Contents lists available at ScienceDirect

Theoretical Population Biology

Persistence of structured populations in random environments

Michel Benaïm [a], Sebastian J. Schreiber [b,c,*]

[a] *Institut de Mathématiques, Université de Neuchâtel, Switzerland*
[b] *Department of Evolution and Ecology, University of California, Davis, USA*
[c] *Center for Population Biology, University of California, Davis, USA*

ARTICLE INFO

ABSTRACT

Environmental fluctuations often have different impacts on individuals that differ in size, age, or spatial location. To understand how population structure, environmental fluctuations, and density-dependent interactions influence population dynamics, we provide a general theory for persistence for density-dependent matrix models in random environments. For populations with compensating density dependence, exhibiting "bounded" dynamics, and living in a stationary environment, we show that persistence is determined by the stochastic growth rate (alternatively, dominant Lyapunov exponent) when the population is rare. If this stochastic growth rate is negative, then the total population abundance goes to zero with probability one. If this stochastic growth rate is positive, there is a unique positive stationary distribution. Provided there are initially some individuals in the population, the population converges in distribution to this stationary distribution and the empirical measures almost surely converge to the distribution of the stationary distribution. For models with overcompensating density-dependence, weaker results are proven. Methods to estimate stochastic growth rates are presented. To illustrate the utility of these results, applications to unstructured, spatially structured, and stage-structured population models are given. For instance, we show that diffusively coupled sink populations can persist provided that within patch fitness is sufficiently variable in time but not strongly correlated across space.

1. Introduction

Many populations of plant and animal species consist of individuals that engage in nonlinear interactions and that experience environmental fluctuations. The effect of these interactions and fluctuations on an individual's fitness can depend on an individual's "state" such as its age, size, physiological condition, or location in space. Understanding how all of these factors influence population persistence is a fundamental issue in many areas of population biology including conservation biology (Caswell, 2001), metapopulation theory (Hanski, 1999), spatial spread of invasive species (Melbourne et al., 2007), and evolution of life history traits (Charlesworth, 1994).

One theoretical approach to understanding persistence is evaluating the "stochastic" per-capita growth rate of a population when rare (Gillespie, 1973; Turelli, 1978; Turelli and Petry, 1980; Chesson and Warner, 1981; Bulmer, 1985; Caswell, 2001). Intuitively, if the stochastic growth rate is positive, the population

tends to increase when rare and, consequently, is more likely to persist. Under suitable conditions, Chesson (1982), Ellner (1984), and Hardin et al. (1988a) placed this heuristic on a mathematically rigorous foundation for populations in serially uncorrelated environments. One of our main goals is to extend these results to correlated random environments.

Unstructured populations living in random environments can be described by models of the form

$$x_{n+1} = a(E_n, x_n)x_n \tag{1}$$

where x_n is the population abundance at time step n, E_n is a random variable describing the environmental state, and $a(E_n, x_n)$ is the per-capita fitness. When population abundance is low, it seems reasonable to approximate the dynamics of (1) with the linear equation

$$x_{n+1} = a(E_n, 0)x_n.$$

For this linear approximation, $x_n = a(E_{n-1}, 0) \ldots a(E_0, 0)x_0$. Under suitable conditions (i.e. $E_0, E_1, E_2, \ldots$ are stationary and ergodic and $\mathbb{E}(|\ln E_1|) < \infty$), the Birkhoff Ergodic Theorem implies that

$$\frac{1}{n} \ln x_n \to \mathbb{E}(\ln a(E_1, 0)) \quad \text{as } n \to \infty \text{ with probability one.}$$

In other words, $x_n \approx x_0 e^{\gamma n}$ where $\gamma = \mathbb{E}(\ln a(E_1, 0))$. Thus for the linearized model, the population grows exponentially, and

thereby persists, provided that $\gamma > 0$. Alternatively, if $\gamma < 0$, the population is driven to extinction at an exponential rate.

While this linear approximation of (1) provided many key insights for theoretical population biology (e.g., Cohen, 1966; Turelli, 1978), its implication for persistence of density-dependent models was not clarified until the early 1980s by the work of Chesson (1982) and Ellner (1984). Under suitable assumptions about the per-capita fitness (e.g. negative compensating density-dependence) and under the assumption that the environmental variables E_n are independent and identically distributed, Chesson and Ellner proved that $\gamma < 0$ implies that populations go extinct with probability one. Alternatively, $\gamma > 0$ implies that the distribution of x_n converges to a *positive* stationary distribution. In particular, $\gamma > 0$ implies that the population has a low probability of reaching low abundances in the long-term (i.e. is *stochastically bounded from below*, Chesson (1978, 1982, 1984)).

To describe the dynamics of structured populations, one can replace the scalar x_n by a vector $X_n = (X_n^1, \ldots, X_n^k)$ where X_n^i is the abundance of the population in state i. Furthermore, replacing the per-capita fitness term $a(E_n, x_n)$ with a non-negative $k \times k$ matrix $A(E_n, X_n)$ yields

$$X_{n+1} = A(E_n, X_n)X_n. \tag{2}$$

When the population abundance is low, it seems reasonable to approximate the dynamics of (2) with the linear equation

$$X_{n+1} = A(E_n, 0)X_n$$

in which case,

$$X_n = A(E_{n-1}, 0)\ldots A(E_0, 0)X_0.$$

Under suitable conditions (i.e. $E_0, E_1, E_2, \ldots$ are stationary and ergodic, $A(E_n, 0)$ are primitive, and $\mathbb{E}(|\ln\|A(E_1, 0)\||) < \infty$), the work of Ruelle (1979) and Arnold et al. (1994) imply that there exists a quantity γ such that

$$\frac{1}{n} \ln \left(X_n^1 + \cdots + X_n^k \right) \to \gamma \quad \text{as } n \to \infty \text{ with probability one.}$$

In other words, the total population size $X_n^1 + \cdots + X_n^k$ grows approximately like $(X_0^1 + \cdots + X_0^k)e^{\gamma n}$. The quantity γ is known as the *dominant Lyapunov exponent* and is also known as the *stochastic growth rate* in theoretical ecology (Tuljapurkar, 1990; Caswell, 2001). For the linearized model, if $\gamma > 0$, the population grows exponentially and persists. Alternatively, if $\gamma < 0$, the population is driven to extinction. To contend with the nonlinearities in structured population models, Hardin et al. (1988a) extended the work of Ellner (1984) to structured populations in serially uncorrelated environments. Under suitable assumptions about the matrices A (e.g. negative compensating density-dependence, primitivity), Hardin et al. (1988a,b) proved that $\gamma > 0$ implies that the distribution of X_n converges to a unique *positive* stationary distribution.

Our main goals are to extend the work of Chesson (1982), Ellner (1984), and Hardin et al. (1988a) in two ways. First, we allow for temporally correlated random environments. There has been a growing recognition amongst ecologists of the importance of long-term correlations in environmental time series (Steele, 1985; Halley, 1996; Vasseur and Yodzis, 2004). These correlations can influence extinction risk (Lawton, 1988; Petchey et al., 1997a; Cuddington and Yodzis, 1999; Heino et al., 2000), inflate population abundances (Gonzalez and Holt, 2002; Holt et al., 2003), and facilitate persistence of couple sink populations (Lawton, 1988; Petchey et al., 1997a; Cuddington and Yodzis, 1999; Heino et al., 2000; Gonzalez and Holt, 2002; Holt et al., 2003; Roy et al., 2005). While there have been some work on computing or approximating the stochastic growth rate with correlated noise (Tuljapurkar, 1982, 1990; Ferriere and Gatto, 1995; Tuljapurkar and Haridas, 2006), there have been no theoretical results connecting these results for the

linear models to the dynamics of the nonlinear models. Here, we extend the results of Chesson, Ellner and Hardin et al. to stationary environments (i.e. $E_1, E_2, \ldots$ is a stationary sequence) and asymptotically stationary environments.

Second, the work of Chesson, Ellner, and Hardin et al. concerns the convergence of the distribution of X_n and, consequently, provides information about X_n upon multiple independent realizations of the population process. Hence, these results can be used to answer questions like "what is the probability the population will fall below a given abundance for a given large n?" However, these results, at least formally, cannot be used to answer questions like "what fraction of time will the population state lie in some range of abundances?" To answer this and related questions, one needs to understand the convergence of empirical measures Π_n which describe the fraction of time the population has spent in a given state up to time n. Here, we prove that if $\gamma > 0$, then the distribution of these empirical measures converges to the unique positive stationary distribution. This later form of convergence is analogous to the strong law of large numbers and is the form of convergence that may be observed empirically in a time series. With this result, questions about the fraction of time spent in a state or the long-term probability of being in a state become interchangeable.

We achieve these goals for a class of models that are described in Section 2. The main results are stated in Section 3. To apply these results to specific models, it is necessary to estimate or compute γ, a topic discussed in Section 4. In Section 5, we illustrate the application of these methods to an unstructured population living in a correlated environment, a spatially structured population, and a biennial plant model. For instance, analytical approximations for the stationary distributions of unstructured models show that reddened noise tends to increase mean population abundance and create bimodal distributions. Alternatively, we show that diffusively coupled sink populations can persist provided that within the patch fitness is sufficiently variable in time but not strongly correlated across space. Proofs of most results are presented in the Appendices.

2. Assumptions

In this section, we present the assumptions on (2) needed to state our main results. A weaker set of assumptions is presented in Appendix B. Prior to stating the assumptions, we introduce some terminology and notation. Assuming that the population has k states, the *population state space* for (2) is

$$\mathbb{R}_+^k = \{x = (x^1, \ldots, x^k) \in \mathbb{R}^k : x^i \geq 0 \text{ for all } i\}$$

where the i-th coordinate x^i of $x \in \mathbb{R}_+^k$ corresponds to the mean abundance or density of individuals in state i. For $x, y \in \mathbb{R}_+^k$ we write $x \geq y$ if $x^i \geq y^i$ for all i, $x > y$ if $x \geq y$ and $x \neq y$, and $x \gg y$ if $x^i > y^i$ for all i.

The population dynamics in (2) are determined by two constructs. First, a sequence of random variables $E_0, E_1, E_2, \ldots$, taking values in an *environmental state space* $\mathscr{S}$. Second, given the current environmental state $e \in \mathscr{S}$ and population state $x \in \mathbb{R}_+^k$, the projection matrix $A(e, x)$ determines the population state in the next time step as follows:

$$f_e(x) := A(e, x)x.$$

2.1. The environmental sequence

Our first assumption concerns the environmental sequence $E_0, E_1, E_2, \ldots$.

A1 $E_0, E_1, E_2, \ldots$ is an *ergodic stationary* sequence in a Polish[1] space $\mathscr{S}$.

---

[1] A Polish space is a complete separable metric space. The assumption that $\mathscr{S}$ is Polish is harmless since most spaces (such as compact spaces, Euclidean spaces,

Stationarity means that for all $i \in \mathbb{N}$ the shifted sequence $E_i, E_{i+1}, E_{i+2}, \ldots$ and the original sequence $E_0, E_1, E_2, \ldots$ have the same distribution. That is

$$\mathbb{P}(E_i \in A_0, \ldots, E_{i+n} \in A_n) = \mathbb{P}(E_0 \in A_0, \ldots, E_n \in A_n)$$

for all Borel sets $A_0, \ldots, A_n$ in $\mathcal{S}$. We relax this assumption in Section 3.3. Ergodicity means, intuitively, the environmental sequence is not a mixture of two other stationary distributions.[2] An important consequence of ergodicity is that if $\mathbb{E}(|h(E_0)|) < \infty$ for some function $h : \mathcal{S} \to \mathbb{R}$, then the temporal average of h, $\frac{h(E_0)+\cdots+h(E_{n-1})}{n}$, converges to its expected value $\mathbb{E}(h(E_0))$ with probability one.

This representation of the environmental dynamics allows for a diversity of possibilities as the following examples illustrate. In Section 3.3, we show that under additional assumptions the stationary assumption can be relaxed.

Example 1 (*Periodic Environments*). Let $\mathcal{S} = \{0, \ldots, m-1\}$. Let E_0 be a uniformly distributed random variable on $\mathcal{S}$ and define $E_{n+1} = E_n + 1 \mod m$ for $n \geq 0$. Then the environmental dynamics correspond to randomly choosing an initial environmental state and successive environmental states changing periodically.

Example 2 (*Quasi-periodic Environments*). Let $\mathcal{S}$ be the unit circle and $\theta/2\pi$ be irrational. Define E_0 to be a uniformly distributed random variable on $\mathcal{S}$ and $E_{n+1} = E_n + \theta \mod 2\pi$ for $n \geq 0$. Then the environmental dynamics correspond to randomly choosing an initial point on the circle and successive environmental states changing in a quasi-periodic fashion.

Example 3 (*Independent and Identically Distributed Environments*). Let $E_0, E_1, E_2, \ldots$ be a sequence of independent and identically distributed random variables. This is the case considered by Chesson (1982), Ellner (1984) and Hardin et al. (1988a,b).

Example 4 (*Finite Markovian Environments*). Suppose the environment only exhibits a finite number of states $\mathcal{S} = \{1, \ldots, m\}$ and the transition probabilities between these states are given by the transition matrix P i.e. the i-j-th entry $P(i, j)$ of P is the probability of going from state i to state j. If P is irreducible, there is a unique stationary distribution μ for P. Namely, there is a row vector μ such that $\mu P = \mu$ with $\sum_i \mu_i = 1$. Let E_0 be a random variable on $\mathcal{S}$ such that $\mathbb{P}(E_1 = i) = \mu_i$. Let E_n for $n \geq 1$ be a sequence of random variables such that $\mathbb{P}(E_n = j | E_{n-1} = i) = P(i, j)$.

Example 5 (*Autoregressive-moving-average (ARMA) Models*). To capture complicated patterns of autocorrelation in a continuous state space, one can consider a sequence of random variables E_n defined by

$$E_n = \sum_{i=1}^{p} b_i E_{p-i} + \sum_{i=0}^{q} c_i Z_{n-i}$$

where $b_1, \ldots, b_p$ and $c_1, \ldots, c_q$ are constants and Z_n are independent standard normal random variables (Box and Jenkins, 1990). The simplest form of these models is a first-order autoregressive model

$$E_n = b E_{n-1} + c_0 Z_n$$

for which the correlation between E_n and E_{n+k} is b^k. This first-order regressive model is stationary if $b < 1$ and E_0 is normal with mean zero and variance $c_0^2/(1-b^2)$.

The five preceding examples are all particular cases of *Markov chains*. This can be generalized as follows.

Example 6 (*General Markovian Environments*). A *Markov kernel* over $\mathcal{S}$ is a family $P = \{P(e, \cdot)\}_{e \in \mathcal{S}}$ of probability measures over $\mathcal{S}$, such that for each Borel set $A \subset \mathcal{S}$ the mapping $e \to P(e, A)$ is Borel. When $\mathcal{S}$ is countable, $P(i, j) = P(i, \{j\})$ is called a *transition matrix*.

A *Markov chain* with kernel P is a stochastic process $(E_n)_{n \geq 0}$ such that

$$\mathbb{P}(E_{n+1} \in A | E_n, E_{n-1}, \ldots, E_0) = P(E_n, A).$$

Hence, if E_0 has distribution μ_0,

$$\mathbb{P}(E_0 \in A_0, E_1 \in A_1, \ldots, E_n \in A_n)$$
$$= \int_{A_0 \times A_1 \times \cdots \times A_n} \mu_0(de_0) P(e_0, de_1) \ldots P(e_{n-1}, de_n).$$

Probability measure μ is called *invariant* for P (or P-invariant) if

$$\mu P = \mu$$

where μP stands for the probability defined by

$$\mu P(A) = \int \mu(de) P(e, A).$$

Given such a measure, the Markov chain $(E_n)_{n \geq 0}$ with kernel P and initial distribution μ (i.e E_0 has distribution μ) is stationary (this follows directly from (6)). If we furthermore assume that there are no other P-invariant probability measures, then $(E_n)_{n \geq 0}$ is ergodic. This later result is proved in Appendix A.

For a general Markov chain, neither the existence nor the uniqueness of invariant measure is automatic. However there is a powerful and well developed theory of Markov chains on general state spaces (see e.g. Duflo (1997) or Meyn and Tweedie (1993)), and numerous fine criteria ensuring both the existence and uniqueness of the invariant probability measure. For instance when $\mathcal{S}$ is finite (Examples 1 and 4) existence always holds and uniqueness follows from the irreducibility of P. Example 2 is a particular case of uniquely ergodic map on an Abelian compact group (Mañé, 1983). Example 5 belongs to the well studied class of iterative Lipschitz Models (see e.g. Duflo (1997), Chapter 6.3 and Theorem 6.3.16).

2.2. The population dynamics

Our remaining assumptions concern the population dynamics. We begin by assuming that the population dynamics tend to remain bounded i.e. there is no tendency for population numbers to explode to infinity. Recall that a continuous map $V : \mathbb{R}_+^k \mapsto \mathbb{R}$ is said to be *proper* provided $\lim_{\|x\| \to \infty} V(x) = \infty$.

A2 Contraction at infinity There exists a proper function $V : \mathbb{R}_+^k \mapsto \mathbb{R}$ and random variables $\alpha, \beta : \mathcal{S} \mapsto [0, \infty)$ such that
1. $V(A(e, x)x) \leq \alpha(e)V(x) + \beta(e)$;
2. $\mathbb{E}(\ln(\alpha)) < 0$;
3. $\mathbb{E}(\ln^+ \beta) < \infty$ where $\ln^+(s) = \max\{\ln s, 0\}$ for $s > 0$.

These conditions roughly require there is a function V that tends to decrease along population trajectories whenever the population abundances are high. For the applications presented in Section 5, $V(x)$ is chosen to be the total population abundance $\sum_{i=1}^{k} x^i$. In the Appendix B, we show that **A2** implies that there exists $C > 0$ such that

$$\liminf_{n \to \infty} \mathbb{P}(X_n \leq M) \geq 1 - \frac{C}{M} \tag{3}$$

for all $M > 0$. Hence **A2** ensures that X_n is *stochastically bounded from above* (Chesson, 1978, 1982, 1984; Ellner, 1984; Chesson and Ellner, 1989).

---

etc.) are Polish and it provides a nice mathematical framework when dealing with random variables and probabilities.

[2] More precisely, for any bounded Borel function $H : \mathcal{S} \times \mathcal{S} \times \cdots \to \mathbb{R}, \mathbb{P}(H(E_0, E_1, \ldots) = H(E_1, E_2, \ldots)) = 1$ implies that $H(E_0, E_1, \ldots)$ is almost surely constant.

Our next assumption, roughly translated, asserts that after sufficiently many time steps, individuals in every state contribute to the abundance of individuals in all other states. For constant environments (i.e. $\mathscr{S}$ consists of a single environmental state), this assumption corresponds to a matrix being primitive (Horn and Johnson, 1990; Caswell, 2001).

A3: **Primitivity** There is a positive integer N such that with probability one

$$A(E_n, X_n) \ldots A(E_1, X_1) \quad \text{has all positive entries}$$

whenever $X_1 \in \mathbb{R}_+^k$ and $n \geq N$.

Our fourth assumption is purely technical, but is met for most models.

A4: **Smoothness** The map $(e, x) \to A(e, x)$ is Borel, $f_e(x)$ is twice continuous differentiable for all $e \in \mathscr{S}, x \in \mathbb{R}_+^k$, and

$$\mathbb{E} \left(\sup_{\|x\| \leq 1} \ln^+ \left(\|f_e(x)\| + \| Df_e(x)\| + \| D^2 f_e(x)\| \right) \right) < +\infty.$$

To account for competition, we assume all of the entries should be non-increasing functions of the population densities. Moreover, for each population state, we assume one of its entries is a decreasing function of its density or the density of one the other population states.

A5: **Intraspecific competition** The matrix entries $A_{ij}(e, x)$ satisfy

$$\frac{\partial A_{ij}}{\partial x_l}(e, x) \leq 0$$

for all e and x. Moreover, for each i there exists some j and l such that this inequality is strict for all e and x.

For our stronger convergence results, we assume that the populations exhibit compensating density dependence.

A6: **Compensating density dependence** All entries of the derivative $Df_e(x)$ of $f_e(x)$ are non-negative for all e and x.

This assumption ensures that the population dynamics are monotone. In other words, if $x \geq y$, then $f_e(x) \geq f_e(y)$. Overcompensating density dependence is ruled out by **A6**. However, Theorem 2 can be used to get persistence results for certain models with overcompensating density dependence.

3. Persistence and extinction

An important quantity to determine persistence is the dominant Lyapunov exponent γ associated with the extinction state 0. Provided that $\mathbb{E}(\ln^+ \|A(E_0, 0)\|) < \infty$ which follows from condition **A4**, Kingman's subadditive ergodic theorem (Kingman, 1973) implies that there exists γ such that

$$\gamma = \lim_{n \to \infty} \frac{1}{n} \ln \|A(E_{n-1}, 0) \ldots A(E_0, 0)\|$$

with probability one. This dominant Lyapunov exponent is the stochastic growth rate of the population when rare. When γ is negative, extinction is expected as the following proposition demonstrates.

Proposition 1 (*Almost Sure Extinction*). *Assume* **A1**, **A4** *and* **A5** *hold. Then*

$$\limsup_{n \to \infty} \frac{1}{n} \ln \|X_n\| \leq \gamma$$

with probability one. In particular, if $\gamma < 0$, then

$$\lim_{n \to \infty} X_n = 0$$

with probability one.

Remark 1. In the case of $\gamma = 0$, we conjecture that X_n converges to 0 in probability. A strategy for proving this conjecture may follow from Theorem 5 of Athreya and Dai (2000).

Proof. Let $X_0 = x \in \mathbb{R}_+^k$. Since

$$X_n = A(E_{n-1}, X_{n-1}) \ldots A(E_1, X_1)A(E_0, X_0)X_0,$$

A5 implies that

$$\limsup_{n \to \infty} \frac{1}{n} \ln \|X_n\| \leq \limsup_{n \to \infty} \frac{1}{n} \ln \|A^n(E_0, 0)\| = \gamma$$

with probability one. $\square$

When $\gamma > 0$, we would expect the population to recover at an exponential rate whenever its abundance gets low and, consequently, persist. To quantify this expectation, there are two ways to think about the asymptotic behavior of X_n. First, one can ask what is the distribution of X_n far into the future. For example, what is the probability that the population abundance of each state is greater than ϵ in the long term i.e. $\mathbb{P}(X_n \geq (\epsilon, \ldots, \epsilon))$ for large n? The answer to this question provides information what happens across many independent realizations of the population dynamics. Section 3.1 takes this approach to the persistence question. Alternatively, one might be interested about the statistics associated with a single realization of the process i.e. a single time series. For instance, one could ask what fraction of the time was the abundance of each population state greater than ϵ? To answer this question, it useful to introduce the *empirical measures*

$$\Pi_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

where δ_{X_i} denotes a Dirac measure at X_i i.e. $\delta_{X_i}(A) = 1$ if $X_i \in A$ and 0 otherwise for any (Borel) set $A \subset \mathbb{R}_+^k$. $\Pi_n(A)$ equals the proportion of time the population spends in A up to time n. Section 3.2 takes this approach to the persistence question. Section 3.3 proves both forms of convergence for certain classes of non-stationary processes.

3.1. Persistence in distribution

If $\gamma > 0$ and all our assumptions are met, our first theorem shows that there is a "globally stable" positive random vector for the population process X_n.

Theorem 1 (*Convergence to a Positive Random Vector*). *Assume* **A1–A6**. *If $\gamma > 0$, then there exists a* **positive** *random vector $\widehat{X} : \Omega \to \mathbb{R}_+^k$ such that the distribution of X_n converges to the distribution $\widehat{X}$ whenever $X_0 = x > 0$.*

Recall that *convergence of X_n to $\widehat{X}$ in distribution* means that for any bounded continuous function $h : \mathbb{R}^k \to \mathbb{R}$, the expected value $\mathbb{E}(h(X_n))$ of $h(X_n)$ converges to the expected value $\mathbb{E}(h(\widehat{X}))$ of $h(\widehat{X})$. Equivalently, X_n converges in distribution to $\widehat{X}$ if $\mathbb{P}(X_n \in A)$ converges to $\mathbb{P}(\widehat{X} \in A)$ for any set A such that the probability of $\widehat{X}$ lying exactly on the boundary of A is zero. Since $\widehat{X} \gg 0$ (i.e. $\mathbb{P}(\widehat{X}_i = 0) = 0$ for all i), it follows that for any $\epsilon > 0$ there exists $x_* \gg 0$ and N such that

$$\mathbb{P}(X_n \geq x_*) \geq 1 - \epsilon$$

for $n \geq N$. In other words, the probability that the population abundances get small far into the future is small. This notion of persistence is a structured version of what Chesson (1978, 1982) calls *stochastically bounded from below*. A proof of Theorem 1 is provided in Appendix C. The proof uses some of the techniques developed in Chueshov's monograph 2002 on monotone random systems. However, under our set of assumptions, we are able to prove stronger results than those proved by Chueshov (e.g., compare to Theorem 4.4.2 or Proposition 4.5.2).

For populations with overcompensating density dependence or without intraspecific competition, one cannot expect convergence to a unique positive vector. Without intraspecific competition (in the strict sense stated in **A5**) or with overcompensating density dependence, deterministic models may exhibit multiple attractors and consequently, their noisy cousins may exhibit multiple positive stationary distributions. The following Theorem provides a way to verify persistence for these, more general, systems. A proof is provided in Appendix D.

Theorem 2. *Assume* **A1**–**A4** *holds for* $X_{n+1} = A(E_n, X_n)X_n$. *Let* $Y_{n+1} = B(E_n, Y_n)Y_n$ *satisfy* **A1**–**A4**, **A6**. *If* $\gamma(B) > 0$ *and* $A(e, x)x \geq B(e, x)x$ *for all* $e \in \mathscr{S}$ *and* $x \in \mathbb{R}_+^k$, *then there exists a positive random vector* $\widehat{X} : \mathscr{S} \to \mathbb{R}_+^k$ *such that*

$$\liminf_{n \to \infty} \mathbb{P}(X_n \geq x) \geq \mathbb{P}(\widehat{X} \geq x)$$

for all $x \geq 0$.

3.2. Persistence via empirical measures

Theorems 1 and 2 provide information about the distribution of the population state far into the future. However, one might be interested in understanding how typical sample paths of the population process are distributed in time. Hence, one can consider the empirical measures Π_n of the state sequence $\{X_n\}_{n=1}^{\infty}$. One can think of Π_n corresponding to placing a weight of $\frac{1}{n}$ at all the population states visited by the population up to time n. Hence Π_n describes the pattern of points in the population state space up to time n. For these empirical measures, we will be interested in whether they converge in distribution to some limiting probability measure, call it μ. When convergence occurs, the pattern of points determined by Π_n are distributed like μ when n is large.

Theorem 3. *Assume the assumptions of Theorem 1 hold. Let* $\widehat{X}$ *be like in Theorem 1. Then, with probability one,* Π_n *converges toward the distribution of* $\widehat{X}$. *That is*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} h(X_i) = \mathbb{E}(h(\widehat{X}))$$

with probability one, for every bounded and continuous function $h : \mathbb{R}^k \to \mathbb{R}$.

3.3. Going beyond stationarity

All of the results thus far assume that the environmental sequence is stationary. When this stationarity is achieved only in the long-term, it seems reasonable to expect that Π_n and the distribution of X_n should converge to the distribution of $\widehat{X}$ even if the initial environmental state is not following the stationary distribution. We are able to confirm this expectation in a quite general setting.

Let $(F_n)_{n \geq 0}$ be a sequence of random variables taking values in $\mathscr{S}$. We shall say that $(F_n)_{n \geq 0}$ is *asymptotically ergodic-stationary* if there exists an ergodic stationary process $(E_n)_{n \geq 0}$ taking values in $\mathscr{S}$ and a sequence $\alpha_n \to 0$ such that, for all $i \in \mathbb{N}$ and all Borel subsets $A_1, \ldots, A_i$ of $\mathscr{S}$

$$|\mathbb{P}(F_{n+1} \in A_1, \ldots, F_{n+i} \in A_i) - \mathbb{P}(E_1 \in A_1, \ldots, E_i \in A_i)| \leq \alpha_n.$$

In other words $(F_n)_{n \geq 0}$ and $(E_n)_{n \geq 0}$ are asymptotically equally distributed. By a mild abuse of terminology we call (E_n) the *limiting process* of (F_n).

An important class of asymptotically ergodic-stationary processes is given by *strongly ergodic Markov chains*.

Example 7 (*Strongly Ergodic Markov Chains*). We use the notation of Example 6 in Section 2. The Markov chain $(F_n)_{n \geq 0}$ (or the kernel P) is said to be *strongly ergodic*[3] if there exists a probability measure μ such that for every initial distribution ν (i.e $\mathbb{P}(F_0 \in A) = \nu(A)$)

$$\lim_{n \to \infty} \sup_A |\mathbb{P}(F_n \in A) - \mu(A)| = 0$$

or; equivalently

$$\lim_{n \to \infty} \sup_A |\nu P^n(A) - \mu(A)| = 0.$$

Here the supremum is taken over all Borel subsets $A \subset \mathscr{S}$; and νP^n stands for the measure recursively defined by $\nu P^n = (\nu P^{n-1})P$.

It is not hard to show that this makes (F_n) asymptotically ergodic-stationary with the limiting process given by the chain with kernel P and initial distribution μ. In particular, μ is P-invariant and unique. Some classical examples of conditions ensuring strong ergodicity are:

- (Finite Markov chains): $\mathscr{S}$ is finite with P irreducible and aperiodic.
- (Countable Markov chains): $\mathscr{S}$ is countable, P is irreducible, aperiodic and admits an invariant probability.
- (Harris chains) There exist a Borel set $A \subset \mathscr{S}$, a probability measure ν over $\mathscr{S}$, and a number $\beta > 0$ such that
  - $\mathbb{P}(\exists n \geq 1 : F_n \in A | F_0 = f) = 1$ for all $f \in \mathscr{S}$, and
  - $\mathbb{P}(F_1 \in B | X_0 = f) \geq \beta \nu(B)$ for all $f \in A$.

We refer to Meyn and Tweedie (1993) and Duflo (1997) for further examples, conditions and details.

Theorem 4. *Let* $(F_n)_{n \geq 0}$ *be an asymptotically ergodic-stationary process with limiting process* $(E_n)_{n \geq 0}$. *Let* (Y_n) *be solution to*

$$Y_{n+1} = A(F_n, Y_n)Y_n, \quad n \geq 0$$

with $Y_0 = y > 0$. *Assume that assumptions of Theorem 1 hold for* A *and* (E_n). *Let* $\widehat{X}$ *be as in Theorem 1. Then the distribution of* Y_n *converges to the distribution of* $\widehat{X}$ *and, with probability one,* $\Pi_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$ *converges to the distribution of* $\widehat{X}$.

4. Verifying persistence

Our results imply that verifying persistence for stochastic structured models boils down to verifying the appropriate assumptions **A1**–**A6** and understanding the sign of the dominant Lyapunov exponent of $A(e, 0)$. The assumptions **A1**–**A6** are either satisfied automatically for most models or require straightforward computations to verify. The hardest assumption to verify is the positivity of γ. Here, we provide some easily computed lower bounds for γ and approaches to analytically estimating γ. For a more comprehensive treatment of Lyapunov exponents, we refer the reader to the monograph of Tuljapurkar (1990) and the article of Ferriere and Gatto (1995).

A key result for random products of non-negative matrices is a random version of the Perron-Frobenius Theorem. Roughly speaking, this theorem asserts that the stochastic growth rate for the linearized system is realized for every positive initial condition. Moreover, this stochastic growth rate depends analytically on the entries of the matrices. These two facts are extremely useful for computing or estimating γ. Ruelle (1979) provided the first random version of this theorem assuming that the environmental state space $\mathscr{S}$ is compact. Arnold et al. (1994) were able to remove this compactness assumption. Note that for primitive random products of matrices (i.e. random products satisfying **A3**), Theorem 5 can be applied by replacing A with A^N.

---

[3] What we call strongly ergodic here is often called ergodic in the literature on Markov chains. However we want to distinguish this notion from the ergodicty of the stationary process.

Theorem 5 (*Arnold et al., 1994*). *Let $A : \mathcal{S} \to$ be a Borel map into positive $k \times k$ matrices. If $\mathbb{E}(\ln^+ \|A\|) < \infty$ and $\mathbb{E}\left(\ln^+ \max 1/A_{ij}\right) < \infty$, then $\lim_{n \to \infty} \frac{1}{n} \ln \|A(E_{n-1}) \dots A(E_0)v\| = \gamma$ with probability one for all $v > 0$. Moreover, $\gamma = \gamma(A)$ is an analytic function in a neighborhood of A in the space of L^1 functions from $\mathcal{S}$ to $k \times k$ matrices where $\|A\|_1 = \mathbb{E}\|A(E_0)\|$.*

While $\gamma = \mathbb{E}(\ln a)$ for unstructured populations due to the commutativity of scalar products, explicitly computing γ for structured populations is often analytically intractable. Two notable exceptions are Roerdink's formula (1987) for γ for a two age class model which is discussed in Section 5.3 and random matrices that share a reproductive value or share a stable structure (Tuljapurkar, 1986, 1990). Other computable examples for random products of 2×2 matrices can be found in Key (1987) and Mannion (1993). Tuljapurkar states that the random matrices $A(e, 0)$ *share a stable structure* if there is a positive vector v and Borel function $\lambda : \Omega \to [0, \infty)$ such that

$$A(e, 0)v = \lambda(e)v$$

for all e. Similarly, the matrices *share a reproductive value* if there exists a positive row vector w such that

$$wA(e, 0)v = \lambda(e)w$$

for all e. In either case, Theorem 5 implies that $\gamma = \mathbb{E} \ln \lambda(E_0)$. A remarkable feature about this observation is that despite the potential non-commutativity of the matrices, the autocorrelation structure of the matrices has no effect on γ. Section 5.2 illustrates a natural way that a shared stable structure arises in spatial models.

Since one cannot typically find explicit formula for γ, it is useful to have lower and upper bounds for γ. Using Theorem 5, we can prove the existence of an easily computed lower bound to γ that involves the *permanent of a matrix* $A = (a_{ij})$

$$\text{perm}(A) = \sum_{\sigma} a_{\sigma(1)1} \dots a_{\sigma(k)k}$$

where the sum is taken over all permutations of the set $\{1, \dots, k\}$. The definition of perm(A) differs from that of the determinant of A in that the signs of the permutations σ are not taken into account. For a 2×2 matrix, for instance, the permanent is given by $a_{11}a_{22} + a_{12}a_{21}$ while the determinant is given by $a_{11}a_{22} - a_{12}a_{21}$. A formula similar to Laplace's for expanding a determinant along a row or column is also valid for the permanent. However, in the case of the permanent, all signs have to be ignored in this expansion. Using the permanents, we get the following estimate for the dominant Lyapunov exponent. Other lower bounds for γ are discussed in chapter 7 of Tuljapurkar's monograph (1990). A proof of this Proposition is given in Appendix G.

Proposition 2. *Let A be a Borel map from $\mathcal{S}$ to the non-negative matrices such that* **A3** *holds, $\mathbb{E}(\ln^+ \|A(E_0)\|) < \infty$ and $\mathbb{E}(\ln \text{perm}(A(E_0))) < \infty$. Then the dominant Lyapunov exponent γ for A satisfies*

$$\gamma \geq \frac{1}{k}\mathbb{E}(\ln(\text{perm}(A(E_0)))).$$

For random products of non-negative 2×2 matrices $A(e)$, this result implies that the dominant Lyapunov exponent is positive provided that $\mathbb{E}(\ln(a_{11}a_{22} + a_{12}a_{21})) > 0$, a condition that can easily be verified. For random Leslie matrices of the form

$$A(e) = \begin{pmatrix} a_1(e) & a_2(e) & \dots & a_{k-1}(e) & a_k(e) \\ s_2(e) & 0 & \dots & 0 & 0 \\ 0 & s_3(e) & \dots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & \dots & 0 & s_k(e) & 0 \end{pmatrix}$$

where s_i is the probability of surviving from stage $i - 1$ to stage i and a_i is the fecundity of stage i, Proposition 2 implies that

$$\gamma \geq \frac{1}{k}\mathbb{E}(\ln(s_2 \dots s_k a_k)).$$

Since $s_2(e) \dots s_k(e)$ is the probability of an individual surviving from stage 1 to stage k for the environmental conditions e, $s_2(e) \dots s_k(e)a_k(e)$ is the mean number of progeny an individual in stage 1 expects to produce at stage k under environmental state e.

Theorem 5 asserts that the dominant Lyapunov exponent for primitive $A(e)$ depends analytically on the entries of $A(e)$. Hence, Lyapunov exponents computed for limiting scenarios provide estimates for the Lyapunov exponents near the limit. Moreover, analyticity allows one to compute Taylor expansions of $\gamma(A + \epsilon B)$ where A is constant matrix (whose dominant Lyapunov exponent is given by the ln of its dominant eigenvalue) and B is a random perturbation of A. This approach to understanding the effects of noise on structured populations is extremely useful and is discussed in Tuljapurkar's monograph 1990. Differentiabily of $\gamma(A)$ can be lost as A approaches a reducible random matrix. However, the following proposition shows that even for non-negative (not necessarily primitive) $A(e)$ that γ depends continuously on the entries of $A(e)$. This fact is used in 5.2 and 5.3. A proof is provided in Appendix H.

Proposition 3. *Assume $\mathcal{S}$ is compact, $\mathcal{A}$ is the set of continuous maps from $\mathcal{S}$ to the non-negative $k \times k$ matrices, and $\gamma(A)$ denotes the dominant Lyapunov exponent for $A \in \mathcal{A}$. Then $A \mapsto \gamma(A)$ is continuous with respect to the topology given by $\|A\| = \sup_{e \in \mathcal{S}} \|A(e)\|$.*

5. Applications

Here we provide applications to unstructured populations in correlated environments, spatially structured populations, and biennial plants to illustrate the utility of our results.

5.1. Unstructured populations in correlated environments

The simplest application of our results is to an unstructured population with Beverton–Holt dynamics in a random environment. Ellner (1984) and Haskell and Sacker (2005) considered such a model for independent and identically distributed random environments. Our results allow, more generally, for stationary environments and also ensure convergence of the empirical measures. If X_n denotes the abundance of the population at time n, then Beverton–Holt dynamics in a random environment are given by

$$X_{n+1} = \frac{\lambda_n X_n}{1 + a_n X_n} \tag{4}$$

where λ_n and a_n are stationary sequences of positive random variables corresponding to the intrinsic fitness and competition intensity of the population at time n. In Appendix I, we prove that assumptions **A1**–**A6** are satisfied provided that $\mathbb{E}\left(\ln^+\left(\lambda_n/a_n + \lambda_n + 2\lambda_n a_n\right)\right) < \infty$.

By the Birkhoff ergodic theorem, $\gamma = \mathbb{E}(\ln \lambda_1)$. Hence, if $\mathbb{E}(\ln \lambda_1) < 0$, then extinction occurs with probability one. If $\mathbb{E}(\ln \lambda_1) > 0$, then there is a unique positive random variable $\widehat{X}$ such that the distribution of X_n and the empirical measures $\Pi_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ converge to the distribution of $\widehat{X}$. Hence, if the dynamics of (4) are simulated for a sufficiently long period of time, then the simulation's histogram is a good approximation to the distribution of $\widehat{X}$ (Fig. 1a–b).

To illustrate how temporal correlations in a_n or λ_n influence the shape of the stationary distribution, we simulated the dynamics of (4) with $a_n = 0.01$, $\lambda_n = \exp(\sigma E_n + \mu)$, and

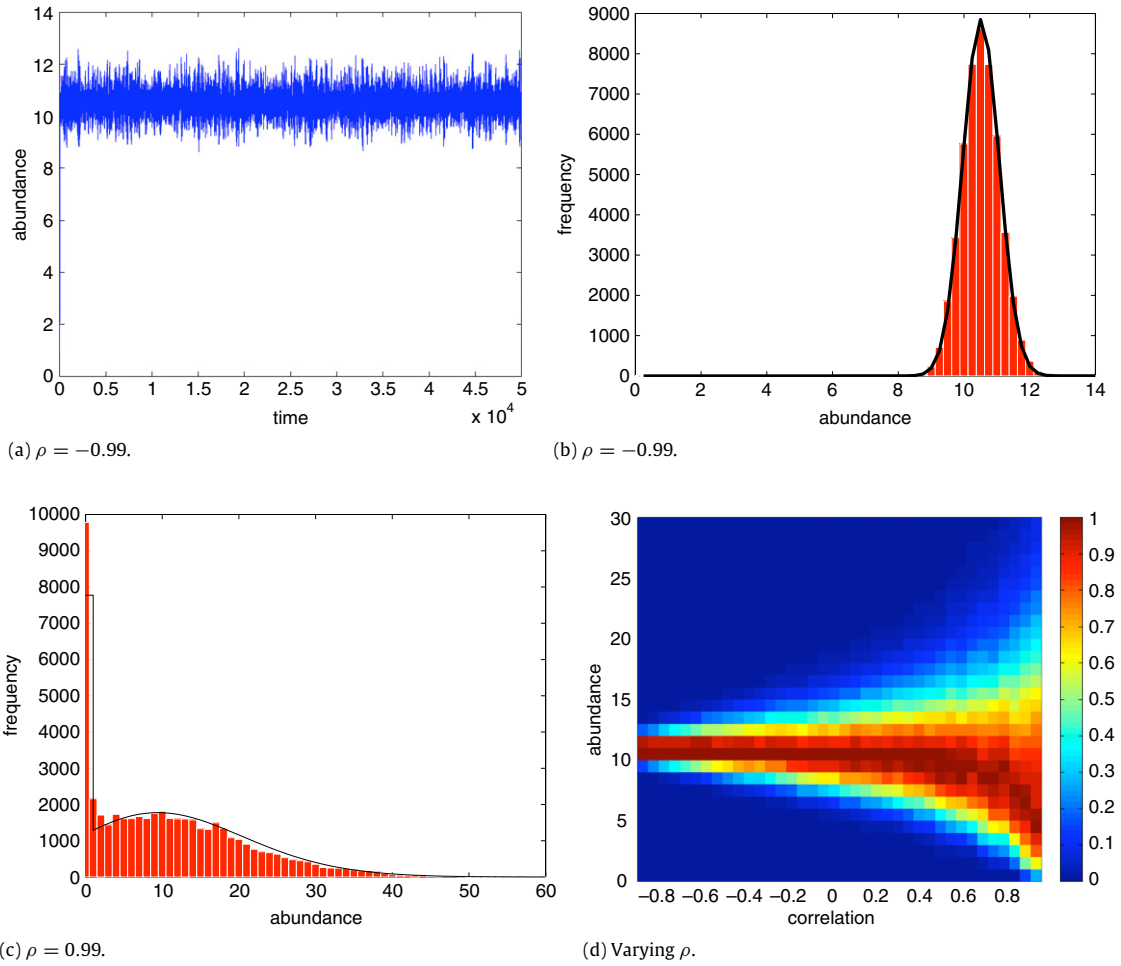$$E_{n+1} = \rho E_n + \sqrt{1 - \rho^2} Z_n$$

(a) $\rho = -0.99$.

(b) $\rho = -0.99$.

(c) $\rho = 0.99$.

(d) Varying ρ.

Fig. 1. Effects of environmental correlations on unstructured populations whose dynamics are given by (4) with $a_n = 0.01$, $\lambda_n = \exp(E_n\sigma + \mu)$, $E_{n+1} = \rho E_n + \sqrt{1 - \rho^2}Z_n$, and Z_n are independent standard normals. In (a), the dynamics of X_n for 50,000 time steps when $\rho = -0.99$. In (b), the histogram corresponding to (a) with the analytic approximation in black. In (c), the histogram for $\rho = 0.99$ with the analytic approximation in black. In (d), the density of the stationary distribution is plotted as a function of the correlation coefficient ρ. Hotter colors correspond to higher densities. In all figures, $\mu = \sigma = 0.1$.

where $-1 \leq \rho \leq 1$ corresponds to the correlation $\mathbb{E}(E_n E_{n+1})$ and Z_n are independent standard normals. This choice of λ_n and a_n clearly satisfies condition (12) in Appendix I. Even though ρ has no effect on the dominant Lyapunov exponent, Fig. 1c illustrates that the stationary distribution gets broader with increasing positive correlation.

When the correlation is nearly perfect i.e. $\rho \approx \pm 1$, one can approximate the distribution of $\widehat{X}$. If the environmental noise is strongly positively correlated i.e. $\rho \approx 1$, then one can derive the following approximation by assuming that the population dynamics are tracking the equilibrium abundance determined by λ_n:

$$\widehat{Y}_+ = \begin{cases} 100\left(\exp(\sigma Z + \mu) - 1\right) & \text{if } Z \geq -\mu/\sigma \\ 0 & \text{if } Z < -\mu/\sigma \end{cases}$$

where Z is a standard normal random variable. This approximation appears to works rather well (Fig. 1d) for very strong positive correlations and suggests the fraction of time spent at low densities increases with coefficient of variation σ/μ of the environmental process. However, since this approximation involves a fundamental phase transition from convergence to a unique positive stationary distribution for $\rho < 1$ to a stationary distribution placing positive weight on 0, more mathematical analysis is needed to understand the validity of this approximation.

When the environmental noise is strongly negatively correlated i.e. $\rho \approx -1$, one can approximate the stationary distribution $\widehat{X}$ by

assuming the population dynamics are tracking period two points determined by successive values of λ_n. This approximation also works rather well (Fig. 1b) and is given by

$$\widehat{Y}_- = \frac{100\left(\exp(2\mu) - 1\right)}{1 + \exp(\sigma Z + \mu)}$$

where Z is a standard normal random variable. Unlike the approximation for the positively correlated environment, the approximation for the negatively correlated environment results in a maximum observable population abundance (i.e. $100\left(e^{2\mu-1}\right)$) and is unimodal with a positive mode. Jensen's inequality implies that

$$\mathbb{E}\widehat{Y}_- > 100(\exp(\mu) - 1)$$

whenever $\sigma > 0$.

In both approximations, environmental noise inflates the mean population abundance. In general, simulations suggest that the magnitude of this inflation increases with ρ. This observation is consistent with recent theoretical and empirical work on the inflationary effects of autocorrelation on sink populations (Gonzalez and Holt, 2002).

5.2. Metapopulation persistence in random environments

Consider a population that lives in k distinct patches and whose abundance in the ith patch at time n is X_n^i. Let $X_n = (X_n^1, \ldots, X_n^k)$ be

the vector of population abundances. If in each time step a fraction d of individuals disperses, dispersing individuals are distributed equally amongst all the patches, and the per-capita fitness on an individual in the i-th patch is given by a Beverton Holt model with random coefficients, then

$$X_{n+1} = ((1-d)I_k + dJ_k) \operatorname{diag}\left\{ \frac{\lambda_n^1}{1+a_1 X_n^1}, \ldots, \frac{\lambda_n^k}{1+a_k X_n^k} \right\} X_n \quad (5)$$

where I_k is the $k \times k$ identity matrix, J_k is the $k \times k$ matrix whose entries all equal $1/k$, $a_i > 0$ measure the intensity of competition in patch i, and λ_n^i is a stationary sequence of positive random variables corresponding to the intrinsic fitness in patch i. In Appendix I, we show assumptions **A1–A6** are satisfied provided that $\mathbb{E}\left(\ln^+ (\max_i \lambda_i/a_i + \max_i \lambda_i + \max_i \lambda_i a_i)\right) < \infty$.

Define

$$A_d(e, x) = ((1-d)I_k + dJ_k) \operatorname{diag}\left\{ \frac{e_1}{1+a_1 x^1}, \ldots, \frac{e_k}{1+a_k x^k} \right\}.$$

and $\gamma(d)$ to be the dominant Lyapunov exponent for populations with dispersal fraction d i.e.

$$\lim_{n\to\infty} \frac{1}{n} \ln \|A_d(E_{n-1}, 0) \ldots A_d(E_0, 0)\| = \gamma(d)$$

with probability one.

When $d = 0$, $A_d(e, 0)$ is a diagonal matrix with diagonal entries $\lambda_i(e)$. Hence,

$$\gamma(0) = \max_i \mathbb{E}(\ln \lambda^i).$$

By continuity of $d \mapsto \gamma(d)$ (assuming $\mathcal{E}$ is compact), $\mathbb{E}(\ln \lambda^i) > 0$ for some patch i implies that the population persists whenever d is sufficiently small. Conversely, if $\mathbb{E}(\ln \lambda^i) < 0$ for all patches, then extinction occurs for low dispersal rates. Hence, quite intuitively, for relatively sedentary populations, persistence requires that at least one of the patches is a source patch in the sense that $\mathbb{E}(\ln \lambda^i) > 0$.

At high dispersal rates, however, the existence of a source patch is no longer required for persistence. To see why, consider the limiting case of $d = 1$ which corresponds to a fully mixed population. By Theorem 5, $\gamma = \lim_{n\to\infty} \frac{1}{n} \ln \|A_1(E_{n-1}, 0) \ldots A_1(E_0, 0)\mathbf{1}\|$ with probability one where $\mathbf{1} = (1, 1, \ldots, 1)^t$. Since

$$A_1(e, 0)\mathbf{1} = \left(\frac{1}{k} \sum_{i=1}^{k} \lambda^i(e) \right) \mathbf{1}$$

these random matrices share a stable structure $\mathbf{1}$ and, consequently,

$$\gamma(1) = \mathbb{E}\left(\ln \left(\frac{1}{k} \sum_{i=1}^{k} \lambda^i \right) \right)$$

By Jensen's inequality

$$\mathbb{E}\left(\ln \left(\frac{1}{k} \sum_i \lambda^i \right) \right) \geq \frac{1}{k} \sum_i \mathbb{E}\left(\ln \left(\lambda^i \right) \right)$$

with a strict inequality if the λ^i have positive variance. Hence, even if every patch is a sink in the sense that $\mathbb{E}(\ln \lambda^i) < 0$ for all i, then $\mathbb{E}\left(\ln \left(\frac{1}{k} \sum_i \lambda^i \right) \right)$ still can be positive in which case the population can persist provided it disperses sufficiently quickly (i.e. d is sufficiently close to one).

To see how these persistence criteria play out, let us assume that all the patches experience the same distribution of environmental variability over time. Define λ^i to be lognormally distributed

random variables with log-mean μ, log-standard deviation σ, and pairwise covariance ρ^2. More precisely,

$$\lambda_n^i = \exp(\sigma E_n^i + \mu) \quad (6)$$
$$E_n^i = \rho Z_n^0 + \sqrt{1 - \rho^2} Z_n^i$$

where Z_n^i are standard normal random variables independent in time and space. All patches are sinks whenever $\mu < 0$. When there are sufficiently many patches and no spatial correlations (i.e. $\rho = 0$), the strong law of large numbers implies that

$$\mathbb{E}\left(\ln \left(\frac{1}{k} \sum_i \lambda^i \right) \right) \approx \ln(\mathbb{E}(\lambda^1)) = \mu + \sigma^2/2.$$

Hence, for highly dispersive populations in patchy environments with no spatial correlations, persistence occurs provided that

$$\sigma^2 > -2\mu. \quad (7)$$

By Jensen's inequality (7) is a necessary condition for persistence when $\rho > 0$. However, (7) may no longer be a sufficient condition as illustrated in Fig. 2. Indeed when the spatial correlation ρ is sufficiently close to 1, the dominant Lyapunov exponent is given by μ and, consequently, extinction is inevitable if all the patches are sinks. Fig. 2 illustrates that the mean population abundance tends to decrease with the spatial correlation and, consequently, we conjecture that γ is a decreasing function of ρ. The maximal variance in population abundance, however, occurs at intermediate levels of spatial correlation.

5.3. Biennial plants

Biennial plants typically flower only in the second year of their existence after which they die. However, for many biennial species, individual plants may exhibit delayed flowering in which they flower in a later year. Delayed flowering can serve as a bet-hedging strategy in uncertain environments. Roerdink (1987) provided a detailed analysis of a density-independent model of delayed flowering. Here, we show how Roerdink's analysis coupled with Theorem 1 can be used to analyze a density-dependent version of his model. Let X_n^1 denote the abundance of one year old individuals in year n and X_n^2 denote the abundance of individuals greater than one year old in year n. Let $X_n = (X_n^1, X_n^2)$. Let p be the probability that a plant flowers during its second year. Let $fs_1(x^1 + x^2)$ with $s_1(x^1, x^2) = \frac{1}{1+b_1(x^1+x^2)}$ be the mean number of germinating seeds produced by a flowering plant. Let $s_2(x^1, x^2) = \frac{a}{1+b_2(x^1+x^2)}$ be the probability that a plant survives to the next year. If we assume that $f = E_n$ at time n, then the plant dynamics are given by

$$X_{n+1} = \begin{pmatrix} 0 & p E_n s_1(X_n) \\ s_2(X_n) & (1-p) s_2(X_n) \end{pmatrix} X_n. \quad (8)$$

Following Roerdink (1987), we assume that $E_0, E_1, E_2, \ldots$ is a sequence of independent random variables that are gamma distributed with probability density function

$$g(t) = \frac{1}{\theta^b \Gamma(k)} t^{k-1} \exp(-t/\theta),$$

scale parameter $\theta > 0$, shape parameter $k > 0$ and $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$. The mean and variance of ξ_1 are given by $k\theta$ and $k\theta^2$. Appendix I shows that assumptions **A1–A6** are satisfied for this model.

Adapting ideas from Tuljapurkar (1984), Roerdink (1987) found an explicit formula for the dominant Lyapunov exponent. For $0 < p < 1$, the dominant Lyapunov exponent is given by

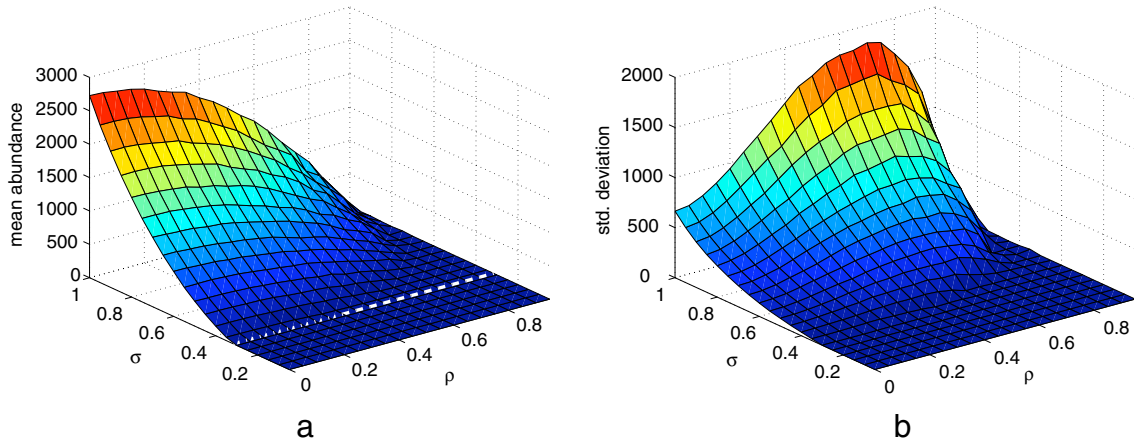$$\gamma = \ln a(1-p) + K^{-1} \int_0^\infty \ln(1+t) \, t^{k-1} (1+t)^{-k} e^{-zt} dt$$

Fig. 2. Effects of environmental noise and spatial correlations on metapopulations whose dynamics are given by (5) and (6) with $a_i = 0.01$, $\mu = -0.05$, and $d = 0.8$. In (a) and (b), mean and standard deviation of metapopulation abundance is plotted as a function of environmental variance σ^2 and spatial correlation ρ. In (a), the dashed line corresponds to extinction threshold ($\mu + \sigma^2/2 = 0$) when there are many independent patches. For each parameter combination, the population dynamics were simulated for 100,000 time steps.
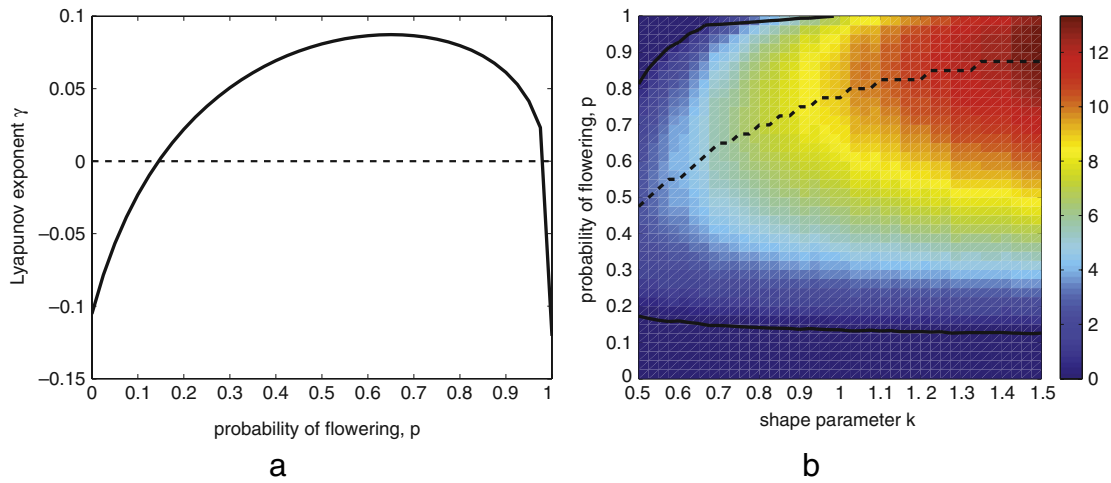

Fig. 3. Biennial plants in a random environment. In (a), the dominant Lyapunov exponent γ for (8) as a function of the fraction p of flowering second year plants. In (b), the long-term mean abundance of $\widehat{X}_1 + \widehat{X}_2$ is plotted as a function of the shape parameter k with a fixed mean fertility $k\theta = 2$. The solid line denotes where $\gamma = 0$ and the dashed line denotes the flowering probability that maximizes γ. In both figures, parameter values are $b_1 = b_2 = 0.01$, mean fertility $k\theta = 2$, and $a = 0.9$. In (a), $\theta = 0.725$.

where $K = \int_0^\infty t^{k-1}(1+t)^{-k}e^{-zt}dt$ and $z = (1-p)^2/(\theta p)$. For $p = 0$, $\gamma = \ln a$, while for $p = 1$, $\gamma = \frac{1}{2}(\ln(a\theta) + \psi(a))$ where $\psi(a)$ is the digamma function. Proposition 3 implies that γ is continuous at $p = 0$ and 1. Furthermore, Roerdink proved that $\frac{\partial \gamma}{\partial p}$ is positive at $p = 0$ and approaches $-\infty$ as p approaches 1. Hence, in the words of Roerdink, "this remarkable result implies that, independent of the... means and variances of [the fertilities]," the stochastic growth rate of the population is maximized by the population playing an appropriate bet hedging strategy for flowering (i.e. p strictly between 0 and 1). Thus, Theorem 1 implies that persistence is more likely for populations playing a bet hedging strategy (e.g., in Fig. 3a the stochastic growth rate is maximized at an intermediate p). When there is constant density dependence (i.e. $b_1 = b_2$) in which density dependence influences all classes in the same way, Fig. 3b illustrates that maximum mean population abundance appears to be maximized when γ is maximized. Moreover, the optimal strategy p decreases with the variance in the fertility (i.e. p increases with the shape parameter k).

6. Discussion

Following Chesson (1978, 1982) and Chesson and Warner (1981), we have used a stochastic boundedness criterion to judge whether a structured population living in an random environment is persistent. In the words of Chesson (1982),

> This criterion requires that the probability of observing a population below any given density, should converge to zero with density, uniformly in time. Consequently it 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.

Chesson's quote highlights two ways that stochastic boundedness from below can be interpreted. First, it places constraints on the probability of the population reaching low densities in the future. This interpretation corresponds to understanding the distribution of X_n over many realizations of the population dynamics. Second, it places constraints on the frequency that the population trajectory reaches low densities. This interpretation corresponds to understanding the behavior of the empirical measures $\Pi_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$. We show that under suitable conditions on

the form of density dependence, persistence with respect to either of these interpretations occurs when the "stochastic growth rate" of the population is positive when rare. Verifying persistence with our methods was illustrated for unstructured populations in environments with correlated noise, spatial structured populations, and biennial plant models where the stochastic growth rate can be explicitly computed. Each of these examples highlights insights that can be gleamed from analytic approaches and also highlight the challenges that confront analytical approaches to tackling the question of persistence in random environments.

For unstructured populations, the stochastic growth rate γ depends only on the distribution of the per-capita fitness values. In particular, γ does not depend on the correlative structure of the noise. However, this correlative structure can exert a strong influence on the shape of the positive stationary distribution $\widehat{X}$ of the population process. For instance, for populations with compensating density dependence, we show that that strong positive autocorrelations in the environmental noise (i.e. "red noise") result in a broad stationary distribution with weight near extinction. In contrast, our analytic approximations show that strong negative autocorrelations in the environmental noise (i.e. "blue noise") result in a stationary distribution that is clustered away from extinction.

These results are consistent with the intuitive argument of Lawton (1988) that reddened noise increases extinction risk. Namely, reddened noise is likely to contain long runs of bad conditions and while a population may survive one setback, it is unlikely to survive a long run of setbacks. Simulation studies have shown that overcompensating density dependence can reverse these trends. Namely, reddened noise can enhance population persistence (Ripa and Lundberg, 1996; Petchey et al., 1997b). However, if the noise and the overcompensating density dependence are too strong, growth-catastrophe dynamics can ensue which increases the risk of extinction (Gyllenberg et al., 1994a,b). Thus, colored noise and density-dependence can have subtle interactions that influence population persistence. To better understand this interaction, there is a need to develop analytical methods to approximate the stationary distribution for populations with overcompensating density dependence in correlated environments.

For structured populations, the stochastic growth rate γ is notoriously difficult to compute. However, in special cases it can be computed explicitly and provide useful insights about persistence. For instance, for the biennial plant model, Roerdink (1987) developed an explicit formula for γ. Despite the complexity of this formula, Roerdink (1987) used it to show that γ is *always* maximized by populations playing a bet-hedging strategy with only a fraction of the population flowering each year. By contrast, for classical unstructured models (e.g. $X_{n+1} = (s(1-p) + a_n p)X_n$ where p is the probability of a seed germinating (or the plant flowering), s is the probability of a seed surviving, and a_n is the fecundity of a plant), bet-hedging with $0 < p < 1$ maximizes the stochastic growth rate $\gamma = \mathbb{E}\left(\ln\left(s(1-p) + pa\right)\right)$ only if the arithmetic and harmonic mean of a plant's fecundity is greater than the likelihood of a seed surviving. Hence, accounting for age-structure can fundamentally change predictions about persistence in random environments. This change may get further magnified by accounting for autocorrelation in the environment.

If density dependence acts on all individuals (e.g. seeds, first years, etc.) in the same manner, it is not hard to show that a population playing a germination strategy that maximizes γ will competitively exclude any population playing a different strategy (Schreiber, unpublished notes). However, for most populations density dependence is likely to act differently on different types of individuals (especially if one considers seeds versus plants) in which case evolutionarily stable strategies

need not maximize γ, may involve of a coalition of coexisting phenotypes, or may not exist at all. To approach this issue from an analytic point of view, results for coexistence and exclusion for competing phenotypes in random environments are needed. Chesson and Ellner (1989) have provided some results in this direction for unstructured models of competing populations in serially uncorrelated environments. They show that coexistence requires mutual invasibility (i.e. each population can invade when the other population is established). Applying these criteria are difficult as they typically requiring knowing the stationary distributions of the isolated populations. In strongly correlated environments, approximations to these stationary distributions are possible. Hence, there may be analytical opportunities if the results of Chesson and Ellner are extended to correlated random environments. We are currently pursuing such results.

For our spatial model, it is possible to estimate the stochastic growth rate γ when populations are either weakly coupled by dispersal or strongly coupled by dispersal. Our results show that persistence for relatively sedentary populations requires, not surprisingly, the existence of at least one source habitat. More surprisingly, at high dispersal rates, our results show that persistence can occur even if all patches are sinks. Hence, our results analytically confirm numerical observations from previous studies of density dependent spatial models in random environments (Jansen and Yoshimura, 1998; Roy et al., 2005). Intuitively, dispersal in random environments can serve as spatial bet-hedging: individuals by spreading their risk across space increase their fitness across time. Mathematically, it follows from the observation that the geometric mean of the spatial means is greater than the spatial mean of the geometric means.

In the "classical" metapopulation scenario (i.e. many identical and independent patches) with highly dispersive populations, we show that stochastic growth rate is $\mu + \sigma^2/2$ where μ and σ^2 are the mean and variance of the logarithmic fitness. Hence, increasing within the patch variability of fitness, increases the stochastic growth rate of the population and the possibility of metapopulation persistence. Our simulation results suggest that positive spatial correlations in fitness tend to reduce mean metapopulation abundance yet increase the coefficient of variation of metapopulation abundnace. Hence, the greatest variability in metapopulation abundance occurs at intermediate levels of spatial correlations.

From the analytic perspective there are several challenging problems to pursue here. First, does the stochastic growth rate for "classical metapopulation scenario" monotonically increase with the dispersal rate as suggested by its values at low and high dispersal rates? This result can be shown for periodic environments (Li and Schreiber, unpublished notes). Furthermore, Hardin et al. (1988b, 1990) have show that for a continuous space analogue of our model that the stochastic growth rate is maximized by a dispersal kernel corresponding to the "go everywhere uniformly" dispersal strategy. Hence, it seems a positive answer to this question is likely. Another challenging question is how can one analytically estimate the mean metapopulation abundance? In the spirit of Tuljapurkar's work (1982; 1990) on low noise approximations for the stochastic growth rate, one might consider low noise approximations of the mean metapopulation abundance by finding appropriate Taylor expansions around the no-noise limit of the equilibrium abundance. However, the feasibility of this approach remains to be determined.

In conclusion, we have shown that under suitable conditions, structured populations persist (in the sense of stochastic boundedness) provided they can increase when rare. Our results also show under what conditions population trajectories converge with probability one to a unique stationary distribution. When these conditions are met, all the statistical information about the asymptotic

population dynamics can be extracted from a single long run of the model. Since the shape of these stationary distributions determines how frequently population trajectories get close to extinction, a fundamental and challenging problem for theoretical population biology is to find analytic methods to approximate these stationary distributions. We have suggested some possibilities but more methods are needed as these analytical approximations may provide general insights that elude simulation studies.

Acknowledgments

Appendix A. Markovian environments

We use the notation of Example 6. Let P be a Markov kernel over $\mathscr{S}$ with an invariant probability μ and let $(E_n)_{n \geq 0}$ the Markov chain with kernel P and initial distribution μ. As already noticed, $(E_n)_{n \geq 0}$ is stationary.

A Borel set $A \subset \mathscr{S}$ is said (P, μ) invariant if $P(x, A) = \mathbf{1}_A(x)$ μ-almost surely. By Theorem 2.1 in Kifer (1986), $(E_n)_{n \geq 0}$ is ergodic if and only if every (P, μ) invariant set has μ measure 0 or 1. In particular

Lemma 1. *Suppose μ is the unique invariant measure of P. Then $(E_n)_{n \geq 0}$ is an ergodic stationary sequence.*

Proof. Let A be (P, μ) invariant with $\mu(A) > 0$. Then the measure $\mu_A(\cdot) = \frac{\mu(A \cap \cdot)}{\mu(A)}$ is also an invariant invariant measure of P. Hence $\mu = \mu_A$ so that $\mu(A) = 1$. □

Appendix B. The random dynamical systems framework

To prove our results, it is useful to embed (2) and assumptions **A1**–**A6** within Arnold's general framework of random dynamical systems (Arnold, 1998). In this framework, let Ω be the set of all possible environmental states, $\mathbb{Q}$ be a probability measure on Ω describing the probabilities of environmental events, and let $\tau : \Omega \to \Omega$ be a invertible map describing the environmental dynamics. More specifically, if the current environmental state is e, then the environmental state in the next time step is τe. Randomness enters by choosing the initial environmental state E_0 randomly from Ω with respect to the probability distribution $\mathbb{Q}$ and defining future environmental states iteratively by

$$E_{n+1} = \tau(E_n).$$

Let A be a Borel map from $\Omega \times \mathbb{R}_+^k$ to the space of $k \times k$ non-negative matrices. Let $X_0 = x$ be the initial state of the population and let E_0 be a randomly chosen environmental state according to the probability law $\mathbb{Q}$. Then the population dynamics are given by

$$\begin{cases} E_{n+1} = \tau(E_n) & E_0 \text{ a random variable with distribution } \mathbb{Q} \\ X_{n+1} = A(E_n, X_n) X_n & X_0 = x \in \mathbb{R}_+^k. \end{cases} \quad (9)$$

Since the system dynamics are determined by $\tau : \Omega \to \Omega$, $\mathbb{Q}$, and A, we call (9) the *random dynamical system determined by* $(\tau, \mathbb{Q}, A)$.

The dynamics of (2) where $E_0, E_1, \ldots$ is an ergodic stationary sequence on $\mathscr{S}$ are recovered in this framework by setting

$$\Omega = \mathscr{S}^{\mathbb{Z}} = \{\{e_i\}_{i=-\infty}^{\infty} : e_i \in \mathscr{S}\} \qquad \tau(\{e_i\}_{i=-\infty}^{\infty}) = \{e_{i+1}\}_{i=-\infty}^{\infty},$$

defining $\mathbb{Q}$ to be the probability measure on Ω satisfying

$$\mathbb{Q}(\{e \in \Omega : e_i \in S_0, \ldots e_{i+n} \in S_n\}) = \mathbb{P}(E_0 \in S_0, \ldots, E_n \in S_n)$$

for any Borel sets $S_0, \ldots, S_n \subset \mathscr{S}$, and requiring that $A(e, x)$ depends only on e_0 and x. Due to the Polish assumption, $\mathscr{S}^{\mathbb{Z}}$ is itself Polish, its Borel sigma field equals its product sigma field; and by the Kolmogorov consistency theorem, the probability $\mathbb{Q}$ is well defined. Furthermore, By a theorem of Rokhlin (1964) τ is ergodic with respect to $\mathbb{Q}$.

To state the assumptions for this more general framework, define

$$f_e(x) = A(e, x) x$$

and

$$f_e^n(x) := (f_{\tau^{n-1}e} \circ \ldots \circ f_{\tau e} \circ f_e)(x) \quad \text{for } n \geq 1.$$

Equivalently, we can write

$$f_e^n(x) = A^n(e, x) x$$

where

$$A^n(e, x) = A(\tau^{n-1}e, f_{\tau^{n-1}x}(x)) \ldots A(\tau e, f_e(x)) A(e, x).$$

Finally, for any Borel function $h : \Omega \to \mathbb{R}$ define

$$\mathbb{E}(h) = \int_{\Omega} h(e) \, d\mathbb{Q}(e)$$

whenever the integral is well defined.

With this notation, the assumptions we place on the random dynamical system $(\tau, \mathbb{Q}, A)$ are as follows.

A1' Ω is a topological space, $\mathbb{Q}$ is a Borel probability measure, and τ is an invertible map that is ergodic with respect to $\mathbb{Q}$

A2' (Dissipativeness) There is a Borel function $r : \Omega \to [0, \infty)$ such that

$$\limsup_{n \to \infty} \|f_{\tau^{-n}e}^n(x)\| \leq r(e)$$

for all $x \in \mathbb{R}_+^k$.

A3' (Primitivity) There is a positive integer N such that $A^n(e, x)$ has all positive entries for all e and x and $n \geq N$.

A4' (Smoothness) A is Borel, the maps $x \mapsto f_e(x)$ are twice continuous differentiable for all $e \in \Omega$, and

$$\mathbb{E}\left(\sup_{\|x\| \leq 1} \ln^+ \left(\|f_e(x)\| + \| Df_e(x)\| + \| D^2 f_e(x)\|\right)\right) < +\infty.$$

A5' (Intraspecific competition) The matrix entries $A_{ij}(e, x)$ satisfy $\frac{\partial A_{ij}}{\partial x_l}(e, x) \leq 0$ for all e and x. Moreover, for each i there exists some j and l such that this inequality is strict for all e and x.

A6' (Compensating density dependence) All entries of the derivative $Df_e(x)$ of $f_e(x)$ are non-negative for all e and x.

Since every ergodic stationary processes on a topological space $\mathscr{S}$ can be described an ergodic measure preserving transformation (see above), **A1'** is less restrictive than **A1**. Assumptions **A3'**–**A6'** are simply restatements of assumptions **A3**–**A6** in the random dynamical systems framework. Since τ is $\mathbb{Q}$-invariant, we have that

$$\mathbb{Q}(\|f_{\tau^{-n}e}^n(x)\| \leq r(e)) = \mathbb{Q}(\|f_e^n(x)\| \leq r(\tau^n e)).$$

Consequently, **A2'** implies that

$$\lim_{n \to \infty} \mathbb{Q}(\|f_e^n(x)\| \leq r(\tau^n e)) = 1.$$

Moreover, since $X_n = f_{E_0}^n x$, **A2** implies that for all $M > 0$

$$\liminf_{n \to \infty} \mathbb{Q}(X_n \leq M) \geq \mathbb{Q}(r(E_0) \leq M) \geq 1 - \frac{\mathbb{E}(r)}{M} \quad (10)$$

where the last inequality is just Markov inequality. Next we show that **A2'** is less restrictive than **A2**.

Proposition 4. *Assume there exists a proper function* $V : \mathbb{R}_+^k \mapsto \mathbb{R}$ *and Borel maps* $\alpha, \beta : \Omega \mapsto [0, \infty)$ *such that*

1. $V(f_e(x)) \leq \alpha(e)V(x) + \beta(e)$;
2. $\mathbb{E}(\ln(\alpha)) < 0$
3. $\mathbb{E}(\ln^+ \beta) < \infty$

Then **A2**$'$ *is satisfied.*

Proof. Without loss of generality we assume that $V \geq 0$. Let $\alpha_i = \alpha(\tau_{-i}(e))$ and $\beta_i = \beta(\tau_{-i}(e))$. Condition (1) and an elementary recursion gives

$$V(f_{\tau^{-n}(e)}^n(x)) \leq \left(\prod_{i=1}^n \alpha_i\right) V(x) + \sum_{i=1}^n \eta_i$$

where

$$\eta_i = \beta_i \prod_{j<i} \alpha_j.$$

Condition (2) and the Birkhoff ergodic theorem give

$$\lim_{i \to \infty} \frac{\ln\left(\prod_{j<i} \alpha_j\right)}{i} = \mathbb{E}(\ln(\alpha)) < 0.$$

By condition (3) and stationarity

$$\mathbb{P}(\ln(\beta_i) \geq i\epsilon) = \mathbb{P}(\ln_+(\beta_i) \geq i\epsilon) = \mathbb{P}(\ln_+(\beta_1) \geq i\epsilon).$$

Hence

$$\sum_{i=1}^\infty \mathbb{P}(\ln(\beta_i) \geq i\epsilon) = \sum_{i=1}^\infty \mathbb{P}(\ln_+(\beta_1) \geq i\epsilon)$$

$$\leq \frac{\mathbb{E}(\ln_+(\beta)) + 1}{\epsilon} < \infty,$$

so that, by the Borel Cantelli lemma,

$$\limsup_i \frac{\ln(\beta_i)}{i} \leq 0$$

$\mathbb{Q}$-almost surely. Hence

$$\limsup_{n \to \infty} V(f_{\tau^{-n}(e)}^n(x)) \leq \sum_{i=1}^\infty \eta_i < \infty$$

$\mathbb{Q}$-almost surely. Let, for all $R > 0$, V_R denotes the supremum of $\|x\|$ over the set $\{x : V(x) \leq R\}$. Since V is proper this later set is compact and $V_R < \infty$. It now suffices to set $R(e) = \sum_{i=1}^\infty \eta_i + 1$ and $r(e) = V_{R(e)}$ to conclude that

$$\limsup_{n \to \infty} \|(f_{\tau^{-n}(e)}^n(x))\| \leq r(e)$$

$\mathbb{Q}$-almost surely. $\quad\square$

Appendix C. Proof of Theorem 1

We prove Theorem 1 using the random dynamical systems framework described in Appendix B under assumptions **A1**$'$–**A6**$'$. The idea of the proof is as follows. Imagine for each environmental state $e \in \Omega$ starting the population process in state x at time $-n$ and examining the population state at time 0. In other words, we want to understand the behavior of $f_{\tau^{-n}e}^n(x)$ for large n. The map $x \mapsto f_{\tau^{-1}e}(x)$ will be shown to be a contraction. Consequently for large n, $f_{\tau^{-n}e}^n(x)$ as $n \to \infty$ becomes asymptotically independent of $x \gg 0$. Call the limiting value $\widehat{X}(e)$. Since $\gamma > 0$, the origin tends to repel population trajectories and, consequently, we will show that $\widehat{X}(e) \gg 0$ with probability one. Pointwise convergence (in e) of $f_{\tau^{-n}e}^n(x)$ to $\widehat{X}(e)$ implies convergence of the distribution of

$f_{\tau^{-n}E_0}^n(x)$ to the distribution of $\widehat{X}$. Stationarity of the environment process implies that the distribution of $f_{\tau^{-n}E_0}^n x$ is the same as the distribution of $X_n = f_{\tau^n E_0}^n(x)$. Our proof is motivated by the approach taken by Chueshov (2002), but yields a stronger result for the models being considered here. In particular, Theorem 4.4.2 and Proposition 4.5.2 of Chueshov (2002) do not use the dominant Lyapunov exponent to determine whether $\widehat{X} \gg 0$ or not, and if $\widehat{X} \gg 0$, Theorem 4.4.2 and Proposition 4.5.2 of Chueshov only ensure convergence of X_0 provided that there exists $\lambda > 1$ such that $\widehat{X}(e)/\lambda < X_0(e) < \lambda \widehat{X}(e)$ for almost every $e \in \Omega$.

Assume $\gamma > 0$. Assumption **A6**$'$ implies that f_e is order preserving i.e $f_e(y) \geq f_e(x)$ for all $e \in \Omega$ and $y \geq x$. A key result that we need is the following proposition which follows from Ruelle's unstable manifold theorem (Ruelle, 1982). We let $\text{int } \mathbb{R}_+^k = \{x \in \mathbb{R}_+^k : x_i > 0\}$.

Proposition 5. *Assume* **A1**$'$–**A4**$'$ *and* $\gamma > 0$. *Then there exists an* τ-*invariant set* $\widetilde{\Omega} \subset \Omega$ *such that* $\mathbb{Q}(\Omega) = 1$, *and Borel functions,* $u_n : \Omega \to \text{int } \mathbb{R}_+^k, \alpha : \Omega \to (0, \infty)$ *such that*

- $f_{\tau^{-n}e} u_n(e) = u_{n-1}(e)$. *In particular,* $f_{\tau^{-n}e}^n u_n(e) = u_0(e)$.
- $\|u_n(e)\| \leq \alpha(e) \exp(-n0.99\gamma)$ *for all* $e \in \widetilde{\Omega}$ *and* $n \geq 0$.

Since this proposition does not follow directly from any of the theorem statements in Ruelle's paper, we provide an outline of the proof.

Proof. Assumption **A4**$'$ implies that

$$\mathbb{E}(\ln^+ \|A(e, 0)\|) < +\infty.$$

Hence, by Kingman's subadditive ergodic theorem

$$\gamma = \lim_{n \to \infty} \frac{1}{n} \ln \|A^n(e, 0)\|$$

with probability one. Define $B^n(e, 0) = A^*(\tau^{-n}e, 0) \dots A^*(\tau^{-1}e, 0)$ where $*$ denotes the transpose of a matrix. Since $B^n(e, 0)^* = A^n(\tau^{-n}e, 0)$, the dominant Lyapunov exponent of B equals γ i.e.

$$\gamma = \lim_{n \to \infty} \frac{1}{n} \ln \|B^n(e, 0)\|$$

with probability one. Theorem 5 and assumption **A3**$'$ imply that there exists a Borel set $\Omega_1 \subset \Omega$ and Borel function $V : \Omega \to \mathbb{R}_+^k$ such that $\mathbb{Q}(\Omega_1) = 1, \tau\Omega_1 = \Omega_1, \|V(e)\| = 1, V(e) \gg 0$, $V(\tau^{-1}e) = \frac{B(e,0)V(e)}{\|B(e,0)V(e)\|}$ for all $e \in \Omega_1, \gamma = \lim_{n\to\infty} \frac{1}{n} \ln \|B^n(e, 0)w\|$ for all $e \in \Omega_1, w > 0$, and

$$\lim_{n \to \infty} \frac{B^n(e, 0)w}{\|B^n(e, 0)w\|} \cdot V(\tau^{-n}e)^\perp = 0$$

for all $e \in \Omega_1$ and $w > 0$. Propositions 3.2 and 2.1 of Ruelle (1982), and Birkoff's ergodic theorem in conjunction with assumption **A4**$'$ imply that there exists a Borel set $\Omega_2 \subset \Omega_1$ such that

- $\mathbb{Q}(\Omega_2) = 1$ and $\tau\Omega_2 = \Omega_2$
- conditions (S) and equations (5.1)–(5.2) in Ruelle hold for $T_e^n = B^n(e, 0)$ for all $e \in \Omega_2$, and
- $\lim_{n\to\infty} \frac{1}{n} \sup_{\|x\|=1} \ln^+(\|f_{\tau^{-n}e}(x)\| + \|Df_{\tau^{-n}e}(x)\| + \|D^2 f_{\tau^{-n}e}(x)\|)$ $= 0$ for all $e \in \Omega_2$.

For any $t > 0$, let $D(t) = \{x \in \mathbb{R}_+^k : \|x\| \leq t\}$. It follows from the proof of Theorem 6.1 of Ruelle (see the last sentence of Section 6 and the Remark 5.3a) for each $e \in \Omega_2$ there exists $\alpha(e)$ and $\beta(e)$ such that the set

$$\mathcal{V}_e = \{u_0 \in D(\beta(e)) : \exists (u_n)_{n\geq0} \text{ with } f_{\tau^{-n}e} u_n = u_{n-1}$$
$$\text{and } \|u_n\| \leq \alpha(e) \exp(-n\,0.99\,\gamma)\}$$

is a one-dimensional submanifold of $D(\beta(e))$ tangent at 0 to $V(e)$ and (u_n) is uniquely determined by u_0. Let $u_0(e)$ be $\mathcal{V}_e \cap D(\beta(e))$ and

for $n \geq 1$, let $u_n(e)$ be the unique sequence such that $f^n_{\tau^{-n}e} u_n(e) = u_0(e)$ and $\|u_n(e)\| \leq \alpha(e) \exp(-n\, 0.99\, \gamma)$ for all $n \geq 0$. By the (essentially) explicit constructions of $\mathcal{V}_e$, $u_n(e)$, $\alpha(e)$ and $\beta(e)$ are measurable. $\quad\square$

Let $x \gg 0$ be given and define

$$\Gamma_x(e) = \cap_{n \geq 1} \overline{\cup_{m \geq n} f^m_{\tau^{-m}e} x}$$

to be the omega-limit set of x. Assumption **A2$'$** and Proposition 1.6.2 in Chueshov (2002) imply that $\Gamma_x(e)$ is a compact invariant set i.e. $f_e(\Gamma_x(e)) = \Gamma_x(\tau e)$ and $\Gamma_x(e)$ is compact. Let $\widetilde{\Omega} \subset \Omega$ and $u_n : \Omega \to \mathrm{int}\, \mathbb{R}^k_+$ be as given by Proposition 5. For each $e \in \widetilde{\Omega}$, Proposition 5 and our assumption that $\gamma > 0$ imply that there exists an $M = M(e) \geq 0$ such that $u_n(e) \leq x$ for all $n \geq M$. By monotonicity,

$$f^n_{\tau^{-n}e}(x) \geq f^n_{\tau^{-n}e}(u_n(e)) = u_0(e) \gg 0 \quad \text{for } n \geq M(e), e \in \widetilde{\Omega}.$$

Hence, $\Gamma_x(e) \geq u_0(e) \gg 0$ for all $e \in \widetilde{\Omega}$.

Define the Birkhoff part metric p on $\mathrm{int}\, \mathbb{R}^k_+$ by

$$p(y, z) = \max_i |\ln y_i/z_i| = \min\{\ln a : z/a \leq y \leq a\, z\}.$$

We need the following proposition which is a standard result in the theory of monotone dynamical systems.

Proposition 6. *Let $g : \mathbb{R}^k_+ \to \mathbb{R}^k_+$ be an order-preserving C^1 map. If $Dg(y)y \ll g(y)$ for all $y \gg 0$, then*

$$p(g(y), g(z)) < p(y, z)$$

for all $y, z \in \mathrm{int}\, \mathbb{R}^k_+$ such that $z \neq y$.

Proof. $Dg(y)y \ll g(y)$ for all $y \gg 0$ implies that

$$\frac{d}{da}\frac{g(ay)}{a} = \frac{1}{a^2}\left(Dg(ay)ay - g(ay)\right) \ll 0$$

for all $y \gg 0$ and $a > 0$. Hence, $g(y) - g(by)/b = \int_b^1 \frac{d}{da}\frac{g(ay)}{a} da \ll 0$ for $b \in (0, 1)$ and g is strongly sublinear i.e. $b\, g(y) \ll g(by)$ for any $y \gg 0$ and $b \in (0, 1)$. Let $y \neq z \in \mathrm{int}\, \mathbb{R}^k_+$. Let $a > 1$ be such that $p(y, z) = \ln a$. Then $z/a \leq y \leq a\, z$. Since g is order preserving and strongly sublinear, it follows that $g(y) \geq g(z/a) \gg g(z)/a$ and $g(z) \geq g(y/a) \gg g(y)/a$. Hence, $g(z)/a \ll g(y) \ll a\, g(z)$ and $p(g(y), g(z)) < p(y, z)$. $\quad\square$

To apply this proposition to $g = f_e$, assumption **A5** implies that for $x \gg 0$,

$$Df_e(y)y = \sum_i y_i \frac{\partial A}{\partial x_i}(e, y)y + A(e, y)y$$
$$\ll A(e, y)y.$$

Hence Proposition 6 implies that for all $n \geq 1$, $p(f^n_e y, f^n_e z) \leq p(y, z)$ with equality if and only if $y = z$. Define the Borel function $h : \Omega \to \mathbb{R}$ by

$$h(e) = \max\{p(y, z) : y, z \in \Gamma_x(e)\}.$$

Since the part metric is contracting and Γ_x is invariant, $h(\tau e) \leq h(e)$ for all e and with equality if and only if $h(e) = 0$. By τ-invariance of $\mathbb{Q}$,

$$\int h(e) - h(\tau e) d\mathbb{Q}(e) = 0.$$

Hence, $h(e) = 0$ $\mathbb{Q}$-almost surely. Thus, $\Gamma_x(e)$ is a single point $\mathbb{Q}$-almost surely. Moreover, for our chosen x, the definition of $\Gamma_x(e)$ implies that

$$\lim_{n \to \infty} f^n_{\tau^{-n}e}(x) = \Gamma_x(e)$$

for almost every e.

Define $\widehat{X}(e) = \Gamma_x(e)$ whenever $\Gamma_x(e)$ is a point and 0 otherwise. $\widehat{X}(e)$ is strongly positive (i.e. $\widehat{X}(e) \gg 0$ almost surely) and is a random equilibrium i.e. $f_e \widehat{X}(e) = \widehat{X}(\tau e)$ almost surely. To see that $\widehat{X}(e)$ is unique (i.e. independent of the choice of $x \gg 0$), assume that $\widehat{Y}(e)$ is a strongly positive random equilibrium. Define

$$h(e) = p(\widehat{X}(e), \widehat{Y}(e)).$$

Since the part metric is a contraction and $\widehat{X}$ and $\widehat{Y}$ are random equilibria, $h(\tau e) \leq h(e)$ almost surely and with equality if and only if $h(e) = 0$. As $\int h(e) - h(\tau e) d\mathbb{Q}(e) = 0$, $h(e) = 0$ almost surely and $\widehat{X}$ is unique up to a set of measure zero. Thus, we have shown that there exists $\widehat{X}$ such that $\widehat{X} \gg 0$ almost surely and

$$\lim_{n \to \infty} f^n_{\tau^{-n}e}(e)x = \widehat{X}(e)$$

almost surely. Since τ is measure preserving and $X_n = f^n_{E_0}(x)$ where E_0 is a random variable with distribution $\mathbb{Q}$,

$$\int h(X_n) d\mathbb{Q}(e) = \int h(f^n_{\tau^{-n}e} xB) d\mathbb{Q}(e)$$

for any continuous bounded function $h : \mathbb{R}^k_+ \to R$, $x \in \mathbb{R}^k_+$, and $n \geq 0$. Hence,

$$\lim_{n \to \infty} \int h(X_n) d\mathbb{Q}(e) = \int h(\widehat{X}(e)) d\mathbb{Q}(e)$$

for any continuous bounded function i.e. X_n converges to $\widehat{X}$ in distribution.

Appendix D. Proof of Theorem 2

We prove Theorem 2 using the random dynamical systems framework described in Appendix B. We assume that **A1$'$–A4$'$** holds for $(\tau, \mathbb{Q}, A)$ and **A1$'$–A4$'$**, **A6$'$** hold for $(\tau, \mathbb{Q}, B)$. Furthermore, we assume that $\gamma > 0$ for $(\tau, \mathbb{Q}, B)$ and $A(e, x)x \geq B(e, x)x$ for all $e \in \Omega$ and $x \in \mathbb{R}^k_+$.

Let $f_e(x) = A(e, x)x$ and $g_e(x) = B(e, x)x$. The proof of Theorem 1 implies that there is a random positive vector $\widehat{X} : \Omega \to \mathbb{R}^k_+$ such that

$$\lim_{n \to \infty} f^n_{\tau^{-n}e}(x) = \widehat{X}(e)$$

for almost every e. The assumption that $B(e, x) \geq A(e, x)$ for all x, e implies that $g^n_e(x) \geq f^n_e(x)$ for all x, e. Indeed, $g_e(x) = B(e, x)x \geq A(e, x)x = f_e(x)$. Next, assume that $g^n_e(x) \geq f^n_e(x)$ for all e, x. Since f_e is order preserving, it follows that

$$g^{n+1}_e(x) = g_{\tau^n e} \circ g^n_e(x)$$
$$\geq f_{\tau^n e} \circ g^n_e(x)$$
$$\geq f_{\tau^n e} \circ f^n_e(x) = f^{n+1}_e(x).$$

Therefore

$$\liminf_{n \to \infty} g^n_{\tau^{-n}e}(x) \geq \widehat{X}(e) \quad \text{for almost every } e$$

where the lim inf is taken component-wise. Let $h : \mathbb{R}^k_+ \to \mathbb{R}$ be a bounded continuous function such that $h(y) \geq h(x)$ whenever $y \geq x$. Since τ is measure preserving, $X_n = g^n_{E_0}(x)$ has the same distribution as $g^n_{\tau^{-n}E_0}(x)$. Fatou's lemma and monotonicity of h implies

$$\liminf_{n \to \infty} \mathbb{E}\left(h(X_n)\right) = \liminf_{n \to \infty} \mathbb{E}\left(h(g^n_{\tau^{-n}E_0}(x))\right)$$
$$\geq \mathbb{E}\left(\liminf_{n \to \infty} h(g^n_{\tau^{-n}E_0}(x))\right)$$
$$\geq \mathbb{E}\left(h(\liminf_{n \to \infty} g^n_{\tau^{-n}E_0}(x))\right) \geq \mathbb{E}\left(h(\widehat{X})\right).$$

Appendix E. Proof of Theorem 3

We prove Theorem 3 using the random dynamical systems framework described in Appendix B under assumptions **A1$'$–A6$'$**.

Lemma 2. *Let $h : \mathbb{R}^k_+ \mapsto \mathbb{R}$ be a continuous function with compact support contained in $\mathrm{int}\, \mathbb{R}^k_+$, and*

$$\widetilde{\Omega} = \{e \in \Omega : \forall x \in \mathrm{int}\, \mathbb{R}^k_+ \lim_{n \to \infty} |h(f^n_e(x)) - h(\widehat{X}(\tau^n(e)))| = 0\}.$$

Then $\mathbb{Q}(\widetilde{\Omega}) = 1$.

Proof. • We first prove that $\lim_{n\to\infty} p(f_e^n(x)), \widehat{X}(\tau^n(e)) = 0$ almost surely for $x \gg 0$.

Set $V_n(e) = \min(1, p(f_e^n(x), \widehat{X}(\tau^n(e))))$. By Proposition 6, (V_n) decreases almost surely, so that we can define $V(e) = \lim_{n\to\infty} V_n(e) \geq 0$. By monotone convergence $\mathbb{E}(V) = \lim_n \mathbb{E}(V_n)$. Now, by invariance of $\mathbb{Q}$ under τ,

$$\mathbb{E}(V_n) = \int \min(1, p(f_e^n(x), \widehat{X}(\tau^n(e))))d\mathbb{Q}(e)$$
$$= \int \min(1, p(f_{\tau^{-n}(e)}^n(x), \widehat{X}(e)))d\mathbb{Q}(e).$$

Thus $\lim_{n\to\infty} \mathbb{E}(V_n) = 0$ by Lebesgue convergence theorem and the fact that $\lim_{n\to\infty} f_{\tau^{-n}(e)}^n(x) = \widehat{X}(e)$ as previously shown. It follows that $\mathbb{E}(V) = 0$, hence V, is null.

• Let, for $i \in \mathbb{N}^*$, $\underline{x}(i) = (1/i, \ldots, 1/i) \in \mathbb{R}^k$ and $\overline{x}(i) = (i, \ldots, i) \in \mathbb{R}^k$. From what precedes, the set

$$\tilde{\Omega} = \{e \in \Omega : \forall i \in \mathbb{N}^*, \lim_{n\to\infty} p(f_e^n(\underline{x}(i), \widehat{X}(\tau^n(e))))$$
$$= \lim_{n\to\infty} p(f_e(\overline{x}(i), \widehat{X}(\tau^n(e)))) = 0\}$$

has full measure as a countable intersection of full measure sets.

For any $x \gg 0$ there is some $i \in \mathbb{N}^*$ such that $\underline{x}(i) \leq x \leq \overline{x}(i)$, so that, by monotonicity,

$$f_e^n(\underline{x}(i)) \leq f_e^n(x) \leq f_e^n(\overline{x}(i)).$$

Now, it is easily seen that p enjoys the two following properties:

1.
   $$\forall u, v, w \in \text{int } \mathbb{R}_+^k, p(u, w) \leq p(u, v) + p(v, w).$$
2.
   $$\forall \underline{u}, u, \overline{u} \in \text{int } \mathbb{R}_+^k, \underline{u} \leq u \leq \overline{u} \Rightarrow p(\underline{u}, u) \leq p(\underline{u}, \overline{u})$$

Hence

$$p(f_e^n(x), \widehat{X}(\tau^n(e))) \leq p(f_n^e(x), f_n^e(\underline{x}(i))) + p(f_n^e(\underline{x}(i)), \widehat{X}(\tau^n(e)))$$
$$\leq 2p(f_n^e(\underline{x}(i)), \widehat{X}(\tau^n(e))) + p(f_n^e(\overline{x}(i)), \widehat{X}(\tau^n(e))).$$

It then follows that

$$\lim_{n\to\infty} p(f_e^n(x), \widehat{X}(\tau^n(e))) = 0$$

for all $e \in \tilde{\Omega}$ and **for all $x \gg 0$.**

• For $R > 0$, set $B_R = [1/R, R]^k$. Elementary properties of p imply that:
1. For all $x, y \in B_R$,
   $$p(x, y) \geq 1/R \|x - y\|$$
   with $\|x\| = \max_i |x_i|$;
2.
   $$\{x \gg 0 : \exists y \in B_R \text{ such that } p(x, y) \leq 1\} \subset B_{R\exp(1)}.$$

Let now h be a continuous function with support in B_R. Uniform continuity of h and these latter properties imply that for all $\epsilon > 0$ there exists $\alpha > 0$ such that

$$p(x, y) \leq \alpha \Rightarrow |h(x) - h(y)| \leq \epsilon.$$

Hence

$$\lim_{n\to\infty} |h(f_e^n(x)) - h\widehat{X}(\tau^n(e))| = 0$$

for all $e \in \tilde{\Omega}$ and **for all $x \gg 0$.** $\square$

Lemma 3. *Let (ν_n) be a sequence of probability measures on $\mathbb{R}_+^k$ and ν a probability measure on $\mathbb{R}_+^k$ such that $\nu(\text{int } \mathbb{R}_+^k) = 1$. Then $\nu_n \to \nu$ weakly if and only if $\nu_n h \to \nu h$ for every continuous function h with compact support contained in $\text{int } \mathbb{R}_+^k$. Here νh stands for $\int h d\nu$.*

Proof. The first implication is obvious. We prove the second. Let g_R be any continuous function with $0 \leq g_R \leq 1, g_R = 1$ on $B_R = [1/R, R]^k$ and $g_R = 0$ on $\mathbb{R}_+^k \setminus B_{R+1}$. Fix $\epsilon > 0$. Define $\partial \mathbb{R}_+^k = \{x \in \mathbb{R}_+^k : \prod_i x^i = 0\}$. Since $\nu(\partial \mathbb{R}_+^k) = 0$ there exists $R > 0$ such that $\nu(1 - g_R) \leq \epsilon$. Since $\nu_n g_R \to \nu g_R$ we get that $\nu_n(1 - g_R) \leq \epsilon$ for n large enough. Let now h be any bounded and continuous function. One has for n large enough

$$|\nu_n h - \nu h| \leq |\nu_n(hg_R) - \nu(hg_R)| + 2\|h\|\epsilon.$$

By hypothesis, the first term in the right-hand side goes to zero as $n \to \infty$ and since ϵ is arbitrary, this proves that

$$\nu_n h \to \nu h. \quad \square$$

We now prove the theorem. Let h be continuous with compact support contained in $\text{int } \mathbb{R}_+^k$. By the Birkhoff ergodic theorem

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n h(\widehat{X}(\tau^n(e))) = \mathbb{E}(h(\widehat{X}))$$

so that, by Cesaro convergence, and Lemma 2 we get

$$\lim_{n\to\infty} \Pi_n h = \mathbb{E}(h(\widehat{X}))$$

almost surely. Using Lemma 3 concludes the proof.

Appendix F. Proof of Theorem 4

In this section it is assumed that $\Omega = \mathcal{S}^{\mathbb{Z}}$, $(E_n)_{n\in\mathbb{Z}}$ is an ergodic stationary process having the law $\mathbb{Q}$, $(F_n)_{n\geq 0}$ is an asymptotically ergodic-stationary process taking values in $\mathcal{S}$ with limiting process $(E_n)_{n\geq 0}$; and $f_e(x) = A(e_0, x)x$ for $e = \{e_i\}_{i=-\infty}^\infty$. The key tool to the proof of Theorem 4 is Goldstein's maximal coupling theorem (see Lindvall (2002), Theorem 14.10), from which the following proposition follows directly.

Proposition 7. *There exists a probability space $(\Omega', \mathcal{F}', \mathbb{Q}')$, two $\mathcal{S}$-valued random processes $(E_n')_{n\geq 0}$, $(F_n')_{n\geq 0}$ and a $\mathbb{N}$-valued random variable T all defined on $(\Omega', \mathcal{F}', \mathbb{Q}')$ such that*

1. *Processes $(E_n)_{n\geq 0}$ and $(E_n')_{n\geq 0}$ (respectively $(F_n)_{n\geq 0}$ and $(F_n')_{n\geq 0}$) are equally distributed;*
2. *$F_n' = E_n'$ for $n \geq T$.*
3. *For all Borel sets $A_1, \ldots, A_i$,*

   $$|\mathbb{Q}(F_{n+1} \in A_1, \ldots, F_{n+i} \in A_i) - \mathbb{Q}(E_1 \in A_1, \ldots, E_i \in A_i)|$$
   $$= |\mathbb{Q}'(F_{n+1}' \in A_1, \ldots, F_{n+i}' \in A_i)$$
   $$- \mathbb{Q}'(E_1' \in A_1, \ldots, E_i' \in A_i)| = 2\mathbb{Q}'(T > n).$$

Without loss of generality we may assume that $E' = E, F' = F, \mathbb{Q}' = \mathbb{Q}$, etc.

Let now (Y_n) be solution to the recursion

$$Y_{n+1} = A(F_n, Y_n)(Y_n),$$

with $Y_0 = y$.

Let h be a continuous function with compact support in $\text{int } \mathbb{R}_+^k$ and $\Omega^* = \bigcap_{i\in\mathbb{N}} \tau^{-i}(\tilde{\Omega})$ where $\tilde{\Omega}$ is like in Lemma 2. By Lemma 2 and invariance of $\mathbb{Q}$ under τ, Ω^* has full measure and for all $e \in \Omega^*$, $i \in \mathbb{N}$ and $z \gg 0$

$$\lim_{n\to\infty} h(f_{\tau^i(e)}^{n-i}(z)) - h(\widehat{X}(\tau^n(e))) = 0.$$

Thus

$$\lim_{n\to\infty} h(f_{\tau^i(E)}^{n-i}(f_F^i(y))) - h(\widehat{X}(\tau^n(E))) = 0$$

$\mathbb{Q}$ almost surely for all $y \gg 0$.

Since f_F^i depends only on $F_0, \ldots F_i$ and $f_{\tau^i(E)}^{n-i}$ only on $E_i, \ldots, E_n$,

$$f_{\tau^i(E)}^{n-i}(f_F^i(y)) = f_{\tau^i(F)}^{n-i}(f_F^i(y))$$

almost surely on the event $T = i$. Hence

$$\lim_{n \to \infty} h(Y_n) - h(\widehat{X} \circ \tau^n) = 0$$

almost surely. Convergence of distribution of (Y_n) then follows from the stationarity of $(\widehat{X} \circ \tau^n)$ and Lemma 3. Convergence almost sure of the empirical measure follows from the Birkhoff theorem applied to $(\widehat{X} \circ \tau^n)$ (exactly like in the proof of Theorem 3).

Appendix G. Proof of Proposition 2

Let $X_0 > 0$ be any positive vector. Define

$$X_n = A^n(e)X_0 = A(\tau^n e) \ldots A(e)X_0$$

and

$$a(n) = \prod_{i=1}^k X_n^i.$$

Since $A(e)$ and X_n are non-negative,

$$
\begin{aligned}
a(n+1) &= \prod_{i=1}^k \sum_{j=1}^k a_{ij}(\tau^n e)X_n^j \\
&\geq \operatorname{perm}(A(\tau^n e))a(n) \\
&\geq \prod_{i=0}^n \operatorname{perm}(A(\tau^i e))a(0)
\end{aligned}
$$

where the final inequality follows from induction on n. By the Birkhoff ergodic theorem

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^n \ln(\operatorname{perm}(A(\tau^i e))) = \mathbb{E}(\ln(\operatorname{perm}(A)))$$

with probability one. On the other hand, Theorem 5 implies that

$$\lim_{n \to \infty} \frac{1}{n} \ln X_i(n) = \gamma$$

for all $1 \leq i \leq k$ with probability one. Hence,

$$\mathbb{E}(\ln(\operatorname{perm}(A))) \leq \lim_{n \to \infty} \frac{1}{n} \ln a(n) = k\gamma$$

with probability one.

Appendix H. Proof of Proposition 3

Kingman's subadditive ergodic theorem implies that

$$\gamma(A) = \inf_{n \geq 1} \frac{1}{n} \mathbb{E}(\ln \|A^n\|).$$

Hence, $\gamma(A)$ is upper semi-continuous. To see lower semi-continuity, notice that for any $\epsilon > 0$

$$
\begin{aligned}
\gamma(B) &= \inf_{n \geq 1} \frac{1}{n} \mathbb{E}(\ln \|B^n\|) \geq (1-\epsilon) \inf_{n \geq 1} \frac{1}{n} \mathbb{E}(\ln \|A^n\|) \\
&\geq (1-\epsilon)\gamma(A)
\end{aligned}
$$

for any $B : \Omega \to$ non-negative matrices such that $B(e) \geq (1-\epsilon)A(e)$ for all $e \in \Omega$.

Appendix I. Verification of A1–A6 for the applications

For the Beverton–Holt model, we verify the assumptions **A1**–**A6** as follows. To verify boundedness, notice that $\frac{\lambda x}{1+ax} \leq \frac{\lambda}{a}$ for all $x \geq 0$. Hence, if

$$\mathbb{E}\left(\frac{\lambda_n}{a_n}\right) < +\infty \tag{11}$$

then Assumption **A2** is satisfied with $V(x) = x$, $\alpha(e) = 0$ and $\beta(e) = e^1/e^2$. Since $A(e, x) = \frac{e^1}{1+e^2 x} > 0$ for all x and e, **A3** is satisfied. A quick computation reveals that

$$\left|\frac{d}{dx}\frac{\lambda x}{1+ax}\right| \leq \lambda \quad \text{and} \quad \left|\frac{d^2}{dx^2}\frac{\lambda x}{1+ax}\right| \leq 2\lambda a$$

for all $x \geq 0$. Hence, **A4** is satisfied provided that

$$\mathbb{E}\left(\ln^+(\lambda_n/a_n + \lambda_n + 2\lambda_n a_n)\right) < \infty. \tag{12}$$

Assumptions **A5** and **A6** follow from the observations that $\frac{d}{dx}\frac{\lambda}{1+ax} < 0$ and $\frac{d}{dx}\frac{\lambda x}{1+ax} > 0$. By Jensen's inequality, (12) implies (11). Thus, our theorems apply provided that (12) holds.

For the metapopulation model, we verify the assumptions **A1**–**A6** as follows. Define

$$A_d(e, x) = ((1-d)I_k + dJ_k) \operatorname{diag}\left\{\frac{e_1}{1+a_1 x^1}, \ldots, \frac{e_k}{1+a_k x^k}\right\}.$$

Using $V(x) = x^1 + \cdots + x^k$, $\alpha(e) = 0$ and $\beta(e) = \max_i e_i/a_i$, **A2** is satisfied provided that

$$\mathbb{E}\left(\max_i \lambda_i/a_i\right) < \infty.$$

Provided that $d > 0$, the primitivity assumption **A3** is satisfied with $N = 1$. The smoothness assumption **A4** is satisfied provided that

$$\mathbb{E}\left(\ln^+\left(\max_i \lambda_i/a_i + \max_i \lambda_i + \max_i \lambda_i a_i\right)\right) < \infty.$$

Assumptions **A5** is satisfied provided that $d > 0$. **A6** is always satisfied.

For the biennial model, we verify the assumptions as follows. To verify (8) **A2**, we follow the notation of Example 3 in Section 2 and define

$$f_e(x^1, x^2) = \begin{pmatrix} 0 & p\, e s_1(x^1, x^2) \\ s_2(x^1, x^2) & (1-p)s_2(x^1, x^2) \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \end{pmatrix}$$

Set $V(x^1, x^2) = x^1 + x^2$ and $s(t) = \frac{\max(1,a)}{1+\min(b_1,b_2)t}$. A rough estimate gives

$$V(f_e(x^1, x^2)) \leq (1+e)V(x^1, x^2)s(V(x^1, x^2)).$$

Let $\epsilon > 0$ be such that

$$\ln \epsilon + \mathbb{E}E_0 = \ln \epsilon + k\theta < 0$$

and $C > 0$ such that $s(t) < \epsilon$ whenever $t > C$. Then

$$V(f_e(x^1, x^2)) \leq \epsilon(1+e)V(x^1, x^2) + (1+e)C\, s(C).$$

Hence, **A2** holds with $\alpha(e) = \epsilon(1+e)$ and $\beta(e) = (1+e)Cs(C)$. The remaining assumptions **A1**, **A3**–**A6**, are easily verified for this model.

References

Arnold, L., 1998. Random dynamical systems. In: Springer monographs in mathematics. Springer-Verlag, Berlin.

Arnold, L., Gundlach, V.M., Demetrius, L., 1994. Evolutionary formalism for products of positive random matrices. Ann. Appl. Probab. 4 (3), 859–901.

Athreya, K.B., Dai, J., 2000. Random logistic maps I. J. Theoret. Probab. 13, 595–608.

Box, G.E.P., Jenkins, G., 1990. Times Series Analysis, Forecasting and Control. Holden-Day, Incorporated.

Bulmer, M.G., 1985. Selection for iteroparity in a variable environment. Amer. Nat. 126 (1), 63–71. URL: http://www.jstor.org/stable/2461562.

Caswell, H., 2001. Matrix Population Models. Sinauer, Sunderland, Massachuesetts.

Charlesworth, B., 1994. Evolution in age-structured populations. In: Cambridge studies in Mathematical Biology, 2nd ed. Cambridge University Press.

Chesson, P.L., 1978. Predator-prey theory and variability. Annu. Rev. Ecol. Syst. 9, 323–347.

Chesson, P.L., 1982. The stabilizing effect of a random environment. J. Math. Biol. 15 (1), 1–36.

Chesson, P.L., 1984. Persistence of a Markovian population in a patchy environment. Z. Wahrsch. Verw. Gebiete 66 (1), 97–107.

Chesson, P.L., Ellner, S., 1989. Invasibility and stochastic boundedness in monotonic competition models. J. Math. Biol. 27, 117–138.

Chesson, P.L., Warner, R.R., 1981. Environmental variability promotes coexistence in lottery competitive systems. Amer. Nat. 117 (6), 923. URL: http://www.journals.uchicago.edu/doi/abs/10.1086/283778.

Chueshov, I., 2002. Monotone random systems theory and applications. In: Lecture Notes in Mathematics, Vol. 1779. Springer-Verlag, Berlin.

Cohen, D., 1966. Optimizing reproduction in a randomly varying environment. J. Theoret. Biol. 12, 119–129.

Cuddington, K.M., Yodzis, P., 1999. Black noise and population persistence. Proc Biol Sci. 266, 969–973.

Duflo, M., 1997. Random Iterative Models. Springer Verlag.

Ellner, S.P., 1984. Asymptotic behavior of some stochastic difference equation population models. J. Math. Biol. 19, 169–200.

Ferriere, R., Gatto, M., 1995. Lyapunov exponents and the mathematics of invasion in oscillatory or chaotic populations. Theoret. Popul. Biol. 48, 126–171.

Gillespie, J.H., 1973. Polymorphism in random environments. Theoret. Popul. Biol. 4, 193–195.

Gonzalez, A., Holt, R.D., 2002. The inflationary effects of environmental fluctuations in source-sink systems. Proc. Nat. Acad. Sci. 99, 14872–14877.

Gyllenberg, M., Hognas, G., Koski, T., 1994a. Null recurrence in a stochastic Ricker model. In: Analysis, Algebra, and Computers in Mathematical Research (Lulea, 1992). In: Lecture Notes in Pure and Applied Mathematics, Decker, New York, pp. 147–164.

Gyllenberg, M., Hognas, G., Koski, T., 1994b. Population models with environmental stochasticity. J. Math. Biol. 32, 93–108.

Halley, J.M., 1996. Ecology, evolution and noise. Trends in Ecology & Evolution 11, 33–37.

Hanski, I., 1999. Metapopulation Ecology. In: Oxford Series in Ecology and Evolution. Oxford University Press.

Hardin, D.P., Takáč, P., Webb, G.F., 1988a. Asymptotic properties of a continuous-space discrete-time population model in a random environment. J. Math. Biol. 26 (4), 361–374.

Hardin, D.P., Takáč, P., Webb, G.F., 1988b. A comparison of dispersal strategies for survival of spatially heterogeneous populations. SIAM J. Appl. Math. 48 (6), 1396–1423.

Hardin, D.P., Takáč, P., Webb, G.F., 1990. Dispersion population models discrete in time and continuous in space. J. Math. Biol. 28 (1), 1–20.

Haskell, C., Sacker, R.J., 2005. The stochastic Beverton-Holt equation and the M. Neubert conjecture. J. Dynam. Differential Equations 17 (4), 825–844.

Heino, M., Ripa, J., Kaitala, V., 2000. Extinction risk under coloured environmental noise. Ecography 23, 177–184.

Holt, R.D., Gomulkiewicz, R., Barfield, M., 2003. The phenomenology of niche evolution via quantitative traits in a 'black-hole' sink. Proc. Roy. Soc. B 270, 215–224.

Horn, R.A., Johnson, C.R., 1990. Matrix Analysis. Cambridge University Press, Cambridge, corrected reprint of the 1985 original.

Jansen, V.A.A., Yoshimura, J., 1998. Populations can persist in an environment consisting of sink habitats only. Proc. Natl. Acad. Sci. USA 95, 3696–3698.

Key, E.S., 1987. Computable examples of the maximal lyapunov exponent. Probab. Th. Rel. Fields 75, 97–107.

Kifer, Yuri, 1986. Ergodic Theory of Random Transformations. In: Progress in Probability and Statistics, vol. 10. Birkhäuser Boston Inc., Boston, MA, ISBN: 0-8176-3319-7, p. x+210.

Kingman, J.F.C., 1973. Subadditive ergodic theory. Ann. Probab. 1, 883–909.

Lawton, J.H., 1988. More time means more variation. Nature 334, 563.

Lindvall, T., 2002. Lectures on the Coupling Method. Dover.

Mañé, R., 1983. Ergodic Theory and Differentiable Dynamics. Springer-Verlag, New York.

Mannion, D., 1993. Products of 2×2 random matrices. Annals Appl. Probab. 3 (4), 1189–1218. URL: http://www.jstor.org/stable/2245205.

Melbourne, B.A., Cornell, H.V., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A.L., Hall, R.J., Harrison, S., Hasting, A., Holland, M., Holyoak, M., Lambrinos, J., Moore, K., Yokomizo, H., 2007. Invasion in a heterogeneous world: Resistance, coexistence or hostile takeover? Ecology Letters 10 (1), 77–94. URL: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1461-0248.2006.00987.x.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer.

Petchey, O.L., Gonzalez, A., Wilson, H.B., 1997a. Effects on population persistence: The interaction between environmental noise colour, intraspecific competition and space. Proc. Biol. Sci. 264 (1389), 1841–1847. URL: http://www.jstor.org/stable/51122.

Petchey, O.L., Gonzalez, A., Wilson, H.B., 1997b. Effects on population persistence: The interaction between environmental noise colour, intraspecific competition and space. Proc. Biol. Sci. 264 (1389), 1841–1847. URL: http://www.jstor.org/stable/51122.

Ripa, J., Lundberg, P., 1996. Noise colour and the risk of population extinctions. Proc. Biol. Sci. 263 (1377), 1751–1753. URL: http://www.jstor.org/stable/50667.

Roerdink, J.B.T.M., 1987. The biennial life strategy in a random environment. J. Math. Biol. 26, 309–320.

Rokhlin, V.A., 1964. Exact endomorphism of a lebesgue space. Izv. Acad. Sci USSR, Ser Mat 25, 499–530.

Roy, M., Holt, R.D., Barfield, M., 2005. Temporal autocorrelation can enhance the persistence and abundance of metapopulations comprised of coupled sinks. Amer. Nat. 166, 246–261.

Ruelle, D., 1979. Ergodic theory of differentiable dynamical systems. IHES Publ. Math. 50, 27–58.

Ruelle, D., 1982. Characteristic exponents and invariant manifolds in Hilbert space. Ann. of Math. 115, 243–290.

Steele, J.H., 1985. A comparison of terrestrial and marine ecological systems. Nature 313, 355–358.

Tuljapurkar, S., 1982. Population dynamics in variable environments. II. Correlated environments, sensitivity analysis and dynamics. Theoret. Popul. Biol. 21, 114–140.

Tuljapurkar, S., 1984. Demography in stochastic environments. I Exact distributions of age structure. J. Math. Biol. 19, 335–350.

Tuljapurkar, S., 1986. Demography in stochastic environments. II Growth and convergence rates. J. Math. Biol. 24, 569–581.

Tuljapurkar, S., 1990. Population Dynamics in Variable Environments. Springer-Verlag, New York.

Tuljapurkar, S., Haridas, C.V., 2006. Temporal autocorrelation and stochastic population growth. Ecol. Lett. 9, 327–337.

Turelli, M., 1978. Does environmental variability limit niche overlap? Proc. Nat. Acad. Sci. USA 75 (10), 5085–5089. URL: http://www.jstor.org/stable/69064.

Turelli, M., Petry, D., 1980. Density-dependent selection in a random environment: An evolutionary process that can maintain stable population dynamics. Proc. Nat. Acad. Sci. USA 77 (12), 7501–7505. URL: http://www.jstor.org/stable/9794.

Vasseur, D.A., Yodzis, P., 2004. The color of environmental noise. Ecology 85, 1146–1152.