Revised manuscript latex file Click here to view linked References

Temporal variability can promote migration between habitats

Harman Jaggi*[1], David Steinsaltz[2], and Shripad Tuljapurkar[1]

[1]Department of Biology, Stanford University, Stanford, CA 94305-5020, USA

[2]Department of Statistics, University of Oxford, Oxford OX1 2HB, United Kingdom

Keywords

migration, evolution of migration, variable environments, reduction principle, dispersal, stochastic growth rates, temporal variation.

*corresponding author

Abstract

Understanding the conditions that promote the evolution of migration is important in ecology and evolution. When environments are fixed and there is one most favorable site, migration to other sites lowers overall growth rate and is not favored. Here we ask, can environmental variability favor migration when there is one best site on average? Previous work suggests that the answer is yes, but a general and precise answer remained elusive. Here we establish new, rigorous inequalities to show (and use simulations to illustrate) how stochastic growth rate can increase with migration when fitness (dis)advantages fluctuate over time across sites. The effect of migration between sites on the overall stochastic growth rate depends on the difference in expected growth rates and the variance of the fluctuating difference in growth rates. When fluctuations (variance) are large, a population can benefit from bursts of higher growth in sites that are worse on average. Such bursts become more probable as the between-site variance increases. Our results apply to many (≥ 2) sites, and reveal an interplay between the length of paths between sites, the average differences in site-specific growth rates, and the size of fluctuations. Our findings have implications for evolutionary biology as they provide conditions for departure from the reduction principle, and for ecological dynamics: even when there are superior sites in a sea of poor habitats, variability and habitat quality across space determine the importance of migration.

1 Introduction

Migration is important to the life history of many species, with consequences for range shifts, genetic diversity, and population dynamics [49]. Traits governing migration are subject to selection and can significantly influence eco-evolutionary dynamics of spatially structured populations [16, 38, 9, 50]. In these papers, the term migration may mean 'regular movements by animals between locations' while dispersal may refer only to 'movement of an individual from site of birth to site of reproduction' and thus may be different behaviours [41, 11]. But we are concerned with a conceptually simpler problem: understanding the conditions that promote evolution of migration (here considered the same as dispersal) in the simplest sense of 'moving from one spatial unit to another and back' [6].

To explain why this is an important and interesting issue, consider what happens in the absence of variability, i.e., when population growth rate at each site is fixed. Qualitatively, for two sites, any dispersal from the better site (call it site 1) to the poorer site (call it site 2) would decrease the overall growth rate of the total population. Hence we would not expect dispersal away from the most favorable site to be favored by natural selection. The surprisingly complicated mathematical proof of this for multiple sites was given by

[32]. Karlin proved that the overall population growth rate will be determined by the rate prevailing at the best site. Thus in constant environments, dispersal away from the best site can reduce population growth. That has since become established as one example of a general reduction principle which we may describe as ‘mixing reduces growth’ and ‘differential mixing selects for reduced growth’ in constant environments [7, 29, 23, 1, 3, 2, 4].

Past work suggests that environmental variability may select for dispersal, as it allows populations to cope with variation in resources, escape unfavorable environments and benefit from new environments [53, 8, 5, 52], thus providing conditions for a departure from the reduction principle [42, 2]. Environmental or resource variability has been incorporated by accounting for temporal variation in carrying capacities [39, 15] or growth rates [63, 34], using approximations and/or numerical simulations to examine the evolution of dispersal. Many other studies have also shown that dispersal is relevant in spatially and temporally varying environments [62, 24, 31, 15, 43, 56, 42, 51, 30, 46, 40].

In this paper we answer a simple question: are there conditions that would favor the evolution of dispersal among sites with distinct stochastic growth rates? The use of stochastic growth rate as the appropriate invasion fitness into a monomorphic population has been established in both a genetic context [59] and a general ecological context [44]. We are interested in migration evolution and invasion dynamics, and therefore do not consider the extensive work on how population density affects dispersal [57].

The initial condition for an evolutionary analysis is the case where there is no dispersal. In the neighborhood of that initial condition, a central challenge for this analysis is that the stochastic growth rate changes very rapidly and is non-differentiable (a singularity) in the case of no dispersal. What does this mean? Let m be the probability of migration between two sites. The system of matrices will not satisfy at $m = 0$ the conditions under which the population growth rate is known [64] to be analytic. Our results show that under some conditions (as we describe further below) there is indeed a singularity at $m = 0$, with the derivative going to ∞. Thus, a typical approach such as a perturbation analysis cannot predict the exact nature of migration rate near $m = 0$. This can be seen in simple numerical simulations of migration between two sites. Such a singularity has been pointed out in special models for plants [48], life cycle delays [61], dispersal [63] and structured plant models [10]. Note that kin selection and density-dependence are not able to alter this singularity. We address the challenge by establishing rigorous inequalities for stochas-

tic growth rate.

To describe the problem and our model, we first consider the simple case of two sites with density-independent exponential population growth rates. We assume that population growth rates change over time and the logarithm of growth rates is normally distributed. We characterize the stochasticity of environmental conditions by the variance (σ^2) of difference in site-specific population log growth rates (allowing for covariation across sites). Migration is parameterized by a probability m of changing site and $(1-m)$ of staying. We assume there is a superior site (site 1), meaning that its expected logarithmic growth rate μ_1 is greater than the expected logarithmic growth rate μ_2 at site 2. Thus, populations have to overcome an average growth rate difference between sites, given by $\tilde{\mu} = \mu_1 - \mu_2 > 0$. Our metric for fitness is the (long-run) stochastic population growth rate, a [58]. We examine the asymptotic limit, (*i.e.*, small migration limit) of the long-run population growth rate $a(m)$.

We find the change $a(m) - a(0)$ in the stochastic population growth rate with migration relative to no migration is bounded above and below by functions that behave as a power m^γ of the migration rate. When $\gamma < 1$ there is, as mentioned earlier, a singularity at 0: the derivative of m^γ is proportional to $\left(1/m^{1-\gamma}\right)$ (which blows up as m approaches 0). Thus, when $\gamma < 1$ the function m^γ is non-differentiable at 0, and will dominate any linear function for m sufficiently close to 0. Under general assumptions we prove that the power γ of the migration rate is bounded above by $4\tilde{\mu}/\sigma^2$. Therefore, small migration is favored relative to zero migration when the variance of difference of growth rates between sites is large enough to overcome a fixed difference in expected growth rates. The intuition is that when there are independent fluctuations in growth rates across sites, populations may use dispersal to exploit this environmental variability (σ^2) to their benefit. As the difference $\tilde{\mu}$ in average quality increases, even larger fluctuations are required to make migration favorable.

Note that these results refer only to the variance of the difference in log growth rates between the sites. Whether these rates are positively correlated, negatively correlated, or uncorrelated depends in addition on the variance at each site individually. These individual variances will affect the baseline growth rate $a(0)$, but have no bearing on the results we discuss here concerning the difference $a(m) - a(0)$. In addition, we show that migration can be advantageous in spite of costs proportional to the total migration rate — if, for instance, we suppose that some proportion of those migrating are lost or fail to reproduce. We consider as well the setting where there are multiple sites, where it may not be possible to migrate in one step between

the best and second-best sites. We consider the impact of path lengths on the long-term growth rate during migration and find that short paths to return to the best site are favored. We also find that path lengths matter only when sites are similar (on average).

The paper provides an approach to studying migration patterns (such as those of birds or mammals) by examining variation in environmental conditions, and describing the relationship with site-specific metrics (such as population growth rates), to help bridge the gap between theory and empirical studies [40]. The next section of the paper describes the model in constant and variable environments. We follow with the theorem and simulations for the two-site case. Next, we describe the general theorem for any number of sites and simulations for 4 sites. We conclude with further discussion of the relevance of our results.

2 Model in constant environments

We describe the model by first considering constant environments, and then introduce variable environments. The growth rates of the population at each site are described by a diagonal matrix $\mathbf{G}$, where each diagonal element represents site-specific growth rate. There is one "best" site at which the growth rate is higher than at every other site. We assume no individuals die or reproduce during migration, and dispersal matrix $\mathbf{M}$ is stochastic (the rows are probabilities of moving from or staying at a site and must sum to 1). The individuals can get from the one site to any other site in a finite number of steps, so the migration matrix $\mathbf{M}$ is irreducible. (Throughout bold symbols refer to matrices and vectors.)

Let the components of vector $\mathbf{N}(t) = (N_1(t), N_2(t))^{\mathrm{T}}$ denote the population size at each site and T denotes the transpose. Then the population dynamics can be described by

$$\mathbf{N}(t+1) = [(1-m)\mathbf{I} + m\mathbf{M}]\ \mathbf{G}\ \mathbf{N}(t), \tag{1}$$

where the migration rate m scales the pattern of dispersal described by $\mathbf{M}$.

When there is migration, that is $m \neq 0$, the overall growth rate is given by the largest eigenvalue (also called spectral radius **s**) of the product of $\mathbf{M}(m) = [(1-m)\mathbf{I} + m\mathbf{M}]$ and diagonal growth matrix $\mathbf{G}$. [32] used the Donsker–Varadhan formula for principal eigenvalues to prove that the spectral radius is strictly decreasing in migration rate m, so in this case migration will not favored. We would like to note that Karlin's Theorem 5.2 was independently derived by [35], using a very different proof (though still deeply con-

nected to Donsker–Varadhan theory).

Interpreted biologically, the Karlin's theorem states: in such a system, the greater the level of mixing of population between sites, the lower the long-term growth rate of the total population. Such selection for reduced dispersal under constant selection is called the *reduction principle*, studied primarily in a genetic context by [21, 22, 23]. The reduction principle has since been extended to the evolution of mutation rates, sexual reproduction, and cultural transmission [3, 2].

Incorporating time-varying growth rates

Thus dispersal is not favored in a model with fixed growth rates. Now assume population growth depends on site-specific environmental factors that change over time. To account for this variation we treat growth rates as a random process and equation (1) is modified using a time-dependent matrix specifying growth-rates ($\mathbf{G_t}$) as:

$$\mathbf{N}(t+1) = \mathbf{M}\,\mathbf{G_t}\,\mathbf{M}\,\mathbf{G_{t-1}} \cdots \mathbf{M}\,\mathbf{G_1}\,\mathbf{N}(0). \tag{2}$$

Analysis of such a system is challenging for many reasons [17, 48, 2]. For structured populations, the stochastic growth rate is notoriously difficult to compute. Thus, we look beyond the domain of classical probability to the theory of large deviations. That theory examines extreme events that emerge from multiple rare occurrences. In our setting, the rare migration events can accumulate over time to dominate the aggregate dynamics, as described in the model and analysis below.

3 Model for heterogeneous environments

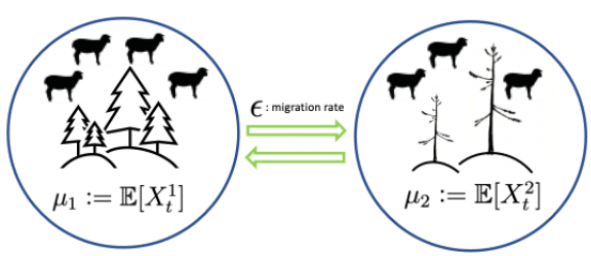


Figure 1: Site 1 (best site) has higher expected growth rate (μ_1) than at site 2 (μ_2).

To analyze dispersal evolution, we assume density independence, and for simplicity assume no age or stage structure (i.e., a scalar population) at each site. We start with two sites (1 and 2), where population grows at a rate $G_t^1 = \mathrm{e}^{X_t^{(1)}}$ and $G_t^2 = \mathrm{e}^{X_t^{(2)}}$, respectively. The log growth rate vector $\mathbf{X}_t = (X_t^{(1)}, X_t^{(2)})$ at t

is assumed to be serially independent; however, the components within each $\mathbf{X}_t$, corresponding to growth rates at different sites, may be dependent.

The expected values of log growth rates are given by $\mu_1 := \mathbb{E}[X_t^{(1)}]$ and $\mu_2 := \mathbb{E}[X_t^{(2)}]$ for site 1 and 2, respectively. We assume site 1 is the best site on average, meaning μ_1 is strictly larger μ_2. So the difference $\tilde{\mu} := \mu_1 - \mu_2$ between average growth rates is always positive ($\tilde{\mu} > 0$).

We assume the difference in growth rates between site 1 and site 2 at each time t is distributed normally, and $\widetilde{X}_t := X_t^{(2)} - X_t^{(1)}$ has variance σ^2. Note that $\sigma > 0$ so long as the two sites are not perfectly positively correlated. This allows us to vary site-specific variances and co-variances between sites. The arguments and theorems hold whether growth rates $\widetilde{X}_t$ are assumed to be Gaussian or sub-Gaussian. Sub-Gaussian refers to a class of random variables whose tails decay at least as fast as a Gaussian (normal) distribution, meaning that extreme values are no more likely than for a Gaussian variable. This means that our results are not tied to the details of the Gaussian distribution, but only to the general shape of the extreme-event likelihood. At the same time, while the results do still hold when tails are strictly **thinner** than Gaussian tails — for instance, when the growth-rate fluctuations are bounded — the result is much less interesting in this case, as the both bounds are then strictly negative for small m.

The probabilities of migration to another site and of staying at a site are given by m and $(1-m)$, respectively. The populations at the two sites are governed by the system of equations

$$\begin{aligned} N_{t+1}^1 &= (1-m)\, \mathrm{e}^{X_t^{(1)}} N_t^1 + \; m\, \mathrm{e}^{X_t^{(2)}} N_t^2 \,, \\ N_{t+1}^2 &= m\, \mathrm{e}^{X_t^{(1)}} N_t^1 + \; (1-m)\, \mathrm{e}^{X_t^{(2)}} N_t^2 \,. \end{aligned}$$

In matrix form the equations are written as

$$\begin{pmatrix} N_{t+1}^1 \\ N_{t+1}^2 \end{pmatrix} = \left(\mathbf{I} + m \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \right) \begin{pmatrix} \mathrm{e}^{X_t^{(1)}} & 0 \\ 0 & \mathrm{e}^{X_t^{(2)}} \end{pmatrix} \begin{pmatrix} \mathbf{N}_t^1 \\ \mathbf{N}_t^2 \end{pmatrix}$$

$$\mathbf{N}_{t+1} = (\mathbf{I} + m\mathbf{A})\, \mathrm{e}^{\mathrm{diag}(\mathbf{X}_t)} \mathbf{N}_t. \tag{3}$$

In the last equation we define $\mathbf{D}_t(m) := (\mathbf{I} + m\mathbf{A})\,\mathrm{e}^{\left(\mathrm{diag}(\mathbf{X}_t)\right)}$, where the matrix $\mathbf{X}_t$ is diagonal with elements being the population growth rates $X_t^{(1)}, X_t^{(2)}$ at the two sites. We assume that the sequence $\mathbf{X}_t$, $\mathbf{X}_{t-1}$, ..., $\mathbf{X}_1$ is i.i.d. in time, although there can be any correlation between sites at the same time. Also, the pattern (but not the scale m) of dispersal is given by the matrix $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$. In the main manuscript, we assume the dispersal matrix $\mathbf{A}$ to be constant but the appendix generalizes the results to $\mathbf{A}_t$ matrices that are i.i.d with time.

Long-run stochastic growth rate

Owing to the multiplicative nature of population growth (survival and reproduction), in a stochastic environment, the long-term growth converges to a limit as:

$$a_{ij}(m) := \lim_{t\to\infty} \frac{1}{t} \log(\mathbf{D}_t(m) \cdot \mathbf{D}_{t-1}(m) \cdot \cdots \cdot \mathbf{D}_1(m))_{ij}. \tag{4}$$

For $m > 0$ the i.i.d. sequence $\mathbf{D}_t(m) = (\mathbf{I} + m\mathbf{A})\,\mathbf{G_t} = (\mathbf{I} + m\mathbf{A})\,\mathrm{e}^{\mathrm{diag}(\mathbf{X}_t)}$ satisfies the conditions for the existence of a stochastic growth rate independent of starting condition [17] and so $a_{ij}(m)$ is the same for all i and j; we call this value $a(m)$. When $m = 0$, the limit $a(0)_{ij}$ in (4) exists, but is equal to μ_i at place i on the diagonal and $-\infty$ off the diagonal. Thus, when there is no migration site 1 dominates the total population.

4 Main result for two sites in variable environments

4.1 Asymptotic result for two sites

For migration to be favored, the stochastic growth rate $a(m)$ must increase as a function of migration rate relative to no migration, $a(0)$. Our theorem below describes lower and upper bounds for $a(m) - a(0)$ in the small-m limit in terms of the expected growth rates between two sites, $\tilde{\mu} = \mu_1 - \mu_2$ and the variance σ^2 describing the fluctuations.

The probabilistic reasoning for how the theorem works is as follows: since the environment is fluctuating, there is always a fraction of times when the site with lower growth rate is doing better than the best site. The proof enumerates sequences of possible excursions (and their probabilities) for which the less favorable site does better than the favorable site. Here, excursions are defined as the set of paths in the migration graph that start and end at the best site 1, with no intervening returns to site 1. We would like to note

that for the two-site case, the only possibility is to move from site 1 to site 2 and back. Excursions become more complex when we discuss the case for more than two sites in section 5. Once the excursions are defined we are able to balance the exponentially growing number of excursions against the exponentially large probability penalty incurred by repeated migration events to find a growth rate that will be achieved with probability 1.

Note that to ease the exposition the results for the case of two sites and more than two sites are presented separately in the main manuscript. In the Appendix, we present the general theorem for any number of sites. The mathematical details of the proof are presented in Appendix B.

Theorem 1. *The difference between stochastic growth rates, $a(m)$ and $a(0)$ is bounded below and above (scaled by a constant) as follows: For any positive δ,*

$$m^{4\tilde{\mu}/\sigma^2+\delta} - (1+\delta)m \leq a(m) - a(0) \leq m^{\left(\frac{4\tilde{\mu}/\sigma^2}{2\tilde{\mu}/\sigma^2+1}\right)-\delta} - m$$

for positive m sufficiently close to 0.

Here m is the probability of migration, $\tilde{\mu} = \mu_1 - \mu_2$ is the difference in expected growth rates between site 1 and site 2, $\sigma^2 = \mathrm{Var}(X_t^{(2)} - X_t^{(1)})$ is the variance in difference between growth rates at site 1 and site 2 (at time t), $a(m)$ is the stochastic growth rate with migration, $a(0)$ is the stochastic growth rate without migration, and $\delta > 0$ is a small positive quantity.

Note that the upper bound is dominated by the lower power of m, and the lower bound is dominated by the higher power of m. We are interested in the lower bound $m^{4\rho+\delta}$ for small m, where $\rho = \tilde{\mu}/\sigma^2$. Note that the increase in growth rates is at least as large as $m^{4\rho+\delta}$. The smaller the ρ, the more steeply the power $m^{4\rho}$ jumps up from $m = 0$. For the ratio ρ to be small either the numerator (the expected inter-site difference $\tilde{\mu}$) must be small or the denominator (the variance of inter-site difference σ^2) must be large. To put it differently, the evolution of migration will be favored when the variance of inter-site log growth rate difference (σ^2) is large enough to overcome a fixed mean difference ($\tilde{\mu}$). In the Appendix A, we present a figure (Figure 2) and a table with upper and lower powers, 4ρ and $\frac{4\rho}{2\rho+1}$, respectively. We simulate the slope as $\frac{\log(a(m_1)/a(m_2))}{\log(m_1/m_2)}$ by taking two values of migration rate ($m_1 = 1 \times 10^{-12}$ and $m_2 = 1 \times 10^{-10}$). Most of the simulated slopes lie within lower and upper powers as shown in Figure 2 in the Appendix.

Note that this increase ($m^{4\rho+\delta}$) is opposed by a linear penalty $-(|\mathbf{A}(0,0)|+\delta)m$, corresponding to the direct effect of population lost to the steady stream of out-migration from the optimal site. The linear penalty is of the form $|\mathbf{A}(0,0)|$ since δ can be made arbitrarily small. Since the dispersal matrix is constant $\mathbf{A}$, $-(|\mathbf{A}(0,0)|+\delta)$ is just $(1+\delta)$. In the Appendix B, $\mathbf{A}$ can vary with time ($\mathbf{A}_t$) and hence the linear term is of the form of expectation of absolute value of $\mathbf{A}_t(0,0)$. When ρ is smaller than $\frac{1}{4}$, the positive term will definitely overwhelm the linear penalty for sufficiently small positive values of m. For large values of ρ the linear migration penalty will definitely predominate, and no positive amount of migration will be favorable. The theorem does not exclude the possibility that this transition may occur at an even higher level, up to $\rho = \frac{1}{2}$ (when the power on the right-hand side is 1). But once the power of m is known to be below 1, this migration benefit persists for sufficiently small m regardless of any additional migration costs proportional to m.

In the next section, we use simulations to illustrate and examine the effects of $\tilde{\mu}$ and σ^2 through simulations. Following that, we use our simulation results to present an intuitive argument (section 4.3) for our analysis.

4.2 Simulations for two sites

We illustrate the theorem by examining the effects of varying fluctuations (σ^2), the difference in average growth rates between sites ($\tilde{\mu}$), or both on stochastic growth rate $a(m)$. The stochastic growth rate can be calculated from numerical simulations of many time steps of a population.

4.2.1 Effect of varying fluctuations on stochastic growth rate

We first find that when the difference in mean growth rates between the sites is small ($\tilde{\mu} = 0.05$ in Figure 2), a small positive migration rate will benefit from the variability between sites (described by σ^2) relative to no migration. As predicted by the theorem, increasing the variability σ^2 of the difference between $\widetilde{X}_t^{(1)} = X_t^{(1)} - X_t^{(0)}$ increases the benefit from migration. This may enable populations to grow faster than would have been possible at the single best site.

Our finding suggests that a species will benefit from the ability to wander between two similarly good territories, where environmental conditions vary separately between them from year to year. We discuss the intuition for the ratio $\rho = \frac{\tilde{\mu}}{\sigma^2}$ at the end of the section.

Figure 2 describes stochastic growth rate a as a function of migration rate m for different levels of variabil-

ity, characterized by σ^2. For low values of σ^2, the gain from migration is not sufficient to cause an overall increase in the long-term growth rate. However, for higher values of σ^2, introduction of even small level of migration leads to a sharp increase in stochastic growth rate. The sharp increase with migration (neighborhood of $m = 0$) is evident in Figure 2 and in the log plot as shown in Appendix A. Therefore, given enough variability (a high enough value of σ^2), migration rate m can increase $a(m)$ relative to $a(0)$.

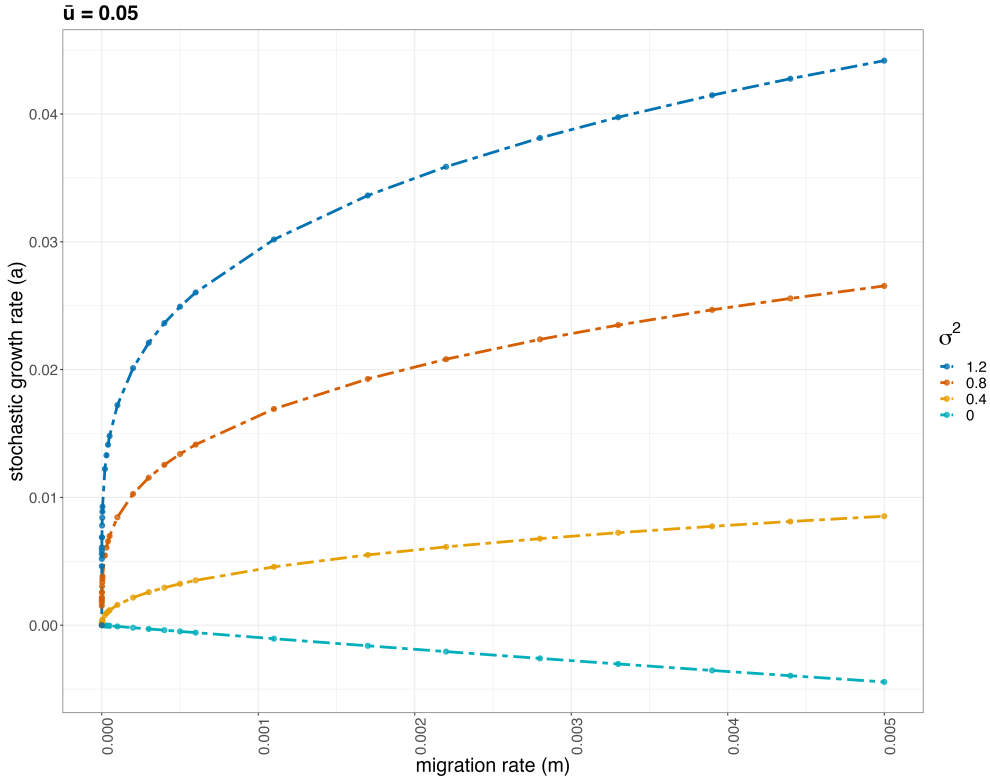


Figure 2: Stochastic growth rate as a function of migration rate. Different colors indicate different values of variability, described by σ^2. Higher values of σ^2 correspond to increase in stochastic growth rate for a fixed value of difference between average site-specific growth rates ($\tilde{\mu} = 0.05$). Here, $a(0)$ is 0.

4.2.2 Effect of varying difference in mean growth rates between sites

Next we consider the effect of changing the difference $\tilde{\mu} = \mu_1 - \mu_2$ between the average growth rates at sites 1 and 2. We find that the larger the difference between the expected growth rate in sites ($\tilde{\mu}_1$ increased from 0 to 0.45 in Figure 3), the smaller the benefit from migration, in terms of increase in long-term growth rate.

Biologically, a greater difference $\tilde{\mu}$ will make the rare events when growth at the inferior site exceeds the growth at the best site even rarer, while a greater variance (σ^2) will make such exceptional deviations more likely. Thus, high variance may compensate for a large difference in mean growth rate but only to an ex-

tent. The ratio determines whether populations may be better off without migration.

Figure 3 depicts the effect of changing $\tilde{\mu}$ on stochastic log growth rate a as a function of m. (The growth rate at site 1 is set to 0.) The first panel (top-left) presents the case where both sites have same expected growth rate ($\tilde{\mu} = 0$). In this case, when variability σ^2 equal to 0 the stochastic log growth rate is always 0, the common growth rate of the two sites (light blue line). For any nonzero σ^2, a jumps up rapidly as m departs from 0. (Note that the case $\tilde{\mu} = 0$ is not actually covered by the results of this paper. It is shown in [54] the increase in $a(m)$ as m increases from 0 is faster than any power law.) In the last panel, by contrast, $\tilde{\mu}$ is so large that the stochastic growth rate does not increase for any of the given fluctuation sizes. There is no benefit from migration, as the sites differ too greatly in their mean quality.

We illustrate the combined effect of changing site quality $\tilde{\mu}$ and variability σ^2 on the difference in stochastic rates (with and without migration) $a(m) - a(0)$ in Figures 4 (a) and (b). To emphasize the role of fluctuations in increasing the difference $a(m) - a(0)$ (denoted by "a_diff" in the legend in Figure 4 (a)), the limits on the x-axis are above 0.5. In Figure 4 (a) it is evident from the contours that for a fixed difference in expected growth rates between sites $\tilde{\mu}$, with enough variability σ^2, populations overcome the effect of migrating to a site with poorer habitat quality. In Figure 4 (b), the blue and orange color contrast the gain in growth rate from migration due to higher σ^2.

4.3 Heuristic arguments for the two-site case

Theorem 1 gives the lower and upper bound for the difference in stochastic growth rates (with and without migration), $a(m) - a(0)$. Since the lower bound is positive and increasing as a function of m, migration increases the overall growth rate on the long run. The bounds depend on the ratio $\rho = \tilde{\mu}/\sigma^2$. We discuss an intuition for the ratio. A similar ratio is discussed in [39] based on the mean and variance of the lognormal distribution.

In the case of no stochasticity, $X^{(1)} > X^{(2)}$, and dispersal should not be favored as the population achieves a higher growth rate in site 1. When some individuals disperse during a time interval from site 1 to site 2, the disperser's growth rate relative to a non-disperser is $Y_t = X_t^{(2)} - X_t^{(1)}$. The geometric growth rate, per unit time, is changed by a factor $\exp(Y_t)$. For $m = 0$, the expected growth rate, $\mathbb{E}(Y_t) = \mu_2 - \mu_1 = -\tilde{\mu} < 0$ and $\mathrm{Var}(Y_t) = \sigma^2 > 0$. Since we are assuming $X_t^{(1)}$ and $X_t^{(2)}$ are normal, the distribution of Y_t is lognormal and $\log \mathbb{E}(\exp(Y_t)) = -\tilde{\mu} + \frac{1}{2}\sigma^2$. So the expected relative growth rate is greater than 1 if the ratio $\frac{\tilde{\mu}}{\sigma^2} < \frac{1}{2}$.

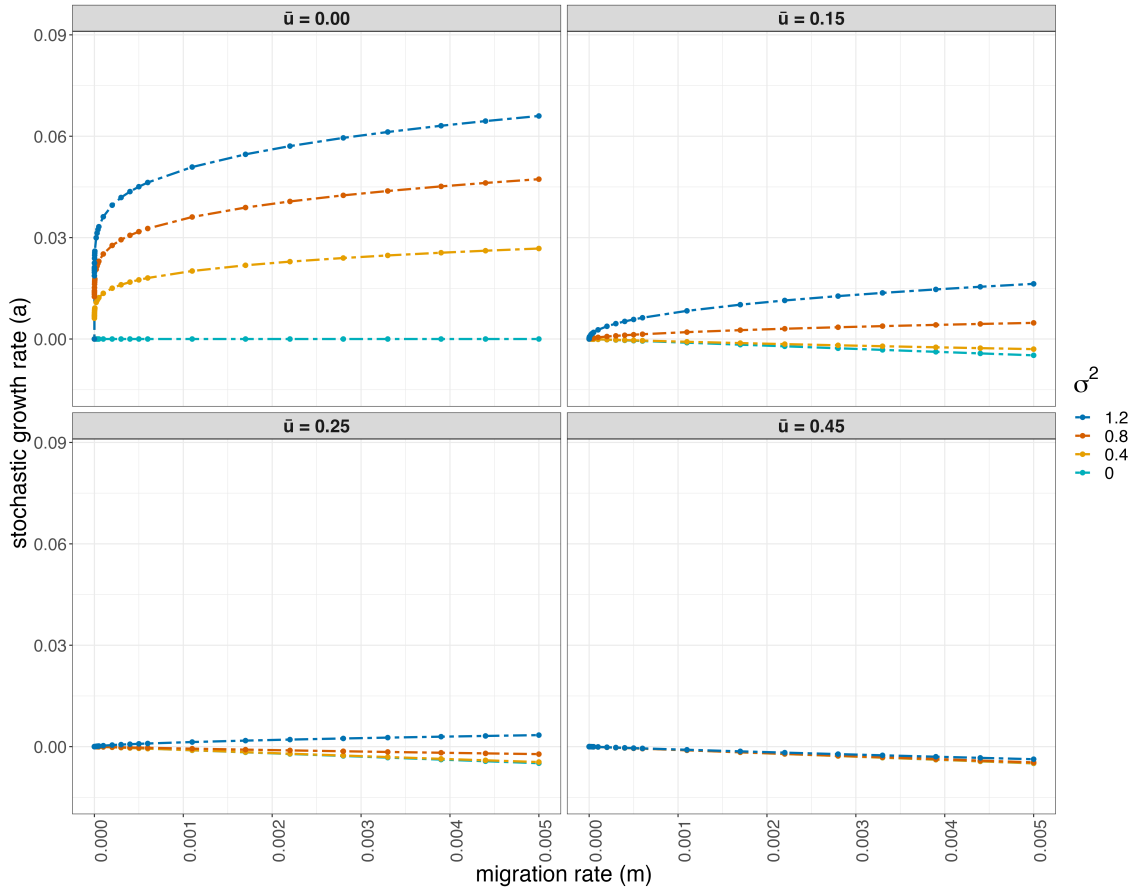


Figure 3: Stochastic growth rate as a function of migration rate. On the x-axis is the migration rate, on the y-axis is the stochastic growth rate. Each panel corresponds to a fixed value of difference in expected growth rates between two sites, $\tilde{\mu} = \mu_0 - \mu_1$. Different colors indicate values of variability, σ^2.

This is one approach to understand why the bounds depend on ratio of $\rho = \tilde{\mu}/\sigma^2$, and also why $\frac{1}{2}$ is an important threshold. However, the actual relative rate depends on the time period T over which a migrating ancestry experiences this relative rate, a concept that is formalised in section as the duration of an excursion (see Appendix B).

4.3.1 Effect of varying migration penalty

Dispersal may be expensive if dispersing individuals are vulnerable to incremental risks of death [16, 9]. Therefore, we examined the effect of increasing migration penalty or costs on the long-run growth rate. The costs are increased by changing the elements in the migration matrix to account for individuals lost during migration. We make the row sum $\sum_i \mathbf{A}(j,i)$ negative to account for loss of individuals as individuals leave the best site at the rate -m.

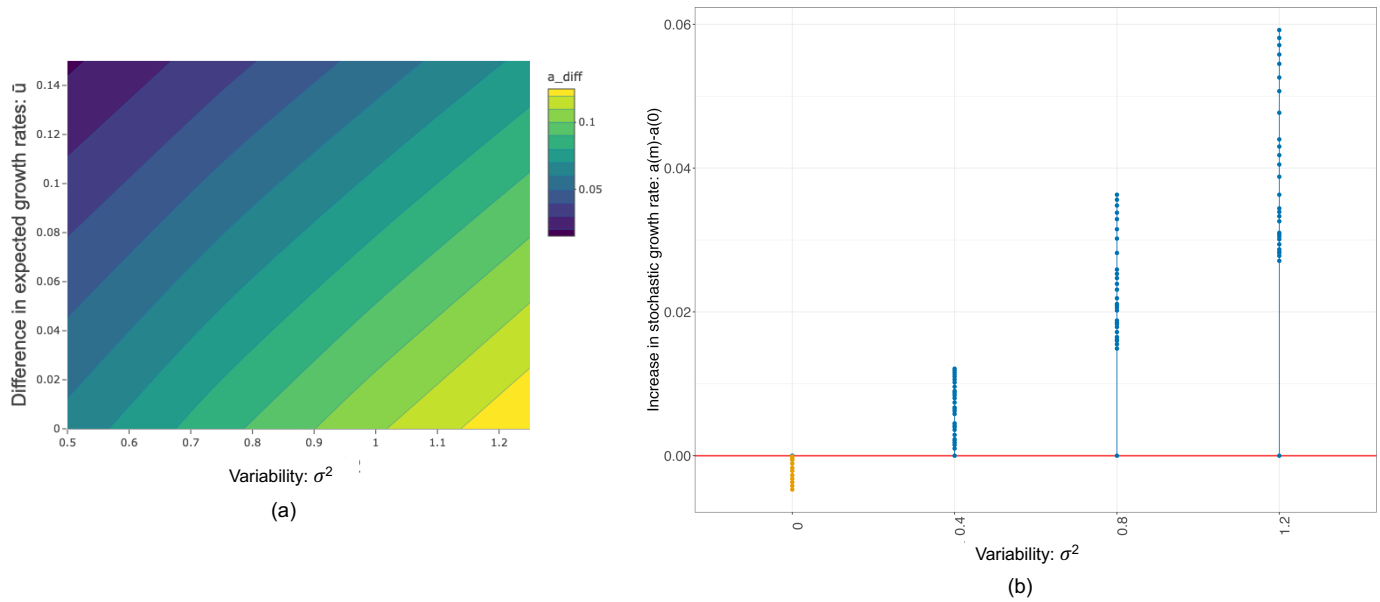


Figure 4: (a) The plot depicts the combined effect of $\tilde{\mu}$ and σ^2 on stochastic growth rate. On the y-axis is the difference in expected growth rates $\tilde{\mu}$. On the x-axis is the variability, σ^2. Each color in the legend ("a_diff") corresponds to values of $a(m) - a(0)$ (we fix a value of migration rate m at 0.005). (b) Here, on the y-axis is difference in stochastic growth rate, with and without migration, $a(m) - a(0)$. On the x-axis is the variability, σ^2 at the fixed value of $\tilde{\mu} = 0.1$. On the x-axis is the variability σ^2. The blue and orange color is to contrast the gain in growth rate from migration due to higher σ^2
.

This imposes an additional linear cost (as a function of migration rate m) to migration, which may make itself felt when m is large. Recall the lower bound in Theorem 1 is $m^{4\tilde{\mu}/\sigma^2+\delta} - (1+\delta)m$. For small m the balance of costs and benefits for migration at rate m depends essentially on $\tilde{\mu}/\sigma^2$. When this exponent $4\tilde{\mu}/\sigma^2$ is smaller than 1 the benefit will dominate and growth will be increased for sufficiently small positive m, regardless of the size of the linear migration penalty. That is, assuming the costs of migration (borne by the individuals migrating) are linear in m, the advantage of small migration will persist even when migration is inherently costly. When the power of m is greater than 1 the migration penalty will dominate for small m, and the effect of migration on log growth rate will be negative.

The important point regarding introducing a cost to migration is that the effect in **A** is always nearly linear in m, while the increase of a near 0 is often superlinear, growing as m^γ, where $\gamma = 4\tilde{\mu}/\sigma^2$. If the power γ

is strictly less than 1, the rapid increase in a near 0 will be qualitatively unaffected by a linear term for m sufficiently small.

Figure 5 shows two panels with increasing levels of migration penalty at fixed $\tilde{\mu}$ and σ^2. The two plots in the left and right panel of Figure 5 correspond to high and low variability ($\sigma^2 = 1$ and $\sigma^2 = 0.25$). The penalty is introduced by making $\sum_i \mathbf{A}(j,i)$ negative by 20% and 10% as this implies individuals were lost during migration. In the top panel, we apply a migration penalty of 20% for a fixed $\tilde{\mu}$ whereas the bottom panel corresponds to migration penalty of 10%. Despite a higher penalty in the top panel, we see that benefit from migration is attained when σ^2 is higher.

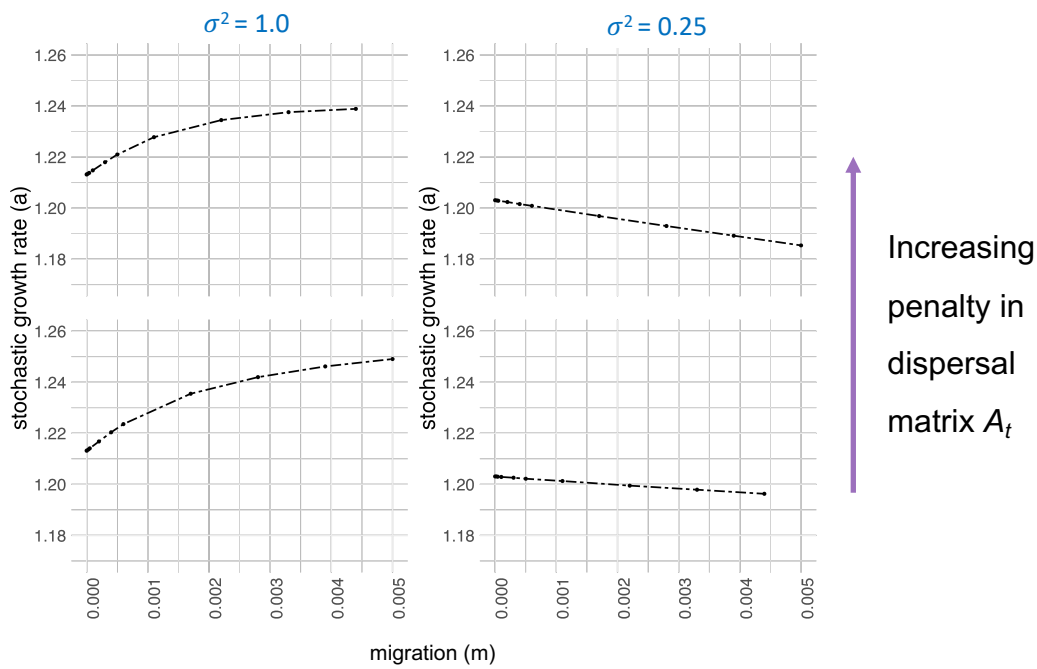


Figure 5: The figure has migration rate on the x-axis and the stochastic growth rate on the y-axis. The two plots in the left panel correspond to high variability ($\sigma^2 = 1$) while the right panel corresponds to low variability ($\sigma^2 = 0.25$). The top panel corresponds to migration penalty of 20% for a fixed $\tilde{\mu}$ whereas the bottom panel corresponds to migration penalty of 10%. The penalty is introduced by making $\sum_i \mathbf{A}(j,i)$ negative by 20% and 10% as this implies individuals were lost during migration.

5 Main result for many sites in variable environments

We now extend the system to any number of sites. We will index the sites from 0 instead of 1. Consider sites $0, 1, \ldots, d-1$ ($d > 3$) with log growth rates $X_t^{(0)}, X_t^{(1)}, \ldots, X_t^{(d-1)}$ during time period t. The expected growth rates at each site are μ_0, μ_1, $\cdots$, μ_{d-1}, respectively. Since we are interested in the difference in expected growth rates between of any site j from the unique best site 1, we define $\tilde{\mu}_j = \mu_0 - \mu_j$. As discussed

in earlier sections, the random variables $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ are assumed Gaussian with variance σ_i^2 and describe the fluctuations of the difference in growth rates at time t. The requirement that the log growth rates have Gaussian distribution is for simplicity of presentation; in the Appendix this assumption is relaxed to allow $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ to be sub-Gaussian with variance factor τ_j. The probabilistic reasoning for the theorem for multiple sites is analogous to that for the theorem for two sites.

The theorem establishes sequences of possible excursions (and their probabilities) for which the less favorable site does better than the favorable site. However, when there are more than two sites, there can be many different paths. The paths may involve no direct transition between the best site and the second-best site. Recall that the growth-rate bonus due to migration arises from small subpopulations migrating to an inferior site just as that site is experiencing an accidental run of exceptionally good conditions; and then depositing a fraction of its greatly expanded offspring back into site 0 to continue the run of normal (optimal) growth. Since migration involves only a fraction m of the population, the fraction making the round trip will be the even tinier m^2 for a two-site system, but in general m^κ, where κ is the number of migration events required to bring descendants back to site 0.

When considering a model with more than two sites we thus need to define for each site j a new variable κ_j, defined as the smallest length of a cycle in migration graph $\mathcal{M}$ that starts and ends at site 0, and passes through site j. For two sites, κ is always 2, since the possible transitions can be from site 0 to site 1 and site 1 to site 0. We incorporate κ by modifying the paths between sites, reflected in the dispersal matrix $\mathbf{A}$.

Theorem 2. *Let $\rho_j = \frac{\tilde{\mu}_j}{\sigma_j^2}$ and κ_j be the smallest length of a cycle in migration graph $\mathcal{M}$ that starts and ends at site 0, and passes through site j. Define $\underline{\eta} = \min_j 2\kappa_j\rho_j$, $\bar{\eta} = \min_j 2\kappa_j\rho_j/(1+2\rho_j)$. Then for any $\delta > 0$, we have for all $m > 0$ sufficiently small:*

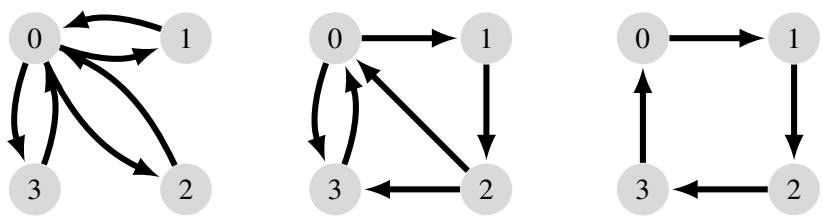
$$m^{\underline{\eta}+\delta} - (1+\delta)m \leq a(m) - a(0) \leq m^{\bar{\eta}-\delta} - m$$

Note that the lower bound is dominated by the higher power of m as $m \to 0$. As σ_j^2 increases, the power decreases, yielding conditions that make small deviations from the case of zero migration advantageous despite the costs that are merely proportional on the order of m. Note that κ features in the two-site case as well, but since $\kappa = 2$ is fixed for the two site case it is absorbed as a constant.

As discussed previously, the increase of a near 0 is often superlinear, growing as $m^{\underline{\eta}}$. If the power $\underline{\eta}$ is strictly less than 1, the rapid increase in a near 0 will be qualitatively unaffected by a linear term for m sufficiently small.

5.1 Example of movement patterns

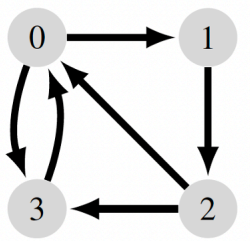
Consider the case for 4 sites- populations dispersing away from site 0 can return in a cycle of length 2, 3 or 4. For simplicity in the examples below we assume there is no loss of individuals. These matrices represent models where no net population is lost in migration – hence the rows sum to 0 – but nothing would be changed in the calculations below if we penalised migration. In Figure 6, the migration graphs and corresponding matrices are shown. Each graph corresponds to different length for κ_j.



$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -1 & 0.5 & 0.3 & 0.2 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -1 & 0.5 & 0 & 0.5 \\ 0 & -1 & 1 & 0 \\ 0.5 & 0 & -1 & 0.5 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array}$$

Figure 6: The migration graph $\mathcal{M}$ for four-sites and their respective migration matrices. For graph (a): κ_j is 2 for all sites j, for graph (b): κ_1 is 3 for site 1, and for graph (c): κ_j is 4 for all sites j.

For migration graph $\mathcal{M}$ in Figure 7, we can calculate $\rho = \tilde{\mu}_j/\sigma_j^2$ for the four-site graph. From the theorem, it is clear that to increase $a(m) - a(0)$, $\kappa\rho$ must be minimized. We see that $\kappa\rho$ is minimized at site 1, despite the fact that it is not in the shortest cycle, nor does it have the smallest mean difference in log growth rate from the best site 0, yielding a lower bound on the sensitivity to migration on the order of $m^{0.6}$. The upper bound $m^{0.41}$ comes from migration to site 3, which has the lowest mean log growth rate, but compensates for this with a shorter cycle length.



(a) Migration graph

j	$\mathbb{E}[X_t^{(j)}]$	σ_j^2	ρ_j	κ_j	$\kappa_j\rho_j$	$\frac{\kappa_j\rho_j}{1+2\rho_j}$
0	1					
1	0.4	3	0.2	3	0.6	0.43
2	0.7	0.6	0.5	3	1.5	0.75
3	0.3	2	0.35	2	0.7	0.41

(b) Selecting the second site

Figure 7: Example of calculating exponents for the migration graph given in Figure 6(b). The values of $\mathbb{E}[X_t^{(j)}]$ and σ_j^2 are arbitrary, and have been selected for illustrative purposes.

5.2 Simulations for a 4-site model

5.2.1 Effect of varying fluctuations and difference in mean growth rates

For four sites (labelled 0, 1, 2, and 3), we assume site 0 has highest average growth rate μ_0. We define the difference for sites from the best site as follows: $\tilde{\mu}_1 = \mu_0 - \mu_1$, $\tilde{\mu}_2 = \mu_0 - \mu_2$, and $\tilde{\mu}_3 = \mu_0 - \mu_3$ and vary the variance of the fluctuations σ^2. For the simulations we consider two cases: first, we consider the vector $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3) = (0.05, 0.10, 0.15)$. Since $\tilde{\mu}$ refers to differences, the vector corresponds to a case when the three sites are similar to the best site on average because the difference between sites is atmost 0.15. Now, we consider $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ is $(0.35, 0.40, 0.45)$ so that the difference in expected growth rate from the best site is at least 0.35. Please note that these numbers are chosen for purposes of illustration, and have no special significance. Other combinations of parameters would produce qualitatively similar results.

As shown in Figure 8, similar to the results for 2 sites, we find that increasing fluctuations increases the gain in stochastic growth rate relative to no migration. However, as the gap grows between mean growth at the best site and at the other sites, the benefits of migration for stochastic growth rate shrink, and ultimately migration will be detrimental despite high fluctuations. Hence populations cannot benefit from variability in this case (last panel Figure 8). We consistently find that the smallest kappa yields the greatest increase in stochastic growth rate, as discussed in the next section.

5.2.2 Effect of varying path length between sites

We now examine how κ affects the long-term growth rate. Recall that κ_j is defined to be the length of shortest cycle in the migration graph that starts and ends at 0, passing through j.

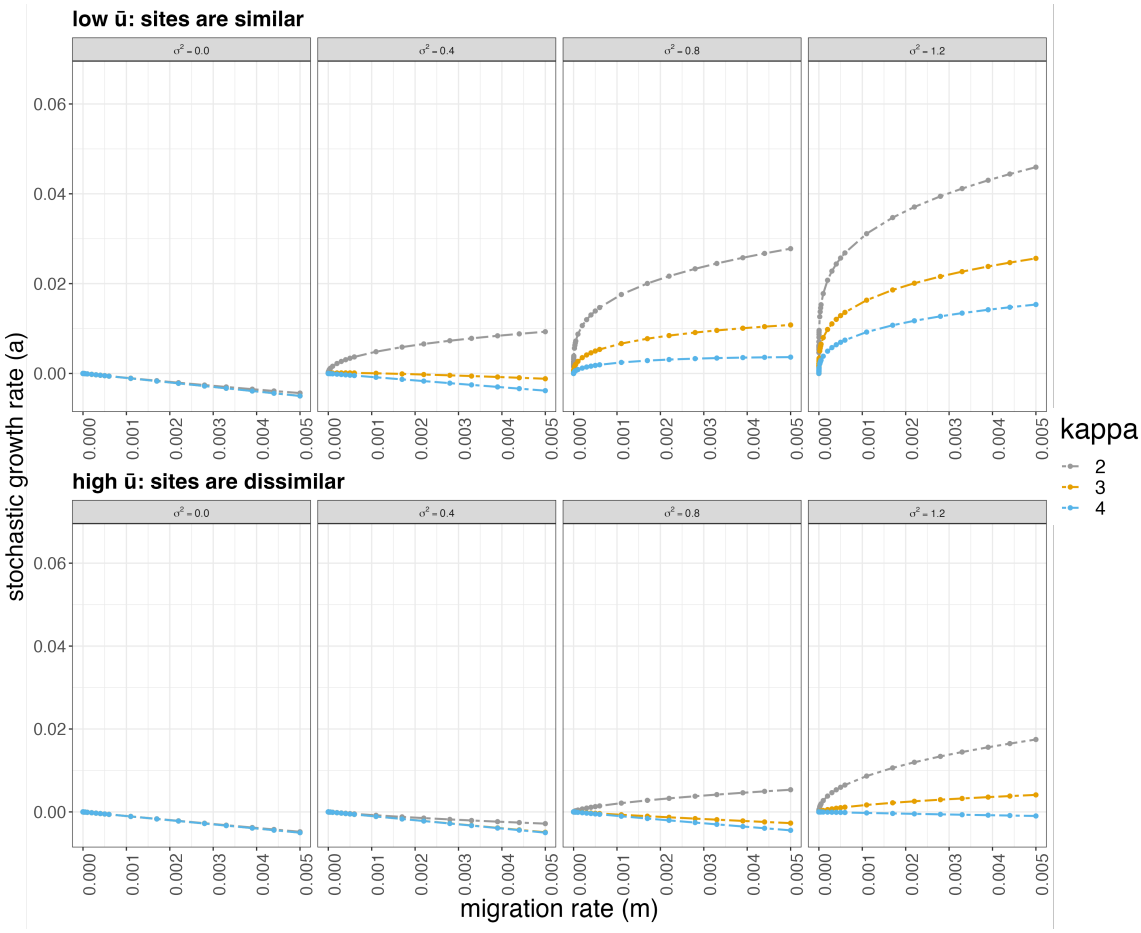


Figure 8: The x-axis is migration rate and y-axis is the stochastic growth rate: $a(m)$. The top panel corresponds to values for the vector $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ such that the sites are similar. The panel on the bottom corresponds to values such that sites are dissimilar. Each color corresponds to a value of κ (for four sites) since an excursion can be made in steps of 2, 3 and 4.

In Figure 8, each color corresponds to a value of κ for four sites since an excursion can be made in lengths of 2, 3 and 4. From the theorem, we expected the shortest excursion with $\kappa = 2$ to lead to the maximum gain in stochastic growth rate, and that is what we find in Figure 8. The longest excursion with $\kappa = 4$ leads to the lowest increase in stochastic growth rate. Although populations sample more variability by increasing the number of possible paths and sites, it is important to return to the best site in the shortest sequence possible to maximize the gain in overall growth rate. That is, increase in growth rate is determined by being able to gain access to the best site and return with a smaller number of low-probability migration events (lower κ).

In addition, as evidenced by the last two panels in Figure 8, the length of κ matters when the fluctuations are high and when the sites are more similar to the best site. The bottom two panels of Figure 9 show how the stochastic growth rates for different path lengths remain similar when the difference between sites is

increased (high μ). Hence, path length matters most when the sites are similar, as illustrated in top two panels of Figure 9.

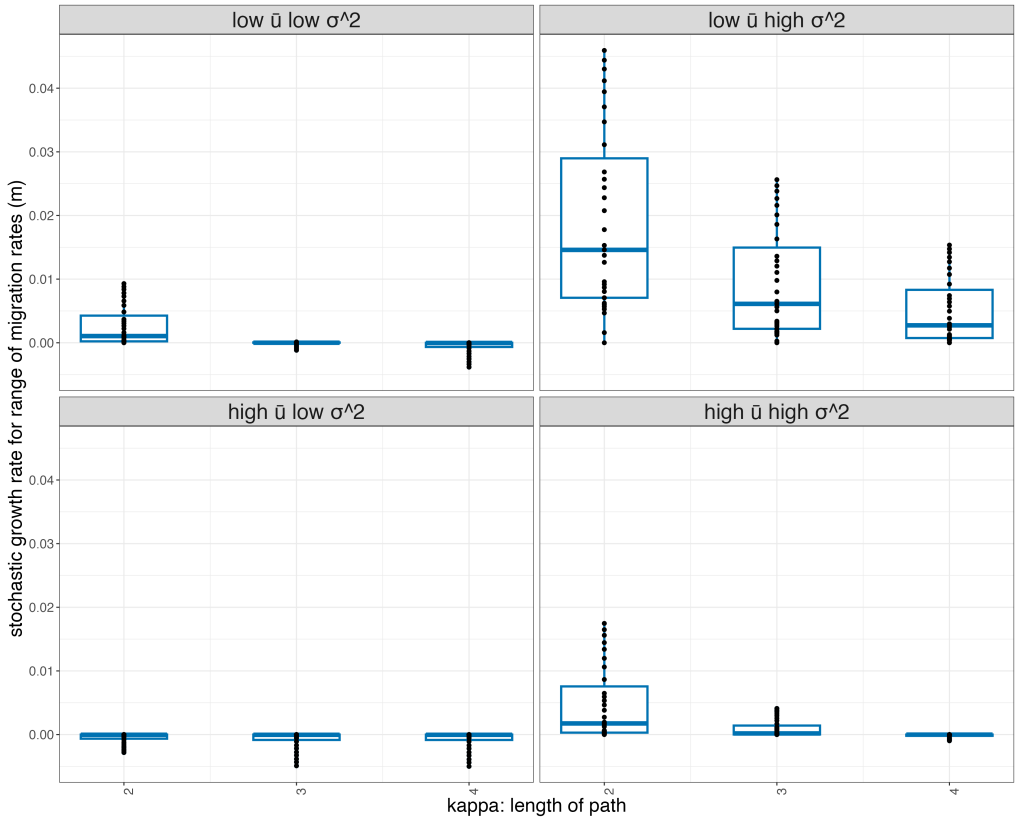


Figure 9: The effect of changing topology on growth rates in a 4-site model. The x-axis is the length of the shortest path to the best site (site 0) in steps of 2, 3 and 4. The y-axis are the values of stochastic growth rate for a range of m. Each panel corresponds to a fixed level of σ^2, and varies in low and high difference in expected growth rates from best site.

6 Discussion

We provide conditions under which introducing a small amount of migration will increase the growth rate of the population relative to the zero-migration case. Our research contributes fundamentally to the literature by proving that the increase in long-run growth rate $a(m) - a(0)$ with migration at rate m is bounded above and below by powers of m for m in a neighborhood of 0, in a simple model of multi-site population growth that takes environmental variability into account. These powers depend on the ratio $\tilde{\mu}/\sigma^2$, where $\tilde{\mu}$ is the difference in expected growth rates and σ^2 is the variance of the difference of site-specific growth rates at time t. Our results complement the analysis in [20] of optimal migration rates for populations with stochastic growth rates at different sites.

We illustrate our results (through simulations) by examining a wide range of fluctuations, differences in

average growth rates between sites, and path lengths. We find that as the average (over time) qualities of the sites diverge, the increase in long-run growth rate due to introduction of migration becomes less pronounced. Environmental fluctuations compensate (to an extent) for the fixed difference in mean log growth rates. Biologically, this means that habitat quality and fluctuations are both important determinants of migration evolution and population persistence.

By changing how each site is connected to the 'best' site (varying κ), we find that when variability σ^2 is low, excursions of any length are unlikely to overcome the deficit in average growth (refer to Figure 9). When variability σ^2 is high, excursions may overcome average differences when the length of κ is small and will yield the maximum increase in stochastic growth rate in that case. Recent studies on seabirds [19, 33] show that a longer migration route does not necessarily have carry-over effects on subsequent reproduction and apparent survival for little auk *(Alle alle)* and lesser black-backed gulls *(Larus fuscus)*, respectively.

Our results build on earlier work that emphasized that the role of migration in allowing populations to keep pace with spatial and temporal variation in the habitat [53, 8, 61]. [5] formulated the concept of "spreading of risk" as a way to survive and reproduce in a spatially variable environment so that any one failure will not be decisively harmful. This concept recognizes the importance of migration as an escape from unfavorable conditions, and adaptation to changing environments by exploiting new sites. [53] intuited that the rate of migration should be highest in species inhabiting the most temporary habitats and found that migratory movement in terrestrial Arthropoda is positively correlated with impermanence of the habitat. Numerous empirical studies have found that dispersal is more common in rapidly-changing habitats; e.g., [18] studied 35 species of plant hoppers living in habitats of varying persistence, and found that winged morphs were more common in ephemeral habitats while flightlessness was favored in persistent habitats.

A similar argument holds in the case of diapause, which is a delay in biological development, often at an early life stage. For example, many plant species produce seeds that do not germinate immediately, but remain dormant for extended times forming seed banks [60, 61]. The paper by [54] considers a two-year life-cycle model of a species, where the first year is spent growing from a seed to an adult, and the second year is spent producing seeds. They examined the effect on growth rates if a small fraction of the seeds remains dormant for one year. When vital rates remain constant over time, delaying individual growth reduces population growth. However, when environmental conditions are variable, even a tiny amount of

dormancy can boost population growth. This finding parallels ours — it is equivalent to the present model with all sites having exactly the same average growth rate — but requires a substantially different analysis. Taken together, these results determine conditions under which there are departures from the reduction principle, that is, the selection for reduced dispersal (under constant selection) [3, 4].

Our findings have implications in this era of global habitat degradation, altitudinal range-shifts, and species redistribution [37, 25] which provide directions for future research. Importantly, our work provides a way to study migration patterns (of birds or mammals) by examining variation in environmental conditions, and metrics describing site-specific growth rates. These results may also be important in conservation biology and low-density populations: when there are superior sites in a sea of poor habitats, variability, habitat quality across space, and length of path may be key to determining the importance of migration.

Theoretical work on the evolution of dispersal has a long history [27, 39, 43, 28, 56, 2]. Indeed, there is a large and rich literature in dispersal and metapopulation models [56, 12, 13, 55, 36, 26, 45, 47] that remain important in most applications. Studies that unify theory and experiment, as discussed by [40], will provide ways to test or parameterize theoretical models [49]. It is pertinent to account for density-dependence when there are variations in population growth rates across space and analyze how adding more individuals that migrate impacts stochastic growth rate. Importantly, our research does not distinguish between different patterns of movement, such for breeding dispersal or seasonal migratory patterns and will be an important avenue for further research [41, 11]. Finally, future work may also examine how variability in dispersal rate impacts overall populations, and account for structure in spatial sites.

7 Conclusions

We solve a fundamental question in evolutionary ecology: can dispersal between sites evolve in variable environments? We use a simple model in which one site is on average the best (highest expected growth rate), and dispersal moves individuals between the best site and other sites. We show that sufficient variability means that an individual who disperses can exploit favorable runs of the environment in other sites. We provide precise inequalities that demonstrate the trade-off between differences in expected growth rates between sites, and variability between sites. We illustrate our results (with simulations) for two and four sites. For the multiple (≥ 2) site case, we consider the effect of alternative paths to return to the best site. This fundamental idea about the evolution of dispersal is old but had not been demonstrated mathemati-

cally even in a simple model.

Contributions

HJ assisted with the mathematics, planned and carried out the simulations, and was primarily responsible for writing the paper. DS conceived and carried out the mathematics, and assisted with the simulations. ST proposed, led, and assisted with all phases of the project. All authors contributed to shaping the final version.

Acknowledgement

HJ would like to thank Siddharth Srinivasan for the very helpful discussions that enabled a better understanding of the mathematics and large deviation theory. HJ would also like to thank Nikunj Goel for providing constructive feedback for editing the draft. We thank Stanford University and University of Oxford as our funding sources.

References

[1] Altenberg, L. A generalization of theory on the evolution of modifier genes. (Citeseer, 1984)

[2] Altenberg, L. The evolution of dispersal in random environments and the principle of partial control. *Ecological Monographs*. **82**, 297-333 (2012)

[3] Altenberg, L. & Feldman, M. Selection, generalized transmission and the evolution of modifier genes. I. The reduction principle. *Genetics*. **117**, 559-572 (1987)

[4] Altenberg, L., Liberman, U. & Feldman, M. Unified reduction principle for the evolution of mutation, migration, and recombination. *Proceedings Of The National Academy Of Sciences*. **114**, E2392-E2400 (2017)

[5] Andrewartha, H. & Birch, L. The ecological web: more on the distribution and abundance of animals. (University of Chicago Press,1986)

[6] Baker, R. & Others Evolutionary ecology of animal migration. (1978)

[7] Balkau, B. & Feldman, M. Selection for migration modification. *Genetics*. **74**, 171-174 (1973)

[8] Den, B. PJ 1968. Spreading of risk and stabilization of animal numbers. *Acta Biotheor*. **18** pp. 165-194

[9] Bonte, D. & Dahirel, M. Dispersal: a central and independent trait in life history. *Oikos*. **126**, 472-479 (2017)

[10] Benaim, M. & Schreiber, S. Persistence of structured populations in random environments. *Theoretical Population Biology*. **76**, 19-34 (2009)

[11] Boulinier, T., Kada, S., Ponchon, A., Dupraz, M., Dietrich, M., Gamble, A., Bourret, V., Duriez, O., Bazire, R., Tornos, J. & Others Migration, prospecting, dispersal? What host movement matters for infectious agent circulation?. *Integrative And Comparative Biology*. **56**, 330-342 (2016)

[12] Cadet, C., Ferriere, R., Metz, J. & Van Baalen, M. The evolution of dispersal under demographic stochasticity. *The American Naturalist*. **162**, 427-441 (2003)

[13] Cheptou, P., Carrue, O., Rouifed, S. & Cantarel, A. Rapid evolution of seed dispersal in an urban environment in the weed Crepis sancta. *Proceedings Of The National Academy Of Sciences*. **105**, 3796-3799 (2008)

[14] Matthysen, E., Clobert, J., Baguette, M., Benton, T. & Bullock, J. Dispersal ecology and evolution. (Oxford University Press, Oxford,2012)

[15] Cohen, D. & Levin, S. Dispersal in patchy environments: the effects of temporal and spatial structure. *Theoretical Population Biology*. **39**, 63-99 (1991)

[16] Clobert, J. Dispersal ecology and evolution. (Oxford University Press,2012)

[17] Cohen, J. Ergodic theorems in demography. *Bulletin Of The American Mathematical Society*. **1**, 275-295 (1979)

[18] Denno, R., Roderick, G., Olmstead, K. & Dobel, H. Density-related migration in planthoppers (Homoptera: Delphacidae): the role of habitat persistence. *The American Naturalist*. **138**, 1513-1541 (1991)

[19] Dufour, P., Wojczulanis-Jakubas, K., Lavergne, S., Renaud, J., Jakubas, D. & Descamps, S. A two-fold increase in migration distance does not have breeding consequences in a long-distance migratory seabird with high flight costs. *Marine Ecology Progress Series*. **676** pp. 117-126 (2021)

[20] Evans, S., Ralph, P., Schreiber, S. & Sen, A. Stochastic population growth in spatially heterogeneous environments. *Journal Of Mathematical Biology*. **66**, 423-476 (2013)

[21] Feldman, M. Selection for linkage modification: I. Random mating populations. *Theoretical Population Biology*. **3**, 324-346 (1972)

[22] Feldman, M., Christiansen, F. & Brooks, L. Evolution of recombination in a constant environment. *Proceedings Of The National Academy Of Sciences*. **77**, 4838-4841 (1980)

[23] Feldman, M. & Liberman, U. An evolutionary reduction principle for genetic modifiers.. *Proceedings Of The National Academy Of Sciences*. **83**, 4824-4827 (1986)

[24] Gadgil, M. Dispersal: population consequences and evolution. *Ecology*. **52**, 253-261 (1971)

[25] Goel, N., Guttal, V., Levin, S. & Staver, A. Dispersal increases the resilience of tropical savanna and forest distributions. *The American Naturalist*. **195**, 833-850 (2020)

[26] Grilli, J., Barabás, G. & Allesina, S. Metapopulation persistence in random fragmented landscapes. *PLoS Computational Biology*. **11**, e1004251 (2015)

[27] Hamilton, W. & May, R. Dispersal in stable habitats. *Nature*. **269**, 578-581 (1977)

[28] Hanski, I. & Simberloff, D. The metapopulation approach, its history, conceptual domain, and application to conservation. *Metapopulation Biology*. pp. 5-26 (1997)

[29] Hastings, A. Can spatial variation alone lead to selection for dispersal?. *Theoretical Population Biology*. **24**, 244-251 (1983)

[30] Henriques-Silva, R., Boivin, F., Calcagno, V., Urban, M. & Peres-Neto, P. On the evolution of dispersal via heterogeneity in spatial connectivity. *Proceedings Of The Royal Society B: Biological Sciences*. **282**, 20142879 (2015)

[31] Johnson, M. & Gaines, M. Evolution of dispersal: theoretical models and empirical tests using birds and mammals. *Annual Review Of Ecology And Systematics*. **21**, 449-480 (1990)

[32] Karlin, S. & Others Classifications of selection-migration structures and conditions for a protected polymorphism. *Evol. Biol*. **14**, 204 (1982)

[33] Kentie, R., Morgan Brown, J., Camphuysen, K. & Shamoun-Baranes, J. Distance doesn't matter: migration strategy in a seabird has no effect on survival or reproduction. *Proceedings Of The Royal Society B*. **290**, 20222408 (2023)

[34] Khaladi, M., Lebreton, J. & Khermjioui, A. The evolution of dispersal in random environment. *Acta Biotheoretica*. **60** pp. 155-165 (2012)

[35] Kirkland, S., Li, C. & Schreiber, S. On the evolution of dispersal in patchy landscapes. *SIAM Journal On Applied Mathematics*. **66**, 1366-1382 (2006)

[36] Legrand, D., Trochet, A., Moulherat, S., Calvez, O., Stevens, V., Ducatez, S., Clobert, J. & Baguette, M. Ranking the ecological causes of dispersal in a butterfly. *Ecography*. **38**, 822-831 (2015)

[37] Lenoir, J., Gégout, J., Marquet, P., Ruffray, P. & Brisse, H. A significant upward shift in plant species optimum elevation during the 20th century. *Science*. **320**, 1768-1771 (2008)

[38] Lowe, W. & McPeek, M. Is dispersal neutral?. *Trends In Ecology & Evolution*. **29**, 444-450 (2014)

[39] Levin, S., Cohen, D. & Hastings, A. Dispersal strategies in patchy environments. *Theoretical Population Biology*. **26**, 165-191 (1984)

[40] Lustenhouwer, N., Moerman, F., Altermatt, F., Bassar, R., Bocedi, G., Bonte, D., Dey, S., Fronhofer, E., Rocha, É., Giometto, A. & Others Experimental evolution of dispersal: Unifying theory, experiments and natural systems. *Journal Of Animal Ecology*. **92**, 1113-1123 (2023)

[41] Matthysen, E. & Clobert, J. Multicausality of dispersal: a review. *Dispersal Ecology And Evolution*. **27** pp. 3-18 (2012)

[42] McNamara, J. & Dall, S. The evolution of unconditional strategies via the multiplier effect. *Ecology Letters*. **14**, 237-243 (2011)

[43] McPeek, M. & Holt, R. The evolution of dispersal in spatially and temporally varying environments. *The American Naturalist*. **140**, 1010-1027 (1992)

[44] Metz, J., Nisbet, R. & Geritz, S. How should we define fitness for general ecological scenarios?. *Trends In Ecology & Evolution*. **7**, 198-202 (1992)

[45] Miller, Z. & Allesina, S. Metapopulations with habitat modification. *Proceedings Of The National Academy Of Sciences*. **118**, e2109896118 (2021)

[46] Oldfather, M., Van Den Elzen, C., Heffernan, P. & Emery, N. Dispersal evolution in temporally variable environments: implications for plant range dynamics. *American Journal Of Botany*. **108**, 1584-1594 (2021)

[47] Poli, C., Meyer, K., Darby, P., Dudek, S., Kent, G. & Fletcher Jr, R. Foray movements are common and vary with natal habitat for a highly mobile bird. *Ecology And Evolution*. **14**, e11096 (2024)

[48] Roerdink, J. The biennial life strategy in a random environment. *Journal Of Mathematical Biology*. **26** pp. 199-215 (1988)

[49] Ronce, O. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annu. Rev. Ecol. Evol. Syst.*. **38** pp. 231-253 (2007)

[50] Saastamoinen, M., Bocedi, G., Cote, J., Legrand, D., Guillaume, F., Wheat, C., Fronhofer, E., Garcia, C., Henry, R., Husby, A. & Others Genetics of dispersal. *Biological Reviews*. **93**, 574-599 (2018)

[51] Schreiber, S. The evolution of patch selection in stochastic environments. *The American Naturalist*. **180**, 17-34 (2012)

[52] Schreiber, S. & Li, C. Evolution of unconditional dispersal in periodic environments. *Journal Of Biological Dynamics*. **5**, 120-134 (2011)

[53] Southwood, T. Migration of terrestrial arthropods in relation to habitat. *Biological Reviews*. **37**, 171-211 (1962)

[54] Steinsaltz, D. & Tuljapurkar, S. Stability of fixed life histories to perturbation by rare diapause. *ArXiv Preprint ArXiv:1809.04027*. (2018)

[55] Strasser, C., Neubert, M., Caswell, H. & Hunter, C. Contributions of high-and low-quality patches to a metapopulation with stochastic disturbance. *Theoretical Ecology*. **5** pp. 167-179 (2012)

[56] Travis, J. & Dytham, C. Habitat persistence, habitat availability and the evolution of dispersal. *Proceedings Of The Royal Society Of London. Series B: Biological Sciences*. **266**, 723-728 (1999)

[57] Travis, J., Murrell, D. & Dytham, C. The evolution of densitydependent dispersal. *Proceedings Of The Royal Society Of London. Series B: Biological Sciences*. **266**, 1837-1842 (1999)

[58] Tuljapurkar, S. Population dynamics in variable environments. (Springer Science & Business Media,2013)

[59] Tuljapurkar, S. Population dynamics in variable environments. III. Evolutionary dynamics of r-selection. *Theoretical Population Biology*. **21**, 141-165 (1982)

[60] Tuljapurkar, S. An uncertain life: demography in random environments. *Theoretical Population Biology*. **35**, 227-294 (1989)

[61] Tuljapurkar, S. Delayed reproduction and fitness in variable environments.. *Proceedings Of The National Academy Of Sciences*. **87**, 1139-1143 (1990)

[62] Van Valen, L. Group selection and the evolution of dispersal. *Evolution*. pp. 591-598 (1971)

[63] Wiener, P. & Tuljapurkar, S. Migration in variable environments: exploring life-history evolution using structured population models. *Journal Of Theoretical Biology*. **166**, 75-90 (1994)

[64] Peres, Y. Analytic dependence of Lyapunov exponents on transition probabilities. *Lyapunov Exponents: Proceedings Of A Conference Held In Oberwolfach, May 28June 2, 1990*. pp. 64-80 (2006)

SUPPLEMENTARY INFORMATION: TEMPORAL VARIABILITY CAN PROMOTE MIGRATION BETWEEN HABITATS

Appendix A

Approximately linear relationship near 0 in the Log-Log Plot for stochastic growth rate (a) versus migration rate (m).

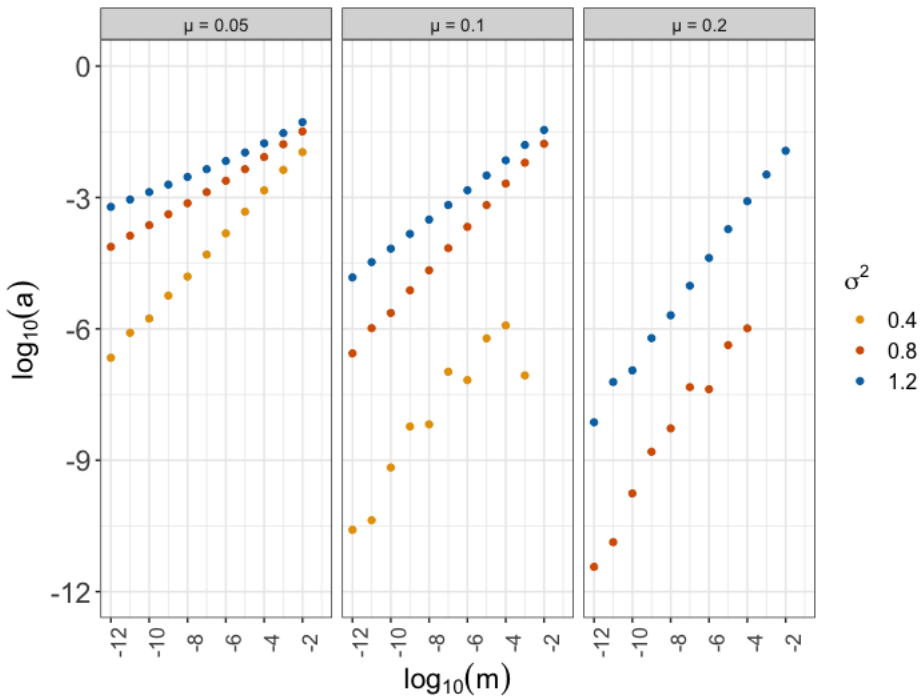


Figure 1. The y-axis represents logarithm of the stochastic growth rate (a). On the x-axis is the logarithm of the migration rate (m). Each veritcal panel corresponds to expected growth rates between sites $\tilde{\mu} = \mu_1 - \mu_2 = 0.05, 0.1, 0.2$. Colours indicate values of variability, σ^2. There is no value for $\sigma^2 = 0$ as the stochastic growth rate (a) is negative and thus the logarithm for it is not defined.

Simulated slopes within bounds from the Theorem.

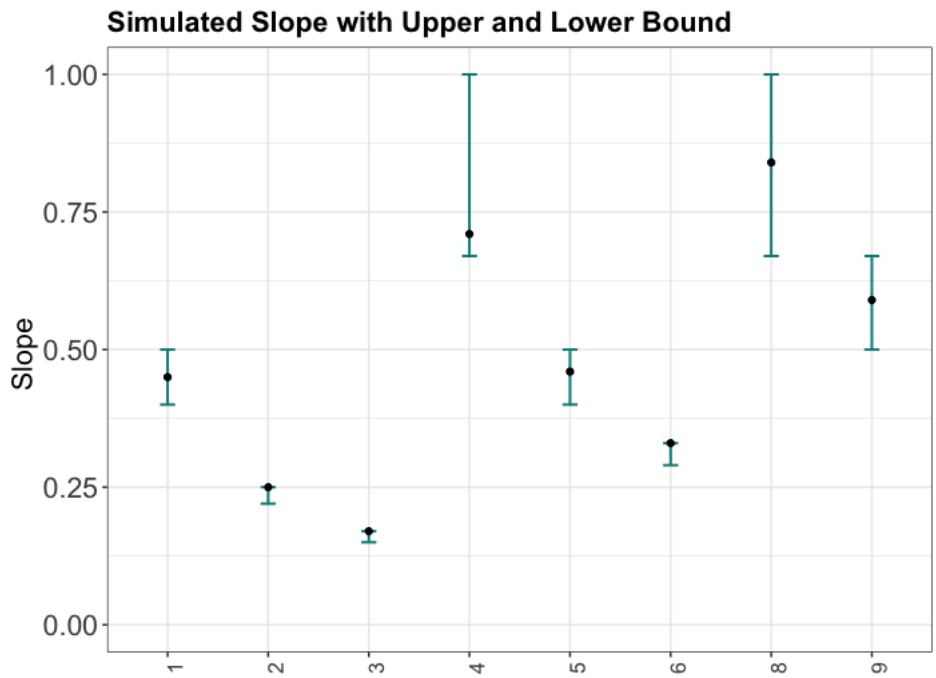


FIGURE 2. The y-axis represents the slope, upper bound and lower bound for different combinations of μ and σ^2 shown in Figure 2. The slope (equivalent to the power of m) is calculated between two values of migration rate ($m_1 = 1 \times 10^{-12}$ and $m_2 = 1 \times 10^{-10}$) as $\frac{\log(a(m_1)/a(m_2))}{\log(m_1/m_2)}$. The blue-green bars represent the upper and lower bounds predicted by the theorem asymptotically as $\log m \to -\infty$, and the black points represent the simulated slope value. The x-axis refers to different combinations of $\tilde{\mu}$ and σ^2 as defined in Appendix Figure 1, and tabulated in Table 1. Note that for simulation 7 the power is greater than 1, hence the change in growth rate is negative, hence there is no log plot and no slope.

TABLE 1. Table for slopes and corresponding upper and lower bounds

Sim No.	$\tilde{\mu}$	σ^2	Slope	Upper bound	Lower bound
1	0.05	0.40	0.448	0.500	0.400
2	0.05	0.80	0.247	0.250	0.222
3	0.05	1.20	0.169	0.167	0.154
4	0.10	0.40	0.711	1.000	0.167
5	0.10	0.80	0.460	0.500	0.400
6	0.10	1.20	0.328	0.333	0.286
7	0.20	0.40	NA	2.000	1.000
8	0.20	0.80	0.838	1.000	0.667
9	0.20	1.20	0.592	0.667	0.500

Appendix B

The goal of the appendix is to describe and prove Theorem 1 and 2 discussed in the main manuscript. Note that for simplicity, the theorems for the case of two sites and more than two sites are presented separately in the main manuscript. In the Appendix, we present the general theorem for any number of sites and the 2-site theorem is a special case of the general theorem.

First, we state the key notations and assumptions. We define sub-Gaussian random variables and state Theorem 1 and 2.1. Next we define *ancestries* and *excursions* — sequences of sites that are allowed by the migration graph — that will be enumerated in the derivation of upper (in section 4) and lower bounds (in section 5).

We summarize some key assumptions and definitions from the main manuscript below:

1. Notation and basic assumptions

- Let there be d sites such that populations $0, \ldots, d-1$ have log growth rates $X_t^{(i)}$ during time period t at site i. The vectors $X_t = (X_t^{(0)}, \ldots, X_t^{(d-1)})$ are assumed independent, but the components within each X_t — corresponding to growth rates at different sites at the same time — may be dependent.

- We write $\mu_i := \mathbb{E}[X_t^{(i)}]$, and assume that μ_0 is strictly the largest of these. Writing $\tilde{\mu}_i := \mu_0 - \mu_i$, these are all positive. We write $\mathbf{X}$ for the complete collection of all $X_t^{(j)}$ for $j = 0, 1, \ldots, d-1$, $0 \leq t < \infty$.

- The random variables $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ are assumed sub-Gaussian with variance factor τ_i, defined below in (2). When they are Gaussian, as we have assumed for the simulations, τ_i is then simply the variance.

- Suppose $G_1, G_2, \ldots$ is an i.i.d. sequence of $d \times d$ diagonal matrices. The diagonal elements of G_t are the log growth rates at each site at time t such that $X_t^{(i)}$ all have finite mean μ_i and finite variance.

- We let A_t be an i.i.d. sequence of $d \times d$ matrices with negative entries on the diagonal and non-negative entries off the diagonal. We assume that the pairs (G_t, A_t) are all independent, but not necessarily that A_t and G_t are independent for any given t. The entries of A_t are all assumed to be almost-surely bounded.

1.1. **Migration graph:** We define the migration graph $\mathcal{M}$ to be an irreducible directed graph whose vertices are the sites $\{0, \ldots, d-1\}$, representing the transitions that have nonzero probability. We assume that $\mathcal{M}$ is connected.

Let m parameterize the overall rate of migration. Thus, migration rates from site i to site j in $mA_t(j, i)$. Population distributions are thus column vectors, and are updated from time $t-1$ to time t by left multiplication.

When $i \to j$ (i is directly linked to j in $\mathcal{M}$) we assume $\mathbb{E}[(\log A_t(j,i))_+]$ exists and is finite, though $A_t(j,i)$ may still have nonzero probability of being 0. We write $\mu_A := \max_{i,j} \mathbb{E}[(\log A_t(j,i))_+]$.

We will assume that $A_t(i,i)$ is almost surely bounded below, and we will henceforth always be assuming that m is sufficiently small that $1 + mA_t(i,i) > 0$ for all i with probability 1. We write $\Delta^i := -\mathbb{E}[A_t(i,i)]$, which we assume also to be finite.

Using the system of equations defined in the main manuscript in Equation (3), we define

$$D_t(m) := (I_d + mA_t) \exp\big(\mathrm{diag}(X_t)\big) .$$

For $m > 0$ the i.i.d. sequence $D_t(m)$ satisfies the conditions for the existence of a stochastic growth rate independent of starting condition [Coh79]. That is, if we define the partial products

$$R_T(m) := D_{T-1}(m) \cdot D_{T-2}(m) \cdot \dots \cdot D_0(m)$$

then

$$a(m) := \lim_{T\to\infty} T^{-1} \log R_T(m)_{ij}$$

are well defined deterministic quantities, in the sense that the limit exists almost surely, is almost-surely constant, and is the same for any $0 \le i, j \le d-1$.

The quantity we are trying to approximate and provide the bounds for is

$$\boxed{a(m) - a(0) = \lim_{T\to\infty} T^{-1}\Big(\log R_T(0,0) - \sum_{t=1}^{T} X_t^{(0)}\Big).} \tag{1}$$

1.2. **Sub-Gaussian variables:** Although we describe our results using Gaussian random variables, the results are more general and apply to sub-Gaussian random variables. The random variables $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ are assumed sub-Gaussian with variance factor τ_i. This is a bigger class of distributions which includes Gaussian distribution. Please note that in the main document, all simulations are done assuming Gaussian.

A random variable Z is said to be *sub-Gaussian* if it has finite *variance factor* (the term applied in [BLM13]) $\tau = \tau(Z)$, defined by

$$\tau := \inf\left\{ c \ge 0 \,:\, \log \mathbb{E}\left[\mathrm{e}^{\lambda(Z - \mathbb{E}[Z])}\right] \le \frac{c\lambda^2}{2} \,\forall \lambda \in \mathbb{R} \right\}. \tag{2}$$

($\sqrt{\tau(Z)}$ is called the *sub-Gaussian standard* in [BK00].)

Biological Intuition: This may be thought of as an upper bound on the scale of the tails, and it is this that determines the upper bound on the sensitivity of the stochastic growth rate to migration.

Note that the Legendre transform for the moment generating function $I_Z(z) := \sup_\lambda\{\lambda z - \log \mathbb{E}[\mathrm{e}^{\lambda(Z-\mathbb{E}[Z])}]\}$ satisfies $I_Z(z) \geq z^2/2\tau(Z)$ for all z. Applying Markov's inequality implies the standard sub-Gaussian concentration inequality

$$\mathbb{P}\big\{Z \geq \mathbb{E}[Z] + z\big\} \leq \mathrm{e}^{-z^2/2\tau(Z)}. \tag{3}$$

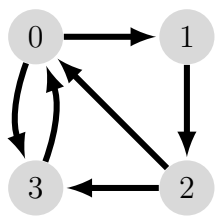
To describe the lower bounds we define a lower subvariance $\underline{\tau}(Z)$ by

$$\underline{\tau}(Z) := \sup\left\{c \geq 0 \,:\, \log \mathbb{E}\left[\mathrm{e}^{\lambda(Z-\mathbb{E}[Z])}\right] \geq \frac{c^2\lambda^2}{2}\,\forall \lambda \in \mathbb{R}\right\}. \tag{4}$$

Then for all z we have $I_Z(z) \leq z^2/2\tau(Z)$. We define also

$$\rho_j := \frac{\tilde{\mu}_j}{\tau_j} \qquad \text{and} \qquad \underline{\rho}_j := \frac{\tilde{\mu}_j}{\underline{\tau}_j}. \tag{5}$$

If Z is Gaussian with variance σ^2 then of course $\tau(Z) = \underline{\tau}(Z) = \sigma$. The calculation of ρ is illustrated in Figure 3.



(a) The migration graph $\mathcal{M}$

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1 & -0.5 & 0.7 & 0.5 \\ -0.5 & 1 & 0 & 0 \\ 0.7 & 0 & 1 & -0.3 \\ 0.5 & 0 & -0.3 & 2 \end{pmatrix} \end{array}$$

(b) $\mathrm{Cov}(X_t^{(i)}, X_t^{(j)})$

j	$\mathbb{E}[X_t^{(j)}]$	τ_j	ρ_j	κ_j	$\kappa_j\rho_j$	$\frac{\kappa_j\rho_j}{1+2\rho_j}$
0	1					
1	0.4	3	0.2	3	0.6	0.43
2	0.7	0.6	0.5	3	1.5	0.75
3	0.3	2	0.35	2	0.7	0.41

(c) Selecting the second site

FIGURE 3. Calculating ρ on a four-site graph. For simplicity, we assume the variables X_t are all Gaussian, so that $\tau^2 = \underline{\tau}^2$ is the variance, and $\rho = \underline{\rho}$. We see that $\kappa\rho$ is minimized at site 1, despite the fact that it is not in the shortest cycle, nor does it have the smallest mean difference in log growth rate from the optimal site 0, yielding a lower bound on the sensitivity to migration on the order of $m^{0.6}$. The upper bound $m^{0.41}$ comes from migration to site 3, which has lower mean log growth rate, but compensates for this with a shorter cycle length.

Like variances, sub-Gaussian variance factors are sub-additive for sums of independent random variables, as stated in [BK00, Lemma 1.7]: If $Z_1, \ldots, Z_k$ are independent random sub-Gaussian random variables with variance factors $\tau_1, \ldots, \tau_k$. Then $Z_1 + \cdots + Z_k$ is sub-Gaussian with variance factor less than or equal to $\tau_1 + \cdots + \tau_k$.

2. Theorem

In the Appendix, we present the general theorem for any number of sites and the two-site as a special case after the describing the theorem. For simplicity, the theorems for the case of two sites and more than two sites are presented separately in the main manuscript. This is because when there are more than two sites ($d > 2$), the result is complicated for two reasons: First, the effect of migration will be dominated by the site that has the second fastest growth; second, the increment to growth will be smaller if direct transition between the two leading sites is impossible. For this purpose, for each $1 \leq j \leq d-1$ we define κ_j to be the smallest length of a cycle in $\mathcal{M}$ that starts and ends at 0, and passes through j. (Thus $2d-2 \geq \kappa_j \geq 2$, and κ_j is equal to 2 when there are direct paths from 0 to j and from j to 0.)

Theorem 1. *Under the assumptions of section 1, let* $\underline{\eta} = \min_j 2\kappa_j \underline{\rho}_j$, *and* $\bar{\eta} = \min_j 2\kappa_j \rho_j/(1+2\rho_j)$. *Then for any* $\delta > 0$*:*

$$m^{\underline{\eta}+\delta} - (\mathbb{E}\left[-\mathbf{A}_t(0,0)\right] + \delta)m \leq a(m) - a(0) \leq m^{\bar{\eta}-\delta} - \mathbb{E}\left[-\mathbf{A}_t(0,0)\right] m \tag{6}$$

for $m > 0$ *sufficiently small.*

Note that the upper bound is dominated by the lower power of m, and the lower bound is dominated by the higher power of m and κ_j is the smallest length of a cycle in $\mathcal{M}$ that starts and ends at 0, and passes through j. From the theorem, it is clear that to increase $a(m) - a(0)$, $\kappa_j \underline{\rho}_j$ must be minimized.

Since the dispersal matrix is not constant the linear term has a coefficient that is the expectation of $\mathbf{A}_t(0,0)$. In the main manuscript, this is just -1. (It has been written as $-\mathbb{E}\left[-\mathbf{A}_t(0,0)\right]$ to make clear that it is generally understood to be negative.

The important point is that the effect of the diagonal terms in $\mathbf{A}_t$ is always nearly linear in m, while the increase of a near 0 is often superlinear, growing as a power of m smaller than 1. If the power $\rho_j = \tilde{\mu}_j/\underline{\tau}_j$ is strictly less than 1, the rapid increase in a near 0 will be qualitatively unaffected by a linear term for m sufficiently small. Even when the linear term is negative (as we will generally be assuming it to be), the growth rate a will still be increasing on a small interval of $m > 0$.

The cost of migration is governed by the row sums of matrix $\mathbf{A}_t$. Since individuals who migrate are lost from the originating site, it would be reasonable to require that the total of all migration from a site $\sum_i A_t(j,i)$ is no more than 0; if individuals are lost in the process of migration then the sum will be negative. For the section on migration penalty in the main manuscript, we make the $\sum_i A_t(j,i)$ negative to account for loss of individuals.

2.1. **Case when there are 2 sites.** This is a special case for the Theorem above since $\kappa = 2$ when there are only two sites. In this case, the difference between stochastic growth rates, $a(m)$ and $a(0)$ is bounded below and above (scaled by a constant) as follows: For any positive δ,

$$m^{4\tilde{\mu}/\underline{\tau}+\delta} - (\mathbb{E}\left[-\mathbf{A}_t(0,0)\right] + \delta)m \leq a(m) - a(0) \leq m^{\left(\frac{4\tilde{\mu}/\tau}{2\tilde{\mu}/\tau+1}\right)-\delta} - \mathbb{E}\left[-\mathbf{A}_t(0,0)\right] m$$

for positive m sufficiently close to 0.

Now that the theorems have been stated, we will define the excursion decomposition that will be used to prove the bounds for $a(m) - a(0)$.

3. Excursion decompositions (path length)

Since we are assuming the unique maximum average growth rate is at site 0, the maximum growth for the perturbed process will arise from rare excursions away from 0; in particular, from those that include the (not necessarily unique) site that minimises $\rho\kappa$ in (5).

Define $\mathcal{E}$ to be the set — called the *excursions* — of cycles in the migration graph that start and end at 0, with no intervening returns to 0. For an excursion $\mathbf{e}$ we write $|\mathbf{e}|$ for the length of the cycle minus 2 — that is, the number of time steps spent away from 0.

As is standard in statistical mechanics, the change in stochastic growth rate will ultimately be determined by the tradeoff between the penalty for migration and for spending time away from the optimal site, and the entropy gain from increasing the number of possible paths as the number of allowed transitions grows. To calculate this tradeoff quantitatively we need a notation that will allow us to enumerate the paths — defined by a sequence of excursions — that will draw similar penalties. Such a sequence may be thought of as an *ancestry*, as it records the locations of the ancestors of a generation T individual in each of the preceding generations. For purposes of enumeration the essential features will be the number of excursions in the ancestry, total length of the excursions, and the number of transitions between sites.

For a given excursion $\mathbf{e}$, comprising sites $0 = \mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{|\mathbf{e}|-1}, \mathbf{e}_{|\mathbf{e}|}, \mathbf{e}_{|\mathbf{e}|+1} = 0$, we define

$$\Xi(\mathbf{e}) := \left\{0 \leq t \leq |\mathbf{e}| + 1 \,:\, \mathbf{e}_t \neq \mathbf{e}_{t+1}\right\}$$

Biological Interpretation: $\Xi(\mathbf{e})$ denotes the set of change points in an excursion, i.e, it collects the times at which a transition is made, meaning $e_t \neq e_{t+1}$. The total number of transitions made from one site to the next, i.e, $\#\Xi(\mathbf{e})$ is denoted by s.

$$\kappa(\mathbf{e}) := \max\left\{\kappa_j \,:\, j \in \mathbf{e}\right\};$$

Biological Interpretation: Here κ_j=min {no. of edges of $P|P$ is a closed path passing through 0 and j}. That is, κ_j is the length of the smallest cycle that passes through vertex j and starts and ends at 0. The smallest number of steps possible in a cycle containing j and 0 can not be strictly greater than the actual number of transitions in an excursion passing through j, which defines such a cycle. Therefore, $\kappa(\mathbf{e}) \leq \#\Xi(\mathbf{e}) = s$ since $\kappa(\mathbf{e})$ is the maximum of all κ_j for $j \in \mathbf{e}$. We call $\kappa(\mathbf{e})$ the diameter of $\hat{e}$.

$$\rho(\mathbf{e}) := \min\left\{\rho_j \,:\, j \in \mathbf{e}\right\}.$$

Biological Interpretation: This is the loss in expected log growth at the site that might be thought of as the "target" of the excursion — the site where the loss is the smallest, scaled by the variance, which turns out to be the appropriate combination for measuring the influence of the fluctuations at the non-optimal site j, given that an excursion reaches that site.

We write $\hat{\mathcal{E}}_T$ for the collection of ancestries — sequences of excursions — that can be fit into time $\{1, \ldots, T\}$. That is, an element $\hat{\mathbf{e}} \in \hat{\mathcal{E}}_T$ has an *excursion count* $k(\hat{\mathbf{e}})$, such that each $i \in \{1, \ldots, k(\hat{\mathbf{e}})\}$ there is a pair $(t_i, \hat{\mathbf{e}}_i)$ with $t_i \in \{2, \ldots, T-1\}$ and $\hat{\mathbf{e}}_i \in \mathcal{E}$ satisfying

$$t_i + |\hat{\mathbf{e}}_i| < t_{i+1},$$
$$t_{k(\hat{\mathbf{e}})} + |\hat{\mathbf{e}}_{k(\hat{\mathbf{e}})}| \leq T.$$

t_i may be understood as the starting time of excursion i, which then continues for $|\hat{\mathbf{e}}_i|$, and so is definitely back at 0 before the next excursion starts at time t_{i+1}. We write the total length of an ancestry as $\|\hat{\mathbf{e}}\| := \sum_{i=1}^{k(\hat{\mathbf{e}})} |\hat{\mathbf{e}}_i|$. We also write $\hat{\mathcal{E}}_{T;k,n,s}$ for the subset of $\hat{\mathcal{E}}_T$ comprising ancestries whose excursion count is k, whose total length is n, and the sum of whose change-point counts $\#\Xi(\hat{\mathbf{e}}_i)$ is s. That is, $s = \#\Xi(\hat{\mathbf{e}})$, where $\Xi(\hat{\mathbf{e}}) := \bigcup_{i=1}^{k}\{t_i + t \,:\, t \in \Xi(\hat{\mathbf{e}}_i)\}$. The *null ancestry* is the element of $\hat{\mathcal{E}}_T$ with $k(\hat{\mathbf{e}}) = \|\hat{\mathbf{e}}\| = 0$.

The $(0,0)$ entry of the product R_T will be a sum of terms that are enumerated by elements of $\hat{\mathcal{E}}_T$, corresponding to paths through the sites. We define new random variables as a function of the realizations of $\mathbf{X}$ and of $\mathbf{A}$ (the collection of all matrices A)

$$\alpha_t(i,j) := \widetilde{X}_t^{(j)} + \log\frac{1 + mA_t(j,j)}{1 + mA_t(0,0)} + \begin{cases} \log m + \log A_t(j,i) & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases} \tag{7}$$

Given an excursion $\mathbf{e}$ and a starting time $t_0 \in \{0, \ldots, T - |\mathbf{e}| - 1\}$ we define the random variables

$$\mathbf{e}[t_0; \mathbf{X}, \mathbf{A}] := \sum_{0 \leq t \leq |\mathbf{e}|} \alpha_{t+t_0}(\mathbf{e}_t, \mathbf{e}_{t+1}) \tag{8}$$

Of course, this sum may be $-\infty$, if it includes a transition at which the corresponding entry of A is 0. But our assumptions imply that it is finite with nonzero probability if $\mathbf{e} \in \mathcal{E}$. Given an ancestry $\hat{\mathbf{e}} = \big((t_i, \hat{\mathbf{e}}_i)\big)_{i=1}^{k} \in \hat{\mathcal{E}}_T$, we define

$$\hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] := \sum_{i=1}^{k} \hat{\mathbf{e}}_i[t_i; \mathbf{X}, \mathbf{A}]. \tag{9}$$

To find the bounds for equation (1), we will use Lemma 2 to find an upper bound for $\log R_T(0,0)$. The proof invokes summing over ancestries.

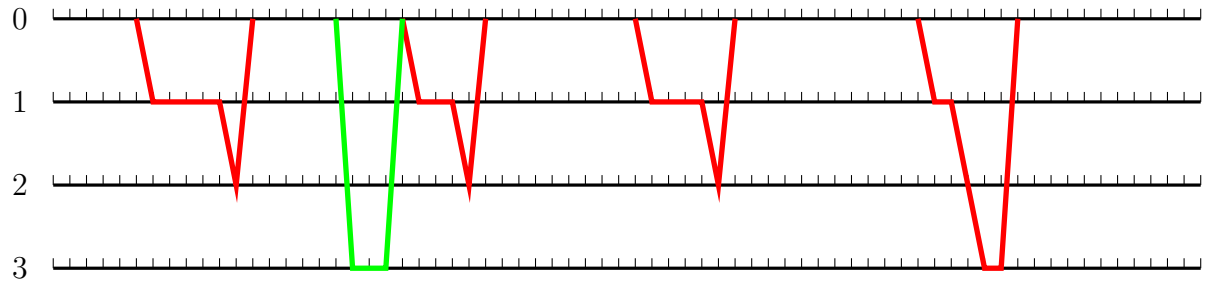


FIGURE 4. An ancestry for $T = 70$ comprising $k = 5$ excursions. This is based on the migration graph example from Figure 3. Three excursions (red) have diameter 3, and one (green) has diameter 2. Note that one timepoint ($t = 21$) is included in two different excursions. The red excursions all have $\rho(\hat{\mathbf{e}}) = 0.2$, and the green excursion has $\rho(\hat{\mathbf{e}}) = 0.35$. The lengths are 6, 3, 4, 5, 5, giving the sequence a total length $n = 23$. The change-point counts are 3, 2, 3, 3, 4, summing to $s = 15$.

Lemma 2.

$$\log R_T(0,0) = \sum_{t=1}^{T} X_t^{(0)} + m \sum_{t=1}^{T} \log\left(1 + mA_t(0,0)\right) + \log\Big(1 + \sum_{\hat{\mathbf{e}} \in \hat{\mathcal{E}}_T \setminus \{\hat{\mathbf{e}}^0\}} \mathrm{e}^{\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]}\Big), \tag{10}$$

where $\hat{\mathbf{e}}^0$ *is the null ancestry.*

Proof. We have, by definition,

$$R_T(0,0) = \sum_{(x_0,\ldots,x_T)} \prod_{t=1}^{T} D_t(m)\left(x_t, x_{t-1}\right)$$

where the summation is over $(x_0, \ldots, x_T) \in \{0, \ldots, d-1\}^{T+1}$ with $x_0 = x_T = 0$. We may restrict the summation to $(T+1)$ -tuples such that $D_t(m)\left(x_t, x_{t-1}\right) > 0$, which will only be true when (x_{t-1}, x_t) is an edge of $\mathcal{M}$. Such sequences of states map one-to-one onto ancestries.

By the definition of matrix products,

$$R_T(0,0) = \sum_{k=0}^{T} \sum_{\hat{e} \in \hat{\mathcal{E}}_{T,k}} \prod_{i=1}^{k+1} \left(\prod_{t=t_{i-1}+|\hat{e}_{i-1}|+1}^{t_i} D_t(m)(0,0) \right) \cdot \prod_{i=1}^{k} \left(\prod_{t=1}^{|\hat{e}_i|} D_{t+t_i}(m) \left((\hat{\mathbf{e}}_i)_{t+1}, (\hat{\mathbf{e}}_i)_t \right) \right),$$

where $t_0 = 0$ and $t_{k+1} = T$. The logarithm of the summand in the above expression is

$$\begin{aligned} &\sum_{i=1}^{k+1} \sum_{t=t_{i-1}+|e_{i-1}|+1}^{t_i} \log D_t(m)(0,0) + \sum_{i=1}^{k} \sum_{t=1}^{|e_i|} \log \frac{D_{t+t_i}(m)\left((\hat{\mathbf{e}}_i)_{t+1}, (\hat{\mathbf{e}}_i)_t\right) D_t(m)(0,0)}{D_t(m)(0,0)} \\ &\quad = \sum_{t=0}^{T} D_t(m)(0,0) + \sum_{i=1}^{k} \sum_{t=1}^{|e_i|} \log \frac{D_{t+t_i}(m)\left((\hat{\mathbf{e}}_i)_{t+1}, (\hat{\mathbf{e}}_i)_t\right)}{D_t(m)(0,0)}. \end{aligned}$$

We have $D_t(m)(0,0) = \mathrm{e}^{X_t^{(0)}}(1 + mA_t(0,0))$. Taking the logarithm of the expression, we get for the expression above

$$\sum_{t=1}^{T}\left(X_t^{(0)} + \log(1 + mA_t(0,0))\right) - \sum_{i=1}^{k}\sum_{t=1}^{|\hat{\mathrm{e}}_i|}\log\frac{D_{t+t_i}(m)\left((\hat{\mathbf{e}}_i)_t\,,(\hat{\mathbf{e}}_i)_{t+1}\right)}{D_{t+t_i}(m)(0,0)}$$
$$= \sum_{t=0}^{T}\left(X_t^{(0)} + \log\big(1 + mA_t(0,0)\big)\right) + \sum_{i=1}^{k}\sum_{t=1}^{|\hat{\mathrm{e}}_i|}\alpha_{t+t_i}(m)\left((\hat{\mathbf{e}}_i)_t\,,(\hat{\mathbf{e}}_i)_{t+1}\right),$$

since

$$\log\frac{D_t(m)(i,j)}{D_t(m)(0,0)} = \alpha_t(i,j) := \widetilde{X}_t^{(\mathbf{e}_t)} + \log\frac{1 + mA_t(\mathbf{e}_t,\mathbf{e}_t)}{1 + mA_t(0,0)} + \begin{cases}\log m + \log A_t(j,i) & \text{if } i \neq j,\\ 0 & \text{if } i = j.\end{cases}$$

The result then follows from splitting off the null excursion and invoking the definition of the random variable $\hat{\mathbf{e}}[X,A]$. □

We have then

$$\begin{aligned}
&\log R_T(0,0) - \sum_{t=1}^{T}\log\big(1 + mA_t(0,0)\big) - \sum_{t=1}^{T}X_t^{(0)}\\
&= \log\left(1 + \sum_{\hat{\mathrm{e}}\in\hat{\mathcal{E}}_T\setminus\{\hat{\mathrm{e}}^0\}}\mathrm{e}^{\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right)\\
&= \log\left(\sum_{\hat{\mathrm{e}}\in\hat{\mathcal{E}}_T}\mathrm{e}^{\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right)\\
&= \log\left(\sum_{1\le k,n,s\le T}\sum_{\hat{\mathrm{e}}\in\hat{\mathcal{E}}_{T;k,n,s}}\mathrm{e}^{\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right)\\
&\le \log\left(\sum_{1\le k,n,s\le T}\#\hat{\mathcal{E}}_{T;k,n,s}\mathrm{e}^{\max_{\hat{\mathrm{e}}\in\hat{\mathcal{E}}_T}\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right)\\
&\le \log T^3\max_{1\le k,n,s\le T}\#\hat{\mathcal{E}}_{T;k,n,s}\mathrm{e}^{\max_{\hat{\mathrm{e}}\in\hat{\mathcal{E}}_T}\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\\
&= \log T^3 + \max_{1\le k,n,s\le T}\log\left(\#\hat{\mathcal{E}}_{T;k,n,s}\mathrm{e}^{\max_{\hat{\mathrm{e}}\in\hat{\mathcal{E}}_T}\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right)\\
&= 3\log T + \max_{1\le k,n,s\le T}\left(\log(\#\hat{\mathcal{E}}_{T;k,n,s}) + \max_{\hat{\mathrm{e}}\in\hat{\mathcal{E}}_T}\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]\right).
\end{aligned}\tag{11}$$

By the Strong Law of Large Numbers,

$$\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\log\big(1 + mA_t(0,0)\big) = \mathbb{E}[\log\big(1 + mA_t(0,0)\big)] \text{ almost surely.}$$

As we have assumed $A_t(0,0)$ to be almost-surely bounded, we have for any positive δ

$$m\mathbb{E}[A_t(0,0)] - \delta \leq \mathbb{E}[\log\big(1 + mA_t(0,0)\big)] \leq m\mathbb{E}[A_t(0,0)],$$

for m sufficiently small.

Since

$$a(m) - a(0) = \lim_{T\to\infty} T^{-1}\log R_T(0,0) - T^{-1}\sum_{t=1}^{T} X_t^{(0)},$$

we have the bounds

$$\boxed{a(m) - a(0) \geq \liminf_{T\to\infty} T^{-1} \max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_T} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + m\left(\mathbb{E}[A_t(0,0)] - \delta\right),} \tag{12}$$

and

$$\boxed{a(m) - a(0) \leq \limsup_{T\to\infty} T^{-1} \max_{1\leq k,n,s\leq T}\Big(\log \#\hat{\mathcal{E}}_{T;k,n,s} + \max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_{T;k,n,s}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\Big) + m\mathbb{E}[A_t(0,0)].} \tag{13}$$

4. Derivation of the upper bound

We prove the upper bound in (6). We may replace $A_t(i,j)$ by $A_t(i,j) \vee 1$ for any $(j,i) \in \mathcal{M}$, since decreasing A_t can only decrease $a(m) - a(0)$. That is, we put a floor under those off-diagonal elements which are allowable migrations. This avoids the nuisance of having entries be sometimes 0 , and an upper bound that holds under these conditions will hold *a fortiori* under the original conditions. Indeed, we may assume without loss of generality that all $A_t(j,i) = 1$ identically for $i \neq j$, since $a(m)$ — the stochastic growth rate with the true values of A_t — is no larger than $a(A,m;1)$, the stochastic growth rate where all values of A_t that are permitted to be nonzero are replaced by 1. This changes our upper bound only by a constant, which may be absorbed into the constant of the theorem. Thus, we will proceed under this assumption.

An element of $\hat{\mathcal{E}}_{T;k,n,s}$ may be determined by the following choices:

(i) Choose k points out of T where the excursions begin, yielding no more than $\binom{T}{k}$ possibilities;

(ii) Choose k numbers for the lengths of the excursions that add up to n, yielding no more than $\binom{n}{k}$ possibilities;

(iii) Choose $s - 2k$ time points within these excursions as times when there is a change of site, yielding at most $\binom{n}{s-2k}$ possibilities;

(iv) There are no more than d^s ways to choose the sites to which the excursions move at the s times when there is a change.

Then using the points above with the a crude bound based on Stirling's Formula (derived in section 6.1), we get:

$$\log \#\hat{\mathcal{E}}_{T;k,n,s} \leq s \log d + (s-2k) \log \frac{n}{s-2k} + k \log \frac{n}{k} + T \log \frac{T}{k}. \tag{14}$$

Claim 3. *Suppose that ρ_j and κ_j are each minimised at site $j = 1$. For any $\delta > 0$ and*

$$z = m^{\bar{\eta}} \cdot (\log m^{-1})^{\kappa_1 + 1},$$

we have

$$\limsup_{T \to \infty} T^{-1} \log \mathbb{P}\Big\{ \max_{\hat{\mathbf{e}} \in \hat{\mathcal{E}}_{T;k,n,s}} \big(\hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] + \log \#\hat{\mathcal{E}}_{T;k,n,s} \big) \geq zT \Big\} < 0 \tag{15}$$

for $m > 0$ sufficiently small, meaning $m \in (0, m_0)$ where m_0 is a constant, depending only on the model parameters and δ.

The proof of Claim 3 is provided at the end of the section for Derivation of the Lower Bound 5. Assuming the Claim,

$$\sum_{T=T_0}^{\infty} \mathbb{P} \left\{ T^{-1} \max_{1 \leq k,n,s \leq T} \Big(\log \#\hat{\mathcal{E}}_{T;k,n,s} + \max_{\hat{\mathbf{e}} \in \hat{\mathcal{E}}_{T;k,n,s}} \hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] \Big) \geq z \right\} < \infty.$$

By the Borel–Cantelli Lemma, this implies that with probability 1 this event occurs only finitely often. It follows that the limsup is no greater than z almost surely, and hence, by (13), that

$$a(m) - a(0) \leq m^{\bar{\eta}} \cdot (\log m^{-1})^{\kappa_1 + 1}. \tag{16}$$

It remains only to clear away the assumption that that κ_j and ρ_j are both minimized at site 1. We do this by stratifying the excursions further by their diameter (recall the definition from section 3). Define

$$\breve{\rho}(\kappa) := \min\big\{ \rho_j \,:\, \kappa_j \leq \kappa \big\}.$$

If $\mathbf{e}$ is an excursion with diameter κ, then any site j included in $\mathbf{e}$ has $\kappa_j \leq \kappa$, hence also $\rho_j \geq \breve{\rho}(\kappa)$. Furthermore,

$$\min_{1 \leq j \leq d-1} 2\rho_j \kappa_j / (1 + 2\rho_j) = \min_{2 \leq \kappa \leq d-1} 2\breve{\rho}(\kappa)\kappa / (1 + 2\breve{\rho}(\kappa)) \tag{17}$$

The maximum in (13) may be written as a maximum over $(k_2, \ldots, k_{d-1})$, representing the number of excursions whose diameter is $2, 3, \ldots, d-1$, with the constraint $\sum k_\kappa = k$. We write $\hat{\mathcal{E}}^{(\kappa)}_{T;k,n,s}$ for the ancestries consisting of k excursions, all of which have diameter κ; and $\hat{\mathcal{E}}_{T;(k_\kappa),n,s}$ for the set of ancestries that have exactly k_κ excursions with diameter κ. Then $\hat{\mathcal{E}}_{T;(k_\kappa),n,s}$ is naturally included in the direct sum of $\hat{\mathcal{E}}^{(\kappa)}_{T;k_\kappa,n,s}$. (A sequence of mixed diameters $\hat{\mathbf{e}} \in \hat{\mathcal{E}}_{T;(k_\kappa),n,s}$ may be decomposed into sequences $\hat{\mathbf{e}}_{(\kappa)} \in \hat{\mathcal{E}}^{(\kappa)}_{T;k_\kappa,n,s}$ of excursions with each particular diameter. Referring back to the example in Figure 4,

this would entail making one ancestry by dropping out the green excursions, and a separate one by dropping out the red excursions.) Thus

$$\begin{aligned}\max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_{T;k,n,s}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] &= \max_{\sum k_\kappa = k} \max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_{T;(k_\kappa),n,s}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] \\ &\le \max_{\sum k_\kappa = k} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathcal{E}}^{(\kappa)}_{T;k_\kappa,n,s}} \sum_\kappa \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}] \\ &\le \sum_\kappa \max_{1\le k_\kappa\le T} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathcal{E}}^{(\kappa)}_{T;k_\kappa,n,s}} \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}],\end{aligned}$$

using the general fact that the maximum of a sum is smaller than the sum of maxima. Thus we have

$$\max_{1\le k,n,s\le T}\Big(\log \#\hat{\mathcal{E}}_{T;k,n,s} + \max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_{T;k,n,s}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\Big) \le \sum_\kappa \max_{1\le k_\kappa,n_\kappa,s_\kappa\le T} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathcal{E}}^{(\kappa)}_{T;k_\kappa,n_\kappa,s_\kappa}} \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}].$$

Here n_κ and s_κ simply index the choice of total length and number of change-points among all of the excursions of length κ.

Because all excursions in $\hat{\mathcal{E}}^{(\kappa)}_{T;k_\kappa,n_\kappa,s_\kappa}$ pass through only sites j with $\rho_j \ge \breve{\rho}(\kappa)$, the same argument used for the upper bound in (16) may be applied to show that almost surely for $m > 0$ sufficiently small

$$\limsup_{T\to\infty} T^{-1} \max_{1\le k_\kappa,n_\kappa,s_\kappa\le T} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathcal{E}}^{(\kappa)}_{T;k_\kappa,n_\kappa,s_\kappa}} \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}] \le (\log m^{-1})^{1+\delta} m^{2\kappa\breve{\rho}(\kappa)/(1+2\breve{\rho}(\kappa))}.$$

It follows that

$$\begin{aligned}\limsup_{T\to\infty} T^{-1} \max_{1\le k,n,s\le T}\Big(\log \#\hat{\mathcal{E}}_{T;k,n,s} + \max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_{T;k,n,s}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\Big) \\ &\le \sum_\kappa (\log m^{-1})^{1+\delta} m^{2\kappa\breve{\rho}(\kappa)/(1+2\breve{\rho}(\kappa))} \\ &\le (2d-2)(\log m^{-1})^{1+\delta} m^{\min_{1\le j\le d-1} 2\kappa_j\rho_j/(1+2\rho_j)}\end{aligned}$$

by (17), which completes the proof.

5. Derivation of the lower bound

We show that the lower bound applies for each j; it will then hold in particular for the j at which $\kappa_j\rho_j$ attains its minimum. We may assume without loss of generality that this "second-best" site is $j = 1$, and we will write simply κ, $\tilde{\mu}$, and $\underline{\rho}$ for κ_1, $\tilde{\mu}_1$, and $\underline{\rho}_1$. We assume that the row sums are identically 0, hence there is no cost to migration.

Let $0 = j_0, j_1, j_2, \ldots, j_I = 1, j_{I+1}, \ldots, j_{\kappa-1}, j_\kappa = 0$, be a cycle from 0 in $\mathcal{M}$, passing through 1. We may fix a real number A_* and $p > 0$ such that

$$\mathbb{P}\Big\{\sum_{i=0}^{\kappa-1}\big(\widetilde{X}^{(j_i)}_{t+i} + \log A_{t+i}(j_{i+1},j_i)\big) > -\kappa A_*\Big\} \ge p. \tag{18}$$

Assume that $m \le \mathrm{e}^{-1}$ and $T > \log m^{-1}$. Defining $k = \lfloor T/m\rfloor$ and $s = \lfloor \kappa \log m^{-1}/\tilde{\mu}\rfloor + \kappa$, we will apply (12) by considering only excursions of length exactly $s-1$, which proceed exactly through the sequence of sites $0 = j_0, j_1, \ldots, j_{I-1}, 1, \ldots, 1, j_{I+1}, \ldots, j_{\kappa-1}, j_\kappa = 0$,

where site 1 is repeated exactly $s-\kappa+1$ times. The idea is that the excursion fills a time block of length s, proceeding as quickly as possible from 0 to 1, remaining $s-\kappa+1$ time units at 1, and then returning as quickly as possible to 0.

We define the standard excursion $\mathbf{e}_\circ := (j_1, \ldots, j_{I-1}, 1, \ldots, 1, j_{I+1}, \ldots, j_{\kappa-1})$, with $s-\kappa+1$ repetitions of site 1; and an ancestry $\hat{\mathbf{e}}_\circ$ consisting of those pairs $(\ell s+1, \mathbf{e}_\ell)$ for which

$$Y_\ell := \mathbf{e}_\circ\big[\ell s+1; \mathbf{X}, \mathbf{A}\big] > 0. \tag{19}$$

That is, $\hat{\mathbf{e}}_\circ$ is assembled from identical excursions of form $\mathbf{e}_\circ$ which can start only at times $\ell s+1$. Each one of the k possible excursions is included precisely when its contribution to the sum would be positive.

We have

$$\hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] = \sum_{\ell=0}^{k-1} (Y_\ell)_+ \,.$$

Since the excursion contributions Y_ℓ are all independent, combining (12) with the Strong Law of Large Numbers yields

$$a(m) - a(0) \geq \frac{\mathbb{E}[(Y_\ell)_+]}{s}. \tag{20}$$

We now observe that for any ℓ

$$Y_\ell = \sum_{i=0}^{I-1} \alpha_{\ell s+i}(j_i, j_{i+1}) + \sum_{i=I}^{\kappa-1} \alpha_{(\ell+1)s-\kappa+i}(j_i, j_{i+1}) \; + \sum_{t=\ell s+I}^{(\ell+1)s-\kappa+I-1} \widetilde{X}_t^{(1)} + m\left(A_t(1,1) - A_t(0,0)\right).$$

Note that the α_t terms are independent of the $\widetilde{X}_t^{(1)}$ terms, because they draw their component random variables from different times.

By (18) and the definition of s, which makes $\kappa \log m^{-1} \leq \tilde{\mu}(s-\kappa+1)$, for any $y > 0$

$$\begin{aligned}\mathbb{P}\big\{Y_\ell > (s-\kappa+1)y\big\} &\geq p\mathbb{P}\Big\{(s-\kappa+1)Z > (s-\kappa+1)(y+\tilde{\mu}) + \kappa\big(\log m^{-1} + A_*\big)\Big\} \\ &\geq p\mathbb{P}\Big\{Z > \big(y+\tilde{\mu}\big) + \frac{\kappa}{s-\kappa+1} \cdot \frac{A_*}{\log m^{-1}}\Big\},\end{aligned}$$

where

$$Z := (s-\kappa+1)^{-1} \sum_{t=\ell s+I}^{(\ell+1)s-\kappa+I-1} (\widetilde{X}_t^{(1)} + \tilde{\mu}).$$

is the mean of independent mean-zero random variables whose moment generating functions have Legendre transform $I(z) \leq \frac{z^2}{2\underline{\tau}^2} = \frac{\rho z^2}{2\tilde{\mu}}$. By Cramér's Theorem [DZ09, Theorem 2.2.3]

$$\begin{aligned}\lim_{s\to\infty} (s-\kappa+1)^{-1} \log \mathbb{P}\Big\{Z > \big(y + \tilde{\mu}(2+\delta)\big)\Big\} &\geq -I\left(y + 2\tilde{\mu}\right) \\ &\geq -\frac{\rho}{2\tilde{\mu}}(y+2\tilde{\mu})^2.\end{aligned} \tag{21}$$

Holding y fixed, it follows that for any $\delta > 0$ we have for s sufficiently large (which is to say, for m sufficiently small)

$$\mathbb{P}\Big\{\frac{Y_\ell}{s} > y\Big\} \geq p \exp\left\{-\frac{s(y+2\tilde{\mu})^2\underline{\rho}}{2\tilde{\mu}} - s\frac{\tilde{\mu}\delta}{4\kappa}\right\}.$$

Taking m sufficiently small that $\delta \log m^{-1} \geq 1/2s\tilde{\mu}\underline{\rho}$ and $\log m^{-1} \leq m^{-\delta/2}$, it follows, by Lemma 4 (proved next), that

$$\begin{aligned}\mathbb{E}\big[(Y_\ell)_+\big] &\geq \frac{p}{s\underline{\rho}} \exp\left\{-\frac{s(2\tilde{\mu})^2\underline{\rho}}{2\tilde{\mu}} - \frac{\delta}{4}\log m^{-1} - \frac{1}{8s\tilde{\mu}\underline{\rho}}\right\} \\ &\geq \frac{p}{s\underline{\rho}} \exp\left\{-\left(2\kappa\underline{\rho} + \frac{\delta}{2}\right)\log m^{-1}\right\} \\ &\geq \frac{p\kappa}{\mu\underline{\rho}} m^{2\kappa\underline{\rho}+\delta}.\end{aligned}$$

Combining this with (9) and (20) completes the proof.

Lemma 4. *Let Z be a random variable with tail bound*

$$\mathbb{P}\big\{Z > z\big\} \geq \mathrm{e}^{-(z_0+z)^2/2\tau_Z} \text{ for all } z > 0.$$

Then

$$\mathbb{E}[Z_+] \geq \mathrm{e}^{-z_0^2/2\tau_Z} \cdot \mathrm{e}^{-\tau_Z/2z_0^2} \cdot \frac{\tau_Z}{\mathrm{e}z_0}.$$

Proof.

$$\begin{aligned}\mathbb{E}[Z_+] &= \int_0^\infty \mathbb{P}\big\{Z > z\big\}\,\mathrm{d}z \\ &\geq \mathrm{e}^{-z_0^2/2\tau_Z} \int_0^\infty \mathrm{e}^{-z(2z_0+z)/2\tau_Z}\,\mathrm{d}z \\ &\geq z_* \mathrm{e}^{-z_0^2/\tau_Z} \mathrm{e}^{-z_*(2z_0+z_*)/2\tau_Z}\end{aligned}$$

for any fixed $z_* > 0$, since the integrand is decreasing. Taking $z_* = \tau_Z/z_0$ gives us the result. □

The statements for general A follow then from the linear contribution of $m\sum A_t(0,0)$ to (20). The rest follows as in the conclusion of section 4 by considering the lower bound, applied to a modified $\widetilde{X}_t^{(j)}$, taken as $X_t^{(j)} - X_t^{(0)} + m\left(A_t(j,j) - A_t(0,0)\right)$, showing that the contribution of $A(j,j) - A(0,0)$ over all non-trivial excursions can shift the power $\underline{\eta}$ by an arbitrarily small amount, which may be incorporated into δ.

6. Proofs of claims

6.1. Derivation of Stirling's-Formula bound.

$$\log\binom{n}{j} \leq j + j\log\frac{n}{j}, \tag{22}$$

which holds for all positive integers n and $0 \leq j \leq n$, following the convention $0 \cdot \log 0 = 0 \cdot \log \infty = 0$. This is because: Since $j \geq 0$,

$$e^j = \sum_s \frac{j^s}{s!} > \frac{j^j}{j!}, \text{ so } (1/j!) < \frac{e^j}{j^j}.$$
$$\binom{n}{j} = \frac{n(n-1)\dots(n-j+1)}{j!} < \frac{n^j e^j}{j^j},$$
$$\log \binom{n}{j} < j + j\log n - j\log j = j + j\log\frac{n}{j}.$$

6.2. **Proof of Claim 3.**

Proof. We define

$$\zeta := \frac{k}{T}, \quad \nu := \frac{n}{k}, \quad \beta := \frac{s}{k}.$$

That is, ζ is the rate of excursions per unit time; ν is the average length of excursions; and β is the average number of migration events over the course of the excursions. We have the constraints $1/\zeta \geq \nu \geq \beta \geq \kappa_1 \geq 2$ (since κ_1 is the minimum κ_j, hence the minimum number of changes in each excursion). The bound (14) may be written as

$$\log \#\hat{\mathcal{E}}_{T;k,n,s} \leq \zeta T \left(\beta \log d + (\beta-2)\log\nu - (\beta-2)\log(\beta-2) - \log\zeta\right). \tag{23}$$

Suppose now we fix some element $\hat{\mathbf{e}}$ of $\hat{\mathcal{E}}_{T;k,n,s}$, and list all the states of all the excursions in order as $j_1, \dots, j_n$, and the associated times as $t_1 < t_2 < \dots < t_n$. Remember that here $j_i = 1$ for every i. Now, using the definition for an excursion defined above:

$$\begin{aligned}
\mathbb{E}\big[\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\big] &= \mathbb{E}\left(\sum_{i=1}^{k}\sum_{0\leq t\leq|\mathbf{e}_i|} \alpha_{t+t_i}\left(\mathbf{e}_{k,t},\mathbf{e}_{k,t+1}\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{k}\sum_{0\leq t\leq|\mathbf{e}_i|} \left(\widetilde{X}_t^{(\mathbf{e}_{k,t})}{}_{t+t_i} + \mathbf{1}_{\{\mathbf{e}_{k,t}\neq\mathbf{e}_{k,t+1}\}}\log m + (\log A_t(\mathbf{e}_{k,t},\mathbf{e}_{k,t+1}))_+\right)\right) \\
&\leq \mathbb{E}\Big[\sum_{i=1}^{k}\sum_{0\leq t\leq|\mathbf{e}_i|} \left(\widetilde{X}_{t+t_i}^{(\mathbf{e}_{k,t})} + \log m + \mu_A\right)\Big] \\
&\leq -\sum_{i=1}^{n}\tilde{\mu}^{(j_i)} + s(\mu_A + \log m)
\end{aligned}$$

Therefore we have

$$\mathbb{E}\big[\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\big] \leq s(\mu_A + \log m) - \sum_{i=1}^{n}\tilde{\mu}^{(j_i)}.$$

The random variable $Y := \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] - \mathbb{E}[\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]]$ is a sum of sub-Gaussian random variables, hence is also sub-Gaussian. The variance factor is bounded by the sum of

the variance factors of the summands, hence no bigger than $\sum_{i=1}^n \tau_{j_i} + s\tau_A$. We write

$$\tau_A := \max_{j\in\{0,\dots,d-1\}} \frac{\tilde\mu_j}{\tau_j} \max_{i\neq\ell} \tau\Big(\widetilde{X}_t^{(i)} + \big(\log A_t(\ell,i)\big)_+\Big). \tag{24}$$

This definition makes the variance factor of $\widetilde{X}_t^{(i)} + \log A_t(\ell,i)$ for any $i \neq 0$ will be less than or equal to τ_A/ρ_i. We are assuming that ρ_j is minimized at $j=1$, so $\rho_1 \le \tilde\mu_{j_i}/\tau_{j_i}$ and $\tau_{j_i} \le \tilde\mu_{j_i}/\rho_1$. We have

$$\begin{aligned}\sum_{i=1}^n \tau_{j_i} + s\frac{\tau_A}{\rho_1} &\leqslant \sum_{i=1}^n \frac{\tilde\mu_{j_i}}{\rho_1} + \frac{s\tau_A}{\rho_1}\\ &= \frac{1}{\rho_1}\sum_{i=1}^n \big(\tilde\mu_{j_i} + s\tau_A\big).\end{aligned} \tag{25}$$

By (3): $\mathbb{P}\big\{Z \ge \mathbb{E}[Z] + z\big\} \le \mathrm{e}^{-z^2/2\tau(Z)}$, for any $x, z > 0$,

$$\begin{aligned}\mathbb{P}\big\{\big(\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + x\big) \ge zT\big\} &\le \mathbb{P}\Big\{Y \ge \sum_{i=1}^n \tilde\mu_{j_i} - s\mu_A + s\log m^{-1} + zT - x\Big\}\\ &\le \exp\Bigg\{-\frac{1}{2}\left(\sum_{i=1}^n \tau_{j_i} + s\frac{\tau_A}{\rho_1}\right)^{-1}\left(\sum_{i=1}^n \tilde\mu_{j_i} - s\mu_A + s\log m^{-1} + zT - x\right)^2\Bigg\}.\end{aligned}$$

Taking $x = \log \#\hat{\mathcal{E}}_{T;k,n,s}$ and substituting (23) and (25), we get

$$\begin{aligned}\log\mathbb{P}\big(\max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_{T;k,n,s}} \big(\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + \log\#\hat{\mathcal{E}}_{T;k,n,s}\big) \ge zT\big)&\\ \le \log\#\hat{\mathcal{E}}_{T;k,n,s} + \max_{\hat{\mathbf{e}}\in\hat{\mathcal{E}}_{T;k,n,s}} \log\mathbb{P}\big\{\big(\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + \log\#\hat{\mathcal{E}}_{T;k,n,s}\big) \ge zT\big\}&\\ \le -\frac{\rho_1}{2}\cdot T \inf_{S\ge 2(\tau_A+\tilde\mu_*)} \inf_{u\ge z} \frac{z}{u}\Theta(S,\beta,u)&\end{aligned}$$

where

$$u := z/\zeta, \qquad \tilde\mu_* := \min_{1\le j\le d-1} \tilde\mu_j,$$

$$S := \beta\tau_A + k^{-1}\sum_{i=1}^n \tilde\mu^{(j_i)} \ge \beta(\tau_A + \tilde\mu_*) \ge \kappa_1(\tau_A + \tilde\mu_*), \text{ and}$$

$$\begin{aligned}\Theta &= \Theta(S,\beta,u)^2\\ &:= S^{-1}\left(S + \beta\log m^{-1} + (\beta-2)\log(\beta-2) - \beta\log\frac{d(S-\beta\tau_A)}{\tilde\mu_*} - \beta\tau_A + \log z - \log u + u\right)^2\\ &\qquad + \frac{2}{\rho_1}\big((\beta-2)\log(\beta-2) - \beta\log d(S-\beta\tau_A)\tilde\mu_* - \log u\big).\end{aligned}$$

Here we have taken advantage of the fact that $\log\zeta < 0$, and of the bound $\nu \le (S-\beta\tau_A)/\tilde\mu_*$.

We will now show that there is a positive constant m_0 — constant in the sense of being expressible in terms only of the model parameters ρ_1, κ_1, $\tilde\mu_*$, d, τ_A, and μ_A — such that

for any fixed $m \in (0, m_0)$ and

$$z = m^{2\kappa_1\rho_1/(1+2\rho_1)}(\log m^{-1})^{\kappa_1+1},$$

and any $S > 0$, $u > 0$, and $\beta \geq \kappa_1$,

$$\Theta(S, \beta, u) \geq u.$$

The result then follows immediately, since then for all T,

$$T^{-1} \log \mathbb{P}\big(\max_{\hat{\mathbf{e}} \in \hat{\mathcal{E}}_{T;k,n,s}} \big(\hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] + \log \#\hat{\mathcal{E}}_{T;k,n,s}\big) \geq zT\big) \leq -\frac{\rho_1 z}{2}$$

We consider separately the minimization of Θ for large and small values of S. Suppose first that $S > \kappa_1 \log m^{-1}$. We use the bound $S^{-1}(S+x)^2 \geq S + 2x$. We have then

$$\begin{aligned} \Theta \geq S + (\beta - 2)\left(\left(2 + \frac{2}{\rho_1}\right)\left(\log(\beta-2) - \log\frac{dS}{\tilde{\mu}_*}\right) - 2\tau_A + 2\log m^{-1}\right) \\ + 4\log m^{-1} - 4\tau_A - 4\log\frac{dS}{\tilde{\mu}_*} + \frac{2+2\rho_1}{\rho_1}(\log z - \log u) + 2u. \end{aligned} \tag{26}$$

(We have dropped $-\beta\tau_A$ from the log terms, as this would only increase the value of Θ.) By taking the partial derivative with respect to β, we see that this is minimized at

$$\log(\beta - 2) = \log\frac{dS}{\tilde{\mu}_*} + \frac{\rho}{\rho+1}(\tau_A + \log m) - 1.$$

(This value of β may be less than the smallest allowable value κ_1, but we may still use it for a lower bound on Θ.) For m sufficiently small $S + \log z > 0$, hence

$$\begin{aligned} \Theta &\geq S\left(1 - \frac{d}{\tilde{\mu}_*}\mathrm{e}^{\tau_A} m^{\rho_1/(\rho_1+1)}\right) - 2\log\frac{dS}{\tilde{\mu}_*} - 4\tau_A \\ &\quad + 4\log m^{-1} + \frac{2+2\rho_1}{\rho_1}(\log z - \log u) + 2u \\ &\geq 4\log m^{-1} + 2u - \frac{2+2\rho_1}{\rho_1}\log u \\ &\geq u \text{ for all } u > 0, \qquad \text{for } m \text{ sufficiently small.} \end{aligned}$$

Now suppose that $S \leq \kappa_1 \log m^{-1}$. We will use the general bound $S^{-1}(S+x)^2 \geq 4x$, which holds for any $S > 0$ and any y. We have now a slightly altered version of (26)

$$\begin{aligned} \Theta \geq (\beta - 2)\left(\left(4 + \frac{2}{\rho_1}\right)\left(\log(\beta-2) - \log\frac{dS}{\tilde{\mu}_*}\right) - 4\tau_A + 4\log m^{-1}\right) \\ + 8\log m^{-1} - 8\tau_A - 8\log\frac{dS}{\tilde{\mu}_*} + \frac{4+2\rho_1}{\rho_1}(\log z - \log u) + 4u. \end{aligned}$$

As before, the minimum of this expression with respect to β occurs at

$$\beta_{\min} = 2 + \mathrm{e}^{-1}\frac{dS}{\tilde{\mu}_*}\left(\exp(\tau_A)m\right)^{2\rho_1/(2\rho_1+1)},$$

and it is increasing on $(\beta_{\min}, \infty)$. Since we have assumed $S \leq \kappa_1 \log m^{-1}$, we may choose the constant m_0 such that this $\beta_{\min}$ is strictly less than 3. This means that the minimum will occur at $\beta = \beta_{\min}$ when $\kappa_1 = 2$, and at $\beta = \kappa_1$ otherwise.

Consider first the case $\kappa_1 = 2$, so that $\log z = \frac{4\rho_1}{1+2\rho_1} \log m$. Substituting $\beta = \beta_{\min}$ we have

$$\begin{aligned}
\Theta &\geq -\frac{\tilde{\mu}_*}{\mathrm{e}}^{\tau_A} m^{2\rho/(2\rho+1)} S + 8 \log m^{-1} - 8\tau_A + \frac{2+4\rho_1}{\rho_1} (\log z - \log u) + 4u \\
&\geq \Big(8 - \frac{2+4\rho_1}{\rho_1} \cdot \frac{4\rho_1}{1+2\rho_1}\Big) \log m^{-1} + 4u - \frac{2+2\rho_1}{\rho_1} \log u \\
&\qquad + 4 \log\log m^{-1} - \frac{d}{\tilde{\mu}_*} \kappa_1 \mathrm{e}^{\tau_A} m^{2\rho/(2\rho+1)} \log m^{-1} \\
&= u + \Big[\Big(3u - \frac{2+2\rho_1}{\rho_1} \log u\Big) \frac{d}{\tilde{\mu}_*} \kappa_1 \mathrm{e}^{\tau_A} m^{2\rho/(2\rho+1)} \log m^{-1} + 4 \log\log m^{-1}\Big] \\
&\geq u + \Big[\Big(\frac{2}{\rho_1} + 2\Big)\Big(1 - \log \frac{2+2\rho_1}{3\rho_1}\Big) - \frac{d}{\tilde{\mu}_*} \kappa_1 \mathrm{e}^{\tau_A} m^{2\rho/(2\rho+1)} \log m^{-1} + 4 \log\log m^{-1}\Big].
\end{aligned}$$

This expression in brackets depends only on model parameters — in particular, not on u — and is positive for m sufficiently small.

Finally, we consider the case $\kappa_1 \geq 3$. Substituting $\beta = \kappa_1$ and simplifying we obtain

$$\begin{aligned}
\Theta &\geq -\left(4 + \frac{2}{\rho_1}\right) \kappa_1 \log \frac{dS}{\tilde{\mu}_*} + 4\kappa_1 \log m^{-1} - 4\kappa_1 \tau_A + \frac{2+4\rho_1}{\rho_1} (\log z - \log u) + 4u \\
&\geq -C - \left(\frac{4\rho_1 + 2}{\rho_1}\right) \kappa_1 \log\log m^{-1} + \left(4\kappa_1 - \frac{4+2\rho_1}{\rho_1} \cdot \frac{2\kappa_1\rho_1}{1+2\rho_1}\right) \log m^{-1} \\
&\qquad + \frac{4\rho_1+2}{\rho_1} (\kappa_1 + 1) \log\log m^{-1} - \frac{4+2\rho_1}{\rho_1} \log u + 4u \qquad \text{where } C \text{ is a constant} \\
&\geq -C + \frac{4\rho_1+2}{\rho_1} \log\log m^{-1} - \frac{4+2\rho_1}{\rho_1} \log u + 4u.
\end{aligned}$$

As before, this will be $> u$ for all $u > 0$, when m is sufficiently small, which completes the proof. □

References

[BK00] V. V. Buldygin and Yu. V. Kozachenko. *Metric characterization of random variables and random processes*, volume 188 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2000. Translated from the 1998 Russian original by V. Zaiats.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[Coh79] Joel E. Cohen. Ergodic theorems in demography. *Bulletin of the American Mathematical Society*, 1:275–95, 1979.

[DZ09] A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*. Springer Verlag, 2nd edition, 2009.