Regulation of stem cell dynamics through volume exclusion

PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: García-Tejera R, Schumacher L, Grima R. 2022 Regulation of stem cell dynamics through volume exclusion. *Proc. R. Soc. A* **478**: 20220376.
https://doi.org/10.1098/rspa.2022.0376

Received: 1 June 2022
Accepted: 27 September 2022

Subject Areas:
mathematical physics, statistical physics

Keywords:
stem cells, volume exclusion, master equation, renormalized system-size expansion, transient bimodality, competition for space

Author for correspondence:
Rodrigo García-Tejera
e-mail: rodrigo.garcia@ed.ac.uk

Regulation of stem cell dynamics through volume exclusion

Rodrigo García-Tejera[1,2], Linus Schumacher[1,2] and Ramon Grima[2]

[1]Centre for Regenerative Medicine, University of Edinburgh, 5 Little France Dr, Edinburgh EH16 4UU, UK
[2]School of Biological Sciences, University of Edinburgh, Kings Buildings, Mayfield Road, Edinburgh EH9 3JF, UK

RG-T , 0000-0003-3427-3587; LS, 0000-0003-0797-3406; RG , 0000-0002-1266-8169

The maintenance and regeneration of adult tissues rely on the self-renewal of stem cells. Regeneration without over-proliferation requires precise regulation of the stem cell proliferation and differentiation rates. The nature of such regulatory mechanisms in different tissues, and how to incorporate them in models of stem cell population dynamics, is incompletely understood. The critical birth-death (CBD) process is widely used to model stem cell populations, capturing key phenomena, such as scaling laws in clone size distributions. However, the CBD process neglects regulatory mechanisms. Here, we propose the birth-death process with volume exclusion (vBD), a variation of the birth-death process that considers crowding effects, such as may arise due to limited space in a stem cell niche. While the deterministic rate equations predict a single non-trivial attracting steady state, the master equation predicts extinction and transient distributions of stem cell numbers with three possible behaviours: long-lived quasi-steady state (QSS), and short-lived bimodal or unimodal distributions. In all cases, we approximate solutions to the vBD master equation using a renormalized system-size expansion, QSS approximation and the Wentzel–Kramers–Brillouin method. Our study suggests that the size distribution of a stem cell population bears signatures that are useful to detect negative feedback mediated via volume exclusion.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

Stem cells (SCs) are a population of cells capable of self-renewing and differentiating into all the cells in a particular lineage. In adult tissue homeostasis, SCs slowly self-renew and differentiate to compensate for the death of other cells while maintaining a constant average population size [1–4]. Upon injury, however, the SC proliferation and differentiation rates increase dramatically to repair the tissue, only settling back into homeostasis when regeneration is completed [5–7]. Such tight control over the SC proliferation and differentiation rates requires regulatory mechanisms providing feedback to the SCs.

Against the backdrop of recent advances in experimental techniques in stem cell biology, a broad range of SC regulatory mechanisms have been reported, such as negative feedback exerted by the more differentiated cells [6–8], competition between SCs for fate determinants [9,10], or mechanical feedback [11–13]. An additional plausible mechanism stems from the confinement of SCs to a particular microenvironment, the stem cell niche [9,14–16]. This microenvironment plays a key role in maintaining cell stemness and promoting SC quiescence [17–19], self-renewal, or differentiation, according to tissue requirements; however, it also triggers a competition between SCs for niche access [10,20,21].

Crowding effects are often associated with volume exclusion [22–26]. The non-negligible volume of particles restricts their movement, thus obstructing their access to available free space [22]. As a consequence, the accessible phase space can be greatly reduced. If cells are dividing without reducing their size, crowding effects can have an impact on the SC proliferation and death (or differentiation) rates. For example, a proliferation event reduces the available space, which in turn reduces the proliferation rate, thus creating a negative feedback loop. Volume exclusion has also been suggested to play a role in the regulation of cancer stem cells and tumour growth [27]. However, it is not yet clear how to distinguish between crowding effects and other regulatory mechanisms from observations of the population evolution (e.g. from snapshots of the SC population at different times).

Stem cell division and differentiation have been previously modelled stochastically by the simple chemical reaction network $S \to 2S$, $S \to \emptyset$ [1,28], where differentiation is equivalent to death if differentiated progeny do not self-renew. To prevent the population from diverging or vanishing, the birth and death rates must be equal, thus obtaining a critical birth-death (CBD) process. For stem cell population dynamics, the CBD process has been frequently treated under well-mixed and dilute gas conditions, which facilitate its computational implementation, e.g. Gillespie algorithm [29,30] and mathematical analysis through the master equation formalism [31,32]. This approach has been successfully employed to illustrate key features of SC populations, such as population asymmetry (the maintenance of a constant average population via symmetric divisions that are balanced at the population level, instead of asymmetric divisions), neutral competition [1,28,33] and scaling properties of clone size distributions [28,33–35]. However, the CBD ignores the finite-size nature of cells and thus disregards the role of available space in cell division.

Here, we present a modification of the birth-death process that includes competition for niche access, the birth-death process with volume exclusion (vBD). We subdivide the space within a niche into N voxels (small volumes of space); each voxel is either occupied by a stem cell or else is empty. Assuming well-mixed conditions, the effective chemical reaction network describing this process is $S + E \to 2S$, $S \to E$, where S and E describe stem cells and empty voxels in the niche, respectively. The first reaction reflects the need for a stem cell to find an empty voxel to divide, while the second one represents the birth of an empty voxel after stem cell death or differentiation (assuming differentiated progeny leave the niche space). Naturally, the system obeys the conservation law $n_S + n_E = N$, where N is the niche carrying capacity, and n_S, n_E are the number of stem cells and empty voxels, respectively. Note that the vBD resembles a stochastic SIS model, with the number of infectious given by the species S, and susceptible by E. However, while the SIS model is usually treated for $N \to \infty$, we are interested in the low N behaviour. In terms of the vBD model parameters, the basic reproduction number for the equivalent SIS

model is given by $R_0 = Nk_1/k_2$, where k_1 and k_2 are the rates of proliferation and differentiation, respectively. The deterministic rate equations for the vBD process predict a logistic convergence to a non-trivial attracting steady state. From a microscopic perspective (i.e. the master equation's solutions), however, this prediction is not realized, and the vBD process relaxes to extinction, irrespective of the parameter values and initial conditions.

Our analysis reveals the three different behaviours of the stochastic vBD process that are absent in its deterministic counterpart. When the birth rate is much larger than the death rate, the system quickly takes the form of a long-lived, quasi-steady state (QSS) and very slowly relaxes to extinction through a transient bimodal distribution. Conversely, for death rates much larger than the birth rates, the system quickly converges to extinction through a unimodal transient. Lastly, when the birth and death rates are comparable, the transient distribution is bimodal but the convergence to extinction is fast. For these three different parameter regimes, we approximate the solution of the vBD master equation using a quasi-stationary approximation, a renormalized system-size expansion (SSE) (including finite size corrections to the linear-noise approximation (LNA)), and the Wentzel–Kramers–Brillouin (WKB) method. In particular, the renormalized SSE is a recent modification of the original van Kampen's SSE that has not been widely used yet, but proves useful for tackling master equations of non-linear birth-death processes. Finally, we derive an expression for the expected extinction time based on Kolmogorov's backward equation and first-passage time theory. Our analytical solutions provide insights into the rich behaviours of the vBD model.

2. Birth-death process with volume exclusion model

The vBD is defined by the chemical reaction network

$$S + E \xrightarrow{k_1} 2S; \quad S \xrightarrow{k_2} E, \tag{2.1}$$

where S and E represent stem cells and empty voxels, respectively (see figure 1*a* for an illustration). For the deterministic system to have a non-trivial steady state, we require that $k_1 > k_2$. Note that the two species are coupled by the conservation law $n_S + n_E = N$. Assuming mass-action kinetics and defining the dimensionless time $\tau = k_2 t$, the rate equation for the average stem cell concentration $\phi = n_S/N$ adopts the logistic form

$$\frac{\partial \phi}{\partial \tau} = \frac{(\phi^* - \phi)\phi}{1 - \phi^*}, \tag{2.2}$$

where $\phi^* = 1 - k_2/k_1$ is the non-trivial steady state. The deterministic evolution of the stem cell concentration has the form $\phi(\tau) = (\phi^*\phi_0)/[(\phi^* - \phi_0)\,\mathrm{e}^{-\phi^*\tau/(1-\phi^*)} + \phi_0]$, where $\phi(0) = \phi_0$ is the initial condition, and we can appreciate that $\phi \xrightarrow{\tau\to\infty} \phi^*$ when $\phi_0 \neq 0$ (figure 1*b*). In this deterministic system, the extinction state $\phi = 0$ is never reached unless $\phi(0) = 0$.

The stochastic behaviour of the vBD differs from the deterministic predictions. Trajectories of the vBD generated using the stochastic simulation algorithm (SSA [29]) fluctuate in the vicinity of the deterministic steady state for some finite period of time (figure 1*b*), but fluctuations eventually drive the stem cell number to extinction. The ensemble average of stochastic trajectories thus converges to zero (see blue line in figure 1*b*), disagreeing with the deterministic model's prediction.

A stochastic treatment of the vBD process is provided by its chemical master equation (CME), i.e. Kolmogorov's forward equation. The CME describes the time-evolution of the probability that the system is in one of its states [31,36–38]. To construct the CME, we first note that the vBD model only involves reactions that increase or reduce the number of stem cells by one unit. Hence the stochastic process underlying the reaction network (2.1) takes the form of the Markov chain depicted in figure 2*a*, where the states $0, 1, \ldots, N$ represent the number of stem cells. Note that

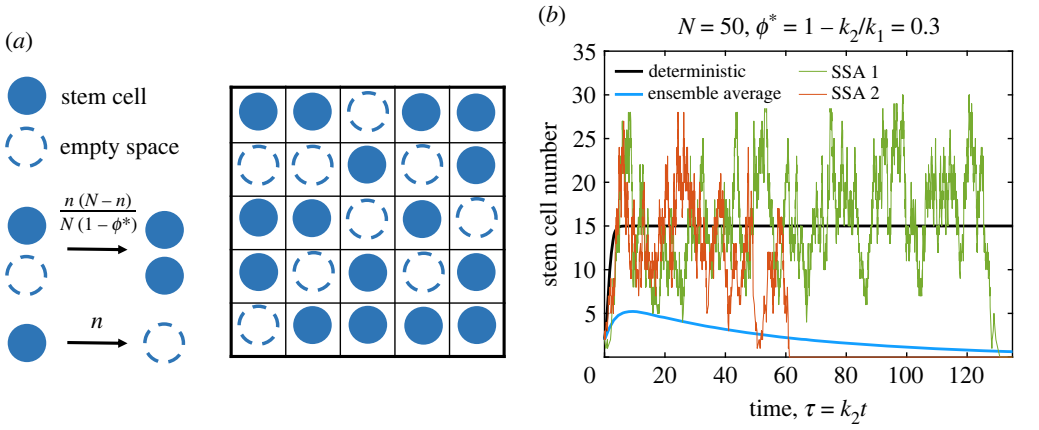

Figure 1. The vBD model and its deterministic approximation. (*a*) Stem cells (solid circles) move randomly in the niche (grid) by switching positions with empty voxels (dashed, empty circles). At any time, a stem cell can 'react' with a neighbouring empty voxel to create two stem cells (cell division), and a single stem cell can differentiate or die, leaving a new empty voxel in return. The propensities follow mass-action kinetics. (*b*) Under well-mixed conditions, the deterministic approximation predicts the evolution of the mean stem cell concentration satisfying the logistic equation (2.2), thus portraying a logistic convergence to the average stem cell number $N\phi^*$ (black line). Stochastic trajectories (green and orange lines) obtained by the SSA, however, are driven to the extinction state by fluctuations. The ensemble average (blue line) of 2×10^4 stochastic trajectories reveals the eventual distinction—in contrast to the deterministic prediction of a non-zero steady state. (Online version in colour.)

the vBD process features a reflecting boundary at $n = N$ and an absorbing boundary at $n = 0$. The propensities are determined by the law of mass-action and in dimensionless units read:

$$\left.\begin{aligned} a_n &= (n-1)(N-n+1)/[N(1-\phi^*)], \quad n \geq 1 \\ \text{and} \qquad b_n &= n+1, \quad n \geq 0. \end{aligned}\right\} \tag{2.3}$$

Let $P(n, \tau \mid P(\tau_0) = P_0)$ be the probability of finding the system in a state of n cells at time τ, given that it was P_0 at time τ_0, which we will abbreviate as $P(n, \tau)$. Defining the probability vector $\mathbf{P}(\tau) = (P(0, \tau), P(1, \tau), \ldots, P(N, \tau))^\intercal$, where $\intercal$ denotes the vector transpose, the CME can be expressed as $\mathrm{d}\mathbf{P}/\mathrm{d}\tau = \mathcal{M}\mathbf{P}$, where $\mathcal{M}$ is the operator defined by

$$\mathcal{M} = \begin{bmatrix} 0 & b_0 & 0 & 0 & \ldots \\ 0 & -a_2 - b_0 & b_1 & 0 & \ldots \\ 0 & a_2 & -a_3 - b_1 & b_2 \ldots & \\ 0 & 0 & a_3 & -a_4 - b_2 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{2.4}$$

The nth row of the master equation reads

$$\frac{\mathrm{d}P}{\mathrm{d}\tau} = a_n P(n-1, \tau) + b_n P(n+1, \tau) - (a_{n+1} + b_{n-1}) P(n, \tau), \tag{2.5}$$

with $a_0 = b_{-1} = 0$. Note that the only parameters present are the carrying capacity, N, and the steady state from the deterministic equations, ϕ^*, as per (2.3).

The solution of the master equation, for an initial probability distribution $\mathbf{P}(0)$, is given by $\mathbf{P}(\tau) = \mathrm{e}^{\mathcal{M}\tau}\mathbf{P}(0)$. The main properties of the solution are captured by the eigenvectors and eigenvalues of $\mathcal{M}$. It is easy to prove that $\lambda_0 = 0$ is always an eigenvalue associated with the eigenvector $[1, 0, \ldots, 0]^\intercal$ (the extinction state), while the other eigenvalues are real and negative [39]. Therefore, the extinction state is always reached, irrespective of the parameter values and initial conditions. Moreover, the expected extinction time is the inverse of the spectral gap, $|\lambda_1 - \lambda_0|^{-1}$, where λ_1 is the smallest (in absolute value) non-zero eigenvalue. The third eigenvalue,

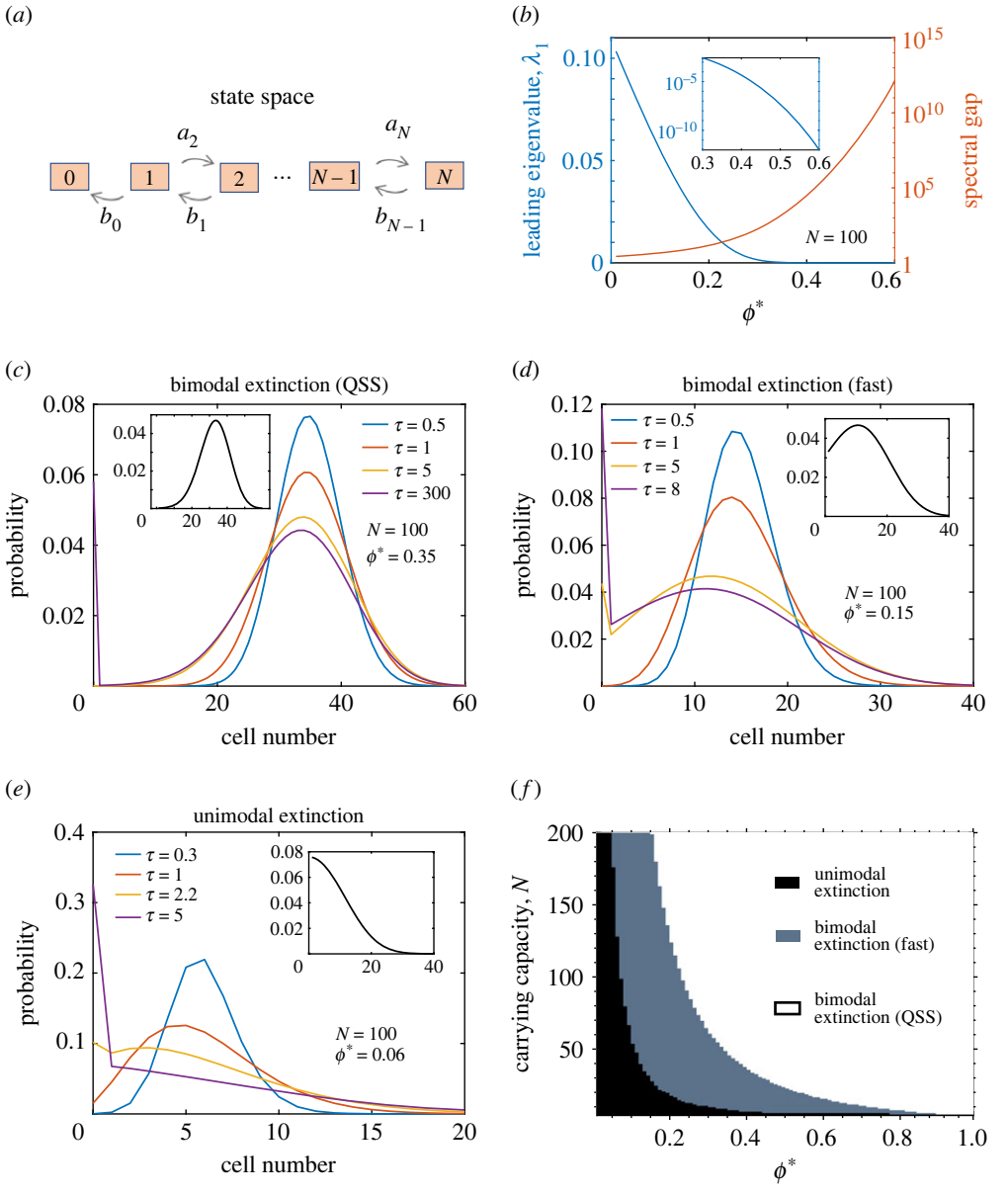

Figure 2. Stochastic properties of the vBD model. (*a*) The vBD system jumps between states of 1, 2, . . . , N stem cells in a Markov process, with propensities a_n and b_n. (*b*) The leading eigenvalue of the operator $\mathcal{M}$, λ_1, vanishes as the deterministic equation's steady state solution, ϕ^*, increases (blue line, inset for logarithmic scale). Correspondingly, the expected extinction time, $|\lambda_1|^{-1}$, is large for high ϕ^*, indicating the presence of a QSS. The relative spectral gap of the system conditioned on non-extinction, i.e. $|(\lambda_2 - \lambda_1)/\lambda_2|$, increases with ϕ^* (orange line). In general, $|\lambda_2|$ is at least 2.6 times higher than $|\lambda_1|$, supporting a solution of the form of equation (2.6) for the vBD master equation. (*c–e*) Probability distributions of the vBD model for different ϕ^* values and times. All time-dependent PDFs are found by numerically solving the master equation $\mathrm{d}\mathbf{P}/\mathrm{d}\tau = \mathcal{M}\mathbf{P}$ using direct matrix exponentiation. For high ϕ^* a QSS is achieved before relaxation into extinction through transient bimodality (*c*). For intermediate ϕ^* transient bimodality is present but a long-lived QSS is absent (*d*). For low ϕ^* extinction is fast via unimodal transient (panel *e*). In all parameter regimes, the probability distributions of the surviving trajectories reach steady states (insets in *c–e*). (*f*) Behaviours of the vBD model for different ϕ^* values and carrying capacities; see main text for explanation of how the phase boundary lines are numerically determined. Note that the discrete jumps of the transition lines reveal the finite-size nature of the niche since the concentrations can only be of the form n/N, with n integer, and the carrying capacity N can only adopt integer values. (Online version in colour.)

λ_2, has a considerably higher absolute value than λ_1, as we can appreciate from the spectral gap of the reduced system obtained by eliminating the extinction state (orange line in figure 2*b*)—we observe that the smallest gap is $\lambda_2 \approx 2.6\lambda_1$ which is achieved in the limit of small ϕ^*. Hence, after an initial transient, the PDF is dominated by the eigenvectors associated with λ_0 and λ_1, leading to

$$P(n,\tau) \approx \mathbf{e}_0 + \mathrm{e}^{-\lambda_1 \tau}\mathbf{e}_1, \tag{2.6}$$

where $\mathbf{e}_0 = [1,0,\ldots,0]^\intercal$ is the extinction state and $\mathbf{e}_1 = [-1, f(1), f(2), \ldots, f(N)]^\intercal$ is the leading eigenvector. The first element of $\mathbf{e_1}$ comes from the lower boundary of the state space. The first row in the CME reads $\mathrm{d}P(0,\tau)/\mathrm{d}\tau = b_0 P(1,\tau) = b_0 f(1)\,\mathrm{e}^{-\lambda_1\tau}$, which leads to (i) $\lambda_1 = b_0 f(1)$ and (ii) $P(0,\tau) = 1 - \mathrm{e}^{-\lambda_1\tau}$. Note that $f(n)$ is the PDF of stem cell numbers conditioned on non-extinction. It follows that the expected extinction time is

$$\mathbb{T} = [b_0 f(1)]^{-1}. \tag{2.7}$$

The distribution of the surviving trajectories is defined as

$$\tilde{\mathbf{P}}(\tau) = [\tilde{P_1}(\tau), \ldots, \tilde{P_N}(\tau)]^\intercal = \frac{\mathbf{P}'(\tau)}{\sum_{k=1}^{N} P_k(\tau)} = \frac{\mathbf{P}'(\tau)}{1 - P_0(\tau)}, \tag{2.8}$$

where $\mathbf{P}'(\tau) = [P(1,\tau), \ldots, P(N,\tau)]^\intercal$. Hence, the master equation for the PDF conditioned on non-extinction reads

$$\frac{\partial \tilde{\mathbf{P}}}{\partial \tau} = \frac{\partial \mathbf{P}'}{\partial \tau}\frac{1}{1-P_0} + \frac{\mathbf{P}'}{(1-P_0)^2}\frac{\partial P_0}{\partial \tau} = \tilde{\mathcal{M}}\tilde{\mathbf{P}} + b_0 \tilde{P}_1 \tilde{\mathbf{P}}, \tag{2.9}$$

where the operator $\tilde{\mathcal{M}}$ results from eliminating the first row and column of $\mathcal{M}$. Note that even though the steady state in the original CME is the extinction state, the CME for the surviving trajectories presents a non-trivial steady state (figure 2*c–e*). In effect, the steady state of equation (2.9) yields the entries $f(k)$ of $\mathbf{e_1}$.

Numerical experiments reveal the presence of three different behaviours in the solution given by equation (2.6). For high ϕ^* values, the system relaxes to a long-lived QSS, $f(n)$, before slowly relaxing again to extinction through a transient bimodal distribution (figure 2*c*). The two modes given by the eigenstates $\mathbf{e}_0$ and $\mathbf{e}_1$ in equation (2.6) are located at the extinction state and near (but not necessarily at) $N\phi^*$, respectively. The leading eigenvalue vanishes for high ϕ^* (blue line in figure 2*b*), indicating the presence of a QSS. For intermediate ϕ^* values, the transient bimodality is still present, but the relaxation to extinction is faster than the time to reach the QSS (figure 2*d*). For low ϕ^* values, the system rapidly goes extinct, and the mode around $N\phi^*$ is absent (figure 2*e*). The phase diagram in figure 2*f* summarizes the three types of behaviour of the vBD model as a function of ϕ^* and the carrying capacity N. To determine the parameter regimes for these three behaviours, we compute numerically the solution to the master equation, $\mathbf{P}(\tau) = \mathrm{e}^{\mathcal{M}\tau}\mathbf{P}(0)$, choosing $P(n,0) = \delta_{nk}$, where δ_{ij} is the Kronecker delta and $k = \lceil N\phi^* \rceil$. We then extract the PDF conditioned on non-extinction after the initial transient. The black region in figure 2*f* corresponds to the case in which such distribution does not present maximum for $n > 1$. To distinguish between the bimodal extinction and QSS regions, we compute numerically the second eigenvalue of $\mathcal{M}$, λ_1, and set a tolerance $\alpha = 1 \times 10^{-2}$. The grey region in figure 2*f* corresponds to $\lambda_1 > \alpha$, and the white region to $\lambda_1 \leq \alpha$. In the next two sections, we derive approximate solutions for $f(n)$ and $\mathbb{T}$ in the three different regimes.

3. Approximate solutions

Here, we present approximate solutions for the vBD master equation. From equation (2.6), it follows that the solution adopts the form $P(n,\tau) = f(n)\,\mathrm{e}^{-\lambda_1\tau}$, $\forall\ n \geq 1$, and $P(0,\tau) = 1 - \mathrm{e}^{-\lambda_1\tau}$. Hence, the asymptotic solution is determined by the PDF conditioned on non-extinction, $f(n)$, and the inverse of the expected extinction time, λ_1. In this section, we tackle the problem of approximating $f(n)$, assuming λ_1 is known, while in the next section, we derive an accurate expression for the expected extinction time and correspondingly for λ_1.

Substituting equation (2.6) into the master equation, we obtain

$$a_n f(n-1) + b_n f(n+1) + (\lambda_1 - a_{n+1} - b_{n-1}) f(n) = 0, \tag{3.1}$$

where $1 \leq n < N$. Defining $u_n = f(n)/f(n-1)$ leads to the recurrence relations

$$u_k = \frac{a_k}{b_{k-1} + a_{k+1} - \lambda_1 - b_k u_{k+1}}, \quad k \geq 2, \tag{3.2}$$

with the boundary conditions $u_2 = (a_2 + b_0 - \lambda_1)/b_1$ and $u_N = a_N/(b_{N-1} - \lambda_1)$. Equation (3.2) can be solved iteratively, leading to a solution in terms of the continued fraction

$$u_k = \cfrac{a_k}{c_k} \Bigg| - \cfrac{b_k a_{k+1}}{c_{k+1}} \Bigg| \cdots - \cfrac{b_{N-2} a_{N-1}}{c_{N-1}} \Bigg| - \cfrac{b_{N-1} a_N}{b_{N-1} - \lambda_1}, \tag{3.3}$$

where we have defined $c_k = \lambda_1 - b_{k-1} - a_{k+1}$. To obtain an expression for $f(n)$, we apply the boundary condition $f(1) = \lambda_1/b_0$, thus leading to

$$f(n) = \frac{\lambda_1}{b_0} \prod_{k=2}^{n} u_k. \tag{3.4}$$

Equations (3.3) and (3.4) are exact solutions for the post-transient dynamics given by equation (2.6). However, the slow convergence of the continued fraction and the difficulty in applying truncation methods render equations (3.3) and (3.4) unsuitable for the analysis of the system's dynamics. On the other hand, the continued fraction solution offers a fast computational estimation for the master equation's time-dependent solution, often less demanding than direct matrix exponentiation to estimate $\mathbf{P}(n, \tau) = e^{\mathcal{M}\tau}\mathbf{P}(0)$. In what follows, we derive three different approximate solutions to the master equation of the vBD process, one for each region of the phase diagram shown in figure 2*f*.

(a) QSS approximation

For high ϕ^* values, $\lambda_1 \ll 1$ and correspondingly the mean extinction time is very large (see figure 1*b*). Since $\lambda_1 = b_0 f(1)$, we can impose a QSS condition by disregarding the second term in the r.h.s. of equation (2.9), leading to $\partial\tilde{\mathbf{P}}/\partial\tau \approx \tilde{\mathcal{M}}\tilde{\mathbf{P}}$. The steady state of the CME conditioned on non-extinction can be obtained by solving $\tilde{\mathcal{M}}\tilde{\mathbf{P}} = 0$; the Nth row yielding $f(N-1) = (b_{N-1}/a_N) f(N)$. This relationship can be iterated to obtain

$$f(k) = \prod_{i=k}^{N-1} \frac{b_i}{\hat{a}_{i+1}} f(N), \quad 1 \leq k < N, \tag{3.5}$$

where $\hat{a}_i = a_i + \delta_{i2} b_0$, and δ_{i2} is the Kronecker delta. Finally, to find $f(N)$, we make use of the normalization condition which leads us to a closed-form solution for the QSS PDF

$$f(N) = \left[1 + \sum_{k=1}^{N-1} \prod_{i=k}^{N-1} \frac{b_i}{\hat{a}_{i+1}} \right]^{-1}. \tag{3.6}$$

Substituting the propensities a_n and b_n from equation (2.3) yields the approximate QSS solution

$$f(k) = \frac{(\Sigma^2)^{N-k} \left[1 + \delta_{k1} \frac{(N-1)}{N-1+\Sigma^2} \right] / (N-k)!}{1 + \sum_{i=1}^{N-1} \frac{(\Sigma^2)^{N-i}}{(N-i)!} \left[1 + \delta_{i1} \frac{(N-1))}{N-1+\Sigma^2} \right]}, \tag{3.7}$$

for $1 \leq k \leq N$, where we defined $\Sigma^2 = N(1-\phi^*)$. The QSS approximation accurately describes the PDF conditioned on non-extinction for high ϕ^* values (figures 3*a* and 4*a*).

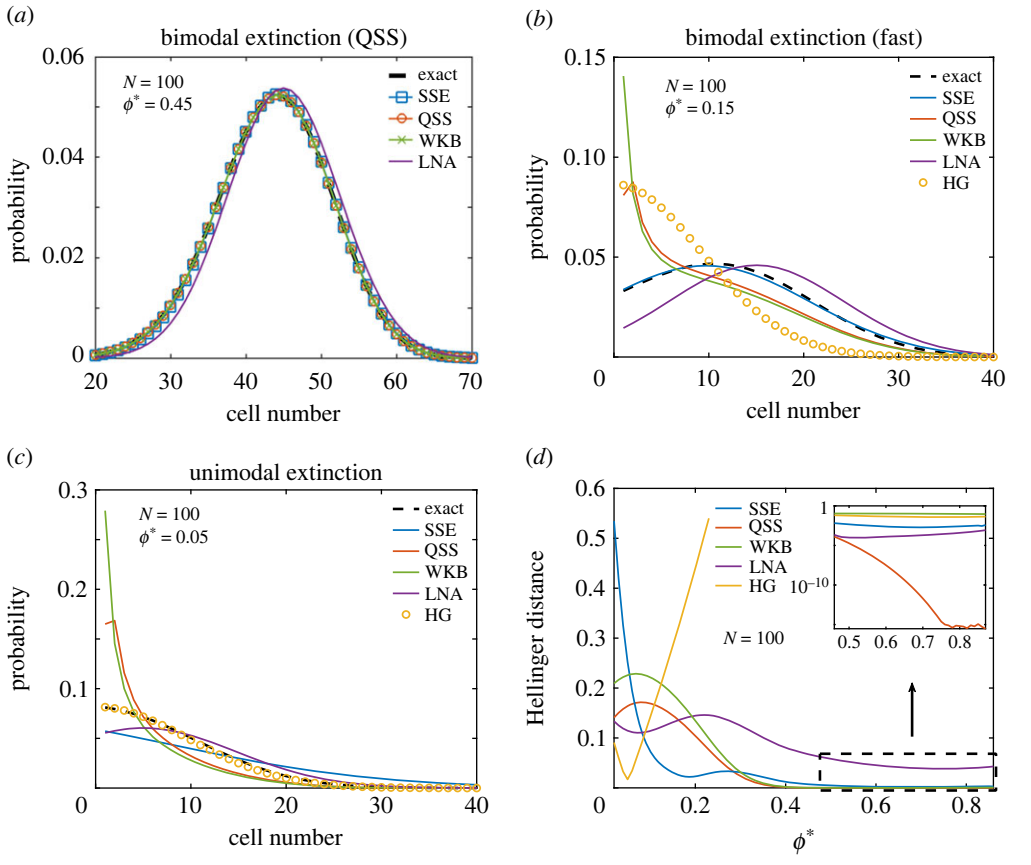

Figure 3. Approximate solutions of the vBD master equation. The SSE, QSS, WKB, LNA and HG lines correspond to the renormalized SSE (up to order N^{-1}), QSS approximation, WKB approximation, linear-noise and half Gaussian approximations, respectively. (*a*) For (N, ϕ^*) within the bimodal extinction region the QSS, SSE and WKB provide good approximations of the probability distribution conditioned on non-extinction, $f(n)$, although the QSS is the most accurate one (see inset in *d*). (*b*) In the bimodal extinction region, the renormalized SSE is the only accurate approximation. (*c*) In the unimodal extinction region, the HG provides the best approximation. (*d*) The Hellinger distance between the various distribution approximations and the exact distribution confirms that the QSS, SSE and HG are the best approximations for the QSS, bimodal extinction and unimodal extinction regions, respectively. (Online version in colour.)

From the QSS approximation, we can calculate the position for the non-zero mode. To do so, it is convenient to study the discrete first derivative $\partial f(n)/\partial n = f(n+1) - f(n)$. For a maximum or minimum to take place, it is necessary that $f(n+1)/f(n) = 1$. According to equation (3.7), we have

$$\frac{f(n+1)}{f(n)} = \frac{n(N-n)}{N(1-\phi^*)(n+1)} \quad \forall\, n \geq 2. \tag{3.8}$$

Solving for n, we have

$$n^{*\pm} = \tfrac{1}{2}\left[N\phi^* \pm \sqrt{(N\phi^*)^2 - 4N(1-\phi^*)}\right]. \tag{3.9}$$

The sign of the discrete second derivative, $\partial^2 f/\partial n^2 = f(n+1) + f(n-1) - 2f(n)$ reveals that, when both solutions exist, n^{*+} corresponds to a maximum and n^{*-} to a minimum. Note that, for $n^{*\pm}$ to adopt real values, $(N\phi^*)^2 \geq 4N(1-\phi^*)$, which is the case for all parameter sets in the QSS region. Hence, in general, the position of the non-zero mode differs from the deterministic model's prediction, i.e. $n^{*+} \neq N\phi^*$. As an example, for a carrying capacity of $N = 100$ and $\phi^* = 0.4$, then $n^{*+} \approx 38$. Note that when $\phi^* \to 1$ or $N \to \infty$, $n^{*+} \to N\phi^*$ and $n^{*-} \to 0$. Thus, in these limits, the

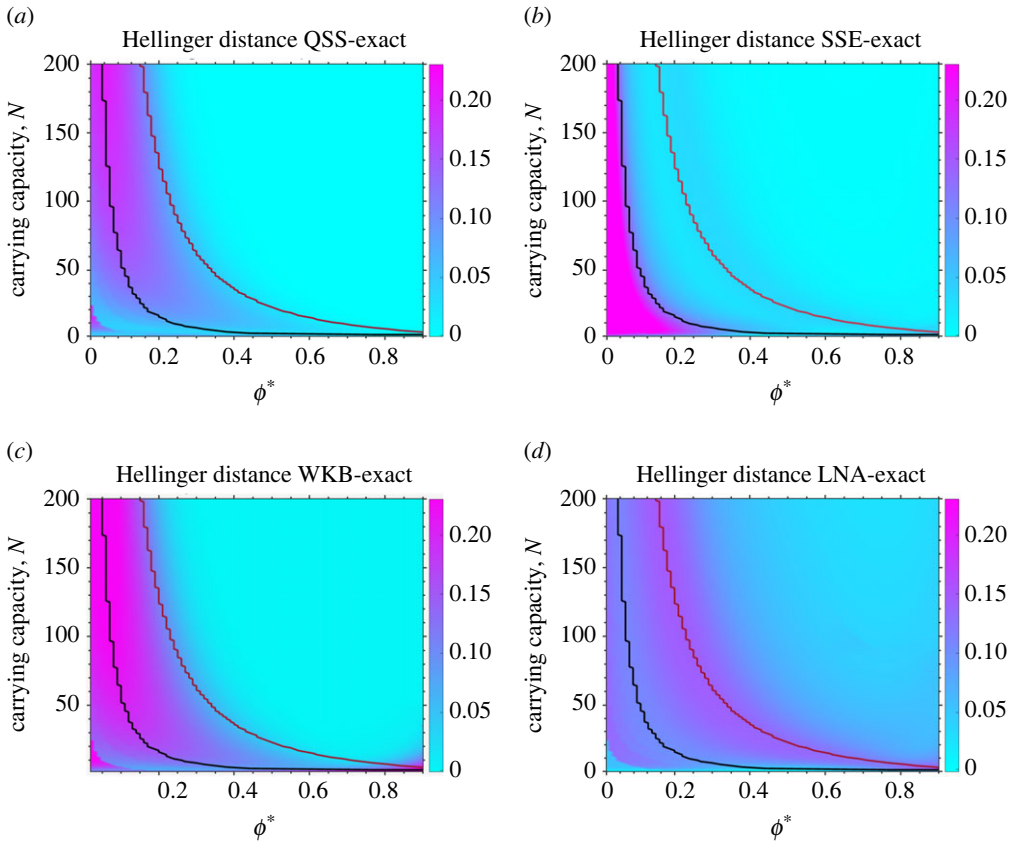

Figure 4. Hellinger distance between the exact and approximate probability distributions conditioned on non-extinction, as a function of the carrying capacity, N, and the deterministic rate equation's steady state, ϕ^*. Red lines represent the interfaces between the QSS and the non-QSS regions, while the black lines are the interfaces between the transient unimodal and bimodal extinction regions; note these lines demarcate the three different phases shown in figure 2*f*. For high carrying capacities, the QSS approximation performs well at capturing the PDF conditioned on non-extinction (bimodal extinction region; see (*a*)). The renormalized SSE is accurate in both fast and QSS bimodal extinction regions (see *b*). The WKB has a similar range of validity as the QSS approximation (see *c*). The LNA is noticeably less accurate than the other three approximations except for low ϕ^* (unimodal extinction region; see *d*). (Online version in colour.)

distribution has a single mode sitting at the deterministic model's prediction of the mean stem cell number.

(b) Renormalized system size expansion

For parameter sets (ϕ^*, N) in the bimodal and unimodal extinction regions (grey and black areas in figure 2*f*), the QSS approximation is unable to capture the PDF conditioned on non-extinction (as it can be seen in figures 3*b*,*c* and 4*a*). In such cases, the condition $\lambda_1 \ll 1$ does not hold, and hence the second term in the r.h.s. of equation (2.9) can no longer be considered negligible (which is needed to find the steady state of the CME conditioned on non-extinction iteratively, as we did in the QSS case). Hence a different approximation is needed. In what follows, we derive a new approximate PDF conditioned on non-extinction based on a high-order renormalized SSE of the vBD's master equation. In §3(i), we provide a general analytical recipe to obtain an expression for the PDF conditioned on non-extinction, equations (3.20) and (3.21) being the main results. While self-contained, this derivation is mathematically lengthy, and the reader can skip it should they

be solely interested in its application to the vBD process. We point the reader to [40] for a detailed derivation of the method. In §3(ii), we apply the renormalized SSE results to the vBD system, and obtain an analytical expression for the PDF conditioned on non-extinction.

(i) Derivation of the renormalized SSE formula

The van Kampen SSE approximates a master equation by splitting the random variable that describes the stem cell number, n, into deterministic and non-deterministic components, to then obtain a master equation for the non-deterministic components (usually taken as a continuous random variable). The resultant CME can be expanded in powers of $N^{-1/2}$, truncated to the desired order and solved, leading to a hierarchy of approximate solutions [41]. Whilst the bulk of applications are centred in truncating the CME after the order N^0 to obtain the LNA, which considers the particle concentration equal to its deterministic value, other works have used higher order truncation schemes to obtain corrections to the mean concentration [42,43]. The deterministic component is commonly assumed to be given by the solution of the rate equations. However, the disagreement between the deterministic and stochastic predictions (see figure 1*b*) makes it sensible to introduce a correction term to the mean stem cell concentration in the original ansatz. To do so, we follow the procedure from [40] that starts by considering the ansatz

$$\frac{n}{N} = \phi + N^{-1/2}\langle\epsilon\rangle + N^{-1/2}\hat{\epsilon}, \tag{3.10}$$

where the first term in the r.h.s. is the zero-order mean stem cell concentration obtained from the deterministic rate equations, the second term is a correction to the mean concentration due to fluctuations, and the third term in the r.h.s. represents fluctuations about the corrected mean concentration. This renormalization of the mean concentration leads to a different SSE than the conventional one by van Kampen, which we refer to as the renormalized SSE.

Next, we briefly describe how to compute the corrections to the mean concentration as a series in powers of $N^{-1/2}$:

$$\langle\epsilon\rangle = \sum_{j=0}^{\infty} N^{-j/2} a_1^{(j)}. \tag{3.11}$$

The expansion coefficients for the correction term to the mean concentration are calculated iteratively as follows:

$$a_n^{(j)} = -\frac{1}{n\mathcal{J}} \sum_{k=1}^{j} \sum_{s=0}^{\lceil k/2 \rceil} \sum_{p=1}^{k-2(s-1)} \mathcal{D}_{p,s}^{k-p-2(s-1)} \sum_{m=0}^{3(j-k)} a_m^{(j-k)} \mathcal{I}_{mn}^{p,k-p-2(s-1)}, \tag{3.12}$$

where $\mathcal{J}$ is the Jacobian of the deterministic rate equations, and we assume $a_m^{(0)} = 0$. To define the operators $\mathcal{D}_{p,s}^{q}$, we assume that the propensity functions for a birth or death event when the system is in a state of n stem cells (a_{n+1} and b_{n-1}), expressed in terms of the concentrations, can be expanded in power series of the inverse carrying capacity (N^{-1}) as

$$a(N\phi, N) = N \sum_{s=0}^{\infty} N^{-s} g_1^{(s)}(\phi) \tag{3.13}$$

and

$$b(N\phi, N) = N \sum_{s=0}^{\infty} N^{-s} g_2^{(s)}(\phi), \tag{3.14}$$

where $g_r^{(s)}(\phi)$ are the expansion coefficients. For example, $g_1^{(s)}(\phi)$ can be obtained by defining $z = N^{-1}$, transforming $a(N\phi, N)/N \to z a(\phi/z, 1/z)$, and Taylor-expanding around $z = 0$. Following

we define

$$\mathcal{D}^{q}_{p,s} = \sum_{r=1}^{2} (S_r)^p \frac{\partial^q g_r^{(s)}(\phi)}{\partial \phi^q}, \tag{3.15}$$

where S_r is the net change in the number of stem cells when the rth reaction occurs, namely $S_1 = 1$ and $S_2 = -1$. The functions $\mathcal{I}^{\alpha\beta}_{mn}$ in equation (3.12) are defined as

$$\mathcal{I}^{\alpha\beta}_{mn} = \frac{\sigma^{\beta-\alpha+n-m}}{\alpha!} \sum_{s=0}^{\min(n-\alpha,m)} \binom{m}{s} \frac{[\beta+\alpha+2s-(m+n)-1]!!}{[\beta+\alpha+2s-(m+n)]!(n-\alpha-s)!}, \tag{3.16}$$

for $(\alpha+\beta)-(m+n)$ even, and zero otherwise. We have introduced the notation $(2k-1)!! = (2k)!/(2^k k!)$ for the double factorial, and σ is the standard deviation of the concentration according to the standard LNA [41]

$$\frac{\partial \sigma^2}{\partial \tau} = 2\mathcal{J}\sigma^2 + \mathcal{D}^2_{0,0}. \tag{3.17}$$

The variance of the fluctuations can also be computed as a series in powers of $N^{-1/2}$

$$\hat{\sigma}^2 = \sigma^2 + \sum_{j=1}^{\infty} N^{-j/2} \hat{\sigma}_j^2, \tag{3.18}$$

where the expansion coefficients are given by

$$\hat{\sigma_j^2} = 2\left[a_2^{(j)} - \frac{\mathcal{B}_{j,2}(\{\chi!a_1^{(\chi)}\}_{\chi=1}^{j-1})}{j!} \right], \tag{3.19}$$

with $\mathcal{B}_{j,k}$ being the partial Bell polynomials, where $\{\cdot\}$ denotes the set of arguments [44]. For example, $\mathcal{B}_{4,2}(\{\chi!a_1^{(\chi)}\})$ has as arguments $1!a_1^{(1)}$, $2!a_1^{(2)}$, and $3!a_1^{(3)}$.

To summarize, the SSE procedure involves expanding the master equation for the fluctuations about the mean concentration (assuming the fluctuations are a continuous random variable) in powers of $N^{-1/2}$, and truncating after $\mathcal{O}(N^0)$ to obtain a Fokker–Planck equation that yields the LNA, thus describing Gaussian fluctuations around the mean concentration. The higher-order approximate solutions can then be expressed in terms of the first-order approximation. However, this approach often leads to non-physically meaningful distributions for truncations of the SSE beyond the LNA level of approximation, e.g. yielding negative probabilities or oscillatory behaviour. Such effects can be greatly reduced by introducing a discrete formulation of the SSE's approximate solutions [40].

The discrete formulation replaces the continuous-variable LNA approximate solution with the discrete approximation

$$P_0(n,\tau) = \frac{1}{2} \frac{\mathrm{e}^{-(x^2/2\Sigma^2)}}{\sqrt{2\pi}\,\Sigma} \left[\operatorname{erf}\left(\frac{\mathrm{i}x + \pi\Sigma^2}{\sqrt{2}\Sigma} \right) - \operatorname{erf}\left(\frac{\mathrm{i}x - \pi\Sigma^2}{\sqrt{2}\Sigma} \right) \right], \tag{3.20}$$

where erf is the error function, $x = n - N\phi - N^{1/2}\langle\epsilon\rangle$ is the stem cell number centred about its (corrected) deterministic value, and $\Sigma^2 = N\hat{\sigma}^2$ its variance. Note that the time dependence is implicit in the temporal change of the mean concentration and the variance of concentration fluctuations, $\phi(\tau)$ and $\sigma(\tau)$, respectively. Equation (3.20) is the discrete version of a Gaussian in the sense that every moment of the distribution coincides with their corresponding of a continuous-variable Gaussian. Next, the expansion of the SSE approximate solution up to any order can be expressed in terms of $P_0(n,\tau)$ and its derivatives

$$P(n,\tau) = P_0(n,\tau) + \sum_{j=1}^{\infty} N^{-j/2} \sum_{m=1}^{3j} \hat{a}_m^{(j)} (-N^{1/2}\partial_n)^m P_0(n,\tau), \tag{3.21}$$

where the new expansion coefficients $\hat{a}_m^{(j)}$ are related to the coefficients found earlier ($a_m^{(j)}$ in equation (3.12)) by the relation

$$\hat{a}_m^{(j)} = \sum_{k=0}^{j} \sum_{n=0}^{3k} a_n^{(k)} k_{m-n}^{(j-k)}, \tag{3.22}$$

with

$$k_j^{(n)} = \frac{1}{n!} \sum_{m=0}^{\lfloor j/2 \rfloor} (-1)^{j+m} \sum_{k=j-2m}^{n-m} \binom{n}{k} \times \mathcal{B}_{k,j-2m}(\{\chi! a_1^{(\chi)}\}_{\chi=1}^{k-j+2m+1}) \times \mathcal{B}_{n-k,m}\left(\left\{\frac{\chi!}{2}\hat{\sigma}^2_{(\chi)}\right\}_{\chi=1}^{n-k-m+1}\right). \tag{3.23}$$

To calculate the derivatives of $P_0(n,\tau)$, it is possible to prove by induction the following formula:

$$\frac{\partial^m P_0}{\partial y^m} = -\frac{(m-1)}{\Sigma^2}\frac{\partial^{m-2} P_0}{\partial y^{m-2}} - \frac{y}{\Sigma^2}\frac{\partial^{m-1} P_0}{\partial y^{m-1}} + (-1)^{\lceil m/2 \rceil + 1} F(y,m)\pi^{m-2}\frac{\mathrm{e}^{-(\pi^2\Sigma^2/2)}}{\Sigma^2}, \tag{3.24}$$

where $\lceil \cdot \rceil$ denotes the ceiling function, and $F(y,m)$ is $\sin(\pi y)$ for m odd and $\cos(\pi y)$ for m even.

Calculation of the correction terms to the mean concentration and the variance of fluctuations, $\langle \epsilon \rangle$ and $\hat{\sigma}^2$, respectively, by truncation of equations (3.11) and (3.18) to any desired order, followed by substitution into equation (3.21) and truncation, provides a means to systematically obtain approximate solutions to the CME. Note that the order of the approximate solution is determined by order of truncation of $\langle \epsilon \rangle$ and $\hat{\sigma}^2$. For example, to obtain an approximation of order $(s+1)/2$ with $s = 0, 1, \ldots$, both $\langle \epsilon \rangle$ and $\hat{\sigma}^2$ are expanded up to that order (assuming that at least one of the coefficients is non-zero) using equations (3.11) and (3.18), and then the approximate distribution is obtained by truncating equation (3.21) after the order $(s+1)/2$.

(ii) The expansion of the vBD master equation

We now employ the renormalized discrete formulation of the SSE to approximate the probability distributions conditioned on non-extinction in the bimodal extinction region, i.e. the grey zone in figure 2*f*. To do so, we observe that the absorbing boundary, which is the factor that renders the deterministic rate equations inaccurate for calculating the evolution of the mean stem cell number, is absent when conditioning on non-extinction. In effect, the rate equations capture meaningful information about the mean stem cell number of the surviving stochastic trajectories. Thus, to approximate the steady states of equation (2.9) we can apply the renormalized SSE described in the previous subsection, under stationary conditions, to the non-conditional master equation (2.5), which includes the extinction state. We then recover the PDF conditioned on non-extinction by removing the extinction state and multiplying the resulting distribution by a normalization constant $1/(1-P(0))$, where $P(0)$ is the probability of being in the extinction state obtained by the SSE. In particular, to capture non-Gaussian fluctuations, we truncate the renormalized SSE after terms of $\mathcal{O}(N^{-1})$.

The expansion coefficients for the vBD propensities ($g_r^{(s)}$ in equation (3.14)) are given by

$$g_1^{(0)}(\phi) = \frac{\phi(1-\phi)}{(1-\phi^*)}; \quad g_2^{(0)}(\phi) = \phi, \tag{3.25}$$

and $g_r^{(s)}(\phi) = 0 \; \forall \; s \geq 1$, which allows us to calculate the values of $D_{p,s}^q$.

From the standard LNA, we obtain the order N^0 approximation for the first two moments of the distributions

$$\frac{\partial \phi}{\partial \tau} = \frac{(\phi^* - \phi)\phi}{1-\phi^*} \tag{3.26}$$

and

$$\frac{\partial \sigma^2}{\partial \tau} = \frac{1}{1-\phi^*}[2(\phi^* - 2\phi)\sigma^2 + \phi(2 - \phi - \phi^*)], \tag{3.27}$$

royalsocietypublishing.org/journal/rspa Proc. R. Soc. A 478: 20220376

where the first line is the rate equation and the second one is equation (3.17) in the vBD case. These equations predict the first two moments to be $\phi = \phi^*$ and $\sigma^2 = 1 - \phi^*$ under stationary conditions. The Jacobian in stationary conditions reads $\mathcal{J} = -\phi^*/(1-\phi^*)$. We can now calculate the first few expansion coefficients $a_n^{(j)}$ from equation (3.12):

$$\left.\begin{aligned} a_1^{(1)} &= \phi^* - 1, \\ a_3^{(1)} &= \frac{(1-\phi^*)^2}{6\phi^*}[(2-3\phi^*)(1-\phi^*) - 2], \\ a_2^{(2)} &= \frac{1}{4\phi^*}[a_1^{(1)}(3\phi^* - 4) - 12a_3^{(1)} - (1-\phi^*)], \\ a_4^{(2)} &= \frac{(1-\phi^*)}{8\phi^*}\left[a_3^{(1)}\frac{(5\phi^*-8)}{1-\phi^*} - a_1^{(1)}\phi^* - \frac{\phi^*}{6} - 1\right] \\ \text{and} \quad a_6^{(2)} &= \frac{1}{2}(a_3^{(1)})^2, \end{aligned}\right\} \tag{3.28}$$

whilst $a_n^{(j)} = 0$ for $n + j$ odd. Next, we expand the corrections to the mean concentration and fluctuation's variance ($\langle\epsilon\rangle$ and $\hat{\sigma}^2$ in equations (3.11) and (3.18)) up to order N^{-1}: $\langle\epsilon\rangle \approx N^{-1/2}a_1^{(1)} + N^{-1}a_1^{(2)}$, and $\hat{\sigma}^2 \approx \sigma^2 + N^{-1/2}\hat{\sigma}_1^2 + N^{-1}\hat{\sigma}_2^2$. Equation (3.19) yields $\hat{\sigma}_1^2 = 0$ and $\hat{\sigma}_2^2 = 2[a_2^{(2)} - (a_1^{(1)})^2]$. Thus, we arrive at

$$\left.\begin{aligned} \langle\epsilon\rangle &\approx N^{-1/2}a_1^{(1)} \\ \text{and} \quad \hat{\sigma}^2 &\approx 2N^{-1}\left[a_2^{(2)} - (a_1^{(1)})^2\right], \end{aligned}\right\} \tag{3.29}$$

which, upon substitution on equation (3.20) leads to our order N^0 (LNA) approximation:

$$f_0(n) = \frac{P_0(n)}{1 - P_0(0)}. \tag{3.30}$$

Note that, since we have imposed the stationary conditions $\phi = \phi^*$ and $\sigma = 1 - \phi^*$, there is no longer a time dependence in P_0. To calculate the higher order approximate solutions we make use of equation (3.21). The expansion coefficients, corrected by renormalization, are given by equation (3.22). The first few coefficients are

$$\left.\begin{aligned} \hat{a}_1^{(1)} &= \hat{a}_2^{(1)} = \hat{a}_1^{(2)} = \hat{a}_2^{(2)} = \hat{a}_3^{(2)} = \hat{a}_5^{(2)} = 0, \\ \hat{a}_3^{(1)} &= \frac{(1-\phi^*)^2}{6\phi^*}[(2-3\phi^*)(1-\phi^*) - 2], \\ \hat{a}_4^{(2)} &= \frac{(1-\phi^*)^2}{2\phi^*}\left(3\phi^* - \frac{5}{4}\right) - \frac{1-\phi^*}{48} - \phi^* + \frac{(2-\phi^*)(8-5\phi^*)}{48\phi^*(1-\phi^*)} \\ \text{and} \quad \hat{a}_6^{(2)} &= \frac{1}{72}(1-\phi^*)^2. \end{aligned}\right\} \tag{3.31}$$

Remarkably, there are fewer non-zero renormalized coefficients than regular ones; hence, the analytical expressions for the SSE distributions corrected by renormalization adopt a simpler form.

Next, equation (3.21) yields the following expression for the order $N^{-1/2}$ approximate solution

$$\left.\begin{aligned} P_1(n) &= P_0(n) - N^{-1/2}\left\{\hat{a}_1^{(1)}N^{1/2}\frac{\partial P_0}{\partial n} + \hat{a}_3^{(1)}N^{3/2}\frac{\partial^3 P_0}{\partial n^3}\right\} \\ \text{and} \quad f_1(n) &= \frac{P_1(n)}{1 - P_1(0)}. \end{aligned}\right\} \tag{3.32}$$

Finally, the order N^{-1} approximate solution is

$$\left.\begin{aligned} P_2(n) &= P_1(n) + N^{-1}\left\{\hat{a}_4^2 N^2 \frac{\partial^4 P_0}{\partial n^4} + \hat{a}_6^2 N^3 \frac{\partial^6 P_0}{\partial n^6}\right\} \\ \text{and} \qquad f_2(n) &= \frac{P_2(n)}{1 - P_2(0)}. \end{aligned}\right\} \tag{3.33}$$

The renormalized SSE accurately describes the probability conditioned on non-extinction for ϕ^* in the QSS region (figure 3*a*), as well as in the fast bimodal extinction region (figure 3*b*), where the QSS approximation breaks down, and accurately captures the distribution skewness. Moreover, this result is robust to changes in the carrying capacity (figure 4*b*). However, the accuracy of the renormalized SSE decreases dramatically for very low ϕ^* values, as can be seen in figure 3*c*. For non-linear birth-death processes featuring non-Gaussian fluctuations, the renormalized SSE consistently performs better than the LNA, which is unable to capture the distribution skewness under stationary conditions.

(iii) Approximation for low ϕ^* values

The phases with extinction through a bimodal and unimodal transient are separated by a critical curve, which also accurately demarcates the regions of parameter space where the renormalized SSE expansion is accurate and where it is not (see figure 4*b*). We attribute the inaccuracy of the SSE in the low ϕ^* region to the fact that the PDF conditioned on non-extinction no longer features a mode around $N\phi^*$, which is the basis of the SSE approximations.

In the unimodal extinction region—black region in figure 2*f*—the time-dependent probability distribution features a mode at $n = 0$ as extinction is approached. This suggests that we can approximate the cell number concentration by the steady-state of the deterministic equations, ϕ^*, and the cell number concentration fluctuations (conditioned on non-extinction) by means of the LNA, equation (3.27), yielding $\sigma^2 = 1 - \phi^*$ under stationary conditions. Note that the latter conditions naturally arise from the conditioning of the distribution on non-extinction. Given the mode at zero, a Gaussian is clearly not a good approximation and hence instead we try a half-Gaussian approximation with the aforementioned first two moments

$$f(n) = \sqrt{\frac{2}{N(1-\phi^*)\pi}}\, \mathrm{e}^{-n^2/[2N(1-\phi^*)]}. \tag{3.34}$$

The half-Gaussian provides an excellent approximation to the PDF conditioned on non-extinction in the low ϕ^* unimodal region (figure 3*c*). However, as expected, this approximation breaks down for higher ϕ^* values, where the non-trivial mode is present (figure 3*b*,*d*).

(c) Wentzel–Kramers–Brillouin approximation

An alternative way to obtain a QSS approximation is the popular WKB approximation [45–47]. The WKB approximation often leads to simpler expressions for a QSS than equation (3.5). A detailed derivation of the WKB approach to solve master equations in QSS conditions can be found in [45], although we repeat the main ideas in what follows.

The WKB approximation starts by transforming the vBD's PDF of observing n cells, $P(n)$ to a continuous PDF for the cell concentration, $\phi = n/N$, assuming N sufficiently large. The vBD master equation reads

$$\frac{\partial P}{\partial \tau} = a_n P(n-1,\tau) + b_n P(n+1,\tau) - (a_{n+1} + b_{n-1})P(n,\tau). \tag{3.35}$$

Defining $\phi = n/N$, we can transform the propensities to $\Omega_+(\phi) = a_{n+1}/N$ and $\Omega_-(\phi) = b_{n-1}/N$, arriving at

$$\Omega_+(\phi) = \frac{\phi(1-\phi)}{1-\phi^*}; \quad \Omega_-(\phi) = \phi. \tag{3.36}$$

We can now write the master equation for the PDF of the continuous variable ϕ, $\Pi(\phi,\tau) = P(N\phi,\tau)$. It follows that $P(n\pm1,\tau) = P(N(\phi\pm1/N),\tau) = \Pi(\phi\pm1/N,\tau)$. Multiplying the CME by $1/N$ and applying the quasi-stationary condition $\partial P/\partial\tau = 0$ yields

$$\Omega_+\left(\phi-\frac{1}{N}\right)\Pi\left(\phi-\frac{1}{N}\right)+\Omega_-\left(\phi+\frac{1}{N}\right)\Pi\left(\phi+\frac{1}{N}\right)-(\Omega_-(\phi)+\Omega_+(\phi))\Pi(\phi)=0. \tag{3.37}$$

Next, the WKB approximation amounts to assuming a solution of the form,

$$\Pi(\phi) = K(\phi)\,\mathrm{e}^{-NS(\phi)}\left[1+\mathcal{O}\left(\frac{1}{N}\right)\right], \tag{3.38}$$

where $S(\phi)$ and $K(\phi)$ are of the order of unity. Substituting in the quasi-stationary master equation, expanding with respect to N^{-1}, and collecting the leading order terms yields

$$\Omega_+(\phi)\,\mathrm{e}^{S'(\phi)-1}+\Omega_-(\phi)\,\mathrm{e}^{-S'(\phi)-1}=0, \tag{3.39}$$

where $S'(\phi) = \mathrm{d}S(\phi)/\mathrm{d}\phi$. From here, we note that the above equation corresponds to a stationary Hamilton–Jacobi equation ($H(\phi,S'(\phi))=0$), for an action S with Hamiltonian

$$H(\phi,p) = \Omega_+(\phi)\,\mathrm{e}^{p-1}+\Omega_-(\phi)\,\mathrm{e}^{-p-1}, \tag{3.40}$$

with $p = S'(\phi)$. The corresponding Hamilton equations are

$$\left.\begin{aligned}\dot\phi &= \frac{\partial H}{\partial p} = \Omega_+(\phi)\,\mathrm{e}^{p}+\Omega_-(\phi)\,\mathrm{e}^{-p}\\ \text{and}\qquad \dot p &= -\frac{\partial H}{\partial\phi} = (1-\mathrm{e}^{p})\frac{\partial\Omega_+(\phi)}{\partial\phi}+(1-\mathrm{e}^{-p})\frac{\partial\Omega_-(\phi)}{\partial\phi}.\end{aligned}\right\} \tag{3.41}$$

Since we are interested in the zero-energy solution ($H=0$), with initial conditions $\phi(t_0)=\phi^*$, the action along a fluctuation trajectory will be given by

$$S = \int_{t_0}^{t} p\dot\phi\,\mathrm{d}t'. \tag{3.42}$$

Hence, we can find S by solving Hamilton's equations and integrating $p\dot\phi$. From Hamilton's equations (3.41), we note that there is a trivial solution with $p=0$. This solution leads to the deterministic rate equations and hence is of no interest to us. The other solution comes from setting $H=0$ in the Hamiltonian expression (3.40), solving for p, and substituting in Hamilton's equations, to yield

$$p = \log\left(\frac{\Omega_-(\phi)}{\Omega_+(\phi)}\right);\quad \dot\phi = \Omega_-(\phi)-\Omega_+(\phi). \tag{3.43}$$

We are now ready to calculate the action $S(\phi)$. Substituting in equation (3.42) and integrating, we obtain

$$S(\phi)-S(\phi^*) = (1-\phi)\log\left(\frac{1-\phi}{1-\phi^*}\right)+\phi-\phi^*. \tag{3.44}$$

It is possible to prove, using the next order contributions to the WKB expansion, that the prefactor $K(\phi)$ is

$$K(\phi) = A(\Omega_+(\phi)\Omega_-(\phi))^{-1/2}, \tag{3.45}$$

where A is later determined by the normalization condition [45]. Finally, the WKB expansion for the vBD system yields the probability distribution

$$\Pi(\phi) = \sqrt{\frac{N}{2\pi(1-\phi)}}\,\frac{\phi^*}{\phi}\,\mathrm{e}^{N(\phi^*-\phi)}\left(\frac{1-\phi}{1-\phi^*}\right)^{N(\phi-1)}. \tag{3.46}$$

While the WKB is a good approximation for large values of ϕ^* (figure 3*a*), it fails for small and intermediate values of ϕ^* (3*b*,*c*). In general, the range of validity of the WKB approximation coincides with that of the QSS approximation (compare figure 4*a*,*c*). An advantage of the WKB approximation stems from its simple analytical form. However, the QSS approximation is generally more accurate than the WKB approximation (see figure 3*d*).

4. Calculation of the extinction time

In the previous section, we have shown different approximations for the leading eigenvector of the vBD model's master equation solution, equation (2.6). The time-dependent component of the solution is determined by the leading eigenvalue, λ_1, which is the inverse of the expected extinction time. The most direct method for estimating this eigenvalue would be to calculate the probability conditioned on non-extinction, $f(n)$, and making use of equation (2.7) to yield $\lambda_1 = b_0 f(1)$. However, the absorbing boundary at $n = 0$ leads to an inaccurate estimation of the probability of having $n = 1$ stem cells (using all approximation methods considered) and thus we cannot use this method to estimate the extinction time (see figure 5*a*). Hence, we present an alternative calculation for the expected extinction time that is based on averaging the mean extinction time starting from any state, and makes use of Kolmogorov's backward equation. This approach has been effectively used for other similar problems (see, for example, [48,49]). Instead of relying solely on the estimation of $f(1)$, this method involves averaging among all the $f(n)$ values, which significantly improves the accuracy with respect to direct application of $\lambda_1 = b_0 f(1)$.

Given an estimate for the probability distribution of the surviving trajectories, $f(n)$, the expected extinction time $\mathbb{T}$ is simply the average among initial conditions of the mean first passage times to hit the extinction state

$$\mathbb{T} = \sum_{n=1}^{N} \tau_n^* f(n), \tag{4.1}$$

where τ_n^* is the mean first passage time to hit the extinction state, starting from the state with n cells. Hence, to estimate $\mathbb{T}$ we need to find the mean first passage times τ_n^*. To do so, we make use of the discrete-time Kolmogorov's backward equation

$$Q_{0,n}(\tau + \Delta\tau) = a_{n+1}\Delta\tau Q_{0,n+1}(\tau) + b_{0,n-1}\Delta\tau Q_{0,n-1}(\tau) + (1 - a_{n+1}\Delta\tau - b_{n-1}\Delta\tau)Q_{0,n}(\tau), \tag{4.2}$$

where $Q_{0,n}(\tau)$ is the probability of being extinct at time τ, given that initially there were n cells. The backward equation just states that the total probability of becoming extinct from state n at time $\tau + \Delta\tau$ equals the probability of jumping to the state with $n + 1$ cells and then going extinct, plus the probability of jumping to state with $n - 1$ cells and then going extinct, plus the probability of staying in the same state and going extinct. The probability density of becoming extinct at time $\tau + \Delta\tau$ is $Q_{0,n}(\tau + \Delta\tau) - Q_{0,n}(\tau)$. Hence, the mean first passage time to extinction, starting from a state of n cells, reads

$$\tau_n^* = \sum_{k=0}^{\infty} k\Delta\tau[Q_{0,n}(k\Delta\tau) - Q_{0,n}((k-1)\Delta\tau)]. \tag{4.3}$$

Here, we assume $Q_{0,n}(-\Delta\tau) = 0$. To simplify the notation, let us denote with $Q_n(\tau)$ the cumulative probability of becoming extinct at time τ given that the system started in state with n cells. Substituting $\tau + \Delta\tau \to \tau$ in equation (4.2), we have $Q_n(\tau) = a_{n+1}\Delta\tau Q_{n+1}(\tau - \Delta\tau) + b_{n-1}\Delta\tau Q_{n-1}(\tau - \Delta\tau) + (1 - \Delta\tau(a_{n+1} + b_{n-1}))Q_n(\tau - \Delta\tau)$. Subtracting both expressions, multiplying by $\tau = k\Delta\tau$ and integrating over time (which in this case amounts to sum over all k), leads to

$$\begin{aligned}\sum_{k=0}^{\infty} k\Delta\tau[Q_n((k+1)\Delta\tau) - Q_n(k\Delta\tau)] = &\sum_{k=0}^{\infty} a_{n+1}k(\Delta\tau)^2[Q_{n+1}(\tau) - Q_{n+1}(\tau - \Delta\tau)] \\ &+ b_{n-1}k(\Delta\tau)^2[Q_{n-1}(\tau) - Q_{n-1}(\tau - \Delta\tau)] \\ &+ k(\Delta\tau)(1 - \Delta\tau(a_{n+1} + b_{n-1}))[Q_n(\tau) - Q_n(\tau - \Delta\tau)].\end{aligned} \tag{4.4}$$

We can rewrite the l.h.s. as $\sum_{k=0}^{\infty}(k-1)\Delta\tau[Q_n(k\Delta\tau) - Q_n((k-1)\Delta\tau)] = \tau_n - \Delta\tau$, where we have used that $Q_n(t) = 0\ \forall\tau < 0$, and the normalization condition of the probability density. The r.h.s.

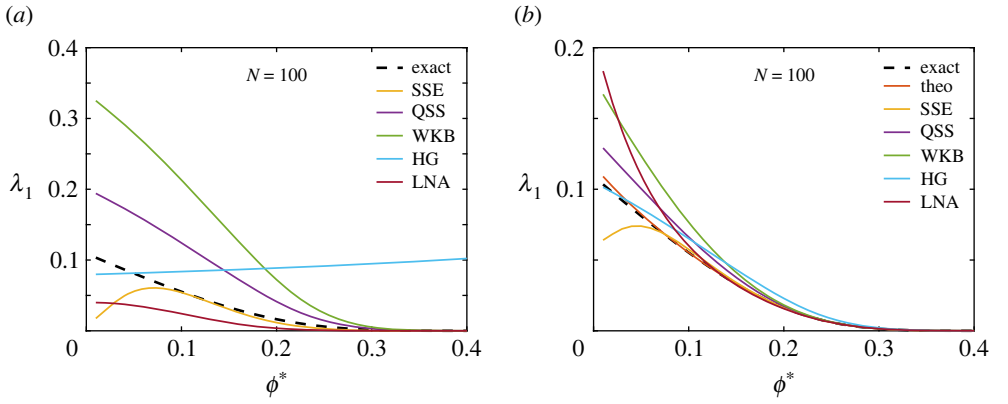

Figure 5. Inverse of the expected extinction time, λ_1, as a function of ϕ^* for a fixed carrying capacity of $N = 100$. (*a*) λ_1 estimated as $b_0 f(1)$, following equation (2.7), with $f(1)$ extracted from the PDF approximations shown in figure 3. (*b*) λ_1 calculated from equation (4.1) and equation (4.8), using different approximations of the probability distribution conditioned on non-extinction. The black dashed line represents the exact λ_1 value, numerically obtained from the leading eigenvalue of the master operator. The λ_1 estimation via equation (4.1) with $f(n)$ calculated numerically (labelled 'Theo', orange line) is in good agreement with the exact λ_1. For very low ϕ^*, equation (4.1) with the half Gaussian approximation is the best fit, while for higher ϕ^* the SSE performs better. For high ϕ^* all approximations except the half Gaussian lead to good estimates of λ_1. The application of equation (4.1) outperforms the direct approximation of λ_1 using equation (2.7) (contrast (*a*) and (*b*)).

of the expression is $a_{n+1}\Delta\tau\tau^*_{n+1} + b_{n-1}\Delta\tau\tau^*_{n-1} + (1 - \Delta\tau(a_{n+1} + b_{n-1}))\tau^*_n$. Thus, we obtain

$$a_{n+1}\tau^*_{n+1} + b_{n-1}\tau^*_{n-1} - (a_{n+1} + b_{n-1})\tau^*_n + 1 = 0. \tag{4.5}$$

The first boundary condition for this recurrence relation is $\tau^*_0 = 0$, i.e. the mean first passage time to extinction starting from extinction is zero. The second condition is $\tau^*_N = \tau^*_{N-1} + 1/b_{N-1}$, i.e. the mean first passage time to extinction from N cells is the corresponding one from $N-1$ cells plus the meantime in which the system hops from N cells to $N-1$ cells.

To solve for τ^*_n we define $v_n = \tau^*_n - \tau^*_{n-1}$. Thus, the recurrence relation becomes

$$\left.\begin{aligned} a_{n+1}v_{n+1} - b_{n-1}v_n + 1 = 0 \\ \text{and} \qquad v_N = \frac{1}{b_{N-1}}, \end{aligned}\right\} \tag{4.6}$$

which can be solved iteratively (applying the corresponding boundary condition), to yield

$$v_{n-k} = \sum_{j=1}^{k+1} \frac{1}{b_{N-j}} \sum_{i=j}^{k} \frac{a_{N-i+1}}{b_{N-i-1}}. \tag{4.7}$$

Finally, solving for τ^*_n yields

$$\tau^*_n = \sum_{k=N-n}^{N-1} \sum_{j=1}^{k+1} \frac{1}{b_{N-j}} \prod_{i=j}^{k} \frac{a_{N-i+1}}{b_{N-i-1}}. \tag{4.8}$$

Naturally, the accuracy of the expected extinction time $\mathbb{T}$ (and its inverse, λ_1) is sensitive to the accuracy of the approximate PDF conditioned on non-extinction, $f(n)$. When calculating λ_1 from equations (4.1) and (4.8) (with $f(n)$ estimated numerically), the result is in very good agreement with λ_1 numerically calculated from the master operator's eigenvalues (compare the orange and black dashed lines in figure 5*b*). The approximate λ_1 calculated using equation (4.8), where $f(n)$ is obtained using one of the approximations for the probability distribution conditioned on non-extinction (figure 5*b*), is much closer to its real value than λ_1 calculated via equation (4.1) (figure 5*a*). For high ϕ^*, all $f(n)$ approximations lead to accurate λ_1 estimations, with the

exception of the half-Gaussian. On the other hand, the half-Gaussian approximation performs well for very low ϕ^*, when every other approximation of $f(n)$ yields inaccurate λ_1 estimations. In the intermediate region, the renormalized SSE turns out to be the best choice.

Extinction times vary significantly between parameter regimes, which we can think of as corresponding to different stem cell niches. Assuming that the system is in QSS, the average number of stem cells is roughly given by $\langle n \rangle \approx N\phi^*$. Hence, given the average number of stem cells in a niche, we can estimate the effective carrying capacity for different ϕ^* values from the QSS parameter regime. For example, for adult human crypt stem cell niches harbouring 4–6 cells with division period of 24–30 h [50,51], the extinction times fall between 13 days and 50 years when comparing the results for different ϕ^* values spanning all the QSS region. A predicted extinction time on this scale could point to the need of additional regulatory mechanisms preventing extinction of a tissue's stem cell population within the lifetime of an organism. For slow-cycling stem cells such as haematopoietic stem cells, the lowest estimates yield the presence of 10^4 cells dividing once every 40 weeks [52]. More recent estimates point to the number of hematopoietic stem cells in the human body falling within $5 \times 10^4 - 2 \times 10^5$ [53]. It is unclear how these stem cells are distributed in individual niches, but even for niches hosting 100 stem cells, the extinction times fall within 10^{11}–10^{13} years, much higher than the human lifetime. For niches hosting a larger number of stem cells, the extinction time is even higher.

5. Discussion and conclusion

We have introduced the vBD, a variation of the birth-death process that incorporates crowding effects due to the finite size of stem cell niches. For cell division to occur there needs to be free space in the niche to accommodate the newborn cell—hence the effective proliferation rate is higher (lower) when the niche is less (more) populated. In effect, the expected stem cell number in the vBD model is independent of the initial condition, in contrast to the CBD case, in which the expected number of stem cells is constant in time. Regulation through volume exclusion could also affect clone size distributions and their scaling with average clone size. In contrast to the CBD process, in which the size evolution of different clones is independent, in the vBD all clones are coupled through the empty space species. When one clone grows in size, the probability of other clones growing is decreased, thus affecting the clone-size distribution. Hence, it might be possible to detect evidence of volume exclusion effects from snapshots of clone size distributions, which are commonly measured experimentally [54–57]. We will explore this in follow-up research.

At the stochastic level, the predictions of the vBD master equation differ significantly from those of its deterministic counterpart. While the deterministic rate equations feature a stable steady state, ϕ^*, to which the system converges logistically, the master equation's solution predicts the vBD model converging to extinction for all parameter sets. However, for ϕ^* sufficiently large, a long-lived QSS appears, and convergence to extinction is very slow. Hence, for high ϕ^*, a single stochastic trajectory representing a real system might fluctuate around the QSS's non-trivial mode for all its lifetime.

We have shown the vBD model to have three phases with different behaviours that, to the best of our knowledge, have not been analysed before: fast extinction dynamics through a unimodal transient, fast extinction dynamics through a bimodal transient and slow extinction dynamics with a QSS. Transient bimodality rarely occurs in chemical reaction networks, but has been reported in recent works [58,59]. For the QSS region, we have shown two independent approximate solutions to the master equation, the QSS and WKB approximations. The QSS provides a more accurate approximation, whilst the WKB adopts a simpler mathematical expression. Moreover, the QSS approximation allowed us to prove that the position of the non-trivial mode generally differs from the deterministic prediction. For the bimodal extinction region, we have derived an approximate time-dependent solution to the master equation by making use of a renormalized SSE, which is particularly useful for solving master equations of non-linear birth-death processes, but has not been widely applied. Remarkably, the expression obtained by the renormalized SSE is simpler than the one from the regular SSE.

The vBD model is mathematically similar to susceptible-infected-susceptible (SIS) models in epidemiology [60–63]. Most of the stochastic SIS models studies consider a very large carrying capacity, or let it tend to infinity. Here, we have instead focused on solving the master equation for low carrying capacities, motivated by the application to stem cell population dynamics in niches that can have a carrying capacity as low as a few tens of cells. The approximate solutions to the master equations we present here, thus, can shed light into the behaviour of SIS-like systems when the effect of having a finite carrying capacity becomes evident. In particular, the solutions of the vBD master equation can be interpreted as the time-dependent PDF of the number of infected individuals, and our approximation of the expected extinction time as the population's expected time to recovery. Another similar model can be found when studying the role of positive feedback in cluster formation of signalling molecules [64]. In this case, the transition from quasi-steady to extinction states is interpreted as a switch from clustered to non-clustered states. A difference between such a model and the vBD is that transitions to non-clustered to clustered states are allowed, whilst in the vBD model it is not possible to exit the extinction state.

Our study proposes and describes a minimal model for stem cell dynamics with regulation through competition for space. In this context, the system parameters should be taken as effective parameters encompassing many different features. Let us conclude by discussing aspects of biological realism that could be represented explicitly in future extensions of our model. The vBD model assumes Markovian dynamics, which implies exponentially distributed waiting times between consecutive cell divisions. There is increasing evidence showing that the cell-cycle times are not exponentially distributed [65–69]. The inclusion of realistic cell-cycle time distributions in the vBD model will require further research. Spatially, our model considers cells distributed on a grid, assuming that the cell shapes remain unaltered. This hypothesis ignores the mechanical plasticity of cells, their mechanical response to pressure in different environments, and the irregular geometry of the stem cell niches [16,70]. Moreover, here we have modelled cell populations as well-mixed systems. This well-mixing assumption may be inappropriate when a niche is highly occupied. Since our aim is to study the hallmarks of regulation through volume exclusion, we have also neglected other plausible regulatory mechanisms that might be acting in parallel, such as competition for other resources [10,71,72] or cell–cell communication pathways between more differentiated (for example, transit amplifying) cells and stem cells [6,73]. In open niches these more differentiated cells might also compete with stem cells for space [10,72,74,75]. The role of multiple cell types in competition for space and the interaction between different plausible regulatory mechanisms are interesting avenues for future research.

Data accessibility. Code can be found at https://github.com/RodrigoGarciaTejera/vBD_simulations.git.

Authors' contributions. R.G.-T.: conceptualization, formal analysis, investigation, methodology, visualization, writing—original draft, writing—review and editing; L.S.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing—original draft, writing—review and editing; R.G.: conceptualization, formal analysis, investigation, methodology, resources, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. L.S. was supported by Chancellor's Fellowship from the University of Edinburgh. R.G-T. was supported by Chancellor's Fellow PhD. Studentship and Edinburgh Global Scholarship from the University of Edinburgh. R.G. was supported by Leverhulme Trust Grant number RPG-2018-423. For the purpose of open access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

1. Till JE, McCulloch EA, Siminovitch L. 1964 A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells. *Proc. Natl Acad. Sci. USA* **51**, 29–36. (doi:10.1073/pnas.51.1.29)
2. Tothova Z, Gilliland DG. 2007 Foxo transcription factors and stem cell homeostasis: insights from the hematopoietic system. *Cell Stem Cell* **1**, 140–152. (doi:10.1016/j.stem.2007.07.017)

3. Morrison SJ, Scadden DT. 2014 The bone marrow niche for haematopoietic stem cells. *Nature* **505**, 327–334. (doi:10.1038/nature12984)
4. North TE *et al.* 2007 Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011. (doi:10.1038/nature05883)
5. Becker NB, Günther M, A Jolly CL, Höfer T. 2019 Stem cell homeostasis by integral feedback through the niche. *J. Theor. Biol.* **481**, 100–109. (doi:10.1016/j.jtbi.2018.12.029)
6. Lander AD, Gokoffski KK, Wan FYM, Nie Q, Calof AL. 2009 Cell lineages and the logic of proliferative control. *PLoS Biol.* **7**, e1000015. (doi:10.1371/journal.pbio.1000015)
7. Renardy M, Jilkine A, Shahriyari L, Chou C-S. 2018 Control of cell fraction and population recovery during tissue regeneration in stem cell lineages. *J. Theor. Biol.* **445**, 33–50. (doi:10.1016/j.jtbi.2018.02.017)
8. Rodriguez-Brenes IA, Wodarz D, Komarova NL. 2013 Stem cell control, oscillations, and tissue regeneration in spatial and non-spatial models. *Front. Oncol.* **3**, 82. (doi:10.3389/fonc.2013.00082)
9. Yoshida S. 2018 Open niche regulation of mouse spermatogenic stem cells. *Dev. Growth Differ.* **60**, 542–552. (doi:10.1111/dgd.12574)
10. Kitadate Y *et al.* 2019 Competition for mitogens regulates spermatogenic stem cell homeostasis in an open niche. *Cell Stem Cell* **24**, 79–92. (doi:10.1016/j.stem.2018.11.013)
11. Hannezo E, Coucke A, Joanny J-F. 2016 Interplay of migratory and division forces as a generic mechanism for stem cell patterns. *Phys. Rev. E* **93**, 022405. (doi:10.1103/PhysRevE.93.022405)
12. Vining KH, Mooney DJ. 2017 Mechanical forces direct stem cell behaviour in development and regeneration. *Nat. Rev. Mol. Cell Biol.* **18**, 728–742. (doi:10.1038/nrm.2017.108)
13. Zhang Y, Wei H, Wen W. 2021 Phase separation and mechanical forces in regulating asymmetric cell division of neural stem cells. *Int. J. Mol. Sci.* **22**, 10267. (doi:10.3390/ijms221910267)
14. Tumbar T, Guasch G, Greco V, Blanpain C, Lowry WE, Rendl M, Fuchs E. 2004 Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363. (doi:10.1126/science.1092436)
15. Zhang J *et al.* 2003 Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841. (doi:10.1038/nature02041)
16. Li LL. Xie T. 2005 Stem cell niche: structure and function. *Annu. Rev. Cell Dev. Biol.* **21**, 605–631. (doi:10.1146/annurev.cellbio.21.012704.131525)
17. So W-K. Cheung TH. 2018 Molecular regulation of cellular quiescence: a perspective from adult stem cells and its niches. *Cell. Quiescence* **1686**, 1–25. (doi:10.1007/978-1-4939-7371-2_1)
18. Sampath SC *et al.* 2018 Induction of muscle stem cell quiescence by the secreted niche factor oncostatin M. *Nat. Commun.* **9**, 1–9. (doi:10.1038/s41467-018-03876-8)
19. Arai F, Suda T. 2008 Quiescent stem cells in the niche. *StemBook [Internet]*.
20. Stine RR, Matunis EL. 2013 Stem cell competition: finding balance in the niche. *Trends Cell Biol.* **23**, 357–364. (doi:10.1016/j.tcb.2013.03.001)
21. Corominas-Murtra B, Scheele CLGJ, Kishi K, Ellenbroek SIJ, Simons BD, Van Rheenen J, Hannezo E. 2020 Stem cell lineage survival as a noisy competition for niche access. *Proc. Natl Acad. Sci. USA* **117**, 16969–16975. (doi:10.1073/pnas.1921205117)
22. Smith S, Cianci C, Grima R. 2017 Macromolecular crowding directs the motion of small molecules inside cells. *J. R. Soc. Interface* **14**, 20170047. (doi:10.1098/rsif.2017.0047)
23. Wilson D, Byrne H, Bruna M. 2017 Reactions, diffusion and volume exclusion in a heterogeneous system of interacting particles. (http://arxiv.org/abs/1705.00004)
24. Dyson L, Baker RE. 2015 The importance of volume exclusion in modelling cellular migration. *J. Math. Biol.* **71**, 691–711. (doi:10.1007/s00285-014-0829-0)
25. Flory PJ, Fisk S. 1966 Effect of volume exclusion on the dimensions of polymer chains. *J. Chem. Phys.* **44**, 2243–2248. (doi:10.1063/1.1727029)
26. Segall DE, Nelson PC, Phillips R. 2006 Volume-exclusion effects in tethered-particle experiments: bead size matters. *Phys. Rev. Lett.* **96**, 088306. (doi:10.1103/PhysRevLett.96.088306)
27. Hillen T, Enderling H, Hahnfeldt P. 2013 The tumor growth paradox and immune system-mediated selection for cancer stem cells. *Bull. Math. Biol.* **75**, 161–184. (doi:10.1007/s11538-012-9798-x)
28. Klein AM, Simons BD. 2011 Universal patterns of stem cell fate in cycling adult tissues. *Development* **138**, 3103–3111. (doi:10.1242/dev.060103)
29. Gillespie DT. 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434. (doi:10.1016/0021-9991(76)90041-3)

30. Schnoerr D, Sanguinetti G, Grima R. 2015 Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.* **143**, 11B610_1. (doi:10.1063/1.4934990)
31. Harris TE. 1963 *The theory of branching processes*, vol. 6. Berlin: Springer.
32. Grima R. 2015 Linear-noise approximation and the chemical master equation agree up to second-order moments for a class of chemical systems. *Phys. Rev. E* **92**, 042124. (doi:10.1103/PhysRevE.92.042124)
33. Simons BD, Clevers H. 2011 Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **145**, 851–862. (doi:10.1016/j.cell.2011.05.033)
34. Yamaguchi H, Kawaguchi K, Sagawa T. 2017 Dynamical crossover in a stochastic model of cell fate decision. *Phys. Rev. E* **96**, 012401. (doi:10.1103/PhysRevE.96.012401)
35. Robertson NA *et al.* 2021 Longitudinal dynamics of clonal hematopoiesis identifies gene-specific fitness effects. *bioRxiv*. (doi:10.1101/2021.05.27.446006)
36. Gardiner C. 1963 *Stochastic methods: a handbook for the natural and social sciences*, 4th edn. Berlin: Springer.
37. Schnoerr D, Sanguinetti G, Grima R. 2017 Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A: Math. Theor.* **50**, 093001. (doi:10.1088/1751-8121/aa54d9)
38. Van Kampen NG. 1992 *Stochastic processes in physics and chemistry*, vol. 1. Amsterdam: Elsevier.
39. Smith S, Shahrezaei V. 2015 General transient solution of the one-step master equation in one dimension. *Phys. Rev. E* **91**, 062119. (doi:10.1103/PhysRevE.91.062119)
40. Thomas P, Grima R. 2015 Approximate probability distributions of the master equation. *Phys. Rev. E* **92**, 012120. (doi:10.1103/PhysRevE.92.012120)
41. Van Kampen NG. 1976 The expansion of the master equation. *Adv. Chem. Phys.* **34**, 245–311.
42. Grima R. 2009 Noise-induced breakdown of the Michaelis–Menten equation in steady-state conditions. *Phys. Rev. Lett.* **102**, 218103. (doi:10.1103/PhysRevLett.102.218103)
43. Grima R. 2010 An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions. *J. Chem. Phys.* **133**, 07B604. (doi:10.1063/1.3454685)
44. Andrews GE. 1998 *The theory of partitions*, vol. 2. Cambridge, UK: Cambridge University Press.
45. Cianci C, Fanelli D, McKane AJ. 2015 WKB versus generalized van Kampen system-size expansion: the stochastic logistic equation. (http://arxiv.org/abs/1508.00490)
46. Assaf M, Meerson B. 2006 Spectral formulation and WKB approximation for rare-event statistics in reaction systems. *Phys. Rev. E* **74**, 041115. (doi:10.1103/PhysRevE.74.041115)
47. Assaf M, Meerson B. 2017 WKB theory of large deviations in stochastic populations. *J. Phys. A: Math. Theor.* **50**, 263001. (doi:10.1088/1751-8121/aa669a)
48. Ashcroft P. 2015 *The statistical physics of fixation and equilibration in individual-based models*. United Kingdom: The University of Manchester.
49. Holehouse J, Pollitt H. 2021 Non-equilibrium time-dependent solution to discrete choice with social interactions. (http://arxiv.org/abs/2109.09633)
50. Barker N *et al.* 2009 Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611. (doi:10.1038/nature07602)
51. Umar S. 2010 Intestinal stem cells. *Curr. Gastroenterol. Rep.* **12**, 340–348. (doi:10.1007/s11894-010-0130-3)
52. Catlin SN, Busque L, Gale RE, Guttorp P, Abkowitz JL. 2011 The replication rate of human hematopoietic stem cells *in vivo*. *Blood* **117**, 4460–4466. (doi:10.1182/blood-2010-08-303537)
53. Lee-Six H *et al.* 2018 Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478. (doi:10.1038/s41586-018-0497-0)
54. Buckingham ME, Meilhac SM. 2011 Tracing cells for tracking cell lineage and clonal behavior. *Dev. Cell* **21**, 394–409. (doi:10.1016/j.devcel.2011.07.019)
55. Kretzschmar K, Watt FM. 2012 Lineage tracing. *Cell* **148**, 33–45. (doi:10.1016/j.cell.2012.01.002)
56. Kester L, van Oudenaarden A. 2018 Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* **23**, 166–179. (doi:10.1016/j.stem.2018.04.014)
57. Blanpain C, Simons BD. 2013 Unravelling stem cell dynamics by lineage tracing. *Nat. Rev. Mol. Cell Biol.* **14**, 489–502. (doi:10.1038/nrm3625)
58. Jia C, Grima R. 2020 Dynamical phase diagram of an auto-regulating gene in fast switching conditions. *J. Chem. Phys.* **152**, 174110. (doi:10.1063/5.0007221)
59. Holehouse J, Sukys A, Grima R. 2020 Stochastic time-dependent enzyme kinetics: closed-form solution and transient bimodality. *J. Chem. Phys.* **153**, 164113. (doi:10.1063/5.0017573)

60. Kryscio RJ, Lefévre C. 1989 On the extinction of the S–I–S stochastic logistic epidemic. *J. Appl. Probab.* **26**, 685–694. (doi:10.2307/3214374)
61. Jacquez JA, Simon CP. 1993 The stochastic SI model with recruitment and deaths I. Comparison with the closed sis model. *Math. Biosci.* **117**, 77–125. (doi:10.1016/0025-5564(93)90018-6)
62. Cao B, Shan M, Zhang Q, Wang W. 2017 A stochastic sis epidemic model with vaccination. *Physica A* **486**, 127–143. (doi:10.1016/j.physa.2017.05.083)
63. Nåsell I. 2011 *Extinction and quasi-stationarity in the stochastic logistic SIS model*. Berlin: Springer.
64. Jilkine A, Angenent SB, Wu LF, Altschuler SJ. 2011 A density-dependent switch drives stochastic clustering and polarization of signaling molecules. *PLoS Comput. Biol.* **7**, e1002271. (doi:10.1371/journal.pcbi.1002271)
65. Perez-Carrasco R, Beentjes C, Grima R. 2020 Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance. *J. R. Soc. Interface* **17**, 20200360. (doi:10.1098/rsif.2020.0360)
66. Weber TS, Jaehnert I, Schichor C, Or-Guil M, Carneiro J. 2014 Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. *PLoS Comput. Biol.* **10**, e1003616. (doi:10.1371/journal.pcbi.1003616)
67. Stumpf PS *et al.* 2017 Stem cell differentiation as a non-Markov stochastic process. *Cell Syst.* **5**, 268–282. (doi:10.1016/j.cels.2017.08.009)
68. Gavagnin E, Ford MJ, Mort RL, Rogers T, Yates CA. 2019 The invasion speed of cell migration models with realistic cell cycle time distributions. *J. Theor. Biol.* **481**, 91–99. (doi:10.1016/j.jtbi.2018.09.010)
69. Chao HX *et al.* 2019 Evidence that the human cell cycle is a series of uncoupled, memoryless phases. *Mol. Syst. Biol.* **15**, e8604. (doi:10.15252/msb.20188604)
70. Bonakdar N, Gerum R, Kuhn M, Spörrer M, Lippert A, Schneider W, Aifantis KE, Fabry B. 2016 Mechanical plasticity of cells. *Nat. Mater.* **15**, 1090–1094. (doi:10.1038/nmat4689)
71. Jörg DJ, Kitadate Y, Yoshida S, Simons BD. 2021 Stem cell populations as self-renewing many-particle systems. *Annu. Rev. Condens. Matter Phys.* **12**, 135–153. (doi:10.1146/annurev-conmatphys-041720-125707)
72. Kitadate Y, Yoshida S. 2022 Regulation of spermatogenic stem cell homeostasis by mitogen competition in an open niche microenvironment. *Genes Genet. Syst.* **97**, 15–25. (doi:10.1266/ggs.21-00062)
73. Lo W-C, Chou C-S, Gokoffski KK, Wan FY-M, Lander AD, Calof AL, Nie Q. 2009 Feedback regulation in multistage cell lineages. *Math. Biosci. Eng.* **6**, 59.
74. Crane GM, Jeffery E, Morrison SJ. 2017 Adult haematopoietic stem cell niches. *Nat. Rev. Immunol.* **17**, 573–590. (doi:10.1038/nri.2017.53)
75. Kokkaliaris KD, Kunz L, Cabezas-Wallscheid N, Christodoulou C, Renders S, Camargo F, Trumpp A, Scadden DT, Schroeder T. 2020 Adult blood stem cell localization reflects the abundance of reported bone marrow niche cell types and their combinations. *Blood* **136**, 2296–2307. (doi:10.1182/blood.2020006574)