J. Math. Biol.
DOI 10.1007/s00285-017-1153-2

Mathematical Biology

Stochastic population growth in spatially heterogeneous environments: the density-dependent case

Alexandru Hening[1,2] · **Dang H. Nguyen**[3] · **George Yin**[3]

Received: 29 May 2016 / Revised: 16 June 2017
© The Author(s) 2017. This article is an open access publication

Abstract This work is devoted to studying the dynamics of a structured population that is subject to the combined effects of environmental stochasticity, competition for resources, spatio-temporal heterogeneity and dispersal. The population is spread throughout n patches whose population abundances are modeled as the solutions of a system of nonlinear stochastic differential equations living on $[0, \infty)^n$. We prove that r, the stochastic growth rate of the total population in the absence of competition, determines the long-term behaviour of the population. The parameter r can be expressed as the Lyapunov exponent of an associated linearized system of stochastic differential equations. Detailed analysis shows that if $r > 0$, the population abundances converge polynomially fast to a unique invariant probability measure on $(0, \infty)^n$, while when $r < 0$, the population abundances of the patches converge almost surely to 0 exponentially fast. This generalizes and extends the results of Evans et al. (J Math Biol

A. Hening was in part supported by EPSRC Grant EP/K034316/1.
The research of D. Nguyen and G. Yin was supported in part by the National Science Foundation under grant DMS-1710827.

✉ Alexandru Hening
Alexandru.Hening@tufts.edu; a.hening@imperial.ac.uk

Dang H. Nguyen
dangnh.maths@gmail.com

George Yin
gyin@math.wayne.edu

1 Department of Mathematics, Tufts University, Bromfield-Pearson Hall, 503 Boston Avenue, Medford, MA 02155, USA

2 Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

3 Department of Mathematics, Wayne State University, Detroit, MI 48202, USA

Published online: 03 July 2017

66(3):423–476, 2013) and proves one of their conjectures. Compared to recent developments, our model incorporates very general density-dependent growth rates and competition terms. Furthermore, we prove that persistence is robust to small, possibly density dependent, perturbations of the growth rates, dispersal matrix and covariance matrix of the environmental noise. We also show that the stochastic growth rate depends continuously on the coefficients. Our work allows the environmental noise driving our system to be degenerate. This is relevant from a biological point of view since, for example, the environments of the different patches can be perfectly correlated. We show how one can adapt the nondegenerate results to the degenerate setting. As an example we fully analyze the two-patch case, $n = 2$, and show that the stochastic growth rate is a decreasing function of the dispersion rate. In particular, coupling two sink patches can never yield persistence, in contrast to the results from the nondegenerate setting treated by Evans et al. which show that sometimes coupling by dispersal can make the system persistent.

Keywords Stochastic population growth · Density-dependence · Ergodicity · Spatial and temporal heterogeneity · Lotka–Volterra model · Lyapunov exponent · Habitat fragmentation · Stochastic environment · Dispersion

Mathematics Subject Classification 92D25 · 37H15 · 60H10

Contents

1 Introduction . . .
2 Model and results . . .
2.1 Degenerate noise . . .
2.2 Case study: $n = 2$ . . .
2.3 Robust persistence and extinction . . .
3 Theoretical and numerical examples . . .
3.1 The degenerate case $\sigma_1 = \sigma_2, \alpha = \beta$ . . .
3.2 The degenerate case when the conditions of Theorem 2.6 are violated . . .
4 Discussion and generalizations . . .
4.1 k species competing and dispersing in n patches . . .
Appendix A: The case $r > 0$ . . .
Appendix B: The case $r < 0$ . . .
Appendix C: Degenerate diffusion in $\mathbb{R}^n$ . . .
C.1: Case study: $n = 2$ . . .
Appendix D: Robustness of the model . . .
D.1: Continuous dependence of r on the coefficients . . .
D.2: Robust persistence and extinction . . .
References . . .

1 Introduction

The survival of an organism is influenced by both biotic (competition for resources, predator-prey interactions) and abiotic (light, precipitation, availability of resources) factors. Since these factors are space-time dependent, all types of organisms have to choose their dispersal strategies: If they disperse they can arrive in locations with dif-

ferent environmental conditions while if they do not disperse they face the temporal fluctuations of the local environmental conditions. The dispersion strategy impacts key attributes of a population including its spatial distribution and temporal fluctuations in its abundance. Individuals selecting more favorable habitats are more likely to survive or reproduce. When population densities increase in these habitats, organisms may prosper by selecting habitats that were previously unused. There have been numerous studies of the interplay between dispersal and environmental heterogeneity and how this influences population growth; see Hastings (1983), Gonzalez and Holt (2002), Schmidt (2004), Roy et al. (2005), Schreiber (2010), Cantrell et al. (2012), Durrett and Remenik (2012), Evans et al. (2013) and references therein. The mathematical analysis for stochastic models with density-dependent feedbacks is less explored. In the setting of discrete-space discrete-time models there have been thorough studies by Benaïm and Schreiber (2009); Schreiber (2010); Schreiber et al. (2011). Continuous-space discrete-time population models that disperse and experience uncorrelated, environmental stochasticity have been studied by Hardin et al. (1988a, b, 1990). They show that the leading Lyapunov exponent r of the linearization of the system around the extinction state almost determines the persistence and extinction of the population. For continuous-space continuous-time population models Mierczyński and Shen (2004) study the dynamics of random Kolmogorov type PDE models in bounded domains. Once again, it is shown that the leading Lyapunov exponent r of the linarization around the trivial equilibrium 0 almost determines when the population goes extinct and when it persists. In the current paper we explore the question of persistence and extinction when the population dynamics is given by a system of stochastic differential equations. In our setting, even though our methods and techniques are very different from those used by Hardin et al. (1988a) and Mierczyński and Shen (2004), we still make use of the system linearized around the extinction state. The Lyapunov exponent of this linearized system plays a key role throughout our arguments.

Evans et al. (2013) studied a linear stochastic model that describes the dynamics of populations that continuously experience uncertainty in time and space. Their work has shed some light on key issues from population biology. Their results provide fundamental insights into "ideal free" movement in the face of uncertainty, the evolution of dispersal rates, the single large or several small (SLOSS) debate in conservation biology, and the persistence of coupled sink populations. In this paper, we propose a density-dependent model of stochastic population growth that captures the interactions between dispersal and environmental heterogeneity and complements the work of Evans et al. (2013). We then present a rigorous and comprehensive study of the proposed model based on stochastic analysis.

The dynamics of a population in nature is stochastic. This is due to *environmental stochasticity*—the fluctuations of the environment make the growth rates random. One of the simplest models for a population living in a single patch is

$$dU(t) = U(t)(a - bU(t))dt + \sigma U(t)dW(t), t \geq 0, \tag{1.1}$$

where $U(t)$ is the population abundance at time t, a is the mean per-capita growth rate, $b > 0$ is the strength of intraspecific competition, σ^2 is the infinitesimal variance of fluctuations in the per-capita growth rate and $(W(t))_{t\geq 0}$ is a standard Brownian

motion. The long-term behavior of (1.1) is determined by the *stochastic growth rate* $a - \frac{\sigma^2}{2}$ in the following way (see Evans et al. 2015; Dennis and Patil 1984):

- If $a - \frac{\sigma^2}{2} > 0$ and $U(0) = u > 0$, then $(U(t))_{t\geq 0}$ converges weakly to its unique invariant probability measure ρ on $(0, \infty)$.
- If $a - \frac{\sigma^2}{2} < 0$ and $U(0) = u > 0$, then $\lim_{t\to\infty} U(t) = 0$ almost surely.
- If $a - \frac{\sigma^2}{2} = 0$ and $U(0) = u > 0$, then $\liminf_{t\to\infty} U(t) = 0$ almost surely, $\limsup_{t\to\infty} U(t) = \infty$ almost surely, and $\lim_{t\to\infty} \frac{1}{t}\int_0^t U(s)\,ds = 0$ almost surely.

Organisms are always affected by temporal heterogeneities, but they are subject to spatial heterogeneities only when they disperse. Population growth is influenced by spatial heterogeneity through the way organisms respond to environmental signals (see Hastings 1983; Cantrell and Cosner 1991; Chesson 2000; Schreiber and Lloyd-Smith 2009). There have been several analytic studies that contributed to a better understanding of the separate effects of spatial and temporal heterogeneities on population dynamics. However, few theoretical studies have considered the combined effects of spatio-temporal heterogeneities, dispersal, and density-dependence for discretely structured populations with continuous-time dynamics.

As seen in both the continuous (Evans et al. 2013) and the discrete (Palmqvist and Lundberg 1998) settings, the extinction risk of a population is greatly affected by the spatio-temporal correlation between the environment in the different patches. For example, if spatial correlations are weak, one can show that populations coupled via dispersal can survive even though every patch, on its own, would go extinct (see Evans et al. 2013; Jansen and Yoshimura 1998; Harrison and Quinn 1989). Various species usually exhibit spatial synchrony. Ecologists are interested in this pattern as it can lead to the extinction of rare species. Possible causes for synchrony are dispersal and spatial correlations in the environment (see Legendre 1993; Kendall et al. 2000; Liebhold et al. 2004). Consequently, it makes sense to look at stochastic patch models coupled by dispersion for which the environmental noise of the different patches can be strongly correlated. We do this by extending the setting of Evans et al. (2013) by allowing the environmental noise driving the system to be degenerate.

The rest of the paper is organized as follows. In Sect. 2, we introduce our model for a population living in a patchy environment. It takes into account the dispersal between different patches and density-dependent feedback. The temporal fluctuations of the environmental conditions of the various patches are modeled by Brownian motions that are correlated. We start by considering the relative abundances of the different patches in a low density approximation. We show that these relative abundances converge in distribution to their unique invariant probability measure asymptotically as time goes to infinity. Using this invariant probability measure we derive an expression for r, the stochastic growth rate (Lyapunov exponent) in the absence of competition. We show that this r is key in analyzing the long-term behavior of the populations. In Appendix A we show that if $r > 0$ then the abundances converge weakly, polynomially fast, to their unique invariant probability measure on $(0, \infty)^n$. In Appendix B, we show that if $r < 0$ then all the population abundances go extinct asymptotically, at an exponential rate (with exponential constant r). Appendix C is dedicated to the case when the noise

driving our system is degenerate (that is, the dimension of the noise is lower than the number of patches). In Appendix D, we show that r depends continuously on the coefficients of our model and that persistence is robust—that is, small perturbations of the model do not make a persistent system become extinct. We provide some numerical examples and possible generalizations in Sect. 4.

2 Model and results

We study a population with overlapping generations, which live in a spatio-temporally heterogeneous environment consisting of n distinct patches. The growth rate of each patch is determined by both deterministic and stochastic environmental inputs. We denote by $X_i(t)$ the population abundance at time $t \geq 0$ of the ith patch and write $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ for the vector of population abundances. Following Evans et al. (2013), it is appropriate to model $\mathbf{X}(t)$ as a Markov process with the following properties when $0 \leq \Delta t \ll 1$:

- the conditional mean is

$$\mathbb{E}\left[X_i(t+\Delta t) - X_i(t) \mid X_i(t) = x_i\right] \approx \left[a_i x_i - x_i b_i(x_i) + \sum_{j \neq i} \left(x_j D_{ji} - x_i D_{ij}\right)\right] \Delta t,$$

  where $a_i \in \mathbb{R}$ is the per-capita growth rate in the ith patch, $b_i(x_i)$ is the per-capita strength of intraspecific competition in patch i when the abundance of the patch is x_i, and $D_{ij} \geq 0$ is the dispersal rate from patch i to patch j;
- the conditional covariance is

$$\mathrm{Cov}\left[X_i(t+\Delta t) - X_i(t), X_j(t+\Delta t) - X_j(t) \mid \mathbf{X} = \mathbf{x}\right] \approx \sigma_{ij} x_i x_j \Delta t$$

  for some covariance matrix $\Sigma = (\sigma_{ij})$.

The difference between our model and the one from Evans et al. (2013) is that we added density-dependent feedback through the $x_i b_i(x_i)$ terms.

We work on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ with filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. We consider the system

$$\begin{aligned} dX_i(t) = &\left(X_i(t)\left(a_i - b_i(X_i(t))\right) + \sum_{j=1}^{n} D_{ji} X_j(t)\right) dt \\ &+ X_i(t) dE_i(t),\ i = 1, \ldots, n, \end{aligned} \tag{2.1}$$

where $D_{ij} \geq 0$ for $j \neq i$ is the per-capita rate at which the population in patch i disperses to patch j, $D_{ii} = -\sum_{j\neq i} D_{ij}$ is the total per-capita immigration rate out of patch i, $\mathbf{E}(t) = (E_1(t), \ldots, E_n(t))^T = \Gamma^\top \mathbf{B}(t)$, Γ is a $n \times n$ matrix such that $\Gamma^\top \Gamma = \Sigma = (\sigma_{ij})_{n\times n}$ and $\mathbf{B}(t) = (B_1(t), \ldots, B_n(t))$ is a vector of independent standard Brownian motions adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Throughout the paper, we work with the following assumption regarding the growth of the instraspecific competition rates.

Assumption 2.1 For each $i = 1, \dots, n$ the function $b_i : \mathbb{R}_+ \mapsto \mathbb{R}$ is locally Lipschitz and vanishing at 0. Furthermore, there are $M_b > 0$, $\gamma_b > 0$ such that

$$\frac{\sum_{i=1}^n x_i(b_i(x_i) - a_i)}{\sum_{i=1}^n x_i} > \gamma_b \text{ for any } x_i \geq 0, i = 1, \dots, n \text{ satisfying } \sum_{i=1}^n x_i \geq M_b \tag{2.2}$$

Remark 2.1 Note that if we set $x_j = x \geq M_b$ and $x_i = 0, i \neq j$, we get from (2.2) that

$$b_j(x) - a_j > \gamma_b,\ x \geq M_b,\ j = 1, \dots, n.$$

Remark 2.2 Note that condition (2.2) is biologically reasonable because it holds if the b_i's are sufficiently large for large x_i's. We provide some simple scenarios when Assumption 2.1 is satisfied.

a) Suppose $b_i : [0, \infty) \to [0, \infty), i = 1, \dots, n$ are locally Lipschitz and vanishing at 0. Assume that there exist $\gamma_b > 0$, $\tilde{M}_b > 0$ such that

$$\inf_{x \in [\tilde{M}_b, \infty)} b_i(x) - a_i - \gamma_b > 0,\ i = 1, \dots, n$$

It is easy to show that Assumption 2.1 holds.

b) Particular cases of (a) are for example, any $b_i : \mathbb{R}_+ \mapsto \mathbb{R}$ that are locally Lipschitz, vanishing at 0 such that $\lim_{x \to \infty} b_i(x) = \infty$.

c) One natural choice for the competition functions, which is widely used throughout the literature, is $b_i(x) = \kappa_i x, x \in (0, \infty)$ for some $\kappa_i > 0$. In this case the competition terms become $-x_i b(x_i) = -\kappa_i x_i^2$.

Remark 2.3 Note that if we have the SDE

$$\begin{aligned} dX_i(t) = &\left(X_i(t) f_i(X_i(t)) + \sum_{j=1}^n D_{ji} X_j(t) \right) dt \\ &+ X_i(t) dE_i(t),\ i = 1, \dots, n, \end{aligned} \tag{2.3}$$

where f_i are locally Lipschitz this can always be rewritten in the form (2.1) with

$$a_i := f_i(0) \text{ and } b_i(x) := f_i(0) - f_i(x),\ i = 1, \dots, n.$$

Therefore, our setting is in fact very general and incorporates both nonlinear growth rates and nonlinear competition terms.

The drift $\tilde{f}(\mathbf{x}) = (\tilde{f}_1(\mathbf{x}), \dots, \tilde{f}_n(\mathbf{x}))$ where $\tilde{f}_i(\mathbf{x}) = x_i(a_i - b_i(x_i)) + \sum_{j=1}^n D_{ji} X_j(t)$ is sometimes said to be *cooperative*. This is because $f_i(\mathbf{x}) \leq f_i(\mathbf{y})$ if $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^n$ such that $x_i = y_i, x_j \leq y_j$ for $j \neq i$. A distinctive property of *cooperative systems* is that comparison arguments are generally satisfied. We refer to Chueshov (2002) for more details.

Remark 2.4 If the dispersal matrix (D_{ij}) has a normalized dominant left eigenvector $\alpha = (\alpha_1, \ldots, \alpha_n)$ then one can show that the system

$$dX_i(t) = \left(X_i(t)\,(a_i - b_i X_i(t)) + \delta \sum_{j=1}^{n} D_{ji} X_j(t) \right) dt + X_i(t) dE_i(t),\ i = 1, \ldots, n,$$

converges as $\delta \to \infty$ to a system $(\tilde{X}_1(t), \ldots, \tilde{X}_n(t))$ for which

$$\tilde{X}_i(t) = \alpha_i \tilde{X}(t),\ t \geq 0,\ i = 1, \ldots, n,$$

where $\tilde{X}(t) = \tilde{X}_1(t) + \cdots + \tilde{X}_n(t)$ and $\tilde{X}$ is an autonomous Markov process that satisfies the SDE

$$d\tilde{X}(t) = \tilde{X}(t) \sum_{i=1}^{n} \alpha_i (a_i - b_i \alpha_i \tilde{X}(t))\, dt + \tilde{X}(t) \sum_{i=1}^{n} \alpha_i\, dE_i(t).$$

As such, our system is a general version of the system treated in Evans et al. (2015). One can recover the system from Evans et al. (2015) as an infinite dispersion limit of ours.

We denote by $\mathbf{X}^{\mathbf{x}}(t)$ the solution of (2.1) started at $\mathbf{X}(0) = \mathbf{x} \in \mathbb{R}^n_+$. Following Evans et al. (2013), we call matrices D with zero row sums and non-negative off-diagonal entries *dispersal matrices*. If D is a dispersal matrix, then it is a generator of a continuous-time Markov chain. Define $P_t := \exp(tD), t \geq 0$. Then $P_t, t \geq 0$ is a matrix with non-negative entries that gives the transition probabilities of a Markov chain: The (i, j)th entry of P_t gives the proportion of the population that was initially in patch i at time 0 but has dispersed to patch j at time t and D is the generator of this Markov chain. If one wants to include mortality induced because of dispersal, one can add cemetery patches in which dispersing individuals enter and experience a killing rate before moving to their final destination. Our model is a density-dependent generalization of the one by Evans et al. (2013). We are able to prove that the linearization of the density-dependent model fully determines the non-linear density-dependent behavior, a fact which was conjectured by Evans et al. (2013). Furthermore, we prove stronger convergence results and thus extend the work of Evans et al. (2013). Analogous results for discrete-time versions of the model have been studied by Benaïm and Schreiber (2009) for discrete-space and by Hardin et al. (1988a, b) for continuous-space.

We will work under the following assumptions.

Assumption 2.2 The dispersal matrix D is *irreducible*.

Assumption 2.3 The covariance matrix Σ is non-singular.

Assumption 2.2 is equivalent to forcing the entries of the matrix $P_t = \exp(tD)$ to be strictly positive for all $t > 0$. This means that it is possible for the population to

disperse between any two patches. We can always reduce our problem to this setting by working with the maximal irreducible subsets of patches. Assumption 2.3 says that our randomness is non-degenerate, and thus truly n-dimensional. We show in Appendix C how to get the desired results when Assumption 2.3 does not hold.

Throughout the paper we set $\mathbb{R}^n_+ := [0,\infty)^n$ and $\mathbb{R}^{n,\circ}_+ := (0,\infty)^n$. We define the total abundance of our population at time $t \geq 0$ via $S(t) := \sum_{i=1}^n X_i(t)$ and let $Y_i(t) := \frac{X_i(t)}{S(t)}$ be the proportion of the total population that is in patch i at time $t \geq 0$. Set $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_n(t))$. An application of Itô's lemma to (2.1) yields

$$\begin{aligned} dY_i(t) &= Y_i(t)\left(a_i - \sum_{j=1}^n a_j Y_j(t) - b_i(S(t)Y_i(t)) + \sum_{j=1}^n Y_j(t) b_j(S(t)Y_j(t))\right)dt \\ &\quad + \sum_{j=1}^n D_{ji}Y_j(t)dt \; + \; Y_i(t)\left(\sum_{j,k=1}^n \sigma_{kj}Y_k(t)Y_j(t)) - \sum_{j=1}^n \sigma_{ij}Y_j(t)\right)dt \\ &\quad + Y_i(t)\left[dE_i(t) - \sum_{j=1}^n Y_j(t)dE_j(t)\right] \\ dS(t) &= S(t)\left(\sum_{i=1}^n (a_i Y_i(t) - Y_i(t)b_i(S(t)Y_i(t)))\right)dt + S(t)\sum_{i=1}^n Y_i(t)dE_i(t) \end{aligned} \tag{2.4}$$

We can rewrite (2.4) in the following compact equation for $(\mathbf{Y}(t), S(t))$ where $\mathbf{b}(\mathbf{x}) = (b_1(x_1), \ldots, b_n(x_n))$.

$$\begin{aligned} d\mathbf{Y}(t) &= \Big(\mathrm{diag}(\mathbf{Y}(t)) - \mathbf{Y}(t)\mathbf{Y}^\top(t)\Big)\Gamma^\top d\mathbf{B}(t) \\ &\quad + \mathbf{D}^\top \mathbf{Y}(t)dt + \Big(\mathrm{diag}(\mathbf{Y}(t)) - \mathbf{Y}(t)\mathbf{Y}^\top(t)\Big) \\ &\quad \times (\mathbf{a} - \Sigma\mathbf{Y}(t) - \mathbf{b}(S(t)\mathbf{Y}(t)))dt \\ dS(t) &= S(t)\left[\mathbf{a} - \mathbf{b}(S(t)\mathbf{Y}(t))\right]^\top \mathbf{Y}(t)dt + S(t)\mathbf{Y}(t)^\top \Gamma^\top d\mathbf{B}(t), \end{aligned} \tag{2.5}$$

where $\mathbf{Y}(t)$ lies in the simplex $\Delta := \{(y_1, \ldots, y_n) \in \mathbb{R}^n_+ : y_1 + \cdots + y_n = 1\}$. Let $\Delta^\circ = \{(y_1, \ldots, y_n) \in \mathbb{R}^{n,\circ}_+ : y_1 + \cdots + y_n = 1\}$ be the interior of Δ.

Consider Equation (2.5) on the boundary $((\mathbf{y}, s) : \mathbf{y} \in \Delta, s = 0)$ (that is, we set $S(t) \equiv 0$ in the equation for $\mathbf{Y}(t)$). We have the following system

$$\begin{aligned} d\tilde{\mathbf{Y}}(t) &= \Big(\mathrm{diag}(\tilde{\mathbf{Y}}(t)) - \tilde{\mathbf{Y}}(t)\tilde{\mathbf{Y}}^\top(t)\Big)\Gamma^\top d\mathbf{B}(t) \\ &\quad + \mathbf{D}^\top \tilde{\mathbf{Y}}(t)dt + \Big(\mathrm{diag}(\tilde{\mathbf{Y}}(t)) - \tilde{\mathbf{Y}}(t)\tilde{\mathbf{Y}}^\top(t)\Big)(\mathbf{a} - \Sigma\tilde{\mathbf{Y}}(t))dt \end{aligned} \tag{2.6}$$

on the simplex Δ. We also introduce the linearized version of (2.1), where the competition terms $b_i(x_i)$ are all set to 0,

$$d\mathcal{X}_i(t) = \left(\mathcal{X}_i(t)a_i + \sum_{j=1}^{n} D_{ji}\mathcal{X}_j(t)\right)dt + \mathcal{X}_i(t)dE_i(t), \quad i = 1, \ldots, n. \tag{2.7}$$

and let $\mathcal{S}(t) = \sum_{i=1}^{n} \mathcal{X}_i(t)$ be the total population abundance, in the absence of competition. The processes $(\mathcal{X}_1(t), \ldots, \mathcal{X}_n(t))$, $\tilde{\mathbf{Y}}(t)$ and $\mathcal{S}(t)$ have been studied by Evans et al. (2013).

Evans et al. (2013, Proposition 3.1) proved that the process $(\tilde{\mathbf{Y}}(t))_{t\geq 0}$ is an irreducible Markov process, which has the strong Feller property and admits a unique invariant probability measure ν^* on Δ. Let $\tilde{\mathbf{Y}}(\infty)$ be a random variable on Δ with distribution ν. We define

$$\begin{aligned} r &:= \int_{\Delta} \left(\mathbf{a}^{\top}\mathbf{y} - \frac{1}{2}\mathbf{y}^{\top}\Sigma\mathbf{y}\right)\nu^*(d\mathbf{y}) \\ &= \sum_i a_i \mathbb{E}\left[\tilde{Y}_i(\infty)\right] - \frac{1}{2}\mathbb{E}\left[\sum_{ij}\sigma_{ij}\tilde{Y}_i(\infty)\tilde{Y}_j(\infty)\right] \end{aligned} \tag{2.8}$$

Remark 2.5 We note that r is the stochastic growth rate (or Lyapunov exponent) of the total population $\mathcal{S}(t)$ in the absence of competition. That is,

$$\mathbb{P}\left\{\lim_{t\to\infty}\frac{\ln \mathcal{S}^{\mathbf{x}}(t)}{t} = r\right\} = 1.$$

The expression (2.8) for r coincides with the one derived by Evans et al. (2013).

We use superscripts to denote the starting points of our processes. For example $(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t))$ denotes the solution of (2.4) with $(\mathbf{Y}(0), S(0)) = (\mathbf{y}, s) \in \Delta \times (0, \infty)$. Fix $\mathbf{x} \in \mathbb{R}^n_+$ and define the *normalized occupation measures*,

$$\Pi_t^{(\mathbf{x})}(\cdot) = \frac{1}{t}\int_0^t \mathbf{1}_{\{\mathbf{X}^{\mathbf{x}}(u)\in\cdot\}}du. \tag{2.9}$$

These random measures describe the distribution of the observed population dynamics up to time t. If we define the sets

$$S_\eta := \{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^{n,\circ}_+ : |x_i| \leq \eta \text{ for some } i = 1, \ldots, n\},$$

then $\Pi_t^{(\mathbf{x})}(S_\eta)$ is the fraction of the time in the interval $[0, t]$ that the total abundance of some patch is less than η given that our population starts at $\mathbf{X}(0) = \mathbf{x}$.

Definition 2.1 One can define a distance on the space of probability measures living on the Borel measurable subsets of $\mathbb{R}^n_+$, that is on the space $(\mathbb{R}^n_+, \mathcal{B}(\mathbb{R}^n_+))$. This is done by defining $\|\cdot,\cdot\|_{\text{TV}}$, the *total variation norm*, via

$$\|\mu, \nu\|_{\text{TV}} := \sup_{A\in\mathcal{B}(\mathbb{R}^n_+)} |\mu(A) - \nu(A)|.$$

Theorem 2.1 *Suppose that Assumptions 2.2 and 2.3 hold and that $r > 0$. The process $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))_{t\geq 0}$ has a unique invariant probability measure π on $\mathbb{R}^{n,\circ}_+$ that is absolutely continuous with respect to the Lebesgue measure and for any $q^* > 0$,*

$$\lim_{t\to\infty} t^{q^*} \| P_{\mathbf{X}}(t, \mathbf{x}, \cdot) - \pi(\cdot)\|_{TV} = 0, \ \ \mathbf{x} \in \mathbb{R}^{n,\circ}_+, \tag{2.10}$$

and $P_{\mathbf{X}}(t, \mathbf{x}, \cdot)$ is the transition probability of $(\mathbf{X}(t))_{t\geq 0}$. Moreover, for any initial value $\mathbf{x} \in \mathbb{R}^n_+\backslash\{\mathbf{0}\}$ and any π-integrable function f we have

$$\mathbb{P}\left\{\lim_{T\to\infty} \frac{1}{T}\int_0^T f\left(\mathbf{X}^{\mathbf{x}}(t)\right) dt = \int_{\mathbb{R}^{n,\circ}_+} f(\mathbf{u})\pi(d\mathbf{u})\right\} = 1. \tag{2.11}$$

Remark 2.6 Theorem 2.1 is a direct consequence of Theorem A.2, which will be proved in Appendix A. As a corollary we get the following result.

Definition 2.2 Following Roth and Schreiber (2014), we say that the model (2.1) is stochastically persistent if for all $\varepsilon > 0$, there exists $\eta > 0$ such that with probability one,

$$\Pi_t^{(\mathbf{x})}(S_\eta) \leq \varepsilon$$

for t sufficiently large and $\mathbf{x} \in S_\eta\backslash\{\mathbf{0}\}$.

Corollary 2.1 *If Assumptions 2.2 and 2.3 hold, and $r > 0$, then the process $\mathbf{X}(t)$ is stochastically persistent.*

Proof By Theorem 2.1, we have that for all $\mathbf{x} \in \mathbb{R}^{n,\circ}_+$,

$$\mathbb{P}\left\{\Pi_t^{(\mathbf{x})} \Rightarrow \pi \text{ as } t \to \infty\right\} = 1.$$

Since π is supported on $\mathbb{R}^{n,\circ}_+$, we get the desired result. □

Biological interpretation of Theorem 2.1 *The quantity r is the Lyapunov exponent or stochastic growth rate of the total population process $(\mathcal{S}(t))_{t\geq 0}$ in the absence of competition. This number describes the long-term growth rate of the population in the presence of a stochastic environment. According to (2.8) r can be written as the difference $\overline{\mu} - \frac{1}{2}\overline{\sigma}^2$ where*

- *$\overline{\mu}$ is the average of per-capita growth rates with respect to the asymptotic distribution $\tilde{\mathbf{Y}}(\infty)$ of the population in the absence of competition.*

- $\overline{\sigma}^2$ *is the infinitesimal variance of the environmental stochasticity averaged according to the asymptotic distribution of the population in the absence of competition.*

We note by (2.8) *that* r *depends on the dispersal matrix, the growth rates at 0 and the covariance matrix of the environmental noise. As such, the stochastic growth rate can change due to the dispersal strategy or environmental fluctuations.*

When the stochastic growth rate of the population in absence of competition is strictly positive (i.e. $r > 0$*) our population is persistent in a strong sense: for any starting point* $(X_1(0), \ldots, X_n(0)) = (x_1, \ldots, x_n) \in \mathbb{R}_+^{n,\circ}$ *the distribution of the population densities at time* t *in the* n *patches* $(X_1(t), \ldots, X_n(t))$ *converges as* $t \to \infty$ *to the unique probability measure* π *that is supported on* $\mathbb{R}_+^{n,\circ}$.

Definition 2.3 We say the population of patch i goes extinct if for all $\mathbf{x} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$

$$\mathbb{P}\left\{\lim_{t\to\infty} X_i^{\mathbf{x}}(t) = 0\right\} = 1.$$

We say the population goes extinct if the populations from all the patches go extinct, that is if for all $\mathbf{x} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$

$$\mathbb{P}\left\{\lim_{t\to\infty} \mathbf{X}^{\mathbf{x}}(t) = \mathbf{0}\right\} = 1.$$

Theorem 2.2 *Suppose that Assumptions 2.2 and 2.3 hold and that* $r < 0$*. Then for any* $i = 1, \ldots, n$ *and any* $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$,

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{\ln X_i^{\mathbf{x}}(t)}{t} = r\right\} = 1. \tag{2.12}$$

Biological interpretation of Theorem 2.2 *If the stochastic growth rate of the population in the absence of competition is negative (i.e.* $r < 0$*) the population densities of the* n *patches* $(X_1(t), \ldots, X_n(t))$ *go extinct exponentially fast with rates* $r < 0$ *with probability* 1 *for any starting point* $(X_1(0), \ldots, X_n(0)) = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$.

In Appendix A, we prove Theorem 2.1 while Theorem 2.2 is proven in Appendix B.

2.1 Degenerate noise

We consider the evolution of the process $(\mathbf{X}(t))_{t\geq 0}$ given by (2.1) when Assumption 2.3 does not hold. If the covariance matrix $\Sigma = \Gamma^T\Gamma$ coming for the Brownian motions $\mathbf{E}(t) = (E_1(t), \ldots, E_n(t))^T = \Gamma^\top \mathbf{B}(t)$ is singular, the environmental noise driving our SDEs has a lower dimension than the dimension n of the underlying state space. It becomes much more complex to prove that our process is Feller and irreducible. In order to verify the Feller property, we have to verify the so-called Hörmander condition, and to verify the irreducibility, we have to investigate the controllability of a related control system.

We are able to prove the following extinction and persistence results.

Theorem 2.3 *Assume that* $\tilde{\mathbf{Y}}(t)$ *has a* ***unique*** *invariant probability measure* ν^*. *Define* r *by* (2.8). *Suppose that* $r < 0$. *Then for any* $i = 1, \dots, n$ *and any* $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n_+$

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{\ln X_i^{\mathbf{x}}(t)}{t} = r\right\} = 1. \tag{2.13}$$

In particular, for any $i = 1, \dots, n$ *and any* $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n_+$

$$\mathbb{P}\left\{\lim_{t\to\infty} X_i^{\mathbf{x}}(t) = 0\right\} = 1.$$

Remark 2.7 The extra assumption in this setting is that the Markov process describing the proportions of the populations of the patches evolving without competition, $\tilde{\mathbf{Y}}(t)$, has a unique invariant probability measure. In fact, we conjecture that $\tilde{\mathbf{Y}}(t)$ always has a unique invariant probability measure. We were able to prove this conjecture when $n = 2$—see Remark 3.1 for details.

Theorem 2.4 *Assume that* $\tilde{\mathbf{Y}}(t)$ *has a* ***unique*** *invariant probability measure* ν^*. *Define* r *by* (2.8). *Suppose that Assumption 2.2 holds and that* $r > 0$. *Assume further that there is a sufficiently large* $T > 0$ *such that the Markov chain* $(\mathbf{Y}(kT), S(kT))_{k\in\mathbb{N}}$ *it is irreducible and aperiodic, and that every compact set in* $\Delta^\circ \times (0, \infty)$ *is petite for this Markov chain.*

The process $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))_{t\geq 0}$ *has a unique invariant probability measure* π *on* $\mathbb{R}^{n,\circ}_+$ *that is absolutely continuous with respect to the Lebesgue measure and for any* $q^* > 0$,

$$\lim_{t\to\infty} t^{q^*} \| P_{\mathbf{X}}(t, \mathbf{x}, \cdot) - \pi(\cdot)\|_{TV} = 0, \ \mathbf{x} \in \mathbb{R}^{n,\circ}_+, \tag{2.14}$$

where $\|\cdot, \cdot\|_{TV}$ *is the total variation norm and* $P_{\mathbf{X}}(t, \mathbf{x}, \cdot)$ *is the transition probability of* $(\mathbf{X}(t))_{t\geq 0}$. *Moreover, for any initial value* $\mathbf{x} \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}$ *and any* π*-integrable function* f, *we have*

$$\mathbb{P}\left\{\lim_{T\to\infty} \frac{1}{T}\int_0^T f\left(\mathbf{X}^{\mathbf{x}}(t)\right) dt = \int_{\mathbb{R}^{n,\circ}_+} f(\mathbf{u})\pi(d\mathbf{u})\right\} = 1. \tag{2.15}$$

Remark 2.8 We require as before that $\tilde{\mathbf{Y}}(t)$ has a unique invariant probability measure. Furthermore, we require that there exists some time $T > 0$ such that if we observe the process $(\mathbf{Y}(t), S(t))$ at the fixed times $T, 2T, 3T, \dots, kT, \dots$ it is irreducible (loosely speaking this means that the process can visit any state) and aperiodic (returns to a given state occur at irregular times).

2.2 Case study: $n = 2$

Note that the two Theorems above have some extra assumptions. We exhibit how one can get these conditions explicitly as functions of the various parameters of the model.

For the sake of a clean exposition we chose to fully treat the case when $n = 2$ and $b_i(x) = b_i x, x \geq 0, i = 1, 2$ for some $b_1, b_2 > 0$ (each specific case would have to be studied separately as the computations change in each setting). As a result, (2.1) becomes

$$\begin{cases} dX_1(t) = \big(X_1(t)(a_1 - b_1 X_1(t)) - \alpha X_1(t) + \beta X_2(t)\big)dt + \sigma_1 X_1(t) dB(t) \\ dX_2(t) = \big(X_2(t)(a_2 - b_2 X_2(t)) + \alpha X_1(t) - \beta X_2(t)\big)dt + \sigma_2 X_2(t) dB(t), \end{cases}$$

where σ_1, σ_2 are non-zero constants and $(B(t))_{t\geq 0}$ is a one dimensional Brownian motion. The Lyapunov exponent can now be expressed as (see Remark 3.1)

$$\begin{aligned} r = a_2 - \frac{\sigma_2^2}{2} + (a_1 - a_2 + \sigma_2^2) \int_0^1 y \rho_1^*(y)\, dy \\ - \frac{(\sigma_1 - \sigma_2)^2}{2} \int_0^1 y^2 \rho_1^*(y)\, dy \end{aligned} \tag{2.16}$$

where ρ_1^* is given in (3.5) later.

If $\sigma_1 = \sigma_2 =: \sigma$, one has (see Remark 3.1)

$$r = a_2 - \frac{\sigma^2}{2} + (a_1 - a_2 + \sigma^2) y^\star. \tag{2.17}$$

Theorem 2.5 *Define r by (2.16) if $\sigma_1 \neq \sigma_2$ and by (2.17) if $\sigma_1 = \sigma_2 = \sigma$. If $r < 0$ then for any $i = 1, 2$ and any $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2_+$*

$$\mathbb{P}\left\{ \lim_{t\to\infty} \frac{\ln X_i^{\mathbf{x}}(t)}{t} = r \right\} = 1. \tag{2.18}$$

Theorem 2.6 *Suppose that $\sigma_1 \neq \sigma_2$ or $\beta + (b_2/b_1)(a_1 - a_2 - \alpha + \beta) - \alpha(b_2/b_1)^2 \neq 0$. Define r as in Theorem 2.5. If $r > 0$ then the conclusion of Theorem 2.4 holds.*

Remark 2.9 Once again the parameter r tells us when the population goes extinct and when it persists. To obtain the conclusion of Theorem 2.4 when $r > 0$, we need $\sigma_1 \neq \sigma_2$ or $\beta + (b_2/b_1)(a_1 - a_2 - \alpha + \beta) - \alpha(b_2/b_1)^2 \neq 0$. The condition $\sigma_1 \neq \sigma_2$ tells us that the noise must at least differ through its variance. If $\sigma_1 = \sigma_2$ then we require

$$a_1 + \beta \frac{b_1 + b_2}{b_2} \neq a_2 + \alpha \frac{b_1 + b_2}{b_1}.$$

The term $\beta \frac{b_1+b_2}{b_2}$ measures the dispersion rate of individuals from patch 2 to patch 1 averaged by the inverse relative competition strength of patch 2. In particular, if $b_1 = b_2$ we have that

$$2(\beta - \alpha) \neq a_2 - a_1,$$

that is twice the difference of the dispersal rates cannot equal the difference of the growth rates. The dynamics of the system is very different if these conditions do not hold (see Sect. 3.2 and Theorem 2.7).

Theorem 2.7 *Suppose that $\sigma_1 = \sigma_2 = \sigma$, $b_1 = b_2$ and $2(\beta - \alpha) = a_2 - a_1$. In this setting one can show that the stochastic growth rate is given by $r = a_1 - \alpha + \beta - \frac{\sigma^2}{2}$. Assume that $(X_1(0), X_2(0)) = \mathbf{x} = (x_1, x_2) \in \mathbb{R}^{2,\circ}_+$ and let $U(t)$ be the solution to*

$$dU(t) = U(t)(a_1 - \alpha + \beta - bU(t))\,dt + \sigma U(t)dB(t), U(0) = x_2.$$

Then we get the following results

1) *If $x_1 = x_2$ then $\mathbb{P}(X_1^{\mathbf{x}}(t) = X_2^{\mathbf{x}}(t) = U(t), t \geq 0) = 1$.*
2) *If $x_1 \neq x_2$ then $\mathbb{P}(X_1^{\mathbf{x}}(t) \neq X_2^{\mathbf{x}}(t), t \geq 0) = 1$.*
3) *If $r < 0$ then $X_1(t)$ and $X_2(t)$ converges to 0 exponentially fast. If $r > 0$ then*

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{X_1^{\mathbf{x}}(t)}{U^{\mathbf{x}}(t)} = \lim_{t\to\infty} \frac{X_2^{\mathbf{x}}(t)}{U^{\mathbf{x}}(t)} = 1\right\} = 1.$$

Thus, both $X_1(t)$ and $X_2(t)$ converge to a unique invariant probability measure ρ on $(0, \infty)$, which is the invariant probability measure of $U(t)$. The invariant probability measure of $(X_1(t), X_2(t))_{t\geq 0}$ is concentrated on the one-dimensional manifold $\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^{2,\circ}_+ : x_1 = x_2\}$.

The proof of Theorem 2.7 is presented in Sect. 3.2.

2.3 Robust persistence and extinction

The model we work with is an approximation of the real biological models. As a result, it is relevant to see if 'close models' behave similarly to ours. This reduces to studying the robustness of our system. Consider the process

$$d\widehat{X}_i = \widehat{X}_i\left(\widehat{a}_i - \widehat{b}_i(X_i)\right)dt + \widehat{D}_{ij}(\widehat{\mathbf{X}})\widehat{X}_i dt + \widehat{X}_i\widehat{\Gamma}(\widehat{\mathbf{X}})d\mathbf{B}(t) \tag{2.19}$$

where $\widehat{\mathbf{b}}(\cdot)$, $\widehat{D}(\cdot)$, $\widehat{\Gamma}(\cdot)$ are locally Lipschitz functions and $\widehat{D}_{ij}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n_+$, $i \neq j$ and $\widehat{D}_{ii}(\mathbf{x}) = -\sum_{j\neq i} D_{ij}(\mathbf{x})$. If there exists $\theta > 0$ such that

$$\sup_{\mathbf{x}\in\mathbb{R}^{n,\circ}_+} \left\{\|\mathbf{a} - \widehat{\mathbf{a}}\|, \|\mathbf{b}(\mathbf{x}) - \widehat{\mathbf{b}}(\mathbf{x})\|, \|D - \widehat{D}(\mathbf{x})\|, \|\Gamma - \widehat{\Gamma}(\mathbf{x})\|\right\} < \theta, \tag{2.20}$$

then we call $\widehat{\mathbf{X}}$ a θ-*perturbation* of $\mathbf{X}$.

Theorem 2.8 *Suppose that the dynamics of $(\mathbf{X}(t))_{t\geq 0}$ satisfy the assumptions of Theorem 2.1. Then there exists $\theta > 0$ such that any θ-perturbation $(\widehat{\mathbf{X}}(t))_{t\geq 0}$ of $(\mathbf{X}(t))_{t\geq 0}$*

is persistent. Moreover, the process $(\widehat{\mathbf{X}}(t))_{t\geq 0}$ *has a unique invariant probability measure* $\widehat{\pi}$ *on* $\mathbb{R}^{n,\circ}_+$ *that is absolutely continuous with respect to the Lebesgue measure and for any* $q^* > 0$

$$\lim_{t\to\infty} t^{q^*} \| P_{\widehat{\mathbf{X}}}(t, \mathbf{x}, \cdot) - \widehat{\pi}(\cdot)\|_{TV} = 0, \ \mathbf{x} \in \mathbb{R}^{n,\circ}_+,$$

where $P_{\widehat{\mathbf{X}}}(t, \mathbf{x}, \cdot)$ *is the transition probability of* $(\widehat{\mathbf{X}}(t))_{t\geq 0}$.

Biological interpretation of Theorem 2.8 *As long as the perturbation of our model is small, persistence does not change to extinction. Our model, even though it is only an approximation of reality, can provide relevant information regarding biological systems. Small enough changes in the growth rates, the competition rates, the dispersion matrix and the covariance matrix leave a persistent system unchanged.*

3 Theoretical and numerical examples

This subsection is devoted to some theoretical and numerical examples. We choose the dimension to be $n = 2$, so that we can compute the stochastic growth rate explicitly.

Remark 3.1 If an explicit expression for r is desirable, one needs to determine the first and second moments for the invariant probability measure ν^*. One can show that ρ^*, the density of ν^* with respect to Lebesgue measure, satisfies

$$-\sum_i \frac{\partial}{\partial y_i}[\mu_i(\mathbf{y})\rho^*(\mathbf{y})] + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j}[v_{ij}(\mathbf{y})\rho^*(\mathbf{y})] = 0, \ \mathbf{y} \in \Delta, \tag{3.1}$$

where $\mu_i(\mathbf{y})$ and $v_{i,j}(\mathbf{y})$ are the entries of

$$\begin{aligned}
\mu(\mathbf{y}) &= D^\top \mathbf{y} + \left(\mathrm{diag}(\mathbf{y}) - \mathbf{y}\mathbf{y}^\top\right)(\mathbf{a} - \Sigma\mathbf{y}), \\
v(\mathbf{y}) &= \left(\mathrm{diag}(\mathbf{y}) - \mathbf{y}\mathbf{y}^\top(t)\right)\Gamma^\top\Gamma\left(\mathrm{diag}(\mathbf{y}) - \mathbf{y}\mathbf{y}^\top(t)\right),
\end{aligned}$$

and ρ^* is constrained by $\int_\Delta \rho^*(\mathbf{y})d\mathbf{y} = 1$ with appropriate boundary conditions. The boundary conditions are usually found by characterizing the domain of the infinitesimal generator of the Feller diffusion process $\tilde{\mathbf{Y}}(t)$, which is usually a very difficult problem.

However, following Evans et al. (2013), in the case of two patches ($n = 2$) and non-degenerate noise the problem is significantly easier. Let $\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$. The system becomes

$$\begin{cases}
dX_1(t) = \big(X_1(t)(a_1 - bX_1(t)) - \alpha X_1(t) + \beta X_2(t)\big)dt + \sigma_1 X_1(t)dB_1(t) \\
dX_2(t) = \big(X_2(t)(a_2 - bX_2(t)) + \alpha X_1(t) - \beta t X_2(t)\big)dt + \sigma_2 X_2(t)dB_2(t).
\end{cases} \tag{3.2}$$

It is easy to find the density ρ_1^* of $\tilde{Y}_1(\infty)$ explicitly (by solving (3.1)) and noting that 0, 1 are both entrance boundaries for the diffusion $\tilde{Y}_1(t)$). Then

$$\rho_1^*(x) = Cx^{\beta-\alpha_1}(1-x)^{-\beta-\alpha_2}\exp\left(-\frac{2}{\sigma_1^2+\sigma_2^2}\left(\frac{\beta}{x}+\frac{\alpha}{1-x}\right)\right),\ x\in(0,1)$$

where $C>0$ is a normalization constant and

$$\alpha_i := \frac{2\sigma_i^2}{\sigma_1^2+\sigma_2^2},\ i=1,2$$
$$\beta := \frac{2}{\sigma_1^2+\sigma_2^2}(a_1-a_2+\beta-\alpha).$$

One can then get the following explicit expression for the Lyapunov exponent

$$r = a_2 - \frac{\sigma_2^2}{2} + (a_1-a_2+\sigma_2^2)\int_0^1 y\rho_1^*(y)\,dy$$
$$-\frac{\sigma_1^2+\sigma_2^2}{2}\int_0^1 y^2\rho_1^*(y)\,dy. \tag{3.3}$$

Next, consider the degenerate case

$$\begin{cases} dX_1(t) = \big(X_1(t)(a_1-b_1X_1(t)) - \alpha X_1(t) + \beta X_2(t)\big)dt + \sigma_1 X_1(t)dB(t) \\ dX_2(t) = \big(X_2(t)(a_2-b_2X_2(t)) + \alpha X_1(t) - \beta X_2(t)\big)dt + \sigma_2 X_2(t)dB(t), \end{cases} \tag{3.4}$$

where σ_1,σ_2 are non-zero constants and $(B(t))_{t\geq 0}$ is a one dimensional Brownian motion. Since $\tilde{Y}_1(t)+\tilde{Y}_2(t)=1$, to find the invariant probability measure of $\tilde{\mathbf{Y}}(t)$, we only need to find the invariant probability measure of $\tilde{Y}_1(t)$.

If $\sigma_2\neq\sigma_2$ we can find the invariant density ρ_1^* of $\tilde{Y}_1(\infty)$ explicitly (by solving (3.1)). Then

$$\rho_1^*(x) = Cx^{\widehat{\beta}-\widehat{\alpha}_1}(1-x)^{-\widehat{\beta}-\widehat{\alpha}_2}\exp\left(-\frac{2}{(\sigma_1-\sigma_2)^2}\left(\frac{\beta}{x}+\frac{\alpha}{1-x}\right)\right),$$
$$x\in(0,1) \tag{3.5}$$

where $C>0$ is a normalization constant and

$$\widehat{\alpha}_1 := \frac{-2\sigma_1}{(\sigma_1-\sigma_2)},\quad \widehat{\alpha}_2 := \frac{2\sigma_2}{(\sigma_1-\sigma_2)},$$
$$\widehat{\beta} := \frac{2}{(\sigma_1-\sigma_2)^2}(a_1-a_2+\beta-\alpha).$$

The Lyapunov exponent can now be expressed as

$$r = a_2 - \frac{\sigma_2^2}{2} + (a_1 - a_2 + \sigma_2^2)\int_0^1 y\rho_1^*(y)\,dy$$
$$-\frac{(\sigma_1 - \sigma_2)^2}{2}\int_0^1 y^2\rho_1^*(y)\,dy.$$

We note that the structure of the stochastic growth rate r for non-degenerate noise (3.3) and for degenerate noise (2.16) with $\sigma_1 \neq \sigma_2$ is the same. The only difference is that one needs to make the substitution $\sigma_1^2 + \sigma_2^2 \mapsto (\sigma_1 - \sigma_2)^2$ and the changes in $\widehat{\alpha}_i$.

If $\sigma_1 = \sigma_2 =: \sigma$ the system (2.6) for $\tilde{\mathbf{Y}}(t) = (\tilde{Y}_1(t), \tilde{Y}_2(t))$ can be written as

$$\begin{cases} d\tilde{Y}_1(t) = & \Big(\tilde{Y}_1(t)(a_1 - a_1\tilde{Y}_1(t) - a_2\tilde{Y}_2(t)) - \alpha\tilde{Y}_1(t) + \beta\tilde{Y}_2(t)\Big)dt \\ & +\sigma^2\tilde{Y}_1(t)\Big[(\tilde{Y}_1(t) + \tilde{Y}_2(t))^2 - (\tilde{Y}_1(t) + \tilde{Y}_2(t))^2\Big]dt \\ d\tilde{Y}_1(t) = & \Big(\tilde{Y}_2(t)(a_2 - a_1\tilde{Y}_1(t) - a_2\tilde{Y}_2(t)) - \beta\tilde{Y}_2(t) + \alpha\tilde{Y}_1(t)\Big)dt \\ & +\sigma^2\tilde{Y}_2(t)\Big[(\tilde{Y}_1(t) + \tilde{Y}_2(t))^2 - (\tilde{Y}_1(t) + \tilde{Y}_2(t))^2\Big]dt. \end{cases} \tag{3.6}$$

Using the fact that $\tilde{Y}_1(t) + \tilde{Y}_2(t) = 1$ this reduces to

$$d\tilde{Y}_1(t) = \Big((a_1 - a_2)[1 - \tilde{Y}_1(t)]\tilde{Y}_1(t) + \beta - (\alpha + \beta)\tilde{Y}_1(t)]\Big)dt. \tag{3.7}$$

The unique equilibrium of 3.7 in [0,1] is the root $y^\star$ in [0,1] of $(a_1 - a_2)(1 - y)y + \beta - (\alpha + \beta)y = 0$. Hence, the unique invariant probability measure of $\tilde{\mathbf{Y}}(t)$ in this case is the Dirac measure concentrated in $(y^\star, 1 - y^\star)$. Thus

$$r = a_2 - \frac{\sigma^2}{2} + (a_1 - a_2 + \sigma^2)y^\star.$$

3.1 The degenerate case $\sigma_1 = \sigma_2, \alpha = \beta$

Consider the following system, where $\alpha, \sigma, a_i, b_i, i = 1, 2$ are positive constants.

$$\begin{cases} dX_1(t) = \big(X_1(t)(a_1 - b_1X_1(t)) - \alpha X_1(t) + \alpha X_2(t)\big)dt + \sigma X_1(t)dB(t) \\ dX_2(t) = \big(X_2(t)(a_2 - b_2X_2(t)) + \alpha X_1(t) - \alpha X_2(t)\big)dt + \sigma X_2(t)dB(t). \end{cases} \tag{3.8}$$

Suppose that $a_1 \neq a_2$ or that $b_1 \neq b_2$. This system is degenerate since both equations are driven by a single Brownian motion. In this case, the unique equilibrium of (3.7) in [0,1] is the root $y^\star$ in [0,1] of $(a_1 - a_2)(1 - y)y + \alpha(1 - 2y) = 0$. Solving this quadratic equation, we have $y^\star = \dfrac{a_1 - a_2 - 2\alpha + \sqrt{(a_1 - a_2)^2 + 4\alpha^2}}{2(a_1 - a_2)}$ if $a_1 \neq a_2$ and $y^\star = \frac{1}{2}$ if $a_1 = a_2$.

It can be proved easily that this equilibrium is asymptotically stable and that $\lim_{t\to\infty} \tilde{Y}_1(t) = y^\star$. Thus, if $a_1 \neq a_2$

$$\begin{aligned} r &= a_1 y^\star + a_2(1-y^\star) - \frac{\sigma^2}{2} \\ &= a_2 + \frac{a_1 - a_2 - 2\alpha + \sqrt{(a_1-a_2)^2 + 4\alpha^2}}{2} - \frac{\sigma^2}{2} \\ &= \frac{a_1 + a_2 - 2\alpha + \sqrt{(a_1-a_2)^2 + 4\alpha^2}}{2} - \frac{\sigma^2}{2}. \end{aligned}$$

As a result

$$r = \begin{cases} \dfrac{a_1 + a_2 - 2\alpha + \sqrt{(a_1-a_2)^2 + 4\alpha^2}}{2} - \dfrac{\sigma^2}{2} & \text{if } a_1 \neq a_2, b_1 = b_2 \\ a_1 - \dfrac{\sigma^2}{2} & \text{if } a_1 = a_2, b_1 \neq b_2. \end{cases} \tag{3.9}$$

Note that if $a_1 \neq a_2$ and $b_1 = b_2$

$$\alpha + (b_2/b_1)(a_1 - a_2) - \alpha(b_2/b_1)^2 = a_1 - a_2 \neq 0$$

and that if $a_1 = a_2$ and $b_1 \neq b_2$

$$\alpha + (b_2/b_1)(a_1 - a_2) - \alpha(b_2/b_1)^2 = \alpha\,(1 - b_2/b_1) \neq 0.$$

Therefore, the assumptions of Theorem 2.6 hold. If $r < 0$, by Theorem 2.5 the population goes extinct, while if $r > 0$, the population persists by Theorem 2.6.

3.2 The degenerate case when the conditions of Theorem 2.6 are violated

We analyse the system

$$\begin{cases} dX_1(t) = \big(X_1(t)(a_1 - bX_1(t)) - \alpha X_1(t) + \beta X_2(t)\big)dt + \sigma X_1(t)dB(t) \\ dX_2(t) = \big(X_2(t)(a_2 - bX_2(t)) + \alpha X_1(t) - \beta X_2(t)\big)dt + \sigma X_2(t)dB(t), \end{cases} \tag{3.10}$$

when $2(\beta - \alpha) = a_2 - a_1$. In this case $\sigma_1 = \sigma_2 = \sigma$,

$$\beta + (b_2/b_1)(a_1 - a_2 - \alpha + \beta) - \alpha(b_2/b_1) = 0$$

and

$$r = a_1 - \alpha + \beta - \frac{\sigma^2}{2}.$$

If $r < 0$ then $\lim_{t\to\infty} X_1(t) = \lim_{t\to\infty} X_2(t) = 0$ almost surely as the result of Theorem 2.5.

We focus on the case $r > 0$ and show that some of the results violate the conclusions of Theorem 2.6.

If we set $Z(t) = X_1(t)/X_2(t)$ then (see (C.6))

$$dZ(t) = \Big((1 - Z(t))Z(t)X_2(t) + \beta + \widehat{a}_1 Z(t) - \alpha Z^2(t)\Big)dt.$$

Noting that $\widehat{a}_1 = a_1 - a_2 - \alpha + \beta = \alpha - \beta$ yields

$$d(Z(t) - 1) = \Big(- (Z(t) - 1)Z(t)X_2(t) - (Z(t) - 1)(\alpha Z(t) + \beta)\Big)dt.$$

Assume $Z(0) \neq 1$ and without loss of generality suppose $Z(0) > 1$. This implies

$$Z(t) - 1 = (Z(0) - 1) \exp\left(- \int_0^t [Z(s)X_2(s) + (\alpha Z(s) + \beta)]\, ds \right). \quad (3.11)$$

Since $Z(t)$ and $X_2(t)$ do not explode to $\pm\infty$ in finite time we can conclude that if $Z(0) \neq 0$ then $Z(t) \neq 0$ for any $t \geq 0$ with probability 1. In other words, if $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^{2,\circ}_+$ with $x_1 \neq x_2$ then

$$\mathbb{P}(X_1^{\mathbf{x}}(t) = X_2^{\mathbf{x}}(t), t \geq 0) = 0.$$

One can further see from (3.11) that $Z(t) - 1$ tends to 0 exponentially fast. If $Z(0) = 1$ let $X_1(0) = X_2(0) = x > 0$. Similar arguments to the above show that

$$\mathbb{P}(X_1^{\mathbf{x}}(t) \neq X_2^{\mathbf{x}}(t), t \geq 0) = 0.$$

To gain more insight into the asymptotic properties of $(X_1(t), X_2(t))$, we study

$$\begin{aligned} dX_2(t) &= X_2(t)\Big((\widehat{a}_2 - bX_2(t)) + \alpha Z(t)\Big)dt + \sigma X_2(t)dB(t) \\ &= X_2(t)\Big(a_1 - \alpha + \beta - bX_2(t)) + \alpha(Z(t) - 1)\Big)dt + \sigma X_2(t)dB(t) \end{aligned}$$

We have from Itô's formula that,

$$\begin{aligned} d\frac{1}{X_2(t)} = &\left(b + (-a_1 + \alpha - \beta + \sigma^2 - \alpha(Z(t) - 1))\frac{1}{X_2(t)}\right) dt \\ &- \sigma \frac{1}{X_2(t)} dB(t). \end{aligned}$$

By the variation-of constants formula (see Mao 1997, Section 3.4), we have

$$\frac{1}{X_2(t)} = \phi^{-1}(t)\left[\frac{1}{x_2} + b\int_0^t \phi(s)ds\right]$$

where

$$\phi(t) := \exp\left[rt + \alpha \int_0^t (Z(s) - 1)ds + \sigma B(t)\right].$$

Thus,

$$X_2(t) = \frac{\phi(t)}{x_2^{-1} + b\int_0^t \phi(s)ds}.$$

It is well-known that

$$U(t) := \frac{e^{rt+\sigma B(t)}}{x_2^{-1} + b\int_0^t e^{rs+\sigma B(s)}ds},$$

is the solution to the stochastic logistic equation

$$dU(t) = U(t)(a_1 - \alpha + \beta - bU(t))\,dt + \sigma U(t)dB(t), U(0) = x_2.$$

By the law of the iterated logarithm, almost surely

$$\lim_{t\to\infty} \phi(t) = \lim_{t\to\infty} e^{rt+\sigma B(t)} = \infty. \tag{3.12}$$

We have

$$\frac{X_2(t)}{U(t)} = \frac{\exp\left(\alpha\int_0^t (Z(s) - 1)ds\right)\left[x_2^{-1} + b\int_0^t e^{rt+\sigma B(t)}ds\right]}{x_2^{-1} + b\int_0^t \phi(s)ds}.$$

In view of (3.12), we can use L'hospital's rule to obtain

$$\begin{aligned}
&\lim_{t\to\infty} \frac{X_2(t)}{U(t)} \\
&= \lim_{t\to\infty} \frac{\exp\left(\alpha\int_0^t (Z(s) - 1)ds\right) e^{rt+\sigma B(t)}}{\phi(t)} \\
&\quad + \lim_{t\to\infty} \frac{\alpha(Z(t) - 1)\exp\left(\alpha\int_0^t (Z(s) - 1)ds\right)\left[x_2^{-1} + b\int_0^t e^{rt+\sigma B(t)}ds\right]}{b\phi(t)} \\
&= 1 + \lim_{t\to\infty} \frac{\alpha(Z(t) - 1)\left[x_2^{-1} + b\int_0^t e^{rt+\sigma B(t)}ds\right]}{be^{rt+\sigma B(t)}}
\end{aligned} \tag{3.13}$$

almost surely. By the law of the iterated logarithm, $\lim_{t\to\infty} \frac{e^{rt+\sigma B(t)}}{e^{(r-\varepsilon)t}} = \infty$ and $\lim_{t\to\infty} \frac{e^{rt+\sigma B(t)}}{e^{(r+\varepsilon)t}} = 0$ for any $\varepsilon > 0$. Applying this and (3.11) to (3.13), it is easy to show that with probability 1

$$\lim_{t\to\infty} \frac{X_2(t)}{U(t)} = 1.$$

Since $\lim_{t\to\infty} Z(t) = 1$ almost surely, we also have $\lim_{t\to\infty} \frac{X_1(t)}{U(t)} = 1$ almost surely. Thus, the long term behavior of $X_1(t)$ and $X_2(t)$ is governed by the one-dimensional diffusion $U(t)$. In particular, both $X_1(t)$ and $X_2(t)$ converge to a unique invariant probability measure ρ on $(0, \infty)$, which is the invariant probability measure of $U(t)$. In this case, the invariant probability measure of $\mathbf{X}(t) = (X_1(t), X_2(t))_{t\geq 0}$ is not absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{2,\circ}_+$. Instead, the invariant probability measure is concentrated on the one-dimensional manifold $\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^{2,\circ}_+ : x_1 = x_2\}$.

Biological interpretation *The stochastic growth rate in this degenerate setting is given by* $r = a_1 - \alpha + \beta - \frac{\sigma^2}{2}$. *We note that this term is equal to the stochastic growth rate of patch* 1, $a_1 - \frac{\sigma^2}{2}$, *to which we add* β, *the rate of dispersal from patch* 1 *to patch* 2, *and subtract* α, *the rate of dispersal from patch* 2 *to patch* 1. *When*

$$a_1 - \frac{\sigma^2}{2} > \alpha - \beta$$

one has persistence, while when

$$a_1 - \frac{\sigma^2}{2} < \alpha - \beta$$

one has extinction. In particular, if the patches on their own are sink patches so that $a_1 - \frac{\sigma^2}{2} < 0$ *and* $a_2 - \frac{\sigma^2}{2} < 0$ *dispersion cannot lead to persistence since*

$$a_1 - \frac{\sigma^2}{2} > \alpha - \beta \text{ and } a_2 - \frac{\sigma^2}{2} > \beta - \alpha$$

cannot hold simultaneously. The behavior of the system when $r > 0$ *is different from the behavior in the non-degenerate setting of Theorem 2.1 or the degenerate setting of Theorem 2.6. Namely, if the patches start with equal populations then the patch abundances remain equal for all times and evolve according to the one-dimensional logistic diffusion* $U(t)$. *If the patches start with different population abundances then* $X_1(t)$ *and* $X_2(t)$ *are never equal but tend to each other asymptotically as* $t \to \infty$. *Furthermore, the long term behavior of* $X_1(t)$ *and* $X_2(t)$ *is once again determined by the logistic diffusion* $U(t)$ *as almost surely* $\frac{X_i(t)}{U(t)} \to 1$ *as* $t \to \infty$. *As such, if* $r > 0$ *we have persistence but the invariant measure the system converges to does not have* $\mathbb{R}^{2,\circ}_+$ *as its support anymore. Instead the invariant measure has the line* $\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^{2,\circ}_+ : x_1 = x_2\}$ *as its support.*

Example 3.1 We discuss the case when $a_1 \neq a_2$ and $\sigma_1 = \sigma_2$. The stochastic growth rate can be written by the analysis in the sections above as

$$r = \begin{cases} \dfrac{a_1 + a_2 - 2\alpha + \sqrt{(a_1 - a_2)^2 + 4\alpha^2}}{2} - \dfrac{\sigma^2}{2} & \text{if } \alpha = \beta, b_1 = b_2 \\ a_1 - \alpha + \beta - \frac{\sigma^2}{2} & \text{if } a_2 - a_1 = 2(\beta - \alpha), b_1 = b_2. \end{cases} \tag{3.14}$$

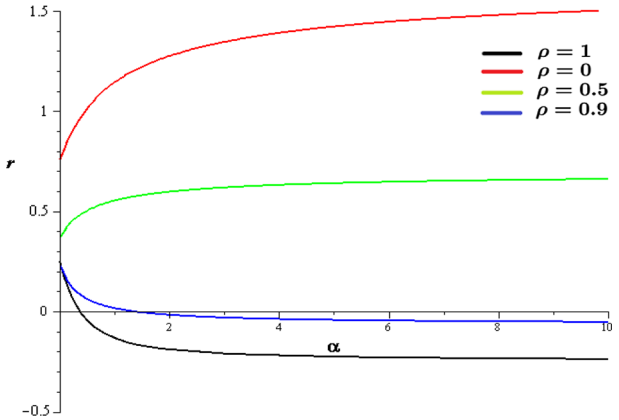


Fig. 1 Consider (3.2) when $\alpha = \beta$ and the Brownian motions B_1 and B_2 are assumed to have correlation ρ. The *graphs* show the stochastic growth rate r as a function of the dispersal rate α for different values of the correlation. Note that if $\rho = 0$ we get the setting when the Brownian motions of the two patches are independent while when $\rho = 1$ we get that one Brownian motion drives the dynamics of both patches. The parameters are $\alpha = \beta, a_1 = 3, a_2 = 4, \sigma_1^2 = \sigma_2^2 = 7$

Biological interpretation *In the case when $a_1 = a_2, \sigma_1 = \sigma_2$ and $b_1 \neq b_2$ (so that the two patches only differ in their competition rates) the stochastic growth rate r does not depend on the dispersal rate α. The system behaves just as a single-patch system with stochastic growth rate $a_1 - \frac{\sigma^2}{2}$. In contrast to* Evans et al. (2013, Example 1) *coupling two sink patches by dispersion cannot yield persistence.*

However, if the growth rates of the patches are different $a_1 \neq a_2$ then the expression for r given in (3.14) *yields for $\alpha \gg |a_1 - a_2|$ that*

$$r \approx \frac{a_1 + a_2}{2} - \frac{\sigma^2}{2} + \frac{(a_1 - a_2)^2}{8\alpha}.$$

In particular

$$\lim_{\alpha \to \infty} r(\alpha) = \frac{a_1 + a_2}{2} - \frac{\sigma^2}{2}.$$

We note that r is a decreasing function of the dispersal rate α for large values of α (also see Fig. 1). This is different from the result of Evans et al. (2013, Example 1) *where r was shown to be an increasing function of α. In contrast to the non-degenerate case, coupling patches by dispersal decreases the stochastic growth rate and as such makes persistence less likely. This highlights the negative effect of spatial correlations on population persistence and why one may no longer get the rescue effect. This is one of your main biological conclusions. Furthermore, we also recover that dispersal has a negative impact on the stochastic growth rate when there is spatial heterogeneity (i.e. $a_1 \neq a_2$). This fact has a long history, going back to the work by* Karlin (1982).

4 Discussion and generalizations

For numerous models of population dynamics it is natural to assume that time is continuous. One reason for this is that often environmental conditions change continuously with time and therefore can naturally be described by continuous time models. There have been a few papers dedicated to the study of stochastic differential equation models of interacting, unstructured populations in stochastic environments (see Benaïm et al. 2008; Schreiber et al. 2011; Evans et al. 2015). These models however do not account for population structure or correlated environmental fluctuations.

Examples of structured populations can be found by looking at a population in which individuals can live in one of n patches (e.g. fish swimming between basins of a lake or butterflies dispersing between meadows). Dispersion is viewed by many population biologists as an important mechanism for survival. Not only does dispersion allow individuals to escape unfavorable landscapes (due to environmental changes or lack of resources), it also facilitates populations to smooth out local spatio-temporal environmental changes. Patch models of dispersion have been studied extensively in the deterministic setting (see for example Hastings 1983; Cantrell et al. 2012). In the stochastic setting, there have been results for discrete time and space by Benaïm and Schreiber (2009), for continuous time and discrete space by Evans et al. (2013) and for structured populations that evolve continuously both in time and space.

We analyze the dynamics of a population that is spread throughout n patches, evolves in a stochastic environment (that can be spatially correlated), disperses among the patches and whose members compete with each other for resources. We characterize the long-term behavior of our system as a function of r—the growth rate in the absence of competition. The quantity r is also the Lyapunov exponent of a suitable linearization of the system around 0. Our analysis shows that $r < 0$ implies extinction and $r > 0$ persistence. The limit case $r = 0$ cannot be analyzed in our framework. We expect that new methods have to be developed in order to tackle the $r = 0$ scenario.

Since mathematical models are always approximations of nature it is necessary to study how the persistence and extinction results change under small perturbations of the parameters of the models. The concept of robust persistence (or permanence) has been introduced by Hutson and Schmitt (1992). They showed that for certain systems persistence holds even when one has small perturbations of the growth functions. There have been results on robust persistence in the deterministic setting for Kolmogorov systems by Schreiber (2000) and Garay and Hofbauer (2003). Recently, robust permanence for deterministic Kolmogorov equations with respect to perturbations in both the growth functions and the feedback dynamics has been analyzed by Patel and Schreiber (2016). In the stochastic differential equations setting results on robust persistence and extinction have been proven by Schreiber et al. (2011) and Benaïm et al. (2008). We prove analogous results in our framework where the populations are coupled by dispersal. For robust persistence we show in Appendix D that even with *density-dependent* perturbations of the growth rates, dispersion matrix and environmental covariance matrix, if these perturbations are sufficiently small and if the unperturbed system is persistent then the perturbed system is also persistent. In the case of extinction we can prove robustness when there are small *constant* perturbations of the growth rates, dispersal matrices and covariance matrices.

In ecology there has been an increased interest in the spatial synchrony present in population dynamics. This refers to the changes in the time-dependent characteristics (i.e. abundances etc) of structured populations. One of the mechanisms which creates synchrony is the dependence of the population dynamics on a synchronous random environmental factor such as temperature or rainfall. The synchronizing effect of environmental stochasticity, or the so-called *Moran effect*, has been observed in multiple population models. Usually this effect is the result of random but correlated weather effects acting on spatially structured populations. Following Legendre (1993) one could argue that our world is a spatially correlated one. For many biotic and abiotic factors, like population density, temperature or growth rate, values at close locations are usually similar. For an in-depth analysis of spatial synchrony see Kendall et al. (2000) and Liebhold et al. (2004). Most stochastic differential models appearing in population dynamics treat only the case when the noise is non-degenerate (although see Rudnicki 2003; Dieu et al. 2016). This simplifies the technical proofs significantly. However, from a biological point of view it is not clear that the noise should never be degenerate. For example if one models a system with multiple populations then all populations can be influenced by the same factors (a disease, changes in temperature and sunlight etc). Environmental factors can intrinsically create spatial correlations and as such it makes sense to study how these degenerate systems compare to the non-degenerate ones. In our setting the n different patches could be strongly spatially correlated. Actually, in some cases it could be more realistic to have the same one-dimensional Brownian motion $(B_t)_{t\geq 0}$ driving the dynamics of all patches. We were able to find conditions under which the proofs from the non-degenerate case can be generalized to the degenerate setting. This is a first step towards a model that tries to explain the complex relationship between dispersal, stochastic environments and spatial correlations.

We fully analyze what happens if there are only two patches, $n = 2$, and the noise is degenerate. Our results show unexpectedly, and in contrast to the non-degenerate results by Evans et al. (2013), that coupling two sink patches cannot yield persistence. More generally, we show that the stochastic growth rate is a decreasing function of the dispersal rate. In specific instances of the degenerate setting, even when there is persistence, the invariant probability measure the system converges to does not have $\mathbb{R}^{2,\circ}_+$ as its support. Instead, the abundances of the two patches converge to an invariant probability measure supported on the line $\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^{2,\circ}_+ : x_1 = x_2\}$. These examples shows that degenerate noise is not just an added technicality—the results can be completely different from those in the non-degenerate setting. The negative effect of spatial correlations (including the fully degenerate case) has been studied in several papers for discrete-time models (see Schreiber 2010; Harrison and Quinn 1989; Palmqvist and Lundberg 1998; Bascompte et al. 2002; Roy et al. 2005). The negative impact of dispersal on the stochastic growth rate r when there is spatial heterogeneity (i.e. $a_1 \neq a_2$) has a long history going back to the work of Karlin (1982) on the *Reduction Principle*. Following Altenberg (2012) the reduction principle can be stated as the widely exhibited phenomenon that mixing reduces growth, and differential growth selects for reduced mixing. The first use of this principle in the study of the evolution of dispersal can be found in Hastings (1983). The work of

Kirkland et al. (2006) provides an independent proof of the Reduction Principle and applications to nonlinear competing species in discrete-time, discrete-space models. In the case of continuous-time, discrete-space models (given by branching processes) a version of the Reduction Principle is analysed by Schreiber and Lloyd-Smith (2009).

4.1 k species competing and dispersing in n patches

Real populations do not evolve in isolation and as a result much of ecology is concerned with understanding the characteristics that allow two species to coexist, or one species to take over the habitat of another. It is of fundamental importance to understand what will happen to an invading species. Will it invade successfully or die out in the attempt? If it does invade, will it coexist with the native population? Mathematical models for invasibility have contributed significantly to the understanding of the epidemiology of infectious disease outbreaks (Cross et al. 2005) and ecological processes (Law and Morton 1996; Caswell 2001). There is widespread empirical evidence that heterogeneity, arising from abiotic (precipitation, temperature, sunlight) or biotic (competition, predation) factors, is important in determining invasibility (Davies et al. 2005; Pyšek and Hulme 2005). However, few theoretical studies have investigated this; see, e.g., Schreiber and Lloyd-Smith (2009), Schreiber and Ryan (2011) and Schreiber (2012).

In this paper we have considered the dynamics of one population that disperses through n patches. One possible generalization would be to look at k populations $(\mathbf{X}^1, \ldots, \mathbf{X}^k)$ that compete with each other for resources, have different dispersion strategies and possibly experience the environmental noise differently. Looking at such a model could shed light upon fundamental problems regarding invasions in spatio-temporally heterogeneous environments.

The extension of our results to competition models could lead to the development of a stochastic version of the treatment of the evolution of dispersal developed for patch models in the deterministic setting by Hastings (1983) and Cantrell et al. (2012). In the current paper we have focused on how spatio-temporal variation influences the persistence and extinction of structured populations. In a follow-up paper we intend to look at the dispersal strategies in terms of *evolutionarily stable strategies* (ESS) which can be characterized by showing that a population having a dispersal strategy (D_{ij}) cannot be invaded by any other population having a different dispersal strategy $(\tilde{D}_{ij})$. The first thing to check would be whether this model has ESS and, if they exist, whether they are unique. One might even get that there are no ESS in our setting. For example, Schreiber and Li (2011) show that there exist no ESS for periodic non-linear models and instead one gets a coalition of strategies that act as an ESS. We expect to be able to generalize the results of Cantrell et al. (2012) to a stochastic setting using the methods from this paper.

Acknowledgements We thank Sebastian J. Schreiber and three anonymous referees for their detailed comments which helped improve this manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix A: The case $r > 0$

The next sequence of lemmas and propositions is used to prove Theorem 2.1. We start by showing that our processes are well-defined Markov processes.

Proposition A.1 *The SDE (stochastic differential equation) defined by (2.1) has unique strong solutions* $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))$, $t \geq 0$ *for any* $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n_+$. *Furthermore,* $\mathbf{X}(t)$ *is a strong Markov process with the Feller property, is irreducible on* $\mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ *and* $\mathbb{P}\{X_i(t) > 0,\ t > 0, i = 1, \dots, n\} = 1$ *if* $\mathbf{X}(0) \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$.

Proof Since the coefficients of (2.1) are locally Lipschitz, there exists a unique local solution to (2.1) with a given initial value. In other words, for any initial value, there is a stopping time $\tau_e > 0$ and a process $(\mathbf{X}(t))_{t \geq 0}$ satisfying (2.1) up to τ_e and $\lim_{t \to \tau_e} \|\mathbf{X}(t)\| = \infty$ (see e.g. Khasminskii 2012, Section 3.4). Clearly, if $\mathbf{X}(0) = 0$ then $\mathbf{X}(t) = 0,\ t \in [0, \tau_e)$ which implies that $\tau_e = \infty$. By a comparison theorem for SDEs (see Geiß and Manthey (1994, Theorem 1.2) and Remark A.2 below),

$$\mathbb{P}\{X_i(t) < \mathcal{X}_i(t),\ t \in (0, \tau_e), i = 1, \dots, n\} = 1 \text{ if } X_i(0) = \mathcal{X}_i(0) \geq M_b \quad \text{(A.1)}$$

where $(\mathcal{X}_i(t))_{t \geq 0}$ is given by (2.7). Since (2.7) has a global solution due to the Lipschitz property of its coefficients, we have from (A.1) that $\tau_e = \infty$ almost surely. Define the process

$$d\overline{\mathcal{X}}_i(t) = \left(-\left|\frac{3a_i}{2}\right| \overline{\mathcal{X}}_i(t) + \sum_{j=1}^{n} D_{ji} \overline{\mathcal{X}}_j(t) \right) dt + \overline{\mathcal{X}}_i(t) dE_i(t),\ \ i = 1, \dots, n.$$

Since the b_is are continuous and vanish at 0, there exists $r > 0$ such that for $|\mathbf{x}| \leq r$ we have

$$-\left|\frac{3a_i}{2}\right| \leq a_i - b_i(x_i),\ i = 1, \dots, n. \quad \text{(A.2)}$$

Let τ be the stopping time

$$\tau := \inf\left\{t : \left|\overline{\mathcal{X}}(t)\right| > r\right\} \quad \text{(A.3)}$$

Now, consider the case $\mathbf{X}(0) \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$. By Evans et al. (2013, Proposition 3.1), (A.2), (A.3) and a comparison argument (see Remark A.2 and the proof of Evans et al. (2015, Theorem 4.1)), we can show that

$$\mathbb{P}\left\{X_i \geq \overline{\mathcal{X}}_i(t) > 0, t \in (0, \tau)\right\} = 1,$$

which implies

$$\mathbb{P}\{X_i(t) > 0, t \in (0,\infty)\} = 1 \text{ for all } \mathbf{X}(0) \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}. \tag{A.4}$$

Moreover, since $\mathbb{P}\{0 \le X_i(t) < \mathcal{X}_i(t) \text{ for all } t \ge 0, i = 1,\dots,n\} = 1$, we can use standard arguments (e.g., Mao 1997, Theorem 2.9.3) to obtain the Feller property of the solution to (2.1). □

Remark A.1 There are different possible definitions of "Feller" in the literature. What we mean by Feller is that the semigroup $(T_t)_{t\ge 0}$ of the process maps the set of bounded continuous functions $C_b(\mathbb{R}^n_+)$ into itself i.e.

$$T_t(C_b(\mathbb{R}^n_+)) \subset C_b(\mathbb{R}^n_+),\ t \ge 0.$$

Definition A.1 We call a mapping $f : \mathbb{R}^d \to \mathbb{R}^d$ **quasi-monotonously increasing**, if for $j = 1,\dots,d$

$$f_j(x) \le f_j(y),$$

whenever $x_j = y_j$ and $x_l \le y_l, l \ne j$.

Remark A.2 One often wants to apply the well-known comparison theorem for one-dimensional SDEs (see Ikeda and Watanabe 1989) to a multidimensional setting. Below we explain why we can make use of comparison theorems for stochastic differential equations in our setting. Consider the following two systems

$$dR_j(t) = a_j(t, R(t))\,dt + \sum_{k=1}^{r} \sigma_{jk}(t, R(t))dW_k(t) \tag{A.5}$$

and

$$dS_j(t) = b_j(t, S(t))\,dt + \sum_{k=1}^{r} \sigma_{jk}(t, S(t))dW_k(t) \tag{A.6}$$

for $j = 1,\dots,d, t \ge 0$ together with the initial condition

$$R_j(0) \le S_j(0),\ j = 1,\dots,d \quad \mathbb{P}-\text{a.s.}, \tag{A.7}$$

where $W = (W_1(t),\dots,W_r(t))_{t\ge 0}$ is an r-dimensional standard Brownian motion, and the coefficients a_i, b_i, σ_{jk} are continuous mappings on $\mathbb{R}_+ \times \mathbb{R}^d$. Suppose (A.5) and (A.6) have explosion times θ_R, θ_S.

Let (C0), (C1), and (C2) be the following conditions.

(C0) The solution to (A.5) is pathwise unique and the drift coefficient $a(t,x)$ is quasi-monotonously (see Definition A.1) increasing with respect to x.

(C1) For every $t \geq 0$, $j = 1, \ldots, d$ and $x \in \mathbb{R}^d$ the following inequality holds

$$a_j(t, x) \leq b_j(t, x).$$

(C2) There exists a strictly increasing function $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ with $\rho(0) = 0$ and

$$\int_{0+}^{\infty} \frac{1}{\rho^2(u)}\, du = \infty$$

such that for each $j = 1, \ldots, d$

$$\sum_{k=1}^{r} |\sigma_{jk}(t, x) - \sigma_{jk}(t, y)| \leq \rho(|x_j - y_j|) \text{ for all } t \geq 0, x, y \in \mathbb{R}^d.$$

Sometimes it is assumed incorrectly that conditions (C1) and (C2) suffice to conclude that $\mathbb{P}\{R(t) \leq Y(t), t \in [0, \theta_R \wedge \theta_S)\} = 1$. Some illuminating counterexamples regarding this issue can be found in Assing and Manthey (1995, Section 3). However, if in addition to conditions (C1) and (C2), one also has condition (C0), then Geiß and Manthey (1994, Theorem 1.2) indicates that $\mathbb{P}\{R(t) \leq Y(t), t \in [0, \theta_R \wedge \theta_S)\} = 1$. Note that, in the setting of our paper, the drift coefficient of (2.7) is quasi-monotonously increasing and we can pick $\rho(x) = x, x \in \mathbb{R}_+$. Therefore, conditions (C0), (C1), and C(2) hold, which allows us to use the comparison results. In special cases one can prove comparison theorems even when quasi-monotonicity fails; see Evans et al. (2015, Theorem 6.1) and Nlath et al. (2007, Corollary A.2).

To proceed, let us recall some technical concepts and results needed to prove the main theorem. Let $\Phi = (\Phi_0, \Phi_1, \ldots)$ be a discrete-time Markov chain on a general state space $(E, \mathcal{E})$, where $\mathcal{E}$ is a countably generated σ-algebra. Denote by $\mathcal{P}$ the Markov transition kernel for Φ. If there is a non-trivial σ-finite positive measure φ on $(E, \mathcal{E})$ such that for any $A \in \mathcal{E}$ satisfying $\varphi(A) > 0$ we have

$$\sum_{n=1}^{\infty} \mathcal{P}^n(x, A) > 0, \ x \in E$$

where $\mathcal{P}^n$ is the n-step transition kernel of Φ, then the Markov chain Φ is called *φ-irreducible*. It can be shown (see Nummelin 1984) that if Φ is φ-irreducible, then there exists a positive integer d and disjoint subsets $E_0, \ldots, E_{d-1}$ such that for all $i = 0, \ldots, d-1$ and all $x \in E_i$, we have

$$\mathcal{P}(x, E_j) = 1 \text{ where } j = i + 1 \ (\text{mod } d)$$

and

$$\varphi\left(E \setminus \bigcup_{i=0}^{d-1} E_i\right) = 0.$$

The smallest positive integer d satisfying the above is called the period of Φ. An *aperiodic* Markov chain is a chain with period $d = 1$.

A set $C \in \mathcal{E}$ is called *petite*, if there exists a non-negative sequence $(a_n)_{n\in\mathbb{N}}$ with $\sum_{n=1}^{\infty} a_n = 1$ and a nontrivial positive measure ν on $(E, \mathcal{E})$ such that

$$\sum_{n=1}^{\infty} a_n \mathcal{P}^n(x, A) \geq \nu(A), \;\; x \in C, A \in \mathcal{E}.$$

The following theorem is extracted from Jarner and Roberts (2002, Theorem 3.6).

Theorem A.1 *Suppose that Φ is irreducible and aperiodic and fix $0 < \gamma < 1$. Assume that there exists a petite set $C \subset E$, positive constants κ_1, κ_2 and a function $V : E \to [1, \infty)$ such that*

$$\mathcal{P}V \leq V - \kappa_1 V^{\gamma} + \kappa_2 \mathbf{1}_C.$$

Then there exists a probability measure π on $(E, \mathcal{E})$ such that

$$(n+1)^{\frac{\gamma}{1-\gamma}} \|\mathcal{P}(x, \cdot) - \pi(\cdot)\|_{TV} \to 0 \text{ as } n \to \infty \text{ for all } x \in E.$$

The next series of lemmas and propositions are used to show that we can construct a function V satisfying the assumptions of Theorem A.1.

Lemma A.1 *For any $T > 0$, there exists an open set $N_0 \subset \mathbb{R}^{n,\circ}_+$ such that the Markov chain $\{(\mathbf{Y}(kT), S(kT)), k \in \mathbb{N}\}$ on $\Delta \times (0, \infty)$ is φ-irreducible and aperiodic, where $\varphi(\cdot) = m(\cdot \cap N_0)$ and $m(\cdot)$ is Lebesgue measure. Moreover, every compact set $K \subset \Delta \times (0, \infty)$ is petite. Similarly, Δ is a petite set of the Markov chain $\{\tilde{\mathbf{Y}}(kT), k \in \mathbb{N}\}$.*

Proof To prove this lemma, it is more convenient to work with the process $\mathbf{X}(t)$ that lives on $\mathbb{R}^n_+ \backslash \{\mathbf{0}\}$. Since $(\mathbf{X}(t))_{t\geq 0}$ is a nondegenerate diffusion with smooth coefficients in $\mathbb{R}^{n,\circ}_+$, by Rey-Bellet (2006, Corollary 7.2), the transition semigroup $P_{\mathbf{X}}(t, \mathbf{x}, \cdot)$ of $(\mathbf{X}(t))_{t\geq 0}$ has a smooth, positive density $(0, \infty) \times \mathbb{R}^{2n,\circ}_+ \ni (t, \mathbf{x}, \mathbf{x}')) \mapsto p_{\mathbf{X}}(t, \mathbf{x}, \mathbf{x}') \in [0, \infty)$. Fix a point $\mathbf{x}_0 \in \mathbb{R}^{n,\circ}_+$. Since $\int_{\mathbb{R}^{n,\circ}_+} p(t, \mathbf{x}_0, \mathbf{x}) d\mathbf{x} = 1$ there exists $\mathbf{x}_1 \in \mathbb{R}^{n,\circ}_+$ such that $p_{\mathbf{X}}\left(\frac{T}{2}, \mathbf{x}_0, \mathbf{x}_1\right) > 0$. There exist bounded open sets $N_0 \ni \mathbf{x}_0, N_1 \ni \mathbf{x}_1$ satisfying

$$\widehat{p} := \inf \left\{ p_{\mathbf{X}}\left(\frac{T}{2}, \mathbf{x}, \mathbf{x}'\right) > 0 : \mathbf{x} \in N_0, \mathbf{x}' \in N_1 \right\} > 0. \tag{A.8}$$

Slightly modifying the proof of Evans et al. (2013, Proposition 3.1) (the part proving the irreducibility of the solution process), we have that $\tilde{p}_{\mathbf{x}} := P_{\mathbf{X}}\left(\frac{T}{2}, \mathbf{x}, N_0\right) > 0$ for all $\mathbf{x} \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}$. Since $(\mathbf{X}(t))_{t\geq 0}$ has the Feller property, there is a neighborhood $N_{\mathbf{x}} \ni \mathbf{x}$ such that

$$P_{\mathbf{X}}\left(\frac{T}{2}, \mathbf{x}', N_0\right) > \frac{\tilde{p}_{\mathbf{x}}}{2}, \; \mathbf{x}' \in N_{\mathbf{x}}. \tag{A.9}$$

For any compact set $K \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}$, there are finite $\mathbf{x}_2, \ldots, \mathbf{x}_k$ such that $K \subset \bigcup_{i=2}^k N_{\mathbf{x}_i}$. As a result,

$$P_{\mathbf{X}}\left(\frac{T}{2}, \mathbf{x}', N_0\right) > \tilde{p}_K := \min\left\{\frac{\tilde{p}_{\mathbf{x}_i}}{2}, i = 2, \ldots, k\right\}. \tag{A.10}$$

In view of (A.8), (A.9), and (A.10), an application of the Chapman-Kolmogorov equations yields that for any $\mathbf{x} \in K$ and any measurable set $A \subset \mathbb{R}^{n,\circ}_+$,

$$\begin{aligned} P_{\mathbf{X}}\big(T, \mathbf{x}, A\big) &\geq \int_{N_0} P_{\mathbf{X}}\left(\frac{T}{2}, \mathbf{x}, d\mathbf{x}'\right) P_{\mathbf{X}}\left(\frac{T}{2}, \mathbf{x}', A\right) \\ &\geq \tilde{p}_K \widehat{p} m(A \cap N_1), \end{aligned}$$

where $m(\cdot)$ is Lebesgue measure on $\mathbb{R}^{n,\circ}_+$. Since the measure $\nu(\cdot) = m(\cdot \cap N_1)$ is non-trivial, we can easily obtain that K is a petite set of the Markov chain $\{(\mathbf{X}(kT)), k \in \mathbb{N}\}$. Moreover, K can be chosen arbitrarily. Hence, for any $\mathbf{x} \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}$ there is $\overline{p}_{\mathbf{x}} > 0$ such that

$$P_{\mathbf{X}}\big(T, \mathbf{x}, \cdot\big) \geq \overline{p}_{\mathbf{x}} m(\cdot \cap N_1), \tag{A.11}$$

which means that $\{(\mathbf{X}(kT)), k \in \mathbb{N}\}$ is irreducible.

Suppose that $\{(\mathbf{X}(kT)), k \in \mathbb{N}\}$ is not aperiodic. Then there are disjoint subsets of $\mathbb{R}^n_+ \backslash \{\mathbf{0}\}$, denoted by $A_0, \ldots, A_{d-1}$ with $d > 1$ such that for any $\mathbf{x} \in A_i$,

$$P_{\mathbf{X}}(T, \mathbf{x}, A_j) = 1 \text{ where } j = i + 1 \ (\text{mod } d).$$

Since $P(T, \mathbf{x}, \cdot)$ has a density, $m(A_i) > 0$ for $i = 0, \ldots, d-1$. In view of (A.11), we must have $m(N_0 \cap A_i) = 0$ for any $i = 0, \ldots, d-1$. This contradicts the fact that

$$m\left(N_0 \bigcap \left(E \backslash \bigcup_{i=0}^{d-1} A_i\right)\right) = 0.$$

This contradiction implies that $\{\mathbf{X}(kT), k \in \mathbb{N}\}$ is aperiodic. In the same manner, we can prove that $\tilde{\mathbf{Y}}(t)$ is irreducible, aperiodic and its state space, Δ, is petite. □

Lemma A.2 *There exists a positive constant K_1 such that*

$$\mathbb{E}S^{\mathbf{y},s}(t) \leq e^{-\gamma_b t} s + K_1, \ \ (\mathbf{y}, s) \in \Delta \times (0, \infty), t \geq 0. \tag{A.12}$$

Moreover, for any $H > 0$, $T > 0$, and $\varepsilon > 0$, there is a $\tilde{k} = \tilde{k}(H, T, \varepsilon) > 0$ such that

$$\mathbb{P}\{S^{\mathbf{y},s}(t) < \tilde{k}, \ t \in [0, T]\} > 1 - \varepsilon, (\mathbf{y}, s) \in \Delta \times (0, H]. \tag{A.13}$$

Proof In view of (2.2), if $s \geq M_b$ then $-[\mathbf{b}(s\mathbf{y})]^\top \mathbf{y} + \mathbf{a}^\top \mathbf{y} + \gamma_b \leq 0$. Let

$$\tilde{K}_1 = \sup_{\mathbf{y}\in\Delta, s\leq M_b} \left\{ s(-[\mathbf{b}(s\mathbf{y})]^\top \mathbf{y} + \mathbf{a}^\top \mathbf{y} + \gamma_b) \right\} < \infty.$$

For $k \in \mathbb{N}$, define the bounded stopping time

$$\eta_k^{\mathbf{y},s} = \inf\{t \geq 0 : S^{\mathbf{y},s}(t) \geq k\}. \tag{A.14}$$

Dynkin's formula for the function $f(t,s) := e^{\gamma_b t} s$ and the bounded stopping time $t \wedge \eta_k^{\mathbf{y},s}$ yield

$$\begin{aligned}\mathbb{E}[e^{\gamma_b t\wedge\eta_k^{\mathbf{y},s}} S^{\mathbf{y},s}(t \wedge \eta_k^{\mathbf{y},s})] &= s + \mathbb{E}\int_0^{t\wedge\eta_k^{\mathbf{y},s}} e^{\gamma_b u} S^{\mathbf{y},s}(u) \\ &\quad \left(\gamma_b + [\mathbf{a} - \mathbf{b}(S^{\mathbf{y},s}(u)\mathbf{Y}^{\mathbf{y},s}(u)]^\top \mathbf{Y}^{\mathbf{y},s}(u)\right) du \\ &\leq s + \mathbb{E}\int_0^{t\wedge\eta_k^{\mathbf{y},s}} \tilde{K}_1 e^{\gamma_b u} du \leq s + \frac{\tilde{K}_1}{\gamma_b}(e^{\gamma_b t} - 1).\end{aligned} \tag{A.15}$$

The claim (A.13) follows directly from (A.15). Moreover, by letting $k \to \infty$ in (A.15), we obtain from Fatou's lemma that

$$\mathbb{E}e^{\gamma_b t} S^{\mathbf{y},s}(t) \leq s + K_1 e^{\gamma_b t} \text{ for } K_1 = \frac{\tilde{K}_1}{\gamma_b}, \tag{A.16}$$

which implies (A.12). □

Proposition A.2 *For any $\varepsilon > 0$ and $T > 0$, there is a $\delta = \delta(\varepsilon, T) > 0$ such that*

$$\mathbb{P}\left\{ \|(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t)) - (\tilde{\mathbf{Y}}^{\mathbf{y}}(t), 0)\| \leq \varepsilon,\ 0 \leq t \leq T \right\} > 1 - \varepsilon$$

given that $(\mathbf{y}, s) \in \Delta \times (0, \delta)$.

Proof In view of (A.13), for any $\varepsilon > 0, T > 0$, there is $\tilde{k} = \tilde{k}(\varepsilon, T) > 0$ such that

$$\mathbb{P}\{\eta_{\tilde{k}}^{\mathbf{y},s} < T\} \geq 1 - \frac{\varepsilon}{3},\ (\mathbf{y}, s) \in \Delta \times (0, \varepsilon) \tag{A.17}$$

where $\eta_k^{\mathbf{y},s}$ is defined by (A.14). Since the coefficients of equation (2.4) are locally Lipschitz, using the arguments from Mao (1997, Lemma 9.4) and noting $S^{\mathbf{y},0}(t) \equiv 0$, we obtain for any $(\mathbf{y}, s) \in \Delta \times (0, \varepsilon)$ that

$$\mathbb{E}\left(\sup_{0\leq t\leq T\wedge\eta_{\tilde{k}}^{\mathbf{y},s}\wedge\eta_{\tilde{k}}^{\mathbf{y},0}} \left\| \left(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t)\right) - \left(\mathbf{Y}^{\mathbf{y},0}(t), 0\right) \right\|^2 \right) \leq Cs^2, \tag{A.18}$$

where C is a constant that depends on $H, T, \tilde{k}$. Applying Chebyshev's inequality to (A.18), there is a $\delta \in (0, \varepsilon)$ such that for all $(\mathbf{y}, s) \in \Delta \times (0, \delta)$

$$\mathbb{P}\left\{\sup_{0\le t\le T\wedge\eta_{\tilde{k}}^{\mathbf{y},s}\wedge\eta_{\tilde{k}}^{\mathbf{y},0}}\left\|\left(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t)\right)-\left(\mathbf{Y}^{\mathbf{y},0}(t),0\right)\right\|<\varepsilon\right\}>1-\frac{\varepsilon}{3}. \tag{A.19}$$

Combining (A.18) and (A.19) yields

$$\mathbb{P}\left\{\sup_{0\le t\le T}\left\|\left(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t)\right)-\left(\mathbf{Y}^{\mathbf{y},0}(t),0\right)\right\|<\varepsilon\right\}>1-\varepsilon. \tag{A.20}$$

for any $(\mathbf{y}, s) \in \Delta \times (0, \delta)$. The desired result is obtained by noting that $\mathbf{Y}^{\mathbf{y},0}(t) = \tilde{\mathbf{Y}}^{\mathbf{y}}(t)$, $t \ge 0$. □

Lemma A.3 *There are positive constants K_2 and K_3 such that for any $(\mathbf{y}, s) \in \Delta \times (0, \infty)$, $T \ge 0$, one has*

$$\mathbb{E}\left([\ln S^{\mathbf{y},s}(T)]^2\right) \le ((\ln s)^2 + 1)K_2 \exp\{K_3 T\}, \tag{A.21}$$

Proof In view of Itô's formula,

$$\begin{aligned} d\ln^2 S(t) = &\left(\mathbf{Y}(t)^\top \Sigma \mathbf{Y}(t) + 2\ln S(t)\left(\mathbf{a}^\top \mathbf{Y}(t) - [\mathbf{b}(S(t)\mathbf{Y}(t))]^\top\right.\right. \\ &\left.\left.\mathbf{Y}(t) - \frac{1}{2}\mathbf{Y}(t)^\top \Sigma \mathbf{Y}(t)\right)\right) dt \; + \; 2\ln S(t)\mathbf{Y}(t)^\top d\mathbf{E}(t). \end{aligned} \tag{A.22}$$

Now, we estimate $g(\mathbf{y}, s) = \mathbf{y}^\top \Sigma \mathbf{y} + 2\ln s\left(\mathbf{a}^\top \mathbf{y} - \mathbf{b}(s\mathbf{y}) - \frac{1}{2}\mathbf{y}^\top \Sigma \mathbf{y}\right)$ for $(\mathbf{y}, s) \in \Delta \times (0, \infty)$. Let M_b be as in (2.2). If $s > M_b$ then $\ln s > 0$ and $\left(\mathbf{a}^\top \mathbf{y} - \mathbf{b}(s\mathbf{y}) - \frac{1}{2}\mathbf{y}^\top \Sigma \mathbf{y}\right) < 0$. Letting

$$M_1 := \sup_{\{(\mathbf{y},s)\in\Delta\times(0,M_b]\}}\left\{\left|\left(\mathbf{a}^\top \mathbf{y} - \mathbf{b}(s\mathbf{y}) - \frac{1}{2}\mathbf{y}^\top \Sigma \mathbf{y}\right)\right|\right\} < \infty$$

and

$$\begin{aligned} \|\Sigma\| &:= \sup\{\mathbf{y}^\top \Sigma \mathbf{y} : \mathbf{y} \in \Delta\}, \\ g(\mathbf{y}, s) &\le \|\Sigma\| + M_1|\ln s| \le M_1 \ln^2 s + 2M_1 \\ &\quad + \|\Sigma\| \text{ for all } (\mathbf{y}, s) \in \Delta \times (0, \infty). \end{aligned}$$

With this estimate, we can apply Dynkin's formula to (A.22) and use standard arguments (e.g., Mao 1997, Theorem 2.4.1) to obtain

$$\mathbb{E}\left([\ln S^{\mathbf{y},s}(T)]^2 \mathbf{1}_A\right) \le K_2(\ln s)^2 \exp\{K_3 T\} \text{ for all } (\mathbf{y}, s) \in \Delta \times (0, \infty)$$

for some positive constants K_2 and K_3. □

Lemma A.4 *There is a positive constant K_4 such that for any $(\mathbf{y}, s) \in \Delta \times (0,1)$, and $A \in \mathcal{F}$,*

$$\mathbb{E}\left([\ln S^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s})]_-^2\right) \leq (\ln s)^2 + K_4\sqrt{\mathbb{P}(A)}(T+1)[\ln s]_- + K_4 T^2, \tag{A.23}$$

where $[\ln x]_- := \max\{0, -\ln x\}$, *and*

$$\zeta^{\mathbf{y},s} := \inf\{t \geq 0 : S^{\mathbf{y},s}(t) = 1\}. \tag{A.24}$$

Proof Let

$$M_2 = \sup_{\{(\mathbf{y},s)\in\Delta\times(0,1)\}} \left\{(-\mathbf{a}^\top \mathbf{y} + \frac{1}{2}\mathbf{y}^\top \Sigma \mathbf{y} + \mathbf{b}(s\mathbf{y})^\top \mathbf{y}\right\} < \infty.$$

Using Dynkin's formula,

$$\begin{aligned} -\ln S^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s}) &= -\ln s - M^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s}) \\ &\quad + \int_0^{T\wedge\zeta^{\mathbf{y},s}} \Big(-\mathbf{a}^\top \mathbf{Y}^{\mathbf{y},s}(t) + \mathbf{b}(S^{\mathbf{y},s}(t)\mathbf{Y}^{\mathbf{y},s}(t))^T \mathbf{Y}^{\mathbf{y},s}(t) \\ &\quad + \frac{1}{2}\mathbf{Y}^{\mathbf{y},s}(t)^\top \Sigma \mathbf{Y}^{\mathbf{y},s}(t)\Big) dt \\ &\leq [\ln s]_- + M_2 T + |M^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s})|, \end{aligned} \tag{A.25}$$

where

$$M^{\mathbf{y},s}(t) = \int_0^t \mathbf{Y}(t)^\top d\mathbf{E}(t) = \int_0^t \mathbf{Y}(t)^\top \Gamma d\mathbf{B}(t). \tag{A.26}$$

It follows from (A.25) that

$$\begin{aligned} [\ln S^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s})]_-^2 \mathbf{1}_A &\leq \left([\ln s]_- + M_2 T + |M^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s})|\right)^2 \mathbf{1}_A \\ &\leq \left([\ln s]_-^2 + 2(M_2 T + |M^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s})|)[\ln s]_-\right) \mathbf{1}_A \\ &\quad + \left(2(M_2 T)^2 + 2|M^{\mathbf{y},s}(T \wedge \zeta^{\mathbf{y},s})|^2\right) \mathbf{1}_A \end{aligned} \tag{A.27}$$

An application of Itô's isometry yields

$$\mathbb{E}[|M_{z,y}(T \wedge \zeta^{\mathbf{y},s})|^2 \mathbf{1}_A] \leq \mathbb{E}|M_{z,y}(T \wedge \zeta^{\mathbf{y},s})|^2 \leq \|\Sigma\| T. \tag{A.28}$$

By a straightforward use of Hölder's inequality and (A.28),

$$\begin{aligned}\mathbb{E}[|M_{z,y}(T\wedge\zeta^{\mathbf{y},s})|\mathbf{1}_A] \leq& \left(\mathbb{P}(A)\mathbb{E}|M_{z,y}(T\wedge\zeta^{\mathbf{y},s})|^2\right)^{1/2}\\ \leq& \sqrt{\mathbb{P}(A)}\sqrt{\|\Sigma\|T} \leq \sqrt{\mathbb{P}(A)\|\Sigma\|}(T+1).\end{aligned} \tag{A.29}$$

Taking expectation on both sides of (A.27) and using the estimates from (A.28) and (A.29), we have

$$\mathbb{E}\left[[\ln S^{\mathbf{y},s}(T\wedge\zeta^{\mathbf{y},s})]_-^2\mathbf{1}_A\right] \leq [\ln s]_-^2\mathbb{P}(A) + K_4(T+1)\sqrt{\mathbb{P}(A)}[\ln s]_- + K_4T^2,$$

for some positive constant K_4. □

Let M_3 be a positive constant such that

$$\left|\mathbf{a}^\top\mathbf{y} - \frac{1}{2}\mathbf{y}^\top\Sigma\mathbf{y} - \mathbf{a}^\top\mathbf{y}' - \frac{1}{2}\mathbf{y}'^\top\Sigma\mathbf{y}'\right| \leq M_3\|\mathbf{y}'-\mathbf{y}|,\ \mathbf{y},\mathbf{y}'\in\Delta. \tag{A.30}$$

From now on, we assume that $\varepsilon\in(0,1)$ is chosen small enough to satisfy the following

$$\begin{aligned}&(M_3+2)\,\varepsilon + \sup_{\{0\leq s\leq\varepsilon,\mathbf{y}\in\Delta\}}\{\mathbf{b}(s\mathbf{y})^\top\mathbf{y}\} < \frac{r}{4}\\ &-\frac{3r}{2}(1-3\varepsilon) + 2K_4\sqrt{\varepsilon} < -r\end{aligned} \tag{A.31}$$

Lemma A.5 *For ε satisfying (A.31), there is $\delta(\varepsilon)=\delta\in(0,1)$ and $T^*(\varepsilon)=T^*>1$ such that*

$$\mathbb{P}\left\{\ln s + \frac{3rT^*}{4} \leq \ln S^{\mathbf{y},s}(T^*) < 0\right\} \geq 1-3\varepsilon \tag{A.32}$$

for all $(\mathbf{y},s)\in\Delta\times(0,\delta)$.

Proof Since Δ is a petite set of $\{\tilde{\mathbf{Y}}(t): t\geq 0\}$, in view of Meyn and Tweedie (1993, Theorem 6.1), there are γ_1 and $\gamma_2>0$ such that

$$\|\tilde{P}(t,\mathbf{y},\cdot)-\nu^*\|_{TV} \leq \gamma_1\exp(-\gamma_2 t),\ \mathbf{y}\in\Delta, t\in[0,\infty). \tag{A.33}$$

where $\tilde{P}(t,\mathbf{y},\cdot)$ is the transition probability of $\{\tilde{\mathbf{Y}}(t): t\geq 0\}$. Let $M_4 = \max_{\mathbf{y}\in\Delta}\{|\mathbf{a}^\top\mathbf{y} - \frac{1}{2}\mathbf{y}^\top\Sigma\mathbf{y}|\} < \infty$. In view of (2.8) and (A.33), we have

$$\begin{aligned}&\frac{1}{T}\mathbb{E}\left|\int_0^T\left(\mathbf{a}^\top\tilde{\mathbf{Y}}^{\mathbf{y}}(t) - \frac{1}{2}\tilde{\mathbf{Y}}^{\mathbf{y}}(t)^\top\Sigma\tilde{\mathbf{Y}}^{\mathbf{y}}(t)\right)dt - rT\right|\\ &\leq \frac{1}{T}\int_0^T\int_\Delta\left|\left(\mathbf{a}^\top\mathbf{y}' - \frac{1}{2}\mathbf{y}'^\top\Sigma\mathbf{y}'\right)\left(\tilde{P}(t,\mathbf{y},d\mathbf{y}') - \nu^*(d\mathbf{y}')\right)\right|\\ &\leq \frac{M_4}{T}\int_0^T\|\tilde{P}(t,\mathbf{y},\cdot)-\nu^*\|_{TV}dt \leq \frac{M_4\gamma_1}{T}.\end{aligned} \tag{A.34}$$

On one hand, letting $M^{\mathbf{y},s}(T)$ be defined as (A.26), we have from Itô's isometry that

$$\mathbb{E}\left[\frac{M^{\mathbf{y},s}(T)}{T}\right]^2 = \frac{1}{T^2}\mathbb{E}\int_0^T \mathbf{Y}^{\mathbf{y},s}(t)^\top \Sigma \mathbf{Y}^{\mathbf{y},s}(t)dt \le \frac{\|\Sigma\|}{T}. \tag{A.35}$$

With standard estimation techniques, it follows from (A.34) and (A.35) that for any $\varepsilon > 0$, there is a $T^* = T^*(\varepsilon)$ such that

$$\mathbb{P}\left\{\left|\frac{1}{T^*}\int_0^{T^*}\left(\mathbf{a}^\top \tilde{\mathbf{Y}}^{\mathbf{y}}(t) - \frac{1}{2}\tilde{\mathbf{Y}}^{\mathbf{y}}(t)^\top \Sigma \tilde{\mathbf{Y}}^{\mathbf{y}}(t)\right)dt - r\right| < \varepsilon\right\} > 1-\varepsilon,\ \mathbf{y}\in\Delta, \tag{A.36}$$

and

$$\mathbb{P}\left\{\left|\frac{M^{\mathbf{y},s}(T^*)}{T^*}\right| < \varepsilon\right\} > 1-\varepsilon,\ (\mathbf{y},s)\in\Delta\times(0,\infty). \tag{A.37}$$

By virtue of Proposition A.2, (A.30), and (A.36), there is $\delta = \delta(\varepsilon, T^*) \in (0,\varepsilon)$ such that

$$\mathbb{P}(\Omega_1^{\mathbf{y},s}) > 1-2\varepsilon,\ (\mathbf{y},s)\in\Delta\times(0,\delta)$$

where

$$\begin{aligned}\Omega_1^{\mathbf{y},s} := &\left\{\int_0^{T^*}\left(\mathbf{a}^\top \mathbf{Y}^{\mathbf{y},s}(t) - \frac{1}{2}\mathbf{Y}^{\mathbf{y},s}(t)^\top \Sigma \mathbf{Y}^{\mathbf{y},s}(t)\right)dt > T^*(r-(M_3+1)\varepsilon)\right\}\\ &\cap\left\{S^{\mathbf{y},s}(t) < \varepsilon,\ t\in[0,T^*]\right\}.\end{aligned}$$

Using $\mathbf{y}^\top \mathbf{b}(s\mathbf{y}) < \frac{r}{4}$ for all $(\mathbf{y},s)\in\Delta\times(0,\varepsilon)$ from (A.31) we have that on the set $\Omega_2^{\mathbf{y},s} := \Omega_1^{\mathbf{y},s}\bigcap\left\{\left|\frac{M^{\mathbf{y},s}(T)}{T}\right| < \varepsilon\right\}$ the following holds

$$\begin{aligned}0 > \ln\varepsilon \ge \ln S^{\mathbf{y},s}(T^*) &= \ln s + M^{\mathbf{y},s}(T^*) - \int_0^{T^*}\mathbf{Y}^{\mathbf{y},s}(t)^\top \mathbf{b}(S^{\mathbf{y},s}(t)\mathbf{Y}^{\mathbf{y},s}(t))dt\\ &\quad + \int_0^{T^*}\left(\mathbf{a}^\top \mathbf{Y}^{\mathbf{y},s}(t) - \frac{1}{2}\mathbf{Y}^{\mathbf{y},s}(t)^\top \Sigma \mathbf{Y}^{\mathbf{y},s}(t)\right)dt\\ &\ge \ln s + \left(r-(M_3+2)\varepsilon - \sup_{\{0\le s\le\varepsilon,\mathbf{y}\in\Delta\}}\{\mathbf{b}(s\mathbf{y})^\top\mathbf{y}\}\right)T^*\\ &\ge \ln s + \frac{3r}{4}T^*.\end{aligned} \tag{A.38}$$

Noting

$$\mathbb{P}\left(\Omega_2^{\mathbf{y},s}\right) \ge 1-3\varepsilon \text{ for all } (\mathbf{y},s)\in\Delta\times(0,\delta),$$

the proof is complete. □

Proposition A.3 *Assume $r > 0$. Let δ and T^* be as in Lemma A.5. There exists a positive constant $K^* = K^*(\delta, T^*)$ such that*

$$\mathbb{E}[\ln S^{\mathbf{y},s}(T^*)]_-^2 \leq [\ln s]_-^2 - rT^*[\ln s]_- + K^* \tag{A.39}$$

for any $(\mathbf{y}, s) \in \Delta \times (0, \infty)$.

Proof We look at three cases of the initial data $(\mathbf{y}, s)$.

Case I $s \in (0, \delta)$. We have from Lemma A.5 that $\mathbb{P}(\Omega_2^{\mathbf{y},s}) \geq 1 - 3\varepsilon$ where $\Omega_2^{\mathbf{y},s}$ is defined as in the proof of Lemma A.5. On $\Omega_2^{\mathbf{y},s}$, we have

$$-\ln s - \frac{3rT^*}{4} \geq -\ln S^{\mathbf{y},s}(T^*) > 0.$$

Hence,

$$0 \leq [\ln S^{\mathbf{y},s}(T^*)]_- \leq [\ln s]_- - \frac{3rT^*}{4}.$$

Squaring both sides yields

$$[\ln S^{\mathbf{y},s}(T^*)]_-^2 \leq [\ln s]_-^2 - \frac{3rT^*}{2}[\ln s]_- + \frac{9r^2T^{*2}}{16},$$

which implies that

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\Omega_2^{\mathbf{y},s}}[\ln S^{\mathbf{y},s}(T^*)]_-^2\right] \leq \mathbb{P}(\Omega_2^{\mathbf{y},s})[\ln s]_-^2 - \frac{3rT^*}{2}\mathbb{P}(\Omega_2^{\mathbf{y},s})[\ln s]_- \\ + \frac{9r^2T^{*2}}{16}\mathbb{P}(\Omega_2^{\mathbf{y},s}).\end{aligned} \tag{A.40}$$

On $\Omega_3^{\mathbf{y},s} := \{\zeta^{\mathbf{y},s} < T^*\}$ with $\zeta^{\mathbf{y},s}$ defined in (A.24), since $\ln S^{\mathbf{y},s}(\zeta^{\mathbf{y},s}) = 0$, we have from Lemma A.3 and the strong Markov property of $(\mathbf{Y}(t), S(t))$ that

$$\mathbb{E}\left[\mathbf{1}_{\Omega_3^{\mathbf{y},s}}[\ln S^{\mathbf{y},s}(T^*)]_-^2\right] \leq \mathbb{P}(\Omega_3^{\mathbf{y},s})K_2\exp(K_3T^*). \tag{A.41}$$

On the set $\Omega_4^{\mathbf{y},s} := \Omega\backslash(\Omega_2^{\mathbf{y},s} \cup \Omega_3^{\mathbf{y},s})$, applying Lemma A.4 and noting that $\zeta^{\mathbf{y},s} > T^*$ in $\Omega_4^{\mathbf{y},s}$ and $T^* > 1$, we obtain

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\Omega_4^{\mathbf{y},s}}[\ln S^{\mathbf{y},s}(T^*)]_-^2\right] \leq [\ln s]_-^2\mathbb{P}(\Omega_4^{\mathbf{y},s}) \\ + 2K_4T^*\sqrt{\mathbb{P}(\Omega_4^{\mathbf{y},s})}[\ln s]_- + K_4T^{*2}.\end{aligned} \tag{A.42}$$

Adding (A.40), (A.41), and (A.42) side by side, we get

$$\begin{aligned}\mathbb{E}[\ln S^{\mathbf{y},s}(T^*)]_-^2 &\le [\ln s]_-^2 + \Big(-\frac{3r}{2}(1-3\varepsilon) + 2K_4\sqrt{\varepsilon}\Big)T^*[\ln s]_- + K_5^*(T^*)\\ &\le [\ln s]_-^2 - rT^*[\ln s]_- + K_5^*(T^*),\end{aligned} \tag{A.43}$$

where $K_5^*(T^*)$ is a positive constant independent of $(\mathbf{y}, s) \in \Delta \times (0, \delta)$.

Case II $s \in [\delta, 1]$. We have from Lemma A.3 that

$$\begin{aligned}\mathbb{E}[\ln S^{\mathbf{y},s}(T^*)]_-^2 &\le \mathbb{E}[\ln S^{\mathbf{y},s}(T^*)]^2 \le [\ln s]^2 + K_2\exp(K_3T^*)\\ &\le ([\ln\delta]^2 + 1)K_2\exp(K_3T^*).\end{aligned} \tag{A.44}$$

Case III $s \in (1, \infty)$. Note that if $\zeta^{\mathbf{y},s} > T^*$, then $[\ln S^{\mathbf{y},s}(T^*)]_-^2 = 0$. Thus, using Lemma A.3 and the strong Markov property of $(\mathbf{Y}(t), S(t))$ once more, we obtain

$$\mathbb{E}[\ln S^{\mathbf{y},s}(T^*)]_-^2 = \mathbb{E}\left(\mathbf{1}_{\{\zeta^{\mathbf{y},s}<T^*\}}[\ln S^{\mathbf{y},s}(T^*)]_-^2\right) \le K_2\exp(K_3T^*). \tag{A.45}$$

Combing (A.43), (A.44), and (A.45), and setting $K^* = \max\{K_5^*(T^*), ([\ln\delta]^2 + 1)K_2\exp(K_3T^*)\}$, the proof is concluded. □

Theorem A.2 *Suppose that Assumptions 2.2 and 2.3 hold and that $r > 0$. Let $P(t, (\mathbf{y}, s), \cdot)$ be the semigroup of the process $((\mathbf{Y}(t), S(t))_{t\ge0}$. Then, there exists an invariant probability measure π^* of the process $((\mathbf{Y}(t), S(t))_{t\ge0}$ on $\Delta \times (0, \infty)$. Moreover, $\pi^*(\Delta^\circ \times (0, \infty)) = 1$, π^* is absolutely continuous with respect to the Lebesgue measure on $\Delta \times (0, \infty)$ and*

$$\lim_{t\to\infty} t^{q^*}\|P(t, (\mathbf{y}, s), \cdot) - \pi^*(\cdot)\|_{TV} = 0,\ (\mathbf{y}, s) \in \Delta^\circ \times (0, \infty), \tag{A.46}$$

where $\|\cdot\|_{TV}$ is the total variation norm and q^* is any positive number. In addition, for any initial value $(\mathbf{y}, s) \in \Delta \times (0, \infty)$ and any π^*-integrable function f, we have

$$\begin{aligned}&\mathbb{P}\left\{\lim_{T\to\infty}\frac{1}{T}\int_0^T f\left(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t)\right)dt\right.\\ &\left.= \int_{\Delta^\circ\times(0,\infty)} f(\mathbf{y}', s')\pi^*(d\mathbf{y}', s')\right\} = 1.\end{aligned} \tag{A.47}$$

Proof By virtue of Lemma A.2, there is an $h_1 := 1 - \exp(-\gamma_b T^*) > 0$ satisfying

$$\begin{aligned}\mathbb{E}S^{\mathbf{y},s} + 1 &\le s + 1 - h_1 s + K_1 \le s + 1 - h_1\sqrt{s+1} + K_1\\ &\quad + h_1,\ (\mathbf{y}, s) \in \Delta \times (0, \infty).\end{aligned} \tag{A.48}$$

Let $V(s) = s + 1 + [\ln s]_-^2$. In view of Proposition A.3 and (A.48),

$$\begin{aligned}\mathbb{E}V\left(S^{\mathbf{y},s}(T^*)\right) &\leq s + 1 - h_2\left(\sqrt{s+1} + [\ln s]_-\right) + H_2 \\ &\leq V(s) - \frac{h_2}{2}\sqrt{V(s)} + H_2 \text{ for all } (\mathbf{y}, s) \in \Delta \times (0, \infty),\end{aligned} \tag{A.49}$$

where $h_2 = \min\{h_1, rT^*\}$, $H_2 = H_1 + h_1 + K_1$. Let $\kappa > 1$ such that

$$\frac{\sqrt{V(s)}}{2} \geq H_2 \text{ for all } s \notin [\kappa^{-1}, \kappa]. \tag{A.50}$$

Combining (A.49) and (A.50), we arrive at

$$\begin{aligned}\mathbb{E}V\left(S^{\mathbf{y},s}(T^*)\right) \leq V(s) - \frac{h_2}{4}\sqrt{V(s)} + H_2\mathbf{1}_{\{(\mathbf{y},s)\in\Delta\times[\kappa^{-1},\kappa]\}} \\ \text{for all } (\mathbf{y}, s) \in \Delta \times (0, \infty).\end{aligned} \tag{A.51}$$

Using the estimate (A.51), Lemma A.1, and Theorem A.1, the Markov chain $(\mathbf{Y}(kT^*), S(kT^*))_{k\geq 0}$ has a unique invariant probability measure π^* and

$$k\|P(kT^*, (\mathbf{y}, s), \cdot) - \pi^*\|_{TV} \to 0 \text{ as } k \to \infty. \tag{A.52}$$

As a direct consequence, for fixed $\mathbf{y}_0, s_0$, the family $\{P(kT^*, (\mathbf{y}_0, s_0), \cdot), k \in \mathbb{N}\}$ is tight, that is, for any $\theta > 0$, there is a compact set $K_\theta \subset \Delta \times (0, \infty)$ such that

$$P(kT^*, (\mathbf{y}_0, s_0), K_\theta) > 1 - \theta \text{ for all } k \in \mathbb{N}. \tag{A.53}$$

Since $s^2 + \ln^2 s \to \infty$ as $s \to 0$ or $s \to \infty$, in view of Lemmas A.2 and A.3 and a standard estimate, there is a $\kappa_\theta > 1$ such that

$$\mathbb{P}\left\{S^{\mathbf{y},s}(t) \in [\kappa_\theta^{-1}, \kappa_\theta]\right\} > 1 - \theta, \text{ for all } (\mathbf{y}, s) \in K_\theta, t \in [0, T^*],$$

or equivalently,

$$P\left(t, (\mathbf{y}, s), \Delta \times [\kappa_\theta^{-1}, \kappa_\theta]\right) > 1 - \theta \text{ for all } (\mathbf{y}, s) \in K_\theta, t \in [0, T^*]. \tag{A.54}$$

Using the Chapman-Kolmogorov relation together with (A.53) and (A.54) yields

$$P\left(u, (\mathbf{y}_0, s_0), \Delta \times [\kappa_\theta^{-1}, \kappa_\theta]\right) > 1 - 2\theta \text{ for all } u \geq 0,$$

which implies that the family of empirical measures $\left\{\frac{1}{T}\int_0^T P(u, (\mathbf{y}_0, s_0), \cdot)du, T > 0\right\}$ is tight in $\Delta \times (0, \infty)$. Thus $(\mathbf{Y}(t), S(t))$ has an invariant probability measure π_* on $\Delta \times (0, \infty)$ (see e.g., Evans et al. 2015, Proposition 6.4). As a result, the Markov chain $(\mathbf{Y}(kT^*), S(kT^*))_{k\in\mathbb{N}}$ has an invariant probability measure π_*. In view of (A.52), π_*

must coincide with π^*. Thus, π^* is an invariant probability measure of the process $(\mathbf{Y}(t), S(t))_{t\geq 0}$ on $\Delta \times (0,\infty)$.

In the proofs, we used the function $[\ln s]_-^2$ for the sake of simplicity. In fact, we can treat $[\ln s]_-^{1+q}$ for any small $q \in (0,1)$ in the same manner. We can show that there are $h_q, H_q > 0$, and a compact set $K_q \subset \Delta \times (0,\infty)$ satisfying

$$\begin{aligned}\mathbb{E}V_q(S^{\mathbf{y},s}(T^*)) \leq V_q(s) - h_q[V_q(s)]^{\frac{1}{1+q}} + H_q\mathbf{1}_{\{(\mathbf{y},s)\in K_q\}},\\ (\mathbf{y},s) \in \Delta \times (0,\infty),\end{aligned} \tag{A.55}$$

where $V_q(s) := s + 1 + [\ln s]_-^{1+q}$. Then applying Theorem A.1, we obtain

$$k^{1/q}\|P(kT^*, (\mathbf{y},s),\cdot) - \pi^*\| \to 0 \text{ as } k \to \infty. \tag{A.56}$$

Let $f : \Delta \times (0,\infty) \mapsto [-1,1]$ be a measurable function. Since π^* is an invariant measure, then for any $u \geq 0$,

$$\begin{aligned}&\int_{\Delta\times(0,\infty)} f(\mathbf{y}',s')\pi^*(d\mathbf{y}',ds')\\ &\quad= \int_{\Delta\times(0,\infty)} \pi^*(d\mathbf{y}_1,ds_1)\int_{\Delta\times(0,\infty)} P(u,(\mathbf{y}_1,s_1),(\mathbf{y}',ds'))f(\mathbf{y}',s').\end{aligned}$$

Using this equality and the Chapman–Kolmogorov equation, we have

$$\begin{aligned}&|f(\mathbf{y}',s')(P(t+u,(\mathbf{y},s),d\mathbf{y}',ds') - \pi^*(d\mathbf{y}',ds')|\\ &\quad= \Big|\int_{\Delta\times(0,\infty)} \big(P(t,(\mathbf{y},s),d\mathbf{y}_1,ds_1) - \pi^*(d\mathbf{y}_1,ds_1)\big)\\ &\qquad\times \int_{\Delta\times(0,\infty)} (f(\mathbf{y}',s')P(t,(\mathbf{y}_1,s_1),(d\mathbf{y}',ds'))\Big|\\ &\quad\leq \|P(t,(\mathbf{y},s),\cdot) - \pi^*\|_{TV}\\ &\qquad\left(\text{since } \Big|\int_{\Delta\times(0,\infty)} (f(\mathbf{y}',s')P(t,(\mathbf{y}_1,s_1),(d\mathbf{y}',ds')\Big| \leq 1 \text{ for all } \mathbf{y}_1,s_1\right),\end{aligned}$$

which means that $\|P(t,(\mathbf{y},s),\cdot) - \pi^*\|_{TV}$ is decreasing in t. As a result, we deduce from (A.55) that

$$t^{q^*}\|P(t,(\mathbf{y},s),\cdot) - \pi^*\|_{TV} \to 0 \text{ as } t \to \infty,$$

where $q^* = 1/q \in (1,\infty)$.

In view of Proposition A.1, for any $t > 0$, $\mathbb{P}\{\mathbf{Y}^{\mathbf{y},s}(t) \in \Delta^\circ\} = 1$. Thus,

$$\pi^*(\Delta^\circ \times (0,\infty)) = \int_{\Delta\times(0,\infty)} \mathbb{P}\{\mathbf{Y}^{\mathbf{y},s}(t) \in \Delta^\circ\}\pi^*(d\mathbf{y},ds) = \pi^*(\Delta\times(0,\infty)) = 1.$$

By Kallenberg (2002, Theorem 20.17), our process $(\mathbf{Y}(t), S(t))_{t\geq 0}$ is either Harris recurrent or uniformly transient on $\Delta^\circ \times (0,\infty)$. Using Kallenberg (2002, Theorem 20.21), our process cannot be uniformly transient and also have an invariant probability measure. Therefore, our process is Harris recurrent. Kallenberg (2002, Theorem 20.17) further indicates that any Harris recurrent Feller process on $\Delta^\circ \times (0,\infty)$ with strictly positive transition densities has a locally finite invariant measure that is equivalent to Lebesgue measure and is unique up to normalization. Since we already know that $(\mathbf{Y}(t), S(t))_{t\geq 0}$ has a unique invariant probability measure, this probability measure has an almost everywhere strictly positive density with respect to the Lebesgue measure. □

Appendix B: The case $r < 0$

Theorem B.1 *Suppose that $r < 0$. Then for any $i = 1,\dots,n$ and any $\mathbf{x} = (x_1,\dots,x_n) \in \mathbb{R}^n_+$,*

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{\ln X_i^{\mathbf{x}}(t)}{t} = r\right\} = 1. \tag{B.1}$$

In particular, for any $i = 1,\dots,n$ and any $\mathbf{x} = (x_1,\dots,x_n) \in \mathbb{R}^n_+$

$$\mathbb{P}\left\{\lim_{t\to\infty} X_i^{\mathbf{x}}(t) = 0\right\} = 1.$$

Proof Let $\theta > 0$ and $\check{a}_i = a_i + \theta$, and define the process $\check{\mathbf{X}}^{\mathbf{x}}(t) = (\check{X}_1^{\mathbf{x}}(t),\dots,\check{X}_n^{\mathbf{x}}(t))$ as the solution to

$$d\check{X}_i(t) = \left(\check{X}_i(t)(\check{a}_i) + \sum_{j=1}^n D_{ji}\check{X}_j(t)\right)dt + \check{X}_i(t)dE_i(t),\ i = 1,\dots,n \tag{B.2}$$

started at $\mathbf{x} = (x_1,\dots,x_n) \in \mathbb{R}^n_+$. Letting $\check{S}(t) = \sum \check{X}_i(t)$ and $\check{\mathbf{Y}}(t) = \dfrac{\check{\mathbf{X}}(t)}{S(t)}$, we have

$$\begin{aligned} d\check{\mathbf{Y}}(t) &= \left(\operatorname{diag}(\check{\mathbf{Y}}(t)) - \check{\mathbf{Y}}(t)\check{\mathbf{Y}}^\top(t)\right)\Gamma^\top d\mathbf{B}(t) \\ &\quad + \mathbf{D}^\top\check{\mathbf{Y}}(t)dt + \left(\operatorname{diag}(\check{\mathbf{Y}}(t)) - \check{\mathbf{Y}}(t)\check{\mathbf{Y}}^\top(t)\right)(\check{\mathbf{a}} - \Sigma\check{\mathbf{Y}}(t))dt \\ d\ln\check{S}(t) &= \left(\check{\mathbf{a}}^\top\check{\mathbf{Y}}(t) - \frac{1}{2}\check{\mathbf{Y}}(t)^\top\Sigma\check{\mathbf{Y}}(t)\right)dt + \check{\mathbf{Y}}(t)^\top\Gamma^\top d\mathbf{B}(t) \end{aligned} \tag{B.3}$$

Let $(\check{\mathbf{Y}}^{\mathbf{y}}(t), \check{S}^{\mathbf{y},s}(t))$ be the solution to (B.3) with initial condition $(\mathbf{y}, s)$. Note that $\check{\mathbf{Y}}^{\mathbf{y}}(t)$ does not depend on s. First, fix $\mathbf{y}_0 \in \Delta$. We have that

$$\lim_{t\to\infty} \frac{1}{t}\left(\int_0^t \left(\check{\mathbf{a}}^\top \check{\mathbf{Y}}^{\mathbf{y}_0}(u) - \frac{1}{2}\check{\mathbf{Y}}^{\mathbf{y}_0}(u)^\top \Sigma \check{\mathbf{Y}}^{\mathbf{y}_0}(u)\right) du + \int_0^t \check{\mathbf{Y}}^{\mathbf{y}_0}(u)^\top \Gamma^\top d\mathbf{B}(u)\right)$$
$$= \check{r} := \int_\Delta \left(\check{\mathbf{a}}^\top \mathbf{y} - \frac{1}{2}\mathbf{y}^\top \Sigma \mathbf{y}\right) \check{\nu}^*(d\mathbf{y}), \mathbb{P} - \text{a.s.}, \tag{B.4}$$

where $\check{\nu}^*$ is the unique invariant probability measure of $(\check{\mathbf{Y}}(t))_{t\geq 0}$. By the continuous dependence of r on the coefficients (established in the Proposition D.1), there is $\theta > 0$ such that $\check{r} < \frac{r}{2} < 0$. Let $\delta > 0$ such that $\sup\{-b_i(x) : x < \delta, i = 1, \ldots, n\} < \theta$ (this is possible since the b_i's are continuous and vanish at 0). Because $\check{r} < 0$, it follows from (B.4) that

$$\sup_{t\in[0,\infty)} \left(\int_0^t \left(\check{\mathbf{a}}^\top \check{\mathbf{Y}}^{\mathbf{y}_0}(s) - \frac{1}{2}\check{\mathbf{Y}}^{\mathbf{y}_0}(s)^\top \Sigma \check{\mathbf{Y}}^{\mathbf{y}_0}(s)\right) ds + \int_0^t \check{\mathbf{Y}}^{\mathbf{y}_0}(s)^\top \Gamma^\top d\mathbf{B}(s)\right) < \infty \ \mathbb{P} - \text{a.s.}.$$

As a result, for any $\varepsilon > 0$, there is an $H_\varepsilon > 0$ satisfying

$$\mathbb{P}\left\{\sup_{t\in[0,\infty)} \left(\int_0^t \left(\check{\mathbf{a}}^\top \check{\mathbf{Y}}^{\mathbf{y}_0}(u) - \frac{1}{2}\check{\mathbf{Y}}^{\mathbf{y}_0}(u)^\top \Sigma \check{\mathbf{Y}}^{\mathbf{y}_0}(u)\right) du + \int_0^t \check{\mathbf{Y}}^{\mathbf{y}_0}(u)^\top \Gamma^\top d\mathbf{B}(u)\right) < H_\varepsilon\right\} > 1 - \varepsilon,$$

which combined with (B.3) implies that

$$\mathbb{P}\left\{\sup_{t\in[0,\infty)} \check{S}^{\mathbf{y}_0,s_0}(t) < \delta\right\} > 1 - \varepsilon \text{ if } s_0 < \delta \exp(-H_\varepsilon). \tag{B.5}$$

Then, a comparison argument shows (see Remark A.2) that for $\mathbf{x}_0 = s_0\mathbf{y}_0 \in \mathbb{R}^n_+$ and $i = 1, \ldots, n$

$$\mathbb{P}\left\{X_i^{\mathbf{x}_0}(t) \leq \check{X}_i^{\mathbf{x}_0}(t),\ t \in [0, \xi^{\mathbf{x}_0})\right\} = 1 \tag{B.6}$$

where $\xi^{\mathbf{x}_0} = \inf\{t \geq 0 : \sum_{i=1}^n \check{X}_i^{\mathbf{x}_0}(t) \geq \delta\}$. By virtue of (B.5), $\mathbb{P}\{\xi^{\mathbf{x}_0} = \infty\} > 1 - \varepsilon$ if $s_0 < \delta \exp(-H_\varepsilon)$. Using (B.4) and (B.6) yields that

$$\mathbb{P}\left\{\limsup_{t\to\infty} \frac{\ln S^{\mathbf{y}_0,s_0}}{t} \leq \check{r} < 0\right\} > 1 - \varepsilon \text{ if } s < \delta \exp(-H_\varepsilon). \tag{B.7}$$

Thus, the process $(\mathbf{Y}(t), S(t))_{t\geq 0}$ is not a recurrent diffusion process in $\Delta \times (0,\infty)$. Hence, it must be transient with probability 1, that is, for any compact $K \in (0,\infty)$ and any initial value $(\mathbf{y}, s) \in \Delta \times (0,\infty)$ we have

$$\mathbb{P}\left\{\lim_{t\to\infty} \mathbf{1}_{\{S^{\mathbf{y},s}(t)\in K\}} = 0\right\} = 1. \tag{B.8}$$

In view of Lemma A.2,

$$\mathbb{P}\left\{\lim_{t\to\infty} S^{\mathbf{y},s}(t) = \infty\right\} = 0. \tag{B.9}$$

It follows from (B.8) and (B.9) that $\mathbb{P}\{\lim_{t\to\infty} S^{\mathbf{y},s}(t) = 0\} = 1$ for any $(\mathbf{y}, s) \in \Delta \times (0,\infty)$. Moreover, since $(\tilde{\mathbf{Y}}(t))_{\{t\geq 0\}}$ has a unique invariant probability measure ν^*, on the boundary $\Delta \times \{0\}$, $(\mathbf{Y}(t), S(t))$ has a unique invariant probability measure $\nu^* \times \delta_0^*$, where δ_0^* is the Dirac measure concentrated on $\{0\}$. Fix $(\mathbf{y}, s) \in \Delta \times (0,\infty)$, and define the normalized occupation measures,

$$\Pi_t(\cdot) = \frac{1}{t}\int_0^t \mathbf{1}_{\{(\mathbf{Y}^{\mathbf{y},s}(u), S^{\mathbf{y},s}(u))\in\cdot\}} du.$$

Since $\mathbb{P}\left\{\lim_{t\to\infty} S^{\mathbf{y},s}(t) = 0\right\} = 1$, the family $\{\Pi_k(\cdot), k \in \mathbb{N}\}$ is tight in the space $\Delta \times [0,\infty)$ for almost every ω. In view of the proofs of Evans et al. (2015, Theorem 4.2) or Schreiber et al. (2011, Theorems 4, 5) the set of weak* limit points of $\{\Pi_k, k \in \mathbb{N}\}$ is a nonempty set of invariant probability measures of the process $(\mathbf{Y}(t), S(t))$. As pointed out above, the process $(\mathbf{Y}(t), S(t))$ has only one invariant probability measure, namely, $\nu^* \times \delta_0^*$. Thus, for almost every $\omega \in \Omega$, $\{\Pi_k(\cdot), k \in \mathbb{N}\}$ converges weakly to $\nu^* \times \delta_0^*$ as $k \to \infty$. As a result, for any bounded continuous function $g(\cdot,\cdot) : \Delta \times [0,\infty) \mapsto \mathbb{R}$ we have $\lim_{k\to\infty} \frac{1}{k}\int_0^k g(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t))dt = \int_\Delta g(\mathbf{y}', 0)\nu^*(d\mathbf{y}')$ $\mathbb{P}$-a.s. Since $g(\cdot,\cdot)$ is bounded, we easily obtain

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T g(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t))dt = \int_\Delta g(\mathbf{y}', 0)\nu^*(d\mathbf{y}') \ \mathbb{P}\text{-a.s.} \tag{B.10}$$

Consequently,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \left(\mathbf{a}^\top \mathbf{Y}^{\mathbf{y},s}(t) - \frac{1}{2}\mathbf{Y}^{\mathbf{y},s}(t)^\top \Sigma \mathbf{Y}^{\mathbf{y},s}(t)\right) dt = r \ \mathbb{P}\text{-a.s} \tag{B.11}$$

Since $\mathbb{P}\{\lim_{t\to\infty} S^{\mathbf{y},s}(t) = 0\} = 1$ and $b_i(0) = 0, i = 1,\ldots,n$, we have by Dominated Convergence that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{Y}^{\mathbf{y},s}(t)^\top \mathbf{b}(S^{\mathbf{y},s}(t)\mathbf{Y}^{\mathbf{y},s}(t))dt = 0 \ \mathbb{P}\text{-a.s.} \tag{B.12}$$

Applying the strong law of large numbers for martingales to the process $(M^{\mathbf{y},s}(t))_{t\geq 0}$ defined by (A.26), we deduce

$$\lim_{T\to\infty}\frac{M^{\mathbf{y},s}(T)}{T}=0\ \mathbb{P}\text{-a.s.} \tag{B.13}$$

Note that

$$\begin{aligned}\frac{\ln S^{\mathbf{y},s}(T)}{T}=&\frac{\ln s}{T}+\frac{M^{\mathbf{y},s}(T)}{T}-\frac{1}{T}\int_0^T\mathbf{Y}^{\mathbf{y},s}(t)^\top\mathbf{b}(S^{\mathbf{y},s}(t)\mathbf{Y}^{\mathbf{y},s}(t))dt\\&+\frac{1}{T}\int_0^T\left(\mathbf{a}^\top\mathbf{Y}^{\mathbf{y},s}(t)-\frac{1}{2}\mathbf{Y}^{\mathbf{y},s}(t)^\top\Sigma\mathbf{Y}^{\mathbf{y},s}(t)\right)dt\end{aligned} \tag{B.14}$$

Applying (B.11), (B.12), and (B.13) to (B.14), we obtain

$$\lim_{T\to\infty}\frac{\ln S^{\mathbf{y},s}(T)}{T}=r,\ \mathbb{P}\text{-a.s.} \tag{B.15}$$

In light of (B.15), to derive $\mathbb{P}\left\{\lim_{T\to\infty}\frac{\ln X_i^{\mathbf{x}}(T)}{T}=r\right\}=1$, it suffices to show $\mathbb{P}\left\{\lim_{T\to\infty}\frac{\ln Y_i^{\mathbf{y},s}(T)}{T}=0\right\}=1$ for each $i=1,\dots,n$. In view of Itô's lemma,

$$\begin{aligned}\frac{\ln Y_i^{\mathbf{y},s}(T)}{T}=&\frac{\ln y_i}{T}+\frac{1}{T}\int_0^T\left(a_i-\sum_{j=1}^n a_jY_j^{\mathbf{y},s}(t)-D_{ii}-\frac{\sigma_{ii}}{2}\right.\\&\left.+\sum_{j,k=1}^n\frac{\sigma_{kj}}{2}Y_k^{\mathbf{y},s}(t)Y_j^{\mathbf{y},s}(t)\right)dt\\&+\frac{1}{T}\int_0^T\left(-b_i(S^{\mathbf{y},s}(t)Y_i^{\mathbf{y},s}(t))+\sum_{j=1}^n Y_j^{\mathbf{y},s}(t)b_j(S^{\mathbf{y},s}(t)Y_j^{\mathbf{y},s}(t))\right)dt\\&+\frac{1}{T}\int_0^T\left(\sum_{j=1,j\neq i}^n D_{ji}\frac{Y_j^{\mathbf{y},s}(t)}{Y_i^{\mathbf{y},s}(t)}\right)dt\\&+\frac{1}{T}\int_0^T\left[dE_i(t)-\sum_{j=1}^n Y_j^{\mathbf{y},s}(t)dE_j(t)\right],\end{aligned} \tag{B.16}$$

and

$$\begin{aligned}\frac{\ln\tilde{Y}_i^{\mathbf{y}}(T)}{T}=&\frac{\ln y_i}{T}+\frac{1}{T}\int_0^T\left(a_i-\sum_{j=1}^n a_j\tilde{Y}_j^{\mathbf{y}}(t)-D_{ii}\right.\\&\left.-\frac{\sigma_{ii}}{2}+\sum_{j,k=1}^n\frac{\sigma_{kj}}{2}\tilde{Y}_k^{\mathbf{y}}(t)\tilde{Y}_j^{\mathbf{y}}(t)\right)dt\end{aligned}$$

$$+\frac{1}{T}\int_0^T\left(\sum_{j=1,j\neq i}^n D_{ji}\frac{\tilde{Y}_j^{\mathbf{y}}(t)}{\tilde{Y}_i^{\mathbf{y}}(t)}\right)dt$$
$$+\frac{1}{T}\int_0^T\left[dE_i(t)-\sum_{j=1}^n\tilde{Y}_j^{\mathbf{y}}(t)dE_j(t)\right]. \tag{B.17}$$

By the strong laws of large numbers for martingales,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\left[dE_i(t)-\sum_{j=1}^n\tilde{Y}_j^{\mathbf{y}}(t)dE_j(t)\right]=0,\ \mathbb{P}\text{-a.s.} \tag{B.18}$$

Let $G_i=\sup_{\mathbf{y}\in\Delta}\left\{\left|a_i-\sum_{j=1}^n a_jy_j-D_{ii}-\frac{\sigma_{ii}}{2}+\sum_{j,k=1}^n\frac{\sigma_{kj}}{2}y_ky_j\right|\right\}<\infty$. As a result of (B.17) and (B.18) and the fact that $\limsup_{T\to\infty}\frac{\ln\tilde{Y}_i^{\mathbf{y}}(T)}{T}\leq 0$ almost surely, we obtain

$$\limsup_{T\to\infty}\frac{1}{T}\int_0^T\left(\sum_{j=1,j\neq i}^n D_{ji}\frac{\tilde{Y}_j^{\mathbf{y}}(t)}{\tilde{Y}_i^{\mathbf{y}}(t)}\right)dt\leq G_i,\ \mathbb{P}\text{-a.s.} \tag{B.19}$$

For any $k>0$, it follows from (B.18) and the strong law of large numbers that

$$\int_\Delta k\wedge\left(\sum_{j=1,j\neq i}^n D_{ji}\frac{y_j}{y_i}\right)\nu^*(d\mathbf{y})=\lim_{T\to\infty}\frac{1}{T}\int_0^T k\wedge\left(\sum_{j=1,j\neq i}^n D_{ji}\frac{\tilde{Y}_j^{\mathbf{y}}(t)}{\tilde{Y}_i^{\mathbf{y}}(t)}\right)dt$$
$$\leq G_i$$

Letting $k\to\infty$ we have

$$\rho_i:=\int_\Delta\sum_{j=1,j\neq i}^n D_{ji}\frac{y_j}{y_i}\nu^*(d\mathbf{y})\leq G_i,$$

which implies

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\left(\sum_{j=1,j\neq i}^n D_{ji}\frac{\tilde{Y}_j^{\mathbf{y}}(t)}{\tilde{Y}_i^{\mathbf{y}}(t)}\right)dt=\rho_i. \tag{B.20}$$

Using (B.18), (B.20), and applying the strong law of large numbers for the process $(\tilde{\mathbf{Y}}(t))_{t\geq 0}$, we arrive at

$$\lim_{T\to\infty}\frac{\ln\tilde{Y}_i^{\mathbf{y}}(T)}{T}=\beta_i+\rho_i\leq 0,\ \mathbb{P}\text{-a.s.}, \tag{B.21}$$

where

$$\beta_i := \int_\Delta \left(a_i - \sum_{j=1}^n a_j y_j - D_{ii} - \frac{\sigma_{ii}}{2} + \sum_{j,k=1}^n \frac{\sigma_{kj}}{2} y_k y_j \right) \nu^*(d\mathbf{y}).$$

If $\beta_i + \rho_i < 0$, then $\tilde{Y}_i^{\mathbf{y}}(T) \to 0$ almost surely as $T \to \infty$, which contradicts the fact that $\tilde{\mathbf{Y}}^{\mathbf{y}}(T)$ converges weakly to ν^* that is concentrated on Δ°. As a result, $\beta_i + \rho_i = 0$. For any $\theta > 0$, there is $k_\theta > 0$ such that

$$\int_\Delta k_\theta \wedge \left(\sum_{j=1, j\neq i}^n D_{ji} \frac{y_j}{y_i} \right) \nu^*(d\mathbf{y}) > \rho_i - \theta.$$

Using (B.10), we have with probability 1 that

$$\begin{aligned} &\liminf_{T\to\infty} \frac{1}{T} \int_0^T \left(\sum_{j=1, j\neq i}^n D_{ji} \frac{Y_j^{\mathbf{y},s}(t)}{Y_i^{\mathbf{y},s}(t)} \right) dt \\ &\geq \lim_{T\to\infty} \frac{1}{T} \int_0^T k_\theta \wedge \left(\sum_{j=1, j\neq i}^n D_{ji} \frac{Y_j^{\mathbf{y},s}(t)}{Y_i^{\mathbf{y},s}(t)} \right) dt \\ &\geq \rho_i - \theta. \end{aligned} \tag{B.22}$$

and

$$\begin{aligned} &\lim_{T\to\infty} \frac{1}{T} \int_0^T \left(a_i - \sum_{j=1}^n a_j Y_j^{\mathbf{y},s}(t) - D_{ii} - \frac{\sigma_{ii}}{2} \right. \\ &\left. + \sum_{j,k=1}^n \frac{\sigma_{kj}}{2} Y_k^{\mathbf{y},s}(t) Y_j^{\mathbf{y},s}(t)) \right) dt = \beta_i. \end{aligned} \tag{B.23}$$

Applying (B.22), (B.23), and the fact $\mathbb{P}\left\{ \lim_{T\to\infty} S^{\mathbf{y},s}(T) = 0 \right\} = 1$ to (B.16), we obtain that

$$\liminf_{T\to\infty} \frac{\ln Y_i^{\mathbf{y},s}(T)}{T} \geq \beta_i + \rho_i - \theta = -\theta, \ \mathbb{P}\text{-a.s.}$$

Since it holds for any $\theta > 0$, we have

$$\lim_{T\to\infty} \frac{\ln Y_i^{\mathbf{y},s}(T)}{T} = 0, \ \mathbb{P}\text{-a.s.}$$

The above equality combined with (B.15) and $X_i(T) = Y_i(T)S(T)$ yield the desired result. □

Appendix C: Degenerate diffusion in $\mathbb{R}^n$

If the correlation matrix Σ is degenerate, the diffusion $\tilde{\mathbf{Y}}(t)$ from (2.6) still has an invariant probability measure ν^* since it is a Feller-Markov process in a compact set. Moreover, $\nu^*(\Delta^\circ) = 1$ because the property that $\mathbb{P}\left\{\tilde{\mathbf{Y}}(t) \in \Delta^\circ,\ t > 0\right\} = 1$ is satisfied as long as Assumption 2.2 holds, that is, the dispersion matrix (D_{ij}) is irreducible. It is readily seen that the following is true.

Theorem C.1 *Assume that $\tilde{\mathbf{Y}}(t)$ has a **unique** invariant probability measure ν^*. Define r by (2.8). Suppose that $r < 0$. Then for any $i = 1, \ldots, n$ and any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_+$*

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{\ln X_i^{\mathbf{x}}(t)}{t} = r\right\} = 1. \tag{C.1}$$

In particular, for any $i = 1, \ldots, n$ and any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_+$

$$\mathbb{P}\left\{\lim_{t\to\infty} X_i^{\mathbf{x}}(t) = 0\right\} = 1.$$

Remark C.1 The Markov process $\{\tilde{\mathbf{Y}}(t), t \geq 0\}$ has a unique invariant probability measure if it is irreducible. Moreover, since $\mathbb{P}\{\tilde{\mathbf{Y}}^{\mathbf{y}}(t) > 0 \text{ for all } t > 0\} = 1$ for any $\mathbf{y} \in \Delta$, we need only check its irreducibility in Δ°. To prove that the diffusion $\{\tilde{\mathbf{Y}}(t), t \geq 0\}$ is irreducible in Δ°, we pursue the following approach:

- First, we show that the process $\{\tilde{\mathbf{Y}}(t), t \geq 0\}$ verifies Hörmander's condition. As a result, the process $\{\tilde{\mathbf{Y}}(t), t \geq 0\}$ has a smooth density function for any $t > 0$; see e.g., Rey-Bellet (2006).
- Next, we show that there is an open set $N \subset \Delta^\circ$ such that for any open set $N_0 \subset N$, and $\mathbf{y} \in \Delta^\circ$, there is a $t_0 > 0$ such that $\mathbb{P}\{\tilde{\mathbf{Y}}^{\mathbf{y}}(t_0) \in N_0\} > 0$. This claim is usually proved by analyzing the control systems corresponding to the diffusion and using the support theorem. We refer to Kliemann (1987) and Rey-Bellet (2006) for more details. This then shows that the process $\{\tilde{\mathbf{Y}}(t), t \geq 0\}$ is irreducible in Δ°.

Now we consider the case $r > 0$. We still assume that $\{\tilde{\mathbf{Y}}(t) : t \geq 0\}$ has a unique invariant probability measure. In order to obtain Theorem 2.1 for our degenerate process, we have to show that there is a sufficiently large $T > 0$ such that the Markov chain $(\mathbf{Y}(kT), S(kT))_{k\in\mathbb{N}}$ is irreducible and aperiodic and every compact subset of $\Delta^\circ \times (0, \infty)$ is petite for this Markov chain. Note that if every compact subset of $\Delta^\circ \times (0, \infty)$ is petite with respect to $(\mathbf{Y}(kT), S(kT))_{k\in\mathbb{N}}$, then any compact subset of $\Delta \times (0, \infty)$ is petite with respect to $(\mathbf{Y}(kT), S(kT))_{k\in\mathbb{N}}$ by the arguments in the proof of Lemma A.1.

Sufficient conditions for the above properties can be obtained by verifying the well-known Hörmander condition as well as investigating the control systems associated with the diffusion (2.4). Once we have the Markov chain $(\mathbf{Y}(kT), S(kT))_{k\in\mathbb{N}}$ being irreducible and aperiodic, and every compact subset of $\Delta^\circ \times (0,\infty)$ being petite for sufficiently large T, we can follow the steps from Appendix A to obtain the following result.

Theorem C.2 *Assume that $\tilde{\mathbf{Y}}(t)$ has a **unique** invariant probability measure ν^*. Define r by (2.8). Suppose that Assumption 2.2 holds and that $r > 0$. Assume further that there is a sufficiently large $T > 0$ such that the Markov chain $(\mathbf{Y}(kT), S(kT))_{k\in\mathbb{N}}$ is irreducible and aperiodic, and that every compact set in $\Delta^\circ \times (0,\infty)$ is petite for this Markov chain.*

The process $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))_{t\geq 0}$ has a unique invariant probability measure π on $\mathbb{R}^{n,\circ}_+$ that is absolutely continuous with respect to the Lebesgue measure and for any $q^* > 0$,

$$\lim_{t\to\infty} t^{q^*} \| P_{\mathbf{X}}(t, \mathbf{x}, \cdot) - \pi(\cdot)\|_{TV} = 0, \ \ \mathbf{x} \in \mathbb{R}^{n,\circ}_+, \tag{C.2}$$

where $\|\cdot, \cdot\|_{TV}$ is the total variation norm and $P_{\mathbf{X}}(t, \mathbf{x}, \cdot)$ is the transition probability of $(\mathbf{X}(t))_{t\geq 0}$. Moreover, for any initial value $\mathbf{x} \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}$ and any π-integrable function f, we have

$$\mathbb{P}\left\{ \lim_{T\to\infty} \frac{1}{T} \int_0^T f\left(\mathbf{X}^{\mathbf{x}}(t)\right) dt = \int_{\mathbb{R}^{n,\circ}_+} f(\mathbf{u})\pi(d\mathbf{u}) \right\} = 1. \tag{C.3}$$

C.1: Case study: $n = 2$

In what follows, we show that if $r > 0$, there is a sufficiently large $T > 0$ such that the Markov chain $(\mathbf{Y}(kT), S(kT))_{k\in\mathbb{N}}$ is irreducible and aperiodic, and that every compact set in $\Delta^\circ \times (0,\infty)$ is petite for the Markov chain.

For simplicity of presentation, we restrict ourselves to the $n = 2$ case, and assume that $b_i(x) = b_i x, x \geq 0, i = 1, 2$ for some $b_1, b_2 > 0$. As a result, (2.1) becomes

$$\begin{cases} dX_1(t) = \big(X_1(t)(a_1 - b_1 X_1(t)) - \alpha X_1(t) + \beta X_2(t)\big)dt + \sigma_1 X_1(t) dB(t) \\ dX_2(t) = \big(X_2(t)(a_2 - b_2 X_2(t)) + \alpha X_1(t) - \beta X_2(t)\big)dt + \sigma_2 X_2(t) dB(t), \end{cases} \tag{C.4}$$

where σ_1, σ_2 are non-zero constants and $(B(t))_{t\geq 0}$ is a one dimensional Brownian motion.

Setting $S(t) = X_1(t) + X_2(t)$ and $Y_i(t) = X_i(t)/S(t)$, $i = 1, 2$, we have from Itô's Lemma,

$$\begin{aligned} dY_i(t) &= Y_i(t)\left(a_i - \sum_{j=1}^{2} a_j Y_j - b_i S(t) Y_i(t) + S(t)\sum_{j=1}^{2} b_j Y_j^2(t))\right) dt \\ &\quad + (-1)^i \left(\alpha Y_1(t) - \beta Y_2(t)\right) dt \\ &\quad + Y_i(t)\left(\sum_{j,k=1}^{2} \sigma_k \sigma_j Y_k(t) Y_j(t)) - \sum_{j=1}^{2} \sigma_i \sigma_j Y_j(t)\right) dt \\ &\quad + (-1)^i (\sigma_2 - \sigma_1) Y_1(t) Y_2(t) dB(t) \\ dS(t) &= S(t)\left(\sum_{i=1}^{2} (a_i Y_i(t) - Y_i(t) b_i S(t) Y_i(t)\right) dt \\ &\quad + S(t)(\sigma_1 Y_1(t) + \sigma_2 Y_2(t)) dB(t). \end{aligned} \tag{C.5}$$

We use the process $(Y_1(t), Y_2(t), S(t))_{t\geq 0}$ to construct a Lyapunov function for a suitable skeleton $(Y_1(kT^*), Y_2(kT^*), S(kT^*))_{k\in\mathbb{N}}$ as we have done in Appendix A. However, to simplify the computations when verifying the hypotheses of Theorems C.1 and C.2, instead of working with $(Y_1(t), Y_2(t), S(t))$, we treat the system $(Z(t), X_2(t))$ where $Z(t) := X_1(t)/X_2(t)$. An application of Itô's Lemma yields

$$\begin{aligned} dZ(t) &= \Big((b_2 - b_1 Z(t)) Z(t) X_2(t) + \beta + \widehat{a}_1 Z(t) - \alpha Z^2(t)\Big) dt \\ &\quad + Z(t)[\sigma_1 - \sigma_2] dB(t) \\ dX_2(t) &= X_2(t)\Big((\widehat{a}_2 - b_2 X_2(t)) + \alpha Z(t)\Big) dt + \sigma_2 X_2(t) dB(t), \end{aligned} \tag{C.6}$$

where $\widehat{a}_1 = a_1 - a_2 - \alpha + \beta + \sigma_2^2 - \sigma_1\sigma_2$ and $\widehat{a}_2 = a_2 - \beta$.

To proceed, we first convert (C.6) to Stratonovich form to facilitate the verification of Hörmander's condition. System (C.6) can be rewritten as

$$\begin{aligned} dZ(t) =& \Big((b_2 - b_1 Z(t)) Z(t) X_2(t) + \beta + \left(\widehat{a}_1 - \frac{(\sigma_1 - \sigma_2)^2}{2}\right) Z(t) - \alpha Z^2(t)\Big) dt \\ &+ Z(t)[\sigma_1 - \sigma_2] \circ dB(t) \\ dX_2(t) =& X_2(t)\left(\left(\widehat{a}_2 - \frac{\sigma_2^2}{2} - b_2 X_2(t)\right) + \alpha Z(t)\right) dt + \sigma_2 X_2(t) \circ dB(t). \end{aligned} \tag{C.7}$$

Let

$$A_0(z, y) = \begin{pmatrix} (b_2 - b_1 z) zy + \beta + \left(\widehat{a}_1 - \dfrac{(\sigma_1 - \sigma_2)^2}{2}\right) z - \alpha z^2 \\ y\left(\widehat{a}_2 - \dfrac{\sigma_2^2}{2} - b_2 y\right) + \alpha zy \end{pmatrix},$$

and

$$A_1(z, y) = \begin{pmatrix} (\sigma_1 - \sigma_2)z \\ \sigma_2 y \end{pmatrix}.$$

Recall that the diffusion (C.7) is said to satisfy *Hörmander's condition* if the set of vector fields A_1, $[A_1, A_0]$, $[A_1, [A_1, A_0]]$, $[A_0, [A_1, A_0]]$, ... spans $\mathbb{R}^2$ at every $(z, y) \in \mathbb{R}^{2,\circ}_+$, where $[\cdot, \cdot]$ is the Lie bracket, which is defined as follows (see Rey-Bellet 2006 for more details). If $\Phi(z, y) = (\Phi_1(z, y), \Phi_2(z, y))^\top$ and $\Psi(z, y) = (\Psi_1(z, y), \Psi_2(z, y))^\top$ are vector fields on $\mathbb{R}^2$ (where $z^\top$ denotes the transpose of z), then the Lie bracket $[\Phi, \Psi]$ is a vector field given by

$$[\Phi, \Psi]_j(z, y) = \left(\Phi_1(z, y)\frac{\partial \Psi_j}{\partial z}(z, y) - \Psi_1(z, y)\frac{\partial \Phi_j}{\partial z}(z, y)\right) + \left(\Phi_2(z, y)\frac{\partial \Psi_j}{\partial y}(z, y) - \Psi_2(z, y)\frac{\partial \Phi_j}{\partial y}(z, y)\right), \quad j = 1, 2.$$

Proposition C.1 *Suppose that $\sigma_1 \neq \sigma_2$ or $\beta + (b_2/b_1)(a_1 - a_2) - \alpha(b_2/b_1)^2 \neq 0$. Then Hörmander's condition holds for the diffusion $(Z(t), X_2(t))_{t\geq 0}$ given by (C.7). As a result, the transition probability $P(t, (z, y), \cdot)$ of $(Z(t), X_2(t))_{t\geq 0}$ has a smooth density $\mathbb{R}_+ \times \mathbb{R}^4_+ \ni (t, z, y, z', y') \mapsto p(t, z, y, z', y') \in \mathbb{R}_+$ with respect to Lebesgue measure.*

Proof Set $\sigma := \dfrac{\sigma_1 - \sigma_2}{\sigma_2}$. By a direct calculation,

$$A_2(z, y) := \frac{1}{\sigma_2}[A_0, A_1](z, y) = \begin{pmatrix} \sigma(\beta + \alpha z^2) + (\sigma + 1)b_1 z^2 y - zyb_2 \\ -\sigma\alpha zy + b_2 y^2 \end{pmatrix},$$

and for $k > 2$, we have

$$A_k(z, y) := \frac{1}{\sigma_2}[A_1, A_{k-1}](z, y) = \begin{pmatrix} \sigma^{k-1}(\beta + (-1)^k \alpha z^2) + (-1)^k(\sigma + 1)^2 b_1 z^2 y + (-1)^{k+1} zyb_2 \\ (-1)^{k+1}\sigma^2\alpha zy + (-1)^k b_2 y^2. \end{pmatrix}.$$

If $\sigma \neq 0$ or equivalently $\sigma_1 \neq \sigma_2$, a straightforward but tedious computation shows that the rank of the matrix with columns A_1, A_2, A_3, A_4 is always 2 for any $(z, y) \in \mathbb{R}^{2,\circ}_+$. As a result, if $\sigma_1 \neq \sigma_2$, Hörmander's condition is satisfied for the diffusion (C.7). Therefore, the transition probability $P(t, (z, y), \cdot)$ of $(Z(t), X_2(t))$ has a smooth density function, denoted by $p(t, z, y, z', y')$; see Rey-Bellet (2006, Corollary 7.2).

Now, we show that Hörmander's condition holds if $\sigma_1 = \sigma_2$ and $\beta + (b_2/b_1)(a_1 - a_2 - \alpha + \beta) - \alpha(b_2/b_1)^2 \neq 0$. In this case,

$$A_2(z, y) = [A_0, A_1](z, y) = \begin{pmatrix} -\alpha yz(b_2 - b_1 z) \\ \alpha b_2 y^2 \end{pmatrix},$$

and

$$C(z,y)=\begin{pmatrix}C_1(z,y)\\C_2(z,y)\end{pmatrix}:=\left[A_0,\frac{1}{\alpha b_2}A_2\right](z,y),$$

where

$$\begin{aligned}C_1(z,y)=&\,y(2b_1z/b_2-1)A_{0,1}(z,y)+yz(1-zb_1/b_2)\frac{\partial A_{0,1}(z,y)}{\partial z}\\&+z(zb_1/b_2-1)A_{0,2}(z,y)+y^2z(zb_1/b_2-1).\end{aligned}$$

With $A_{0,i}(z,y)$ denoting the i-th component of $A_0(z,y)$. Observe that $A_1(x,y)$, $A_2(z,y)$ span $\mathbb{R}^2$ for any $(z,y)\in\mathbb{R}^{2,\circ}_+$ satisfying $z\neq b_2/b_1$. If $z=b_2/b_1$ we have $C_1(b_2/b_1,y)=yA_{0,1}(b_2/b_1,y)=y\big[\beta+(b_2/b_1)(a_1-a_2-\alpha+\beta)-\alpha(b_2/b_1)^2\big]\neq 0$ hence $C(b_2/b_1,y)$ and $A_2(b_2/b_1,y)$ span $\mathbb{R}^2$ for all $y>0$. As a result, we obtain the desired result.

□

To proceed, we consider the following control system, which is associated with (C.7).

$$\begin{cases}dz_\phi(t)=&(b_2-b_1z_\phi(t))z_\phi(t)y_\phi(t)+\beta\\&+\left(\widehat{a}_1-\dfrac{(\sigma_1-\sigma_2)^2}{2}\right)z_\phi(t)-\alpha z_\phi^2(t)+(\sigma_1-\sigma_2)z_\phi\phi(t)\\dy_\phi(t)=&y_\phi(t)\left(\widehat{a}_2-\dfrac{\sigma_2^2}{2}-b_2y_\phi(t)\right)+\alpha z_\phi(t)y_\phi(t)+\sigma_2y_\phi(t)\phi(t)\end{cases}\tag{C.8}$$

Let $(z_\phi(t,z,y),\ y_\phi(t,z,y))$ be the solution to equation (C.8) with control ϕ and initial value (z,y). Denote by $\mathcal{O}_1^+(z,y)$ the reachable set from (z,y), that is the set of $(z',y')\in\mathbb{R}^{2,\circ}_+$ such that there exists a $t\geq 0$ and a control $\phi(\cdot)$ satisfying $z_\phi(t,z,y)=z'$, $y_\phi(t,z,y)=z'$. We first recall some concepts introduced in Kliemann (1987). Let U be a subset of $\mathbb{R}^{2,\circ}_+$ satisfying $u_2\in\overline{\mathcal{O}_1^+(u_1)}$ for any $u_1,u_2\in U$. Then there is a unique maximal set $V\supset U$ such that this property still holds for V. Such V is called a *control set*. A control set C is said to be *invariant* if $\overline{\mathcal{O}_1^+(w)}\subset\overline{C}$ for all $w\in C$.

Finding invariant control sets for (C.8) is facilitated by using a change of variables argument. Put $w_\phi(t)=z_\phi(t)y_\phi^{r+1}(t)$ with $r=\frac{-\sigma_1}{\sigma_2}$. We have

$$\begin{cases}dw_\phi(t)=&h(w_\phi(t),y_\phi(t))dt\\dy_\phi(t)=&y_\phi(t)\left(\widehat{a}_2-\dfrac{\sigma_2^2}{2}-b_2y_\phi(t)\right)+\alpha w_\phi(t)y_\phi^{-r}(t)+\sigma_2y_\phi(t)\phi(t),\end{cases}\tag{C.9}$$

where

$$h(w,y) = w\left(a_1 - \frac{\sigma_1^2}{2} + r\left(a_2 - \frac{\sigma_2^2}{2}\right) + r\beta - \alpha \right.$$
$$\left. -b_1 w y^r - b_2 r y + \beta y^{1-r} w^{-1} + \alpha r w y^{r-1}\right).$$

Denote by $\mathcal{O}_2^+(w,y)$ the set of $(w',y') \in \mathbb{R}_+^{2,\circ}$ such that there is a $t>0$ and a control $\phi(\cdot)$ such that $w_\phi(t,w,y)=w'$, $z_\phi(t,w,y)=w'$.

Lemma C.1 *The control system (C.9) has only one invariant control set $\tilde{\mathcal{C}}$ and $\overline{\mathcal{O}_2^+(w,y)} \supset \tilde{\mathcal{C}}$ for any $(w,y) \in \mathbb{R}_+^{2,\circ}$, The set $\tilde{\mathcal{C}}$ is defined by $\tilde{\mathcal{C}} = \{(w,y) \in \mathbb{R}_+^{2,\circ} : w < c^*\}$, where*

$$c^* = \sup\left\{ w : \sup_{y>0}\{h(w',y)\} \geq 0 \text{ for all } w' < w \right\}.$$

Consequently, the control system (C.8) has only one invariant control set $\mathcal{C}$ and $\overline{\mathcal{O}_1^+(z,y)} \supset \mathcal{C}$ for any $(w,y) \in \mathbb{R}_+^{2,\circ}$, where $\mathcal{C} := \{(z,y) \in \mathbb{R}_+^{2,\circ} : z y^{r+1} \leq c^*\}$. Moreover, by Kliemann (1987, Lemma 4.1), $(Z(t), X_2(t))$ has at most one invariant probability measure whose support is $\mathcal{C}$.

Proof First, we need to show that c^* is well-defined (although it can be $+\infty$). Since $\lim_{w\to 0} h(w,y) = \infty$, which implies that $\left\{ w : \sup_{y>0}\{h(w',y)\} \geq 0 \text{ for all } w' \leq w \right\}$ is a nonempty set. Hence c^* is well-defined. The claim that $\overline{\mathcal{O}_2^+(w,y)} \supset \tilde{\mathcal{C}}$ for any $(w,y) \in \mathbb{R}_+^{2,\circ}$ can be proved by standard arguments. Let us explain the main ideas here. On the phase space $(w,y) \in \mathbb{R}_+^{2,\circ}$, since the control $\phi(t)$ only appears in the equation of y_ϕ, we can easily control vertically, that is, for any initial points y_0 and w_0, there is a control so that y_ϕ can reach any given point y_1 while w_ϕ stays in a given neighborhood of w_0. If $h(w_0,y_0) < 0$, we can choose a feedback control such that $(w_\phi(t), u_\phi(t))$ reaches a point to the 'left' (w_1, y_0) with $w_1 < w_0$ as long as $h(w,y_0) < 0$ for $w \in [w_1, w_0]$. Likewise, for $h(w_0,y_0) > 0$, we can choose a feedback control such that $(w_\phi(t), u_\phi(t))$ can reach a point to the 'right' (w_1, y_0) with $w_1 > w_0$ as long as $h(w,y_0) > 0$ for $w \in [w_0, w_1]$. We also have that $\inf_{y>0}\{h(w,y)\} = -\infty$ for any $w > 0$. Using these facts, we can follow the steps from Du et al. (2016, Section 3) to obtain the desired results. □

Lemma C.2 *There is a point $(z^*, y^*) \in \mathcal{C}$ such that for any open set $N^* \ni (z^*, y^*)$ and $T > 0$, there is an open neighborhood $W^* \ni (z^*, y^*)$ and a control ϕ^* such that*

$$(z_{\phi^*}(t,z,y), y_{\phi^*}(t,z,y)) \in N^* \text{ for all } (z,y) \in W^*, t \in [0,T].$$

Proof To obtain the result, we work on (C.9), which is equivalent to (C.8). By the definition of $\tilde{\mathcal{C}}$ and the fact that $\lim_{y\to\infty} h(w, y) = -\infty$ if $r > 0$ and $\lim_{y\to 0} h(w, y) = -\infty$ if $r < 0$, there is a point $(w^*, y^*) \in \tilde{\mathcal{C}}$ such that $h(w^*, y^*) = 0$. We can design a feedback control ϕ^* such that

$$\begin{cases} dw_{\phi^*}(t) = & h(w_{\phi^*}(t), y^*)dt \\ dy_{\phi^*}(t) = & 0. \end{cases} \tag{C.10}$$

If $w_{\phi^*}(t) = w^*$ then $w_{\phi^*}(t) = w^*$ for all $t > 0$. By the continuous dependence on initial data of solutions to differential equations, for any given neighborhood $\tilde{N}^*$ of (w^*, y^*), we can find a neighborhood $\tilde{W}^*$ of (w^*, y^*) such that $(w_{\phi^*}(t, w, y), y_{\phi^*}(t, w, y)) \in \tilde{N}^*$ for any $t \in [0, T]$ and $(w, y) \in \tilde{W}^*$, which proves the lemma. □

Proposition C.2 *Suppose $\sigma_1 \neq \sigma_2$ or $\beta + (b_2/b_1)(a_1 - a_2) - \alpha(b_2/b_1)^2 \neq 0$. For any $T > 0$, every compact set $K \subset \mathbb{R}^{2,\circ}_+$ is petite set with respect to the Markov chain $(Z(kT), X_2(kT))_{k\in\mathbb{N}}$.*

Proof Let (z^*, y^*) be as in Lemma C.2. Pick $(z^\diamond, y^\diamond) \in \mathbb{R}^{2,\circ}_+$ such that $p(T, z^*, y^*, z^\diamond, y^\diamond) > 0$. By the smoothness of $p(T, \cdot, \cdot, \cdot, \cdot)$, there exists a neighborhood N^* and an open set $N^\diamond \ni (z^\diamond, y^\diamond)$ such that

$$p(1, z, y, z', y') \geq p^\diamond > 0 \text{ for all } (z, y) \in N^*, (z', y') \in N^\diamond. \tag{C.11}$$

Let W^* be a neighborhood of (z^*, y^*) satisfying

$$(z_{\phi^*}(t, z, y), y_{\phi^*}(t, z, y)) \in N^* \text{ for all } (z, y) \in W^*, t \in [0, T]. \tag{C.12}$$

For each $(z, y) \in \mathbb{R}^{2,\circ}_+$, noting that $(z^*, y^*) \in \mathcal{C} \subset \overline{\mathcal{O}^+_1(z, y)}$, there is a control ϕ and $t_{z,y} > 0$ such that

$$(z_\phi(t_{z,y}, z, y), y_\phi(t_{z,y}, z, y)) \in W^*. \tag{C.13}$$

Let $n_{z,y} \in \mathbb{N}$ such that $(n_{z,y} - 1)T < t_{z,y} \leq n_{n,y}T$ and $\tilde{\phi}$ be defined as $\tilde{\phi}(t) = \phi(t)$ if $t < t_{z,y}$ and $\tilde{\phi}(t) = \phi^*(t)$ if $t > t_{z,y}$. Using the control $\tilde{\phi}$, we obtain from (C.12) and (C.13) that

$$\left(z_{\tilde{\phi}}(n_{z,y}T, z, y), y_{\tilde{\phi}}(n_{z,y}T, z, y)\right) \in N^*. \tag{C.14}$$

In view of the support theorem (see Ikeda and Watanabe 1989, Theorem 8.1, p. 518),

$$P(n_{z,y}T, z, y, N^*) := 2\rho_{z,y} > 0.$$

Since $(Z_{z,y}(t), Y_{z,y}(t))$ is a Markov–Feller process, there exists an open set $V_{z,y} \ni (z, y)$ such that $P(n_{z,y}T, z', y', N^*) \geq \rho_{u,v}$ for all $(z', y') \in V_{z,y}$. Since K is a compact set, there is a finite number of V_{z_i,y_i}, $i = 1, \ldots, k_0$ satisfying $K \subset \bigcup_{i=1}^{k_0} V_{z_i,y_i}$. Let $\rho_K = \min\{\rho_{z_i,y_i}, \ i = 1, \ldots, k_0\}$. For each $(z, y) \in K$, there exists n_{z_i,y_i} such that

$$P(n_{z_i,y_i}T, z, y, N^*) \geq \rho_K. \tag{C.15}$$

From (C.11) and (C.15), for all $(z, y) \in K$, there exists n_{z_i,y_i} such that

$$p((n_{z_i,y_i} + 1)T, z, y, z', y') \geq \rho_K p^\diamond \ \text{ for all } \ (z', y') \in N^\diamond. \tag{C.16}$$

It follows from (C.16) that

$$\frac{1}{k_0}\sum_{i=1}^{k_0} P((n_{z_i,y_i} + 1)T, z, y, A) \geq \frac{1}{k_0}\rho_K p^\diamond m(N^\diamond \cap A) \quad \text{for all } A \in \mathcal{B}\left(\mathbb{R}_+^{2,\circ}\right), \tag{C.17}$$

where $m(\cdot)$ is the Lebesgue measure on $\mathbb{R}_+^{2,\circ}$. Equation (C.17) implies that every compact set $K \subset \mathbb{R}_+^{2,\circ}$ is petite for the Markov chain $(Z(kT), X_2(kT))_{k\in\mathbb{N}}$. □

We have shown in the beginning of Sect. 2.2. that $\tilde{\mathbf{Y}}(t)$ has a **unique** invariant probability measure ν^*. Having Proposition C.2, we note that the assumptions, and therefore the conclusions, of Theorems C.1 and C.2 hold for model (C.4). This argument proves Theorems 2.5 and 2.6.

Appendix D: Robustness of the model

The robustness is studied from several angles, including continuous dependence of r on the coefficients of the stochastic differential equation, robustness of persistence, and robust attenuation against extinction. They are presented in a couple subsections.

D.1: Continuous dependence of r on the coefficients

We show that r depends continuously on the coefficients of the stochastic differential equation (2.6). Consider the equation

$$\begin{aligned} d\widehat{\mathbf{Y}}(t) = &\left(\operatorname{diag}(\widehat{\mathbf{Y}}(t)) - \widehat{\mathbf{Y}}(t)\widehat{\mathbf{Y}}^\top(t)\right)\widehat{\Gamma}^\top d\mathbf{B}(t) \\ &+ \widehat{\mathbf{D}}^\top\widehat{\mathbf{Y}}(t)dt + \left(\operatorname{diag}(\widehat{\mathbf{Y}}(t)) - \widehat{\mathbf{Y}}(t)\widehat{\mathbf{Y}}^\top(t)\right)(\widehat{\mathbf{a}} - \widehat{\Sigma}\widehat{\mathbf{Y}}(t))dt \end{aligned} \tag{D.1}$$

on the simplex Δ. Suppose that $\widehat{\Sigma}$ is positive definite. In this case, $(\widehat{\mathbf{Y}}(t))_{t\geq 0}$ has a unique invariant probability measure $\widehat{\nu}^*$. Define

$$\widehat{r} := \int_{\Delta} \left(\widehat{\mathbf{a}}^\top \mathbf{y} - \frac{1}{2} \mathbf{y}^\top \widehat{\Sigma} \mathbf{y} \right) \widehat{\nu}^*(d\mathbf{y}) \tag{D.2}$$

Fix the coefficients of 2.6.

Proposition D.1 *For any $\varepsilon > 0$, there is a $\theta_2 > 0$ such that if*

$$\max \left\{ \|\mathbf{a} - \widehat{\mathbf{a}}\|, \|D - \widehat{D}\|, \|\Gamma - \widehat{\Gamma}\| \right\} < \theta_2$$

then

$$|r - \widehat{r}| < \varepsilon.$$

Proof First, let $\theta_1 > 0$ such that if $\max \left\{ \|\mathbf{a} - \widehat{\mathbf{a}}\|, \|D - \widehat{D}\|, \|\Gamma - \widehat{\Gamma}\| \right\} < \theta_1$, then

$$\left| \left(\widehat{\mathbf{a}}^\top \mathbf{y} - \frac{1}{2} \mathbf{y}^\top \widehat{\Sigma} \mathbf{y} \right) - \left(\mathbf{a}^\top \mathbf{y} - \frac{1}{2} \mathbf{y}^\top \Sigma \mathbf{y} \right) \right| < \frac{\varepsilon}{3} \text{ for all } \mathbf{y} \in \Delta. \tag{D.3}$$

Let $\gamma_1, \gamma_2, M_3, M_4$ be defined as in the proof of Lemma A.5. Pick $T = T(\varepsilon) > 0$ such that

$$\|\tilde{P}(T, \mathbf{y}, \cdot) - \nu^*\|_{TV} \leq \gamma_1 \exp(-\gamma_2 T) < \frac{\varepsilon}{3M_4} \text{ for all } \mathbf{y} \in \Delta. \tag{D.4}$$

By standard arguments, there is a $\theta_2 \in (0, \theta_1)$ such that if $\max \left\{ \|\mathbf{a} - \widehat{\mathbf{a}}\|, \|D - \widehat{D}\|, \|\Gamma - \widehat{\Gamma}\| \right\} < \delta_2$, then

$$\mathbb{P} \left\{ \|\tilde{\mathbf{Y}}^{\mathbf{y}}(T) - \widehat{\mathbf{Y}}^{\mathbf{y}}(T)\| < \frac{\varepsilon}{6M_3} \right\} > \frac{\varepsilon}{6M_4} \text{ for all } \mathbf{y} \in \Delta \tag{D.5}$$

Let $\mathbf{y}^*$ be a Δ-valued and $\mathcal{F}_0$-measurable random variable whose distribution is $\widehat{\nu}^*$. Clearly,

$$\int_{\Delta} \left(\mathbf{a}^\top \mathbf{y} - \frac{1}{2} \mathbf{y}^\top \Sigma \mathbf{y} \right) \widehat{\nu}^*(d\mathbf{y}) = \mathbb{E} \left(\mathbf{a}^\top \widehat{\mathbf{Y}}^{\mathbf{y}^*}(T) - \frac{1}{2} (\widehat{\mathbf{Y}}^{\mathbf{y}^*}(T))^\top \Sigma \widehat{\mathbf{Y}}^{\mathbf{y}^*}(T) \right). \tag{D.6}$$

In view of (D.4),

$$\begin{aligned} \left| \mathbb{E} \left(\mathbf{a}^\top \tilde{\mathbf{Y}}^{\mathbf{y}^*}(T) - \frac{1}{2} (\tilde{\mathbf{Y}}^{\mathbf{y}^*}(T))^\top \Sigma \tilde{\mathbf{Y}}^{\mathbf{y}^*}(T) \right) - r \right| &\leq M_4 \sup_{\mathbf{y} \in \Delta} \{ \|\tilde{P}(t, \mathbf{y}, \cdot) - \mu^*\| \} \\ &\leq \frac{\varepsilon}{3}. \end{aligned} \tag{D.7}$$

It follows from (D.5) that

$$\mathbb{E}\left|\mathbf{a}^\top\widehat{\mathbf{Y}}^{\mathbf{y}^*}(T)-\frac{1}{2}(\widehat{\mathbf{Y}}^{\mathbf{y}^*}(T))^\top\Sigma\widehat{\mathbf{Y}}^{\mathbf{y}^*}(T)-\mathbf{a}^\top\tilde{\mathbf{Y}}^{\mathbf{y}^*}(T)+\frac{1}{2}(\tilde{\mathbf{Y}}^{\mathbf{y}^*}(T))^\top\Sigma\tilde{\mathbf{Y}}^{\mathbf{y}^*}(T)\right|$$
$$\leq M_3\frac{\varepsilon}{6M_3}\mathbb{P}\left\{\|\tilde{\mathbf{Y}}^{\mathbf{y}^*}-\widehat{\mathbf{Y}}^{\mathbf{y}^*}\|<\frac{\varepsilon}{6M_3}\right\}+M_4\mathbb{P}\left\{\|\tilde{\mathbf{Y}}^{\mathbf{y}^*}-\widehat{\mathbf{Y}}^{\mathbf{y}^*}\|\geq\frac{\varepsilon}{6M_3}\right\}\leq\frac{\varepsilon}{3}. \tag{D.8}$$

In view of (D.2), (D.3), (D.6), (D.7), and (D.8), if

$$\max\left\{\|\mathbf{a}-\widehat{\mathbf{a}}\|,\|D-\widehat{D}\|,\|\Gamma-\widehat{\Gamma}\|\right\}<\theta_2$$

then $|r-\widehat{r}|<\varepsilon$, which completes the proof. □

Remark D.1 The continuous dependence of r on the coefficients can also be proved by generalizing the arguments from the proof of Evans et al. (2013, Proposition 3). Since Evans et al. (2013, Proposition 3) focuses only on the continuity for a specific parameter rather than all parameters, we provided an alternative proof for the sake of completeness.

D.2: Robust persistence and extinction

Sketch of proof of Theorem 2.8 As usual, we work with

$$\begin{aligned}d\widehat{\mathbf{Y}}(t)&=\left(\mathrm{diag}(\widehat{\mathbf{Y}}(t))-\widehat{\mathbf{Y}}(t)\widehat{\mathbf{Y}}^\top(t)\right)\widehat{\Gamma}^\top(\widehat{S}(t)\widehat{\mathbf{Y}}(t))d\mathbf{B}(t)+\widehat{\mathbf{D}}(\widehat{S}(t)\widehat{\mathbf{Y}}(t))^\top\widehat{\mathbf{Y}}(t)dt\\&\quad+\left(\mathrm{diag}(\widehat{\mathbf{Y}}(t))-\widehat{\mathbf{Y}}(t)\widehat{\mathbf{Y}}^\top(t)\right)(\widehat{\mathbf{a}}-\widehat{\Sigma}(\widehat{S}(t)\widehat{\mathbf{Y}}(t))\widehat{\mathbf{Y}}(t)-\widehat{\mathbf{b}}(\widehat{S}(t)\widehat{\mathbf{Y}}(t)))dt\\d\widehat{S}(t)&\\&=\widehat{S}(t)\left[\widehat{\mathbf{a}}-\widehat{\mathbf{b}}(\widehat{S}(t)\widehat{\mathbf{Y}}(t))\right]^\top\widehat{\mathbf{Y}}(t)dt+\widehat{S}(t)\widehat{\mathbf{Y}}(t)^\top\widehat{\Gamma}^\top(\widehat{S}(t)\widehat{\mathbf{Y}}(t))d\mathbf{B}(t),\end{aligned} \tag{D.9}$$

where $\widehat{S}(t):=\sum_i\widehat{X}_i(t)$, $\widehat{\mathbf{Y}}(t):=\frac{\widehat{\mathbf{X}}(t)}{\widehat{S}(t)}$. In order to have a complete proof for this proposition one can follow the steps from Appendix A. First, since Σ is positive definite then so is $\widehat{\Sigma}(\mathbf{x}):=\widehat{\Gamma}(\mathbf{x})^\top\widehat{\Gamma}(\mathbf{x})$ if $\sup_{\mathbf{x}\in\mathbb{R}^{n,\circ}_+}\|\widehat{\Gamma}(\mathbf{x})-\Gamma\|$ is sufficiently small. As a result, $(\widehat{\mathbf{X}}(t))_{t\geq0}$ is a nondegenerate diffusion in $\mathbb{R}^{n,\circ}_+$ and Lemma A.1 holds for $(\widehat{\mathbf{Y}}(nT),\widehat{S}(nT))_{n\in\mathbb{N}}$. We also have the following results: there exist positive constants $\widehat{K}_i:i=1,\ldots,4$, which do not depend on θ as long as θ is sufficiently small, such that

$$\mathbb{E}\widehat{S}^{\mathbf{y},s}(t)\leq e^{-\gamma_b t/2}s+\widehat{K}_1,\ \ (\mathbf{y},s)\in\Delta\times(0,\infty),t\geq0. \tag{D.10}$$

$$\mathbb{E}\left([\ln\widehat{S}^{\mathbf{y},s}(T)]^2\right)\leq((\ln s)^2+1)\widehat{K}_2\exp\{\widehat{K}_3T\},\ \ (\mathbf{y},s)\in\Delta\times(0,\infty),T\geq0, \tag{D.11}$$

and

$$\mathbb{E}\left([\ln \widehat{S}^{\mathbf{y},s}(T \wedge \widehat{\zeta}^{\mathbf{y},s})]_-^2\right) \leq (\ln s)^2 + \widehat{K}_4\sqrt{\mathbb{P}(A)}(T+1)[\ln s]_- + \widehat{K}_4 T^2 \tag{D.12}$$

for all $(\mathbf{y}, s) \in \Delta \times (0,1)$, $A \in \mathcal{F}$ where

$$\widehat{\zeta}^{\mathbf{y},s} := \inf\{t \geq 0 : \widehat{S}^{\mathbf{y},s}(t) = 1\}.$$

On the other hand, standard arguments show that for any $\varepsilon > 0, T > 0$, there is a $\theta = \theta(\varepsilon, T) > 0$ such that

$$\mathbb{P}\left\{\left\|(\mathbf{Y}^{\mathbf{y},s}(t), S^{\mathbf{y},s}(t)) - (\widehat{\mathbf{Y}}^{\mathbf{y},s}(t), \widehat{S}^{\mathbf{y},s}(t))\right\| \leq \varepsilon,\ 0 \leq t \leq T\right\} > 1 - \varepsilon$$

given that $(\mathbf{y}, s) \in \Delta \times [0,1]$. Combining this fact with Proposition A.2, one can find $\delta = \delta(\varepsilon, T) > 0$ and $\theta = \theta(\varepsilon, T) > 0$ such that

$$\mathbb{P}\left\{\left\|(\tilde{\mathbf{Y}}^{\mathbf{y},s}(t), 0) - (\widehat{\mathbf{Y}}^{\mathbf{y},s}(t), \widehat{S}^{\mathbf{y},s}(t))\right\| \leq \varepsilon,\ 0 \leq t \leq T\right\} > 1 - \varepsilon$$

given that $(\mathbf{y}, s) \in \Delta \times (0, \delta)$ and (2.20) holds. With this fact, we can use Lemma A.5 with slight modification to show that, for any $\varepsilon > 0$, there is a $T^* = T^*(\varepsilon)$ and $\delta = \delta(\varepsilon, T^*), \theta = \theta(\varepsilon, T^*)$ such that

$$\mathbb{P}\left\{\ln s + \frac{3rT^*}{4} \leq \ln \widehat{S}^{\mathbf{y},s}(T^*) < 0\right\} \geq 1 - 3\varepsilon \text{ for all } (\mathbf{y}, s) \in \Delta \times (0, \delta) \tag{D.13}$$

given that (2.20) holds. Having (D.10), (D.11), (D.12), and (D.13), we can use the arguments from Proposition A.3 and Theorem A.2 to finish the proof. □

Remark D.2 If $r < 0, \mathbf{X}(t)$ converges to $\mathbf{0}$ with probability 1. By virtue of Proposition D.1, if $\widehat{D}, \widehat{\Gamma}$ are constant matrices and max $\left\{\|\mathbf{a} - \widehat{\mathbf{a}}\|, \|D - \widehat{D}\|, \|\Gamma - \widehat{\Gamma}\|\right\}$ is sufficiently small then $\widehat{\mathbf{X}}(t)$ converges to $\mathbf{0}$ with an exponential rate almost surely. We conjecture that this result holds for any θ-perturbation of $\mathbf{X}(t)$ defined by (2.20). However, when $\widehat{D} := \widehat{D}(\mathbf{x}), \widehat{\Gamma} := \widehat{\Gamma}(\mathbf{x})$, comparison arguments may be not applicable. Moreover, it is also difficult to analyze the asymptotic behavior of the equation without competition terms, namely

$$\begin{aligned} d\widehat{\mathcal{X}}(t) = &\left(\operatorname{diag}(\widehat{\mathcal{X}}(t))\widehat{\mathbf{a}} + \widehat{D}(\widehat{\mathcal{X}}(t))^\top \widehat{\mathcal{X}}(t)\right) dt \\ &+ \operatorname{diag}(\widehat{\mathcal{X}}(t))\widehat{\Gamma}(\widehat{\mathcal{X}}(t))^\top d\mathbf{B}(t). \end{aligned} \tag{D.14}$$

References

Altenberg L (2012) The evolution of dispersal in random environments and the principle of partial control. Ecol Monogr 82(3):297–333

Assing S, Manthey R (1995) The behavior of solutions of stochastic differential inequalities. Probab Theory Relat Fields 103(4):493–514

Bascompte J, Possingham H, Roughgarden J (2002) Patchy populations in stochastic environments: critical number of patches for persistence. Am Nat 159(2):128–137

Benaïm M, Schreiber SJ (2009) Persistence of structured populations in random environments. Theor Popul Biol 76(1):19–34

Benaïm M, Hofbauer J, Sandholm WH (2008) Robust permanence and impermanence for stochastic replicator dynamics. J Biol Dyn 2(2):180–195

Blath J, Etheridge A, Meredith M (2007) Coexistence in locally regulated competing populations and survival of branching annihilating random walk. Ann Appl Probab 17(5–6):1474–1507 ISSN 1050-5164

Cantrell RS, Cosner C (1991) The effects of spatial heterogeneity in population dynamics. J Math Biol 29(4):315–338

Cantrell RS, Cosner C, Lou Y (2012) Evolutionary stability of ideal free dispersal strategies in patchy environments. J Math Biol 65(5):943–965

Caswell H (2001) Matrix population models. Wiley Online Library, New York

Chesson P (2000) General theory of competitive coexistence in spatially-varying environments. Theor Popul Biol 58(3):211–237

Chueshov I (2002) Monotone random systems theory and applications, vol 1779. Springer Science & Business Media, Berlin

Cross PC, Lloyd-Smith JO, Johnson PLF, Getz WM (2005) Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. Ecol Lett 8(6):587–595

Davies KF, Chesson P, Harrison S, Inouye BD, Melbourne B, Rice KJ (2005) Spatial heterogeneity explains the scale dependence of the native-exotic diversity relationship. Ecology 86(6):1602–1610

Dennis B, Patil GP (1984) The gamma distribution and weighted multimodal gamma distributions as models of population abundance. Math Biosci 68(2):187–212

Dieu NT, Nguyen DH, Du NH, Yin G (2016) Classification of asymptotic behavior in a stochastic sir model. SIAM J Appl Dyn Syst 15(2):1062–1084

Du NH, Nguyen DH, Yin G (2016) Conditions for permanence and ergodicity of certain stochastic predator-prey models. J Appl Probab 53:187–202

Durrett R, Remenik D (2012) Evolution of dispersal distance. J Math Biol 64(4):657–666

Evans SN, Ralph PL, Schreiber SJ, Sen A (2013) Stochastic population growth in spatially heterogeneous environments. J Math Biol 66(3):423–476

Evans SN, Hening A, Schreiber SJ (2015) Protected polymorphisms and evolutionary stability of patch-selection strategies in stochastic environments. J Math Biol 71(2):325–359 ISSN 0303-6812

Garay BM, Hofbauer J (2003) Robust permanence for ecological differential equations, minimax, and discretizations. SIAM J Math Anal 34(5):1007–1039

Geiß C, Manthey R (1994) Comparison theorems for stochastic differential equations in finite and infinite dimensions. Stoch Process Appl 53(1):23–35

Gonzalez A, Holt RD (2002) The inflationary effects of environmental fluctuations in source-sink systems. Proc Nat Acad Sci 99(23):14872–14877

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

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

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

Harrison S, Quinn JF (1989) Correlated environments and the persistence of metapopulations. Oikos 56(3):293–298

Hastings A (1983) Can spatial variation alone lead to selection for dispersal? Theor Popul Biol 24(3):244–251

Hutson V, Schmitt K (1992) Permanence and the dynamics of biological systems. Math Biosci 111(1):1–71

Ikeda N, Watanabe S (1989) Stochastic differential equations and diffusion processes. North-Holland Publishing Co., Amsterdam
Jansen VAA, Yoshimura J (1998) Populations can persist in an environment consisting of sink habitats only. Proc Nat Acad Sci 95(7):3696–3698
Jarner SF, Roberts GO (2002) Polynomial convergence rates of Markov chains. Ann Appl Probab 12(1):224–247 ISSN 1050-5164
Kallenberg O (2002) Foundations of modern probability. Springer, Berlin
Karlin S (1982) Classifications of selection-migration structures and conditions for a protected polymorphism. Evol Biol 14(61):204
Kendall BE, Bjørnstad ON, Bascompte J, Keitt TH, Fagan WF (2000) Dispersal, environmental correlation, and spatial synchrony in population dynamics. Am Nat 155(5):628–636
Khasminskii R (2012) Stochastic stability of differential equations, volume 66 of Stochastic modelling and applied probability, 2nd edn. Springer, Heidelberg (With contributions by G. N. Milstein and M. B. Nevelson)
Kirkland S, Li C-K, Schreiber SJ (2006) On the evolution of dispersal in patchy landscapes. SIAM J Appl Math 66(4):1366–1382
Kliemann W (1987) Recurrence and invariant measures for degenerate diffusions. Ann Probab 15:690–707
Law R, Morton RD (1996) Permanence and the assembly of ecological communities. Ecology 77:762–775
Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74(6):1659–1673
Liebhold A, Koenig WD, Bjørnstad ON (2004) Spatial synchrony in population dynamics. Annu Rev Ecol Evolut Syst 35:467–490
Mao X (1997) Stochastic differential equations and their applications. Horwood publishing series in mathematics & applications. Horwood Publishing Limited, Chichester
Meyn SP, Tweedie RL (1993) Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. Adv Appl Probab 24(3):518–548
Mierczyński J, Shen W (2004) Lyapunov exponents and asymptotic dynamics in random Kolmogorov models. J Evol Equ 4(3):371–390
Nummelin E (1984) General irreducible Markov chains and nonnegative operators, volume 83 of Cambridge tracts in mathematics. Cambridge University Press, Cambridge
Palmqvist E, Lundberg P (1998) Population extinctions in correlated environments. Oikos 83(2):359–367
Patel S, Schreiber SJ (2016) Robust permanence for ecological equations with internal and external feedbacks. arXiv:1612.06554
Pyšek P, Hulme PE (2005) Spatio-temporal dynamics of plant invasions: linking pattern to process. Ecoscience 12(3):302–315
Rey-Bellet L (2006) Ergodic properties of Markov processes. In: Attal S, Joye A, Pillet CA (eds) Open quantum systems. II, volume 1881 of Lecture notes in mathematics, pp 1–39. Springer, Berlin
Roth G, Schreiber SJ (2014) Persistence in fluctuating environments for interacting structured populations. J Math Biol 69(5):1267–1317
Roy M, Holt RD, Barfield M (2005) Temporal autocorrelation can enhance the persistence and abundance of metapopulations comprised of coupled sinks. Am Nat 166(2):246–261
Rudnicki R (2003) Long-time behaviour of a stochastic preypredator model. Stoch Process Appl 108(1):93–107
Schmidt KA (2004) Site fidelity in temporally correlated environments enhances population persistence. Ecol Lett 7(3):176–184
Schreiber SJ (2000) Criteria for Cr robust permanence. J Differ Equ 162(2):400–426
Schreiber SJ (2010) Interactive effects of temporal correlations, spatial heterogeneity and dispersal on population persistence. Proc R Soc Lond B Biol Sci. http://rspb.royalsocietypublishing.org/content/early/2010/02/12/rspb.2009.2006.full. Accessed 01 Dec 2016
Schreiber SJ (2012) The evolution of patch selection in stochastic environments. Am Nat 180(1):17–34
Schreiber SJ, Li C-K (2011) Evolution of unconditional dispersal in periodic environments. J Biol Dyn 5(2):120–134
Schreiber SJ, Lloyd-Smith JO (2009) Invasion dynamics in spatially heterogeneous environments. Am Nat 174(4):490–505
Schreiber SJ, Ryan ME (2011) Invasion speeds for structured populations in fluctuating environments. Theor Ecol 4(4):423–434
Schreiber SJ, Benaïm M, Atchadé KAS (2011) Persistence in fluctuating environments. J Math Biol 62(5):655–683