

ESTIMATING POPULATION SPREAD: WHAT CAN WE FORECAST AND HOW WELL?

JAMES S. CLARK,^{1,3} MARK LEWIS,² JASON S. MCLACHLAN,¹ AND JANNEKE HILLERISLAMBERS¹

¹Department of Biology and University Program in Ecology, Duke University, Durham, North Carolina 27708 USA

²Department of Mathematical Sciences and Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

Abstract. Recent literature on plant population spread advocates quantification of long-distance dispersal (LDD). These estimates could provide insights into rates of migration in response to climate change and rates of alien invasions. LDD information is not available for parameterization of current models because it is hard to obtain. We combine a new stochastic model with a flexible framework that permits assimilation of evidence that might be derived from a range of sources. Results are consistent with the prediction of traditional diffusion that population spread has a finite asymptotic velocity. Unlike traditional diffusion, spread is not well described by the mean; it is erratic.

In contrast with deterministic models, our results show that inherent uncertainty, rather than parameter sensitivity, thwarts informative forecasts of spread velocity. Analysis shows that, because LDD is inherently unpredictable, even full knowledge of LDD parameters might not provide informative estimates of velocity for populations characterized by LDD. Although predictive distributions are too broad to provide precise estimates of spread rate, they are valuable for comparing spread potential among species and for identifying potential for invasion.

Using combinations of dispersal data and the estimates provided by dispersal biologists that derive from multiple sources, the model predicts spread rates that are much slower than those from traditional (deterministic) fat-tailed models and from simulation models of spread, but for different reasons. Deterministic fat-tailed models overestimate spread rate, because they assume that fractions of individuals can rapidly occupy distant sites. Stochastic models recognize that distant colonization is limited to discrete individuals. Stochastic simulations of plant migration overestimate migration of trees, because they typically assume values of R_0 that are too large.

Key words: Bayes; climate change; dispersal; invasion; migration; population spread.

INTRODUCTION

Increasing efforts to estimate long-distance dispersal (LDD) is an understandable reaction to growing realization that rare events can control the rate of population spread (Kot et al. 1996, Clark 1998, Higgins and Richardson 1999, Cain et al. 2000, Clark et al. 2001c). Scenarios of climate change and spread of exotic species are based on dispersal estimates and migration potential. Models used to calculate spread require estimates of offspring production (e.g., net reproductive rate R_0), a dispersal kernel, and a time scale. The dispersal kernel describes the scatter of offspring about the parent plant in the form of a probability density function. Estimation requires a full accounting of seed in terms of proportions of R_0 offspring that travel various distances. The time scale is controlled by schedules of fecundity, mortality, and growth and is often summarized in models by generation time T .

Before investing heavily in the study of LDD, it is worth asking how well such data would allow us to anticipate spread. Dispersal studies typically target a particular vector, such as extreme winds (Snow et al. 1995), vertebrates (Storm and Montgomery 1975, Johnson et al. 1997, Shilton et al. 1999), and flowing waters (Thebaud and Debussche 1991, Kubitzki and Ziburski 1994) and a spatial scale consistent with that vector. But traditional migration models require a probability density function describing dispersal of all seed used to estimate R_0 . Observations of individual or groups of LDD events can be used in these models only if they are combined with dispersal estimates for all seeds produced over the lifetime of a “typical” plant; it is not enough to know that some seeds could travel long distances. Nor is it obvious how studies could be designed to simultaneously estimate production and dispersal for seed borne by a range of vectors over a range of distance classes over the lifetimes of perennial plants and the range of conditions encountered over multiple generations.

The potential value of LDD information further depends on sensitivity. If traditional deterministic models are accurate, then the extreme sensitivity of spread ve-

Manuscript received 6 November 2001; revised 16 August 2002; accepted 3 September 2002. Corresponding Editor: M. L. Cain. For reprints of this Special Feature, see footnote 1, p. 1943.

³ E-mail: jimclark@duke.edu

locity to the shape of the kernel tail (Kot et al. 1996, Clark 1998) means that any additional information on LDD could add substantially to predictive capacity. But new models that are stochastic (because they recognize random behavior of discrete individuals) do not predict this extreme sensitivity (Clark et al. 2001c). Rather, these models suggest that velocity of spread is inherently unpredictable (see also Mollison 1972, Lewis 1997, 2000). Thus, even full knowledge of a dispersal kernel might not permit an informative forecast. A model structure is needed that (1) allows for assimilation of accumulating data sets that come from a range of sources, including studies of different dispersal vectors, distance classes, locations, and environmental conditions; and (2) permits analysis of stochastic population spread based on these different sources.

Here, we integrate new stochastic models of population spread with the types of LDD information that ecologists are likely to collect. Such data will rarely (perhaps never) be available for full parameterization of the kernel, because we cannot track all LDD events and not simultaneously for all seeds. We anticipate that research will continue to target specific vectors and distance classes. They will derive from multiple sources. To place our estimates in the context of population spread, we develop a framework for assimilating different types of observational and experimental data in models that can be used to gauge the impact of LDD on spread potential. A key element of the approach is a (multinomial) distribution of probabilities for seeds dispersed by different vectors that fall in different distance classes. The framework permits assimilation of new evidence as it accumulates. To illustrate the approach, our application of the model makes use of survey data from dispersal biologists, who, in turn, rely on knowledge of specific vectors and distance classes.

The analysis demonstrates two aspects of invasion rate that are not apparent from the existing literature. First, in contrast with populations that are not characterized by LDD (primarily vertebrates), plant life history makes spread velocity inherently variable. In this context, inherent uncertainty is associated with the stochastic aspects of population spread and is distinct from parameter uncertainty, which depends, in large part, on quantity of data (Clark et al. 2001a). When R_0 is large and combines with a fat-tailed kernel, informative forecasts are not feasible. No amount of LDD sampling can overcome the inherent uncertainty of spread. This analysis provides background on the degree of insight ecologists can expect to gain from extensive study of LDD. Second, use of deterministic models and misconceptions about the meaning of R_0 have resulted in overly optimistic estimates of spread potential. For many species, spread in response to climate change could be much slower than many models now predict.

METHODS

Our approach combines dispersal data that could be obtained by a range of approaches within stochastic

models that can accommodate LDD. We first summarize classical methods for estimating spread rates. We then describe two ways to estimate the rate of spread when dispersal is fat-tailed; these are the basis for the generalized approach derived here. The first method is parametric, stochastic, and uses extreme value distributions derived from fitted dispersal kernels to determine the velocity of the furthest-forward individual in an expanding population (Clark et al. 2001c). This method permits a solution in the common situation where a fitted kernel is available, but raw data are not. The tail of the fitted dispersal kernel may be poorly identified by data (Kot et al. 1996, Clark 1998), but, unlike deterministic models, the predicted velocity is not extremely sensitive to it (Clark et al. 2001c). The second method is nonparametric, deterministic, and does not make explicit assumptions about the tail shape, proceeding directly from raw data to an estimate of velocity (Clark et al. 2001b). It has the disadvantages that (1) raw data are typically only available to the investigator, and (2) the estimate is sensitive to the amount of data collected, particularly those collections that are most distant. The two methods are complementary, and both are accommodated by the generalized approach derived here.

We develop a framework for integrating data with more qualitative information on extreme dispersal that can have a range of sources. The approach permits assimilation of new insights as they accumulate. Finally, we integrate the two ways to estimate spread within this general framework.

Assumptions of spread potential with LDD

Classical diffusion models can provide useful predictions of spread for populations not characterized by LDD. The aggregation of animals can act against the establishment of outlying populations, thus promoting a diffusive pattern of spread. Simple diffusion adequately characterizes observations of spread for organisms such as muskrats (Skellam 1951), sea otters (Lubina and Levin 1988), House Finches (Okubo 1988, Wikle 2002) and some insects. Andow et al. (1990) provide an overview. Diffusion will continue to provide a valuable paradigm for spatial dynamics of populations for which dispersal is local (Okubo 1980).

LDD is common for plants, perhaps more so than for animals. Like the spread of epidemics in humans (Mollison 1977), LDD has precipitated model development that admits more variable dispersal patterns and can be addressed with a diffusion model. These variable patterns have given rise to new classes of models where spread occurs on multiple scales. Such models employ concepts such as stratified diffusion or coalescing colonies in order to explain the spatial patterns of spread (Shigesada and Kawasaki 1997). Existing models of plant population spread with climate change use one of several assumptions:

1) Assume an equilibrium between climate and vegetation. Because LDD data are unavailable, models assume that vegetation tracks climate change (Neilson 1993, Sykes et al. 1996, Iverson and Prasad 1998, McKenney-Easterling et al. 2000). These studies are intended to estimate potential changes in distribution. Some studies explore the magnitude of migration lags with simulations based on contrasting assumptions that vegetation is in equilibrium with changing environmental conditions vs. those with no migration (e.g., Kirilenko et al. 2000, Van Minnen et al. 2000).

2) Design dispersal kernels to give the fossil pollen answer. The extremely high rates of spread inferred from paleoecological data (100–1000 m/yr) suggest potential for LDD (Davis 1986, Johnson 1989). To insure that models predict rates this high, dispersal functions are designed such that, when embedded in simulations, they produce the rates inferred from fossil pollen (Malanson and Cairns 1997, Iverson et al. 1999, Van Minnen et al. 2000).

3) Use “literature” estimates. The literature does not contain dispersal kernels with tails that are well supported by data. A number of models (e.g., Dyer 1995, Collingham et al. 1996, He and Mladenoff 1999, Higgins and Richardson 1999) base dispersal kernels on forestry or other publications as a first approximation, but generally modify these in some way.

Classical methods for estimating wave speeds

Early models for population spread used partial differential equations (Fisher 1937). The expected spread rate is given compactly as $c = 2\sqrt{rD}$, where $r = \ln R_0/T$ is population growth rate, and D is a diffusion coefficient. Skellam (1955) demonstrated that this model apparently underestimates prehistoric tree spread by an order of magnitude. Relaxing the assumption of diffusive (Brownian) motion yields a new class of models that can include rare, LDD events. When expressed in the deterministic format these models give estimates of spread that are high (Kot et al. 1996), sometimes even higher than paleodata reconstructions (Clark et al. 1998).

Consider a population having expected lifetime seed production of R_0 seeds, the average offspring of which has mean parent age of T years (T is generation time), and a dispersal kernel $f(x)$. The integrodifference equation model of population growth,

$$N(x, t + T) = \int_{-\infty}^{\infty} R_0 N(y, t) f(x - y) dy \quad (1)$$

says that density N at a location x one generation from now results from seed production that disperses to x from all other locations. For exponentially bounded kernels (Wienberger 1982, see also Kot et al. 1996) the wave speed that can be estimated as

$$c = Z(s')$$

where $Z(s') = 1/s \ln[R_0 M(s)]$, and where

$$M(s) = \int_{-\infty}^{\infty} f(x)e^{sx} dx \quad (2)$$

is the moment generating function (MGF) for the dispersal kernel. s' is the value of s that minimizes the function $Z(s)$, i.e., $\partial Z(s)/\partial s|_{s'}$. However, when tails of the kernel are “fat” (not exponentially bounded), accelerating invasions are predicted (Kot et al. 1996), and speeds are asymptotically infinite. We extend the integrodifference wave speed calculation to include a nonparametric estimate of the kernel from raw dispersal data. In this case, asymptotically infinite speeds are not possible.

Two methods to estimate wave speed

A recent parametric method uses information on seed production, dispersal, and generation time to estimate the change in the location of furthest forward individuals of a population. For a population that is neither growing nor declining, $R_0 = 1$. Population spread requires that $R_0 > 1$ somewhere outside the current population range. For a population expanding “to the right,” we track the successive locations of the furthest forward individual. The velocity of spread, in terms of the furthest forward individual derives from a density of extreme values from a single parent, is

$$\varphi(c) = R_0 f(c) \left[\int_{-\infty}^c f(y) dy \right]^{R_0 - 1} \quad (3)$$

This density describes the probability density that one individual will arrive at location c and remaining ($R_0 - 1$) individuals settle somewhere to the left of c . The leading R_0 coefficient normalizes the density, representing the number of ways in which we could obtain this density (there are R_0 different candidates for the extreme disperser).

This model is stochastic because C is a random variate drawn at each generation from $\varphi(c)$. The stochasticity describes “inherent” uncertainty in population spread, because increased knowledge of parameters does not reduce variability beyond that contained in $\varphi(x)$. If the furthest forward individual in one generation gives rise to the furthest forward individual at the next generation, the expected velocity of spread is

$$E[C] = \frac{1}{T} \int_{-\infty}^{\infty} xp(x) dx \approx \frac{1}{T} \sqrt{\frac{\pi u R_0}{2}} \quad (4)$$

where the kernel is taken to be the two-dimensional t ($2Dt$) dispersal kernel which fits seed dispersal with kernel parameter u (in m^2):

$$f(x) = \frac{1}{\pi u \left(1 + \frac{x^2}{u} \right)^2}$$

(Clark et al. 1999). This estimate can be generalized

to the case where an individual other than that furthest forward gives rise to the next generation (Clark et al. 2001c). The predicted rate of spread lies above that predicted by diffusion models (Skellam 1951), but below that predicted by fat-tailed deterministic models (Kot et al. 1996). Clark et al. (2001c) examine consequences of variability in R_0 on the stochastic model (Eq. 3).

The nonparametric extension of the deterministic wave speed estimate described in Eqs. (1) and (2) uses the empirical moment generating function (MGF),

$$M_n(s) = \frac{1}{n} \sum_{i=1}^n \exp[sX_i] \tag{5}$$

for dispersal distances $\{X_i\}$, $i = 1, \dots, n$ and, for the discrete MGF,

$$M_m(s) = \sum_{j=1}^m f(x_j) \exp[sx_j] \tag{6}$$

where $f(x_j)$ is the fraction of n seeds recovered in the j th distance interval, x_j is the distance to the j th interval, and there are m total intervals. The wave speed can now be determined from the MGF, i.e.,

$$c_n = Z_n(s') \tag{7a}$$

$$c_m = Z_m(s') \tag{7b}$$

for empirical and discrete distributions, respectively. Simulation shows that the two methods result in similar estimates of wave speed.

Both parametric (Eq. 2) and nonparametric (Eqs. 5, 6) versions of this model are deterministic, because they do not include environmental or individual variability in the calculation of the spread rate c . The model assumes that a plant “samples” the environment R_0 times, but uncertainty in the point estimate $E[c_n]$ depends on n , not R_0 . The uncertainty in c vanishes as sample size n becomes large (Clark et al. 2001b provide asymptotics and recommend a nonparametric bootstrap). Unlike the stochastic version (Eq. 3), there is no “inherent” uncertainty in this model. Thus, we do not use this model to assess the relative importances of inherent vs. parameter uncertainty.

Estimating the effect of LDD events

Because they are rare, LDD events cannot be estimated precisely. We describe a method for the definition and updating of density functions that define the probabilities θ_j for a multinomial kernel, having distance classes j . Based on observations of foraging and/or caching behavior (Johnson et al. 1997), gut passage times (Yumoto et al. 1999), or observations following storms (Snow et al. 1995) we could define broad classes and assign tentative probabilities. Our method permits successive refinement of initial θ_j estimates to accommodate accumulating evidence. Dispersal distances are summarized by a dispersal kernel with discrete classes. This kernel provides a flexible structure for analysis of

spread. Products of this analysis include posterior densities of parameters θ_j and estimates of wave speed c .

A semiformal summary.—Following a slightly technical outline of the underlying theory, we provide a simple recipe for application. Consider a sample of n seeds described by the discrete multinomial dispersal kernel $X \sim \text{Multinom}(n, \theta_1, \dots, \theta_m)$ with likelihood

$$p(\mathbf{x} | \theta) \propto \prod_{j=1}^m \theta_j^{x_j} \tag{8}$$

where parameters θ_j sum to 1 for m distance classes, and x_j is the number of seeds that settle in class j . A Bayesian framework for the analysis is based on the conjugate multinomial/Dirichlet likelihood/prior and provides a basis for calculation of wave speeds that integrate traditional seed trap data and extreme LDD estimate information. We assume that distance categories are ordered from near to far, with $j = m$ being most distant to the right. Densities of probability classes are Dirichlet, the prior being $\theta \sim \text{Dirichlet}(a_1, \dots, a_m)$ or

$$p(\theta) \propto \prod_{j=1}^m \theta_j^{a_j-1}. \tag{9}$$

The marginal prior for the j th distance class is Beta,

$$\theta_j \sim \text{Beta}\left(a_j, \sum_i a_i - a_j\right). \tag{10}$$

A noninformative prior is obtained by setting $a_j = 1$ for all j . The posterior is

$$\theta | \mathbf{x} \sim \text{Dirichlet}(A_1, \dots, A_m) \tag{11}$$

where parameters

$$A_j = a_j + x_j \tag{12}$$

represent the sum of prior and data contributions to class j . The marginal posterior for class j is

$$p(\theta_j | \mathbf{x}) = \text{Beta}\left(\theta_j | A_j, \sum_{i=1}^m A_i - A_j\right). \tag{13}$$

If the fitted dispersal kernel $f(x)$ is available, then that kernel can be included within this framework, becoming, say, the conditional density of distance x given that a seed settles in the first distance class with probability θ_1 . Then the (marginal) density for a distance $x | (x \in x_1)$ is

$$\begin{aligned} f(x) & \int_0^1 \text{Bin}(x_1 | n, \theta_1) \text{Beta}\left(\theta_1 | a_1, \sum_i a_i - a_1\right) d\theta_1 \\ & = f(x) \int_0^1 \theta_1^{x_1+a_1-1} (1 - \theta_1)^{n-x_1+\sum_i a_i - a_1-1} d\theta_1 \\ & = f(x) \text{BetaBin}\left(x_1 | n, a_1, \sum_i a_i - a_1\right) \end{aligned} \tag{14}$$

where

$$\begin{aligned}
 & \text{BetaBin}(x_1 | n, a, b) \\
 &= \frac{\Gamma(n + 1)}{\Gamma(x_1 + 1)\Gamma(n - x_1 + 1)} \\
 & \times \frac{\Gamma(a_1 + x_1)\Gamma(n - x_1 + b)}{\Gamma\left(n + \sum_i a_i\right)} \frac{\Gamma\left(\sum_i a_i\right)}{\Gamma(a_1)\Gamma\left(\sum_i a_i - a_1\right)}.
 \end{aligned} \tag{15}$$

The posterior for assimilation of successive new data sets obtains by straightforward extension. For k data sets, the summed contribution to each distance class becomes

$$\bar{A}_j = a_j + \sum_{i=1}^k x_{ij}$$

where x_{ij} is the number of seeds observed in the j th class from the i th data set. The marginal posterior for, say, the first class is

$$\begin{aligned}
 p(\theta_1 | \mathbf{x}) = & \text{Beta}\left(\theta_1 \left| \sum_{i=1}^k x_{i1} + a_1, \right. \right. \\
 & \left. \left. \sum_{i=1}^k (n_i - x_{i1}) + \sum_{i=1}^m a_i - a_1 \right.\right). \tag{16}
 \end{aligned}$$

A simple recipe.—The foregoing technical description collapses to a rather simple approach. Note that the Beta marginal posteriors (Eq. 16) require two parameters. For the i th marginal, the first parameter is the sum of the prior and data for class i . The second parameter is the sum of prior and data for all remaining classes. These parameters can be used to compute Beta densities using many standard software packages.

Application

We used the model to determine the rates of spread that would be estimated from the combination of fitted data and the assumptions of dispersal biologists and the contribution of estimation error. Our prior distribution was taken from the parametric kernel fitted to dispersal data for *Acer rubrum* (Clark et al. 1999). This prior is used for demonstration, because it is the assumption that is closest to the common practice of “extrapolation.” It can be weighted to have large or inconsequential effect on the posterior, depending on the quantity and quality of LDD dispersal information. The prior estimate for the j th class is $a_j = \int_{x_j}^{x_{j+1}} f(x) dx$ with the upper integration limit for distance class m taken to be $x_{m+1} = \infty$.

For purposes of illustration, the “data” are best estimates by dispersal biologists of the fraction of seed expected to disperse according to broad distance classes. Log-scale distance classes were used to accommodate the approximate nature of LDD information.

Because we do not believe that surveys provide precise information, we did not explore consequences of alternative distance classes. We surveyed dispersal biologists for their estimates of the fraction of seed, for each of several dispersal types, that is dispersed to different distance classes. We weighted their opinions (relative to one other and to the prior) based on their own assessments; each respondent choose one of four classes that best describe their degree of confidence in their estimates. We calculated marginal posteriors using Eq. 16.

The furthest forward estimate of spread rate is given as the expected distance jumped per generation, based on a simplified stochastic process where the furthest forward is an offspring of the furthest forward at the last step. We used simulations of this simplified stochastic process to also evaluate variability in the distance jumped per generation. As we will show, this variability can dominate estimate errors of θ .

Confidence intervals on predictive distributions of wave speed were estimated to include inherent uncertainty, estimation uncertainty, and both. Betabinomial posteriors were used with measured dispersal data and to estimate composite dispersal kernels and wave speeds predicted by empirical MGFs (Eq. 6) and extreme values (Eq. 2). We propagated parameter uncertainty by drawing parameter estimates from marginal posteriors using Monte Carlo simulation. For the i th iteration of the algorithm, class probabilities θ_{ij} were drawn from marginal beta posteriors (Eq. 13). Then multinomial class values x_{ij} were drawn sequentially from the marginal binomial distributions (Gelman et al. 1995). Under the assumption that settling of seed far from the source is little affected by the precise distance, within a class j , we drew exponential random exponential distances. Clearly, other assumptions could have been applied. For the first distance class, x_{i1} random t variates were drawn from the standard t distribution $Z \sim t_2(v)$ with $v = 2$ df with the marginal (one-dimensional) distances $x = Z\sqrt{u}$, which accommodates the Jacobian for the variable change from standard variates to those having scale parameter u . In summary, the algorithm draws R_0 seeds from a multinomial distribution of classes. Within class 1, distances are $2Dt$ distributed. Within all other classes, they are exponential. For each Monte Carlo iteration, we determined extreme values and spread estimates for the empirical MGF. The contribution of parameter vs. inherent uncertainty was determined by comparing confidence intervals on spread rates using point estimates of parameters with those obtained when estimation error is included.

For demonstration, we used data from two species with contrasting dispersal properties and fecundities. *Acer rubrum* is increasing rapidly in the second-growth forests of the southern Appalachians and currently has a large reproductive value (Clark 1998). It serves as an example of a mid-sized wind-dispersed seed. The

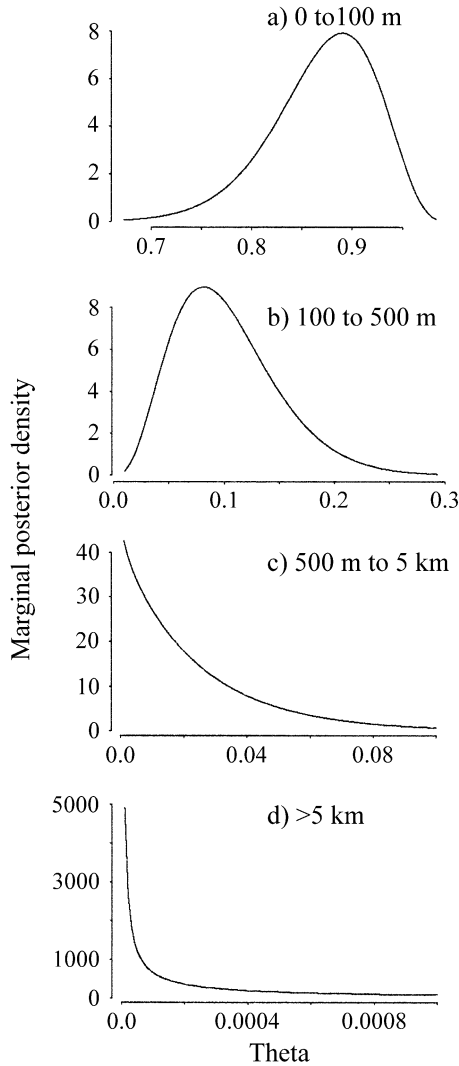


FIG. 1. Example of marginal posterior (beta) densities for surveys of dispersal distance classes for *Acer rubrum*. Densities describe the degree of confidence that can be assigned to different values for each probability (see Eq. 14).

dispersal kernel has a fat tail with a dispersal parameter estimate of $u = 602 \text{ m}^2$ and a broad confidence interval (Clark et al. 1999). *Carya glabra* grows slowly, has no known long-distance dispersal vector, $u = 8.82 \text{ m}^2$, and it is not increasing in these forests. Dispersal estimates come from a set of 100 seed traps located in five stands with collections spanning eight years (Clark et al. 1999).

RESULTS

Prior probabilities taken from the fitted kernel place nearly 98% of the seed at distances $< 100 \text{ m}$, and on the order of 10 out of a million beyond 5 km. A sample of 24 dispersal biologists, only two of whom have estimated dispersal for *Acer rubrum* from data, estimated far more LDD dispersal than is suggested by the fitted

kernel. Marginal posteriors for probabilities θ_j describe fractions of seed dispersed to distance classes (Fig. 1). The mean for each class amounts to a weighted mean, the weights being self-assigned. The spread of each marginal represents parameter uncertainty and decreases with sample size, in this case, the number of respondents. A mean fraction of 0.87 was estimated to remain within 100 m of the parent plant, and 0.003 were estimated to travel beyond 5 km.

The corresponding kernel is a mixture, with the first interval (0 to 100 m) showing the kernel fitted to seed trap data and the long distance classes derived from surveys (Fig. 2). The long tail is apparent only on a log scale (Fig. 2a, b). The tail governs the rate of spread for species having high R_0 , because dispersal biologists tend to believe that the probability of LDD is in the range where both methods predict dominance by the tail.

Wave speeds predicted by the furthest forward model are high and erratic. The high rate results from the combined effect of a fat tail and high R_0 . The confidence intervals explode with lead time (Fig. 2c) due to the inherent uncertainty of the process. Recall that this rate is contingent on R_0 continuing to assume this high value at the leading edge for the duration of the simulation.

Although parameter uncertainty is not small (survey respondents disagree, and they have limited basis for estimates) (Fig. 1), the uncertainty in spread velocity is dominated by inherent uncertainty, rather than by parameter uncertainty. To demonstrate the effect we compared spread estimates computed using the uncertainty represented by posteriors in Fig. 1 with those obtained under the assumption of full knowledge of the kernel (i.e., point estimates for multinomial parameters). We used a value of $R_0 = 5$. Two thousand simulated predictions of spread for both cases are nearly identical (Fig. 3). Thus, reducing uncertainty in parameters has little effect on predictions. This result does not mean that study of LDD provides no insight (see Discussion).

A value of R_0 closer to 1, which would apply if, say, the population was near constant density, and success in the new environment was not substantially better than in the current location, predicts much slower spread. At $R_0 = 2$, the model predicts a velocity of $44 \pm 314 \text{ m/yr}$ (mean $\pm 1 \text{ SE}$), with a mode of only 3.8 m/yr (Fig. 4a, see also Fig. 2d).

For a species that lacks LDD vectors and environments outside the current ranges substantially better than the existing one, spread predictions are extremely low. Using estimates for *Carya* and $R_0 = 2$, the predicted rate of spread is $0.55 \pm 3.4 \text{ m/yr}$ (Fig. 4b). This low rate results despite a value of R_0 that allows for a doubling of population density each generation.

DISCUSSION

Our method for estimating invasion rate admits the disparate evidence that ecologists use as basis for mod-

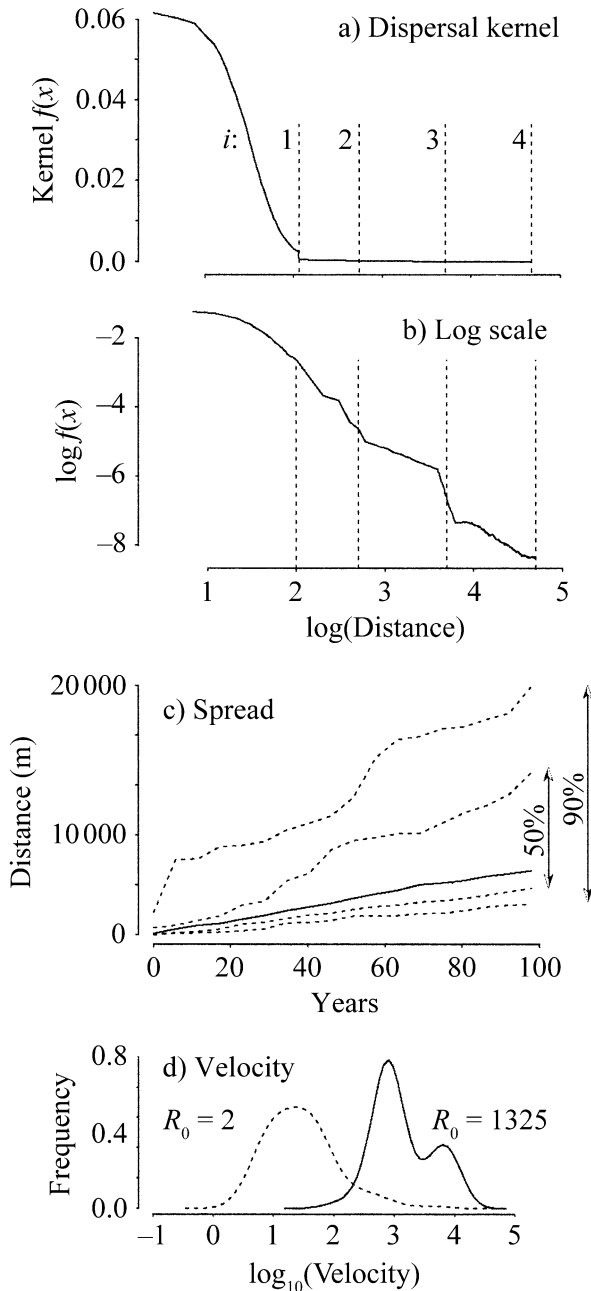


FIG. 2. Dispersal kernel and spread predictions for *Acer rubrum* using a value of $R_0 = 1325$ based on resampling from posteriors in Fig. 1. The four distance classes are based on the fitted kernel ($i = 1$), with fractions corresponding to each of the four parts of Fig. 1. Part (c) shows median, 50% CI, and 90% CI for migration distance. In (d), a distribution of velocities (m/yr) is compared with the same dispersal kernel, but $R_0 = 2$.

els and the stochastic nature of population spread. Analysis shows that, despite the apparent importance of estimating LDD, parameter uncertainty is not the principle source of forecast uncertainty. Moreover, values of R_0 that could be expected for many species in

modern landscapes will not provide for the rapid spread predicted by models calibrated to produce the seemingly rapid postglacial rates. We begin by placing our approach in the context of current methods. We then consider feasible goals for LDD research. Finally, we consider why spread rates of many species will be much lower than predicted by most current models.

Predictions from uncertain data

Current estimates of the potential for vegetation to track regional climate change and for aliens to invade new regions attempt to combine disconnected data sets, observations, and anecdote. Disparate evidence is pieced together and used to explore scenarios of change (assumptions 2 and 3 in *Assumptions of spread potential with LDD*). Each analysis tends to have a unique methodology. It would be difficult to define a confidence interval for such predictions, because there are no formal procedures that can be traced from data to prediction. Investigators might disagree on the importance or pattern of dispersal types, but those disagreements cannot be pursued within the current heterogeneous framework. Nonetheless, projections from such models are influential (Intergovernmental Panel on Climate Change [IPCC] 1996, National Assessment Synthesis Team [NAST] 2000).

Our approach does not reduce the uncertainty in LDD dispersal estimation. Rather, we provide a formal basis for acknowledging it and incorporating it in models that can be evaluated, assimilated, and extended with each new data set. The approach permits explicit treatment of uncertainty in terms of confidence intervals on predictions that are traced to uncertainty in inputs. It is reproducible. Investigators can disagree, and they can evaluate the consequences of that disagreement. The development and analysis of new scenarios can be pursued without changing model structure or abandoning accumulated information that may have been gath-

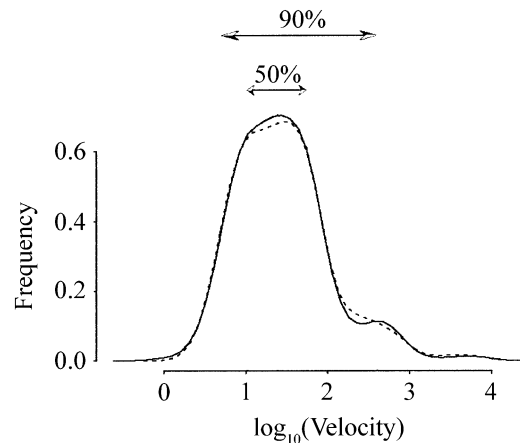


FIG. 3. Comparison of the distribution of the furthest forward spread rates using point estimates for parameters (dashed line) and including estimation error (solid line). Velocity was measured as meters per year.

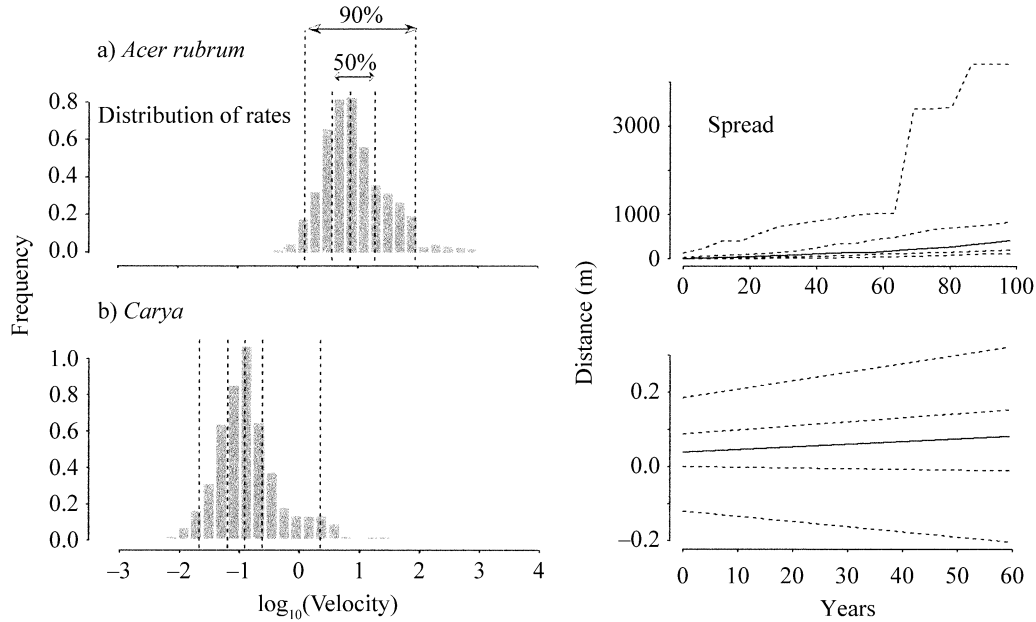


FIG. 4. Distribution of rates and spread based on $R_0 = 2$. Distributions of rates show quantiles and medians (central dashed line). Rates of spread (right) are uncertain (confidence intervals are broad), and extremely slow for *Carya*.

ered under the assumption of an alternative model structure. We focused on the degree of confidence that can be expected from predictions of population spread, but the general structure can be applied broadly.

The benefits of LDD estimation

In contrast with traditional deterministic models that would predict large benefit, in terms of predictive potential, from any LDD measurement (reduction in parameter uncertainty), a more realistic stochastic approach leads us to the opposite conclusion. Measurement of LDD is labor-intensive at best, and, typically, impossible. Our analysis demonstrates that the information yield on measurement, in terms of predictive capacity, is low (Fig. 2c). A typical index of information content is the standard error (from Fisher Information) or confidence interval (Clark et al. 2001a). Our confidence intervals that result from LDD are conservative, in that we ignore estimation error of dispersal parameter u and process error in R_0 , which can be large (Clark et al. 2001c). While estimation error is reduced by increased sampling (standard errors are asymptotically proportional to $1/\sqrt{n}$), the inherent stochasticity of the process will remain large (Figs. 2, 4). The 50% confidence intervals that span several orders of magnitude after several generations (Figs. 2c, 4) indicate that we cannot expect precise predictions even if we could obtain precise LDD measurements.

The fact that reduced parameter uncertainty does little to improve forecast information does not imply that LDD research has no benefit. Where detailed characterization of wind fields is possible, short-term predictions can be reasonable for intermediate distances (e.g.,

seasonal values for up to several hundred meters) (Nathan et al. 2002). Tree migrations span generations, during which extreme events, such as hurricanes and tornadoes, carry seeds much further distances and are not readily incorporated into predictive mechanistic models. Experimental and observational data can identify vectors that were previously unknown and their relative importances (Bullock and Clarke 2000, Nathan et al. 2000). They can help define the potential for rapid spread (Davis 1986), if not precise estimates of rate.

Why is spread potential overestimated?

Our analysis that admits LDD information in a stochastic way does not predict rates of spread as high as do deterministic models (Kot et al. 1996, Clark 1998) and simulation models of response to climate change (Collingham et al. 1996, Iverson et al. 1999). Indeed spread could be rapid for some species, but some predictions may be overestimates.

There are three factors that contribute to the tendency to overestimate spread velocity. First is R_0 . Net reproductive rate can be viewed as the number of individuals in the next generation that will replace the current generation. If population density is more or less constant, then $R_0 \approx 1$. A value of $R_0 > 1$ implies population growth, and it is required for a population to spread by "reaction diffusion." If R_0 is unity, then spread would be exceedingly slow. In successional forests, R_0 can be temporarily large for some species, but it cannot remain large in a closed forest over successive generations.

In models of plant population spread, ecologists assign values for R_0 that seem like reasonable guesses of

seed output of a large tree; a population must be producing “millions of seeds” (Skellam 1955). Moreover, fecundity of trees in closed forests is much lower than most ecologists think, as observation is biased toward fecund, open-grown individuals. Fecundity is hard to observe for most trees in forests, and it is lower than in the open. The fact that R_0 must integrate probability of survival is sometimes overlooked; a rare seed may disperse far, but it still may face overwhelming odds of mortality (Clark et al. 2001c). Likewise, generation times of late successional species in closed forests are much longer than typically used in population models. Maturation is slow in the understory, and generation time is longer than maturation time.

Large R_0 values lead to predictions of rapid spread. In fact, unless potential success is greater in the new environment than in the existing one, a R_0 value close to 1 is more realistic; on average, a tree will often produce one replacement. Conditions for growth and survival may indeed be superior in the new environment as, say, climate ameliorates. In many cases, the environment will be worse (the population does not spread). If environments are about the same, the population will slowly invade, because R_0 might be somewhat greater than 1. For many species (e.g., some late successional ones) that do not exploit the novel environments created by humans, future climate change might rapidly outpace the capacity for spread (Fig. 4). In modern landscapes, species that exploit manipulated landscapes and are fecund in open and edge environments might spread rapidly to new environments.

A second factor that leads to overestimates results from bias in the parameter estimates for fecundity and dispersal. Inverse methods based on classical approaches (e.g., maximum likelihood; Ribbens et al. 1994, Clark et al. 1998, 1999) overestimate both. This bias results from the fact that the model cannot resolve changes in the fecundity schedule with age, and it does not accommodate the large process error associated with year-to-year variability in seed production. Including both of these sources of stochasticity within a hierarchical Bayes format results in large variability in fecundity, but overall low mean values, and shorter (and more realistic) dispersal parameter estimates (J. S. Clark, S. LaDeau, and I. Ibanez, *unpublished manuscript*).

A third contributor to overestimates of spread potential is calibration to produce the early Holocene rates inferred from fossil pollen (1×10^2 to 1×10^3 m/yr; assumption 3 in *Assumptions of spread potential with LDD*). Fossil pollen rates do not apply to modern circumstances, and they could be wrong. It is especially hard to invoke LDD for genera like *Carya* (Fig. 4b), which has no known modern or prehistoric LDD vector, with the possible exception of humans. Glacial populations in eastern North America might have already been further north than previously interpreted from fossil pollen data. Low pollen percentages might be mis-

interpreted (Bennett 1985), and minimum winter temperatures may not have been extremely low (Wright 1992). There is increasing macrofossil (Jackson et al. 2000), pollen (Russell and Stanford 2000), and molecular (J. S. McLachlan, J. S. Clark, and P. Manos, *unpublished manuscript*) evidence that Glacial populations of many “temperate” species extended to mid latitudes. If so, then velocity of spread is overestimated by traditional interpretations of the pollen record.

ACKNOWLEDGMENTS

This study was supported by NSF grants DEB-9632854 and DEB 9981392 to J. S. Clark and DEB-9632854, DEB-9981392, DMS-9973212 and the Canada Research Chair program to M. Lewis.

LITERATURE CITED

- Andow, D., P. Kareiva, S. Levin, and A. Okubo. 1990. Spread of invading organisms. *Landscape Ecology* **4**:177–188.
- Bennett, K. D. 1985. The spread of *Fagus grandifolia* across eastern North America during the last 18 000 years. *Journal of Biogeography* **12**:147–164.
- Bullock, J. M., and R. T. Clarke. 2000. Long distance seed dispersal by wind: measuring and modelling the tail of the curve. *Oecologia* **124**:506–521.
- Cain, M. L., B. G. Milligan, and A. E. Strand. 2000. Long-distance seed dispersal in plant populations. *American Journal of Botany* **87**:1217–1227.
- Clark, J. S. 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleo record. *American Naturalist* **152**:204–224.
- Clark, J. S., et al. 2001a. Ecological forecasts: an emerging imperative. *Science* **293**:657–660.
- Clark, J. S., L. Horvath, and M. Lewis. 2001b. On the estimation of spread for a biological population. *Statistics and Probability Letters* **51**:225–234.
- Clark, J. S., M. Lewis, and L. Horvath. 2001c. Invasion by extremes: population spread with variation in dispersal and reproduction. *American Naturalist* **157**:537–554.
- Clark, J. S., E. Macklin, and L. Wood. 1998. Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecological Monographs* **68**:213–235.
- Clark, J. S., M. Silman, R. Kern, E. Macklin, and J. HilleRisLambers. 1999. Seed dispersal near and far: generalized patterns across temperate and tropical forests. *Ecology* **80**:1475–1494.
- Collingham, Y. C., M. O. Hill, and B. Huntley. 1996. The migration of sessile organisms: a simulation model with measurable parameters. *Journal of Vegetation Science* **7**:831–846.
- Davis, M. B. 1986. Climatic instability, time lags, and community disequilibrium. Pages 269–284 in J. Diamond and T. J. Case, editors. *Community ecology*. Harper and Row, New York, New York, USA.
- Dyer, J. M. 1995. Assessment of climatic warming using a model of forest species migration. *Ecological Modeling* **79**:199–219.
- Fisher, R. A. 1937. The wave of advance of advantageous genes. *Annals of Eugenics* **7**:355–369.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian data analysis*. Chapman and Hall, London, UK.
- Hew, H. S., and D. J. Mladenoff. 1999. The effects of seed dispersal on the simulation of long-term forest landscape change. *Ecosystems* **2**:308–319.
- Higgins, S. I., and D. M. Richardson. 1999. Predicting plant migration rates in a changing world: the role of long-distance dispersal. *American Naturalist* **153**:464–475.

- IPCC (Intergovernmental Panel on Climate Change). 2001. Third Assessment Report. Cambridge University Press, Port Chester, New York, USA.
- Iverson, L. R., and A. M. Prasad. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs* **68**:465–485.
- Iverson, L. R., A. M. Prasad, and M. W. Schwartz. 1999. Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with *Pinus virginiana*. *Ecological Modeling* **115**:77–93.
- Jackson, S. T., R. S. Webb, K. H. Anderson, J. T. Overpeck, T. Webb, J. W. Williams, and B. C. S. Hansen. 2000. Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews* **19**:489–508.
- Johnson, W. C., C. S. Adkisson, T. R. Crow, and M. D. Dixon. 1997. Nut caching by Blue Jays (*Cyanocitta cristata* L.): implications for tree demography. *American Midland Naturalist* **138**:357–370.
- Johnson, W. C., and T. Webb III. 1989. The role of Blue Jays in the postglacial dispersal of fagaceous trees in eastern North America. *Journal of Biogeography* **16**:561–571.
- Kirilenko, A. P., N. V. Belotelov, and B. G. Bogatyrev. 2000. Global model of vegetation migration: incorporation of climatic variability. *Ecological Modeling* **132**:125–133.
- Kot, M., M. A. Lewis, and P. van den Driessche. 1996. Dispersal data and the spread of invading organisms. *Ecology* **77**:2027–2042.
- Kubitzki, K., and A. Ziburski. 1994. Seed dispersal in flood plain forest of Amazonia. *Biotropica* **26**:30–43.
- Lewis, M. A. 1997. Variability, patchiness and jump dispersal in the spread of an invading population. Pages 46–69 in D. Tilman and P. Kareiva, editors. *Spatial ecology: the role of space in population dynamics and interspecific interactions*. Princeton University Press, Princeton, New Jersey, USA.
- Lewis, M. A. 2000. Spread rate for a nonlinear stochastic invasion. *Journal of Mathematical Biology* **41**:430–454.
- Lubina, J. A., and S. A. Levin. 1988. The spread of a reinvading species: range expansion in the California sea otter. *American Naturalist* **131**:526–543.
- Malanson, G. P., and D. M. Cairns. 1997. Effects of dispersal, population delays, and forest fragmentation on tree migration rates. *Plant Ecology* **131**:67–79.
- McKenney-Easterling, M., D. R. DeWalle, L. R. Iverson, A. M. Prasad, and A. R. Buda. 2000. The potential impacts of climate change and variability on forests and forestry in the Mid-Atlantic Region. *Climate Research* **14**:195–206.
- Mollison, D. 1972. The rate of spatial propagation of simple epidemics. *Proceedings of the Sixth Berkeley Symposium on Mathematics, Statistics, and Probability* **3**:579–614.
- NAST (National Assessment Synthesis Team). 2000. Climate change impacts on the United States: the potential consequences of climate variability and change. Overview report. National Assessment Synthesis Team. The US Global Change Research Program. Cambridge University Press, Port Chester, New York, USA.
- Nathan, R., G. Katul, H. S. Horn, S. M. Thomas, R. Oren, R. Avissar, S. W. Pacala, and S. A. Levin. 2002. Mechanisms of long-distance dispersal of seeds by wind. *Nature* **418**:409–413.
- Nathan, R., U. N. Safriel, I. Noy-Meir, and G. Schiller. 2000. Spatiotemporal variation in seed dispersal and recruitment near and far from *Pinus halepensis* trees. *Ecology* **81**:2156–2169.
- Neilson, R. P. 1993. Vegetation redistribution—a possible biosphere source of CO₂ during climatic change. *Water Air and Soil Pollution* **70**:659–673.
- Okubo, A. 1980. *Diffusion and ecological problems: mathematical models*. Springer-Verlag, New York, New York, USA.
- Okubo, A. 1988. Diffusion type models for avian range expansion. Pages 1038–1049 in H. Quillet, editor. *Acta XIX Congress Internationalis Ornithologici I*. National Museum of Natural Sciences, University of Ottawa Press, Ottawa, Ontario, Canada.
- Ribbens, E., J. A. Silander, and S. W. Pacala. 1994. Seedling recruitment in forests: calibrating models to predict patterns of tree seedling dispersion. *Ecology* **75**:1794–1806.
- Russell, E. W. B., and S. D. Stanford. 2000. Late-glacial environmental changes south of the Wisconsinan terminal moraine in the eastern United States. *Quaternary Research* **53**:105–113.
- Shigesada, N., and K. Kawasaki. 1997. *Biological invasions: theory and practice*. Oxford University Press, Oxford, UK.
- Shilton, L. A., J. D. Altringham, S. G. Compton, and R. S. Whittaker. 1999. Old World fruit bats can be long-distance seed dispersers through extended retention of viable seeds in the gut. *Proceedings of the Royal Society of London Series B* **266**:219–223.
- Skellam, J. G. 1951. Random dispersal in theoretical populations. *Biometrika* **38**:196–218.
- Snow, J. T., A. L. Wyatt, A. K. McCarthy, and E. K. Bishop. 1995. Fallout of debris from tornadic thunderstorms: a historical perspective and two examples from VORTEX. *Bulletin of the American Meteorological Society* **76**:1777–1790.
- Storm, G. L., and G. G. Montgomery. 1975. Dispersal and social contact among red foxes: results from telemetry and computer simulation. Pages 237–246 in M. W. Fox, editor. *The wild canids: their systematics, behavioral ecology, and evolution*. Van Nostrand Reinhold, New York, New York, USA.
- Sykes, M. T., I. C. Prentice, and W. Cramer. 1996. A bioclimatic model for the potential distributions of north European tree species under present and future climates. *Journal of Biogeography* **23**:203–233.
- Thebaud, C., and M. Debussche. 1991. Rapid invasion of *Fraxinus ornus* L along the Herault river system in southern France—the importance of seed dispersal by water. *Journal of Biogeography* **18**:7–12.
- Van Minnen, J. G., R. Leemans, and F. Ihle. 2000. Defining the importance of including transient ecosystem responses to simulate C-cycle dynamics in a global change model. *Global Change Biology* **6**:595–611.
- Weinberger, H. F. 1982. Long-time behavior of a class of biological models. *SIAM Journal of Mathematical Analysis* **13**:353–396.
- Wikle, C. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **84**:1382–1394.
- Wright, H. E., Jr. 1992. Patterns of Holocene climatic change in the midwestern United States. *Quaternary Research* **38**:129–134.
- Yumoto, T., K. Kimura, and A. Nishimura. 1999. Estimation of retention times and distances of seed dispersal by two monkey species, *Aouatta seniculus* and *Legothrix lagotricha*, in a Columbia forest. *Ecological Research* **12**:179–191.