data over 50 identically prepared realisations. (d) and (h) Calibrated solution of the classical logistic growth model superimposed with the total density profile computed using the extended model. For both of experimental cell lines, red represents the fastest-proliferating subpopulation (subpopulation 1); green represents the intermediate subpopulation (subpopulation 2); and blue represents the slowest-proliferating subpopulation (subpopulation 3). For cell line 1, $P_p^{(1)} = 0.01$, $P_p^{(2)} = 0.0067$, and $P_p^{(3)} = 0.0032$. For cell line 2, $P_p^{(1)} = 0.01$, $P_p^{(2)} = 0.0046$, and $P_p^{(3)} = 0.0026$.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We first model the heterogeneous population growth for cell line 1 and 2 using the discrete model. For each cell line, agents are initially distributed uniformly on the lattice so that 10% of lattice sites are occupied. We are also careful to ensure that the initial proportions of subpopulation 1, 2 and 3 correspond to the proportions of the three subpopulations in the histograms in Figure 5(a) and (e). Snapshots of the growing cell populations for both cell lines are given in Figure 5(b) and (f). These snapshots immediately reveal some interesting features. For cell line 1, the distribution of the three subpopulations remains similar over time, as there appears to be roughly equal proportions of red, green and blue agents at $t = 10$ as there are initially, at $t = 0$. However, we observe very different behaviour for cell line 2, as the relative abundance of the three subpopulations changes dramatically over time. For example, at $t = 0$, we see that subpopulation 1 is the least abundant subpopulation. However, by the end of the growth process, at $t = 10$, subpopulation 1 is the most abundant subpopulation. These qualitative trends are also clear in Figure 5(c) and (g) where we compare averaged discrete data from repeated simulations of the stochastic model and the solution of the corresponding continuum model. In addition to quantifying the behaviour we see in the discrete snapshots, the solution of the continuum model in Figure 5(c) and (g) confirm that the continuum model is an accurate approximation of the discrete model.

To examine the implications of taking a standard approach and neglecting the role of heterogeneity, we also calibrate the solution of the classical logistic model to the total density data in Figure 5(c) and (g). Following the same approach described in Section 4.2, results in Figure 5(d) and (h) show the evolution of the total cell density profile superimposed with the best-fit classical logistic growth curves for cell line 1 and 2, respectively. Interestingly, the quality of match between the classical logistic growth model and the heterogenous population growth curve is relatively good for cell line 1, whereas the quality of match for cell line 2 is poor.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

These results show that the consequences of neglecting the role of heterogeneity is subtle. In particular, under some circumstances it is possible to accurately predict the growth of a heterogeneous cell population using the classical logistic growth model, whereas in other circumstances the classical logistic growth model provides a poor match.

6 Analytical insight for two subpopulations, $N = 2$

To support our numerical solutions of the continuum model developed in Section 3, we now provide some simple analysis. This analysis provides both mathematical insight into the extended logistic growth model, as well as providing biological insight into the effects of heterogeneity. For brevity we concentrate on the case in which there are two subpopulations present, with densities $C_1(t)$ and $C_2(t)$. In this case the extended model, Equation (7), simplifies to

$$\begin{aligned} \frac{\mathrm{d}C_1(t)}{\mathrm{d}t} &= C_1(t)\Big[1 - S(t)\Big], \\ \frac{\mathrm{d}C_2(t)}{\mathrm{d}t} &= r_2 C_2(t)\Big[1 - S(t)\Big], \end{aligned} \tag{11}$$

where $r_2 \leq 1$. The governing equation for the evolution of the total density is

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = \Big[C_1(t) + r_2 C_2(t)\Big]\Big[1 - S(t)\Big]. \tag{12}$$

The solutions for both $C_1(t)$ and $C_2(t)$ are sigmoid curves that monotonically increase from the initial densities, $C_1(0)$ and $C_2(0)$, provided that $C_1(0) + C_2(0) < 1$. In the long time limit the solution of Equation (11) reaches a steady state solution, where $S(t) \to 1$ as $t \to \infty$. To analyse this long time behaviour we denote the steady state densities as,

$$\mathcal{C}_1 = \lim_{t\to\infty} C_1(t), \qquad \mathcal{C}_2 = \lim_{t\to\infty} C_2(t), \tag{13}$$

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

so that we have $\mathcal{C}_1 + \mathcal{C}_2 = 1$.

6.1 Exact steady state concentrations

It is not immediately clear what the steady state values $\mathcal{C}_1$ and $\mathcal{C}_2$ are from Equation (11) without first solving the transient model for the long time behaviour (Simpson et al., 2007). Furthermore, it is unclear how these steady state densities depend on the initial condition, $C_1(0)$ and $C_2(0)$, or on the proliferation rate r_2. To provide insight into this question, we can solve for $\mathcal{C}_1$ and $\mathcal{C}_2$ directly by first dividing one of the equations in (11) by the other, separating variables, and integrating to give

$$C_2(t) = \frac{C_2(0)}{\Big[C_1(0)\Big]^{r_2}} \Big[C_1(t)\Big]^{r_2}. \tag{14}$$

This relationship holds for all t. By substituting Equation (14) into Equation (11), we eliminate $C_2(t)$ to give

$$\frac{\mathrm{d}C_1(t)}{\mathrm{d}t} = r_1 C_1(t) \left(1 - C_1(t) - \frac{C_2(0)}{\Big[C_1(0)\Big]^{r_2}} \Big[C_1(t)\Big]^{r_2} \right). \tag{15}$$

This equation is a now a direct analogue of the classical logistic growth model. For general values of $r_2 < 1$, Equation (15) has no exact solution. However, in this form it is easy to read off the steady-state value by setting the time derivative to zero, resulting in

$$1 - \mathcal{C}_1 - \frac{C_2(0)}{\Big[C_1(0)\Big]^{r_2}} \left(\mathcal{C}_1\right)^{r_2} = 0. \tag{16}$$

Equation (16) is a simple algebraic equation which can be solved using any iterative numerical method, such as MATLAB's *fsolve* function. The value of $\mathcal{C}_2$ is then given by $\mathcal{C}_2 = 1 - \mathcal{C}_1$.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results in Figure 6(a)-(b) show two examples where we have used this approach to directly calculate $\mathcal{C}_1$ and $\mathcal{C}_2$. These predictions are superimposed on the associated transient solutions of Equation (11), showing that the direct method provides a simple and accurate way to calculate the long-time steady solution, without needing to use numerical integration to evaluate the long time limit of the transient solution.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

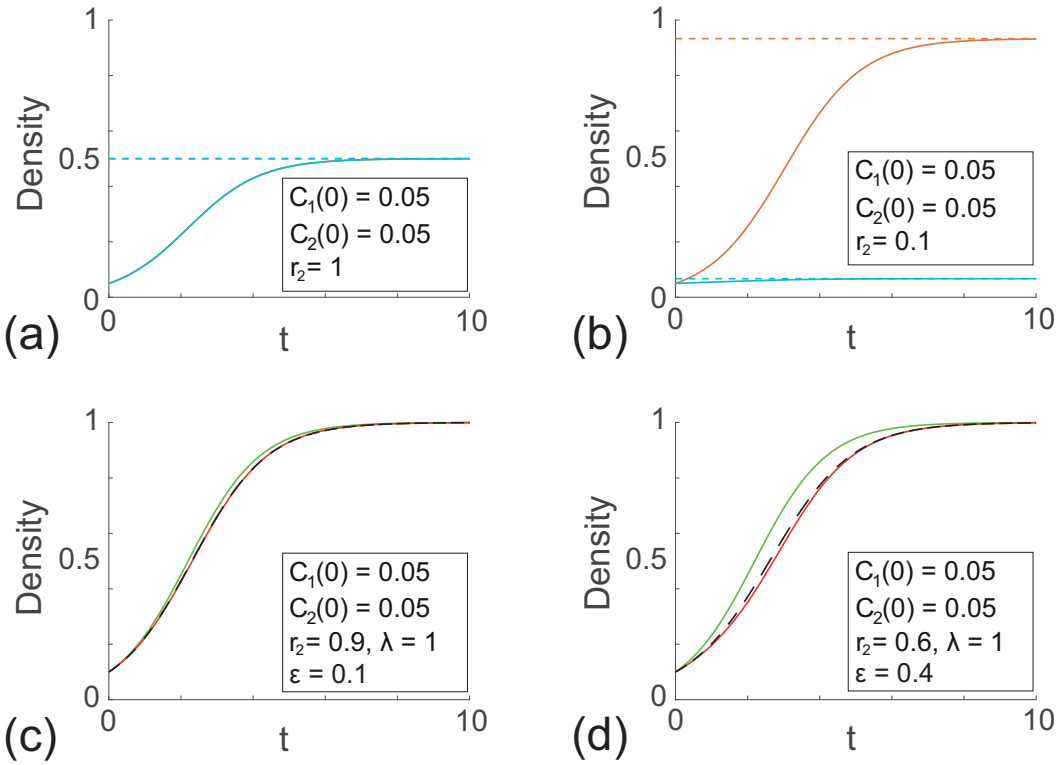

Fig. 6. **Analytical insight into the extended logistic model with two subpopulations,** $N = 2$**.** (a)-(b) Steady state results (dashed) compared to the full transient numerical solutions (solid). Here, the first subpopulation is shown in orange, and the second subpopulation is shown in blue. (c)-(d) Comparison of the transient numerical solutions of total density (black) with the two-term perturbation solutions in the limit of small heterogeneity. The $\mathcal{O}(1)$ perturbation solution is plotted in green and the two-term $\mathcal{O}(\varepsilon)$ perturbation solution is plotted in red.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6.2 Approximate results for small heterogeneity

Although the extended logistic growth model given by Equation (11) does not have an exact solution, we can obtain approximate results in the limit of small heterogeneity. To explore this we consider $r_2 = 1 - \varepsilon$, where $\varepsilon \ll 1$, and propose the perturbation solution (Murray, 2012)

$$\begin{aligned} C_1(t) &= C_1^{(0)}(t) + \varepsilon C_1^{(1)}(t) + \mathcal{O}(\varepsilon^2), \\ C_2(t) &= C_2^{(0)}(t) + \varepsilon C_2^{(1)}(t) + \mathcal{O}(\varepsilon^2), \end{aligned} \tag{17}$$

where the superscripts (0) and (1) represent the leading order and first correction terms, respectively. The asymptotic solution for the total population is obtained by summing over the solutions for the two subpopulations:

$$S(t) = S^{(0)}(t) + \varepsilon S^{(1)}(t) + \mathcal{O}(\varepsilon^2). \tag{18}$$

Substituting Equation (17) into the extended logistic model, given by Equation (11), gives the system

$$\begin{aligned} \frac{\mathrm{d}C_1^{(0)}(t)}{\mathrm{d}t} &= C_1^{(0)}(t)\Big[1 - S^{(0)}(t)\Big], \\ \frac{\mathrm{d}C_2^{(0)}(t)}{\mathrm{d}t} &= C_2^{(0)}(t)\Big[1 - S^{(0)}(t)\Big], \end{aligned} \tag{19}$$

with $C_1^{(0)}(0) = C_1(0)$, $C_2^{(0)}(0) = C_2(0)$. Correspondingly, the $\mathcal{O}(1)$ equation for the total population is

$$\frac{\mathrm{d}S^{(0)}(t)}{\mathrm{d}t} = S^{(0)}(t)\Big[1 - S^{(0)}(t)\Big], \tag{20}$$

with $S^{(0)}(0) = S(0)$. Equation (20) is the standard logistic growth model with

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the explicit solution (Murray, 2002)

$$S^{(0)}(t) = \frac{S(0)}{\Big[1 - S(0)\Big]\mathrm{e}^{-t} + S(0)}. \tag{21}$$

This is to be expected since our leading order problem holds for $\epsilon = 0$, in which case both populations have the same proliferation rate, so effectively there is one population. Solving for the individual subpopulations gives the leading order solutions

$$\begin{aligned} C_1^{(0)}(t) &= \frac{C_1(0)}{\Big[1 - S(0)\Big]\mathrm{e}^{-t} + S(0)}, \\ C_2^{(0)}(t) &= \frac{C_2(0)}{\Big[1 - S(0)\Big]\mathrm{e}^{-t} + S(0)}. \end{aligned} \tag{22}$$

For $\varepsilon \ll 1$, we proceed to solve for the correction terms. The governing equations for the individual populations are

$$\begin{aligned} \frac{\mathrm{d}C_1^{(1)}(t)}{\mathrm{d}t} &= C_1^{(1)}(t)\Big[1 - S^{(0)}(t)\Big] - C_1^{(0)}(t)S^{(1)}(t), \\ \frac{\mathrm{d}C_2^{(1)}(t)}{\mathrm{d}t} &= \Big[C_2^{(1)}(t) - C_2^{(0)}(t)\Big]\Big[1 - S^{(0)}(t)\Big] - C_2^{(0)}(t)S^{(1)}(t), \end{aligned} \tag{23}$$

while the corresponding equation for the total density is

$$\frac{\mathrm{d}S^{(1)}(t)}{\mathrm{d}t} = S^{(1)}(t)\Big[1 - 2S^{(0)}(t)\Big] - C_2^{(0)}(t)\Big[1 - S^{(0)}(t)\Big]. \tag{24}$$

Equation (24) has an explicit solution

$$S^{(1)}(t) = \frac{C_2(0)\Big[S(0) - 1\Big]t\mathrm{e}^{-t}}{\Big\{-\mathrm{e}^{-t}\Big[1 - S(0)\Big] - S(0)\Big\}^2}. \tag{25}$$

Neglecting higher order terms we obtain the two-term perturbation solution

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for the total density

$$S(t) = \frac{S(0)}{\Big[1 - S(0)\Big]\mathrm{e}^{-t} + S(0)} + \varepsilon \frac{C_2(0)\Big[S(0) - 1\Big]t\mathrm{e}^{-t}}{\Big\{ - \mathrm{e}^{-t}\Big[1 - S(0)\Big] - S(0)\Big\}^2} + \mathcal{O}(\varepsilon^2). \quad (26)$$

To demonstrate the effectiveness of this approximation, the $\mathcal{O}(1)$ and $\mathcal{O}(\varepsilon)$ perturbation solutions for the total density are plotted in Figure 6(c)-(d). The corresponding full numerical solution is also presented. For a small amount of heterogeneity in the population ($\varepsilon = 0.1$, Figure 6(c)), the leading order term (solid green) provides a reasonably good approximation of the numerical solution (dashed black). However, for larger heterogeneity ($\varepsilon = 0.4$, Figure 6(d)), we see that the leading order term is no longer close to the numerical solution, and instead the full two-term perturbation solution Equation (26) (solid red) is required to provide a good approximation. These plots provide further evidence that when the population is almost homogenous, then the classical logistic model provides a good approximation. However, as heterogeneity becomes more pronounced, then our extended logistic growth model does a much better job at describing the dynamics. Furthermore, provided the heterogeneity is not too great, our two-term perturbation solution acts as a very good analytical approximation.

Results in Figure 6(c)-(d) correspond to one particular choice of initial condition, $C_1(0)$ and $C_2(0)$. Additional results (Supplementary Material, Figure S3) show that Equation (26) also provides a good approximation when we vary the initial condition, provided that ε is sufficiently small.

7 Conclusions

In this study we develop discrete and continuum models of cell migration and cell proliferation that allow us to explicitly investigate the role of heterogeneity,

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with a particular emphasis on the role of heterogeneity in the proliferation rate. Despite the fact that heterogeneity is commonly observed in cell populations, and is thought to play an important role in disease progression and tissue repair (Evan and Vousden, 2001; Haridas et al., 2017; Pavlath et al., 1998), standard mathematical models of cell migration and cell proliferation neglect to account for heterogeneity. Indeed, most standard mathematical models of cell proliferation simply treat the proliferation rate as a constant.

To explore the role of heterogeneity, we start by developing a discrete modelling framework to simulate cell migration and cell proliferation, modulated by crowding effects. The key point of the model is to deliberately introduce heterogeneity in the individuals within the population. The continuum limit description of the discrete model leads to a system of coupled, nonlinear ODEs. It is of interest to note that in the simplest case where the proliferation rates of each subpopulation are identical, the system of ODEs simplifies to the classical logistic growth model. Therefore, we call the new model the extended logistic growth model. Averaged data obtained from repeated simulations of the discrete model compare very well with the solution of the extended logistic growth model.

To explore the consequences of applying the logistic growth model and neglecting the role of heterogeneity, we perform a set of *in silico* experiments and generate density profiles describing the growth of a heterogeneous population of cells. We calibrate the solution of the classical logistic equation to match that data. Interestingly, while the classical logistic growth model can provide an accurate prediction of the growth of some kinds of heterogeneous populations, we also find that in some circumstances the classical approach can not make accurate predictions. We also generate *in silico* data by parameterising the extended logistic growth model with a set of heterogeneous proliferation rates from recent experimental measurements. Again, we find that the classical logistic model performs very well under some conditions, but it performs

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

poorly for others. Overall, we find that when the heterogeneous population contains a small proportion of relatively fast proliferating cells, the classical logistic equation performs poorly. Therefore, we suggest that care ought to be exercised when modelling the growth of certain cell populations. For example, when modelling a population of cells that might involve mutations that act to increase the proliferation rate in a small subpopulation (e.g. Davis et al., 2017), the extended logistic growth model might be more accurate than the classical logistic equation.

Unlike the classical logistic equation, the extended logistic growth model does not have an exact solution. Therefore, in general, we have to rely on numerical solutions. However, we also show how to provide some further insight by obtaining analytical solutions in the case where there are just two subpopulations present, $N = 2$. We obtain exact expressions for the long-time steady state solution, and show that these exact expressions can be solved numerically to predict the steady state solution without using numerical integration to solve the full transient model. Furthermore, we also obtain approximate insight by constructing perturbation solutions in the limit that the degree of heterogeneity is small. The perturbation solutions are insightful since the $\mathcal{O}(1)$ perturbation solution for the total cell density is the classical logistic equation. The $\mathcal{O}(\varepsilon)$ perturbation solution provides a correction term that is accurate even when we consider a relatively large degree of heterogeneity in the system.

Although we have focused here on the question of developing mathematical tools and mathematical insight into the role of heterogeneity in population dynamics associated with populations of cells, it seems likely that the ideas explored here will have consequences beyond the cell biology literature. For example, classical logistic models, with constant growth rates, are also commonly used in mathematical ecology (e.g. Chan and Kim, 2013), and there is also an awareness in the ecology literature that ecological populations do not

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

always grow logistically (e.g. Taylor and Hastings, 2005). Therefore, perhaps some of the ideas developed here might also play a role in our understanding of ecological population dynamics. Another feature of our work is that we have focused exclusively on heterogeneity in cell proliferation rates. However, we note that there is also considerable interest in developing quantitative, predictive mathematical models which incorporate heterogeneity in cell migration (e.g. Read et al., 2016). Again, it seems likely that the kind of approach taken here would also be of interest in the context of exploring heterogeneity in cell migration. These open questions could be considered in future studies.

Acknowledgments. This work is supported by the Australian Research Council (DP140100249, DP170100474). Wang Jin is supported by a QUT Vice Chancellor's Research Fellowship. We thank two anonymous reviewers for their helpful comments.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

[1] F.Q. An, M. Matsuda, H. Fujii, R.F. Tang, H. Amemiya, Y.M. Dai, Y. Matsumoto. Tumor heterogeneity in small hepatocellular carcinoma: Analysis of tumor cell proliferation, expression and mutation of p53 and β-catenin. International Journal of Cancer. 93 (2001) 468–474.

[2] J. Arino, L. Wang, G.S. Wolkowicz. An alternative formulation for a delayed logistic equation. Journal of Theoretical Biology. 241 (2006) 109–119.

[3] B. Azzarone, A. Macieira-Coelho. Heterogeneity of the kinetics of proliferation within human skin fibroblastic cell populations. Journal of Cell Science. 57 (1982) 177–187.

[4] B.T. Bajar, A.J. Lam, R.K. Badiee, Y.-H. Oh, J. Chu, X.X. Zhou, N. Kim, B.B. Kim, M. Chung, A.L. Yablonovitch, B.F. Cruz, K.Kulalert, J.J. Tao, T. Meyer, X.-D. Su, and M.Z. Lin. Fluorescent indicators for simultaneous reporting of all four cell cycle phases. Nature Methods 13 (2016) 993–996.

[5] A.P. Browning, S.W. McCue, R.N. Binny, M.J. Plank, E.T. Shah, M.J. Simpson. Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. Journal of Theoretical Biology. 437 (2018) 251–260.

[6] A.Q. Cai, K.A. Landman, B.D. Hughes. Multi-scale modeling of a wound-healing cell migration assay. Journal of Theoretical Biology. 245 (2007) 576–594.

[7] T. Callaghan, E. Khain, L.M. Sander, R.M. Ziff. (2006) A stochastic model for wound healing. Journal of Statistical Physics. 122 (2006) 909–924.

[8] M.H. Chan, P.S. Kim. Modelling a Wolbachia invasion using a slow-fast dispersal reaction-diffusion approach. Bulletin of Mathematical Biology. 75 (2015) 1501–1523.

[9] D. Chowdhury, A. Schadschneider, K. Nishinari. Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms. Physics of Life Reviews, 2 (2005) 318–352.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[10] A. Davis, R. Gao, N. Navin. Tumor evolution: Linear, branching, neutral or punctuated? Biochimica et Biophysica Acta - Reviews on Cancer. 1867 (2017) 151–161.

[11] C. Deroulers, M. Aubert, M. Badoual, B. Grammaticos. Modeling tumor cell migration: From microscopic to macroscopic models. Physical Review E. 79 (2009) 031917.

[12] F.D. di Fagagna, P.M. Reaper, L. Clay-Farrace, H. Fiegler, P. Carr, T. von Zglinicki, G. Saretzki, N.P. Carter, S.P. Jackson. A DNA damage checkpoint response in telomere-initiated senescence. Nature. 426 (2003) 194–198.

[13] L. Dyson, R. E. Baker. The importance of volume exclusion in modelling cellular migration. Journal of Mathematical Biology. 71 (2015) 691–711.

[14] A. Eladdadi, D. Isaacson. A mathematical model for the effects of HER2 overexpression on cell proliferation in breast cancer. Bulletin of Mathematical Biology. 70 (2008) 1707–1729.

[15] G.I. Evan, K.H. Vousden. Proliferation, cell cycle and apoptosis in cancer. Nature, 411 (2001) 342–348.

[16] F. Frascoli, B.D. Hughes, M.H. Zaman, K.A. Landman. A computational model for collective cellular motion in three dimensions: general framework and case study for cell pair dynamics. PLOS ONE. 8 (3), e59249.

[17] J. García-Algarra, J. Galeano, J.M. Pastor, J.M. Iriondo, J.J. Ramasco. Rethinking the logistic approach for population dynamics of mutualistic interactions. Journal of Theoretical Biology. 363 (2014) 332–343.

[18] P. Gerlee. The model muddle: in search of tumor growth laws. Cancer Research. 73 (2013) 2407–2411.

[19] Z. Guan, S. Jia, Z. Zhu, M. Zhang, C.J. Yang. Facile and rapid generation of large-scale microcollagen gel array for long-term single-cell 3D culture and cell proliferation heterogeneity analysis. Analytical Chemistry. 86 (2014) 2789–2797.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[20] N.K. Haass, K.A. Beaumont, D.S. Hill, A. Anfosso, P. Mrass, M.A. Munoz, I. Kinjyo, W. Weninger. Real-time cell cycle imaging during melanoma growth, invasion, and drug response. Pigment Cell & Melanoma Research. 27 (2014) 764–776.

[21] P. Haridas, C.J. Penington, J.A. McGovern, D.L.S. McElwain, M.J. Simpson. Quantifying rates of cell migration and cell proliferation in co-culture barrier assays reveals how skin and melanoma cells interact during melanoma spreading and invasion. Journal of Theoretical Biology. 423 (2017) 13–25.

[22] L. Hayflick. The limited *in vitro* lifetime of human diploid cell strains. Experimental Cell Research. 37 (1965) 614–636.

[23] W. Jin, E.T. Shah, C.J. Penington, S.W. McCue, L.K. Chopin, M.J. Simpson. Reproducibility of scratch assays is affected by the initial degree of confluence: experiments, modelling and model selection. Journal of Theoretical Biology. 390 (2016) 136–145.

[24] W. Jin, C.J. Penington, S.W. McCue, M.J. Simpson. Stochastic simulation tools and continuum models for describing two–dimensional collective cell spreading with universal growth functions. Physical Biology. 13 (2016) 056003.

[25] W. Jin, C.J. Penington, S.W. McCue, M.J. Simpson. A computational modelling framework to quantify the effects of passaging cell lines. PLOS ONE. 12 (2017) e0181941.

[26] W. Jin, E.T. Shah, C.J. Penington, S.W. McCue, P.K. Maini, M.J. Simpson. Logistic proliferation of cells in scratch assays is delayed. Bulletin of Mathematical Biology. 79 (2017) 1028–1050.

[27] P. Jolicoeur, J. Pontier. Population growth and decline: a four-parameter generalization of the logistic curve. Journal of Theoretical Biology. 141 (1989) 563–571.

[28] H. Kaneoka, G. Perez-Rojas, T. Sasasuki, C.J. Benike, E.G. Engleman. Human T lymphocyte proliferation induced by a pan-T monoclonal antibody (anti-Leu

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4): heterogeneity of response is a function of monocytes. Journal of Immunology. 131 (1983) 158–164.

[29] P.K. Maini, D.L.S. McElwain, D.I. Leavesley. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. Tissue Engineering. 10 (2004) 475–482.

[30] J.D. Murray. Asymptotic analysis. (2012) Springer Science & Business Media, New York.

[31] J.D. Murray. Mathematical biology. I: An introduction, 3rd edn. (2002) Springer, Heidelberg.

[32] J.T. Nardini, D.A. Chapnick, X. Liu, D.M. Bortz. Modeling keratinocyte wound healing dynamics: cell-cell adhesion promotes sustained collective migration. Journal of Theoretical Biology. 400 (2016) 103–117.

[33] Z. Neufeld, W. von Witt, D. Lakatos, J. Wang, B. Hegedus, A. Czirok. The role of Allee effect in modelling post resection recurrence of glioblastoma. PLoS Computational Biology. 13(11) (2017) e1005818.

[34] G.K. Pavlath, D. Thaloor, T.A. Rando, M. Cheong, A.W. English, B. Zheng. Heterogeneity among muscle precursor cells in adult skeletal muscles with differing regenerative capacities. Developmental Dynamics. 212 (1998) 495–508.

[35] M.J. Plank, M.J. Simpson. Models of collective cell behaviour with crowding effects: comparing lattice–based and lattice–free approaches. Journal of the Royal Society Interface. 9 (2012) 2983–2996.

[36] C.D. Powell, S. López, A. Dumas, D.P. Bureau, S.E. Hook, J. France. Mathematical descriptions of indeterminate growth. Journal of Theoretical Biology. 425 (2017) 88–96.

[37] V. Pozzobon, P. Perré. Han's model parameters for microalgae grown under intermittent illumination: Determined using particle swarm optimization. Journal of Theoretical Biology. 437 (2018) 29–35.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[38] M.N. Read, J. Bailey, J. Timmis, T. Chtanova. Leukocyte motility models assessed through simulation and multi-objective optimization-based model selection. PLOS Computational Biology. 12(9) (2016) e1005082.

[39] A. Sakaue-Sawano, H. Kurokawa, T. Morimura, A. Hanyu, H. Hama, H. Osawa, S. Kashiwagi, K. Fukami, T. Miyata, H. Miyoshi, T. Imamura, M. Ogawa, H. Masai, A. Miyawaki. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell 132 (2008) 487–498.

[40] E.A. Sarapata, L.G. de Pillis. A comparison and catalog of intrinsic tumor growth models. Bulletin of Mathematical Biology. 76 (2014) 2010–2024.

[41] B.G. Sengers, C.P. Please, R.O. Oreffo RO. Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration. Journal of the Royal Society Interface. 4 (2007) 1107–1117.

[42] L. Sewalt, K. Harley, P. van Heijster, S. Balasuriya. Influences of Allee effects in the spreading of malignant tumours. Journal of Theoretical Biology. 394 (2016) 77–92.

[43] H. Sheardown, Y.L. Cheng. Mechanisms of corneal epithelial wound healing. Chemical Engineering Science. 51 (1996) 4517–4529.

[44] J.A. Sherratt, J.D. Murray JD. Models of epidermal wound healing. Proceedings of the Royal Society Series B. 241 (1990) 29–36.

[45] M.J. Simpson, K.A. Landman, K. Bhaganagarapu. Coalescence of interacting cell populations. Journal of Theoretical Biology. 247 (2007) 525–543.

[46] M.J. Simpson, K.A. Landman, B.D. Hughes. Multi-species simple exclusion processes. Physica A: Statistical Mechanics and its Applications. 388 (2009) 399–406.

[47] M.J. Simpson, K.A. Landman, B.D. Hughes. Cell invasion with proliferation mechanisms motivated by time–lapse data. Physica A: Statistical Mechanics and its Applications. 389 (2010) 3779–3790.

[48] C.M. Taylor, A. Hastings. Allee effects in biological invasions. Ecological Letters. (2005) 895–908.

bioRxiv preprint doi: https://doi.org/10.1101/231100; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[49] K.K. Treloar, M.J. Simpson, D.L.S. McElwain, R.E. Baker. Are *in vitro* estimates of cell diffusivity and cell proliferation rate sensitive to assay geometry? Journal of Theoretical Biology. 356 (2014) 71–84.

[50] A. Tsoularis, J. Wallace. Analysis of logistic growth models. Mathematical Biosciences. 179 (2002) 21–55.

[51] S.T. Vittadello, S.W. McCue, G. Gunasingh, N.K. Haass, M.J. Simpson. Mathematical models for cell migration with real-time cell cycle dynamics. Biophysical Journal. https://doi.org/10.1016/j.bpj.2017.12.041 (2018).

[52] B.N. Vo, C.C. Drovandi, A.N. Pettitt, M.J. Simpson MJ. Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. Mathematical Biosciences. 263 (2015) 133–142.

[53] D.J. Warne, R.E. Baker, M.J. Simpson. Optimal quantification of contact inhibition in cell populations. Biophysical Journal. 113 (2017) 1920–1934.

[54] W. Wagner, S. Bork, P. Horn, D. Krunic, T. Walenda, A. Diehlmann, V. Benes, J. Blake, F.X. Huber, V. Eckstein, P. Boukamp. Aging and replicative senescence have related effects on human stem and progenitor cells. PLOS. ONE. 4 (2002) e5846.

[55] G.B. West, J.H. Brown, B.J. Enquist BJ. A general model for ontogenetic growth. Nature. 413 (2001) 628–631.

[56] J.M. Willaime, F.E. Turkheimer, L.M. Kenny, E.O. Aboagye. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. Physics in Medicine and Biology. 58 (2013) 187–203.