



# How to avoid an extinction time paradox

Alexander G. Strang<sup>1</sup> · Karen C. Abbott<sup>2</sup> · Peter J. Thomas<sup>1</sup>

Received: 29 May 2018 / Accepted: 11 February 2019  
© Springer Nature B.V. 2019

## Abstract

An essential topic in theoretical ecology is the extinction of populations subject to demographic stochasticity. Mechanistic models of demographic stochasticity, such as birth-death processes, can be analytically intractable, so are frequently approximated with stochastic differential equations (SDEs). Here, we consider two pitfalls in this type of approximation. First, familiar deterministic models are not always appropriate for use in an SDE. Second, the common practice of starting directly from an SDE without explicitly constructing a mechanistic model leaves the noise term up to the modeler's discretion. Since the stability of stochastic models depends on the global properties of both the noise and the deterministic model, overly phenomenological deterministic models, or heuristic choices of noise, can lead to models that are unrealistically stable. The goal of this article is to provide an example of how both of these effects can undermine seemingly reasonable models. Following Dennis et al. (Theor Ecol 9:323–335 2016) and Levine and Meerson (PRE, 87:032127 2013), we compare the persistence of stochastic extensions of standard logistic and Allee models. We show that, for common choices of noise, stochastic logistic models become exponentially *less* extinction prone when a strong Allee effect is introduced. This apparent paradox can be resolved by recognizing that common models of an Allee effect introduce overcompensation that dominates the extinction dynamics, even when the deterministic model is rescaled to account for overcompensation. These problems can be resolved by mechanistic treatment of the deterministic model and the noise.

**Keywords** Extinction time · Allee effects · Stochastic differential equations · Quasi-potentials

## Introduction: the paradox

In this article, we will compare the extinction process of four different stochastic population models. All four models are constructed directly from deterministic models, which are then modified by an additional noise term. We will consider two different deterministic models and two different choices of noise. These models are intentionally chosen to illustrate difficulties that arise in the construction

of stochastic models, even when starting from very familiar and well-understood deterministic models.

Our deterministic models are defined:

$$\frac{d}{dt}x(t) = m(x) \quad (1)$$

where  $m(x)$  is the rate of change of the population  $x(t)$ .

The first deterministic model we consider is the logistic model, defined by setting:

$$m(x) = \frac{r}{K}x(K - x). \quad (2)$$

Here,  $r$  represents the per capita growth rate in the population as  $x$  goes to zero and  $K$  is the carrying capacity. This model has two fixed points: the extinction state  $x = 0$  and the carrying capacity  $x = K$ . The extinction state is unstable, and  $x(t)$  approaches  $K$  from any  $x(0) \neq 0$ . Notice that the per capita growth rate,  $\frac{r}{K}(K - x)$ , is largest when the population  $x$  is small.

The second model we consider introduces an Allee effect. An Allee effect modifies the logistic model so that, for sufficiently small  $x$ , the per capita growth rate decreases as  $x$  decreases (Stephens et al. 1999). These effects often arise from mechanisms that suppress the per

✉ Alexander G. Strang  
ags61@case.edu

Karen C. Abbott  
kcabbott@case.edu

Peter J. Thomas  
pjthomas@case.edu

<sup>1</sup> Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, Yost Hall, 2049 Martin Luther King Jr. Drive, Cleveland, OH 44106, USA

<sup>2</sup> Department of Biology, Case Western Reserve University, 2080 Adelbert Road, Cleveland, OH 44106, USA

capita growth rate at low densities (Boukal and Berec 2002; McCarthy 1997). Classic examples include increased difficulty finding mates, decreased chance of pollination, loss of genetic diversity, or the collapse of social groups (Courchamp et al. 1999; Stephens et al. 1999). A strong Allee effect reduces the per capita growth rate enough that it falls below zero for  $x$  less than some threshold  $A$ . Beneath this threshold the population tends to extinction. Therefore, the introduction of a strong Allee effect should greatly increase a population's risk of extinction by reducing the scale of fluctuation away from  $K$  needed to reach extinction (Courchamp et al. 1999; Stephens and Sutherland 1999). As a result, Allee effects are central to the conservation of populations at low density (Stephens and Sutherland 1999).

Allee effects have been modeled extensively (Boukal and Berec 2002; Stephens and Sutherland 1999). Here we consider the simplest Allee model. A strong Allee effect introduces an Allee threshold  $A$  such that for  $x < A$  the population necessarily tends to extinction. This can be accomplished by replacing the quadratic rate equation used in the logistic model with a cubic rate equation:

$$m(x) = \frac{r}{AK}x(x - A)(K - x) \quad (3)$$

where  $A$  is the Allee threshold (Courchamp et al. 1999; Poggiale 1998).

Although simple, this model is widely used in the literature (Amarasekare 1998; Courchamp et al. 1999; Dennis et al. 2016; Gruntfest et al. 1997; Keitt et al. 2001; Kussell and Vucelja 2014; Lewis and Kareiva 1993; Levine and Meerson 2013; Poggiale 1998). It is often chosen for its tractability and familiarity. In fact, it is so ubiquitous that in some reviews, it is the only Allee model shown explicitly (Courchamp et al. 1999). Our use follows a pair of recent papers (Dennis et al. 2016; Lewis and Kareiva 1993) that adopt this cubic model to study the persistence of fluctuating populations. Note that this model provides the desired asymptotic behavior without addressing an explicit underlying mechanism. It is usually used as a "phenomenological" model when a specific mechanism is either unknown or not of interest (Boukal and Berec 2002). Alternative models based on specific mechanisms are studied in Dennis (1989), Dennis (2002), and McCarthy (1997) and addressed in Appendix A.6.

While multiplication by the monomial  $(x - A)/A$  introduces a strong Allee effect in the vicinity of  $A$ , it also increases the rate of return to  $K$  when  $x$  is near  $K$ . This overcompensation effect is usually counterbalanced by picking the characteristic growth rate  $r$  so that the growth rate in the Allee model,  $r_A$ , is less than the growth rate in the logistic model,  $r_L$  (Dennis 2002; Gruntfest et al. 1997; Lewis and Kareiva 1993; Poggiale 1998). We start with the direct comparison  $r_A = r_L$  since the results are most apparent in this case. Our analysis generalizes easily to situations when  $r_A < r_L$ , for which the qualitative results are largely unchanged (see Section "Does rescaling resolve the paradox?").

To study the extinction process we replace our deterministic process  $x(t)$  with stochastic processes  $X(t)$ . This is accomplished by replacing the deterministic ordinary differential equation (1) with stochastic differential equations of the form:

$$dX = m(X)dt + \sqrt{2v(X)}dW. \quad (4)$$

Here,  $m(X)$  is the infinitesimal change in the expected population and  $v(X)$  is the infinitesimal change in the population variance. These are the mean drift and diffusion coefficient respectively. The Wiener increment,  $dW$ , is the accumulation of Gaussian white noise over the interval  $dt$ . Technically (4) is just a shorthand for an integral equation, which cannot be performed without first choosing an interpretation of integration over  $dW$  (Van Kampen 1980). Here we follow the Ito convention. As in Dennis (2002) and Stephens et al. (1999) we do not consider demographic stochasticity to be an Allee effect.

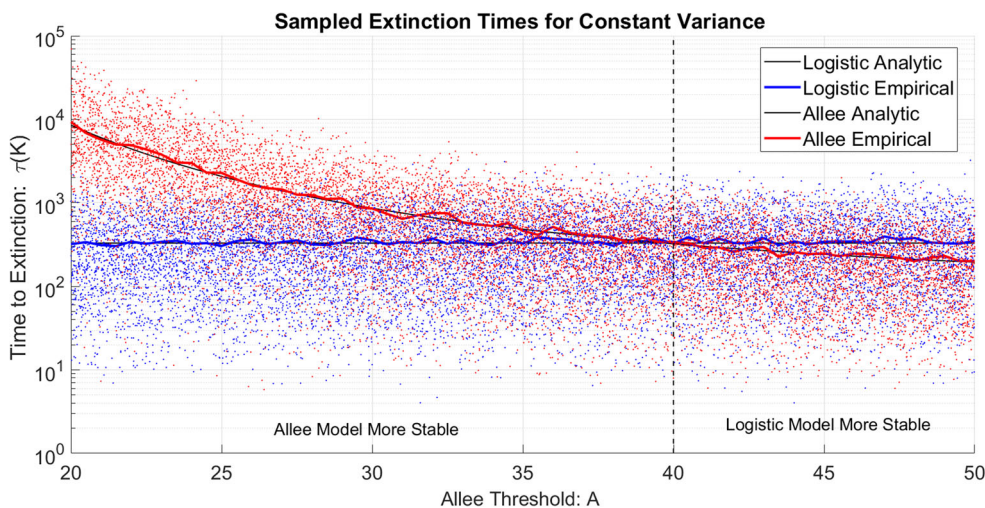
In order to specify a specific stochastic model for the process  $X(t)$  we need to pick the noise variance  $v(x)$ . We will start by considering two common choices, constant variance  $v(x) = \frac{1}{2}\sigma^2$ , and linear variance  $v(x) = \frac{1}{2}\sigma^2x$ . The first represents random, density independent immigration and emigration, and is often chosen for analytic convenience (as in Nolting and Abbott 2016). The second is more realistic in certain general settings (Desharnais Robert et al. 2006), and is commonly used to model demographic noise (Dennis et al. 2016; Kamenev et al. 2008; Tier and Hanson 1981).

To summarize, we consider the four models:

$$\begin{aligned} 1a. \text{ Logistic with Constant Variance: } dX &= \frac{r}{K}X(K - X)dt + \sigma dW \\ 1b. \text{ Logistic with Linear Variance: } dX &= \frac{r}{K}X(K - X)dt + \sigma\sqrt{X}dW \\ 2a. \text{ Allee with Constant Variance: } dX &= \frac{r}{AK}X(X - A)(K - X)dt + \sigma dW \\ 2b. \text{ Allee with Linear Variance: } dX &= \frac{r}{AK}X(X - A)(K - X)dt + \sigma\sqrt{X}dW. \end{aligned} \quad (5)$$

To illustrate the danger of constructing SDEs from phenomenological models with ad hoc noise terms, we use

these four models to ask: how does the introduction of an Allee effect change the risk of extinction for populations



**Fig. 1** Sampled first passage times to extinction for models 1a and 2a (constant variance)  $r = 0.1$ ,  $K = 160$ ,  $\sigma = 20$  and  $A \in [20, 50]$ . The extinction time for each individual simulation is shown with a colored dot, blue for logistic, red for Allee. The corresponding means are

plotted as thick colored lines (blue for logistic, red for Allee). The analytic mean times to extinction are shown in black. The intercept in the analytic first passage times is shown as a dotted black line. For all  $A$  beneath the intercept the Allee model is more stable than the Logistic model. This is evidence of the paradox

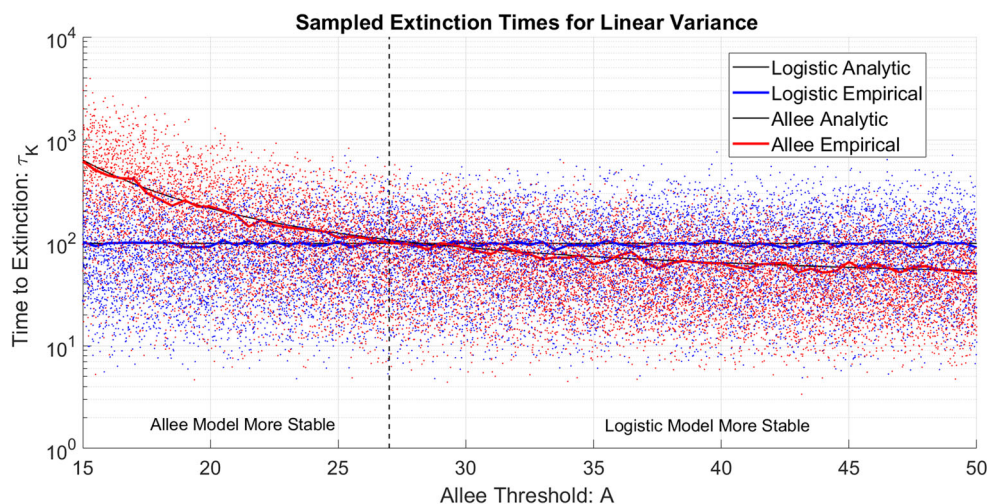
starting at carrying capacity? To answer this question, we compute the mean time to extinction,  $\tau(K)$ , in all four models and compare 1a with 2a, and 1b with 2b.

To illustrate the persistence,  $\tau(K)$ , of the four models, we considered the following example. We set  $r = 0.1$  and  $K = 160$ . For models 1a and 2a (constant variance), we set  $\sigma = 20$  and varied  $A$  from 20 to 50 in increments of 0.5. For each increment, we ran 200 simulations starting from  $K$ . To simulate the population, we used the Milstein method with time step  $dt = 0.01$  (Higham 2001). Each simulation was terminated once the stochastic process reached an extinction threshold ( $X(t) < \epsilon$  for  $\epsilon = 10^{-8}$ ). The time increment was chosen by gradually decreasing  $dt$  until the empirical mean stopped changing (within uncertainty in the mean). We then computed the mean time to extinction and variance

in time to extinction for each  $A$ . The mean and variance in simulated extinction times closely matched analytic predictions from an exponentially distributed extinction process (for analytical details see Sections “Theory”—“How to avoid the paradox”). Results are shown in Fig. 1.

Figure 1 reveals the paradox: for  $A \lesssim 37$  the mean time to extinction in the Allee model is *greater* than the mean time to extinction in the logistic model. This difference becomes larger as  $A$  becomes smaller. For  $A \lesssim 25$  the mean time to extinction in the Allee model is ten to a hundred times *larger* than the mean time to extinction in the logistic model. This directly contradicts the intuition that the introduction of an Allee effect should make a population at carrying capacity less stable. It also contradicts results for discrete time population models, where an Allee effect has

**Fig. 2** Sampled times to extinction for models 1b and 2b (linear variance)  $r = 0.1$ ,  $K = 160$ ,  $\sigma = 20$  and  $A \in [20, 50]$ . The extinction time for each individual simulation is shown with a colored dot, blue for logistic, red for Allee. The corresponding means are plotted as thick colored lines (blue for logistic, red for Allee). The analytic mean times to extinction are shown in black. The intercept is shown as a dotted black line. For all  $A$  beneath the intercept the Allee model is more stable. This is further evidence of the paradox



been shown to greatly decrease the mean time to extinction (Stephan and Wissel 1994).

The same comparison was repeated for the two linear variance models (1b and 2b) to much the same effect. Here the parameters  $r$  and  $K$  were kept the same, the noise intensity was reduced to  $\sigma = 3$ , and we let  $A$  range from [15, 50] (Fig. 2). While the effect is less pronounced, the difference in mean extinction times is well beyond the empirical uncertainty in the mean extinction times, and for small  $A$  leads to a separation of close to an order of magnitude.

In sum, these seemingly reasonable stochastic differential equations lead to highly illogical results. Adding an Allee threshold to our models extended their persistence by orders of magnitude, with stronger effect at lower thresholds. This is in direct contrast to the biology the Allee models are meant to represent, wherein the presence of an Allee threshold heightens extinction risk and the significance of having an Allee effect should be greater the farther the threshold is from zero (Stephens et al. 1999).

The rest of this paper is devoted to explaining this paradox and showing that it reflects a more general problem in the construction of stochastic population models. Namely, when introducing an effect to a model one must be careful to not accidentally introduce an additional effect which leads to a wrong conclusion. The analytic method for solving for  $\tau(K)$  is discussed in Section “Theory.” From the exact expression, we derive an intuitive extinction time estimator that simplifies the discussion. This intuition will depend heavily on a physical analogy and offers a clear qualitative understanding of the paradox (see Appendix A.5). The estimators are applied to the four models in Section “Analysis.” There, we establish an asymptotic bound for  $A$  given large  $K$  beneath which the Allee models are exponentially more stable than the logistic models. With the estimators in hand, we can easily consider other scalings of the Allee model that set  $r_A < r_L$ . We show that rescaling the deterministic model alone may not resolve the paradox, and often introduces new problems. In addition, when the noise is scaled commensurately, the rescaling never resolves the paradox. We then show that the paradox arises both from cavalier treatment of  $v(x)$ , and from differences between the construction of deterministic models and stochastic models. In particular, phenomenological stochastic models should be constructed with global stability in mind, not just the location of saddles and the local stability of equilibria. Based on these considerations we introduce two simple guidelines for constructing models that avoid the paradox (Section “How to avoid the paradox”). These guidelines lead to a natural bound on the asymptotic scaling of mean times to extinction that matches results reported in the literature (Foley 1994; Kamenev et al. 2008; Lande 1993; Ovaskainen and Meerson 2010).

## Theory

We interpret the four models introduced in the previous section (4) using Ito’s integration rule. Assume that  $m(x)$  scales with a parameter  $r$  as in Eq. 2 and  $v(x)$  scales with  $\sigma^2$  as in Eq. 5. Under these assumptions, the probability density  $\pi(x, t)$  that  $X(t) = x$  evolves according to the Ito form of the Fokker-Planck equation (Van Kampen 1980):

$$\frac{\partial}{\partial t} \pi(x, t) = -\frac{\partial}{\partial x} (rm(x)\pi(x, t)) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} (v(x)\pi(x, t)). \quad (6)$$

Technically, the following analysis finds the mean first passage time to a neighborhood of extinction of size  $\epsilon$ ; however, for small  $\epsilon$ , this analysis is biologically equivalent to finding the time to extinction.

The first passage time into  $\epsilon$  is defined as the earliest time at which  $X(t)$  reaches  $\epsilon$ . Since  $X(t)$  is a stochastic process the first passage time is a random variable. As discussed in Section “Introduction: the paradox,” the distribution of extinction times is exponential and thus determined entirely by the mean first passage time to  $\epsilon$ .

Denote the mean first passage time (MFPT) from a state  $x$  to the final state  $\epsilon$  as  $\tau_\epsilon(x)$ . Given infinitesimal mean  $rm(x)$  and variance  $\sigma^2 v(x)/2 > 0$  the MFPT  $\tau_\epsilon(x)$  is the solution to the backward equation:

$$rm(x) \frac{\partial}{\partial x} \tau_\epsilon(x) + \frac{\sigma^2}{2} v(x) \frac{\partial^2}{\partial x^2} \tau_\epsilon(x) = -1 \quad (7)$$

with boundary conditions  $\tau_\epsilon(\epsilon) = 0$  and  $\lim_{x \rightarrow \infty} \partial_x \tau_\epsilon(x) = 0$  (Allen and Allen 2002; Bresloff 2014; Tier and Hanson 1981).

Assume that  $v(x) > 0$  for all  $x > 0$ .<sup>1</sup> Then, Eq. 7 can be rewritten:

$$\frac{\partial^2}{\partial x^2} \tau_\epsilon(x) + \frac{2r}{\sigma^2} \frac{m(x)}{v(x)} \frac{\partial}{\partial x} \tau_\epsilon(x) = -\frac{2}{\sigma^2} \frac{1}{v(x)}. \quad (8)$$

Let  $f(x) = \frac{m(x)}{v(x)}$  and let  $T = \frac{\sigma^2}{2r}$ . These are analogous to forces and temperature in thermodynamic systems governed by the same equation (4) (Hong et al. 2002). For details on the physical analogy, see Section “A.5 Physical analogy.” Loosely speaking, the population tends to grow when  $f(x) > 0$  and to decrease when  $f(x) < 0$ . When the temperature,  $T$ , is large, the process is noisy and is more likely to move against  $f(x)$ . When the temperature is small, the process is close to deterministic.

<sup>1</sup>If extinction is absorbing then  $v(0) = 0$  and the process  $X(t)$  is a singular diffusion (Tier and Hanson 1981). This motivates the choice  $0 < \epsilon \leq 1$ .

Let  $y_\epsilon(x) = \frac{\partial}{\partial x} \tau_\epsilon(x)$ . Then:

$$\frac{\partial}{\partial x} y_\epsilon(x) + \frac{1}{T} f(x) y_\epsilon(x) = -\frac{2}{\sigma^2} \frac{1}{v(x)}. \tag{9}$$

This is an inhomogeneous first-order equation, which can be solved by integrating factors:

$$y_\epsilon(x) = c_1 \exp\left(-\frac{1}{T} \int_\epsilon^x f(s) ds\right) - \exp\left(-\frac{1}{T} \int_\epsilon^x f(s) ds\right) \int_\epsilon^x \frac{2}{\sigma^2} \frac{1}{v(s)} \exp\left(\frac{1}{T} \int_\epsilon^s f(\zeta) d\zeta\right). \tag{10}$$

This solution can be rewritten in terms of a potential function (Dennis et al. 2016),  $S(x)$ , that provides a convenient summary of the models' tendency to move towards  $K$ :<sup>2</sup>

$$S(x) = -\int_0^x f(s) ds. \tag{11}$$

Then:

$$y_\epsilon(x) = \exp\left(\frac{1}{T} (S(x) - S(\epsilon))\right) \times \left[ c_1 - \int_\epsilon^x \frac{2}{\sigma^2} \frac{1}{v(s)} \exp\left(-\frac{1}{T} (S(s) - S(\epsilon))\right) ds \right]. \tag{12}$$

If the population has a carrying capacity (above which  $m(x) < 0$ ) and the ratio  $m(x)/v(x)$  does not converge to zero as  $x$  goes to infinity (the general case provided  $v(x)$  is of the same order in  $x$  as  $m(x)$ ) then  $S(x)$  diverges to positive infinity as  $x$  goes to infinity. Therefore, the integral  $\int_\epsilon^x \frac{1}{v(s)} \exp\left(-\frac{1}{T} (S(s) - S(\epsilon))\right) ds$  converges to a constant as  $x$  goes to infinity. So, to enforce the second boundary condition set  $c_1 = \int_\epsilon^\infty \frac{1}{v(s)} \exp\left(-\frac{1}{T} (S(s) + S(\epsilon))\right) ds$ . Then:

$$y_\epsilon(x) = \int_x^\infty \frac{2}{\sigma^2} \frac{1}{v(s)} \exp\left(\frac{1}{T} (S(x) - S(s))\right) ds. \tag{13}$$

Now, to recover the first passage time integrate  $y_\epsilon(x)$ :

$$\tau_\epsilon(x) = \int_\epsilon^x \int_s^\infty \frac{2}{\sigma^2} \frac{1}{v(\zeta)} \exp\left(\frac{1}{T} (S(s) - S(\zeta))\right) d\zeta ds + c_2 \tag{14}$$

The first boundary condition  $\tau_\epsilon(\epsilon) = 0$  requires  $c_2 = 0$ . Therefore:

$$\tau_\epsilon(x) = \int_\epsilon^x \int_s^\infty \frac{2}{\sigma^2} \frac{1}{v(\zeta)} \exp\left(\frac{1}{T} (S(s) - S(\zeta))\right) d\zeta ds. \tag{15}$$

Before developing the analytic estimators that will be used to explain the paradox, it is helpful to gain some insight into the physical analogy suggested by the potential  $S(x)$ . The deterministic process  $\frac{d}{dt} x = m(x)$  always

moves “downhill” on the quasi-potential surface. Therefore, a valley, or basin of attraction, associated with a local minimum of the potential corresponds to a region inside which the deterministic system is guaranteed to approach a stable fixed point at the minimum (Nolting and Abbott 2016). As a result, probability tends to accumulate at the bottom of wells in the quasi-potential. If the distribution  $\pi(x, t)$  approaches a quasi-stationary distribution  $\pi_q(x)$  such that  $\pi(x > \epsilon, t) \approx \pi_q(x) \exp(-t/\tau)$  for some large  $\tau$  then the potential is large where the quasi-stationary distribution is small, and small where the quasi-stationary distribution is large. In fact, it can be easily shown from the Fokker-Planck equation (6) that if  $\pi(x, t)$  approaches a quasi-stationary distribution  $\pi_q(x)$ , then:

$$S(x) \propto -\lim_{\sigma \rightarrow 0} T \log(\pi_q(x)). \tag{16}$$

Given a notion of energy,  $S(x)$ , it is handy to have a notion of work that is the amount of energy the system needs to walk any trajectory  $X(t)$ . A natural notion of work along a path  $X(t)$ ,  $t \in [0, t_{\text{final}}]$  with respect to a potential function  $S(x)$  is:

$$W(X|S) = \max_{t < s \in [0, t_{\text{final}}]} \{S(X(s)) - S(X(t))\}. \tag{17}$$

Thus, the work needed for a population trajectory to follow a particular path is given by the maximum difference in the potential achieved along the path. Trajectories that manage to escape a basin of attraction (move against  $f(x)$ ) are generally unlikely, with probability proportional to the minimum work needed to escape the basin. As a result, deeper basins are generally more stable (Nolting and Abbott 2016).

We can use these observations about the potential to form analytic approximations to the mean first passage time. In all four models,  $K$  is a stable attractor of the deterministic dynamics, so corresponds to a well in the potential. Therefore, in order for  $X(t)$  to reach extinction starting from  $K$  the process must first escape the basin surrounding  $K$ . This requires climbing the potential, either all the way to  $\epsilon$  or, in the case of the Allee models, until reaching the saddle at  $x = A$ . In either case, if the well is deep enough, this occurrence is extremely unlikely and the mean time to extinction is dominated by the time it takes to escape this well. In general, the time it takes to escape the well is much longer than the time scale of the fluctuations inside the well, and the time scale of the fluctuation that carries  $X(t)$  to extinction, so the corresponding time to extinction follows an approximately exponential distribution.

To turn these observations into an estimator, notice that the exact expression for the mean first passage time (15) is

<sup>2</sup>The potential  $S(x)$  is equivalent to the Friedlin-Wentzell quasi-potential (Nolting and Abbott 2016). It appears in exact solutions to the backward equation (7), so it is important in computing first passage time statistics (Tier and Hanson 1981).

a double integral over an exponential of differences in the potential. It follows that the dominant contribution to the MFPT comes from  $s, \zeta$  where the difference in potential,  $S(s) - S(\zeta)$ , is largest. Since  $S(\zeta)$  is minimized at  $\zeta = K$  this difference is maximized when  $s \in [0, K]$  maximizes  $S(s) - S(K)$ . This difference is precisely the work it takes to escape the well. Therefore, the mean first passage time scales exponentially in the work it takes to escape the well:

$$\tau_\epsilon(K) \approx \exp\left(\frac{1}{T} \max_{\epsilon < s < K} \{S(s) - S(K)\}\right) = \exp\left(\frac{1}{T} W(K \rightarrow 0)\right). \tag{18}$$

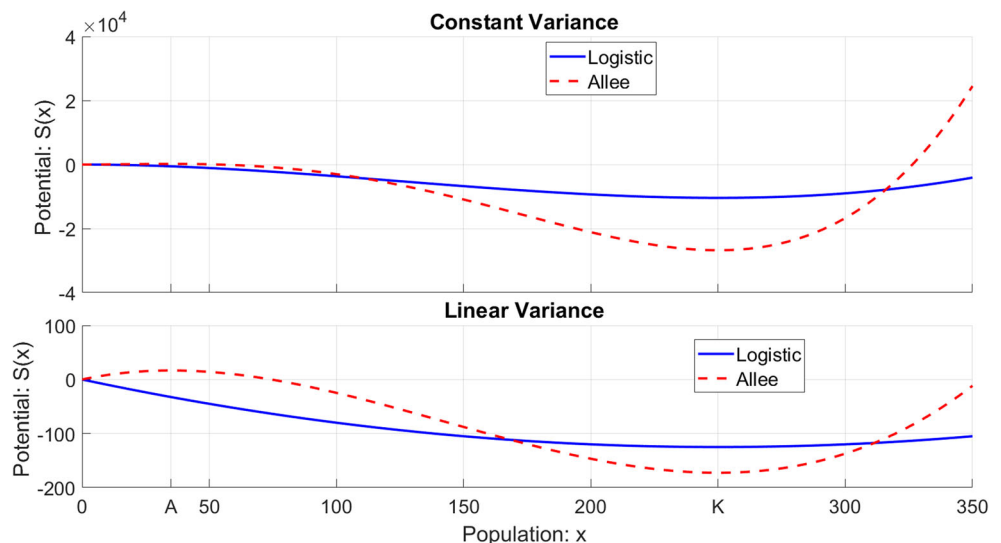
Here,  $W(K \rightarrow 0)$  denotes the work to go from carrying capacity to extinction (the depth of the potential well at carrying capacity).

The difference in potential is maximized by  $s$  in the vicinity of either  $\epsilon$  or  $A$  (depending on the model) and by  $\zeta$  in the neighborhood of  $K$ . Following Bresloff (2014) we can fine tune our estimator by Taylor expanding the potential near 0,  $A$ , and  $K$ . This introduces a prefactor to Eq. 18. Details are provided in Appendix A.1. In general, these estimators can be written in the form:

$$\tau_\epsilon(K) \approx C(K) \exp\left(\frac{1}{T} W(K \rightarrow 0)\right) \tag{19}$$

where  $C(K)$  is a prefactor that is generally much smaller than the exponential term. For all the models we consider, the work to escape carrying capacity will be a polynomial in  $K$ , whose order depends on the choice of  $m(x)$  and  $v(x)$ , while the prefactor  $C(K)$  will be rational in  $K$  and is proportional to  $1/r$ . Since the exponential term dominates for large  $K$  we can study the asymptotic behavior of  $\tau_\epsilon(K)$  purely by considering the work to escape carrying capacity. This greatly simplifies the overall analysis. For completeness, the prefactors are computed in Appendix A.2.

**Fig. 3** Potentials for all four models given  $A = 35$  and  $K = 250$ . In both panels the potential well for the Allee models is deeper at  $x = K$  than the potential well for the corresponding logistic models. This is the source of the paradox; *even though the Allee models introduce a saddle at  $x = A$  they also deepen the potential well at  $x = K$*



Most first passage times to extinction reported in the literature have the form:

$$\tau_\epsilon(K) \approx C(K) \exp\left(\frac{a}{T} K\right) \tag{20}$$

for some constant  $a$  (Foley 1994; Lande 1993; Leigh 1981; Ovaskainen and Meerson 2010; Kamenev et al. 2008). In Foley (1994) and Lande (1993), this exponential scaling is due to the fact that  $m(x)$  is set to be linear in  $x$  (density independent growth) up until  $x = K$ . In Kamenev et al. (2008), this exponential scaling arises naturally from the order of  $m(x)$  relative to  $v(x)$  in a model similar to 1b.

For general  $m(x)$  and  $v(x)$ , there is no reason why the linear exponential scaling law in Eq. 20 must hold. In fact, the work to escape carrying capacity can be quadratic, cubic, or even quartic in  $K$ . This leads to super-exponential scaling of the mean time to extinction. In Section “Theory,” we will compute potentials for the four models and show that it is this super-exponential scaling that allows for the paradox when the cubic Allee model is used. In Section “Analysis,” we will show that super-exponential scaling of this type is unrealistic for SDEs derived from birth-death processes, so is a canary in the coal mine for these models.

### Analysis

Consider the four models introduced in Eq. 5. The associated forces are:

- 1a.  $f(x) = \frac{1}{K} x(K - x)$
- 1b.  $f(x) = \frac{1}{K} (K - x)$
- 2a.  $f(x) = \frac{1}{AK} x(x - A)(K - x)$
- 2b.  $f(x) = \frac{1}{AK} (x - A)(K - x)$ .

The associated potentials are given by integrating the forces (see Eq. 11):

$$\begin{aligned}
 &1a. S(x) = -\frac{1}{K}x^2 \left( \frac{1}{2}K - \frac{1}{3}x \right) \\
 &1b. S(x) = -\frac{1}{K}x \left( K - \frac{1}{2}x \right) \\
 &2a. S(x) = -\frac{1}{AK}x^2 \left( -\frac{1}{4}x^2 + \frac{1}{3}(A + K)x - \frac{1}{2}AK \right) \\
 &2b. S(x) = -\frac{1}{AK}x \left( -\frac{1}{3}x^2 + \frac{1}{2}(A + K)x - AK \right).
 \end{aligned} \tag{22}$$

Examples of these potential functions are shown in Fig. 3. Notice that the Allee model (red dashed line) has a deeper potential well at carrying capacity for both constant and linear variance.

Evaluating the potentials at extinction,  $K$ , and  $A$  respectively:

$$\begin{aligned}
 &1a. S(0) = 0, \quad S(K) = -\frac{1}{6}K^2 \\
 &1b. S(0) = 0, \quad S(K) = -\frac{1}{2}K \\
 &2a. S(A) = \frac{1}{6AK}A^3 \left( K - \frac{1}{2}A \right), \quad S(K) = \frac{1}{6AK}K^3 \left( A - \frac{1}{2}K \right) \\
 &2b. S(A) = \frac{1}{2AK}A^2 \left( K - \frac{1}{3}A \right), \quad S(K) = \frac{1}{2AK}K^2 \left( A - \frac{1}{3}K \right).
 \end{aligned} \tag{23}$$

Therefore, the work needed to move from  $K$  to extinction is:

$$\begin{aligned}
 &1a. W(K \rightarrow 0) = \frac{1}{6}K^2 \\
 &1b. W(K \rightarrow 0) = \frac{1}{2}K \\
 &2a. W(K \rightarrow 0) = \frac{1}{6AK} \left[ A^3 \left( K - \frac{1}{2}A \right) - K^3 \left( A - \frac{1}{2}K \right) \right] \\
 &2b. W(K \rightarrow 0) = \frac{1}{2AK} \left[ A^2 \left( K - \frac{1}{3}A \right) - K^2 \left( A - \frac{1}{3}K \right) \right].
 \end{aligned} \tag{24}$$

This gives extinction time estimates:

$$\begin{aligned}
 &1a. \tau(K) \approx C(K) \exp\left(\frac{r}{3\sigma^2}K^2\right) \\
 &1b. \tau(K) \approx C(K) \exp\left(\frac{r}{\sigma^2}K\right) \\
 &2a. \tau(K) \approx C(A, K) \exp\left(\frac{r}{3\sigma^2} \left[ A^2 \left( 1 - \frac{1}{2} \frac{A}{K} \right) + K^2 \left( \frac{1}{2} \frac{K}{A} - 1 \right) \right] \right) \\
 &2b. \tau(K) \approx C(A, K) \exp\left(\frac{r}{\sigma^2} \left[ A \left( 1 - \frac{1}{3} \frac{A}{K} \right) + K \left( \frac{1}{3} \frac{K}{A} - 1 \right) \right] \right).
 \end{aligned} \tag{25}$$

The exact value of the  $C(K)$  varies depending on the model, in particular, with the curvature of the potential at  $x = K$  (see Appendix A.2).

There are clear differences in the scaling of each model’s potential in  $K$ . The logistic model with constant variance scales quadratically in  $K$ , while the Allee model with constant variance scales cubically in  $K$ . The logistic model with linear variance scales linearly in  $K$ , while the Allee model with linear variance scales quadratically in  $K$ . The depth of the potential well in the Allee model always scales by a polynomial of higher order than the equivalent logistic

model. This is a natural and inevitable consequence of modeling the Allee effect by multiplying the logistic  $m(x)$  by a factor of  $(x - A)/A$ . It follows that for sufficiently large  $K - A$  (that is, a sufficiently small Allee threshold relative to carrying capacity), the Allee model is not just more stable than the equivalent logistic model, it is much more stable, with an exponentially longer mean time to extinction.

This is the paradox. The presence of an Allee threshold makes these models substantially *less* extinction prone and the Allee and logistic models become *less* similar in their behavior as the Allee threshold approaches zero. This directly contradicts the intended biology.

While the models introduced so far are simplistic, they are not special cases. Uncontrolled scalings of this type are difficult to avoid if the variance is chosen independently of the deterministic model, or if  $m(x)$  is defined only by the location of its roots in mind.

For example, suppose we wanted to model a process with multiple stable equilibria  $x = [a_1, a_2, \dots, a_n]$  separated by saddles at  $x = [b_1, b_2, \dots, b_{n-1}]$ . The easiest way to construct an ODE of this type would be to set:

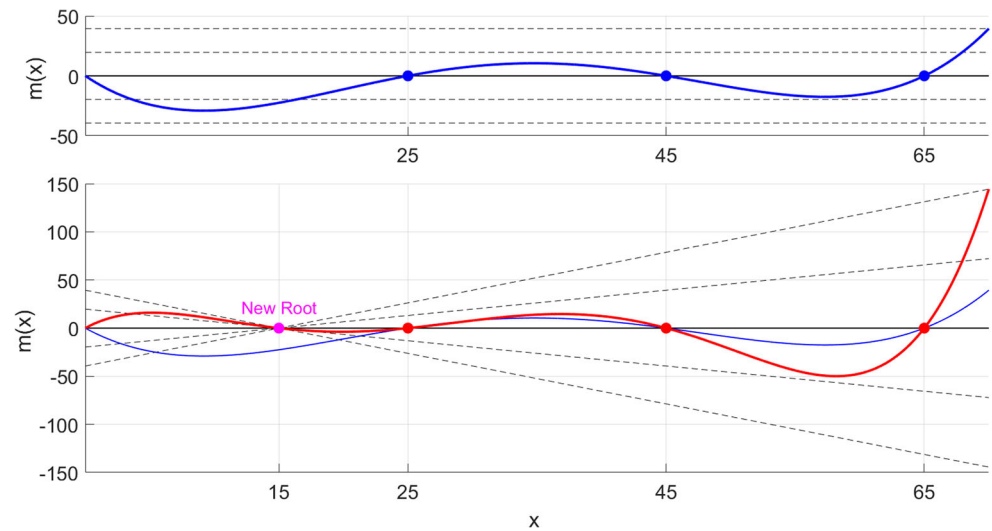
$$m(x) = r \prod_{j=1}^n (a_j - x) \prod_{j=1}^{n-1} (x - b_j). \tag{26}$$

The Allee model, Eq. 3, is defined in exactly this way, with  $n = 2$ ,  $a_1 = 0$ ,  $a_2 = K$ , and  $b_1 = A$ .

In some cases, this is sufficient for phenomenological models of deterministic processes since it guarantees the process will have basins of attraction  $(-\infty, b_1)$ ,  $(b_1, b_2)$ , ...,  $(b_{n-1}, \infty)$  each associated with a stable equilibrium  $a_j$ . This fixes the distance we can perturb  $x$  away from a given equilibrium  $a_j$  while ensuring the process will return to  $a_j$ . Note that the introduction of a new saddle or stable equilibrium only changes the boundaries of the nearest neighboring basins of attraction, so the introduction of new saddles or nodes only changes the resilience of the closest equilibria. In this sense, introducing each root to  $m(x)$  only affects the qualitative dynamics in a local way.

In a stochastic setting, the location of the stable equilibria and saddles remains important, but it is the depth of each potential well that is essential. The polynomial model (26) does not leave any freedom to vary the relative depths of the individual wells, since these are entirely specified by the locations of the wells. The introduction of a saddle or node at any point  $x$  changes the depth of every other well. Even worse, the farther removed the newest saddle or stable equilibrium is from any given well, the more dramatically it changes the depth of that well. This follows naturally from the fact that  $|x - a_i|$  is monotonically increasing the farther

**Fig. 4** Sample polynomial  $m(x)$  before and after the introduction of a new root (extremum in the potential). Notice that the introduction of the new root does not change the roots of  $m(x)$ , but does stretch  $m(x)$  so that its value away from the new root is generally much larger than its value before the root was introduced. The first panel shows the original  $m(x)$  in blue. The second compares the original (thin blue line) to the new  $m(x)$  thick red line. The new root is introduced at  $x = 15$



$x$  is from  $a_i$ . As a result, the introduction of a root in  $m(x)$  affects the global stochastic dynamics of the entire process, and has a larger effect the farther we move away from the root (Fig. 4). This will necessarily lead to unintended side-effects if the root was introduced with the intent of modeling dynamics only in its vicinity, not over the entire domain.

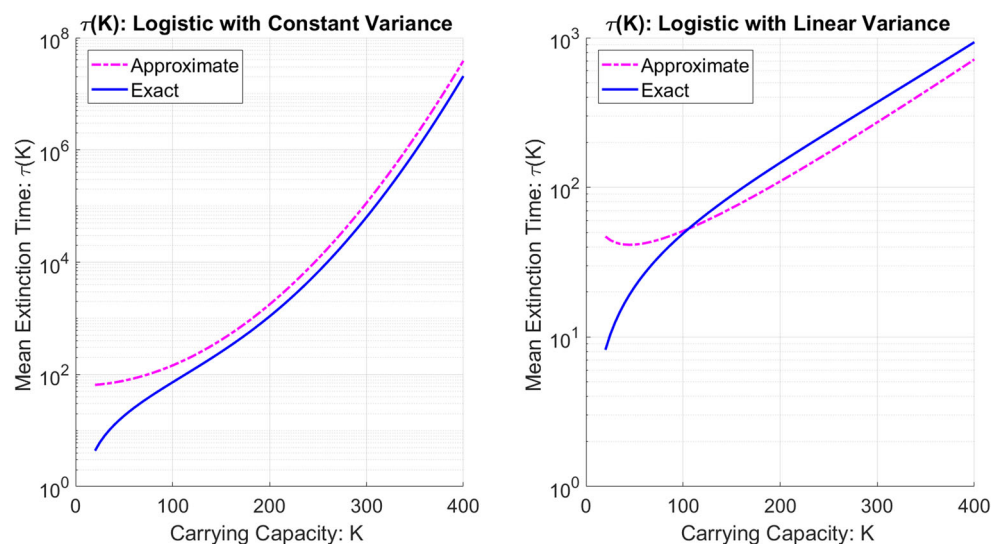
This is the heart of the paradox. Introducing the Allee threshold at  $A$  by multiplying  $m(x)$  with a monomial increases the polynomial order of  $m(x)$  thereby deepening the potential well at  $x = K$ . This makes the carrying capacity exponentially more stable.

To test our analytic approximations (25) numerically, we computed the exact solution to the backward equation (7) for  $\epsilon = 1$ . The exact solution is expressed as a double integral. These integrals were computed numerically in MatLab using built-in numerical integrators. This was

also the technique used to solve for the analytic mean first passage times shown in the Figs. 1–2. As expected, our analytic expressions capture the essential scaling of the mean first passage time to extinction as  $K$  becomes large (see Fig. 5).

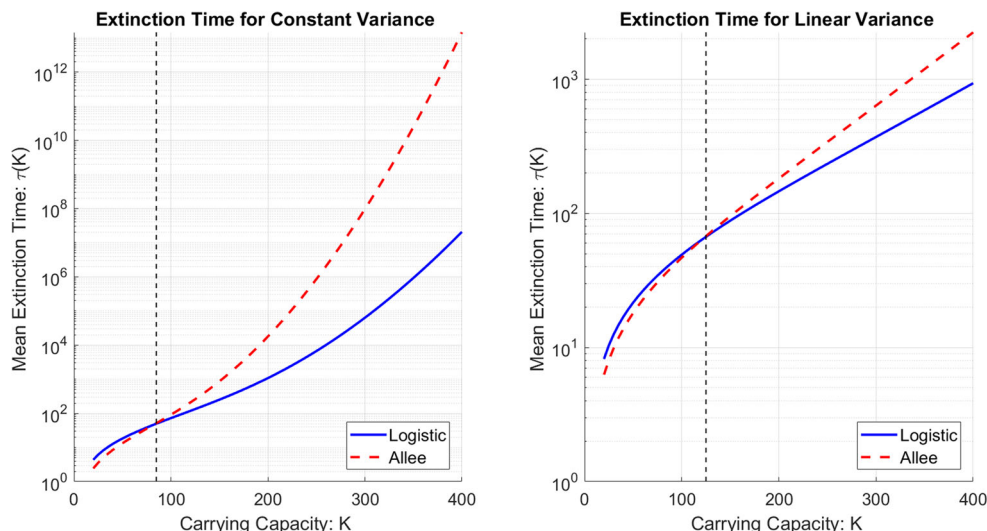
To start, we considered the constant variance case for  $A = K/5$ ,  $r = 0.1$ , and  $\sigma = 20$ . The noise variance was intentionally set large so that the  $\tau_\epsilon(K)$  did not exceed floating point accuracy for large  $K$ . The mean extinction time as a function of  $K \in [10, 400]$  is shown for model 1a and 1b along with the corresponding analytic estimators in Fig. 5. The exact mean extinction time is compared in Fig. 6. For  $K > 90$ , the Allee model is more stable than the logistic model, and the difference in extinction time between the two models becomes more exaggerated for larger  $K$ . This reflects the difference in scaling between the two models.

**Fig. 5** Exact mean time to extinction  $\tau_\epsilon(K)$  (MTE) and analytic approximations for the logistic model. Left panel: constant variance and  $r = 0.1$  and  $\sigma = 20$ . Right panel: linear variance with  $r = 0.1$  and  $\sigma = 3$





**Fig. 6** Exact mean time to extinction for the logistic and Allee models. Left Panel: constant variance,  $A = K/6$ ,  $r = 0.1$ , and  $\sigma = 20$ . Notice that the Allee model has exponentially longer extinction times for  $K > 90$  (marked with a dashed black line). Right Panel: linear variance,  $A = K/6$ ,  $r = 0.1$ , and  $\sigma = 3$ . Notice that the Allee model has exponentially longer extinction times for  $K > 125$  (marked with a dashed black line)

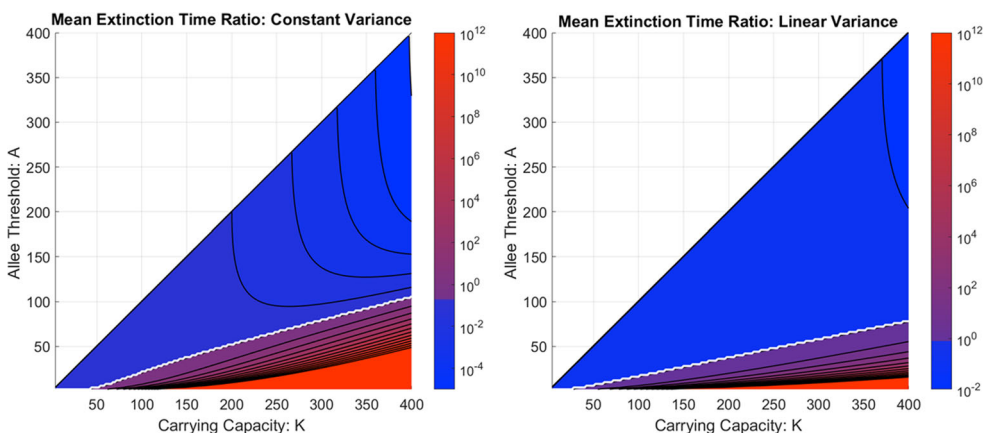


The same analysis was repeated for the linear variance case, for  $A = K/10$ ,  $r = 0.1$ , and  $\sigma = 3$ . Results are shown in Fig. 6. Here, the noise variance does not need to be as large as in the constant variance case to prevent the MTE from exceeding floating point accuracy. Once again, the Allee model is exponentially more stable than the logistic model for large  $K$ , with larger MTE for  $K > 150$ .

Define the ratio  $R(A, K)$  of the mean extinction time of the Allee model to the mean extinction time of the logistic model, given Allee threshold  $A$  and carrying capacity  $K$ . If  $R(A, K) < 1$ , then the logistic model is less extinction prone, and if  $R(A, K) > 1$ , the Allee model is less extinction prone. The ratio  $R(A, K)$  was computed for all  $A \leq K$  and  $K \in [0, 400]$ . A contour plot of  $R(A, K)$  is provided in Fig. 7 for both constant and linear variance.

The blue region above the white line corresponds to  $A, K$  such that the logistic model is more stable (less extinction prone) and the purple to red region beneath the white line corresponds to  $A, K$  such that the Allee model is more stable. By comparing the depths of the potential wells, we find that, for constant variance and large  $K$ , the Allee model is exponentially more stable if  $A < aK$  where  $a \approx 0.2574$  satisfies:  $a^4 - 2a^3 + 4a - 1 = 0$ .

Notice that, regardless of the variance, for any choice of  $A$ , there exists a sufficiently large  $K$  such that the Allee model is orders of magnitude more stable than the logistic model. Also notice that once  $K$  is sufficiently large the extinction time for the Allee model increases much faster than the extinction time for the corresponding logistic model.



**Fig. 7** Contours for the ratio of the exact mean time to extinction (MTE) for Allee and logistic models for  $A \leq K \in [0, 400]$ ,  $r = 0.1$ . The left panel corresponds to constant variance with  $\sigma = 20$ . The right panel corresponds to linear variance with  $\sigma = 3$ . In both the white line represents all  $A, K$  such that the two models have the same first passage time. Each black line above the white line represents a factor

of ten increase in the MTE of the logistic model relative to the Allee model, and each black line beneath the white line represents a factor of ten increase in the MTE of the Allee model relative to the logistic model. The purple-red region beneath the white line is the region where the paradox is observed. The ratio is only calculated up to  $10^{12}$  to avoid numerical errors. This introduces the solid red plateau

This completes the story when  $r_A = r_L = r$ . Multiplying the logistic model by the monomial  $(x - A)/A$  increases the depth of the potential well, thereby increasing the passage time to extinction. This effect can be predicted from the deterministic models simply by considering the per capita growth rate away from  $A$ . In the cubic model, the introduction of the monomial increases the magnitude of the per capita growth rate for  $x - A > A$ . Therefore, if  $K > 2A$  the per capita growth rate around  $x = K$  is increased when moving from the logistic model to the Allee model. In a deterministic setting, this overcompensation effect is usually resolved by rescaling  $r_A$  so that  $r_A < r_L$ . In the next section, we discuss the effects of rescaling on mean time to extinction and show that rescaling does not offer a satisfactory resolution to the paradox in a stochastic setting.

### Does rescaling resolve the paradox?

Here, we generalize the analysis to  $r_A < r_L$ . Our goal is to determine whether or not the paradox can be resolved based on the standard choices of rescaling used to avoid excess overcompensation in deterministic Allee models.

In principle, there are three choices of rescaling that can be motivated by deterministic reasoning. First, if  $r_A$  is interpreted as the maximum per capita growth rate, then  $r_A$  should be chosen so that the Allee model has the same maximum per capita growth rate as the logistic model (Gruntfest et al. 1997; Lewis and Kareiva 1993). This requires setting  $r_A = \frac{4AK}{(K-A)^2}r_L$  (see Appendix A.3 for details). Plugging into Eq. 24 and considering the limit of large  $K - A$ , we find that the work to escape  $K$  is twice as large in the Allee model than in the logistic model when  $v(x)$  is constant, and is 4/3rds times as large in the Allee model when  $v(x)$  is linear. The prefactors also depend on the rescaling; however, for large  $K - A$ , the exponential term still dominates. Therefore, since the work to escape the carrying capacity in the Allee model is greater than the work to escape the carrying capacity in the logistic model, the Allee model is still more persistent than the logistic model. Moreover, since the depth of both wells scales quadratically in  $K$ , the difference in the extinction times still grows super-exponentially in  $K$ . Therefore, rescaling in this way to account for overcompensation in the deterministic model *does not* resolve the paradox.

Given that this rescaling failed to resolve the paradox, it is natural to consider stronger rescalings. Instead of choosing  $r_A$  so that the Allee model has the same maximum per capita growth rate as the logistic model,  $r_A$  can be chosen so that both have the same maximum growth rate. For large  $K - A$ , this scaling is approximated by  $r_A = \frac{16}{27} \frac{(K+\frac{1}{2}A)}{K} \frac{(K-A)^2}{AK} r_L$ . Then, for large  $K$  and constant noise,

the work to escape the carrying capacity in the Allee model is approximately 7/8 times the work to escape the carrying capacity in the logistic model. For linear noise, the work to escape in the Allee model is approximately 9/16 times the work to escape in the logistic potential. Since the work to escape the Allee model is smaller than the work to escape the logistic model, the Allee model is less stable and has shorter mean time to extinction. However, the work to escape the carrying capacity in the Allee model now converges to a fixed proportion of the work to escape the carrying capacity in the logistic model for large  $K$ . This is biologically unreasonable since it implies that for arbitrarily small  $A$ , and arbitrarily large  $K$ , the introduction of an Allee threshold reduces the extinction time by a fixed power. This means that the purportedly local Allee effect is globally destabilizing. A truly local Allee effect would only destabilize the population when  $A$  is reasonably near  $K$ , and for large  $K - A$ , both models should have close to the same persistence. This is the behavior observed in the mechanistic Allee models examined in the next section and Appendix A.6. Therefore, although this rescaling resolves the paradox, it still does not produce a stochastic model that behaves reasonably.

Finally, instead of requiring that the two models share the same maximum growth rates, we could require that they share the same local stability at  $K$ . Then, in the deterministic sense, the two models would be equally stable. This requires setting  $r_A = \frac{A}{K-A}r_L$ . For large  $K - A$  and constant noise, the work to escape carrying capacity in the Allee model converges to one half the work required in the logistic model. For linear noise, the work to escape in the Allee model converges to one-third the work required for the logistic model. Therefore, for large  $K$ , the Allee model is less persistent than the logistic model so the paradox is resolved. However, as in the previous case, the persistence of the Allee model converges to a fixed power of the persistence of the logistic model when  $K - A$  is large. Therefore, this rescaling resolves the original paradox, but also does not produce a model that behaves reasonably.

In all three cases, rescaling reduced the order of the work to escape the Allee potential so that it was comparable to a fixed factor times the work to escape the logistic potential. This resolves the most egregious scaling error apparent in the direct comparison but *does not* necessarily resolve the paradox, and still fails to restrict the impact of the Allee effect to small populations. Moreover, without stochastic models, it is likely impossible to predict a priori whether or not a rescaling will resolve the paradox, since the answer may well depend on the choice of variance. This is because the rescalings were based on deterministic criteria that fixed some feature of the models to be the same at some specific point (or pair of points). The persistence of the stochastic models depends globally on the ratio  $m(x)/v(x)$

so introducing a global rescaling based on a deterministic local criteria cannot be expected to correctly treat the persistence of the stochastic models. This is apparent in the last rescaling, where the two models have the same local stability near  $K$ , but completely different persistences.

Finally, if the noise is derived implicitly from a pair of birth and death rates, as will be illustrated in the following section, then a rescaling of  $m(x)$  is typically due to a rescaling of the underlying birth and death rates that give rise to  $r$ . In that case, the noise intensity is the sum of these underlying rates (Bresloff 2014; Samuel and Taylor Howard 1981), so  $\sigma^2$  should scale with  $r$  (see Section “How to avoid the paradox”). In particular, if the birth and death rates are scaled by some constant  $\alpha$ , then both  $r$  and  $\sigma^2$  are scaled by  $\alpha$ . Since the potential depends on the ratio of the deterministic part to the noise, this rescaling does not change the potential. If we revisit any of the above scalings and properly rescale both  $r$  and  $\sigma^2$ , then for all three scalings, the work to escape carrying capacity in the Allee model would remain an order larger in  $x$  than the corresponding work for the logistic model. It follows that scaling would only resolve the paradox if it changed  $C(A, K)$  enough to counter the exponential of the difference in work to escape the two wells. This is only feasible for small  $K - A$ , since for large  $K - A$ , it would require an unreasonably large, noise-dependent, rescaling that could not be justified on mechanistic grounds.

In conclusion, rescaling the deterministic part alone may account for overcompensation in the deterministic model, but cannot be regarded as a satisfactory resolution of the paradox.

### How to avoid the paradox

There are two simple ways to avoid the extinction time paradox.

The first is to model  $m(x)$  with global stability in mind. If one constructs  $m(x)$  in a way that fixes both the locations and the depths of the potential wells, then the global stability of each well is controlled a priori. This would require using a more complicated  $m(x)$  than the polynomial (26). It would be natural to use saturating functions instead of monomials when introducing new extrema, which would only affect  $m(x)$  locally, thereby avoiding accidental global changes.

Most mechanistic Allee models make this type of local modification by introducing a term that only effects the growth rate at low populations. For example, if the Allee effect arises from difficulty-finding mates, then it is natural to introduce a function  $P(x)$  which gives the probability of a female finding a mate (per unit time) given population size  $x$ . It is usually assumed that  $P(0) = 0$ ,  $P(\infty) = 1$  and  $P(x)$  is monotonically increasing (Boukal and Berec 2002). The specific choice of  $P(x)$  depends on the model for mate finding; however, common choices are  $P(x) =$

$1 - \exp(-ax)$ ,  $P(x) = 1 - (1 - z)^x$ , and  $P(x) = \frac{x}{\theta+x}$  for some  $a, z, \theta > 0$  (Boukal and Berec 2002; Dennis 1989; 2002; Stephan and Wissel 1994). If each mated female has fecundity  $\beta$ , then we can modify the logistic  $m(x)$  so that:

$$m(x) = \beta P(x)x - \delta_0 x - \delta_1 x^2 = \delta_1 \left( \frac{\beta - \delta_0}{\delta_1} - x \right) x - \beta(1 - P(x))x. \tag{27}$$

This resolves the extinction time paradox by restricting the modification to the logistic model to small populations. That said, it does not constrain the scaling of mean time to extinction in any way, so allows for mean extinction times that increase super-exponentially in  $K$ . We will show that this sort of super-exponential scaling is unrealistic for an SDE that is intended to approximate a birth-death process. Therefore, while this modification of  $m(x)$  resolves the paradox, it does not resolve the underlying problem with our models. The potentials corresponding to specific  $P(x)$  are explored in the Appendix A.6.

The second, and more complete, solution is to choose  $v(x)$  in a way that reflects the mechanisms that give rise to the stochastic process. Since we are modeling populations subject to demographic stochasticity, it is natural to model birth and death as independent random events. Under minimal assumptions (Patrick 1995) these events occur at exponentially distributed intervals with density dependent rates. Then, the population,  $X(t)$ , is an integer-valued continuous-time stochastic process. Assume that the birth and death rates,  $\lambda$  and  $\mu$ , are non-negative continuously differentiable functions of population size. Let  $p_i(t)$  be the probability that the population has  $i$  individuals at time  $t$ . Then,  $p_i(t)$  obeys the master equation:

$$\frac{d}{dt} p_i(t) = \lambda(i - 1)p_{i-1}(t) - (\lambda(i) + \mu(i))p_i(t) + \mu(i + 1)p_{i+1}(t). \tag{28}$$

The first term accounts for birth events from  $x = i - 1$ , the second for birth and death events at  $x = i$ , and the third for death from  $x = i + 1$ .

These discrete population models can be difficult to analyze, so it is natural to make a diffusion approximation that approximates their behavior with an SDE. Methods for developing diffusion approximations are outlined in Allen and Allen (2002), Bresloff (2014), and Higham (2008). Suppose we refine our integer-valued population model to include fractional populations, spaced evenly in intervals  $\Delta x$ . When refining the discrete model, modify the birth and death rates so that  $\lambda(x) - \mu(x)$  is proportional to  $\Delta x$  but  $\lambda(x) + \mu(x)$  is unchanged.<sup>3</sup> Then, Taylor expanding  $p$  in small  $\Delta x$ ,  $\frac{1}{\Delta x} p_i(t)$  converges to  $\pi(x, t)$  where  $\pi(x, t)$

<sup>3</sup>This is the standard assumption made when taking the hydrodynamic limit of a random walk. For details see Appendix A.4.

obeys a Fokker-Planck equation with infinitesimal mean  $m(x)$  and variance  $v(x)$  given by:

$$\begin{aligned} m(x) &= \lambda(x) - \mu(x) \\ v(x) &= \lambda(x) + \mu(x). \end{aligned} \quad (29)$$

For details on this approximation, see Appendix A.4. The observation that the difference in birth and death rates is associated with  $m(x)$  while the sum of the two rates is associated with  $v(x)$  is a familiar feature of birth-death models (Bresloff 2014; Doering et al. 2005; Samuel and Taylor Howard 1981; Nisbet and Gurney 1982; Turelli 1977; Wilkinson 2012).

If one starts with a deterministic model and jumps directly to an SDE,  $v(x)$  is not specified by any underlying noise model. As a result,  $v(x)$  is often chosen for analytic convenience (Allen et al. 2008), or based on familiar examples in the literature. It is commonly assumed that processes subject to demographic noise have instantaneous variance  $v(x) = \sigma^2 x$ , and processes driven by environmental noise have instantaneous variance  $v(x) = \sigma^2 x^2$  (Dennis 2002; Desharnais Robert et al. 2006; Tier and Hanson 1981); however, these expressions depend on particular modeling assumptions and by no means apply to all birth-death processes.<sup>4</sup> In contrast, by deriving an SDE from a diffusion limit of a specified birth-death process the infinitesimal variance  $v(x)$  is determined implicitly by the choice of birth and death rates. When  $v(x)$  is derived implicitly from birth and death rates, the corresponding  $v(x) = \lambda(x) + \mu(x)$  is just as nonlinear as  $m(x) = \lambda(x) - \mu(x)$ , so would rarely be chosen if pulled out of a hat. In particular, there is no reason why a birth-death process needs to converge to an SDE with linear  $v(x)$  if  $\lambda(x)$  and  $\mu(x)$  are nonlinear (for examples see Cresson and Sonner 2018). This result generalizes to other mechanistic models of demographic stochasticity, such as discrete time models for populations with annual birth and death pulses (Desharnais Robert et al. 2006).<sup>5</sup>

<sup>4</sup>Note that  $v(x)$  will typically depend on both  $x$  and the model parameters. Therefore, when considering scaling in system size it is more clear to write  $v(x, K)$  where  $v(x, K)$  is the instantaneous variance at  $x$  given a carrying capacity  $K$ . It is generally true that  $v(x, K)$  scales linearly in  $K$  both for physical systems and for populations subject to demographic stochasticity (Bresloff 2014; Desharnais Robert et al. 2006); however, this does not require that  $v(x, K)$  is also linear in  $x$ .

<sup>5</sup>For example, consider a discrete time model representing a population with annual birth and death stages. Suppose that the expected change in population after one year is  $\mathbb{E}[X_{t+1} - X_t] = m(X_t)$ . Suppose that, as in Desharnais Robert et al. (2006), the actual change is Poisson distributed so that  $X_{t+1} = X_t + Z_t$  where  $Z_t$  is a Poisson distributed random variable with mean  $m(X_t)$ . Then, the variance in  $X_{t+1} - X_t$  is  $m(X_t)$ . So, if  $m(x)$  is nonlinear then the variance  $v(x)$  is also nonlinear, and, as in the birth-death process,  $v(x)$  scales with  $m(x)$ .

Suppose that we pick  $\lambda(x), \mu(x)$  to give a particular  $m(x)$ . Then, the noise will often inherit much of the behavior of  $m(x)$ . In particular, if  $m(x)$  is large, then the noise will be larger (after scaling by  $T$ ) since the sum of two positive numbers is always greater than their difference. As a result, the forces  $f$  are bounded by:

$$|f(x)| = \left| \frac{\lambda(x) - \mu(x)}{\lambda(x) + \mu(x)} \right| \leq 1 \quad (30)$$

with equality if and only if either  $\lambda(x)$  or  $\mu(x)$  is zero.

Constructed this way, the forces are saturating in both  $\lambda(x)$  and  $\mu(x)$ . If either birth or death dominates, then the forces approach 1 or  $-1$ . Since the force  $f$  is a derivative of the potential, it follows that the potential is Lipschitz continuous with a constant one (has slope of magnitude  $< 1$  for all  $x$ ):

$$|S(x+h) - S(x)| \leq h. \quad (31)$$

This translates immediately into a limit on how the mean extinction time  $\tau(K)$  can scale with  $K - A$ . Equation 31 implies  $S(0) - S(K) \leq K$  and  $S(A) - S(K) \leq K - A < K$  so  $W(K \rightarrow 0) \leq K$ . This means that  $\tau(K)$  is, at most, exponential in  $K$  as suggested in Eq. 20:

$$\tau(K) \lesssim C(K) \exp\left(\frac{1}{T}K\right) \quad (32)$$

for some rational  $C(K)$ . This is not quite as restrictive as it seems since the temperature  $T$  depends on the limiting procedure used to approximate the birth-death model with an SDE. However, it is enough to suggest that the original four models (except perhaps the logistic model with linear variance and  $r$  fixed) have unrealistic noise assumptions. This underscores the importance of deriving an SDE from an explicit mechanistic model (in this case, specific birth and death rates) (Nisbet and Gurney 1982; Wilkinson 2012).

For an example of how modeling birth and death separately produces noise variance that bounds the associated first passage time to extinction, consider the birth and death rates (Cresson and Sonner 2018; Doering et al. 2005):

$$\begin{aligned} \lambda(x) &= \beta x \\ \mu(x) &= \delta_0 x + \delta_1 x^2. \end{aligned} \quad (33)$$

These lead to logistic  $m(x)$  and quadratic noise variance  $v(x) = (\beta + \delta_0)x + \delta_1 x^2$ .<sup>6</sup> Because we assume that all

<sup>6</sup>Note that there are infinitely many ways to decompose  $m(x) = \lambda(x) - \mu(x)$  into a particular  $\lambda(x)$  and  $\mu(x)$ , and that these models will not all share the same first passage time statistics (Allen and Allen 2002). For example in the logistic model derived from a simple S-I-S infection model the quadratic competition term is moved into birth rather than death (Doering et al. 2005). We follow Nisbet and Gurney (1982) and assume that any positive term in  $m(x)$  is associated with birth and any negative is associated with death. This differs from the implementation in Doering et al. (2005) and Kamenev et al. (2008) but ensures all the rates are positive for all  $x$ . An equivalent model is studied in both Tier and Hanson (1981) and Leigh (1981), albeit derived from different assumptions. A careful analysis of extinction

parameters have constant values, this model includes only demographic noise, not environmental noise. Importantly, had we used the usual assumption that demographic noise gives linear  $v(x)$  (as in model 1b) we would have been studying an SDE that *did not correspond* to the underlying birth-death process.

The potential associated with Eq. 33 is:

$$S(x) = x - \frac{2\delta_0}{\delta_1} \log \left( 1 + \frac{\delta_1}{\beta + \delta_0} x \right). \tag{34}$$

The work needed to escape the well is still linear in  $K = \frac{\beta - \delta_0}{\delta_1}$ :

$$S(0) - S(K) = \left[ (1 + T) \log \left( \frac{1 + T}{T} \right) - 1 \right] K \leq \frac{1}{T} K \tag{35}$$

with  $T = \frac{\sigma^2}{2r} = \frac{\beta + \delta_0}{\beta - \delta_0}$ . It follows that the mean first passage time to extinction scales exponentially, not super-exponentially, in  $K$ .

This example reveals that the usual assumption that  $v(x)$  is linear for demographic noise is not true for general birth-death processes, and had we assumed  $v(x)$  was linear, we would have vastly overestimated the persistence of the population. In fact, the only way that  $v(x)$  can be lower order than  $\lambda(x)$  and  $\mu(x)$  is if  $\lambda(x)$  and  $\mu(x)$  are chosen with a particular symmetry so that  $\lambda(x) + \mu(x)$  is lower order than either  $\lambda(x)$  or  $\mu(x)$  (Kamenev et al. 2008). This symmetry is hard to motivate biologically since it is hard to imagine a realistic scenario in which the birth and death processes conspire to cancel their highest order term, and it typically leads to either a negative birth rate or a negative death rate for large  $x$ .

Therefore, for most birth-death processes, it is likely that the scaling law (20), given by Ovaskainen and Meerson (2010), holds. This can serve as a check to make sure the assumed  $v(x)$  is reasonable and motivates modeling from a birth-death process before moving to an SDE.

### Conclusion

For naively chosen mean population change rate  $m(x)$  and instantaneous variance  $v(x)$ , the introduction of an Allee effect to a stochastic logistic model can exponentially increase the mean first passage time to extinction. This apparent paradox illustrates two major conceptual differences between deterministic and stochastic modeling. First, stability of a stochastic model depends on the global behavior of the model, so phenomenological models, and local corrections, may lead to unexpected dynamics. Second, it

is important to build stochastic models that are mechanistically plausible. Modeling the process from separate birth and death mechanisms resolves the paradox by introducing a natural limit on the steepness of the potential wells. This limit introduces a scaling law for first passage times that can be used to check whether or not a particular choice of  $v(x)$  could plausibly arise from a birth-death process, and suggests that heuristic choice of the noise may lead to extreme overconfidence in the persistence of fluctuating populations.

**Acknowledgements** We thank Samantha Catella, Fang Ji, Brian Lerch, Amy Patterson, Claire Plunkett, Robin Snyder, Alexa Wagner, and two anonymous reviewers for helpful feedback on the project and the manuscript.

**Funding information** This work was supported by NSF DEB-1654989 to KCA and PJT.

## Appendix

### A.1 Extinction time estimators

In Section “Theory,” we derived an exact expression for the first passage time from carrying capacity,  $K$ , to extinction, then introduced a series of estimators based on the exact solution. These estimators were formed by Taylor expanding the potential near populations 0,  $A$ , and  $K$ . The details are provided here.

First, consider the logistic models. Let  $\zeta = K + h$  for some small  $h < \delta_\zeta$  and  $\epsilon < s < \delta_s$  be small. Following the approach of Bresloff (2014) we can approximate the mean first passage time (15) by:

$$\tau_\epsilon(x) \approx \int_\epsilon^{\delta_s} \int_{K-\delta_\zeta}^{K+\delta_\zeta} \frac{2}{\sigma^2} \frac{1}{v(\zeta)} \exp \left( \frac{1}{T} (S(s) - S(\zeta)) \right) d\zeta ds. \tag{36}$$

Now, since the intervals do not overlap we can separate the two integrals:

$$\tau_\epsilon(x) \approx \int_\epsilon^{\delta_s} \exp \left( \frac{1}{T} S(s) \right) ds \times \dots \cdot \int_{\delta_\zeta}^{\delta_\zeta} \frac{2}{\sigma^2} \frac{1}{v(K+h)} \exp \left( \frac{1}{T} (-S(K+h)) \right) dh. \tag{37}$$

Since  $S(x)$  is minimized at  $x = K$  and is twice differentiable, it has a Taylor expansion of the form:

$$S(K+h) = S(K) + \frac{1}{2} S''(K) h^2 + \mathcal{O}(h^3). \tag{38}$$

Therefore, the second integral can be rewritten:

$$\exp \left( -\frac{1}{T} S(K) \right) \int_{-\delta_\zeta}^{\delta_\zeta} \frac{2}{\sigma^2} \frac{1}{v(K+h)} \exp \left( \frac{1}{T} \left( -\frac{1}{2T} S''(K) h^2 \right) \right) dh \tag{39}$$

times for birth-death processes of this type is available in Doering et al. (2005).

Provided  $S''(K)$  is large enough, this integrand decays rapidly for large  $h$ . As a result, we can approximate the integral by replacing the bounds with  $-\infty$  and  $\infty$  and replacing  $\frac{1}{v(K+h)}$  with  $\frac{1}{v(K)} - \frac{v'(K)}{v(K)}h$ :

$$\frac{2}{\sigma^2} \frac{\exp\left(-\frac{1}{T}S(K)\right)}{v(K)} \int_{-\infty}^{\infty} \left[1 - \frac{v'(K)}{v(K)}h\right] \exp\left(\frac{1}{T}\left(-\frac{1}{2T}S''(K)h^2\right)\right) dh. \tag{40}$$

Notice that the exponential term is Gaussian. Therefore, the integral can be rewritten:

$$\frac{2}{\sigma^2} \frac{\exp\left(-\frac{1}{T}S(K)\right)}{v(K)} \sqrt{\frac{2\pi T}{|S''(K)|}} \mathbb{E}\left[1 - \frac{v'(K)}{v(K)}h + \mathcal{O}(h^2)\right] \tag{41}$$

where the expected value is evaluated over a Gaussian distribution. Since the Gaussian distribution is symmetric, all odd order terms in the expansion vanish. Therefore, to second order accuracy in  $h$ :

$$\tau_\epsilon(x) \approx \frac{2}{\sigma^2} \sqrt{\frac{2\pi T}{|S''(K)|}} \frac{\exp\left(-\frac{1}{T}S(K)\right)}{v(K)} \int_\epsilon^{\delta_s} \exp\left(\frac{1}{T}S(s)\right) ds. \tag{42}$$

The remaining integral can be handled in much the same way. If the potential is approaching a local maximum at  $x = 0$  then replace  $S(s)$  with a second order Taylor expansion about  $s = 0$ . This gives:

$$\tau_\epsilon(x) \approx \frac{2}{\sigma^2} \frac{2\pi T}{\sqrt{|S''(0)||S''(K)|}} \frac{\exp\left(\frac{1}{T}(S(0) - S(K))\right)}{v(K)}. \tag{43}$$

If  $S(s)$  is not approaching a local maximum as  $s$  goes to zero then the derivative of  $S(s)$  does not vanish at  $s = 0$ , so the Taylor expansion is dominated by the first order term.

This approximates  $S(s)$  with an exponentially decaying function, and gives:

$$\tau_\epsilon(x) \approx \frac{2}{\sigma^2} \frac{T}{|S'(0)|} \sqrt{\frac{2\pi T}{|S''(K)|}} \frac{\exp\left(\frac{1}{T}(S(0) - S(K))\right)}{v(K)}. \tag{44}$$

The same technique can be applied for the Allee models, expanding about  $s = A$  instead of  $s = 0$ . Since  $S(x)$  is maximized at  $x = A$  for both Allee models we approximate  $S(s)$  near  $A$  with a second order Taylor expansion to get (Bresloff 2014; Nolting and Abbott 2016):

$$\tau_\epsilon(x) \approx \frac{2}{\sigma^2} \frac{2\pi T}{\sqrt{|S''(A)||S''(K)|}} \frac{\exp\left(\frac{1}{T}(S(0) - S(K))\right)}{v(K)}. \tag{45}$$

These three Eqs. 43, 44, 45 all fit the general form Eq. 19.

### A.2 Prefactors

In Section “Theory” we derived first passage time estimators of the general form Eq. 19. The qualitative behavior of the extinction times reflects the exponential scaling in this formula, independent of the prefactors  $C(K)$  or  $C(A, K)$ . Nevertheless, these factors are important for quantitative agreement between theory and simulation (see Fig. 5). Here we compute the prefactors for each of the four models. The prefactors generally depend on the first or second derivatives of the potential at 0,  $A$ , or  $K$ . For the four models, the prefactors are:

- 1a.  $C(K) = \frac{2}{\sigma^2} \frac{1}{v(K)} \frac{2\pi T}{\sqrt{|S''(0)||S''(K)|}}$
- 1b.  $C(K) = \frac{2}{\sigma^2} \frac{1}{v(K)} \frac{T}{|S'(0)|} \sqrt{\frac{2\pi T}{|S''(K)|}}$
- 2a.  $C(A, K) = \frac{2}{\sigma^2} \frac{1}{v(K)} \frac{2\pi T}{\sqrt{|S''(A)||S''(K)|}}$
- 2b.  $C(A, K) = \frac{2}{\sigma^2} \frac{1}{v(K)} \frac{2\pi T}{\sqrt{|S''(A)||S''(K)|}}$ .

The derivatives evaluated at 0,  $A$ , and  $K$  are:

- 1a.  $S''(x) = \frac{1}{K}(2x - K), \quad |S''(0)| = 1, \quad |S''(K)| = 1$
- 1b.  $S'(x) = \frac{1}{K}(x - K), \quad |S'(0)| = 1, \quad S''(x) = \frac{1}{K} = |S''(K)|$
- 2a.  $S''(x) = \frac{1}{AK} [(x - A)(K - x) + 2x - (K + A)], \quad |S''(A)| = \frac{1}{AK} |K - A| = |S''(K)|$
- 2b.  $S''(x) = \frac{1}{AK} [2x - (K + A)], \quad |S''(A)| = \frac{1}{AK} |K - A| = |S''(K)|$ .

Therefore, the prefactors are:

- 1a.  $C(K) = \frac{4\pi T}{\sigma^2}$
- 1b.  $C(K) = \frac{2T}{\sigma^2} \sqrt{\frac{2\pi T}{K}}$
- 2a.  $C(A, K) = \frac{4\pi T}{\sigma^2} \frac{AK}{K - A}$
- 2b.  $C(A, K) = \frac{4\pi T}{\sigma^2} \frac{A}{K - A}$ .

Or, substituting  $T = \frac{\sigma^2}{2r}$ :

- 1a.  $C(K) = \frac{2\pi}{r}$
- 1b.  $C(K) = \frac{1}{r} \sqrt{\frac{2\pi T}{K}}$
- 2a.  $C(A, K) = \frac{2\pi}{r} \frac{AK}{K - A}$
- 2b.  $C(A, K) = \frac{2\pi}{r} \frac{A}{K - A}$ .

Notice that the prefactors for both Allee models diverge if  $A$  approaches  $K$ . The prefactor diverges because models 2a and 2b have quartic and cubic potentials (respectively). As  $A$  approaches  $K$  the second derivative of the potential at  $A$  and  $K$  vanishes, so a second order Taylor expansion about  $S(A)$  and  $S(K)$  is not accurate. Therefore we restrict to  $A \lesssim K/2$ . More accurate approximations could be made by carrying the Taylor expansions out to higher order; however, since the potentials are polynomials this is equivalent to solving for the first passage times exactly.

### A.3 Rescaling

In Section “Analysis,” we discussed three rescalings of  $m(x)$ . This Appendix derives the necessary rescalings, and analyzes their effect on the persistence of the Allee model. In general, consider  $r_A = s(K, A)r_L$  where  $s(K, A)$  is some

scaling factor. We will consider  $s(K, A)$  such that the two models have the same:

1. maximum per capita growth rate (Gruntfest et al. 1997),
2. maximum absolute growth rate, or
3. linear stability at carrying capacity.

The per capita growth rate for the logistic model is  $r_L(K - x)/K$  and is maximized at  $x_* = 0$ . Therefore, the maximum per capita growth rate for the logistic model is  $r_L$ . The per capita growth rate for the Allee model is the quadratic  $r_A(x - A)(K - x)/(AK)$  and is maximized halfway between the Allee threshold and the carrying capacity at  $x_* = (K + A)/2$ . Therefore, the maximum per capita growth rate for the Allee model is  $r_L(K - A)^2/(4AK)$ . It follows that the Allee model has the same maximum per capita growth rate as the logistic model if  $s(A, K) = 4 \frac{AK}{(K - A)^2}$ .

Plugging into Eq. 25, the arguments of the exponential term in the extinction times (work to extinction divided by temperature) become:

$$\begin{aligned}
 \text{2a. } \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \frac{2r_L}{\sigma^2} \frac{4AK}{(K-A)^2} \frac{1}{6AK} \left[ A^3 \left( K - \frac{1}{2}A \right) - K^3 \left( A - \frac{1}{2}K \right) \right] \\
 \text{2b. } \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \frac{2r_L}{\sigma^2} \frac{4AK}{(K-A)^2} \frac{1}{2AK} \left[ A^2 \left( K - \frac{1}{3}A \right) - K^2 \left( A - \frac{1}{3}K \right) \right]
 \end{aligned}
 \tag{50}$$

For large  $K - A$ , both of these polynomials are dominated by the second term in the brackets, so taking the limit as  $K$  goes to infinity:

$$\begin{aligned}
 \text{2a. } \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{2r_L}{\sigma^2} \frac{4AK}{(K-A)^2} \frac{1}{6AK} K^3 \left( \frac{1}{2}K - A \right) \\
 \text{2b. } \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{2r_L}{\sigma^2} \frac{4AK}{(K-A)^2} \frac{1}{2AK} K^2 \left( \frac{1}{3}K - A \right)
 \end{aligned}
 \tag{51}$$

Replacing  $K - A$  with  $K$ ,  $\frac{1}{2}K - A$  with  $\frac{1}{2}K$ , and canceling constants gives:

$$\begin{aligned}
 \text{2a. } \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{2}{3} K^2 \\
 \text{2b. } \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{4}{3} K.
 \end{aligned}
 \tag{52}$$

Notice that both of these arguments are one order lower when compared to the same terms in the unscaled models. On the other hand, the arguments of the exponential terms in the logistic models are:

$$\begin{aligned}
 \text{1a. } \lim_{K \rightarrow \infty} \frac{2r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{1}{3} K^2 \\
 \text{1b. } \lim_{K \rightarrow \infty} \frac{2r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} K.
 \end{aligned}
 \tag{53}$$

Therefore, the argument of the exponential term in the Allee models converges to twice the argument of the

exponential term in the logistic model when the noise is constant, and four-thirds the argument when the noise is linear. The prefactors for the rescaled Allee models are  $C(A, K) = \frac{2\pi}{r_L}(K - A)$  and  $C(A, K) = \frac{2\pi}{r_L} \frac{K - A}{K}$ . As usual these are rational in  $K$  and  $A$ , so are dominated by the exponential terms when  $K$  is large. Therefore, while the scaling successfully reduces the order of the argument of the exponential term in the Allee models, it does not resolve the paradox.

The same analysis can be repeated for the second rescaling. The logistic growth rate is quadratic and is maximized at  $x_* = K/2$ . Therefore, the maximum absolute growth rate for the logistic model is  $r_L K/4$ . The Allee growth rate is cubic, and is maximized at  $x_* = \frac{1}{3}[(K + A) + \sqrt{K^2 + KA - A^2}]$ . For large  $K - A$ , this approaches

$x_* \approx \frac{2}{3}K + \frac{1}{3}A$ . Therefore, the maximum absolute growth rate of the Allee model is approached (from below) by

$\frac{2r_A}{27AK}(K - A)^2(2K + A)$ . Therefore, the scaling  $s(A, K)$  approaches  $\frac{27}{8} \frac{AK^2}{(K-A)^2(2K+A)}$  from below when  $K$  is large.

Plugging into Eq. 25 and taking the limit of large  $K$  gives:

$$\begin{aligned} 2a. \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{27}{96} K^2 = \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{9}{32} K^2 \\ 2b. \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{27}{48} K = \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{9}{16} K. \end{aligned} \tag{54}$$

Once again, rescaling lowers the order of the argument of exponential term in the Allee models. Compared to the equivalent logistic models, the argument of the exponential term in the Allee models converge to  $28/32 = 7/8$  times the argument in the logistic model with constant noise, and  $9/16$  times the argument in the logistic model with linear noise. Since both of these factors are less than one, the exponential term in the Allee model is smaller than in the logistic model, resolving the paradox when  $K$  is large enough that  $C(A, K)$  can be ignored.

Equating the linear stability of the two models at carrying capacity requires rescaling the Allee model so that the derivative of the absolute growth rate at  $x = K$  is equal to the derivative of the absolute growth rate of the logistic model at  $x = K$ . At  $x = K$  the logistic growth rate has slope  $-r_L$ . At  $x = K$  the growth rate of the Allee model has slope  $-r_A \frac{K-A}{A}$ . Therefore  $s(A, K) = \frac{A}{K-A}$ .

Plugging into Eq. 25 and taking the limit of large  $K$  gives:

$$\begin{aligned} 2a. \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{1}{6} K^2 \\ 2b. \lim_{K \rightarrow \infty} \frac{2s(A,K)r_L}{\sigma^2} W(K \rightarrow 0) &= \lim_{K \rightarrow \infty} \frac{r_L}{\sigma^2} \frac{1}{3} K. \end{aligned} \tag{55}$$

The comparable terms in the logistic model are twice and three times as large; therefore, the logistic model is more stable when  $K$  is large enough that  $C(A, K)$  can be ignored.

### A.4 Diffusion approximation

In Section “How to avoid the paradox” we showed that the extinction time paradox is resolved by deriving the SDE models from birth-death processes, without discussing the details of the approximations involved. This is a well studied area, and there are multiple ways of constructing such an approximation. Accordingly, this appendix will distinguish which of these methods is most appropriate in the context of this paper.

The most familiar method for approximating a birth-death process with an SDE is a system size expansion (Van Kampen 1992). The system size expansion assumes that both the birth and death parameter scale naturally in some system size,  $\Omega$  taken to represent a characteristic population. In our case  $\Omega$  could be set to either the carrying capacity or the Allee threshold. By rescaling the population in the system size it is often possible to perform an asymptotic expansion of the birth-death process. As  $\Omega$  becomes large, the master equation converges to a Fokker-Planck equation

(Van Kampen 1992). This Fokker-Planck equation governs the diffusion of probability for an SDE with instantaneous drift  $m(x) = \lambda(x) - \mu(x)$  and diffusion coefficient  $v(x) = \frac{1}{2}(\lambda(x) + \mu(x))$  (Bresloff 2014).

Notice that this method assumes that the system size is large. As a result, it is only appropriate if the dynamics of interest occur at large populations. This is not the case for an extinction process since a population must necessarily be small before going extinct. Albeit, if the Allee threshold,  $A$ , is large, then the extinction process can be separated into two phases. First, the population escapes from the carrying capacity to the Allee threshold; then, after crossing the Allee threshold, it goes extinct with high probability. In general, the escape process is much slower than the process of descending from the Allee threshold to extinction. Therefore, the mean extinction time of an Allee model is dominated by the model’s behavior at populations near, or greater than, the Allee threshold. In that case, if  $A$  is large, it is possible to approximate the mean extinction time based solely on the behavior at large populations. Then, a system size expansion is appropriate. However, our goal is to compare extinction times between Allee models and logistic models. The logistic model does not have a threshold past which extinction becomes highly likely. As a result, mean extinction times for the logistic model depend critically on the behavior of the model at small populations, even if the carrying capacity is large. Therefore, a system size expansion is not appropriate. In some cases, it is possible to use a system size expansion coupled to a small system correction (Ovaskainen and Meerson 2010); however, this is unnecessarily complicated in the context of this paper.

Also notice that the mean drift and diffusion coefficient scale differently in  $\Omega$ . Unlike the mean, the diffusion coefficient is vanishing as  $\Omega$  becomes large (Bresloff 2014).<sup>7</sup> As a result, large system size implies low temperature, and long extinction times. This limit is counterproductive since our goal was to study the roles of  $K$ ,  $A$ , and temperature separately. Moreover, if the system size is greater than one it leads to an SDE that is artificially less noisy than the original birth-death process. One solution is to take the system size limit to derive expressions for  $m(x)$  and  $v(x)$  in terms of  $\lambda(x)$ ,  $\mu(x)$  and  $\Omega$ , then set  $\Omega = 1$ . This, however, defeats

<sup>7</sup>This motivates the use of WKB approximations in Dykman et al. (1994) and Kamenev et al. (2008).



the purpose of the system size expansion since the system size expansion is only accurate<sup>8</sup> if it is assumed that  $\Omega$  is large.

Lastly, there exists a discrete potential for any birth-death process (see Eq. 64) analogous to the continuous potentials discussed in Section “Analysis.” This discrete potential does not match the continuous potential at small populations when the continuous potential is derived from a system size expansion (Bresloff 2014). As a result, both the potentials and the passage time statistics of the SDE differ from the potential and passage time statistics of the original birth-death process (Doering et al. 2005).

In summary, the system size approximation constructs a family of discrete birth-death processes indexed by a parameter that converge to an SDE in a limit. The discrete process is closely approximated by an SDE when the parameter is near its limiting value. This ensures that the SDE is a good approximation for the system in a particular limit; however, it does not guarantee that the birth-death process in that limit closely approximates the original birth-death process. In particular, there is no guarantee that a birth-death process with a large system size gives a good approximation for a birth-death process at small or finite system sizes.

An alternative approach is proposed in Allen and Allen (2002) and Cresson and Sonner (2018). Instead of approximating the discrete process in a particular limit where it converges to an SDE, we could attempt to approximate the original process with an SDE without taking a limit. There is no guarantee that this approximation will be accurate, since the discrete process is not an SDE; however, this approach avoids introducing errors due to an artificial system size. In fact, since the discrete process is not an SDE there will necessarily be error in any approximation, so the choice of SDE is not unique. In Allen and Allen (2002) Allen proposes a tried-and-true heuristic for picking a SDE that is conceptually consistent with the original birth-death process, which we now describe.

Any SDE is uniquely specified by its mean drift  $m(x)$  and diffusion coefficient  $v(x)$ . Let  $\bar{x} = \mathbb{E}[X]$ . If the initial state of the system is known then  $m(x) = \frac{d}{dt}\mathbb{E}[X]$  and  $v(x) = \frac{d}{dt}\mathbb{E}[(X - \bar{x})^2]$ . Given the initial state it is possible to compute both the rate of change in the expected population, and the rate of change of the diffusion coefficient in population for a birth-death process. Therefore, it is natural to approximate the birth-death process with an SDE that has the same mean drift and diffusion coefficient.

Suppose the discrete process is in state  $x$  at time 0. Then, from the master equation, it is easy to check that:

$$\begin{aligned} m(x) &= \frac{d}{dt}\mathbb{E}[X|X(0) = x] = \lambda(x) - \mu(x) \\ v(x) &= \frac{d}{dt}\mathbb{E}[(X - x)^2|X(0) = x] = \lambda(x) + \mu(x). \end{aligned} \tag{56}$$

<sup>8</sup>Accurate in the sense that the master equation converges to the Fokker-Planck equation

This motivates (29). This is philosophically different than the system size approximation since it uses an SDE to approximate the discrete birth-death process as it is, not as it is in a particular limit. When formulated in this way the passage time statistics for the SDE and the birth-death processes are generally similar (Allen and Allen 2002).<sup>9</sup>

Like the system size expansion, it is possible to construct a family of discrete birth-death processes that converge to this SDE in a particular limit. Suppose that we refine the discrete process by introducing fractional populations. By modifying the birth and death rates at the same time, it is possible to construct a sequence of refined birth-death processes whose mean drift and diffusion coefficient do not depend on the size of the refinement.

Consider a discrete birth-death process where each event produces  $0 \leq \Delta x \leq 1$  individuals. Assume that the birth and death rates are smooth continuous functions of  $x$ , and let  $p(x, t)$  denote the probability that the population has  $x$  individuals at time  $t$ . Define the modified rates  $\tilde{\lambda}(x|\Delta x)$  and  $\tilde{\mu}(x|\Delta x)$  such that  $\tilde{\lambda}(x|1) = \lambda(x)$  and  $\tilde{\mu}(x|1) = \mu(x)$ . For concision, we will repress the dependence on  $\Delta x$  unless necessary. Then, the refined master equation at  $x$  reads:

$$\begin{aligned} \frac{d}{dt}p(x, t) &= \tilde{\lambda}(x - \Delta x)p(x - \Delta x, t) - (\tilde{\lambda}(x) + \tilde{\mu}(x))p(x, t) \\ &\quad + \tilde{\mu}(x + \Delta x)p(x + \Delta x, t). \end{aligned} \tag{57}$$

Now let  $\Delta x$  be small. Taylor expanding in  $\Delta x$  gives:

$$\begin{aligned} \tilde{\lambda}(x - \Delta x)p(x - \Delta x, t) &= \tilde{\lambda}(x)p(x, t) - \frac{\partial}{\partial x}\tilde{\lambda}(x)p(x, t)\Delta x \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}\tilde{\lambda}(x)p(x, t)\Delta x^2 + \mathcal{O}(\Delta x^3) \\ \tilde{\mu}(x + \Delta x)p(x + \Delta x, t) &= \tilde{\mu}(x)p(x, t) + \frac{\partial}{\partial x}\tilde{\mu}(x)p(x, t)\Delta x \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}\tilde{\mu}(x)p(x, t)\Delta x^2 + \mathcal{O}(\Delta x^3). \end{aligned} \tag{58}$$

The zeroth order terms cancel, leaving:

$$\begin{aligned} \frac{d}{dt}p(x, t) &= -\frac{\partial}{\partial x}\left(\tilde{\lambda}(x) - \tilde{\mu}(x)\right)p(x, t)\Delta x \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\tilde{\lambda}(x) + \tilde{\mu}(x)\right)p(x, t)\Delta x^2 + \mathcal{O}(\Delta x^3). \end{aligned} \tag{59}$$

Finally, to rewrite in terms of density,  $\rho(x, t)$ , divide across by  $\Delta x$ :

$$\begin{aligned} \frac{\partial}{\partial t}\rho(x, t) &= -\frac{\partial}{\partial x}\left(\tilde{\lambda}(x) - \tilde{\mu}(x)\right)\rho(x, t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\tilde{\lambda}(x) \right. \\ &\quad \left. + \tilde{\mu}(x)\right)\rho(x, t)\Delta x + \mathcal{O}(\Delta x^2). \end{aligned} \tag{60}$$

<sup>9</sup>This approximation ensures that the discrete process and the continuous process have the same mean drift and diffusion coefficient; however, this does not ensure they will have the same higher order moments. The higher order moments of the SDE are Gaussian, while the higher order moments of the discrete process are not. Therefore, this approximation can be viewed as an approximate solution to the moment closure problem via truncation of the moment hierarchy. Any error in the approximation arises from the difference in the higher order moments.

In the limit of vanishing  $\Delta x$ , Eq. 60 approaches the Ito form of the Fokker-Planck equation with infinitesimal mean and diffusion coefficient:

$$\begin{aligned} m(x|\Delta x) &= \tilde{\lambda}(x|\Delta x) - \tilde{\mu}(x|\Delta x) \\ v(x|\Delta x) &= \Delta x \left( \tilde{\lambda}(x|\Delta x) + \tilde{\mu}(x|\Delta x) \right). \end{aligned} \tag{61}$$

Now, to ensure the noise intensity,  $\sigma^2$  does not vanish as  $\Delta x$  goes to zero set:

$$\begin{aligned} \tilde{\lambda}(x|\Delta x) &= \frac{1}{2\Delta x} (\lambda(x) + \mu(x)) + \frac{1}{2} (\lambda(x) - \mu(x)) \\ \tilde{\mu}(x|\Delta x) &= \frac{1}{2\Delta x} (\lambda(x) + \mu(x)) - \frac{1}{2} (\lambda(x) - \mu(x)). \end{aligned} \tag{62}$$

Then,  $\tilde{\lambda}(x|1) = \lambda(x)$ ,  $\tilde{\mu}(x|1) = \mu(x)$  and the infinitesimal mean and diffusion coefficient of the SDE match the infinitesimal mean and diffusion coefficient of the original birth-death process. Defining the modified birth and death rates in this way ensures that the infinitesimal mean and diffusion coefficient of the refined process scale identically in  $\Delta x$ . This is accomplished whenever:<sup>10</sup>

$$\frac{\tilde{\lambda}(x|\Delta x) - \tilde{\mu}(x|\Delta x)}{\tilde{\lambda}(x|\Delta x) + \tilde{\mu}(x|\Delta x)} = \mathcal{O}(\Delta x). \tag{63}$$

Condition (63) guarantees that the potential of the birth-death process converges to the potential of the limiting SDE (Allen and Allen 2002). This is not true for all diffusion approximations (Doering et al. 2005), including the systems size expansion (Bresloff 2014). See Appendix A.5 for an example of a discrete birth-death process that, when refined, obeys condition (63) automatically.

A discrete birth-death process has potential:

$$\phi(x) = -\frac{1}{2} \sum_{y=0}^x \log \left( \frac{\tilde{\lambda}(y)}{\tilde{\mu}(y + \Delta x)} \right) \tag{64}$$

and mean times to extinction (Allen and Allen 2002; Leigh 1981):

$$\tau(K) = \sum_{x=0}^K \sum_{y=x}^{\infty} \exp(2(\phi(x) - \phi(y))). \tag{65}$$

Notice the similarity between (65) and (15). Taking  $\Delta x$  to zero:

$$\frac{\tilde{\lambda}(x)}{\tilde{\mu}(x + \Delta x)} = \frac{v(x) + m(x)\Delta x}{v(x) - m(x)\Delta x} = 1 + 2\frac{m(x)}{v(x)}\Delta x + \mathcal{O}(\Delta x^2). \tag{66}$$

Therefore:

$$\frac{1}{2} \log \left( \frac{\tilde{\lambda}(x)}{\tilde{\mu}(x + \Delta x)} \right) = \frac{m(x)}{v(x)}\Delta x + \mathcal{O}(\Delta x^2). \tag{67}$$

<sup>10</sup>This is a standard assumption used in the “hydrodynamic” limit of a random walk and is a key part of the “delicate balance” (Gillespie John 1989) discussed in Section “How to avoid the paradox.”

Then:

$$\begin{aligned} \lim_{\Delta x \rightarrow 0} \sum_{i=0}^j \frac{1}{2} \log \left( \frac{\tilde{\lambda}(x)}{\tilde{\mu}(x + \Delta x)} \right) &= \lim_{\Delta x \rightarrow 0} \sum_{i=0}^j \frac{m(x)}{v(x)}\Delta x \\ &+ \mathcal{O}(\Delta x^2) = \int_0^x \frac{m(s)}{v(s)} ds = \int_0^x f(s) ds. \end{aligned} \tag{68}$$

Now, from Eq. 11, it follows that the discrete potential converges to the continuous potential:

$$\lim_{\Delta x \rightarrow 0} \phi = S. \tag{69}$$

This same technique is used in Leigh (1981) to approximate  $\phi$  analytically. As  $\Delta x$  goes to zero, the double sum is replaced with a double integral and the extra factor of 2 is absorbed into the temperature. Then, the mean extinction time for the discrete model converges to the mean extinction time for the continuous model.

### A.5 Physical analogy

In Section “Theory,” we asserted that the ratios  $f(x) = m(x)/v(x)$  and  $T = \frac{\sigma^2}{2r}$  acted like forces and temperature. Here, we develop a specific physical system which is mathematically identical to the population models we consider. We then show that  $f(x)$  and  $T$  are force and temperature in the analogous physical system.

Consider an ion with charge  $q$  diffusing down a narrow channel. The channel could be a pore in a cell membrane. The ion is subject to constant thermal noise and, in the absence of a driving potential, would follow a Wiener process (Brownian motion). Assume that the channel is kept at fixed temperature  $T$ , and there is only one charged particle in the channel. The channel is not kept at constant voltage, so for any  $x$ , there is a voltage  $V(x)$ . Therefore, if the ion is at  $x$  it has electrostatic potential energy  $U(x) = qV(x)$ .

Coarse grain the channel into a sequence of discrete intervals with width  $\Delta x$ . These could correspond to a sequence of binding sites that are evenly spaced along the channel. Then, the motion of the ion is closely approximated by a discrete random walk with exponentially distributed waiting times. For a given  $x$  let  $\lambda(x, \Delta x)$ , denote the rate at which the ion hops to  $x + \Delta x$  and  $\mu(x, \Delta x)$  denote the rate at which the ion hops to  $x - \Delta x$ . For concision, we will suppress  $\lambda$  and  $\mu$ ’s dependence on  $\Delta x$  except where necessary. Following (Schnakenberg 1976):

$$\frac{\lambda(x)}{\mu(x + \Delta x)} = \exp \left( \frac{1}{k_B T} (U(x) - U(x + \Delta x)) \right) \tag{70}$$

where  $k_B$  is the Boltzmann constant. This is the essential physics that will link forces and temperature to ratios of  $m(x)$ ,  $v(x)$ ,  $\sigma^2$ , and  $r$ . In order to study the relationship

between physical quantities and the SDE, we take a diffusion limit one side at a time.

The derivative of the potential energy  $U(x)$  is a force  $f(x)$ , so, by the intermediate value theorem:

$$U(x) - U(x + \Delta x) = - \int_x^{x+\Delta x} f(s)ds = f(\zeta)\Delta x \quad (71)$$

for some  $\zeta \in [x, x + \Delta x]$ . In the limit as  $\Delta x$  goes to zero  $f(\zeta)$  converges to  $f(x)$ . Now, Taylor expanding the exponential in small  $\Delta x$ :

$$\exp\left(-\frac{1}{k_B T} f(\zeta)\Delta x\right) = 1 + \frac{1}{k_B T} f(\zeta)\Delta x + \mathcal{O}(\Delta x^2). \quad (72)$$

This implies that, in the limit as  $\Delta x$  goes to zero the ratio  $\lambda(x, \Delta x)/\mu(x + \Delta x, \Delta x)$  converges to 1. This, in turn, implies that  $(\lambda(x, \Delta x) - \mu(x + \Delta x, \Delta x))/(\lambda(x, \Delta x) + \mu(x + \Delta x, \Delta x))$  is  $\mathcal{O}(\Delta x)$  (converges to zero proportional to  $\Delta x$ ). Therefore, this discrete process is mathematically identical to the birth-death process considered in Appendix A.3. It follows that:

$$\frac{\lambda(x)}{\mu(x + \Delta x)} = \frac{\sigma^2 v(x) + rm(x)\Delta x}{\sigma^2 v(x) - rm(x)\Delta x} = 1 + 2 \frac{rm(x)}{\sigma^2 v(x)} \Delta x + \mathcal{O}(\Delta x^2). \quad (73)$$

Equating the left and right hand sides:

$$\frac{2r}{\sigma^2} \frac{m(x)}{v(x)} = \frac{1}{k_B T} f(x) \quad (74)$$

This leads immediately to the association:

$$\frac{m(x)}{v(x)} = f(x) \quad (75)$$

$$\frac{\sigma^2}{2r} = k_B T.$$

### A.6 Analysis for a saturating Allee effect

In Section “How to avoid the paradox,” we proposed two different modifications to our Allee model that resolve the passage time paradox: either modify  $m(x) = \frac{r}{K}x(K - x)$  with a saturating function or derive  $v(x)$  from a birth-death model. Here, we show that these two modifications lead to a potential with the same functional form.

First, suppose we set:

$$m(x) = \frac{r}{K}x(K - x) - \beta(1 - P(x))x, \quad P(x) = \frac{x}{\theta + x} \quad (76)$$

as in Eq. 27. This model can be derived by assuming fecundity  $\beta$ , probability of finding a mate  $P(x)$ , background per capita death rate  $\delta_0$  and additional death rate due to competition  $\delta_1$  (Boukal and Berec 2002; Dennis 2002; 1989; McCarthy 1997). To convert into  $r$  and  $K$  set  $K = \frac{\beta - \delta_0}{\delta_1}$  and  $r = \beta - \delta_0$  (see Eq. 27).

The corresponding deterministic model has root where  $m(x) = 0$ . Therefore, the roots occur at 0, some  $A$ , and some  $K'$  near  $K$ . Since the subsequent model will be analyzed in terms of birth and death rates, and since the

Allee threshold  $A$  and carrying capacity  $K'$  are no longer explicit parameters of the model it will be convenient to work with the parameters  $\beta, \delta_0, \delta_1, \theta$  from now on. The corresponding birth-death model is:

$$\lambda(x) = \beta \frac{x^2}{\theta + x} \quad (77)$$

$$\mu(x) = \delta_0 x + \delta_1 x^2.$$

Therefore, the forces are:

$$f(x) = \frac{\beta \frac{x}{\theta + x} - \delta_0 - \delta_1 x}{\beta \frac{x}{\theta + x} + \delta_0 + \delta_1 x} \quad (78)$$

which can be simplified by multiplying through by  $\theta + x$ :

$$f(x) = \frac{\beta x - \delta_0(\theta + x) - \delta_1 x(\theta + x)}{\beta x + \delta_0(\theta + x) + \delta_1(\theta + x)}$$

$$= - \frac{(-\beta + \delta_0 + \delta_1\theta)x + \delta_0\theta + \delta_1\theta^2}{(\beta + \delta_0 + \delta_1\theta)x + \delta_0\theta + \delta_1\theta^2}. \quad (79)$$

Let  $b = \frac{\beta}{\delta_1}$ ,  $c = \frac{\delta_0}{\delta_1}\theta$ , and  $d = \frac{\delta_0 + \delta_1\theta}{\delta_1}$ . Then, after dividing through by  $\delta_1$  and rearranging:

$$f(x) = - \frac{x^2 + (d - b)x + c}{x^2 + (b + d)x + c}. \quad (80)$$

The corresponding potential  $S(x)$  is:

$$S(x) = \int_0^x \frac{s^2 + (d - b)s + c}{s^2 + (b + d)s + c} ds. \quad (81)$$

To integrate, replace the numerator with  $s^2 + (d + b)s + c - 2bs$ . Then:

$$S(x) = \int_0^x 1 - \frac{2bs}{s^2 + (b + d)s + c} ds$$

$$= x - 2b \int_0^x \frac{s}{s^2 + (b + d)s + c} ds. \quad (82)$$

Notice that the integrand is vanishing for large  $s$ , so the potential is close to linear in  $x$  for large  $x$ . This is a natural feature of  $f(x)$  derived from a birth-death model, and reflects the saturating behavior of  $f(x)$ . To perform the integral factor the denominator:

$$s^2 + (b + d)s + c = (s - r_1)(s - r_2) \quad (83)$$

where  $r_1, r_2$  are the roots of the polynomial given by:

$$r_{1,2} = - \frac{(b + d) \pm \sqrt{(b + d)^2 - 4c}}{2}. \quad (84)$$

Notice that since  $b$  and  $d$  are both positive, the real part of the roots is always negative.

Next, we can expand the integrand using partial fractions:

$$\frac{s}{s^2 + (b + d)s + c} = \frac{r_1}{r_1 - r_2} \frac{1}{s - r_1} + \frac{r_2}{r_2 - r_1} \frac{1}{s - r_2}. \quad (85)$$

Now, the integration is easy:

$$\frac{r_1}{r_1 - r_2} \int_0^x \frac{1}{s - r_1} ds + \frac{r_2}{r_2 - r_1} \int_0^x \frac{1}{s - r_2} ds = \frac{r_1}{r_1 - r_2} \log \left( -\frac{(x - r_1)}{r_1} \right) + \frac{r_2}{r_2 - r_1} \log \left( -\frac{(x - r_2)}{r_2} \right). \quad (86)$$

Therefore:

$$S(x) = x - \left[ \frac{r_1}{r_1 - r_2} \log \left( -\frac{(x - r_1)}{r_1} \right) + \frac{r_2}{r_2 - r_1} \log \left( -\frac{(x - r_2)}{r_2} \right) \right]. \quad (87)$$

Notice the immediate similarity with Eq. 35. This is actually a general form, since any forces that are rational functions whose numerator and denominator are of the same order can be analyzed in the same way. Note that if  $v(x)$  is derived from an underlying birth-death model, the birth and death rates are polynomials or rational functions, and the higher order terms in the birth and death rates do not cancel when added or subtracted, then the forces are rational with numerator and denominator of the same order. For example, suppose we had started with the birth-death model (Brassil 2001):

$$\begin{aligned} \lambda(x) &= \beta_1 x^2 \\ \mu(x) &= \delta_0 x + \delta_2 x^3 \end{aligned} \quad (88)$$

Then:

$$\begin{aligned} m(x) &= \beta_1 x^2 - \delta_0 x - \delta_2 x^3 \\ v(x) &= \beta_1 x^2 + \delta_0 x + \delta_2 x^3. \end{aligned} \quad (89)$$

Notice that  $m(x)$  is cubic, so could be rewritten  $m(x) = \frac{r}{AK}(x - A)(K - x)x$  as in models 2a and 2b. The corresponding forces are:

$$f(x) = -\frac{\delta_2 x^2 - \beta_1 x + \delta_0}{\delta_2 x^2 + \beta_1 x + \delta_0} \quad (90)$$

Let  $b = \frac{\beta_1}{\delta_2}$ ,  $c = \frac{\delta_0}{\delta_2}$ . Then:

$$f(x) = -\frac{x^2 - bx + c}{x^2 + bx + c} = -1 + 2b \frac{x}{x^2 + bx + c} \quad (91)$$

This is essentially the same as the forces we analyzed in the previous case, only with different coefficients  $b$  and  $c$ . The potential can be derived following the same steps as before, only with roots:

$$r_{1,2} = -\frac{b \pm \sqrt{b^2 - 4c}}{2}. \quad (92)$$

In either case, the potential is linear in  $x$  with a logarithmic correction. It follows that the passage time from carrying capacity to extinction is, at worst, exponential in  $K$ .

## References

- Allen LJS, Allen EJ (2002) A comparison of three different stochastic population models with regard to persistence time. *Theor Popul Biol* 64:439–449
- Allen EJ, Allen LJS, Arciniega A, Greenwood PE (2008) Construction of equivalent stochastic differential equation models. *Stoch Anal Appl* 26:274–297
- Amarasekare P (1998) Allee effects in metapopulation dynamics. *Am Nat* 152:298–302
- Patrick B (1995) *Probability and measure*, 3rd edn. Wiley-Interscience Publication, New York, pp 297–308
- Brassil CE (2001) Mean time to extinction of a metapopulation with an Allee effect. *Ecol Model* 143:9–16
- Bresloff PC (2014) *Stochastic processes in cell biology*. Springer, New York, pp 133–138
- Boukal DS, Berec L (2002) Single species models of the Allee effect: extinction boundaries, sex-ratios and mate encounters. *J Theor Biol* 218:375–394
- Courchamp F, Clutton-Brock T, Grenfell B (1999) Inverse density dependence and the Allee effect. *Trends Ecol Evol* 14:405–410
- Cresson J, Sonner S (2018) A note on a derivation method for SDE models: applications in biology and viability criteria. *Stoch Anal Appl* 36:224–239
- Dennis B (1989) Allee effects: population growth, critical density and the chance of extinction. *Nat Resour Model* 3:481–538
- Dennis B (2002) Allee effects in stochastic populations. *Oikos* 96:389–401
- Dennis B, Laila A, Saber E, Eddy K, George L (2016) Allee effects and resilience in stochastic populations. *Theor Ecol* 9:323–335
- Desharnais Robert A, Constantino RF, Cushing JM, Henson Shandelle M, Brian D, King Aaron A (2006) Experimental support of the scaling rule for demographic stochasticity. *Ecol Lett* 9:537–545
- Doering C, Sargsyan K, Sander LM (2005) Extinction times for birth-death processes: exact results, continuum asymptotics and the failure of the Fokker-Planck approximation. *Multi-Scale Model Simul* 3:283–299
- Dykman MI, Mori E, Hunt JRPM (1994) Large fluctuations and optimal paths in chemical kinetics. *J Chem Phys* 100:5735–5750
- Foley P (1994) Predicting extinction times from environmental stochasticity and carrying capacity. *Conserv Biol* 8:124–137
- Gillespie John H (1989) When not to use diffusion processes in population genetics. *Math Evol Theory* 57:70
- Gruntfest Y, Arditi R, Dombrovksy Y (1997) A fragmented population in a varying environment. *J Theor Biol* 185:539–547
- Higham DJ (2001) An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev* 43:525–546

- Higham DJ (2008) Modeling and simulating chemical reactions. *SIAM Rev* 50:347–368
- Kamenev A, Meerson B, Shklovskii B (2008) How colored environmental noise affects population extinction. *PRL* 101:268103–268107
- Samuel K, Taylor Howard M (1981) *A second course in stochastic processes*. Academic Press, New York, pp 169–191
- Keitt TH, Lewis MA, Holt RD (2001) Allee effects, invasion pinning, and species' borders. *Am Nat* 157:203–216
- Kussell E, Vucelja M (2014) Non-equilibrium physics and evolution - adaptation, extinction, and ecology: a key issues review. *Rep Prog Phys* 77:102602
- Lande R (1993) Risks of population extinction from demographic and environmental stochasticity and random catastrophes, vol 142
- Leigh GC (1981) The average lifetime of a population in a varying environment. *J Theor Biol* 90:213–239
- Lewis MA, Kareiva P (1993) Allee dynamics and the spread of invading organisms. *Theor Popul Biol* 43:141–158
- Levine EY, Meerson B (2013) Impact of colored environmental noise on the extinction of a long-lived stochastic population: role of the Allee effect. *PRE* 87:032127
- McCarthy MA (1997) The Allee effect, finding mates and theoretical models. *Ecol Model* 103:99–102
- Nisbet RM, Gurney WCS (1982) *Modelling fluctuating populations*, vol 167–214. The Blackburn Press, New Jersey
- Nolting B, Abbott K (2016) Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. *Ecology* 97:850–864
- Ovaskainen O, Meerson B (2010) Stochastic models of population extinction. *Trends Ecol Evol* 25:643–652
- Poggiale JC (1998) From behavioural to population level: growth and competition. *Math Comput Modell* 27:41–49
- Hong Q, Min Q, Xiang T (2002) Thermodynamics of the general diffusion process: time-reversibility and entropy production. *J Stat Phys* 107:1129–1141
- Schnakenberg J (1976) Network theory of microscopic and macroscopic behaviour of master equation systems. *Rev Mod Phys* 48:571–586
- Stephan T, Wissel C (1994) Stochastic extinction models discrete in time. *Ecol Model* 75:183–192
- Stephens PA, Sutherland WJ, Freckleton RP (1999) What is the Allee effect. *Oikos* 87:185–190
- Stephens PA, Sutherland WJ (1999) Consequences of the Allee effect for behaviour, ecology, and conservation. *Trends Ecol Evol* 14:401–405
- Tier C, Hanson FB (1981) Persistence in density dependent stochastic populations. *Math Biosci* 53:89–117
- Turelli M (1977) Random environments and stochastic calculus. *Theor Popul Biol* 12:140–178
- Van Kampen NG (1980) Ito versus Stratonovich. *J Stat Phys* 24:175–187
- Van Kampen NG (1992) *Stochastic processes in physics and chemistry*. Elsevier Science, New York, pp 134–153
- Wilkinson DJ (2012) *Stochastic modelling for systems biology*. Chapman and Hall CRC Press, USA, pp 3–20,123–168