

Stochastic population growth in spatially heterogeneous environments

Steven N. Evans · Peter L. Ralph ·
Sebastian J. Schreiber · Arnab Sen

Received: 11 May 2011 / Revised: 14 January 2012 / Published online: 18 March 2012
© Springer-Verlag 2012

Abstract Classical ecological theory predicts that environmental stochasticity increases extinction risk by reducing the average per-capita growth rate of populations. For sedentary populations in a spatially homogeneous yet temporally variable environment, a simple model of population growth is a stochastic differential equation $dZ_t = \mu Z_t dt + \sigma Z_t dW_t$, $t \geq 0$, where the conditional law of $Z_{t+\Delta t} - Z_t$ given $Z_t = z$ has mean and variance approximately $z\mu\Delta t$ and $z^2\sigma^2\Delta t$ when the time increment Δt is small. The long-term stochastic growth rate $\lim_{t \rightarrow \infty} t^{-1} \log Z_t$ for such a population equals $\mu - \frac{\sigma^2}{2}$. Most populations, however, experience spatial as well as temporal variability. To understand the interactive effects of environmental stochasticity, spatial heterogeneity, and dispersal on population growth, we study an analogous model $\mathbf{X}_t = (X_t^1, \dots, X_t^n)$, $t \geq 0$, for the population abundances in n patches: the conditional law of $\mathbf{X}_{t+\Delta t}$ given $\mathbf{X}_t = x$ is such that the conditional mean of $X_{t+\Delta t}^i - X_t^i$ is approximately $[x^i \mu_i + \sum_j (x^j D_{ji} - x^i D_{ij})] \Delta t$ where μ_i is the

Electronic supplementary material The online version of this article (doi:10.1007/s00285-012-0514-0) contains supplementary material, which is available to authorized users.

S. N. Evans
Department of Statistics #3860, University of California,
367 Evans Hall, Berkeley, CA 94720-3860, USA
e-mail: evans@stat.berkeley.edu

P. L. Ralph · S. J. Schreiber (✉)
Department of Evolution and Ecology, University of California, Davis, CA 956116, USA
e-mail: sschreiber@ucdavis.edu

P. L. Ralph
e-mail: plralph@ucdavis.edu

A. Sen
Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road,
Cambridge CB3 0WB, UK
e-mail: A.Sen@statslab.cam.ac.uk

per capita growth rate in the i th patch and D_{ij} is the dispersal rate from the i th patch to the j th patch, and the conditional covariance of $X_{t+\Delta t}^i - X_t^i$ and $X_{t+\Delta t}^j - X_t^j$ is approximately $x^i x^j \sigma_{ij} \Delta t$ for some covariance matrix $\Sigma = (\sigma_{ij})$. We show for such a spatially extended population that if $S_t = X_t^1 + \dots + X_t^n$ denotes the total population abundance, then $\mathbf{Y}_t = \mathbf{X}_t/S_t$, the vector of patch proportions, converges in law to a random vector \mathbf{Y}_∞ as $t \rightarrow \infty$, and the stochastic growth rate $\lim_{t \rightarrow \infty} t^{-1} \log S_t$ equals the space-time average per-capita growth rate $\sum_i \mu_i \mathbb{E}[Y_\infty^i]$ experienced by the population minus half of the space-time average temporal variation $\mathbb{E}[\sum_{i,j} \sigma_{ij} Y_\infty^i Y_\infty^j]$ experienced by the population. Using this characterization of the stochastic growth rate, we derive an explicit expression for the stochastic growth rate for populations living in two patches, determine which choices of the dispersal matrix D produce the maximal stochastic growth rate for a freely dispersing population, derive an analytic approximation of the stochastic growth rate for dispersal limited populations, and use group theoretic techniques to approximate the stochastic growth rate for populations living in multi-scale landscapes (e.g. insects on plants in meadows on islands). Our results provide fundamental insights into “ideal free” movement in the face of uncertainty, the persistence of coupled sink populations, the evolution of dispersal rates, and the single large or several small (SLOSS) debate in conservation biology. For example, our analysis implies that even in the absence of density-dependent feedbacks, ideal-free dispersers occupy multiple patches in spatially heterogeneous environments provided environmental fluctuations are sufficiently strong and sufficiently weakly correlated across space. In contrast, for diffusively dispersing populations living in similar environments, intermediate dispersal rates maximize their stochastic growth rate.

Keywords Stochastic population growth · Spatial and temporal heterogeneity · Dominant Lyapunov exponent · Ideal free movement · Evolution of dispersal · Single large or several small debate · Habitat fragmentation

Mathematics Subject Classification 92D25 · 37H15 · 60H10

1 Introduction

Environmental conditions (e.g. light, precipitation, nutrient availability) vary in space and time. Since these conditions influence survivorship and fecundity of an organism, all organisms whether they be plants, animals, or viruses are faced with a fundamental quandary of “Should I stay or should I go?” On the one hand, if individuals disperse in a spatially heterogeneous environment, then they may arrive in locations with poorer environmental conditions. On the other hand, if individuals do not disperse, then they may fare poorly due to temporal fluctuations in local environmental conditions. The consequences of this interaction between dispersal and environmental heterogeneity for population growth has been studied extensively from theoretical, experimental, and applied perspectives (Hastings 1983; Petchey et al. 1997; Lundberg et al. 2000; Gonzalez and Holt 2002; Schmidt 2004; Roy et al. 2005; Boyce et al. 2006; Matthews and Gonzalez 2007; Schreiber 2010; Durrett and Remenik 2012). Here, we provide a mathematically rigorous perspective on these interactive effects using spatially explicit models of stochastic population growth.

Population growth is inherently stochastic due to numerous unpredictable causes. For a single, unstructured population with overlapping generations, the simplest model accounting for these fluctuations is a linear stochastic differential equation of the form

$$dZ_t = \mu Z_t dt + \sigma Z_t dB_t, \quad (1)$$

where Z_t is the population abundance at time t , μ is the mean per-capita growth rate (that is, $\mathbb{E}[Z_{t+\Delta t} - Z_t | Z_t = z] \approx z\mu\Delta t$), σ^2 is the “infinitesimal” variance of fluctuations in the per-capita growth rate (that is, $\mathbb{E}[(Z_{t+\Delta t} - Z_t - z\mu\Delta t)^2 | Z_t = z] \approx z^2\sigma^2\Delta t$), and B_t is a standard Brownian motion. Equivalently, the log population abundance $\log Z_t$ is normally distributed with mean $\log Z_0 + (\mu - \sigma^2/2)t$ and variance $\sigma^2 t$. Hence, even if the mean per-capita growth rate μ is positive these populations decline exponentially towards extinction when $\sigma^2/2 > \mu$ due to the predominance of the stochastic fluctuations. Despite its simplicity, the model (1) is used extensively for projecting future population sizes and estimating extinction risk (Dennis et al. 1991; Foley 1994; Lande et al. 2003). For example, Dennis et al. (1991) estimated μ and σ for six endangered species. These estimates provided a favorable outlook for the continued recovery of the Whooping Crane (i.e. $\mu \gg \sigma^2/2$), but unfavorable prospects for the Yellowstone Grizzly Bear.

Individuals cannot avoid being subject to temporal heterogeneity, but it is only when they disperse that they are affected by spatial variation in the environment. The effect of spatial heterogeneity on population growth depends, intuitively, on how individuals respond to environmental cues (Hastings 1983; Cantrell and Cosner 1991; Dockery et al. 1998; Chesson 2000; Cantrell et al. 2006; Kirkland et al. 2006; Schreiber and Lloyd-Smith 2009). When movement is towards regions with superior habitat quality, the presence of spatial heterogeneity increases the rate of population growth (Chesson 2000; Schreiber and Lloyd-Smith 2009). The most extreme form of this phenomenon occurs when individuals are able to disperse freely and ideally; that is, they can move instantly to the locations that maximize their per-capita growth rate (Fretwell and Lucas 1970; Cantrell et al. 2007). Anthropogenically altered habitats, however, can cause a disassociation between cues used by organisms to assess habitat quality and the actual habitat quality. This disassociation can result in negative associations between movement patterns and habitat quality and a corresponding reduction in the rate of population growth (Remeš 2000; Delibes et al. 2001; Schreiber and Lloyd-Smith 2009). For “random diffusive movement” (that is, no association between movement patterns and habitat quality), spatial heterogeneity increases population growth rates due to the influence of patches of higher quality. However, this boost in growth rate is most potent for sedentary populations (Hastings 1983; Dockery et al. 1998; Kirkland et al. 2006; Schreiber and Saltzman 2009). This dilutionary effect of dispersal on population growth was observed in the invasion of a woody weed, *Mimosa pigra*, into the wetlands of tropical Australia (Lonsdale 1993). A relatively fast disperser, this weed had a population doubling time of 1.2 years on favorable patches, but it exhibited much slower growth at the regional scale (doubling time of 6.7 years) due to the separation of suitable wetland habitats by unsuitable eucalyptus savannas.

Despite these substantial analytic advances in understanding separately the effects of spatial and temporal heterogeneity on population growth, there are few analytic

studies that consider the combined effects. For well-mixed populations with non-overlapping generations living in patchy environments, Metz et al. (1983) showed that population growth is determined by the geometric mean in time of the spatially (arithmetically) averaged per-capita growth rates. A surprising consequence of this expression is that populations coupled by dispersal can persist even though they are extinction prone in every patch (Jansen and Yoshimura 1998). This “rescue effect”, however, only occurs when spatial correlations are sufficiently weak (Harrison and Quinn 1989). Schreiber (2010) extended these results by deriving an analytic approximation for stochastic growth rates for partially mixing populations. This approximation reveals that positive temporal correlations can inflate population growth rates at intermediate dispersal rates, a conclusion consistent with simulation and empirical studies (Roy et al. 2005; Matthews and Gonzalez 2007). For example, Matthews and Gonzalez (2007) manipulated metapopulations of *Paramecium aurelia* by varying spatial-temporal patterns of temperature. In spatially uncorrelated environments, the populations coupled by dispersal always persisted for the duration of the experiment, while some of the uncoupled populations went extinct. Moreover, metapopulations experiencing positive temporal correlations exhibited higher growth rates than metapopulations living in temporally uncorrelated environments.

Here, we introduce and analyze stochastic models of populations that continuously experience uncertainty in time and space. For these models, our analysis answers some fundamental questions in population biology such as:

- How is the long-term spatial distribution of a population related to its rate of growth?
- When are population growth rates maximized at low, high, or intermediate dispersal rates for populations exhibiting diffusive movement?
- What is ideal free movement for individuals constantly facing uncertainty about local environmental conditions?
- To what extent do spatial correlations in temporal fluctuations hamper population persistence?
- How do multiple spatial scales of environmental heterogeneity influence population persistence?

In Sect. 2 we introduce our model for population growth in a patchy environment. It describes temporal fluctuations in the qualities of the various patches using multivariate Brownian motions with correlated components.

In Sect. 3, we first consider the vector-valued stochastic process given by the proportions of the population in each patch. These proportions converge in distribution to a (random) equilibrium at large times. The probability that this equilibrium spatial distribution is in some given subset of the set of possible patch proportions is just the long-term average amount of time that the process spends in that subset. We derive a simple expression for the stochastic growth of the population in terms of the first and second moments of this equilibrium spatial distribution. We also show that this equilibrium spatial distribution is characterized by a solution of a PDE that we solve in the case of two patches and use to examine how the equilibrium spatial distribution depends on the dispersal mechanism. We then present some numerical simulations to

give a first indication of the interesting range of phenomena that can occur when there is spatial heterogeneity in per-capita growth rates and biased movement between patches.

We use the results from Sect. 3 in Sect. 4 to investigate ideal free dispersal in stochastic environments. That is, we determine which forms of dispersal maximize the stochastic growth rate for given mean per-capita growth rates in each of the patches and given infinitesimal covariances for their temporal fluctuations.

We consider the effect of constraints on dispersal in Sect. 5. We suppose that the dispersal rates are fixed up to a scalar multiple δ and establish an analytic approximation for the stochastic growth rate of the form $a + b/\delta$ for large δ . We use this approximation to give criteria for whether low, intermediate, or high dispersal rates maximize the stochastic growth rate. In particular, we combine this analysis with tools from group representation theory to obtain results on the stochastic growth rate for environments with multiple spatial scales.

We discuss how our results relate to existing literature in Sect. 6. We end with a collection of Appendices where, for the sake of streamlining the presentation of our results in the remainder of the paper, we collect most of the proofs.

2 The model

We consider a population with overlapping generations living in a spatially heterogeneous environment consisting of n distinct patches and suppose that the per-capita growth rates within each patch are determined by a mixture of deterministic and stochastic environmental inputs. Let X_t^i denote the abundance of the population in the i th patch at time t and write $\mathbf{X}_t = (X_t^1, \dots, X_t^n)^T$ for the resulting column vector (we will use the superscript T throughout to denote the transpose of a vector or a matrix). If there was no dispersal between patches, it is appropriate to model \mathbf{X} as a Markov process with the following specifications for Δt small:

$$\mathbb{E}[X_{t+\Delta t}^i - X_t^i \mid \mathbf{X}_t = x] \approx \mu_i x^i \Delta t,$$

where μ_i is the mean per-capita growth rate in patch i , and

$$\text{Cov}[X_{t+\Delta t}^i - X_t^i, X_{t+\Delta t}^j - X_t^j \mid \mathbf{X}_t = x] \approx \sigma_{ij} x^i x^j \Delta t,$$

where $\Sigma = (\sigma_{ij})$ is a covariance matrix that captures the spatial dependence between the temporal fluctuations in patch quality. More formally, we consider the system of stochastic differential equations of the form

$$dX_t^i = X_t^i(\mu_i dt + dE_t^i),$$

where $\mathbf{E}_t = \Gamma^T \mathbf{B}_t$, Γ is an $n \times n$ matrix such that $\Gamma^T \Gamma = \Sigma$, and $\mathbf{B}_t = (B_t^1, \dots, B_t^n)^T$, $t \geq 0$, is a vector of independent standard Brownian motions.

In order to incorporate dispersal that couples the dynamics between patches, let $D_{ij} \geq 0$ for $j \neq i$ be the per-capita rate at which the population in patch i disperses to patch j . Define $-D_{ii} := \sum_{j \neq i} D_{ij}$ to be the total per-capita immigration rate out

of patch i . The resulting matrix D has zero row sums and non-negative off-diagonal entries. We call such matrices *dispersal matrices*. It is worth noting that any dispersal matrix D can be viewed as a generator of a continuous time Markov chain; that is, if we write $P_t := \exp(tD)$ for $t \geq 0$, so that $P_t, t \geq 0$, solves the matrix-valued ODE

$$\frac{d}{dt} P_t = P_t D,$$

then the matrix P_t has nonnegative entries, its rows sum to one, and the Chapman–Kolmogorov relations $P_s P_t = P_{s+t}$ hold for all $s, t \geq 0$. The (i, j) th entry of P_t gives the proportion of the population that was originally in patch i at time 0 but has dispersed to patch j at time t .

Adding dispersal to the regional dynamics leads to the system of stochastic differential equations

$$dX_t^i = X_t^i(\mu_i dt + dE_t^i) + \sum_{j=1}^n D_{ji} X_t^j dt. \tag{2}$$

We can write this system more compactly as the vector-valued stochastic differential equation

$$\begin{aligned} d\mathbf{X}_t &= \text{diag}(\mathbf{X}_t) (\mu dt + d\mathbf{E}_t) + D^T \mathbf{X}_t dt \\ &= \text{diag}(\mathbf{X}_t) (\mu dt + \Gamma^T d\mathbf{B}_t) + D^T \mathbf{X}_t dt, \end{aligned} \tag{3}$$

where $\mu := (\mu_1, \dots, \mu_n)^T$, and, given a vector u , we write $\text{diag}(u)$ for the diagonal matrix that has the entries of u along the diagonal.

We implicitly assume in the above set-up that all dispersing individuals arrive in some patch on the landscape. To account for dispersal induced mortality, we can add fictitious patches in which dispersing individuals enter and experience a mortality rate before dispersing to their final destination.

Also, our model does not include density-dependent effects on population growth. However, one can view it as a linearization of a density-dependent model about the extinction equilibrium $(0, \dots, 0)^T$ and, therefore, (3) determines how the population grows when abundances are low. Moreover, for discrete-time analogues of our model, positive population growth for this linearization implies persistence in the sense that there exists a unique positive stationary distribution for corresponding models with compensating density-dependence (Benaïm and Schreiber 2009). We conjecture that the same conclusion holds for our continuous time model.

From now on we assume that the dispersal matrix D is *irreducible* (that is, that it can not be put into block upper-triangular form by a re-labeling of the patches). This is equivalent to assuming that the entries of the matrix $P_t = \exp(tD)$ are strictly positive for all $t > 0$, and so it is possible to disperse between any two patches. Also, we will assume that the covariance matrix Σ has full rank (that is, that it is non-singular). This assumption implies that the randomness in the temporal fluctuations is genuinely n -dimensional.

3 The stable patch distribution and stochastic growth rate

3.1 Stable patch distribution

The key to understanding the asymptotic stochastic growth rate of the population is to first examine the dynamics of the spatial distribution of the population. Let $S_t := X_t^1 + \dots + X_t^n$ denote the total population abundance at time t and write $Y_t^i := X_t^i/S_t$ for the proportion of the total population that is in patch i . Set $\mathbf{Y}_t := (Y_t^1, \dots, Y_t^n)^T$. The stochastic process \mathbf{Y} takes values in the probability simplex $\Delta := \{y \in \mathbb{R}^n : \sum_i y_i = 1, y_i \geq 0\}$.

The following proposition, proved in Appendix A, shows that the stochastic process \mathbf{Y} is autonomously Markov; that is, that its evolution dynamics are governed by a stochastic differential equation that does not involve the total population size. Moreover, it says that the law of the random vector \mathbf{Y}_t converges to a unique equilibrium as $t \rightarrow \infty$. Recall, the law of a random vector $\mathbf{Y} \in \mathbb{R}^n$ is the probability measure $\mu_{\mathbf{Y}}$ on \mathbb{R}^n defined by $\mu_{\mathbf{Y}}(A) = \mathbb{P}\{\mathbf{Y} \in A\}$ for all Borel sets $A \subseteq \mathbb{R}^n$. Moreover, for any $\mu_{\mathbf{Y}}$ -integrable function $h : \mathbb{R}^n \rightarrow \mathbb{R}$, the expectation of $h(\mathbf{Y})$ is defined by

$$\mathbb{E}[h(\mathbf{Y})] = \int h(y) \mu_{\mathbf{Y}}(dy).$$

A sequence of random vectors $\mathbf{Y}_1, \mathbf{Y}_2, \dots$ converges in law to a random vector \mathbf{Y}_∞ if

$$\lim_{n \rightarrow \infty} \mathbb{E}[h(\mathbf{Y}_n)] = \mathbb{E}[h(\mathbf{Y}_\infty)]$$

for every continuous, bounded function $h : \mathbb{R}^n \rightarrow \mathbb{R}$. Convergence in law of a sequence of random vectors is also called convergence in distribution of the random vectors and is equivalent to weak convergence of their laws.

Proposition 1 *Suppose that $\mathbf{X}_0 \neq 0$. Then, the stochastic process \mathbf{Y} satisfies the stochastic differential equation*

$$d\mathbf{Y}_t = \left(\text{diag}(\mathbf{Y}_t) - \mathbf{Y}_t \mathbf{Y}_t^T\right) \Gamma^T d\mathbf{B}_t + D^T \mathbf{Y}_t dt + \left(\text{diag}(\mathbf{Y}_t) - \mathbf{Y}_t \mathbf{Y}_t^T\right) (\mu - \Sigma \mathbf{Y}_t) dt. \tag{4}$$

Moreover, there exists a random variable \mathbf{Y}_∞ taking values in the probability simplex Δ such that \mathbf{Y}_t converges in law to \mathbf{Y}_∞ as $t \rightarrow \infty$ and such that the empirical measure $\Pi_t := \frac{1}{t} \int_0^t \delta_{\mathbf{Y}_s} ds$ converges almost surely to the law of \mathbf{Y}_∞ as $t \rightarrow \infty$. The law of \mathbf{Y}_∞ does not depend on X_0 .

The empirical probability measure Π_t appearing in Proposition 1 describes the proportions of the time interval $[0, t]$ that the process \mathbf{Y} spends in the various subsets of its state space Δ . Namely, for a Borel set $A \subseteq \Delta$ of patch occupancy states, $\Pi_t(A)$ equals the fraction of time spent in these states over the time interval $[0, t]$. For example, if $A = \{y \in \Delta : y_1 > 1/2\}$, then $\Pi_t(A)$ equals the fraction of time for which at least 50% of the population is in patch 1 during the time interval $[0, t]$.

3.2 Stochastic growth rates

Recall that $S_t = X_t^1 + \dots + X_t^n$ is the total population size at time t . That is, $S_t = \mathbf{1}^T \mathbf{X}_t$, where $\mathbf{1} = (1, \dots, 1)^T$. Because $D\mathbf{1} = 0$, it follows from (3) that

$$dS_t = \mathbf{X}_t^T \Gamma^T d\mathbf{B}_t + \mu^T \mathbf{X}_t dt = S_t \mathbf{Y}_t^T \Gamma^T d\mathbf{B}_t + S_t \mu^T \mathbf{Y}_t dt.$$

Therefore, by Itô’s lemma (Gardiner 2004),

$$\log S_t = S_0 + \int_0^t \mathbf{Y}_t^T \Gamma^T d\mathbf{B}_t + \int_0^t \mu^T \mathbf{Y}_t dt - \frac{1}{2} \int_0^t \mathbf{Y}_t^T \Gamma^T \Gamma \mathbf{Y}_t dt.$$

Dividing by t , taking the limit as $t \rightarrow \infty$, and applying Proposition 1 yields the following result.

Theorem 1 *Suppose that $\mathbf{X}_0 \neq 0$. Then,*

$$\chi := \lim_{t \rightarrow \infty} t^{-1} \log S_t = \mu^T \mathbb{E}[\mathbf{Y}_\infty] - \frac{1}{2} \mathbb{E} \left[\mathbf{Y}_\infty^T \Sigma \mathbf{Y}_\infty \right] \quad \text{almost surely,} \quad (5)$$

where \mathbf{Y}_∞ is described in Proposition 1.

The limit χ in (5) is generally known as the *Lyapunov exponent* for the Markov process \mathbf{X} . Following Tuljapurkar (1990), we also call χ the *stochastic growth rate* of the population, as it describes the asymptotic growth rate of the population in the presence of stochasticity. To interpret (5), notice that

$$\langle \mu \rangle := \mu^T \mathbb{E}[\mathbf{Y}_\infty] = \sum_i \mu_i \mathbb{E}[Y_\infty^i] = \lim_{t \rightarrow \infty} \sum_i \mu_i \mathbb{E}[Y_t^i] \quad (6)$$

corresponds to weighted average of the per-capita growth rates with respect to the long-term spatial distribution \mathbf{Y}_∞ of the population. To interpret the other component of (5), let $\text{Var}[X]$ denote the variance of a random variable X . Since $\sum_i Y_t^i (E_{t+\Delta t}^i - E_t^i)$ for small $\Delta t > 0$ is approximately the average environmental change experienced by the population over time interval $[t, t + \Delta]$,

$$\begin{aligned} \langle \sigma^2 \rangle &= \mathbb{E} \left[\mathbf{Y}_\infty^T \Sigma \mathbf{Y}_\infty \right] = \lim_{t \rightarrow \infty} \frac{1}{\Delta t} \text{Var} \left[\mathbf{Y}_t^T (\mathbf{E}_{t+\Delta t} - \mathbf{E}_t) \right] \\ &= \lim_{t \rightarrow \infty} \frac{1}{\Delta t} \text{Var} \left[\sum_i Y_t^i (E_{t+\Delta t}^i - E_t^i) \right] \quad \text{for any } \Delta t > 0 \end{aligned} \quad (7)$$

corresponds to the infinitesimal variance of the environmental fluctuations weighted by the long-term spatial distribution.

Biological interpretation of Theorem 1 *The stochastic growth rate $\langle \mu \rangle - \langle \sigma^2 \rangle / 2$ for a spatially structured population is just what we see for an unstructured population*

where $\langle \mu \rangle$ and $\langle \sigma^2 \rangle$ are the per-capita growth rate and the infinitesimal covariances of the temporal fluctuations averaged appropriately with respect to the equilibrium spatial distribution. Hence, as in a spatially homogeneous environment, environmental fluctuations reduce the population growth rate. However, as we show in greater detail below, interactions between dispersal patterns, spatial heterogeneity and environmental fluctuations may increase the stochastic growth rate by increasing $\langle \mu \rangle$ or decreasing $\langle \sigma^2 \rangle$.

To get a more explicit expression for the stochastic growth rate, we need to determine the distribution of the equilibrium \mathbf{Y}_∞ , or at least find its first and second moments. This problem reduces to solving for the time-invariant solution of the Fokker–Planck equations with appropriate boundary conditions (Gardiner 2004), namely, the density $\rho : \Delta \rightarrow [0, \infty)$ of \mathbf{Y}_∞ satisfies

$$-\sum_i \frac{\partial}{\partial y_i} M_i(y) \rho(y) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j} V_{ij}(y) \rho(y) = 0 \quad \text{for } y \in \Delta, \tag{8}$$

where M_i and V_{ij} are the entries of

$$M(y) = D^T y + \left(\text{diag}(y) - yy^T \right) (\mu - \Sigma y) \text{ and}$$

$$V(y) = \left(\text{diag}(y) - yy^T \right) \Gamma^T \Gamma \left(\text{diag}(y) - yy^T \right),$$

respectively, and ρ is constrained to have $\int_\Delta \rho(y) dy = 1$. However, the PDE (8) needs to be supplemented with appropriate boundary conditions. In principle, these are found by characterizing the domain of the infinitesimal generator of the Feller diffusion process Y and thence characterizing the domain of the adjoint of this operator (Khas'minskii 1960; Bhattacharya 1978; Bogachev et al. 2002, 2009). This appears to be a quite difficult problem. However, in the case of two patches, the problem simplifies to solving an ODE on the unit interval.

Example 1 (Stochastic growth in two patch environments) Assume there are two patches. For simplicity, suppose there are no environmental correlations between the patches; that is, that $\sigma_{ii} = \sigma_i^2$ and $\sigma_{ij} = 0$ for $i \neq j$. Proposition 1 gives that $Y_t^1 = X_t^1 / (X_t^1 + X_t^2)$ satisfies the one-dimensional stochastic differential equation

$$dY_t^1 = M_*(Y_t^1) dt + \sqrt{V_*(Y_t^1)} dB_t$$

where

$$M_*(y) := y(1 - y)(\mu_1 - \mu_2 - \sigma_1^2 y + \sigma_2^2(1 - y)) - D_{12}y + D_{21}(1 - y)$$

and

$$V_*(y) := y^2(1 - y)^2(\sigma_1^2 + \sigma_2^2).$$

We can then apply standard tools for one-dimensional diffusions (Gardiner 2004) (checking that the boundaries at 0 and 1 are “entrance”, and hence inaccessible) to find that the density $\rho(x) : [0, 1] \rightarrow [0, \infty)$ of Y_∞^1 is given by

$$\begin{aligned} \rho(y) &= \frac{C_1}{V_*(y)} \exp\left(2 \int \frac{M_*(y)}{V_*(y)} dy\right) \\ &= \frac{C_2}{y^2(1-y)^2} \exp\left(\frac{2}{\sigma_1^2 + \sigma_2^2} \int \frac{\mu_1 - \mu_2}{y(1-y)} - \frac{\sigma_1^2}{1-y} + \frac{\sigma_2^2}{y} - \frac{D_{12}}{y(1-y)^2} + \frac{D_{21}}{y^2(1-y)} dy\right) \\ &= C_3 y^{\beta - \alpha_1} (1-y)^{-\beta - \alpha_2} \exp\left(-\frac{2}{\sigma_1^2 + \sigma_2^2} \left(\frac{D_{21}}{y} + \frac{D_{12}}{1-y}\right)\right), \end{aligned}$$

where the C_i are normalization constants, and

$$\begin{aligned} \alpha_i &:= \frac{2\sigma_i^2}{\sigma_1^2 + \sigma_2^2} \\ \beta &:= \frac{2}{\sigma_1^2 + \sigma_2^2} (\mu_1 - \mu_2 + D_{21} - D_{12}). \end{aligned}$$

Using this expression in (5), we get the following explicit expression for the stochastic growth rate

$$\begin{aligned} \chi &= \mu_1 \int_0^1 y\rho(y) dy + \mu_2 \int_0^1 (1-y)\rho(y) dy - \frac{\sigma_1^2}{2} \int_0^1 y^2\rho(y) dy \\ &\quad - \frac{\sigma_2^2}{2} \int_0^1 (1-y)^2\rho(y) dy \\ &= \mu_2 - \frac{\sigma_2^2}{2} + (\mu_1 - \mu_2 + \sigma_2^2) \int_0^1 y\rho(y) dy - \frac{\sigma_1^2 + \sigma_2^2}{2} \int_0^1 y^2\rho(y) dy. \end{aligned}$$

Despite its apparent complexity, this formula provides insights into how dispersal may influence population growth. For example, consider a population dispersing diffusively between statistically similar but uncorrelated patches (that is, $D_{12} = D_{21} = \delta/2$, $\mu_1 = \mu_2 = \mu$, and $\sigma_1 = \sigma_2 = \sigma$). We claim that the stochastic growth rate χ is an increasing function of the dispersal rate δ . Intuitively, this occurs because increasing δ decreases the variance of the random variable Y_∞ but has no effect on its expectation.

To verify our claim that χ is increasing with δ , write $\rho(\cdot; \delta)$ for the density of Y_∞^1 to emphasize its dependence on δ and notice that in this case

$$\rho(y; \delta) = \frac{1}{C(\delta)} y^{-1} (1-y)^{-1} \exp\left(-\frac{\delta}{2\sigma^2 y(1-y)}\right), \quad y \in (0, 1),$$

where $C(\delta) = \int_0^1 y^{-1}(1 - y)^{-1} \exp\left(-\frac{\delta}{2\sigma^2 y(1-y)}\right) dy$ is the normalization constant and

$$\chi(\delta) = \mu - \sigma^2/2 + \sigma^2 \int_0^1 y(1 - y) \rho(y; \delta) dy. \tag{9}$$

It suffices to show that

$$\begin{aligned} \int_0^1 y(1 - y) \rho(y; 2\delta\sigma^2) dy &= \frac{\int_0^1 \exp\left(-\frac{\delta}{y(1-y)}\right) dy}{C(2\delta\sigma^2)} \\ &= \frac{\int_0^1 \exp\left(-\frac{\delta}{y(1-y)}\right) dy}{\int_0^1 y^{-1}(1 - y)^{-1} \exp\left(-\frac{\delta}{y(1-y)}\right) dy} \end{aligned}$$

is an increasing function of $\delta > 0$. Differentiating with respect to δ and carrying the differentiation inside the integral sign, we obtain

$$\begin{aligned} C(2\sigma^2\delta)^{-2} \times &\left[\int_0^1 y^{-2}(1 - y)^{-2} \exp\left(-\frac{\delta}{y(1-y)}\right) dy \times \int_0^1 \exp\left(-\frac{\delta}{y(1-y)}\right) dy \right. \\ &\left. - \left(\int_0^1 y^{-1}(1 - y)^{-1} \exp\left(-\frac{\delta}{y(1-y)}\right) dy \right)^2 \right]. \end{aligned}$$

This quantity is the variance of the random variable $(Y_\infty^1(1 - Y_\infty^1))^{-1}$ and is thus nonnegative.

For the purpose of comparison with general asymptotic approximations that we develop later, we note that after a change of variable

$$\frac{\int_0^1 \exp\left(-\frac{\delta}{2\sigma^2 y(1-y)}\right) dy}{\int_0^1 y^{-1}(1 - y)^{-1} \exp\left(-\frac{\delta}{2\sigma^2 y(1-y)}\right) dy} = \frac{\int_0^\infty e^{-z} z^{-\frac{1}{2}} \left(\frac{2\sigma^2 z}{\delta} + 4\right)^{-\frac{3}{2}} dz}{\int_0^\infty e^{-z} z^{-\frac{1}{2}} \left(\frac{2\sigma^2 z}{\delta} + 4\right)^{-\frac{1}{2}} dz}.$$

Upon expanding the two functions $w \mapsto (w + 4)^{-\frac{1}{2}}$ and $w \mapsto (w + 4)^{-\frac{3}{2}}$ in Taylor series around 0 and integrating, we find that the ratio of integrals is of the form

$$\frac{1}{4} - \frac{1}{\delta} \frac{\sigma^2}{16} + O\left(\frac{1}{\delta^2}\right)$$

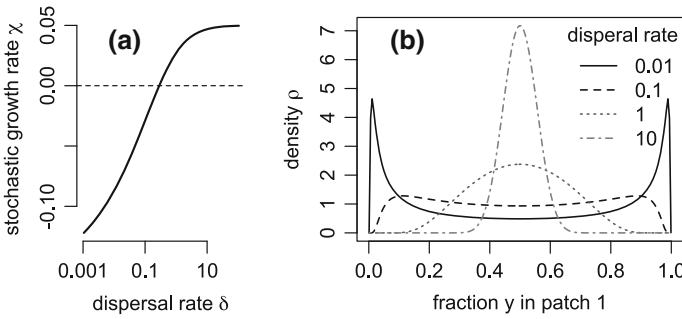


Fig. 1 Spatial distribution and population growth in a two patch environment. **a** The stochastic growth rate χ is plotted as a function of the dispersal rate δ . **b** The stationary density of the fraction of individuals in patch 1 is plotted for different dispersal rates. Parameter values are $\mu_1 = \mu_2 = 0.3$, $\sigma_1 = \sigma_2 = 1$, and $D_{12} = D_{21} = \delta$

as $\delta \rightarrow \infty$, so that

$$\chi(\delta) \approx \mu - \frac{\sigma^2}{4} - \frac{1}{\delta} \frac{\sigma^4}{16} \tag{10}$$

as $\delta \rightarrow \infty$.

Approximation (10) implies, as we prove more generally in Proposition 2, that $\lim_{\delta \rightarrow \infty} \chi(\delta) = \mu - \sigma^2/4$.

Biological interpretation of Example 1 *Even if two patches are unable to sustain a population in the absence of dispersal, connecting the patches by dispersal can permit persistence. This phenomenon occurs only at intermediate levels of environmental stochasticity (i.e. $2\mu < \sigma^2 < 4\mu$). Moreover, when this phenomenon occurs, there is a critical dispersal threshold $\delta^* > 0$ such that the metapopulation decreases to extinction whenever its dispersal rate is too low (i.e. $\delta \leq \delta^*$) and persists otherwise (Fig. 1).*

Because there do not appear to be closed-form expressions for the law of the stable patch distribution \mathbf{Y}_∞ when there are more than two patches, we must seek other routes to understanding the stochastic growth rate in such cases. One approach would be to solve the PDE (8) numerically. A second approach would be to simulate the stochastic process \mathbf{Y} for long time intervals and derive approximate values for the first and second moments of the equilibrium distribution. To give an indication of the range of phenomena that can occur in even relatively simple systems where there is biased movement between patches, we adopt the even simpler approach of simulating the stochastic process \mathbf{X} directly for long time intervals to obtain an approximate value of the stochastic growth rate. We implemented the simulations in a manner similar to that of Talay (1991), and the R code used is provided as supplementary material.

Example 2 (Spatially heterogeneous environments with biased emigration) For these simulations, we consider a metapopulation with either $n = 8$ or $n = 40$ patches of

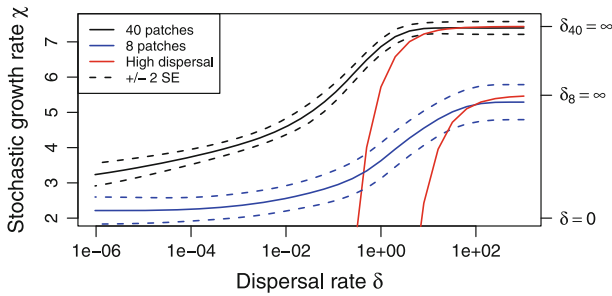


Fig. 2 The effect of dispersal rate δ on populations emigrating more rapidly out of lower quality patches than higher quality patches. Shown is the stochastic growth rate χ estimated from simulation of the SDE for 100 time units, across a range of values of δ , for both a 40-patch and a 8-patch model. Standard errors are estimated using the standard deviation of the stochastic growth rates across nonoverlapping time segments of a given simulation. Details of the dispersal matrix and parameter values are described in the main text. The right-hand axis shows asymptotic values for $\delta = 0$ and $\delta = \infty$, which are: $\chi(0) = \max_i \mu_i$ and $\chi(\infty) = \mu^T \pi - \frac{1}{2} \pi^T \Sigma \pi$ (Proposition 2). “High dispersal” shows the approximation of the form $\chi(\delta) \approx a + b/\delta$ for large δ calculated from formula (19) of Theorem 2

which one quarter are higher quality ($\mu_i = 10$ in these patches) and the remainder are lower quality ($\mu_i = 1$ in the remaining patches). All patches have the same level of spatially uncorrelated environmental noise ($\sigma_{ii} = 16$ for all i and $\sigma_{ij} = 0$ for $i \neq j$). When an organism exits a patch it chooses from the other patches with equal probability, but the emigration rate from a patch depends on the patch quality.

First, we consider the case in which emigration is “adaptive” in the sense that individuals emigrate more rapidly out of low quality patches than high quality patches:

$$D_{ij} = \begin{cases} \delta, & \text{for } i = 1, \dots, n/4 \text{ and } i \neq j, \\ 10\delta, & \text{for } i = n/4 + 1, \dots, n \text{ and } i \neq j. \end{cases}$$

Here, the parameter $\delta > 0$ scales the emigration rate, so that doubling δ doubles the emigration rate from all patches. As expected, since in this case dispersal is “adaptive”, Fig. 2 shows that stochastic growth rate $\chi = \chi(\delta)$ as a function of δ increases with δ . Moreover, Fig. 2 shows asymptotic values at $\delta = \infty$ for each case, and illustrates that the analytic approximation developed later in Theorem 2 works reasonably well for large values of δ . The Figure also shows extremely slow convergence as $\delta \rightarrow 0$ to $\chi(0) = \max_i \mu_i - (1/2)\sigma_i^2$ (note the logarithmic scale on the horizontal axis), indicating that although χ is continuous at $\delta = 0$ by Proposition 3 below, it may not be differentiable there.

Next we consider a case in which emigration is “maladaptive”, in the sense that individuals emigrate more rapidly out of high quality patches than out of low quality patches:

$$D_{ij} = \begin{cases} 10\delta, & \text{for } i = 1, \dots, n/4 \text{ and } i \neq j, \\ \delta, & \text{for } i = n/4 + 1, \dots, n \text{ and } i \neq j. \end{cases}$$

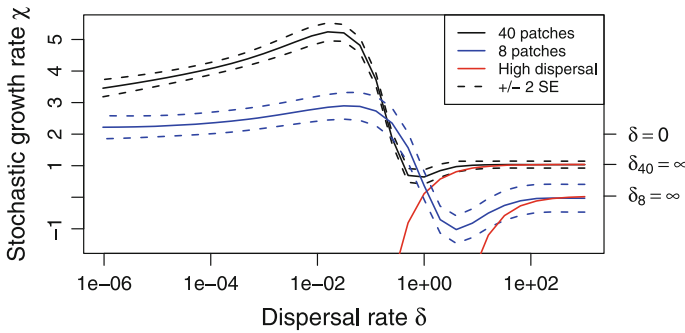


Fig. 3 The effect of dispersal rate δ on populations emigrating more rapidly out of higher quality patches than lower quality patches. Details are as in Fig. 2, but with different dispersal scheme; parameter values are described in the main text

It is possible to show using the results of Sect. 5 below that in this regime, high dispersal rates lead to a *lower* stochastic growth rate than sedentary populations (that is, $\lim_{\delta \rightarrow \infty} \chi(\delta)$ is dominated by $\lim_{\delta \rightarrow 0} \chi(\delta)$), and yet $\chi(\delta)$ increases with δ when δ is large. As illustrated in Fig. 3, the stochastic growth rate $\chi(\delta)$ exhibits a rather complex dependence on δ : increasing at low dispersal rates, declining at higher dispersal rates, and finally increasing again at the highest dispersal rates.

In a conservation framework, increasing δ corresponds to facilitating movement between patches by increasing the size or number of dispersal corridors between patches.

Biological interpretation of Example 2 *For populations exhibiting adaptive movement, increasing the size or number of dispersal corridors between patches enhances metapopulation growth rates. For populations exhibiting maladaptive movement, however, increasing dispersal rates can either increase or decrease metapopulation growth rates.*

4 Ideal free dispersal in a stochastic environment

A basic quandary in evolutionary ecology is, “For a given set of environmental conditions, what dispersal pattern maximizes fitness?” Since fitness in our context corresponds to the stochastic growth rate of the population, we can rephrase this question as, “Given μ and Σ , what form of the dispersal matrix D maximizes χ ?” Following Fretwell and Lucas (1970), we call such an optimal dispersal mechanism *ideal free dispersal* as individuals have no constraints on their dispersal (i.e. are “free”) and have complete knowledge about the distribution of spatial-temporal fluctuations (i.e. are “ideal”).

Equation (5) provides a means to answer this question. Because Σ has full rank, the function $y \mapsto \frac{1}{2}y^T \Sigma y$ is strictly convex, and so Jensen’s inequality implies that

$$\mathbb{E}[\mathbf{Y}_\infty^T \Sigma \mathbf{Y}_\infty] \geq \mathbb{E}[\mathbf{Y}_\infty]^T \Sigma \mathbb{E}[\mathbf{Y}_\infty],$$

with equality if and only if the random vector \mathbf{Y}_∞ is almost surely constant. Hence, to maximize the stochastic growth rate χ , we need to eliminate the variability in \mathbf{Y}_∞ , so that $\mathbf{Y}_\infty = y$ almost surely for a constant y that is chosen to maximize

$$\mu^T y - \frac{1}{2} y^T \Sigma y \tag{11}$$

subject to the constraint $y \in \Delta$. Under our standing non-degeneracy assumptions on D and Σ , the law of \mathbf{Y}_∞ is supported on all of Δ , and so we cannot actually achieve a situation in which \mathbf{Y}_∞ is a constant. However, the following result, which we prove in Appendix B, shows that we can approach this regime arbitrarily closely. Recall that the *stationary distribution* π for an irreducible dispersal matrix Q is a probability vector $\pi \in \Delta$ such that $\pi^T Q = 0$. We note that any vector π in the interior of Δ is the stationary distribution for some irreducible dispersal matrix Q . For example, given π , we can define $Q = \mathbf{1}\pi^T - I$ where I denotes the identity matrix.

Proposition 2 *Consider a vector π in the interior of Δ and an irreducible dispersal matrix Q that has π as its unique stationary distribution. Let $\mathbf{Y}_\infty(\delta)$ be the equilibrium patch distribution and $\chi(\delta)$ be the stochastic growth rate for (3) with $D = \delta Q$. Then $\mathbf{Y}_\infty(\delta)$ converges in law to the constant vector π as $\delta \rightarrow \infty$, and $\chi(\delta)$ converges to $\mu^T \pi - \frac{1}{2} \pi^T \Sigma \pi$ as $\delta \rightarrow \infty$.*

In the absence of population growth due to deterministic or stochastic effects, each of the dispersal matrices δQ in Proposition 2 sends the patch distribution to the vector π regardless of the initial conditions, and the speed at which this happens increases with δ , so that it becomes effectively instantaneous for large δ . Proposition 2 says that this push towards a deterministic equilibrium overcomes any disruptive effects introduced by population growth provided δ is sufficiently large, and so it is possible to produce random equilibrium patch distributions that are arbitrarily close to any given vector π in the interior of Δ . If we further approximate vectors π on the boundary of Δ by ones in the interior, we see that it is possible to produce equilibrium patch distributions that are arbitrarily close to any given vector in Δ .

Given that any patch distribution can be approximated arbitrary closely by the equilibrium patch distribution of a suitable population of rapidly dispersing individuals, the problem of optimizing χ reduces, as we have already noted, to maximizing the strictly concave function $g(y) = \mu^T y - \frac{1}{2} y^T \Sigma y$ over the compact, convex set Δ . This concavity implies there exists at most one local maximum. Denote this unique maximizer by $y^* = (y_1^*, \dots, y_n^*)^T$.

It is optimal for all individuals to remain in the single patch k (that is, $y_k^* = 1$) only if

$$\frac{\partial g}{\partial y_i}(e_k) - \frac{\partial g}{\partial y_k}(e_k) = \mu_i - \sigma_{ik} - \mu_k + \sigma_{kk} < 0 \quad \text{for all } i \neq k,$$

where e_k is the k th element of the standard basis of \mathbb{R}^n , or, equivalently,

$$\mu_k - \mu_i > \sigma_{kk} - \sigma_{ik} \quad \text{for all } i \neq k. \tag{12}$$

Biological interpretation of equation (12) *If the variances of environmental fluctuations are sufficiently large in all patches and the spatial covariances in these environmental fluctuations are sufficiently small, then ideal free dispersers occupy multiple patches.*

When it is optimal to disperse between several patches, we can solve for the optimal dispersal strategy y^* by using the method of Lagrange multipliers. Without loss of generality, assume that the optimal strategy y^* makes use of all patches, that is, that y^* is in the interior of Δ . Indeed, if the optimal strategy does not make use of all patches, then we can consider analogous problems on the faces of the convex polytope Δ of the form $\{y \in \Delta : y_i = 0, i \in A\}$, where A is a subset of $\{1, \dots, n\}$. Because

$$\nabla g(y) = \mu - \Sigma y \text{ and } \nabla \left(\sum_i y_i \right) = \mathbf{1},$$

the optimal y^* must satisfy

$$\mu - \Sigma y^* = \lambda \mathbf{1}, \tag{13}$$

where λ is a Lagrange multiplier. Notice that

$$(\Sigma y)_i = \frac{1}{\Delta t} \mathbb{E} \left[(E_{t+\Delta t}^i - E_t^i) \sum_j y_j (E_{t+\Delta t}^j - E_t^j) \right].$$

Hence, we get the following interpretation.

Biological interpretation of equation (13) *Ideal free populations using multiple patches are distributed across the patches in such a way that the differences between the mean per-capita growth rates and the covariances between the within patch noise and the noise experienced on average by an individual are equal in all occupied patches. In particular, the local stochastic growth rates $\mu_i - \sigma_{ii}/2$ need not be equal in all occupied patches.*

Now,

$$y^* = \Sigma^{-1}(\mu - \lambda \mathbf{1}), \tag{14}$$

and the constraint $\mathbf{1}^T y = 1$ yields

$$1 = \mathbf{1}^T \Sigma^{-1}(\mu - \lambda \mathbf{1}),$$

so that

$$\lambda = \frac{\mathbf{1}^T \Sigma^{-1} \mu - 1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \tag{15}$$

and

$$y^* = \Sigma^{-1} \left(\mu - \frac{\mathbf{1}^T \Sigma^{-1} \mu - 1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \mathbf{1} \right). \tag{16}$$

The right-hand side of equation (16) is the optimal vector y^* we seek, provided that it belongs to the interior of Δ . Otherwise, as we remarked above, we need to perform similar analyses on the faces of the simplex Δ .

To illustrate the utility of this formula, we examine two special cases: when the environmental noise between patches is uncorrelated, and when the patches experience the same individual levels of noise but they are spatially correlated.

Example 3 (Spatially uncorrelated environments) Suppose that there are no spatial correlations in the environmental noise, so that Σ is a diagonal matrix with diagonal entries $\sigma_{ii} = \sigma_i^2$. It follows from equation (16) that the ideal free patch distribution is

$$y_i^* = \frac{1}{\sigma_i^2 \sum_j 1/\sigma_j^2} \left[\sum_j \frac{\mu_i - \mu_j}{\sigma_j^2} + 1 \right], \tag{17}$$

provided that $\sum_j (\mu_j - \mu_i) / \sigma_j^2 < 1$ for all i .

Biological interpretation of equation (17) *In the absence of spatial correlations in environmental fluctuations, ideal free dispersers visit all patches whenever the environmental variation is sufficiently great relative to differences in the mean per-capita growth rates. In particular, if all mean per-capita growth rates are equal, then the fraction of individuals in a patch is inversely proportional to the variation in temporal fluctuations in the patch; that is, $y_i^* = (1/\sigma_i^2) / (\sum_j 1/\sigma_j^2)$.*

Example 4 (Spatially correlated environments) Suppose that the infinitesimal variance of the temporal fluctuations in each patch is σ^2 and that the correlation between the fluctuations in any pair of patches is ρ . Thus, $\Sigma = \sigma^2(1 - \rho)I + \sigma^2\rho J$, where $J = \mathbf{1}\mathbf{1}^T$ is the matrix in which every entry is 1. Provided that $-\frac{1}{n-1} < \rho < 1$, the matrix Σ is non-singular with inverse

$$\Sigma^{-1} = \frac{1}{(1 - \rho)\sigma^2} I - \frac{\rho}{(1 - \rho)(1 + (n - 1)\rho)\sigma^2} J.$$

Denoting by $\bar{\mu} = \frac{1}{n} \sum_i \mu_i$ the average across the patches of the mean per-capita growth rates, the optimal dispersal strategy is given by

$$y_i^* = \frac{\mu_i - \bar{\mu}}{\sigma^2(1 - \rho)} + \frac{1}{n} \tag{18}$$

provided that $y_*^i > 0$ for all i . Notice that (18) agrees with (17) when $\rho = 0$ and $\sigma_i = \sigma$.

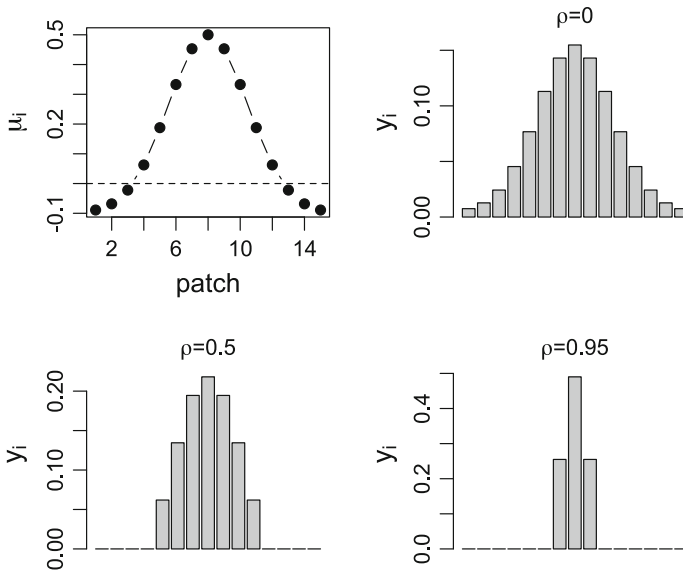


Fig. 4 Effects of spatial correlations on the ideal free patch distribution in a 15 patch environment. Per-capita growth rates μ_i are plotted in the top left. The ideal free patch distribution y^* is plotted at three levels of spatial correlation ρ . Covariances are $\sigma_{ii} = 2$ and $\sigma_{ij} = 2\rho$ for $i \neq j$

Biological interpretation of equation (18) *If environmental fluctuations have a sufficiently large variance σ^2 , then ideal free dispersers visit all patches and spend more time in patches that support higher mean per-capita growth rates. Increasing the common spatial correlation ρ results in ideal free dispersers spending more time in patches whose mean per-capita growth rate is greater than the average of the mean per-capita growth rates and less time in other patches (Fig. 4). When the spatial correlations are sufficiently large, it is no longer optimal to disperse to the patches with lower mean per-capita growth rates ($\rho = 0.5$ and $\rho = 0.95$ in Fig. 4).*

5 The effect of constraints on dispersal

While the ideal free patch distribution is a useful idealization to investigate how organisms should disperse in the absence of constraints, organisms in the natural world have limits on their ability to disperse and to collect and interpret environmental information. Recall from Sect. 4 that if the optimal patch distribution y^* for an ideal free disperser is in the interior of the probability simplex Δ , then, loosely speaking, the ideal free disperser achieves the maximal stochastic growth rate by using a strategy for which dispersal rate matrix is of the form $D = \delta Q$, where Q is any irreducible dispersal matrix with $(y^*)^T Q = 0$ and $\delta = \infty$. At the opposite extreme, if y^* assigns all of its mass to a single patch, then an ideal free disperser never leaves that single most-favored patch.

To get a better understanding of how constraints on dispersal influence population growth, we consider dispersal matrices of the form $D = \delta Q$, where $\delta \geq 0$ and Q is a fixed irreducible dispersal matrix Q with a stationary distribution π that is not

necessarily the optimal patch distribution for an ideal free disperser in the given environmental conditions. We write $\chi(\delta)$ for the stochastic growth rate of the population as a function of the dispersal parameter δ and ask which choice of δ maximizes $\chi(\delta)$. In particular, we are interested in conditions under which some intermediate $\delta > 0$ maximizes the stochastic growth rate $\chi(\delta)$.

We know from Proposition 2 that $\chi(\delta)$ approaches $\pi^T \mu - \frac{1}{2} \pi^T \Sigma \pi$ as $\delta \rightarrow \infty$. We therefore set $\chi(\infty) = \pi^T \mu - \frac{1}{2} \pi^T \Sigma \pi$. On the other hand, if there is no dispersal ($\delta = 0$), then $\lim_{t \rightarrow \infty} \frac{1}{t} \log X_t^i = \mu_i - \frac{\sigma_i^2}{2}$ with probability one whenever $X_0^i > 0$, and so $\lim_{t \rightarrow \infty} \frac{1}{t} \log S_t = \max_i \{ \mu_i - \frac{\sigma_i^2}{2} \}$ whenever $X_0^i > 0$ for all i . Hence, it is reasonable to set $\chi(0) = \max_i \{ \mu_i - \frac{\sigma_i^2}{2} \}$. The following result, which we prove in Appendix C, implies that the function $\delta \mapsto \chi(\delta)$ is continuous on $[0, \infty)$.

Proposition 3 *The function $\delta \mapsto \chi(\delta)$ is analytic on the interval $(0, \infty)$ and continuous at the point $\delta = 0$.*

One way to establish that $\chi(\delta)$ is maximized for an intermediate value of δ is to show that $\chi(0) < \chi(\infty)$ and that $\chi(\delta) > \chi(\infty)$ for all sufficiently large δ . The following theorem provides an asymptotic approximation for $\chi(\delta)$ when δ is large that allows us to check when the latter condition holds. We prove the theorem under the hypothesis that the dispersal matrix Q is *reversible* with respect to its stationary distribution π ; that is, that $\pi_i Q_{ij} = \pi_j Q_{ji}$ for all i, j . Reversibility implies that at stationarity the Markov chain defined by Q exhibits “balanced dispersal in the absence of local demography”. Namely, if a large number of individuals are independently executing the equilibrium movement dynamics, then the rate at which individuals move from patch i to patch j equals the rate at which individuals move from patch j to patch i . We note that diffusive movement (that is, the matrix Q is symmetric) and any form of movement along a one-dimensional landscape (that is, the matrix Q is tridiagonal) are examples of reversible Markov chains. We provide a proof of the theorem in Appendix D. Corollary 1 below, which we prove in Appendix E, provides a more readily computable expression for the asymptotics of the stochastic growth rate under further assumptions.

Theorem 2 *Suppose that Q is reversible with respect to its stationary distribution π . Then,*

$$\begin{aligned} \chi(\delta) = & \left(\mu^T \pi - \frac{1}{2} \pi^T \Sigma \pi \right) + \frac{1}{\delta} \left[(\mu - \Sigma \pi)^T v \right. \\ & - \frac{1}{2} \int_0^\infty \text{Tr} \left(\exp(Q^T s) \left(\text{diag}(\pi) - \pi \pi^T \right) \Sigma \left(\text{diag}(\pi) - \pi \pi^T \right) \right. \\ & \left. \left. \exp(Qs) \Sigma \right) ds \right] + O(\delta^{-\frac{5}{4}}) \end{aligned} \tag{19}$$

as $\delta \rightarrow \infty$, where v is the unique vector satisfying $\mathbf{1}^T v = 0$ and $Q^T v = -(\text{diag}(\pi) - \pi \pi^T)(\mu - \Sigma \pi)$.

When the dispersal matrix $D = \delta Q$ is consistent with ideal dispersal in the limit $\delta \rightarrow \infty$, Eq. (13) implies that $(\mu - \Sigma\pi)^T v = \lambda \mathbf{1}^T v = 0$. On the other hand, the proof of Theorem 2 shows that

$$\int_0^\infty \text{Tr} \left(\exp(Q^T s) \left(\text{diag}(\pi) - \pi\pi^T \right) \Sigma \left(\text{diag}(\pi) - \pi\pi^T \right) \exp(Qs) \Sigma \right) ds = \text{Tr} \left(\mathbb{E}[\mathbf{V}_\infty \mathbf{V}_\infty^T] \Sigma \right) > 0$$

where \mathbf{V}_∞ is a Gaussian random vector. Hence, as expected, $\chi(\delta)$ is an increasing function for large δ when π corresponds to the ideal free distribution associated with μ and Σ . However, when π does not correspond to the ideal free distribution, $\chi(\delta)$ may be increasing or decreasing for large δ as we illustrate below.

When Q and Σ commute, the asymptotic expression (19) for $\chi(\delta)$ simplifies a great deal.

Corollary 1 *Suppose that Q is symmetric and $Q\Sigma = \Sigma Q$. Let $\lambda_1 \leq \dots \leq \lambda_{n-1} < \lambda_n = 0$ be the eigenvalues of Q with corresponding orthonormal eigenvectors ξ_1, \dots, ξ_n . Then, the eigenvalues $\theta_1, \dots, \theta_n$ of Σ can be ordered so that $\Sigma\xi_k = \theta_k\xi_k$, for each $1 \leq k \leq n$, and the approximation (19) reduces to*

$$\chi(\delta) = \left(\bar{\mu} - \frac{1}{2n}\theta_n \right) - \frac{1}{\delta n} \left[\sum_{k=1}^{n-1} \frac{1}{\lambda_k} \left((\xi_k^T \mu)^2 - \frac{1}{4n}\theta_k^2 \right) \right] + O(\delta^{-5/4}) \quad (20)$$

as $\delta \rightarrow \infty$, where $\bar{\mu} = \frac{1}{n} \sum \mu_i$.

To illustrate the utility of this latter approximation, we develop more explicit formulas for three scenarios: diffusive movement in a landscape where all patches are equally connected (that is, a classic ‘‘Levins’’ style landscape (Levins 1969)), diffusive movement in a landscape consisting of a ring of patches, and diffusive movement in a landscape with multiple spatial scales (that is, a hierarchical Levins landscape).

Example 5 (Fully connected metapopulations with unbiased movement) Consider a population in which individuals disperse at the same per-capita rate δ/n between all pairs of patches. Let σ^2 be the variance of the within patch fluctuations and ρ be the correlation in these fluctuations between any pair of patches. Under these assumptions, the dispersal matrix is $Q = J/n - I$ and the environmental covariance matrix is $\Sigma = (1 - \rho)\sigma^2 I + \rho\sigma^2 J$, where recall that $J = \mathbf{1}\mathbf{1}^T$ is the matrix of all ones. Because Q is symmetric, the stationary distribution of Q is uniform; that is, $\pi_1 = \dots = \pi_n = \frac{1}{n}$. Hence, in the absence of population growth there would be equal numbers of individuals in each patch at large times.

Because the matrices I and J commute, the matrices Q and Σ also commute. Recall the notation of Corollary 1. The eigenvector ξ_n is $\frac{1}{\sqrt{n}}\mathbf{1}$. If ξ is any vector of length one orthogonal to ξ_n , then $J\xi = 0$, and so $Q\xi = -\xi$ and $\Sigma\xi = (1 - \rho)\sigma^2\xi$.

We may thus take ξ_1, \dots, ξ_{n-1} to be any orthonormal set of vectors orthogonal to ξ_n . Moreover, $\lambda_1 = \dots = \lambda_{n-1} = -1$ and $\theta_1 = \dots = \theta_{n-1} = (1 - \rho)\sigma^2$.

Now, $(\xi_n^T \mu)^2 = (1/n) (\sum_{k=1}^n \mu_k)^2 = n(\bar{\mu})^2$, and so Parseval's identity implies that $\sum_{k=1}^{n-1} (\xi_k^T \mu)^2 = \sum_{k=1}^n \mu_k^2 - n(\bar{\mu})^2 = \mu^T \mu - n(\bar{\mu})^2$. Denote the variance of the vector μ by

$$\text{Var}[\mu] = \frac{1}{n} \mu^T \mu - (\bar{\mu})^2 = \frac{1}{n} \sum_{k=1}^{n-1} (\xi_k^T \mu)^2.$$

Substituting these observations into equation (20), we get that

$$\chi(\delta) = \bar{\mu} - \frac{\sigma^2}{2n} (1 + (n - 1)\rho) + \frac{1}{\delta} \left[\text{Var}[\mu] - \frac{(n - 1)((1 - \rho)\sigma^2)^2}{4n^2} \right] + O(\delta^{-\frac{5}{4}}). \tag{21}$$

Recall that for the special case of two uncorrelated patches with $D_{12} = D_{21} = \delta/2$, $\mu_1 = \mu_2 = \mu$, and $\sigma_1 = \sigma_2 = \sigma$, we showed from our exact formula for $\chi(\delta)$ in the two patch case that

$$\chi(\delta) \approx \mu - \frac{\sigma^2}{4} - \frac{1}{\delta} \frac{\sigma^4}{16}$$

as $\delta \rightarrow \infty$, see (10). Hence, this approximation agrees with (21).

Approximation (21) implies that $\chi(\delta)$ is decreasing for large δ whenever

$$\frac{n}{\sqrt{n-1}} \sqrt{\text{Var}[\mu]} > \frac{(1 - \rho)\sigma^2}{2}, \tag{22}$$

and that $\chi(\delta)$ is increasing if the opposite inequality holds. We have remarked that, in general, an intermediate dispersal rate is optimal when $\chi(0) < \chi(\infty)$ and $\chi(\delta) > \chi(\infty)$ for all sufficiently large δ . This will occur for individuals in this diffusive dispersal regime when

$$\frac{(1 - \rho)\sigma^2}{2} > \frac{\max_i \mu_i - \bar{\mu}}{1 - 1/n} \tag{23}$$

and (22) holds. In particular, when there are many patches (that is, $n \rightarrow \infty$), inequalities (23) and (22) are both satisfied if

$$(1 - \rho)\sigma^2/2 > \max_i \mu_i - \bar{\mu} > 0.$$

Biological interpretation of equations (22) and (23) *Highly diffusive movement has a negative impact on population growth whenever there are sufficiently many*

patches and there is sufficient spatial variation in the mean per-capita growth rates. Alternatively, if there is no spatial variation in the mean per-capita rates and stochastic fluctuations are not perfectly correlated, then the population growth rate continually increases with higher dispersal rates. This latter observation is consistent with individuals being distributed equally across the landscape in the optimal patch distribution. In contrast, if there is some spatial variation in the mean per-capita growth rates and there are sufficiently large, but not perfectly correlated environmental fluctuations, then an intermediate dispersal rate maximizes the stochastic growth rate for diffusively dispersing populations.

In order to apply Corollary 1, we need to simultaneously diagonalize the matrices Q and Σ . A situation in which this is possible and the resulting formulas provide insight into biologically relevant scenarios is when the dispersal mechanism and the covariance structure of the noise both exhibit the symmetries of an underlying group. Example 5 above is a particular instance of this situation.

More specifically, we suppose that the patches can be labeled with the elements of a finite group G in such a way that the migration rate $Q_{g,h}$ and environmental covariance $\Sigma_{g,h}$ between patches g and h both only depend on the “displacement” gh^{-1} from g to h in G . That is, we assume there exist functions q and s on G such that $Q_{gh} = q(gh^{-1})$ and $\Sigma_{gh} = s(gh^{-1})$. For instance, if G is the group of integers modulo n , then the habitat has n patches arranged in a circle, and the dispersal rate and environmental covariance between two patches only depends on the distance between them, measured in steps around the circle. We do not require that the vector μ of mean per-capita growth rates satisfies any symmetry conditions.

The matrices Q and Σ will commute if q and s are *class functions*, that is, if $q(gh) = q(hg)$ and $s(gh) = s(hg)$ for all $g, h \in G$. We assume this condition holds from now on. Note that if G is Abelian (that is, the group operation is commutative), then any function is a class function.

5.1 Background on group representations

We now record a few facts about representation theory, the tool that will enable us to find the eigenvalues and eigenvectors of Q and Σ , resulting in Theorem 3. We refer readers interested in more detail to (Serre 1977; Diaconis 1988), while readers interested in less mathematical detail may skip directly to Examples 6 and 7 without loss of continuity.

A *unitary representation* of a group G is a homomorphism ρ from G into the group of $d_\rho \times d_\rho$ unitary matrices, where d_ρ is called the *degree* of the representation. Two representations ρ' and ρ'' are *equivalent* if there exists a unitary matrix U such that $\rho''(g) = U\rho'(g)U^{-1}$ for all $g \in G$. A representation ρ' is *irreducible* if it is not equivalent to some representation ρ'' for which $\rho''(g)$ is of the same block diagonal form for all $g \in G$. A finite group has a finite set of inequivalent, irreducible, unitary representations, which we denote by \hat{G} . The simplest representation is the *trivial representation* ρ_{tr} of degree one, for which $\rho_{\text{tr}}(g) = 1$ for all g .

For a simple example that we will return to, let $G = \mathbb{Z}_n$, the group of integers modulo n . Since \mathbb{Z}_n is Abelian, all the irreducible representations are one-dimensional ($d_\rho = 1$ for all $\rho \in \hat{G}$), and are of the form $\rho^{(m)}(j) = \exp(2\pi imj/n)$, so that $\hat{G} = \{\rho_m^{(m)} : 0 \leq m \leq n - 1\}$.

The matrix entries of irreducible representations are *orthogonal*: for $\rho', \rho'' \in \hat{G}$,

$$\sum_{g \in G} \rho'_{ij}(g)\rho''_{k\ell}(g)^* = \begin{cases} \frac{\#G}{d_\rho}, & \text{if } \rho' = \rho'' \text{ and } (i, j) = (k, \ell), \\ 0, & \text{otherwise,} \end{cases} \tag{24}$$

where z^* denotes the complex conjugate of a complex number z , and $\#G$ is the number of elements of G .

The *Fourier transform* of a function $f : G \rightarrow \mathbb{C}$ is a function \hat{f} on \hat{G} defined by

$$\hat{f}(\rho) := \sum_{g \in G} f(g)\rho(g) \quad \text{for } \rho \in \hat{G}. \tag{25}$$

Note that $\hat{f}(\rho)$ is a $d_\rho \times d_\rho$ matrix. It follows from the orthogonality properties of the matrix entries of the irreducible representations recorded above that the Fourier transform may be inverted, giving f explicitly as the linear combination of matrix entries of \hat{f} . The inversion formula is

$$f(g) = \frac{1}{\#G} \sum_{\rho \in \hat{G}} d_\rho \operatorname{Tr}(\rho(g^{-1})\hat{f}(\rho)).$$

For $G = \mathbb{Z}_n$, this is the familiar *discrete Fourier transform*, for which orthogonality of matrix entries is the fact that $(1/n) \sum_{j=0}^{n-1} \exp(2\pi ij(\ell - m)/n) = \delta_{\ell m}$. The transform is given by $\hat{f}(\rho^{(m)}) = \sum_{k=0}^{n-1} f(k) \exp(2\pi imk/n)$ for $0 \leq m \leq n - 1$, and $f(k) = (1/n) \sum_{m=0}^{n-1} \hat{f}(\rho^{(m)}) \exp(-2\pi imk/n)$. The trivial character is $\kappa_{\text{tr}} = \rho^{(0)}$.

Associated with a representation $\rho \in \hat{G}$ is its *character* κ , defined by $\kappa(g) := \operatorname{Tr} \rho(g)$. We write \tilde{G} for the set of characters of irreducible representations. The characters are class functions, and form an orthogonal basis for the subspace of class functions on G and all have the same norm: $\sum_{g \in G} |\kappa(g)|^2 = \#G$, where $|z| = \sqrt{zz^*}$ is the modulus of the complex number z . For $\rho \in \hat{G}$ with character $\kappa \in \tilde{G}$, the Fourier transform of a class function f satisfies

$$\hat{f}(\rho) = \frac{1}{d_\rho} \tilde{f}(\kappa)I$$

where I is the $d_\rho \times d_\rho$ identity matrix and

$$\tilde{f}(\kappa) := \sum_{g \in G} f(g)\kappa(g). \tag{26}$$

Consequently,

$$f(g) = \frac{1}{\#G} \sum_{\kappa \in \tilde{G}} \kappa(g)^* \tilde{f}(\kappa). \tag{27}$$

As noted above, if $G = \mathbb{Z}_n$ then all irreducible representations are one-dimensional, so in this case we may identify the characters with the irreducible representations, $\hat{G} = \tilde{G}$. Since \mathbb{Z}_n is Abelian, all functions on \mathbb{Z}_n are class functions, so that the two Fourier transforms (25) and (26) are equal.

Finally, given a function f on G and character κ , define

$$\|f\|_{\kappa}^2 := \frac{d_{\rho}}{\#G} \sum_{g,h \in G} \kappa(gh^{-1}) f(g) f(h)^*.$$

The following theorem is proved in Appendix F.

Theorem 3 *Suppose that the n patches are labeled by a finite group G in such a way that $Q_{gh} = q(gh^{-1})$ and $\Sigma_{gh} = s(gh^{-1})$, where q and s are class functions. Suppose further that $q(g) = q(g^{-1})$, $g \in G$, so that the matrix Q is symmetric. Let $\bar{\mu} = \frac{1}{\#G} \sum_{g \in G} \mu(g)$ and $\bar{s} = \frac{1}{\#G} \sum_{g \in G} s(g)$. Then,*

$$\chi(\delta) = \left(\bar{\mu} - \frac{1}{2} \bar{s} \right) - \frac{1}{\delta n} \sum_{\kappa \in \tilde{G} \setminus \{\kappa_{\text{tr}}\}} \frac{d_{\kappa}}{\tilde{q}(\kappa)} \left(\|\mu\|_{\kappa}^2 - \frac{1}{4n} \tilde{s}(\kappa)^2 \right) + O(\delta^{-5/4}) \tag{28}$$

as $\delta \rightarrow \infty$. Furthermore, $\tilde{q}(\kappa) < 0$ for all $\kappa \in \tilde{G} \setminus \{\kappa_{\text{tr}}\}$.

Roughly speaking, this expression tells us about the respective roles of variance of patch quality (μ) and covariance of environmental noise (s). The fact that $\tilde{q}(\kappa)$ is negative for all κ leads to the following.

Biological interpretation of equation (28) *If variability in patch quality at a certain scale is larger than the correlation in environmental noise at that scale, in a sense made precise above, then the stochastic growth rate decreases with increasing dispersal rates at that scale. Conversely, if environmental noise is strongly correlated between patches and the mean patch quality is similar, then more dispersal is expected to be better. The relevant sense of “at that scale” is in the sense of the Fourier transform, analogous to the “frequency domain” in Fourier analysis.*

Example 6 (Circle of Patches) Suppose that the n patches of a habitat are arranged in a circle and are labeled by $\mathbb{Z}_n = \{0, 1, \dots, n - 1\}$, the group of integers modulo n with identity element 0. As reviewed above, the Fourier transform is the familiar discrete Fourier transform.

If we assume that individuals disperse only to neighboring patches and these dispersal rates are equal, then $q(1) = q(n - 1) = 1/2$, $q(0) = -1$ and $q(2) = \dots = q(n - 2) = 0$. Assume the environmental noise is independent between patches and

has variance σ^2 i.e. $s(0) = \sigma^2$ and $0 = s(1) = \dots = s(n - 1)$. Finally, suppose that patch quality as measured by the average per-capita growth rates is spatially periodic, so that $\mu(k) = \bar{\mu} + c \cos(2\pi k\ell/n)$ for some $c > 0$, $\bar{\mu}$, and $1 \leq \ell < n/2$.

Under this set of assumptions, we can compute that for $m \neq 0$, $\tilde{q}(m) = \cos(2\pi m/n) - 1$ and $\tilde{s}(m) = \sigma^2$. Furthermore, $\|\mu\|_{\kappa_\ell}^2 = \|\mu\|_{\kappa_{n-\ell}}^2 = nc^2/4$ and $\|\mu\|_{\kappa_m}^2 = 0$ otherwise. From these computations, Theorem 3 implies that

$$\chi(\delta) \approx \bar{\mu} - \sigma^2/2 - \frac{1}{\delta n} \left(\frac{nc^2}{2(\cos(2\pi\ell/n) - 1)} - \sum_{m=1}^{n-1} \frac{\sigma^2}{4n(\cos(2\pi m/n) - 1)} \right)$$

for large δ . Using the identity $\sum_{k=1}^{n-1} (1 - \cos(2\pi k/n))^{-1} = (n^2 - 1)/6$ (see equation 1.381.1 in Gradshteyn and Ryzhik (2007)'s table of integrals and series), this approximation simplifies to

$$\chi(\delta) \approx \bar{\mu} - \sigma^2/2 + \frac{1}{4\delta n^2} \left(\frac{2n^2c^2}{1 - \cos(2\pi\ell/n)} - \frac{1}{6}(n^2 - 1)\sigma^4 \right). \tag{29}$$

Since $\chi(0) = \bar{\mu} + c - \sigma^2/2$, high dispersal is better than no dispersal if $\chi(\infty) - \chi(0) = \sigma^2(1 - 1/n)/2 - c > 0$. When the number of patches is sufficiently large, this inequality implies that highly dispersive populations grow faster than sedentary populations provided that the temporal variation is sufficiently greater than the spatial variation in per-capita growth rates i.e. $\sigma^2 > 2c$. On the other hand, $\chi(\delta)$ is decreasing for large δ if the coefficient of $1/\delta$ is positive i.e.

$$4c^2 > \frac{1}{3}(1 - \cos(2\pi\ell/n))(1 - n^{-2})\sigma^4.$$

Hence, if ℓ/n is small enough, then $\chi(\delta)$ is decreasing for large δ .

Biological interpretation of equation (29) *In a circular habitat with nearest-neighbor dispersal and sinusoidally varying patch quality, intermediate dispersal rates maximize the stochastic growth rate provided that spatial heterogeneity occurs on a short scale (i.e. ℓ/n sufficiently small) and temporal variability is sufficiently large.*

Example 7 (Multi-scale patches) Suppose now that our organism lives in a hierarchically structured habitat. For example, individuals might live on bushes, the bushes grow around the edges of clearings, and the clearings are scattered across an archipelago of islands. We label each bush with an ordered triple recording on which island, in which clearing, and in what bush around the clearing it lives, so that for instance $(2, 1, 4)$ denotes the fourth bush in the first clearing of the second island. To make the mathematical picture a pretty one, we suppose that each of the I islands has the same number C of clearings and each clearing has the same number B of bushes. This enables us identify the habitat structure with the group $\mathbb{Z}_I \otimes \mathbb{Z}_C \otimes \mathbb{Z}_B$, where, as above, \mathbb{Z}_m is the group of integers modulo m . We will get particularly simple and interpretable results if we also assume that dispersal rates and environmental covariances only

depend on the scale at which the movement occurs—between bushes, clearings, or islands.

Although it requires imaginative work to find examples with many more scales than this (do the organism’s fleas have fleas?) it does not cost us anything to work in greater generality. Suppose, then, that the patches in the habitat are labeled with the group $G = G_1 \otimes \cdots \otimes G_k$, where $G_j = \mathbb{Z}_{n_j}$ for $1 \leq j \leq k$.

Thus, one patch is labeled with the identity element $\mathbf{id}_G = (\mathbf{id}_1, \dots, \mathbf{id}_k)$ and every other patch is labeled by the displacement required to get there from \mathbf{id}_G . The later coordinates are understood to be at finer “scales”, so that if $g_i = h_i$ for all $1 \leq i \leq j - 1$, then g and h represent patches in the same *metapatch at scale* j . For instance, in our example above, the archipelago of islands is the single metapatch at scale 1 and the metapatches at scales 2 and 3 are, respectively, the islands and the clearings. We label the metapatches at scale r with the set $Z_r := \{g \in G : g_r = \mathbf{id}_r, \dots, g_k = \mathbf{id}_k\}$, with the convention that $Z_{k+1} := G$. Because a label $g = (g_1, \dots, g_k) \in G$ represents displacement, the coordinate of the leftmost non-identity element of g , denoted by

$$\ell(g) := \min\{j : g_j \neq \mathbf{id}_j\} \text{ and } \ell(\mathbf{id}_G) = k + 1,$$

tells us the scale on which the motion occurs: $g \in G$ corresponds to a displacement that moves between patches within the same metapatch at scale $\ell(g)$ but moves from a patch within a metapatch at scale $\ell(g) + 1$ to a patch within some other metapatch at that scale. Note that $1 \leq \ell(g) \leq k + 1$.

We assume that the dispersal rate and the environmental covariance between two patches only depends on the scale of the displacement necessary to move between the two patches. That is, we suppose there are numbers q_1, \dots, q_{k+1} and s_1, \dots, s_{k+1} such that $q(g) = q_{\ell(g)}$ and $s(g) = s_{\ell(g)}$.

In Appendix G we show that the Fourier transforms appearing in Theorem 3 depend on the following quantities. Let $N_r := \#Z_r = \prod_{j=1}^{r-1} n_j$ be the number of metapatches at scale r . Write $\bar{Z}_r := \{g \in G : g_j = \mathbf{id}_j, j \leq r\}$ for the subgroup of displacements that move from one patch to another within the same metapatch at scale $r + 1$ and set $\bar{N}_r := \#\bar{Z}_r = \prod_{j=r+1}^k n_j$. Set

$$v_\mu(r) := \frac{1}{N_r} \sum_{g \in Z_r} \left(\frac{1}{n_r} \sum_{h \in G_r} \left(\frac{1}{\bar{N}_r} \sum_{z \in \bar{Z}_r} \mu(ghz) \right)^2 - \left(\frac{1}{n_r} \sum_{h \in G_r} \frac{1}{\bar{N}_r} \sum_{z \in \bar{Z}_r} \mu(ghz) \right)^2 \right).$$

We can interpret this quantity as follows. There are N_r metapatches at scale r . Each one has within it n_r metapatches at scale $r + 1$. First, compute the average of μ over all the patches within each metapatch at scale $r + 1$, then compute the variance of these averages within each metapatch at scale r , and finally average these variances across all the metapatches at scale r to produce $v_\mu(r)$. Thus, $v_\mu(r)$ measures the variability in μ that can be attributed to scale $r + 1$. Set

$$\tilde{s}(r) = \sum_{\ell=r}^k (s_{\ell+1} - s_\ell) \bar{N}_\ell$$

and

$$\tilde{q}(r) = - \sum_{\ell=1}^r q_\ell (\bar{N}_{\ell-1} - \bar{N}_\ell) - q_r \bar{N}_r.$$

The following result agrees with Eq. (21), which describes the special case where there is a single scale.

Theorem 4 *For a habitat with the above multi-scale structure, equation (19) reduces to*

$$\chi(\delta) = \left(\bar{\mu} - \frac{1}{2} \bar{s} \right) - \frac{1}{\delta} \sum_{r=1}^k \frac{1}{\tilde{q}(r)} \left(v_\mu(r) - \frac{N_{r+1} - N_r}{4N_{k+1}^2} \tilde{s}(r)^2 \right) + O(\delta^{-5/4}) \quad (30)$$

as $\delta \rightarrow \infty$. Furthermore, $\tilde{q}(r) < 0$ for all $1 \leq r \leq k$.

Note that if s_ℓ increases with ℓ (that is, two patches within the same metapatch have a higher environmental covariance than two patches in different metapatches at that scale), then $\tilde{s}(r)$ decreases with r . Also, if q_ℓ increases with ℓ (that is, there is a higher rate for dispersing to a patch within the same metapatch at some scale than to a patch in another metapatch at that scale), then $\tilde{q}(r)$ is negative and decreases with r . Using these observations, we may read off several things from (30).

First, consider a simple example with a fixed, large number n of patches distributed among a variable number of islands. Now $k = 2$, and let the number of islands $n_1 = 1/\alpha$, with $\alpha \geq 1$, so that the number of patches on each island is $n_2 = \alpha n$. In this case, $N_1 = 1$, $N_2 = 1/\alpha$, and $N_3 = n$, while $\bar{N}_0 = n$, $\bar{N}_1 = \alpha n$, and $\bar{N}_2 = 1$, so (30) reads

$$\begin{aligned} \chi(\delta) &\approx \left(\bar{\mu} - \frac{1}{2} \bar{s} \right) - \frac{1}{\delta} \left\{ -\frac{v_\mu(1)}{q_1 n} + \frac{(1 - \alpha)((s_3 - s_2) + \alpha n(s_2 - s_1))^2}{\alpha q_1 n^2} \right. \\ &\quad \left. - \frac{v_\mu(2) - (\alpha n - 1)(s_3 - s_2)^2}{\alpha n^2 (q_2 \alpha + q_1 (1 - \alpha))} \right\} \\ &= \left(\bar{\mu} - \frac{1}{2} \bar{s} \right) - \frac{\alpha(1 - \alpha)(s_2 - s_1)^2}{\delta q_1} + O(n^{-1}). \end{aligned} \quad (31)$$

The effect of higher dispersal depends on the difference in covariances between patches on the same island and on different islands, and on the number of islands.

Biological interpretation of equation (31) *If a sufficiently large number of patches are distributed equally across a number of islands, then for a given dispersal pattern, the stochastic growth rate increases with the dispersal rate (at high levels of dispersal). This effect is strongest if there are only two islands (i.e. $\alpha = 1/2$).*

Secondly, imagine a fixed ensemble of patches with varying mean per-capita growth rates and consider the following two possibilities for assignment of these patches to metapatches at scale 2 (the islands in our bush-clearing-island example). One possibility is that some islands are assigned patches that are primarily of high quality, whereas

other islands are mostly assigned poor patches. The other possibility is that patches of different quality are evenly spread across the islands, with the range of quality within an island similar to the range of quality between islands. In the first case, the variance across islands of within-island means is comparable to the variance across all patches, so $v_\mu(1) \approx v_\mu(k)$. In the second case, the within-island means are approximately constant, so that $v_\mu(1)$ will be small. Therefore, since $\tilde{q}(r)$ is negative for all r , having local positive association of μ at nearby patches leads to higher stochastic growth rates, at least for large enough values of the dispersal parameter δ .

Biological interpretation of equation (30) *All other things being equal, the species will do better if the good habitat is concentrated on particular islands, rather than spread out across many.*

Finally, we can observe that adding new scales of metapatch may change the situation from one in which $\chi(\delta)$ is maximal at high values of the dispersal parameter δ to one in which $\chi(\delta)$ is maximal at intermediate values of δ , or vice-versa. If $n_1 = 1$, then $\tilde{s}(1)$ and $v_\mu(1)$ are both zero, and changing n_1 (for example, going from one to several islands in our example) will increase $\tilde{s}(1)$. Changing n_1 will also add the quantity $-q_1(n_1 - 1)\tilde{N}_1$ to all values of $\tilde{q}(r)$. The result of this could be to change the sign of the coefficient of $\frac{1}{\delta}$ in (19).

Biological interpretation of equation (30) *The optimal level of dispersal for a subpopulation, and the growth rate at that level of dispersal, may differ drastically depending on whether it is connected (or connectable) by dispersal to other subpopulations.*

6 Discussion

Classical ecology theory predicts that environmental stochasticity increases extinction risk by reducing the long term per-capita growth rate of populations (May 1975; Turelli 1978). For sedentary populations in a spatially homogeneous yet temporally variable environment, a simple model of their growth is given by the stochastic differential equation $dZ_t = \mu Z_t dt + \sigma Z_t dB_t$, where B is a standard Brownian motion. The stochastic growth rate for such populations equals $\mu - \frac{\sigma^2}{2}$; the reduction in the growth rate is proportional to the infinitesimal variance of the noise. Here, we show that a similar expression describes the growth of populations dispersing in spatially and temporally heterogeneous environments. More specifically, if average per-capita growth rate in patch i is μ_i and the infinitesimal spatial covariance between environmental noise in patches i and j is σ_{ij} , then the stochastic growth rate equals the average of the mean per-capita growth rate $\langle \mu \rangle = \sum_i \mu_i \mathbb{E}[Y_\infty^i]$ experienced by the population when the proportions of the population in the various patches have reached equilibrium minus half of the average temporal variation $\langle \sigma^2 \rangle = \mathbb{E}[\sum_{i,j} \sigma_{ij} Y_\infty^i Y_\infty^j]$ experienced by the population in equilibrium. The law of \mathbf{Y}_∞ , the random equilibrium spatial distribution of the population which provides the weights in these averages, is determined by interactions between spatial heterogeneity in mean per-capita growth

rates, the infinitesimal spatial covariances of the environmental noise, and population movement patterns. To investigate how these interactions effect the stochastic growth rate, we derived analytic expressions for the law of \mathbf{Y}_∞ , determined what choice of dispersal mechanisms resulted in optimal stochastic growth rates for a freely dispersing population, and considered the consequences on the stochastic growth rate of limiting the population to a fixed dispersal mechanism. As we now discuss, these analytic results provide fundamental insights into “ideal free” movement in the face of uncertainty, the persistence of coupled sink populations, the evolution of dispersal rates, and the single large or several small (SLOSS) debate in conservation biology.

In spatially heterogeneous environments, “ideal free” individuals disperse to the patch or patches that maximize their long term per-capita growth rate (Fretwell and Lucas 1970; Harper 1982; Oksanen et al. 1995; van Baalen and Sabelis 1999; Schreiber et al. 2000; Schreiber and Vejdani 2006; Kirkland et al. 2006; Cantrell et al. 2007). In the absence of environmental stochasticity and density-dependent feedbacks, ideal free dispersers only select the patches supporting the highest per-capita growth rate. Here, we show that uncertainty due to environmental stochasticity can overturn this prediction. Provided environmental stochasticity is sufficiently strong and spatial correlations are sufficiently weak, equation (16) implies that ideal free dispersers occupy all patches despite spatial variation in the local stochastic growth rates $\mu_i - \sigma_i^2/2$. Intuitively, by spending time in multiple patches, including those that in isolation support lower stochastic growth rates, individuals reduce the net environmental variation ($\langle \sigma^2 \rangle$) they experience and, thereby, increase their stochastic growth rate. Hence, dispersing to lower quality patches is a form of bet-hedging against environmental uncertainty (Slatkin 1974; Philippi and Seger 1989; Wilbur and Rudolf 2006). When environmental fluctuations in higher quality patches are sufficiently strong, this spatial bet-hedging can result in ideal free dispersers occupying sink patches; patches that are unable in the absence of immigration to sustain a population. This latter prediction is consistent with Holt’s analysis of a discrete-time two patch model (Holt 1997). Spatial correlations in environmental fluctuations, however, can disrupt spatial bet-hedging. Movement between patches exhibiting strongly covarying environmental fluctuations has little effect on the net environmental variation ($\langle \sigma^2 \rangle$) experienced by individuals and, therefore, movement to lower quality patches may confer little or no advantage to individuals. Indeed, when the spatial covariation is sufficiently strong, ideal free dispersers only occupy patches with the highest local stochastic growth rates $\mu_i - \sigma_i^2/2$, similar to the case of deterministic environments. In deterministic environments, density dependent feedbacks can result in ideal-free dispersers occupying multiple patches including sink patches (Fretwell and Lucas 1970; Cantrell et al. 2007; Holt and McPeck 1996). Our results show that even density-independent processes can result in populations occupying multiple patches. However, both of these processes are likely to play important roles in the evolution of patch selection.

A sink population is a local population that is sustained by immigration (Holt 1985; Pulliam 1988; Dias 1996). Removing immigration results in a steady decline to extinction. In contrast, source populations persist in the absence of immigration. Empirical studies have shown that landscapes often partition into mosaics of source and sink populations (Murphy 2001; Kreuzer and Huntly 2003; Keagy et al. 2005). For discrete-time two-patch models, Jansen and Yoshimura (1998) showed, quite surprisingly, that sink

populations coupled by dispersal can persist, a prediction supported by recent empirical studies with protozoan populations (Matthews and Gonzalez 2007) and extended to discrete-time multi-patch models (Roy et al. 2005; Schreiber 2010). Here, we show a similar phenomena occurs for populations experiencing continuous temporal fluctuations. For example, if the stochastic growth rates in all patches equal $\mu - \sigma^2/2$ and the spatial correlation between patches is ρ , then equations (5) and (18) imply that populations dispersing freely between n patches persist whenever $\mu - ((n - 1)\rho + 1)\sigma^2/2n > 0$. Hence, ideal free movement mediates persistence whenever local environmental fluctuations produce sink populations (i.e., $\sigma^2/2 > \mu > 0$), environmental fluctuations aren't fully spatially correlated (i.e. $\rho < 2\mu/\sigma^2$) and there are sufficiently many patches (i.e., $n > ((1 - \rho)\sigma^2)/(2\mu - \rho\sigma^2)$). This latter expression for the necessary number of patches to mediate persistence is an exact, continuous time counterpart to an approximation by Bascompte et al. (2002) for discrete time models. When two patches are sufficient to mediate persistence, equation (9) reveals that there is a critical dispersal rate below which the population is extinction prone and above which it persists. Our high dispersal approximation (see equation (21) with $\text{Var}[\mu] = 0$) suggests this dispersal threshold also exists for an arbitrary number of patches.

While ideal free movement corresponds to the optimal dispersal strategy for species without any constraints on their movement or their ability to collect information, many organisms experience these constraints. For instance, in the absence of information about environmental conditions in other patches, individuals may move randomly between patches, in which case the rate of movement (rather than the pattern of movement) is subject to natural selection (Hastings 1983; Levin et al. 1984; McPeck and Holt 1992; Holt and McPeck 1996; Dockery et al. 1998; Hutson et al. 2001; Kirkland et al. 2006). When density-dependent feedbacks are weak and certain symmetry assumptions are met, our high dispersal approximation in (20) implies there is selection for higher dispersal rates whenever

$$\sum_{k=1}^{n-1} \frac{1}{|\lambda_k|} \frac{1}{4n} \theta_k^2 > \sum_{k=1}^{n-1} \frac{1}{|\lambda_k|} (\xi_k^T \mu)^2 \tag{32}$$

where, recall, $\lambda_k < 0$, ξ_k are the eigenvalues/vectors of the dispersal matrix, μ is the vector of per-capita growth rates, and θ_k are the eigenvalues of the covariance matrix for the environmental noise. Roughly speaking, Eq. (32) asserts that if temporal variation (averaged in the appropriate manner) exceeds spatial variation, then there is selection for faster dispersers; a prediction consistent with the general consensus of earlier studies (Levin et al. 1984; McPeck and Holt 1992; Hutson et al. 2001). More specifically, in the highly symmetric case where the temporal variation in all patches equals σ^2 and the spatial correlation between patches is ρ , Eq. (32) simplifies to

$$\frac{(1 - \rho)\sigma^2}{2} > \frac{n}{\sqrt{n - 1}} \sqrt{\text{Var}[\mu]}, \tag{33}$$

in which case lower spatial correlations and larger number of patches also facilitate selection for faster dispersers. Another important constraint influencing the evolution

of dispersal are travel costs that reduce fitness of dispersing individuals. While the effect of these costs have been investigated for deterministic models (DeAngelis et al. 2011), it remains to be seen how these traveling costs interact with environmental stochasticity in determining optimal dispersal strategies.

Previous studies have shown that spatial heterogeneity in per-capita growth rates increases the net population growth rate for deterministic models with diffusive movement (Adler 1992; Schreiber and Lloyd-Smith 2009). Intuitively, spatial heterogeneity provides patches with higher per-capita growth rates that boost the population growth rate, a boost that gets diluted at higher dispersal rates. Our high dispersal approximation (20) shows that this boost also occurs in temporally heterogeneous environments, i.e. the correction term $-\sum_{k=1}^{n-1} \frac{1}{\lambda_k} (\xi_k^T \mu)^2 / \delta$ is positive. More importantly, the multi-scale version of this correction term (30) implies this boost is larger when the variation in the per-capita growth rates occurs at multiple spatial scales. For example, for insects living on plants in meadows on islands, the largest boost occurs when the higher quality plants (i.e. the plants supporting the largest μ_i values) occur on the same island in the same meadow. This analytic conclusion is consistent with numerical simulations showing that habitat fragmentation (e.g. distributing high quality plants more evenly across islands and meadows) increases extinction risk (Fahrig 1997, 2002). Intuitively, spatial aggregation of higher quality patches increases the chance of individuals dispersing away from a high quality patch arriving in another high quality patch. Even without spatial variation in per-capita growth rates, equation (30) implies that strong spatial aggregation of patches maximizes stochastic growth rates for dispersive populations living in environments where temporal correlations decrease with spatial scale. This finding promotes the view that a single large (SL) reserve is a better for conservation than several small (SS) reserves. This finding is consistent with many arguments in the SLOSS debate (Diamond 1975; Wilcox and Murphy 1985; Gilpin 1988; Cantrell and Cosner 1989, 1991). For example, using reaction-diffusion equations, Cantrell and Cosner (1991) found that even in deterministic environments “[it] is better for a population to have a few large regions of favorable habitat than a great many small ones closely intermingled with unfavorable regions.” However, our results run contrary to a numerical simulation study of Quinn and Hastings (1987) that, unlike ours, applies to sedentary populations experiencing independent environments (Gilpin 1988).

While our work provides a diversity of analytical insights into the interactive effects of temporal variability, spatial heterogeneity, and movement on long-term population growth, many challenges remain. Most notably, are there analytic approximations for relatively sedentary populations? What effect do correlations in the temporal fluctuations have on the stochastic growth rate? Can the explicit formulas for stochastic growth rates in two patch environments be extended to special classes of higher dimensional models? Can one extend the analysis to account for density-dependent feedbacks? Answers to these questions are likely to provide important insights into the evolution of dispersal and metapopulation persistence.

Acknowledgments AS was supported by EPSRC grant EP/G055068/1, SNE was supported in part by NSF grant DMS-0907630, SJS was supported in part by NSF grants EF-0928987 and DMS-1022639, and PLR was supported by funds to SJS from the Dean’s Office of the College of Biological Sciences, University of California, Davis, and by NIH fellowship F32-GM096686.

Appendix A: Proof of Proposition 1

Define the matrix R by

$$R := \text{diag}(\mu) + D.$$

Equation (3) becomes

$$d\mathbf{X}_t = \text{diag}(\mathbf{X}_t)\Gamma^T d\mathbf{B}_t + R^T \mathbf{X}_t dt.$$

Recall that $Y_t^j = X_t^j / (X_t^1 + \dots + X_t^n)$ for each $1 \leq j \leq n$ and $\mathbf{Y}_t = (Y_t^1, \dots, Y_t^n)^T$. Fix j and define $f_j(x_1, \dots, x_n) := x_j / (x_1 + \dots + x_n)$, so that $Y^j = f_j(\mathbf{X})$. Using ∂_k to denote differentiation with respect to x_k , observe that

$$\begin{aligned} \partial_j f_j(x_1, \dots, x_n) &= \left(\sum_{\ell \neq j} x_\ell \right) / \left(\sum_{\ell} x_\ell \right)^2, \\ \partial_k f_j(x_1, \dots, x_n) &= -x_j / \left(\sum_{\ell} x_\ell \right)^2, \quad k \neq j. \end{aligned}$$

Moreover,

$$\begin{aligned} \partial_{jj} f_j(x_1, \dots, x_n) &= -2 \left(\sum_{\ell \neq j} x_\ell \right) / \left(\sum_{\ell} x_\ell \right)^3, \\ \partial_{jk} f_j(x_1, \dots, x_n) &= -1 / \left(\sum_{\ell} x_\ell \right)^2 + 2x_j / \left(\sum_{\ell} x_\ell \right)^3, \quad k \neq j \end{aligned}$$

and

$$\partial_{km} f_j(x_1, \dots, x_n) = 2x_j / \left(\sum_{\ell} x_\ell \right)^3, \quad k, m \neq j.$$

It follows from Itô’s Lemma (Gardiner 2004) that for each $1 \leq j \leq n$,

$$\begin{aligned} dY_t^j &= \sum_{k=1}^n \partial_k f_j(\mathbf{X}_t) X_t^k \Gamma_{*k}^T d\mathbf{B}_t + \sum_{k=1}^n \partial_k f_j(\mathbf{X}_t) \mathbf{X}_t^T R_{*k} dt \\ &\quad + (1/2) \sum_{k,m=1}^n \partial_{km} f_j(\mathbf{X}_t) X_t^k X_t^m (\Sigma)_{km} dt, \end{aligned}$$

where Γ_{*k} and R_{*k} denote the k^{th} columns of the matrices Γ and R respectively. Substituting in the derivatives of f_j gives

$$\begin{aligned} dY_t^j &= - \sum_{k \neq j} Y_t^j Y_t^k \Gamma_{*k}^T d\mathbf{B}_t + \sum_{k \neq j} Y_t^j Y_t^k \Gamma_{*j}^T d\mathbf{B}_t \\ &\quad - \sum_{k \neq j} Y_t^j Y_t^T R_{*k} dt + \sum_{k \neq j} Y_t^k Y_t^T R_{*j} dt \\ &\quad + (1/2) \sum_{k,m \neq j} 2Y_t^j Y_t^k Y_t^m \Sigma_{km} dt - (1/2) \sum_{k \neq j} 2Y_t^k (Y_t^j)^2 \Sigma_{jj} dt \\ &\quad + (1/2) \times 2 \sum_{k \neq j} \left(- Y_t^j Y_t^k + 2Y_t^k (Y_t^j)^2 \right) \Sigma_{kj} dt \\ &= -Y_t^j \sum_k Y_t^k \Gamma_{*k}^T d\mathbf{B}_t + Y_t^j \Gamma_{*j}^T d\mathbf{B}_t - Y_t^j \sum_k Y_t^k R_{*k} dt + Y_t^T R_{*j} dt \\ &\quad + Y_t^j \sum_{k,m} Y_t^k Y_t^m \Sigma_{km} dt - Y_t^j \sum_k Y_t^k \Sigma_{kj} dt. \end{aligned}$$

Since $D\mathbf{1} = 0$, we have $\sum_k R_{*k} = R\mathbf{1} = \text{diag}(\mu)\mathbf{1} = \mu$, and the above system of SDEs can be written in the following compact way

$$\begin{aligned} d\mathbf{Y}_t &= -\mathbf{Y}_t \mathbf{Y}_t^T \Gamma^T d\mathbf{B}_t + \text{diag}(\mathbf{Y}_t) \Gamma^T d\mathbf{B}_t \\ &\quad - \mathbf{Y}_t \mathbf{Y}_t^T \mu dt + R^T \mathbf{Y}_t dt + \mathbf{Y}_t \mathbf{Y}_t^T \Sigma \mathbf{Y}_t dt - \text{diag}(\mathbf{Y}_t) \Sigma \mathbf{Y}_t dt \\ &= \left(\text{diag}(\mathbf{Y}_t) - \mathbf{Y}_t \mathbf{Y}_t^T \right) \Gamma^T d\mathbf{B}_t + D^T \mathbf{Y}_t dt \\ &\quad + \left(\text{diag}(\mathbf{Y}_t) - \mathbf{Y}_t \mathbf{Y}_t^T \right) (\mu - \Sigma \mathbf{Y}_t) dt. \end{aligned}$$

Now that the SDE (4) is established, we will prove the ergodicity of the Markov process $(\mathbf{Y}_t)_{t \geq 0}$ defined in (4).

Existence Clearly $(\mathbf{Y}_t)_{t \geq 0}$ is a Feller process. Since for each $t \geq 0$, the random vector \mathbf{Y}_t takes values in the compact state space Δ , it trivially follows that the family of probability measures $\{\mathbb{P}^y\{\mathbf{Y}_t \in \cdot\} : t > 0\}$ is uniformly tight for any fixed $y \in \Delta$, where \mathbb{P}^y denotes the law of the process with $\mathbf{Y}_0 = y$. Hence, by the Krylov–Bogolyubov Theorem (see, for example, (Da Prato and Zabczyk, 1996, Corollary 3.1.2)), there exists at least one probability measure μ on Δ which is an invariant measure for the process $(\mathbf{Y}_t)_{t \geq 0}$, that is,

$$\int_{\Delta} \mu(dy) \mathbb{P}^y\{\mathbf{Y}_t \in \cdot\} = \mu\{\cdot\}.$$

Uniqueness The uniqueness of the invariant measure for $(\mathbf{Y}_t)_{t \geq 0}$ is ensured by the Doob–Khasminskii theorem (see, for example, (Da Prato and Zabczyk, 1996, Chapter 7)), provided this process satisfies the following two properties:

- $(\mathbf{Y}_t)_{t \geq 0}$ is *irreducible*, that is, $\mathbb{P}^y\{\mathbf{Y}_t \in V\} > 0$ for any $t > 0$ and any open set V in the simplex Δ .
- $(\mathbf{Y}_t)_{t \geq 0}$ is *strong Feller*, that is, $\Delta \ni y \mapsto \int_{\Delta} \mathbb{P}^y\{\mathbf{Y}_t \in dz\} f(z)$ is continuous for any bounded measurable function $f : \Delta \rightarrow \mathbb{R}$.

These conditions also ensure that $(\mathbf{Y}_t)_{t \geq 0}$ converges in law to the unique invariant measure. We next establish irreducibility and the strong Feller property of $(\mathbf{Y}_t)_{t \geq 0}$ separately.

(a) Irreducibility It clearly suffices to show that the process $(\mathbf{X}_t)_{t \geq 0}$ as defined by (3) is irreducible, that is, that $\mathbb{P}^x\{\mathbf{X}_t \in U\} > 0$ for each $t > 0$, $x \in \mathbb{R}_+^n \setminus \{0\}$ and open set $U \subseteq \mathbb{R}_+^n$.

We will first prove that $\mathbb{P}^x\{X_t^i > 0 \forall i\} = 1$ for all $t > 0$ and all $x \in \mathbb{R}_+^n \setminus \{0\}$, by induction on the size of the set $G := \{1 \leq i \leq n : x_i = 0\}$. First consider the case $\#G = 0$. By a suitable comparison theorem for SDEs (Geiß and Manthey, 1994, Theorem 1.1), $\mathbb{P}^x\{\mathbf{X}_t \geq \hat{\mathbf{X}}_t \text{ for all } t \geq 0\} = 1$, where \hat{X} is defined by

$$d\hat{X}_t^i = \mu_i \hat{X}_t^i dt + \hat{X}_t^i dE_t^i + D_{ii} \hat{X}_t^i dt, \quad 1 \leq i \leq n.$$

This SDE has the unique solution $\hat{X}_t^i = x^i \exp(E_t^i + (\mu + D_{ii} - \frac{1}{2}\Sigma_{ii})t) > 0$, so

$$\mathbb{P}^x\{X_t^i > 0 \forall i \text{ for all } t > 0\}, \quad x \in (0, \infty)^n. \tag{34}$$

Now suppose $\#G = k < n$. By the irreducibility of the infinitesimal generator matrix D , there exist $i_0 \in G, j_0 \notin G$ such that $D_{j_0, i_0} > 0$. Consider the new SDE

$$d\tilde{X}_t^i = \mu_i \tilde{X}_t^i dt + \tilde{X}_t^i dE_t^i + D_{ii} \tilde{X}_t^i dt, \quad i \neq i_0,$$

and

$$d\tilde{X}_t^{i_0} = \mu_{i_0} \tilde{X}_t^{i_0} dt + \tilde{X}_t^{i_0} dE_t^{i_0} + (D_{j_0 i_0} \tilde{X}_t^{j_0} + D_{i_0 i_0} \tilde{X}_t^{i_0}) dt.$$

By the same comparison theorem, $\mathbb{P}^x\{\mathbf{X}_t \geq \tilde{\mathbf{X}}_t \text{ for all } t \geq 0\} = 1$. Clearly, $\mathbb{P}^x\{\tilde{X}_t^i > 0\} = 1$ for all $i \notin G$ and for all $t > 0$. Since $\tilde{X}_0^{i_0} = 0$ and $\tilde{X}_0^{j_0} > 0$, at time $t = 0$ the diffusion component of $\tilde{X}_t^{i_0}$ vanishes but its drift coefficient is strictly positive. It follows that $\mathbb{P}^x\{\tilde{X}_t^{i_0} > 0\} = 1$ for all $t > 0$. Hence, at any positive time t , almost surely $\tilde{\mathbf{X}}_t$ has at most $k - 1$ zero coordinates, and, by the comparison theorem, so does \mathbf{X}_t . Using the Markov property and the induction hypothesis, we deduce that $\mathbb{P}^x\{X_t^i > 0 \forall i\} = 1$ for all $t > 0$. This proves that each component of \mathbf{X} is strictly positive with probability 1 for each $t > 0$.

Let $\varphi : (0, \infty)^n \rightarrow \mathbb{R}^n$ be the homeomorphism given by $\varphi(x) = (\log x_1, \dots, \log x_n)$. Set $\mathbf{H}_t = \varphi(\mathbf{X}_t)$, with $\mathbf{H}_t = (H_t^1, \dots, H_t^n)^T$. By (34), this stochastic process is well defined provided $\mathbf{X}_0 \in (0, \infty)^n$. Note that $(\mathbf{H}_t)_{t \geq 0}$ satisfies the following SDE,

$$dH_t^i = (\mu_i - \frac{1}{2}\Sigma_{ii})dt + dE_t^i + e^{-H_t^i} \sum_{j=1}^n D_{ji} e^{H_t^j} dt, \quad 1 \leq i \leq n.$$

By Girsanov’s theorem (see (Ikeda and Watanabe, 1989, Section 4 of Chapter IV)), the law of $(\Gamma^T)^{-1}\mathbf{H}_t$ (and hence the law of \mathbf{H}_t) is absolutely continuous with respect to the law of \mathbf{B}_t for any $t > 0$. Thus, $\mathbb{P}^x\{\mathbf{H}_t \in V\} > 0$ for any open set $V \subseteq \mathbb{R}^n$. Finally, for any $x \in \mathbb{R}^n \setminus \{0\}$,

$$\begin{aligned} \mathbb{P}^x\{\mathbf{X}_t \in U\} &= \int_{\mathbb{R}_+^n} \mathbb{P}^x\{\mathbf{X}_{t/2} \in dy\} \mathbb{P}^y\{\mathbf{X}_{t/2} \in U\} \\ &= \int_{(0,\infty)^n} \mathbb{P}^x\{\mathbf{X}_{t/2} \in dy\} \mathbb{P}^y\{\mathbf{X}_{t/2} \in U\} \\ &= \int_{(0,\infty)^n} \mathbb{P}^x\{\mathbf{X}_{t/2} \in dy\} \mathbb{P}^{\varphi(y)}\{\mathbf{H}_{t/2} \in \varphi(U)\} > 0. \end{aligned}$$

(b) Strong Feller property Note that \mathbf{H} satisfies a SDE of the form $d\mathbf{H}_t = \Gamma^T d\mathbf{B}_t + b(\mathbf{H}_t)dt$ for some smooth function $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$. For each $K \geq 1$, consider a new SDE

$$d\mathbf{H}_t^K = \Gamma^T d\mathbf{B}_t + b^K(\mathbf{H}_t)dt,$$

where $b^K : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a smooth bounded function with bounded derivative such that $b^K(x) = b(x)$ on $[-K, K]^n$. Since the matrix Γ is nonsingular, the associated Fisk–Stratonovich type generator of $(\mathbf{H}_t^K)_{t \geq 0}$ is trivially hypoelliptic, which in turn implies that $(\mathbf{H}_t^K)_{t \geq 0}$ is strong Feller for every $K \geq 1$ (see (Ikeda and Watanabe, 1989, Section 8 of Chapter V)). If we define a sequence of stopping times $\tau_K := \inf\{t : \|X_t\|_\infty \geq K\}$, then $\mathbf{H}_0^K = \mathbf{H}_0 = x \in [-K, K]^n$ implies $\mathbf{H}_t^K = \mathbf{H}_t$ for $t \in [0, \tau_K]$. Let $t > 0$ and f be a bounded measurable function. Fix $\epsilon > 0$. Then for any $x \in \mathbb{R}^n$,

$$|\mathbb{E}^x[f(\mathbf{H}_t)] - \mathbb{E}^x[f(\mathbf{H}_t^K)]| \leq 2\|f\|_\infty \mathbb{P}^x\{\tau_K < t\}.$$

Hence, for any open neighborhood $U(x)$ of x ,

$$\begin{aligned} |\mathbb{E}^y[f(\mathbf{H}_t)] - \mathbb{E}^x[f(\mathbf{H}_t)]| &\leq |\mathbb{E}^y[f(\mathbf{H}_t^K)] - \mathbb{E}^x[f(\mathbf{H}_t^K)]| \\ &\quad + 4\|f\|_\infty \sup_{z \in U(x)} \mathbb{P}^z\{\tau_K < t\} \quad \text{for all } y \in U(x). \end{aligned}$$

Since almost surely $\tau_K \uparrow \infty$, we can choose K large enough such that $\mathbb{P}^x\{\tau_K < t\} < \epsilon(8\|f\|_\infty)^{-1}$. Moreover, by the Feller property of $(\mathbf{H}_t)_{t \geq 0}$, there exists a neighborhood $U^1(x)$ of x such that $\sup_{z \in U^1(x)} \mathbb{P}^z\{\tau_K < t\} < \epsilon(8\|f\|_\infty)^{-1}$. From the strong Feller property of $(\mathbf{H}_t^K)_{t \geq 0}$, there exists a neighborhood $U^2(x)$ of x such that $|\mathbb{E}^y[f(\mathbf{H}_t^K)] - \mathbb{E}^x[f(\mathbf{H}_t^K)]| < \epsilon/2$ for all $y \in U^2(x)$. Thus, $|\mathbb{E}^y[f(\mathbf{H}_t)] - \mathbb{E}^x[f(\mathbf{H}_t)]| < \epsilon$ for all $y \in U^1(x) \cap U^2(x)$. Hence, $x \mapsto \mathbb{E}^x[f(\mathbf{H}_t)]$ is continuous.

Now, for $t > 0$ and a bounded measurable function $g : \mathbb{R}_+^n \rightarrow \mathbb{R}$,

$$\mathbb{E}^x [g(\mathbf{X}_t)] = \int_{(0,\infty)^n} \mathbb{P}^x \{ \mathbf{X}_{t/2} \in dy \} \mathbb{E}^{\varphi(y)} [g(\varphi^{-1}(\mathbf{H}_{t/2}))], \quad x \in \mathbb{R}_+^n.$$

Therefore, the map $x \mapsto \mathbb{E}^x [g(\mathbf{X}_t)]$ is continuous, and so $(\mathbf{X}_t)_{t \geq 0}$ is a strong Feller process. It follows easily that $(\mathbf{Y}_t)_{t \geq 0}$ is also a strong Feller process. \square

Appendix B: Proof of Proposition 2

By rescaling time $\tau := \delta t$ and setting $\epsilon := 1/\delta$, (4) becomes

$$d\mathbf{Y}_\tau^\epsilon = \sqrt{\epsilon} f(\mathbf{Y}_\tau^\epsilon) d\mathbf{B}_\tau + \epsilon g(\mathbf{Y}_\tau^\epsilon) dt + Q^T \mathbf{Y}_\tau^\epsilon dt \tag{35}$$

where $f(y) := (\text{diag}(y) - yy^T) \Gamma^T$, $g(y) := (\text{diag}(y) - yy^T) (\mu - \Sigma y)$, and $\mathbf{Y}_\tau^\epsilon := \mathbf{Y}_{\tau/\epsilon}$.

For $\epsilon > 0$, let ν_ϵ be the unique invariant probability measure for (35) guaranteed by Proposition 1. The irreducibility of Q implies that π is the unique stable point for the ODE

$$\frac{d}{d\tau} y_\tau^x = Q^T y_\tau^x, \quad y_0^x = x \in \Delta,$$

and that $\lim_{\tau \rightarrow \infty} y_\tau^x = \pi$ for any $x \in \Delta$. Write ν_0 for the Dirac measure at the point $\pi \in \Delta$. By the compactness of Borel probability measures on Δ in the topology of weak convergence, it suffices to show if ν_{ϵ_k} converges weakly to ν for some sequence $\epsilon_k \downarrow 0$, then $\nu = \nu_0$, and hence it is sufficient to check that

$$\int_{\Delta} h(y_\tau^x) \nu(dx) = \int_{\Delta} h(x) \nu(dx)$$

for every $\tau \geq 0$ and Lipschitz function $h : \Delta \rightarrow \mathbb{R}$.

Set $\mathbf{Y}_\tau^k = \mathbf{Y}_\tau^{\epsilon_k}$ and $\nu_k = \nu_{\epsilon_k}$ for ease of notation. Let L be the Lipschitz constant for the function h . Then,

$$\begin{aligned} \left| \int_{\Delta} (h(y_\tau^x) - h(x)) \nu(dx) \right| &= \lim_{k \rightarrow \infty} \left| \int_{\Delta} (h(y_\tau^x) - h(x)) \nu_k(dx) \right| \\ &\leq \limsup_{k \rightarrow \infty} \underbrace{\left| \int_{\Delta} \left(\mathbb{E}^x \left[h(\mathbf{Y}_\tau^k) \right] - h(x) \right) \nu_k(dx) \right|}_{=0 \text{ by invariance of } \nu_k} \end{aligned}$$

$$\begin{aligned}
 & + \limsup_{k \rightarrow \infty} \left| \int_{\Delta} \mathbb{E}^x \left[h(y_{\tau}^x) - h(\mathbf{Y}_{\tau}^k) \right] \nu_k(dx) \right| \\
 & \leq \limsup_{k \rightarrow \infty} L \int_{\Delta} \mathbb{E}^x \left[\|y_{\tau}^x - \mathbf{Y}_{\tau}^k\| \right] \nu_k(dx),
 \end{aligned}$$

where $\|\cdot\|$ is the usual Euclidean norm on \mathbb{R}^n .

It remains to show that $\lim_{k \rightarrow \infty} \sup_{x \in \Delta} \mathbb{E}^x [\|y_{\tau}^x - \mathbf{Y}_{\tau}^k\|] = 0$. Fix $x \in \Delta$ and set $\mathbf{Z}_{\tau}^k := y_{\tau}^x - \mathbf{Y}_{\tau}^k$. By Itô’s formula,

$$\begin{aligned}
 \mathbb{E}^x \left[\|\mathbf{Z}_{\tau}^k\|^2 \right] & = \mathbb{E} \left[\int_0^{\tau} 2 \langle \mathbf{Z}_s^k, \mathcal{Q}^T \mathbf{Z}_s^k \rangle - 2 \epsilon_k \langle \mathbf{Z}_s^k, g(\mathbf{Y}_s^k) \rangle + \epsilon_k \text{Tr}(f(\mathbf{Y}_s^k) f(\mathbf{Y}_s^k)^T) ds \right] \\
 & \leq 2 \|\mathcal{Q}^T\| \int_0^{\tau} \mathbb{E}^x \left[\|\mathbf{Z}_s^k\|^2 \right] ds + \epsilon_k C \tau,
 \end{aligned}$$

for some constant C that does not depend on x or τ , where we write $\langle \cdot, \cdot \rangle$ for the usual Euclidean inner product on \mathbb{R}^n , and $\|\mathcal{Q}^T\| = \sup_{\|z\|=1} |\langle z, \mathcal{Q}^T z \rangle|$. Gronwall’s inequality implies that

$$\mathbb{E}^x \left[\|\mathbf{Z}_{\tau}^k\|^2 \right] \leq \epsilon_k C e^{2\|\mathcal{Q}^T\| \tau},$$

and so, by Jensen’s inequality,

$$\mathbb{E}^x \left[\|\mathbf{Z}_{\tau}^k\| \right] \leq \sqrt{\epsilon_k C} e^{\|\mathcal{Q}^T\| \tau}.$$

It follows that $\lim_{k \rightarrow \infty} \sup_{x \in \Delta} \mathbb{E}^x [\|y_{\tau}^x - \mathbf{Y}_{\tau}^k\|] = 0$, and hence $v = v_0$, as required.

In particular,

$$\begin{aligned}
 \chi(\delta) & = \int_{\Delta} \mu^T y \nu_{1/\delta}(dy) - \frac{1}{2} \int_{\Delta} y^T \Sigma y \nu_{1/\delta}(dy) \\
 & \rightarrow \mu^T \pi - \frac{1}{2} \pi^T \Sigma \pi
 \end{aligned}$$

as $\delta \rightarrow \infty$. □

Appendix C: Proof of Proposition 3

Fix $\delta \in [0, \infty)$, and denote our underlying probability space by $(\Omega, \mathcal{F}, \mathbb{P})$. Define

$$\Phi_{s,t}^{\delta} : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}^n, \quad 0 \leq s \leq t,$$

by $\Phi_{s,t}^\delta(\mathbf{x}, \omega) = \mathbf{X}_t^\delta(\omega)$, where $(\mathbf{X}_u^\delta)_{u \geq s}$ is the unique (strong) solution of

$$\mathbf{X}_u^\delta = \mathbf{x} + \int_s^u \text{diag}(\mathbf{X}_v^\delta) \Gamma^T d\mathbf{B}_v + \int_s^u (R_\delta)^T \mathbf{X}_v^\delta dv$$

with $R_\delta := \text{diag}(\mu) + \delta Q$.

Note that for all $0 \leq s \leq w \leq t$,

$$\Phi_{s,t}(\cdot, \omega) = \Phi_{w,t}^\delta(\cdot, \omega) \circ \Phi_{s,w}^\delta(\cdot, \omega). \tag{36}$$

It is easy to see that $\Phi_{s,t}^\delta(\cdot, \omega)$ is a linear map from \mathbb{R}^n to \mathbb{R}^n and thus can be represented by a matrix $\mathbf{M}_{s,t}^\delta(\omega)$. From (36), it follows that

$$\mathbf{M}_{s,t}^\delta(\omega) = \mathbf{M}_{w,t}^\delta(\omega) \mathbf{M}_{s,w}^\delta(\omega) \quad \text{for all } 0 \leq s \leq w \leq t.$$

Since $\mathbf{M}_{s,t}^\delta$ is constructed from $(\mathbf{B}_u - \mathbf{B}_s)_{u \in [s,t]}$, the matrices $\{\mathbf{M}_{k,k+1}^\delta\}_{k \in \mathbb{N}}$ are independent. Moreover, since the drift and the diffusion coefficients do not depend on time, $\{\mathbf{M}_{k,k+1}^\delta\}_{k \in \mathbb{N}}$ is a stationary sequence.

We note that the Lyapunov exponent $\chi(\delta)$ of $(\mathbf{X}_t^\delta)_{t \geq 0}$ is the same as

$$\lim_{k \rightarrow \infty} \mathbb{E} \left[k^{-1} \log \|\mathbf{M}_{0,k}^\delta\| \right] = \inf_{k \geq 1} \mathbb{E} \left[k^{-1} \log \|\mathbf{M}_{0,k}^\delta\| \right],$$

where we set

$$\|A\| := \sup \left\{ \sum_{i,j} A_{ij} x_j : \sum_k x_k = 1, x_k \geq 0 \forall k \right\}$$

for a matrix A with nonnegative entries.

Set $\mathbb{R}_+^n := \{\mathbf{x} \in \mathbb{R}^n : x \geq 0\}$. If $\delta > 0$, then it follows from the irreducibility of Q that

$$\mathbf{M}_{s,t}^\delta(\mathbb{R}_+^n) \subseteq \{x \in \mathbb{R}^n : x_i > 0 \text{ for all } 1 \leq i \leq n\} \cup \{0\} \tag{37}$$

and hence $\chi(\delta)$ is analytic on $(0, \infty)$ by (Ruelle, 1979, Theorem 3.1).

The condition (37) fails to hold when $\delta = 0$ and so we must proceed differently. We first claim that for fixed $t > 0$ the map $\delta \mapsto t^{-1} \mathbb{E}[\log \|\mathbf{M}_{0,t}^\delta\|]$ is upper semi-continuous on $[0, \infty)$. To see this, fix $\delta \in [0, \infty)$. Set $\log^+ x = \max(0, \log x)$ and $\log^- x = \min(0, \log x)$. It follows from the continuous dependence of the solution of a SDE on its parameters (Gardiner, 2004, 4.3.2), that $\mathbf{X}_t^{\delta'} \rightarrow \mathbf{X}_t^\delta$ almost surely as $\delta' \rightarrow \delta$, which implies that $\|\mathbf{M}_{0,t}^{\delta'}\| \rightarrow \|\mathbf{M}_{0,t}^\delta\|$ almost surely as $\delta' \rightarrow \delta$. An application of Gronwall’s lemma gives that $\mathbb{E}[\sup_{0 \leq \delta \leq c} \|\mathbf{X}_t^\delta\|] < \infty$ for each $c > 0$. Hence,

$$\mathbb{E} \left[\log^+ \|\mathbf{M}_{0,t}^{\delta'}\| \right] \rightarrow \mathbb{E} \left[\log^+ \|\mathbf{M}_{0,t}^\delta\| \right] \quad \text{as } \delta' \rightarrow \delta.$$

On the other hand, by Fatou’s lemma,

$$\mathbb{E} \left[-\log^- \|\mathbf{M}_{0,t}^\delta\| \right] \leq \liminf_{\delta' \rightarrow \delta} \mathbb{E} \left[-\log^- \|\mathbf{M}_{0,t}^{\delta'}\| \right].$$

Combining these two inequalities gives

$$\limsup_{\delta' \rightarrow \delta} \mathbb{E} \left[\log \|\mathbf{M}_{0,t}^{\delta'}\| \right] \leq \mathbb{E} \left[\log \|\mathbf{M}_{0,t}^\delta\| \right],$$

and the claim follows.

Since $\chi(\delta) = \inf_{t>0} t^{-1} \mathbb{E} \log \|\mathbf{M}_{0,t}^\delta\|$ is the infimum of a family of upper semicontinuous functions, it is itself upper semicontinuous, or equivalently, $\limsup_{\delta' \rightarrow \delta} \chi(\delta') \leq \chi(\delta)$. In particular, $\limsup_{\delta \rightarrow 0} \chi(\delta) \leq \chi(0)$.

We now prove the opposite inequality that $\liminf_{\delta \rightarrow 0} \chi(\delta) \geq \chi(0)$. Fix $\delta > 0$, and without loss of generality suppose that $\max_i -Q_{ii} = 1$, so that if $x_i \geq z_i \geq 0$ for $1 \leq i \leq n$, then $(Qx)_i \geq -z_i$ for $1 \leq i \leq n$. Consider the two SDEs

$$d\mathbf{X}_t^\delta = \text{diag}(\mathbf{X}_t^\delta) \Gamma^T d\mathbf{B}_t + (\text{diag}(\mu) + \delta Q^T) \mathbf{X}_t^\delta dt$$

and

$$d\mathbf{Z}_t^\delta = \text{diag}(\mathbf{Z}_t^\delta) \Gamma^T d\mathbf{B}_t + \text{diag}(\mu - \delta) \mathbf{Z}_t^\delta dt.$$

If $\mathbf{X}_0^\delta = \mathbf{Z}_0^\delta$, then, by the comparison theorem,

$$\mathbf{X}_t^\delta \geq \mathbf{Z}_t^\delta \quad \text{for all } t \geq 0$$

almost surely.

Thus, the Lyapunov exponent of $(\mathbf{X}_t^\delta)_{t \geq 0}$ dominates that of $(\mathbf{Z}_t^\delta)_{t \geq 0}$. Note that the coordinates of \mathbf{Z}^δ are decoupled and hence the Lyapunov exponent of this process is the maximum of the stochastic growth rates for the individual coordinate processes. Therefore,

$$\chi(\delta) \geq \max_j \left(\mu_j - \frac{1}{2} \sum_k \sigma_{kj}^2 \right) - \delta.$$

In particular,

$$\liminf_{\delta \rightarrow 0+} \chi(\delta) \geq \max_j \left(\mu_j - \frac{1}{2} \sum_k \sigma_{kj}^2 \right) = \chi(0), \tag{38}$$

as required. □

Appendix D: Proof of Theorem 2

Recall that

$$d\mathbf{Y}_t = \left(\text{diag}(\mathbf{Y}_t) - \mathbf{Y}_t \mathbf{Y}_t^T\right) \Gamma^T d\mathbf{B}_t + D^T \mathbf{Y}_t dt + \left(\text{diag}(\mathbf{Y}_t) - \mathbf{Y}_t \mathbf{Y}_t^T\right) (\mu - \Sigma \mathbf{Y}_t) dt,$$

where D is of the form δQ , with Q an irreducible infinitesimal generator matrix and $\delta > 0$. Moreover, Q is assumed to be reversible with respect to the unique probability vector π satisfying $Q^T \pi = 0$; that is, that $\pi_i Q_{ij} = \pi_j Q_{ji}$ for all i, j .

Define an inner product on \mathbb{R}^n by $\langle u, v \rangle_\pi := \sum_i \frac{1}{\pi_i} u_i v_i = u^T \text{diag}(\pi)^{-1} v$. It follows from reversibility that the linear operator $v \mapsto Q^T v$ is self-adjoint with respect to this inner product; that is, that $\langle u, Q^T v \rangle_\pi = \langle Q^T u, v \rangle_\pi$ for all u, v .

From the spectral theorem and the Perron-Frobenius theorem, the linear operator $v \mapsto Q^T v$ has eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-1} < \lambda_n = 0$ and corresponding orthonormal eigenvectors ξ_1, \dots, ξ_n with $\xi_n = \pi$ such that

$$Q^T v = \sum_{k=1}^{n-1} \lambda_k \xi_k \langle v, \xi_k \rangle_\pi, \quad v \in \mathbb{R}^n.$$

Note that

$$\mathbf{1}^T v = \langle v, \pi \rangle_\pi = 0 \implies \langle v, Q^T v \rangle_\pi \leq -\kappa \|v\|_\pi^2, \tag{39}$$

where $\kappa := -\lambda_{n-1} > 0$ and $\|\cdot\|_\pi$ is the norm associated with the inner product $\langle \cdot, \cdot \rangle_\pi$.

Note also that if $\mathbf{1}^T v = 0$, then

$$w := \sum_{k=1}^{n-1} \lambda_k^{-1} \xi_k \langle v, \xi_k \rangle_\pi$$

is the unique vector with the properties

$$\langle w, \pi \rangle_\pi = 0 \quad \text{and} \quad Q^T w = v.$$

In particular,

$$\mathbf{1}^T \left(\text{diag}(\pi) - \pi \pi^T\right) (\mu - \Sigma \pi) = \left(\pi^T - \pi^T\right) (\mu - \Sigma \pi) = 0,$$

and so there is a unique vector we denote v such that

$$\mathbf{1}^T v = \langle v, \pi \rangle_\pi = 0 \quad \text{and} \quad Q^T v = -\left(\text{diag}(\pi) - \pi \pi^T\right) (\mu - \Sigma \pi). \tag{40}$$

We emphasize that v does not depend on δ .

Consider the stochastic process

$$\mathbf{U}_t := \delta^{\frac{1}{2}} \left(\mathbf{Y}_{t/\delta} - \pi - \delta^{-1} \nu \right),$$

so that

$$\mathbf{Y}_t = \delta^{-\frac{1}{2}} \mathbf{U}_{\delta t} + \pi + \delta^{-1} \nu.$$

Observe that $\pi + \delta^{-1} \nu$ is indeed a probability vector for δ sufficiently large. Because we are only interested in the equilibrium law of the process \mathbf{Y} , we assume that $\mathbf{Y}_0 = \pi + \delta^{-1} \nu$ and hence $\mathbf{U}_0 = 0$. Note that $0 = \mathbf{1}^T \mathbf{U}_t = \langle \mathbf{U}_t, \pi \rangle_\pi$ for all $t \geq 0$.

We have for the standard Brownian motion $\tilde{\mathbf{B}}_t := \delta^{\frac{1}{2}} \mathbf{B}_{t/\delta}$ that

$$\begin{aligned} d\mathbf{U}_t = & \left(\text{diag}(\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu) - (\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu)(\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu)^T \right) \Gamma^T d\tilde{\mathbf{B}}_t \\ & + \delta^{-\frac{1}{2}} \delta Q^T (\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu) dt \\ & + \delta^{-\frac{1}{2}} \left(\text{diag}(\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu) - (\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu)(\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu)^T \right) \\ & \times \left(\mu - \Sigma(\delta^{-\frac{1}{2}} \mathbf{U}_t + \pi + \delta^{-1} \nu) \right) dt. \end{aligned}$$

Using $Q^T \pi = 0$ and (40), we get

$$\begin{aligned} d\mathbf{U}_t = & \left[\text{diag}(\pi) - \pi \pi^T \right] \Gamma^T d\tilde{\mathbf{B}}_t + Q^T \mathbf{U}_t dt \\ & + \left[\delta^{-\frac{1}{2}} A_{\frac{1}{2}}(\mathbf{U}_t) + \delta^{-1} A_1(\mathbf{U}_t) + \delta^{-\frac{3}{2}} A_{\frac{3}{2}}(\mathbf{U}_t) + \delta^{-2} A_2(\mathbf{U}_t) \right] d\tilde{\mathbf{B}}_t \\ & + \left[\delta^{-1} b_1(\mathbf{U}_t) + \delta^{-\frac{3}{2}} b_{\frac{3}{2}}(\mathbf{U}_t) + \delta^{-2} b_2(\mathbf{U}_t) + \delta^{-\frac{5}{2}} b_{\frac{5}{2}}(\mathbf{U}_t) + \delta^{-3} b_3(\mathbf{U}_t) \right. \\ & \left. + \delta^{-\frac{7}{2}} b_{\frac{7}{2}}(\mathbf{U}_t) \right] dt, \end{aligned}$$

where

$$\begin{aligned} A_{\frac{1}{2}}(u) & := \left[\text{diag}(u) - u \pi^T - \pi u^T \right] \Gamma^T \\ A_1(u) & := \left[-u u^T + \text{diag}(\nu) - \pi \nu^T - \nu \pi^T \right] \Gamma^T \\ A_{\frac{3}{2}}(u) & := \left[-u \nu^T - \nu u^T \right] \Gamma^T \\ A_2(u) & := -\nu \nu^T \Gamma^T \end{aligned}$$

and

$$\begin{aligned} b_1(u) & := -\pi u^T \mu - u \pi^T \mu + \pi u^T \Sigma \pi + u \pi^T \Sigma \pi + \pi \pi^T \Sigma u \\ & \quad + \text{diag}(u) \mu - \text{diag}(\pi) \Sigma u - \text{diag}(u) \Sigma \pi \\ b_{\frac{3}{2}}(u) & := -u u^T \mu - \pi \nu^T \mu - \nu \pi^T \mu \end{aligned}$$

$$\begin{aligned}
 & +\pi u^T \Sigma u + u \pi^T \Sigma u + uu^T \Sigma \pi + \pi v^T \Sigma \pi + v \pi^T \Sigma \pi + \pi \pi^T \Sigma v \\
 & -\text{diag}(\pi) \Sigma v - \text{diag}(u) \Sigma u + \text{diag}(v) \mu - \text{diag}(v) \Sigma \pi \\
 b_2(u) & := -uv^T \mu - vu^T \mu \\
 & +uu^T \Sigma u + u \pi^T \Sigma v + uv^T \Sigma \pi + \pi u^T \Sigma v + \pi v^T \Sigma u + vu^T \Sigma \pi + v \pi^T \Sigma u \\
 & -\text{diag}(u) \Sigma v - \text{diag}(v) \Sigma u \\
 b_{\frac{5}{2}}(u) & := -v v^T \mu + uu^T \Sigma v + uv^T \Sigma u + vu^T \Sigma u + \pi v^T \Sigma v + v \pi^T \Sigma v + v v^T \Sigma \pi \\
 & -\text{diag}(v) \Sigma v \\
 b_3(u) & := uv^T \Sigma v + vu^T \Sigma v + v v^T \Sigma u b_{\frac{7}{2}}(u) := v v^T \Sigma v.
 \end{aligned}$$

By Itô’s lemma,

$$\begin{aligned}
 d\|\mathbf{U}_t\|_{\pi}^2 & = 2\mathbf{U}_t^T \text{diag}(\pi)^{-1} \left[\text{diag}(\pi) - \pi \pi^T \right] \Gamma^T d\tilde{\mathbf{B}}_t + 2\langle \mathbf{U}_t, Q^T \mathbf{U}_t \rangle_{\pi} dt \\
 & + 2 \sum_{\ell=1}^4 \delta^{-\frac{\ell}{2}} \mathbf{U}_t^T \text{diag}(\pi)^{-1} A_{\frac{\ell}{2}}(\mathbf{U}_t) d\tilde{\mathbf{B}}_t \\
 & + 2 \sum_{\ell=2}^7 \delta^{-\frac{\ell}{2}} \mathbf{U}_t^T \text{diag}(\pi)^{-1} b_{\frac{\ell}{2}}(\mathbf{U}_t) dt \\
 & + \text{Tr} \left(\text{diag}(\pi)^{-1} \left[\text{diag}(\pi) - \pi \pi^T \right] \Gamma^T \Gamma \left[\text{diag}(\pi) - \pi \pi^T \right] \right) dt \\
 & + \text{Tr} \left(\text{diag}(\pi)^{-1} \sum_{\ell=1}^4 \delta^{-\frac{\ell}{2}} A_{\frac{\ell}{2}}(\mathbf{U}_t) \times \sum_{\ell=1}^4 \delta^{-\frac{\ell}{2}} A_{\frac{\ell}{2}}(\mathbf{U}_t)^T \right) dt.
 \end{aligned}$$

Note also that

$$|U_t^i| \leq C \delta^{\frac{1}{2}} \quad 1 \leq i \leq n, \tag{41}$$

for an appropriate constant C because $0 \leq Y_t^i \leq 1, 1 \leq i \leq n$. Each function

$$u \mapsto u^T \text{diag}(\pi)^{-1} b_{\frac{\ell}{2}}(u), \quad 2 \leq \ell \leq 7,$$

is a polynomial in u with total degree at most ℓ , and each function

$$u \mapsto \text{Tr} \left(\text{diag}(\pi)^{-1} A_{\frac{\ell'}{2}}(u) A_{\frac{\ell''}{2}}(u)^T \right), \quad 1 \leq \ell', \ell'' \leq 4,$$

is a polynomial in u with total degree at most $\ell' + \ell''$.

It follows that

$$\frac{d}{dt} \mathbb{E} \left[\|\mathbf{U}_t\|_{\pi}^2 \right] \leq -2\kappa \mathbb{E} \left[\|\mathbf{U}_t\|_{\pi}^2 \right] + C' \tag{42}$$

for all $t \geq 0$ for a suitable constant C' that does not depend on δ . Hence,

$$\sup_{t \geq 0} \mathbb{E} \left[\|\mathbf{U}_t\|_\pi^2 \right] \leq \frac{C'}{2\kappa} \tag{43}$$

(recall that $\mathbf{U}_0 = \mathbf{0}$).

Let $(\mathbf{V}_t)_{t \geq 0}$ be the solution of the stochastic differential equation

$$d\mathbf{V}_t = \left[\text{diag}(\pi) - \pi\pi^T \right] \Gamma^T d\tilde{\mathbf{B}}_t + Q^T \mathbf{V}_t dt$$

with $\mathbf{V}_0 = \mathbf{U}_0 = \mathbf{0}$. Note that $d(\mathbf{1}^T \mathbf{V}_t) = 0$ for all $t \geq 0$, and so $\langle \mathbf{V}_t, \pi \rangle_\pi = \mathbf{1}^T \mathbf{V}_t = 0$ for all $t \geq 0$. It is readily checked that

$$\mathbf{V}_t = \int_0^t \exp(Q^T(t-s)) \left[\text{diag}(\pi) - \pi\pi^T \right] \Gamma^T d\tilde{\mathbf{B}}_s.$$

So \mathbf{V} is a Gaussian process for which $\mathbf{E}[\mathbf{V}_t] = \mathbf{0}$ and

$$\mathbb{E}[\mathbf{V}_t \mathbf{V}_t^T] = \int_0^t \exp(Q^T s) \left(\text{diag}(\pi) - \pi\pi^T \right) \Sigma \left(\text{diag}(\pi) - \pi\pi^T \right) \exp(Qs) ds \tag{44}$$

for all $t \geq 0$. Consequently,

$$\sup_{t \geq 0} \mathbb{E} \left[|V_t^i|^p \right] < \infty \tag{45}$$

for $1 \leq i \leq n$ and $p \geq 0$.

In the notation above,

$$d(\mathbf{U}_t - \mathbf{V}_t) = Q^T (\mathbf{U}_t - \mathbf{V}_t) dt + \left[\sum_{\ell=1}^4 \delta^{-\frac{\ell}{2}} A_{\frac{\ell}{2}}(\mathbf{U}_t) \right] d\tilde{\mathbf{B}}_t + \left[\sum_{\ell=2}^7 \delta^{-\frac{\ell}{2}} b_{\frac{3}{2}}(\mathbf{U}_t) \right] dt.$$

Applying Itô's lemma and a combination of (41), (43) and (45), we can argue along the lines we followed to establish (42) to see that

$$\frac{d}{dt} \mathbb{E} \left[\|\mathbf{U}_t - \mathbf{V}_t\|_\pi^2 \right] \leq -2\kappa \mathbb{E} \left[\|\mathbf{U}_t - \mathbf{V}_t\|_\pi^2 \right] + \delta^{-1} C''$$

for all $t \geq 0$ for a suitable constant C'' that does not depend on δ . Hence,

$$\sup_{t \geq 0} \mathbb{E} \left[\|\mathbf{U}_t - \mathbf{V}_t\|_\pi^2 \right] \leq \delta^{-1} \frac{C''}{2\kappa}. \tag{46}$$

Now let \mathbf{Y}_∞ , \mathbf{U}_∞ and \mathbf{V}_∞ be random vectors that are distributed according to the equilibrium laws of $(\mathbf{Y}_t)_{t \geq 0}$, $(\mathbf{U}_t)_{t \geq 0}$ and $(\mathbf{V}_t)_{t \geq 0}$, respectively. Also let \hat{U}^i and \hat{V}^i be the i -th component of the vectors \mathbf{U}_∞ and \mathbf{V}_∞ respectively.

From (41), (43) and the linearity of the function b_1 ,

$$0 = Q^T \mathbb{E}[\mathbf{U}_\infty] + \delta^{-1} b_1(\mathbb{E}[\mathbf{U}_\infty]) + O(\delta^{-\frac{3}{2}}).$$

Noting that $\langle \mathbb{E}[\mathbf{U}_\infty], \pi \rangle_\pi = 0$ because $\langle \mathbf{U}_t, \pi \rangle_\pi = 0$ for all $t \geq 0$, we have from (39) that

$$\begin{aligned} \kappa \|\mathbb{E}[\mathbf{U}_\infty]\|_\pi^2 &\leq -\langle \mathbb{E}[\mathbf{U}_\infty], Q^T \mathbb{E}[\mathbf{U}_\infty] \rangle_\pi \\ &= \delta^{-1} \langle \mathbb{E}[\mathbf{U}_\infty], b_1(\mathbb{E}[\mathbf{U}_\infty]) \rangle_\pi + O(\delta^{-\frac{3}{2}}) \\ &\leq C''' \delta^{-1} \|\mathbb{E}[\mathbf{U}_\infty]\|_\pi^2 + O(\delta^{-\frac{3}{2}}) \end{aligned}$$

for a suitable constant C''' , and hence,

$$\mathbb{E}[\hat{U}^i] = O(\delta^{-\frac{3}{4}}), \quad 1 \leq i \leq n. \tag{47}$$

From (43), (45) and (46),

$$\left| \mathbb{E}[\hat{U}^i \hat{U}^j] - \mathbb{E}[\hat{V}^i \hat{V}^j] \right| = O(\delta^{-\frac{1}{2}}), \quad 1 \leq i, j \leq n. \tag{48}$$

Recall that $\chi(\delta)$ is the Lyapunov exponent, and that

$$\begin{aligned} \chi(\delta) &= \mu^T \mathbb{E}[\mathbf{Y}_\infty] - \frac{1}{2} \mathbb{E}[\mathbf{Y}_\infty^T \Sigma \mathbf{Y}_\infty] \\ &= \mu^T \mathbb{E}[\delta^{-\frac{1}{2}} \mathbf{U}_\infty + \pi + \delta^{-1} \nu] \\ &\quad - \frac{1}{2} \mathbb{E}\left[\left(\delta^{-\frac{1}{2}} \mathbf{U}_\infty + \pi + \delta^{-1} \nu \right)^T \Sigma \left(\delta^{-\frac{1}{2}} \mathbf{U}_\infty + \pi + \delta^{-1} \nu \right) \right] \\ &= \delta^{-\frac{1}{2}} \mu^T \mathbb{E}[\mathbf{U}_\infty] + \mu^T (\pi + \delta^{-1} \nu) \\ &\quad - \delta^{-1} \frac{1}{2} \mathbb{E}[\mathbf{U}_\infty^T \Sigma \mathbf{U}_\infty] - 2\delta^{-\frac{1}{2}} \frac{1}{2} \mathbb{E}[\mathbf{U}_\infty^T] \Sigma (\pi + \delta^{-1} \nu) \\ &\quad - \frac{1}{2} (\pi + \delta^{-1} \nu)^T \Sigma (\pi + \delta^{-1} \nu). \end{aligned}$$

Substituting in (47) and (48), and noting from (44) that the random vector \mathbf{V}_∞ is Gaussian with mean vector 0 and covariance matrix

$$\int_0^\infty \exp(Q^T s) \left(\text{diag}(\pi) - \pi \pi^T \right) \Sigma \left(\text{diag}(\pi) - \pi \pi^T \right) \exp(Qs) ds,$$

we conclude that

$$\begin{aligned} \chi(\delta) &= \left(\mu^T \pi - \frac{1}{2} \pi^T \Sigma \pi \right) \\ &\quad + \delta^{-1} \left[(\mu - \Sigma \pi)^T \nu - \frac{1}{2} \text{Tr} \left(\mathbb{E}[\mathbf{V}_\infty \mathbf{V}_\infty^T] \Sigma \right) \right] + \mathcal{O}(\delta^{-\frac{5}{4}}) \\ &= \left(\mu^T \pi - \frac{1}{2} \pi^T \Sigma \pi \right) + \delta^{-1} \left[(\mu - \Sigma \pi)^T \nu \right. \\ &\quad \left. - \frac{1}{2} \int_0^\infty \text{Tr}(\exp(Q^T s) (\text{diag}(\pi) - \pi \pi^T) \Sigma (\text{diag}(\pi) - \pi \pi^T) \exp(Qs) \Sigma) ds \right] \\ &\quad + \mathcal{O}(\delta^{-\frac{5}{4}}) \end{aligned}$$

as $\delta \rightarrow \infty$. □

Appendix E: Proof of Corollary 1

We now assume that the matrices Q and Σ are both real symmetric (Σ is, of course, always symmetric) and that they commute. Hence, as noted in the statement of the corollary, if $\lambda_1 \leq \dots \leq \lambda_{n-1} < \lambda_n = 0$ are the eigenvalues of Q with corresponding orthonormal eigenvectors ξ_1, \dots, ξ_n , where $\xi_n = \frac{1}{\sqrt{n}} \mathbf{1}$, then

$$Q = \sum_{k=1}^n \lambda_k \xi_k \xi_k^T$$

and it is possible to write the eigenvalues $\theta_1, \dots, \theta_n$ of Σ in some order so that

$$\Sigma = \sum_{k=1}^n \theta_k \xi_k \xi_k^T.$$

By the assumption that Q is symmetric, $\pi = \frac{1}{n} \mathbf{1} = \frac{1}{\sqrt{n}} \xi_n$. Therefore,

$$\mu^T \pi - \frac{1}{2} \pi^T \Sigma \pi = \bar{\mu} - \frac{1}{2n} \theta_n$$

where $\bar{\mu} = \frac{1}{n} \sum_i \mu_i$.

To find the unique vector ν that solves

$$\mathbf{1}^T \nu = 0 \quad \text{and} \quad Q^T \nu = - \left(\text{diag}(\pi) - \pi \pi^T \right) (\mu - \Sigma \pi),$$

write $\nu = \sum_{k=1}^n a_k \xi_k$. The condition $\mathbf{1}^T \nu = 0$ dictates that $a_n = 0$. The second condition becomes

$$\begin{aligned} \sum_{k=1}^{n-1} a_k \lambda_k \xi_k &= -\frac{1}{n} \left(I - \xi_n \xi_n^T \right) \left(\mu - \frac{1}{\sqrt{n}} \theta_n \xi_n \right) \\ &= -\frac{1}{n} \left(\sum_{k=1}^{n-1} \xi_k \xi_k^T \right) \left(\mu - \frac{1}{\sqrt{n}} \theta_n \xi_n \right) \\ &= -\frac{1}{n} \sum_{k=1}^{n-1} (\xi_k^T \mu) \xi_k, \end{aligned}$$

so that $a_k = -(\xi_k^T \mu)/(n\lambda_k)$ for $1 \leq k \leq n - 1$. It follows that

$$\begin{aligned} (\mu - \Sigma \pi)^T v &= -\left(\mu - \frac{1}{\sqrt{n}} \theta_n \xi_n \right)^T \left(\sum_{k=1}^{n-1} \frac{\xi_k^T \mu}{n\lambda_k} \xi_k \right) \\ &= -\sum_{k=1}^{n-1} \frac{(\xi_k^T \mu)^2}{n\lambda_k}. \end{aligned}$$

Lastly, the matrices inside the trace in the integral

$$\int_0^\infty \text{Tr} \left(\exp(Q^T s) \left(\text{diag}(\pi) - \pi \pi^T \right) \Sigma \left(\text{diag}(\pi) - \pi \pi^T \right) \exp(Qs) \Sigma \right) ds$$

commute and so the integral is

$$\begin{aligned} &\int_0^\infty \text{Tr} \left(\left(\text{diag}(\pi) - \pi \pi^T \right)^2 \Sigma^2 \exp(2Qs) \right) ds \\ &= \frac{1}{n^2} \int_0^\infty \text{Tr} \left(\left(I - \xi_n \xi_n^T \right) \left(\sum_{k=1}^n \theta_k^2 \xi_k \xi_k^T \right) \left(\sum_{k=1}^n \exp(2s\lambda_k) \xi_k \xi_k^T \right) \right) ds \\ &= \frac{1}{n^2} \int_0^\infty \text{Tr} \left(\sum_{k=1}^{n-1} \theta_k^2 \exp(2s\lambda_k) \xi_k \xi_k^T \right) ds \\ &= \frac{1}{n^2} \int_0^\infty \left(\sum_{k=1}^{n-1} \theta_k^2 \exp(2s\lambda_k) \right) ds \\ &= -\frac{1}{n^2} \sum_{k=1}^{n-1} \frac{\theta_k^2}{2\lambda_k}. \end{aligned}$$

Therefore, our asymptotic approximation of $\chi(\delta)$ is

$$\left(\bar{\mu} - \frac{1}{2n}\theta_n\right) - \frac{1}{\delta} \left[\sum_{k=1}^{n-1} \frac{1}{n\lambda_k} \left((\xi_k^T \mu)^2 - \frac{1}{4n}\theta_k^2 \right) \right] + O(\delta^{-5/4})$$

as $\delta \rightarrow 0$.

□

Appendix F: Proof of Theorem 3

To show that Theorem 3 follows from Corollary 1, we show that the matrix entries of each irreducible representation belong to a common eigenspace of Q and Σ . Suppose that c is a class function and the matrix C is given by $C_{g,h} = c(gh^{-1})$. Recall from (27) that

$$c(g) = \frac{1}{\#G} \sum_{\kappa \in \tilde{G}} \tilde{c}(\kappa)\kappa(g)^*.$$

Therefore,

$$C_{g,h} = \frac{1}{\#G} \sum_{\kappa \in \tilde{G}} \tilde{c}(\kappa)\kappa(gh^{-1})^*.$$

If κ is associated with the irreducible representation $\rho \in \hat{G}$, then

$$\kappa(gh^{-1}) = \text{Tr}(\rho(gh^{-1})) = \text{Tr}(\rho(g)\rho(h)^\dagger) = \sum_{i,j=1}^{d_\rho} \rho_{ij}(g)\rho_{ij}(h)^* =: (\Xi(\kappa))_{gh},$$

where \dagger denotes the Hermitian conjugate of a matrix. Set $\Pi_\kappa := (d_\kappa/\#G)\Xi(\kappa)$. The $\#G \times \#G$ matrix Π_κ is Hermitian, and it follows from (24) that $\Pi_\kappa^2 = \Pi_\kappa$, so that Π_κ is the projection onto a d_κ^2 -dimensional subspace. Again by (24), the matrices $\Pi_{\kappa'}$ and $\Pi_{\kappa''}$ are orthogonal for distinct κ', κ'' . Thus,

$$C = \sum_{\kappa \in \tilde{G}} \frac{\tilde{c}(\kappa)}{d_\kappa} \Pi_\kappa.$$

This expression is nothing other than the spectral decomposition of the matrix C . It shows that $\tilde{c}(\kappa)/d_\kappa$ is an eigenvalue of C with multiplicity d_κ^2 . In summary, for each $\kappa \in \tilde{G}$ there are eigenvalues $\tilde{q}(\kappa)/d_\kappa$ of Q and $\tilde{s}(\kappa)/d_\kappa$ of Σ , each with multiplicity d_κ^2 .

Therefore, in the notation of Corollary 1,

$$\sum_{k=1}^{n-1} \frac{\theta_k^2}{\lambda_k} = \sum_{\kappa \neq \kappa_{\text{tr}}} d_\kappa^2 \left(\frac{\tilde{s}(\kappa)}{d_\kappa} \right)^2 \frac{d_\kappa}{\tilde{q}(\kappa)} = \sum_{\kappa \neq \kappa_{\text{tr}}} d_\kappa \frac{\tilde{s}(\kappa)^2}{\tilde{q}(\kappa)}.$$

Similarly, we can split the sum

$$\sum_{k=1}^{n-1} \frac{1}{\lambda_k} (\xi_k^T \mu)^2$$

up into contributions from each non-trivial character κ that are of the form

$$\frac{d_\kappa}{\tilde{q}(\kappa)} \sum_k (\xi_k^T \mu)^2,$$

where the sum is over the indices that correspond to eigenvectors in the range of the projection Π_κ . By pairwise orthogonality of the matrices Π_κ and the fact the μ is real, this last quantity is equal to

$$\begin{aligned} \frac{d_\kappa}{\tilde{q}(\kappa)} \|\Pi_\kappa \mu\|^2 &= \frac{d_\kappa}{\tilde{q}(\kappa)} \left(\frac{d_\kappa}{\#G} \right) \sum_{g,h \in G} \mu(g) \kappa(gh^{-1}) \mu(h) \\ &= \frac{d_\kappa}{\tilde{q}(\kappa)} \|\mu\|_\kappa^2, \end{aligned}$$

by definition of $\|\mu\|_\kappa$. □

Appendix G: Proof of Theorem 4

We first recall some notation. For $0 \leq r, \ell \leq k + 1$,

$$\begin{aligned} Z_r &= G_1 \otimes \cdots \otimes G_{r-1} \otimes \{\mathbf{id}_r\} \otimes \cdots \otimes \{\mathbf{id}_k\}, \\ \bar{Z}_\ell &= \{\mathbf{id}_1\} \otimes \cdots \otimes \{\mathbf{id}_\ell\} \otimes G_{\ell+1} \otimes \cdots \otimes G_k \end{aligned}$$

and

$$\ell(g) := \min\{j : g_j \neq \mathbf{id}_j\}.$$

The displacement associated with $g \in G$ moves between two patches that are in the same metapatch at scale $\ell(g)$ but different metapatches at scales $\ell(g) + 1, \ell(g) + 2, \dots$. Recall also that $\#G_r = n_r$, $N_r = \#Z_r = \prod_{j=1}^{r-1} n_j$ and $\bar{N}_\ell = \#\bar{Z}_\ell = \prod_{j=\ell+1}^k n_j$.

Writing $\mathbf{1}_j$ for the trivial character on G_j , put

$$\begin{aligned} \tilde{Z}_r &:= \tilde{G}_1 \otimes \cdots \otimes \tilde{G}_{r-1} \otimes \{\mathbf{1}_r\} \otimes \cdots \otimes \{\mathbf{1}_k\} \\ &= \{\kappa \in \tilde{G} : \kappa(g) = 1 \ \forall g \in \tilde{Z}_{r-1}\} \end{aligned}$$

and

$$r(\kappa) := \max\{j : \kappa \notin \tilde{Z}_j\}.$$

The following orthogonality property of characters:

$$\sum_{g \in G} \kappa'(g)\kappa''(g)^* = \begin{cases} \#G, & \text{if } \kappa' = \kappa'', \\ 0, & \text{otherwise.} \end{cases}$$

leads to the relation

$$\sum_{g \in \tilde{Z}_r} \kappa(g) = \begin{cases} \tilde{N}_r, & \text{if } \kappa \in \tilde{Z}_{r+1}, \\ 0, & \text{otherwise.} \end{cases}$$

We denote this quantity, as a function of κ , by $\tilde{N}_r \delta_{\tilde{Z}_{r+1}}(\kappa)$.

Define the function $f_\ell : G \rightarrow \mathbb{C}$ by setting $f_\ell(g) = 1$ if $\ell(g) = \ell$ and $f_\ell(g) = 0$ otherwise. Then,

$$\begin{aligned} \tilde{f}_\ell(\kappa) &= \sum_{g:\ell(g)=\ell} \kappa(g) \\ &= \sum_{g \in \tilde{Z}_{\ell-1}} \kappa(g) - \sum_{g \in \tilde{Z}_\ell} \kappa(g) \\ &= \tilde{N}_{\ell-1} \delta_{\tilde{Z}_\ell}(\kappa) - \tilde{N}_\ell \delta_{\tilde{Z}_{\ell+1}}(\kappa). \end{aligned}$$

Our assumption that $s(g) = s_{\ell(g)}$ implies that $s(g) = \sum_{\ell=1}^{k+1} s_\ell f_\ell(g)$. Since $\kappa \in \tilde{Z}_\ell$ if and only if $r(\kappa) + 1 \leq \ell$, it follows by linearity that

$$\begin{aligned} \tilde{s}(\kappa) &= \sum_{\ell=1}^{k+1} s_\ell \left(\tilde{N}_{\ell-1} \delta_{\tilde{Z}_\ell}(\kappa) - \tilde{N}_\ell \delta_{\tilde{Z}_{\ell+1}}(\kappa) \right) \\ &= \sum_{\ell=r(\kappa)+1}^{k+1} s_\ell \tilde{N}_{\ell-1} - \sum_{\ell=r(\kappa)}^{k+1} s_\ell \tilde{N}_\ell \\ &= \sum_{\ell=r(\kappa)}^k s_{\ell+1} \tilde{N}_\ell - \sum_{\ell=r(\kappa)}^k s_\ell \tilde{N}_\ell \\ &= \sum_{\ell=r(\kappa)}^k (s_{\ell+1} - s_\ell) \tilde{N}_\ell, \end{aligned}$$

where we used the convention $\bar{N}_{k+1} = 0$.

Turning to q , we have $q(g) = q_{\ell(g)}$ for $g \neq \mathbf{id}_G$ and $q(\mathbf{id}_G) = q_{k+1} = -\sum_{\ell=1}^k q_{\ell}(\bar{N}_{\ell-1} - \bar{N}_{\ell})$. By the same argument as above,

$$\begin{aligned} \tilde{q}(\kappa) &= \sum_{\ell=r(\kappa)+1}^{k+1} q_{\ell} \bar{N}_{\ell-1} - \sum_{\ell=r(\kappa)}^{k+1} q_{\ell} \bar{N}_{\ell} \\ &= \sum_{\ell=r(\kappa)+1}^{k+1} q_{\ell}(\bar{N}_{\ell-1} - \bar{N}_{\ell}) - q_{r(k)} \bar{N}_{r(k)} \\ &= \sum_{\ell=r(\kappa)+1}^k q_{\ell}(\bar{N}_{\ell-1} - \bar{N}_{\ell}) - \sum_{\ell=1}^k q_{\ell}(\bar{N}_{\ell-1} - \bar{N}_{\ell}) - q_{r(k)} \bar{N}_{r(k)} \\ &= -\sum_{\ell=1}^{r(\kappa)} q_{\ell}(\bar{N}_{\ell-1} - \bar{N}_{\ell}) - q_{r(k)} \bar{N}_{r(k)} \\ &= -\sum_{\ell=1}^{r(\kappa)-1} q_{\ell}(\bar{N}_{\ell-1} - \bar{N}_{\ell}) - q_{r(k)} \bar{N}_{r(k)-1}. \end{aligned}$$

Lastly, for an arbitrary function μ we need to evaluate

$$\frac{1}{\#G} \sum_{\kappa:r(\kappa)=r} \|\mu\|_{\kappa}^2.$$

We do that by using the following lemma that follows immediately from orthogonality of characters.

Lemma 1 *Let H and K be two finite Abelian groups. For $f : H \otimes K \rightarrow \mathbb{C}$,*

$$\sum_{\kappa \in \tilde{H}} \left| \sum_{(h,k) \in H \otimes K} f(h,k) \kappa(h) \right|^2 = \#H \sum_{h \in H} \left| \sum_{k \in K} f(h,k) \right|^2.$$

Using Lemma 1 applied to the decomposition of G as $Z_r \otimes \bar{Z}_{r-1}$, we get

$$\sum_{\kappa \in \tilde{Z}_r} \|\mu\|_{\kappa}^2 = \frac{N_r}{\#G} \sum_{g \in Z_r} \left(\sum_{z \in \bar{Z}_{r-1}} \mu(gz) \right)^2.$$

Further decomposing Z_{r+1} as $Z_r \otimes G_r$ and \bar{Z}_{r-1} as $\bar{Z}_r \otimes G_r$, and using $N_{r+1} = n_r N_r$ gives

$$\begin{aligned} \sum_{\kappa:r(\kappa)=r} \|\mu\|_{\kappa}^2 &= \sum_{\kappa \in \tilde{Z}_{r+1}} \|\mu\|_{\kappa}^2 - \sum_{\kappa \in \tilde{Z}_r} \|\mu\|_{\kappa}^2 \\ &= \frac{N_{r+1}}{\#G} \sum_{g \in Z_{r+1}} \left(\sum_{z \in \tilde{Z}_r} \mu(gz) \right)^2 - \frac{N_r}{\#G} \sum_{g \in Z_r} \left(\sum_{z \in \tilde{Z}_{r-1}} \mu(gz) \right)^2 \\ &= \frac{n_r N_{r+1}}{\#G} \sum_{g \in Z_r} \left(\frac{1}{n_r} \sum_{h \in G_r} \left(\sum_{z \in \tilde{Z}_r} \mu(ghz) \right)^2 \right. \\ &\quad \left. - \left(\frac{1}{n_r} \sum_{h \in G_r} \sum_{z \in \tilde{Z}_r} \mu(ghz) \right)^2 \right). \end{aligned}$$

To turn the remaining sums into averages, we need to pull out a factor of $N_r \bar{N}_r^2$, leaving us with $n_r N_{r+1} N_r \bar{N}_r^2 = \prod_{\ell=1}^k n_{\ell}^2 = \#G^2$. Therefore, recalling that

$$v_{\mu}(r) = \frac{1}{N_r} \sum_{g \in Z_r} \left(\frac{1}{n_r} \sum_{h \in G_r} \left(\frac{1}{\bar{N}_r} \sum_{z \in \tilde{Z}_r} \mu(ghz) \right)^2 - \left(\frac{1}{n_r} \sum_{h \in G_r} \frac{1}{\bar{N}_r} \sum_{z \in \tilde{Z}_r} \mu(ghz) \right)^2 \right),$$

we have

$$\sum_{\kappa:r(\kappa)=r} \|\mu\|_{\kappa}^2 = \#G \times v_{\mu}(r).$$

The theorem follows once we note that

$$\#\{\kappa : r(\kappa) = r\} = \#(\tilde{Z}_{r+1} \setminus \tilde{Z}_r) = N_{r+1} - N_r.$$

□

References

Adler FR (1992) The effects of averaging on the basic reproduction ratio. *Math Biosci* 111:89–98

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

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

Bhattacharya RN (1978) Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *Ann Probab* 6(4):541–553. ISSN 00911798. <http://www.jstor.org/stable/2243121>

Bogachev VI, Röckner M, Stannat W (2002) Uniqueness of solutions of elliptic equations and uniqueness of invariant measures of diffusions. *Sbornik: Math* 193(7):945. <http://stacks.iop.org/1064-5616/193/i=7/a=A01>

Bogachev Vladimir I, Krylov Nikolai V, Röckner Michael (2009) Elliptic and parabolic equations for measures. *Russ Math Surv* 64(6):973. <http://stacks.iop.org/0036-0279/64/i=6/a=R02>

- Boyce MS, Haridas CV, Lee CT, the NCEAS Stochastic Demography Working Group (2006) Demography in an increasingly variable world. *Trends Ecol Evol* 21:141–148
- Cantrell RS, Cosner C (1991) The effects of spatial heterogeneity in population dynamics. *J Math Biol* 29:315–338
- Cantrell RS, Cosner C, Lou Y (2006) Movement toward better environments and the evolution of rapid diffusion. *Math Biosci* 204(2):199–214
- Cantrell RS, Cosner C (1989) Diffusive logistic equations with indefinite weights: population models in disrupted environments. *Proc Royal Soc Edinburgh Section A Math* 112(3–4):293–318
- Cantrell RS, Cosner C, Deangelis DL, Padron V (2007) The ideal free distribution as an evolutionarily stable strategy. *J Biol Dyn* 1:249–271
- Chesson PL (2000) General theory of competitive coexistence in spatially-varying environments. *Theor Popul Biol* 58:211–237
- Da Prato G, Zabczyk J (1996) *Ergodicity for Infinite-Dimensional Systems*, London Mathematical Society Lecture Note Series, vol 229. Cambridge University Press, Cambridge
- DeAngelis DL, Wolkowicz GSK, Lou Y, Jiang Y, Novak M, Svanbäck R, Araújo MS, Jo YS, Cleary EA (2011) The effect of travel loss on evolutionarily stable distributions of populations in space. *Am Nat* 178:15–29
- Delibes M, Gaona P, Ferreras P (2001) Effects of an attractive sink leading into maladaptive habitat selection. *Am Nat* 158:277–285
- Dennis B, Munholland PL, Scott JM (1991) Estimation of growth and extinction parameters for endangered species. *Ecol Monogr* 61:115–143
- Diaconis P (1988) *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol 11. Institute of Mathematical Statistics, Hayward
- Diamond JM (1975) The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. *Biol Conserv* 7:129–146
- Dias PC (1996) Sources and sinks in population biology. *Trends Ecol Evol* 11:326–330
- Dockery J, Hutson V, Mischaikow K, Pernarowski M (1998) The evolution of slow dispersal rates: a reaction diffusion model. *J Math Biol* 37:61–83
- Durrett R, Remenik D (2012) Evolution of the dispersal distance. *J Math Biol* (in press)
- Fahrig L (1997) Relative effects of habitat loss and fragmentation on population extinction. *J Wildl Manag* 61:603–610
- Fahrig L (2002) Effect of habitat fragmentation on the extinction threshold: a synthesis. *Ecol Appl* 12:346–353
- Foley P (1994) Predicting extinction times from environmental stochasticity and carrying capacity. *Conserv Biol* 8:124–137
- Fretwell SD, Lucas HL Jr. (1970) On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica* 19:16–36
- Gardiner CW (2004) *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*, Series in synergetics, vol 13, 4th edn. Springer, Berlin
- Geiß C, Manthey R (1994) Comparison theorems for stochastic differential equations in finite and infinite dimensions. *Stoch Processes Appl* 53:23–35
- Gilpin ME (1988) A comment on Quinn and Hastings: extinction in subdivided habitats. *Conserv Biol* 2:290–292
- Gonzalez A, Holt RD (2002) The inflationary effects of environmental fluctuations in source-sink systems. *Proc Natl Acad Sci* 99:14872–14877
- Gradshteyn IS, Ryzhik IM (2007) *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, 7th edn, 2007. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger
- Harper DGC (1982) Competitive foraging in mallards: “ideal free” ducks. *Animal Behav* 30:575–584
- Harrison S, Quinn JF (1989) Correlated environments and the persistence of metapopulations. *Oikos* 56:293–298
- Hastings A (1983) Can spatial variation alone lead to selection for dispersal?. *Theor Popul Biol* 24:244–251
- Holt RD (1985) Patch dynamics in two-patch environments: some anomalous consequences of an optimal habitat distribution. *Theor Popul Biol* 28:181–208
- Holt RD (1997) On the evolutionary stability of sink populations. *Evol Ecol* 11:723–731
- Holt RD, McPeck MA (1996) Chaotic population dynamics favors the evolution of dispersal. *Am Nat* 148(44):709–718

- Hutson V, Mischaikow K, Poláčik P (2001) The evolution of dispersal rates in a heterogeneous time-periodic environment. *J Math Biol* 43:501–533
- Ikeda N, Watanabe S (1989) Stochastic differential equations and diffusion processes, North-Holland Mathematical Library, vol 24, 2nd edn. North-Holland Publishing Co, Amsterdam
- Jansen VAA, Yoshimura J (1998) Populations can persist in an environment consisting of sink habitats only. *Proc Natl Acad Sci USA* 95:3696–3698
- Keagy J, Schreiber SJ, Cristol DA (2005) Replacing sources with sinks: when do populations go down the drain?. *Restor Ecol* 13:529–535
- Khas'minskii RZ (1960) Ergodic properties of recurrent diffusion processes and stabilization of the solution to the cauchy problem for parabolic equations. *Theory Probab Appl* 5(2):179–196, 1960. ISSN 0040585X. doi:[10.1137/1105016](https://doi.org/10.1137/1105016).
- Kirkland S, Li CK, Schreiber SJ (2006) On the evolution of dispersal in patchy landscapes. *SIAM J Appl Math* 66:1366–1382
- Kreuzer MP, Huntly NJ (2003) Habitat-specific demography: evidence for source-sink population structure in a mammal, the pika. *Oecologia* 134:343–349
- Lande R, Engen S, Sæther BE (2003) Stochastic population dynamics in ecology and conservation: an introduction. Oxford University Press, Oxford
- Levin SA, Cohen D, Hastings A (1984) Dispersal strategies in patchy environments. *Theor Popul Biol* 26:165–191
- Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull ESA* 15:237–240
- Lonsdale WM (1993) Rates of spread of an invading species-mimosa pigra in northern Australia. *J Ecol* 81:513–521
- Lundberg P, Ranta E, Ripa J, Kaitala V (2000) Population variability in space and time. *Trends Ecol Evol* 15:460–464
- Matthews DP, Gonzalez A (2007) The inflationary effects of environmental fluctuations ensure the persistence of sink metapopulations. *Ecology* 88:2848–2856
- May RM (1975) Stability and complexity in model ecosystems, 2nd edn. Princeton University Press, Princeton
- McPeck MA, Holt RD (1992) The evolution of dispersal in spatially and temporally varying environments. *Am Nat* 6:1010–1027
- Metz JAJ, de Jong TJ, Klinkhamer PGL (1983) What are the advantages of dispersing; a paper by Kuno extended. *Oecologia* 57:166–169
- Murphy MT (2001) Source-sink dynamics of a declining eastern kingbird population and the value of sink habitats. *Conserv Biol* 15:737–748
- Oksanen T, Power ME, Oksanen L (1995) Ideal free habitat selection and consumer-resource dynamics. *Am Nat* 146:565–585
- Petchey OL, Gonzalez A, Wilson HB (1997) Effects on population persistence: the interaction between environmental noise colour, intraspecific competition and space. *Proc Biol Sci* 264:1841–1847
- Philippi T, Seger J (1989) Hedging one's evolutionary bets, revisited. *Trends Ecol Evol* 4:41–44
- Pulliam HR (1988) Sources, sinks, and population regulation. *Am Nat* 132:652–661
- Quinn JF, Hastings A (1987) Extinction in subdivided habitats. *Conserv Biol* 1:198–209
- Remeš V (2000) How can maladaptive habitat choice generate source-sink population dynamics?. *Oikos* 91:579–582
- Roy M, Holt RD, Barfield M (2005) Temporal autocorrelation can enhance the persistence and abundance of metapopulations comprised of coupled sinks. *Am Nat* 166:246–261
- Ruelle D (1979) Analyticity properties of the characteristic exponents of random matrix products. *Adv Math* 32:68–80
- Schmidt KA (2004) Site fidelity in temporally correlated environments enhances population persistence. *Ecol Lett* 7:176–184
- Schreiber SJ (2010) Interactive effects of temporal correlations, spatial heterogeneity, and dispersal on population persistence. *Proc Royal Soc Biol Sci* 277:1907–1914
- Schreiber SJ, Lloyd-Smith JO (2009) Invasion dynamics in spatially heterogeneous environments. *Am Nat* 174:490–505
- Schreiber SJ, Saltzman E (2009) Evolution of predator and prey movement into sink habitats. *Am Nat* 174:68–81

- Schreiber SJ, Vejdani M (2006) Handling time promotes the coevolution of aggregation in predator–prey systems. *Proc Royal Soc Biol Sci* 273:185–191
- Schreiber SJ, Fox LR, Getz WM (2000) Coevolution of contrary choices in host–parasitoid systems. *American Naturalist* 637–648
- Serre JP (1977) Linear representations of finite groups. Springer, New York. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, vol 42
- Slatkin M (1974) Hedging one’s evolutionary bets. *Nature* 250:704–705
- Talay Denis (1991) Approximation of upper Lyapunov exponents of bilinear stochastic differential systems. *SIAM J Numer Anal* 28(4):1141–1164. ISSN 00361429. <http://www.jstor.org/stable/2157791>
- Tuljapurkar S (1990) Population dynamics in variable environments. Springer, New York
- Turelli M (1978) Random environments and stochastic calculus. *Theor Popul Biol* 12:140–178
- van Baalen M, Sabelis MW (1999) Nonequilibrium population dynamics of “ideal and free” prey and predators. *Am Nat* 154:69–88
- Wilbur HM, Rudolf VHW (2006) Life-history evolution in uncertain environments: bet hedging in time. *Am Nat* 168:398–411
- Wilcox BA, Murphy DD (1985) Conservation strategy: the effects of fragmentation on extinction. *Am Nat* 125:879–887