Cell size statistics in cell lineages and population snapshots with different growth regimes and division strategies

Niccolò Totis, César Nieto, Armin Küper, César Vargas-García, Abhyudai Singh and Steffen Waldherr

Abstract— Growing populations of bacteria control their growth and division reaching narrow distributions of cell-sizes. In this paper we explored how different combinations of growth regimes and division mechanisms lead to different cell-size statistics in these populations. Deterministic and stochastic modeling were used to describe the size distribution of a population of cells that is observed from two different perspectives: as single cell lineages, i.e. random paths in the lineage tree, or as snapshots, at given times, of a population in which all descendants of a single ancestor cell are observed. Our time-dependent approaches allowed us to obtain both the transient dynamics and the steady state values for the main statistical moments of the cell-size distribution. Also, we established mathematical relationships among the statistics in the two considered perspectives, thus improving our knowledge of how cells control their growth and proliferation.

I. INTRODUCTION

Bacteria are able to regulate growth and division in order to maintain their size within a defined range, a state defined as *homeostasis*[1], [2]. Traditionally, experimental studies on cell division observe a cell population from two main alternative perspectives[3]. Either they (i) track *single lineages* as observed in the mother machine[4], either (ii) they take *population snapshots* of descendant cells of a common ancestor, like in flow cytometry[5] experiments. Understanding the differences between these experimental viewpoints [6] has become an expanding area of research[7], [8], [9], [10].

In the present work deterministic and stochastic computational modeling approaches were used to represent these two perspectives and to analyze how the heterogeneity of a cell population with respect to cell size evolves as a result of growth and symmetric division. To observe single lineages, at each cell division the model keeps track of only one newborn cell. Consequently, the total cell number of the considered sub-population remains constant over the simulated time. To observe population snapshots, instead, the model tracks all progenies and thus the evolution of the whole population over time. In our analysis we considered some scenarios of experimental relevance characterized by different growth regimes and division mechanisms, we estimated the dynamics of the main statistical properties of the size distribution and compared the obtained results.

In Section II, we introduce the two modeling frameworks we used to describe size dynamics: the Population Balance Equation (PBE), a population-based deterministic approach, and Monte Carlo simulations, using a modified stochastic simulation algorithm. Furthermore, we present how these were exploited to represent the considered scenarios of growth and division. Details of the analytical derivations as well as of the numerical solutions are provided. In Section III we present how four biologically relevant scenarios characterize the components of the models. The comparison of the results obtained with the different modeling approaches is portrayed with figures and further discussed.

II. METHODS

A. Modeling frameworks

a) Population-based deterministic modeling: the population balance equation (PBE): In general, the framework of PBEs describes the evolution of particulate systems with respect to an extensive property, like size, mass or molecule counts [11]. Here, PBEs allow us to represent the distribution of cell sizes s in the population as a number density function $n(t, s)$ (or NDF) and to study how it changes over time as a result of the processes of cell growth and cell division.

The dynamics of $n(t, s)$ is given by the following integro-partial differential equation nnnnnnnn:

$$\frac{\partial n}{\partial t}(t, s) + \text{div}(f(s)n(t, s)) =$$
$$= -\delta(s)n(t, s) + \int_0^\infty \nu\delta(\zeta)\Phi(s, \zeta)n(t, \zeta)\text{d}\zeta \quad (1)$$

The second term on the left-hand side, the divergence term, describes how the NDF moves through the cell state space. Here, $f(s) = \dot{s}$ describes the single-cell growth rate, ν denotes the number of offsprings generated from each cell division, $\delta(s)$ is the division rate function (in PBE literature, this is often referred to as the breakage rate) and $\Phi(s, \zeta)$ is the kernel function, which defines the probability that the division of a mother cell of size ζ generates a daughter cell of size s. $\Phi(s, \zeta)$ satisfies the property $\int \Phi(s, \zeta)\text{d}s = 1$. In our analysis, these functions are defined as follows:

- ν. We set $\nu = 1$ when, at each division, one single descendant is tracked. We set $\nu = 2$ otherwise.
- $f(s)$. The single-cell growth rate is modeled as $f(s) = gs^a, a \geq 0$. A linear growth regime is obtained by setting $a = 0$, while, for $a = 1$, cells grow exponentially. A general $a > 0$ can be used to explore other division mechanisms[12].

- $\delta(s)$. The division rate is modeled as $\delta(s) = ks^b, b \geq 0$. In this paper, we limit our analysis to $b \in \mathbb{N}$. The division rate is thus constant for $b = 0$, size-dependent otherwise.
- $\Phi(s, \zeta)$. The Dirac-delta function is used here to model an equal repartitioning of volume after cell division, so that $\Phi(s, \zeta) = 2\delta(2s - \zeta)$.

In this way (1) becomes:

$$\frac{\partial n}{\partial t}(t, s) + \frac{\partial (f(s)n(t, s))}{\partial s} = \\ - \delta(s)n(t, s) + 2\nu\delta(2s)n(t, 2s). \quad (2)$$

In our analysis, the evolution of the cell size distribution is described through the changes in its raw moments. Starting from the PBE in (1), we provide here an analytical derivation of the moment dynamics. The ith raw and central moments are denoted as $m_i = \int_0^\infty s^i n(t, s)\mathrm{d}s$ and $cm_i = \int_0^\infty (s - m_1)^i n(t, s)\mathrm{d}s$ respectively. In the following derivations, we assume that all raw moments remain finite at finite times, which implies $\lim_{s \to +\infty} s^i n(t, s) = 0$. While moments $m_i(t)$ are always considered as time-dependent variables, we will make use of m_i instead, so that notations are simplified.

The total number of cells at a given time is equivalent to the zeroth moment m_0 of the NDF:

$$N(t) = \int_0^\infty n(t, s)\mathrm{d}s. \quad (3)$$

Introducing the normalized NDF

$$\bar{n}(t, s) = \frac{n(t, s)}{N(t)} \quad (4)$$

the dynamics of $N(t)$ can be written as

$$\dot{N} = \rho(t)N, \quad (5)$$

with $\rho(t)$, the average specific proliferation rate, defined as

$$\rho(t) = \int_0^\infty \underbrace{-\frac{\partial(f(s)\bar{n}(t, s))}{\partial s}}_{G_\rho} \underbrace{-\delta(s)\bar{n}(t, s)}_{D_\rho} + \\ \underbrace{+2\nu\delta(2s)\bar{n}(t, 2s)}_{B_\rho} \mathrm{d}s. \quad (6)$$

It is important to notice that that ρ differs from the average specific growth rate μ, normally used to compute the dynamics $\dot{X} = X\mu(t)$, where $X = \int_0^\infty xn(t, s)$ is an extensive property, like size, integrated over the whole population. It should be noted that ρ and μ coincide only when $\dot{s} = 0$ is assumed.

The first term in (6), nullifies. This can be shown by using the product rule on the derivative and integration by parts.

$$G_\rho = -\int_0^\infty gas^{a-1}\bar{n}(t, s)\mathrm{d}s \\ - \left(gs^a\bar{n}(t, s)\big|_0^\infty - \int_0^\infty gas^{a-1}\bar{n}(t, s)\mathrm{d}s\right) = 0 \quad (7)$$

We see that the first and last term in (7) cancel out. We notice that the second term is the integrand function of the ath raw moment. This goes to zero at $+\infty$, for our previous assumption, and $\lim_{s \to 0} s^a \bar{n}(t, s) = 0$, as $\bar{n}(t, s)$ is finite $\forall s$. Using the substitution $2s = z$ in the B_ρ term, and noticing that this does not affect the limits of integration, the B_ρ and D_ρ terms can be summed up leaving

$$\rho(t) = (\nu - 1)\int_0^\infty \delta(s)\bar{n}(t, s)\mathrm{d}s = (\nu - 1)km_b. \quad (8)$$

We can already see here that if we choose to model the dynamics of a single lineage, for which $\nu = 1$, a value of $\rho(t) = 0$ is obtained, as expected.

If we rewrite (1) by replacing $n(t, s) = \bar{n}(t, s)N(t)$ and introduce the expression for the kernel function, we obtain

$$\frac{\partial(\bar{n}(t, s)N(t))}{\partial t} = N(t)\frac{\partial\bar{n}}{\partial t}(t, s) + N(t)\rho(t)\bar{n}(t, s) \\ = -N(t)\frac{\partial(f(s)\bar{n}(t, s))}{\partial s} - \delta(s)N(t)\bar{n}(t, s) \\ + 2\nu\delta(2s)N(t)\bar{n}(t, 2s), \quad (9)$$

If we cancel $N(t)$ from both sides, the normalized PBE can be obtained as

$$\frac{\partial\bar{n}}{\partial t}(t, s) = -\rho(t)\bar{n}(t, s) - \frac{\partial(f(s)\bar{n}(t, s))}{\partial s} - \\ - \delta(s)\bar{n}(t, s) + 2\nu\delta(2s)\bar{n}(t, 2s). \quad (10)$$

Taking the ith moment m_i, for $i \geq 1$, of the normalized PBE (10) results in

$$\dot{m}_i(t) = -\rho(t)m_i \underbrace{-\int_0^\infty s^i \frac{\partial(f(s)\bar{n}(t, s))}{\partial s}\mathrm{d}s}_{G_{m_i}} \\ \underbrace{-\int_0^\infty s^i\delta(s)\bar{n}(t, s)\mathrm{d}s}_{D_{m_i}} + \underbrace{\int_0^\infty s^i 2\nu\delta(2s)\bar{n}(t, 2s)\mathrm{d}s}_{B_{m_i}}. \quad (11)$$

Applying the product rule followed by integration by parts, and exploiting our assumptions, as we did in (7), the term G_{m_i} resolves in

$$G_{m_i} = -\int_0^\infty gas^{(i+a-1)}\bar{n}(t, s)\mathrm{d}s - \int_0^\infty gs^{(i+a)}\frac{\partial\bar{n}(t, s)}{\partial s}\mathrm{d}s \\ = -gam_{(i+a-1)} - g\left(s^{i+a}\bar{n}\big|_0^\infty - (i+a)m_{(i+a-1)}\right) \\ = igm_{(i+a-1)}. \quad (12)$$

After the substitution $2s = z$ in B_{m_i}, the terms D_{m_i} and B_{m_i} can be summed up, leaving us with

$$D_{m_i} + B_{m_i} = \left(\frac{\nu}{2^i} - 1\right)\int_0^\infty s^i\delta(s)\bar{n}(t, s)\mathrm{d}s \quad (13)$$

$$= \left(\frac{\nu}{2^i} - 1\right)km_{i+b}. \quad (14)$$

From equations (12) and (14) we notice that, for $a > 1$ or $b > 0$, the moment dynamics are not closed. This dependency

is overcome through the moment closure method [13], for which all central moments higher than cm_i are approximated to zero. Introducing $p = \max((a-1), b)$ we write:

$$cm_{i+\alpha} = \int_0^\infty (s - m_1)^{i+a} \bar{n}(t,s) \mathrm{d}s \approx 0, \quad \text{for } \alpha = 1, ..., p. \quad (15)$$

This approximation offers us algebraic expressions to compute higher moments, and thus the term in (12) and (14). As an example, the approximated expressions for $m_{i+\alpha}$ when $i = 2, 3$ and $p = 1$, are provided here below:

if $(i = 2, p = 1)$, $\qquad m_3 \approx 3m_2 m_1 - 2m_1^3$

if $(i = 3, p = 1)$, $\quad m_4 \approx 4m_3 m_1 - 6m_2 m_1^2 + 3m_1^4$ (16)

Putting all together, (11) becomes:

$$\dot{m}_i = -\rho(t)m_i + igm_{(i+a-1)} + \left(\frac{\nu}{2^i} - 1\right)km_{i+b}. \quad (17)$$

b) Stochastic Simulation Algorithm (SSA): The Stochastic Simulation Algorithm is a numerical approach for generating sample paths of a continuous-time Markov process whose probability distribution evolves through transition rates between its states. For each cell i in the population, the algorithm first computes $P_i(t)$, the probability that cell i divides at time t, by integrating the division rate δ over time (see explanation in Supplementary Information)[12]. Then, these $P_i(t)$ are compared to uniformly generated random numbers r_i, and the next division is implemented only for cell j, with $P_j > r_j$. The algorithm is detailed as follows:

Result: $[s_1(t), \cdots, s_N(t)]$
Define ν; generate $r_1 \sim U(0,1)$;
$t = 0$, $N = 1$, $s_1 = s(0)$, $P_1 = 0$;
while $t < t_f$ **do**
$\quad$ **while** $P_i(t + t') < r_i, \forall i \in \{1, 2, ..., N\}, t' > 0$ **do**
$\quad\quad$ Solve:
$$\frac{ds_i}{dt'} = f(s_i(t + t')),$$
$$\frac{dP_i}{dt'} = \delta(s_i(t + t'))(1 - P_i(t + t'))$$
$\quad$ **end**
$\quad$ $t = t + t'$;
$\quad$ **if** $\nu = 1$ **then**
$\quad\quad$ $s_1(t) = s_1(t)/2$;
$\quad\quad$ $P_1(t) = 0$;
$\quad$ **else**
$\quad\quad$ $N = N + 1$,
$\quad\quad$ Identify $j : P_j(t) > r_j$, $s_N(t) = s_j(t)/2$
$\quad\quad$ $s_j(t) = s_j(t)/2$;
$\quad\quad$ $P_i(t) = 0, \forall i \in \{1, 2, ..., N\}$;
$\quad$ **end**
$\quad$ Generate new $r_i \sim U(0,1), \forall i \in \{1, 2, ..., N\}$;
end
Algorithm 1: Stochastic Simulation Algorithm.

Giving us the sizes $[s_1 \cdots s_N]$ of the cells at the time $t \in (0, t_f)$ for each replica. The single-cell growth rate $f(s)$, the division rate $\delta(s)$ and ν, the number of cells tracked after

division, are used consistently with their definition in the PBE (see (1)). Let L be the number of replicas, $N_l(t)$ the number of cells in replica l at time t, s_i^l the size of cell i in replica l. The mean population number $N(t)$, the statistical mean $\langle s \rangle = m_1$ and variance $\mathrm{var}(s) = m_2 - m_1^2$ of the cell size distribution are computed as follows:

$$N(t) = \frac{1}{L} \sum_{l=1}^L N_l(t); \quad \langle s(t) \rangle \approx \frac{\sum_{l=1}^L \sum_{i=1}^{N_l(t)} s_i^l(t)}{\sum_{l=1}^L N_l(t)};$$

$$\mathrm{var}(s;t) \approx \frac{\sum_{l=1}^L \sum_{i=1}^{N_l(t)} (s_i^l(t) - \langle s(t) \rangle)^2}{\sum_{l=1}^L N_l(t)};$$

$$C_V^2(s;t) = \frac{\mathrm{var}(s;t)}{\langle s(t) \rangle^2} \quad (18)$$

In this work we make use of the squared coefficient of variation $C_V^2(s)$, a dimensionless quantity that allows us to estimate the degree of heterogeneity in the distribution. For ease of notation, in the rest of the paper the $C_V^2(s)$ of the distribution of sizes s is replaced by C_V^2.

III. RESULTS AND DISCUSSION

A. Biological scenarios

For our analysis we define four biologically relevant scenarios, that differ for the cell growth regime and division mechanisms assumed. The single-cell growth rate f is thus defined as either $f(s) = g$ (linear growth), or $f(s) = gs$ (exponential growth). Similarly, the division rate d is considered either constant, $\delta(s) = k$, or linear, $\delta(s) = ks$, with respect to size.

For each scenario, the parameter values for g and k shown in table I were chosen so that the mean doubling time τ satisfies $\tau = 1$, and $\overline{s_b}$, defined the mean size at birth, satisfies $\overline{s_b} = 1$. This $\overline{s_b}$, introduced in previous studies[12] (see Supplementary Information), corresponds to the size that is perfectly doubled when division occurs. Importantly, specific values for both τ and $\overline{s_b}$ can be determined experimentally, so that experimental data, once normalized to these values, can be compared to model predictions.

TABLE I
DESCRIPTION OF THE SCENARIOS DEFINED BY THEIR GROWTH REGIME f
AND DIVISION MECHANISM d.

Sc.	p	a	$f(s)$	$\delta(s)$	ρ
I	0	0	$g = 1$	$k = 1$	$\nu - 1$
II	1	0	$gs = \ln(2)s$	$k = 2\ln(2)$	$\nu - 1$
III	0	1	$g = 1$	$ks = 0.562s$	$(\nu - 1)0.562m_1$
IV	1	1	$gs = \ln(2)s$	$ks = \ln(2)s$	$(\nu - 1)\ln(2)m_1$

B. Computational results

Here, we compare the results of the two approaches presented in II, with respect to the dynamics of the mean size $\langle s \rangle$ and of the coefficient of variation C_V^2. The initial size distribution was defined as a Dirac-Delta with either $m_1(0) = 0.5\overline{s_b}$ or $m_1(0) = 4\overline{s_b}$. These two initial values can be considered as the extremes of a range of biologically realistic cell sizes. All the higher moments have initial

conditions $m_i(0) = (m_1(0))^i$. The initial number of cells in the population is set to $N(0) = 1$.

a) Population Balance Equation: The analytical expressions derived from the general PBE, presented in Section II, are used to compute the dynamics for m_1 and m_2, and either $\nu = 1$ or $\nu = 2$, and reported in table II. For $b = 1$, the moment closure approximation is used and $cm_3 \approx 0$.

The PBE is here solved also numerically, and the moment dynamics are calculated from the dynamics of the full distribution. This offers us a comparison to the results obtained through the analytical derivations. In the literature, different numerical schemes have been proposed to solve PBEs. We exploited the cell average technique, a spatial discretization method that, following the implementation in [14], preserves the first two moments m_0 and m_1 of the distribution.

b) Stochastic Simulation Algorithm (SSA): To study the behavior of single lineages ($\nu = 1$), the dynamics of 20K replicas are simulated. To study the population snapshot perspective ($\nu = 2$), 1K replicas are run, all starting from one cell with the same initial size ($s(0)$). Growth and division rates are defined according to the specific scenarios and the values of g and k. At the predefined time points the mean size $\langle s \rangle$ and the coefficient of variation C_V^2 are computed considering the sizes of all cells across all replicas using expressions (18). To integrate the differential equations in Algorithm 1, we use Euler integration.

Figures 1-4 show the resulting dynamics of mean size $\langle s \rangle$, in units of $\overline{s_b}$, and the variability in terms of C_V^2. In figure 5 we report the comparison of the dynamics of the total number of cells in the simulation.

Fig. 2. Size dynamics for cells growing and dividing in the scenario II ($f = gs = \ln(2)s$, $b = k = 2\ln(2)$). Analytical, Numerical and stochastic simulation (SSA) are compared. (a) Mean size $\langle s \rangle$ dynamics (b) Size variability dynamics (C_V^2). Single lineage is considered ($\nu = 1$) for both (a) and (b). (c) Mean size dynamics (d) Size variability. Population snapshot is considered ($\nu = 2$) or both (c) and (d). Width of the line of SSA represents the 95% confidence interval for 20K cell for $\nu = 1$ and $1K$ population replicas for $\nu = 2$. Two different initial conditions ($m_1(0) = 4$ and $m_1(0) = 0.5$) was considered in all cases.
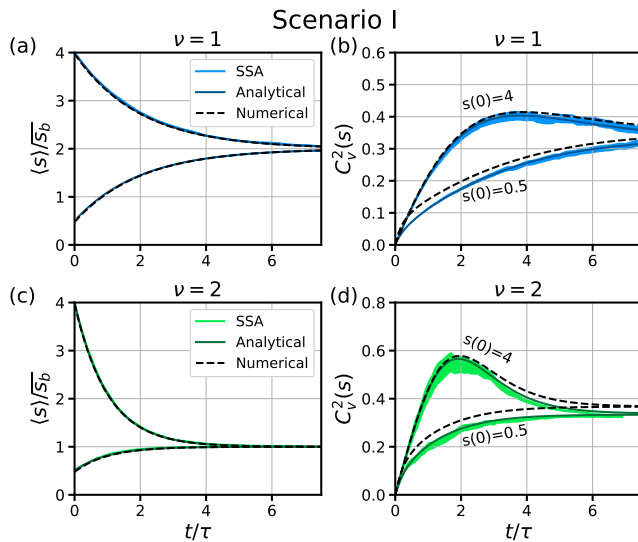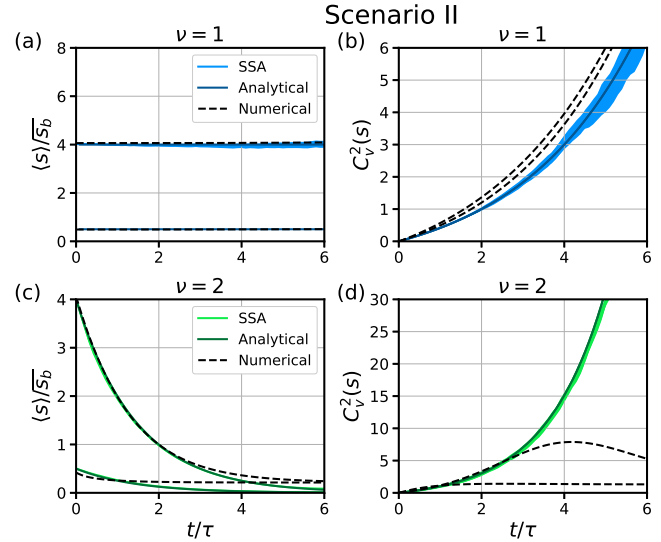

Fig. 1. Size dynamics for cells growing and dividing in the scenario I ($f = g = 1$, $b = k = 1$). Analytical, Numerical and stochastic simulation (SSA) are compared. (a) Mean size $\langle s \rangle$ dynamics (b) Size variability dynamics (C_V^2). Single lineage is considered ($\nu = 1$) for both (a) and (b). (c) Mean size dynamics (d) Size variability. Population snapshot is considered ($\nu = 2$) or both (c) and (d). Width of the line of SSA represents the 95% confidence interval for 20K cell for $\nu = 1$ and $1K$ population replicas for $\nu = 2$. Two different initial conditions ($m_1(0) = 4$ and $m_1(0) = 0.5$) was considered in all cases.
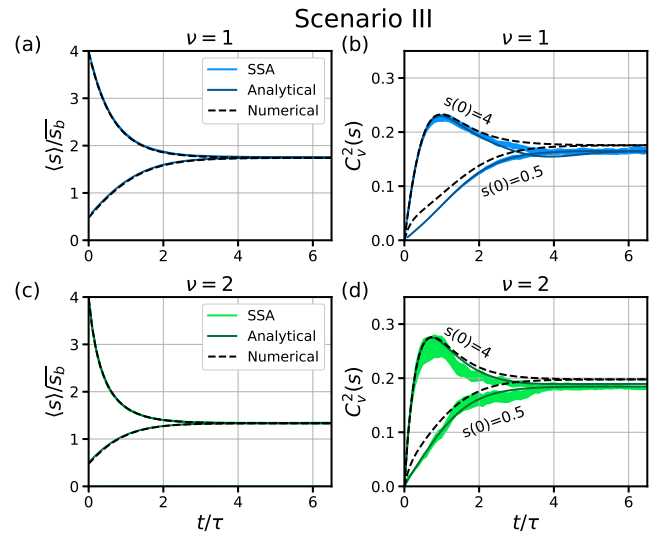

Fig. 3. Size dynamics for cells growing and dividing in the scenario III ($f = g = 1$, $b = ks = 0.562s$). Analytical, Numerical and stochastic simulation (SSA) are compared. (a) Mean size $\langle s \rangle$ dynamics (b) Size variability dynamics (C_V^2). Single lineage is considered ($\nu = 1$) for both (a) and (b). (c) Mean size dynamics (d) Size variability. Population snapshot is considered ($\nu = 2$) or both (c) and (d). Width of the line of SSA represents the 95% confidence interval for 20K cell for $\nu = 1$ and $1K$ population replicas for $\nu = 2$. Two different initial conditions ($m_1(0) = 4$ and $m_1(0) = 0.5$) was considered in all cases.

TABLE II

DYNAMICS OF THE FIRST AND SECOND RAW MOMENTS USING THE MOMENT CLOSURE APPROXIMATION AT THE THIRD MOMENT.

| Sc	$\nu = 1$		$\nu = 2$	
	Moment dynamics	Equilibrium points $[\tilde{m}_1, \tilde{m}_2]$	Moment dynamics	Equilibrium points $[\tilde{m}_1, \tilde{m}_2]$
I	$\dot{m}_1 = g - \frac{1}{2}km_1$ $\dot{m}_2 = 2gm_1 - \frac{3}{4}km_2$	$\left[\frac{2g}{k}, \frac{16g^2}{3k^2}\right]$, stable $\forall g, k > 0$	$\dot{m}_1 = g - km_1$ $\dot{m}_2 = 2gm_1 - \frac{3}{2}km_2$	$\left[\frac{g}{k}, \frac{4g^2}{3k^2}\right]$, stable $\forall g, k > 0$
II	$\dot{m}_1 = gm_1 - \frac{1}{2}km_1$ $\dot{m}_2 = (2g - \frac{3}{4}k)m_2$	$\left[0, 0\right]$, stable if $g < \frac{3k}{8}$	$\dot{m}_1 = (g - k)m_1$ $\dot{m}_2 = (2g - \frac{3}{2}k)m_2$	$\left[0, 0\right]$, stable if $g < \frac{3k}{4}$
III	$\dot{m}_1 = g - \frac{1}{2}km_2$ $\dot{m}_2 = 2gm_1 + \frac{3}{2}km_1^3 - \frac{9}{4}km_1m_2$	$\left[\sqrt{\frac{5g}{3k}}, \frac{2g}{k}\right]$, stable $\forall g, k > 0$ $\left[0, \frac{2g}{k}\right]$, unstable $\forall g, k > 0$	$\dot{m}_1 = g - km_1^2$ $\dot{m}_2 = 2gm_1 + km_1^3 - \frac{5}{2}km_1m_2$	$\left[\sqrt{\frac{g}{k}}, \frac{6g}{5k}\right]$, stable $\forall g, k > 0$
IV	$\dot{m}_1 = gm_1 - \frac{1}{2}km_2$ $\dot{m}_2 = 2gm_2 + \frac{3}{2}km_1^3 - \frac{9}{4}km_1m_2$	$\left[0, 0\right]$, unstable $\forall g, k > 0$	$\dot{m}_1 = gm_1 - km_1^2$ $\dot{m}_2 = 2gm_2 + km_1^3 - \frac{5}{2}km_1m_2$	$\left[\frac{g}{k}, \frac{2g^2}{k^2}\right]$, stable $\forall g, k > 0$ $\left[0, 0\right]$, unstable $\forall g, k > 0$
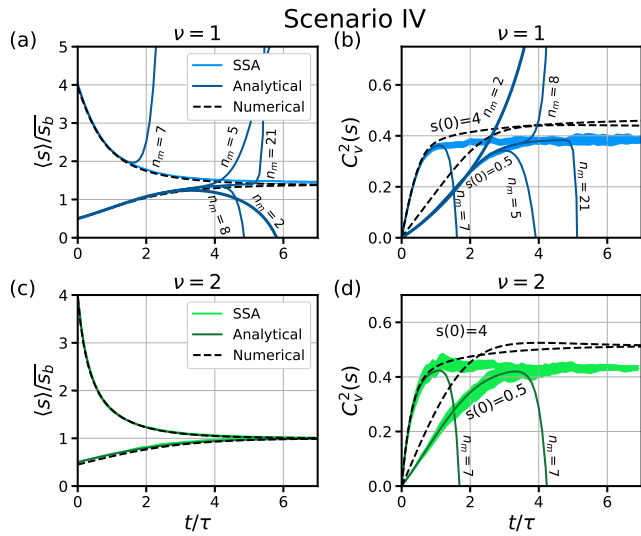
Fig. 4. Size dynamics for cells growing and dividing in the scenario IV ($f = gs = \ln(2)s$, $b = ks = \ln(2)s$). Analytical, Numerical and stochastic simulation (SSA) are compared. ((a) Mean size $\langle s \rangle$ dynamics (b) Size variability dynamics (C_V^2). Single lineage is considered ($\nu = 1$) for both (a) and (b). Different levels (n_m) of ODE truncation using moment closure are shown in (a) and (b). (c) Mean size dynamics (d) Size variability. Population snapshot is considered ($\nu = 2$) or both (c) and (d). Width of the line of SSA represents the 95% confidence interval for 20K cell for $\nu = 1$ and $1K$ population replicas for $\nu = 2$. Two different initial conditions ($m_1(0) = 4$ and $m_1(0) = 0.5$) was considered in all cases.
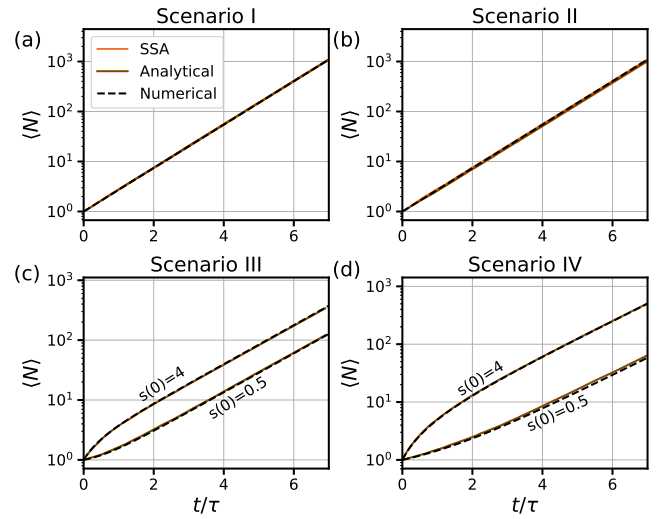

Fig. 5. Total number of particles in the population. Scenarios 1-4 are labeled respectively. Width of the line of SSA represents the 95% confidence interval for $1K$ population replicas.

C. Discussion

Cell size heterogeneity can be quantified across single lineages or taking snapshots of the whole population. This crucial difference is represented, in the equations shown so far, by the value of ν. Depending on the organism and the availability of extracellular nutrients, cells can adopt different growth regimes and division strategies. Considering these, we defined the four scenarios presented in table I.

Figures 1 - 4 show that the four scenarios result in clearly distinct dynamics for $\langle s \rangle$ and C_V^2 . Taking into consideration all three trajectories, we see that in scenarios I, III and IV, a stable steady state is reached. In scenario II, instead, no stable steady state is reached, confirming the analytical results in table II and findings reported in other studies [15]. It is fair

to consider that this scenario is not realistic, as cells with excessively different sizes are generated.

Transient dynamics reveal that, for all scenarios and ν values, the steady state values $\langle \tilde{s} \rangle$ and $\tilde{C}_V^2$ do not depend on the initial conditions.

The comparison between single lineages ($\nu = 1$) and population snapshots ($\nu = 2$), shows us that the computed steady-state values $\langle \tilde{s} \rangle$ are consistently higher for single lineages and lower for population snapshots. As reported in similar studies [6], this is expected, because, independently of the division mechanism, cells that proliferate faster are on average smaller and over-represented in the whole population. Moving one step further this observation, we argue that our computational framework and our results can be used to connect the moment dynamics in the two perspectives we examined. More clearly, if, for instance, one of the scenarios in table I can be considered an accurate representation of an experiment on cell lineages, then, exploiting our framework, the data collected from cell lineages could be used to infer the dynamics of population snapshot. Vice versa, data from

population snapshots could be used to infer the dynamics in single lineages. For scenarios I, II and III, the analytical formulas in table II would make the inference straightforward. For scenario IV, where the analytical dynamics are inaccurate, one would need to first use the numerical approach or the SSA to infer g and k parameters and then, after changing the value of ν, use these to simulate the unknown dynamics.

Despite the clear differences in $\langle \tilde{s} \rangle$, it is important to notice that the steady-state values $\tilde{C}_V^2$, when finite, do not show appreciable differences when different νs are considered. Overall, we can conclude that experimentally, different average sizes, but not significantly different variabilities should be expected for single lineages versus population snapshots.

The use of different modeling approaches and solution schemes offers us an indication on the accuracy of the results: if two trajectories out of three superimpose, their accuracy is reinforced. In scenarios I and II, no moment closure approximation is introduced, thus the analytical solutions are considered as the correct dynamics. As the stochastic simulations lie in close proximity to the analytical curves, we consider that the SSA is accurate. A similar closeness is found again in scenario III, where, however, the analytical moment dynamics are not closed. We can thus conclude that the moment closure approximation [13] is here accurate. In scenario IV, instead, the approximation is unreliable. From Figures 4 (a) and 4 (b) we see that, even if we increase significantly the order of the moment (n_m) where the moment-closure approximation is applied, the accuracy of the analytic trajectory is preserved only for a limited time. In this case, SSA predictions are corroborated by the numerical solution. Implementing the cell average technique as in [14], the accuracy of numerical results is only guaranteed for $\langle \tilde{s} \rangle$ dynamics. This approach looses precision, however, when the C_V^2 rapidly diverges, like in scenario II, as the discretization grid can no more cover the higher cell sizes in the population. Despite these limitations, the numerical solution of the PBE offers a useful comparison at a much lower computational cost than the stochastic approach.

Figure 5 shows that, for constant division rates, the exponential increase in N, the total number of cells in the population, is unaffected by the initial condition. For size-dependent division rates, instead, if $\langle s \rangle|_{t=0} > \langle \tilde{s} \rangle$ the population starts expanding rapidly, while, if $\langle s \rangle|_{t=0} < \langle \tilde{s} \rangle$, the expansion displays a small lag. In both cases an exponential regime is reached after a few doubling times.

IV. CONCLUSIONS

Cell size heterogeneity in growing populations is tightly controlled and generally low, a phenomenon defined as size homeostasis. In this work, we used multiple mathematical modeling approaches to provide a further understanding on how studies on size homeostasis are influenced by changes in different model components. On one side, we considered different combinations of growth regimes and division strategies, which lead us to define four different biologically relevant scenarios. On the other, we studies the variability of cell sizes in a population observed from two perspectives: single lineages and population snapshots.

We used a deterministic population-based framework, the population balance equation (PBE), which models growth and division process trough continuous rates. This approach allowed us to derive analytically a system of ordinary differential equations (ODE) that describes the dynamics of the raw moments of the cell size distribution. For some of the considered scenarios, the moment closure approximation was used to truncate the system and avoid dependencies on higher-order moments. The solution of this ODE system was compared with a numerical solution of the PBE and with stochastic simulations. Our comparison showed that the analytical derivations offer a good approximation in three out of four scenarios.

Secondly, the mean cell-size for single lineages is consistently higher than for population snapshots, while their variability (C_V^2) is not significantly different. In some of the scenarios, we show that the ODEs allow us to establish mathematical relationships among the two considered perspectives. With this, we propose that the steady state values for population snapshots can be inferred from results or data on single lineages, and vice versa.
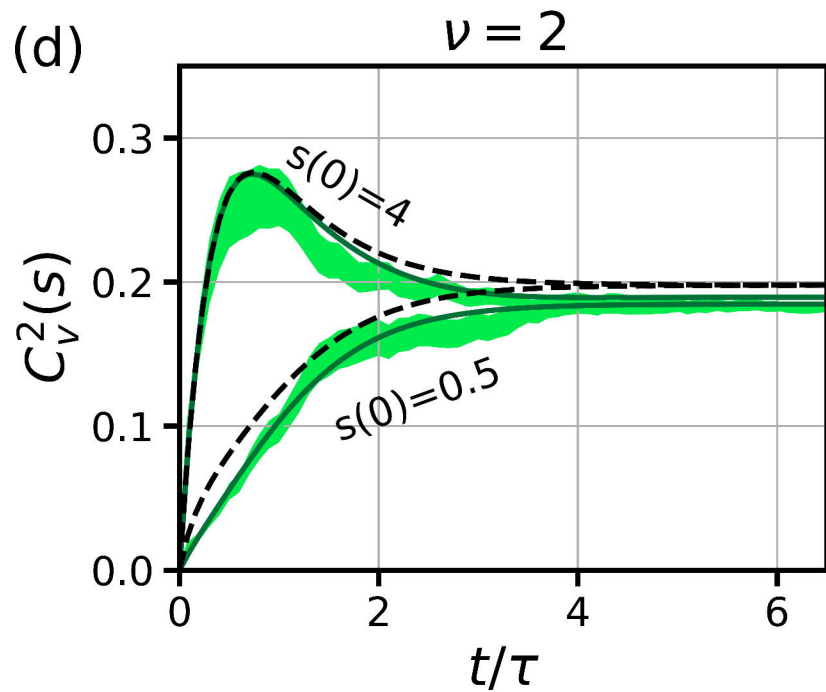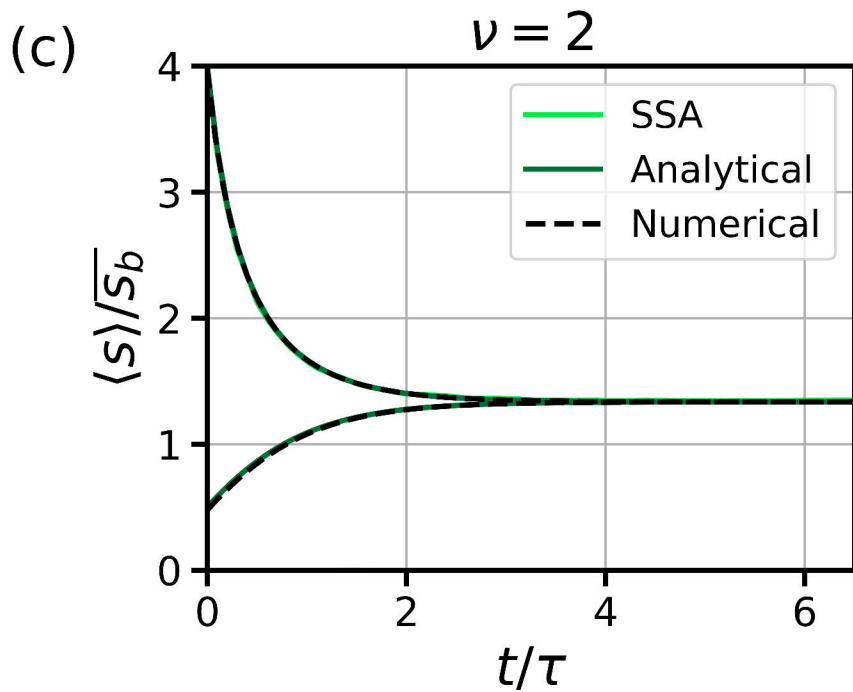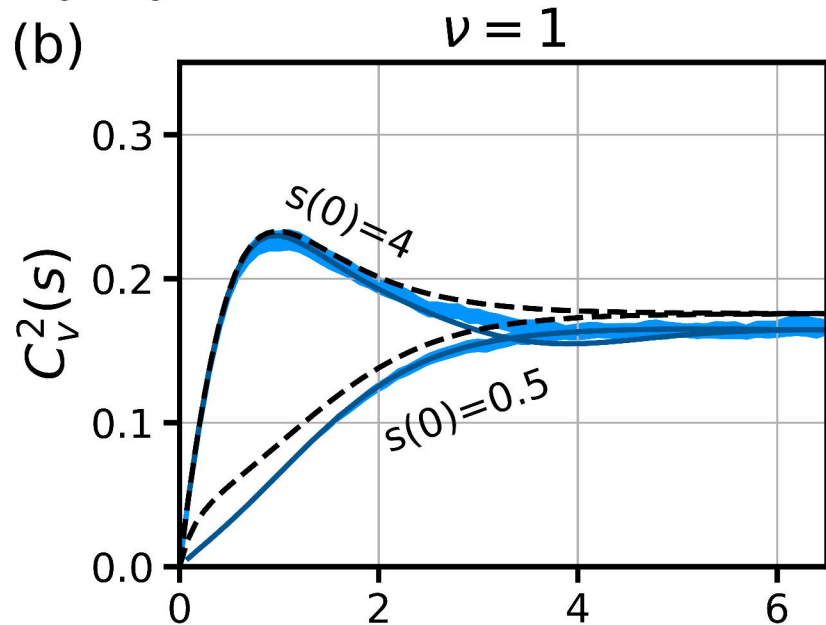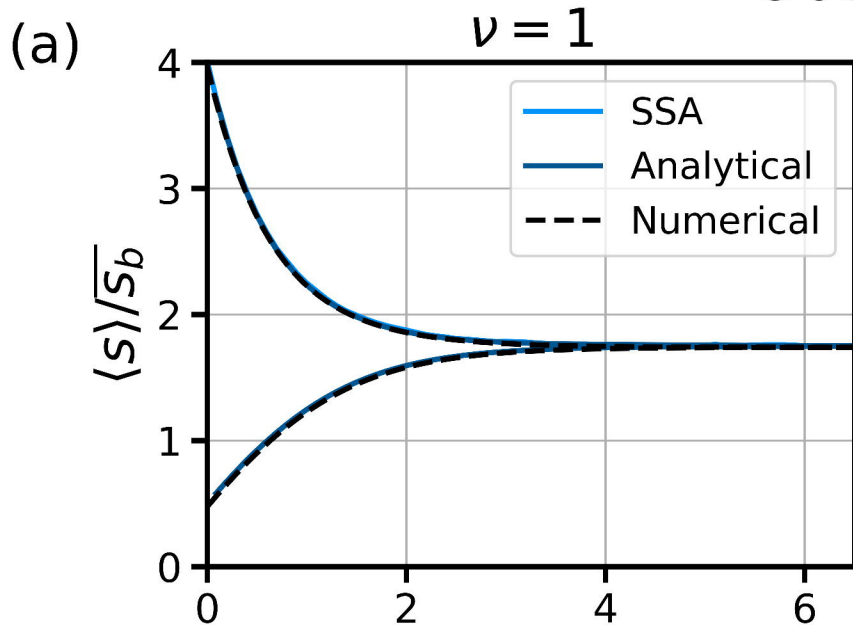
V. ACKNOWLEDGMENTS

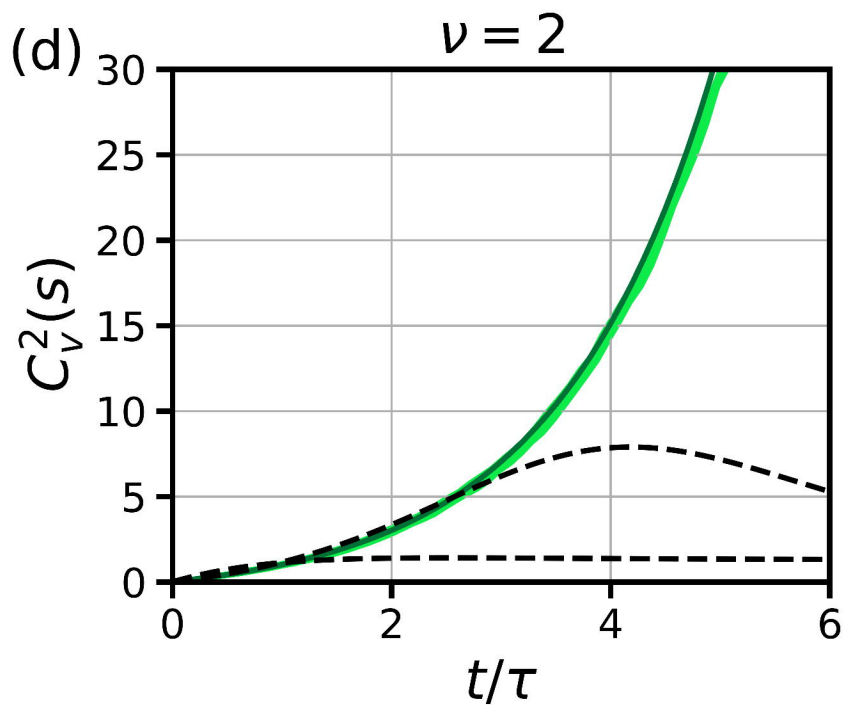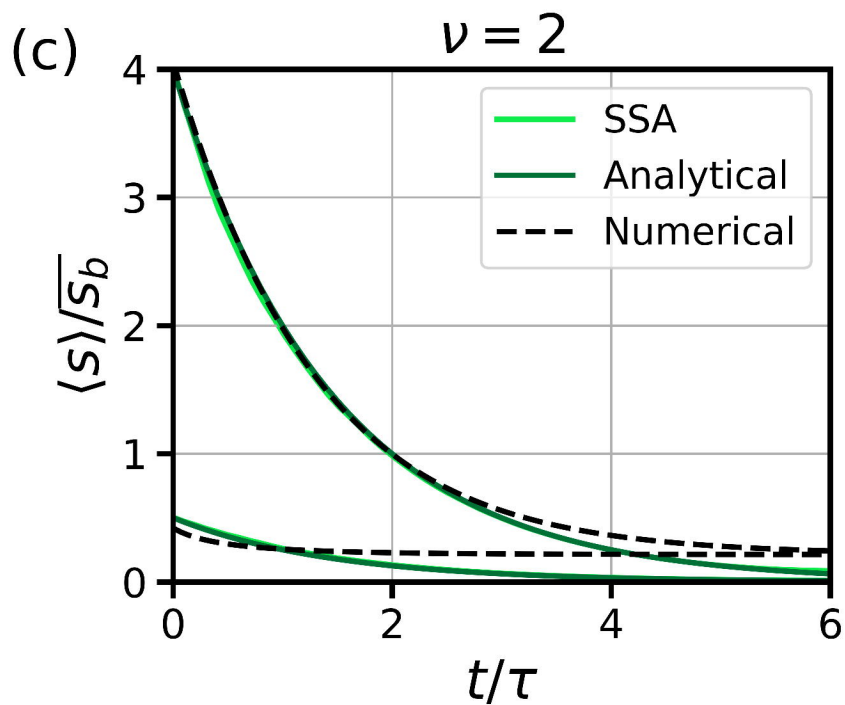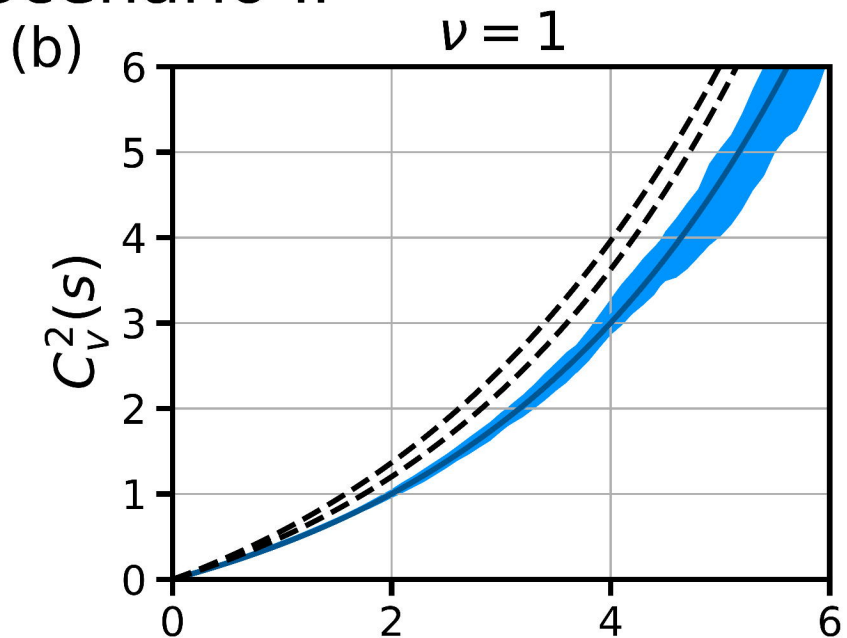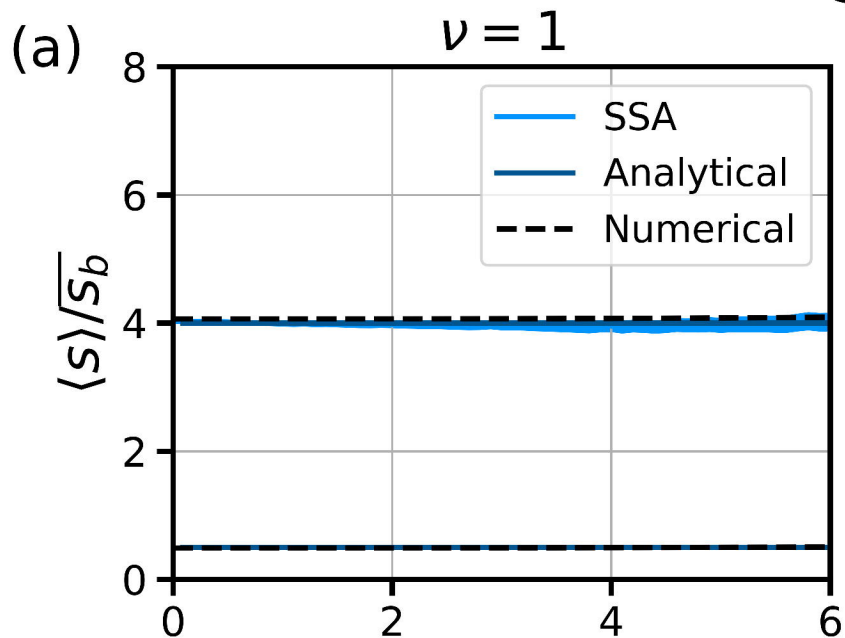## REFERENCES

[1] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, "Robust growth of escherichia coli," *Current biology*, vol. 20, no. 12, pp. 1099–1103, 2010.

[2] W. F. Marshall, K. D. Young, M. Swaffer, E. Wood, P. Nurse, A. Kimura, J. Frankel, J. Wallingford, V. Walbot, X. Qu, *et al.*, "What determines cell size?," *BMC biology*, vol. 10, no. 1, p. 101, 2012.

[3] P. Thomas, "Intrinsic and extrinsic noise of gene expression in lineage trees," *Scientific reports*, vol. 9, no. 1, pp. 1–16, 2019.

[4] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, "Cell-size control and homeostasis in bacteria," *Current Biology*, vol. 25, no. 3, pp. 385–391, 2015.

[5] F. F. Mandy, M. Bergeron, and T. Minkus, "Principles of flow cytometry," *Transfusion science*, vol. 16, no. 4, pp. 303–314, 1995.

[6] P. Thomas, "Making sense of snapshot data: ergodic principle for clonal cell populations," *Journal of The Royal Society Interface*, vol. 14, no. 136, p. 20170467, 2017.

[7] C. A. Nieto-Acuna, C. A. Vargas-Garcia, A. Singh, and J. M. Pedraza, "Efficient computation of stochastic cell-size transient dynamics," *BMC bioinformatics*, vol. 20, no. 23, pp. 1–6, 2019.

[8] P. Thomas, "Analysis of cell size homeostasis at the single-cell and population level," *Frontiers in Physics*, vol. 6, p. 64, 2018.

[9] R. García-García, A. Genthon, and D. Lacoste, "Linking lineage and population observables in biological branching processes," *Physical Review E*, vol. 99, no. 4, p. 042413, 2019.

[10] A. Marguet, M. Lavielle, and E. Cinquemani, "Inheritance and variability of kinetic gene expression parameters in microbial cells: modeling and inference from lineage tree data," *Bioinformatics*, vol. 35, no. 14, pp. i586–i595, 2019.

[11] D. Ramkrishna, *Population balances: Theory and applications to particulate systems in engineering*. Elsevier, 2000.

[12] C. Nieto, J. Arias-Castro, C. Sánchez, C. Vargas-García, and J. M. Pedraza, "Unification of cell division control strategies through continuous rate models," *Physical Review E*, vol. 101, no. 2, p. 022401, 2020.

[13] M. Soltani, C. A. Vargas-Garcia, and A. Singh, "Conditional moment closure schemes for studying stochastic dynamics of genetic circuits," *IEEE transactions on biomedical circuits and systems*, vol. 9, no. 4, pp. 518–526, 2015.

[14] J. Kumar, M. Peglow, G. Warnecke, and S. Heinrich, "An efficient numerical technique for solving population balance equation involving aggregation, breakage, growth and nucleation," *Powder Technology*, vol. 182, no. 1, pp. 81–104, 2008.

[15] C. A. Vargas-Garcia, M. Soltani, and A. Singh, "Conditions for cell size homeostasis: A stochastic hybrid system approach," *IEEE Life Sciences Letters*, vol. 2, no. 4, pp. 47–50, 2016.
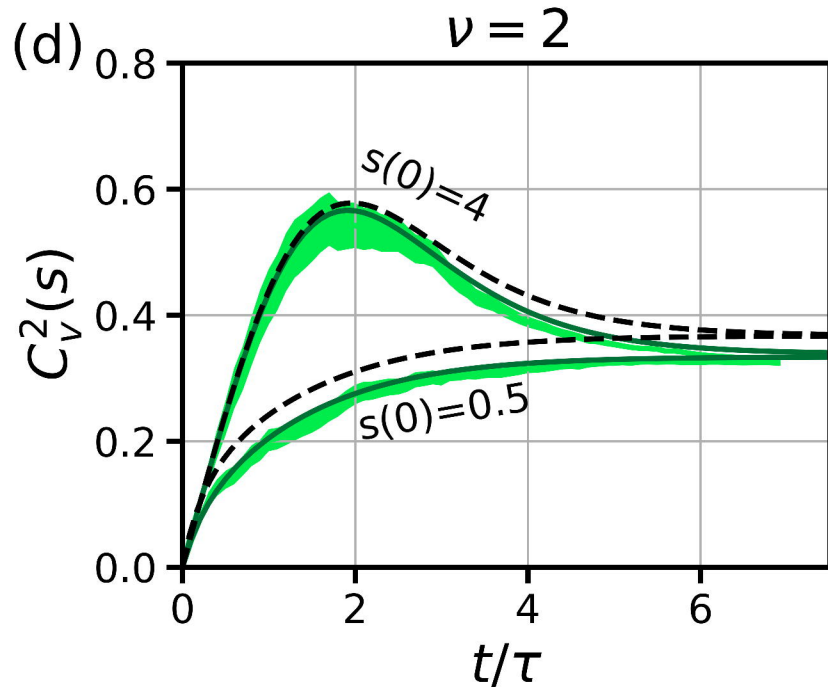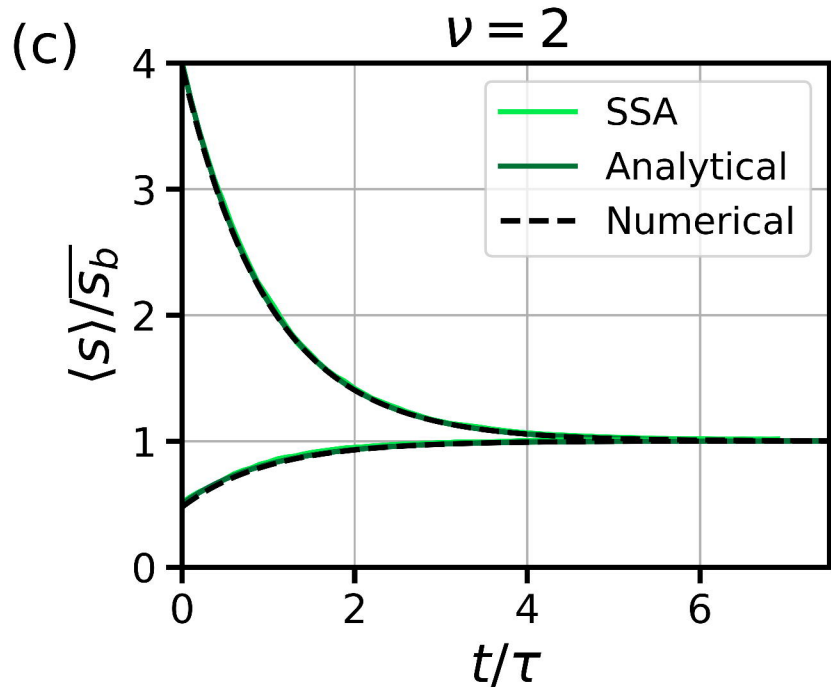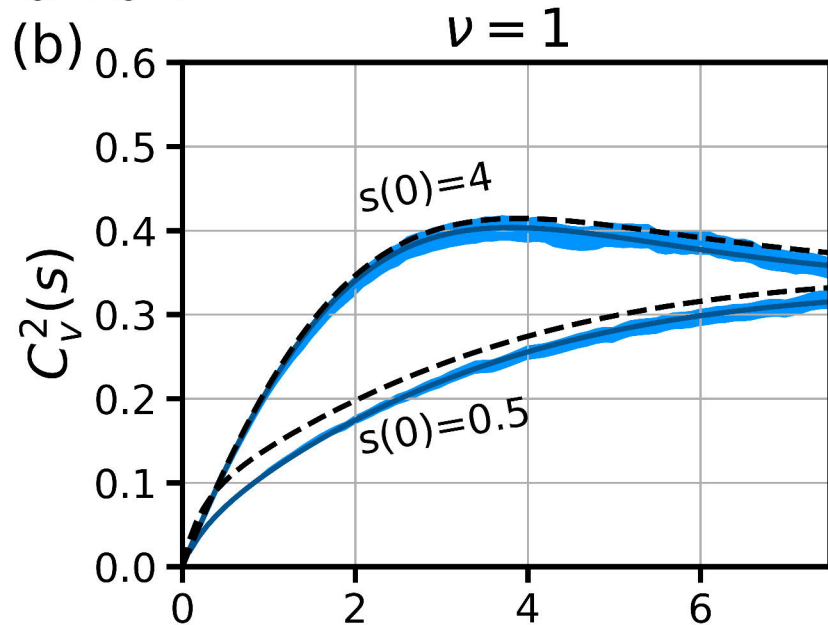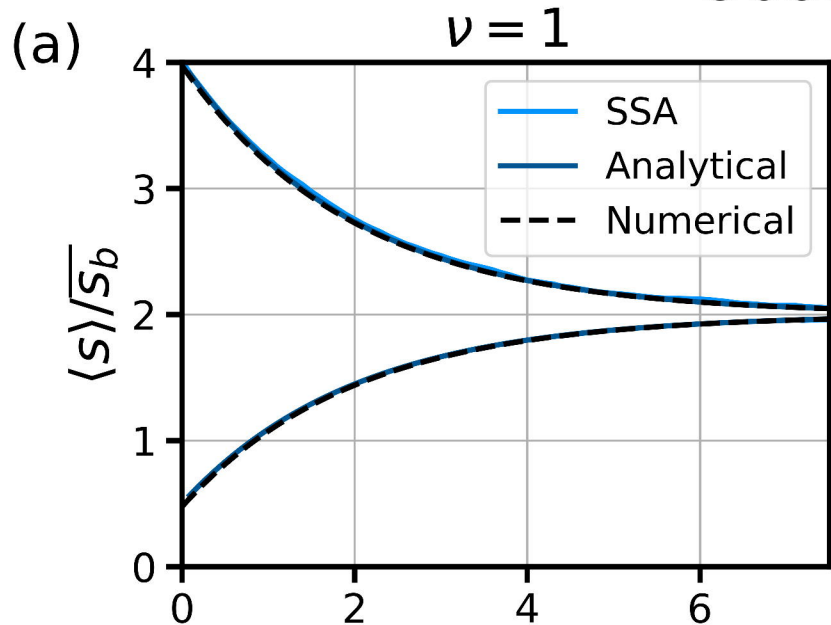
Scenario III

(a) $\nu = 1$

(b) $\nu = 1$

(c) $\nu = 2$

(d) $\nu = 2$

Scenario II

(a) $\nu = 1$

$\langle s \rangle / \overline{s_b}$

SSA
Analytical
Numerical

(b) $\nu = 1$

$C_v^2(s)$

(c) $\nu = 2$

$\langle s \rangle / \overline{s_b}$

SSA
Analytical
Numerical

(d) $\nu = 2$

$C_v^2(s)$

t/τ

Scenario I

(a) $\nu = 1$

(b) $\nu = 1$
- SSA
- Analytical
- Numerical
- $s(0) = 4$
- $s(0) = 0.5$

(c) $\nu = 2$
- SSA
- Analytical
- Numerical

(d) $\nu = 2$
- $s(0) = 4$
- $s(0) = 0.5$

Scenario IV

(a) $\nu = 1$

$\langle s \rangle / \overline{s_b}$

SSA
Analytical
Numerical

$n_m = 7$
$n_m = 5$
$n_m = 21$
$n_m = 8$
$n_m = 2$

(b) $\nu = 1$

$C_v^2(s)$

$s(0) = 4$
$n_m = 2$
$n_m = 8$
$s(0) = 0.5$
$n_m = 7$
$n_m = 5$
$n_m = 21$

(c) $\nu = 2$

$\langle s \rangle / \overline{s_b}$

SSA
Analytical
Numerical

t/τ

(d) $\nu = 2$

$C_v^2(s)$

$s(0) = 4$
$s(0) = 0.5$
$n_m = 7$
$n_m = 7$

t/τ

Figure showing $\langle N \rangle$ versus t/τ for four scenarios. Panels (a) Scenario I, (b) Scenario II, (c) Scenario III, (d) Scenario IV. Legend: SSA, Analytical, Numerical. Panels (c) and (d) show curves labeled $s(0)=4$ and $s(0)=0.5$.