

Evolution of unconditional dispersal in periodic environments

Sebastian J. Schreiber & Chi-Kwong Li

To cite this article: Sebastian J. Schreiber & Chi-Kwong Li (2011) Evolution of unconditional dispersal in periodic environments, Journal of Biological Dynamics, 5:2, 120-134, DOI: 10.1080/17513758.2010.525667

To link to this article: <https://doi.org/10.1080/17513758.2010.525667>



Copyright Taylor and Francis Group, LLC



Published online: 19 Feb 2011.



Submit your article to this journal [↗](#)



Article views: 474



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

Evolution of unconditional dispersal in periodic environments

Sebastian J. Schreiber^{a,b*} and Chi-Kwong Li^c

^a*Department of Evolution and Ecology, University of California, One Shields Avenue, Davis, CA 95616, USA;* ^b*Center for Population Biology, University of California, One Shields Avenue, Davis, CA 95616, USA;* ^c*Department of Mathematics, College of William and Mary, Williamsburg, VA, USA*

(Received 13 November 2009; final version received 24 August 2010)

Organisms modulate their fitness in heterogeneous environments by dispersing. Prior work shows that there is selection against ‘unconditional’ dispersal in spatially heterogeneous environments. ‘Unconditional’ means individuals disperse at a rate independent of their location. We prove that if within-patch fitness varies spatially and between two values temporally, then there is selection for unconditional dispersal: any evolutionarily stable strategy (ESS) or evolutionarily stable coalition (ESC) includes a dispersive phenotype. Moreover, at this ESS or ESC, there is at least one sink patch (i.e. geometric mean of fitness less than one) and no source patches (i.e. geometric mean of fitness greater than one). These results coupled with simulations suggest that spatial-temporal heterogeneity is due to abiotic forcing result in either an ESS with a dispersive phenotype or an ESC with sedentary and dispersive phenotypes. In contrast, the spatial-temporal heterogeneity due to biotic interactions can select for higher dispersal rates that ultimately spatially synchronize population dynamics.

Keywords: ecology and evolutionary biology; population dynamics

1. Introduction

Plants and animals often live in landscapes where environmental conditions vary in space and time. These environmental conditions may include abiotic factors such as light, space, and nutrient availability or biotic factors such as prey, competitors, and predators. Since the fecundity and survivorship of an individual depends on these factors, an organism may decrease or increase its fitness by dispersing across the environment. Understanding how natural selection acts on dispersal in heterogeneous environments has been the focus of much theoretical and empirical work [1–3,5–8,10,11,14,15,18–20,22,24–28,30].

In spatially heterogeneous environments, the general consensus is that there is selection against unconditional dispersal [7,15,22,26,27]. Here, unconditional refers to the assumption that individuals disperse at a rate independent of their location. Using reaction-diffusion equations, Dockery *et al.* [7] proved that for two competing populations only differing in their diffusion constant, the population with the larger diffusion constant is excluded. In [22,27], similar results were proved for populations with non-overlapping generations living in a patchy environment. Alternatively,

*Corresponding author Email: sschreiber@ucdavis.edu

in temporally but not spatially heterogeneous environments, Hutson *et al.* [20] proved that dispersal rates are a selectively neutral trait for reaction-diffusion models and, thereby, confirmed the numerical observations of McPeck and Holt [26] for discrete-time, two-patch models. These results imply that the slightest cost of dispersal would result in selection against dispersal in purely temporally heterogeneous environments.

When there is a mixture of spatial and temporal heterogeneity, the interaction between competing dispersal phenotypes becomes more subtle. Numerically simulating discrete-time, two-patch models, McPeck and Holt [26], Parvinen [27], and Mathias *et al.* [25] found that more dispersive phenotypes could displace more sedentary phenotypes for certain forms of spatial-temporal heterogeneity, while evolutionarily stable coalitions (ESCs) of sedentary and dispersal phenotypes are possible for other forms of spatial-temporal heterogeneity. Hutson *et al.* [20] proved that similar phenomena could occur for reaction diffusion equations. However, analytical criteria distinguishing these outcomes remain elusive.

In this article, we consider the evolution of dispersal for a general class of multi-patch difference equations varying periodically in time. This periodic variation can be either due to biotic interactions or abiotic forcing. Our main goals are to analytically identify potential evolutionarily stable strategies (ESSs) or coalitions for dispersal, characterize the spatial-temporal patterns of fitness generated by populations playing these dispersal strategies, and use our results to compare evolutionary outcomes for oscillations due to abiotic forcing versus oscillations due to biotic interactions.

2. Models and assumptions

To understand the formation of ESCs of dispersive phenotypes, we consider a population consisting of m phenotypes dispersing in an environment consisting of n patches. Let $x_i^j(t)$ denote the abundance of phenotype i in patch j . The fitness of an individual in patch j is assumed to be of the form $f^j(t, \sum_{i=1}^m x_i^j(t))$. In particular, this assumption implies that phenotypes only differ demographically in their propensity to disperse. Moreover, we assume that $f^j(t, \cdot)$ is of period p . This periodicity may arise from exogenous forcing or biological interactions (e.g. over compensating density dependence or predator-prey interactions). While we do not explicitly model interactions with other species, our formulation is sufficiently general to be viewed as the dynamics of a particular species embedded within a web of interacting species.

We assume that the fraction of phenotype i dispersing from any given patch is d_i . Of the individuals dispersing from patch j , a fraction S_{kj} go to patch k . We call $S = (S_{kj})$ the dispersal matrix and it characterizes how dispersing individuals are redistributed across the landscape. Under these assumptions, the interacting phenotypes exhibit the following population dynamics:

$$x_i^j(t+1) = (1 - d_i) f^j \left(t, \sum_{l=1}^m x_l^j(t) \right) x_i^j(t) + d_i \sum_{k=1}^n S_{jk} f^k \left(t, \sum_{l=1}^m x_l^k(t) \right) x_i^k(t).$$

To express this model more succinctly, let $\mathbf{x}^j = (x_1^j, \dots, x_m^j)$ be the vector of abundances of the m phenotypes in patch j , $\mathbf{x}_i = (x_i^1, \dots, x_i^n)^T$ (where T denotes transpose) be the vector of abundances of strategy i across the n patches, and $\|\mathbf{x}^j\| = \sum_{i=1}^m x_i^j$ denote the total abundance of individuals in patch j . Let \mathbf{x} be the matrix with entries x_i^j , $\mathbf{F}(t, \mathbf{x})$ be a diagonal matrix whose j th diagonal element is $f^j(t, \|\mathbf{x}^j\|)$, and $\mathbf{S}(d_i) = (1 - d_i)I + d_i S$ where I is the identity matrix. With this notation, the model is represented more succinctly as

$$\mathbf{x}_i(t+1) = \mathbf{S}(d_i) \mathbf{F}(t, \mathbf{x}(t)) \mathbf{x}_i(t), \quad i = 1, \dots, m. \quad (1)$$

About this model, we make three assumptions throughout this manuscript.

A1: The dispersal matrix \mathbf{S} is irreducible and column stochastic (i.e. S has non-negative entries, and the entries of each column sum up to one).

A2: Equation (1) has a positive period- p point $\hat{\mathbf{x}}(1), \dots, \hat{\mathbf{x}}(p)$, i.e. $\|\hat{\mathbf{x}}_i(t)\| > 0$ and $\|\hat{\mathbf{x}}^j(t)\| > 0$ for all i, j, t , and

$$\prod_{t=1}^p \mathbf{S}(d_i) \mathbf{F}(t, \hat{\mathbf{x}}(t)) \hat{\mathbf{x}}_i(1) = \hat{\mathbf{x}}_i(1) \quad \text{for all } i. \quad (2)$$

A3: $\mathbf{F}(t, \mathbf{x})$ is continuous in \mathbf{x} .

Assumption A1 ensures that individuals or their decedents can move from any patch to any other patch after sufficiently many generations and there is no direct cost to dispersal. Assumption A2 implies that the phenotypes are coexisting in a periodic fashion and occupying all the patches. Assumption A3 is a basic regularity assumption met by most models.

3. Main results

We are primarily interested in understanding when a periodically fluctuating collection of phenotypes cannot be invaded by any other phenotype, and what are the spatial-temporal patterns of fitness at these potential evolutionary end states. To state our main results precisely, we need two sets of definitions.

Invasion rates, Nash equilibria, and evolutionary stability

Let $\mathbf{d} = (d_1, \dots, d_m)$ the coalition of strategies played by the resident population. If we added a ‘mutant’ phenotype with dispersal rate \tilde{d} into the population and this mutant population $\mathbf{y} = (y^1, \dots, y^n)$ is very small, then the resident’s population dynamics are initially barely influenced by the mutant’s population dynamics and the mutant’s dynamics are approximately given by a linear model

$$\mathbf{y}(t+1) = \mathbf{S}(\tilde{d}) \mathbf{F}(t, \mathbf{x}(t)) \mathbf{y}(t).$$

By standard linearization theorems (see, e.g. [21]), this approximation is valid when the size of the mutant population is small and the periodic point in A2 is hyperbolic.

The initial fate of the mutant population depends on the *invasion rate of strategy \tilde{d} against the resident population playing strategies \mathbf{d}* :

$$\mathcal{I}(\mathbf{d}; \tilde{d}) = \varrho \left(\prod_{t=1}^p \mathbf{S}(\tilde{d}) \mathbf{F}(t, \hat{\mathbf{x}}(t)) \right)^{1/p},$$

where $\mathbf{x}(t)$ has period p and $\varrho(\mathbf{A})$ corresponds to the largest eigenvalue for a non-negative matrix \mathbf{A} . If the invasion rate $\mathcal{I}(\mathbf{d}; \tilde{d})$ is greater than one, then the mutant population grows when its size is small. The ultimate fate of the mutant and resident after the mutant increases depends on the details of the full nonlinear model of the resident and invader dynamics. In particular, following an invasion the asymptotic dynamics in general may no longer be periodic (i.e. satisfy A2). There are many cases where post-invasion dynamics will remain periodic (e.g. periodically forced competitive systems as discussed in Section 4.1 or when there is attractor inheritance for a sufficiently small mutation [13]).

If the invasion rate $\mathcal{I}(\mathbf{d}; \tilde{d})$ is less than one, then the mutant population declines exponentially when rare and it cannot invade. Finally, if $\mathcal{I}(\mathbf{d}; \tilde{d}) = 1$, then a mutant may increase or decrease when rare depending on the details of the nonlinearities of the full model. However, if it increases, then it does so at a subexponential rate and, therefore, may be highly vulnerable to stochastic extinction. An important consequence of our assumption A2 is that the invasion rate of mutants with the same strategy as a resident equals one, i.e. $\mathcal{I}(\mathbf{d}; \tilde{d}) = 1$ whenever $\tilde{d} = d_i$ for some i .

Using the invasion rates of mutant strategies, we can define several concepts associated with evolutionary stability. A coalition of strategies \mathbf{d} with $m > 1$ is a *mixed Nash equilibrium* provided that

$$\mathcal{I}(\mathbf{d}; \tilde{d}) \leq 1 \quad \text{for all } \tilde{d} \in [0, 1]. \tag{3}$$

In other words, a mixed Nash equilibrium is a set of strategies in which all mutant strategies are unlikely to invade due to vulnerability to stochastic extinction. When $m = 1$, a strategy satisfying Equation (3) is called simply a *Nash equilibrium*. Under the stronger assumption that rare mutants decline exponentially to extinction (i.e. $\mathcal{I}(\mathbf{d}; \tilde{d}) < 1$ for all $\tilde{d} \notin \{d_1, \dots, d_m\}$), \mathbf{d} is an *ESC* if $m > 1$ or an *ESS* if $m = 1$ [4]. More generally, every ESS (respectively, ESC) is a (mixed) Nash equilibrium.

Sources, sinks, and balanced patches

Pulliam [28] introduced the notion of sources and sink patches for a population at equilibrium. In source patches, the birth rates exceed the death rates, while in sink patches the death rates exceed the birth rates. Here, we extend Pulliam’s definition to population exhibiting periodic fluctuations in abundance. A patch is a source if births exceed deaths ‘on average’ across years, while a patch is a sink if deaths exceed births ‘on average’ across years. For fitnesses varying in time, the appropriate ‘average’ is the geometric mean:

$$\bar{f}_j = \left(\prod_{t=1}^p f^j(t, \|\hat{\mathbf{x}}^j(t)\|) \right)^{1/p}.$$

If $\bar{f}_j < 1$, then patch j is a *sink*. If $\bar{f}_j > 1$, then patch j is a *source*. Following McPeck and Holt [26], we say patch j is *balanced* if $\bar{f}_j = 1$. Individuals remaining in a balanced patch, on average exactly replace themselves.

Main results

We have two main results. Our first result implies that if there is spatial heterogeneity in the within-patch fitnesses under equilibrium conditions, then no coalition of distinctive phenotypes can coexist, there is selection for slower dispersers, and the landscape supports source and sink patches. In particular, this result implies that for populations at equilibrium and playing a Nash equilibrium, all patches must be balanced. This result follows from [22] and generalizes the earlier work of Hastings [15] and Parvinen [27] by allowing for non-diffusive patterns of dispersal (i.e. S need not be symmetric).

PROPOSITION 1 *If $p = 1$ and $\mathbf{F}(1, \hat{\mathbf{x}}(1))$ is not scalar (i.e. there is spatial heterogeneity), then all the d_i are equal and positive, $\mathcal{I}(\mathbf{d}; \tilde{d}) > 1$ for all $0 \leq \tilde{d} < d_1$, and at least one patch is a sink and at least one patch is a source.*

Proof Let \mathbf{A} be the diagonal matrix $\mathbf{F}(1, \hat{\mathbf{x}}(1))$ with diagonal entries $A_{jj} = f^j(1, \|\hat{\mathbf{x}}^j(1)\|)$. Since \mathbf{A} is not scalar, Theorem 3.1 in [22] implies that $\varrho(\mathbf{S}(d)\mathbf{A})$ is a strictly decreasing function

of $d \in [0, 1]$. Assumption A2 implies that $\varrho(\mathbf{S}(d_i)\mathbf{A}) = 1$ for all i . Hence, all of the d_i are equal and $\mathcal{I}(\mathbf{d}; \tilde{d}) > 1$ for all $0 \leq \tilde{d} < d_1$. The d_i cannot equal zero as this would imply that $\mathbf{F}(1, \hat{\mathbf{x}}(1))$ is the identity matrix contradicting the assumption that it is non-scalar. Since $\mathbf{S}(d_1)$ is column stochastic and \mathbf{A} is non-scalar, $\max_i A_{ii} > \varrho(\mathbf{S}(d_1)\mathbf{A}) = 1 > \min_i A_{ii}$. Hence, there is a source patch and a sink patch. ■

Our second result concerns period 2 environments. In contrast to populations at equilibrium, this result implies that any ESS or ESC (or more generally, Nash equilibrium) includes a dispersive phenotype, supports at least one sink patch and supports no source patches. We are able to prove this result only under the restriction that the dispersal matrix S is *diagonally similar to a symmetric matrix*: there exists an invertible diagonal matrix D such that DSD^{-1} is a symmetric matrix. This allows for a diversity of movement patterns including diffusive movement (i.e. symmetric S) and any form of local movement along a one-dimensional gradient (i.e. tridiagonal S). A proof is given in Appendix 1. It is an open problem whether this result extends to all irreducible stochastic matrices S .

THEOREM 2 *If $p = 2$, S is diagonally similar to a symmetric matrix, $\mathbf{F}(2, \hat{\mathbf{x}}(2)) \neq \mathbf{F}(1, \hat{\mathbf{x}}(1))$ (i.e. there is temporal heterogeneity), and $\mathbf{F}(t, \hat{\mathbf{x}}(t))$ is non-scalar for some t (i.e. there is some spatial heterogeneity), then*

- (i) *If $\max_i d_i = 0$, then $\mathcal{I}(\mathbf{d}; \tilde{d}) > 1$ for all $0 < \tilde{d} \leq 1$.*
- (ii) *If \mathbf{d} is a (possibly mixed) Nash equilibrium, then $\max_i d_i > 0$, the set $\text{Sinks} \subset \{1, \dots, n\}$ of sink patches is non-empty, and the remaining patches $\{1, \dots, n\} \setminus \text{Sinks}$ are balanced.*

4. Applications

To illustrate the utility of our results, we present two applications. The first application considers populations with compensating density dependence in a periodically forced environment. We use Theorem 2 and its proof to determine under what conditions there is an ESC of sedentary and dispersive phenotypes. The second application considers populations with sufficiently overcompensatory density dependence to create oscillatory dynamics. We use the proof of Theorem 2 to illustrate how the evolution of higher dispersal rates can synchronize initially asynchronous population dynamics.

4.1. Evolution of dispersal dimorphisms in periodic environments

We consider competing dispersive phenotypes whose within patch dynamics are given by a periodically forced Beverton–Holt model. For simplicity, we assume that only the intrinsic fitness λ_t^j of an individual living in patch j varies in time and space:

$$f^j(t, \|\mathbf{x}^j\|) = \frac{\lambda_t^j}{1 + a\|\mathbf{x}^j\|}, \quad (4)$$

where a measures the intensity of competition within a patch. Also for simplicity, we assume that $S_{kj} = 1/n$ for all j, k . In other words, dispersing individuals are uniformly distributed across the landscape.

When there is only one dispersive phenotype (i.e. $m = 1$), Equation (1) is a sublinear monotone map (see, e.g. [17] for definitions). Consequently, [17, Theorem 6.1] implies that the fate of

the population depends on the linearization of the system at the extinction state. The dominant eigenvalue associated with this linearization is given by

$$\lambda = \varrho \left(\prod_{t=1}^p \mathbf{S}(d_1) \mathbf{F}(t, 0) \right)^{1/p}.$$

If $\lambda \leq 1$, then population goes deterministically towards extinction, i.e. $\mathbf{x}(t)$ converges to 0 for all $\mathbf{x}(0) \geq 0$. Alternatively, if $\lambda > 1$, then the populations increases when rare and ultimately converges to a periodic orbit. More precisely, there exists a periodic orbit, $\{\hat{\mathbf{x}}(1), \dots, \hat{\mathbf{x}}(p)\}$ with $\hat{\mathbf{x}}(t) \gg 0$ for all t , such that $\mathbf{x}(t)$ converges to this periodic orbit whenever $\mathbf{x}(0) \gg 0$. A sufficient condition ensuring $\lambda > 1$ is

$$\left(\prod_{t=1}^p \lambda_t^j \right)^{1/p} > 1 \quad \text{for all } j. \tag{5}$$

In other words, all the patches can support a population in the absence of immigration. While this condition is stronger than what is necessary, for simplicity, we assume that Equation (5) holds for the remainder of this section.

For period 2 environments where the intrinsic fitness vary in space and time, Theorem 2 implies that a sedentary strategy (i.e. $d_1 = 0$) is not a Nash equilibrium as it can be invaded by more dispersive phenotypes. In contrast, for a fully dispersive phenotype (i.e. $d_1 = 1$), there is a periodic point $\hat{\mathbf{x}}(t)$ given by

$$\hat{x}_1^j(t) = \frac{\bar{\lambda}_2 \bar{\lambda}_1 - 1}{a(1 + \bar{\lambda}_t)},$$

where $\bar{\lambda}_t = (1/n) \sum_{j=1}^n \lambda_t^j$ is the spatial average of the intrinsic fitnesses. Along this periodic orbit, a computation reveals that the within-patch fitnesses satisfy

$$\prod_{t=1}^2 f^j(t, \|\hat{\mathbf{x}}^j(t)\|) = \prod_{t=1}^2 \frac{\lambda_t^j}{\bar{\lambda}_t}.$$

Thus, Theorem 2 implies that a necessary condition for $d_1 = 1$ to be a Nash equilibrium is

$$\prod_{t=1}^2 \lambda_t^j \leq \prod_{t=1}^2 \bar{\lambda}_t \quad \text{for all } j \tag{6}$$

with a strict inequality for at least one j . This condition for a Nash equilibrium requires that the geometric mean of the fitness within each patch is no greater than the geometric mean of the spatially averaged fitness.

To illustrate the utility of Equation (6), consider an environment where the fitness in each patch fluctuates between a low value λ_{bad} in ‘bad’ years and a higher value λ_{good} in ‘good’ years i.e. $\lambda_t^j \in \{\lambda_{\text{good}}, \lambda_{\text{bad}}\}$, $\lambda_{\text{bad}} < \lambda_{\text{good}}$, and $\lambda_t^j \neq \lambda_{t+1}^j$. To ensure that all dispersal phenotypes can persist, we assume that the geometric mean $\sqrt{\lambda_{\text{good}} \lambda_{\text{bad}}}$ is greater than one. Provided that there is some spatial asynchrony (i.e. $\lambda_t^j \neq \lambda_t^k$ for some k, j), the necessary condition (6) for a Nash equilibrium of highly dispersive phenotypes simplifies to

$$\sqrt{\lambda_{\text{good}} \lambda_{\text{bad}}} \leq \frac{1}{2} (\lambda_{\text{good}} + \lambda_{\text{bad}}).$$

This inequality holds strictly as the geometric mean is less than the arithmetic mean. Furthermore, a computation reveals that the geometric mean of fitness within patch j satisfies

$$\sqrt{\prod_{t=1}^2 f^j(t, \|\hat{\mathbf{x}}^j(t)\|)} = \sqrt{\frac{\lambda_{\text{good}}\lambda_{\text{bad}}}{\frac{1}{4}(\lambda_{\text{good}} + \lambda_{\text{bad}})^2}} < 1$$

for all patches j . Hence, Theorem A1 in Appendix 1 implies that $\mathcal{I}(1; \tilde{d}) < 1$ for all $\tilde{d} \in [0, 1)$. Therefore, for this environment, a highly dispersive phenotype ($d_1 = 1$) always is an ESS and all patches are sinks for populations playing this ESS.

When the necessary condition (6) for a highly dispersive phenotype to be a Nash equilibrium is not met, sedentary dispersers can invade the patches whose average fitness $\sqrt{\prod_{t=1}^2 \lambda_t^j}$ exceeds the spatially averaged fitness $\sqrt{\prod_{t=1}^2 \bar{\lambda}_t}$. To understand the implications of this invasion, we consider a two patch environment where environmental fluctuations are spatially asynchronous. More precisely, we assume that spatial average of fitness does not vary in time (i.e. $\bar{\lambda}_t = \bar{\lambda}$) and the temporal fluctuations $\epsilon_t \in (-\bar{\lambda}, \bar{\lambda})$ around this spatial average are asynchronous, i.e. $\lambda_t^1 = \bar{\lambda} + \epsilon_t$ and $\lambda_t^2 = \bar{\lambda} - \epsilon_t$ for $t = 1, 2$. The necessary condition (6) for a highly dispersive phenotype to be a Nash equilibrium becomes

$$\bar{\lambda}|\epsilon_1 + \epsilon_2| \leq |\epsilon_1\epsilon_2| \quad \text{and} \quad \epsilon_1\epsilon_2 < 0. \quad (7)$$

When Equation (7) is not met, extensive numerical simulations suggest that after the successful invasion of the sedentary dispersers, the populations approach a period 2 orbit $\hat{\mathbf{x}}(t)$. Moreover, these simulations suggest that $\mathbf{d} = (1, 0)$ is an ESC and, consequently, a potential evolutionary end point consisting of a dimorphism of sedentary and highly dispersive phenotypes. At this ESC, one patch is balanced and occupied by both phenotypes, while the other patch is a sink and only occupied by the dispersive phenotype. Equation (7) implies that the ESC occurs when the temporal correlations of within patch fitness are not too negative (i.e. ϵ_1 is not too close to $-\epsilon_2$ in Figure 1). Pairwise invasibility plots (see, e.g. [12] for a discussion of the interpretation of these plots and the associated terminology) suggest that these ESCs can be reached by small mutational steps when $d = 0$ is a convergently stable branching point (Figure 2(a) and (b)). On the other hand, when Equation (7) is satisfied, the highly dispersive phenotype may or may not be an ESS in the strict sense (Figure 2(c) and (d)). Although numerical simulations confirm that the highly dispersive phenotype resists invasion attempts by nearby phenotypes (i.e. $\mathcal{I}(1, \tilde{d}) \leq 1$ whenever \tilde{d} is sufficiently close to 1), relatively sedentary phenotypes still may be able to invade (Figure 2(d)).

4.2. Evolution of synchronicity

Biotic interactions can generate oscillatory dynamics and, thereby, temporal variation in fitness. To illustrate the feedbacks between evolution of dispersal and biotically generated oscillations, we consider an extension of the coupled logistic map introduced by Hastings [16]. In this model, the local dynamics are determined by the logistic fitness function $rx(1-x)$. A fraction d of all individuals disperses randomly to all patches, i.e. a fraction d/k of individuals from patch j arrives in all patches. Under these assumptions, the dynamics of a single dispersive phenotype is given by

$$x_{t+1}^j = (1-d)r x_t^j(1-x_t^j) + \frac{d}{n} \sum_{k=1}^n r x_t^k(1-x_t^k).$$

We note that in the case of $n = 2$ patches, Hasting's d corresponds to our $d/2$.

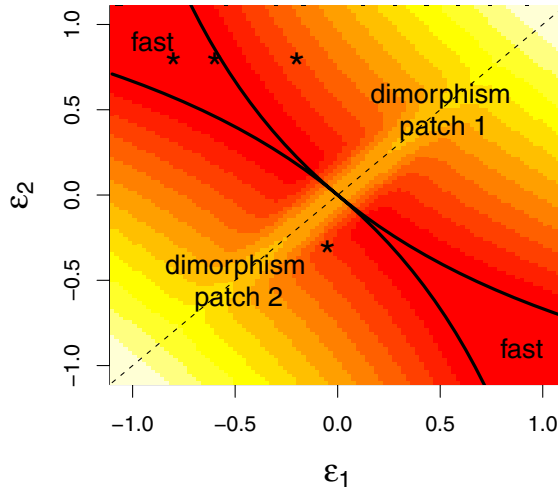


Figure 1. Nash equilibria for a Beverton–Holt two patch model with $\lambda_t^1 = 2 + \epsilon_t$ and $\lambda_t^2 = 2 - \epsilon_t$. The strategy $d_1 = 1$ is a local ESS in the regions denoted ‘fast’. The mixed strategy $\mathbf{d} = (1, 0)$ is an ESC in the regions denoted ‘dimorphism’. For these regions, the sedentary population only resides in the indicated patch. Shading corresponds to the average abundance of the sedentary strategy with lighter colours corresponding to higher abundance. At the dashed line, all (coalitions of) strategies are a (mixed) Nash equilibrium. *s refers to parameter values for which pairwise invasibility plots are shown in Figure 2.

When $r > 3$ and there are two patches, Hastings [16] has shown that there is an interval of dispersal rates between 0 and 1 such that there is an out-of-phase stable period two point (Figure 3(a), (c), and (d)). To apply our results, let $m = 1$, $d_1 = d$ for which there is a stable out-of-phase periodic point, $\hat{\mathbf{x}}(1), \hat{\mathbf{x}}(2)$ (an explicit formula for this orbit can be found in the Appendix of [16]), $S_{jk} = 1/2$ for $1 \leq j, k \leq 2$, and $f^j(t, \|\mathbf{x}^j\|) = r(1 - \|\mathbf{x}^j\|)$. Along this out-of-phase periodic orbit, $\prod_t \mathbf{F}(t, \hat{\mathbf{x}}(t))$ is a scalar matrix. Hence, Theorem A1 implies that along this out-of-phase periodic orbit $\mathcal{I}(d_1, \tilde{d}) > 1$ for all $d_1 < \tilde{d} \leq 1$. Hence, there would be selection for higher dispersal rates. For this two patch case, this implication of Theorem A1 also follows from a proposition of Doebeli and Ruxton [8, p. 1740] in which they performed a direct calculation of the eigenvalues. For $3 < r \leq 3.7$, numerical simulations suggest that this selection for higher dispersal rates ultimately results in a dispersal rate that spatially synchronizes the dynamics (Figure 3(a)–(c)) at which point all dispersal rates are Nash equilibria. However, at higher r values such as $r = 3.95$ (Figure 3(d)), the destabilization of the out-of-phase period 2 point (at $d_1 \approx 0.38$) results in more complex asynchronous dynamics in which case our results are not applicable and evolution of dispersal may no longer synchronize the dynamics.

When there are more than two patches and $r > 3$, out-of-phase two cycles can take on a greater diversity of forms. In particular, one can divide the landscape into two sets of patches such that patches are synchronous within each set and asynchronous across sets. Due to this potential spatial asymmetry in these out-of-phase cycles, we have not been able to show the geometric mean of fitness equals one in all patches. However, numerical simulations for $n = 40$ patches and $3 < r < 3.45$, suggest that this does occur. In which case, Theorem A1 implies that there would be selection for higher dispersal rates along these asynchronous cycles. In fact, numerical simulations for $3 < r < 3.45$ show that there would be selection for higher dispersal rates until the dynamics are spatially synchronized (Figure 4). Moreover, these simulations show, quite intuitively, greater initial asynchrony for the dynamics of the sedentary phenotype result in a stronger selection gradient (Figure 4(b)) and require the evolution of higher dispersal rates to regionally synchronize the dynamics (Figure 4(a)).

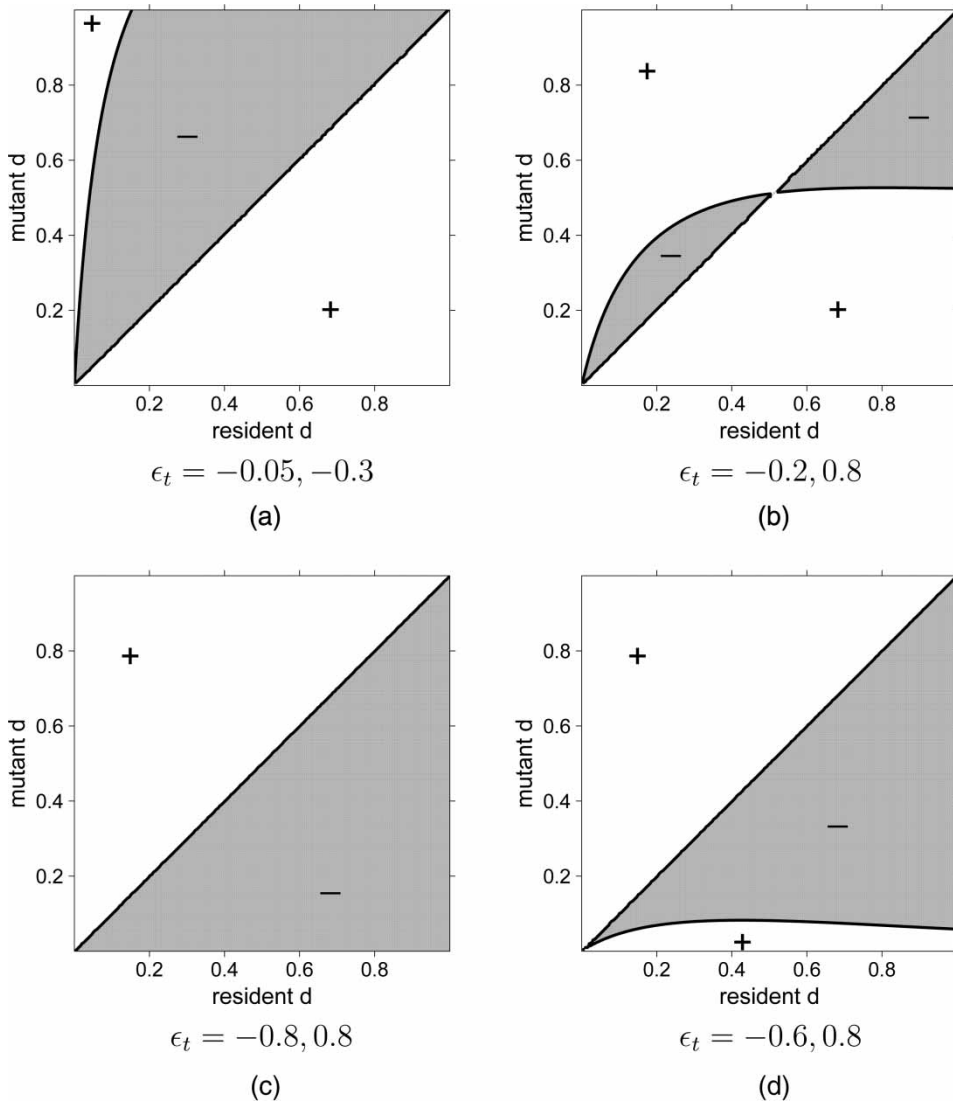


Figure 2. Pairwise invasibility plots (PIPs) for a two patch Beverton–Holt model with $\lambda_t^1 = 2 + \epsilon_t$ and $\lambda_t^2 = 2 - \epsilon_t$. The horizontal and vertical axes correspond to resident d_1 and mutant \tilde{d} dispersal rates, respectively. The dark lines, shaded regions, and unshaded regions correspond to where $\mathcal{I}(d_1, \tilde{d}) = 1$, $\mathcal{I}(d_1, \tilde{d}) < 1$, and $\mathcal{I}(d_1, \tilde{d}) > 1$, respectively. In (a) and (b), $d_1 = 0$ is an evolutionary branching point. In (b), $d_1 \approx 0.5$ is a convergently unstable singular point. In (c), $d_1 = 1$ is convergently stable and evolutionarily stable. In (d), $d_1 = 1$ is convergently stable and a local ESS, but invadable by sufficiently sedentary phenotypes.

5. Discussion

We analysed the evolution of dispersal in spatially and temporally variable environments. When there is spatial variation in fitness and within patch fitness varies in time between a lower and higher value, we proved that any ESS or ESC includes a dispersive phenotype. In particular, our results imply a sedentary population can always be invaded by more dispersive phenotypes. These results are particularly remarkable in light of earlier analysis showing generically the only ESS for spatially heterogeneous environments without temporal heterogeneity is a

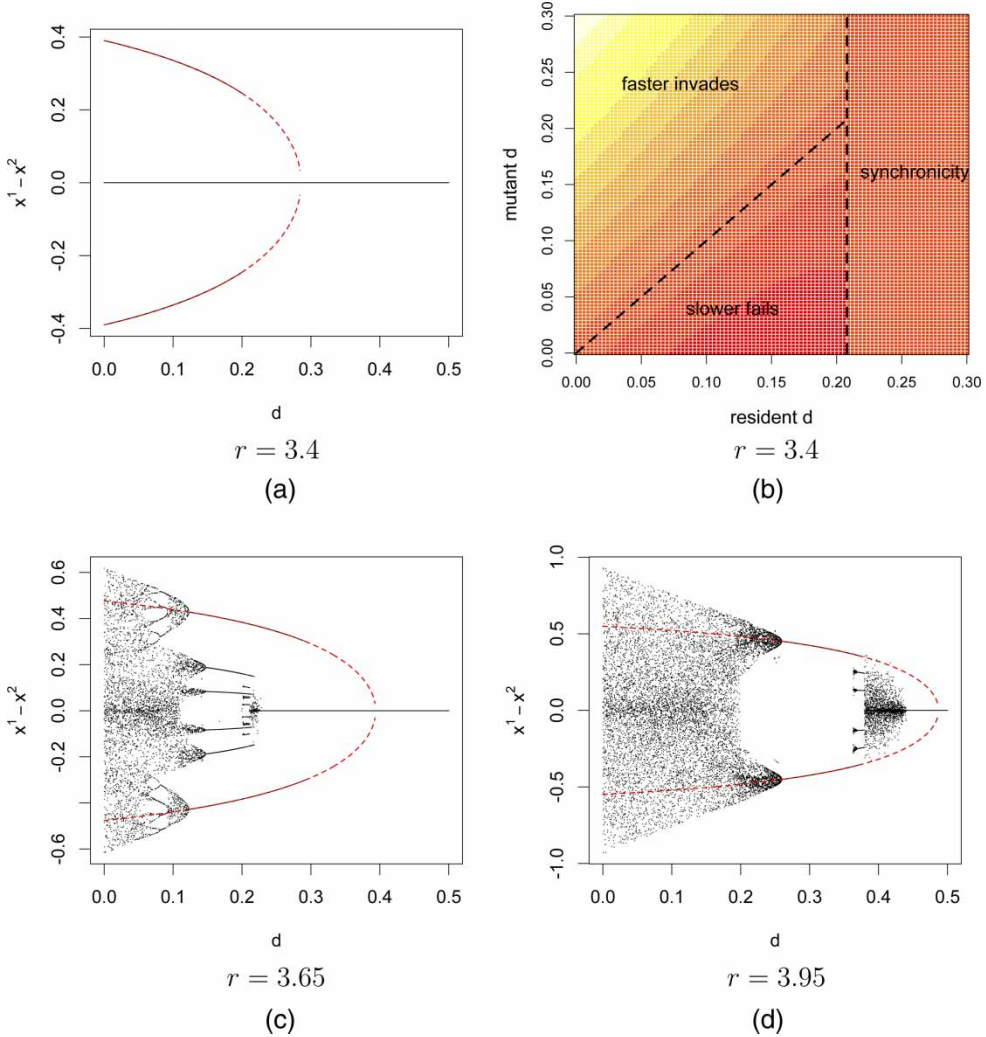


Figure 3. Evolution of spatial synchronization in a two-patch logistic model. In (a), (c), and (d), orbital bifurcation diagrams for phase difference $x^1 - x^2$ are shown for the two-patch logistic equation. All phase differences along attractors are plotted in black. The red-dashed curve corresponds to the out-of-phase, period two orbit which is stable only when it overlaps the black regions. In (b), the contours of $\mathcal{I}(d; \tilde{d})$ are plotted along the out-of-phase periodic point whenever it is stable.

sedentary phenotype [7,15,22,27]. Hence, our results imply this earlier work on the discrete-time models is not generic: arbitrarily small periodic perturbations result in selection for dispersal.

Our results extend a result of [27] who proved sedentary phenotypes could be invaded by phenotypes that disperse uniformly to all patches in every generation. Moreover, they are consistent with the numerical work of Mathias *et al.* [25] who found that ESCs always included highly dispersive phenotypes in discrete-time models with random variation in the vital rates. In contrast, our results are only partially consistent with the analysis of reaction-diffusion models by Hutson *et al.* [20]. Hutson *et al.* proved that certain forms of spatio-temporal heterogeneities select for the higher dispersal rate. However, if the frequency of spatial oscillations (i.e. spatial variation) is too large or too small, then phenotypes with higher dispersal rates are driven to extinction.

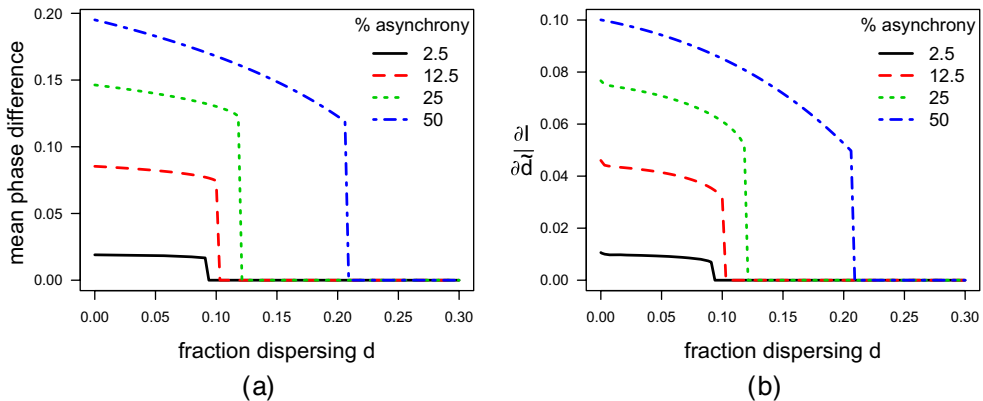


Figure 4. Evolution of spatial synchronization for a 40 patch logistic model. In (a), the mean phase difference in spatial abundance (i.e. $(1/40^2) \sum_{j \neq k} |x^j - x^k|$) for a single population exhibiting a two-cycle is plotted as function of its dispersal rate. In (b), the strength of selection $(\partial I / \partial \tilde{d})(d, d)$ for higher dispersal rates is plotted as a function of its dispersal rate. Different lines correspond to different percentages of initial asynchrony for sedentary phenotype, e.g. 50% implies that half the patches at $d = 0$ were out of sync with the other set of patches.

A partial explanation for this discrepancy is the continuity of the per-capita growth rates in the reaction–diffusion models results in positive correlations in time which may select for slower dispersal rates.

Our analysis and numerical simulations suggest that there are two evolutionary end states for environments where spatial-temporal heterogeneity is generated by abiotic periodic forcing: either an ESS consisting of a highly dispersive phenotype or an ESC consisting of a highly dispersive phenotype and a sedentary phenotype. This is partially consistent with the numerical work of Mathias *et al.* [25] who always found ESCs of high and low dispersal phenotypes. Mathias *et al.* found that these ESCs always could be achieved by small mutational steps leading to an intermediate phenotype (a convergent singular strategy) at which coalitions of faster and slower dispersers could invade and coexist (evolutionary branching). Our numerical and analytic results differ from these conclusions in two ways. First, evolutionary branching occurs at sedentary phenotypes not intermediate phenotypes. Consequently, prior to the branching event, one of the strategies in the ESC is already present. Second, while evolution by small mutational states may culminate in a highly dispersive phenotype, an ESC of highly dispersive and sedentary phenotypes may arise by the invasion of sufficiently slow dispersers, i.e. large mutational steps are required to realize the ESC.

When spatial and temporal heterogeneity is created purely by biotic interactions and initial conditions (e.g. the coupled logistic map), we have shown there can be selection for higher dispersal rates that ultimately may synchronize the population dynamics across space. Following this synchronization event, dispersal would become selectively neutral. Our analytic results extend Doebeli and Ruxton’s [8] two-patch analysis to multiple patches. In a multi-species context, Dercole *et al.* [5] found a related result. Numerical simulations of tritrophic communities with chaotic dynamics showed that evolution of dispersal often drove these spatial networks to weak forms of synchronization.

Our analysis states that an ESS or a ESC for dispersal result in some patches being sinks (i.e. geometric mean of fitness less than one) and the remaining patches being balanced (i.e. geometric mean of fitness equal to one). In the case of ESCs of highly dispersive and sedentary phenotypes, the sedentary phenotypes only reside in the balanced patches. In the case of ESS of highly dispersive phenotypes, all patches may be sinks despite the population persisting. In contrast, ESSs for conditional dispersal in purely spatially heterogeneous environments result in all occupied

patches being balanced [2,22] and, thereby, exhibiting an ideal and free distribution [10]. By appropriately modifying empirical methods for distinguishing between source-sink dynamics and balanced-dispersal [6,9], one might be able to find empirical support for these alternative, evolutionarily stable, spatial-temporal patterns of fitness.

Despite extensive progress, there are many mathematical challenges to overcome if we want to have a better analytical understanding of the evolution of dispersal. Two challenges, of particular interest here, are going beyond the assumption period 2 environments, and accounting for various forms of costs for dispersal. While period 2 environments can be viewed as a crude cartoon of seasonal environments, most environmental fluctuations exhibit multiple modes in their Fourier decomposition and have a significant stochastic component [31]. Despite some progress [29,32], a detailed analytic understanding of the interactive effects of this environmental stochasticity and dispersal on regional fitness remains largely elusive. Alternatively, the ability to disperse or the act of dispersing often is accompanied by costs to the individual. Dispersing individuals may die before reaching their destination. Alternatively, there may trade-offs between dispersal ability and competitive abilities [23,33]. If these costs or trade-offs are sufficiently strong, they can substantially alter the predictions presented here. For example, [27] has shown, quite intuitively, that if costs to dispersal are sufficiently strong, there is always selection against dispersal. However, we need the development of new mathematical methods to unravel the implications of intermediate costs on the evolution of dispersal in environments with spatio-temporal variation.

Acknowledgements

We thank two anonymous reviewers and Fabio Dercole for their extensive and helpful comments. The research of the first author was partially supported by the US National Science Foundation Grants DMS-0517987 and EF-0928987. The research of the second author was partially supported by the U.S. National Science Foundation and the William and Mary Plumeri Award. He is an honorary professor of the University of Hong Kong and an honorary professor of the Taiyuan University of Technology.

References

- [1] R.S. Cantrell, C. Cosner, and Y. Lou, *Movement toward better environments and the evolution of rapid diffusion*, Math. Biosci. 204(2) (2006), pp. 199–214.
- [2] R.S. Cantrell, C. Cosner, D.L. Deangelis, and V. Padron, *The ideal free distribution as an evolutionarily stable strategy*, J. Biol. Dyn. 1 (2007), pp. 249–271.
- [3] D. Cohen and S.A. Levin, *Dispersal in patchy environments: The effects of temporal and spatial structure*, Theor. Popul. Biol. 39 (1991), pp. 63–99.
- [4] R. Cressman, *Evolutionary Dynamics and Extensive Form Games*, MIT Press, Cambridge, MA, 2003.
- [5] F. Dercole, D. Loiacono, and S. Rinaldi, *Synchronization in ecological networks: A byproduct of Darwinian evolution?* Int. J. Bifur. Chaos 17(7) (2007), pp. 2435–2446.
- [6] J.E. Diffendorfer, *Testing models of source-sink dynamics and balanced dispersal*, Oikos 81(3) (1998), pp. 417–433.
- [7] J. Dockery, V. Hutson, K. Mischaikow, and M. Pernarowski, *The evolution of slow dispersal rates: A reaction diffusion model*, J. Math. Biol. 37 (1998), pp. 61–83.
- [8] M. Doebeli and G.D. Ruxton, *Evolution of dispersal rates in metapopulation models: Branching and cyclic dynamics in phenotype space*, Evolution 51(6) (1997), pp. 1730–1741.
- [9] C.P. Doncaster, J. Clobert, B. Doligez, E. Danchin, and L. Gustafsson, *Balanced dispersal between spatially varying local populations: An alternative to the source-sink model*, Am. Nat. 150(4) (1997), pp. 425–445.
- [10] S.D. Fretwell and H.L. Jr. Lucas, *On territorial behavior and other factors influencing habitat distribution in birds*, Acta Biotheor. 19 (1970), pp. 16–36.
- [11] N.A. Friedenberg, *Experimental evolution of dispersal in spatiotemporally variable microcosms*, Ecol. Lett. 6 (2003), pp. 953–959.
- [12] S.A.H. Geritz, J.A.J. Metz, E. Kisdi, and G. Meszina, *Dynamics of adaptation and evolutionary branching*, Phys. Rev. Lett. 78 (1997), pp. 2024–2027.
- [13] S.A.H. Geritz, M. Gyllenberg, F.J.A. Jacobs, and K. Parvinen, *Invasion dynamics and attractor inheritance*, J. Math. Biol. 44(6) (2002), pp. 548–560.
- [14] W.D. Hamilton and R. May, *Dispersal in stable habitats*, Nature 269 (1977), pp. 578–581.
- [15] A. Hastings, *Can spatial variation alone lead to selection for dispersal?* Theor. Popul. Biol. 24 (1983), pp. 244–251.

- [16] A. Hastings, *Complex interactions between dispersal and dynamics: Lessons from coupled Logistic equations*, Ecology 74 (1993), pp. 1362–1372.
- [17] M.W. Hirsch and H.L. Smith, *Monotone maps: A review*, J. Difference Equ. Appl. 11 (2005), pp. 379–398.
- [18] R.D. Holt, *Patch dynamics in two-patch environments: Some anomalous consequences of an optimal habitat distribution*, Theor. Popul. Biol. 28 (1985), pp. 181–208.
- [19] R.D. Holt, *Demographic constraints in evolution: Towards unifying the evolutionary theories of senescence and niche conservatism*, Evol. Ecol. 10 (1996), pp. 1573–8477.
- [20] V. Hutson, K. Mischaikow, and P. Poláčik, *The evolution of dispersal rates in a heterogeneous time-periodic environment*, J. Math. Biol. 43(6) (2001), pp. 953–959.
- [21] A. Katok and B. Hasselblatt, *Modern Theory of Dynamical Systems*, Cambridge University Press, Cambridge, 1995.
- [22] S. Kirkland, C.K. Li, and S.J. Schreiber, *On the evolution of dispersal in patchy landscapes*, SIAM J. Appl. Math. 66(4) (2006), pp. 1366–1382.
- [23] R. Levins and D. Culver, *Regional coexistence of species and competition between rare species*, Proc. Natl Acad. Sci. USA 68(6) (1971), pp. 1246–1248.
- [24] S.A. Levin, D. Cohen, and A. Hastings, *Dispersal strategies in patchy environments*, Theor. Popul. Biol. 26(2) (1984), pp. 165–191.
- [25] A. Mathias, E. Kisdi, and I. Iivieri, *Divergent evolution of dispersal in a heterogeneous landscape*, Evolution 55(2) (2001), pp. 246–259.
- [26] M.A. McPeck and R.D. Holt, *The evolution of dispersal in spatially and temporally varying environments*, Am. Nat. 6 (1992), pp. 1010–1027.
- [27] K. Parvinen, *Evolution of migration in a metapopulation*, Bull. Math. Biol. 61(3) (1999), pp. 531–550.
- [28] H.R. Pulliam, *Sources, sinks, and population regulation*, Am. Nat. 132 (1988), pp. 652–661.
- [29] S.J. Schreiber, *Interactive effects of temporal correlations, spatial heterogeneity, and dispersal on population persistence*, Proc. R. Soc. Biol. Sci., in press.
- [30] S.J. Schreiber and E. Saltzman, *Evolution of predator and prey movement into sink habitats*, Am. Nat. 174 (2009), pp. 68–81.
- [31] D.A. Vasseur and P. Yodzis, *The color of environmental noise*, Ecology 85 (2004), pp. 1146–1152.
- [32] P. Wiener and S. Tuljapurkar, *Migration in variable environments: Exploring life-history evolution using structured population models*, J. Theor. Biol. 166 (1994), pp. 75–90.
- [33] D.W. Yu and H.B. Wilson, *The competition-colonization trade-off is dead; long live the competition-colonization trade-off*, Am. Nat. 158 (2001), pp. 49–63.

Appendix 1. Proof of Theorem 2

The proof of Theorem 2 depends on the following key result which is proven in Appendix 2. We do not impose the assumption that S is irreducible in this result.

THEOREM A1 *Let $D = \text{diag}(d_1, \dots, d_n)$ with $d_1, \dots, d_n \in (0, \infty)$, and let S be an $n \times n$ column stochastic matrix such that RSR^{-1} is symmetric for some diagonal matrix R . For $t \in [0, 1]$, denote by $\varrho(F(t))$ the Perron (largest) eigenvalue of*

$$F(t) = D[(1-t)I_n + tS]D^{-1}[(1-t)I_n + tS].$$

Then $\varrho(F(t))$ is either an increasing function on $[0, 1]$ or a constant function on $[0, 1]$. The latter case happens if and only if D and S commute, equivalently, there is a permutation matrix P such that $PS P^T = S_1 \oplus S_2 \oplus \dots \oplus S_k$ and $PDP^T = d_1 I_{n_1} \oplus d_2 I_{n_2} \oplus \dots \oplus d_k I_{n_k}$, where $S_j \in M_{n_j}$ for $j = 1, \dots, k$, and $n_1 + \dots + n_k = n$.

Remark The above result covers the case of symmetric S . It also covers the case when S is an irreducible tridiagonal column stochastic matrix; one can use a simple continuity argument to extend the result to reducible tridiagonal column stochastic matrices.

To prove the first assertion of Theorem 2, assume that $\max_i d_i = 0$. Assumption A2 implies that $\mathbf{F}(2, \hat{\mathbf{x}}(2))\mathbf{F}(1, \hat{\mathbf{x}}(1)) = I$. Theorem A1 implies $\mathcal{I}(\mathbf{d}; \tilde{d}) > 1$ for all $\tilde{d} \in (0, 1]$.

To prove the second assertion of Theorem 2, assume that \mathbf{d} is a (possibly mixed) Nash equilibrium. The first assertion of the theorem implies that $\max_i d_i > 0$. Next, suppose to the contrary that there exists j such that $f^j(1, \|\hat{\mathbf{x}}^j(1)\|)f^j(2, \|\hat{\mathbf{x}}^j(2)\|) > 1$, i.e. there is a source patch. Then

$$\mathcal{I}(\mathbf{d}; 0) = \max_j \sqrt{f^j(1, \|\hat{\mathbf{x}}^j(1)\|)f^j(2, \|\hat{\mathbf{x}}^j(2)\|)} > 1$$

and by continuity $\mathcal{I}(\mathbf{d}; \tilde{d}) > 1$ for all $\tilde{d} \geq 0$ sufficiently small. As this contradicts the assumption that \mathbf{d} is a Nash coalition, it follows that $f^j(1, \|\hat{\mathbf{x}}^j(1)\|)f^j(2, \|\hat{\mathbf{x}}^j(2)\|) \leq 1$ for all j . Finally, suppose to the contrary that $f^j(1, \|\hat{\mathbf{x}}^j(1)\|)f^j(2, \|\hat{\mathbf{x}}^j(2)\|) = 1$ for all j . Theorem A1 implies that $\mathcal{I}(\mathbf{d}; \max_i d_i) > 1$ which contradicts the fact that $\mathcal{I}(\mathbf{d}; \max_i d_i) = 1$.

Appendix 2. Proof of Theorem A1

Denote by $\|A\|$ the operator norm of the matrix A . The proof of Theorem A1 depends on the following.

THEOREM A2 Suppose $A \in M_n$ is nonzero and satisfies $\|I + A\| \geq 1$. Then $\|I + tA\| \geq \|I + A\|$ for all $t \geq 1$ and one of the following condition holds.

- (a) The function $t \mapsto \|I + tA\|$ is increasing for $t \geq 1$.
- (b) There is a unitary matrix U such that $U^*AU = 0_k \oplus \tilde{A}$, where $\tilde{A} \in M_{n-k}$ is invertible and satisfies $\|I_{n-k} + \tilde{A}\| < 1$. Consequently, there is $t^* > 1$ such that $\|I_{n-k} + t^*\tilde{A}\| = 1$ and the function $t \mapsto \|I + tA\|$ has constant value 1 for $t \in [1, t^*]$ and is increasing for $t > t^*$.

Proof Let u be a unit vector such that $\|I + A\| = \|(I + A)u\|$ and $Au = \alpha u + \beta v$, where $\{u, v\}$ is an orthonormal family. By our assumption,

$$\|(I + A)u\| = \|(1 + \alpha)u + \beta v\| = |1 + \alpha|^2 + |\beta|^2 \geq 1,$$

i.e.

$$2\text{Re}(\alpha) + |\alpha|^2 + |\beta|^2 \geq 0.$$

Thus, for $t > 1$,

$$\begin{aligned} \|I + tA\| &\geq \|(I + tA)u\| \\ &= |1 + t\alpha|^2 + |t\beta|^2 \\ &= |1 + \alpha|^2 + |\beta|^2 + 2(t - 1)\text{Re}(\alpha) + (t^2 - 1)(|\alpha|^2 + |\beta|^2) \\ &= \|I + A\| + (t - 1)[2\text{Re}(\alpha) + |\alpha|^2 + |\beta|^2] + t(t - 1)(|\alpha|^2 + |\beta|^2) \\ &\geq \|I + A\|. \end{aligned}$$

(a) Suppose there is unit vector u in the above calculation satisfying $Au \neq 0$, i.e. $(\alpha, \beta) \neq (0, 0)$. Then the last inequality in the calculation is a strict inequality. Thus, $\|I + tA\| > \|I + A\|$.

Now, for any $t_0 > 1$, we have $\|I + t_0A\| > \|I + A\| \geq 1$. If u_0 is a unit vector satisfying $\|(I + t_0A)u_0\| = \|I + t_0A\|$, then $t_0Au_0 \neq 0$; otherwise, $\|I + t_0A\| = 1$. Consequently, if we replace A by t_0A in the above proof, we have

$$\|I + t(t_0A)\| > \|I + t_0A\|$$

for any $t \geq 1$. Thus the function $t \mapsto \|I + tA\|$ is increasing for $t > 1$.

(b) Suppose $Au = 0$ for every unit vector u satisfying $\|(I + A)u\| = \|I + A\|$. In particular, we have

$$\|I + A\| = \|(I + A)u\| = \|u\| = 1.$$

Let U be unitary such that U^*AU is lower triangular form with the first k diagonal entries equal to zero, and the last $n - k$ diagonal entries non-zero. Since

$$\|I + A\| = \|I + U^*AU\| = 1,$$

we see that

$$e_j^*(I + U^*AU)^*(I + U^*AU)e_j = \|(I + U^*AU)e_j\|^2 \leq 1$$

for $j = 1, \dots, k$. (Here $\{e_1, \dots, e_n\}$ is the standard basis for \mathbb{C}^n .) As a result, we see that $U^*AU = 0_k \oplus \tilde{A}$, where $\tilde{A} \in M_{n-k}$ is invertible.

Since $Au = 0$ for every unit vector u satisfying $\|(I + A)u\| = \|I + A\| = 1$, we conclude that $\|(I_{n-k} + \tilde{A})v\| < 1$ for any unit vector $v \in \mathbb{C}^{n-k}$. Thus, $\|I_{n-k} + \tilde{A}\| < 1$.

Note that $A \neq 0$ implies \tilde{A} is non-trivial, i.e. $k \neq n$. For sufficiently large t , we have $\|I_{n-k} + t\tilde{A}\| \geq 1$. Let t^* be the smallest real number in $(1, \infty)$ such that $\|I_{n-k} + t^*\tilde{A}\| = 1$. Since \tilde{A} is invertible, case (a) must hold and the function $t \mapsto \|I_{n-k} + t\tilde{A}\|$ is increasing for $t \geq t^*$. Hence for $t \geq t^*$, we have $\|I + tA\| = \|I_{n-k} + \tilde{A}\|$ and the function $t \mapsto \|I + tA\|$ is increasing. ■

Proof of Theorem A1 Note that $F(t) = D[(1-t)I_n + tS]D^{-1}[(1-t)I_n + tS]$ and

$$\begin{aligned}\tilde{F}(t) &= RF(t)R^{-1} \\ &= RDR^{-1}[(1-t)I_n + tRSR^{-1}]RD^{-1}R^{-1}[(1-t)I_n + tRSR^{-1}] \\ &= D[(1-t)I_n + tRSR^{-1}]D^{-1}[(1-t)I_n + tRSR^{-1}]\end{aligned}$$

have the same eigenvalues and hence the same spectral radius. Here, we use the fact that $RDR^{-1} = D$ as R is a diagonal matrix. So, $\tilde{S} = RSR^{-1}$ is symmetric with largest eigenvalue equal to 1, and all other eigenvalues lying in $[-1, 1]$. Moreover, if

$$B(t) = D^{1/2}[(1-t)I_n + t\tilde{S}]D^{-1/2},$$

then $D^{-1/2}F(t)D^{1/2} = B(t)B(t)^T$ so that $\varrho(F(t)) = \varrho(D^{-1/2}F(t)D^{1/2}) = \|B(t)\|^2$.

Suppose $0 \leq t_0 < t_0 + t \leq 1$. Let $A_0 = t_0D^{1/2}(\tilde{S} - I_n)D^{-1/2}$. Then

$$\|I + A_0\| = \|B(t_0)\| \geq \varrho(B(t_0)) = 1.$$

By Theorem A2,

$$\|B(t_0 + t)\| = \|I + (1 + t/t_0)A_0\| \geq \|I + A_0\| = \|B(t_0)\|. \quad (1)$$

Thus, $\varrho(F(t)) = \|B(t)\|^2$ is non-decreasing on $[0, 1]$. Moreover, by Theorem A2 the inequality in Equation (1) is an equality if and only if A_0 is unitarily similar to $0_k \oplus \tilde{A}_0$ where \tilde{A}_0 is invertible. Thus,

- (1) the null space of A_0 has dimension k , and
- (2) $A_0v = 0$ if and only if $v^T A_0 = 0$.

From Equation (1), we see that the eigenspace of the Perron root of the matrix $D^{1/2}\tilde{S}D^{-1/2}$ has dimension k . So, the matrix is permutationally similar to a $(k+1) \times (k+1)$ upper triangular block matrix so that each of the first k diagonal block is square irreducible and has eigenvalue 1, and the last block has eigenvalue less than 1. Since \tilde{S} is symmetric and D is diagonal, there is a permutation matrix P such that $P\tilde{S}P^T = \tilde{S}_1 \oplus \cdots \oplus \tilde{S}_{k+1}$ and $PDP^{-1} = D_1 \oplus \cdots \oplus D_{k+1}$ with $\tilde{S}_j, D_j \in M_{n_j}$ and $n_1 + \cdots + n_{k+1} = n$. But then $D_{k+1}^{1/2}\tilde{S}_{k+1}D_{k+1}^{-1/2}$ cannot have spectral radius less than 1. So, S_{k+1} must be vacuous.

Next, we show that each D_j is a scalar matrix. To this end, note that we can construct a null vector of PA_0P^T by extending a Perron vector $u_1 \in \mathbb{R}^{n_1}$ of \tilde{S}_1 to a vector v_1 in \mathbb{R}^n by adding $n - n_1$ zeros. Then $PA_0P^T v_1 = 0$. By (2), we see that $v_1^T PA_0P^T = 0$. It follows that $D_1^{1/2}\tilde{S}_1D_1^{-1/2}u_1 = u_1$ and $u_1^T D_1^{1/2}\tilde{S}_1D_1^{-1/2} = u_1^T$. Equivalently, $\tilde{S}_1D_1^{-1/2}u_1 = D_1^{-1/2}u_1$ and $u_1^T D_1^{1/2}\tilde{S}_1 = u_1^T D_1^{1/2}$. Since \tilde{S}_1 is symmetric, nonnegative, and irreducible, the eigenspace of the Perron eigenvalue is one dimensional and there is a positive vector w such that $\tilde{S}_1 w = w$ and $w^T \tilde{S}_1 = w^T$. Hence $D_1^{-1/2}u_1$ is a multiple of w and $u_1^T D_1^{1/2}$ is a multiple of w^T . Hence D_1 is a scalar matrix. Similarly, one can prove that D_2, \dots, D_k are scalar matrices. Since $P\tilde{S}P^T = \tilde{S}_1 \oplus \cdots \oplus \tilde{S}_k$, we have $PSP^T = S_1 \oplus \cdots \oplus S_k$.

Conversely, if S and D commute, then $F(t) = (tI + (1-t)S)(tI + (1-t)S)$ is column stochastic for all $t \in [0, 1]$. So, $\varrho(F(t)) = 1$ for all $t \in [0, 1]$. ■

Remark Note that the conclusion of Theorem A1 fails if S is not diagonally similar to a symmetric matrix. For example, if $S = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \oplus I_{n-3}$, and $D = \text{diag}(3, 2, 1) \oplus I_{n-3}$, then $\varrho(F(0)) = \varrho(F(1)) = 1$, but $\varrho(F(1/2)) > 1$. One may perturb S slightly, say, replacing it by $(1-\varepsilon)S + \varepsilon J$ for a small $\varepsilon > 0$ so that $\varrho(F(t))$ is not monotone, where J is the matrix with all entries equal to $1/n$.