Published in final edited form as:

Phys Biol. ; 14(3): 035002. doi:10.1088/1478-3975/aa6e5a.

The finite state projection approach to analyze dynamics of heterogeneous populations

Rob Johnson¹ and Brian Munsky^{2,3}

¹Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

²Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO 80523, USA

³School of Biomedical Engineering, Colorado State University, Fort Collins, CO 80523, USA

Abstract

Population modeling aims to capture and predict the dynamics of cell populations in constant or fluctuating environments. At the elementary level, population growth proceeds through sequential divisions of individual cells. Due to stochastic effects, populations of cells are inherently heterogeneous in phenotype, and some phenotypic variables have an effect on division or survival rates, as can be seen in partial drug resistance. Therefore, when modeling population dynamics where the control of growth and division is phenotype dependent, the corresponding model must take account of the underlying cellular heterogeneity. The finite state projection (FSP) approach has often been used to analyze the statistics of independent cells. Here, we extend the FSP analysis to explore the coupling of cell dynamics and biomolecule dynamics within a population. This extension allows a general framework with which to model the state occupations of a heterogeneous, isogenic population of dividing and expiring cells. The method is demonstrated with a simple model of cell-cycle progression, which we use to explore possible dynamics of drug resistance phenotypes in dividing cells. We use this method to show how stochastic single-cell behaviors affect population level efficacy of drug treatments, and we illustrate how slight modifications to treatment regimens may have dramatic effects on drug efficacy.

Keywords

stochastic gene expression; chemical master equation; single-cell heterogeneity; population dynamics

1. Introduction

Cell populations expand and contract through progressions of cell growth, division and death that give rise to fluctuating numbers of genetically identical, but phenotypically heterogeneous cells. The behaviors at the macro, or population, level emerge in part as a result of the heterogeneous intracellular conditions driving cell dynamics at the micro, or individual, level. These conditions are heterogeneous due to the inherent stochasticity of biochemical systems coupled with asymmetric single-cell divisions. Therefore, in order to

understand the mechanisms of population growth at the macro level, it is necessary to connect these to the mechanisms of cellular variability at the micro level [1].

One expedient example is the public health threat of antimicrobial resistance, whose emergence is underpinned by both molecular events and population dynamics [2]. Other related examples include temporary resistance of individual tumor cells to radiation or chemotherapy and epigenetically controlled latency of viral infections to escape antiretroviral therapy. In order to design effective strategies to prevent or avert resistance in these situations, it is necessary to understand how heterogeneous populations respond to toxic conditions. Variability within the population is crucial to these dynamics, underlying the emergence of persister populations [3] and support of mutability [4]. Efforts to predict and mitigate these resistance dynamics can gain much from modeling how internal factors interplay with external conditions (e.g., drug treatment regimens) to control population-level dynamics [1].

In this article, we propose an adaptation to the finite state projection (FSP, [5]) approach, which allows us to couple discrete stochastic dynamics at the cellular level to the overall dynamics at the population level. In the next section, we review previous approaches for the modeling of cell population growth as well as the FSP approach to study single-cell heterogeneity. We then introduce a modification to the FSP to allow for the study of coupled population growth and heterogeneity. We illustrate this approach on an example of eukaryotic cell-cycle progression and growth coupled with time-varying drug treatments, and we show how stochastic single-cell heterogeneity can interact with the cell cycle to elude certain drug treatment regimens. Finally, we summarize our results and discuss their implications for biological and biomedical applications.

2. Methods

Methods previously presented for modeling the growth of cell populations have included both analytical formulations and computational simulations. Delay differential equations impose a time lag for a newly born cell to mature before it is able to divide, acting as a proxy for cell-cycle stages that are not explicitly modeled [6]. Branching processes are used to derive analytical expressions for the existence of a stable birth-type distribution, where new cells inherit some ‘type’ from their mother that determines their propensities to reproduce [7, 8]. These processes lend themselves to questions of cell differentiation [9], for example, and take account of the type as a random variable, but do not allow for a mechanistic underpinning for this variability. Where analytical models are unavailable, researchers have used stochastic simulations, agent-based models, and cellular automata to simulate population trajectories for different models of growth [4, 10, 11, 12, 13, 14].

Physiologically structured population models stratify the population under consideration according to some physiological attribute(s) [15], such as age [16], size [17], cell-cycle stage [18], DNA content [19] and protein. In these models, the growth rate for each individual depends on that individual's realization of the controlling attribute. For example, in age- or size-structured models, older (or larger, respectively) cells have a greater propensity to divide. This structuring prevents the artifact of newly born cells immediately dividing. Such

models can also be used with label structuring to infer rates of death and division from fluorescence data [20].

The present work most closely relates to protein-structured models, in which division depends upon protein content [21, 22, 23]. For example, a *population-expression* analysis was developed in [21] to examine clonal selection dynamics in lymphocytes upon pathogen exposure. The authors showed that models that consider molecular dynamics alone, or population dynamics alone (where the populations are differentiated phenotypic states defined according to an artificial discretization of the molecular state space), fail to capture population dynamics. To simulate the interconnected effects of expression dynamics and population differentiation, the authors developed continuous-valued partial differential equations (PDEs) with non-local effects. In a similar approach, other analyses have focused on coupling of expression and population growth. In [22], the population is structured by cyclin and additionally by age, whereas in [23], delay differential equations are employed to prevent immediate re-entry into the cell cycle. These growth models typically define two cell-cycle stages, proliferative and quiescent. Propensities to divide and transition to quiescence depend on age and cyclin content, whereas transition to proliferation depends on population density. Each of these analyses results in multiple coupled partial differential equations, one for each sub-population.

These studies recognize the role of cell-cycle stages in population dynamics: population growth fundamentally depends on which cells are dividing, and their daughters' propensities to progress through the cell cycle to the next division. A promising development would therefore be to include many discrete cell-cycle stages, as proteins leverage control at checkpoints throughout the cell cycle [14]. Therefore, inclusion of multiple cell-cycle stages allows fine-grained modeling of the control of intracellular molecular dynamics. Here, we present a general formulation for modeling cell population growth, which will allow for an arbitrary number of cell-cycle stages coupled to discrete fluctuations in the number of regulatory protein molecules. Different internal molecular dynamics will be defined for each stage, which may parametrize transition rates between stages and allow for complex control of cell cycle progression. Our goal is to compute the full joint probability distributions for the coupled discrete dynamics of cell-cycle stage and internal molecular composition, which will correspond to protein content or DNA damage in Sections 3.1 and 3.2, respectively.

2.1. The finite state projection

At the level of single cells, biochemical processes are subject to a host of discrete molecular events involving the creation, interaction, and degradation of DNA, RNA, and protein molecules. Because the order and timing of these discrete events are subject to thermal fluctuations and complex or hidden dynamical processes, these biochemical processes appear to occur at random. Although exact dynamics for such biochemical processes are inherently unpredictable, their probability distributions have been shown to be well described by the infinite set of linear ordinary differential equations (ODEs), known as the Chemical Master Equation (CME, [24]). For example, consider a model comprising N_S cell-cycle stages, $m = 1, 2, \dots$, coupled with production and degradation of a single protein product, which could be present in the system at any integer value, $n = 0, 1, 2, \dots$. If we let

$P_{m,n}(t)$ denote the probability that the cell is in the m^{th} stage with n protein molecules at time t , then the CME can be written

$$\frac{d}{dt}P_{m,n}(t) = \sum_{\mu=1}^{N_{\text{Rxn}}} w_{\mu}(m'_{\mu}, n'_{\mu}, t)P_{m'_{\mu}, n'_{\mu}}(t) - w_{\mu}(m, n, t)P_{m,n}(t), \quad (1)$$

where the μ^{th} reaction corresponds to the transition from (m'_{μ}, n'_{μ}) to (m, n) , and where $w_{\mu}(m'_{\mu}, n'_{\mu}, t)$ is the propensity (or stochastic reaction rate) of that reaction at time t .

Because n can take on any integer value, the CME is infinite in dimension, and it cannot be solved exactly except for very simple forms of $w_{\mu}(m, n, t)$. However, for finite times, most of the probability mass of the CME is confined to a finite region (e.g., a finite value of n). The finite state projection (FSP) approach [5] uses this fact to truncate the CME solution into a finite, solvable set of linear ODEs. By sacrificing a small yet quantified proportion of total probability mass, the FSP yields a precise, tractable differential equation for the probability distribution of the most probable realizations.

For convenience, and assuming that there are exactly N_S cell-cycle stages, the CME can be recast into matrix vector form by defining the probability mass vector,

$$\mathbf{P}(t) = [P_{1,0}(t), \dots, P_{N_S,0}(t), P_{1,1}(t), \dots, P_{N_S,1}(t), P_{1,2}(t), \dots]^T, \quad (2)$$

allowing the entire CME to be written as $\frac{d}{dt}\mathbf{P}(t) = \mathbf{A}(t)\mathbf{P}(t)$. In this formulation, the matrix $\mathbf{A}(t)$ is known as the *infinitesimal generator*, and each of its elements, $A_{ij}(t)$, corresponds to the instantaneous rate of transition from state j to state i at time t . In the matrix CME formulation, the sum of $\mathbf{P}$ is unity, and the sum of each column of $\mathbf{A}$ is zero (ensuring that probability mass that leaves one state enters another state).

In the FSP analysis, one specifies a maximum biomolecule content N_P , resulting in a total of $N_T = N_S(N_P + 1)$ possible states a cell can exhibit,

$$\mathbf{P}^{\text{FSP}}(t) = [P_{1,0}^{\text{FSP}}(t), \dots, P_{N_S, N_P+1}^{\text{FSP}}(t)]^T. \quad (3)$$

The resulting finite dimensional differential equation $\frac{d}{dt}\mathbf{P}^{\text{FSP}} = \mathbf{A}^{\text{FSP}}(t)\mathbf{P}^{\text{FSP}}(t)$ is integrated over time to give the distribution over $\mathbf{P}^{\text{FSP}}$ at some time t . The particular advantage of the FSP approach is that it can directly and systematically compare stochastic biochemical models to the type of single-cell data that is readily measured using flow

cytometry or fluorescence microscopy [25], and as a result the FSP has seen substantial success in the fitting and predicting of single-cell molecular distributions in bacterial [26, 27], yeast [28], and human [29] cells. However, in each of these cases, the population size has been assumed to be constant, and little attention has been given to adapt the FSP approach to explore the coupling of population growth and discrete stochastic molecular dynamics.

The basic FSP approach, as described above, tracks the probability distributions for a single cell, and this probability is implicitly assumed to be independent of the overall population dynamics. As such, the processes of population growth and decay pose a challenge to the FSP analysis. For example, in mitosis events, one cell divides to produce two daughter cells, and in cell death events, one cell is removed from the population. In these cases, the probability distribution of molecules over the population as a whole is affected by each individual cell's dynamics. These effects are analogous to non-local interactions in a PDE analysis, as discussed in [21] for continuous valued systems and deterministic partitioning of proteins during division. In the next section, we introduce a new method to extend the FSP approach to circumvent this obstacle when dealing with a discrete-valued, stochastic partitioning process.

2.2. The Population-Based Finite State Projection Method

A possible solution to the problem of stochastic molecular dynamics coupled with changes in cell population size is to describe the population size with a continuous- valued mean-field variable, $Q(t)$. This assumption can be justified when the population is sufficiently large for the mean field approximation to be valid or when the population is artificially kept at a fixed size through continual removal or addition of cells at random to maintain the same overall number.

To introduce our approach, we first define some convenient notation to describe the cell cycle, cell birth, and cell death dynamics. To describe the cell cycle, we adopt a simplified four-stage model including separate stages for early cell growth (G1), DNA synthesis (S), late cell growth (G2) and mitosis (M) as depicted in Figure 1. We assume that the time spent in each stage, $s \in \{G1, S, G2, M\}$, is represented by an exponentially- distributed waiting time. However, we note that more complex waiting times could be achieved with inclusion of additional substages within G1, S, G2 or M. We consider dynamics for a single important biomolecule (e.g., DNA, RNA, or protein), whose quantity is denoted by $n \in \{0, 1, 2, \dots\}$. Extension to multiple bio-molecular species is straightforward, yet often computationally expensive. The overall state of any given cell is described by the combination of the cell-cycle stage and the bio-molecular count, $\mathbf{x} = [s, n]$. These discrete states can be enumerated as $\mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$, where the j^{th} state is given by $\mathbf{x}_j = (S_j, n_j)$. Rates for transition from one cell-cycle stage, s , to the next may depend upon the biomolecule count, n , and is denoted by $k_s(n)$. Production and degradation of biomolecules are functions of the cell-cycle stage and the number of biomolecules and are given by the rates $\beta_s(n)$ and $\delta_s(n)$, respectively. In practice, the rates $k_s(n)$, $\beta_s(n)$, and $\delta_s(n)$ may be assumed to be linear or nonlinear functions of n .

Cell birth is described by the flux through mitosis for a single state, (M, n) . The flux from this state is the product of the probability of being in that state, $P_{M, n}$, and the rate of mitosis for cells in that state, $k_M(n)$. At division, the bio-molecular content is distributed among daughter cells according to the distribution $g_0(n, n^*)$, where n is the biomolecule content of the parent and n^* that of the daughter. For random and symmetric partitioning of biomolecules among daughter cells, $g_0(n, n^*)$ is defined by a binomial distribution.

With the above definitions of the system reactions, we can form a non-conservative infinitesimal generator, $\mathbf{A}$, to describe the transitions between all possible cellular states. We assume a model with the four cell-cycle stages $\{s_0, s_1, s_2, s_3\} = \{G1, S, G2, M\}$. To enumerate all possible cell states according to $j = 1, 2, \dots$, we define the number of

biomolecules as the quotient $n_j = \left\lfloor \frac{j-1}{N_S} \right\rfloor$ and the cell-cycle stage as $s_j = (j-1) \bmod(N_S)$, where N_S is the number of cell-cycle stages. For a general model, transitions out of state $\mathbf{x}_j$ include: biomolecule creation with rate $\beta_{sj}(n_j)$, biomolecule decay with rate $\delta_{sj}(n_j)$, transition through the cell-cycle stages with rate $k_{sj}(n_j)$, and cell death with rate $z_{sj}(n_j)$. These reaction rates can all be combined to form a non-conservative infinitesimal generator as:

$$A_{ij} = \begin{cases} -k_{sj}(n_j) - \beta_{sj}(n_j) - \delta_{sj}(n_j) - z_{sj}(n_j) & i=j \\ k_{sj}(n_j) & i=j+1, s_j \neq M \\ 2k_M(n_j)g_0(n_j, n_i) & s_j=M, s_i=G1, n_i \leq n_j \\ \beta_{sj}(n_j) & i=j+N_S \\ \delta_{sj}(n_j) & i=j-N_S \end{cases} \quad (4)$$

An illustration of the construction of the $\mathbf{A}$ matrix is shown in the next subsection.

Our formulation of $\mathbf{A}$ differs from the standard CME formulation in two regards. First, the cell birth reactions are nonconservative in that the flux leaving a single mother cell in stage M is doubled before arriving as two daughter cells in stage G1. The average population size change per cell due to these mitosis events (i.e., the sum of fluxes through mitosis over all of the states) can be written as

$$\Delta_{\text{birth}}(\mathbf{P}) = \sum_n k_M(n) P_{M,n}. \quad (5)$$

Second, cell death causes a net decrease in the overall population, and the cell death flux, $z_{sj}(n_j)$, also breaks the conservation of probability. The average population change per cell due to death can be written as

$$\Delta_{\text{death}}(\mathbf{P}) = \sum_j z_{sj}(n_j) P_{sj,nj}. \quad (6)$$

Taken together, the birth and death of cells results in a net mean population flux per cell of $\Delta_{\text{net}}(\mathbf{P}) = \text{birth}(\mathbf{P}) - \text{death}(\mathbf{P})$. Although the populations of cells in specific cell states change as a result of these reactions, the total probability mass for all cell states must be conserved. As a result, cell birth and cell death reactions can concentrate or dilute the probability density for all states, not just those involved in these particular reactions. To achieve a net balance, probability mass is drawn from or added to all states, with rates dependent upon each state's instantaneous probability. To accomplish this goal, the differential equation for state transitions is modified with an additive flux term to give the nonlinear, but conservative master equation,

$$\frac{d}{dt}\mathbf{P} = \mathbf{A}\mathbf{P} - \Delta_{\text{net}}(\mathbf{P})\mathbf{P}, \quad (7)$$

which we call the population CME. Like the original CME, this formulation can be truncated to yield a finite dimensional ODE known as the population FSP (pFSP).

$$\frac{d}{dt}\mathbf{P}^{\text{FSP}} = \mathbf{A}^{\text{FSP}}\mathbf{P}^{\text{FSP}} - \Delta_{\text{net}}(\mathbf{P}^{\text{FSP}})\mathbf{P}^{\text{FSP}}, \quad (8)$$

Finally, the mean total population number, $Q(t)$, can be estimated via integrating over the total fluxes through death and division, as follows:

$$\frac{dQ(t)}{dt} = \Delta_{\text{net}}(\mathbf{P}^{\text{FSP}})Q(t) \quad (9)$$

In the following section, we will illustrate the construction of the pFSP approach for a four-stage cell cycle model.

2.2.1. Four-stage cell-cycle model—To illustrate the pFSP approach, we consider a four-stage, biomolecule-dependent cell cycle. The full $\mathbf{A}$ matrix for the four-stage system shown in Figure 1, where the cell-cycle stages are $(s_0, s_1, s_2, s_3) = \{\text{G1}, \text{S}, \text{G2}, \text{M}\}$, is given by:

$$\mathbf{A}^{\text{FSP}} = \begin{pmatrix} A_0 - B_0 & A_{0,1} + D_1 & \cdots & A_{0,N_P} \\ B_0 & A_1 - B_1 - D_1 & \cdots & A_{1,N_P} \\ 0 & B_1 & \cdots & A_{2,N_P} \\ 0 & 0 & \cdots & A_{3,N_P} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{N_P} - B_{N_P} - D_{N_P} \end{pmatrix}. \quad (10)$$

In this formulation A_n , B_n , D_n and A_{n*} , n are 4×4 matrices defined as follows:

$$B_n = \begin{pmatrix} \beta_0(n) & 0 & 0 & 0 \\ 0 & \beta_1(n) & 0 & 0 \\ 0 & 0 & \beta_2(n) & 0 \\ 0 & 0 & 0 & \beta_3(n) \end{pmatrix} \quad (11)$$

is the matrix of biomolecule creation rates;

$$D_n = \begin{pmatrix} \delta_0(n) & 0 & 0 & 0 \\ 0 & \delta_1(n) & 0 & 0 \\ 0 & 0 & \delta_2(n) & 0 \\ 0 & 0 & 0 & \delta_3(n) \end{pmatrix} \quad (12)$$

is the matrix of biomolecule decay rates;

$$A_n = \begin{pmatrix} -k_0(n) - z_0(n) & 0 & 0 & k_3(n)g_0(n, n) \\ k_0(n) & -k_1(n) - z_1(n) & 0 & 0 \\ 0 & k_1(n) & -k_2(n) - z_2(n) & 0 \\ 0 & 0 & k_2(n) & -k_3(n) - z_3(n) \end{pmatrix} \quad (13)$$

is the matrix of cell cycle progression and cell death rates for exactly n biomolecules; and

$$A_{n*,n} = \begin{pmatrix} 0 & 0 & 0 & 2k_3(n)g_0(n, n*) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (14)$$

gives the transition rates to daughter cells at mitosis.

The net flux of probability due to birth or death is computed from Eqs. 5 and 6 to give

$$\Delta_{\text{Net}}(\mathbf{P}^{\text{FSP}}) = \sum_{n=0}^{N_p} k_3(n)P_{3,n}^{\text{FSP}} - \sum_{j=1}^{N_s(N_{p+1})} z_{sj}(n_j)P_{sj,nj}^{\text{FSP}}, \quad (15)$$

and the final pFSP formulation can be written:

$$\begin{aligned}\frac{d}{dt}\mathbf{P}^{\text{FSP}}(t) &= \mathbf{A}^{\text{FSP}}\mathbf{P}^{\text{FSP}} - \Delta_{\text{net}}(\mathbf{P}^{\text{FSP}})\mathbf{P}^{\text{FSP}}, \\ \frac{d}{dt}Q(t) &= \Delta_{\text{net}}(\mathbf{P}^{\text{FSP}})Q(t).\end{aligned}\quad (16)$$

This formulation can now be integrated in time using standard routines. In the examples below, we use Matlab's ode23s routine.

3. Results

We illustrate the method with two four-stage cell-cycle models: one in which single-cell fluctuations of the protein p21 determines cell-cycle stage progression, and one in which a hypothetical chemical therapy causes cells to accumulate DNA damage in S stage and potentially to die in G2 stage. The results of the models are compared with those of a population-tracking stochastic simulation algorithm (SSA, [30]) and a simplified ODE analysis, which are described in Appendix A and Appendix B, respectively.

3.1. Protein dynamics

We first explore a model of protein p21 as a variable upon which cell-cycle progression depends, following [31]. For this analysis, the number of p21 proteins in a cell is described by the variable n , and the matrices in Equations 11 and 12 correspond to protein creation and degradation, respectively. In this model, the rate of p21 production is assumed to depend upon the cell-cycle stage, with high p21 expression in G1 and M stages ($\beta = 80 \text{ hr}^{-1}$), moderate p21 expression in G2 stage ($\beta = 40 \text{ hr}^{-1}$), and low p21 expression in S stage ($\beta = 20 \text{ hr}^{-1}$). p21 degradation is assumed to be first order in p21 for cell-cycle stages S, G2, and M. In cell-cycle stage G1, p21 degradation is assumed to be a saturated function of CDK2

activity according to the degradation rate $\delta(n) = n \left(1 + \frac{8C_G^4}{1+C_G^4} \right) \text{hr}^{-1}$, and the CDK2 level is assumed to be algebraically related to the p21 concentration, n , according to the formula [31]:

$$C_G(n) = \frac{1.75}{\frac{10n^4}{(35^4+n^4)} + 0.75}.$$

The level of p21 imposes a checkpoint on the cell-cycle progression from stage G1 to S according to the nonlinear repression function

$$k_{G1}(n) = \frac{35^2}{35^2 + n^2} \text{hr}^{-1},$$

such that cells containing high levels of p21 (e.g., $n \gg 35$) will be much slower to progress than those with lower levels of p21. All other cell-cycle progression rates are assumed to be

independent of the p21 level. In this model, we assume that the cell death rate is negligible. Table 1 summarizes the full parametrization of this model.

Figure 2 shows the single-cell distributions for the cell-cycle stages and p21 content as predicted by the above described model using the proposed pFSP approach (black) and compared to the results of SSA simulations (blue). From the comparison, it is clear that the pFSP approach accurately captures the probability distribution for the cell-cycle stages (Fig. 2A), the overall p21 population distribution (Fig. 2B) and for the joint distribution of cell-cycle stage and p21 content (Fig. 2C). Figure 3 shows the comparison of the pFSP (Fig. 3A) and SSA (Fig. 3B) analyses for the number of cells over time during the cell growth process.

3.2. DNA Damage and Cell Death Dynamics

As a second illustration of the pFSP approach, we propose a model of drug-induced DNA damage, DNA damage repair, and damage-induced cell mortality, in which these processes are all dependent upon the cell-cycle stage. In this model, the variable n is used to denote the discrete number of drug-induced DNA replication errors. These DNA damage events are assumed to occur during DNA replication only in stage S with the saturated rate

$$\beta_s(D) = 100 \frac{D^{10}}{2^{10} + D^{10}} \text{hr}^{-1},$$

where $D(t)$ is the amount of drug in the system at time t . The DNA damage is assumed to be repaired in all stages as a first order process with constant rate, $\delta(n) = n \text{hr}^{-1}$. Cells die in stage G2 with a rate that depends upon the extent of DNA damage according to the saturated rate

$$z_{G2}(n) = 999 \frac{n^{10}}{40^{10} + n^{10}} \text{hr}^{-1}.$$

We assume that DNA damage instances are independent and that, at division, each daughter cell inherits an equal amount of DNA damage from the parent, i.e. $g_0(n, n^*) = 1$ for $n^* = n$ and zero otherwise. To model the drug concentration profile, we assume a standard one-compartment pharmacokinetic model, and we assume that drug is given in a series of pulses of effective dosage concentrations $\{D_i\}$ at times $\{t_i\}$. Assuming an effective drug clearance rate of $k_e = 1 \text{hr}^{-1}$, the total drug concentration at any time is given by:

$$D(t) = \sum_i D_i \exp(-k_e(t - t_i)),$$

where the summation is over all i such that $t_i \leq t$. All cell-cycle progression rates are assumed to have rates $k_s(n) = 1 \text{hr}^{-1}$. The complete parametrization of the model is given in Table 2.

Figure 4 shows the computed distributions of DNA damage at a time of 5 hr for a particular drug dosage profile where $D_i = 4 \mu\text{g/mL}$ was applied at each time $t_i \in \{0, 2, 4\}$ hr. Figure 4A shows the joint distributions of the cell-cycle stage and the damage content, and Figure 4B shows the total DNA damage. Figures 4C and 4D show the corresponding cumulative distributions for the same conditions. From the figure, it is clear that the pFSP analysis matches closely to the SSA results. For further verification of the pFSP, we explored eight different dosage strategies, each with a different number of drug pulses or different durations over which the drug pulses were given. For each case, we computed the average number of cells at a time of 10 hours using the pFSP, the SSA, and an ODE analysis with the same mechanisms and parameters. It is clear from this comparison that the pFSP approach accurately captures the mean level of cells at the final time. Moreover, the pFSP outperforms the ODE analysis, especially for situations involving many rapid pulses over a short time period. In these circumstances, the ODE model predicts near complete eradication of the population, whereas the pFSP and SSA analyses show that population heterogeneity can lead to a considerable level of population survival. Because population expansion dramatically increases the computational effort needed to compute the full SSA trajectories, we will use the pFSP approach for all subsequent analyses.

Having verified the accuracy of the pFSP approach, we next explore how different drug treatment regimens might affect the efficacy of a treatment. To enable a fair comparison, we explore different regimens in which a total of $100 \mu\text{g/mL}$ is applied, but this application is assumed to be given over different treatment lengths from zero to 100 hrs and in the form of between 1 and 50 equal dosages. For each combination of treatment length and number of pulses, we computed the number of cells in the final population after 100 hours. Figure 5 shows a contour plot of the final cell count versus the number of pulses and the total length of the drug treatment. If the population level falls below one cell at any time, we identify that population as having been eradicated. For this model, population eradication is possible (see the dark blue region of Figure 5), but only if the drug is given at a sufficiently high quantity over a sufficiently long time. If the treatment is spread over too long a time (right side of Fig. 5), or if the drug is given too quickly (left or bottom regions of Fig. 5), the treatment fails.

To further illustrate the importance of the specific treatment protocol, Figure 6 shows trajectories for the drug concentration, DNA damage levels, population count, and cell stages for two treatment regimens. These specific regimens correspond to the circles in Figure 5; both correspond to the same total application and temporal average of the drug, and to the same number of drug pulses, but one regimen results in eradication and the other leads to population explosion. The main difference is that during the longer treatment period, the drug periodically drops below the efficacy threshold. Cell populations that manage to redistribute their cell cycles to be out of phase with the treatment frequency (see Fig. 6 right) can pass through the S cell-cycle stage at low drug periods and proceed with relatively little harm. Similarly, if the overall treatment period were too short, the drug may be cleared before random cells with slow cycles complete their cycles to reach the S stage.

4. Discussion and Conclusion

We have presented a new methodology to model population growth and death, where the dynamics of single cells depend upon a coupling of stochastic biomolecule dynamics and cell-cycle progression. This approach uses an extension of the FSP method [5], which captures the discrete stochastic nature of biomolecule dynamics and allows for complicated behaviors such as non-Gaussian distributions, bimodality, and random synchronization or desynchronization, which are typically averaged away to the detriment of mean-valued ODE models. The methodology requires solving only a sparse system of ODEs, which can be considerably more efficient than performing large numbers of coupled simulations. Moreover, the method directly provides the precise probability distributions of cell responses, which can be used directly to compute the likelihood that observed experimental heterogeneity data is from a proposed combination of model and parameters [25, 28, 29]. For these reasons, it is expected that the pFSP will become a valuable tool to fit and predict experimentally measured single-cell heterogeneity in fluctuating population sizes.

Although the pFSP methods presented here allow for coupling of stochastic molecular reaction and cell-cycle progression and their influence on population growth and death, the method is currently limited to two cases where (i) the single-cell dynamics are independent of population size as in Eq. 16, or (ii) where the population level follows a deterministic and continuous value, $Q(t)$. In the latter case, the overall process can be written in the form:

$$\begin{aligned}\frac{d}{dt}\mathbf{P}(t) &= \mathbf{f}_1(\mathbf{P}(t), Q(t)), \\ \frac{d}{dt}Q(t) &= \Delta_{\text{net}}(\mathbf{P})Q(t),\end{aligned}\quad (17)$$

and the *average* cell population level influences the microscopic cellular dynamics and vice versa. However, in its current form, the pFSP approach is not yet able to examine the distribution of the population level itself as an additional stochastic variable, which has some distribution described by $\mathbf{P}_Q$. Such effects might be observed in a homeostatic system [32] or a population in which local cell density impacts growth rates, through higher-order organization [33], quorum sensing [34], or space or resource restriction [35]. It remains to be seen in future work if the formalism of the FSP approach can be extended, potentially through the application of hybrid or switching methods, to handle models and parameter regimes where the population size statistics are poorly described by its average.

Despite the above limitation, we have shown that the pFSP successfully captures the quantitative behavior of population growth under the influence of microscopic cellular dynamics and coupled to time-varying environmental inputs. We have illustrated how the pFSP approach can be used to explore different dynamics of cell-cycle progression, cell death, and intracellular biochemical reactions and how these mechanisms can influence population size dynamics. Our analyses revealed that the interplay of stochastic behavior at

the single-cell level can have substantial and non-intuitive impacts on the average behavior of a population. In particular, by coupling cell population dynamics to a simple one-compartment pharmacokinetic model with temporally varying drug inputs, we showed that stochastic single-cell dynamics and cell growth or death can have substantial implications for drug resistance. Specifically, we showed that stochastic single-cell dynamics could enable cells to evade treatments in manners that cannot be captured by deterministic models. Analyses with the pFSP approach could be used with more complex pharmacokinetic models to identify optimal treatment regimens to maximize efficacy while minimizing adverse effects, both of which could be affected by cellular heterogeneity. Similar coupling of single-cell dynamics with population growth could eventually allow for fuller pictures of the complex multi-scale dynamics that are omnipresent in the analysis of evolutionary fitness, the exploration of organismal development, and the systematic design of medical treatments.

Acknowledgments

This study began as a project at the 2015 q-bio Summer School (qbSS, Fort Collins) and was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R25GM105608. RJ was supported by a BBSRC PhD studentship and the LoLa grant, number BB/N003608/1. BM was supported under the W. M. Keck Foundation and DTRA FRCALL 12-3-2-0002.

Appendix A: Stochastic simulation

To validate the proposed pFSP method, we compare its results to many stochastic simulations. We implement the stochastic simulation algorithm (SSA, [30]) for a biomolecule-dependent cell cycle by simulating one cell at a time. For the four-stage system, a cell has seven possible state transitions: the four cell-cycle stage transitions (of which only one at a time can be realizable), biomolecule creation, biomolecule degradation, and cell death. When a cell exits from M to G1, its biomolecule content is chosen via the binomial distribution and the simulation continues. Then the sister cell is also simulated, starting at the time of the mitosis event, and with the complement biomolecule content. This process is illustrated in Figure A1. Comparison of the method with the SSA is shown in Figures 4 and 6 and in Table 3.

Figure A1.

Stochastic simulation of the four-stage model. A: results of a simulation that started with two cells, where each line represents a cell. Cells go through four cell-cycle stages, G1, S, G2, and M. Upon division, a new cell is started from that time point with the appropriate biomolecule complement. E.g., at time $t \approx 5.5$, cell 1 exits mitosis, and at the same time, cell 3 begins in G1. At the end, there were 36 cells. B: biomolecule-content trajectory of one cell, with cell-cycle stage shown below.

Appendix B: ODE model

The ODE implementation for the four-stage system consists of eight ODEs: one for each of the stages and one for the DNA damage count for each stage. Variables $G1$, S , $G2$ and M count the number of cells in each cell-cycle stage, and variables n_{G1} , n_S , n_{G2} and n_M the number of total damage events in all cells in the respective stages. We count the total cell number Q as the sum of all cells in each state: $Q = G1 + S + G2 + M$.

The rate equations for the stages are:

$$\frac{dG1}{dt} = 2k_M M - k_{G1} G1, \quad (\text{B.1})$$

$$\frac{dS}{dt} = k_{G1} G1 - k_S S, \quad (\text{B.2})$$

$$\frac{dG2}{dt} = k_S S - k_{G2} G2 - z_{G2} \left(\frac{n_{G2}}{G2} \right) G2, \quad (\text{B.3})$$

$$\frac{dM}{dt} = k_{G2} G2 - k_M M. \quad (\text{B.4})$$

Cell-stage transition rates are linear in the variable that transitions as in Section 3.2, where $k_i(n) = 1$ for all i and all n . Cells die only in stage G2, at a rate, $z_{G2}(n_{G2}/G2)$, dependent on the average amount of DNA damage in each G2 cell. At division, one M cell gives rise to two G1 cells.

The rate equations for the total number of damages in each cell stage are:

$$\frac{dn_{G1}}{dt} = 2k_M n_M - k_{G1} n_{G1} - \delta n_{G1}, \quad (\text{B.5})$$

$$\frac{dn_S}{dt} = k_{G1} n_{G1} - k_S n_S - \delta n_S + \beta_S(D) S, \quad (\text{B.6})$$

$$\frac{dn_{G2}}{dt} = k_S n_S - k_{G2} n_{G2} - \delta n_{G2} - z_{G2} \left(\frac{n_{G2}}{G2} \right) n_{G2}, \quad (\text{B.7})$$

$$\frac{dn_M}{dt} = k_{G2} n_{G2} - k_M n_M - \delta n_M. \quad (\text{B.8})$$

The rates are as defined in Section 3.2. Instances of damage accumulate through damage events, $\beta_S(D)$, that occur in a drug-dependent fashion in stage S only. They diminish due to DNA repair events that occur in all stages, at rate δn_i , where n_i is the total number of damages in all cells in stage i . They diminish in stage G2 due to cell death.

Instances of DNA damage transition from stage to stage along with the cells themselves, i.e. cells ‘carry’ their damages. For example, for n_{G1} damages in G1 cells in stage G1, there are $n_{G1}/G1$ damages on average in each G1 cell. Cells in stage G1 transition to stage S at rate $G1$, so damages pass from stage G1 to stage S at a rate $G1 \cdot n_{G1}/G1 = n_{G1}$. The same reasoning holds for cell death. Finally, damages are duplicated at division.

References

1. Shahrezaei V, Marguerat S. Current Opinion in Microbiology. 2015; 25:127–135. [PubMed: 26093364]
2. The Review on Antimicrobial Resistance. Tech rep. The Review on Antimicrobial Resistance; 2016. Tackling drug-resistant infections globally: final report and recommendations.
3. Allen R, Waclaw B. Physical biology. 2016; 13:045001. [PubMed: 27510596]
4. Tadrowski AC, Evans MR, Waclaw B. arXiv. 2016 (*Preprint* 1601.04600v2).
5. Munsky B, Khammash M. The Journal of Chemical Physics. 2006; 124:044104. [PubMed: 16460146]
6. Villasana M, Radunskaya A. Journal of Mathematical Biology. 2003; 47:270–294. [PubMed: 12955460]
7. Taib Z. Communications in Statistics. Stochastic Models. 1999; 15:719–729.
8. Alexandersson M. Journal of Applied Probability. 2001; 38:685–695.
9. Nordon RE, Ko KH, Odell R, Schroeder T. Journal of Theoretical Biology. 2011; 277:7–18. [PubMed: 21333658]
10. Charlebois DA, Kaern M. Communications in Computational Physics. 2013; 14:461–476.
11. Murphy, JT., Walshe, R. Modeling Antibiotic Resistance in Bacterial Colonies Using Agent-Based Approach. In: Dubitzky, W. Southgate, J., Fuß, H., editors. Understanding the Dynamics of Biological Systems. New York, NY: Springer New York; 2011. p. 131–154.
12. Nevozhay D, Adams RM, Van Itallie E, Bennett MR, Balázsi G. PLoS Computational Biology. 2012; 8:e1002480. [PubMed: 22511863]
13. Schwabe A, Bruggeman FJ. Biophysical Journal. 2014; 107:301–313. [PubMed: 25028872]
14. Altinok A, Gonze D, Lévi F, Goldbeter A. Interface Focus. 2011; 1:36–47. [PubMed: 22419973]
15. Banasiak J, Pichor K, Rudnicki R. Acta Applicandae Mathematicae. 2012; 119:149–166.
16. Dyson J, Vilella-Bressan R, Webb GF. Mathematical Biosciences. 2002; 177–178. 73–83.
17. Ellermeyer SF, Pilyugin SS. Journal of Biological Dynamics. 2012; 6:131–147. [PubMed: 22873584]
18. Massie TM, Ryabov A, Blasius B, Weithoff G, Gaedke U. The American Naturalist. 2013; 182:103–119.
19. Basse B, Baguley BC, Marshall ES, Joseph WR, van Brunt B, Wake G, Wall DJ. Journal of Mathematical Biology. 47(2003):295–312. [PubMed: 14523574]
20. Luzyanina T, Roose D, Schenkel T, Sester M, Ehl S, Meyerhans A, Bocharov G. Theoretical Biology & Medical Modelling. 2007; 4:26. [PubMed: 17650320]
21. Stromberg SP, Antia R, Nemenman I. Physical Biology. 2013; 10:035010. [PubMed: 23735782]
22. Bekkal Brikci F, Clairambault J, Ribba B, Perthame B. Journal of Mathematical Biology. 2007; 57:91–110. [PubMed: 18064465]
23. Borges R, Calsina À, Cuadrado S, Diekmann O. Journal of Evolution Equations. 2014; 14:841–862.
24. McQuarrie DA. Journal of Applied Probability. 1967; 4:413.
25. Fox Z, Neuert G, Munsky B. The Journal of Chemical Physics. 2016; 145:074101. [PubMed: 27544081]
26. Munsky B, Trinh B, Khammash M. Molecular Systems Biology. 2009; 5:318. [PubMed: 19888213]
27. Lou C, Stanton B, Chen YJ, Munsky B, Voigt CA. Nature Biotechnology. 2012; 30:1137–1142.
28. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, van Oudenaarden A. Science. 2013; 339:584–587. [PubMed: 23372015]
29. Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, Mueller F, Darzacq X. Cell Reports. 2014; 8:75–83. [PubMed: 24981864]
30. Gillespie DT. J Phys Chem. 1977; 81:2340–2361.
31. Overton KW, Spencer SL, Noderer WL, Meyer T, Wang CL. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:E4386–93. [PubMed: 25267623]

32. Foley C, Bernard S, Mackey MC. *Journal of Theoretical Biology*. 2006; 238:754–763. [PubMed: 16115650]
33. Volfson D, Cookson S, Hasty J, Tsimring LS. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:15346–15351. [PubMed: 18832176]
34. Whitehead NA, Barnard AML, Slater H, Simpson NJL, George PC, Salmond. *FEMS Microbiology Reviews*. 2001; 25:365–404. [PubMed: 11524130]
35. Shraiman BI. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:3318–3323. [PubMed: 15728365]

Figure 1. Four-stage cell-cycle model

Cells begin in stage G1, progress through stages S and G2 to M, and divide, giving rise to two cells in stage G1. Cells can die in any cell cycle stage, possibly with varying rates.

Figure 2.

Protein distributions in each cell-cycle stage resulting of the four- stage model. (A) Cell-cycle stage distribution. (B) Overall protein distribution. (C) Protein distribution in each cell-cycle stage. Black dashed: FSP results. Blue: SSA results, starting with 40 cells and ending at time $t = 30$ hr with 7,156 cells.

Figure 3. The evolution of cell number over time for the four-stage model. (A) FSP model. (B) SSA, starting with 40 cells.

Figure 4.

DNA damage distributions in each cell-cycle stage resulting of the four-stage DNA-damage model. A) Joint probability densities for protein distribution and cell-cycle stage. B) Overall protein probability density. Panels C and D show the cumulative distributions corresponding to A and B, respectively. The drug dosage regimen corresponds to $4 \mu\text{g/mL}$ pulses of drug applied at $t \in (0, 2, 4)$ hr, and results are shown for the time $t = 5$ hr. The FSP results are shown in black dashed lines, and SSA results starting with 1,000 cells are shown in blue with an additional run shown in orange for the cumulative distribution.

Figure 5.

Contour plot of final population number from the DNA damage pFSP model at time $t = 100$ hours for varying numbers of pulses of drug and duration of treatment. The same total level of drug has been applied in all cases. Circles indicate points displayed individually in Figure 6.

Figure 6.

Individual trajectories over time for points indicated in Figure 5. Top: 25 pulses over 50 hours. Bottom: 25 pulses over 25 hours. From left to right, trajectories show: the drug profile, the average amount of damage per cell, the number of cells in the population, and the proportion of cells in each state.

Table 1

Functional parametrization of p21-dependent cell-cycle model.

Equation	Units	Meaning
$k(n) = \left\{ \frac{35^2}{35^2 + n^2}, 1, 1, 1 \right\}$	cell hr ⁻¹	Cell-stage transition rate
$g_0(n, n^*) = \binom{n}{n^*} \frac{1}{2^n}$	–	p21 inheritance
$\beta = (80, 20, 40, 80)$	molecules hr ⁻¹	p21 creation rate
$\delta_i(n) = n$ for $i = 0$	molecules hr ⁻¹	p21 decay rate in stages S, G2, M
$\delta_0(n) = n \left(1 + \frac{8C_G^4}{1 + C_G^4} \right)$	molecules hr ⁻¹	p21 decay rate in stage G1
$C_G(n) = \frac{1.75}{\frac{10n^4}{(35^4 + n^4)} + 0.75}$	molecules	CDK2 activity
$\lambda(n) = (0, 0, 0, 0)$	cell hr ⁻¹	Cell death rate

Table 2

Functional parametrization of DNA-damage-dependent cell-cycle model.

Equation	Units	Meaning
$k(n) = \{1, 1, 1, 1\}$	cells hr ⁻¹	Cell-stage transition rate
$\beta = \left\{0, 100 \frac{D^{10}}{2^{10} + D^{10}}, 0, 0\right\}$	molecules hr ⁻¹	DNA damage rate
$\delta(n) = \{n, n, n, n\}$	molecules hr ⁻¹	DNA repair rate
$z(n) = \left\{0, 0, 999 \frac{n^{10}}{n^{10} + 40^{10}}, 0\right\}$	cells hr ⁻¹	Cell death rate

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Comparison of total final population numbers, predicted with the FSP method, stochastic simulations, and the ODE model. The model is the DNA damage model of Section 3.2, the total drug is 10 $\mu\text{g/mL}$ and the length of the simulation is 10 hours. The SSA gives the mean and standard deviation of 100 simulations.

Pulses	Duration, hr	FSP	SSA	ODE
20	1.5	82.6156	83.5800 $\pm$ 19.5955	4.0872
45	1.5	82.7006	83.1700 $\pm$ 18.4473	4.0556
20	3.5	86.5031	86.2300 $\pm$ 16.4070	2.8159
45	3.5	86.9441	89.5100 $\pm$ 16.2434	3.0326
20	6	207.9694	207.3200 $\pm$ 14.3948	10.4704
45	6	208.2892	208.4100 $\pm$ 16.1114	13.3587
20	9	208.4521	208.8600 $\pm$ 16.5450	208.4520
45	9	208.4521	207.3200 $\pm$ 13.3635	208.4521