

## RESEARCH ARTICLE

# Solving the stochastic dynamics of population growth

Loïc Marrec<sup>1,2</sup>  | Claudia Bank<sup>1,2</sup> | Thibault Bertrand<sup>3</sup> <sup>1</sup>Institut für Ökologie und Evolution, Universität Bern, Bern, Switzerland<sup>2</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland<sup>3</sup>Department of Mathematics, Imperial College London, London, UK**Correspondence**Loïc Marrec, Institut für Ökologie und Evolution, Universität Bern, Baltzerstrasse 6, Bern CH-3012, Switzerland.  
Email: [loic.marrec@unibe.ch](mailto:loic.marrec@unibe.ch)**Funding information**

H2020 European Research Council, Grant/Award Number: 804569829; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (MiCo4Sys), Grant/Award Number: 315230\_204838/1

**Abstract**

Population growth is a fundamental process in ecology and evolution. The population size dynamics during growth are often described by deterministic equations derived from kinetic models. Here, we simulate several population growth models and compare the size averaged over many stochastic realizations with the deterministic predictions. We show that these deterministic equations are generically bad predictors of the average stochastic population dynamics. Specifically, deterministic predictions overestimate the simulated population sizes, especially those of populations starting with a small number of individuals. Describing population growth as a stochastic birth process, we prove that the discrepancy between deterministic predictions and simulated data is due to unclosed-moment dynamics. In other words, the deterministic approach does not consider the variability of birth times, which is particularly important with small population sizes. We show that some moment-closure approximations describe the growth dynamics better than the deterministic prediction. However, they do not reduce the error satisfactorily and only apply to some population growth models. We explicitly solve the stochastic growth dynamics, and our solution applies to any population growth model. We show that our solution exactly quantifies the dynamics of a community composed of different strains and correctly predicts the fixation probability of a strain in a serial dilution experiment. Our work sets the foundations for a more faithful modeling of community and population dynamics. It will allow the development of new tools for a more accurate analysis of experimental and empirical results, including the inference of important growth parameters.

**KEYWORDS**

community dynamics, fixation probability, master equation, moment closure approximation, Population dynamics

**TAXONOMY CLASSIFICATION**

Community ecology, Demography, Population ecology, Theoretical ecology

## 1 | INTRODUCTION

Population growth is at the heart of fundamental processes in cell biology, evolution, and ecology, from the expansion of bacteria colonies and large-scale animal populations to the propagation of an advantageous

mutation. Predicting population growth dynamics is thus paramount in multiple fields. The outspoken goal of population modeling is to accurately describe the variation in the number of individuals in a population.

Historically, deterministic models are most commonly used to describe population dynamics (Anderson, 2014; Brauer &

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Castillo-Chavez, 2013; Hannon et al., 1997). In these models, the population size is generically described by a continuous variable whose temporal dynamics are governed by an ordinary differential equation. Whereas most of these models are nonlinear—which means that analytical progress is not impossible but limited in some cases (Tsoularis & Wallace, 2002)—it is often relatively simple and computationally fast to obtain accurate numerical solutions, possibly explaining their widespread use. A paradigmatic model of population growth in theoretical ecology is the well-known logistic equation whose study traces back to as early as the middle of the nineteenth century (Verhulst, 1838). The logistic differential equation was initially derived from introducing a self-limiting property in the growth of a biological population to the unconstrained Malthusian exponential growth model (Malthus, 1798). It was re-discovered independently later on (Lotka, 1925; McKendrick & Pai, 1912; Pearl & Reed, 1920). Verhulst's logistic growth model derivation stemmed from the observation that unhindered exponential population growth is unrealistic. Even in the absence of predation relations, intraspecies competition for environmental resources such as food or habitat will lead to a characteristic saturation level, an upper bound on the population size known as the *carrying capacity*. Owing to its ease of use, the simplest logistic growth was used to model biological systems at all scales, from the population growth of micro-organisms (Carlson, 1913; Pearl, 1927) to that of large mammal herds (Morgan, 1976) and fish schools (Krebs, 1985).

Further refinements to the logistic growth function led to the development of a *generalized logistic growth model* (Tsoularis & Wallace, 2002), which captures several commonly used population growth models including Blumberg (Blumberg, 1968), Richards (Richards, 1959) and Gompertz (Gompertz, 1825) growth models. Whereas amenable to easy progress and qualitative predictions, these deterministic models are not entirely faithful to the growth of a natural population, which is inherently stochastic (Bartlett, 1960; McKendrick, 1925). This stochasticity results from both intrinsic (e.g., demography) and extrinsic (e.g., environmental change) noise (Haefner, 2012; Lande et al., 2003). More recent studies have shown that deterministic and stochastic approaches yield critically different results (Allen & Burgin, 2000; Baker & Simpson, 2010; Wakano & Iwasa, 2013; Wilson, 1998).

Although it is often assumed that a deterministic equation can describe the dynamics of a large-volume stochastic system, it is clear that this criterion alone is not sufficient (Gustafsson & Sternad, 2013; Kurtz, 1972). The range of validity of deterministic models is put in question. Even if new conditions for a deterministic equation to describe well the stochastic dynamics of a population have been outlined (Gustafsson & Sternad, 2013), they are not exhaustive and quantitative methods to overcome this discrepancy are missing. Recognizing these limitations, stochastic models have proved helpful in ecology for the past decades (Black & McKane, 2012; Bolker & Pacala, 1997; Keeling, 2000b; Marion et al., 1998; Matis & Kiffe, 1999).

Many studies use deterministic equations to fit experimental and empirical data, allowing the estimation of essential biological

parameters. For example, logistic growth models have been used in microbiology to estimate microbial division rates (Kahm et al., 2010; Sprouffske & Wagner, 2016). A prediction based on deterministic models carries the risk of poorly estimating parameters of interest, such as the division rate of antimicrobial-resistant bacteria, crucial to implementing political measures to slow the spread of antimicrobial resistance. Deterministic models were also used recently to estimate the basic reproductive number during viral outbreaks; those used logistic-like equations (Aviv-Sharon & Aharoni, 2020; Pelinovsky et al., 2020; Shen, 2020; Wu et al., 2020), as well as compartmental models such as SEIR (Shen et al., 2020; Sunhwa & Moran, 2020; Wan et al., 2020). Identifying when a deterministic equation does not correctly describe the average dynamics of stochastic population growth, understanding the reasons for this disagreement, and proposing solutions to remedy it, is thus of paramount importance.

In this work, we leap forward by solving the stochastic dynamics of population growth in the absence of deaths analytically. This resolution allows us to identify the extent to which a deterministic approach is a good approximation of the growth dynamics and to lay the foundation for future inference methods of growth parameters based, for example, on the likelihood function calculation. We consider several classical population growth models. First, we model the population growth as a stochastic birth process and simulate stochastic realizations of these kinetics. We compare their ensemble average to the predictions of the respective deterministic models. We show that the deterministic approach generically overestimates the average population size. This prediction error is larger when the initial number of individuals is very low. To explain the reason behind this discrepancy, we derive a master equation formalism describing the stochastic population growth dynamics and the moment equations. We find that the difference between the population size estimated by the deterministic equation and the mean of the simulated data is due to unclosed moment dynamics. We show that some moment-closure approximations reduce the difference, although it remains globally substantial. Instead, we derive an exact solution to the stochastic population growth. Finally, we apply our results and show that our solution leads to a better prediction of the dynamics of two competing strains and the probability of fixation of a mutant in a serial passage experiment.

## 2 | BIAS OF DETERMINISTIC APPROACHES

### 2.1 | Pure-birth models

Given the inherent stochasticity of population growth processes, we first establish whether a deterministic equation correctly describes the mean trajectory of stochastic growth. We consider four distinct growth models belonging to the *generalized logistic growth models*: Blumberg, Gompertz, Logistic, and Richards models (Tsoularis & Wallace, 2002). Our choice was motivated by their widespread use

to fit experimental or empirical data to estimate growth parameters in microbiology and ecological communities (Ghenu et al., 2022; López et al., 2004; Sprouffske & Wagner, 2016). These kinetic models differ by their per capita growth rates,  $b_N$ .

Under Malthusian growth, the per capita growth rate is constant and independent of the population size  $N$ ; we denote as  $b$  this intrinsic birth rate—also called *exponential birth rate* or *Malthusian parameter*. Note that simple unbounded exponential growth would occur if no restrictions were imposed on the population size (e.g., nutrients, available space). In the growth models considered here, the per capita growth rate is explicitly dependent on the population size. To model environmental constraints such as availability of space or food, one then generically introduces a carrying capacity  $K$  that limits the population size  $N$  to the range  $N_0 \leq N \leq K$  assuming no deaths, where  $N_0$  is the initial population size. Specifically, the per capita birth rate  $b_N$  decreases as the population size increases and vanishes when  $N = K$ .

For a general Logistic growth model, the deterministic equation describing the dynamics of  $N$  reads

$$\frac{dN}{dt} = B_N = bN^\alpha \left[ 1 - \left( \frac{N}{K} \right)^\beta \right]^\gamma, \quad (1)$$

where  $B_N$  denotes the population growth rate. The population growth rates for each of the four nonlinear models considered here are provided in Table 1 along with those of the exponential growth model.

As seen in Figure 1, the four population growth models chosen here display very different birth rate curves. Birth rates in all models are non-monotonic and vanish when  $N \rightarrow 0$  and  $N \rightarrow K$  by construction. In other words, these models impose that a population of size zero cannot grow, and no population can grow beyond the carrying capacity. We note that the Logistic model displays a symmetry around the population size  $N = K/2$ , whereas in both the Blumberg and Richards models, the exponents  $\beta$  and  $\gamma$ , respectively, offer an extra degree of freedom to tune the shape of the growth rate curve and in particular, its asymmetry. We note that the population size at inflection, that is, when the population growth rate is maximum, is given by

$$N^* = K \left( 1 + \frac{\beta\gamma}{\alpha} \right)^{-1/\beta}, \quad (2)$$

TABLE 1 Population growth models

Exponential	Logistic	Blumberg	Richards	Gompertz <sup>a</sup>
$\alpha = 1$	$\alpha = 1$	$\alpha$	$\alpha = 1$	$\alpha = 1$
$\beta = 0$	$\beta = 1$	$\beta = 1$	$\beta$	$\beta \rightarrow 0$
$\gamma = 0$	$\gamma = 1$	$\gamma$	$\gamma = 1$	$\gamma = 1$
$B_N = bN$	$B_N = bN(1 - N/K)$	$B_N = bN^\alpha(1 - N/K)^\gamma$	$B_N = bN(1 - (N/K)^\beta)$	$B_N = bN \log(K/N)$

Note: List of specific population growth models used with their associated exponents and population birth rates  $B_N$ . These are derived from the *generalized logistic growth model* introduced in Equation (1). The exponents  $\alpha$ ,  $\beta$ , and  $\gamma$  allow tuning the symmetry, maximum, and inflection of the population growth rate  $B_N$  (see Figure 1). Birth rate, population size, and carrying capacity are denoted by  $b$ ,  $N$ , and  $K$ , respectively.

<sup>a</sup>The generalized logistic growth model converges to the Gompertz model when the per capita growth rate is divided by  $\beta$  and the limit  $\beta \rightarrow 0$  is taken (Tsoularis & Wallace, 2002).

and is thus dependent on exponents  $\beta$  and  $\gamma$  for the Blumberg and Richards models, respectively. The Gompertz model shows the fastest growth of all at small population sizes. As shown in Figure 2a, the population size in all deterministic models follows a sigmoid curve—also called an S-shape curve—with its characteristic initial phase with slow growth, exponential growth phase, finally followed by a stabilization of the population size at a finite steady-state population size

$$t \infty \lim N(t) = K. \quad (3)$$

For the Logistic, Richards, and Gompertz models, full analytical solutions are available (Tsoularis & Wallace, 2002). These are given by

$$N_L(t) = \frac{K}{1 + (K/N_0 - 1)e^{-bt}}, \quad (4a)$$

$$N_R(t) = \frac{K}{\left[ 1 + \left( (K/N_0)^\beta - 1 \right) e^{-\beta bt} \right]^{1/\beta}}, \quad (4b)$$

$$N_G(t) = K(N_0/K)^{e^{-bt}}. \quad (4c)$$

Those are represented in Figure 2a. Note that no analytical (i.e., closed-form) solution exists for the Blumberg model; in this last case, we proceeded with a direct numerical solution of Equation (1).

However, as pointed out earlier, population growth is inherently stochastic. We, therefore, evaluated the validity range of the above deterministic descriptions by simulating individual stochastic trajectories for the four growth models introduced above using a Gillespie algorithm (Gillespie, 1976, 1977) (see Appendix A). To this end, we recast our problem into a pure-birth process for which population growth results from an individual  $A$  reproducing at a size-dependent rate  $b_N$  following the elementary reaction

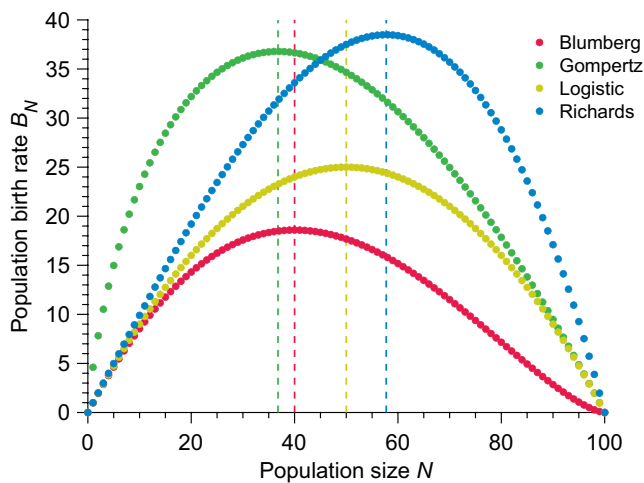


with  $B_N = b_N \times N$ . Naturally, multiple stochastic models may lead to the same deterministic model under a mean-field approximation. Here, we focus on one particular microscopic scenario. Still, as we will argue in the next section, other formulations, including those where birth and death processes are taken into account explicitly, lead to even more drastic disagreement. We average all our results over  $10^5$  independent

stochastic trajectories to obtain the time-dependent average population size. Note that we only consider the case of  $\alpha = 1$  to ensure that we can faithfully match the stochastic models to their deterministic limits. Indeed, these are the only cases for which we obtain a well-defined deterministic limit (see Appendix B for details).

As shown in Figure 2a, we observe a substantial difference between the deterministic predictions and the stochastic mean population size starting from a single individual. To quantify this disagreement, we calculated the relative error  $\eta$  defined as

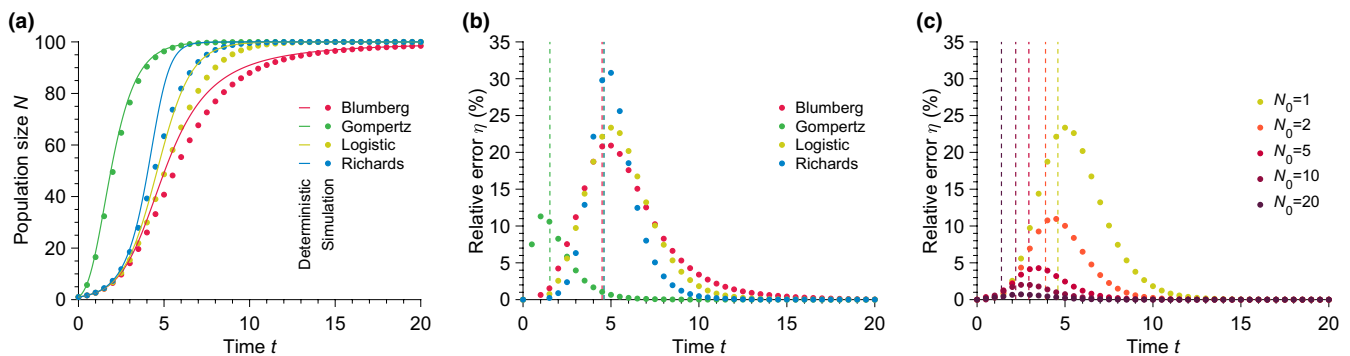
$$\eta(t) = \frac{N_d(t) - N_s(t)}{N_s(t)}, \quad (6)$$



**FIGURE 1** Different growth models display different population birth rates. Population birth rate  $B_N$  versus population size  $N$  for different population growth models. Vertical dashed lines show the location of the optimal birth rate for each model. Parameter values: carrying capacity  $K = 100$ , birth rate  $b = 1$ , exponents  $\alpha = 1$ ,  $\beta = 2$ , and  $\gamma = 1.5$ .

where  $N_d(t)$  and  $N_s(t)$  are, respectively, the time-dependent population sizes predicted by the deterministic model and measured in our stochastic simulations. We observed relative errors as large as 30% independent of the carrying capacity (see Appendix D, Figure A2). Figure 2b shows that the largest error for the parameters chosen is obtained for the Richards model, whereas the smallest is for the Gompertz model (still at a substantial  $\max(\eta) \approx 12\%$ ). We note that the quantitative value of the relative errors measured for the Blumberg and Richards models depend on the choice of exponents  $\beta$  and  $\gamma$ , respectively. Nevertheless, the measured errors remain substantial over a wide range of exponents (see Appendix C, Figure A1). For all the growth models studied here, the deterministic dynamics given by Equation (1) overestimate the population size at all times. It is interesting to note that the maximum error is located around the inflection point (i.e., at  $t$  such that  $d^2N/dt^2 = 0$ ) predicted by the deterministic equation. The inflection point corresponds to the population size where the population birth rate reaches its maximum and starts decreasing.

Furthermore, we observe that the error uniformly decreases as  $N_0$ , the initial number of individuals in the population, increases (see Figure 2c). This discrepancy limits the range of validity of the deterministic models as the initial number of individuals in the population is assumed to be small in many applications, for example, patient zero in a disease spreading scenario, single cell mutation in a mutation fixation experiment, small number of cells in a bacterial colony expansion, etc. To summarize, the deterministic equation fails to describe the average stochastic trajectory. Although the relative error depends on the specific growth model, and thus on the per capita birth rate, it remains substantial in all cases tested and increases with decreasing initial population size. Importantly, the relative error between deterministic predictions and measured mean stochastic population size is independent of the carrying capacity. Thus the discrepancy does not vanish in the limit of large but finite population sizes (see Appendix D, Figures A2 and A3).



**FIGURE 2** Solutions obtained from deterministic models fail to predict the mean stochastic population size. (a) Population size  $N$  versus time  $t$  for different growth models. The data points show simulated data averaged over  $10^4$  stochastic realizations. The solid lines correspond to the analytical solutions given in Equation (4a)–(4c) for the Logistic, Richards, and Gompertz models and a direct numerical resolution of the Blumberg model (see Equation (1)). (b) Relative error  $\eta$  versus time  $t$  for different growth models calculated using Equation (6) and data from (a). The vertical dashed lines represent the inflection times at which  $d^2N/dt^2 = 0$  or equivalently when the population growth rate reaches a maximum. (c) Relative error  $\eta$  versus time  $t$  for different initial population sizes  $N_0$  for the Logistic model. The data points show simulated data averaged over  $10^5$  stochastic realizations. The vertical dashed lines represent the inflection times. Parameter values: carrying capacity  $K = 100$ , initial population size  $N_0 = 1$  (in (a) and (b)), birth rate  $b = 1$ , exponents  $\alpha = 1$ ,  $\beta = 2$  and  $\gamma = 1.5$ .

Stochastic population growth is a Markovian jump process. Therefore, the times between jumps from size  $N$  to  $N + 1$  are exponentially distributed with parameter  $B_N$ . For small initial population sizes  $N_0$ , the rates at which the population initially grows are low (see Figure 1), e.g., the rate of the first reproduction is given by  $bN_0[1 - (N_0/K)^\beta]^\gamma$  for the generalized logistic model. In turn, this implies that early in the process, the distributions of reproduction times lead to a large variance in the population size. We postulate that this large variance accumulated over the growth process is responsible for the disagreement between the deterministic and the mean stochastic trajectories. This postulate is consistent with our observations that: (1) the relative error increases when the initial population size decreases, and (2) the relative error is maximal around the inflection point where the exponential distribution of reproduction times is the tightest. Interestingly, when  $K \rightarrow \infty$  (large volume limit), the rate of first reproduction converges to  $bN_0$  and so is entirely controlled by the initial population size confirming that the observed disagreement remains valid in this limit. Finally, we note that the above deterministic models are mean-field models which intrinsically assume an underlying population size distribution peaked around its mean in contrast to the wide population size distributions observed in the stochastic models.

## 2.2 | Birth-death processes

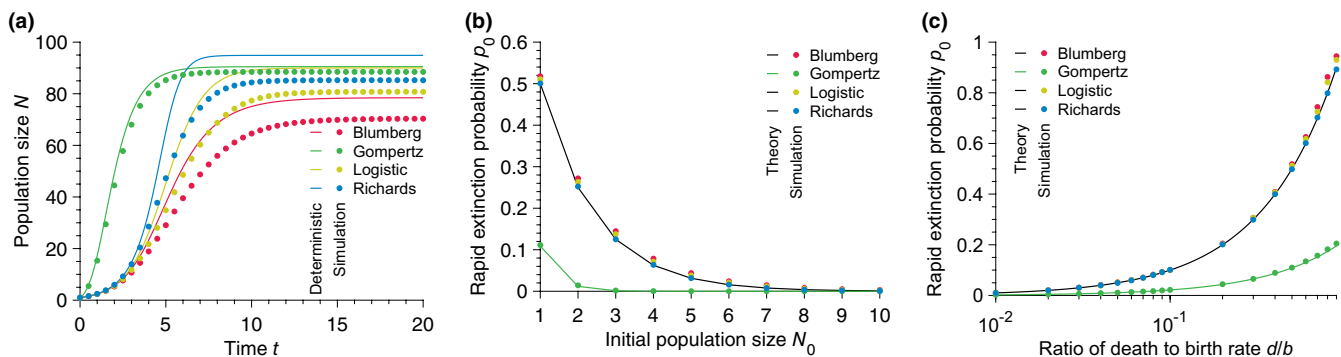
In the previous section, we focused on populations that can only increase in size over time. This assumption, which entails neglecting the death of individuals, is predominant in microbiology, where the models used to fit population growth data do not explicitly include death rates (Kahm et al., 2010; Ram et al., 2019; Sprouffske & Wagner, 2016). Similarly, pharmacodynamic models, which aim at quantifying how antibiotics inhibit growth or kill cells, commonly replace the birth rate with a net birth rate (i.e., birth rate minus death rate) in Equation (1) (Czock & Keller, 2007; Regoes et al., 2004). In this way, the population grows if the net birth rate is positive,

decreases if it is negative, and remains constant for zero net growth rates. However, stochastic population growth can also be modeled as a Markovian jump process where births and deaths are distinct events leading to distinct changes in population size:  $N \rightarrow N + 1$  and  $N \rightarrow N - 1$ , respectively.

We do not expect deterministic models to fare better in the presence of explicit death events with small initial population sizes. For the sake of simplicity and without loss of generality, we add to Equation (1) a linear death term leading to the modified differential equation  $dN/dt = bN^\alpha[1 - (N/K)^\beta]^\gamma - dN$ . We further simulate a stochastic birth-death process known to lead to this deterministic equation in the mean-field limit using once again a Gillespie algorithm (Gillespie, 1976, 1977). We show that a deterministic approach fails to describe the dynamics of the average population growth over many stochastic realizations, as in the death-free case. Strikingly, we demonstrate that it also fails to predict the correct steady-state population size.

Indeed, Figure 3a shows that deterministic models do not predict quantitatively either time-dependent population sizes or steady-state population sizes averaged over many stochastic realizations. We argue that this difference is due to rapid extinctions, which are frequent occurrences when considering low initial population sizes and large ratios of death rate to birth rate (see Figures 3b,c). Specifically, demographic stochasticity leads to extinction with probability  $p_0 = (d/b)^{N_0}$  (or  $p_0 = (d/(\log(K)))^{N_0}$  for the Gompertz model), which is non-zero as long as the death rate is strictly positive. Note that the probability of extinction is obtained from the linear birth-death process (Kendall, 1948), assuming density independence at the beginning of the growth when the population size is very small compared to the carrying capacity. For nonlinear models, these early extinction events are not taken into account in deterministic approaches.

One may wonder whether a fairer approach to matching the stochastic models with the deterministic limits should be to first condition the master equation on survival. One can show that this does not lead to a substantial reduction in the error in predicting



**FIGURE 3** Deterministic approach performs worse with non-zero death rates. (a) Population size  $N$  versus time  $t$  for different population growth models. (b) Rapid extinction probability  $p_0$  versus initial population size  $N_0$  for different population growth models. (c) Rapid extinction probability  $p_0$  versus ratio of death to birth rate  $d/b$  for different population growth models. In every panel, the solid lines the deterministic predictions, and each point results from simulated data averaged over  $10^5$  stochastic realizations. Parameter values: carrying capacity  $K = 100$ , birth rate  $b = 1$ , death rate  $d = 0.1$  (in (a)) and  $0.5$  (in (b)), and initial population size  $N_0 = 1$  (in (a) and (c)).

the transient population size but allows one to correctly predict the quasi-steady-state population size (see Appendix D for details). Furthermore, the linear birth-death process, whose analytical solution is known (Kendall, 1948), has a deterministic limit giving the population size averaged over the survival and early extinction trajectories (see Appendix D for details).

In summary, we conclude that deterministic formalism is not a good predictor of the average population growth dynamics, even for large carrying capacities in the presence or the absence of explicit death events. We also note that the discrepancy between deterministic and average stochastic population sizes worsens as the initial population size decreases. We further conclude that deterministic approaches fail at predicting the steady-state population size when death events are explicitly introduced. In the following, we focus on the pure-birth process, which is already of great interest in microbiology, as we pointed out earlier. In the next section, we adopt a stochastic approach to describe the population growth and obtain an exact analytical solution for the population size distribution at all times.

### 3 | ERROR REDUCTION BY MOMENT-CLOSURE APPROXIMATIONS

To identify the reasons behind the poor performance of the deterministic equation, we return to a stochastic formalism. Generically, any population growth in the absence of death may be described by a stochastic birth process whose rates are defined by the model (Kendall, 1949) (see Table 1 for examples). Let us consider a population whose number of individuals at time  $t$  is denoted by  $N$ , whereas its initial population size is  $N_0$ . As in Equation (5), we consider that each individual in the population replicates with the same per capita rate  $b_N$ . Here, the population size increases from  $N$  to  $N + 1$  individuals at a total rate  $B_N$ , where  $B_N$  was defined in Table 1 for several population growth models.

We focus on finite-sized populations that grow in a constant environment with a carrying capacity  $K$ . To fully account for the stochasticity inherent to demographic noise, we use a microscopic and probabilistic description in continuous time of the birth events within the population. More specifically, we write a system of differential equations describing the probabilities  $P_{N_0,N}(t)$  that a population has a given size  $N$  at a given time  $t$  knowing that it started with  $N_0$  individuals. Because the growth rates vanish when  $N \rightarrow K$ , the size of our population is at most  $K$ , with the state  $N = K$  being an absorbing state, that is, the population indefinitely remains in this state once reaching it for the first time. Put simply, our stochastic process, while continuous in time, has a finite discrete number of possible states. Here, we assume that the population jumps between successive sizes with a rate dependent on its current size leading to a fully coupled system of equations.

This system of differential equations, formally called the *master equation*, governs the time evolution of the probabilities  $P_{N_0,N}(t)$ . Writing the master equation for a stochastic jump process requires

one to think about gain and loss terms to the probabilities  $P_{N_0,N}(t)$ ; for our system, it reads (Gardiner, 2009; Van Kampen, 2011)

$$\frac{dP_{N_0,N}(t)}{dt} = B_{N-1}P_{N_0,N-1}(t) - B_N P_{N_0,N}(t), \quad (7)$$

when  $N_0 \leq N \leq K - 1$  and the probability normalization condition imposes  $\sum_{N=N_0}^K P_{N_0,N}(t) = 1 - \sum_{N=N_0}^{K-1} P_{N_0,N}(t)$ , when  $N = K$ . The first term on the right-hand side of Equation (7) is a gain term corresponding to an increase in the population size from  $N-1$  to  $N$  individuals via a birth event, whereas the second term is a loss term corresponding to the population size transitioning from  $N$  to  $N + 1$  individuals. Although writing down the master equation for a stochastic jump process is often easy, computing the formal solution of master equations is arduous and has been an active field of investigation for decades (Gardiner, 2009; Van Kampen, 2011; Weber & Frey, 2017).

Importantly, the master Equation (7) contains all information about the growth dynamics; in particular, as it governs the probability distribution  $P_{N_0,N}(t)$ , it contains all information to compute the averaged population size trajectory over time. Rather than solving directly Equation (7), we derive an equation governing the moments of  $P_{N_0,N}(t)$ , where the  $m^{\text{th}}$  moment of the population size is defined as

$$\langle N^m \rangle = \sum_{N=N_0}^K N^m P_{N_0,N}(t), \quad (8)$$

with  $m$  a positive integer. For instance, the equation governing the time-evolution of the first moment (i.e.,  $m = 1$ ), which corresponds to the mean population size, reads

$$\frac{d\langle N \rangle}{dt} = \langle B_N \rangle, \quad (9)$$

where as the equation for the second moment (i.e.,  $m = 2$ ) satisfies

$$\frac{d\langle N^2 \rangle}{dt} = \langle (2N + 1)B_N \rangle, \quad (10)$$

where again averages are defined as  $\langle \cdot \rangle = \sum_{N=N_0}^K (\cdot) P_{N_0,N}(t)$ .

The moment equations are closed for linear population birth rates as in the Malthusian (or exponential) growth model. Taking for the sake of simplicity and without loss of generality  $b = 1$ , the population birth rate of the exponential model is given by  $B_N = N$ . Writing out the moment equations, we obtain, for instance,

$$\frac{d\langle N \rangle}{dt} = \langle N \rangle, \quad (11a)$$

$$\frac{d\langle N^2 \rangle}{dt} = \langle N \rangle + 2\langle N^2 \rangle, \quad (11b)$$

$$\frac{d\langle N^3 \rangle}{dt} = \langle N \rangle + 3\langle N^2 \rangle + 3\langle N^3 \rangle. \quad (11c)$$

First, the average population size given by Equation (11a) shows that the deterministic model accurately describes the average

stochastic population size dynamics. Secondly, the equation for the second moment (11b) is closed as it only depends on the second moment itself and the first moment, which can be obtained analytically by solving Equation (11a). Similarly, the third moment (11c) is closed as well; the equation for the  $m^{\text{th}}$  moment only depends on the  $m$  first moments, and so by solving the moment equations sequentially, we obtain a closed equation for any moment of the distribution. This way, important distribution properties such as its variance or skewness can be studied analytically. Note that the master equation (7) is analytically solvable for linear population birth rates, also called Malthusian growth (Taylor et al., 1998), which allows one to obtain distribution properties without solving the ordinary differential equations satisfied by the moments.

Meanwhile, the moment dynamics are unclosed for nonlinear population birth rates, which is often the case for finite populations. In other words, the equation for  $m^{\text{th}}$  moment may involve higher-order moments, leading to an infinite hierarchy of moment equations that is not solvable. Approximation techniques to get around this problem exist, and here, we discuss the accuracy of the most commonly used.

The most basic way to get around this problem is to apply a so-called *mean-field approximation*. The mean-field approximation relies on the crucial assumption that the distribution of population sizes is well-peaked so that  $P_{N_0,N} \approx \delta(N - \langle N \rangle)$ , where  $\delta$  is the Kronecker-delta function (i.e.,  $\delta(N - \langle N \rangle) = 1$  if  $N = \langle N \rangle$  and 0 otherwise). This approximation naturally leads to the approximations  $\langle N \rangle \approx N$ , and  $\langle B_N \rangle \approx B_N$ , which enable us to recover the deterministic limit given by Equation (1). More formally, a Kramers-Moyal expansion in combination with a diffusion approximation of the master equation (7) and a mean-field approximation leads to the same resulting deterministic equation (Gardiner, 2009) (see Appendix B for details). As we argued earlier, this deterministic model has been very popular, for instance, in population genetics theory focusing on the evolutionary effects of natural selection and mutation, as it simplifies calculations and circumvents the need for the master equation framework (Crow & Kimura, 1970; Ewens, 2004; Kingman, 1979; Nagylaki, 2013; Nei, 1987; Rice, 2004). Importantly, we argued in Section 2 that the distributions of population sizes were wide (see also Figure 5). So unsurprisingly, the mean-field approximation fails to satisfactorily describe the population size averaged over stochastic realizations.

Going beyond the mean-field approximation requires us to close the hierarchy of moment equations; these methods are called *moment-closure approximations*. They have been extensively used to provide analytic approximations to nonlinear stochastic population growth models (Krishnarajah et al., 2005; Näsell, 2003b; Singh & Hespanha, 2007). In the following, we focus on the Logistic growth model and proceed to several common moment-closure approximations. Writing the first few moment equations for the Logistic growth model, we obtain

$$\frac{d\langle N \rangle}{dt} = \langle N \rangle - \frac{\langle N^2 \rangle}{K}, \quad (12a)$$

$$\frac{d\langle N^2 \rangle}{dt} = \langle N \rangle + \left[ 2 - \frac{1}{K} \right] \langle N^2 \rangle - \frac{2}{K} \langle N^3 \rangle, \quad (12b)$$

$$\begin{aligned} \frac{d\langle N^3 \rangle}{dt} = \langle N \rangle - \left[ 3 - \frac{1}{K} \right] \langle N^2 \rangle \\ + \left[ 3 - \frac{3}{K} \right] \langle N^3 \rangle - \frac{3}{K} \langle N^4 \rangle. \end{aligned} \quad (12c)$$

Notably, the equation for the first moment  $\langle N \rangle$  depends on the second moment  $\langle N^2 \rangle$ , whereas the equation for the second moment  $\langle N^2 \rangle$  depends on the third moment  $\langle N^3 \rangle$ , and so on. To close this hierarchy of moment equations, two routes are often employed: (i) closure methods can rely on a cumulant truncation procedure, in which the  $k$  first cumulant equations are approximated by setting all cumulants of order higher than  $k$  to 0 (Matis & Kiffe, 1996, 1999), (ii) closure methods can also be based on assumptions on the form of the underlying distribution of population sizes  $P_{N_0,N}(t)$  (Isham, 1991; Krishnarajah et al., 2005; Marion et al., 1998; Näsell, 2003b; Whittle, 1957). Most recently, the latter method has been extensively used; in these latter approximations, one often focuses only on the first two moments.

Here, we test several common moment-closure approximations that express  $\langle N^3 \rangle$  as a function of the first two moments  $\langle N \rangle$  and  $\langle N^2 \rangle$ , allowing us to close the first two moment equations (12a) and (12b). We report all moment-closure approximations tested here in Table 2. As shown in Figure 4, all moment-closure approximations tested here show a substantial disagreement with the simulated stochastic mean population. Whereas the Binomial, separable derivative-matching (SDM), and mean-field approximations overestimate the mean population size, the New-Poisson, Näsell-Poisson, and Normal moment-closure approximations underestimate it. We report absolute relative errors ranging from  $\sim 6\%$  for the Binomial approximation to  $\sim 26\%$  for the normal approximation. Indeed, classical moment-closure approximations fail in problems with very skewed underlying probability distributions, for which accurate knowledge of the higher order moments is needed (Krishnarajah et al., 2005). For instance, the third moment of the distribution (the one being approximated in the methods presented here) is directly related with the lopsidedness of the distribution. Figure 5 shows that our population growth dynamics are plagued by large skewness. In other words, Figure 5 shows that the population size can have many different values at a given time during stochastic growth.

Notably, some moment-closure approximations perform better than the mean-field approximation, although the latter is the most widely used. While these moment-closure approximations generally reduce the error in predicting the dynamics of population growth, we note that they cannot generically be applied to the other growth models considered in this paper. Indeed, the right-hand side of Equation (9) depends on the terms  $\langle (1 - N/K)^N \rangle$ ,  $\langle \log(N)N \rangle$  and  $\langle N^{\beta+1} \rangle$  for the Blumberg, Gompertz, and Richards models, respectively. These are not easily expressed in terms of combinations of moments of the distribution. Since these

TABLE 2 Moment-closure approximations

Moment-closure approximation	Third moment $\langle N^3 \rangle$
Binomial (Nåsell, 2003a)	$2(\langle N^2 \rangle - \langle N \rangle^2)^2 / \langle N \rangle - \langle N^2 \rangle + \langle N \rangle^2 + 3\langle N^2 \rangle \langle N \rangle - 2\langle N \rangle^3$
Lognormal (Keeling, 2000a, 2000b)	$\langle N^2 \rangle^3 / \langle N \rangle^3$
Nåsell-Poisson (Nåsell, 2003a)	$\langle N \rangle + 3\langle N^2 \rangle \langle N \rangle - 2\langle N \rangle^3$
New-Poisson (Nåsell, 2003a)	$\langle N^2 \rangle - \langle N \rangle^2 + 3\langle N^2 \rangle \langle N \rangle - 2\langle N \rangle^3$
Normal (Matis & Kiffe, 1996; Nåsell, 2003b; Whittle, 1957)	$3\langle N^2 \rangle \langle N \rangle - 2\langle N \rangle^3$
Separable Derivative-Matching (Singh & Hespanha, 2007)	$\langle N^2 \rangle^3 / \langle N \rangle^3$

Note: Common moment-closure approximations where the third moment  $\langle N^3 \rangle$  is expressed as a function of the first moment  $\langle N \rangle$  and the second moment  $\langle N^2 \rangle$ .

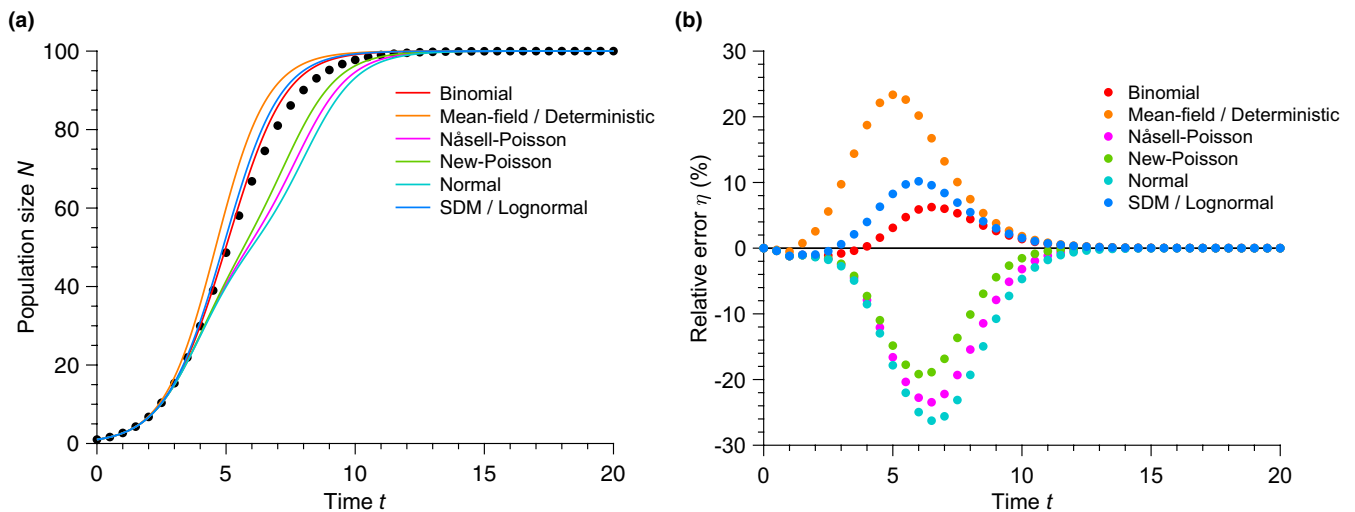


FIGURE 4 Moment-closure approximations do not satisfactorily reduce the error. (a) Population size  $N$  versus time  $t$  for moment-closure approximations with the Logistic model. The data points show simulated data averaged over  $10^5$  stochastic realizations. The solid lines correspond to the moment-closure approximations (see Table 2). (b) Relative error  $\eta$  versus time  $t$  for different moment-closure approximations.  $\eta$  is calculated using data from Panel (a). Parameter values: carrying capacity  $K = 100$ , initial population size  $N_0 = 1$  and birth rate  $b = 1$ .

nonlinear population growth models are widely used, we must devise a generically applicable method, which we do in what follows. However, we recognize that moment-closure approximations allow growth dynamics to be described from one or two ordinary differential equations, which may be easier than a fully probabilistic description.

## 4 | TOWARD AN EXACT SOLUTION TO STOCHASTIC POPULATION GROWTH

### 4.1 | First approach: Transition rate matrix

A first approach to try and solve the master equation directly is to recast it in the language of Markov chains. Namely, stochastic population growth can be interpreted as a continuous-time,  $K$ -state Markov jump process with transition rate matrix  $\mathbf{R} = (R_{ij})_{1 \leq i \leq K, 1 \leq j \leq K}$  where the matrix element  $R_{ij}$  with  $i \neq j$  is understood to be the rate at which the population switches from value  $N = j$  to  $N = i$  (note that

other conventions exist). The diagonal elements of the transition rate matrix are generically fixed by enforcing conservation of total probability, which implies that the columns of  $\mathbf{R}$  sum up to zero, that is,  $\sum_{i=1}^K R_{ij} = 0$ , so that  $R_{ij} = -\sum_{i=1, i \neq j}^K R_{ij}$ .

As the population sizes potentially vary from 1 to  $K$ , we first define  $\mathbf{P}(t)$  as the column vector of dimension  $K$

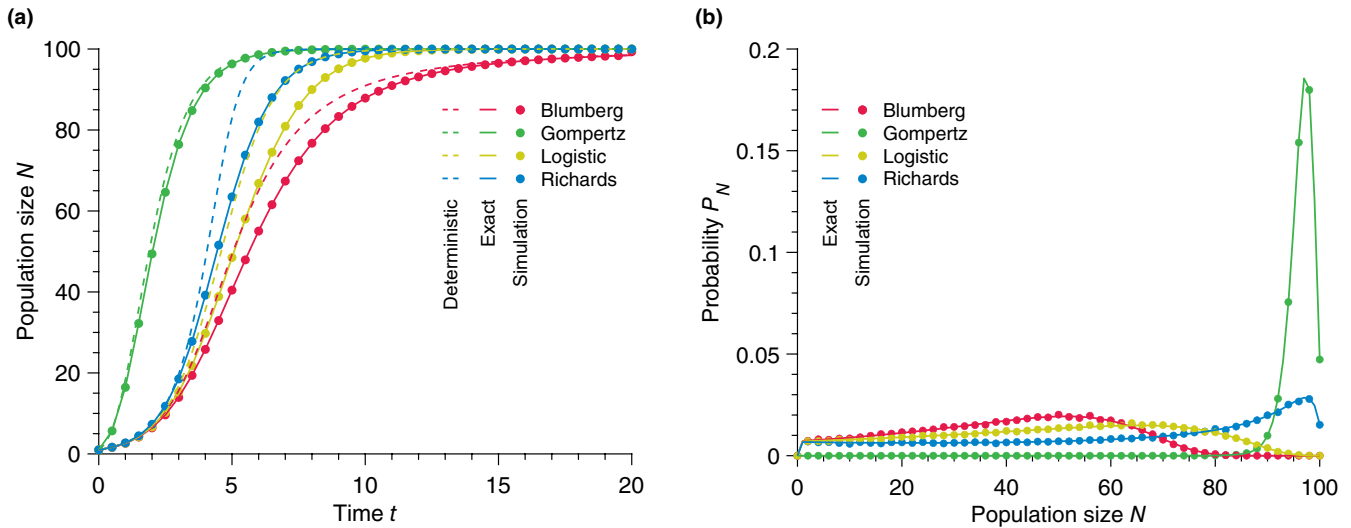
$$\mathbf{P}(t) = \begin{pmatrix} P_{N_0,1}(t) \\ P_{N_0,2}(t) \\ \vdots \\ P_{N_0,K}(t) \end{pmatrix}, \quad (13)$$

with the  $i$ th-component corresponding to the time-dependent probability of having a population of size  $i$ . We then rewrite Equation (7) in matrix form as

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{R} \cdot \mathbf{P}(t), \quad (14)$$

where the  $K \times K$  transition rate matrix  $\mathbf{R}$  is defined as





**FIGURE 5** Exact solution to the nonlinear population growth problem. (a) Population size  $N$  versus time  $t$  for different population growth models. The data points show simulated data averaged over  $10^5$  stochastic realizations. The solid lines correspond to our solution, whereas the dashed lines represent the deterministic equation. (b) Probability  $P_N$  of having  $N$  individuals at  $t = 5$ . Parameter values: carrying capacity  $K = 100$ , initial population size  $N_0 = 1$ , birth rate  $b = 1$ , exponents  $\alpha = 1$ ,  $\beta = 2$  and  $\gamma = 1.5$ .

$$R_{ij} = \begin{cases} -B_j & \text{if } i=j, \\ B_j & \text{if } i=j+1, \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

leading to

$$R = \begin{pmatrix} -B_1 & 0 & 0 & \dots & 0 \\ B_1 & -B_2 & 0 & \dots & 0 \\ 0 & B_2 & -B_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & -B_K \end{pmatrix} \quad (16)$$

In the case where the transition rate matrix does not have repeated entries, the solution to Equation (14) can be written as a weighted superposition of the  $K$  eigenvectors  $\mathbf{v}_k$  of the transition rate matrix multiplied by an exponential function whose rate is given by the associated eigenvalue  $\mu_k$ . Namely, we write

$$\mathbf{P}(t) = \sum_{k=1}^K \mathbf{v}_k c_k e^{\mu_k t}, \quad (17)$$

a vector whose components give us the time-dependent probability that the population size is  $N$  given that it was  $N_0$  at  $t = 0$ ,

$$P_{N_0,N}(t) = \sum_{k=1}^K v_{k,N} c_k e^{\mu_k t}. \quad (18)$$

Note that the coefficients  $c_k$  are obtained by the imposition of the initial conditions and here must satisfy  $P_{N_0,N}(0) = \sum_{k=1}^K v_{k,N} c_k = \delta_{N,N_0}$ , where  $\delta$  is the Kronecker delta (i.e.,  $\delta_{N,N_0} = 1$  if  $N = N_0$  and 0 otherwise). Furthermore,  $c_k = 0$  for  $1 \leq N < N_0$  since a population size

lower than  $N_0$  cannot be reached. Indeed, death events are not considered here, so the population size can only increase over time.

As the transition rate matrix  $\mathbf{R}$  is lower triangular, the eigenvalues are equal to the diagonal entries of the matrix, and we obtain  $\mu_k = -B_k$ ,  $1 \leq k \leq K$ . Finally, we compute the eigenvectors as

$$v_{k,N} = \begin{cases} 0 & \text{if } 1 \leq N \leq k-1, \\ \prod_{q=k}^{N-1} \frac{B_q}{B_{q+1} - B_k} & \text{otherwise.} \end{cases} \quad (19)$$

Notably, the eigenvectors  $\mathbf{v}_k$  are ill-defined when the birth rates degenerate (i.e., when  $B_k = B_{k'}$  for  $k \neq k'$ ). Indeed, Equation (19) shows that some of the components of the eigenvectors diverge when birth rates degenerate. The degeneracy typically happens when the population birth rate curve displays particular symmetries (see Figure 1). For instance, the Logistic growth model presents a mirror symmetry with respect to  $K/2$ ; for all  $1 \leq k \leq K$ , we obtain  $B_k = B_{K-k}$ . Diagonalization of the transition rate matrix would thus not apply to the Logistic growth model.

In the case where every birth rate is distinct (i.e.,  $B_k \neq B_{k'}$  if  $k \neq k'$ ), we compute the constants  $c_k$ , and we finally get

$$P_{N_0,N}(t) = \frac{1}{B_N} \sum_{k=N_0}^N B_k e^{-B_k t} \begin{pmatrix} \prod_{q=N_0}^N \frac{B_q}{B_q - B_k} \\ q=N_0 \\ q \neq k \end{pmatrix}. \quad (20)$$

Note that the condition on the transition rate matrix not having repeated entries translates to imposing on the growth model not to lead to degenerate birth rates. The solution to the stochastic dynamics of population growth models with degenerate birth rates

cannot be directly obtained via Equation (17); indeed the degeneracy in the birth rates leads to degenerate eigenvalues  $\{\mu_i\}_{i \in [1, M]}$ , with  $M < K$ . This means that the solution would then be written as a linear combination of exponential terms with parameter  $\mu_i$ , whose coefficients are polynomial factors. In general, identifying these polynomial factors can be a very tedious task. To circumvent this issue, and thus the lack of universality of Equation (17), we develop in the next section a new resolution method that applies to any population growth model.

## 4.2 | Exact solution: distribution of waiting times between birth events

As we just saw, the method based on the transition rate matrix is unsatisfactory because it does not yield a closed-form solution for Equation (14) if birth rates are degenerate. Here, we suggest an approach based on waiting times between birth events. The underlying idea is that in the absence of deaths, population growth can be interpreted as a succession of events (births) happening in a well-defined order separated by waiting times, which are random variables. To reach size  $K$  from its initial size  $N_0$ , the population has to grow one individual at a time and go from  $N_0$  to  $N_0 + 1$ , then from  $N_0 + 1$  to  $N_0 + 2$ , etc. in this precise order. Therefore, all the information needed to derive the time-dependent probability  $P_{N_0, N}(t)$  should be contained in the distribution of the time between two birth events. In other words, in the approach based on the master equation and the transition rate matrix, the reasoning is based on population sizes. In contrast, in this approach, our reasoning is based on waiting times between successive events.

We denote by  $\tau_N$  the time elapsed between two births where the population size increases from  $N$  to  $N + 1$  individuals. Owing to the Markovian nature of the process,  $\tau_N$  is a stochastic variable exponentially distributed with mean  $B_N^{-1}$ . Then, the probability of having a given number  $N$  of individuals at time  $t$  must be equal to the probability that  $N - N_0$  births occurred by  $t$  and not  $N - N_0 + 1$  yet. Quantitatively speaking,

$$P_{N_0, N}(t) = \text{Prob} \left( \sum_{k=N_0}^{N-1} \tau_k < t < \sum_{k=N_0}^N \tau_k \right). \quad (24)$$

The sum of  $n$  exponentially distributed random variables with rates  $\{\lambda_i\}_{1 \leq i \leq n}$  follows a hypoexponential distribution (Amari & Misra, 1997; Ross, 2007), which we denote  $H(t; \{\lambda_i\}_{1 \leq i \leq n})$ . Hypoexponential distributions were previously studied in the context of population genetics (Strimmer & Pybus, 2001) but also cell and systems biology (Chao et al., 2019; Gavagnin et al., 2019; Golubev, 2016; Yates et al., 2017).

Using the expression for the probability density function for the hypoexponential distribution (see Table 3 for the three cases to consider), we write the exact solution to Equation (7) in the form

$$P_{N_0, N}(t) = \frac{1}{B_N} H(t; \{B_{N_0}, B_{N_0+1}, \dots, B_N\}). \quad (25)$$

TABLE 3 Definition of the probability density function for the hypoexponential distribution.

Case 1 if all  $n$  rates are identical, then  $\lambda_i = \lambda$ , for  $1 \leq i \leq n$  and the PDF reads

$$H(t; \{\lambda_i\}_{1 \leq i \leq n}) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \equiv \text{Erlang}(t; \lambda, n) \quad (21)$$

Case 2 if all rates  $\{\lambda_i\}_{1 \leq i \leq n}$  are distinct, then the PDF reads

$$H(t; \{\lambda_i\}_{1 \leq i \leq n}) = \sum_{i=1}^n \lambda_i e^{-\lambda_i t} \left[ \prod_{\substack{j=1 \\ j \neq i}}^n \frac{\lambda_j}{\lambda_j - \lambda_i} \right]. \quad (22)$$

Case 3 if the rates present some amount of degeneracy, we denote the multiplicities of the unique rates  $\{\lambda_i\}_{1 \leq i \leq m}$  as  $\{n_i\}_{1 \leq i \leq m}$  with the constraint that  $\sum_{i=1}^m n_i = n$ , where  $m$  is the number of unique rates. In this last case, the PDF reads

$$H(t; \{\lambda_i\}_{1 \leq i \leq n}) = \left( \prod_{i=1}^m \lambda_i^{n_i} \right) \sum_{i=1}^m \sum_{k=0}^{n_i-1} \frac{C(i, k)}{k! (n_i - k - 1)!} t^{n-k-1} e^{-\lambda_i t}, \quad (23)$$

with  $C(i, k) = \left[ \frac{d^k}{ds^k} \prod_{j=1}^m (s + \lambda_j)^{-n_j} \right]_{s=-\lambda_i}$ .

Notably, in the case where all population growth rates are distinct, the hypoexponential distribution takes the form (22), and we recover exactly the solution introduced in the previous section (see Equation (20)). In Figure 5, we compare our exact solution (25) to the mean population size measured in simulations of the stochastic process for our four nonlinear growth models and show perfect agreement in all cases. We also confirm in Figure 5 that the full probability distributions measured from simulations agree with our exact solution.

## 5 | APPLICATIONS

Finally, we cover two examples of applications in which we show that using a deterministic model instead of an exact solution to stochastic population growth leads to quantitatively very different results and may misinterpret important experimental results.

### 5.1 | Population growth dynamics within a community

First, we extend our results to the study of a community composed of multiple strains evolving in the same environment. Community dynamics are at the heart of theoretical ecology and have been recently applied to the growing field of microbiome studies, where predicting the relative abundance of each microbial strain in the gut microbiota represents an opportunity for medical diagnosis and treatment (Gebreyal et al., 2022). For simplicity and without loss of generality, we focus on the case of two competing strains. Consider, for instance, the

population dynamics of a wild-type (W) strain and a mutant strain (M) competing in a batch culture environment; we denote their intrinsic birth rates  $b_W$  and  $b_M$ , respectively. As before, we define  $N$  as the size of the community, whereas  $n$  (resp.  $N - n$ ) denotes the number of M (resp. W) individuals. We assume that the size of the community is limited by a single carrying capacity  $K$ . Note that our approach is easily generalizable to cases with multiple strains or with different carrying capacities and with explicit interaction parameters.

Furthermore, we introduce the relative fitness of the two strains, defined as  $r = b_M / b_W$ ; this ratio indicates which strain is favored by natural selection. Specifically, if  $r > 1$  (resp.  $r < 1$ ), then strain M is beneficial (resp. deleterious), with  $r = 1$  corresponding to the neutral case. Each time an individual reproduces, the probability that this individual is of strain M is then given by

$$\Gamma_{N,n} = \frac{rn}{(r-1)n + N}. \quad (26)$$

Here, we start with an initial community size  $N_0$ , composed of  $n_0$  individuals from strain M and  $N_0 - n_0$  individuals from strain W. The probability  $P_{N_0,N}(t)$  that the community has a total size  $N$  at time  $t$ , knowing that the initial size of the community was  $N_0$ , is given directly by Equation (25) with population reproduction rates

$$B_N = \sum_{n=n_0}^{N-N_0+n_0} P(N, n | N_0, n_0) (B_{N,n}^{(M)} + B_{N,n}^{(W)}), \quad (27)$$

where  $B_{N,n}^{(M)}$  and  $B_{N,n}^{(W)}$  are the rates at which each population increases. For the Logistic model, these rates are for instance given by

$$B_{N,n}^{(M)} = b_M(1 - N/K)n, \quad (28a)$$

$$B_{N,n}^{(W)} = b_W(1 - N/K)(N - n). \quad (28b)$$

Furthermore, the probability  $P(N, n | N_0, n_0)$  of finding  $n$  individuals of type M when the total number of individuals is  $N$  satisfies

$$P(N+1, n | N_0, n_0) = (1 - \Gamma_{N,n})P(N, n | N_0, n_0) + \Gamma_{N,n-1}P(N, n-1 | N_0, n_0), \quad (29)$$

subject to the initial conditions  $P(N_0, n | N_0, n_0) = \delta_{n,n_0}$ . Note that Equation (29) has been extensively studied in (Houchmandzadeh, 2018). By definition, the probability to observe  $n$  individuals of type M at time  $t$ , knowing that we had initially  $n_0$  such individuals, is

$$P_{n_0,n}(t) = \sum_{N=N_0}^K P_{N_0,N}(t)P(N, n | N_0, n_0), \quad (30)$$

which we exactly compute using Equations (25), and (29). Then, using Equations (25), (27), and (29), we compute the average stochastic community and population sizes

$$\langle N \rangle(t) = \sum_{N=N_0}^K NP_{N_0,N}(t), \quad (31a)$$

$$\langle n \rangle(t) = \sum_{n=n_0}^K nP_{n_0,n}(t). \quad (31b)$$

On the contrary, a deterministic description of the community dynamics leads to the system of differential equations

$$\frac{dn}{dt} = b_M n \left(1 - \frac{N}{K}\right), \quad (32a)$$

$$\frac{dN}{dt} = [b_M n + b_W(N - n)] \left(1 - \frac{N}{K}\right). \quad (32b)$$

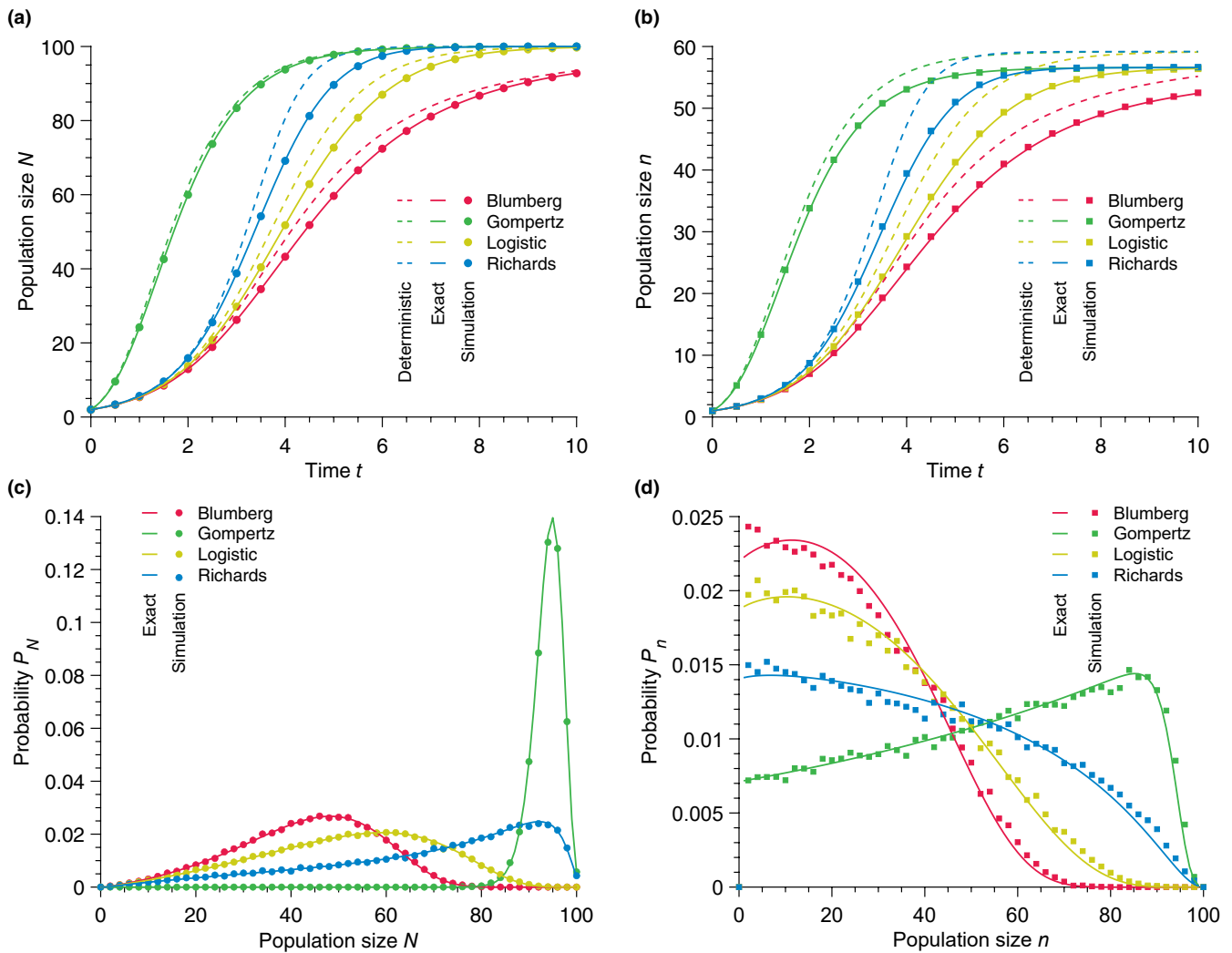
We once again compare the results of stochastic simulations quantitatively to the predictions of Equation (31) on the one hand and Equation (32) on the other hand. We show in Figure 6a,b that the deterministic model grossly overestimates the size of the community, whereas our stochastic solution perfectly matches the simulated data. Strikingly, the deterministic model is shown to overestimate the equilibrium population size of strain M. Our stochastic approach provides an exact prediction of the average community and population sizes and, more importantly, yields the full time-dependent probability distributions of community and population sizes, which is not possible with a deterministic approach (see Figure 6c,d).

## 5.2 | Fixation probability in a serial passage experiment

Finally, we show that our exact stochastic solution yields the fixation probability of a strain in a serial passage experiment. For this, let us assume that we start an experiment with the same number of individuals M and W. As usual, the initial size of the community formed by strains M and W is given by  $N_0$ ; we introduce the dilution rate  $D$  defined as the ratio of the initial size of the community to the carrying capacity,  $D = N_0 / K$ .

In a serial passage experiment, the community grows for a time  $\tau$  before one applies a bottleneck by taking a random sample of  $N_0$  individuals; mathematically, this corresponds to selecting  $N_0$  individuals from the community following a binomial law. One then proceeds with a new growth phase of length  $\tau$  before applying a new bottleneck. This process is repeated until only a single strain is left in the community. In these experiments, a quantity of interest is the probability  $p_{\text{fix}}$  that the strain M fixes and that the strain W goes extinct. In particular, optimizing the fixation probability as a function of the dilution ratio  $D$  and the waiting time  $\tau$  has attracted much attention, especially in the context of directed evolution (LeClair & Wahl, 2017; Wahl et al., 2002; Wahl & Gerrish, 2001). To our knowledge, existing studies answered this question using only deterministic approaches for modeling population dynamics.

To calculate  $p_{\text{fix}}$ , we model the system as a Markov chain on the number of individuals M after each bottleneck event. At the moment of applying a bottleneck, the population contains  $N(\tau)$  individuals, including  $n(\tau)$  individuals of strain M. If picking a single individual



**FIGURE 6** Exact time-dependent population growth and steady-state population sizes in community dynamics. (a) Total population size  $N$  versus time  $t$  for different population growth models. (b) Population size  $n$  of strain  $M$  versus time  $t$  for different population growth models. (c) Probability  $P_N$  of finding  $N$  individuals at time  $t = 4$ . (d) Probability  $P_n$  of finding  $n$  individuals at time  $t = 4$ . In every panel, the solid lines represent our stochastic solution. In (a) and (b), the dashed lines show the deterministic predictions whereas each point results from simulated data averaged over  $10^5$  stochastic realizations. Parameter values: carrying capacity  $K = 100$ , mutant birth rate  $b_M = 1.1$ , wild-type birth rate  $b_W = 1$ , initial community size  $N_0 = 2$ , and initial wild-type population size  $n_0 = 1$ .

randomly from the community, the probability that this individual is of type  $M$  is thus given by  $n(\tau)/N(\tau)$ . The probability  $\Pi_{k \rightarrow l}$  that the number of individuals  $M$  goes from  $k$  to  $l$  when a bottleneck is applied follows the binomial distribution

$$\Pi_{k \rightarrow l} = \binom{N_0}{l} \left( \frac{n(\tau)}{N(\tau)} \right)^l \left( 1 - \frac{n(\tau)}{N(\tau)} \right)^{N_0 - l}, \quad (33)$$

where  $n(\tau)$  depends on its starting value  $k$ . Note that here we follow the convention used in (Wahl et al., 2002; Wahl & Gerrish, 2001) and model bottlenecks as draws from a population with replacement leading to binomial distributions for  $\Pi_{k \rightarrow l}$ ; one could also model bottlenecks as draws without replacement, naturally leading to hypergeometric distributions. Under a deterministic approach,  $n(\tau)$  and  $N(\tau)$  are obtained by solving Equation (32) with the initial conditions  $n_0 = k$  and  $N_0 = DK$ . However, in a stochastic approach, we write the  $\Pi_{k \rightarrow l}$  as a

sum over all possible pairs  $(N, n)$  at time  $\tau$  weighted by their respective probabilities,

$$\Pi_{k \rightarrow l} = \sum_{N=N_0}^K \sum_{n=k}^N P_{N_0, N(\tau)} P(N, n | N_0, k) \times \binom{N_0}{l} \left( \frac{n}{N} \right)^l \left( 1 - \frac{n}{N} \right)^{N_0 - l}, \quad (34)$$

where  $P(N, n | N_0, n_0)$  is governed by Equation (29). Finally, we note that conservation of probabilities imposes that  $\Pi_{l \rightarrow j} = 1 - \sum_{k \neq l} \Pi_{k \rightarrow l}$ .

We define  $\mathbf{P} = (P_0, P_1, \dots, P_{N_0})$  as the column vector of probabilities  $P_i$  to have  $i$  individuals of strain  $M$  in the random sample of  $N_0$  individuals from the community following a bottleneck event. As such, the serial passage experiment defines a discrete-time Markov process, in which time is measured in units of  $\tau$  with  $t$  thus corresponding to the number of bottlenecks. Here we follow reference

Marrec and Bitbol (2018) and take the limit of continuous time to write the master equation governing  $\mathbf{P}$  as

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{R} \cdot \mathbf{P}(t). \quad (35)$$

Here, the elements of the transition rate matrix  $\mathbf{R}$  are given by

$$R_{ij} = \Pi_{j \rightarrow i}, \quad (36a)$$

$$R_{jj} = - \sum_{i \neq j} \Pi_{j \rightarrow i}, \quad (36b)$$

in which (36b) ensures conservation of probability.

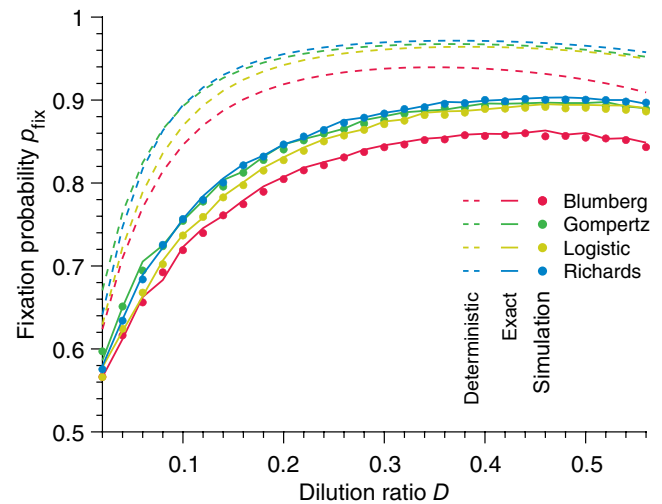
The Markov process thus defined possesses two absorbing states, namely  $n = 0$  and  $n = N_0$ , which correspond to the extinction and fixation of strain M, respectively. By definition, once the system reaches one of these states, it remains there indefinitely. Mathematically, this implies that the first and last columns of the transition rate matrix are filled with zeros as these columns contain the transition rates out of the  $n = 0$  and  $n = N_0$  states, respectively. As the matrix  $\mathbf{R}$  is not invertible, we introduce the reduced transition rate matrix  $\tilde{\mathbf{R}}$  in which the rows and the columns corresponding to the absorbing states are removed. According to this definition,  $\tilde{\mathbf{R}}$  is invertible and we can write the fixation probability as

$$p_{\text{fix}} = - \sum_{i=1}^{N_0-1} \Pi_{i \rightarrow N_0} \left( \tilde{\mathbf{R}}^{-1} \right)_{i, n_0}, \quad (37)$$

where  $n_0$  is the number of individuals M at the beginning of the experiment. As shown in Figure 7, the deterministic approach grossly overestimates the fixation probability for all dilution rates  $D$ . Here, we confirm this result for all nonlinear growth models studied. In contrast, we show that our stochastic solution exactly matches the results of stochastic simulations. This result has substantial consequences as we show in particular that the dilution ratio predicted by a deterministic approach to optimize the fixation probability is far from being the *actual* optimal dilution ratio, although it is commonly used in the literature, as was already argued.

## 6 | DISCUSSION

In conclusion, we showed that a deterministic approach to population growth leads to biased predictions of the average behavior of this inherently stochastic process. More precisely, deterministic models overestimate the population size averaged over large numbers of stochastic realizations. This overestimation increases with decreasing initial population size. Qualitatively, the bias of the deterministic approach is due to the variability of the waiting times between reproduction events, which is particularly important at small population sizes. Quantitatively, the bias of the deterministic approach is due to unclosed-moment dynamics. Importantly, we showed that moment-closure approximations are insufficient to substantially reduce the relative difference between analytical



**FIGURE 7** Fixation probability in a serial passage experiment. Fixation probability  $p_{\text{fix}}$  versus dilution ratio  $D$  for different population growth models. The solid lines represent our stochastic solution, the dashed lines the deterministic predictions, and each point results from simulated data averaged over  $10^5$  stochastic realizations. Parameter values: carrying capacity  $K = 100$ , mutant birth rate  $b_M = 1.1$ , wild-type birth rate  $b_W = 1$ , time between each bottleneck  $\tau = 3$ , initial mutant population size  $n_0 = 1, 2, \dots, 28$ , initial community size  $N_0 = 2, 4, \dots, 56$ .

predictions and average population sizes and do not apply to all population growth models.

It should be noted that methods other than moment-closure approximations have been used to describe the average stochastic population dynamics by a simple ordinary differential equation. For example, the results of Kurtz (1972) on the convergence of average stochastic trajectories to a deterministic model in the case of large carrying capacity and large initial numbers of individuals have been extended more recently in this context of the so-called fluid limit. This limit theorem shows that the dynamics of a population with a small initial size compared to the carrying capacity can be accurately described by a deterministic equation with a random initial size (Baker et al., 2018; Barbour et al., 2016). However, it is important to note that while Barbour et al. (2016) explicitly show that the whole distribution of the population size is converging to that of the fluid limit with random initial conditions, this often involves a non-explicit function. In practice, as obtaining the full time-dependent population size distribution is essential for the inference of growth parameters, the fluid limit has limited applicability. Nonetheless, for some growth models, it can provide explicit results for the average population size. Furthermore, by construction, the fluid limit provides an accurate description of the long-time population growth dynamics but fails at describing the short-time transient growth accurately. This is particularly limiting, for instance in the context of epidemics where the goal is to infer important parameters such as the reproduction rate as early as possible. On a more technical note, the random initial size distribution can be tedious to derive for some population growth models. In this context, we would be remiss not to mention a method based on a theorem by Cohn and Jagers (1994), which also allows one to describe the average

stochastic population dynamics, but in this case, at any time (Uecker & Hermisson, 2011). However, as in the fluid limit, this method does not provide a way to obtain the population size distribution.

In contrast, we proposed two methods to derive exact solutions to the stochastic population growth dynamics: either by solving the master equation directly, which requires the diagonalization of the transition rate matrix or by tracking reproduction times instead of population sizes. The first method was shown to be valid only in situations where the reproduction rates are distinct, whereas the second is generic. Our solution has revealed that the temporal distribution of population sizes is proportional to a hypoexponential distribution.

Note that as argued earlier, stochasticity in population growth can generically stem from intrinsic and extrinsic noises. Here, we focused on the effect of intrinsic noise and more specifically on the case where the only source of stochasticity resides in the waiting times between births, in an otherwise homogeneous population. A natural extension of our work would be to consider heterogeneous populations in which individuals display growth variability, that is, different intrinsic birth rates  $b$ . In this context, it would be interesting to study the effect of this intrapopulation variability on the growth dynamics both computationally and analytically by extending the present work.

It is also interesting to note that the role of extrinsic noise on population growth dynamics has also attracted attention in the past. A case in point is the study of the effect of stochastic temporal environmental fluctuations on growth (and other vital) rates and eventually population growth dynamics Steiner (2020). While early works focused on homogeneous populations (Lande et al., 2003), the works of Tuljapurkar and Orzack in the early 1980s extended these results to structured populations, with a particular focus on age-structured matrix population models (Tuljapurkar & Orzack, 1980). In a series of seminal papers, they provided analytical insights into population structure and long-run growth rates under stochastic environmental fluctuations showing that the distribution of population size is lognormal, and thus skewed (Tuljapurkar, 1982a). In particular, Tuljapurkar's small noise approximation, a perturbation formula used to calculate the population growth rate in stochastic environments, showed that population growth in constant and fluctuating environments differ, even when the mean growth rates are designed to be equal (Tuljapurkar, 1982b). Formally, it quantifies the decrease in the population growth rate in structured populations in response to the increased variance in vital rates due to randomly varying environments. In an interesting parallel with our own study, this result disproves the then commonly accepted assumption that averaging growth rates across realizations of the fluctuating environment would lead to an accurate estimate of the effective growth rates. Following on from Tuljapurkar's original work on sensitivity analyses to variance in vital rates, recent studies performed an age-specific sensitivity analysis of stable, stochastic, and transient growth in structured populations (Giaino & Traulsen, 2022).

Finally, we showed that our solution provides a more accurate description (than a deterministic approach) of the time-dependent and steady-state population sizes in a community composed of competing strains and the fixation probability of a mutation in a serial passage

experiment. Thus, our theory offers an opportunity to quantify the dynamics of microbial communities from colonization to coexistence and thus contributes to the growing field of microbial eco-evolutionary dynamics. For example, the gut of *C. elegans* worms colonized by two neutrally-competing strains was shown to transition from a single-strain composition at a low colonization rate to coexistence at a high colonization rate (Vega & Gore, 2017). In previous work, a deterministic approach was used to predict this transition (Vega & Gore, 2017). Our exact solution enables the expansion of this work by quantifying the abundance distribution of either strain within each worm's gut.

Our work opens new perspectives on population dynamics, ecology, and evolution. Importantly, our theory yields the full time-dependent probability distribution of population sizes (see Equation (25)). Based on this result, an interesting future research direction would be to improve inference methods for growth parameters, as current methods suffer from substantial limitations (Ghenu et al., 2022). We propose that the present work constitutes the first step toward an exact inference method since it allows for the exact calculation of the likelihood function (Cranmer et al., 2020).

#### AUTHOR CONTRIBUTIONS

**Loïc Marrec:** Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); resources (lead); software (lead); validation (lead); visualization (lead); writing – original draft (equal); writing – review and editing (equal). **Claudia Bank:** Funding acquisition (lead); supervision (lead); validation (equal); writing – review and editing (equal). **Thibault Bertrand:** Formal analysis (supporting); investigation (supporting); methodology (supporting); resources (supporting); software (supporting); validation (supporting); visualization (supporting); writing – original draft (equal); writing – review and editing (equal).

#### ACKNOWLEDGMENTS

The authors thank the THEE Group and Stephan Peischl at UniBe for discussion and feedback on the manuscript.

#### FUNDING INFORMATION

CB is grateful for funding from ERC Starting Grant 804569829 (FIT2GO) and SNSF Project Grant 315230\_204838/1 (MiCo4Sys).

#### CONFLICT OF INTEREST STATEMENT

We declare we have no competing interests.

#### DATA AVAILABILITY STATEMENT

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article and Appendix. Annotated Matlab implementations of numerical simulations are available on Zenodo (doi: <https://doi.org/10.5281/zenodo.8083898>).

#### ORCID

Loïc Marrec  <https://orcid.org/0000-0003-0941-6603>

Thibault Bertrand  <https://orcid.org/0000-0001-9812-7904>

## REFERENCES

- Allen, L. J. S., & Burgin, A. M. (2000). Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences*, 163(1), 1–33.
- Amari, S. V., & Misra, R. B. (1997). Closed-form expressions for distribution of sum of exponential random variables. *IEEE Transactions on Reliability*, 46(4), 519–522.
- Anderson, R. M. (2014). *The population dynamics of infectious diseases: Theory and applications*. Springer.
- Aviv-Sharon, E., & Aharoni, A. (2020). Generalized logistic growth modeling of the covid-19 pandemic in asia. *Infectious Disease Modelling*, 5, 502–509.
- Baker, J., Chigansky, P., Hamza, K., & Klebaner, F. C. (2018). Persistence of small noise and random initial conditions. *Advances in Applied Probability*, 50(A), 67–81.
- Baker, R., & Simpson, M. (2010). Correcting mean-field approximations for birth-death-movement processes. *Physical Review E*, 82, 041905.
- Barbour, A. D., Chigansky, P., & Klebaner, F. C. (2016). On the emergence of random initial conditions in fluid limits. *Journal of Applied Probability*, 53(4), 1193–1205.
- Bartlett, M. (1960). *Stochastic population models in ecology and epidemiology. Methuen's monographs on applied probability and statistics*. Methuen.
- Black, A. J., & McKane, A. J. (2012). Stochastic formulation of ecological models and their applications. *Trends in Ecology & Evolution*, 27(6), 337–345.
- Blumberg, A. A. (1968). Logistic growth rate functions. *Journal of Theoretical Biology*, 21(1), 42–44.
- Bolker, B., & Pacala, S. W. (1997). Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, 52(3), 179–197.
- Brauer, F., & Castillo-Chavez, C. (2013). *Mathematical models in population biology and epidemiology. Texts in applied mathematics*. Springer.
- Carlson, T. (1913). Über geschwindigkeit und grösse der hefevermehrung in würze. *Biochemische Zeitschrift*, 57, 313–334.
- Chao, H. X., Fakhreddin, R. I., Shimerov, H. K., Kedziora, K. M., Kumar, R. J., Perez, J., Limas, J. C., Grant, G. D., Cook, J. G., Gupta, G. P., & Purvis, J. E. (2019). Evidence that the human cell cycle is a series of uncoupled, memoryless phases. *Molecular Systems Biology*, 15(3), e8604.
- Cohn, H., & Jagers, P. (1994). General branching processes in varying environment. *The Annals of Applied Probability*, 4(1), 184–193.
- Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences of the United States of America*, 117(48), 30055–30062.
- Crow, J. F., & Kimura, M. (1970). *An Introduction to population genetics theory*. Burgess Publishing Company.
- Czock, D., & Keller, F. (2007). Mechanism-based pharmacokinetic-pharmacodynamic modeling of antimicrobial drug effects. *Journal of Pharmacokinetics and Pharmacodynamics*, 34(6), 727–751.
- Ewens, W. J. (2004). *Mathematical population genetics 1: Theoretical introduction. Interdisciplinary applied mathematics*. Springer.
- Gardiner, C. (2009). *Stochastic methods: A handbook for the natural and social sciences. Springer series in synergetics*. Springer.
- Gavagnin, E., Ford, M. J., Mort, R. L., Rogers, T., & Yates, C. A. (2019). The invasion speed of cell migration models with realistic cell cycle time distributions. *Journal of Theoretical Biology*, 481, 91–99.
- Gebrayel, P., Nicco, C., Al Khodor, S., Bilinski, J., Caselli, E., Comelli, E. M., Egert, M., Giaroni, C., Karpinski, T. M., Loniewski, I., Mulak, A., Reygnier, J., Samczuk, P., Serino, M., Sikora, M., Terranegra, A., Ufnal, M., Villegier, R., Pichon, C., ... Edeas, M. (2022). Microbiota medicine: Towards clinical revolution. *Journal of Translational Medicine*, 20(1), 111.
- Ghenu, A.-H., Marrec, L., & Bank, C. (2022). Challenges and pitfalls of inferring microbial growth rates from lab cultures. *bioRxiv*. <https://doi.org/10.1101/2022.06.24.497412>
- Gaiimo, S., & Traulsen, A. (2022). Age-specific sensitivity analysis of stable, stochastic and transient growth for stage-classified populations. *Ecology and Evolution*, 12(12), e9561.
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4), 403–434.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25), 2340–2361.
- Golubev, A. (2016). Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression. *Journal of Theoretical Biology*, 393, 203–217.
- Gompertz, B. (1825). XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. F.R.S. *Philosophical Transactions of the Royal Society of London*, 115, 513–583.
- Gustafsson, L., & Sternad, M. (2013). When can a deterministic model of a population system reveal what will happen on average? *Mathematical Biosciences*, 243(1), 28–45.
- Haefner, J. W. (2012). *Modeling biological systems: Principles and applications*. Springer US.
- Hannon, B., Ruth, M., & Levin, S. A. (1997). *Modeling dynamics biological systems. Modeling dynamic systems*. Springer.
- Houchmandzadeh, B. (2018). Giant fluctuations in logistic growth of two species competing for limited resources. *Physical Review E*, 98, 042118.
- Isham, V. (1991). Assessing the variability of stochastic epidemics. *Mathematical Biosciences*, 107(2), 209–224.
- Kahm, M., Hasenbrink, G., Lichtenberg-Fraté, H., Ludwig, J., & Kschischo, M. (2010). Grofit: Fitting biological growth curves with *r*. *Journal of Statistical Software*, 33(7), 1–21.
- Keeling, M. J. (2000a). Multiplicative moments and measures of persistence in ecology. *Journal of Theoretical Biology*, 205(2), 269–281.
- Keeling, M. J. (2000b). Metapopulation moments: Coupling, stochasticity and persistence. *Journal of Animal Ecology*, 69(5), 725–736.
- Kendall, D. G. (1948). On the generalized "birth-and-death" process. *The Annals of Mathematical Statistics*, 19(1), 1–15.
- Kendall, D. G. (1949). Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2), 230–282.
- Kingman, J. F. C. (1979). Deterministic and stochastic models in population genetics. *Advances in Applied Probability*, 11(2), 264.
- Krebs, C. J. (1985). *The experimental analysis of distribution and abundance*. Haper and Row.
- Krishnarajah, I., Cook, A., Marion, G., & Gibson, G. (2005). Novel moment closure approximations in stochastic epidemics. *Bulletin of Mathematical Biology*, 67(4), 855–873.
- Kurtz, T. G. (1972). The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7), 2976–2978.
- Lande, R., Engen, S., & Sæther, B.-E. (2003). *Stochastic population dynamics in ecology and conservation. Oxford series in ecology and evolution*. Oxford University Press.
- LeClair, J. S., & Wahl, L. M. (2017). The impact of population bottlenecks on microbial adaptation. *Journal of Statistical Physics*, 172(1), 114–125.
- López, S., Prieto, M., Dijkstra, J., Dhanoa, M., & France, J. (2004). Statistical evaluation of mathematical models for microbial growth. *International Journal of Food Microbiology*, 96(3), 289–300.
- Lotka, A. J. (1925). *Elements of physical biology*. Williams and Wilkins Company.
- Malthus, T. R. (1798). *An essay on the principle of population*. J. Johnson.
- Marion, G., Renshaw, E., & Gibson, G. (1998). Stochastic effects in a model of nematode infection in ruminants. *Mathematical Medicine and Biology: A Journal of the IMA*, 15(2), 97–116.

- Marrec, L., & Bitbol, A.-F. (2018). Quantifying the impact of a periodic presence of antimicrobial on resistance evolution in a homogeneous microbial population of fixed size. *Journal of Theoretical Biology*, 457, 190–198.
- Matis, J. H., & Kiffe, T. R. (1996). On approximating the moments of the equilibrium distribution of a stochastic logistic model. *Biometrics*, 52(3), 980–991.
- Matis, J. H., & Kiffe, T. R. (1999). Effects of immigration on some stochastic logistic models: A cumulant truncation analysis. *Theoretical Population Biology*, 56(2), 139–161.
- McKendrick, A. G. (1925). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44, 98–130.
- McKendrick, A. G., & Pai, M. K. (1912). XLV—The rate of multiplication of micro-organisms: A mathematical study. *Proceedings of the Royal Society of Edinburgh*, 31, 649–655.
- Morgan, B. J. T. (1976). Stochastic models of grouping changes. *Advances in Applied Probability*, 8(1), 30–57.
- Nagylaki, T. (2013). *Introduction to theoretical population genetics. Biomathematics*. Springer.
- Nåsell, I. (2003a). An extension of the moment closure method. *Theoretical Population Biology*, 64(2), 233–239.
- Nåsell, I. (2003b). Moment closure and the stochastic logistic model. *Theoretical Population Biology*, 63(2), 159–168.
- Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.
- Pearl, R. (1927). The growth of populations. *The Quarterly Review of Biology*, 2(4), 532–548.
- Pearl, R., & Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences of the United States of America*, 6(6), 275–288.
- Pelinovsky, E., Kurkin, A., Kurkina, O., Kokoulina, M., & Epifanova, A. (2020). Logistic equation and covid-19. *Chaos, Solitons & Fractals*, 140, 110241.
- Ram, Y., Dellus-Gur, E., Bibi, M., Karkare, K., Obolski, U., Feldman, M. W., Cooper, T. F., Berman, J., & Hadany, L. (2019). Predicting microbial growth in a mixed culture from growth curve data. *Proceedings of the National Academy of Sciences of the United States of America*, 116(29), 14698–14707.
- Regoes, R. R., Wiuff, C., Zappala, R. M., Garner, K. N., Baquero, F., & Levin, B. R. (2004). Pharmacodynamic functions: A multiparameter approach to the design of antibiotic treatment regimens. *Antimicrobial Agents and Chemotherapy*, 48(10), 3670–3676.
- Rice, S. H. (2004). *Evolutionary theory: Mathematical and conceptual foundations*. Sinauer.
- Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, 10(2), 290–301.
- Ross, S. (2007). *Introduction to probability models* (9th ed.). Academic Press.
- Shen, C. Y. (2020). Logistic growth modelling of covid-19 proliferation in China and its international implications. *International Journal of Infectious Diseases*, 96, 582–589.
- Shen, M., Peng, Z., Xiao, Y., & Zhang, L. (2020). Modeling the epidemic trend of the 2019 novel coronavirus outbreak in China. *The Innovation*, 1(3), 100048.
- Singh, A., & Hespánha, J. P. (2007). A derivative matching approach to moment closure for the stochastic logistic model. *Bulletin of Mathematical Biology*, 69(6), 1909–1925.
- Sprouffske, K., & Wagner, A. (2016). Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17(1), 172.
- Steiner, U. K. (2020). Tuljapurkar and Orzack (1980) and Tuljapurkar (1982a,b): Population dynamics in variable environments. *Theoretical Population Biology*, 133, 21–22.
- Strimmer, K., & Pybus, O. G. (2001). Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution*, 18(12), 2298–2305.
- Sunhwa, C., & Moran, K. (2020). Estimating the reproductive number and the outbreak size of covid-19 in Korea. *Epidemiol Health*, 42, e2020011.
- Taylor, H., Karlin, S., & Taylor, H. (1998). *An Introduction to stochastic modeling*. Elsevier Science.
- Tsoularis, A., & Wallace, J. (2002). Analysis of logistic growth models. *Mathematical Biosciences*, 179(1), 21–55.
- Tuljapurkar, S. D. (1982a). Population dynamics in variable environments. II. Correlated environments, sensitivity analysis and dynamics. *Theoretical Population Biology*, 21(1), 114–140.
- Tuljapurkar, S. D. (1982b). Population dynamics in variable environments. III. Evolutionary dynamics of r-selection. *Theoretical Population Biology*, 21(1), 141–165.
- Tuljapurkar, S. D., & Orzack, S. H. (1980). Population dynamics in variable environments I. long-run growth rates and extinction. *Theoretical Population Biology*, 18(3), 314–342.
- Uecker, H., & Hermisson, J. (2011). On the fixation process of a beneficial mutation in a variable environment. *Genetics*, 188(4), 915–930.
- Van Kampen, N. G. (2011). *Stochastic processes in physics and chemistry*. North-Holland Personal Library. Elsevier Science.
- Vega, N. M., & Gore, J. (2017). Stochastic assembly produces heterogeneous communities in the caenorhabditis elegans intestine. *PLoS Biology*, 15(3), 1–20.
- Verhulst, P. F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondence Mathématique et Physique*, 10, 113–121.
- Wahl, L. M., & Gerrish, P. J. (2001). The probability that beneficial mutations are lost in populations with periodic bottlenecks. *Evolution*, 55(12), 2606–2610.
- Wahl, L. M., Gerrish, P. J., & Saika-Voivod, I. (2002). Evaluating the impact of population bottlenecks in experimental evolution. *Genetics*, 162(2), 961–971.
- Wakano, J. Y., & Iwasa, Y. (2013). Evolutionary branching in a finite population: Deterministic branching vs. stochastic branching. *Genetics*, 193(1), 229–241.
- Wan, K., Chen, J., Lu, C., Dong, L., Wu, Z., & Zhang, L. (2020). When will the battle against novel coronavirus end in Wuhan: A seir modeling analysis. *Journal of Global Health*, 10(1), 011002.
- Weber, M. F., & Frey, E. (2017). Master equations and the theory of stochastic path integrals. *Reports on Progress in Physics*, 80(4), 046601.
- Whittle, P. (1957). On the use of the normal approximation in the treatment of stochastic processes. *Journal of the Royal Statistical Society. Series B*, 19(2), 268–281.
- Wilson, W. G. (1998). Resolving discrepancies between deterministic population models and individual-based simulations. *The American Naturalist*, 151(2), 116–134.
- Wu, K., Darcet, D., Wang, Q., & Sornette, D. (2020). Generalized logistic growth modeling of the covid-19 outbreak: Comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear Dynamics*, 101(3), 1561–1581.
- Yates, C. A., Ford, M. J., & Mort, R. L. (2017). A multi-stage representation of cell proliferation as a Markov process. *Bulletin of Mathematical Biology*, 79(12), 2905–2928.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Marrec, L., Bank, C., & Bertrand, T. (2023). Solving the stochastic dynamics of population growth. *Ecology and Evolution*, 13, e10295. <https://doi.org/10.1002/ece3.10295>



## APPENDIX A

## Detailed simulation method

In this work, the population growths are simulated using a Gillespie algorithm.

## A.1. | Single-species population

Let us denote by  $N$  the population size. The single elementary event that can happen is the reproduction of an individual:



Simulations follow these elementary steps:

1. Initialization: The population starts from  $N = N_0$  individuals at time  $t = 0$ .
2. The time increment  $\Delta t$  is sampled randomly from an exponential distribution with mean  $1/B_N$ .
3. The population size increases from  $N$  to  $N + 1$  and the time from  $t$  to  $t + \Delta t$ .
4. We go back to Step 2 and iterate until the total number of individuals is  $K$  or the defined maximum time is reached.

## A.2. | Community

Let us consider two types of individuals, namely  $M$  and  $W$ , whose numbers are denoted by  $N_M$  and  $N_W$ , respectively. The two elementary events that can happen are the reproduction of an individual  $M$  or  $W$ :



Simulations follow these elementary steps:

1. Initialization: The population starts from  $N_M = N_{M,0}$   $M$  individuals and  $N_W = N_{W,0}$   $W$  individuals at time  $t = 0$ .
2. The time increment  $\Delta t$  is sampled randomly from an exponential distribution with mean  $1 / (B_{N_M, N_W}^{(M)} + B_{N_M, N_W}^{(W)})$ . The next event that may occur is chosen randomly, proportionally to its probability. For instance, reproduction of a  $W$  individual is chosen with probability  $B_{N_M, N_W}^{(W)} / (B_{N_M, N_W}^{(M)} + B_{N_M, N_W}^{(W)})$ .
3. The time increases from  $t$  to  $t + \Delta t$  and the event chosen at Step 2 is executed. For instance, if reproduction of a  $W$  individual is chosen, then  $N_W$  increases by one.
4. We go back to Step 2 and iterate until the total number of individuals is  $K$  or the defined maximum time is reached.

## A.3. | Serial passage experiment

Let us consider two types of individuals, namely  $M$  and  $W$ , whose numbers are denoted by  $N_M$  and  $N_W$ , respectively. The two elementary events that can happen are the reproduction of an individual  $M$  or  $W$ :



Simulations follow these elementary steps:

1. Initialization: The population starts from  $N_M = N_{M,0}$   $M$  individuals and  $N_W = N_{W,0}$   $W$  individuals at time  $t = 0$ . The next time when the dilution occurs is stored in the variable  $t_{\text{dilution}}$ , which is initialized at  $t_{\text{dilution}} = \tau$ , the first time when dilution occurs.
2. The time increment  $\Delta t$  is sampled randomly from an exponential distribution with mean  $1 / (B_{N_M, N_W}^{(M)} + B_{N_M, N_W}^{(W)})$ . The next event that may occur is chosen randomly, proportionally to its probability. For instance, reproduction of a  $W$  individual is chosen with probability  $B_{N_M, N_W}^{(W)} / (B_{N_M, N_W}^{(M)} + B_{N_M, N_W}^{(W)})$ .
3. If  $t + \Delta t < t_{\text{dilution}}$ , time is increased to  $t + \Delta t$  and the event chosen at Step 2 is executed. For instance, if reproduction of a  $W$  individual is chosen, then  $N_W$  increases by one.
4. If  $t + \Delta t \geq t_{\text{dilution}}$ , the event chosen at Step 2 is not executed because a dilution has to occur before. The dilution is performed: time is incremented to  $t = t_{\text{dilution}}$ ,  $N_{M,0}$   $M$  individuals are selected from the community following a binomial law  $\mathcal{B}(D \times K, N_M / (N_M + N_W))$ , and  $N_{W,0} = D \times K - N_{M,0}$ . In addition,  $t_{\text{dilution}}$  is incremented to  $t_{\text{dilution}} + \tau$ , and thus stores the next time when the dilution occurs.
5. We go back to Step 2 and iterate until there is only one species left.

Note that we mostly deal with parameter values such that, at the end of each growth phase, the population size is equal to the carrying capacity almost certainly, that is,  $P_{N_0, K}(\tau) \approx 1$ . Otherwise, we would be required to consider  $N_M + N_W$  rather than  $K$  at step 4.

## APPENDIX B

## From the stochastic model to the deterministic limit

Here, we present a full derivation of the deterministic limit of the stochastic pure-birth model. Starting from the master equation satisfied by the probabilities of having a given population size at a given time, we obtain a Fokker-Planck equation, corresponding to the diffusion approximation, and then a deterministic differential equation, in the limits of increasingly large population sizes. Let us first recall the master equation corresponding to the pure-birth process, where  $N$  denotes the number of individuals.

$$\frac{dP_{N_0, N}(t)}{dt} = B_{N-1} P_{N_0, N-1}(t) - B_N P_{N_0, N}(t), \quad (B1)$$

Let us now introduce the reduced variable  $n = N/K$ , as well as  $\rho(n, t | n_0) = K \times P_{N_0, N}(t)$  and  $\tilde{B}(n) = B_N/K$  (e.g.,  $\tilde{B}(n) = b(1-n^\beta)^\gamma n = b(1-(N/K)^\beta)^\gamma N/K = B_N/K$  for the generalized

logistic model with  $\alpha = 1$ ). Then, the master equation (4b) can be rewritten as

$$\frac{1}{K} \frac{\partial \rho(n, t | n_0)}{\partial t} = \tilde{B}(n - 1/K) \rho(n - 1/K) - \tilde{B}(n) \rho(n, t). \quad (\text{B2})$$

### B.1. | Diffusion approximation

For  $K \gg 1$ , considering that jumps are small at each step of the pure-birth process, that is,  $1/K \ll n$ , the probability density  $\rho(n, t | n_0)$  and the transition probability  $\tilde{B}(n)$  can be expanded in a Taylor series around  $n$ . This expansion, known as a Kramers-Moyal expansion, yields, to first order in  $1/K$

$$\frac{\partial \rho(n, t | n_0)}{\partial t} = - \frac{\partial}{\partial n} (\rho(n, t | n_0) \tilde{B}(n)) + \frac{1}{2K} \frac{\partial^2}{\partial n^2} (\rho(n, t | n_0) \tilde{B}(n)). \quad (\text{B3})$$

The previous equation is known as a diffusion equation, or a Fokker-Planck equation, or a Kolmogorov forward equation.

### B.2. | Deterministic limit

In the limit  $K \rightarrow \infty$ , retaining only the zeroth-order terms in  $1/K$ , Equation 5 reduces to

$$\frac{\partial \rho(n, t | n_0)}{\partial t} = - \frac{\partial}{\partial n} (\rho(n, t | n_0) \tilde{B}(n)). \quad (\text{B4})$$

Let us focus on the average value of  $n$ , denoted by  $\langle n \rangle$ . Using Equation 6 yields

$$\frac{d\langle n \rangle}{dt} = \int_0^1 n \frac{\partial \rho(n, t | n_0)}{\partial t} dn, \quad (\text{B5})$$

$$= - \int_0^1 n \frac{\partial}{\partial n} (\rho(n, t | n_0) \tilde{B}(n)) dn, \quad (\text{B6})$$

$$= - [\rho(n, t | n_0) \tilde{B}(n)]_0^1 + \int_0^1 \rho(n, t | n_0) \tilde{B}(n) dn, \quad (\text{B7})$$

$$= \langle \tilde{B}(n) \rangle. \quad (\text{B8})$$

The first term of right-hand side of Equation 9 vanishes because  $\tilde{B}(0) = \tilde{B}(1) = 0$ . In the limit  $K \rightarrow \infty$ , the distribution of  $n$  can be assumed very peaked around its mean, so  $\langle n \rangle \approx n$  and  $\langle \tilde{B}(n) \rangle \approx \tilde{B}(n)$ , yielding

$$\frac{dn}{dt} = \tilde{B}(n). \quad (\text{B9})$$

Multiplying both sides of the previous equation by the carrying capacity  $K$  leads to

$$\frac{dN}{dt} = B_N. \quad (\text{B10})$$

## APPENDIX C

### Additional figures

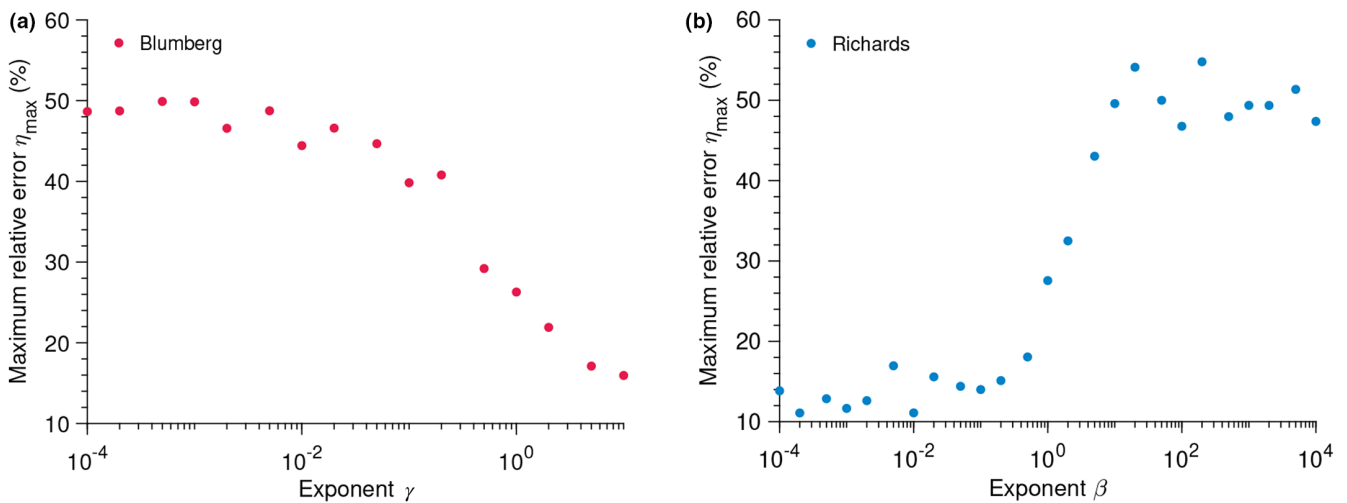


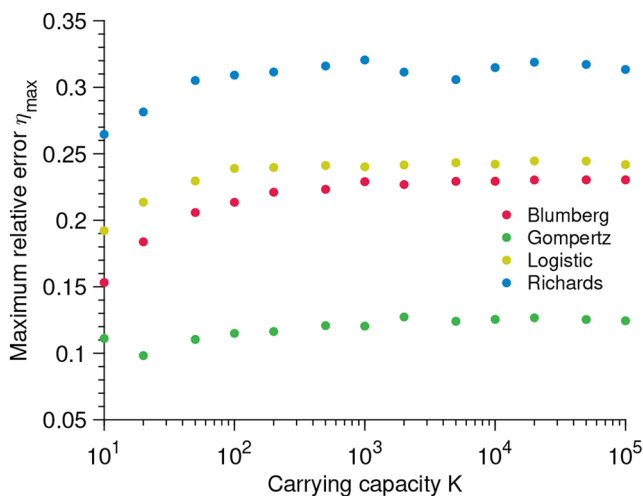
FIGURE A1 The maximum relative error remains high for all parameter values. Maximum relative error  $\eta_{\max}$  versus exponent  $\gamma$  (a) and  $\beta$  (b) for the Blumberg and Richards models, respectively. Parameter values:  $K = 100$ ,  $N_0 = 1$ ,  $b = 1$  and  $\alpha = 1$ .

## APPENDIX D

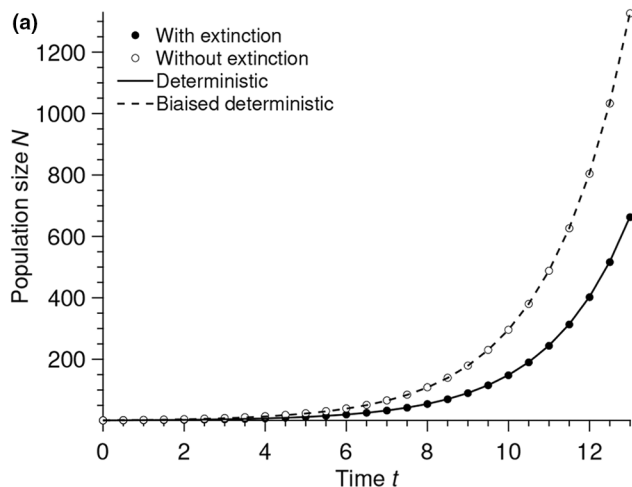
## Birth-death process

Let us consider a population following a linear birth-death process. The population size, the birth rate, and the death rate are denoted by  $N$ ,  $b$ , and  $d$ , respectively. The master equation giving the probability  $P_{N_0,N}(t)$  that the population size has a given size  $N$  at a given time  $t$  starting from  $N_0$  individuals reads.

$$\frac{dP_{N_0,N}(t)}{dt} = b(N-1)P_{N_0,N-1}(t) + d(N+1)P_{N_0,N+1}(t) - (b+d)NP_{N_0,N}(t). \quad (D1)$$



**FIGURE A2** The maximum relative error does not depend on the carrying capacity. Maximum relative error  $\eta_{\max}$  versus carrying capacity  $K$  for different population growth models. Parameter values:  $K = 100$ ,  $N_0 = 1$ ,  $b = 1$ ,  $\alpha = 1$ ,  $\beta = 2$  and  $\gamma = 1.5$ .



Note that the analytical solution of Equation (11c) is known (Kendall, 1948). One can derive the mean population size from the master equation using  $\langle N \rangle = \sum_{N=0}^K N \times P_{N_0,N}(t)$ , which yields

$$\frac{d\langle N \rangle}{dt} = (b-d)\langle N \rangle. \quad (D2)$$

The solution of the previous ordinary differential equation is given by

$$\langle N \rangle(t) = N_0 e^{(b-d)t}. \quad (D3)$$

The solution is similar to the deterministic limit

$$N(t) = N_0 e^{(b-d)t}. \quad (D4)$$

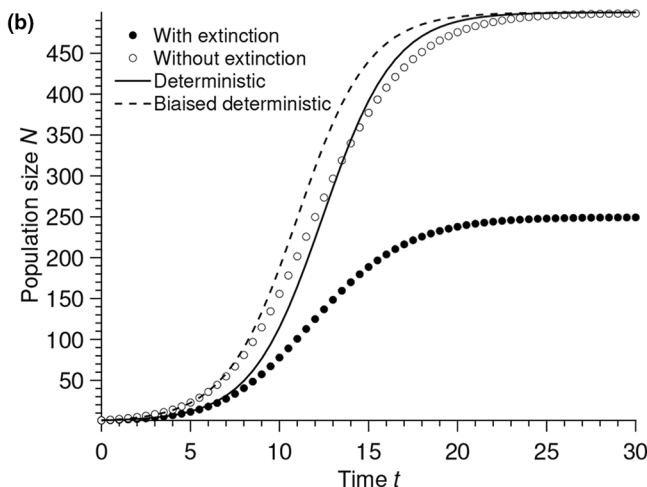
As shown in Figure S3a, the deterministic limit contains both trajectories: survival and extinction.

Now let us consider a master equation biased on the stochastic trajectories that lead to fixation. To do so, let us define  $Q_{N_0,N}(t) = P_{N_0,N}(t) / (1 - P_{N_0,0}(t))$ . The biased master equation reads

$$\frac{dQ_{N_0,N}(t)}{dt} = b(N-1)Q_{N_0,N-1}(t) + d(N+1)Q_{N_0,N+1}(t) - (b+d)NQ_{N_0,N}(t) + \frac{Q_{N_0,N}(t)}{1 - P_{N_0,0}(t)} \frac{dP_{N_0,0}(t)}{dt}. \quad (D5)$$

The mean population size averaged over stochastic trajectories leading to fixation is given by  $\langle N \rangle_{\theta}(t) = \sum_{N=0}^K N \times Q_{N_0,N}(t)$ , which yields

$$\frac{d\langle N \rangle_{\theta}}{dt} = (b-d)\langle N \rangle_{\theta} + \frac{\langle N \rangle_{\theta}}{1 - P_{N_0,0}(t)} \frac{dP_{N_0,0}(t)}{dt}, \quad (D6)$$



**FIGURE A3** The deterministic limit correctly describes the dynamics of a linear death-birth process but not of a nonlinear death-birth process. Population size as a function of time for a linear birth-death process (a) and a nonlinear birth-death process (b). The logistic model was used for the nonlinear birth-death process. The data points are averaged over  $10^4$  stochastic realizations. Filled data points are averaged over all stochastic trajectories, whether they lead to survival or extinction, whereas empty data points are averaged over stochastic trajectories leading to survival only. The solid lines correspond to the deterministic limit, whereas the dashed lines correspond to the "biased" deterministic limit (i.e., conditioned on non-extinction or survival). Parameters values:  $b = 1$ ,  $d = 0.5$ ,  $N_0 = 1$ , and  $K = 10^3$ .

where

$$P_{N_0,0}(t) = \left(1 + \frac{d-b}{b-de^{(d-b)t}}\right)^{N_0}. \quad (\text{D7})$$

Note that the previous equation allows us to recover  $P_{N_0,0}(t) \propto (d/b)^{N_0}$ , which is the probability that an early extinction occurs. After solving Equation 14, one obtains

$$\langle N \rangle_{\theta}(t) = \frac{N_0 e^{(b-d)t}}{1 - P_{N_0,0}(t)}. \quad (\text{D8})$$

Applying the same reasoning to the logistic growth, which is a nonlinear birth-death process, leads to Figure S3b. The only difference between the linear and the nonlinear birth-death processes is that a mean-field approximation is required for the latter to derive the deterministic limit. Note that one can assume that  $P_{N_0,0}(t)$  for the nonlinear birth-death process is the same as for the linear birth-death process as early extinctions occur when the density dependence can be neglected.