A guide to calculating discrete-time invasion rates from data

Mark A. Lewis *
University of Alberta

Michael G. Neubert
Woods Hole Oceanographic Institute

Hal Caswell
Woods Hole Oceanographic Institute

James S. Clark
Duke University

Katriona Shea
Pennsylvania State University

July 6, 2004

1 Introduction

One measure of biological invasiveness is the rate at which an established invader will spread spatially in its new environment. Slow spread signifies slow increase in ecological impact, whereas fast spread signifies the converse. If one can predict spread rates from life history attributes, such as growth rates and dispersal distances, then potential invasiveness can be assessed before the invasion occurs. A prediction of this sort requires models for population spread. As outlined below, such models have a long and distinguished history in quantitative ecology.

Whereas early mathematical models for population spread were primarily conceptual and qualitative in nature, a new generation of realistic models is emerging. These new models are tied directly to the demography and dispersal of individuals. However, there are new challenges in the linking of these models to the biological processes.

As we will illustrate in this chapter, spread rate predictions are very sensitive to assumptions about long-distance dispersal. Are there robust methods for estimating spread rates? This is one question we will address.

Furthermore almost all mathematical models assume that the spread occurs in one spatial dimension, along a line. This is not because mathematicians have not noticed that most

*This work was conducted as part of the Demography and Dispersal Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (DEB-0072909), the University of California, and the Santa Barbara campus. This research was also supported in part by a Canada Research Chair, the Natural Sciences and Engineering Research Council of Canada, and the National Science Foundation (DEB-0315860, DEB-9527400, DEB-9808501, DMS-9973312, DMS-0213698, DEB-0315860). Thanks to the Demography and Dispersal Working Group members for their help and support with the research.

population spread takes place in two dimensions (except cases like dispersal along a coastline or river). Rather, it is because the main qualitative features of invasions are apparent in one-dimensional models, although, as we will show, the quantitative results in specific cases can depend very much on the the dimensionality of the dispersal data and of the model used for analysis. In particular, we will show that the naive application of one dimensional models to two dimensional dispersal data will produce a systematic bias (sometimes positive, sometimes negative) in spread rate estimates. In this chapter we will outline new fitting methods for avoiding these biases.

2 Modeling Background

The issue of spread rate for biological invaders was considered in detail by Skellam (1951) where Fisher's 1937 partial differential equation model was used to describe the rate of change of local population density with time

$$\frac{\partial n}{\partial t} = \rho n \left(1 - \frac{n}{\kappa}\right) + D\frac{\partial^2 n}{\partial x^2}. \tag{1}$$

Here $n(x,t)$ is the local population density, ρ is the intrinsic growth rate, κ is the carrying capacity, D is the diffusion coefficient, x is the one-dimensional space coordinate and t is time. The rate of spread of a population obeying this equation asymptotically approaches $c^* = 2\sqrt{\rho D}$ for large times (Kolmogorov et al. 1937; Aronson and Weinberger 1975) (Figure 1). Whereas the spread rate predictions are on a landscape scale, the parameters r and D can be measured on the individual level, using life table analysis, and mark-recapture. Furthermore, comparisons of historically observed invasive spread rates with the asymptotic spread rate formula have held up for a wide variety of species (Andow et al. 1990), making the interplay between invasion theory and spread data a modern-day success story in quantitative ecology (Shigesada and Kawasaki 1997).

A major problem with Fisher's 1937 model is that it imposes one particular form of dispersal. The assumption of diffusion in equation (1), which implies normally distributed dispersal propagules, is often violated when dispersal is measured for biological populations (Lewis 1997). While there is tremendous variability in such dispersal data, there is a strong tendency for the distribution of dispersal distances to be leptokurtic, with a larger number of distances near the center and in the tails than in a normal distribution with comparable variance. The effect of the long-distance dispersers, as described by the tails of the distribution of distances, is dramatic (Caswell et al. 2003). Predictions for invasion rates can speed up by an order of magnitude or more when the long-distance dispersers are included (Kot, Lewis, and van den Driessche 1996).

When the dispersal distribution is far from normal, the partial differential equation model (1) no longer suffices. It is then necessary to define a dispersal kernel $k(x,y)$, in one spatial dimension, which describes the probability that a propagule that starts at y moves to the interval $(x,\, x+dx)$ by $k(x,y)\,dx$. The units for the dispersal kernel k are per unit length. The spatial region of interest is the interval $\Omega = (-\ell,\, \ell)$. When modelling invasions, we typically assume ℓ be arbitrarily large. In the absence of immigration from outside Ω, every

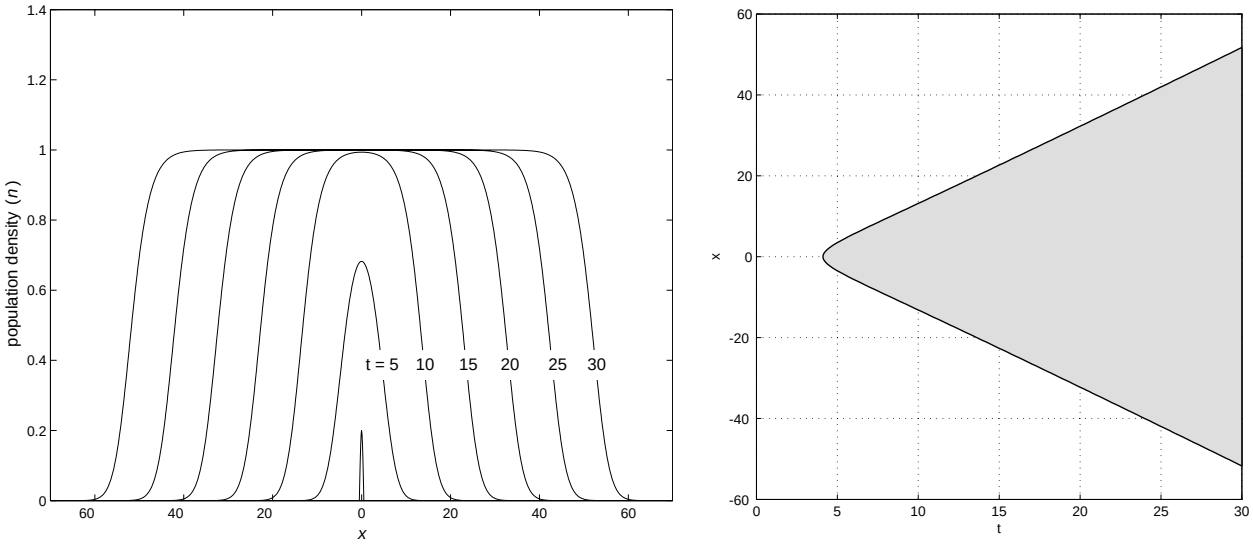

Figure 1: Population spread for Fisher's equation. A typical solution of Fisher's equation (1) illustrates growth and spread in one-dimensional linear space. At left, the solution is plotted for equally-spaced time intervals. At right, the gray area denotes the region in space where the population is larger than a threshold level $n = 0.5$. The boundaries of this area have slopes equal to the asymptotic spread rate $c = 2\sqrt{\rho D}$. For this figure, $\rho = \kappa = D = 1$, and $n(x, 0) = 0.2\cos(\pi x)$ for $|x| \leq 0.5$. Based on Neubert and Parker (2004).

disperser must originate at some other point in Ω, so that

$$\int_{-\ell}^{\ell} k(x, y)\, dy = 1. \tag{2}$$

In a homogeneous habitat, dispersal between two locations will only depend upon upon the relative locations of the start and finish points. In this case, to which we will restrict our attention for the remainder of the chapter, we write $k(x, y) = k(x - y)$. (Shigesada, Kawasaki, and Teramoto (1986) examine invasions in heterogeneous environments.)

To include population dynamics, population growth from one generation to the next can be described with a nonlinear function

$$n_{t+1} = f(n_t) = n_t g(n_t), \tag{3}$$

where g defines the per capita growth rate as a function of local population density n_t. Non-overlapping generations are assumed, but, here and in the subsequent analysis, the assumption can be relaxed to include models with stage-structure (see Neubert and Caswell (2000) for details).

The nonspatial population model (3) is then modified to allow for dispersal between reproduction events. If we designate the population density at location x and time t by $n_t(x)$, these models take the form of a scalar *integrodifference equation*

$$n_t(x) = \int_{-\ell}^{\ell} f(n_t(y))\, k(x - y)\, dy. \tag{4}$$

Early analyses of population spread using integrodifference equations appeared in the mathematical literature primarily in the context of genetics (Weinberger 1982; Lui 1982a; Lui 1982b; Weinberger 1984; Lui 1985; Lui 1986; Lui 1989a; Lui 1989b). However, integrodifference equations, and generalizations of them, are now being used by a growing list of ecologists to investigate the spread rate of ecological populations (Kot 1992; Allen et al. 1996; Allen et al. 1996; Kot et al. 1996; Veit and Lewis 1996; Lewis 1997; Hart and Gardner 1997; Clark et al. 1998; Clark 1998; Higgins and Richardson 1999; Neubert and Caswell 2000; Neubert et al. 2000; Takasu et al. 2000; Woolcock and Cousens 2000; With 2002; Schofield 2002; Caswell et al. 2003; Clark et al. 2003; Marchant 2003; Neubert and Parker 2004; Powell and Zimmermann 2004). Among the results of these investigations are two key findings. First, integrodifference models produce a richer set of invasion dynamics than can be generated by the reaction-diffusion equation (1), including, for example, the possibility of accelerating spread (Kot et al. 1996). Second, the shape of the dispersal kernel—especially the shape of the tails which determine the probability of long-distance dispersal—plays a crucial role in determining the rate of spread.

The effect of long-distance dispersal on spread rates was highlighted by Kot et al. (1996), who fit dispersal kernels to a data set describing the displacement of of genetically marked *Drosophila* (Dobzhansky and Wright 1943), and predictions for the corresponding asymptotic spread rate were linked to the shapes of the kernels. These predicted spread rates varied over an order of magnitude, depending upon the fatness of the tails of the related dispersal kernels (see Figure 2).

The issue of long-distance dispersal, however, goes beyond the choice of a parametric dispersal distribution to describe a set of data. It can also reflect real biological processes. For example, Neubert and Caswell (2000) computed the spread rate for the herbaceous plant *Dipsacus sylvestris* based on data from a seed trap experiment (Werner 1975). Teasel seeds are known to float, but dispersal by streams or rivers was obviously not measured by the seed trap experiment. Neubert and Caswell calculated the asymptotic spread rate resulting from hypothetical mixtures of the seed trap data and dispersal by water with a longer mean distance. They found that long-distance dispersal of even one seed in a million was enough to make the spread rate dependent on the water dispersal alone. Similarly, the seeds of the tropical plant *Calathea ovandensis* are dispersed by at least four species of ant, each with its own typical dispersal distance (Horvitz and Schemske 1986). Neubert and Caswell found that over 99% of the asymptotic spread rate was accounted for by the ant species with the longest dispersal distance, even though it dispersed only 7% of the seeds.

Classical models for population spread like (1) and (4) consider the case where the spatial domain is one-dimensional and linear, and a small beachhead of individuals is introduced locally. This is only directly applicable to cases such as population spread along a roadside, coastline (Lubina and Levin 1988) or a river (Speirs and Gurney 2001; Pachepsky et al. 2004). Later, we will introduce models that describe dispersal in two spatial dimensions. These models produce asymptotic spread rate predictions that can differ in different directions if the dispersal kernel is not radially symmetric.

In both one and two dimensions, the dispersal kernel plays a crucial role in determining the asymptotic rate of spread. There are many methods for estimating dispersal kernels, and for estimating the properties of those kernels that enter into the formulae for spread rate, from data (see, for example, Silverman (1986)). We will discuss some of these methods

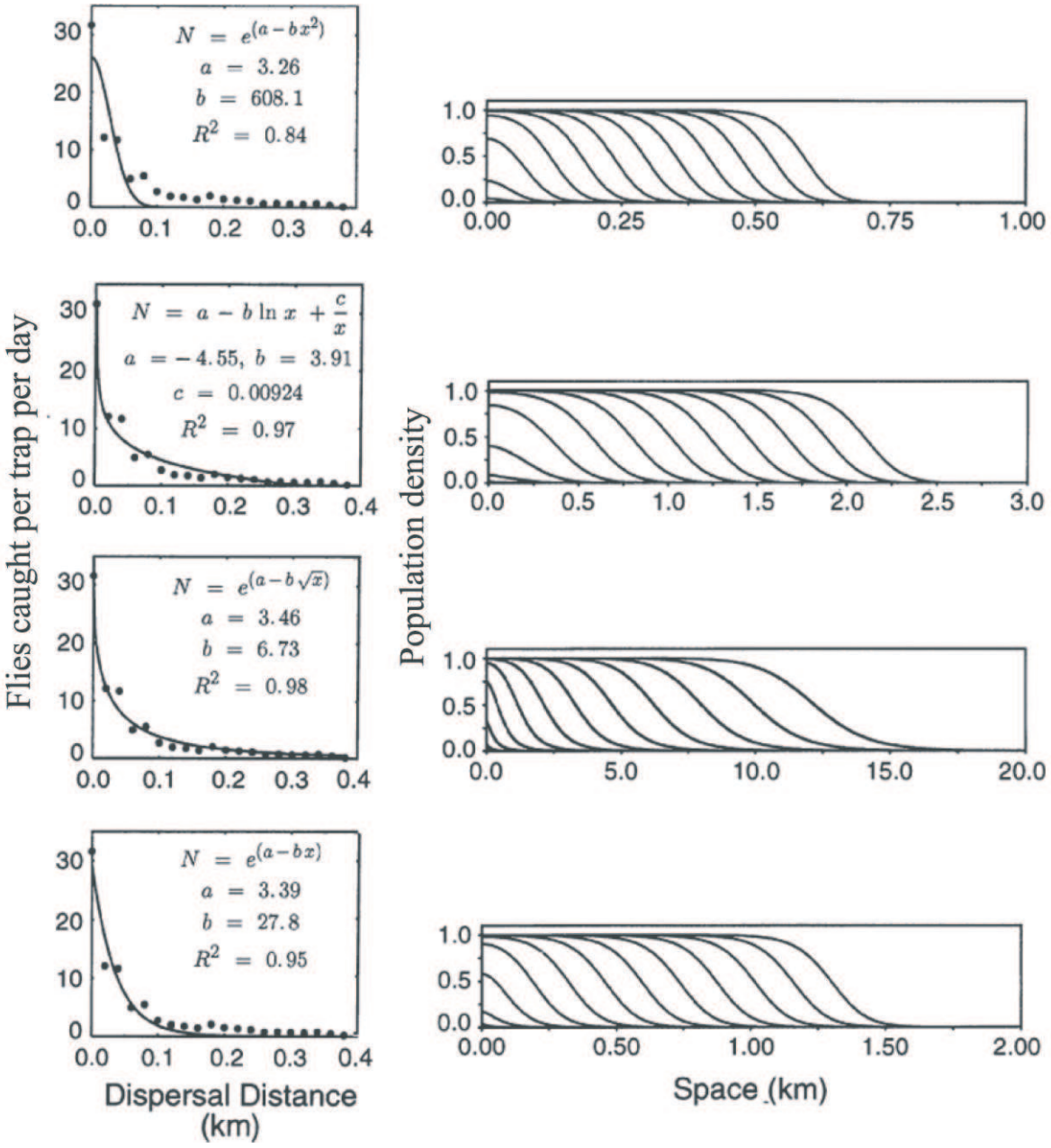

Figure 2: Fitted functions to *D. pseudoobscura* dispersal data provide ingredients for an integrodifference model for insect spread 4. The left panels show average number of insects caught per trap per day in Dobzhansky and Wright's experiments. It was assumed that dispersal was equally likely in both directions, so the dispersal kernels were $k(x) = (g(x) + g(-x))/2$, where g is the fitted function. The right panel shows simulations of the integrodifference equations. Simulations assume Beverton-Holt population dynamics for $f(n)$, with a geometric growth rate of $\lambda = 10$. The carrying capacity was scaled to equal one. Each integrodifference was iterated for 12 generations. Based on Kot et al. (1996). Here is is assumed that all reproduction and dispersal occurs along a one-dimensional strip of suitable habitat. The spread rate can be calculated by dividing the total distance moved by the population front by the 12 generations taken to move the distance.

later in the chapter. First, we discuss some of the kinds of dispersal data that are typically collected and/or published.

3 Dispersal Kernels in One and Two Dimensions

We first consider forms in which data are collected. These fall loosely into two kinds: dispersal data and density data. Dispersal data describe the location of dispersers relative to the parents. These data come from following individual dispersers (eg., banding and recapture of birds, mark and recapture of seeds using coloring and/or radio-tagging) and is recorded as either displacements or displacement distances. In contrast, density data describe the density of dispersers (number per unit area) observed at a given point, typically as a function of distance from a natal site or source of dispersers. These data come from seed traps, pheremone traps for insects and so forth.

If we assume that the population lives along a one-dimensional strip of suitable habitat, along which all dispersal and reproduction occurs (Section 4), the one-dimensional dispersal kernel (number of dispersers per unit length) is needed in equation (4). This kernel can be fitted either directly from the one-dimensional dispersal data or from the density data (number of dispersers per unit area) multiplied by the width of the strip. The constraint that the kernel must be scaled to integrate to 1 (equation (2)), means that the kernel k will actually be independent of the width of the strip—multiplication is only done formally to ensure the correct units for k.

If we assume that the population lives in a two dimensional habitat, it is first necessary to extend the definition of dispersal to two spatial dimensions. Here dispersal is between points $\mathbf{x} = [x_1, x_2]^{\mathrm{T}}$ and $\mathbf{y} = [y_1, y_2]^{\mathrm{T}}$ in two dimensional space. The the two-dimensional dispersal kernel $K(\mathbf{x}, \mathbf{y})$ describes the probability of a propagule which starts at $\mathbf{y}$ moving to the rectangle with corners $\mathbf{x}$ and $\mathbf{x}+\mathbf{dx}$ by $K(\mathbf{x}, \mathbf{y})\, dx_1\, dx_2$.[1] The spatial region of interest is given by Ω. In the case of invasions, this is typically assumed to be arbitrarily large. In the absence of immigration from outside Ω, every disperser must originate at some other point in Ω, so that

$$\int_{\Omega} K(\mathbf{x}, \mathbf{y})\, \mathbf{dy} = 1. \tag{5}$$

If the kernel $K(\mathbf{x}, \mathbf{y})$ depends only upon the relative locations of the start and finish points we write $K(\mathbf{x}, \mathbf{y}) = K(\mathbf{x} - \mathbf{y})$.

To understand the difference between one and two dimensional dispersal kernels we consider the case where the dispersal is isotropic (identical in all directions). In this case the two dimensional dispersal kernel K can be written as a function of the dispersal radius $r = |\mathbf{x} - \mathbf{y}|$. In a linear one-dimensional environment, the scaled distribution of densities (dispersers per unit length) and a distribution of distances that the dispersers travel from the parent are the same and are given by the kernel k. In a two-dimensional environment, the scaled distribution of densities (dispersers per unit area), given by $K(r)$, and the distribution of distances that dispersers travel from from the parent (dispersers per unit length),

[1] Throughout this chapter, we will use boldface Roman characters such as $\mathbf{u}$ and $\mathbf{v}$ to represent vectors. Here, $|\mathbf{u}| = \sqrt{u_1^2 + u_2^2}$ denotes the length of the vector $\mathbf{u}$, and $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 = |\mathbf{u}|\, |\mathbf{v}|\, \cos\theta$ denotes the 'dot product' or projection of one vector onto the other, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

given by $\tilde{K} = 2\pi r K(r)$, are not the same, because there is more area available at distances further from the parent.

4 Population Spread in a One-dimensional Linear Environment

The Dobzhansky and Wright insect data were collected from traps placed along linear transects radiating from a point source. Thus the traps give a relative measure of the density of dispersers (number per unit area), as a function of distance from the release site. The most reasonable assumption is that the transect data describe the radial drop-off in settled insect density in a two-dimensional habitat (Section 5). However, for the sake of illustration, in this paper we first consider the assumption that the insects and transects are found along a one-dimensional strip of suitable habitat, along which all dispersal and reproduction occurs (Section 4). This is a standard assumption in population spread models and is the assumption made in the original analysis of the insect spread rates in Kot et al. (1996). As we will show, the different assumptions about the habitat and dispersal give rise to quite different spread rate estimates.

4.1 Theory

As the introduced beachhead of individuals grows and disperses we expect growth of the range boundary with time. Figure 1 shows a typical progression. A plot of range boundary versus time gives lines whose slope eventually become constant. The eventual slope of these lines—the asymptotic rate of spread of the invasion (henceforth referred to simply as spread rate)—can be predicted using mathematical theory which relates the slope to model parameters. In this section we outline the theory.

It is first necessary to make some assumptions about the growth dynamics. The simplest population dynamics exhibit no overcompensation or Allee effect. This assumption translates into a growth function f in (4) that is monotonically increasing with maximum per capita growth rates at lowest population levels.

The assumption that the maximum per capita growth rate λ occurs at the lowest possible density means $\lambda = g(0) \geq g(n)$ for $n > 0$. A growing population requires $\lambda > 1$. As described above, local introduction of individuals, coupled with a growth rate $\lambda > 1$ and a dispersal kernel $k(z)$, $z = x - y$, in (4), means the population spreads as it grows and disperses (Figure 2). Weinberger (1982) showed that, under the above assumptions on growth dynamics, the population spreads to the right at a rate which approaches speed c as the time since introduction increases, where

$$c = \min_{s>0} \frac{1}{s} \ln\left[\lambda M(s)\right], \tag{6}$$

and $M(s)$ is the moment generating function for the dispersal kernel $k(z)$

$$M(s) = \int_{-\infty}^{\infty} k(z) \exp(sz)\, dz. \tag{7}$$

The parameter s can be understood as a measure of the steepness of the wave: $n \propto \exp(-sz)$ at the leading edge of the rightward-spreading population. To find the speed of the leftward-spreading front, one should use the moment generating function for $k(-z)$ in equation (6).

Thus the population spread rate depends only upon two features: the geometric growth rate of the population λ, and the shape of the dispersal kernel k. The speed for Fisher's equation (1), $c = 2\sqrt{\rho D}$, can be regained from equations (6)–(7) by the choice of $k = N(0, 2D)$ and $\rho = \log(\lambda)$. Details are given in Kot et al. (1996).

In practice, calculation of the spread rate must be done numerically, either by using a standard minimization routine, or by solving the double root condition equations, $\exp(sc) = \lambda M(s)$ and $c\exp(sc) = \lambda M'(s)$, for the wave speed c and wave steepness s. For a simple example, written in Maple code see the Appendix.

Here it is assumed that the kernel k is exponentially bounded so that the moment generating function (7) can be calculated. When the kernel k has tails that are fatter than exponential, there is no asymptotic rate of spread—the spread rate continues to increase with increasing time (Kot et al. 1996) (Figure 2, third panel from the top). In this situation, an alternative definition of spread rate, based on the change in location of the furthest forward individual in the population from generation to generation (furthest forward velocity) is appropriate (Clark, Lewis, and Horvath 2001) (see Discussion).

When the growth rate λ is known, but the dispersal kernel is unknown, estimates for population spread rates using equation (6) can vary widely, depending upon the parametric form of the kernel chosen, as with the simulations shown in Figure 2.

When appropriate data are available, this problem can be addressed by means of a nonparametric estimator for the moment generating function (7) which makes no assumption about the form of the underlying kernel. In this case the moment generating function is estimated from raw one-dimensional linear dispersal displacement data, $z_1 \ldots z_N$.

$$M^E(s) = \frac{1}{N}\sum_{i=1}^{N} \exp(sz_i). \tag{8}$$

The superscript is used to indicate an empirical estimate of the moment generating function (Clark, Horvath, and Lewis 2001). Here the dispersal measurements arise from tracking a series of individuals. As a displacement, x_i gives the distance and direction that the $i-th$ individual moves. By convention, leftward movements are assigned negative values. It is assumed that the tracking effort and tracking efficiency remain constant over the entire linear one-dimensional dispersal region (Fujiwara et al. (2004) consider the effects of changes in sampling effort or detection probability). In the case where $z_1 \ldots z_N$ are (nonnegative) distances, rather than displacements, the assumption of a symmetric dispersal kernel (whereby individuals are as likely to move to the left as to the right) leads to

$$M^E(s) = \frac{1}{N}\sum_{i=1}^{N} \cosh(sz_i). \tag{9}$$

Substitution of $M^E(s)$ instead of $M(s)$ in (6) leads to an empirically estimated wave speed c^E. This empirically estimated wave speed has many nice properties (Clark et al. 2001). For example it is unbiased: c^E converges to to the true population spread rate c as

$N \to \infty$. This is not generally the case when parametric kernels are fitted to the dispersal data. In that case, the true population dispersal kernel is not known, and different fitted kernels can give very different wave speed predictions.

When the number of data points N is finite, each empirically estimated wave speed will be different, as it will depend on the precise data set used. However, the distribution of the empirical wave speed c^E about the true wave speed c is approximately Gaussian and the variance of the Gaussian approaches zero as N approaches infinity (Clark et al. 2001). Although there is no closed form expression for the variance, it can be estimated using bootstrapping methods (Clark et al. 2001).

When the data collection or recording methods do not provide dispersal data, but histogram (density) data are available, it may be necessary to use such a histogram as an estimator for the kernel

$$k^H(z) = \begin{cases} f_i & \text{if} \quad \xi_{i-1} \le z < \xi_i \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where f_i is the bin height for the histogram, $1 \le i \le L$, and

$$\sum_{i=1}^{L} (\xi_i - \xi_{i-1}) f_i = 1. \tag{11}$$

This yields

$$M^H(s) = \frac{1}{s} \sum_{i=1}^{L} f_i \left[\exp(s\xi_i) - \exp(s\xi_{i-1}) \right], \tag{12}$$

and substitution of $M^H(s)$ instead of $M(s)$ in (6) leads to a 'histogram' estimator for the wave speed c^H. Due to the arbitrary nature of location of the histogram bins, it can be shown that the histogram estimator does not provide an unbiased estimator for the true speed c. However, in the absence of other data, this histogram estimator is a useful alternative to the empirical estimator given above and, in practice, gives very similar results. As the sizes of the bins $\xi_i - \xi_{i-1}$ approaches zero the two estimators are identical.

In the case where the histogram data is for distances as opposed to displacements, the assumption of a symmetric redistribution kernel causes (12) to be modified to

$$M^H(s) = \frac{1}{s} \sum_{i=1}^{L} f_i \left[\sinh(s\xi_i) - \sinh(s\xi_{i-1}) \right]. \tag{13}$$

We now apply the histogram spread rate estimate (13) and (6) to the (Dobzhansky and Wright 1943) data, under the assumption that the insects and transects are found along a one-dimensional strip of suitable habitat, along which all dispersal and reproduction occurs (see discussion in Section 2). This estimate, which uses the data shown in Figure 2, gives a spread rate of 0.258 km per year. This is higher than the spread rate prediction made by using the exponential and Gaussian kernels in Figure 2, but is substantially lower than the prediction made by the fat tailed kernel. As with the empirical estimator, bootstrapping gives the distribution of wave speeds, from which confidence intervals can be calculated (Figure 3).

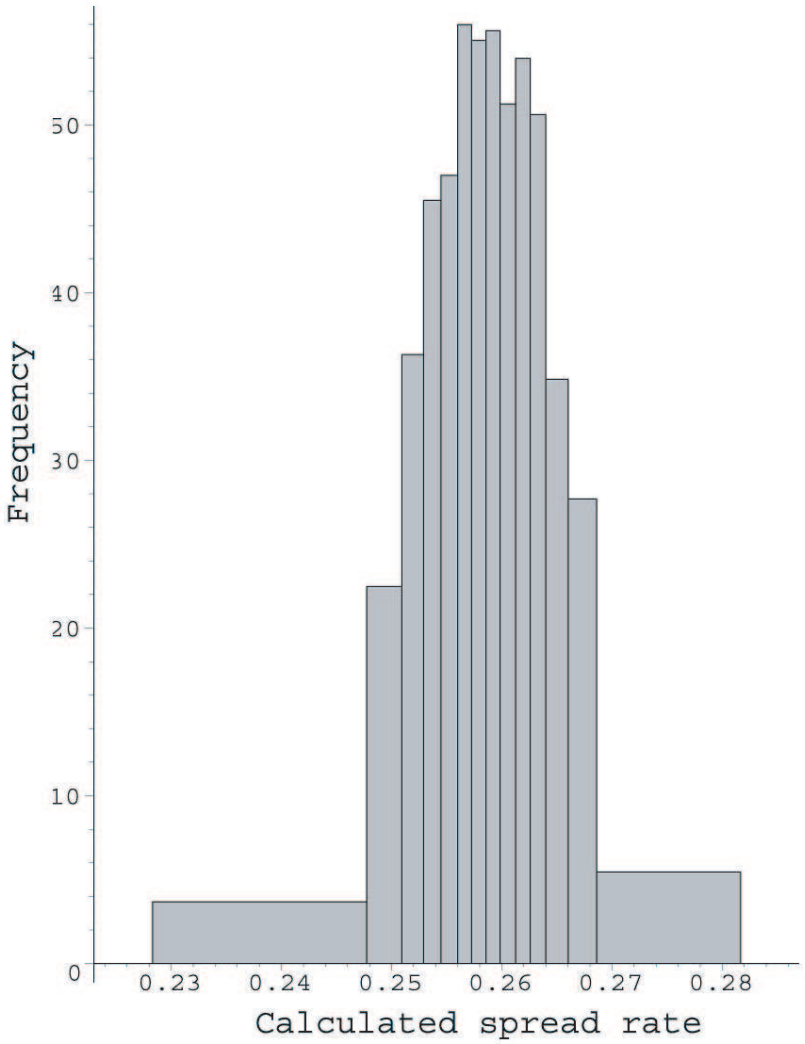

Figure 3: Bootstrapping gives a range of possible values for the histogram wave speed. The histogram wave speed estimator (equations (13) and (6)) is applied to a histogram based on the total number of insects caught per trap, starting at distance zero, finishing at distance 0.38 km, with inter-trap spacing of 0.02 km were: 317, 121, 117, 50, 55, 27, 19, 18, 14, 21, 15, 13, 11, 7, 7, 6, 6, 7, 4, and 2. Bootstrapping was done by re-sampling from the 837 dispersal distances, with replacement, to produce 5000 new data sets. The wave speed was calculated for each of these new data sets, and the distribution of speeds is shown here. Ninety per cent of the speeds fell in the range (0.246 km/year, 0.270 km/year). Compare the estimate here with the spread rates simulated in Figure 2.

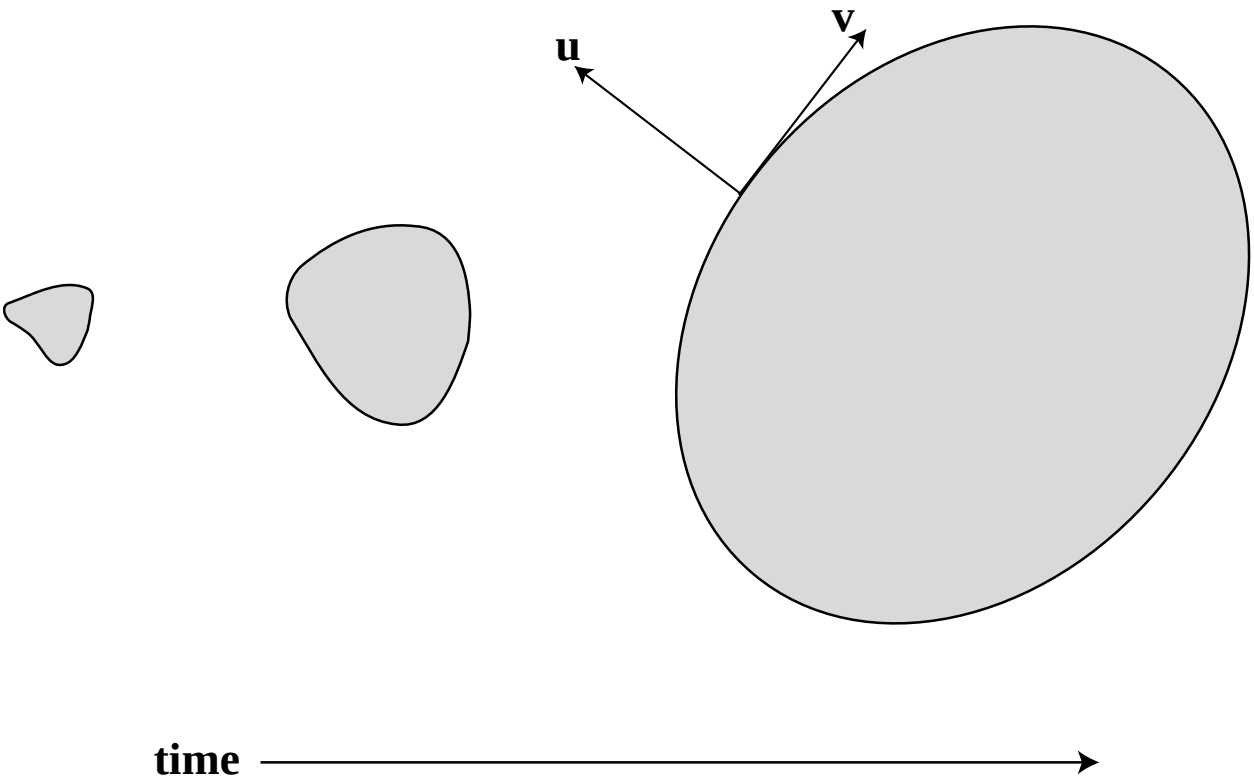

Figure 4: A sketch of the three stages of population spread in a homogeneous environment. Shaded areas indicate invaded habitat. As time progresses, the initial 'beachhead" grows and becomes more elliptically shaped. For long times, the invasion front is approximately planar in all directions. The speed in the direction $\mathbf{u}$, perpendicular to the front, is found using the marginal dispersal kernel in that direction $k_{\mathbf{u}}(u)$ (c. f. equation 28). The marginal kernel, in turn, is found by integrating the original 2-dimensional distribution over the direction $\mathbf{v}$.

5 Population Spread in Two Dimensions

Invasion in two spatial dimensions involves local introduction at a beachhead, followed by growth and spread in space. Here the process can be divided into three stages: the initial establishment, the early radial expansion until well established in space, and the later spread of the established population (Figure 4). Although each of these three stages is of interest biologically, the focus of this chapter is on analyzing the later spread of an established population. At this stage we can approximate the invading front by a planar front, moving with a well-defined speed.

The calculation of spread in two dimensions requires a unit vector $\mathbf{u} = [u_1, u_2]^{\mathrm{T}}$ describing the direction, perpendicular to the wave front, in which the spread is being considered (Figure 4). The asymptotic spread rate in the direction $\mathbf{u}$ is given by

$$c_{\mathbf{u}} = \min_{s>0} \frac{1}{s} \ln \left[\lambda M_{\mathbf{u}}(s)\right], \tag{14}$$

where $M_{\mathbf{u}}(s)$ is the 'directional' moment generating function, the moment generating function of K evaluated in the direction of $\mathbf{u}$ (Appendix). Two approaches for calculation of the directional moment generating function are given in the the Appendix. In general the planar spread rate $c_{\mathbf{u}}$ will depend on the direction $\mathbf{u}$. However, when the dispersal kernel is directionally isotropic (is identical in all directions) the planar spread rate will also be isotropic.

We now consider the case with directional isotropy, so that $K(\mathbf{z})$, $\mathbf{z} = \mathbf{x} - \mathbf{y}$, can be rewritten as $K(r)$, $r = \sqrt{z_1^2 + z_2^2}$. Note that $K(r)$ is a two-dimensional density function which denotes the relative number of seeds *per unit area* falling at distance r from the source. The area under $K(r)$ is equal to one:

$$\int_0^{2\pi} \int_0^{\infty} K(r) r \, dr d\theta = 1. \tag{15}$$

A related kernel denotes the number of seeds *per unit length* falling a distance r from the source. This related kernel is found by multiplying $K(r)$ by the perimeter of a circle of radius r to account for the fact that, at larger radii, there is more available area for seeds to fall $\tilde{K}(r) = 2\pi r K(r)$. If the data have been collected in this form, a simple rescaling by $2\pi r$ will transform $\tilde{K}$ to K.

We now consider how to calculate the planar spread rate (see equation (14)) when the kernel is radially symmetric. We illustrate the two options that are given in the Appendix for calculating the directional moment generating function for this case. Here we consider a wave spreading in the x_1-direction so that $\mathbf{u} = [1, 0]^{\mathrm{T}}$.

1. Evaluate the marginal distribution of K by integrating over the z_2-direction to yield a one dimensional dispersal kernel that describes dispersal in the z_1 direction,

$$k_{\mathbf{u}}(z_1) = \int_{-\infty}^{\infty} K(\sqrt{z_1^2 + z_2^2}) \, dz_2, \tag{16}$$

   and then calculate the directional moment generating function $M_{\mathbf{u}}$ of the kernel $k_{\mathbf{u}}(z_1)$, and thus the speed $c_{\mathbf{u}}$ (14) (see equations (27) and (28) for details). This method effectively reduces the two dimensional spread problem to one spatial dimension, by first taking the marginal distribution of the dispersal kernel, and then proceeding as with the one dimensional case. This approach is conceptually straightforward, but many marginal distributions cannot be calculated analytically, even for simple kernels.

2. Evaluate the moment generating function in the z_1-direction (see equation (26)) directly as

$$\begin{aligned} M_{\mathbf{u}} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(r) \exp(s z_1) \, dz_1 \, dz_2 && (17) \\ &= \int_0^{2\pi} \int_0^{\infty} K(r) \exp(sr \cos\theta) \, r \, dr d\theta && (18) \\ &= 2\pi \int_0^{\infty} K(r) \, r \, I_0(sr) \, dr, && (19) \\ &= \int_0^{\infty} \tilde{K}(r) \, I_0(sr) \, dr, && (20) \end{aligned}$$

   and then use this to calculate of the speed $c_{\mathbf{u}}$ (14). Here I_0 is the modified Bessel function of the first kind and zeroth order (Abramowiz and Stegun 1970).

When there are raw one-dimensional radial dispersal distance data, $r_1 \ldots r_N$, then the empirical moment generating for the planar wave speed calculation (20) becomes

$$M_{\mathbf{u}}^E(s) = \frac{1}{N}\sum_{i=1}^{N} I_0(sr_i), \tag{21}$$

and substitution into (14) gives the empirical estimator for the planar wave speed. Here the r_i measurements arise from tracking a series of individuals as they disperse. It is assumed that the tracking effort per unit area and tracking efficiency per unit area remains constant over the entire dispersal area.

Given a histogram of radial density data

$$K^H(r) = \begin{cases} f_i & \text{if} \quad \rho_{i-1} \le r < \rho_i \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

where $1 \le i \le L$, the constraint that the area under the histogram integrates to one (15) means

$$\pi \sum_{i=1}^{L} (\rho_i^2 - \rho_{i-1}^2) f_i = 1. \tag{23}$$

If the histogram is measured in terms of the relative *number* of dispersers that move a distance r rather than the relative *density*, it is necessary to rescale the kernel by $(2\pi r)^{-1}$, as above.

The directional moment generating function for the histogram comes from substituting (22) into (19) to obtain

$$M_{\mathbf{u}}^H(s) = \frac{2\pi}{s}\sum_{i=1}^{L} f_i(\rho_i I_1(s\rho_i) - \rho_{i-1} I_1(s\rho_{i-1})) \tag{24}$$

(compare with (12)), and substitution into (14) gives the planar wave speed. Here I_1 is the modified Bessel function of the first kind and first order (Abramowiz and Stegun 1970).

For the sake of illustration, Section 4 considered the case where the Dobzhansky and Wright data described insect movement along linear transects of available habitat. Accordingly, the histogram estimator for the spread rate for the situation shown in Figure 2 was calculated, using equations (12) and (6) as 0.258 km per year.

To compare the difference between the results of this section and Section 4, we revisit the Dobzhansky and Wright data, under the more reasonable assumption that the transect data describe the radial drop-off in settled insect density in a two-dimensional habitat. In this case the histogram estimator for the planar spread rate is calculated from equations (24) and (14) as 0.288 km per year, 16% higher than the previous estimate.

Repeating the comparison of spread rates for different λ's shows that the spread rate, assuming radial dispersal and growth in the two-dimensional habitat, is significantly higher than the spread rate, assuming growth and dispersal in a linear one-dimensional habitat, (see Table 1). This difference is most pronounced for a low growth rates λ. This ordering of spread rates (2D speed larger than 1D speed, Table 1) appears only for density data, not dispersal data. In fact, if the original data describe dispersal, then it can be shown that the ordering will be reversed, with the 1D speed (based on growth and dispersal in a linear one-dimensional habitat, eg, equation 8 and 6) larger than 2D speed (based on radial dispersal an growth in a two-dimensional habitat, eg, equation 21 and 14).

λ	1D speed (km/year)	2D speed (km/year)	ratio of speeds
10	0.258	0.288	1.16
4	0.194	0.231	1.19
2	0.130	0.166	1.27
1.1	0.0466	0.0619	1.39

Table 1: Spread rates for varying growth rates λ. Dispersal data are as given by the left panels in Figure 2. In column labeled 1D it is assumed that the growth and dispersal occur in a 1D linear habitat. Spread rates are calculated using equations (12) and (6). In column labeled 2D it is assumed that the dispersal is radial, and that growth and dispersal occur in a 2D habitat. Here the planar spread rate is calculated from equations (24) and (14). The last column shows the ratio of the second to the first columns entries. The difference between the spread rates is most pronounced for low growth rates.

6 Monte Carlo Methods

The simplest estimator of the moment generating function is the empirical estimator calculated from one-dimensional displacement data (equation (8)). It is not only simple to implement numerically, but in many cases is appealingly nonparametric.

It is sometimes useful to generate appropriate 1-D displacement data from some other form, using Monte Carlo methods. The resulting displacements can then be used as input to the empirical estimator.

For example, Caswell et al. (2003) analyzed the invasion rates of several species of European birds, using dispersal data compiled by van van den Bosch et al. (1992). The dispersal data were obtained in the form of histograms of displacement distances, not densities. Thus they were interpreted as giving a set of distances, in two-dimensional space, moved by a set of marked individuals.

Obtaining the appropriate one-dimensional dispersal kernel requires the marginal distribution of the two-dimensional distribution of displacements. This was obtained by generating a large number of random displacement distances from the histogram (assuming a uniform distribution of displacement distance within each histogram bin). Assuming two-dimensional isotropy, each of these distances was assigned a direction uniformly distributed between 0 and 2π. This produced a set of artificial 2-dimensional displacement data, the distance component of which matches the reported histogram. The marginalized distribution was easily generated by taking the z_1 component of each of the points. The resulting set of distances was then input to the empirical moment-generating function to produce an estimate of wave speed.

This Monte Carlo method is appropriate because we know mathematically that, providing the dispersal kernel is exponentially bounded, the empirical estimator for the wave speed is unbiased (ie, approaches the true wave speed as $N \to \infty$). Details on convergence can be found in Clark et al. (2001).

7 Discussion

The focus of this paper is methods to reliably connect population spread rate theory to biological data. There are two features that we focus on: (i) sensitivity of spread rate estimates to model assumptions about long-distance dispersal, and (ii) model fitting issues that arise from fitting two-dimensional dispersal data to one-dimensional models.

The empirical and histogram estimators provide a method to bypass assumptions about long distance dispersal. These are really 'what you see is what you get' estimators. Indeed, formally substituting the observed sum of point dispersal jumps $k(z) = \sum_{i=1}^{N} \delta(z - z_i)$ into equation (7) leads to the empirical moment generating function (8). Whereas the above sum of delta functions provides a poor estimator for the kernel, the process leads to a good estimate for the empirical moment generating function and for the spread rate, which is the quantity of interest. The histogram estimator behaves in a manner which is similar to the empirical estimate, although the location of bins can produce small biases in the spread rate estimate.

We recommend that the empirical and histogram spread rate estimators should be a basic tool for any modeling exercise which involves spread rates where long-distance dispersal plays a role. Fitted parametric dispersal kernels may be preferable when there is reason to believe that extrapolation of the dispersal function beyond the furthest observed dispersal distance can be justified. However, even in this case, it is useful to compare spread rate results from the parametric fitted kernel with the empirical or histogram spread rates. The difference between the spread rates will highlight the implications of assumptions about long-distance dispersal distances that go into the parametric kernel.

The new model fitting issue we focus on is the correct method to calculate planar population spread from radially symmetric dispersal data. Naively fitting the linear one-dimensional dispersal kernel to the radially symmetric dispersal data will give the wrong spread rate when the one-dimensional spread model is used, biased downward for density data, and upward for dispersal data.

In a linear one-dimensional environment, the scaled distribution of densities (dispersers per unit length) and a distribution of distances that the dispersers travel from the parent are the same and are given by the kernel k. In a two-dimensional environment, the scaled distribution of densities (dispersers per unit area), given by K, and the distribution of distances that dispersers travel from from the parent, given by $\tilde{K} = 2\pi r K(r)$, are not the same, because there is more area available at distances further from the parent.

The correct method for calculating the spread in the two-dimensional environment involves a directional moment generating function of the two-dimensional dispersal kernel. This is found either by evaluating the moment generating function of the marginal distribution of the dispersal kernel $K(r)$, or by using $\tilde{K}(r)$ in a modified moment generating function calculation.

Data	MGF $[M(s)]$	Equation
parametric kernel	$\int_{-\infty}^{\infty} k(z)\exp(sz)\,dz$	7
dispersal displacement data	$\frac{1}{N}\sum_{i=1}^{N}\exp(sz_i)$	8
displacement distance data	$\frac{1}{N}\sum_{i=1}^{N}\cosh(sz_i)$	9
displacement histogram	$\frac{1}{s}\sum_{i=1}^{L} f_i\left(\exp(s\xi_i)-\exp(\xi_{i-1})\right)$	12
distance histogram	$\frac{1}{s}\sum_{i=1}^{L} f_i\left[\sinh(s\xi_i)-\sinh(s\xi_{i-1})\right],\ \xi_0=0$	13

Table 2: Methods to calculate the moment generating function needed for the spread rate calculation (equation (6)) when the population lives in a one-dimensional linear habitat.

To see the difference between the linear one-dimensional histogram density kernel and the marginal distribution of it's radially symmetric analog, we can compute the marginal distribution of the radially symmetric histogram density kernel (22) directly

$$k_{\mathbf{u}}^{H}(z_1) = \begin{cases} 2\sum_{i=j}^{L}(f_i - f_{i+1})\sqrt{\rho_i^2 - z_1^2} & \text{if } \rho_{j-1} \le |z_1| < \rho_j \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

The two kernels $k^H(z)$ (10) and $k_{\mathbf{u}}^H(z)$ (25) are shown in Figure 5. As can be seen from the results in Table 1, the increase in number of long-distance dispersers in the marginal distribution speeds up the spread rate significantly.

The correct formulae for calculating spread rates are given in Tables 2 and 3. To the best of our knowledge, these formulae have not been widely developed or discussed elsewhere in the literature. It is our hope that these tables, along with the discussion in the chapter will form a user's guide to calculating spread rates for discrete-time models, and, in turn, will lead to a closer connection between ecological theory and data.

When the dispersal data come from a 'fat-tailed' kernel (kernel with no moment generating function), theory predicts a constantly accelerating wave with an asymptotically infinite speed (Section 4.1). However, the spread rates based on the empirical moment generating function (eg, equation 8) or the histogram moment generating function (eg, equation 12) are finite. This is because the these spread rates are calculated from a finite number of dispersal observations or on densities measured over a finite region. Here additional sampling effort may find rare, long-distance dispersers or may measure nonzero densities of dispersers far from their natal source, and thus the spread rate estimate can increase with sampling effort.

When there is good reason to believe that the dispersal kernel is fat-tailed or nearly fat-tailed, a more appropriate measure of spread rate is the 'furthest-forward' velocity (Clark, Lewis, and Horvath 2001) which remains bounded. A more general discussion of uncertainty in spread rates associated with long-distance dispersal is given in Clark et al. (2003).

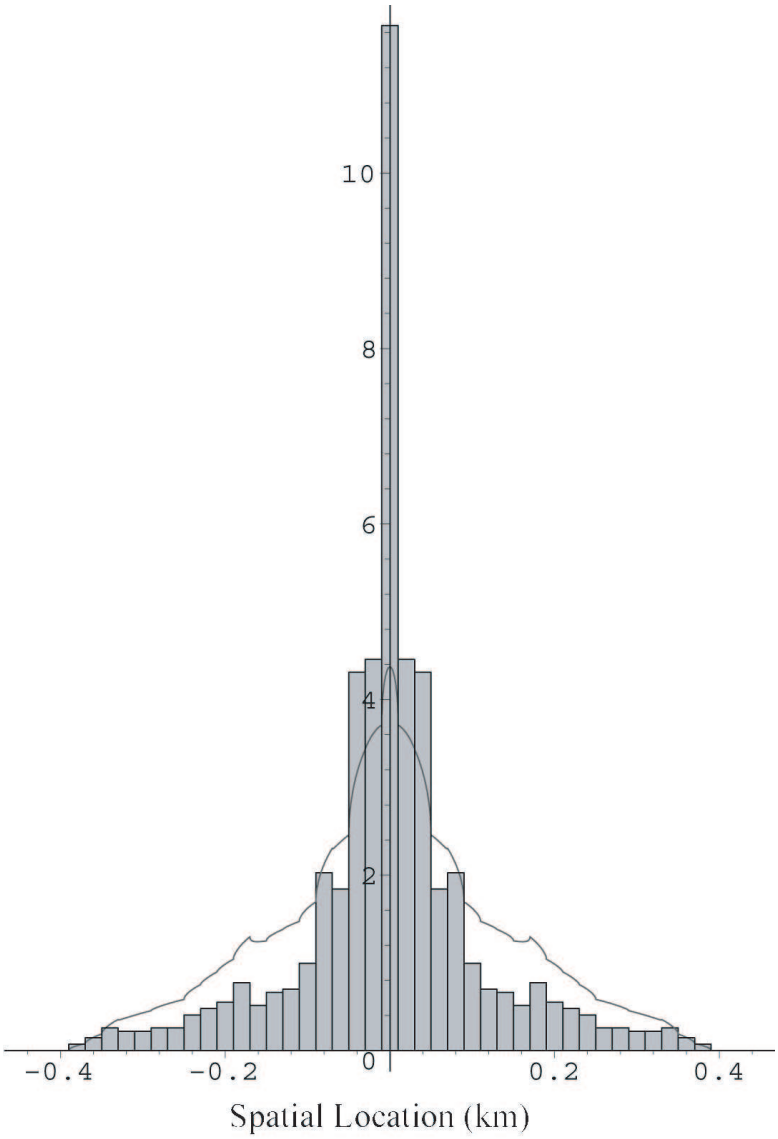

Figure 5: Histogram dispersal kernel (equation 10) forDobzhansky and Wright's *Drosophila* dispersal data under the assumption that dispersal occurs in a linear one-dimensional habitat (grey). Marginal distribution (equation (25) of histogram dispersal kernel (equation 22) under the assumption that dispersal occurs in a two-dimensional habitat and is radially symmetric.

Data	MGF $[M_x(s)]$	Equation
parametric 2D kernel	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(\mathbf{z})\exp(sz_1)\,dz_1dz_2$	26
parametric radial kernel	$2\pi\int_0^{\infty} k(r)\,r\,I_0(sr)\,dr$	19
radial displacement data	$\frac{1}{N}\sum_{i=1}^{N} I_0(r_i s)$	21
radial histogram data	$M_x^H(s)=\frac{2\pi}{s}\sum_{i=1}^{L} f_i(\rho_i I_1(s\rho_i)-\rho_{i-1}I_1(s\rho_{i-1}))$	24

I_0 and I_1 are modified Bessel functions of zero$^{\text{th}}$ and first order (cf. Abramowiz and Stegun (1970)).

Table 3: Methods to calculate the directional moment generating function needed for the planar spread rate calculation (equation (14)) when the population lives in a two-dimensional habitat.

Appendix

Calculating the Spread Rate with Maple

The following is Maple code that can be used to calculate the spread rate for a population with geometric growth rate λ, and the composite Laplace dispersal kernel $k(z) = p\alpha_1 \exp(-\alpha_1|z|) + (1-p)\alpha_2 \exp(-\alpha_2|z|)$.

```
Digits:=20;
#
# Work out speed gam for the following parameter values
#
lambda:=1.2;
alpha_1:=10.;
alpha_2:=1.0;
p:=0.99;
#
# define the moment generating function and exponential
#
v_1:=alpha_1^2/(alpha_1^2-s^2);
v_2:=alpha_2^2/(alpha_2^2-s^2);
fn1:=exp(s*gam);
fn2:=lambda*(p*v_1+(1-p)*v_2);
#
# Find the double root
#
eq1:=fn1=fn2;
eq2:=diff(fn1,s)=diff(fn2,s);
fsolve({eq1,eq2},{s,gam},s=0..alpha_2);
#
```

The Directional Moment Generating Function

The directional moment generating function is

$$M_{\mathbf{u}}(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(\mathbf{z}) \exp[s\mathbf{u}\cdot\mathbf{z}]\, dz_1\, dz_2. \tag{26}$$

Recall that the dispersal kernel $K(\mathbf{z})$ now describes the probability density for jumps whose directions and magnitude are described by the vector $[z_1, z_2]^{\mathrm{T}}$. The term $\mathbf{u}\cdot\mathbf{z} = u_1 z_1 + u_2 z_2$ in equation (26) is the component of the dispersal jump $\mathbf{z}$ that lies in the $\mathbf{u}$-direction. If the kernel has no directional bias (is isotropic), depending only upon distance (and hence $K(\mathbf{z}) = K(\sqrt{z_1^2 + z_2^2})$), then the moment generating function $M_{\mathbf{u}}(s)$ and hence the speed $c_{\mathbf{u}}$ is independent of the direction vector $\mathbf{u}$. We consider this case in Section 5.

The 'directional' moment generating function $M_{\mathbf{u}}$ can be interpreted as the moment generating function of the kernel $K(\mathbf{z})$ marginalized in the direction of the unit vector $\mathbf{v}$ which is perpendicular to $\mathbf{u}$

$$M_{\mathbf{u}}(s) = \int_{-\infty}^{\infty} k_{\mathbf{u}}(u) \exp(su)\, du. \tag{27}$$

where

$$k_{\mathbf{u}}(u) = \int_{-\infty}^{\infty} K(\mathbf{z}) dv. \tag{28}$$

and $u = \mathbf{u} \cdot \mathbf{z}$, $v = \mathbf{v} \cdot \mathbf{z}$.

References

Abramowiz, M. and I. Stegun (1970). *Handbook of Mathematical Functions*. Dover Publications Inc.

Allen, E. J., L. J. S. Allen, and X. Gilliam (1996). Dispersal and competition models for plants. *Journal of Mathematical Biology 34*, 455–481.

Allen, L. J. S., E. J. Allen, and S. Ponweera (1996). A mathematical model for weed dispersal and control. *Bulletin of Mathematical of Biology 58*(5), 815–834.

Andow, D. A., P. M. Kareiva, S. A. Levin, and A. Okubo (1990). Spread of invading organisms. *Landscape Ecology 4*, 177–188.

Aronson, D. G. and H. F. Weinberger (1975). Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation. In J. A. Goldstein (Ed.), *Lecture Notes in Mathematics*, Volume 446, pp. 5–49. Berlin: Springer-Verlag.

Caswell, H., R. Lensink, and M. G. Neubert (2003). Demography and dispersal: Comparing invasion speeds using life table response experiments. *Ecology 84*, 1968–1978.

Clark, J. S. (1998, August). Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord. *The American Naturalist 152*(2), 204–224.

Clark, J. S., C. Fastie, G. Hurtt, S. T. Jackson, C. Johnson, G. A. King, M. Lewis, J. Lynch, S. Pacala, C. Prentice, E. W. Schupp, T. W. III, and P. Wycoff (1998, January). Reid's paradox of rapid plant migration. *BioScience 48*(1), 13–24.

Clark, J. S., L. Horvath, and M. A. Lewis (2001). On the estimation of spread rate for a biological population. *Statistics and Probability Letters 51*, 225–234.

Clark, J. S., M. A. Lewis, and L. Horvath (2001). Invasion by extremes: Population spread with variation in dispersal and reproduction. *American Naturalist 157*, 537–554.

Clark, J. S., M. A. Lewis, J. S. McLachlan, and J. H. RisLambers (2003). Estimating population spread: What can we forecast and how well? *Ecology 84*, 1979–1988.

Dobzhansky, T. and S. Wright (1943). Genetics of natural populations. x. Dispersion rates in *drosophila pseudoobscura*. *Genetics 27*, 304–340.

Fisher, R. A. (1937). The wave of advance of advantageous genes. *Ann. Eugen. London, 37*, 355–369.

Fujiwara, M., K. Anderson, H. Caswell, and M. G. Neubert (2004). On the esitmation of dispersal kernels from individual capture-recapture data. *Environmental and Ecological Statistics In press*.

Hart, D. R. and R. H. Gardner (1997). A spatial model for the spread of invading organisms subject to competition. *J. Math. Biol. 35*, 935–948.

Higgins, S. I. and D. M. Richardson (1999). Predicting plant migration rates in a changing world: the role of long-distance dispersal. *Am. Nat. 153*, 464.

Horvitz, C. C. and D. W. Schemske (1986). Seed dispersal of a neotropical myrmecochore: variation in removal rates and dispersal distance. *Biotropica 18*, 319–323.

Kolmogorov, A., Petrovsky, and N. I. Piscounov (1937). Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application a un problème biologique. *Bull. Moscow Univ. Math. Mech 1*(6), 1–26.

Kot, M. (1992). Discrete-time travelling waves: Ecological examples. *J. Math. Biol. 30*, 413–436.

Kot, M., M. A. Lewis, and P. van den Driessche (1996). Dispersal data and the spread of invading organisms. *Ecology 77*(7), 2027–2042.

Lewis, M. A. (1997). *Variability, Patchiness, and Jump Dispersal in hte Spread of an Invading Population*, Chapter 3, pp. 46–69. Princeton, New Jersey: Princetin Univesity Press.

Lubina, J. A. and S. A. Levin (1988). The spread of a reinvading species: range expansion in the California Sea Otter. *American Naturalist 151*(4), 526–543.

Lui, R. (1982a). A nonlinear integral operator arising from a model in population genetics, I. 1cmonotone initial data. *SIAM J. Math. Anal. 13*(6), 913–937.

Lui, R. (1982b). A nonlinear integral operator arising from a model in population genetics, II Initial data with compact support. *SIAM J. Math. Anal. 13*(6), 938–953.

Lui, R. (1985). A nonlinear integral operator arising from a model in population genetics, III. Initial data with compact support. *SIAM J. Math. Anal. 16*(6), 1180–1206.

Lui, R. (1986). A nonlinear integral operator arising from a model in population genetics IV. Clines. *SIAM J. Math. Anal. 17*(1), 152–168.

Lui, R. (1989a). Biological growth and spread modeled by systems of recursions. I Mathematical theory. *Mathematical Biosciences 93*, 269–295.

Lui, R. (1989b). Biological growth and spread modeled by systems of recursions. II Biological theory. *Mathematical Biosciences 93*, 297–312.

Marchant, S. (2003). Analysis of an integrodifference model for biological invasions with a quasi-local interaction. Master's thesis, University of British Columbia.

Neubert, M. G. and H. Caswell (2000). Demography and dispersal: Calculation and sensitivity analysis of invasion speed for stage-structured populations. *Ecology 81*, 1613–1628.

Neubert, M. G., M. Kot, and M. A. Lewis (2000). Invasion speed in fluctuating environments. *Proc. Roy. Soc. Lond. B 267*, 1603–1610.

Neubert, M. G. and I. Parker (2004). Using integrodifference equations to project rates of spread for invasive species. *Risk Analysis*, in press.

Pachepsky, E., F. Lutscher, and M. A. Lewis (2004). Persistence, spread and the drift paradox. *Theoretical Population Biology*, in press.

Powell, J. and N. E. Zimmermann (2004). Analysis of active seed dispersal contributes to resolving Reid's paradox. *Ecology 85*, 490–506.

Schofield, P. (2002). Spatially explicit models of Turelli-Hoffmann *wolbachia* invasive wave fronts. *Journal of Theoretical Biology 215*, 121–131.

Shigesada, N. and K. Kawasaki (1997). *Biological Invasions: Theory and Practice*. Oxford: Oxford University Press.

Shigesada, N., K. Kawasaki, and E. Teramoto (1986, August). Traveling periodic waves in heterogeneous environments. *Theor. Pop. Biol. 30*(1), 143–160.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Skellam, J. G. (1951). Random dispersal in theoretical populations. *Biometrika 38*, 196–218.

Speirs, D. C. and W. S. C. Gurney (2001). Population persistence in rivers and estuaries. *Ecology 82*, 1219–1237.

Takasu, F., N. Yamamoto, K. Kawasaki, K. Togashi, and N. Shigesada (2000). Modeling the range expansion of an introduced tree disease. *Biological Invasions 2*, 141–150.

van den Bosch, F., R. Hengeveld, and J. A. J. Metz (1992). Analysing the velocity of animal range expansion. *Journal of Biogeography 19*, 135–150.

Veit, R. R. and M. A. Lewis (1996, August). Disperal, population growth and the allee effect: Dynamics of the house finch invasion of eastern north america. *The American Naturalist 148*(2), 255–274.

Weinberger, H. F. (1982, May). Long–time behavior of a class of biological models. *SIAM journal on mathematical analysis 13*(3), 353–396.

Weinberger, H. F. (1984). Long-time behaviour of a class of biological models. In W. E. F. III (Ed.), *In: Research Notes in Mathematics No. 101*, pp. 323–352. Marshfield, MA: Pitman Publishing Ltd.

Werner, P. A. (1975). A seed trap for determining patterns of seed deposition in terrestrial plants. *Canadian Journal of Botany 53*, 810–813.

With, K. A. (2002). The landscape ecology of invasive spread. *Conservation Biology 16*, 1192.

Woolcock, J. L. and R. Cousens (2000). A mathematical analysis of factors affectng the rate of spread of annual weeds in an arable field. *Weed Science 48*, 141–150.