Coexistence in the Face of Uncertainty

Sebastian J. Schreiber

> *Lest men believe your tale untrue, keep probability in view.*
>
> *—John Gay*

Abstract Over the past century, nonlinear difference and differential equations have been used to understand conditions for coexistence of interacting populations. However, these models fail to account for random fluctuations due to demographic and environmental stochasticity which are experienced by all populations. I review some recent mathematical results about persistence and coexistence for models accounting for each of these forms of stochasticity. Demographic stochasticity stems from populations and communities consisting of a finite number of interacting individuals, and often are represented by Markovian models with a countable number of states. For closed populations in a bounded world, extinction occurs in finite time but may be preceded by long-term transients. Quasi-stationary distributions (QSDs) of these Markov models characterize this meta-stable behavior. For sufficiently large "habitat sizes", QSDs are shown to concentrate on the positive attractors of deterministic models. Moreover, the probability extinction decreases exponentially with habitat size. Alternatively, environmental stochasticity stems from fluctuations in environmental conditions which influence survival, growth, and reproduction. Stochastic difference equations can be used to model the effects of environmental stochasticity on population and community dynamics. For these models, stochastic persistence corresponds to empirical measures placing arbitrarily little weight on arbitrarily low population densities. Sufficient and necessary conditions for stochastic persistence are reviewed. These conditions involve weighted combinations of Lyapunov exponents corresponding to "average" per-capita growth rates of rare species. The results are illustrated with how climatic variability influenced the dynamics of Bay checkerspot butterflies, the persistence

S.J. Schreiber (✉)
Department of Evolution and Ecology, University of California, Davis, CA 95616, USA
e-mail: sschreiber@ucdavis.edu

of coupled sink populations, coexistence of competitors through the storage effect, and stochastic rock-paper-scissor communities. Open problems and conjectures are presented.

1 Introduction

A long standing, fundamental question in biology is "what are the minimal conditions to ensure the long-term persistence of a population or the long-term coexistence of interacting species?" The answers to this question are essential for guiding conservation efforts for threatened and endangered species, and identifying mechanisms that maintain biodiversity. Mathematical models have and continue to play an important role in identifying these potential mechanisms and, when coupled with empirical work, can test whether or not a given mechanism is operating in a specific population or ecological community [1]. Since the pioneering work of [35] and [56] on competitive and predator–prey interactions, [41, 54] on host–parasite interactions, and [30] on disease outbreaks, nonlinear difference and differential equations have been used to understand conditions for persistence of populations or communities of interacting species. For these deterministic models, persistence or species coexistence is often equated with an attractor bounded away from the extinction states in which case persistence holds over an infinite time horizon [47]. However (with apologies to John Gay), lest biologists believe this theory untrue, the models need to keep probability in view. That is, all natural populations exhibit random fluctuations due to mixture of intrinsic and extrinsic factors known as demographic and environmental stochasticity. The goal of this chapter is to present models that account for these random fluctuations, review some mathematical methods for analyzing these stochastic models, and illustrate how these random fluctuations hamper or facilitate population persistence and species coexistence.

Demographic stochasticity corresponds to random fluctuations due to populations consisting of a finite number of individuals whose fates aren't perfectly correlated. That is, even if all individuals in a population appear to be identical, some undetectable differences between individuals (e.g. in their physiology or microenvironment) result in some individuals dying while others survive. To capture these "unknowable" differences, models can assign the same probabilities of dying to each individuals and treat survival amongst individuals as independent flips of a coin—heads life, tails death. Similarly, surviving individuals may differ in the number of offspring they produce despite appearing to be identical. To capture these unknowable differences, the number of offspring produced by these individuals are modeled as independent draws from the same probability distribution. The resulting stochastic models accounting for these random fluctuations typically

correspond to Markov chains on a finite or countable state space[1] e.g. the numbers of individuals, 0, 1, 2, 3, ..., in a population. When these models represent populations or communities whose numbers tend to stay bounded and have no immigration, the populations in these models always go extinct in finite time [10]. Hence, unlike deterministic models, the asymptotic behavior of these stochastic models is trivial: eventually no one is left. This raises the following basic question about the relationship between models accounting for demographic stochasticity and their deterministic counterparts:

> "Any population allowing individual variation in reproduction, ultimately dies out–unless it grows beyond all limits, an impossibility in a bounded world. Deterministic population mathematics on the contrary allows stable asymptotics. Are these artifacts or do they tell us something interesting about quasi-stationary stages of real or stochastic populations?"— Peter [29]

As it turns out, there is a strong correspondence between the quasi-stationary behavior of the stochastic models and the attractors of an appropriately defined mean-field model. Moreover, this correspondence highlights a universal scaling relationship between extinction times and the size of the habitat where the species live. These results and their applications are the focus of the first half of this review.

While demographic stochasticity affects individuals independently, environmental stochasticity concerns correlated demographic responses (e.g. increased survival, growth or reproduction) among individuals. These correlations often stem from individuals experiencing similar fluctuations in environmental conditions (e.g. temperature, precipitation, winds) which impact their survival, growth, or reproduction. Models driven by randomly fluctuating parameters or Brownian motions, such as random difference equations or stochastic differential equations, can capture these sources of random fluctuations. Unlike models for demographic stochasticity, these Markov chains always live on uncountable state spaces where the non-negative reals represent densities of populations of sufficiently large size that one can ignore the effects of being discrete and finite. Consequently, like their deterministic counterparts, extinction in these random difference equations only occurs asymptotically, and persistence is equated with tendency to stay away from low densities [11]. Understanding what this exactly means, reviewing methods for verifying this stochastic form of persistence, and applying these methods to gain insights about population persistence and species coexistence are the focus of the second half of this review.

Of course, all population systems experience a mixture of demographic and environmental stochasticity. While the theoretical biology literature is replete with models accounting for each of these forms of noise separately, I know of no studies that rigorously blend the results presented in this review. Hence, I conclude by discussing some open problems and future challenges at this mathematical interface.

---

[1]See, however, the discussion for biologically motivated uncountable state spaces.

2 Demographic Stochasticity

To model finite populations and account for demographic stochasticity, we consider Markov chains on a countable state space which usually is the non-negative cone of the integer lattice. Many of these stochastic models have a deterministic counterpart, sometimes called the "deterministic skeleton" or the "mean field model". As I discuss below, these deterministic models can provide some useful insights about the transient behavior of the stochastic models and when coupled with large deviation theory provide insights into the length of these transients.

To get a flavor of the types of models being considered, lets begin with a stochastic counterpart to the discrete-time Lotka-Volterra equations. This example motivates the main results and will illustrate their applicability.

Example 1 (Poisson Lotka-Volterra Processes)   The continuous time Lotka-Volterra equations form the bedrock for much of community ecology theory. While there are various formulations of their discrete-time counterparts, a particularly pleasing one that retains several key dynamical features of the continuous-time models was studied by [26]. These models keep track of the densities $x_t = (x_{1,t}, \ldots, x_{k,t})$ of k interacting species, where the subscripts denote the species identity i and time t (e.g. year or day). As with the classical continuous time equations, there is a matrix $A = (a_{ij})_{i,j}$ where a_{ij} corresponds to the "per-capita" effect of species j on species i and a vector $r = (r_1, \ldots, r_k)$ of the "intrinsic per-capita growth rates" for all of the species. With this notation, the equations take on the form:

$$x_{i,t+1} = x_{i,t} \exp\left(r_i + \sum_j a_{ij}x_{j,t}\right) =: F_i(x_t) \text{ with } i = 1, 2, \ldots, k. \qquad (1)$$

The state space for these dynamics are given by the non-negative orthant

$$\mathbb{R}^k_+ = \{x \in \mathbb{R}^k : x_i \geq 0 \text{ for all } i\}$$

of the k-dimensional Euclidean space $\mathbb{R}^k$.

To define the Poisson Lotka-Volterra process, let $1/\varepsilon$ be the size of the habitat in which the species live. Let $N^\varepsilon_t = (N^\varepsilon_{1,t}, \ldots, N^\varepsilon_{k,t})$ denote the vector of species abundances which are integer-valued. Then the density of species i is $X^\varepsilon_{i,t} = \varepsilon N^\varepsilon_{i,t}$. Over the next time step, each individual replaces itself with a Poisson number of individuals with mean

$$\exp\left(r_i + \sum_j a_{ij}X^\varepsilon_{j,t}\right).$$

If the individuals update independent of one another, then $N_{i,t+1}^\varepsilon$ is a sum of $N_{i,t}^\varepsilon$ independent Poisson random variables. Thus, $N_{i,t+1}^\varepsilon$ is also Poisson distributed with mean

$$N_{i,t}^\varepsilon \exp\left(r_i + \sum_j a_{ij} X_{j,t}^\varepsilon\right) = F_i(X_t^\varepsilon)/\varepsilon.$$

Namely,

$$\mathbb{P}[X_{i,t+1}^\varepsilon = \varepsilon j | X_t^\varepsilon = x] = \mathbb{P}[N_{i,t+1}^\varepsilon = j | X_t^\varepsilon = x] = \exp(-F_i(x)/\varepsilon)\frac{(F_i(x)/\varepsilon)^j}{j!}. \quad (2)$$

The state space for N_t^ε is the non-negative, k dimensional integer lattice

$$\mathbb{Z}_+^k = \{(z_1, \ldots, z_k) : z_i \text{ are non-negative integers}\}$$

while the state space for X_t^ε is the non-negative, rescaled integer lattice

$$\varepsilon\mathbb{Z}_+^k = \{(\varepsilon z_1, \ldots, \varepsilon z_k) : z_i \text{ are non-negative integers}\}.$$

Now consider a solution to deterministic model x_t and the stochastic process X_t^ε initiated at the same densities $x_0 = X_0^\varepsilon = x$. To see how likely X_t^ε deviates from x_t, we use Chebyshev's inequality. As the mean and variance of a Poisson random variable are equal, Chebyshev's inequality implies

$$\mathbb{P}\left[|X_{i,1}^\varepsilon - x_{i,1}| \geq \delta \Big| X_0^\varepsilon = x_0 = x\right] \leq \frac{\text{Var}[X_{i,1}^\varepsilon]}{\delta^2} = \frac{\varepsilon^2 \text{Var}[N_{i,1}^\varepsilon]}{\delta^2} = \frac{\varepsilon F_i(x)}{\delta^2} \quad (3)$$

where $\text{Var}[X]$ denotes the variance of a random variable X. In words, provided the habitat size $1/\varepsilon$ is sufficiently large, a substantial deviation between X_1^ε and x_1 is unlikely. In fact, one can show that over any finite time interval $[1, T]$, the stochastic dynamics are likely to be close to the deterministic dynamics over the time interval $[1, T]$ provided the habitat size $1/\varepsilon$ is sufficiently large:

$$\lim_{\varepsilon \to 0} \mathbb{P}\left[\max_{1 \leq i \leq k, 1 \leq t \leq T} |X_{i,t}^\varepsilon - x_{i,t}| \geq \delta \Big| X_0^\varepsilon = x_0 = x\right] = 0. \quad (4)$$

Figure 1 illustrates this fact for a Poisson Lotka-Volterra process with two competing species. Equation (4) is the discrete-time analog of a result derived by [34] for continuous-time Markov chains. [34] also provides "second-order" approximations for finite time intervals using Gaussian processes and stochastic differential equations. While these approximations are also useful for discrete-time models, we do not review them here.

Despite X_t^ε stochastically tracking x_t with high probability for long periods of time, eventually their behavior diverges as Poisson Lotka-Volterra processes go extinct in finite time or exhibit unbounded growth.
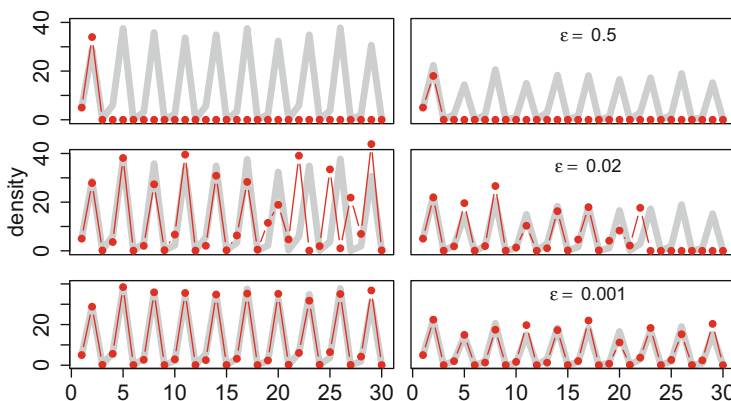
Fig. 1 Realizations of a Poisson Lotka-Voltera process with two competing species (species 1 on the left, species 2 on the right). The deterministic dynamics are shown as a *thick gray line*. Stochastic realizations are shown in *red*. Each row corresponds to a different habitat size $1/\varepsilon$. Parameter values: A is the matrix with rows $(-0.2, 0.1)$, $(-0.15, 0.2)$, and $r = (3.25, 3.25)$ for the model described in Example 1

Proposition 1 *Let X_t^ε be a Poisson Lotka-Volterra process with $\varepsilon > 0$. Then*

$$\mathbb{P}\left[\{X_t^\varepsilon = 0 \text{ for some } t\} \cup \{\lim_{t \to \infty} \sum_i X_{i,t}^\varepsilon = \infty\}\right] = 1$$

Furthermore, if F is pre-compact i.e. $F(\mathbb{R}_+^k) \subset [0, m]^k$ for some $m \geq 0$, then

$$\mathbb{P}\left[\{X_t^\varepsilon = 0 \text{ for some } t\}\right] = 1$$

The strategy used to prove the first statement of the proposition is applicable to many models of closed populations. The key ingredients are that there is a uniform lower bound to the probability of any individual dying, and individuals die independently of one another [10]. Proving, however, that extinction always occurs with probability one requires additional elements which aren't meet by all ecological models, but is meet for "realistic" models.

Proof For the first assertion, take any integer $m > 0$. Let

$$\beta = \min_{x \in [0,m]^k} \mathbb{P}[X_1^\varepsilon = 0 | X_0^\varepsilon = x] = \min_{x \in [0,m]^k} \exp\left(-\sum_{i=1}^k F_i(x)/\varepsilon\right) > 0.$$

Next we use the following standard result in Markov chain theory [17, Theorem 2.3 in Chapter 5].

Proposition 2 *Let X be a Markov chain and suppose that*

$$\mathbb{P}\left[\bigcup_{s=1}^{\infty}\{X_{t+s} \in C\}\Big|X_t\right] \geq \beta > 0 \text{ on } \{X_t \in B\}.$$

Then

$$P\left[\{X_t \text{ enters } B \text{ infinitely often}\} \setminus \{X_t \text{ enters } C \text{ infinitely often}\}\right] = 0.$$

Let $\mathscr{B}_m = \{X_t^\varepsilon$ enters $[0,m]^k$ infinitely often$\}$ and $\mathscr{E} = \{X_t^\varepsilon = 0$ for some $t\}$. Proposition 2 with $B = [0,m]^k$ and $C = \{0\}$ implies that

$$\mathbb{P}\left[\mathscr{B}_m \setminus \mathscr{E}\right] = 0. \tag{5}$$

The complement of the event $\cup_m \mathscr{B}_m$ equals the event $\mathscr{A} = \{\lim_{t\to\infty}\sum_i X_{i,t}^\varepsilon = \infty\}$. As $\mathscr{B}_m$ is an increasing sequence of events,

$$1 = \mathbb{P}\left[\mathscr{A} \cup \{\cup_m \mathscr{B}_m\}\right]$$

$$= \lim_{m\to\infty} \mathbb{P}\left[\mathscr{A} \cup \mathscr{B}_m\right]$$

$$\leq \lim_{m\to\infty} \mathbb{P}\left[\mathscr{A} \cup \mathscr{E}\right]$$

where the final inequality follows from (5). This completes the proof of the first assertion.

To prove the second assertion, assume that there exists $m > 0$ such that $F(x) \in [0,m]^k$ for all $x \in \mathbb{R}_+^k$ i.e. F is pre-compact. Define

$$\beta = \inf_{x\in\mathbb{R}_+^k} \mathbb{P}[X_{t+1}^\varepsilon = 0|X_t = x]$$

$$= \inf_{x\in\mathbb{R}_+^k} \exp\left(-\sum_i F_i(x)/\varepsilon\right)$$

$$\geq \exp(-k\,m/\varepsilon)$$

Applying Proposition 2 with $B = \mathbb{R}_+^k$ and $C = \{0\}$ completes the proof of the second assertion. $\qquad\square$

Equation (4) and Proposition 1 raise two fundamental questions about these stochastic, finite population models: How long before extinction occurs? Prior to extinction what can one say about the transient population dynamics? To get some insights into both of these questions, we build on the work of [21] and [31] on random perturbations of dynamical systems, and [3] on quasi-stationary distributions.

2.1 Random Perturbations and Quasi-Stationary Distributions

The Poisson Lotka-Volterra process (Example 1) illustrates how Markovian models can be viewed as random perturbations of a deterministic model. To generalize this idea, consider a continuous, precompact[2] map $F : \mathscr{S} \to \mathscr{S}$, where $\mathscr{S}$ is a closed subset of $\mathbb{R}^k$. F will be the deterministic skeleton of our stochastic models. *A random perturbation of F* is a family of Markov chains $\{X^\varepsilon\}_{\varepsilon>0}$ on $\mathscr{S}$ whose transition kernels

$$p^\varepsilon(x, \Gamma) = \mathbb{P}\left[X^\varepsilon_{t+1} \in \Gamma \mid X^\varepsilon_t = x\right] \text{ for all } x \in \mathscr{S} \text{ and Borel sets } \Gamma \subset \mathscr{S}$$

enjoy the following hypothesis:

Hypothesis 2.1 *For any $\delta > 0$,*

$$\lim_{\varepsilon \to 0} \sup_{x \in \mathscr{S}} p^\varepsilon\left(x, \mathscr{S} \setminus N^\delta(F(x))\right) = 0$$

where $N^\delta(y) := \{x \in \mathscr{S} : \|y - x\| < \delta\}$ denotes the δ-neighborhood of a point $y \in \mathscr{S}$.

Hypothesis 2.1 implies that the Markov chains X^ε converge to the deterministic limit as $\varepsilon \downarrow 0$ i.e. the probability of X^ε_1 being arbitrarily close to $F(x)$ given $X^\varepsilon_0 = x$ is arbitrarily close to one for ε sufficiently small. Hence, one can view F as the "deterministic skeleton" which gets clothed by the stochastic dynamic X^ε. The next example illustrates how to verify the Poisson Lotka-Volterra process is a random perturbation of the Lotka-Volterra difference equations.

Example 2 (The Poisson Lotka-Volterra Processes Revisited) Consider the Poisson Lotka-Volterra processes from Example 1 where $F(x) = (F_1(x), \ldots, F_k(x))$ and $F_i(x) = x_i \exp(r_i + \sum_j a_{ij}x_j)$ and $\mathscr{S} = \mathbb{R}^k_+$. For many natural choices of r_i and a_{ij}, [26] have shown there exists $C > 0$ such that $F(\mathscr{S}) \subset [0, C]^k$ i.e. F is pre-compact. While the corresponding Lotka-Volterra process X^ε lives on $\varepsilon\mathbb{Z}^k_+$, the process can be extended to all of $\mathscr{S}$ by allowing X^ε_0 to be any point in $\mathscr{S}$ and update with the transition probabilities of (2). With this extension, X^ε_1 always lies in $\varepsilon\mathbb{Z}^k_+$ and p^ε is characterized by the following probabilities

$$p^\epsilon(x, \{y\}) = \prod_{i=1}^k \exp(-F_i(x)/\varepsilon)\frac{(F_i(x)/\varepsilon)^{j_i}}{j_i!} \text{ for } y = \varepsilon(j_1, \ldots, j_k) \in \varepsilon\mathbb{Z}^k_+, x \in \mathscr{S}$$

and 0 otherwise. With this extension, Hypothesis 1 for the Lotka-Volterra process follows from equation (3).

---

[2]Namely, there exists $C > 0$ such that $F(\mathscr{S})$ lies in $[0, C]^k$.

As with the Poisson Lotka-Volterra process, stochastic models of interacting populations without immigration always have absorbing states $\mathscr{S}_0 \subset \mathscr{S}$ corresponding to the loss of one or more populations. Hence, we restrict our attention to models which satisfy the following standing hypothesis:

Hypothesis 2.2 *The state space $\mathscr{S}$ can be written $\mathscr{S} = \mathscr{S}_0 \cup \mathscr{S}_+$, where*

- *$\mathscr{S}_0$ is a closed subset of $\mathscr{S}$;*
- *$\mathscr{S}_0$ and $\mathscr{S}_+$ are positively F-invariant, i.e $F(\mathscr{S}_0) \subseteq \mathscr{S}_0$ and $F(\mathscr{S}_+) \subseteq \mathscr{S}_+$;*
- *the set $\mathscr{S}_0$ is assumed to be absorbing for the random perturbations:*

$$p^\varepsilon(x, \mathscr{S}_+) = 0, \text{ for all } \varepsilon > 0, \ x \in \mathscr{S}_0. \tag{6}$$

- *absorption occurs in finite time with probability one:*

$$\mathbb{P}\left[X_t^\varepsilon \in \mathscr{S}_0 \text{ for some } t \geq 1 | X_0^\varepsilon = x\right] = 1$$

for all $x \in \mathscr{S}$ and $\varepsilon > 0$.

The final bullet point implies that extinction of one or more species is inevitable in finite time. For example, Proposition 1 implies this hypothesis for Poisson Lotka-Volterra processes whenever F is pre-compact.

Despite this eventual absorption, the process X^ε may spend exceptionally long periods of time in the set $\mathscr{S}_+$ of transient states provided that $\varepsilon > 0$ is sufficiently small. This "metastable" behavior may correspond to long-term persistence of an endemic disease, long-term coexistence of interacting species as in the case of the Poisson Lotka-Volterra process, or maintenance of a genetic polymorphism. One approach to examining these metastable behaviors are quasi-stationary distributions which are invariant distributions when the process is conditioned on non-absorption.

Definition 2.1 A probability measure μ_ε on $\mathscr{S}_+$ is a *quasi-stationary distribution (QSD)* for p^ε provided there exists $\lambda_\varepsilon \in (0, 1)$ such that

$$\int_{\mathscr{S}_+} p^\varepsilon(x, \Gamma)\mu_\varepsilon(dx) = \lambda_\varepsilon \mu_\varepsilon(\Gamma) \text{ for all Borel sets } \Gamma \subset \mathscr{S}_+.$$

Equivalently, dropping the ε superscript and subscripts, a QSD μ satisfies the identity

$$\mu(\Gamma) = \mathbb{P}_\mu [X_t \in \Gamma \mid X_t \in \mathscr{S}_+] \text{ for all } t,$$

where $\mathbb{P}_\mu$ denotes the law of the Markov chain $\{X_t\}_{t=0}^\infty$, conditional to X_0 being distributed according to μ.

In the case that the Markov chain has a finite number of states and P is the transition matrix (i.e. $P_{ij} = p(i, \{j\})$), [15] showed that the QSD is given by $\mu(\{i\}) = \pi_i$ where π is the normalized, dominant left eigenvector of the matrix Q given by removing the rows and columns of P corresponding to extinction

states in $\mathscr{S}_0$. In this case, λ is the corresponding eigenvalue of this eigenvector. For the Poisson Lotka-Volterra processes in which the unperturbed dynamic F is pre-compact, Proposition 6.1 from [20] implies the existence of QSDs for these processes. Examples of these QSDs for these processes are shown in Figs. 2, 3, and 4. More generally, the existence of QSDs has been studied extensively by many authors as reviewed by [39].

What do these QSD's and λ tell us about the behavior of the stochastic process? From the perspective of metastability, QSDs often exhibit the following property:

$$\mu(\Gamma) = \lim_{t \to +\infty} \mathbb{P}[X_t \in \Gamma \mid X_t \in \mathscr{S}_+, X_0 = x]$$

where the limit exists and is independent of the initial state $x \in \mathscr{S}_+$. In words, the QSD describes the probability distribution of X_t, conditioned on non-extinction, far into the future. Hence, the QSD provides a statistical description of the meta-stable behavior of the process. The eigenvalue, λ provides information about the length of the metastable behavior of X_t. Specifically, given that the process is following the QSD (e.g. X_0 is distributed like μ), and λ equals the probability of persisting in the next time step. Thus, the mean time to extinction is $\frac{1}{1-\lambda}$. [22] call $\frac{1}{1-\lambda}$, the "intrinsic mean time to extinction" and, convincingly, argue that it is a fundamental statistic for comparing extinction risk across stochastic models.

2.2 Positive Attractors, Intrinsic Extinction Risk, and Metastability

When the habitat size is sufficiently large i.e. ε is small, there is a strong relationship between the existence of attractors in $\mathscr{S}_+$ (i.e. "positive" attractors) for the unperturbed system F and the quasi-stationary distributions of X^ε. This relationship simultaneously provides information about the metastable behavior of the stochastic model and intrinsic probability of extinction, $1 - \lambda_\varepsilon$. To make this relationship mathematically rigorous, we need to strengthen Hypotheses 2.1 and 2.2. [20] presents two ways to strengthen these hypothesis. We focus on their large deviation approach as it is most easily verified. This approach requires identifying a *rate function* $\rho : \mathscr{S} \times \mathscr{S} \to [0, \infty]$ that describes the probability of a large deviation between F and X^ε. That is, for a sufficiently small neighborhood U of a point y, the rate function should have the property

$$\mathbb{P}[X_{t+1}^\varepsilon \in U | X_t^\varepsilon = x] \approx \exp(-\rho(x, y)/\varepsilon).$$

Hypothesis 2.3 provides the precise definition and desired properties of ρ.

Hypothesis 2.3 *There exists a* rate function $\rho : \mathscr{S} \times \mathscr{S} \to [0, +\infty]$ *such that*

(i) ρ *is continuous on* $\mathscr{S}_+ \times \mathscr{S}$,

(ii) $\rho(x, y) = 0$ if and only if $y = F(x)$,

(iii) for any $\beta > 0$,

$$\inf \{\rho(x, y) : \ x \in \mathscr{S}, \ y \in \mathscr{S}, \ \|F(x) - y\| > \beta\} > 0, \tag{7}$$

(iv) for any open set U, there is the lower bound

$$\liminf_{\varepsilon \to 0} \varepsilon \log p^{\varepsilon}(x, U) \geq - \inf_{y \in U} \rho(x, y) \tag{8}$$

that holds uniformly for x in compact subsets of $\mathscr{S}_+$ whenever U is an open ball in $\mathscr{S}$. Additionally, for any closed set C, there is the uniform upper bound

$$\limsup_{\varepsilon \to 0} \sup_{x \in \mathscr{S}} \varepsilon \log p^{\varepsilon}(x, C) \leq - \inf_{y \in C} \rho(x, y). \tag{9}$$

Equations (7) and (9), in particular, imply that Hypothesis 2.1 holds. Furthermore, as $\mathscr{S}_0$ is absorbing, equation (8) implies that $\rho(x, y) = +\infty$ for all $x \in \mathscr{S}_0, y \in \mathscr{S}_+$. Identifying the rate function ρ typically requires making use of the Gärtner-Ellis theorem [16, Theorem 2.3.6] which provides large deviation estimates for sums of independent random variables. Example 3 below describes how this theorem was used for the Poisson Lotka-Volterra processes.

We strengthen Hypothesis 2.2 as follows:

Hypothesis 2.4 For any $c > 0$, there exists an open neighborhood V_0 of $\mathscr{S}_0$ such that

$$\lim_{\varepsilon \to 0} \inf_{x \in V_0} \varepsilon \log p^{\varepsilon}(x, \mathscr{S}_0) \geq -c. \tag{10}$$

Equation (10) implies that

$$\mathbb{P}[X_{t+1}^{\varepsilon} \in \mathscr{S}_0 | X_t \in V_0] \geq \exp(-c/\varepsilon)$$

for $\varepsilon > 0$ sufficiently small. Namely, the probability of absorption near the boundary, at most, decays exponentially with habitat size. The following example discusses why these stronger hypotheses hold for the Poisson Lotka-Volterra process.

Example 3 (Return of the Poisson Lotka-Volterra Process) Using the Gärtner-Ellis theorem [16, Theorem 2.3.6], Faure and Schreiber [20, Proposition 6.4] showed that $\rho(x, y) = \sum_i y_i \log \frac{y_i}{F_i(x)} - y_i$ is the rate function for any Poisson processes with mean $F : \mathbb{R}_+^k \to \mathbb{R}_+^k$ including the Poisson Lotka-Volterra Process of Example 1. To see why Hypothesis 4 holds for the Poisson Lotka-Volterra process, assume x is such that $x_i \leq \delta$ for some $\delta > 0$ and i. Then

$$\varepsilon \log \mathbb{P}[X_{t+1}^{\varepsilon} \in \mathscr{S}_0 | X_t = x] \geq \varepsilon \log \mathbb{P}[X_{i,t+1}^{\varepsilon} = 0 | X_t = x]$$
$$= \varepsilon \log \exp(-F_i(x)/\varepsilon) = -F_i(x)$$

Hence, for any $c > 0$, choose $\delta > 0$ sufficiently small to ensure that for all i, $F_i(x) \leq c$ whenever $x_i \leq \delta$. In which case, choosing $V_0 = \{x \in \mathbb{R}_+^k : x_i \leq \delta$ for some $i\}$ satisfies (10).

As many discrete distributions are used in models with demographic stochasticity (e.g. negative binomial, mixtures of Bernoullis and negative binomials), an important open problem is the following:

Problem 1 For which types of random perturbations of an ecological model F do Hypotheses 3 and 4 hold?

To relate QSDs to the attractors of the deterministic dynamics, we recall the definition of an attractor and weak* convergence of probability measures. A compact set $A \subset \mathscr{S}$ is an attractor for F if there exists a neighborhood U of A such that (i) $\cap_{n \geq 1} F^n(U) = A$ and (ii) for any open set V containing A, $F^n(U) \subset V$ for some $n \geq 1$. A weak* limit point of a family of probability measures $\{\mu_\varepsilon\}_{\varepsilon > 0}$ on $\mathscr{S}$ is a probability measure μ^0 such that there exists a sequence $\varepsilon_n \downarrow 0$ satisfying

$$\lim_{n \to \infty} \int h(x) \mu^{\varepsilon_n}(dx) = \int h(x) \mu^0(dx)$$

for all continuous functions $h : \mathscr{S} \to \mathbb{R}$. Namely, the expectation of any continuous function with respect to μ^{ε_n} converges to its expectation with respect to μ^0 as $n \to \infty$. The following theorem follows from [20, Lemma 3.9 and Theorem 3.12].

Theorem 2.5 *Assume Hypotheses 2.3 and 2.4 hold. Assume for each $\varepsilon > 0$, there exists a QSD μ_ε for X^ε. If there exists a positive attractor $A \subset \mathscr{S}_+$, then*

- *there exists a neighborhood V_0 of $\mathscr{S}_0$ such that all weak* limit points μ^0 of $\{\mu_\varepsilon\}_{\varepsilon > 0}$ are F-invariant and $\mu^0(V_0) = 0$, and*
- *there exists $c > 0$ such that*

$$\lambda_\varepsilon \geq 1 - e^{-c/\varepsilon} \text{ for all } \varepsilon > 0. \tag{11}$$

Alternatively, assume that $\mathscr{S}_0$ is a global attractor for the dynamics of F. Then any weak*-limit point of $\{\mu_\varepsilon\}_{\varepsilon > 0}$ is supported by $\mathscr{S}_0$.

Theorem 2.5 implies the existence of a positive attractor of the deterministic dynamics ensures the stochastic process exhibits metastable behavior for large habitat size, and the probability of extinction $1 - \lambda_\varepsilon$ decreases exponentially with habitat size. Equivalently, the mean time to extinction $1/(1 - \lambda_\varepsilon)$ increases exponential with habitat size. These conclusions are illustrated in Fig. 2 with a one-dimensional Poisson Lotka-Volterra process (the Poisson Ricker process described below in Example 4).

Even if F has no positive attractors, $\mathscr{S}_0$ may not be a global attractor as there might be an unstable invariant set in $\mathscr{S}_+$. For example, single species models with positive feedbacks can have an uncountable number of unstable periodic orbits despite almost every initial condition going to extinction [46]. Hence, the necessary
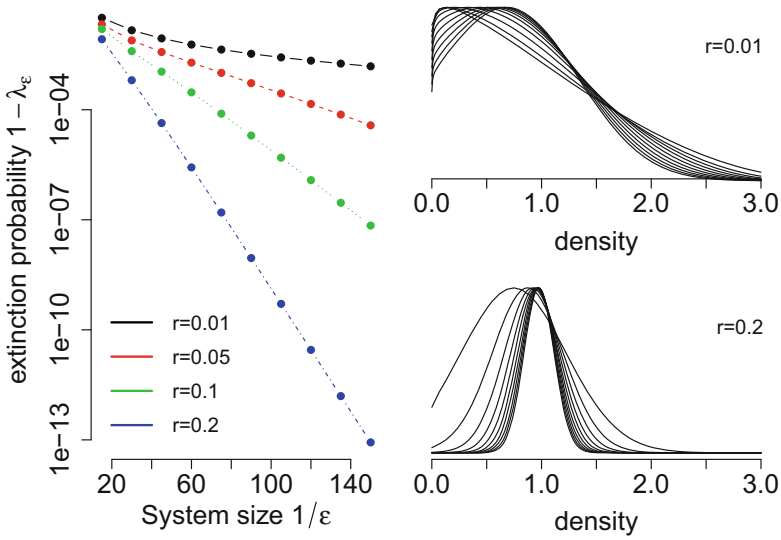
Fig. 2 Extinction probabilities and QSDs for the Poisson Ricker process described in Example 4. In the left panel, the "intrinsic" extinction probability $1 - \lambda_\varepsilon$ plotted as function of the habitat size $1/\varepsilon$ and for different r values. In the right panels, the QSDs plotted for a range of habitat sizes and two r values

and sufficient conditions for metastability in Theorem 2.5 are not equivalent. However, if F has no positive attractors, one can show that all points in $\mathscr{S}_+$ can with arbitrarily small perturbations be "forced" to $\mathscr{S}_0$ [47, 48]. Hence, this raises the following open problem.

Problem 2 If F has no positive attractors, are all the weak*-limit points of the QSDs supported by the extinction set $\mathscr{S}_0$?

While the methodology used to prove Theorem 2.5 provides an explicit expression for $c > 0$, this expression is fairly abstract and only provides a fairly crude lower bound. This suggests the following questions which, if solved, may provide insights into how extinction probabilities depend on the nature of the nonlinear feedbacks within and between populations and the form of demographic stochasticity.

Problem 3 If F has positive attractors, when does the limit

$$\lim_{\varepsilon \to 0} -\frac{1}{\varepsilon} \log(1 - \lambda_\varepsilon) =: c$$

exist? If the limit exists, under what circumstances can we derive explicit expressions for c? or good explicit lower bounds for c?

Theorem 2.5 only ensures that the metastable dynamics concentrates on an invariant set for the deterministic dynamics. However, it is natural to conjecture
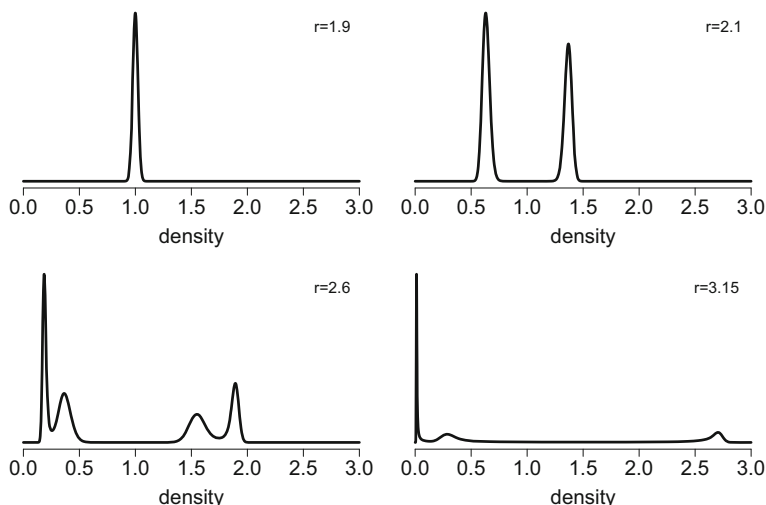
Fig. 3 QSDs for the stochastic Ricker model (see Example 4) for r values where $F(x) = x \exp(r(1-x))$ has a stable periodic orbit. Habitat size $1/\varepsilon$ is $2,500$

that the QSDs μ_ε should concentrate on the positive attractors of F. These positive attractors, however, may coexist with complex unstable behavior. For example, the Ricker equation $F(x) = x \exp(r(1-x))$ can have a stable periodic orbit coexisting with an infinite number of unstable periodic orbits (e.g. the case of a stable period 3 orbit as illustrated in Fig. 3).

To identify when this intuition is correct, a few definitions from dynamical systems are required. For $x \in \mathscr{S}$, let $\omega(x) = \{y : \text{there exists } n_k \to \infty \text{ such that } \lim_{k\to\infty} F^{n_k}(x) = y\}$ be the ω-*limit set for* x and $\alpha(x) = \{y : \text{there exist } n_k \to \infty \text{ and } y_k \in \mathscr{S} \text{ such that } F^{n_k}(y_k) = x \text{ and } \lim_{k\to\infty} y_k = y\}$ be the α-*limit set for* x. Our assumption that F is precompact implies that there exists a global attractor given by the compact, F-invariant set $\Lambda = \cap_{n\geq 0} F^n(\mathscr{S})$. For all $x \in \Lambda$, $\omega(x)$ and $\alpha(x)$ are compact, non-empty, F-invariant sets. A *Morse decomposition* of the dynamics of F is a collection of F-invariant, compact sets $K_1, \ldots, K_\ell$ such that

- K_i is isolated i.e. there exists a neighborhood of K_i such that it is the maximal F-invariant set in the neighborhood, and
- for every $x \in \Lambda \setminus \cup_{i=1}^{\ell} K_i$, there exist $j < i$ such that $\alpha(x) \subset K_j$ and $\omega(x) \subset K_i$.

Replacing the invariant sets K_i by points, one can think of F being gradient-like as all orbits move from lower indexed invariant sets to higher indexed invariant sets. Finally, recall that a compact invariant set K is *transitive* if there exists an $x \in K$ such that $\{x, F(x), F^2(x), \ldots\}$ is dense in K. Faure and Schreiber [20, Theorem 2.7, Remark 2.8, and Proposition 5.1] proved the following result about QSDs not concentrating on the non-attractors of F. The assumptions of this theorem can be verified for many ecological models.

Theorem 2.6 *Assume Hypotheses 2.3 and 2.4 hold. Let $K_1, \ldots, K_\ell$ be a Morse decomposition for F such $K_j, \ldots, K_\ell$ are attractors. If*

- $K_i \subset \mathscr{S}_+$ or $K_i \subset \mathscr{S}_0$ for each i,
- $K_i \subset \mathscr{S}_+$ for some $i \geq j$, and
- K_i with $i \leq j - 1$ is transitive whenever $K_i \subset \mathscr{S}_+$,

then any weak*-limit point of $\{\mu_\varepsilon\}_{\varepsilon > 0}$ is F-invariant and is supported by the union of attractors in $\mathscr{S}_+$.

For random perturbations of deterministic models without absorbing states (e.g. models accounting for immigration or mutations between genotypes), the work of [31] and [21] can be used to show that the stationary distributions often concentrate on a unique attractor. However, due to the singularity of the rate function ρ along the extinction set $\mathscr{S}_0$, the approach used by these authors doesn't readily extend to the stochastic models considered here. This raises the following open problem:

Problem 4 If F has multiple, positive attractors, under what conditions do the QSDs μ_ε concentrate on a unique one of these positive attractors as $\varepsilon \downarrow 0$?

Lets apply some of these results to the Poisson Lotka-Volterra processes from Example 1.

Example 4 (The Ricker Model) The simplest of Poisson Lotka-Volterra processes is the stochastic Ricker model for a single species where $F(x) = x \exp(r(1 - x))$ with $r > 0$. [33] proved that for an open and dense set of $r > 0$ values, the Ricker map has a Morse decomposition consisting of a finite number of unstable, intransitive sets (more specifically, hyperbolic sets) and a unique stable period orbit $\{p, F(p), \ldots, F^T(p)\}$. As the stable periodic orbit is the only attractor, Theorem 2.6 implies the following result.

Corollary 1 *Consider the Ricker process with $r > 0$ such that $F(x) = x \exp (r(1-x))$ has the aforementioned Morse decomposition. Then any weak*-limit point of $\{\mu_\varepsilon\}_{\varepsilon > 0}$ is supported by the unique stable periodic orbit $\{p, F(p), \ldots, F^T(p)\}$.*

Figure 3 illustrates this corollary: QSDs concentrating on the stable periodic orbit of period 1 for $r = 1.9$, period 2 for $r = 2.1$, period 4 for $r = 2.6$, and period 3 for $r = 3.15$. Remarkably, in the case of the stable orbit of period 3, there exists an infinite number of unstable periodic orbits which the QSDs do not concentrate on. We note that [27, 32, 42] proved similar results to Corollary 1 using inherently one dimensional methods.

Example 5 (Revenge of the Poisson Lotka-Volterra Processes) For higher dimensional Lotka-Volterra processes, we can use properties of Lotka-Volterra difference equations in conjunctions with Theorems 2.5 and 2.6 to derive two algebraically verifiable results for the stochastic models. First, if the deterministic map $F = (F_1, \ldots, F_k)$ with $F_i(x) = x_i \exp(\sum_j A_{ij}x_j + r_i)$ is pre-compact and there is no internal fixed point (i.e. there is no strictly positive solution to $Ax = -r$), then [26] proved that the boundary of the positive orthant is a global attractor. Hence, Theorem 2.5 implies the following corollary.

Corollary 2 *Let X^ε be a Poisson Lotka-Volterra process such that F is pre-compact and admits no positive fixed point. Then any weak*-limit point of $\{\mu_\varepsilon\}_{\varepsilon>0}$ is supported by $\mathscr{S}_0$, the boundary of the positive orthant of $\mathbb{R}^k_+$.*

On the other hand, [26] derived a simple algebraic condition which ensures that the deterministic dynamics of F has a positive attractor. Namely, there exist $p_i > 0$ such that

$$\sum_i p_i \left(\sum_j A_{ij} x_j^* + r_i \right) > 0 \tag{12}$$

for any fixed point x^* on the boundary of the positive orthant. Hence, Theorem 2.5 implies the following corollary.

Corollary 3 *Assume $F = (F_1, \ldots, F_k)$ with $F_i(x) = x_i \exp(\sum_j A_{ij} x_j + r_i)$ is pre-compact and satisfies (12) for some choice of $p_i > 0$. If X^ε is the corresponding Poisson Lotka-Volterra process, then any weak*-limit point of $\{\mu_\varepsilon\}_{\varepsilon>0}$ is supported by A where $A \subset \mathscr{S}_+$ is the global, positive attractor for F. Moreover, there exists $c > 0$ such that $\lambda_\varepsilon \geq 1 - \exp(c/\varepsilon)$ for all $\varepsilon > 0$ sufficiently small.*

Figure 4 illustrates the convergence of the QSDs to the attractor of F for a Lotka-Volterra process of two competing species. Even for populations of only hundreds of individuals ($\varepsilon = 0.01$), this figure illustrates that species can coexist for tens of thousands of generations despite oscillating between low and high densities, a key signature of the underlying deterministic dynamics. However, only at much larger habitat sizes (e.g. $1/\varepsilon = 1,000,000$) do the metastable behaviors clearly articulate the underlying deterministic complexities.

3   Environmental Stochasticity

To understand how environmental fluctuations, in and of themselves, influence population dynamics, we shift our attention to models for which the habitat size is sufficiently large that one can approximate the population state by a continuous variable. Specifically, let $X_t \in \mathbb{R}^k_+$ denote the state of the population or community at time t. The components of $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{k,t})$ corresponds to densities or frequencies of subpopulations. To account for environmental fluctuations, let $\mathscr{E} \subset \mathbb{R}^m$ (for some m) be a compact set representing all possible environmental states e.g. all possible precipitation and temperature values. I assume that $E_{t+1} \in \mathscr{E}$ represents the environmental state of the system over the time interval $(t, t + 1]$ that determines how the community state changes over that time interval. If the population or community state X_{t+1} depends continuously on E_{t+1} and X_t, then

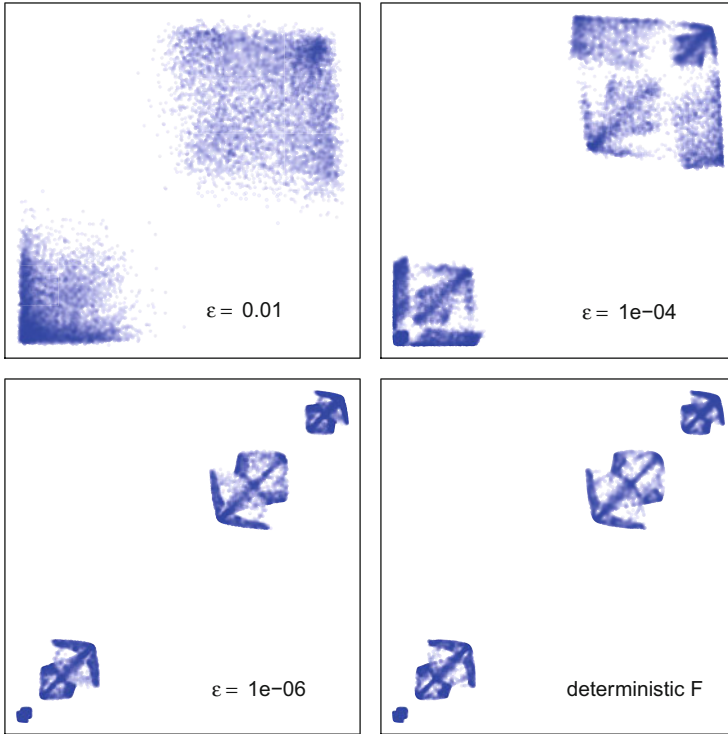$$X_{t+1} = F(X_t, E_{t+1}) \tag{13}$$

Fig. 4 Numerically estimated QSDs for a Poisson Lotka-Volterra process with two competing species, and the global attractor of the deterministic map $F_i(x) = x_i \exp(\sum_j A_{ij} x_j x + r_i)$. The stochastic and deterministic processes were simulated for $50,000$ time steps and the last $17,500$ time steps are plotted in the x_1–x_2 plane. Parameters: A is the matrix with rows $(-0.2, -0.01)$, $(-0.01, -0.2)$ and $r = (2.71, 2.71)$

for a continuous map $F : \mathbb{R}_+^k \times \mathscr{E} \to \mathbb{R}_+^k$. If the E_t are random variables, then (13) is known as a *continuous, random dynamical system*. [2] provides a thorough overview of the general theory of these random dynamical systems.

To state the main hypotheses about (13), recall that a sequence of random variables, $E_1, E_2, \ldots,$ is *stationary* if for every pair of non-negative integers t and s, $E_1, \ldots, E_t$ and $E_{1+s}, \ldots, E_{t+s}$ have the same distribution. The sequence is *ergodic* if with probability one all realizations of the sequence have the same asymptotic statistical properties e.g. time averages (see, e.g., [17] for a more precise definition).

Hypothesis 3.7 $E_1, E_2, \ldots$ *are an ergodic and stationary sequence of random variables taking value in $\mathscr{E}$. Let π be the stationary distribution of this sequence i.e. the probability measure π on $\mathscr{E}$ such that $\mathbb{P}[E_t \in B] = \pi(B)$ for all Borel sets $B \subset \mathscr{E}$.*

This hypothesis is satisfied for a diversity of models of environmental dynamics. For example, E_t could be given by a finite state Markov chain on a finite number

of environmental states, say $e_1, e_2, \ldots, e_m \in E$ (e.g. wet and cool, wet and hot, dry and cool, dry and hot) with transition probabilities $p_{ij} = \mathbb{P}[E_{t+1} = e_j | E_t = e_i]$. If the transition matrix $P = (p_{ij})_{i,j}$ is aperiodic and irreducible, then E_t is asymptotically ergodic and stationary. Alternatively, E_t could be given by a sequence of independent and identically distributed random variables or, more generally, an autoregressive process.

Our second hypothesis simply assumes that population densities remain bounded and allows for the possibility of extinction.

Hypothesis 3.8 *There are compact sets $\mathscr{S} \subset \mathbb{R}_+^k$ and $\mathscr{S}_0 \subset \{x \in \mathscr{S} : \prod_i x_i = 0\}$ such that $F : \mathscr{S} \times \mathscr{E} \to \mathscr{S}$, $F : \mathscr{S}_0 \times \mathscr{E} \to \mathscr{S}_0$, and $F : \mathscr{S}_+ \times \mathscr{E} \to \mathscr{S}_+$ where $\mathscr{S}_+ = \mathscr{S} \setminus \mathscr{S}_0$.*

For example, $\mathscr{S}$ may equal $[0, M]^k$ where M is the maximal density of a species or $\mathscr{S}$ may be the probability simplex $\Delta = \{x \in \mathbb{R}_+^k : \sum_i x_i = 1\}$ where $x \in \mathscr{S}$ corresponds to the vector of genotypic frequencies. As in the case of demographic stochasticity, $\mathscr{S}_0$ corresponds to the set where one or more populations have gone extinct. Invariance of $\mathscr{S}_0$ implies that once the population has gone extinct it remains extinct i.e. the "no cats, no kittens" principle. Invariance of $\mathscr{S}_+$ implies that populations can not go extinct in one time step but only asymptotically. This latter assumption is met by most (but not all) models in the population biology literature.

For these stochastic difference equations, there are several concepts of "persistence" which are reviewed in [49]. Here, we focus on the "typical trajectory" perspective. Namely, "how frequently does the typical population trajectory visit a particular configuration of the population state space far into the future?" The answer to this question is characterized by *empirical measures* for X_t:

$$\Pi_t^x(A) = \frac{\#\{0 \le s \le t : X_s \in A\}}{t + 1}$$

where $X_0 = x$ and A is a Borel subset of $\mathscr{S}$. $\Pi_t(A)$ equals the fraction of time that X_s spends in the set A over the time interval $[0, t]$. Provided the limit exists, the long-term frequency that X_t enters A is given by $\lim_{t\to\infty} \Pi_t^x(A)$. It is important to note that these empirical measures are random measures as they depend on the particular realization of the stochastic process. Figure 5 provides graphical illustrations of empirical measures for a single species model (top row) and a two species model (bottom row). For both models, the empirical measure at time t can be approximated by a histogram describing the frequency X_t spends in different parts (e.g. intervals or hexagons) of the population state space $\mathscr{S}$.

Stochastic persistence corresponds to the typical trajectory spending arbitrarily little time, arbitrarily near the extinction set $\mathscr{S}_0$. More precisely, for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\limsup_{t\to\infty} \Pi_t^x (\{x \in \mathscr{S} : \mathrm{dist}(x, \mathscr{S}_0) \le \delta\}) \le \varepsilon \text{ with probability one for all } x \in \mathscr{S}_+$$
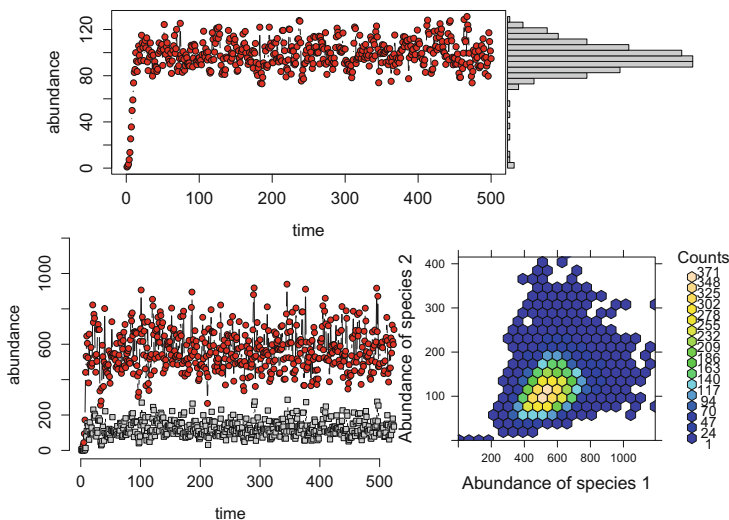
Fig. 5 Visualizing the empirical measures Π_t^x for two models with environmental stochasticity. In top row, the time series of a realization of a stochastic, single species model $X_{t+1} = \frac{\eta_t + 1 X_t}{1 + 0.01 X_t}$ where η_t is a truncated log normal with log-mean $\log 2$ and log-variance 0.01. Histogram to the right of the time series corresponds to $\Pi_{500}^x([a, b])$ for intervals $[a, b]$ of width 10 from 0 to 140. In the bottom row, the time series of a realization of a stochastic predator-prey model $X_{1,t+1} = X_{1,t} \exp(\eta_{t+1} - 0.001 X_{1,t} - 0.001 X_{2,t}), X_{2,t+1} = 0.5 X_{1,t}(1 - \exp(-0.001 X_{2,t}))$ here η_t is a truncated log normal with log-mean $\log 2$ and log-variance 0.04. To the right of the time series, the time spent in each colored hexagon in $\mathbb{R}_+^2$ is shown. $\Pi_{500}^x(H)$ for one of the hexagons $H \subset \mathbb{R}_+^2$ equals the count divided by 500. The truncated normals are used for these models to ensure that dynamics remain in a compact set $\mathscr{S}$

where $\text{dist}(x, S_0) = \min_{y \in \mathscr{S}_0} \|x - y\|$. In contrast to the deterministic notions of uniform persistence or permanence, stochastic persistence allows for trajectories to get arbitrarily close to extinction and only requires the frequency of these events are very small. One could insist that the trajectories never get close to extinction. However, such a definition is too strict for any model where there is a positive probability of years where the population is tending to decline e.g. the models discussed in Sect. 7. Regarding this point, [11] wrote

"This criterion... places restrictions on the expected frequency of fluctuations to low population levels. Given that fluctuations in the environment will continually perturb population densities, it is to be expected that any nominated population density, no matter how small, will eventually be seen. Indeed this is the usual case in stochastic population models and is not an unreasonable postulate about the real world. Thus a reasonable persistence criterion cannot hope to do better than place restrictions on the frequencies with which such events occur."

Conditions for verifying stochastic persistence appear in papers by [5, 44, 49, 50]. As the results by [44] are the most general, we focus on them. We begin with single species models and then expand to multi-species models.

3.1 Single Species Models

Consider a single species for which an individual can be in one of k states. For example, these states may correspond to age where k is the maximal age, living in one of k spatial locations or "patches", discrete behavioral states that an individual can move between, different genotypes in an asexual population coupled by mutation, or finite number of developmental stages or size classes. $X_{i,t}$ corresponds to population density of individuals in state i and $X_t = (X_{1,t}, \ldots, X_{k,t})$ is the population state. The population state is updated by multiplication by a $k \times k$ matrix $A(X_t, E_{t+1})$ dependent on the population and environmental state:

$$X_{t+1} = A(X_t, E_{t+1})X_t =: F(X_t, E_{t+1}). \tag{14}$$

Assume $A(X, E)$ satisfies the following hypothesis.

Hypothesis 3.9 *A is a continuous mapping from $\mathscr{S} \times \mathscr{E}$ to non-negative $k \times k$ matrices. Furthermore, there exists a non-negative, primitive matrix B such that $A(x, E)$ has the same sign structure as B for all x, E i.e. the i–j-th entry of $A(x, E)$ is positive if and only if the i–j-th entry of B is positive.*

The primitivity assumption implies that there is a time, T, such that after T time steps, individuals in every state contribute to individuals in all other states. Specifically, $A(X_{T-1}, E_T)A(X_{T-2}, E_{T-1})\ldots A(X_0, E_1)$ has only positive entries for any $X_0, \ldots, X_{T-1} \in \mathscr{S}$ and $E_1, \ldots, E_T \in \mathscr{E}$. This assumption is met for most models.

   To determine whether or not the population has a tendency to increase or decrease when rare, we can approximate the dynamics of (14) when $X_0 \approx 0$ by the linearized system

$$Z_{t+1} = B_{t+1}Z_t \text{ where } Z_0 = X_0 \text{ and } B_{t+1} = A(0, E_{t+1}). \tag{15}$$

Iterating this matrix equation gives

$$Z_t = B_t B_{t-1} B_{t-2} \ldots B_2 B_1 Z_0.$$

Proposition 3.2 from [45] and Birkhoff's ergodic theorem implies there is a quantity r, the dominant Lyapunov exponent, such that

$$\lim_{t \to \infty} \frac{1}{t} \log \|Z_t\| = r \text{ with probability one}$$

whenever $Z_0 \in \mathbb{R}_+^k \setminus \{0\}$. Following [7–9], we call r the *low-density per-capita growth* of the population. When $r > 0$, Z_t with probability one grows exponentially and we would expect the population state X_t to increase when rare. Conversely when $r < 0$, Z_t with probability one converges to 0. Consistent with these predictions from the linear approximation, Roth and Schreiber [44, Theorems 3.1,5.1] proved the following result.

Theorem 3.10 *Assume Hypotheses 3.7 through 3.9 hold with $\mathscr{S}_0 = \{0\}$. If $r > 0$, then (14) is stochastically persistent. If $r < 0$ and $A(0, E) \geq A(X, E)$ for all X, E, then*

$$\lim_{t \to \infty} X_t = 0 \text{ with probability one.}$$

The assumption in the partial converse is a weak form of negative-density dependence as it requires that the best conditions (in terms of magnitude of the entries of A) occurs at low densities. There are cases where this might not be true e.g. models accounting for positive density-dependence, size structured models where growth to the next stage is maximal at low densities.

Example 6 (The Case of the Bay Checkerspot Butterflies) The simplest case for which Theorem 3.10 applies are unstructured models where $k = 1$. In this case, $B_t = A(0, E_t)$ are scalars and

$$r = \mathbb{E}[\log B_t].$$

The exponential e^r corresponds to the geometric mean of the B_t. By Jensen's inequality, the arithmetic mean $\mathbb{E}[B_t]$ is greater than or equal to this geometric mean e^r, with equality only if B_t is constant with probability one. Hence, environmental fluctuations in the low-density fitnesses B_t reduce r and have a detrimental effect on population persistence.

To illustrate this fundamental demographic principle, we visit a study by [38] on the dynamics of Bay checkerspot butterflies, a critically endangered species. In the 1990s, two populations of this species went extinct in Northern California. The population densities for one of these populations is shown in the left hand side of Fig. 6. Both extinctions were observed to coincide with a change in precipitation variability in the 1970s (right hand side of Fig. 6): the standard deviation in precipitation is approximately 50% higher after 1971 than before 1971.
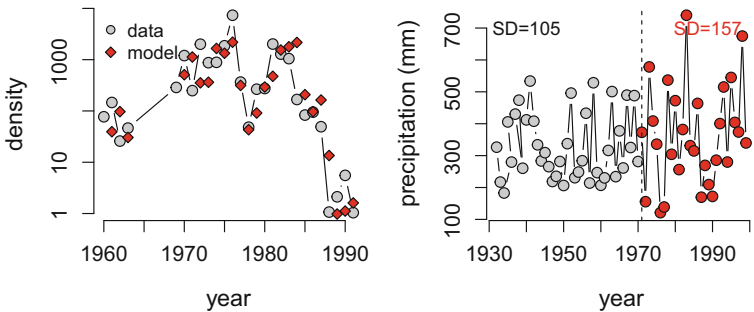

Fig. 6 Checkerspot population dynamics (left) and precipitation (right) from Example 6. Model fit for population dynamics as red diamonds
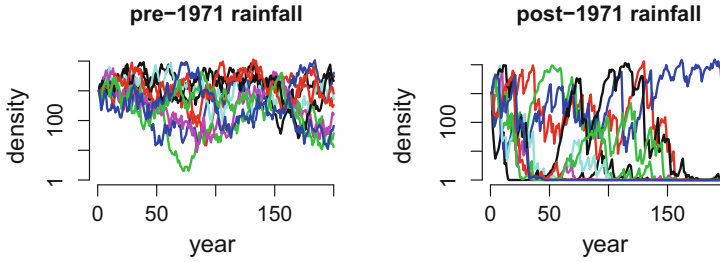
Fig. 7 Simulated checkerspot population dynamics with pre-1971 precipitation data (left) and post-1971 precipitation data (right) from Example 6

To evaluate whether this shift in precipitation variability may have caused the extinction of the checkerspots, [38] developed a stochastic difference equation of the following type

$$n_{t+1} = n_t \exp(a - bn_t + cE_{t+1}^{-2})$$

where E_t is precipitation in year t. Using linear regression on a log-scale yields a model whose fit for one-year predictions are shown as red diamonds in Fig. 6. To compare the pre-1971 and post-1971 population dynamics of the populations, [38] ran their stochastic difference equations with E_t given by independent draws from the corresponding years of precipitation data. The resulting models satisfy all of the assumptions of Theorem 3.10. The model with random draws from the pre-1971 precipitation data yields $r = \mathbb{E}[a + cE_1^{-2}] = 0.04$. Hence, Theorem 3.10 implies stochastic persistence with this form of climatic variability (left hand side of Fig. 7). In contrast, the model with random draws from the post-1971 precipitation data yields $r = -0.049$. Hence, Theorem 3.10 implies the population is extinction bound with this form of climatic variability (right hand side of Fig. 7).

Example 7 (Spatially Structured Populations) To illustrate the application of Theorem 3.10 to structured populations, consider a population in which individuals can live in one of k patches (e.g. butterflies dispersing between heath meadows, pike swimming between the northern and southern basin of a lake, acorn woodpeckers flying between canyons). $X_{i,t}$ is the population density in patch i. Let $E_{t+1} = (E_{1,t+1}, \ldots, E_{k,t+1})$ be the environmental state over $(t, t+1]$ where $E_{i,t}$ be the low-density fitness of individuals in patch i. To account for within-patch competition, let $f_i(X_i, E_i) = E_i/(1 + c_i X_i)$ be the fitness of an individual in patch i where c_i measures the strength of competition within patch i. This fitness function corresponds to the Beverton-Holt model in population biology.

To couple the dynamics of the patches, let d be the fraction of dispersing individuals that go with equal likelihood to any other patch. In the words of Ulysses Everett McGill in *O Brother, Where Art Thou?*

"Well ain't [these patches] a geographical oddity! Two weeks from everywhere!"

Despite this odd geographic regularity, these all-to-all coupling models have proven valuable to understanding spatial population dynamics. Under these assumptions, we get a spatially structured model of the form

$$X_{i,t+1} = (1 - d)f_i(X_{i,t}, E_{i,t+1})X_{i,t} + \frac{d}{k - 1} \sum_{j \neq i} f_j(X_{j,t}, E_{j,t+1})X_{j,t}. \qquad (16)$$

For this model, $A(X, E)$ is the matrix whose i–j-th entry equals $\frac{d}{k-1}f_j(X_{j,t}, E_{j,t+1})$ for $j \neq i$ and $(1 - d)f_i(X_{i,t}, E_{i,t+1})$ for $j = i$.

The low density per-capita growth rate r is the dominant Lyapunov exponent of the random product of the matrices $B_t = A(0, E_t)$. Theorem 3.10 implies this model exhibits stochastic persistence if $r > 0$ and asymptotic extinction with probability one if $r < 0$. In fact, as this spatial model has some special properties (monotonicity and sublinearity), work of Benaïm and Schreiber [5, Theorem 1] implies if $r > 0$, then there is a probability measure m on $\mathscr{S}_+$ such that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} h(X_t) = \int h(x)\, m(dx) \text{ with probability one}$$

for any $x \in \mathscr{S}_+$ and any continuous function $h : \mathscr{S} \to \mathbb{R}$. Namely, for all positive initial conditions, the long-term behavior is statistically characterized by the probability measure m that places no weight on the extinction set. When this occurs, running the model once for sufficiently long describes the long-term statistical behavior for all runs with probability one. The probability measure m corresponds to the marginal of an invariant measure for the stochastic model.

But when is $r > 0$? Finding explicit, tractable formulas for r, in general, appears impossible. However, for sedentary populations ($d \approx 0$) and perfectly mixing populations ($d = 1 - 1/k$), one has explicit expressions for r. In the limit of $d = 0$,

$$r = \max_i \mathbb{E}[\log E_{i,t}]$$

as $f_i(0, E_i) = E_i$. As r varies continuously with d (cf. Benaïm and Schreiber [5, Proposition 3]), it follows that persistence for small d (i.e. mostly sedentary populations) only occurs if $\mathbb{E}[\log E_{i,t}] > 0$. Equivalently, the geometric mean $\exp(\mathbb{E}[\log E_{i,t}])$ of the low-density fitnesses $E_{i,t}$ is greater than one in at least one patch.

When $d = 1 - 1/k$, the fraction of individuals going from any one patch to any other patch is $1/k$. In this case, the model reduces to a scalar model for which

$$r = \mathbb{E}\left[\log \left(\frac{1}{k} \sum_{i=1}^{k} E_{i,t} \right) \right].$$

Namely, e^r is equal to the geometric mean of the spatial means [40]. Applying Jensen's inequality to the outer and inner expressions of r, one gets

$$\log\left(\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}[E_{i,t}]\right) > r > \frac{1}{k}\sum_{i=1}^{k}\mathbb{E}[\log E_{i,t}].$$

Hence, persistence requires that the expected fitness in one patch is greater than one (i.e. $\mathbb{E}[E_{i,t}] > 1$ for some i in the left hand side), but can occur even if all the patches are unable to sustain the population (i.e. $\mathbb{E}[\log E_{i,t}] < 0$ for all i on the right hand side). Hence, local populations which are tending toward extinction (i.e. $\mathbb{E}[\log E_{i,t}] < 0$ in all patches) can persist if they are coupled by dispersal. Even more surprising, [51] shows that stochastic persistence is possible in temporally autocorrelated environments even if $\mathbb{E}[E_{i,t}] < 1$ for all patches.

To better understand how r depends on d, I make raise the following problem which has been proven have an affirmative answer for two-patch stochastic differential equation models by [18].

Problem 5 If $E_{i,t}$ are independent and identically distributed in time and space, then is r an increasing function of d on the interval $(0, 1 - 1/k)$? In particular, if $\mathbb{E}[\log E_{i,t}] < 0 < \mathbb{E}[\log \frac{1}{k}\sum_i E_{i,t}]$, then does there exists a $d^* \in (0, 1-1/k)$ such that the population stochastically persists for $d \in (d^*, 1 - 1/k]$ and goes asymptotically extinct with probability one for $d \in (0, d^*)$?

3.2 Multi-Species Communities

No species is an island as species regularly interact with other species. To account for these interactions, lets extend (14) to account for n species. Within species i, there are k_i states for individuals and $X_{i,t} = (X_{i1,t}, \ldots, X_{ik_i,t})$ is the vector of the densities of individuals in these different states. Then $X_t = (X_{1,t}, \ldots, X_{n,t})$ is the densities of all species in all of their states and corresponds to the community state at time t. Multiplication by a $k_i \times k_i$ matrix $A_i(X_t, E_{t+1})$ updates the state of species i:

$$X_{i,t+1} = A_i(X_t, E_{t+1})X_{i,t} =: F_i(X_t, E_{t+1}) \text{ with } i = 1, 2, \ldots, n. \qquad (17)$$

Assume that each of the A_i satisfy Hypothesis 3.9.

To determine whether each species can increase when rare, consider the scenario where a subset of species are absent from the community (i.e. rare) and the remaining species coexist at an ergodic, stationary distribution μ for (17). Then, as in the single species case, we ask: do the rare species have a tendency to increase or decrease in this community context? Before pursuing this agenda, recall that stationarity means that μ is a probability measure on $\mathscr{S} \times \mathscr{E}$ such that (i) the marginal of μ on $\mathscr{E}$ is π i.e. $\pi(B) = \mu(\mathscr{S} \times B)$ for all $B \subset \mathscr{E}$ and (ii) if X_0, E_0 are drawn randomly from this distribution, then E_t, X_t follows this distribution for all time i.e.

$\mathbb{P}[(X_t, E_t) \in B] = \mu(B)$ for all t and Borel sets $B \subset \mathscr{S} \times \mathscr{E}$. Furthermore, ergodicity means that μ is indecomposable i.e. it can not be written as a convex combination of two other stationary distributions. Due to compactness of $\mathscr{E} \times \mathscr{S}$, stationary distributions always exist see, e.g., Arnold [2, Theorem 1.5.8].

By ergodicity, there exists a set of species $I \subset \{1, 2, \ldots, n\}$ such that μ is only supported by these species i.e. $\mu(\{x \in \mathscr{S} : \|x_i\| > 0 \text{ if and only if } i \in I\} \times \mathscr{E}) = 1$. Suppose $i \notin I$ is one of the species not supported by μ and the sub-community I follows the stationary dynamics i.e. X_0, E_0 is randomly chosen with respect to μ. To determine whether or not species i has a tendency to increase or decrease when introduced at small densities $x_i = (x_{i1}, \ldots, x_{ik_i}) \approx 0$, we can approximate the dynamics of species i with the linearized system

$$Z_{t+1} = B_{t+1}Z_t \text{ where } Z_0 = x_i \text{ and } B_{t+1} = A_i(X_t, E_{t+1}) \tag{18}$$

where X_t, E_t is following the stationary distribution given by μ. Iterating this matrix equation gives

$$Z_t = B_t B_{t-1} B_{t-2} \ldots B_2 B_1 Z_0$$

As before, Proposition 3.2 from [45] and Birkhoff's ergodic theorem implies there is a quantity $r_i(\mu)$ such that

$$\lim_{t \to \infty} \frac{1}{t} \log \|Z_t\| = r_i(\mu) \text{ with probability one.}$$

Lets call $r_i(\mu)$ the *per-capita growth rate of species i when the community is in the stationary state given by μ*. For species $i \in I$ in the sub-community $I, r_i(\mu)$ can be defined in the same manner, but it will always equal zero [44, Proposition 8.19]. Intuitively for species not going extinct or growing without bound, the average per-capita growth rate is zero. In the words of [24],

"a finite world can support only a finite population; therefore, population growth must eventually equal zero."

Using these per-capita growth rates, [44] proved the following theorem.

Theorem 3.11 *Let $\mathscr{S}_0 = \{x \in \mathscr{S} : \prod \|x_i\| = 0\}$. If there exist $p_1, \ldots, p_n > 0$ such that*

$$\sum_i p_i r_i(\mu) > 0 \tag{19}$$

for all ergodic stationary distributions μ supported by $\mathscr{S}_0$, then (17) is stochastically persistent.

The stochastic persistence condition is the stochastic analog of a condition introduced by [25] for ordinary differential equation models. The sum in (19) is effectively only over the missing species as $r_i(\mu) = 0$ for all the species supported

by μ. As the reverse of this condition implies that the extinction set $\mathscr{S}_0$ is an attractor for deterministic models, it is natural to raise the following question:

Problem 6 Let $\mathscr{S}_0 = \{x \in \mathscr{S} : \prod \|x_i\| = 0\}$. If there exist $p_1, \ldots, p_n > 0$ such that

$$\sum_i p_i r_i(\mu) < 0$$

for all ergodic stationary distributions μ supported by $\mathscr{S}_0$, then does it follow that for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\mathbb{P}[\lim_{t \to \infty} \text{dist}(X_t, \mathscr{S}_0) = 0 | X_0 = x] \geq 1 - \varepsilon$$

whenever $\text{dist}(x, \mathscr{S}_0) \leq \delta$?

For stochastic differential equations on the simplex, Benaïm et al. [6, Theorems 4.2,5.1] proved affirmative answers to this problem for systems with small or large levels of noise. In their case, $\mathscr{S}_0$ was shown to be a global attractor with probability one. This stronger conclusion will not hold in general.

We illustrate Theorem 3.11 with applications to competing species and stochastic Lotka-Volterra differences equations. In both examples, the interacting species are unstructured i.e. $k_i = 1$.

Example 8 (Competing Species and the Storage Effect) One of the fundamental principle in ecology is the competitive exclusion principle which asserts that two species competing for a single limiting resource (e.g. space, nutrients) can not coexist at equilibrium. However, many species which appear to be competing for a single resource do coexist. One resolution to this paradox for competing planktonic species was suggested by [28] who wrote

> "The diversity of the plankton [is] explicable primarily by a permanent failure to achieve equilibrium as the relevant external factors changes."

Intuitively, if environmental conditions vary such that each species has a period in which it does better than its competitors, then coexistence should be possible. Understanding exactly when this occurs is the focus of a series of papers by Peter Chesson and his collaborators [7, 11–14]. We illustrate one of the main conclusions from this work using a model from [12].

Consider two competing species with densities $X_t = (X_t^1, X_t^2)$ in year t. Let $E_{i,t}$ be the low-density per-capita reproductive output of species i, $s_i \in (0, 1)$ the probability of adults surviving to the next year, and $f : [0, \infty) \to (0, \infty)$ a continuously differentiable, decreasing function accounting for negative effects of competition on reproduction. If $C_t = E_{1,t}X_{1,t} + E_{2,t}X_{2,t}$ represents the "intensity of competition among the offspring", then we have the following model of competitive interactions

$$X_{i,t+1} = X_{i,t} \underbrace{(E_{i,t+1}f(C_t) + s_i)}_{A_i(X_t, E_{t+1})} \text{ where } C_t = E_{1,t}X_{1,t} + E_{2,t}X_{2,t}. \qquad (20)$$

To ensure that stochastic dynamics eventually enter a compact set $\mathscr{S}$, assume that $\lim_{x\to\infty} f(x) = 0$ and there exists $M > 0$ such that $E_{i,t} \in [0, M]$ for all i and t. The first assumption is satisfied for many models in population biology e.g. $f(x) = \exp(-cx)$ or $\frac{1}{1+cx^b}$ with $c > 0, b > 0$.

To apply Theorem 3.11, we need $p_1, p_2 > 0$ such that $p_1 r_1(\mu) + p_1 r_2(\mu) > 0$ for all ergodic stationary distributions μ supported by $\mathscr{S}_0 = \{x \in S : x_1 x_2 = 0\}$. There are three types of μ to consider: μ supports no species (i.e. $I = \emptyset$), μ only supports species 1 (i.e. $I = \{1\}$), or μ only supports species 2 (i.e. $I = \{2\}$). For μ supported on $\{(0, 0)\} \times \mathscr{E}$ i.e. no species are supported, the persistence condition demands

$$\sum_i p_i r_i(\mu) = \sum_i p_i \mathbb{E}[\log(E_{i,t}f(0) + s_i)] > 0. \tag{21}$$

For μ supported by $\{(x_1, 0) : x_1 > 0\} \times \mathscr{E}$, $r_1(\mu) = 0$ and the persistence criterion requires

$$\sum_i p_i r_i(\mu) = p_2 r_2(\mu) = p_2 \int \log(E_2 f(E_1 X_1) + s_2)\mu(dXdE) > 0. \tag{22}$$

As f is a decreasing function, this condition being satisfied implies

$$\int \log(E_2 f(0) + s_2)\mu(dXdE) = \mathbb{E}[\log(E_{2,t}f(0) + s_2)] > 0.$$

Similarly, for μ supported by $\{(0, x_2) : x_2 > 0\} \times \mathscr{E}$, we need

$$\sum_i p_i r_i(\mu) = p_1 r_1(\mu) = p_1 \int \log(E_1 f(E_2 X_2) + s_1)\mu(dXdE) > 0. \tag{23}$$

which implies

$$\int \log(E_1 f(0) + s_1)\mu(dXdE) = \mathbb{E}[\log(E_{1,t}f(0) + s_1)] > 0.$$

As inequalities (22) and (23) imply inequality (21) for any $p_1, p_2 > 0$, inequalities (22) and (23) imply stochastic persistence. These inequalities correspond to the classical mutual invasibility criterion [55]: if each of the species can increase when rare, the competing species coexist.

To verify whether or not these conditions are satisfied is, in general, a challenging issue. However, [12] developed a formula for the $r_i(\mu)$ when the competition is symmetric. Namely, $s_1 = s_2 =: s$, E_t are independent and identically distributed, and $E_{1,t}, E_{2,t}$ are exchangeable i.e. $P[(E_{1,t}, E_{2,t}) \in B] = P[(E_{2,t}, E_{1,t}) \in B]$ for any Borel $B \subset \mathscr{E} \times \mathscr{E}$. Before describing Chesson's formula, lets examine the dynamics of the deterministic case. Exchangeability and determinism imply there exists a

constant $E > 0$ such that $E_{1,t} = E_{2,t} = E$ for all t. Hence, the deterministic model is given by

$$x_{i,t+1} = x_{i,t} \left(Ef(Ex_{1,t} + Ex_{2,t}) + s \right) \text{ with } i = 1, 2.$$

As $x_{1,t+1}/x_{2,t+1} = x_{1,t}/x_{2,t}$ for all t, all radial lines in the positive orthant are invariant. Provided $Ef(0) + s > 1$ (i.e. each species persists in the absence of competition), there exists a line of equilibria connecting the two axes. Regarding these neutral dynamics, [12] wrote

> "Classically, when faced with a deterministic model of this sort ecologists have concluded that only one species can persist when the likely effects a stochastic environment are taken into account. The reason for this conclusion is the argument that environmental perturbations will cause a random walk to take place in which eventually all but one species becomes extinct."

Dispelling this faulty expectation, [12] derived a formula for the $r_i(\mu)$. To describe this formula, assume inequality (21) holds and μ is an ergodic, stationary distribution supporting species 1. As the E_t are independent in time, μ can be written as a product measure $m \times \pi$ on $\mathscr{S} \times \mathscr{E}$ where π is given by Hypothesis 3.7. Define

$$h(E_1, E_2) = \int \log \left(E_2 f(x_1 E_1) + s \right) m(dx).$$

[12] showed that

$$r_2(\mu) = -\frac{1}{2} \mathbb{E} \left[\int_{E_{1,t}}^{E_{2,t}} \int_{E_{1,t}}^{E_{2,t}} \frac{\partial^2 h}{\partial E_1 \partial E_2} (E_1, E_2) dE_1 dE_2 \right].$$

As f is a decreasing function,

$$\frac{\partial^2 h}{\partial E_1 \partial E_2}(E_1, E_2) = \frac{f'(x_1 E_1) x_1 s}{(E_2 f(x_1 E_1) + s)^2} < 0$$

whenever $s > 0$. Hence, $r_2(\mu) > 0$ provided that $\mathbb{P}[E_{1,t} > E_{2,t}] > 0$ (i.e. there is some variation) and $s > 0$. As this holds for any ergodic μ supporting species 1 and a similar argument yields $r_1(\mu) > 0$ for any ergodic μ supporting species 2, it follows that this symmetric version of the model is stochastically persistent (Fig. 8).

The analysis of this model highlights three key ingredients required for environmental fluctuations to mediate coexistence. First, there must periods of time such that each species has a higher birth rate i.e. $E_{1,t}$ and $E_{2,t}$ vary and are not perfectly correlated. Second, year to year survivorship needs to be sufficiently positive (i.e. $s > 0$ in the model) to ensure species can "store" the gains from one favorable period to the next favorable period. Finally, the increase in fitness due to good conditions for one species is greater in years when those conditions are worse for its competitor i.e. $\frac{\partial^2 h}{\partial E_1 \partial E_2}(E_1, E_2) < 0$. These are the key ingredients of the "storage effect" as introduced by [14].
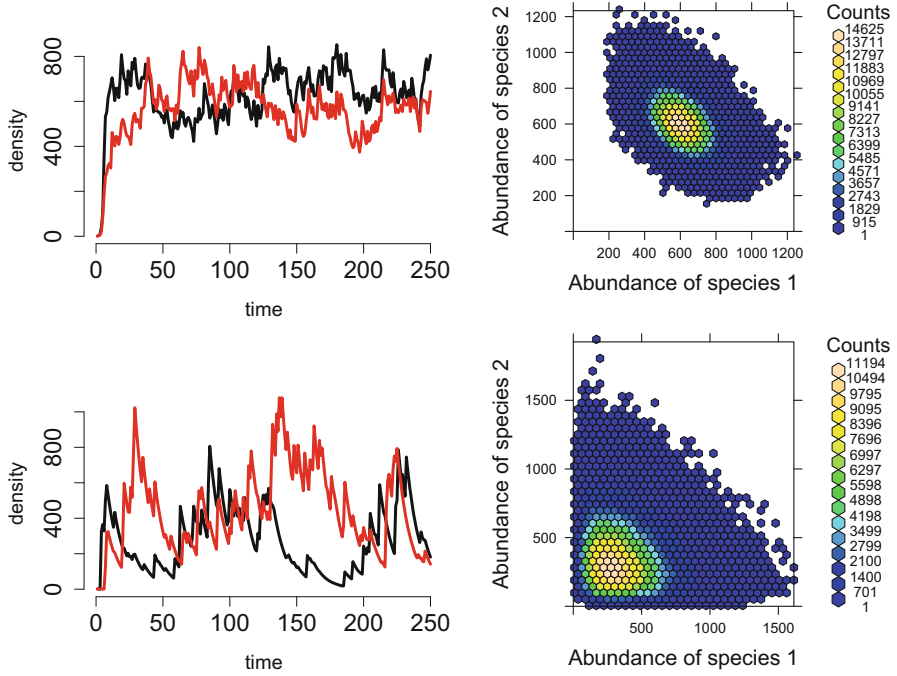
Fig. 8 Stochastic persistence of competing species from Example 8. Two simulations of model (20) with $f(x) = \exp(-0.001x)$, $E_{i,t}$ truncated log normals with log-mean 1 and log-variance 0.25 (upper row) and 25 (lower row). Models were run for $1,000,000$ time steps. Time series on the left show the first 250 time steps. The two dimensional histograms on the right correspond to the last $999,000$ time steps

Example 9 (Stochastic Lotka-Volterra Difference Equations) Previously, we studied the Poisson Lotka-Volterra processes which injected demographic stochasticity into the discrete-time Lotka-Volterra equations (1). Now, we examine the effects of injecting environmental stochasticity into these deterministic equations of n interacting species:

$$X_{i,t+1} = X_{i,t} \exp\left(\sum_{j=1}^{n} A_{ij} X_{j,t} + b_i + E_{i,t}\right) \tag{24}$$

where the matrix $A = (A_{ij})_{i,j}$ describes pairwise interactions between species, $b = (b_1, \ldots, b_n)$ describes the intrinsic rates of growth of each species in the absence of environmental fluctuations, and $E_{i,t}$ describes density-independent fluctuations. [55] used two dimensional versions of (24) to examine niche overlap of competitors in random environments.

The following lemma shows that verifying persistence for these equations reduces to a linear algebra problem. In particular, this lemma implies that the permanence criteria developed by [26] extend to these stochastically perturbed Lotka-Volterra systems.

Lemma 3.1 *Let μ be an ergodic stationary distribution for (24) and $I \subset \{1, \ldots, k\}$ be the species supported by μ i.e. $\mu(\{x \in \mathscr{S} : x^i > 0 \text{ iff } i \in I\} \times \mathscr{E}) = 1$. Define $\beta_i = b_i + \mathbb{E}[E_{i,t}]$. If there exists a unique solution $\hat{x}$ to*

$$\sum_{j \in I} A_{ij}\hat{x}_j + \beta_i = 0 \text{ for } i \in I \text{ and } \hat{x}_i = 0 \text{ for } i \notin I \tag{25}$$

then

$$r_i(\mu) = \begin{cases} 0 & \text{if } i \in I \\ \sum_{j \in I} A_{ij}\hat{x}_j + \beta_i & \text{otherwise.} \end{cases}$$

The following proof of this lemma is nearly identical to the proof given by [50] for the case E_t are independent and identically distributed in time.

Proof Let μ and I be as assumed in the statement of the lemma. We have

$$r_i(\mu) = \sum_{j \in I} A_{ij} \int x_j \, \mu(dxdE) + \beta_i$$

for all i. As $r_i(\mu) = 0$ for all $i \in I$,

$$0 = \sum_{j \in I} A_{ij} \int x_j \, \mu(dxdE) + \beta_i$$

for all $i \in I$. Since we have assumed there is a unique solution $\hat{x}$ to this system of linear equations, it follows that $\int x_i \mu(dxdE) = \hat{x}_i$ for all i and the lemma follows. ∎

This lemma implies that verifying the stochastic persistence condition reduces to finding $p_1, \ldots, p_n > 0$ such that

$$\sum_{i \notin I} p_i \sum_{j \in I} A_{ij}\hat{x}_j + \beta_i > 0$$

for every $I \subset \{1, \ldots, n\}$ and $\hat{x} \in \mathscr{S}_0$ satisfying equation (25). The next example illustrates the utility of this criterion.

Example 10 (Rock-Paper-Scissor Dynamics) The Lotka-Volterra model of rock-paper-scissor dynamics is a prototype for understanding intransitive ecological outcomes [37, 52]. Here, a simple stochastic version of this dynamic is given

by (24) with X_1, X_2, X_3 corresponding to the densities of the rock, paper, and scissors populations, and the matrixes A and b given by

$$A = -1 + \begin{pmatrix} 0 & -\ell_2 & w_3 \\ w_1 & 0 & -\ell_3 \\ -\ell_1 & w_2 & 0 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

with $1 > w_i > 0$ and $\ell_i > 0$. The $-\ell_i$ correspond to a reduction in the per-capita growth rate of the population losing against population i, and w_i corresponds to the increase in the per-capita growth rate of the population winning against population i. Assume that the $E_{i,t}$ in (24) are compactly supported random variables with zero expectation. Under this assumption, β_i as defined in Lemma 3.1 equal 1.

Our assumptions about A and b imply that in pairwise interactions population 1 is excluded by population 2, population 2 is excluded by population 3, and population 3 is excluded by population 1. Hence, there are only four solutions of (25) that need to be considered: $\hat{x} = (0, 0, 0)$, $\hat{x} = (1, 0, 0)$, $\hat{x} = (0, 1, 0)$, and $\hat{x} = (0, 0, 1)$. Hence, verifying stochastic persistence reduces to determining whether there exist positive reals p_1, p_2, p_3 such that

$$p_1 + p_2 + p_3 > 0$$
$$p_1 \cdot 0 + p_2 w_1 - p_3 \ell_1 > 0$$
$$-p_1 \ell_2 + p_2 \cdot 0 + p_3 w_2 > 0$$
$$p_1 w_3 - p_2 \ell_3 + p_3 \cdot 0 > 0$$

where these equation come from evaluating $\sum_i p_i r_i(\mu)$ at ergodic measures corresponding to $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. Solving these linear inequalities implies that there is the desired choice of p_i if and only if $w_1 w_2 w_3 > \ell_1 \ell_2 \ell_3$ i.e. the geometric mean of the fitness payoffs to the winners exceeds the geometric mean of the fitness losses of the losers. Figure 9 illustrates the dynamics of coexistence when $w_1 w_2 w_3 > \ell_1 \ell_2 \ell_3$ and exclusion when $w_1 w_2 w_3 < \ell_1 \ell_2 \ell_3$.

4 Parting Thoughts and Future Challenges

The results reviewed here provide some ways to think about species coexistence or population persistence in the face of uncertainty. In the face of demographic uncertainty, species may coexist for exceptionally long periods of time prior to going extinct. I discussed how this metastable behavior may be predicted by the existence of positive attractors for the underlying deterministic dynamics, in which case the times to extinction increase exponentially with habitat size. Alternatively, in the face
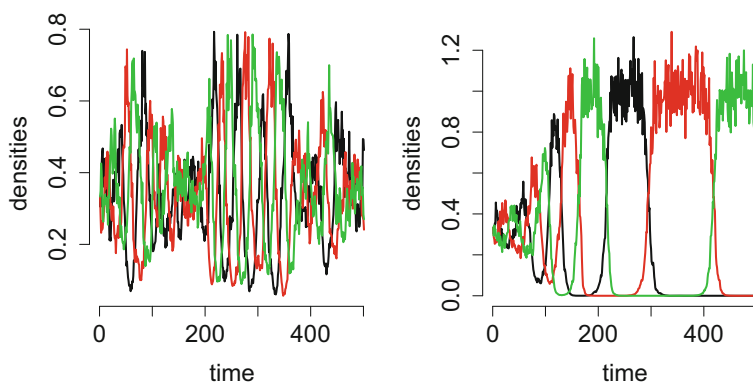
Fig. 9 Stochastic rock-paper-scissor dynamics (Example 10) with stochastic persistence in the left hand panel ($w_i = 0.3 > 0.2 = \ell_i$ for all i) and stochastic exclusion in the right hand panel ($w_i = 0.2 < 0.3 = \ell_i$ for all i)

of environmental stochasticity, species may coexist in the sense of rarely visiting low densities. I discussed how this form of stochastic persistence can be identified by examining species' per-capita growth rates $r_i(\mu)$ when rare. Weighted combinations of these per-capita growth rates can measure to what extent communities move away extinction as one or more species become rare. Despite this progress, many exciting challenges lie ahead.

Many demographic processes and environmental conditions vary continuously in time and are better represented by continuous time models. For continuous-time Markov chains accounting for demographic stochasticity, [36] proved results similar to Theorems 2.5 and 2.6 discussed here. For stochastic differential equations of interacting, unstructured populations in fluctuating environments, there exist some results similar to Theorem 3.11 by [6, 50] and [19]. These stochastic differential equations, however, fail to account for population structure or correlated environmental fluctuations. One step toward temporally correlated environments was recently taken by [4]. They characterized stochastic persistence for continuous-time models of competing species experiencing a finite number of environmental states driven by a continuous-time Markov chain. Generalizing these results to higher dimensional communities and structured populations remains an important challenge. Another exciting possibility is studying stochastic persistence for continuous-time models with stochastic birth or mortality impulses, as often observed in nature.

Biologists often measure continuous traits (e.g. body size or geographical location of an individual) that have important demographic consequences (e.g. larger individuals may produce more offspring and be more likely to survive). Unlike models accounting for discrete traits as considered here, models with continuous traits are infinite-dimensional and, consequently, even stochastic counterparts only accounting for demographic stochasticity correspond to Markov chains on uncount-

able state spaces (see, e.g., [53]). One form of these models, integral projection models (IPMs), have become exceptionally popular in the ecological literature in the past decade as they interface well with demographic data sets (see, e.g., [43] for a recent discussion). Consequently, there is a need for the development of the infinite-dimensional counterparts to the results presented here (see [23] for results for structured populations facing uncorrelated, environmental stochasticity).

For both forms of stochasticity, there are few results for demonstrating that populations are "extinction-prone" (e.g. limiting QSDs being supported by the extinction set in Theorem 2.5 or Benaïm et al. [6, Theorems 4.2,5.1] for stochastic differential equations). No study of persistence or coexistence is complete without understanding this complementary outcome. Hopefully, answers to Problems 2 and 6 will narrow our gap in understanding these outcomes. Furthermore, even when populations aren't extinction prone in the aforementioned sense, extinction is inevitable as all real population are finite. Answers to Problem 3 and their applications to specific models could provide new insights about how feedbacks between nonlinearities and noise determine the "intrinsic" extinction probabilities, quantities of particular importance for conservation biology.

Finally, there is the elephant in the review: what can one say for models accounting for both forms of stochasticity? At this point, all I have to offer is a natural conjecture which combines the results presented here. Namely, let $x_{t+1} = F(x_t, E_{t+1})$ be a random difference equation and $\{X_t^\varepsilon\}_{\varepsilon>0}$ be a family of Markov chains satisfying the environmental dependent versions of Hypotheses 2.3 and 2.4 e.g. the rate function ρ in Hypothesis 2.3 depends on $E \in \mathscr{E}$ as well as $x, y \in \mathscr{S}$. In light of the results presented here, these models lead to the following challenging problem:

Problem 7 Is it true that stochastic persistence of $x_{t+1} = F(x_t, E_{t+1})$ implies the weak* limit points of the QSDS of $\{X_t^\varepsilon\}_{\varepsilon>0}$ are supported by $\mathscr{S}_+$ and $\lambda_\varepsilon \geq 1 - \exp(-c/\varepsilon)$ for some $c > 0$?

I believe there should be an affirmative answer to this question. Namely, stochastic persistence in the face of environmental fluctuations implies long-term, persistent, metastable behavior for communities of interacting populations of finite size, and the extinction probabilities decay exponentially with community "size." Hopefully, this review will inspire work to address this problem as well as for the other challenges posed here.

References

1. P.B. Adler, S.P. Ellner, and J.M. Levine. Coexistence of perennial plants: an embarrassment of niches. *Ecology letters*, 13:1019–1029, 2010.
2. L. Arnold. *Random dynamical systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998. ISBN 3-540-63758-3.
3. A.D. Barbour. Quasi-Stationary Distributions in Markov Population Processes. *Advances in Applied Probability*, 8:296–314, 1976.
4. M. Benaïm and C. Lobry. Lotka Volterra in fluctuating environment or "how good can be bad". *arXiv preprint arXiv:1412.1107*, 2014.
5. M. Benaïm and S.J. Schreiber. Persistence of structured populations in random environments. *Theoretical Population Biology*, 76:19–34, 2009.
6. M. Benaïm, J. Hofbauer, and W. Sandholm. Robust permanence and impermanence for the stochastic replicator dynamics. *Journal of Biological Dynamics*, 2:180–195, 2008.
7. P. Chesson. Multispecies competition in variable environments. *Theoretical Population Biology*, 45(3):227–276, 1994.
8. P. Chesson. General theory of competitive coexistence in spatially-varying environments. *Theoretical Population Biology*, 58:211–237, 2000.
9. P. Chesson. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31:343–366, 2000. ISSN 00664162.
10. P. L. Chesson. Predator-prey theory and variability. *Annu. Rev. Ecol. Syst.*, 9:323–347, 1978.
11. P. L. Chesson. The stabilizing effect of a random environment. *J. Math. Biol.*, 15(1):1–36, 1982.
12. P.L. Chesson. Interactions between environment and competition: how environmental fluctuations mediate coexistence and competitive exclusion. *Lecture Notes in Biomathematics*, 77:51–71, 1988.
13. P.L. Chesson and S. Ellner. Invasibility and stochastic boundedness in monotonic competition models. *Journal of Mathematical Biology*, 27:117–138, 1989.
14. P.L. Chesson and R.R. Warner. Environmental variability promotes coexistence in lottery competitive systems. *The American Naturalist*, 117(6):923, 1981.
15. J.N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing discrete-time finite markov chains. *Journal of Applied Probability*, 2:88–100, 1965.
16. A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*. Applications of Mathematics: Stochastic Modelling and Applied Probability. Springer, 1993.
17. R. Durrett. *Probability: Theory and examples*. Duxbury Press, Belmont, CA, 1996.
18. S.N. Evans, P. Ralph, S.J. Schreiber, and A. Sen. Stochastic growth rates in spatio-temporal heterogeneous environments. *Journal of Mathematical Biology*, 66:423–476, 2013.
19. S.N. Evans, A Hening, and S.J. Schreiber. Protected polymorphisms and evolutionary stability of patch-selection strategies in stochastic environments. *Journal of Mathematical Biology*, 71:325–359, 2015.
20. M. Faure and S. J. Schreiber. Quasi-stationary distributions for randomly perturbed dynamical systems. *Annals of Applied Probability*, 24:553–598, 2014.
21. M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition, 1998. ISBN 0-387-98362-7. Translated from the 1979 Russian original by Joseph Szücs.
22. V. Grimm and C. Wissel. The intrinsic mean time to extinction: a unifying approach to analysing persistence and viability of populations. *Oikos*, 105:501–511, 2004.
23. D. P. Hardin, P. Takáč, and G. F. Webb. Asymptotic properties of a continuous-space discrete-time population model in a random environment. *Journal of Mathematical Biology*, 26:361–374, 1988a. ISSN 0303-6812.
24. G. Hardin. The tragedy of the commons. *Science*, 162:1243–1248, 1968.

25. J. Hofbauer. A general cooperation theorem for hypercycles. *Monatshefte für Mathematik*, 91:233–240, 1981.
26. J. Hofbauer, V. Hutson, and W. Jansen. Coexistence for systems governed by difference equations of Lotka-Volterra type. *Journal of Mathematical Biology*, 25:553–570, 1987.
27. G. Högnäs. On the quasi-stationary distribution of a stochastic Ricker model. *Stochastic Processes and their Applications*, 70:243–263, 1997.
28. G.E. Hutchinson. The paradox of the plankton. *The American Naturalist*, 95:137–145, 1961.
29. P. Jagers. A plea for stochastic population dynamics. *Journal of Mathematical Biology*, 60:761–764, 2010.
30. W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal society of London. Series A*, 115:700–721, 1927.
31. Y. Kifer. *Random perturbations of dynamical systems*. Birkhauser, New York, 1988.
32. F.C. Klebaner, J. Lazar, and O. Zeitouni. On the quasi-stationary distribution for some randomly perturbed transformations of an interval. *Annals of Applied Probability*, 8:300–315, 1998. ISSN 1050-5164.
33. O. S. Kozlovski. Axiom A maps are dense in the space of unimodal maps in the C^k topology. *Annals of Mathematics*, 157:1–43, 2003. ISSN 0003-486X.
34. Thomas G. Kurtz. *Approximation of population processes*, volume 36 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1981. ISBN 0-89871-169-X.
35. A. J. Lotka. *Elements of Physical Biology*. Williams and Witkins, Baltimore, 1925.
36. B. Marmet. Quasi-stationary distributions for stochastic approximation algorithms with constant step size. *arXiv preprint arXiv:1303.7081*, 2013.
37. R. M. May and W. Leonard. Nonlinear aspects of competition between three species. *SIAM Journal of Applied Mathematics*, 29:243–252, 1975.
38. J. F. McLaughlin, J. J. Hellmann, C. L. Boggs, and P. R. Ehrlich. Climate change hastens population extinctions. *Proceeding of the National Academy of Sciences USA*, 99:6070–6074, 2002.
39. S. Méléard and D. Villemonais. Quasi-stationary distributions and population processes. *Probability Surveys*, 9:340–410, 2012.
40. J.A.J. Metz, T.J. de Jong, and P.G.L. Klinkhamer. What are the advantages of dispersing; a paper by Kuno extended. *Oecologia*, 57:166–169, 1983.
41. A. J. Nicholson and V. A. Bailey. The balance of animal populations. *Proceedings of the Zoological Society of London*, pages 551–598, 1935.
42. K. Ramanan and O. Zeitouni. The quasi-stationary distribution for small random perturbations of certain one-dimensional maps. *Stochastic Processes and Applications*, 84:25–51, 1999. ISSN 0304-4149.
43. M. Rees, D.Z. Childs, and S.P. Ellner. Building integral projection models: a user's guide. *Journal of Animal Ecology*, 83:528–545, 2014.
44. G. Roth and S.J. Schreiber. Persistence in fluctuating environments for interacting structured populations. *Journal of Mathematical Biology*, 68:1267–1317, 2014.
45. D. Ruelle. Analycity properties of the characteristic exponents of random matrix products. *Advances in Mathematics*, 32:68–80, 1979. ISSN 0001-8708.
46. S. J. Schreiber. Allee effects, chaotic transients, and unexpected extinctions. *Theoretical Population Biology*, 2003.
47. S. J. Schreiber. Persistence despite perturbations for interacting populations. *Journal of Theoretical Biology*, 242:844–52, 2006.
48. S. J. Schreiber. On persistence and extinction of randomly perturbed dynamical systems. *Discrete and Continous Dynamical Systems B*, 7:457–463, 2007.
49. S. J. Schreiber. Persistence for stochastic difference equations: a mini-review. *Journal of Difference Equations and Applications*, 18:1381–1403, 2012.
50. S. J. Schreiber, M. Benaïm, and K. A. S. Atchadé. Persistence in fluctuating environments. *Journal of Mathematical Biology*, 62:655–683, 2011.

51. S.J. Schreiber. Interactive effects of temporal correlations, spatial heterogeneity and dispersal on population persistence. *Proceedings of the Royal Society B: Biological Sciences*, 277:1907–1914, 2010.
52. S.J. Schreiber and T.P. Killingback. Cycling in space: Persistence of rock-paper-scissor metacommunities. *Theoretical Population Biology*, 86:1–11, 2013.
53. S.J. Schreiber and N. Ross. Individual-based integral projection models: The role of size-structure on extinction risk and establishment success. *Methods in Ecology and Evolution*, in press. URL http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12537/abstract.
54. W. R. Thompson. La théorie mathématique de l'action des parasites entomophages et le facteur du hassard. *Annales Faculte des Sciences de Marseille*, 2:69–89, 1924.
55. M. Turelli. Niche overlap and invasion of competitors in random environments i. models without demographic stochasticity. *Theoretical Population Biology*, 20:1–56, 1981.
56. V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.