# Population genetics meets ecology: a guide to individual-based simulations in continuous landscapes

Elizabeth T. Chevy[1*], Jiseon Min[2*], Victoria Caudill[2], Samuel E. Champer[3], Benjamin C. Haller[3], Clara T. Rehmann[2], Chris C. R. Smith[2], Silas Tittes[2], Philipp W. Messer[3], Andrew D. Kern[2,4], Sohini Ramachandran[1], and Peter L. Ralph[2,5]

[*]These authors contributed equally to the paper.
[1]Center for Computational Molecular Biology, Brown University, Providence RI 02912, USA
[2]Institute of Ecology and Evolution, University of Oregon, Eugene OR 97402, USA
[3]Department of Computational Biology, Cornell University, Ithaca NY 14853, USA
[4]Department of Biology, University of Oregon, Eugene OR 97402, USA
[5]Department of Mathematics, University of Oregon, Eugene OR 97402, USA

July 24, 2024

## Abstract

Individual-based simulation has become an increasingly crucial tool for many fields of population biology. However, implementing realistic and stable simulations in continuous space presents a variety of difficulties, from modeling choices to computational efficiency. This paper aims to be a practical guide to spatial simulation, helping researchers to implement realistic and efficient spatial, individual-based simulations and avoid common pitfalls. To do this, we delve into mechanisms of mating, reproduction, density-dependent feedback, and dispersal, all of which may vary across the landscape, discuss how these affect population dynamics, and describe how to parameterize simulations in convenient ways (for instance, to achieve a desired population density). We also demonstrate how to implement these models using the current version of the individual-based simulator, SLiM. Since SLiM has the capacity to simulate genomes, we also discuss natural selection – in particular, how genetic variation can affect demographic processes. Finally, we provide four short vignettes: simulations of pikas that shift their range up a mountain as temperatures rise; mosquitoes that live in rivers as juveniles and experience seasonally changing habitat; cane toads that expand across Australia, reaching 120 million individuals; and monarch butterflies whose populations are regulated by an explicitly modeled resource (milkweed).

## Introduction

Explicit spatial models are indispensable for understanding how species live, interact, and evolve across geographic landscapes. However, the process of formulating sensible models of demography in continuous space is fraught with pitfalls and choices that are unfamiliar to many researchers interested spatial modeling. For instance, one of the more commonly used non-spatial models is the Wright–Fisher model, in which the number of individuals is fixed – directly specified as a parameter. However, as we describe below, in spatial models with locally defined dynamics the number of individuals is a stochastic, emergent property of the dynamics and it takes some expertise to arrange a model that produces a desired equilibrium size. Even fundamental population genetics concepts such as selection coefficients cease to have a single obvious interpretation in a spatial context.

Implementation of an individual-based simulation requires great specificity – choices must be made regarding a large number of mechanisms and parameters. Simulated organisms must separately give birth and

die, unlike in more abstract theory such as the Wright–Fisher model, where often only the net effects of birth minus death enter (*e.g.*, Cantrell and Cosner, 2004). Here we review various aspects of space that must be modeled, covering both spatial movement and spatial interactions. Spatial movement, such as dispersal, is familiar to population geneticists, but spatial interactions may not be, even though some sort of negative feedback between population density and net reproductive rate is necessary to avoid unbounded growth.

This paper is targeted at researchers who wish to produce simulations that are – at least roughly – modeled on concrete empirical systems. At least parts of this guide will be useful to: empirical researchers wanting to explore plausibility of hypotheses or power of study designs; methods developers who want to test their methods on realistic spatial models; theoreticians wishing to explore spatial models; or managers wishing to explore alternative scenarios. So, we aim to make it easy to build a model starting from those quantities that we generally have good estimates of for particular organisms. This differs from the modeling philosophy in much theoretical work, which often begins from a "simplest possible" model. For instance, population genetics theory often uses the Wright–Fisher model, with the justification that the effects of small fluctuations in population size may be absorbed into the effective population size parameter. Similarly, both evolutionary and ecological theory often rely on (partial) differential equation models that do not explicitly represent individuals, and so rely on various statistical approximations.

Why individual-based simulations, and why in continuous space? Since real individuals are discrete, and live in continuous space, such simulations can in principle more accurately model real-world situations. Some phenomena simply require individual-based simulations (*e.g.*, locally adaptive genotypes). Non-individual-based models can be more computationally efficient, but at an often unknown cost to accuracy (Stillman et al., 2015). For instance, commonly-used discretized spatial landscapes can scale to much larger populations, but such discretized landscapes do not provide a consistent approximation to certain aspects of continuous-space dynamics (in two or more dimensions, Barton et al., 2002; Battey et al., 2020), and so it is difficult to understand which predictions apply to real-world situations. An even more practical consideration is that developing individual-based simulations in continuous space often requires *less* time and modeling expertise. This is because it is usually relatively easy to quickly come up with order-of-magnitude estimates of the various aspects of an organism's life cycle. On the other hand, more abstracted modeling frameworks usually involve analytical approximations, compound parameters, and other technicalities that require careful checking and mathematical experience. For instance, how does one translate "offspring disperse around 100m" to a migration rate between adjacent grid cells in a discretized landscape? (There are good answers to this question, but these are less straightforward than directly inputting a mean dispersal of 100m.) This assumes the modeler has a particular organism in mind, while simulations with fewer choices might feel more general for theoretical work or methods development. However, this can be a false generality, as more work is required after the fact to determine which real-world organisms a given result or method is applicable to.

Explicit individual-based population models are not new to ecology (DeAngelis and Yurek, 2017; Grimm, 1999). Furthermore, a great deal of ecological work has sought to quantify the effects of density-dependent demographic feedback – how demographic processes lead to stability of a population. For a single species, some sort of negative feedback between population density and growth rate is necessary to avoid the population growing without bound, although there are a great many ways to set this up in practice (De Wit, 1960; Beverton and Holt, 1957; Ellner et al., 2016). Demographers have a deep understanding of how to describe and parameterize the statistical properties of birth and death, and what the emergent consequences are for population growth, lifespan, age distribution, and long-term fitness (Tuljapurkar, 2013). Although temporal stochasticity is relatively well-understood in demography (Tuljapurkar, 1989), the consequences of spatial heterogeneity – particularly outside of metapopulation models (Hanski, 1997) – have received less attention.

Geography can have strong effects on patterns of genetic variation (Wright, 1943; Malécot, 1969; Rousset, 2000; Charlesworth et al., 2003; Battey et al., 2020; Min et al., 2022) and on evolutionary processes (Felsenstein, 1976; Uecker et al., 2014; Savolainen et al., 2007). Genetic differentiation is shaped by the movement of individuals, and hence distance and geographical features, as well as the spatio-temporal history of the species (Hewitt, 2011; Rosenberg et al., 2005; Ramachandran et al., 2005). Geography is therefore not only important but also provides a relatively untapped source of information to inform inference (Bradburd and Ralph, 2019). However, it is difficult to obtain analytical predictions from spatial population-genetics models (Felsenstein, 1975; Barton et al., 2002). Most spatial work in population genetics uses partial differential equations that do not represent genetic differentiation (Slatkin, 1973; Barton, 1979; Sedghifar et al., 2016; Etheridge et al., 2024), or specifically look at the fronts of population expansions (Paulose et al., 2019;

Nullmeier and Hallatschek, 2013; Etheridge and Penington, 2022).

Simulation has long been a useful tool in the study of populations (Grimm, 1999; DeAngelis and Yurek, 2017), particularly for the purposes of prediction (for example, for population viability analysis, see Dunning Jr et al., 1995) – even predating common usage of digital computers (Pearson, 1960). Simulations are also useful for inference, ranging from exploratory simulations to training for deep learning. The possible uses depend a great deal on the cost of a single simulation: for instance, many modern machine learning methods are only feasible with relatively fast simulations. Introducing geographic space increases computational complexity, making the task of producing sufficient training data more difficult (but see Smith et al., 2023; Champer et al., 2021).

This paper is intended as a guide to the territory of spatial modeling, with a focus on simulating individual-based models. We focus on discrete-time models of continuous geographical space, in which interactions and (usually) dispersal are local. The paper has two main parts: conceptual discussion of practical aspects of spatial simulation, followed by some biologically-motivated examples. The first topic is population regulation, as this is the topic which in our experience is the greatest barrier to new modelers. Next, we discuss spatial movement and mate choice, highlighting some interesting challenges. Subsequently, we discuss implementation of spatial heterogeneity and natural selection. Within each topic we provide examples chosen to demonstrate a particular technique or opportunity, with accompanying code snippets. Last, we demonstrate several complete, biologically-motivated simulations that incorporate many of the ideas presented earlier. We are particularly motivated by the simulation framework provided by SLiM (Haller and Messer, 2023), a flexible and powerful individual-based eco-evolutionary simulator that now includes a full set of tools for modeling interactions between multiple species across geographical landscapes. A minimal but complete SLiM script is provided in Appendix A, and our main text is accompanied by boxes that describe how to implement the spatial models we describe here in SLiM. This paper also serves to describe new capabilities in the newly-released SLiM v4.2. Code for all models below is available at `https://github.com/kr-colab/spatial_sims_standard`. We recommend that readers open these recipes in SLiM's GUI to experiment with while reading.

## Population dynamics and density

We start by describing how to maintain a stable population. An extremely simple spatial simulation might be as follows: organisms are asexual, and do not move during their lifespan. Each time step, each adult gives birth to a Poisson($f$) number of offspring, which each disperse to a random location, where displacement from their offspring-bearing parent's location is drawn from a Normal distribution with mean zero and standard deviation $\sigma_D$. Then, each individual, independently, dies with a given probability $\mu$.

Were we to simulate such a model, its flaw would be immediately apparent as it runs: either all individuals rapidly disappear (if $f \leq \mu$), or the computer grinds to a halt as the number of individuals explodes (if $f > \mu$). For this reason, we need some kind of *density-dependent feedback* to maintain a stable population that can be simulated for any reasonable length of time – in other words, we need the net population growth rate to change from positive to negative as the population density grows past some point. To achieve that, either birth rates need to decrease or death rates need to increase with local population density.

To provide density-dependent feedback, we first need a notion of "population density" at a point in space, a measure of the number of individuals nearby per unit area. Let $K$ denote the desired equilibrium population density (in individuals per unit area). A general way to define $n(\boldsymbol{x})$, the population density around a location $\boldsymbol{x}$, is to specify an *interaction kernel*, $\rho(\boldsymbol{x})$, which is a nonnegative function with $\int \rho(\boldsymbol{x})\boldsymbol{dx} = 1$; an *interaction scale* $\sigma_X$; and then if the locations of the individuals in the population are $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$, define $n(\boldsymbol{x})$ by

$$n(\boldsymbol{x}) = \frac{1}{\sigma_X^2} \sum_i \rho\left(\frac{\boldsymbol{x} - \boldsymbol{x_i}}{\sigma_X}\right). \tag{1}$$

(This is in two dimensions; in one dimension we would divide by $\sigma_X$ instead of $\sigma_X^2$.) Since $\int \rho(\boldsymbol{x}/\sigma_X)\boldsymbol{dx} = \sigma_X^2$, the value of $n(\boldsymbol{x})$ is in units of individuals per unit area. A common choice for $\rho(\boldsymbol{x})$ is the Gaussian density function. One concrete interpretation is that if $\rho((\boldsymbol{x} - \boldsymbol{y})/\sigma_X)$ gives the proportion of time that an individual at $\boldsymbol{y}$ spends near $\boldsymbol{x}$, then $n(\boldsymbol{x})$ is proportional to the time spent by *all* individuals near $\boldsymbol{x}$. (More concretely, $\int_A n(\boldsymbol{x})\boldsymbol{dx}$ is the total amount of time spent by all individuals in the region $A$.)

Now suppose that in each time step of the model, first each individual produces a random number of offspring; for an individual at $\boldsymbol{x}$ the mean number of offspring is $f(n(\boldsymbol{x})/K)$ where $f$ is the *birth rate* or *fecundity*, and each juvenile disperses to a nearby location whose displacement from $\boldsymbol{x}$ is chosen from a given probability distribution. Then, all individuals (including those just born) survive with probability $1-\mu(n(\boldsymbol{x})/K)$, making $\mu(u)$ the *mortality rate* at scaled density $u$. As written, each rate depends on spatial location $\boldsymbol{x}$ only through the smoothed population density $n(\boldsymbol{x})$ at that location, scaled by a parameter $K$ that controls the equilibrium density. For simplicity, in this model all individuals reproduce (as hermaphrodites), there is no consequence of aging, and we do not explicitly model the movement of an individual during its lifespan (but see subsequent sections).

**Equilibrium population density** Will this model stabilize, and if so, to what density? We expect an equilibrium when births balance deaths. If we define the local *per capita* net reproductive rate (the expected increase due to birth minus the decrease due to death, per individual) to be

$$F(u) = f(u)(1-\mu(u)) - \mu(u), \tag{2}$$

then we expect an equilibrium at a density of $n_*$ solving $F(n_*/K) = 0$. For this reason, it is convenient to choose the functional forms of $f$ and $\mu$ so that $F(1) = 0$, in which case we expect the equilibrium $n_*$ to be roughly equal to $K$. (However, we will see below that often $K$ is not *exactly* the equilibrium density.) In addition to $F(1) = 0$, for the population to be stable we also need $F'(1) < 0$ so that the net reproductive rate decreases with density near to the equilibrium. The argument $u$ in $f(u)$, $\mu(u)$, and $F(u)$ is the *scaled* population density $u = n(\boldsymbol{x})/K$, written this way so that it is easy to control the equilibrium population density by simply changing $K$, independently of other factors.

A brief note on what we have done here: it might seem most natural to set up a simulation using fecundity and mortality rates based on empirical observation. If so, then mean population density would be an emergent property of the simulation – in other words, our script would not have a parameter $K$ that we could directly adjust. However, to do this we need empirical estimates of how fecundity and mortality depend on local density, which are very difficult to obtain. A much more common situation is to have estimates of fecundity and mortality rates *and population density* at equilibrium. The parameterization we outline here, in which $K$ is a directly tunable rather than emergent parameter, is designed to make this use case natural. (Fecundity and mortality rates at equilibrium are also naturally introduced into the functional forms for $f(u)$ and $\mu(u)$.)

**Functional forms** The next question is: what forms should we use for the functions $f(u)$ and/or $\mu(u)$? There is surprisingly little guidance from the theoretical or empirical literature. Population models mostly come in two flavors: "phenomenological" (or "top-down") models that only consider the net reproductive rate, $F(u)$; and "mechanistic" (or "bottom-up") models that explicitly consider birth and death separately (Geritz and Kisdi, 2012). We need a mechanistic model, since our approach here is individual-based, but most literature on density dependence uses phenomenological models (a notable exception, see Coulson et al. (2008), or integral projection models Ellner et al. (2016)). Matrix population models (Caswell, 2000) are (mostly) mechanistic and widely used for management, but rarely incorporate density dependence (for a good reason: management questions are often concerned with short-term predictions, for which linear models are good enough). Furthermore, empirical estimation of functional forms for density dependence is a logistically and statistically daunting task, even without considering environmental variation. Modelers using simulation may therefore wish to simply choose a mathematically convenient form, reducing the problem to estimating parameters given that functional form. Perhaps the most obvious strategy is to multiply the base fecundity or survival by function of $n$ that decreases for sufficiently large $n$. For instance, if we set $\mu$ to a constant and $f(n) = f_0 e^{-n}$, then the equilibrium is at $n_* = K\log(\mu/f_0)$; setting $f_0 = \mu$, this model would conveniently reach equilibrium near $K$. Appendix C summarizes various familiar functional forms for $F(u)$ from the literature, and works through examples.

**Regulation by mortality** Suppose we would like to use density-dependent feedback only on mortality. In this case, the average number of offspring is constant: we may say $f(u) = f$. Recall that $F(u) =$

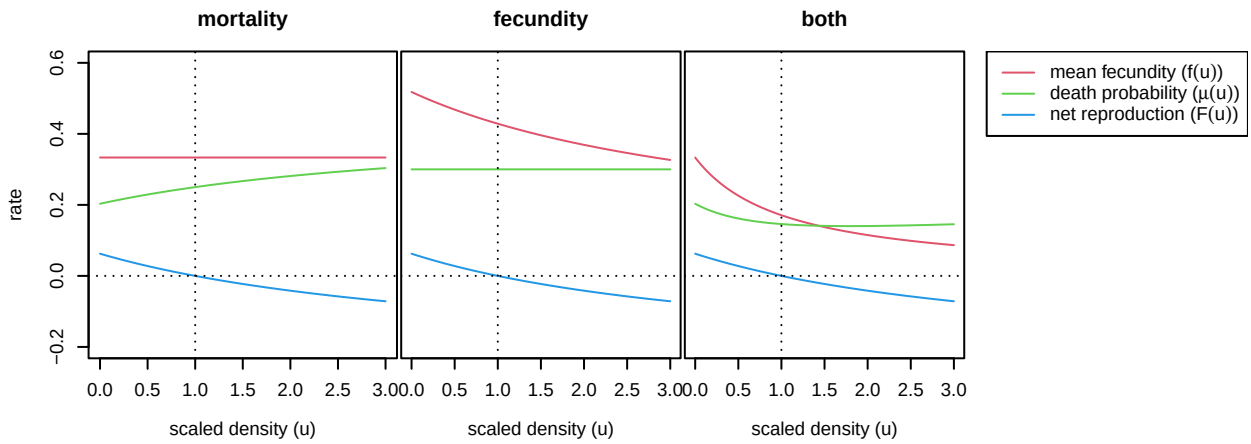



Figure 1: Three example models with "Beverton–Holt" regulation of population density. The relationship between scaled local population density and vital rates: mean fecundity ($f(u)$), probability of death ($\mu(u)$), and net reproduction rate ($F(u) = f(u) - \mu(u)$). In each, the horizontal axis is shown in units of $K$, the parameter controlling equilibrium density ("carrying capacity"). The models are: **mortality** from equation (4), **fecundity** from equation (5), and **both** from equation (6).

$(1 - \mu(u))f - \mu(u)$, and so the survival probability at scaled density $u$ is

$$1 - \mu(u) = \frac{1 + F(u)}{1 + f}. \tag{3}$$

The Beverton–Holt form for the net effects of density (Beverton and Holt, 1957) would have that $F(u)$ is proportional to $(1+a)/(1+au) - 1$, for some constant $a$ that controls the strength of the density-dependent feedback. So, to set up the model to have "Beverton–Holt" feedback, we plug in $F(u) = (1+a)/(1+au) - 1$ and obtain that

$$1 - \mu(u) = \frac{1 + a}{(1 + f)(1 + au)} \qquad \text{if } a \leq f. \tag{4}$$

The leftmost plot in Figure 1 visualizes Equations 3 and 4, and the others will be discussed later in Density dependence, life history, and stochasticity. This is also demonstrated in Box 1, and most of the examples in this paper use this model with $a = f$. The condition that $a \leq f$ is so that this gives us a valid probability: otherwise, the survival probability can be negative. In practice, it could be better to reparameterize with $s_0$, the survival probability at low density, so that $a = s_0(1 + f) + 1$. Note that the strategy used here cannot work with the discrete logistic: since $F(u) = 1 - u$ can get arbitrarily negative, eventually random density fluctuations will lead to negative probabilities, and errors in the simulation.

This example has been simplistic in many ways. Of course, there are many other possible ways to set up $f(u)$ and $\mu(u)$ that result in the same net form of density dependence (see Density dependence, life history, and stochasticity for further discussion). We also discuss there how stochasticity alters the realized equilibrium density. Finally, if there are two sexes and only one sex can bear offspring, then $f$ should not be the total number of offspring, but rather the number of offspring *of the offspring-bearing sex* (e.g., if females are the offspring-bearing sex, half the total assuming equal numbers of male and female offspring, representing just the female offspring). More generally, if vital rates depend on any aspect of the individual (for example, only adults can reproduce, or mortality is age-dependent), then the equivalent calculations must be done with a matrix population model (Caswell, 2000) or an integral projection model (Ellner et al., 2016). Our examples mostly ignore such complications, although several models in later sections have distinct life stages.

**Context and connections** *This section reviews wider connections to population modeling, and can be safely skipped by readers not familiar with that literature.* Historically, most spatial population models

**Box 1: Regulating mortality and fecundity in SLiM**

The Wright–Fisher model is a population model with a *fixed* population size. A spatial Wright–Fisher model, therefore, has *global* population regulation – individuals are affected by others arbitrarily far away, resulting in counter-intuitive, unrealistic consequences (Felsenstein, 1975). SLiM's default model is the Wright–Fisher model ("WF"). In this study, our goal is to model particular species in realistic ways, so we use only SLiM's non-Wright–Fisher ("nonWF") model, which requires explicitly choosing how births and deaths occur.

In a SLiM nonWF model, the "fitness" attribute of an individual is the probability of survival until the next time step. So, controlling mortality in SLiM is simply a matter of setting individuals' fitness. To compute local density, we use the `localPopulationDensity()` function, which computes density just as described in equation (1) (with options for the choice of kernel). To do this we need to first (during setup) define an `InteractionType` object, here using a Normal kernel with a scale of `SX` and a maximum distance of `SX * 3`):

```
initializeInteractionType(1, "xy", maxDistance=SX * 3);
i1.setInteractionFunction("n", 1.0, SX);
```

Now, to use Beverton–Holt regulation on mortality as described around equation 4 (setting $a = 1$ in equation 4 and with $K$ defined elsewhere), in each time step, we use this interaction type to set survival probabilities:

```
inds = p1.individuals;
density = i1.localPopulationDensity(inds);
u = density / ((1 + f) * K);
inds.fitnessScaling = 1 / (1 + f * u);
```

Note that the local population density is measured after reproduction and before death ($N_t^+$ in Appendix B), so we divide by $1 + f$ as well as $K$ when converting it to scaled population density $u$ from the equation (1), so it reflects density *before* reproduction.

On the other hand, offspring are produced by a "`reproduction()` callback", a chunk of code that is executed each time step for each individual, and produces any desired new offspring for the focal individual. For instance, if we would like to use Ricker regulation on fecundity (see equation 19 with $\alpha = 1$ and $\beta = 0$), we might for efficiency pre-compute each individual's number of offspring:

```
inds.tag = rpois(length(inds), lambda=exp(-density/K));
```

and then in the "`reproduction()` callback produce the offspring:

```
p1.addCrossed(individual, mate, count=individual.tag);
```

were either based on partial differential equations and so did not explicitly represent individuals (Skellam, 1951; Beverton and Holt, 1957; Cantrell and Cosner, 2004), used an array of randomly mating populations (*i.e.*, metapopulation models, Hanski, 1997), or had individuals living on a regular grid (Epperson, 2003). Although there are use cases for these, the main reason that individual-based continuous-space models are not more common may simply be convenience, both for mathematical analysis and software programming. The continuous-space formulation we use was introduced by Pacala (1986) and Bolker and Pacala (1997), and has been used in many theoretical studies (Dieckmann and Law, 1999; Snyder and Chesson, 2004; Etheridge et al., 2024).

A more modern extension to matrix population models, integral projection models, more commonly incorporate density dependence (Ellner et al., 2016), and may even estimate functional forms – see for example Adler et al. (2010). Much of the remaining literature on density dependence only considers the form of $F(u)$, rather than separating out the effects of density on different life history components (reviewed in Caswell, 2000; Eskola and Geritz, 2007). There are a wide variety of methods to infer $F(u)$ from time series data (Lande et al., 2003), but statistical issues make the problem difficult (Freckleton et al., 2006). However, the "$F(u)$" thus inferred is not necessarily the same as ours – many of these methods assume no

demographic stochasticity, and hence an infinite population size. In other words, the $F(u)$ that these infer is a *landscape-scale* relationship, averaging the net effects of birth and death over thousands or millions of individuals. Our $F(u)$, on the other hand, is local, and describes how a single individual is affected by having a few more or less neighbors. These are related, but need not have the same functional forms. In particular, the behavior of $F(u)$ for very large or small values of $u$ is usually much more important when applied to an individual, because the local density around an individual might be proportionally much larger or smaller than average, due to random fluctuations, than the total population size across a landscape. These can have a strong effect - for instance, increasing population density can have a positive effect on population growth (for at least low enough densities), a dynamic which can have interesting and important effects (Courchamp et al., 2008).

Furthermore, effects of density dependence are often mediated (at least in part) through interactions with other species, and the nature of these interactions may depend on environmental context. These interactions are often modeled as the net result of pairwise interactions, and various methods are used to estimate these potentially numerous and environmentally-dependent effects (for recent examples, see Weiss-Lehman et al., 2022; Bimler et al., 2023). It is beyond the scope of this article to review the full range of approaches and possibilities – but, note that there is no obstacle to simulation of multiple species whose interactions vary across space and/or time in SLiM (Haller and Messer, 2023).

Huge amounts of observational effort have gone into detailed estimates of the demography of particular species, especially those of conservation or management concern. However, such observations by necessity describe a snapshot of demographic rates in particular conditions (or, averaged over a particular range of conditions). Estimating the functional responses to density necessary to describe long-term dynamics is much more difficult, although methodological progress has been made, for instance, by incorporating many sources of information with "integrated population models" (Zipkin et al., 2023). A great deal of ecological work has also sought to quantify the effects of density-dependent demographic feedback. In a spatial model, this population density is usually measured locally – around the individual in question – yielding concepts such as the "competition kernel" (Bolker and Pacala, 1997) or "crowding index" (Pacala and Silander, 1985). Density-dependent feedback has a key place in the theory of coexistence between species: the Janzen-Connell hypothesis suggests that stronger intra- than inter-specific negative density-dependent effects could provide a mechanism for species coexistence (Janzen, 1970; Connell, 1971; Terborgh, 2012; Hülsmann et al., 2021). A substantial number of studies across a variety of organisms (mostly plants) have quantified these effects, both within species (*e.g.*, Weiner, 1982; Silander, Jr and Pacala, 1985; Specht and Arnold, 2018; Spotswood et al., 2017) and between (*e.g.*, Mack and Harper, 1977; Song et al., 2021; Zaiats et al., 2021). Estimation of density-dependent interactions strengths between many species in a community is challenging, but important for understanding community assembly (*e.g.*, Weiss-Lehman et al., 2022; Bimler et al., 2023). Adding to this complexity, the effects likely often depend strongly on age (Richardson et al., 2024).

Researchers in plant ecology have made the most progress towards empirical understanding of the mechanistic underpinnings of the sort of local density dependence we require. In practice there are a great many possible ways to quantify the cumulative effect of the neighbors of a single individual (Weigelt and Jolliffe, 2003). Our formulation here treats all individuals equivalently, but in practice one could include the effects of age or size. For instance, a common method in forestry modeling defines a "neighborhood competition index" for a given tree as the sum over all neighbors of their diameter (to some power) divided by distance (to another power) (Bella, 1971; Daniels, 1976; Canham et al., 2004). Empirical studies have estimated these kernels in a variety of situations: for instance, Teller et al. (2016) and Adler et al. (2018) use spline methods to flexibly estimate the effects of total area of nearby plants on a target plant's growth and survival.

## Spatial scales and neighborhood sizes

Now that we understand how to obtain stable simulations, we can move on to a more spatial topic: that of spatial scale. It is convenient to parameterize a spatial model using the following quantities, depicted in Figure 2: (*i*) interaction scale, $\sigma_X$, the typical distance over which individuals affect each other ecologically; (*ii*) dispersal scale, $\sigma_D$, which is a typical distance between parent and offspring; (*iii*) mate choice scale, $\sigma_M$, which is a typical distance between mates; and (*iv*) movement scale, $\sigma_V$, the typical distance over which an individual moves each time step. Note that $\sigma_D$ is typically referred to as the "dispersal distance", but here

we use "dispersal scale" for consistency with the other spatial scales. We have already met $\sigma_X$, and will discuss the remaining ones in subsequent sections.

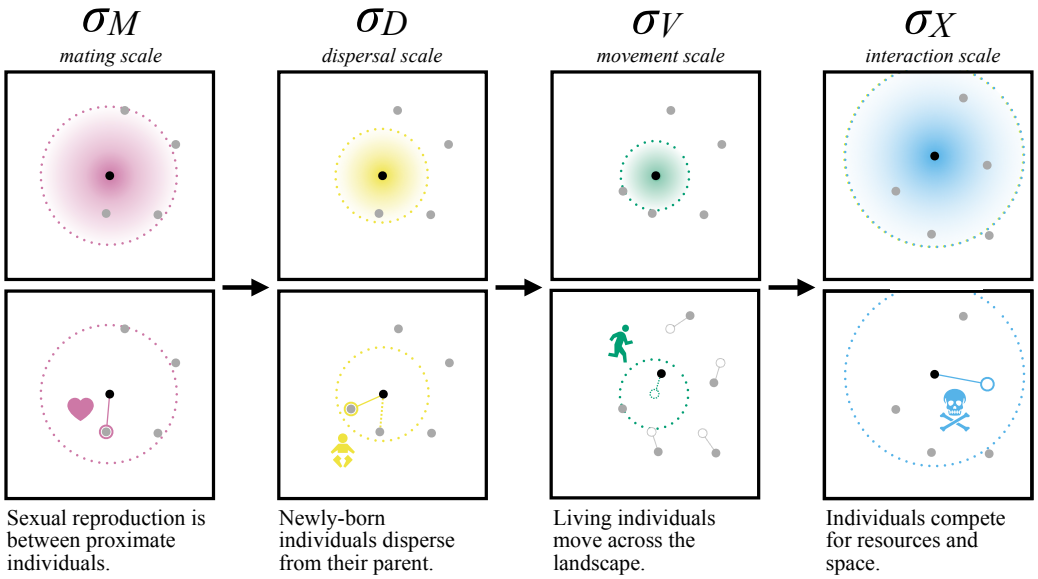



Figure 2: Spatial processes act over spatial scales defined by the $\sigma$ parameters (radii of dotted circles). In one simulation time step, mate choice, offspring dispersal, adult movement and inter-individual interaction (columns) all take place. (The order of these events is shown here as performed in SLiM.) The scale over which these processes occur are parameterized by $\sigma_M$, $\sigma_D$, $\sigma_V$, and $\sigma_X$, respectively; they can be used to calculate neighborhood sizes, to define the scale of the kernels used to draw movement or dispersal vectors (see Movement and dispersal) and to estimate local population density (see Mating and other pairwise interactions).

Each spatial scale parameter determines how many other individuals typically exist in an individual's "neighborhood", and so to describe the general behavior of a model it is helpful to look at a few "neighborhood sizes". First, the *mating neighborhood size*, $N_M = \pi\sigma_M^2 K$ and the *interaction neighborhood size*, $N_X = \pi\sigma_X^2 K$, measure, respectively, the typical number of other individuals in a circle of radius $\sigma_M$ or $\sigma_X$ at equilibrium. Note that these measures are only intended to be order-of-magnitude diagnostics: for instance, depending on the mating kernel, individuals out to $2\sigma_M$ or $3\sigma_M$ are probably also available to mate. These can be useful diagnostics: for instance, $N_M$ should be the same order of magnitude as the number of potential mates, so if it is small then mate limitation may be a problem (as discussed below in Mating and other pairwise interactions). Less obvious but equally important is $N_X$, which measures the typical number of other individuals that will factor into the local density of a given individual, and hence may affect their demographic rates (such as their survival probability, as in Box 1). $N_X$ provides a measure of the individual-to-individual variability in local density: roughly speaking, a larger $N_X$ means that local density is obtained by averaging over a larger area, and is thus less noisy. More concretely, suppose the number of neighbors of an individual within distance $\sigma_X$ is $P$. Birth and death rates depend on local density divided by $K$, which is roughly $P/N_X$ – so, if the standard deviation of $P/N_X$ is small, all individuals experience very similar local densities. The distribution of $P$ depends on the details, but we might expect the standard deviation of $P$ to be around $\sqrt{N_X}$ (if it is Poisson, for example), so the standard deviation of $P/N_X$ is proportional to $1/\sqrt{N_X}$. In other words, if $N_X$ is 100 then density should only vary by 10% or so; while if $N_X$ is around 10, then we expect some individuals to experience substantially lower and higher densities. Further discussion of how $N_X$ and $N_M$ can help diagnose odd model behaviour is given in the Appendix (section B).

Another "neighborhood size" that should be introduced here is a classical one: Wright's neighborhood size, which is $N_W = \pi\sigma_{\text{eff}}^2 K$, where $\sigma_{\text{eff}}^2$ is the (squared) *effective dispersal distance*, the variance of the displacement between parent and offspring along any axis looking back along a lineage (that is, along a chain of parent–offspring relationships). This quantity appears frequently in work on continuous spatial models in population genetics (Wright, 1943, 1946; Barton et al., 2002; Rousset, 1997; Robledo-Arnuncio and Rousset, 2010), and is clearly affected by $\sigma_D$, $\sigma_V$, and $\sigma_M$, as well as the mean generation time, but no explicit expression for $\sigma_{\text{eff}}$ from these parameters is known. $N_W$ gives, roughly, the number of "potential parents" of a given individual, and so is a measure of the rate of local genetic drift – if $N_W$ is small, then local inbreeding (and spatial structure more generally) will be stronger.

**Practical considerations** Although we encourage basing modeling decisions on empirical understanding, this is not always feasible. For instance, suppose one is simulating a species of fairly common shrub that lives widely across a landscape. In practice its local density is determined by local habitat (perhaps it lives on steeper slopes) and complex local interactions with other species. Any concrete estimate of the interaction scale for a plant is probably quite small – the effect of one shrub on another more than a few meters away is (in the short term at least) necessarily quite small. However, implementing a spatial model with an interaction scale of only a few meters (and no other species) will likely lead to a population size that is much too large. One option is to somehow include other species and fine-scale habitat suitability, but doing this in a realistic and efficient way can be a major challenge. A simpler option is to set the interaction scale to be on the order of the mean inter-individual spacing, and adjust the form of density dependence to roughly match the observed population density. The simulated population will probably be more evenly spread out across space than in reality, but it is hopefully at least a better approximation than a nonspatial model. More work is needed to develop appropriate modeling strategies for such situations, and to understand the consequences of these strategies.

# Movement and dispersal

The distance and timing with which individuals move across the landscape influences their spatial distribution, as well as the degree of relatedness expected for nearby individuals. Parent–offspring dispersal and the movement of (adult) organisms during their lifetime are sometimes separated, but can be implemented in similar ways. Clobert et al. (2012) reviews many aspects of dispersal, from parameterization and estimation to implications for ecology and evolution (see also Saastamoinen et al., 2018; Edelaar and Bolnick, 2012). (Clobert et al. (2012) also defined "dispersal" to be any spatial movement that leads to gene flow; this includes the movement of gametes, which we discuss below.) Besides the overall scale of movement, the next most important choice is the shape of the dispersal distribution, or *kernel*.

**Practical considerations** The easiest way to implement dispersal in two dimensions is simply to say that an individual at $(x, y)$ will produce an offspring that lives at $(x + \sigma_D X, y + \sigma_D Y)$, where $X$ and $Y$ are independent and Gaussian (with standard deviation 1). Most simulations in this paper were done in this way, for familiarity rather than any particular reason. How, then, to draw from a different dispersal kernel? The first guess – choose $X$ and $Y$ from a different distribution – does *not* work: the result will not be rotationally symmetric, and dispersal will tend to unrealistically align with the axes. Options are to either move a random distance at a uniformly chosen angle (which is simpler), or to multiply a bivariate Gaussian by a random scaling factor (which has other advantages). What to call a given two-dimensional kernel is not standard – for instance, would a "Student's $t$ kernel" have a $t$-distributed distance? Or, a $t$-shaped cross-section? In simulations below we use the latter convention, as described in Box 3, and discuss these choices more in Appendix D.

Movement in practice often depends on the environment, of course: organisms tend to move within particular habitats, and barriers are ubiquitous on all scales. Small-scale heterogeneity may be averaged out across the time scale simulated, and so incorporated (implicitly) in the movement kernel. However, large-scale heterogeneity can be important. Movement on a heterogeneous landscape still relies on some way of randomly choosing nearby points, and hence a movement kernel. One way to incorporate this is discussed in Box 5.

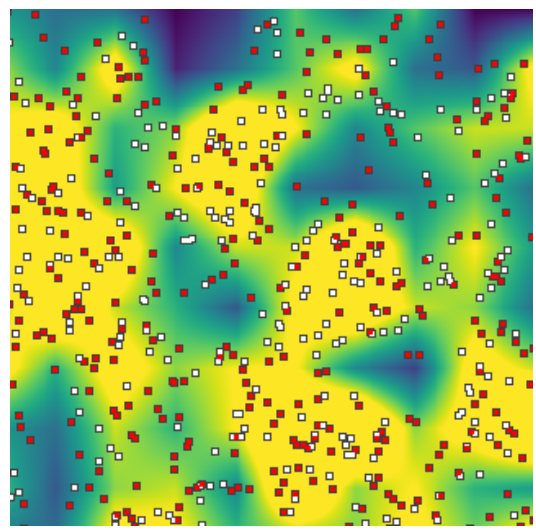
**Box 2: Summarizing the state of the population**

An extremely useful feature in SLiM is its GUI, which lets the user visualize the simulated individuals. Such visualization can, for instance, let one easily see areas of higher fecundity, differences in age structure, or even local adaptation. One way to do this is by setting the color of each individual: by default, individuals are colored by fitness (*i.e.*, probability of survival), but this code snippet will set the color to reflect the sex of the individual:

```
females = p1.subsetIndividuals(sex = "F");
females.color = "white";
males = p1.subsetIndividuals(sex = "M");
males.color = "red";
```

It can also be helpful to summarize the state of the population as a map. This can be done with great flexibility with the `summarizeIndividuals()` function, which creates a rasterized map for which the value of each pixel is some summary of the individuals within that pixel. For instance, the map of density shown as the background for the screenshot here was made using the following code:

```
density = summarizeIndividuals(p1.individuals,
  c(25, 25), p1.spatialBounds,
  operation="individuals.size();",
  empty=0.0, perUnitArea=T);
p1.defineSpatialMap("density",
  "xy", density, T, c(0,K),
  colors(20, "viridis"));
```



In this close-up screenshot of a dioecious simulation, there tend to be more females (white points) where the local density is higher (yellow background); and there tend to be more males (red points) where the density is lower (blue background). We discuss the heterogeneous spatial patterns generated in dioecious simulations below, in "Mating and other pairwise interactions". See below, for instance, in Box 5, for more use of `defineSpatialMap()`. Furthermore, SLiM's GUI provides built-in support for various diagnostic plots, such as the age distribution, population size over time, and allele frequency trajectories, and allows you to make your own custom plots as well.

**Context and connections** The concept of a dispersal distribution is perhaps most well-defined for plants, which (mostly) have only one opportunity to move during their lifetime, as a seed or other propagule. Seeds and pollen can be moved by gravity, wind (Nurminiemi et al., 1998), water (Murray, 2012), or animals (Morales and Morán López, 2022; Pons and Pausas, 2007), and complex models for these have been developed and estimated from empirical data (Neubert et al., 1995; Tufto et al., 1997; Austerlitz et al., 2004; Katul et al., 2005). Of course, many animals have relatively stable locations as well, perhaps depending on the season: for instance, Paradis et al. (1998) review estimates of post-natal and breeding dispersal in many bird species. The combination of these various processes (which often happen across many different spatial scales) has been referred to as the "total dispersal kernel" (Rogers et al., 2019), and can easily lead to "long-tailed" distributions (Cain et al., 2000; Edwards et al., 2007) in which rare, long-distance events are important. Dispersal is notoriously hard to estimate, as it often requires observing rare events, but has been done in a variety of organisms including kangaroo rats (French et al., 1968), mosquitoes (Estep et al., 2014), *Drosophila* (Dobzhansky and Wright, 1943), *Prunus* shrubs (Robledo-Arnuncio and García, 2007), pine

**Box 3: Random dispersal and displacement**

Perturbing a spatial location by adding a displacement drawn from a given kernel is a common operation in simulations – for instance, to choose offspring locations. To do this, SLiM provides the `pointDeviated()` method, which also needs to know the shape of the kernel, the type of boundary condition, and other parameters. For instance, the following code:

```
    locs = subpop.pointDeviated(nOff, individual.spatialPosition, "reprising", INF, "t"
    , DF, SD);
    offsprings = subpop.addCrossed(individual, mate, count=nOff);
    offsprings.setSpatialPosition(locs);
```

creates `nOff` offspring, and sets each offspring's position to a randomly sampled location near the location of the parent (`individual`). The random displacement is drawn with density $f(x) \propto (1 + \|x/\mathrm{SD}\|^2/\mathrm{DF})^{-(\mathrm{DF}+1)/2}$, *i.e.*, from the kernel whose density is formed by rotating the $t$ distribution with scale `SD` and `DF` degrees of freedom about the origin. The `reprising` argument conditions the result on falling within the spatial bounds of the simulation.

pollen (Robledo-Arnuncio and Gil, 2005), and butterflies (Suchan et al., 2024). Dispersal can of course also depend on density (Harman et al., 2020) (for example, if individuals preferentially disperse out of crowded areas).

**Clumping and movement** In many species reproduction is local, and so tends to produce clumps which are spread out by movement and dispersal. This balance can lead to either smoothly uniform populations, or strikingly uneven landscapes of high and low density on various scales: Figure 3 shows two examples. Such patterns can be intriguing or puzzling, and have real consequences for demography and genetic variation.

Although the scale of movement most strongly affects spatial patterns, the shape of the kernel is also important, and can have surprising effects: even very rare long-range movement can have strong effects on the speed of a range expansion (Mollison, 1972; Paulose et al., 2019) or the relationship between genetic and geographic distances (Smith and Weissman, 2023). For most models in this paper we use a Gaussian distribution for movement, mostly for the sake of familiarity. To see the effects of rare, long-range movement, a convenient "fat-tailed" kernel is the Student's $t$: the smaller the degrees of freedom parameter, the more likely are extremely long movements. Appendix D has more discussion, including how we define the kernel in more than one dimension.

Spatial clumpiness is sometimes visually obvious (as in Figure 3B), but more generally the clumping tendency of individuals can be measured by spatial correlations. One informative measure of clumping tendency is the *pair correlation function*, which shows for each distance $x$ how likely an individual is to have a neighbor at that distance relative to the average density (see Baddeley et al., 2015, for discussion and other useful descriptors of point data). Concretely, the pair correlation function estimates the mean density at distance $x$ away from an individual divided by the overall mean density, averaged across individuals; if the points are independently placed, it is constant at 1.0. Figure 3C shows pair correlation functions for simulations in which adults do not move and dispersal follows either a $t$ or a Gaussian distribution, or in which both dispersal and adult movement are Gaussian (with $\sigma_D = \sigma_V$). Otherwise, simulations are Beverton–Holt regulated on mortality.

Figures 3C shows that individuals are more likely to be around a distance of $\sigma_D$ from each other than otherwise, but this tendency is reduced with more long-distance dispersal ($t$ dispersal with lower degrees of freedom). In other words, a little long-range dispersal reduces clumping. Note, however, that in these examples the scale of clumping is quite narrow: correlations only extend out to 2 or 3 multiples of $\sigma_D$; for another visualization of this relatively subtle clumping, see Figure S3. Adding adult movement to a dispersal-only model unsurprisingly reduces correlations as well.

In spatial models neighbors tend to be more related to each other than to distant individuals, a pattern known as "isolation by distance" (Wright, 1943). The scale at which this correlation appears is determined by how far individuals move, and is also affected by the shape of the dispersal kernel: Figures 3C&D show

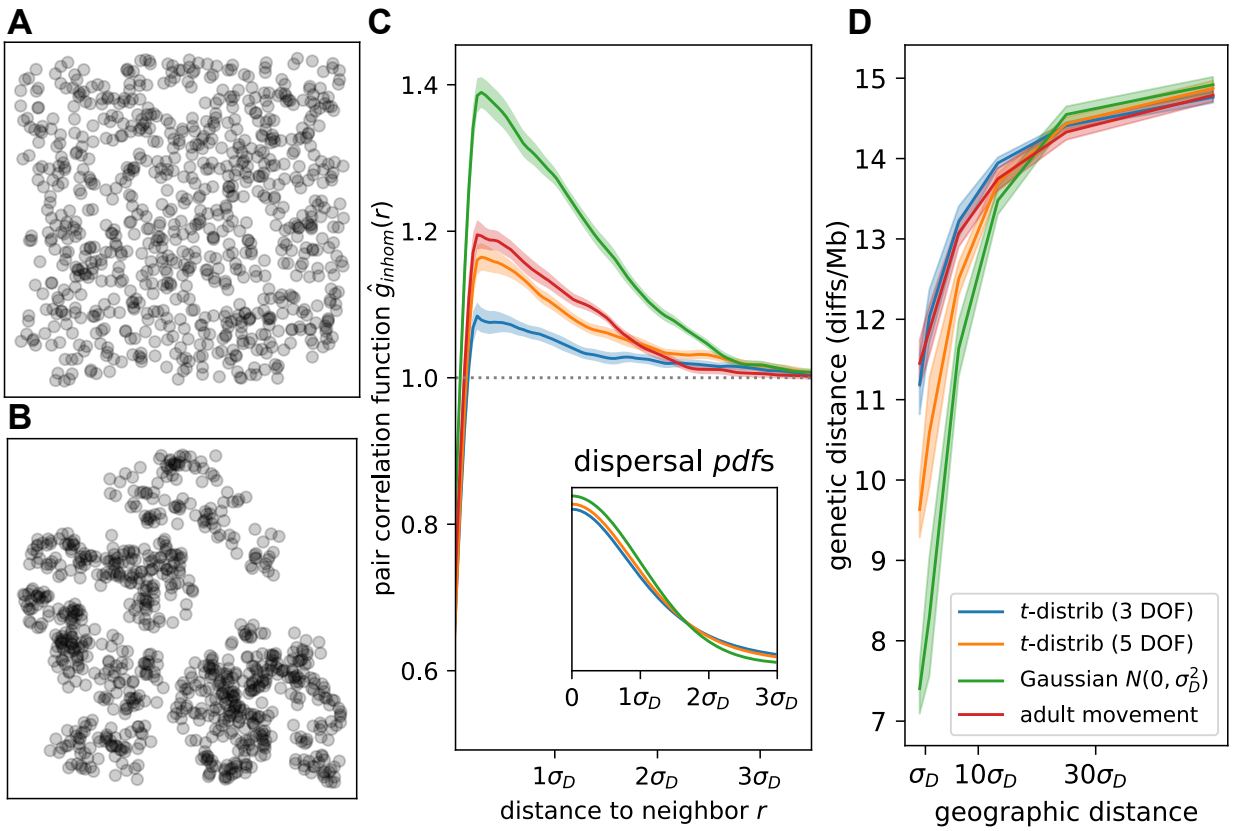



Figure 3: **(A, B)** Examples of simulations showing clumping (B) and not (A). **(C)** Dispersal kernel and movement type affects magnitude and spatial scale of clumping. Pair correlation functions (main panel) show density of pairs of individuals found a particular distance apart, relative to distances expected under a Poisson process (1.0; grey dotted line). Curves show the average across 50 independent time steps. Probability density functions (inset) for dispersal scales under the three dispersal kernels used: $t$-distribution with 3 degrees of freedom, $t$-distribution with 5 degrees of freedom, and Gaussian. The scale of $t$-distributions is equal to the standard deviation of the normal distribution ($\sigma_D = 0.3$). The "adult movement" scenario uses the Gaussian for dispersal as well as for movement at each time step. **(D)** Fat-tailed dispersal and adult movement flatten genetic isolation-by-distance. Plots show mean genetic distance between pairs of individuals at increasing geographic distance, averaged across ten independent replicates.

that simulations with more long-range movement – but with comparable mean dispersal scale – tend to have a weaker relationship between geographic and genetic distance (see also Smith and Weissman, 2023).

## Mating and other pairwise interactions

Mating is a crucial interaction for biological simulations, and there are numerous aspects and choices to consider. We probably don't want to simulate the detailed movement of individual pollen grains or the meanderings of a male moth seeking a female, and instead would like to skip to the realized outcome, *i.e.*, "choose a mate nearby". (We follow the literature on mating systems in calling this "mating", even when referring to plants or broadcast spawners.) As with dispersal, it is easiest to specify what "nearby" means with a kernel: roughly, an individual can be chosen with probability proportional to the kernel. More concretely, if the kernel is $\rho$ and the mating scale is $\sigma_M$, then an individual at $\boldsymbol{x}$ would assign weight $w_i = \rho((\boldsymbol{x_i} - \boldsymbol{x})/\sigma_M)$ to another individual at location $\boldsymbol{x_i}$, and then the probability individual $i$ is chosen is equal to $w_i$ divided by the sum of weights across all (nearby) individuals. This same method can be used for other individual-to-individual interactions, such as predators choosing prey individuals.

So, we can adjust the typical distance between mates with the mating scale, $\sigma_M$, which is directly analogous to parameters used to describe interaction ($\sigma_X$) and dispersal ($\sigma_D$), although each uses an underlying kernel in slightly different ways. Of course, the criteria determining potential mates for a given individual differ widely among species (Shuster, 2003). Relevant questions about the mating system include: How often does selfing occur, and in what circumstances? Are sexes separate (dioecy/gonochory) or not (monoecy/hermaphrodity)? Are there distinct mating types or self-incompatibility systems? An important note is that for dioecious species, calculations to determine stability of population density are easiest if done using only the reproducing sex.

**Box 4: Interactions, and mating**

The mechanism that SLiM uses to mediate most effects that some individuals have on others is called an "interaction type" (see Box 1). We used a symmetric interaction type in Box 1 to compute local density: every individual affected every other. Some interactions are not symmetric: we might, for instance, want each female to be able to find nearby males. To do this, we first set up a sex-specific interaction (again using a Gaussian kernel, with standard deviation `SM`):

```
initializeInteractionType(2, "xy", maxDistance=SM * 3, sexSegregation = "FM");
i2.setInteractionFunction("n", 1.0, SM);
```

The `sexSegregation` parameter value of `"FM"` means that females will receive the interaction and males will exert the interaction; it is asymmetric. Then, we can use the interaction in a `reproduction(NULL, "F")` block as follows (the `NULL` and ``'F'' arguments imply that it applies to all females) to produce a single offspring:

```
mate = i2.drawByStrength(individual, 1);
subpop.addCrossed(individual, mate, count=1);
```

This chooses a male `mate` from among the neighbors of the focal female, `individual`, with probability proportional to a Gaussian kernel with standard deviation `SM`, as described in the text. The interaction type itself guarantees that the chosen mate will be male. It is possible to set up other constraints on interaction types as well, such as a minimum and/or maximum age, to represent other constraints on the reproductive eligibility of individuals.

## Sex-specific spatial structure

The combination of density dependence and mating system can have surprising consequences: for instance, in the dioecious simulations above, clustering is dependent on sex. To see this, Figure 4 shows the *mark connection function* (Baddeley et al., 2015, §14.6.4.2) for female/male, female/female, and male/male neighbors, which shows the proportion of pairs of points at distance $r$ that are one female and one male, two females, or two males, respectively. Curiously, we see that the probability that an individual's neighbor is of the other sex does not depend on that neighbor's distance. However, the probability of having a *same*-sex neighbor *does* change with distance. As shown in Figure 4, within $3\sigma_D$ neighbors of a female are more likely than expected to be female, and nearby neighbors of a male are less likely to be male than expected.

Why does this happen? A simple reason is that we are modeling a dioecious scenario with no adult movement and where offspring are only generated by females and placed nearby. Local density-dependent mortality means that all individuals tend to kill their neighbors, but only females can replace them. Correspondingly, the spatial range of a dioecious system where offspring disperse from a particular sex is determined by the range of that sex – others on the periphery cannot extend the range because offspring do not disperse from them. This is a disadvantage for colonizing new areas (Obbard et al., 2006), and may explain the spatial distribution of dioecious individuals away from a range front (Mirski et al., 2017).

Indeed, Shuster (2003) showed that female aggregation (as observed in Figure 4) is a universal consequence in mating systems with female choice (Shuster, 2003, ch. 2). Other aspects of the simulation may differ by sex: for instance, sex-biased dispersal is pervasive (Trochet et al., 2016) and generates detectable spatial

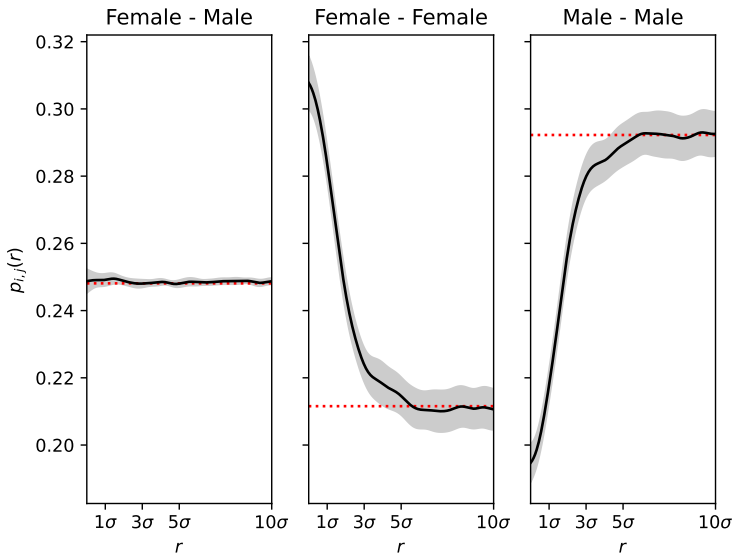



Figure 4: Dioecy generates female underdispersion, male overdispersion. Each panel shows the proportion of pairs of individuals at the distance shown on the horizontal axis that are either **(left)** a female and a male **(middle)** both female, or **(right)** both male. Average and standard deviation of 50 independent simulation ticks are shown. Red dashed lines indicate expectation if individuals' locations were chosen uniformly. Note that the "Male–Female" proportion is be identical to the "Female–Male" proportion, so that twice the left panel plus the other two panels is equal to 1.

patterns of genetic relatedness between sexes (Aguillon et al., 2017; Broquet and Petit, 2009; Laporte and Charlesworth, 2002). Social mating structures can induce sex-biased dispersal and thereby create similar patterns (Pusey and Packer, 1987; Hammond et al., 2005).

## Maps: spatial heterogeneity

Most real populations are far from uniformly distributed in space, and in most cases, the underlying cause is thought to be environmental heterogeneity. Until this point, we have considered simulations of homogeneous landscapes only. Such "flat" landscapes are useful for developing models and/or theory, but incorporating aspects of real landscapes can make simulations better match the complicated real world. Spatial heterogeneity can be introduced simply by making some parameter of the model, such as fecundity or mortality, vary across space – in which case we can visualize the parameter as a map. Such maps might represent specific environmental conditions, habitat boundaries, or abstract habitat quality.

**Practical considerations** Raster-based images provide a convenient way to introduce a map of spatial heterogeneity into a simulation framework such as SLiM that cannot directly read geospatial data formats. As demonstrated in Box 5, values can be read directly from a CSV file. Alternatively, a monochrome PNG file consists of a rectangular grid of pixels with integer values between 0 and 255; however, the values then need to be shifted and scaled to lie in the desired range after being read in. In the simple example shown in Figure 5, the map could represent the altitude of a (conical) mountain, with darker shades indicating higher elevation. High-resolution maps are publicly available from various sources, including NASA's Earthdata platform (2024), ESA's Earth System Data Lab (2024), or PRISM Climate Group. Open-source tools for processing remote sensing data are also available (see Montero et al., 2023).

To use a raster-based image, we must decide how the bounds of the image map onto geographic space. In general we match the (rectangular) image with the (rectangular) spatial area to be simulated. Although

**Box 5: Defining and manipulating maps**

SLiM provides extensive support for defining and manipulating spatial maps. We can begin by reading the values for a spatial map (say, of elevation) from a CSV file containing a rectangular grid of values using the code below:

```
mapValues = readCSV("elevation.csv").asMatrix();
map = p1.defineSpatialMap("elevation", "xy", mapValues);
```

(Images can also be read in as PNG files.) If this is a low-resolution raster then we may wish to smoothly interpolate it to higher resolution, done below with bicubic interpolation:

```
map.interpolate(factor=20, "cubic");
```

Once we have this map, we can extract the values of the map at arbitrary locations – for instance, the elevations at which individuals live:

```
elevs = map.mapValue(inds.spatialPosition);
```

Further operations are available, including blurring and algebraic manipulations of the values. The dispersal method in Box 3 allowed individuals to move equally well in any direction. To guide movement with a map – for instance, to induce a preference for moving uphill, for this example where map values indicate elevation – we can use the `map.sampleNearbyPoint()` method:

```
inds = p1.individuals;
pos = map.sampleNearbyPoint(inds.spatialPosition, INF, "n", SM);
inds.setSpatialPosition(pos);
```

This will move each individual to a new location sampled nearby and weighted by map value: if the original location is $\boldsymbol{x}$, $\rho()$ is a Gaussian kernel (specified as type `"n"`) of width `SM`, and the value of the map at location $y$ is $m(y)$, then the new location is chosen with density proportional to $m(\boldsymbol{y})\rho((\boldsymbol{y} - \boldsymbol{x})/\sigma_M)$.

the image is represented as discrete pixels, we will need to obtain values of the map at arbitrary locations (not just at the center each raster pixel). This can be done, for instance, by either associating each pixel's value with the rectangle it would visually cover (as seen in Box 2), or associating each pixel's value with the corresponding cell's midpoint and interpolating between.

Figure 5 shows results from simulations in which organisms living higher on a simulated mountain are more likely to survive. This is implemented by setting the local carrying capacity, $K(x)$, to be given by the value of the map at $x$.

Figure 5C shows realized densities at various elevations (averaged over 1,000 time steps) with different values of dispersal scale ($\sigma_D$). The first thing we notice, either from a bird's-eye view of the simulation (Figure 5A) or from an empirical map of realized long-term average density (Figure 5B), is that population density at least roughly matches the underlying map (the background in Figure 5A). A more precise comparison (Figure 5C) shows that the local density at $\boldsymbol{x}$ is not exactly $K(\boldsymbol{x})$, and the amount of deviation can depend on various other factors (colors show three choices of $\sigma_D$). In most situations, the equilibrium density is below $K(\boldsymbol{x})$. This is probably for two reasons: first, stochasticity usually reduces equilibrium density (see Appendix B); and second, since the mountain is conical, each location is adjacent to less higher-density area than to lower-density area, and so the net flux of migrants is negative. At the lowest dispersal ($\sigma_D = 0.25$, black points), areas with carrying capacity below about 5 individuals per unit area are not self-sustaining (see Appendix B for more discussion). On the other hand, at the highest dispersal ($\sigma_D = 4$, green points), the overall relationship between density and carrying capacity is flattened as offspring from high-fecundity areas end up across the entire range – however, most offspring are still produced near the top of the mountain, and lower elevations are maintained by source-sink dynamics. Finally, even with the higher carrying capacity, the curves are generally lower than $y = x$ line: we explain why in Appendix B.

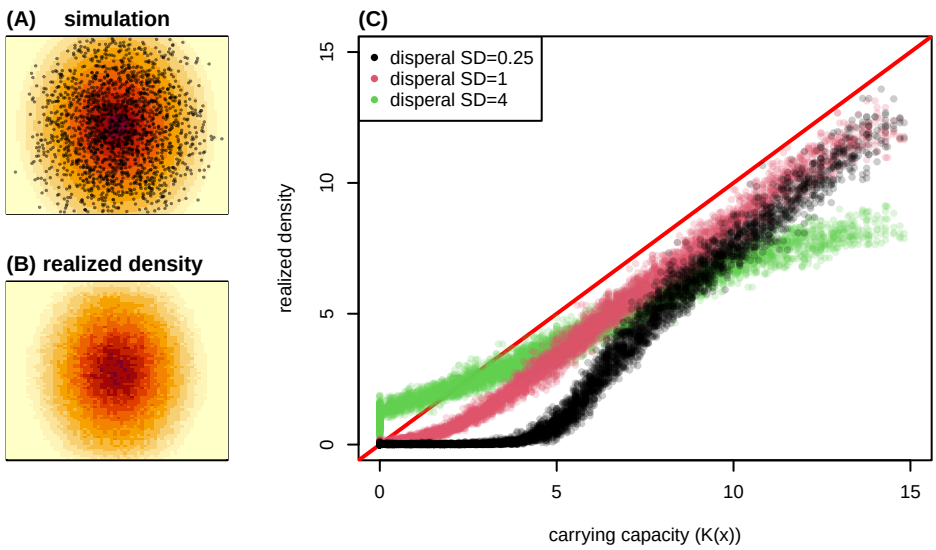



Figure 5: A spatial simulation using a heterogeneous map of carrying capacity and the reprising boundary condition. **(A)** A representative time step during the simulation, with each of the roughly 1500 individuals shown as points and the map of carrying capacity, $K(\boldsymbol{x})$, shown in the background. **(B)** A map of realized density, averaged over $10^4$ time steps. **(C)** Comparison of realized density to carrying capacity from three different runs using different dispersal scales: each point shows the realized density in one of the pixels of the map shown in (B) plotted against the value of $K(\boldsymbol{x})$ in the center of the pixel, for three values of $\sigma_D$ (labeled "SD" in the legend). The dimensions of the map are $25 \times 20$, and the interaction scale is $\sigma_X = 0.3$. The simulation has Ricker regulation of fecundity: the expected number of offspring of an individual at location $\boldsymbol{x}$ is $\exp(-\frac{n(\boldsymbol{x})}{K(\boldsymbol{x})} - 1)\mu/(1-\mu)$, where $\mu$ is the (fixed) probability of survival, $n(\boldsymbol{x})$ is the local density at $\boldsymbol{x}$, and $K(\boldsymbol{x})$ is obtained from the value of the image at $\boldsymbol{x}$.

## Density dependence, life history, and stochasticity

Above we introduced a model of "Beverton–Holt density-dependent feedback on mortality". We next give some other examples that have the same Beverton–Holt form for the net effect of density on population regulation, but differ in other ways.

**Fecundity regulation** The Beverton–Holt density-dependent regulation of equation (4) has the probability of death increase with local density while fecundity stays constant. Alternatively, we can set the probability of death to a constant: $\mu(u) = \mu_0$, and then solve for $f(u)$ to obtain the Beverton–Holt form $F(u) = \alpha((1+a)/(1+au) - 1)$ (it turns out we will need the scaling factor $0 < \alpha < 1$ to make a model with positive birth rates and death probabilities between 0 and 1). Then, fecundity should depend on scaled density $u$ as follows:

$$f(u) = \frac{1}{(1-\mu_0)}\left(\alpha\frac{(1+a)}{(1+au)} + (\mu_0 - \alpha)\right). \tag{5}$$

This is shown in the middle panel of Figure 1.

**Compensatory regulation on juvenile and adult mortality** Although the effect of density on net reproductive rate must be negative to have a stable population, the effect on individual demographic components can go the other way, as long as it is compensated by other components. For instance, the probability of survival $1 - \mu(u)$ can *increase* with density as long as fecundity $f(u)$ decreases faster. One way to do this

**Box 6: Adaptive empirical tuning for emergent parameters**

We often want a simulation to match a given estimated or observed density. However, this is not as simple as setting the value of $K$ in the code, because population density is an *emergent* quantity – a complex consequence of births, deaths, movement, and local interactions. Fortunately, even matching emergent quantities can be quite simple in SLiM. The code here matches population density to a desired value, but it demonstrates a general technique that may be applied to match other quantities such as mean age or degree of clustering.
We first define a global parameter `ADJ` that will be adjusted:

```
defineGlobal("ADJ", 1.0);
```

Then we modify the code from Box 1 so that this parameter adjusts the carrying capacity:

```
inds = p1.individuals;
density = 11.localPopulationDensity(inds);
u = density / ((1 + f) * K);
inds.fitnessScaling = 1 / (1 + ADJ * f * u);
```

Then, each timestep `ADJ` is updated by a factor $\exp[\alpha(Y - K)]$ where $Y$ is global population density (population size divided by total area) and $\alpha$ is an update rate (*e.g.*, $\alpha = 0.01$):

```
obsDensity = p1.individualCount / (WIDTH * HEIGHT);
defineGlobal("ADJ", ADJ * exp((obsDensity - K) * ALPHA));
```

If $Y > K$, we want the density to be lowered in the next cycle, so we decrease carrying capacity by increasing `ADJ`. On the other hand, if $Y < K$, `ADJ` is decreased so that we get a higher $Y$ in the next cycle.
This tuning should stop after an initial "burn-in" period, or the appropriate value for `ADJ`, once found, should be hard-coded into the final model. This is particularly important because if changes occur in the simulation that affect population size (*e.g.*, a reduction in habitat), this adaptive code will force the density back to the chosen value, which is generally not desirable – the population size should, in fact, change in response to such changes in the simulated conditions.

is to set

$$\begin{aligned} f(u) &= \frac{f_0}{1 + bu} \quad \text{and} \\ 1 - \mu(u) &= \frac{1 + F(u)}{1 + f(u)}, \end{aligned} \tag{6}$$

where $F(u) = \alpha((1 + a)/(1 + au) - 1)$, *i.e.*, the Beverton–Holt form again. This produces valid survival probabilities if $f_0 \geq a\alpha$. Depending on the choice of the parameters, $f(u)$ may increase or decrease with $u$, and is not necessarily monotone. One choice is shown in the rightmost panel of Figure 1.

**Beverton–Holt models?** In summary, these three examples (4), (5), and (6) might all be reasonably called "Beverton–Holt" models, although they differ substantially in the underlying mechanism, and could show very different demographic patterns (*e.g.*, age distributions). The three models of equations (4), (5), and (6) are shown in Figure 1. Although they have similar behavior around the equilibrium density, they have quite different life history implications. Most strikingly, in the first model, mortality increases with density, in the second mortality is constant, while in the third, mortality decreases with density. There are corresponding differences in age structure among the models, as shown in Figure S6, although the population dynamics (in terms of total population size) are similar. More examples along these lines are given in Appendix C.

**Stochastic effects** Despite all this theory, in practice equilibrium density is usually not $K$ – it is often lower. (See Figures 5 and S7, for instance.) This is due, effectively, to various forms of stochasticity.

One is random lack of mates (when $N_M$ is small); another is local extinction by fluctuations (when $N_X$ is small). These effects can be quite troublesome when setting up computational experiments across a range of parameters, especially if we want constant total population size.

Apart from those issues, the most common reasons for significant discrepancy between the observed density and the "desired" density (set by $K$) have to do with when and how density is measured. First: at what stage is density measured? In each time step, density increases after birth and decreases after death; $K$ can match at most one of these times. Second: where is density measured? We naturally look at total population size divided by total area; however, individuals (useful, but) necessarily experience on average a higher density, since they themselves count towards their density. So, the correct comparison is of $K$ to local population density averaged across the location of all individuals. Since this over-samples areas of higher density, this density will be higher than "total population size divided by area".

All these issues are discussed in much greater detail in Appendix B. For practical reasons it often suffices to simply be aware that population density is fundamentally an emergent property, determined in complex ways by nearly all parts of the life cycle. If a precise total population size is desired in a particular simulation, a simple solution is to adjust some parameter (*e.g.*, the birth rate) until the desired value is achieved (see Box 6), but of course that will (unavoidably) alter the dynamics of the simulation in other respects.

# Natural selection

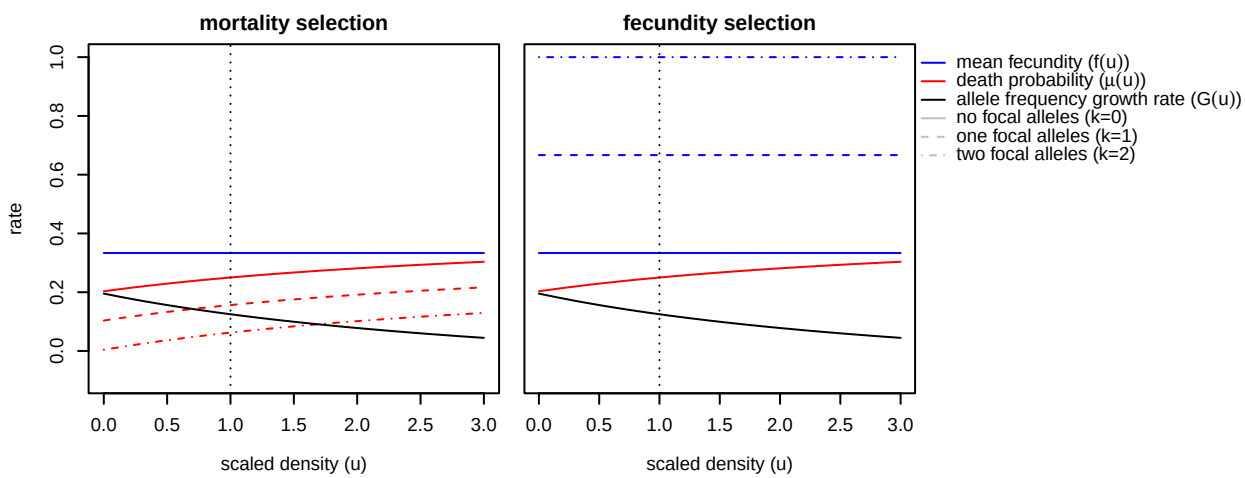



Figure 6: Density dependence of mortality (red lines) and fecundity (blue lines) described in equations (8) and (10), with $s = 0.25$. Line types show the mortality or fecundity for the different genotypes, i.e., for individuals carrying $k = 0$, 1, or 2 copies of the focal allele. The mean growth rate of the focal allele when rare (black line; $G(u)$ from equation (7)) is the same for both types of selection, at all densities. The vertical dotted line is at scaled density $u = 1$. Other parameters are as in Figure 1.

Evolutionary simulations of course need natural selection; it turns out that there are many different ways to implement selection in an individual-based simulation that all map to the same abstract models from population genetics, and yet produce distinct outcomes. We demonstrate this by comparing mortality- and fecundity-based selection. How do we think quantitatively about natural selection? Much of population-genetics theory is expressed in terms of the "selection coefficient" of a variant, usually denoted $s$, which is usually defined in the context of a Wright–Fisher model to be a change in relative probability of being "selected" to provide offspring to the next generation. Implementing natural selection in spatial models, even for the simplest case of a single variant, requires deciding where and how the variant affects the life cycle – for instance, the variant may affect mortality, fecundity, or both. Intuitively, one may achieve the same increase in "fitness" by increasing fecundity or decreasing mortality appropriately. However, these choices are known to produce differences in nonspatial models (Bodmer, 1965; Nagylaki and Crow, 1974), and so analogies between such the dynamics of such alleles and those of a selected allele in a Wright–Fisher model

are necessarily approximate. This is important when attempting to match simulation results to theory: just because a variable that affects survival or fecundity is named $s$ does not mean that using its value in expressions derived from the Wright–Fisher model correctly predicts the probability of establishment, mean frequency, or other classical quantities, even in a non-spatial simulation. Below, we parameterize and simulate from two models, in which selection acts on mortality or fecundity, in each choosing $s$ so that variant's frequency changes on average by a factor of $1 + hs$ per time step, when rare, where $h$ is the dominance coefficient. (This may not be the best definition for use with classical formulas, which often measure change per generation.) For a historical and philosophical review of definitions and measures of selection see Endler (1986). Implementation in SLiM is described in Box 7.

**Parameterization of selection** Recall that $f(u)$ and $\mu(u)$ are, respectively, the per-capita mean number of offspring and probability of death per time step when at scaled density $u$. Now, these can depend on the individual's genotype – so, let $f_k(u)$ be the mean fecundity of an individual with $k$ copies of a focal allele when at scaled density $u$, and similarly $\mu_k(u)$ the probability of death. We'll assume the focal allele is rare, so that most of the population is homozygous (and so use $f_0(u)$ and $\mu_0(u)$), but a few individuals are heterozygous (and so use $f_1(u)$ and $\mu_1(u)$). If so, then the per capita rate at which the number of focal alleles grows when at scaled density $u$ in an outcrossing species is

$$G(u) = (1 - \mu_1(u))(1 + (f_0(u) + f_1(u))/2) - 1. \tag{7}$$

(This is almost the same as the expression for $F(u)$, the expected change in the number of *individuals* in equation (2), except that the fecundity term is $(f_0(u)+f_1(u))/2$; this is because the offspring of a heterozygous parent will only inherit the focal allele from that parent half the time.)

First let's suppose an additive allele (so that $h = 1/2$) simply increases survival by $s/2$ per allele, so that

$$1 - \mu_k(u) = (1 - \mu(u))(1 + ks/2). \tag{8}$$

If the allele doesn't affect fecundity (so that $f_k(u) = f(u)$), then

$$G(u) = (1 - \mu(u))(1 + f(u))(1 + s/2) - 1. \tag{9}$$

We've set things up here so that $s$ means what we want – the growth rate when the new allele is rare, at the population's equilibrium density ($u = 1$), is $G(1) = s/2$.

Suppose instead that the allele increases fecundity. To make this allele have the same growth rate when rare as the allele that increases survival, we will set up $f_k(u)$ so that $G(u)$ has the same form. To do this, we can define

$$f_k(u) = (1 + f(u))(1 + ks) - 1. \tag{10}$$

With this form for fecundity selection, the growth rate when rare at density $u$ is also given by $G(u)$ in equation (9), and so at equilibrium, the per-capita growth rate of the number of alleles is again $s/2$. Roughly speaking, each allele increases fecundity by $s$, rather than the $s/2$ in equation (8) for mortality selection, because here the effects of fecundity are only affecting half of the parents (the offspring-bearing ones).

**That's the theory – how's it work in practice?** Above we laid out deterministic, large-population-size theory that suggests that a variant that affects survivorship as in (8) or fecundity as in (10) might sweep in similar ways. Although the resulting dynamics are similar, there are substantial differences. Figure 7 shows selective sweeps in spatial and nonspatial simulations, with selection acting on either fecundity or mortality. All simulations use density-dependent feedback on mortality, as in equation 4, and with either probability of survival or mean fecundity adjusted based on the genotype of the individual as in equations (8) and (10). The nonspatial simulations work the same, except that the global density (total number of individuals divided by total area) is substituted for local density when computing the probability of survival.

Figure 7A shows the allele frequency trajectories of the sweeping alleles: *de novo* mutations that either increase fecundity (red) or decrease mortality (blue). Comparing the allele-frequency trajectories between mortality- and fecundity-based selection, the speed of both spatial and nonspatial sweeps is unchanged, as

predicted (i.e., $G(u)$ is the same for both). We also see the well-known observation that selective sweeps progress more slowly in spatially structured populations.

However, we also see that sweeping populations behave differently in other ways. Figure 7B shows that the equilibrium population density increases as the beneficial allele fixes (as expected, since the selected alleles increase net fecundity minus mortality), but it increases by a greater amount under fecundity-based selection. Furthermore, Figure 7C shows that as the allele increases in frequency, the mean age drops with fecundity-, but not mortality-based selection. The reduction in mean age under fecundity-based selection once the beneficial allele becomes common is intuitive: population size increases as the sweep progresses, but because the beneficial allele does not confer protection against mortality, individuals are subject to the negative effects of increased population density. However, why fecundity-based selection results in a greater increase in population size is less clear.

To understand why this is, consider the population after the fixation of the focal allele: the mean per-capita change in population size, is now equal to zero. In these simulations, density dependence uses the Beverton–Holt model with $a = f_0$, *i.e.*, $f(u) = f_0$ and $1 - \mu(u) = 1/(1 + f_0 u)$. The new scaled equilibrium after fixation occurs at $u$ solving $(1 - \mu_2(u))(1 + f_2(u)) = 1$; for mortality-based selection this is at $u_\mu = 1 + s(1 + \frac{1}{f_0})$, while for fecundity-based selection this is at $u_f = 1 + 2s(1 + \frac{1}{f_0})$. In other words, the increase in equilibrium density for fecundity-based selection should be roughly twice that of the increase in mortality-based selection.

There are further differences between the spatial and non-spatial simulations, that are likely due to the spatial-density regulation processes that cause realized population sizes in spatial simulations to be smaller than in their nonspatial counterparts (as discussed in the previous section).

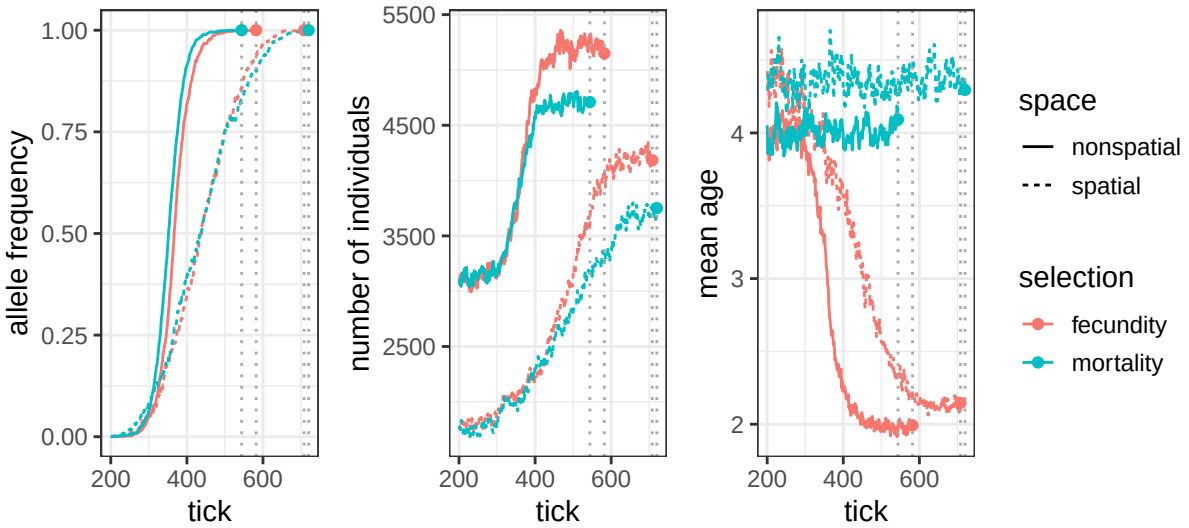



Figure 7: Allele frequency (A), population size (B), and average individual age (C) over time in spatial and nonspatial non-WF simulations. Lines end at the time the selected allele fixed (indicated by dotted vertical lines).

## Case Studies

How can we combine all the strategies we have seen thus far into a useful model of a living system? Here we illustrate how the preceding spatial modeling concepts can be used to model complex scenarios such as (1) temporal change, (2) complex life cycles, (3) continental-scale systems, and (4) competition for resources.

We motivate each scenario with an organism. The vignettes are not intended to be complete or accurate models of the population biology of these organisms; they are meant to illustrate how to apply the concepts

**Box 7: Selection**

In the nonWF SLiM models we use, the "fitness" of an individual as calculated by SLiM is the probability that the individual survives to the next time step. So, to make it so that each mutation affects survival by a factor of `S_MORT`, we simply declare it so:

```
initializeMutationType("m1", 0.5, "f", S_MORT);
```

(The factor of `0.5` is the dominance coefficient: heterozygotes will have fitness multiplied by `1 + S_MORT*0.5`, and homozygotes by `1 + S_MORT`.) To have the same type of mutations also affect fecundity with selection coefficient `S_FEC`, we just need to account for genotype when setting up offspring numbers (as in Box 1):

```
   indiv_s = 2*S_FEC * inds.countOfMutationsOfType(m1) / 2;
   inds.tag = rpois(length(inds), mean=FECUN * (1 + indiv_s) + indiv_s);
```

(As written, the effects of any mutation of type `m1` on fecundity are additive across loci and copies; other arrangements are possible.)

presented in this paper to a study system. Additional methodological information for each vignette can be found in Appendix E and our code example repository `https://github.com/kr-colab/spatial_sims_standard`.

## Changes over time: pikas and environmental change

Modeling spatial heterogeneity (as introduced in section Maps: spatial heterogeneity) makes simulations more realistic and informative, but sometimes *temporal* change in landscape variables is just as important. With this example, we explore how both seasonal fluctuation and globally rising temperatures affect population dynamics of an alpine organism: the pika.

Pikas (*Ochotona daurica*) are adapted to mountainous habitat at relatively high elevation, and they cannot survive in extreme heat (or cold), as would be encountered at lower (or higher) elevations (Beever et al., 2010). Such ecology makes pikas potentially vulnerable as global temperatures are expected to increase over time.

### Modeling approach

We incorporate three types of temporal change in temperature: (1) within-year seasonal change, (2) random, autoregressive fluctuations between years, and (3) steady global temperature rise. Because temperature varies with elevation and elevation varies dramatically in mountainous regions, we model temperature as a function of elevation. A topographic map for a region (40.40363°N to 40.53856°N; 105.7326°W to 105.5977°W) of Rocky Mountain National Park in Colorado was obtained using the `elevatr` package in R (with the WGS84 projection) and subsequently used to calculate the temperature at each point in space and time.

We connect temperature to fitness simply by killing offspring that are exposed to temperatures beyond the minimum (−5°C) or maximum (28°C) sustained by pikas (Beever et al., 2010). Each tick of our simulation represents one year. To account for *within*-year seasonal variation (*i.e.*, winter cold and summer heat), we narrow the viability range by the yearly variation in temperature ($V_S$), defining the probability of survival of a pika at location $x$ in year $t$ to be

$$1 - \mu(\boldsymbol{x}, t) = \begin{cases} 0, & \text{if } T(\boldsymbol{x}, t) < (-5 + V_S/2) \\ 0, & \text{if } T(\boldsymbol{x}, t) > (28 - V_S/2) \\ \frac{1}{1+fu(\boldsymbol{x})}, & \text{otherwise} \end{cases}$$

where the last term is our usual Beverton–Holt mortality regulation with local scaled density $u(\boldsymbol{x}) = n(\boldsymbol{x})/K$ (see equation 4). In this equation, $T(\boldsymbol{x}, t)$ represents the midpoint of seasonal temperatures in year $t$, defined

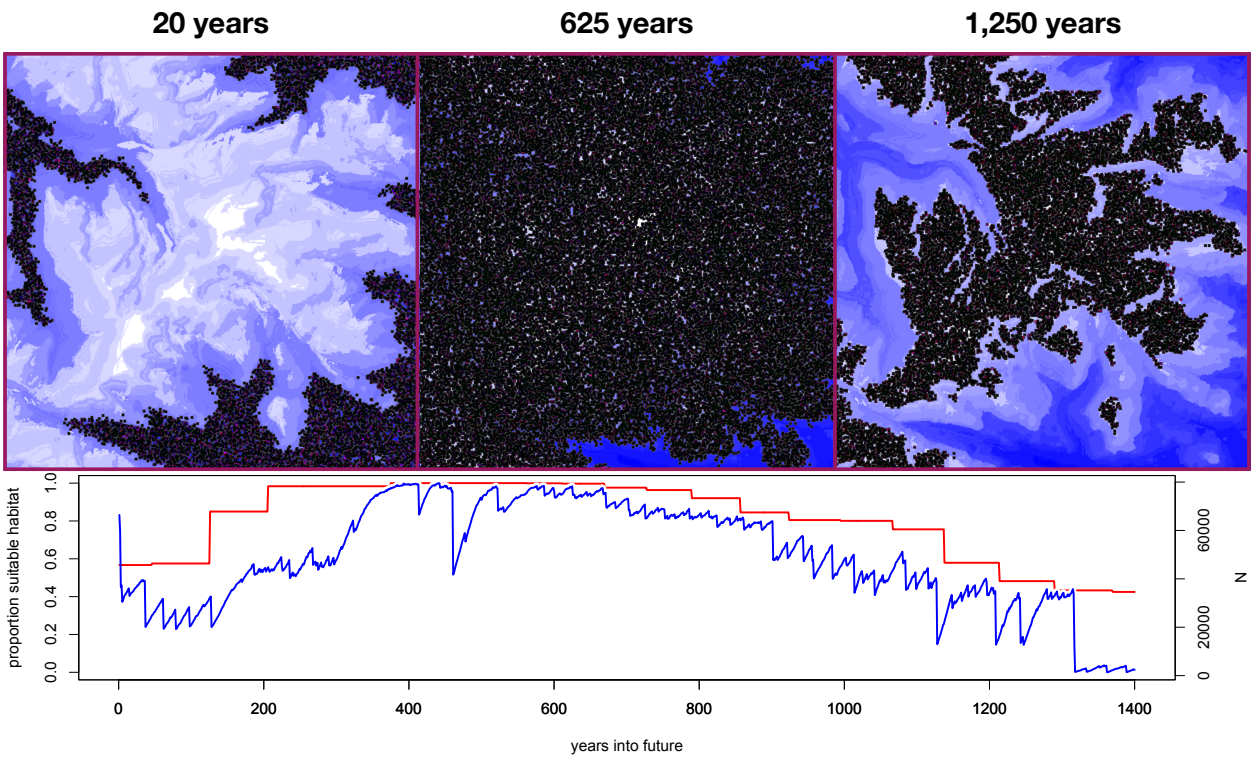



Figure 8: Pika simulation. (*top*) Screenshots of individual spatial positions (black) at different time points. The background image shows elevation, where blue and white correspond to lower and higher elevation, respectively. (*bottom*) The blue line shows population size over time; the red line shows the proportion of habitable space before the addition of random noise.

below, and the within-year variability $V_S$ broadens the temperatures experienced by the pikas around that midpoint. For example, if $T(\boldsymbol{x}, t)$ were $-4°$C (inside the viability range), but seasonal temperatures varied by $4°$C, the winter would be $-6°$C – cold enough to kill.

Between-year temperature changes are modeled by the temperature function $T(\boldsymbol{x}, t) = T^{\text{elev}}(\boldsymbol{x}) + T^{\text{fluct}}(t) + 0.016t$, where $T^{\text{elev}}(x)$ is calculated from the elevation map following fitted models from Collados-Lara et al. (2020), and an autoregressive term: $T^{\text{fluct}}(0) = 0$ and for $t \geq 1$,

$$T^{\text{fluct}}(t) = p^{\text{fluct}} T(t-1) + \epsilon(t),$$

where $\epsilon(t)$ is Normally distributed with mean zero and standard deviation $s^{\text{fluct}}$). The parameters $p^{\text{fluct}}$ and $s^{\text{fluct}}$ are known as persistence and shock, respectively, and determine how correlated the noise is between years. Finally, we increase the global temperature over time by adding $0.016°$C per year (Foster and Rahmstorf, 2011).

### Observations and extensions

The resulting simulation is a habitat suitability model in which the population's geographic distribution moves toward higher elevation as the global temperature increases (Figure 8).

We found that if the magnitude of random variation around the expected annual temperature is large – *e.g.*, standard deviation $5°$C – the probability of extinction increases significantly, particularly for early generations. In other words, one or a few bad years could be devastating for pika populations. This result may provide a useful lesson: even if a species appears to be thriving, the long-term success of the species is not guaranteed.

Surprisingly, simulated population size increased in the intermediate-term future, since for the first few hundred years, habitat losses at lower elevation were more than compensated for by habitat gains at higher

elevation. This illustrates that results from simulation on a specific map may not be generalizable.

Here, we use an empirically-informed linear trend with Gaussian noise as our global temperature change schedule. An alternative strategy for simulating temperature changes would be to pre-process multiple temperature maps reflecting different years, for example, and continually feed the simulation new maps over time. Such maps could perhaps even be generated directly from a climate model, rather than temperature being assumed to be a direct function of elevation.

## Life cycle stages: mosquitos in Burkina Faso

So far, our models have not taken life cycle stages into account; individuals have been able to mate immediately after they are born, and their survival has not been age-dependent. However, for many organisms their patterns of mobility, competition for resources, and mating capability are age-dependent. In this vignette, we demonstrate how to simulate a population with distinct juvenile and adult phases. Specifically, we set up a simulation of mosquitos in Burkina Faso in West Africa, inspired by North et al. (2019).

### Modeling approach

In this model, individuals begin life as juveniles, and mature into adults after a fixed time span. Population regulation and individual behavior depends on the life stage. For adults, survival is independent of the landscape map; adults have a constant survival probability of 0.875 per day (the model takes one time step per day). The population is regulated by density-dependent survival of larvae, that varies across a map. Following the model in North et al. (2019), we set the carrying capacity of larvae based on water availability. Larval carrying capacity was only non-zero at the outlines of water features extracted from GIS data of inland water in Burkina Faso. There, the carrying capacity fluctuated seasonally to mimic rainy and dry seasons. That is, $K(x, y, t) = K_{\text{base}}(x, y) + K_{\text{rain}}(t)$, where $K_{\text{base}}(x, y)$ is obtained from the map of inland water, and $K_{\text{rain}}(t)$ is a sinusoidal function with a period of 365 days and a minimum value of 0. We then use a Beverton-Holt form for the survival probability of larvae. Since there are ten age classes for larvae (and survival of each depends on the total number across all ages), parameterizing the model so that local density is (roughly) $K$ involves solving a system of equations (it is a matrix model), described in Appendix E.2.

A female adult mosquito mates with an adult male within the maximum mating distance, and lays eggs for larvae by sampling a location within the dispersal radius weighted by carrying capacity. Each day, adult mosquitos move by a random displacement sampled from a Gaussian distribution, whereas larvae do not move from their original location until they mature.

### Observations and extensions

This model simulates a mosquito population with structured life history. The population size of larvae and adults fluctuate periodically, following precipitation levels with a slight lag (Figure 9). Recall that adult mortality is simply constant, so the periodic fluctuations of adults are mediated through larval carrying-capacity dynamics.

We envision a few immediate extensions to our simulation. Here, we did not distinguish between perennial and fluctuating water sources. Simulating water sources that appear and disappear (adding a $t$ argument to $K_{\text{base}}(x, y)$) could result in extinction dynamics such as those observed in the pika example above. Similarly, the amount and duration of rainfall could be location-dependent (adding $x, y$ arguments to $K_{\text{rain}}(t)$).

We set the maximum of the $K_{\text{base}}$ and $K_{\text{rain}}$ functions to 0.002 individuals per square meter, which is probably much lower than in nature, to keep the memory usage of our simulation low. To run a simulation with a realistic (huge) number of mosquitoes we would need to make some efficiency improvements. Fortunately, an example of the necessary techniques follows.

## Continental-scale systems: invasion of the cane toads

It is often thought that individual-based spatial models are intrinsically so slow that they are quite limited in the population size that can be modeled. Our goal with this vignette is to demonstrate that this is not necessarily the case, by showing how to model large-scale populations and landscapes with relative ease and efficiency.

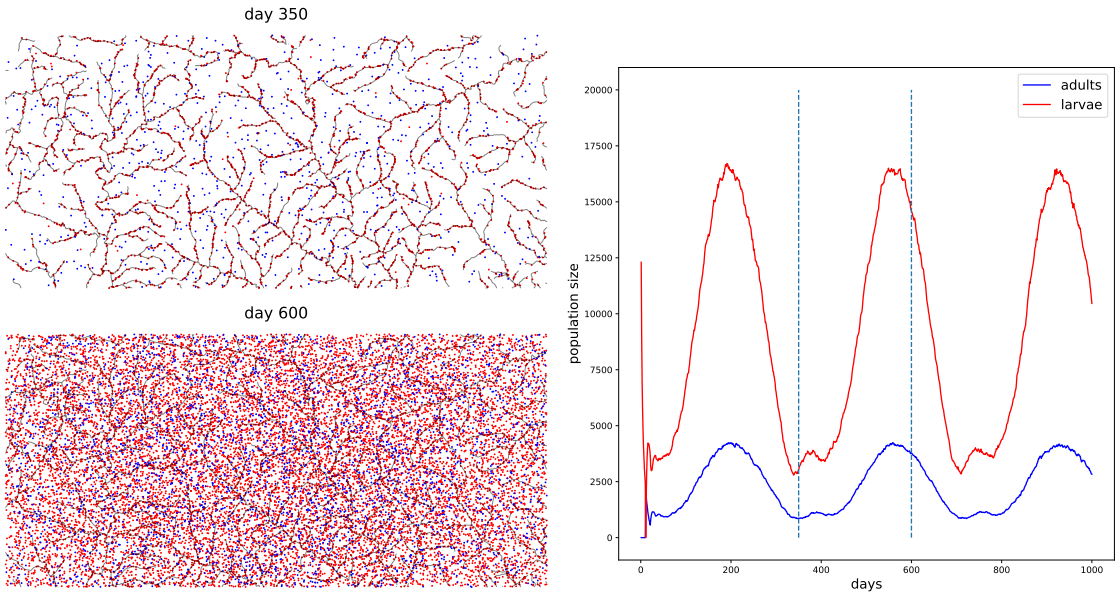



Figure 9: **(left)** Adults (blue) and larvae (red) population on the river map of Burkina Faso, at two time points during the year: in the dry season (*e.g.*, top, day 350), larvae can only survive in bodies of water, while in the wet season (*e.g.*, bottom, day 600), larvae can survive in many places. **(right)** Adults (blue) and larvae (red) population sizes oscillate with the seasonal cycle. Vertical lines indicate the time points when the distribution of the juvenile and adult populations are plotted on the left.

Cane toads (*Bufo marinus* or *Rhinella marina*) are native to Central and South America, and were intentionally introduced to the Northeast coast of Australia in 1935 as a pest-control measure. Since their introduction, cane toads have experienced explosive population growth, with hundreds of millions of individuals spreading over several million square kilometers, resulting in considerable negative economic and ecological impacts (Shine, 2010; Urban et al., 2008).

### Modeling approach

We parameterized the simulation model using information about the biology of cane toads when available. Toad population densities in established populations have been estimated around 8000 per square kilometer (Freeland, 1986). Based on this, and then tailored heuristically over several trial runs to produce the most realistic pattern of range expansion, we settled upon a local carrying capacity of 1000 per square kilometer. Telemetry data shows cane toads can travel up to 0.2 km per day (Shine et al., 2021). In order to interpret simulation time steps roughly as years, we set the spatial scale parameters $\sigma_D, \sigma_X$, and $\sigma_M$ to 20 km. For dispersal, we used a Student's $t$-distribution, which provided more rare long-range dispersal events than a Gaussian kernel would (See Fig 3).

We chose to model survival probability as a function of precipitation due to its visually similar distribution to the empirical occurrence data, which seems reasonable given cane toads' known sensitivity to moisture conditions (Child et al., 2009; Cohen and Alford, 1996). Specifically, we multiplied our typical density-dependent survival probability $n(\boldsymbol{x})/K$ by $1/(1+\exp(-(\alpha+\beta P(\boldsymbol{x})))$, where $\alpha$ and $\beta$ are parameters controlling the intercept and slope of the fitness curve, and $P(\boldsymbol{x})$ is the value of precipitation at location $\boldsymbol{x}$ in kg/m$^2$/year. Mating and interaction neighborhoods were large, which led to prohibitively long runtimes, so we used the fast map-based methods for local density estimation described in Box 8 and for mate choice described in Box 9.

We initiated the invasion with 10,000 individuals, though the number of individuals actually released in 1935 was likely greater (Shine et al., 2020). We chose the starting locations of the simulated toads by randomly sampling locations from the first four years of the observed occurrence data.

To make the simulation more efficient, we followed the approach described in Box 8, modified to measure

the local density of individuals per unit of *habitable* area.

### Observations and extensions

We were able to approach the true spatial and population scale of the cane-toad invasion, with a final census size of about 120 million individuals, nearing the estimated modern census size of Australian cane toads (Australia, 2019), after running for about 5 days using 200 GB of RAM at maximum. We visually compared empirical occurrence data for cane toads to our simulations, with and without annual precipitation's effect on survival (Figure 10). While there are obvious differences in densities and locations between the simulated and observed data, it is clear that modeling annual precipitation's effect on survival greatly improves the likeness.

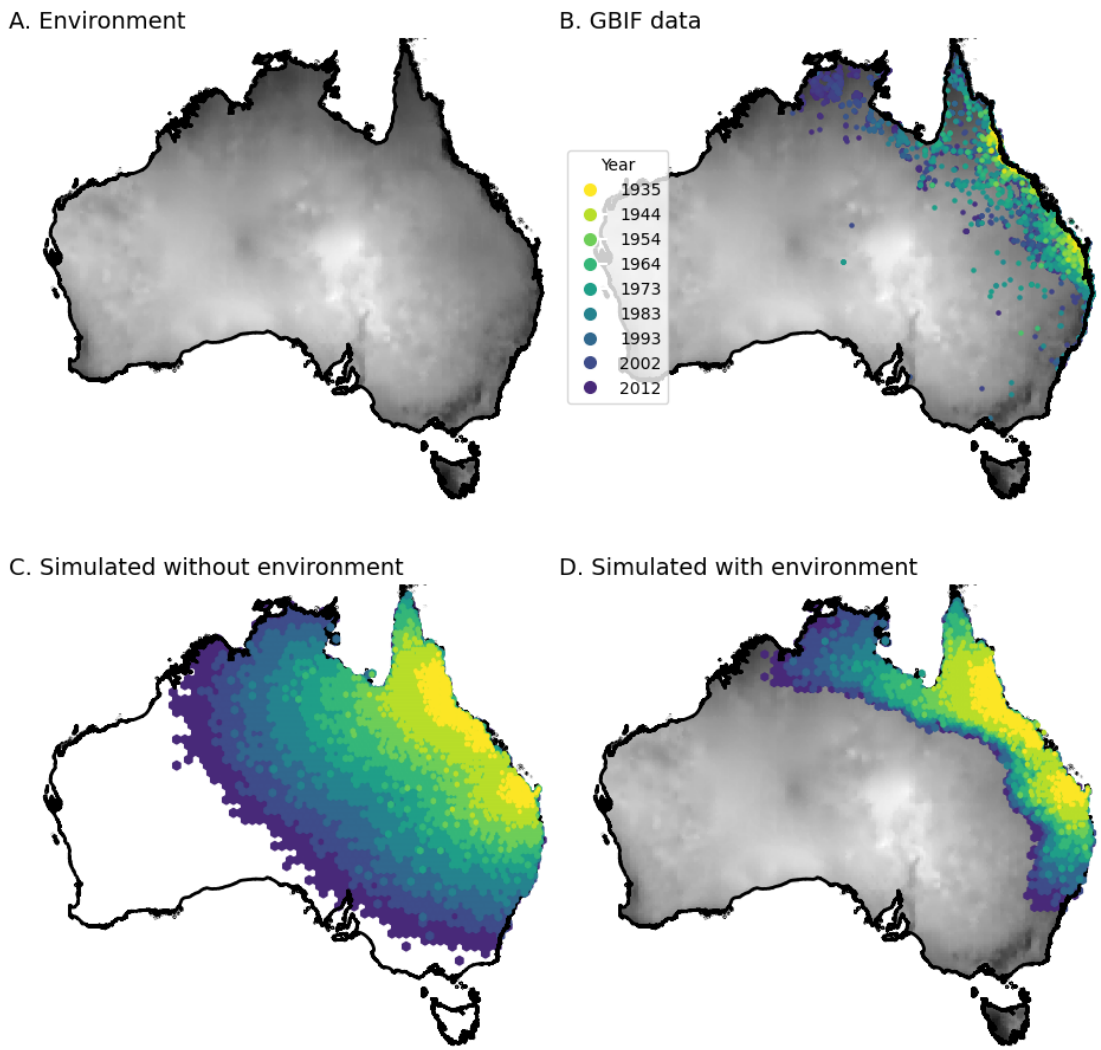



Figure 10: Simulating the cane toad invasion with and without and effect of annual precipitation on survival. **(A)** Map of Australia shaded by Annual Precipitation (kg $m^{-2}$ $year^{-1}$). **(B)** Observed distribution of cane toads from Global Biodiversity Information Facility (GBIF) **(C)** Simulation of cane toad invasion without Annual Precipitation effecting fitness. **(D)** Simulation of cane toad invasion with Annual Precipitation effecting fitness with all else being equal to **(C)**.

The approach described here is similar to classical niche modeling (Peterson, 2001), which has been used extensively to model cane-toad distributions (Shine, 2010), but is somewhat simplified by only using one

environmental variable. (This would be straightforward to extend; multiple environmental variables could be combined into a single habitability map for use in the model, with no impact on runtime performance.) The benefit of this vignette's approach over previous approaches is the way that it combines information from the explicit individual-based simulation and environmental data, producing a simulation at a realistic scale with respect to both landscape size and population size.

**Box 8: Using maps for faster spatial interactions**

In Box 1, we estimated the local population density for each individual with equation (1), via the `localPopulationDensity()` method. For each individual, this method sums the "interaction strength" (*i.e.*, kernel density) for every other individual within the provided maximum distance. So, if the total number of individuals is $T$ and the typical number of "neighbors" (other individuals no more than the maximum distance away) is $N_X$, then the complexity of this operation is $TN_X$. If each individual has a large number of neighbors, this can be quite costly. However, if the number of neighbors is large, it should work just as well to (a) create a (discretized) map of local density, and (b) look up the value of the local density experienced by each individual on that map. Map lookup is quick, so if the cost of creating the map is smaller than $TN_X$, we will have a more efficient model.
We can create a map whose value at $\boldsymbol{x}$ is (approximately) given by equation (1) in two steps: (1) use `summarizeIndividuals()` to measure the number of individuals per unit area in each cell in a grid (see Box 2), and (2) `smooth()` this map using the appropriate kernel, so that the density value at a given point in the resulting map depends upon an appropriate weighted average of the individuals per unit area across nearby cells of the grid. Using the same kernel as in Box 1:

```
    grid_dims = ceil(2 * p1.spatialBounds / SX);
    raw = summarizeIndividuals(p1.individuals, grid_dims, p1.spatialBounds,
        operation="individuals.size();", perUnitArea=T);
    density_map = p1.defineSpatialMap("density", "xy", raw);
    density_map.smooth(SX * 3, "n", SX);
    defineGlobal("DENSITY", density_map);
```

Then, we can modify the code of Box 1 to use the map instead:

```
    inds = p1.individuals;
    density = p1.spatialMapValue(DENSITY, inds.spatialPosition);
    u = density / ((1 + f) * K);
    inds.fitnessScaling = 1 / (1 + f * u);
```

This obtains (nearly) the same value as would `localPopulationDensity()` if the resolution of the map should be finer than the scale over which the kernel changes. In this example, that scale is `SX`, so we have ensured that the map has cells of size smaller than `SX/2`. The approximation is examined in Appendix F.

## Resource-explicit competition: monarchs and milkweed

In the preceding examples we have regulated populations through competitive interactions between individuals, either explicitly, as in Box 1, or in a space-averaged manner, as in Box 8. Population regulation in this model is managed quite differently. This is a "resource-explicit" model, wherein the population is extrinsically regulated by the availability of an external resource (Champer et al., 2024), as outlined in Box 10.

We simulate monarch butterflies (*Danaus plexippus*) and the milkweed (*Asclepias* spp.) plants on which they lay their eggs. Though adult monarchs feed on nectar from numerous types of plants, monarch caterpillars are specialists that eat only milkweed (Oberhauser et al., 2004).

**Box 9: Using maps for faster mate choice**

A similar problem as in Box 8 arises when choosing mates: even though only one mate needs to be chosen, the underlying operation is of order $TN_M$. The same map of density can be used to solve this problem as well: instead of choosing an individual with probability proportional to a kernel, it is (nearly) equivalent to: (1) choose a point in space nearby, with probability proportional to the map multiplied by the kernel, and then (2) take the individual closest to that point. Recall the number of possible mate scales as $N_M = \sigma_M^2 K$ from Section Spatial scales and neighborhood sizes. This implies that the number of neighbors grows linearly with $K$. Thus, we can keep the number of potential mate roughly constant regardless of how large local density is by rescaling the maximum distance in the `InteractionType` used for mate choice by $1/\sqrt{K}$ from the code in Box 4:

```
    initializeInteractionType(2, "xy", maxDistance=5/sqrt(K), sexSegregation = "FM");
    i2.setInteractionFunction("f", 1.0);
```

We've set the maximum distance in the interaction kernel to be a value that should give us around 25 neighbors for each individual; however, if density varies significantly across the landscape, this may make some individuals in low density areas fail to mate. Then, we choose the mate as follows:

```
  mate_location = DENSITY.sampleNearbyPoint(individual.spatialPosition, 3*SM, "n", SM);
  mate = i2.nearestNeighborsOfPoint(mate_location, p1, 1);
```

Here the specification of a Gaussian mate-choice kernel with standard deviation `SM` has moved from the definition of the `InteractionType` to the `sampleNearbyPoint` call: given a location $x$, a map with value $m(y)$ at $y$, and a kernel $\rho()$, this returns a random point $z$ sampled with probability proportional to $m(z)\rho(z)$. We then choose the mate as the individual nearest to that point. The approximation is examined in Appendix F.

### Modeling approach

Monarchs in the model progress through three life-cycle stages: caterpillar, pupa, and butterfly. Mortality is modeled differently in each stage. The food resource, milkweed, is directly included in the model as a second species.

During the first two weeks of their lives (time steps 0 and 1), caterpillars interact with nearby milkweed plants and accumulate resources. The amount of resources collected from a milkweed plant each week is inversely proportional to the number of competing caterpillars within a given interaction scale ($\sigma_X$) of that plant. This means that a plant fed upon by more caterpillars will be depleted more quickly. When individuals reach their third week, they enter the pupa phase. At this time, survival is proportional to the amount of milkweed eaten as a caterpillar. (The survival probability is calculated only once when it becomes a pupa.) Surviving pupae become butterflies during their fifth week. Butterflies disperse, reproduce, and experience mortality at an age-dependent rate. Mortality is not density-dependent during this life cycle stage. Figure 11 shows a snapshot of the simulation, highlighting the dispersal and reproduction of individuals across the landscape.

Reproduction also involves a resource-explicit interaction, since monarchs only lay their eggs on milkweed. To accomplish this, a spatial interaction is evaluated between adult males and milkweed (with a much longer range than the interaction between caterpillars and milkweed); each milkweed caches a list of nearby males from that interaction. Next, a similar spatial interaction is evaluated between adult females and milkweed; females are iterated through, with each female selecting a mate from the males cached at the milkweed plants within the female's interaction range. Finally, the females randomly distribute their eggs at the plants where matings occurred.

### Observations and extensions

In addition to reflecting the life history of monarch butterflies, the resource-explicit modeling approach is highly performant. Each female monarch can lay several hundred eggs, very few of which go on to reach

**Box 10: Resource-explicit foraging and mortality**

In this box, we outline the "resource-explicit" modeling approach described more fully in Champer et al. (2024). This approach implements density-dependent population regulation that is mediated indirectly through the availability of a resource, rather than directly through competitive interactions between individuals. A simple resource-explicit simulation contains two species in a multi-species SLiM model: the focal species, in subpopulation `p1`, and a species representing the resource, in subpopulation `p2` , in which individuals represent patches of the modeled resource.
During their foraging phase, individuals collect resources from nearby patches. Each patch can each support a certain number of individuals per time step. If a patch can support 10 individuals, but there are 100 individuals nearby at a given time step, each individual will receive a resource amount of 1/10 the amount necessary to guarantee survival (but might also forage from other nearby patches). The amount of resource that each individual collects over their foraging phase is stored in their `tagF` property.

```
    i1.evaluate(c(p1, p2));
    for (patch in p2.individuals) {
        customers = i1.nearestNeighbors(patch, p1.individualCount, p1);
        customers.tagF = customers.tagF + inds_fed_per_patch / size(customers);
    }
```

After their foraging phase, individuals that have not consumed enough resources are killed. In the code example below, this occurs when individuals reach age 2.

```
    at_risk_indivs = p1.subsetIndividuals(minAge=2);
    mortality_indices = runif(size(at_risk_indivs)) > at_risk_indivs.tagF;
    dead = at_risk_indivs[mortality_indices];
    species.killIndividuals(dead);
```

If an individual collected at least one unit of the resource, its survival to the next time step is guaranteed. Otherwise, it survives with a probability equal to the total amount it collected. See Appendix F for discussion of the relationship to the method of Box 8.

adulthood. As a result, the number of caterpillars in the model far exceeds the number of milkweed plants. Thus, regulating the size of the caterpillar population by evaluating a spatial interaction between caterpillars and plants is far more efficient than regulating the population by evaluating an interaction directly between caterpillars.

This approach also provides an intuitive way to investigate questions related to resource availability. For example, this model could probe the effects that milkweed habitat loss, caused by urbanization or climate change, could have on the viability of the local monarch population.

# Conclusion

Individual-based simulations are a powerful method for studying how demographic and population-genetic processes operate over continuous geographic space. Modelers must design rules for how individuals in the simulation interact with others nearby, and how forces such as selection operate. Individual-based simulations are well-suited to this problem, since they are very flexible, and can be tailored to a specific research system. However, flexibility can be both a blessing and a curse: it is easy to design a simulation with unstable population dynamics or unrealistic life-history traits. Similarly, stochasticity can cause a population that is intended to equilibrate to instead die out. Such problems likely reveal a flaw in our understanding of the system being modeled. Here, we provide guidance and connections to the ecological literature for researchers interested in designing stable, efficient, and interpretable spatial simulations. Interpretability is a key advantage to spatial, individual-based simulations, since it can take substantial effort to translate the results of more abstracted models back to the domain of interest.

Realism is not a goal of our case studies, but each illustrates the degree of realism that can be obtained

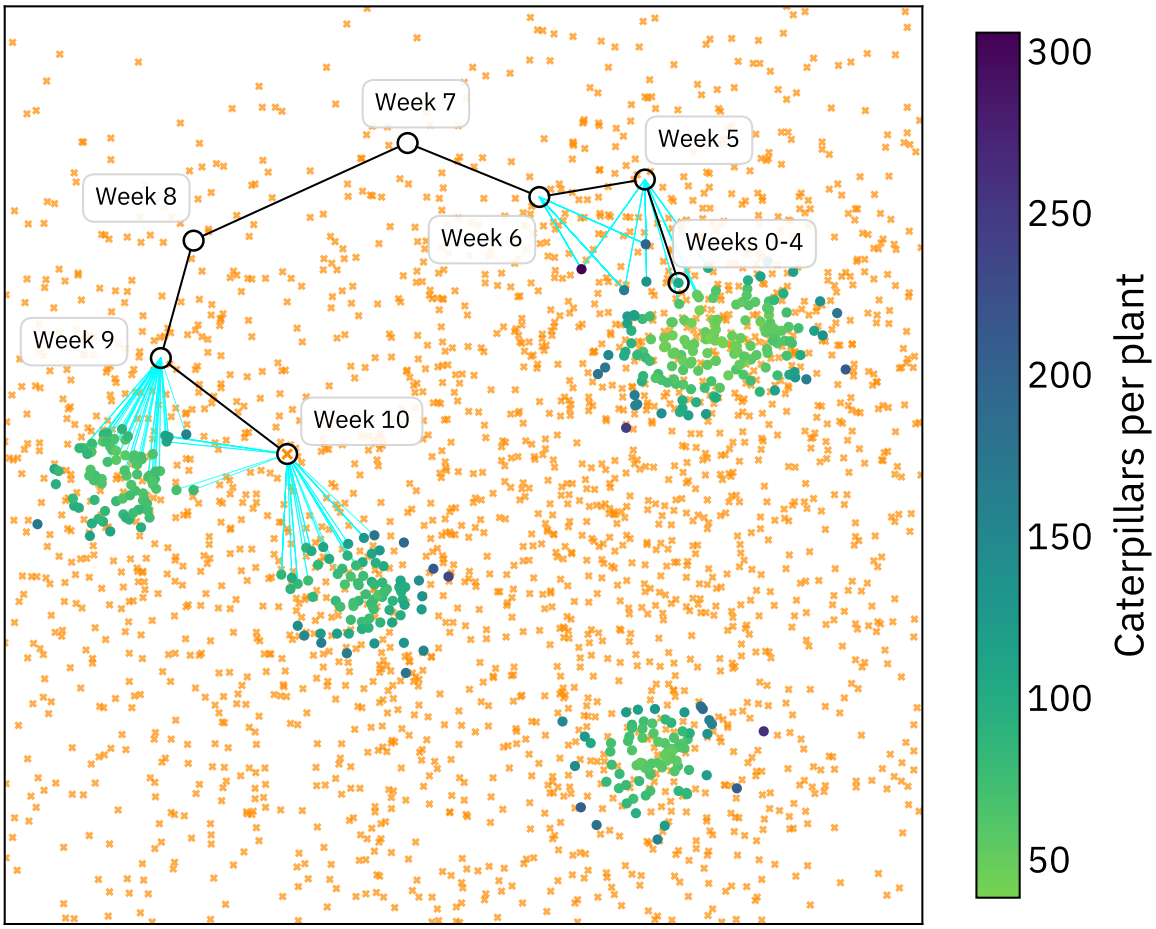



Figure 11: A portion of the landscape with butterflies (orange) and milkweed (green). There are about 40,000 caterpillars and 400 milkweed plants in this area. Each plant provides sufficient resources to allow an average of two caterpillars to survive to adulthood per week. A total of 2,680 butterflies are present in this area. The life history of a single individual is tracked from its origin on a particular milkweed plant, with black lines depicting dispersal, and teal lines showing where this individual laid eggs.

from SLiM without serious effort. Increasing computational efficiency and flexibility of simulation engines are bringing individual-based simulations closer to realistic models of the ecology of specific systems. Careful implementation of ecologically realistic evolutionary models will be important to many applied fields, such as understanding and predicting how climate change affects organisms' ranges, predicting the consequences of a gene-drive release in the wild, and rescuing species close to extinction. As the scale and specificity of *in silico* models improves, individual-based simulations will become valuable tools beyond academic research for management professionals in conservation, management, and public health.

**Modeling density dependence** Setting out to write this paper, we hoped to provide a comprehensive yet simple guide to best practices in implementing density-dependent population regulation. Although we have provided one or two paths forward and elucidated many of the issues (see in particular Appendix B), careful empirical practitioners will soon encounter additional questions. What are some flexible and robust families of functional forms, and what aspects of these matter in practice? How can these be parameterized so that parameters naturally correspond to observable/interpretable quantities? How can these be fit to data? How should local habitat quality and density interact? At first, we imagined that the answers would be found in familiar names, and so the relationship between local density and fitness would be described as

logistic, Beverton–Holt, Ricker, *et cetera*. However, we quickly found that these models were developed to describe population-level net changes, and so not only do not account for individual-level stochasticity, but furthermore do not separate birth from death. This is an area of active work – see, for instance, Aoyama et al. (2022) and Adler et al. (2018) for recent good examples. Full exploration of these questions was too much for this paper.

**Spatial data and niche modeling** Several of our case studies use environmental variables to specify where on the simulated landscape organisms are most likely to live. The explosion in remote-sensing data provides many potential data sources for modeling spatial heterogeneity. However, what is often needed in a model is a composite proxy for "suitability" that can be incorporated into local demographics. The process of predicting where a species might or does live is known as Ecological (or, Environmental) Niche Modeling (Booth et al., 1988; Peterson, 2001). This can be done in a variety of ways; for instance, one might model either the potential or the realized niche, and predict probabilities of occurrence or population densities (reviewed in Sillero, 2011). Environmental niche models are often used to predict suitable habitat either in other locations or other time periods (Werkowska et al., 2017; Yates et al., 2018), but resulting estimates can vary widely in quality and there are a number of statistical pitfalls (Sillero and Barbosa, 2021) that simulation testing could diagnose and simulation-based inference could potentially help avoid.

While ecological niche modeling uses observation data to predict where organisms might live, a collection of landscape-genetics tools try to use genetic relatedness to predict where organisms move. The methods generate a map of "landscape resistance" that aims to describe how easily individuals move over different parts of the landscape (McRae et al., 2008). However, resistance models rely on correlating genetic distance to an abstract notion motivated by electrical circuits (reviewed in Peterson et al., 2019; Cruzan and Hendrickson, 2020). They lack an underlying mechanistic model, so estimates can be problematic in practice (Cushman et al., 2013; Graves et al., 2013), and are expected to mislead in plausible situations such as biased dispersal (Lundgren and Ralph, 2019). Again, simulation-based inference provides a promising route forward (Smith et al., 2023, 2024), since it does not rely on explicit likelihoods or other mathematical descriptions.

**Sampling** Any simulation study that wishes to make comparisons to real data needs to also consider the sampling effort that led to those data, and realistic simulation of many sampling schemes can be daunting. In practice, sampling can strongly affect results (for an example in population genetics, see Battey et al., 2020). However, it is relatively easy to assess the robustness of results to variations in sampling scheme. Furthermore, it is often possible to "over-sample" simulations: as done for instance in Smith et al. (2023), many simulated datasets can be obtained from each costly spatial simulation, simply by repeating the sampling effort (however, if sufficient simulations are not done, model performance will be poor). It would be useful to develop a standard set of tools that implement various sampling schemes for simulated spatial populations.

**The future** Although the spatial simulations we present here incorporate many more aspects of real organisms' lives than does the Wright–Fisher model, there are many things that we have not tried to explicitly model, such as seasonal migration, herding or flocking, territoriality, foraging strategies, microhabitat variation, broadcast spawning, resource storing, pollination, predation, facilitation, and other inter-species interactions. Any of these can be modeled in SLiM with more or less effort, and indeed many are described in the SLiM manual (Haller and Messer, 2024). The decision of which aspects of biology to model is in practice made by prioritizing those aspects expected to substantially affect the question being studied. We are excited to see the wide variety of simulations that researchers develop in the future, as we explore these questions and build on each other's work.

# Acknowledgements

Many thanks to Jeremy Collings for useful and insightful comments on the manuscript. This work was funded in part by NIH/NHGRI grant R01 HG012473 to PLR, ADK, BCH, and PWM, NIH/NIGMS grants R35 GM148253 to ADK, R35 GM139628 to SR and ETC, and R35 GM152242 to PWM. ETC also acknowledges

support as a trainee under the Brown University Predoctoral Training Program in Biological Data Science (NIH T32 GM128596).

## Data availability

SLiM scripts suitable for re-use of all simulations used in this paper are at `https://github.com/kr-colab/spatial_sims_standard` (for SLiM v4.2 Haller and Messer, 2023). Scripts to produce the figures in this manuscript are available at at `https://github.com/kr-colab/spatial_sims`.

## Appendix A Example SLiM script

Here is a complete SLiM script for a spatial simulation with local Beverton-Holt regulation on mortality.

```
initialize() {
  initializeSLiMModelType("nonWF");
  initializeSLiMOptions(dimensionality="xy");

    defineConstant("SD", 0.3);         // sigma_D, dispersal scale
    defineConstant("SX", 0.3);         // sigma_X, interaction scale for measuring local density
    defineConstant("SM", 0.3);         // sigma_M, mate choice scale
    defineConstant("K", 5);            // carrying capacity per unit area
    defineConstant("FECUN", 0.25);     // mean number of offspring per time step
    defineConstant("WIDTH", 25.0);     // width of the simulated area
    defineConstant("HEIGHT", 25.0);    // height of the simulated area
    defineConstant("RUNTIME", 200);    // total number of ticks to run the simulation for

  // Set up constants that depend on externally defined parameters
  defineConstant("RHO", FECUN / ((1 + FECUN) * K));

  // basic neutral genetics
  initializeMutationRate(1e-8);
  initializeMutationType("m1", 0.5, "f", 0.0);
  initializeGenomicElementType("g1", m1, 1.0);
  initializeGenomicElement(g1, 0, 1e8-1);
  initializeRecombinationRate(1e-8);

  // spatial interaction for local density measurement
  initializeInteractionType(1, "xy", reciprocal=T, maxDistance=3 * SX);
  i1.setInteractionFunction("n", 1, SX);

  // spatial interaction for mate choice
  initializeInteractionType(2, "xy", reciprocal=T, maxDistance=3 * SM);
  i2.setInteractionFunction("n", 1, SM);
}

1 first() {
  sim.addSubpop("p1", asInteger(K * WIDTH * HEIGHT));
  p1.setSpatialBounds(c(0, 0, WIDTH, HEIGHT));
  p1.individuals.setSpatialPosition(p1.pointUniform(p1.individualCount));
}

first() {
  // preparation for the reproduction() callback
  i2.evaluate(p1);
}

reproduction() {
  mate = i2.drawByStrength(individual, 1);
  if (mate.size())
    subpop.addCrossed(individual, mate, count=rpois(1, FECUN));
}

early() {
```

```
  // Disperse offspring
  offspring = p1.subsetIndividuals(maxAge=0);
  p1.deviatePositions(offspring, "reprising", INF, "n", SD);

  // Measure local density and use it for density regulation
  i1.evaluate(p1);
  inds = p1.individuals;
  competition = i1.localPopulationDensity(inds);
  inds.fitnessScaling = 1 / (1 + RHO * competition);
}

late() {
  if (p1.individualCount == 0) {
    catn("Population went extinct! Ending the simulation.");
    sim.simulationFinished();
  }
}

RUNTIME late() {
  catn("End of simulation (run time reached)");
  // code for output might go here
  sim.simulationFinished();
}
```

# Appendix B Pitfalls

Even after carefully parameterizing a simulation to equilibrate near a given population density, it is fairly easy in practice to end up with a spatial simulation that mysteriously dies out or behaves oddly in other ways. A less dramatic annoyance is that usually the realized population size is not equal to the desired density, $K$, multiplied by the total area. This section describes the root causes of these issues and ways to diagnose them. The discussion gets into the weeds, so here is a summary of what to check if density is not what you expect (details below):

1. Visualize the simulation to check for odd dynamics or spatial patterns, such as a regular array of clumps.
2. Make sure the neighborhood sizes $N_X$, $N_M$, and $N_D$ are not too small (if in doubt, observe the effects of increasing $\sigma_X$, $\sigma_M$ and/or $\sigma_D$).
3. Look at the mean density *experienced by individuals*, not the total density across the landscape. If you want instead to set the total population size, you'll need a post-hoc adjustment to $K$ as in Box 6.
4. Make sure the *stage* you're measuring density in agrees with the theoretical calculations (*i.e.*, between birth and death or between death and birth).
5. Consider stochasticity: density varies randomly across the landscape, making the realized mean density differ from $K$.

Although it is natural to expect that the realized density of a simulation will be exactly the specified value of $K$, it is important to remember in practice that having a density different from $K$ is not necessarily a "problem": instead, it may reflect the natural biological consequences of the chosen model.

## Why are there weird regular clumps?

At its most extreme, a dispersal scale much smaller than the interaction scale can lead to strange, regular arrays of clumps. (Clumps may appear for many other reasons, but here we're talking about a regular, hexagonal grid of clumps.) Examples are shown in Figure S9. Although such regular patterns formed by this mechanism rare in nature, they are easy to accidentally produce in simulation (and are one reason it is important to visualize the simulation, as in Box 2). For discussion of this strange phenomenon, see Sasaki (1997), Etheridge et al. (2024), or the "Spatial competition and spatial mate choice in a nonWF model"

section of the SLiM manual (Haller and Messer, 2024). These are probably an indication that the dispersal or interaction scale are not well-chosen, but may indicate something more interesting.

## Why does my simulation run so slowly?

The runtime of an individual-based simulation is at least proportional to the total number of individuals. However, it is common for the runtime of spatial simulations to grow *more* than linearly in the number of individuals, because of spatial dynamics that involve a large number of pairwise comparisons or interactions (such as spatial mate choice and spatial competition). Performance problems resulting from this can often be diagnosed by looking for *large* neighborhood sizes: if $N_X$ (the interaction neighborhood size) or $N_M$ (the mating neighborhood size) are large, one may encounter slow run times. Happily, there are solutions that can often be applied.

Perhaps the most obvious solutions are to directly reduce the number of pairwise interactions. One way is to shrink the neighborhood sizes of the model, by reducing $\sigma_M$ and/or $\sigma_X$. However, that will often noticeably change the behavior of the model and sacrifice biological realism. Another way is to shrink the neighborhood size is to cut off the spatial kernel at a shorter distance, with little loss of exactitude; for a Gaussian spatial kernel, for example, cutting off at two standard deviations rather than three can reduce runtime with (perhaps) little change in dynamics, since interacting individuals 2–3 standard deviations from the focal individual interacted with that individual quite weakly anyway. However, this cannot cut runtime by more than about half, so for most models with large neighborhood sizes, another strategy is needed.

A second option is to use a "resource node" approach, as demonstrated in the "monarchs" example, which effectively mediates the many possible individual-individual interactions with a smaller number of interactions between each individual and a nearby node (representing a localized amount of resources). The approach is discussed more fully in Champer et al. (2024).

A third option is to use spatial map operations to approximate the pairwise interactions more efficiently. Effectively, we can pre-compute a smoothed version of the map of density in each tick of the model, then use that map to look up the density for each individual as a summary of all of the pairwise interactions it receives (Box 8). It turns out the same method can be used to efficiently pick nearby mates as well (Box 9). This option is in fact equivalent to using a regular grid of resource nodes, and can be proven to be a good approximation: see Appendix F.

Figure S1 shows that using the map-based approximation methods described in Boxes 8 and 9 makes it much easier to scale simulations to much higher neighborhood sizes. A naive implementation of pairwise interactions would result in runtimes that are quadratic in total population size, and hence totally infeasible for all but very small populations. Standard pairwise interactions in SLiM use efficient data structures ($k$-d trees) and a maximum cutoff (here, of $3\sigma$), but still compute all pairwise interactions out to the maximum distance, and so are quadratic instead in neighborhood size ($N_X$ or $N_M$), shown as solid lines in Figure S1. The spatial map-based methods, on the other hand, scale linearly with neighborhood size. These faster methods are approximate, but correspond closely, especially at high densities, as shown in Figure S2.

All three options have pros and cons, and may alter the behavior of the model. If the natural dynamics of the species are mediated through discrete locations (e.g., feeding or locations or mating sites), then adapting the resource node method is probably the most natural method. (In fact, it has been suggested that uncommon insects gather in discrete locations to find mates for not dissimilar reasons (Alcock, 1987).) If not – and so implementing resource nodes would require inventing new aspects of the model – it may be more natural to use the map-based methods to implement the existing model. Finally, it may be necessary (especially for development purposes!) to simply model a smaller landscape.

## Why does my simulation die out?

There are a variety of reasons why a simulation might die out (or have far fewer individuals than you expect). For instance, this can happen if $\sigma_M$, $\sigma_X$, or $\sigma_D$ are too small. In all cases, "too small" can be diagnosed by looking at the relevant neighborhood sizes: for example, if $N_M = K\pi\sigma_M^2$ is small (less than about 1), there may be a problem related to $\sigma_M$. This problem manifests as individuals being unable to reproduce because they cannot find a mate. This is particularly likely to happen if the mating scale has been made smaller to reduce runtime (see the previous section for discussion). Solutions might be either to increase $\sigma_M$, to allow

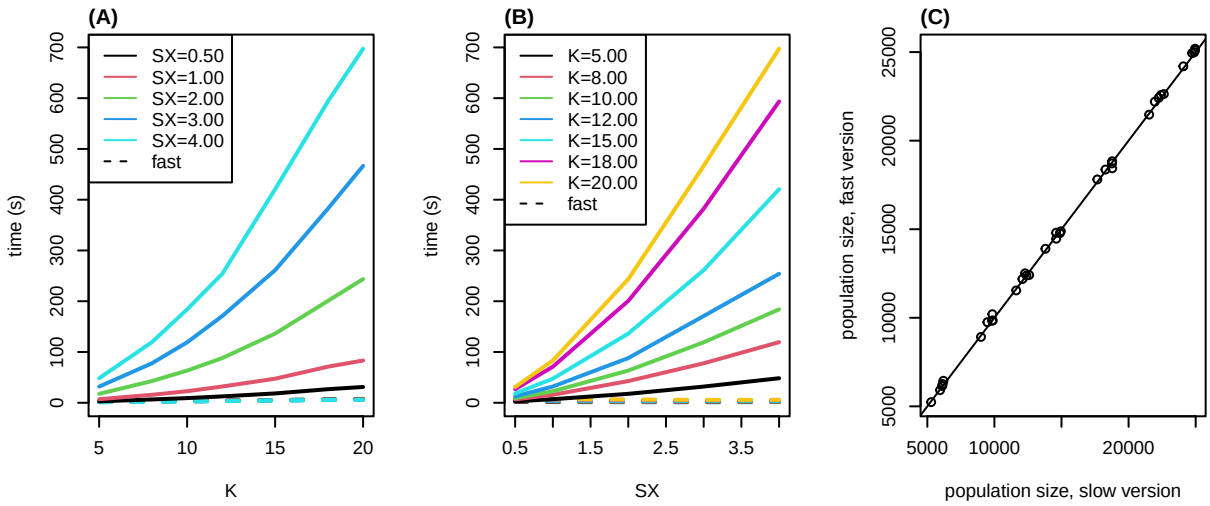



Figure S1: Runtimes for a "minimal" model with local mate choice and density-dependent Beverton–Holt feedback on mortality, plotted against **(A)** carrying capacity ($K$) and **(B)** interaction scale ($\sigma_X$, written `SX`). Runtimes are shown for models that do both mate choice and local density computations (solid lines) using (pairwise) interactions, and (dotted lines, all overlapping) using spatial maps, as described in Boxes 8 and 9. Also, **(C)** final population sizes after 100 time steps for the same combinations of $K$ and $\sigma_X$ shown in (A) and (B), for otherwise equivalent models that use either pairwise interactions (horizontal axis, "slow version"), or spatial map methods (vertical axis, "fast version").

selfing, or to increase the fecundity of those individuals that do reproduce, depending upon the biology of the system being modeled. The result also depends on the mating scheme, as shown in Figure S7.

The reasons that small $\sigma_X$ can be a problem are more subtle. Since $\sigma_X$ determines the range over which density is computed, and each focal individual itself counts towards its local population density, then if $\sigma_X$ is sufficiently small even a single isolated individual can have "local density" greater than the carrying capacity. This effect is demonstrated in Figure S8, in which the population dies out for small $\sigma_X$. The effect also appears in Figure S7, in which selfing simulations die out at low $K$ (and hence low $N_X$) – since they self, they are not dying out due to small $N_M$. Conceptually, this happens if the simulated individuals cannot range over a large enough area to obtain sufficient resources for survival, even in the absence of competition. Since the density of a single individual calculated by equation (1) is $\rho(0)/\sigma_X^2$, then this will occur if $\rho(0)/\sigma_X^2$ is close to or greater than $K$, *i.e.*, roughly if $\sigma_X \leq 1/\sqrt{K}$. This could be avoided by excluding the focal individual from the calculation of local density; however, this leads to the opposite problem: densities can get much larger than $K$. This happens because if $\sigma_X \leq 1/\sqrt{K}$, a single neighbor within scale $\sigma_X$ will lead to a "local density" of more than $K$ – but further away neighbors are unaffected. If local density increases mortality, a cartoon version of the situation is that an individual with a neighbor within distance $\sigma_X$ is killed, but neighbors further away than this are ignored; as a result, the density equilibrates to around $1/\sigma_X^2$, rather than $K$. (The above discussion is in two dimensions; in a one-dimensional model the equilibrium density would be around $1/\sigma_X$.)

The population can die out if $\sigma_D$ (and/or $\sigma_V$) is too small for similar reasons: conceptually, if offspring do not disperse far enough from their parents and local competition is strong, then families reduce their own fitness. For a simple example, suppose that the probability that a new offspring with local density $n$ survives is $e^{-n/K}$ and $\sigma_D$ is much less than $\sigma_X$. Then, a group of $m$ siblings form a clump of $m+1$ individuals with their parent, with density at least $(m+1)\rho(0)/\sigma_X^2$; so, the expected number of surviving offspring is smaller – at most $m\exp(-(m+1)\rho(0)/\sigma_X^2)$ – and the smaller number should be used in calculations of net reproductive output. However, if new individuals move sufficiently far (with $\sigma_V$) before local density effects, then the effect may be avoided.

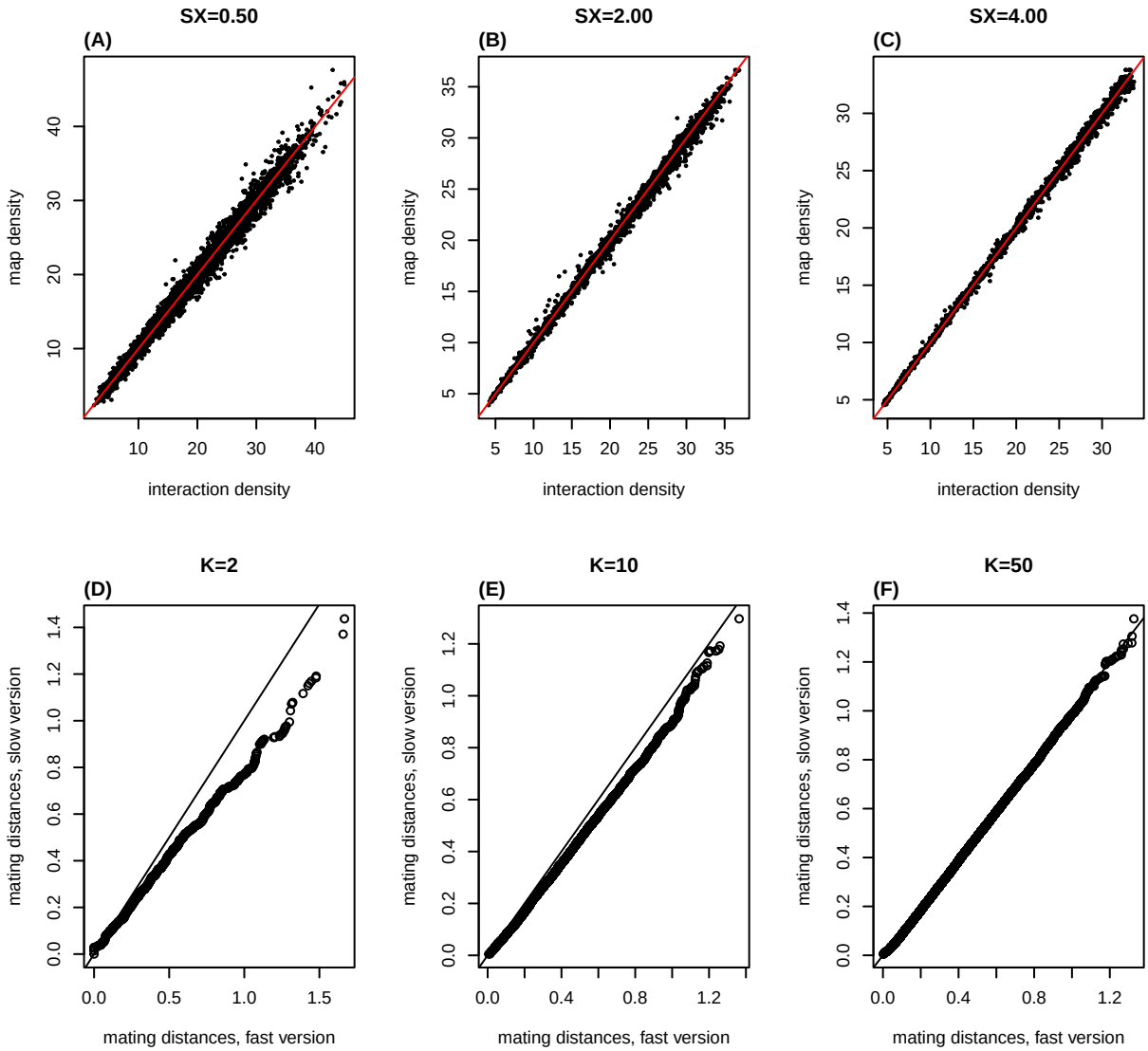



Figure S2: **(A-C)** Local density around each of $10^4$ individuals, measured both using pairwise interactions (equation (1)) and interpolation on a spatial map (Box 8). Each panel shows $10^4$ individuals randomly sampled from a separate simulation of the type described in Figure S1, but with $K$ increasing linearly over 20 time steps, and different values of $\sigma_X$ (labeled `SX`). **(D-F)** Q-Q plots comparing realized distributions of mating distances between simulations as above but either using individual-based mate choice (horizontal axis, "slow version") or using map-based mate choice as in Box 9 (vertical axis, "fast version"), at three values of density ($K$). Shown are the quantiles of the distances to the most recent mate for roughly 30,000 individuals in each simulation; at lower densities, the "fast version" tends to have slightly longer mating distances.

The effects of changing $\sigma_X$, $\sigma_D$, and removing the focal individual from density calculations are shown in an example in Figure S10.

## Why is the realized density not equal to $K$?

We've carefully set things up so that the equilibrium density in a neutral, spatially homogeneous simulation "should be" equal to $K$. If the realized density is very different (*e.g.*, zero) then the problem is probably one of the pitfalls described above. But even if $N_X$, $N_M$, and $N_V$ are not small, realized density still often differs from $K$ by 20% or 30%. Perhaps the simplest reason is "edge effects", but we assume the range is large enough these are unimportant (and, in practice SLiM computes local population density in such a way that local density is unaffected by edges). Another simple reason could be that density differs at different points in the life cycle – see below for more discussion of this.

First, we need to consider: *which* realized population density should we compare to $K$? The first answer that might spring to mind is "number of individuals divided by total area"; however, what matters for equilibrium is the density *experienced by individuals*. In other words, to see theoretical predictions playing out, we should measure local population density for each individual, and average that across individuals – that is after all the density that matters to the dynamics. This is seen in Figure S8, where mean density around individuals is shown in blue and number of individuals divided by area is shown in red, as well as in Figure S7, where mean density around individuals is shown in dotted lines and number of individuals divided by area in solid. Conceptually, if at equilibrium the simulation is very patchy (so individuals tend to be bunched up), then the mean density experienced by individuals could be much higher than the number of individuals divided by total area. In fact, the mean density around individuals is almost always be higher than the number of individuals divided by total area.

This line of reasoning leads to the second point: across individuals, local population density is a distribution, not a single value. It turns out that this stochasticity can also be important. Conceptually, the decrease in net reproduction of individuals with higher than average density may not be balanced by those individuals with lower density; how this happens depends on the shape of $F(u)$ and the shape of the distribution of densities. Below, we work through both these issues in more detail.

**Mean density around individuals** Why is the density experienced by individuals higher than the number of individuals divided by total area? Concretely, suppose that $u_i = n(x_i)/K$ is the scaled local density for individual $i$; the expected change in population size is zero if $\sum_i F(u_i) = 0$. Equivalently, if $U$ is the scaled local density for a randomly chosen individual, equilibrium occurs if $\mathbb{E}[F(U)] = 0$. Suppose instead that we look at the distribution of local densities across *space* instead of across individuals. Heuristically, suppose that we divide the landscape up into many small regions, each of area $\epsilon$, and let $v_j$ be the scaled density in region $j$ (*i.e.*, $n(y_j)/K$ for some point $y_j$ in region $j$, and the regions are small enough that density is constant within each). The number of individuals in region $j$ is $\epsilon K v_j$, and so the net contribution to the next step's population size from region $j$ is $\epsilon K v_j F(v_j)$. This equilibrium occurs if $\sum_j v_j F(v_j) = 0$. Equivalently, if we let $V$ denote the local density around a uniformly chosen point on the landscape, then equilibrium occurs if $\mathbb{E}[VF(V)] = 0$. In fact, the relationship between $U$ and $V$ is that $U$ is a *size-biased* draw from $V$; the relationship between the two is that $\mathbb{E}[f(U)] = \mathbb{E}[Vf(V)]/\mathbb{E}[V]$ for *any* function $f$. In particular, $\mathbb{E}[U] = \mathbb{E}[V^2]/\mathbb{E}[V] > \mathbb{E}[V]$ (by Jensen's inequality), and $\mathbb{E}[V]$ is just the total number of individuals divided by the total area (except for some edge effects).

Sharp-eyed readers of this and the next section will notice that we are sweeping something under the rug: when we measure density using equation (1), we do not include the focal individual. Taking this into account properly when defining $V$ is much less clean, so for illustrative purposes we have omitted this. In fact, if individual locations are independent and uniformly distributed, then mean density around individuals (measured without the focal individual!) is equal to the number of individuals divided by total area. Nonetheless, we think the calculations are informative.

**Stochasticity** Now we can immediately see how stochasticity interacts with nonlinearity in density dependence to increase or decrease equilibrium density. First suppose that $F$ is convex and decreasing, *i.e.*, $F''(u) > 0$ and $F'(u) \leq 0$ for all $u$. Then by Jensen's inequality, $\mathbb{E}[F(U)] > F(\mathbb{E}[U])$, and since at equilibrium, $\mathbb{E}[F(U)] = 0$, we have that $F(\mathbb{E}[U]) < 0$. Since we've assumed that $F(1) = 0$ and $F$ is decreasing, this implies that $\mathbb{E}[U] > 1$, *i.e.*, a convex $F$ *increases* the equilibrium mean density experienced by individuals above $K$. By the same argument if $F$ is concave, $\mathbb{E}[U] < 1$.

We can make the same argument for the total density, $V$: if $G(v) = vF(v)$ is convex, then $\mathbb{E}[V] > 1$,

while if $G$ is concave then $\mathbb{E}[V] < 1$. This at first seems odd: if $F(u)$ is convex and $uF(u)$ is concave, then the mean density experienced by individuals is *higher* than $K$, while the total density is *lower* than $K$. However, this is perfectly possible, and in fact seen in Figure S8.

A Taylor expansion lets us estimate more precisely the deviation of equilibrium size from $K$. Taylor expanding $F(u)$ about $u = 1$, we get that

$$\begin{aligned} 0 = \mathbb{E}[F(U)] &\approx \mathbb{E}\left[F(1) + (U-1)F'(1) + \frac{1}{2}(U-1)^2 F''(1)\right] \\ &= (\mathbb{E}[U]-1)F'(1) + \frac{1}{2}\operatorname{var}[U]F''(1) + \frac{1}{2}(\mathbb{E}[U]-1)^2 F''(1). \end{aligned}$$

If the deviation is small (*i.e.*, $|\mathbb{E}[U]-1| \ll |F'(U)/F''(U)|, (\mathbb{E}[U]-1)^2 \ll \operatorname{var}[U]$), the second order term is negligible. Thus, we can write:

$$\mathbb{E}[U] \approx 1 - \frac{1}{2}\operatorname{var}[U]\frac{F''(1)}{F'(1)}, \tag{11}$$

when $\frac{1}{2}\operatorname{var}[U] \ll \left(\frac{F'(1)}{F''(1)}\right)^2$. Since $F'(1)$ is negative, agrees with the argument above. Figure S11 shows that this prediction bears out well in practice (in two non-spatial models), as long as the population does not go extinct.

Similarly,

$$\begin{aligned} 0 = \mathbb{E}[F(U)] &= \mathbb{E}[VF(V)] \\ &\approx \mathbb{E}\left[1F(1) + (V-1)(F(1) + 1F'(1)) + \frac{1}{2}(V-1)^2(2F'(1) + 1F''(1))\right] \\ &\approx (\mathbb{E}[V]-1)F'(1) + \frac{1}{2}\operatorname{var}[V](2F'(1) + F''(1)), \end{aligned}$$

and hence

$$\mathbb{E}[V] \approx 1 - \operatorname{var}[V]\left(1 + \frac{F''(1)}{2F'(1)}\right),$$

which is smaller than $\mathbb{E}[U]$, and further away from 1, except in extreme circumstances.

What determines the strength of stochasticity? Since this has to do with random variation in "experienced" density across the landscape, stochasticity goes down as interaction neighborhood size $N_X$ increases. In other words, if $\sigma_X$ is larger, then we measure density averaging over larger areas, which is therefore less variable. As in equation (11), stochasticity affects equilibrium by a factor proportional to $\operatorname{var}[U]$. If $Y$ is the number of individuals within distance $\sigma_X$ of a random individual, then $\mathbb{E}[Y]$ is around $N_X$, and if noise is Poisson then also $\operatorname{var}[Y] \approx N_X$. Since $U$ is obtained by dividing $n(x)$ (from equation (1)) by $K$, and $n(x)$ is roughly $Y/\pi\sigma_X^2$, we expect $\operatorname{var}[U]$ to be of order $\operatorname{var}[Y]/(\pi\sigma_X^2 K)^2 = N_X/N_X^2 = 1/N_X$. So, we expect the deviation of realized density from $K$ to be of order $1/N_X$. Again, this is seen in Figure S8: the form of density dependence has a convex $F(u)$, and so for smaller values of $\sigma_X$, the value of $\mathbb{E}[U]$ (blue line) is above $K$. Conversely, the function $uF(u)$ is concave, and so $\mathbb{E}[V]$ (red line) is increasing, but is well below $K$ for other reasons.

The difference between the mean density around individuals ($\mathbb{E}[U]$) and the mean density by area ($\mathbb{E}[V]$, or total population size divided by total area) is well-known: the ratio $\mathbb{E}[U]/\mathbb{E}[V]$ is a scale-dependent measure of clustering known as *mean crowding* (Lloyd, 1967) that increases the more clustered individuals are on the scale used to measure local density.

**Density measurement timing** The other thing to consider is: *when* is density being measured? In each time step there are some births and some deaths; we follow SLiM in taking births first in the time step, but since the two alternate, this choice seems arbitrary. However, having density effects or movement occur in one or the other stage can affect the model (e.g., Taylor, 2010). Following SLiM as we do, the most common

"population size" is at the end of the time step, *i.e.*, after deaths (or equivalently, before births). However, population size may also be measured (and used!) between births and deaths.

Concretely, suppose that $N_t$ is the population size at the start of time step $t$, and $N_t^+$ is the population size after births in time step $t$. Above, we did calculations like this: if the mean fecundity in time step $t$ is $f_t$, and the mean probability of survival is $1 - \mu_t$, then

$$\mathbb{E}[N_t^+] = N_t(1 + f_t) \tag{12}$$

$$\mathbb{E}[N_{t+1}] = N_t^+(1 - \mu_t) \tag{13}$$

$$= N_t(1 + f_t)(1 - \mu_t). \tag{14}$$

*Which* density is used to determine $f_t$ and $\mu_t$? Naturally, $f_t$ can't depend on the number of births, so it will use $N_t$ (*i.e.*, the density at the start of the time step). However, should the density dependence for $\mu_t$ use $N_t$ or $N_t^+$? If the newborn individuals contribute to density, then $\mu_t$ should depend on $N_t^+$, *i.e.*, density computed using the offspring as well. However, this will be larger than the density at the start of the time step by a factor of $1 + f$, and so the equation for the local net per capita reproductive rate analogous to equation (2) is

$$F^+(u) = (1 + f(u))(1 - \mu(u(1 + f(u)))) - 1. \tag{15}$$

As before, equilibrium would be around density $n_*$ solving $F^+(n_*/K) = 0$, and so to arrange as before for the equilibrium density to be around $K$ we'd like to set up the functional forms so that $F^+(1) = 0$. One way to do this is to start with functional forms for $f(u)$ and $\mu(u)$ so that $F(u) = (1+f(u))(1-\mu(u))-1$ satisfies $F(1) = 0$ (*i.e.*, formulated for measuring density between death and birth), and then define the survival probability to be $1 - \mu^+(u) = 1 - \mu(u/(1 + f(1)))$. Then $F^+(u)$ defined using $\mu^+(u)$ satisfies $F^+(1)$. This is the approach taken by Battey et al. (2020). Another approach is to compute a density *map*, as in Box 8, at the start of each time step, and use that map to determine density for juveniles as well.

# Appendix C Parameterization of density dependence

If one wants to use a familiar phenomenological model as the basis for density dependence, there are several popular choices for the function form of $F(u)$. Here, we summarize these, parameterized so that $F(1) = 0$ (and so will have an equilibrium near $n_* = K$):

- $F(n) \propto 1 - n$, (Discrete-logistic model)
- $F(n) \propto \frac{1+a}{1+an} - 1$, (Beverton–Holt model)
- $F(n) \propto \frac{(1+a)^c}{(1+an)^c} - 1$, (Hassell model)
- $F(n) \propto e^{r(1-n)} - 1$. (Ricker model)

(Here, $\propto$ indicates that each can be scaled by a constant, reflecting an overall time scaling.)

Using each one still requires a number of choices. Next, we work through in more detail the steps involved in arranging birth and death rates so that the net birth rate, $F$, has a particular functional form, and give a number of examples that help to show the issues involved. Suppose here that each time step has birth followed by death; death applies in the same way to individuals just born as those previously alive; the mean fecundity of an individual with local density $N$ is $f(N/K)$; and the probability of death of an individual with local density $N$ is $\mu(N/K)$.

Roughly, the net change in population size when at scaled density $u = N/K$ is $F(u) = f(u)(1 - \mu(u)) - \mu(u)$. However, this is not not necessarily right, since it depends when the densities are measured: if the density for mortality is measured after birth, then the density passed to $\mu$ will be different than that passed to $f$. So, we'll always define $F(u)$ to be the mean per-capita change in population size across one step when starting at scaled population size $u$. However, which point in the time step (*i.e.*, after birth and before death or vice-versa) is the reference point will depend on the situation. Our goal is to figure out how to arrive at a given functional form for $F$ in various scenarios.

**Fecundity regulation:** If death probability is constant: $\mu(u) = \mu_0$, then taking $u$ to be the scaled population density before birth, simply

$$F(u) = f(u)(1 - \mu_0) - \mu_0,$$

and so

$$f(u) = \frac{F(u) + \mu_0}{1 - \mu_0}.$$

Since $0 \leq f < \infty$, for this to make sense we need $\mu_0 > 0$ and $F \geq -\mu_0$.

**Beverton–Holt fecundity regulation:** With $F(u) = \alpha((1 + a)/(1 + an) - 1)$, this is

$$f(u) = \frac{1}{(1 - \mu_0)} \left( \alpha \frac{(1 + a)}{(1 + au)} + (\mu_0 - \alpha) \right), \tag{16}$$

and we need $\alpha \leq \mu_0$.

**Ricker fecundity regulation:** With $F(u) = \alpha(e^{r(1-u)} - 1)$, this is

$$f(u) = \frac{1}{1 - \mu_0} \left( \alpha e^{r(1-u)} + (\mu_0 - \alpha) \right),$$

and we again need $\alpha \leq \mu_0$.

**Mortality regulation:** Suppose instead that fecundity is constant: $f(u) = f_0$, and that we measure density for mortality *after* birth (so it includes the new births). Then

$$F(u) = f_0 - \mu(u)(1 + f_0),$$

and so the survival probability is

$$1 - \mu(u) = \frac{1 + F(u)}{1 + f_0}.$$

Since we must have $0 \leq \mu \leq 1$, we require that $-1 \leq F \leq f_0$. Note that in this model (regardless of the form of $F$), the death probability at the stationary point ($u = 1$, since $F(1) = 0$) is $\mu(1) = 1/(1 + f_0)$; and so the mean lifetime is $1 + f_0$. Also note that this produces an equilibrium density of $K$ *before death* (as opposed to before birth, in the previous models); the density will be lower after death.

**Beverton–Holt mortality regulation:** With $F(u) = \alpha((1 + a)/(1 + au) - 1)$, this is

$$1 - \mu(u) = \frac{1}{(1 + f_0)} \left( \alpha \frac{(1 + a)}{(1 + au)} + (1 - \alpha) \right), \tag{17}$$

and we need $\alpha a \leq f_0$.

**Ricker mortality regulation:** With $F(u) = \alpha(e^{r(1-u)} - 1)$, this is

$$1 - \mu(u) = \frac{1}{1 + f_0} \left( \alpha e^{r(1-u)} + (1 - \alpha) \right),$$

and we need $\alpha(e^r - 1) \leq f_0$.

**Mortality and fecundity:** Suppose now we'd like both mortality and fecundity to change with density. So that there's only one density, let's say again that is measured after birth and before death. If the density at this time is $N$, then the mean number of individuals in the next time step is

$$N(1-\mu(N/K))(1+f(N/K)),$$

so that the mean net change is

$$F(u) = (1-\mu(u))(1+f(u)) - 1.$$

Given a desired functional form for $F(u)$ and $\mu(u)$ we would then define

$$f(u) = \frac{F(u)+1}{1-\mu(u)} - 1.$$

Note that for fecundity to remain finite, we need $\mu(u)$ to be bounded away from zero. On the other hand, if we have $f(u)$ then

$$1-\mu(u) = \frac{F(u)+1}{1+f(u)}.$$

**Mixed Beverton–Holt:** Suppose we set $f(u) = f_0/(1+bu)$ and would like $F(u) = \alpha((1+a)/(1+au)-1)$ for some constants $b$ and $\alpha$. Then, we would set

$$1-\mu(u) = \frac{\alpha(1+a)/(1+au) + (1-\alpha)}{f_0/(1+bu)+1}, \tag{18}$$

and we need $\alpha a < f_0$.

**Mixed Ricker:** Now suppose that we set $1-\mu(u) = (1-\mu_0)e^{-su} + \mu_\infty(1-e^{-su})$ and would like $F(u) = \alpha(e^{r(1-u)}-1)$. (The extra parameters are not unnecessary complications: we will need $\alpha < 1$ and $\mu_\infty > 0$ for the following to work out.) Then, we would set

$$f(u) = \frac{\alpha e^{r(1-u)} + (1-\alpha)}{(1-\mu_0)e^{-su} + \mu_\infty(1-e^{-su})} - 1,$$

and we need $s \le r$ for this to remain positive. This has $f(0) = (\alpha(e^r-1)+1)/(1-\mu_0)-1$, which is positive if $r > 0$ (already a requirement). Also, $f(\infty) = (1-\alpha)/\mu_\infty - 1$, so we also need $1-\alpha > \mu_\infty$.

**Hassell with mortality regulation** Let's set up the Hassell, which is $F(u) = b((1+a)/(1+au))^c - 1)$. Note that $F(0) = b((1+a)^c-1)$, $F(\infty) = -b$, and $F'(1) = cab(1+a)^c$. Setting fecundity to be constant at $f_0$ and plugging in to the expression for mortality regulation above,

$$\mu(0) = \frac{f_0 - b((1+a)^c-1)}{1+f_0}$$

and

$$\mu(\infty) = \frac{f_0+b}{1+f_0}.$$

So, if we'd like to fix $\mu(0) = \mu_0$ and $\mu_\infty = \mu(\infty)$, then these determine $a$ and $b$:

$$\begin{aligned} b &= \mu_\infty(1+f_0) - f_0 \\ a &= \left(\frac{f_0 - \mu_0(1+f_0)}{\mu_\infty(1+f_0)-f_0} + 1\right)^{1/c} - 1. \end{aligned}$$

This leaves us with a parameterization in terms of the mean fecundity, $f_0$, the death rate at low density, $\mu_0$, the death rate at high density, $\mu_\infty$, and the exponent $c$ that controls how steep the curve is between. (We could reparameterize $c$ so we have a parameter that is exactly $F'(1)$, but this is less compelling.)

**Ricker parameterization with fecundity regulation** If $\mu(u) = \mu_0$ and we want $F(u) = C(e^{r(1-u)} - 1)$, then

$$f(u) = \alpha e^{-ru} + \beta, \tag{19}$$

where letting $\alpha = Ce^r/(1-\mu_0)$ and $\beta = (\mu_0 - C)/(1-\mu_0)$. So, $f(0) = \alpha + \beta$ and $f(\infty) = \beta$. If we set $C = \mu_0$ then this is $\beta = 0$ and $\alpha = \mu_0 e^r/(1-\mu_0)$, and so $f(u) = \mu_0 e^{-r(u-1)}/(1-\mu_0)$.

**Regulation by juvenile mortality** Now suppose that adult death rate and fecundity are constant (so, $f(u) = f$ and $\mu(u) = \mu$), but that the probability of survival of *juveniles* is density-dependent. (Perhaps density dependence only affects the species during seedling recruitment.) So, if we call $r(u)$ the probability of survival in the first year at scaled density $u$, the net per capita reproduction function is $F(u) = r(u)f - \mu$. To make this proportional to the Beverton–Holt form, we can set

$$r(u) = \frac{\mu(1+a)}{f(1+au)}, \tag{20}$$

which results in $F(u) = \mu((1+a)/(1+au) - 1)$ – *i.e.*, the Beverton–Holt form, scaled by $\mu$.

**Density regulation of both juvenile and adult mortality** Now suppose that the probability of survival of juveniles (*i.e.*, to their first year) is $r(n/K)$, not $1 - \mu(n/K)$. In this case,

$$F(u) = r(u)f(u) - \mu(u),$$

and so if we fix $f(u) = f$ then we have, for instance,

$$fr(u) = F(u) + \mu(u).$$

Suppose we want $F(u) = (1+a)/(1+au) - 1$ and $\mu(u) = bu/(1+bu)$; then we would have

$$\begin{aligned} fr(u) &= \frac{(1+a)}{(1+au)} + \frac{bu}{(1+bu)} - 1 \\ &= \frac{(1+a)}{(1+au)} - \frac{1}{(1+bu)} \\ &= \frac{a}{(1+au)} + \frac{(b-a)u}{(1+au)(1+bu)} \end{aligned}$$

With this definition, $r(0) = a/f$ and $r(\infty) = (b-a)/((b+a)f)$, so we need $a \le \max(b, f)$ (and some other conditions). Note that $r(\infty) > r(0)$ (*i.e.*, *increasing* recruitment with density) if $b - a > a(a+b)$: for instance, take $f = 1$ and $a = 1/4$ and $b = 3/4$.

# Appendix D Parameterization and sampling for dispersal kernels

When sampling a new random vector for displacement or dispersal, it is simplest to think in Cartesian coordinates: to draw the displacement as $(X, Y)$ where $X$ and $Y$ are independent draws from some distribution. This works well if $X$ and $Y$ are Gaussian, but plotting the resulting bivariate distribution quickly shows oddities: dispersers will tend to fall on either around the axes or around the diagonals, depending on the distribution chosen. In fact, displacements in orthogonal directions are independent *only* for the Gaussian distribution. To obtain a rotationally symmetric dispersal kernel, it helps to think in polar/spherical coordinates. This also brings up an issue of terminology: what do we call a given rotationally symmetric bivariate distribution? Natural choices are to name it after the shape of either (a) the profile, $X$, or (b) the distance, $R = \sqrt{X^2 + Y^2}$. These agree only in one dimension. We work through some examples below to make the underlying issues clear.

**Gaussian (and Rayleigh)** Imagine displacements along the $x$ and $y$ axes are sampled from the same Normal distribution with zero mean and variance of $\sigma^2$. The displacement in $x, y$ then has the multivariate normal density

$$p(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).$$

In polar coordinates,

$$p(r, \theta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right),$$

*i.e.*, the density only depends on $r$, and so the kernel is radially symmetric.

To find the distribution of $R$, we can find the cumulative distribution, $\mathbb{P}(r < R)$, and differentiate. The CDF can be found by integrating $p(x, y)$ over rings of circumference $2\pi r$ and infinitesimal width of $dr$:

$$\mathbb{P}\{R < s\} = 2\pi \int_0^s p(x, y) r dr, \tag{21}$$

$$= \int_0^s \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr. \tag{22}$$

Thus, the PDF of $R$ is $p(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$, which is a Rayleigh distribution. In summary,

$$X, Y \sim N(0, \sigma^2) \tag{23}$$

is equivalent to

$$\theta \sim \text{Uniform}(0, 2\pi), R \sim \text{Rayleigh}(\sigma) \tag{24}$$

By the same argument, in three dimensions the angular part would be uniformly distributed on the sphere, and $R$ has density proportional to $r^2 \exp(r^2/2\sigma^2)$.

**Student's $t$** We can do something similar with the Student's $t$ distribution, but it will be clear that we need to be careful with generalizing it to a higher dimension.

First, what is the Student's $t$ distribution? In one dimension, it is the distribution with density

$$p(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}, \tag{25}$$

where $\nu$ is the "degrees of freedom" and $\sigma$ is a scale parameter. If $\nu = 1$, it is a Cauchy distribution, and if $\nu \to \infty$, the distribution converges to a standard Normal distribution. The Student's $t$ distribution is a heavy tailed distribution: all moments of order $\nu$ or higher are not defined.

Now suppose we want to "use the Student's $t$" in two dimensions. A **wrong** way to do this is to sample $x$ and $y$ independently from the same $t$ distribution. If you do that, the joint distribution $p(x, y)$ is $p(x)p(y)$ due to independence, and it looks like

$$p(x, y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)^2}{\nu\pi\sigma^2\Gamma\left(\frac{\nu}{2}\right)^2} \left(\left(1 + \frac{x^2}{\nu\sigma^2}\right)\left(1 + \frac{y^2}{\nu\sigma^2}\right)\right)^{-(\nu+1)/2}. \tag{26}$$

If we plug in $x = r\cos\theta$ and $y = r\sin\theta$ as we did before with the Normal distribution, we realize that $p(x, y)$ depend on both $r$ and $\theta$! In other words, the distribution of $R$ depends on the angle, $\theta$: long distances will be more common in some directions than others.

It turns out that although there are many ways to generalize the $t$ distribution to more than one dimension, there is not a single standard way (Kotz and Nadarajah, 2004). Here are three possibilities for how we might choose a radially symmetric bivariate kernel $p(x, y)$:

1. The distribution of the distance, $R$, is Student's $t$.

2. The distribution of the distance along an arbitrary axis, $X$, is Student's $t$.

3. The shape of the kernel taken along a line through the origin is the Student's $t$ density.

These conditions are, equivalently, that (1) $p(r,\theta) \propto r^{-1}(1+r^2)^{-(\nu+1)/2}$ (where $p(r,\theta)$ is $p(x,y)$ is radial coordinates); (2) $\int p(x,y)dy \propto (1+x^2)^{-(\nu+1)/2}$; and (3) $p(x,0) \propto (1+x^2)^{-(\nu+1)/2}$. If we chose the first option, then we'd be compelled to call the bivariate Gaussian the "bivariate Rayleigh" distribution, so for consistency with the Gaussian, we'll discard that option. Although option (2) is perhaps more elegant, figuring out what the actual density is for an arbitrary kernel shape is more involved, so we have chosen to go with option (3) (both here and in SLiM).

So: to make a radially symmetric Student's $t$ kernel, we'd like a bivariate kernel $p(x,y)$ proportional to $p(\sqrt{x^2+y^2})$, where the second $p()$ is the univariate Student's $t$. With the normalization factor, for $\nu > 1$ this is

$$p(x,y) = \frac{(\nu-1)}{4\pi\nu\sigma^2}\left(1+\frac{x^2+y^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}. \tag{27}$$

The density of $R$, the displacement distance, is equal to $2\pi r p(r,0)$ .

$$p_R(r) \propto \frac{r}{\sigma}\left(1+\frac{r^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}. \tag{28}$$

Note that if $\nu = 1$ (*i.e.*, the Cauchy distribution), the integral $\int_0^\infty p_R(r)dr$ does not converge. This means $\nu = 1$ gives an invalid probability distribution, and so in two dimensional space, the number of degrees of freedom should be greater than 1. (More generally, in $d$ dimensions we'll need $\nu > d-1$.)

**General kernels** To generalize this, begin with a univariate probability distribution with density $f(x)$. Then, we define another distribution for $r$ by $p_R(r) \propto r^{d-1}f(r)$ if in $d$ dimensions. This is equivalent in two dimensions to defining a joint distribution of $x, y$ by $p(x,y) = f(\sqrt{x^2+y^2})$. To sample a random variable $(X,Y)$ in two dimension, we sample $R$ from $p_R(r)$, sample $\theta$ from the uniform distribution on $(0,2\pi)$, and set $X = R\cos\theta$ and $Y = R\sin\theta$. (More generally, we would choose the radial component to be uniform on the $d$-sphere; the easiest way to do that is to let $(X_1,\ldots,X_d) = R(Z_1,\ldots,Z_d)/\sqrt{Z_1^2+\cdots+Z_d^2}$, where $Z_1,\ldots,Z_d$ are independent standard Normal.) This is how the function `pointDeviated()` in SLiM returns a displaced locations from various rotationally symmetric kernel.

## D.1 Sampling from kernels with a covariance matrix

The multivariate Normal is not necessarily rotationally symmetric: it allows a general covariance matrix; if the covariance matrix is not a multiple of the identity, then the contours of the density are ellipses, not circles. How might we introduce covariance matrices to multivariate dispersal? The most convenient way to do this is to use *scale mixtures* of Normals, *i.e.*, just multiply a given multivariate Normal distribution (with some covariance matrix) by a random scaling, choosing the distribution of the random scaling appropriately. For instance, if we'd like a fat-tailed dispersal kernel whose level sets are ellipses rotated by an angle $\theta$ counterclockwise, we can write:

```
U = rgamma(1, NU/2, NU/2);
  dxprime = rnorm(1, 0, SD_X) / sqrt(U);
  dyprime = rnorm(1, 0, SD_Y) / sqrt(U);
pos = individual.spatialPosition + c(dxprime * cos(THETA) - dyprime * sin(THETA),
          dxprime * sin(THETA) + dyprime * cos(THETA));
if (p1.pointInBounds(pos)) {
  offspring = subpop.addCrossed(individual, mate);
  offspring.setSpatialPosition(pos);
}
```

Note that we use `pointInBounds()` to check boundary condition, and so the boundary is absorbing. Here we've generated a *scale mixture of Normals*, by dividing our (correlated) multivariate Gaussian by $\sqrt{U}$, where $U$ is Gamma($\nu/2$,$\nu/2$) distributed. This is in fact another (and arguably better) common definition of the "multivariate Student's $t$" (Kotz and Nadarajah, 2004).

# Appendix E Additional methodological details for case studies

## E.1 Temporal change: pikas

For computational efficiency, we focus on a 266 km$^2$ region of Rocky Mountain National Park (RMNP) in Colorado. This is a consequential choice: by simulating a restricted area of the species range, we could for example overestimate the probability of extinction by missing larger spatial scale population dynamics. In addition, the resulting genetic variation will certainly be affected by modeling a smaller, narrowly distributed population.

To regulate the population we at first aim for a uniform density of 250 individuals per km$^2$ throughout the habitat. This value is informed by the number of pika scat piles observed by Erb et al. (2014) in sites in the Rocky Mountains. For the expected lifetime we use 3.25 years, as reported by Smith (1974). Thus, we set $K = 250$ and fecundity equal to $\frac{1}{3.25}$.

We did not find a published value for competitive interaction distance for pikas, nor for mating distance. Therefore, we decided to use the same value for the spatial scales of competition, parent-offspring dispersal, and mating (*i.e.*, $\sigma_I = \sigma_D = \sigma_M$). Our assumption will introduce error if the competitive interaction scale in pikas is different than the dispersal scale, or if the mating scale varies from the other two values, which are both likely to be true. There is no adult movement in the pika model.

This simplifying assumption might be accurate, as pikas are territorial (Smith, 1974). To choose a value for $\sigma_X$ (and the other, shared interaction scales) we use as a starting point the value of 300 m from Smith (1974) which was the maximum reported distance traveled by juveniles. Assuming that parent-offspring dispersal is Gaussian distributed in each dimension, we calculated a $\sigma_X$ such that three standard deviations from the mean, Euclidean distance is $300 = \sqrt{\frac{\pi}{2}}\sigma_X + 3\sqrt{(2-\frac{\pi}{2})}\sigma_X$ (using formulas for mean and standard deviation of the Rayleigh distribution).

## E.2 Complex life cycles: mosquitoes

In our model of mosquitoes, we assume that only the juveniles' viability is affected by local density, while adult population size is regulated only through a constant mortality. Due to this detail, if we want to control the juvenile density to match carrying capacity, we need to modify the density control function (such as Beverton–Holt model) to reflect the life cycle.

To start our derivation, let's say density of juveniles with age $i$ is $a_i$ with $i = 1, ..., m-1$ where $m$ is maturation age. Let's also define density of adults as $a_m$. The adults have a fixed survival probability, $1 - \mu_a$. We use a variation of Beverton–Holt model where survival probability of juvenile population is a local population density factor, $\frac{1}{1+\rho u}$, multiplied by the baseline survival probability $1 - \mu_j$, where $\mu_j$ is the baseline mortality of juveniles. (here $u$ is population density of juveniles, *i.e.*, $\sum_{i=1}^{m-1} a_i$). Our goal is to find $\rho$.

Then we get a system of $m$ equations for $\mathbf{a}$ (left-hand-side is the $a$ in the next time step, but there is no time dependence by definition of equilibrium.):

$$a_1 = \frac{a_m FEC/2 \cdot (1-\mu_j)}{1 + \rho(\sum_{i=1}^{m-1} a_i + FEC/2 \cdot a_m)} \tag{29}$$

$$a_k = \frac{a_{k-1}(1-\mu_j)}{1 + \rho(\sum_{i=1}^{m-1} a_i + FEC/2 \cdot a_m)} \tag{30}$$

$$a_m = \frac{a_{m-1}(1-\mu_j)}{1 + \rho(\sum_{i=1}^{m-1} a_i + FEC/2 \cdot a_m)} + a_m(1-\mu_a), \tag{31}$$

where $k = 2, ..., m-1$. Notice that we have a factor of 2 for fecundity, FEC, because only female adults (assumed to be half of total adult population) produces FEC new individuals. $\rho$ controls where the equilibrium density is for juveniles, and it is an unknown for now. We want $\rho$ to make $\sum a_{i=1}^{m-1} = K$, where $K$ is carrying capacity of larvae.

Due to recursive relation between $a_i$'s, we can simplify the system of $m+1$ equations for $m+1$ unknowns ($a_i$'s and $\rho$) to two equations with two unknowns, $a_m$ and $\rho$.

To make it a little clear, let's define,

$$r \equiv \frac{1-\mu_j}{1+\rho(\sum_{i=1}^{m-1} a_i + FEC/2 \cdot a_m)} = \frac{1-\mu_j}{1+\rho(K + FEC/2 \cdot a_m)} \tag{32}$$

Now we start from the first equation of $a_1$,

$$a_1 = a_m FEC/2 \cdot r \tag{33}$$

and plug it into the next one to find $a_2$.

$$a_2 = a_m FEC/2 \cdot r^2 \tag{34}$$

and keep going to $a_{m-1}$. We see that $a_i = a_m FEC/2 \cdot r^m$, a nice geometric series. (This makes sense because we tend to see exponential distributed age-structure in simulations.) This is nice because the sum can be simplified very nicely

$$\sum_{i=1}^{m-1} a_i = a_m \cdot FEC/2 \cdot r \cdot \frac{1-r^{m-1}}{1-r} = K. \tag{35}$$

This is one of two equations we will need. The second one comes from plugging in $a_{m-1}$ to the original equation for $a_m$.

$$a_m = a_m(1-\mu_a) + a_m FEC/2 \cdot r^m. \tag{36}$$

Dividing both sides by $a_m$, and rearranging terms, we get

$$r = \left(\frac{\mu_a}{FEC/2}\right)^{1/m}. \tag{37}$$

Plugging this into the first equation, we get $a_m$! And we can find $\rho$ by plugging it into the equation we used to define $r$. Finally, we get $a_m$ and $\rho$ as a function of $a_m$ that we have in our simulation model:

$$a_m = K \frac{1-\left(\frac{\mu_a}{FEC/2}\right)^{1/m}}{1-\left(\frac{\mu_a}{FEC/2}\right)^{1-1/m}} \cdot \mu_a^{-1/m}(FEC/2)^{1-1/m} \tag{38}$$

$$\rho = \frac{\frac{(1-\mu_j)(FEC/2)^{1/m}}{\mu_a^{1/m}} - 1}{K + a_m \cdot FEC/2} \tag{39}$$

In Figure S12 we compare the expected adults to juveniles ratio from the equation above to the simulated value. Even though we didn't consider other factors like spatial aspects (finding mates, migrations, heterogeneity of the river map) and seasonal fluctuation, the simulated ratio stay pretty close to theoretical expectation. In addition, I also plot the average local population density that juveniles measure around themselves through an interaction kernel with width $SX = 20.0$ in Figure S13. The density closely follows the rain factor which is a sinusoidal function added everywhere on the map to set the baseline carrying capacity, except for when the rain factor is close to zero. During the "dry season", we see increase of density because we programmed the adult females to disperse offspring to the locations where carrying capacity is high within the maximum dispersal distance.

## E.3 Continental-scale systems: cane toads

Note that there has already been considerable effort to simulate the Australian cane toad invasion (Kearney et al., 2008), including simulations that incorporate genetic information to infer many biological parameters (Estoup et al., 2010).

Toads were simulated to have a juvenile state of 1 year. Individuals that were at least one year old were allowed to disperse once each year over their entire lifetime. Toads were randomly assigned to be male or female. Offspring initial locations were set to their mother's locations.

Previous studies have shown that the spread rate and distribution of cane toads has been influenced by environmental heterogeneity (Urban et al., 2008). We initially used a homogeneous landscape to model the cane toad invasion, which resulted in glaring discrepancies between the simulated and observed distributions. In efforts to increase the likeness of simulated distributions to the observed data, we incorporated environmental heterogeneity across space.

The kernels controlling competition and mate choice were kept as Gaussians.

The pipeline for downloading the complete data is available (`https://github.com/kr-colab/spatial_sims/blob/main/silas/range_expansion/pipelines/get_data.smk`), and is described here briefly. We downloaded cane toad occurrence data was from the Global Biodiversity Information Facility (`https://doi.org/10.15468/dl.8pukaa`), bioclimatic variables from WorldClim (`https://www.worldclim.org/data/bioclim.html` (Fick and Hijmans, 2017)), and a shapefile for the geographic perimeter of Australia from the Australian Bureau of Statistics (`https://www.abs.gov.au/`). We converted the locations from latitude and longitude values to km using using GeoPandas (Jordahl et al., 2020) (see notebook/vignette for further details), and set the origin of the map to the earliest location in the occurrence data.

**Extensions** Our approach could be extended to use multiple environmental maps along with a multivariate fitness function over each environmental dimension.

If desired, the simulations could be extended to include the ability to use and compare genetic information between observed and simulated data. SLiM's ability to model explicit genomes could also allow for more complexity and realism, such as a heritable component for dispersal ability to model assortative mating for dispersal ability, as well as gene surfing, where deleterious alleles are maintained at the range edge (Miller et al., 2020; Shine et al., 2021). Additional complexity could be incorporated into the life history traits as well. For example, cannibalism has been observed in cane toads (DeVore et al., 2021). This and other modifications could be made related to changes age related competition, fecundity, mortality, and establishment.

### E.4 Resource competition: monarchs

The milkweed is only minimally simulated within this model. A number of patch centroids are randomly spread across an area of the model, and a number of milkweed plants are scattered around each centroid according to a Gaussian distribution. During each year of the model, after the monarchs have migrated south for the winter, the locations of the patch centroids and plants are re-randomized.

## Appendix F Map-based approximations to density

In this section, we provide a formal argument showing that the approximation scheme of Box 8 converges to the local density for a fine enough grid. Conceptually, this works because the approximation effectively computes density as if all points were at the node of the region they are in. Since all the regions are small, this does not change things much. The argument applies to a "resource node" scheme as well, as long as the nodes are sufficiently dense.

Suppose that the positions of individuals on the landscape are recorded as a collection of points $\{x_i\}_{i=1}^N$. It will be helpful for notation to represent the state of the population as a point measure, $\Lambda = \sum_i \delta_{x_i}$. For simplicity in this section, suppose that distances are measured in units of the interaction scale, i.e., $\sigma_X = 1$. The density we would like to compute from equation (1) is then

$$n(x) = \sum_i \rho(x - x_i),$$

the convolution of $\Lambda$ with the kernel $\rho$. Now suppose that we placed a discrete set of nodes on the landscape at locations $\{y_j\}$, and for each $j$ let $A_j$ denote the portion of space that is closer to $y_j$ than to any other node. In other words, $\{A_j\}$ is the Voronoi tesselation associated with $\{y_j\}$; and suppose we assign the boundaries

between regions in some sensible way. Suppose that the diameters of all $A_j$ are less than $\epsilon$; we will show that using these nodes we can approximate $n(x)$ well to within an error that is proportional $\epsilon$ – so, finer meshes of nodes will make better approximations.

Suppose that we're evaluating density at the location of node $y_j$. The approximation outlined in Box 8 seeks to approximate $n(y_j)$ by

$$\hat{n}(y_j) = \sum_k \rho(y_j - y_k)\Lambda(A_k),$$

where $\Lambda(A_k)$ is the number of individuals within the region $A_k$, which we may write as $\Lambda(A_k) = \sum_i 1_{A_k}(x_i)$. So, we can write

$$\begin{aligned} n(y_j) - \hat{n}(y_j) &= \sum_i \left( \rho(y_j - x_i) - \sum_k \rho(y_j - y_k) 1_{A_k}(x_i) \right) \\ &= \sum_k \sum_{i: x_i \in A_k} \left( \rho(y_j - x_i) - \rho(y_j - y_k) \right). \end{aligned} \tag{40}$$

Now note that by the intermediate value theorem and the fundamental theorem of calculus,

$$\rho(y_j - x) - \rho(y_j - y_k) = (x - y_k)\rho'(y_j - y_k + \alpha(y_k - x))$$

for some $0 \leq \alpha \leq 1$. Now, $\rho(x) \to 0$ as $x \to \infty$, so for any $\delta$ we may pick $R$ so that

$$\int_{|x - y_j| > R - \epsilon} \rho'(x) dx < 1/N.$$

Write $N_R = \#\{i : |x_i - y_j| \leq R\}$ for the number of points closer than $R$. Furthermore, suppose the derivative of $\rho$ is bounded above by $C$, *i.e.*, $\|\rho'(x)\| \leq C$. If $|x - y_k| \leq \epsilon$, then

$$|\rho(y_j - x) - \rho(y_j - y_k)| \leq \epsilon C.$$

Since for $x_i \in A_k$, by definition $|x_i - y_k| \leq \epsilon$, plugging this into equation (40), and splitting the sum into regions with nodes further than $R$ away and from $y_j$ and not,

$$|n(y_j) - \hat{n}(y_j)| \leq \epsilon \left(1 + C N_R\right).$$

Since $C$ and $N_R$ are fixed, this goes to zero as $\epsilon \to 0$. This shows that $n(y_j) \approx \hat{n}(y_j)$; since $\hat{n}(x)$ is defined for arbitrary $x$ by interpolation, $n(x) \approx \hat{n}(x)$ as well.

# Supplementary figures

# References

Peter B. Adler, Stephen P. Ellner, and Jonathan M. Levine. Coexistence of perennial plants: an embarrassment of niches. *Ecology Letters*, 13(8):1019–1029, 2010. doi: https://doi.org/10.1111/j.1461-0248.2010.01496.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2010.01496.x`.

Peter B. Adler, Andrew Kleinhesselink, Giles Hooker, Joshua B. Taylor, Brittany Teller, and Stephen P. Ellner. Weak interspecific interactions in a sagebrush steppe? Conflicting evidence from observations and experiments. *Ecology*, 99(7):1621–1632, 2018. doi: https://doi.org/10.1002/ecy.2363. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2363`.

Stepfanie M. Aguillon, John W. Fitzpatrick, Reed Bowman, Stephan J. Schoech, Andrew G. Clark, Graham Coop, and Nancy Chen. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. *PLOS Genetics*, 13(8):e1006911, August 2017. ISSN 1553-7404. doi: 10.1371/journal.pgen.1006911. URL `https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006911`.

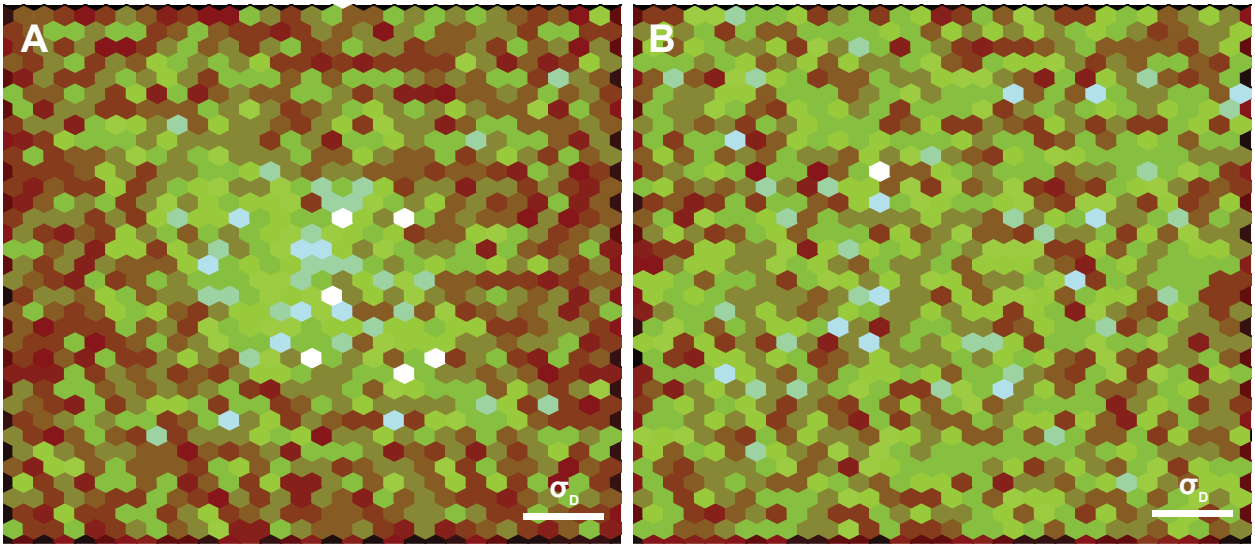



Figure S3: Fatter-tailed dispersal reduces clumping. Fry plots illustrate the distance between each pair of individuals at a given tick. The dense region of points near the origin in the Gaussian panel indicates many pairs of individuals are separated by small distances. Points shown are pairwise distances $\leq 1$ from an arbitrary time step from one replicate. Here, $\sigma_D = 0.3$. **(A)** Gaussian dispersal kernel. **(B)** Student's $t$ dispersal kernel.

John Alcock. Leks and hilltopping in insects. *Journal of Natural History*, 21(2):319–328, 1987. doi: 10.1080/00222938700771041. URL `https://doi.org/10.1080/00222938700771041`.

Lina Aoyama, Lauren G. Shoemaker, Benjamin Gilbert, Sharon K. Collinge, Akasha M. Faist, Nancy Shackelford, Vicky M. Temperton, György Barabás, Loralee Larios, Emma Ladouceur, Oscar Godoy, Catherine Bowler, and Lauren M. Hallett. Application of modern coexistence theory to rare plant restoration provides early indication of restoration trajectories. *Ecological Applications*, 32(7):e2649, 2022. doi: https://doi.org/10.1002/eap.2649. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/eap.2649`.

Frederic Austerlitz, Christopher W. Dick, Cyril Dutech, Etienne K. Klein, Sylvie Oddou-Muratorio, Peter E. Smouse, and Victoria L. Sork. Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, 13(4):937–954, 2004. doi: 10.1111/j.1365-294X.2004.02100.x. URL `http://dx.doi.org/10.1111/j.1365-294X.2004.02100.x`.

WWF Australia. 10 facts about cane toads, 2019. URL `https://wwf.org.au/blogs/10-facts-about-cane-toads/`.

Adrian Baddeley, Ege Rubak, and Rolph Turner. *Spatial Point Patterns: Methodology and Applications with R*. Interdisciplinary Statistics Series. Chapman and Hall/CRC, 2015. ISBN 978-1-4822-1020-0.

Nick H. Barton. The dynamics of hybrid zones. *Heredity*, 43(3):341–359, December 1979. ISSN 0018-067X. URL `http://dx.doi.org/10.1038/hdy.1979.87`.

Nick H. Barton, Frantz Depaulis, and Alison M. Etheridge. Neutral evolution in spatially continuous populations. *Theoretical Population Biology*, 61(1):31–48, February 2002. URL `http://dx.doi.org/10.1006/tpbi.2001.1557`.

C J Battey, Peter L Ralph, and Andrew D Kern. Space is the place: Effects of continuous spatial structure on analysis of population genetic data. *Genetics*, 215(1):193–214, 05 2020. ISSN 1943-2631. doi: 10.1534/genetics.120.303143. URL `https://doi.org/10.1534/genetics.120.303143`.

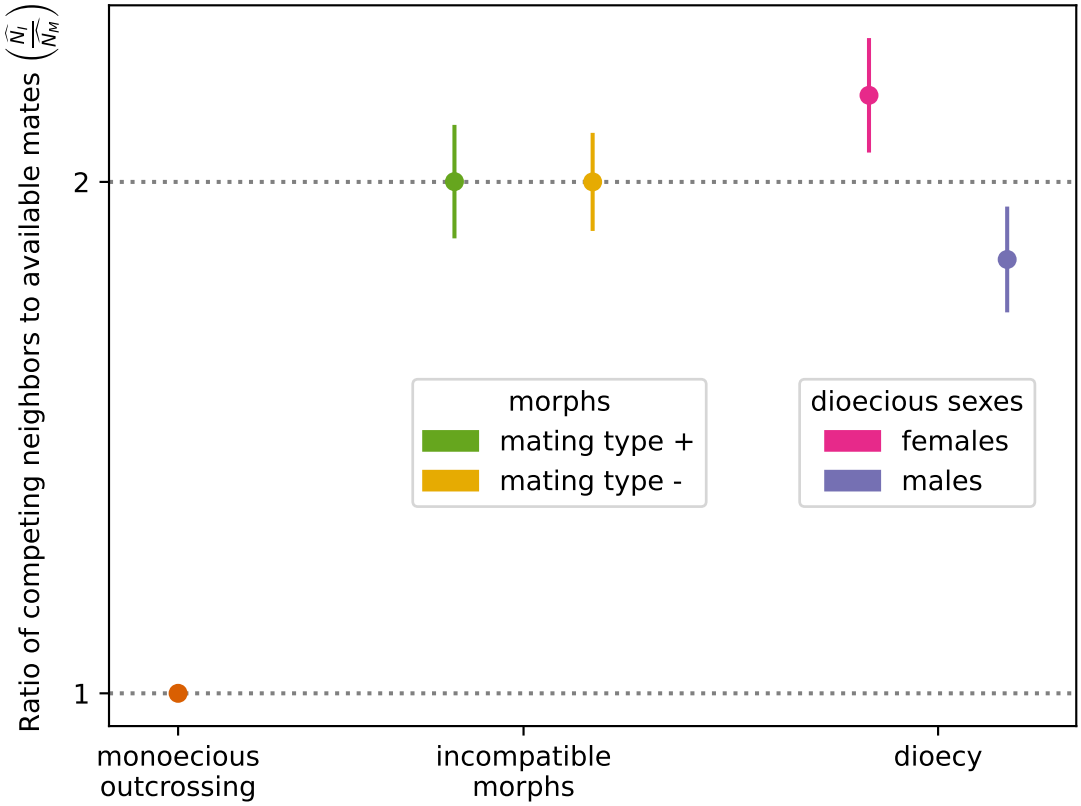



Figure S4: Ratio of competitive interactors to potential mating interactors differs within and among sexual mating systems. We calculate the mean number of competitors (within $3\sigma_X$) experienced by each individual, and divide by the mean number of potential mates (within $3\sigma_M$). Error bars show standard deviation of means across 50 independent time steps. Dashed lines show expected ratios of 1:1 for outcrossing among monoecious individuals, and 2:1 for two distinct mating types ($\pm$ for mutually incompatible morphs, or female/male for dioecy).

Erik A Beever, Chris Ray, Philip W Mote, and Jennifer L Wilkening. Testing alternative models of climate-mediated extirpations. *Ecological applications*, 20(1):164–178, 2010.

I. E. Bella. A new competition model for individual trees. *Forest Science*, 17(3):364–372, 09 1971. ISSN 0015-749X. doi: 10.1093/forestscience/17.3.364. URL `https://doi.org/10.1093/forestscience/17.3.364`.

Raymond J. H. Beverton and Sidney J. Holt. *On the Dynamics of Exploited Fish Populations*. Springer Netherlands, 1957. ISBN 9789401121064. doi: 10.1007/978-94-011-2106-4. URL `http://dx.doi.org/10.1007/978-94-011-2106-4`.

Malyon D. Bimler, Margaret M. Mayfield, Trace E. Martyn, and Daniel B. Stouffer. Estimating interaction strengths for diverse horizontal systems using performance data. *Methods in Ecology and Evolution*, 14(3):968–980, 2023. doi: https://doi.org/10.1111/2041-210X.14068. URL `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14068`.

W F Bodmer. Differential fertility in population genetics models. *Genetics*, 51(3):411–424, March 1965. ISSN 00166731. doi: 10.1093/genetics/51.3.411. URL `https://pubmed.ncbi.nlm.nih.gov/17248243`.

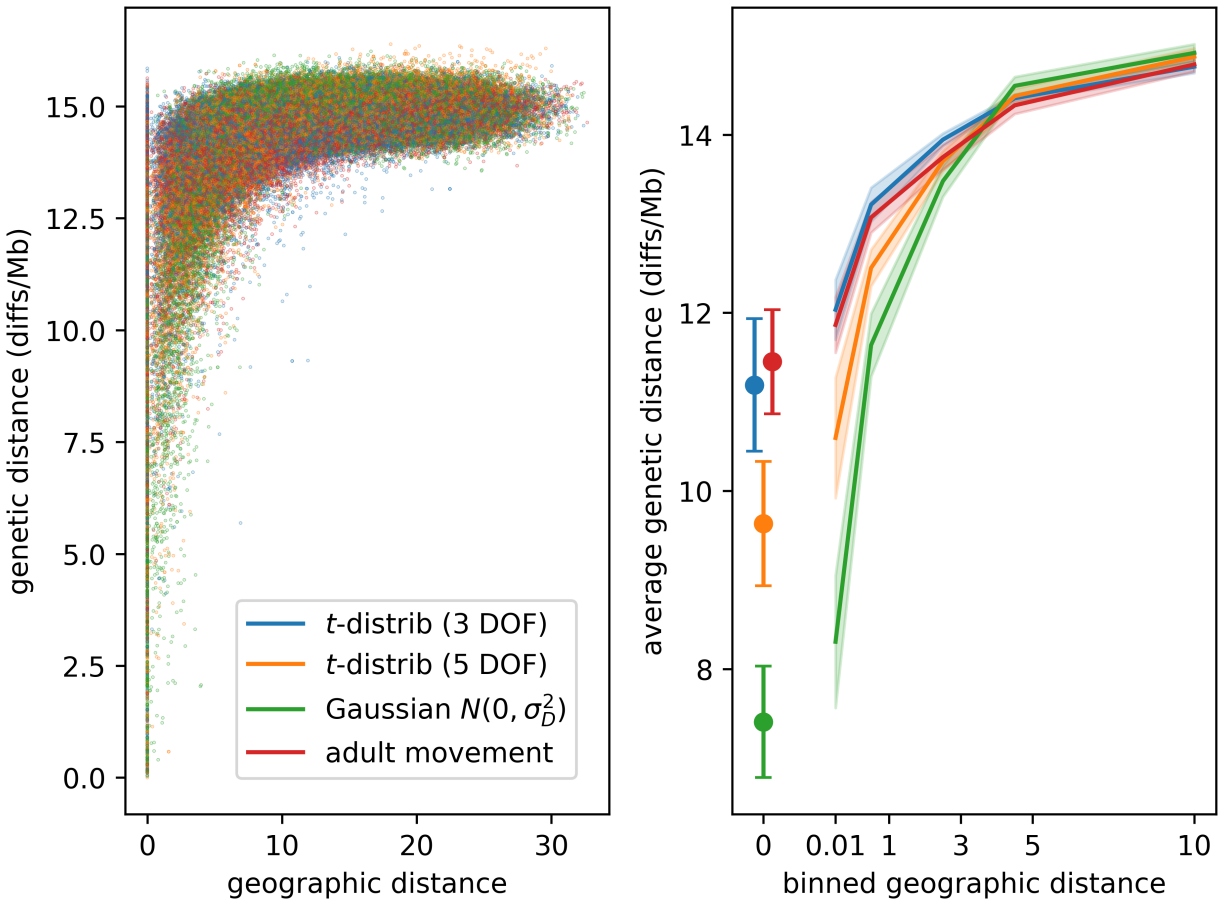



Figure S5: Fat-tailed dispersal and adult movement flatten genetic isolation-by-distance. Plots show genetic distance between pairs of individuals at increasing geographic distance. Points (A) and average curves (B) shown are from ten independent replicates. Error bars in (B) for pairs at $< 0.01$ units show one standard deviation. The scale of $t$-distributions is equal to the standard deviation of the normal distribution ($\sigma_D = 0.3$). Adult movement scenario uses Gaussian dispersal for movement at each time step.

B M Bolker and S W Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Pop. Biol.*, 52(3):179–197, 1997.

Trevor H. Booth, Henry A. Nix, Michael F. Hutchinson, and Tom Jovanic. Niche analysis and tree species introduction. *Forest Ecology and Management*, 23(1):47–59, 1988. ISSN 0378-1127. doi: https://doi.org/10.1016/0378-1127(88)90013-8. URL `https://www.sciencedirect.com/science/article/pii/0378112788900138`.

Gideon S. Bradburd and Peter L. Ralph. Spatial population genetics: It's about time. *Annual Review of Ecology, Evolution, and Systematics*, 50(1):427–449, 2019. doi: 10.1146/annurev-ecolsys-110316-022659. URL `https://doi.org/10.1146/annurev-ecolsys-110316-022659`.

Thomas Broquet and Eric J. Petit. Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):193–216, 2009. doi: 10.1146/annurev.ecolsys.110308.120324. URL `https://doi.org/10.1146/annurev.ecolsys.110308.120324`.

Michael L. Cain, Brook G. Milligan, and Allan E. Strand. Long-distance seed dispersal in plant populations. *Am. J. Bot.*, 87(9):1217–1227, 2000. URL `https://www.jstor.org/stable/2656714`.

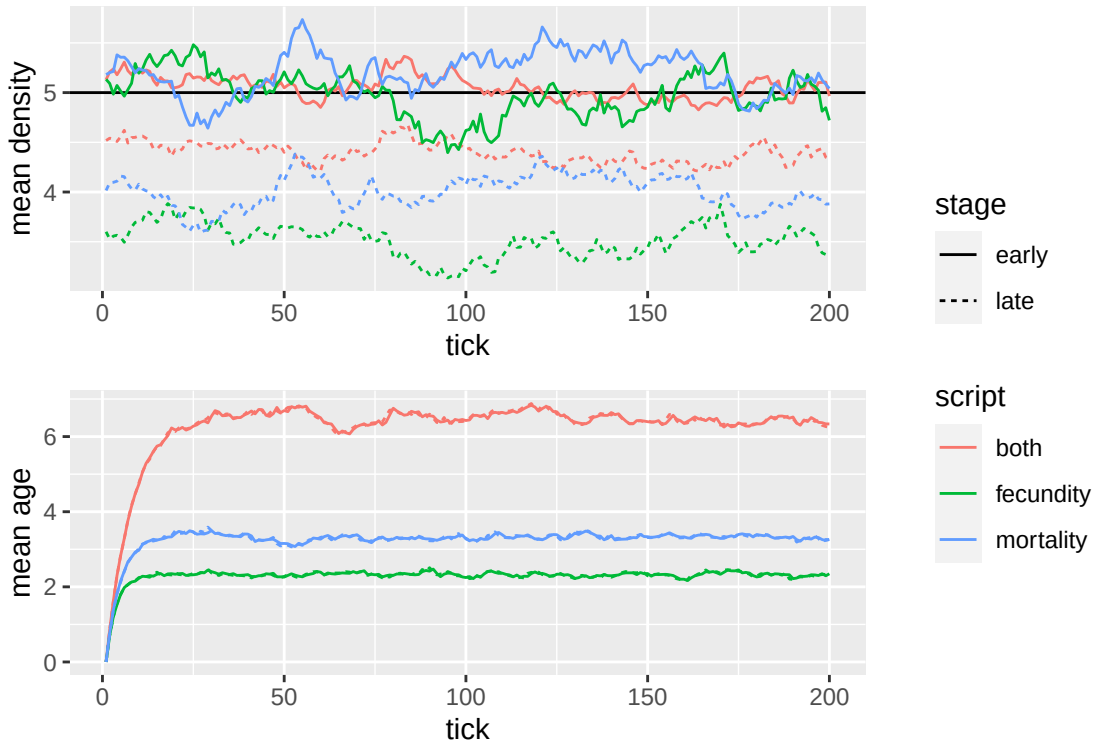



Figure S6: Traces from spatial simulations of the three "Beverton–Holt" models shown in Figure 1. **(Top)** Average local density for all individuals, measured both between birth and death ("early") and between death and birth ("late"), and **(bottom)** mean ages across 200 time steps. Horizontal line shows the value of $K = 5$; local population density determines population regulation between birth and death for all models. Other parameters: $\sigma_D = \sigma_X = 1.2$, and range was a $25 \times 25$ square area.

Charles D Canham, Philip T LePage, and K Dave Coates. A neighborhood analysis of canopy tree competition: effects of shading versus crowding. *Canadian Journal of Forest Research*, 34(4):778–787, 2004. doi: 10.1139/x03-232. URL `https://doi.org/10.1139/x03-232`.

Robert Stephen Cantrell and Chris Cosner. *Spatial ecology via reaction-diffusion equations*. John Wiley & Sons, 2004.

Hal Caswell. *Matrix population models*, volume 1. Sinauer Sunderland, MA, 2000.

Samuel E Champer, Nathan Oakes, Ronin Sharma, Pablo García-Díaz, Jackson Champer, and Philipp W Messer. Modeling CRISPR gene drives for suppression of invasive rodents using a supervised machine learning framework. *PLoS computational biology*, 17(12):e1009660, 2021.

Samuel E. Champer, Bryan Chae, Benjamin C. Haller, Jackson Champer, and Philipp W. Messer. Resource-explicit interactions in spatial population models. *bioRxiv*, 2024. doi: 10.1101/2024.01.13.575512. URL `https://www.biorxiv.org/content/early/2024/01/15/2024.01.13.575512`.

Brian Charlesworth, Deborah Charlesworth, and Nicholas H. Barton. The effects of genetic and geographic structure on neutral variation. *Annual Review of Ecology, Evolution, and Systematics*, 34(1):99–125, 2003. doi: 10.1146/annurev.ecolsys.34.011802.132359. URL `http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.34.011802.132359`.

Travis Child, Benjamin L Phillips, and Richard Shine. Does desiccation risk drive the distribution of juvenile cane toads (*Bufo marinus*) in tropical Australia? *Journal of Tropical Ecology*, 25(2):193–200, 2009.

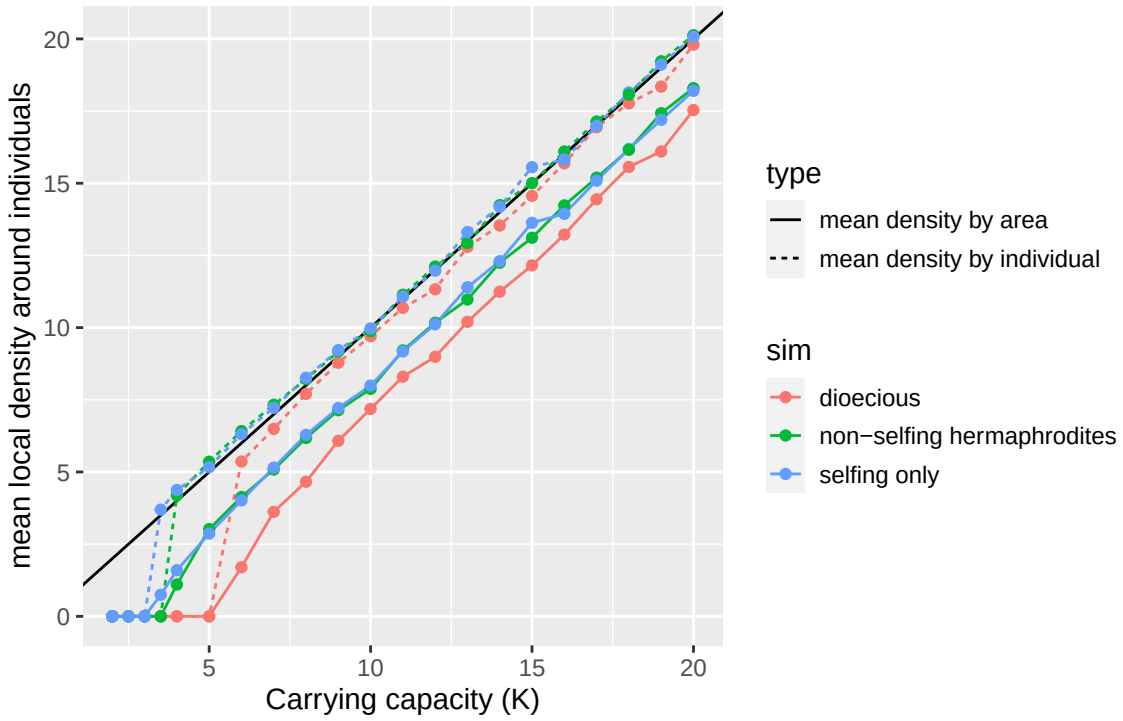



Figure S7: Average realized population density at equilibrium of three different models, plotted against the "carrying capacity" parameter, $K$. Each model uses Beverton–Holt regulation of mortality with the same parameters ($\sigma_D = \sigma_I = \sigma_M = 0.3$). and differ only in that "*dioecious*" individuals are one of two sexes: only females reproduce and only if they mate with a male; "*non-selfing hermaphrodites*" also must mate with another individual, but all individuals can reproduce; and "*selfing*" individuals can all reproduce and have no need for mating. Furthermore, mean fecundity is $f = 0.5$ for the dioecious simulations and $f = 0.25$ for the others. Dotted lines show mean local density experienced by individuals; solid lines show total population size divided by total area. All simulations die out at low density, but those requiring mating die out at higher $K$: in each case, when $N_M$ is around 0.4.

Jean Clobert, Michel Baguette, Tim G. Benton, and James M. Bullock. *Dispersal Ecology and Evolution*. Oxford University Press, 09 2012. ISBN 9780199608898. doi: 10.1093/acprof:oso/9780199608898.001.0001. URL `https://doi.org/10.1093/acprof:oso/9780199608898.001.0001`.

Martin P Cohen and Ross A Alford. Factors affecting diurnal shelter use by the cane toad, *Bufo marinus*. *Herpetologica*, pages 172–181, 1996.

Antonio-Juan Collados-Lara, Steven R Fassnacht, Eulogio Pardo-Igúzquiza, and David Pulido-Velazquez. Assessment of high resolution air temperature fields at Rocky Mountain National Park by combining scarce point measurements with elevation and remote sensing data. *Remote Sensing*, 13(1):113, 2020.

Joseph H Connell. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. *Dynamics of populations*, 298(312), 1971. URL `https://archive.org/details/connell-1971`.

T. Coulson, T. H. G. Ezard, F. Pelletier, G. Tavecchia, N. C. Stenseth, D. Z. Childs, J. G. Pilkington, J. M. Pemberton, L. E. B. Kruuk, T. H. Clutton-Brock, and M. J. Crawley. Estimating the functional form for the density dependence from life history data. *Ecology*, 89(6):1661–1674, 2008. doi: https://doi.org/10.1890/07-1099.1. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/07-1099.1`.

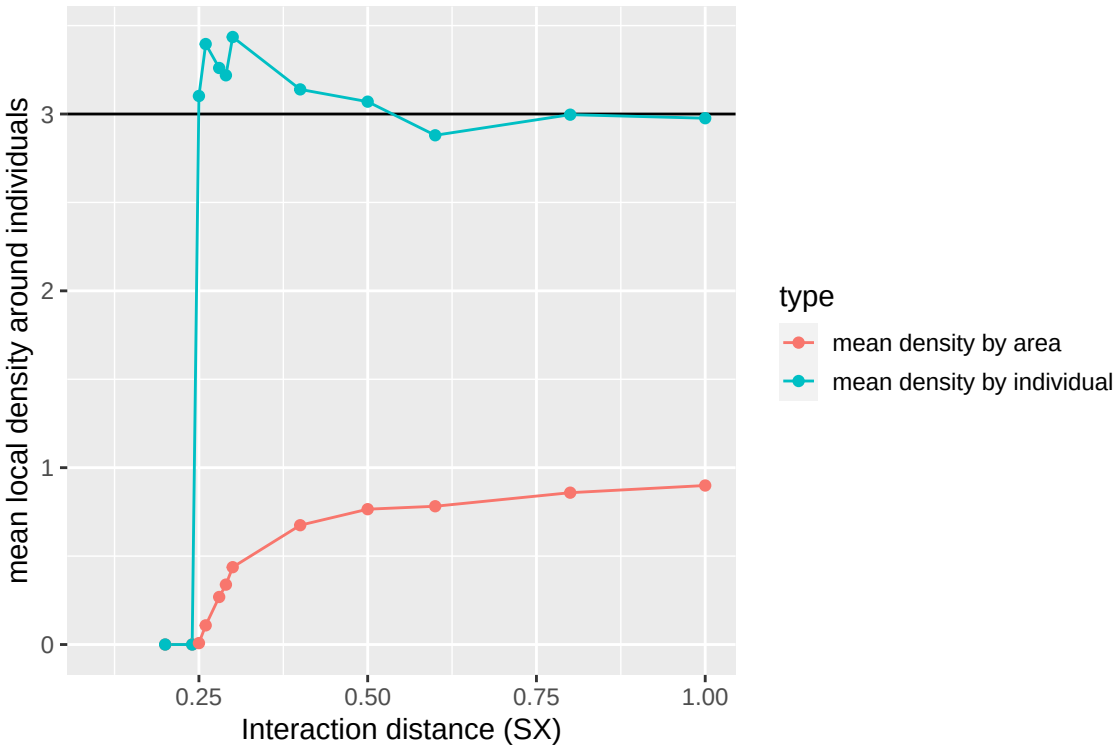



Figure S8: Mean density at equilbrium in simulations with different values of $\sigma_X$ (`SX` in the figure). The blue line shows mean density around individuals (*i.e.*, the average of local density calculated with equation (1) for all individuals), the red line shows total population size divided by area (a $25 \times 25$ box), and the horizontal line is at $K = 3$. The simulation uses Beverton–Holt control of mortality with $\sigma_D = 0.5$, reproduction entirely through selfing, and no adult movement. The population dies out if $\sigma_X$ is below 0.25, while for $0.25 \leq \sigma_X \leq 0.5$, total density (*i.e.*, density averaged by area) increases, while mean density experienced by individuals is larger than $K$ and decreases.

Franck Courchamp, Ludek Berec, and Joanna Gascoigne. *Allee effects in ecology and conservation*. OUP Oxford, 2008.

Mitchell B. Cruzan and Elizabeth C. Hendrickson. Landscape genetics of plants: Challenges and opportunities. *Plant Communications*, 1(6):100100, 2020. ISSN 2590-3462. doi: https://doi.org/10.1016/j.xplc.2020.100100. URL `https://www.sciencedirect.com/science/article/pii/S2590346220301280`.

Samuel A. Cushman, Andrew J. Shirk, and Erin L. Landguth. Landscape genetics and limiting factors. *Conservation Genetics*, 14(2):263–274, April 2013. ISSN 1572-9737. doi: 10.1007/s10592-012-0396-0. URL `https://doi.org/10.1007/s10592-012-0396-0`.

Richard F. Daniels. Simple competition indices and their correlation with annual loblolly pine tree growth. *Forest Science*, 22(4):454–456, 12 1976. ISSN 0015-749X. doi: 10.1093/forestscience/22.4.454. URL `https://doi.org/10.1093/forestscience/22.4.454`.

Cornelis Teunis De Wit. On competition. Technical Report 66.8, Instituot Voor Biologisch en Scheikundig Onderzoek Van Landbouwgewassen - Wageningen, 1960.

Donald L. DeAngelis and Simeon Yurek. Spatially explicit modeling in ecology: A review. *Ecosystems*, 20 (2):284–300, March 2017. ISSN 1435-0629. doi: 10.1007/s10021-016-0066-z. URL `https://doi.org/10.1007/s10021-016-0066-z`.

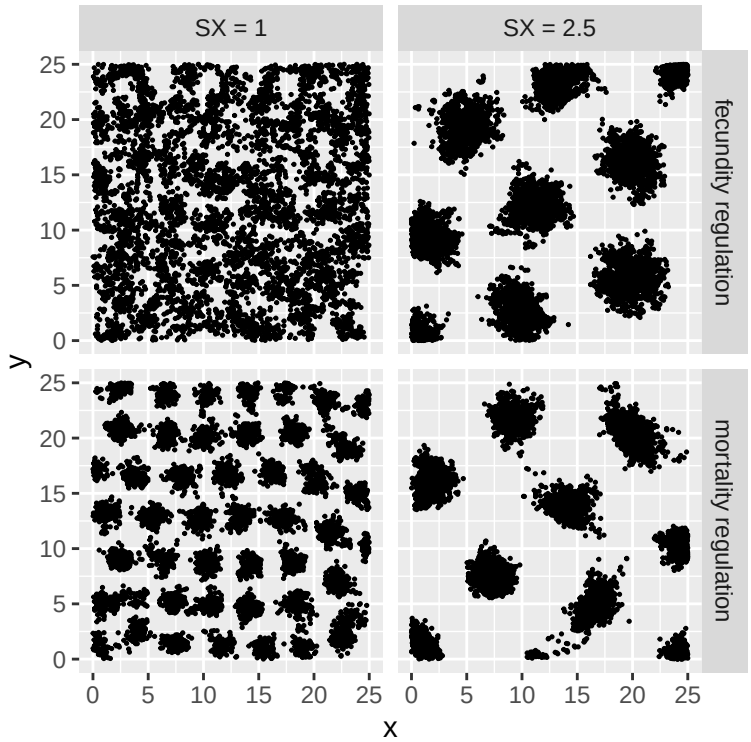



Figure S9: Examples of simulations exhibiting clumping, using $\sigma_D = 0.2$, two different values of $\sigma_X$ (labeled `SX`), and two different types of density-dependent feedback (either mortality or fecundity regulation, as in Figure S6).

Jayna L DeVore, Michael R Crossland, Richard Shine, and Simon Ducatez. The evolution of targeted cannibalism and cannibal-induced defenses in invasive populations of cane toads. *Proceedings of the National Academy of Sciences*, 118(35):e2100765118, 2021.

Ulf Dieckmann and Richard Law. Relaxation projections and the method of moments. In Dieckmann U, Law R, and Metz JA, editors, *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, pages 412–455. Cambridge University Press, 1999.

T Dobzhansky and S Wright. Genetics of natural populations. X. Dispersion rates in *Drosophila Pseudoobscura*. *Genetics*, 28(4):304–340, July 1943. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209213/`.

John B. Dunning Jr, David J. Stewart, Brent J. Danielson, Barry R. Noon, Terry L. Root, Roland H. Lamberson, and Ernest E. Stevens. Spatially explicit population models: Current forms and future uses. *Ecological Applications*, 5(1):3–11, 1995. doi: https://doi.org/10.2307/1942045. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1942045`.

P Edelaar and D I Bolnick. Non-random gene flow: an underappreciated force in evolution and ecology. *Trends Ecol Evol*, 27(12):659–665, December 2012. doi: 10.1016/j.tree.2012.07.009. URL `http://www.ncbi.nlm.nih.gov/pubmed/22884295`.

A M Edwards, R A Phillips, N W Watkins, M P Freeman, E J Murphy, V Afanasyev, S V Buldyrev, M G da Luz, E P Raposo, H E Stanley, and G M Viswanathan. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165):1044–1048, October 2007. doi: 10.1038/nature06199. URL `http://www.ncbi.nlm.nih.gov/pubmed/17960243`.

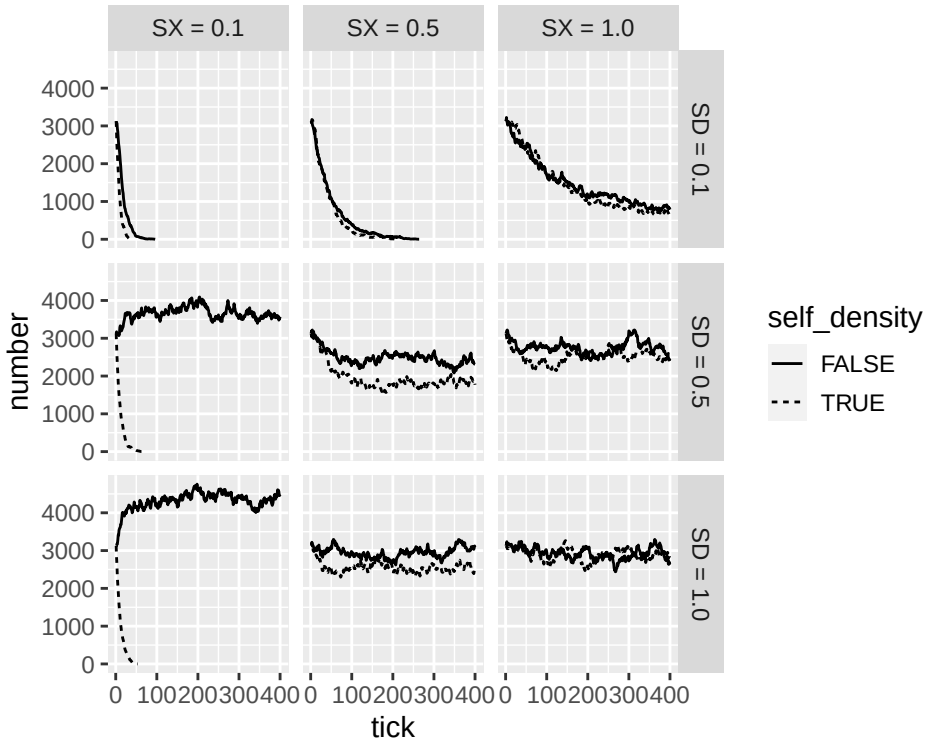



Figure S10: Population sizes through time for simulations with various values of dispersal scale ($\sigma_D$, here `SD`) and interaction scale ($\sigma_X$, here `SX`), with or without inclusion of the focal individual in local density calculations. Each simulation had Beverton–Holt density-dependent feedback on mortality (as in Figure S6), and was run with $K = 5$ on a $25 \times 25$ square area, and were started with $5 \times 25 = 3125$ individuals (so, lines that are roughly flat are fluctuating around a total density of $K = 5$). Solid lines compute "local density" for control of mortality of each individual using equation 1, while dotted lines do the same except excluding the focal individual.

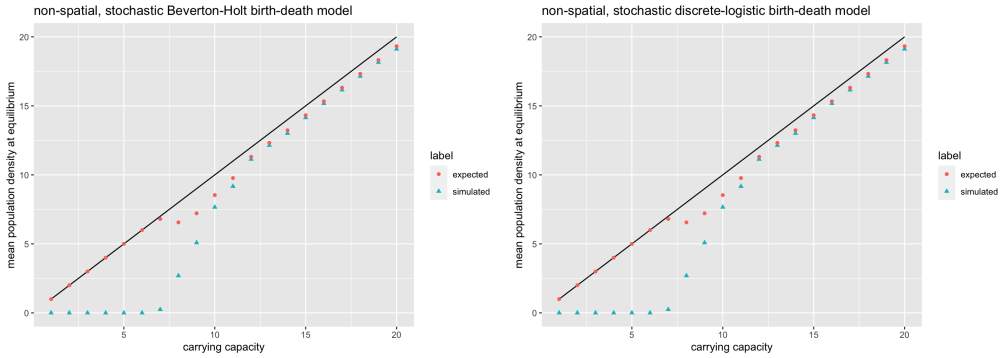



Figure S11: Expected population density obtained using equation (11) and actual average population density in non-spatial models with **(left)** Beverton–Holt regulation of mortality for which "expected" is $K - \frac{Var[U]}{(f+1)K}$, and **(left)** discrete logistic regulation of mortality, for which "expected" is $K - \frac{Var[Y]}{K}$. Non-spatial simulations were run in R. Solid line is where density is equal to carrying capacity, which is true for deterministic model.

Stephen P. Ellner, Dylan Z. Childs, and Mark Rees. *Data-driven Modelling of Structured Populations: A Practical Guide to the Integral Projection Model.* Springer International Publishing,

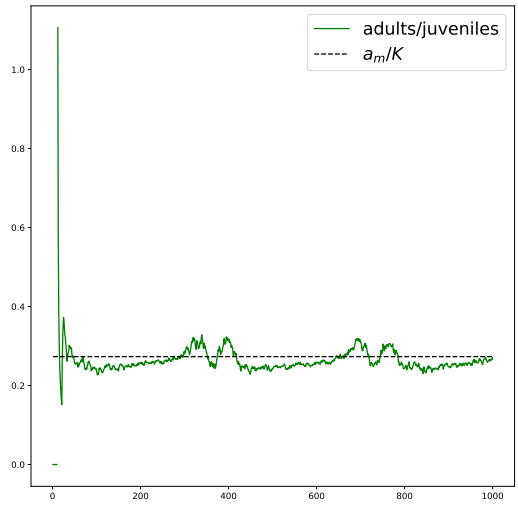



Figure S12: Ratio between adult and juvenile mosquito counts.

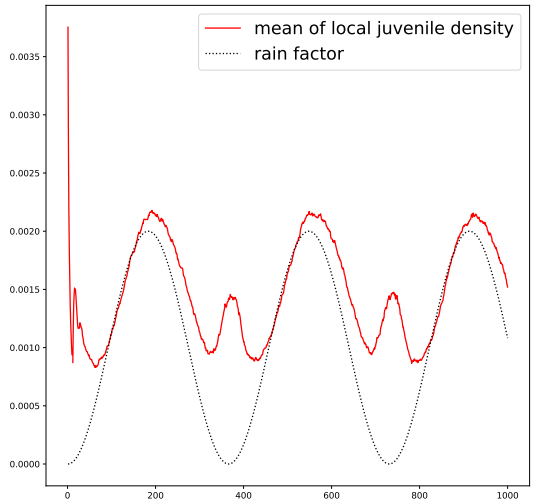



Figure S13: Local population density of juvenile mosquitoes plotted against rain factor, which is minimum carrying capacity at each time point.

2016. ISBN 9783319288932. doi: 10.1007/978-3-319-28893-2. URL `http://dx.doi.org/10.1007/978-3-319-28893-2`.

John A. Endler. *Natural Selection in the Wild*. Princeton University Press, April 1986. ISBN 978-0-691-08387-2.

B.K. Epperson. *Geographical Genetics*. Monographs in Population Biology. Princeton University Press, 2003. ISBN 9781400835621. URL `https://books.google.de/books?id=15p-Wkk1f9cC`.

Liesl P Erb, Chris Ray, and Robert Guralnick. Determinants of pika population density vs. occupancy in the Southern Rocky Mountains. *Ecological Applications*, 24(3):429–435, 2014.

Hanna T. M. Eskola and Stefan A. H. Geritz. On the mechanistic derivation of various discrete-time population models. *Bulletin of Mathematical Biology*, 69(1):329–346, January 2007. ISSN 1522-9602. doi: 10.1007/s11538-006-9126-4. URL `https://doi.org/10.1007/s11538-006-9126-4`.

Laura K. Estep, Nathan D. Burkett-Cadena, Geoffrey E. Hill, Robert S. Unnasch, and Thomas R. Unnasch. Estimation of dispersal distances of *Culex erraticus* in a focus of eastern equine encephalitis virus in the Southeastern United States. *Journal of Medical Entomology*, 47(6):977–986, 12 2014. ISSN 0022-2585. doi: 10.1603/ME10056. URL `https://doi.org/10.1603/ME10056`.

Arnaud Estoup, Stuart JE Baird, Nicolas Ray, Mathias Currat, Jean-Marie Cornuet, Filipe Santos, Mark A Beaumont, and Laurent Excoffier. Combining genetic, historical and geographical data to reconstruct the dynamics of bioinvasions: application to the cane toad *Bufo marinus*. *Molecular Ecology Resources*, 10 (5):886–901, 2010.

Alison Etheridge and Sarah Penington. Genealogies in bistable waves. *Electronic Journal of Probability*, 27 (none):1 – 99, 2022. doi: 10.1214/22-EJP845. URL `https://doi.org/10.1214/22-EJP845`.

Alison M. Etheridge, Thomas G. Kurtz, Ian Letter, Peter L. Ralph, and Terence Tsui Ho Lung. Looking forwards and backwards: Dynamics and genealogies of locally regulated populations. *Electronic Journal of Probability*, 29(none):1 – 85, 2024. doi: 10.1214/24-EJP1075. URL `https://doi.org/10.1214/24-EJP1075`.

European Space Agency. Earth System Data Lab, 2024. URL `https://www.earthsystemdatalab.net/`.

Joseph Felsenstein. A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist*, 109(967):359–368, 1975. ISSN 00030147. URL `http://www.jstor.org/stable/2459700`.

Joseph Felsenstein. The theoretical population genetics of variable selection and migration. *Annu Rev Genet*, 10:253–280, 1976. doi: 10.1146/annurev.ge.10.120176.001345. URL `http://www.ncbi.nlm.nih.gov/pubmed/797310`.

Stephen E Fick and Robert J Hijmans. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12):4302–4315, 2017.

Grant Foster and Stefan Rahmstorf. Global temperature evolution 1979–2010. *Environmental research letters*, 6(4):044022, 2011.

Robert P. Freckleton, Andrew R. Watkinson, Rhys E. Green, and William J. Sutherland. Census error and the detection of density dependence. *Journal of Animal Ecology*, 75(4):837–851, 2006. doi: https://doi.org/10.1111/j.1365-2656.2006.01121.x. URL `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2656.2006.01121.x`.

WJ Freeland. Populations of cane toad, *Bufo marinus*, in relation to time since colonization. *Wildlife Research*, 13(2):321–329, 1986.

Norman R. French, Theodore Y. Tagami, and Page Hayden. Dispersal in a population of desert rodents. *Journal of Mammalogy*, 49(2):272–280, 1968. ISSN 00222372. URL `http://www.jstor.org/stable/1377984`.

Stefan A. H. Geritz and Éva Kisdi. Mathematical ecology: why mechanistic models? *Journal of Mathematical Biology*, 65(6):1411–1415, December 2012. ISSN 1432-1416. doi: 10.1007/s00285-011-0496-3. URL `https://doi.org/10.1007/s00285-011-0496-3`.

T. A. Graves, P. Beier, and J. A. Royle. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Mol Ecol*, 22(15):3888–3903, August 2013.

Volker Grimm. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological modelling*, 115(2-3):129–148, 1999.

Benjamin C. Haller and Philipp W. Messer. SLiM 4: Multispecies eco-evolutionary modeling. *The American Naturalist*, 201(5):E127–E139, 2023. doi: 10.1086/723601. URL https://doi.org/10.1086/723601. PMID: 37130229.

Benjamin C. Haller and Philipp W. Messer. *SLiM: An Evolutionary Simulation Framework (the SLiM manual)*, 2024. URL https://messerlab.org/SLiM.

Robert L Hammond, Lori J Lawson Handley, Bruce J Winney, Michael W Bruford, and Nicolas Perrin. Genetic evidence for female-biased dispersal and gene flow in a polygynous primate. *Proceedings of the Royal Society B: Biological Sciences*, 273(1585):479–484, November 2005. doi: 10.1098/rspb.2005.3257. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2005.3257.

Ilkka Hanski. Metapopulation dynamics: From concepts and observations to predictive models. In Ilkka Hanski and Michael E. Gilpin, editors, *Metapopulation Biology*, pages 69–91. Academic Press, San Diego, 1997. ISBN 978-0-12-323445-2. doi: https://doi.org/10.1016/B978-012323445-2/50007-9. URL https://www.sciencedirect.com/science/article/pii/B9780123234452500079.

Rachel R. Harman, Jerome Goddard II, Ratnasingham Shivaji, and James T. Cronin. Frequency of occurrence and population-dynamic consequences of different forms of density-dependent emigration. *The American Naturalist*, 0(ja):null, 2020. doi: 10.1086/708156. URL https://doi.org/10.1086/708156.

Godfrey M. Hewitt. Quaternary phylogeography: the roots of hybrid zones. *Genetica*, 139(5):617–638, 2011. ISSN 1573-6857. doi: 10.1007/s10709-011-9547-3. URL http://dx.doi.org/10.1007/s10709-011-9547-3.

Lisa Hülsmann, Ryan A. Chisholm, and Florian Hartig. Is variation in conspecific negative density dependence driving tree diversity patterns at large scales? *Trends in Ecology & Evolution*, 36(2):151–163, February 2021. ISSN 01695347. doi: 10.1016/j.tree.2020.10.003. URL https://doi.org/10.1016/j.tree.2020.10.003.

Daniel H. Janzen. Herbivores and the number of tree species in tropical forests. *The American Naturalist*, 104(940):501–528, 1970. doi: 10.1086/282687. URL https://doi.org/10.1086/282687.

Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, maxalbert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe, and François Leblanc. geopandas/geopandas: v0.8.1, July 2020. URL https://doi.org/10.5281/zenodo.3946761.

G. G. Katul, A. Porporato, R. Nathan, M. Siqueira, M. B. Soons, D. Poggi, H. S. Horn, and S. A. Levin. Mechanistic analytical models for long-distance seed dispersal by wind. *The American Naturalist*, 166(3): 368–381, 2005. doi: 10.1086/432589. URL https://doi.org/10.1086/432589. PMID: 16224691.

Michael Kearney, Ben L Phillips, Christopher R Tracy, Keith A Christian, Gregory Betts, and Warren P Porter. Modelling species distributions without using species distributions: the cane toad in australia under current and future climates. *Ecography*, 31(4):423–434, 2008.

Samuel Kotz and Saralees Nadarajah. *Multivariate t-Distributions and Their Applications*. Cambridge University Press, Cambridge, 2004. ISBN 9780521826549. doi: DOI:10.1017/CBO9780511550683. URL https://www.cambridge.org/core/books/multivariate-tdistributions-and-their-applications/463E067E32C230D3205563EDD195DA2D.

Russell Lande, Steinar Engen, and Bernt-Erik Saether. *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, 04 2003. ISBN 9780198525257. doi: 10.1093/acprof:oso/9780198525257.001.0001. URL https://doi.org/10.1093/acprof:oso/9780198525257.001.0001.

Valérie Laporte and Brian Charlesworth. Effective Population Size and Population Subdivision in Demographically Structured Populations. *Genetics*, 162(1):501–519, September 2002. ISSN 1943-2631. doi: 10.1093/genetics/162.1.501. URL https://doi.org/10.1093/genetics/162.1.501.

Monte Lloyd. 'mean crowding'. *Journal of Animal Ecology*, 36(1):1–30, 1967. ISSN 00218790, 13652656. URL http://www.jstor.org/stable/3012.

Erik Lundgren and Peter L. Ralph. Are populations like a circuit? Comparing isolation by resistance to a new coalescent-based method. *Molecular Ecology Resources*, 19(6):1388–1406, 2019. doi: 10.1111/1755-0998.13035. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13035.

Richard N. Mack and John L. Harper. Interference in dune annuals: Spatial pattern and neighbourhood effects. *Journal of Ecology*, 65(2):345–363, 1977. ISSN 00220477, 13652745. URL http://www.jstor.org/stable/2259487.

G. Malécot. *The Mathematics of Heredity*. Freeman, 1969. ISBN 9780716706786. URL https://wellcomecollection.org/works/msfaxgkw. Translated from the French edition, 1948.

Brad H. McRae, Brett G. Dickson, Timothy H. Keitt, and Viral B. Shah. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, 89(10):2712–2724, 2008. ISSN 1939-9170. doi: 10.1890/07-1861.1. URL http://dx.doi.org/10.1890/07-1861.1.

Tom EX Miller, Amy L Angert, Carissa D Brown, Julie A Lee-Yaw, Mark Lewis, Frithjof Lutscher, Nathan G Marculis, Brett A Melbourne, Allison K Shaw, Marianna Szűcs, et al. Eco-evolutionary dynamics of range expansion. *Ecology*, 101(10):e03139, 2020.

Jiseon Min, Misha Gupta, Michael M Desai, and Daniel B Weissman. Spatial structure alters the site frequency spectrum produced by hitchhiking. *Genetics*, 222(3):iyac139, 09 2022. ISSN 1943-2631. doi: 10.1093/genetics/iyac139. URL https://doi.org/10.1093/genetics/iyac139.

P. Mirski, E. Brzosko, I. Jedrzejczyk, J. Kotowicz, B. Ostrowiecka, and A. Wróblewska. Genetic structure of dioecious and trioecious *Salix myrsinifolia* populations at the border of geographic range. *Tree Genetics & Genomes*, 13(1):15, January 2017. ISSN 1614-2950. doi: 10.1007/s11295-016-1096-6. URL https://doi.org/10.1007/s11295-016-1096-6.

Denis Mollison. The rate of spatial propagation of simple epidemics. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory*, pages 579–614, Berkeley, Calif., 1972. Univ. California Press. URL http://projecteuclid.org/euclid.bsmsp/1200514358.

David Montero, César Aybar, Miguel D. Mahecha, Francesco Martinuzzi, Maximilian Söchting, and Sebastian Wieneke. A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research. *Scientific Data*, 10(1):197, April 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-02096-0. URL https://www.nature.com/articles/s41597-023-02096-0.

Juan Manuel Morales and Teresa Morán López. Mechanistic models of seed dispersal by animals. *Oikos*, 2022(2), 2022. doi: https://doi.org/10.1111/oik.08328. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.08328.

David R Murray. *Seed Dispersal by Water*, chapter 2, pages 49–86. Academic Press, 2012.

Thomas Nagylaki and James F. Crow. Continuous selective models. *Theoretical Population Biology*, 5(2):257–283, 1974. ISSN 0040-5809. doi: https://doi.org/10.1016/0040-5809(74)90045-8. URL https://www.sciencedirect.com/science/article/pii/0040580974900458.

NASA. Earthdata, 2024. URL https://www.earthdata.nasa.gov/.

M.G. Neubert, M. Kot, and M.A. Lewis. Dispersal and pattern formation in a discrete-time predator-prey model. *Theoretical Population Biology*, 48(1):7 – 43, 1995. ISSN 0040-5809. doi: http://dx.doi.org/10.1006/tpbi.1995.1020. URL http://www.sciencedirect.com/science/article/pii/S0040580985710209.

Ace R. North, Austin Burt, and H. Charles J. Godfray. Modelling the potential of genetic control of malaria mosquitoes at national scale. *BMC Biology*, 17(1):1–12, December 2019. ISSN 1741-7007. doi: 10.1186/s12915-019-0645-5. URL https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-019-0645-5. Number: 1 Publisher: BioMed Central.

Jens Nullmeier and Oskar Hallatschek. The coalescent in boundary-limited range expansions. *Evolution*, 67 (5):1307–1320, 2013. ISSN 1558-5646. doi: 10.1111/evo.12037. URL http://dx.doi.org/10.1111/evo.12037.

Minna Nurminiemi, Jarle Tufto, Nils-Otto Nilsson, and Odd Arne Rognli. Spatial models of pollen dispersal in the forage grass meadow fescue. *Evolutionary Ecology*, 12(4):487, May 1998. ISSN 1573-8477. doi: 10.1023/A:1006529023036. URL https://doi.org/10.1023/A:1006529023036.

Darren J. Obbard, Stephen A. Harris, and John R. Pannell. Sexual systems and population genetic structure in an annual plant: testing the metapopulation model. *The American Naturalist*, 167(3):354–366, March 2006. ISSN 1537-5323. doi: 10.1086/499546.

Karen Suzanne Oberhauser, Michelle J. Solensky, Journey North, and Monarch Larva Monitoring Project. *Monarch butterfly biology & conservation*. Cornell University Press, 2004. ISBN 9780801441882. URL https://www.cornellpress.cornell.edu/book/9780801441882/the-monarch-butterfly/.

Stephen W. Pacala. Neighborhood models of plant population dynamics. 2. Multi-species models of annuals. *Theoretical Population Biology*, 29(2):262–292, 1986. ISSN 0040-5809. doi: https://doi.org/10.1016/0040-5809(86)90011-0. URL https://www.sciencedirect.com/science/article/pii/0040580986900110.

Stephen W. Pacala and J. A. Silander. Neighborhood models of plant population dynamics. I. Single-species models of annuals. *The American Naturalist*, 125(3):385–411, 1985. ISSN 00030147, 15375323. URL http://www.jstor.org/stable/2461522.

Emmanuel Paradis, Stephen R. Baillie, William J. Sutherland, and Richard D. Gregory. Patterns of natal and breeding dispersal in birds. *Journal of Animal Ecology*, 67(4):518–536, 1998. doi: https://doi.org/10.1046/j.1365-2656.1998.00215.x. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2656.1998.00215.x.

Jayson Paulose, Joachim Hermisson, and Oskar Hallatschek. Spatial soft sweeps: Patterns of adaptation in populations with long-range dispersal. *PLOS Genetics*, 15(2):1–29, 02 2019. doi: 10.1371/journal.pgen.1007936. URL https://doi.org/10.1371/journal.pgen.1007936.

Oliver P. Pearson. A mechanical model for the study of population dynamics. *Ecology*, 41(3):494–508, 1960. ISSN 00129658, 19399170. URL http://www.jstor.org/stable/1933324.

A Townsend Peterson. Predicting species' geographic distributions based on ecological niche modeling. *The condor*, 103(3):599–605, 2001.

Erin E. Peterson, Ephraim M. Hanks, Mevin B. Hooten, Jay M. Ver Hoef, and Marie-Josée Fortin. Spatially structured statistical network models for landscape genetics. *Ecological Monographs*, 89(2):e01355, 2019. doi: 10.1002/ecm.1355. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecm.1355.

Josep Pons and Juli G. Pausas. Acorn dispersal estimated by radio-tracking. *Oecologia*, 153(4):903–911, October 2007. ISSN 1432-1939. doi: 10.1007/s00442-007-0788-x. URL https://doi.org/10.1007/s00442-007-0788-x.

PRISM Climate Group. PRISM Climate Data, 2024. URL https://www.prism.oregonstate.edu/.

Anne E. Pusey and Craig Packer. The Evolution of Sex-Biased Dispersal in Lions. *Behaviour*, 101(4):275–310, January 1987. ISSN 0005-7959, 1568-539X. doi: 10.1163/156853987X00026. URL https://brill.com/view/journals/beh/101/4/article-p275_2.xml.

Sohini Ramachandran, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences*, 102(44):15942–15947, 2005. doi: 10.1073/pnas.0507611102. URL `https://www.pnas.org/doi/abs/10.1073/pnas.0507611102`.

Lea K. Richardson, Scott W. Nordstrom, Amy Waananen, Riley D. Thoen, Amy B. Dykstra, Gretel Kiefer, Drake E. Mullett, Erin G. Eichenberger, Ruth G. Shaw, and Stuart Wagenius. Juvenile survival increases with dispersal distance and varies across years: 15 years of evidence in a prairie perennial. *Ecology*, n/a (n/a):e4331, 2024. doi: https://doi.org/10.1002/ecy.4331. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.4331`.

J J Robledo-Arnuncio and L. Gil. Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity*, 94(1):13–22, 2005. ISSN 13652540. doi: 10.1038/sj.hdy.6800542. URL `https://doi.org/10.1038/sj.hdy.6800542`.

J. J. Robledo-Arnuncio and F. Rousset. Isolation by distance in a continuous population under stochastic demographic fluctuations. *Journal of Evolutionary Biology*, 23(1):53–71, 2010. ISSN 1420-9101. doi: 10.1111/j.1420-9101.2009.01860.x. URL `http://dx.doi.org/10.1111/j.1420-9101.2009.01860.x`.

Juan Jose Robledo-Arnuncio and Cristina García. Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology*, 16(23):5098–5109, 2007. doi: https://doi.org/10.1111/j.1365-294X.2007.03427.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2007.03427.x`.

Haldre S Rogers, Noelle G Beckman, Florian Hartig, Jeremy S Johnson, Gesine Pufal, Katriona Shea, Damaris Zurell, James M Bullock, Robert Stephen Cantrell, Bette Loiselle, Liba Pejchar, Onja H Razafindratsima, Manette E Sandor, Eugene W Schupp, W Christopher Strickland, and Jenny Zambrano. The total dispersal kernel: a review and future directions. *AoB PLANTS*, 11(5):plz042, 09 2019. ISSN 2041-2851. doi: 10.1093/aobpla/plz042. URL `https://doi.org/10.1093/aobpla/plz042`.

Noah A Rosenberg, Saurabh Mahajan, Sohini Ramachandran, Chengfeng Zhao, Jonathan K Pritchard, and Marcus W Feldman. Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure. *PLoS Genetics*, 1(6):e70, December 2005. ISSN 1553-7390. doi: 10.1371/journal.pgen.0010070. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1310579/`.

Rousset. Genetic differentiation between individuals. *Journal of Evolutionary Biology*, 13(1):58–62, 2000. doi: https://doi.org/10.1046/j.1420-9101.2000.00137.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1420-9101.2000.00137.x`.

F Rousset. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145(4):1219–1228, April 1997. URL `http://www.ncbi.nlm.nih.gov/pubmed/9093870`.

M. Saastamoinen, G. Bocedi, J. Cote, D. Legrand, F. Guillaume, C. W. Wheat, E. A. Fronhofer, C. Garcia, R. Henry, A. Husby, M. Baguette, D. Bonte, A. Coulon, H. Kokko, E. Matthysen, K. Niitepold, E. Nonaka, V. M. Stevens, J. M. J. Travis, K. Donohue, J. M. Bullock, and M. Del Mar Delgado. Genetics of dispersal. *Biol Rev Camb Philos Soc*, 93(1):574–599, February 2018.

Akira Sasaki. Clumped distribution by neighbourhood competition. *Journal of Theoretical Biology*, 186(4):415–430, June 1997. doi: 10.1006/jtbi.1996.0370. URL `https://doi.org/10.1006%2Fjtbi.1996.0370`.

Outi Savolainen, Tanja Pyhäjärvi, and Timo Knürr. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, 38(1):595–619, 2007. doi: 10.1146/annurev.ecolsys.38.091206.095646. URL `http://dx.doi.org/10.1146/annurev.ecolsys.38.091206.095646`.

A Sedghifar, Y Brandvain, and P Ralph. Beyond clines: lineages and haplotype blocks in hybrid zones. *Mol Ecol*, 25(11):2559–2576, June 2016. doi: 10.1111/mec.13677. URL `https://www.ncbi.nlm.nih.gov/pubmed/27148805`.

Richard Shine. The ecological impact of invasive cane toads (*Bufo marinus*) in Australia. *The Quarterly review of biology*, 85(3):253–291, 2010.

Richard Shine, Georgia Ward-Fear, and Gregory P Brown. A famous failure: Why were cane toads an ineffective biocontrol in australia? *Conservation Science and Practice*, 2(12):e296, 2020.

Richard Shine, Ross A Alford, Ryan Blennerhasset, Gregory P Brown, Jayna L DeVore, Simon Ducatez, Patrick Finnerty, Matthew Greenlees, Shannon W Kaiser, Samantha McCann, et al. Increased rates of dispersal of free-ranging cane toads (*Rhinella marina*) during their global invasion. *Scientific Reports*, 11 (1):23574, 2021.

Stephen M. Shuster. *Mating systems and strategies*. Monographs in behavior and ecology. University Press, Princeton, 2003. ISBN 978-0-691-04930-4.

John A. Silander, Jr and Stephen W. Pacala. Neighborhood predictors of plant performance. *Oecologia*, 66 (2):256–263, 1985. ISSN 00298549, 14321939. URL `http://www.jstor.org/stable/4217621`.

Neftalí Sillero. What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling*, 222(8):1343–1346, 2011. ISSN 0304-3800. doi: https://doi.org/10.1016/j.ecolmodel.2011.01.018. URL `https://www.sciencedirect.com/science/article/pii/S0304380011000500`.

Neftalí Sillero and A. Márcia Barbosa. Common mistakes in ecological niche models. *International Journal of Geographical Information Science*, 35(2):213–226, 2021. doi: 10.1080/13658816.2020.1798968. URL `https://doi.org/10.1080/13658816.2020.1798968`.

J. G. Skellam. Random dispersal in theoretical populations. *Biometrika*, 38(1/2):196–218, 1951. ISSN 00063444. URL `http://www.jstor.org/stable/2332328`.

Montgomery Slatkin. Gene flow and selection in a cline. *Genetics*, 75(4):733–756, 1973. URL `http://www.genetics.org/cgi/content/abstract/75/4/733`.

Andrew T Smith. The distribution and dispersal of pikas: consequences of insular population structure. *Ecology*, 55(5):1112–1119, 1974.

Chris C R Smith, Silas Tittes, Peter L Ralph, and Andrew D Kern. Dispersal inference from population genetic variation using a convolutional neural network. *Genetics*, 224(2):iyad068, 04 2023. ISSN 1943-2631. doi: 10.1093/genetics/iyad068. URL `https://doi.org/10.1093/genetics/iyad068`.

Chris C. R. Smith, Gilia Patterson, Peter L. Ralph, and Andrew D. Kern. Estimation of spatial demographic maps from polymorphism data using a neural network. *bioRxiv*, 2024. doi: 10.1101/2024.03.15.585300. URL `https://www.biorxiv.org/content/early/2024/03/17/2024.03.15.585300`.

Tyler B Smith and Daniel B Weissman. Isolation by distance in populations with power-law dispersal. *G3 Genes|Genomes|Genetics*, 13(4):jkad023, April 2023. ISSN 2160-1836. doi: 10.1093/g3journal/jkad023. URL `https://doi.org/10.1093/g3journal/jkad023`.

Robin E. Snyder and Peter Chesson. How the spatial scales of dispersal, competition, and environmental heterogeneity interact to affect coexistence. *The American Naturalist*, 164(5):633–650, November 2004. doi: 10.1086/424969. URL `https://doi.org/10.1086%2F424969`.

Xiaoyang Song, Jun Ying Lim, Jie Yang, and Matthew Scott Luskin. When do Janzen-Connell effects matter? A phylogenetic meta-analysis of conspecific negative distance and density dependence experiments. *Ecology Letters*, 24(3):608–620, 2021. doi: https://doi.org/10.1111/ele.13665. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.13665`.

Hannah M. Specht and Todd W. Arnold. Banding age ratios reveal prairie waterfowl fecundity is affected by climate, density dependence and predator–prey dynamics. *Journal of Applied Ecology*, 55(6):2854–2864, 2018. doi: https://doi.org/10.1111/1365-2664.13186. URL `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.13186`.

Erica N. Spotswood, Pierre Mariotte, Emily C. Farrer, Liana Nichols, and Katharine N. Suding. Separating sources of density-dependent and density-independent establishment limitation in invading species. *Journal of Ecology*, 105(2):436–444, 2017. doi: https://doi.org/10.1111/1365-2745.12686. URL `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2745.12686`.

Richard A Stillman, Steven F Railsback, Jarl Giske, Uta Berger, and Volker Grimm. Making predictions in a changing world: The benefits of individual-based ecology. *Bioscience*, 65(2):140–150, February 2015. ISSN 15253244. doi: 10.1093/biosci/biu192. URL `https://pubmed.ncbi.nlm.nih.gov/26955076`.

Tomasz Suchan, Clément P Bataille, Megan S Reich, Eric Toro-Delgado, Roger Vila, Naomi E Pierce, and Gerard Talavera. A trans-oceanic flight of over 4,200 km by painted lady butterflies. *Nature Communications*, 15(1):5205, 2024.

P. Taylor. Birth–death symmetry in the evolution of a social trait. *Journal of Evolutionary Biology*, 23(12): 2569–2578, 12 2010. ISSN 1010-061X. doi: 10.1111/j.1420-9101.2010.02122.x. URL `https://doi.org/10.1111/j.1420-9101.2010.02122.x`.

Brittany J. Teller, Peter B. Adler, Collin B. Edwards, Giles Hooker, and Stephen P. Ellner. Linking demography with drivers: climate and competition. *Methods in Ecology and Evolution*, 7(2):171–183, 2016. doi: https://doi.org/10.1111/2041-210X.12486. URL `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12486`.

John Terborgh. Enemies maintain hyperdiverse tropical forests. *The American Naturalist*, 179(3):303–314, 2012. doi: 10.1086/664183. URL `https://doi.org/10.1086/664183`.

Audrey Trochet, Elodie A. Courtois, Virginie M. Stevens, Michel Baguette, Alexis Chaine, Dirk S. Schmeller, Jean Clobert, and John J. Wiens. Evolution of Sex-Biased Dispersal. *The Quarterly Review of Biology*, 91(3):297–320, August 2016. ISSN 0033-5770. doi: 10.1086/688097. URL `https://www.journals.uchicago.edu/doi/full/10.1086/688097`.

J Tufto, S Engen, and K Hindar. Stochastic dispersal processes in plant populations. *Theor Popul Biol*, 52 (1):16–26, August 1997. URL `http://www.ncbi.nlm.nih.gov/pubmed/9356320`.

Shripad Tuljapurkar. An uncertain life: Demography in random environments. *Theoretical Population Biology*, 35(3):227–294, 1989. ISSN 0040-5809. doi: https://doi.org/10.1016/0040-5809(89)90001-4. URL `https://www.sciencedirect.com/science/article/pii/0040580989900014`.

Shripad Tuljapurkar. *Population Dynamics in Variable Environments*. Lecture Notes in Biomathematics. Springer Berlin Heidelberg, 2013. ISBN 9783642516528. URL `https://books.google.com/books?id=Y6X1CAAAQBAJ`.

H Uecker, S P Otto, and J Hermisson. Evolutionary rescue in structured populations. *Am Nat*, 183(1): 17–35, January 2014. doi: 10.1086/673914. URL `http://www.ncbi.nlm.nih.gov/pubmed/24334746`.

Mark C Urban, Ben L Phillips, David K Skelly, and Richard Shine. A toad more traveled: the heterogeneous invasion dynamics of cane toads in australia. *The American Naturalist*, 171(3):E134–E148, 2008.

Alexandra Weigelt and Peter Jolliffe. Indices of plant competition. *Journal of Ecology*, 91(5):707–720, 2003. ISSN 00220477, 13652745. URL `http://www.jstor.org/stable/3599697`.

Jacob Weiner. A neighborhood model of annual-plant interference. *Ecology*, 63(5):1237–1241, 1982. doi: https://doi.org/10.2307/1938849. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1938849`.

Christopher P. Weiss-Lehman, Chhaya M. Werner, Catherine H. Bowler, Lauren M. Hallett, Margaret M. Mayfield, Oscar Godoy, Lina Aoyama, György Barabás, Chengjin Chu, Emma Ladouceur, Loralee Larios, and Lauren G. Shoemaker. Disentangling key species interactions in diverse and heterogeneous communities: A Bayesian sparse modelling approach. *Ecology Letters*, 25(5):1263–1276, 2022. doi: https://doi.org/10.1111/ele.13977. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.13977`.

Wioletta Werkowska, Ana L. Márquez, Raimundo Real, and Pelayo Acevedo. A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25(1):127–133, 2017. doi: 10.1139/er-2016-0045. URL https://doi.org/10.1139/er-2016-0045.

S Wright. Isolation by distance. *Genetics*, 28(2):114–138, March 1943. URL http://www.genetics.org/cgi/reprint/28/2/114.

S Wright. Isolation by distance under diverse systems of mating. *Genetics*, 31(1):39–59, January 1946. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209315.

Katherine L. Yates, Phil J. Bouchet, M. Julian Caley, Kerrie Mengersen, Christophe F. Randin, Stephen Parnell, Alan H. Fielding, Andrew J. Bamford, Stephen Ban, A. Márcia Barbosa, Carsten F. Dormann, Jane Elith, Clare B. Embling, Gary N. Ervin, Rebecca Fisher, Susan Gould, Roland F. Graf, Edward J. Gregr, Patrick N. Halpin, Risto K. Heikkinen, Stefan Heinänen, Alice R. Jones, Periyadan K. Krishnakumar, Valentina Lauria, Hector Lozano-Montes, Laura Mannocci, Camille Mellin, Mohsen B. Mesgaran, Elena Moreno-Amat, Sophie Mormede, Emilie Novaczek, Steffen Oppel, Guillermo Ortuño Crespo, A. Townsend Peterson, Giovanni Rapacciuolo, Jason J. Roberts, Rebecca E. Ross, Kylie L. Scales, David Schoeman, Paul Snelgrove, Göran Sundblad, Wilfried Thuiller, Leigh G. Torres, Heroen Verbruggen, Lifei Wang, Seth Wenger, Mark J. Whittingham, Yuri Zharikov, Damaris Zurell, and Ana M.M Sequeira. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33(10):790–802, October 2018. ISSN 01695347. doi: 10.1016/j.tree.2018.08.001. URL https://doi.org/10.1016/j.tree.2018.08.001.

Andrii Zaiats, Matthew J. Germino, Marcelo D. Serpe, Bryce A. Richardson, and T. Trevor Caughlin. Intraspecific variation mediates density dependence in a genetically diverse plant species. *Ecology*, 102 (11):e03502, 2021. doi: https://doi.org/10.1002/ecy.3502. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.3502.

Elise F. Zipkin, Jeffrey W. Doser, Courtney L. Davis, Wendy Leuenberger, Samuel Ayebare, and Kayla L. Davis. Integrated community models: A framework combining multispecies data sources to estimate the status, trends and dynamics of biodiversity. *Journal of Animal Ecology*, 92(12):2248–2262, 2023. doi: https://doi.org/10.1111/1365-2656.14012. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2656.14012.