

PREDICTING INVASIONS: PROPAGULE PRESSURE AND THE GRAVITY OF ALLEE EFFECTS

BRIAN LEUNG,¹ JOHN M. DRAKE, AND DAVID M. LODGE

Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556 USA

Abstract. Invasions by nonindigenous species impose large environmental and economic costs. In order to prevent invasions and target monitoring efforts most effectively, we need to forecast locations at the greatest risk of new invasions. To accomplish this, we need to estimate propagule pressure (inoculum size) and consider population processes, including possible Allee effects. Here, we develop a method to estimate the probability of population establishment, based on survival analysis and maximum likelihood techniques. We demonstrate theoretically the validity of this approach, considering environmental heterogeneity, estimation error, and nonlinearity. We then apply this method to zebra mussel (*Dreissena polymorpha*) invasions of Michigan inland lakes. We fit our model using presence/absence data for 1589 lakes between 1992 and 1996 and propagule pressure estimates from gravity models of boater traffic. Using our fitted model, we estimated the probabilities of establishment and demonstrated that Allee effects were present in the zebra mussel system. We validated our model using invasion data from 1997–2001 (data not used to parameterize the model). For the validation time period, we correctly predicted up to nine times as many invasions as a null (random) model. Further, the Allee model assigned average probabilities of invasion four times higher for lakes that became invaded than for uninvaded lakes, whereas the non-Allee model predicted probabilities for invaded lakes only two times higher. Thus, our model demonstrates the importance of considering the Allee effect and improves predictions of invasions.

Key words: *Dreissena polymorpha*; exotic; extinction; gravity model; incidence model; invasive species; nonindigenous; risk; time series; zebra mussel.

INTRODUCTION

Invasions by nonindigenous species are increasing (Cohen and Carlton 1998), causing large losses in biodiversity and other environmental effects (Sala et al. 2000, Lodge 2001) and imposing high economic damages (Pimentel et al. 1999). In order to more effectively target prevention efforts, society requires better estimates of the risk (or probability) of establishment—a critical component of the invasion process.

However, estimation of the probability of establishment represents a major challenge in invasion biology (Sakai et al. 2001). Difficulties occur because of the rarity of invasion events, because the relevant dynamics occur at the beginning of an invasion when population sizes are small and difficult to estimate, and because the numbers of failed introductions are often unknown. Typically, population dynamics are studied in detail long after establishment, when population sizes are large enough to cause damage to industry or to the environment. Extrapolation of population dynamics at large population sizes to the initial establishment conditions would likely be misleading due to processes such as negative density dependence when populations are large and Allee effects when populations are small.

Detailed information on population rates at small population sizes will likely remain difficult to obtain. Thus, sufficient information for a reasonable mechanistic population dynamical model will be rare, especially in the face of environmental heterogeneity where population rates differ between locations.

Because propagule pressure (Kolar and Lodge 2001) and the Allee effect are potentially important determinants of the probability of establishment, both are of great interest to both basic and applied ecologists (Courchamp et al. 1999, Dennis 2002). Consideration of Allee effects may allow better estimates of the risk of invasion and better assessments of the level of management effort necessary to reduce invasion risk. For example, if Allee effects are present, the probability of invasion will remain high above a threshold, but will be low below that threshold. Despite the potential importance of the Allee effect and the abundant theoretical work on it (Dennis 1989, 2002, Lewis and Kareiva 1993, Amaresekare 1998, Lande 1998, Gyllenberg et al. 1999, Etienne et al. 2002), actual demonstrations of the Allee effect in natural populations are scarce (but see Fowler and Baker 1991), and operational methods for its detection are needed.

In this paper, our purpose is twofold: to present a method to estimate the functional relation between propagule pressure and probability of establishment and to develop a test for Allee effects. We estimate

Manuscript received 18 September 2002; revised 28 August 2003; accepted 14 October 2003. Corresponding Editor: J. R. Bence.

¹ E-mail: bleung@nd.edu

probability of establishment, while circumventing explicit measurements of population dynamics. Instead, we present methods to use information that is commonly available. In general, higher propagule pressure should result in a higher probability of establishment (although quantification is difficult [Kolar and Lodge 2001]), even without knowing the details of the underlying population dynamics. Thus, the problem may be framed as one of obtaining enough estimates of propagule pressure over a sufficient range and of determining valid statistical tests for the relation between propagule pressure and invasion success.

We employ only the most commonly available data: time series of invasions (presence/absence data) and indices of propagule pressure from gravity models. Where pathways of transmission are discontinuous (e.g., human-mediated movement between lakes), gravity models may be ideal because they can extract a great deal of information from modest sources of data. We also evaluate potential, realistic, confounding factors: environmental heterogeneity, estimation error, model misspecification, and nonlinear relations between estimated and actual propagule pressure. We first analyze our model in a theoretical context, where relations between propagule pressure versus establishment are generated from known functional forms (exponential and Weibull functions). Second, we apply our method to zebra mussel (*Dreissena polymorpha*) gravity models.

METHODS

Theoretical presence/absence and propagule pressure simulations

We present generic functional forms and methods that will allow us to test for Allee effects and approximate the relation between propagule pressure and the probability of establishment. If propagules each have an independent chance of establishment (i.e., Allee effects not present), the total probability of establishment (E) would be the complement of all propagules failing to establish:

$$E(N_{l,t}) = 1 - (1 - p)^{N_{l,t}} \tag{1}$$

where p is the probability of a single propagule establishing and N is the number of propagules arriving at location l at time t . This can equivalently be written as the standard asymptotic curve:

$$E(N_{l,t}) = 1 - e^{-(\alpha N_{l,t})} \tag{2}$$

where α is a shape coefficient and is equal to $-\ln(1 - p)$. We term this (Eq. 2) the independence model. In contrast, the Allee effect is described by a curve that contains a transition point, below which establishment is disproportionately less likely. We can generate a family of such curves by adding a shape parameter (c) (Dennis 2002):

$$E(N_{l,t}) = 1 - e^{-(\alpha N_{l,t})^c} \tag{3}$$

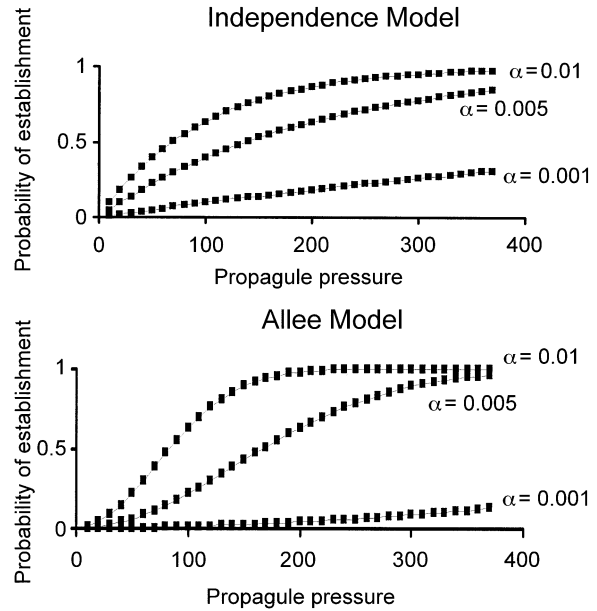


FIG. 1. Functional forms for non-Allee and Allee models: hypothetical examples of families of curves between propagule pressure and probability of establishment when Allee effects are absent (top panel) and present (lower panel). The standard asymptotic curve was used to describe the independence model (Eq. 2), and the Weibull function was used to describe the Allee model (Eq. 3). We used $c = 2$ and three parameter values for α (in Eqs. 2 and 3).

This is simply the Weibull function (Fig. 1). The Weibull function has a transition point when αN equals unity where the values of E in Eqs. 2 and 3 would be identical.

We consider a time-series data set (the matrix \mathbf{D}) with information on propagule pressure and date of invasion (if any) at each location over time. As the invasion of a landscape progresses, we would expect propagule pressures and their distribution to change, and our analysis should incorporate this change. Using logic from survival analysis (Collette 1994), we can determine the likelihood of observing our data set (\mathbf{D}) given our model (Eqs. 2 or 3), and relate the probability of establishment to propagule pressure. First, we consider the probability (H) of observing the pattern at each location (l). Specifically, for locations that become invaded at time t , we want to consider the joint probabilities of becoming invaded at time t and of remaining uninvaded up to time t , given the history of propagule pressure and given a model (M) (Eqs. 2 or 3). The probability of remaining uninvaded during a time interval (i), would simply be the complement of E . Therefore,

$$H_l = E_M(N_{l,t}) \prod_{i=1}^{t-1} [1 - E_M(N_{l,i})]. \tag{4}$$

For locations that do not become invaded for the duration of the study (T) (i.e., “censored” data [Collette

1994]), we simply use the joint probabilities over time of remaining uninvaded given the history of propagule pressure:

$$H_{l,M} = \prod_{i=1}^T [1 - E_M(N_{l,i})]. \quad (5)$$

The log-likelihood (\mathbf{L}) for the entire data set (\mathbf{D}) given a model (M) would be

$$\mathbf{L}(\mathbf{D}|M) = \sum_{l=1}^L \ln(H_{l,M}) \quad (6)$$

where H_l is determined using Eq. 4 for invaded locations, and Eq. 5 for locations that did not become invaded during the study. The model (M) can either be the independence (non-Allee) model (Eq. 2) or the Allee model (Eq. 3). Importantly, we note that Eq. 2 is a special case of Eq. 3 (when $c = 1$). Thus, we can test for Allee effects by using the likelihood ratio test and comparing the Allee model (c_{MLE} fit using maximum likelihood methods) vs. the independence model ($c = 1$):

$$LR = 2[\mathbf{L}(\mathbf{D}|M_{c=1}) - \mathbf{L}(\mathbf{D}|M_{c_{MLE}})]. \quad (7)$$

LR has a chi-square distribution with one degree of freedom (Hilborn and Mangel 1997). The probability (P) of a value equal to or greater than LR can be approximated with the incomplete gamma function (Press et al. 1995). We are specifically interested in the one-tailed hypothesis $c_{MLE} > 1$ (Fig. 1).

We use simulation techniques to examine the efficacy of our methods. Specifically, we examine Type I error rates—the probability of falsely detecting the Allee effect when it is not present—and examine the ability to estimate the underlying population parameters for c and α . We determine if Type I error rates were at the nominal level, i.e., P should be ≤ 0.05 in only 5% of the simulations, when $c = 1$. We modeled environmental heterogeneity, by selecting values of α_i from a uniform distribution with mean α and range r (if $r = 0$, the environment would be homogeneous). The consequence of environmental heterogeneity was that different locations had different probabilities of establishment for a given propagule pressure. We did not consider heterogeneity of the parameter c because this would automatically result in Allee effects being present.

In many situations, direct estimates of propagule pressure do not exist. Often, however, surrogate predictors exist, including some with a mechanistic basis (e.g., estimates from gravity models in our case). We modeled estimation error (ε) of propagule pressure ($N_{ac} = N + \varepsilon$, minimum = 0) using a normal variate with mean zero and standard deviation σ_ε . N_{ac} was the true (actual) propagule pressure and N was the predictor variable that we used as our estimate. We also considered nonlinear relations between actual propagule pres-

sure (N_{ac}) and our predictor (N), with log-normal errors:

$$N_{ac} = aN^b \exp(\varepsilon). \quad (8)$$

We examined the validity of the model when we used a misspecified model (i.e., assumed linearity) and when we incorporated the nonlinearity via statistical fitting, in the presence and absence of estimation error (ε).

For statistical fitting, we modeled the case where we had some actual measures of N_{ac} with which we could calibrate the relation between N_{ac} and N (as in the case for zebra mussels). In this case, we used linear regression on log-log transformed data (i.e., $\ln(N_{ac}) = \ln(a) + b \ln(N) + \varepsilon$) to estimate parameters, and back transformed to obtain a nonlinear estimate of N_{ac} (N'). Note that only the parameter b matters, because parameter a (Eq. 8) would be incorporated into parameter α (Eqs. 2 or 3). We examined Type I errors (falsely rejecting the null hypothesis that the pattern of invasion is generated by the independence model) and Type II error rates (failing to reject a false null hypothesis) associated with this procedure.

We simulated 100 locations across 20 time intervals, and repeated this simulation 1000 times for each parameter set. Initial propagule pressure ($N_{l,1}$) was generated for each location (l) from a uniform distribution between 0 and 100. We assumed that as the invasion progressed and more locations became invaded, propagule pressure would increase ($N_{l,t+1} = N_{l,t} + R$) where R was randomly chosen from a uniform distribution between 0 and 10. At each simulation, we chose α values randomly from a uniform distribution between 0 and 0.01. We also examined α ranging between 0 and 0.001, but these did not change the conclusions of the manuscript other than to reduce the power of detecting Allee effects, and were not presented. It was the relative (rather than the absolute) values of these parameters that were important. For instance, in the absence of Allee effects with maximums of $N_{l,1} = 100$ and $\alpha = 0.01$, all locations would initially have a <63% chance of invasion. For $\alpha = 0.001$, the maximum initial probability would be 9.5%. Thus, we modeled a wide range of probabilities of establishment. To examine Type I errors, we held $c = 1$. To examine power and the ability to detect the Allee effect, we used $c = 2$. We used the simplex algorithm (Press et al. 1995) to fit the models, and determine the maximum likelihood values for the parameters α (and c for the Allee model).

We examined two levels of environmental heterogeneity; $r = 0$ and $2/3\alpha$ (i.e., α_i can range from population value $\alpha \pm 2/3\alpha$). For estimation error, we used $\sigma_\varepsilon = 0$ (no error), 10, and 50. To examine nonlinear relations, we used $a = 4$, $b = 0.7$, $\sigma_\varepsilon = 0, 0.3$, and 1. These values were chosen to generate a large range of measurement errors. We constrained the simulations such that at least 10% of the lakes would be invaded (if they were not, we re-ran that simulation).

Gravity models and propagule pressure

Gravity models provide a potentially powerful means of generating relative estimates of propagule pressure over time. There are several different forms of gravity model (Thomas and Hugget 1980); we chose the production constrained gravity model used by Bossenbroek et al. (2001) as our starting point, as this was logistically the most attractive. Propagules are transmitted to different locations via vectors such as human movement. The number of visitations (T_j) to a destination location (j) depends upon the number of vectors (O_i) at each source location (i), the attraction of the destination location (W_j), the distance between the source and destination (D_{ij} , the greater the distance, the lower the tendency), and the other available locations that each vector could visit instead (A_i). Gravity models provide a formal structure for these relations:

$$U_j = \sum_{i=1}^K A_i O_i W_j D_{ij}^{-d} \quad (9)$$

$$A_i = 1 / \sum_{j=1}^L W_j D_{ij}^{-d} \quad (10)$$

$$T_j = f(U_j) \quad (11)$$

where U_j is a metric of vector traffic from the gravity model, T_j is the full model's prediction of actual traffic and a determinant of propagule pressure, K is the number of spatially separated sources, L is the number of locations, and d is a shape parameter describing the relation between traffic and distance. Note that we described T_j as a function of the output from Eq. 9 to permit generalization and nonlinear relations (the actual function used is given below). For our example system, we modeled boater traffic (our vector) from Michigan counties (sources) to Michigan lakes (destination locations), where O_i was the number of boaters registered in each county, W_j was the area of the lake (in hectares), which has a strong positive relationship to boat usage (Reed-Anderson et al. 2000), and D_{ij} was the distance between destination lakes and source counties (as measured from the central point within lakes and counties). We also modeled traffic to Great Lakes boat ramps (including Lake St. Clair), where the attractiveness was a fitted value (g). We assumed that boater traffic from invaded to uninvaded lakes was directly proportional to propagule pressure, which is reasonable for invaders such as zebra mussels where the primary mode of transmission is boater movement (Johnson et al. 2001). We used this type of gravity model because it required far less effort to obtain data on boater registration than explicit counts of boater traffic for all lakes, allowing us to model a large number of lakes ($N = 1589$).

We parameterized our gravity model values (for d and g) by fitting estimates of boat traffic generated from the model (U_j) to empirical measures of actual boat

traffic (T_j) from creel surveys conducted by the Michigan Department of Natural Resources (available online)² (Lockwood et al. 1999). Boat traffic was calculated based on the number of angler trips reported in the creel surveys. On average, there were 2.5 anglers per boat (Roger Lockwood, *personal communication*). The relation between U_j and T_j was nonlinear, and we assumed a power relation (i.e., in Eq. 11, $T_j = f(U_j) = aU_j^b$). We estimated the parameters of the model by using the simplex algorithm to find the values of d (2.15) and g (16 000 ha) that maximized Pearson's correlation coefficient between the natural log of (U) (the output from Eq. 10) and the natural log of empirically estimated boat traffic. Because we used a power function in Eq. 11, we did not need to directly estimate a and b as part of the nonlinear search. There is a linear relationship between the log transformations of empirically estimated boat traffic and U (i.e., $\ln(T_j^g) = \ln(a) + b \ln(U_j) + \varepsilon$), and the parameter estimates of a and b are implicitly estimated as those that produce the linear relationship that maximizes the Pearson correlation.

This resulted in a best-fit Pearson's correlation value of 0.89 (sample size = 19, almost 80% of the variation explained; $b = 0.53$), suggesting that this was a reasonable model. However, because the relation between T and U were nonlinear, we did not have true balancing factors (A_i ; Fotheringham and O'Kelly 1989), but rather interpreted A_i as a correction to account for other potential destinations.

Having determined parameter values for d and g , we could estimate the boater traffic (T_j) for all 1589 lakes included in the model. However, we still needed to estimate propagule pressure to a lake. To do this, we performed a two-step process (following Bossenbroek et al. 2001). First, we calculated the proportion of boats that could become infested: the proportion of boats from a given county visiting infested lakes (X_i). Second, we calculated the proportion of infested boats visiting uninfested lakes.

For the first step, to maintain tractability given the nonlinear relation between T and U , we made the simplifying assumption that the total number of boats was affected by non-linearity but the composition of sources was not. Thus, the boats visiting a given lake (j) from county (i) was (T_{ij}), where

$$T_{ij} = T_j \frac{U_{ij}}{U_j} \quad (12)$$

The proportion of boat traffic originating in a county that is exposed to infestation (X_i) was estimated by

$$X_i = \frac{\sum_{k=1}^M T_{ik}}{\sum_{l=1}^L T_{il}} \quad (13)$$

where the numerator was summed across M infested

² URL: http://www.michigandnr.com/PUBLICATIONS/PDFS/ifr/ifrilibra/technical/reports/2000-3tr/2000-3tr_app.pdf

lakes and the denominator included all lakes L . We used the proportion of traffic to infested lakes rather than a simpler measure such as fraction of a county infested, because lakes were visited with different frequencies and we wanted to incorporate these differences into the model, and because lakes were natural units of measurement since it is lakes that become invaded.

For the second step, we need to know how many of these potentially infested boats (Eqs. 12 and 13) subsequently visit uninfested lakes. This was calculated as the boater traffic to an uninvaded lake from a source county (T_{ij}) multiplied by the proportion of those boats that were previously infested (X_i), summed over all (K) counties:

$$Q_j = \sum_{i=1}^K T_{ij} X_i. \quad (14)$$

Q_j was used as our proportional estimate of propagule pressure. Q_j changed over time, as more lakes became infested. Of course, the number of boats that were potentially (as opposed to actually) infested that visited a lake would be much larger than the actual propagule pressure to that lake. However, this would not affect our ability to model the probability of the establishment (Eqs. 2 or 3) because the coefficient α incorporates any proportional difference between propagule pressure and infested boater traffic.

Zebra mussels, Allee effects, and the probability of establishment

In the preceding section, we described the methods for constructing the gravity model, and estimating boater traffic and an index of propagule pressure. As explained above, we used creel surveys to calibrate our gravity model and fit gravity model parameter values for d and g . Next, we merged our calibrated gravity model with our maximum-likelihood survival analysis procedure to determine the probability of establishment and assay for the presence of Allee effects (using Eqs. 1–7). To do this, we needed to relate propagule pressure (from the gravity model) to invasion status over time and fit parameter values for α and c (Eqs. 2 and 3).

We restricted our analysis to the state of Michigan because it provided the best data, and because the majority of zebra mussel invasions have occurred here (Kraft and Johnson 2000). We obtained data for boater registration, lake characteristics (location, size, and chemistry) including only lakes larger than 25 ha, and the location of Great Lakes (including Lake St. Clair) boat ramps from Bossenbroek et al. (2001). We used lake chemistry (pH and calcium) to identify lakes that did not have the appropriate habitat for zebra mussels (Ramcharan et al. 1992), and excluded these lakes from the analysis of invasion status vs. propagule pressure. However, these lakes were included in the gravity model because they provided alternative destinations and would therefore affect traffic to lakes that could be

invaded. We obtained the date of invasion of Michigan lakes from Hokanson (2002), and fit parameters α and c (Eqs. 2 and 3) using invasion data between 1992 and 1996. Following distribution maps available online,³ we assumed that the Great Lake boat ramps were infested by 1992, except those on Lake Superior.

To validate our model, we examined the ability to predict known invasions that occurred during 1997–2001 (Hokanson 2002), but using the parameter values of α and c generated from data between 1992 and 1996 (i.e., the data used to generate the model differed from the data used to test it). We calculated the likelihood ratios for Allee effects ($c > 1$) and non-Allee effects ($c = 1$), for each year separately. We also examined the benefits of using our models by comparing their results to those of a null (random) model: the expectation if we did not have lake specific information. In the null model, all lakes were assigned equal probabilities of invasion, calculated as the number of new lakes invaded divided by the lakes available to be invaded. We used our model to predict the top 10, 100, and 200 lakes at risk of being newly invaded and compared our predictions with actual invasions during 1997–2001 (i.e., we excluded lakes invaded during 1992–1996).

RESULTS

Theoretic simulations

We detected Allee effects, by testing whether our parameter c was significantly greater than 1 (Eqs. 1–7), and we determined the validity of our analysis by examining Type I error rates. We found Type I error rates approximately at the nominal (5%) or conservative levels (Table 1, column 4). This method was robust to environmental heterogeneity (r) (the probabilities of establishment differed between lakes) and estimation error (σ_e) (Table 1, column 4, $r > 0$ and/or $\sigma_e > 0$). Thus, we should not find an unacceptable number of spurious Allee effects using this method.

When we considered nonlinear relations, we obtained conservative Type I error rates (Table 1, column 4) if we did not conduct transformations (model linearity labeled NL, NF). As mentioned, we specifically modeled $b < 1$ (Eq. 8), as this was observed for our real-world example (if $b > 1$, Type I error rates would be inflated). Transformation using regression yielded appropriate Type I error rates at the 5% level (Table 1, column 4), even in the presence of measurement error and heterogeneity (model linearity labeled NL, F). Thus, our fitting procedure was appropriate.

Measurement error ($\sigma_e > 0$) and model misspecification (model linearity labeled NL, NF) had the potential to reduce the power to detect the Allee effect (Table 1, column 7), if the magnitude of error was great enough. For our highest measurement error, Allee effects were barely detectable (e.g., Table 1, column 7,

³ URL: <http://nas.er.usgs.gov>

TABLE 1. Results from simulations.

Row	Model linearity	Parameters	$c = 1$			$c = 2$		
			Type I error rate	$(\hat{\alpha} - \alpha) \times 10^{-3}$	$\hat{c} - c$	Power	$(\hat{\alpha} - \alpha) \times 10^{-3}$	$\hat{c} - c$
1	L	$r = 0, \sigma_e = 0$	5.0	0.14 (1.26)	0.04 (0.23)	93	0.07 (0.74)	0.05 (0.38)
2	L	$r = 2/3, \sigma_e = 0$	3.7	1.26 (1.63)	-0.04 (0.23)	91	0.76 (0.89)	0.26 (0.53)
3	L	$r = 0, \sigma_e = 10$	5.2	0.02 (1.36)	0.01 (0.24)	92	-0.06 (0.78)	-0.01 (0.39)
4	L	$r = 0, \sigma_e = 50$	0.7	-1.78 (1.91)	-0.27 (0.24)	18	-1.84 (1.50)	-0.83 (0.39)
5	L	$r = 2/3, \sigma_e = 10$	5.2	1.16 (1.65)	-0.04 (0.24)	89	0.66 (0.90)	0.14 (0.50)
6	L	$r = 2/3, \sigma_e = 50$	0.2	-1.03 (1.93)	-0.35 (0.22)	16	-1.13 (1.31)	-0.90 (0.37)
7	NL, NF	$r = 0, \sigma_e = 0$	0.1	-0.70 (1.58)	-0.28 (0.20)	47	-0.80 (1.01)	-0.56 (0.32)
8	NL, F	$r = 0, \sigma_e = 0$	4.9	0.10 (1.28)	0.04 (0.30)	86	0.10 (0.83)	0.07 (0.45)
9	NL, NF	$r = 0, \sigma_e = 0.3$	0.1	-0.75 (1.61)	-0.30 (0.20)	42	-0.60 (1.06)	-0.59 (0.33)
10	NL, F	$r = 0, \sigma_e = 0.3$	5.3	-0.04 (1.30)	0.03 (0.28)	72	-0.39 (1.40)	-0.14 (0.57)
11	NL, NF	$r = 0, \sigma_e = 1$	0.1	-0.50 (1.84)	-0.38 (0.19)	4.7	-0.71 (1.40)	-1.06 (0.28)
12	NL, F	$r = 0, \sigma_e = 1$	4.4	-1.41 (1.84)	0.02 (0.37)	28	-1.76 (1.70)	-0.62 (0.53)
13	NL, NF	$r = 2/3, \sigma_e = 1$	0	0.83 (2.42)	-0.43 (0.18)	2.1	0.23 (1.59)	-1.11 (0.24)
14	NL, F	$r = 2/3, \sigma_e = 1$	3.9	-0.67 (2.03)	-0.03 (0.37)	25	-1.32 (1.85)	-0.70 (0.48)

Notes: Analyses are based on 1000 simulations for each parameter set (values of r and σ_e); r is the range for α_i values that determines how heterogeneous different locations are, and it is expressed here as a proportion of the average population value (α) (Eqs. 2 and 3). σ_e = estimation error (additive for the linear model, multiplicative for the nonlinear model). L = linear and NL = nonlinear relations between estimated and actual propagule pressure. For our nonlinear models, NF = nonfitted models, which simulated model misspecification, and F = fitted statistical corrections using log-log transformed regressions. The nominal Type I error rate was $\alpha = 0.05$ (5% false detection of Allee effects when the independence model was true, $c = 1$, column 4). We modeled Allee effects by setting $c = 2$ and assaying power (column 7). Average differences between estimated and actual parameter values ($\hat{\alpha} - \alpha$ or $\hat{c} - c$), as well as the standard deviation (in parentheses) are presented (columns 5, 6, 8, and 9). Negative differences indicate underestimated parameter values (for parameter c , less likely to detect the Allee effect); positive differences indicate the converse.

reducing power from 93% [row 1] to 18% [Row 4]). Thus, we had modeled the relevant range of measurement errors. For nonlinear relations, power was greatly improved if transformations were used (Table 1, column 7, model linearity labeled F). Thus, transformations should be used for the nonlinear relation modeled, because they provide appropriate Type I error rates and improved power.

The estimation of parameters α and c was reasonable, with average deviations ($\hat{\alpha} - \alpha, \hat{c} - c$) typically much smaller than the variability (one standard deviation of the difference is presented) (Table 1, columns 5, 6, 8, and 9). Further, the average deviations were generally less than 10% of the actual population parameter value for α and c . Exceptions to this included environmental heterogeneity, which increased deviations from α (Table 1, columns 5 and 8, rows 2 and 5). However, this should be expected as there was no single value of α and instead there was a distribution of α values in the heterogeneous simulations. Further, for our highest measurement error and for the model misspecifications examined, biases occurred resulting in systematic underestimates of parameter values (Table 1, columns 5, 6, 8, and 9, rows 4 and 6 and rows labeled NF). This was consistent with the above observation—underestimating c would result in decreased ability to detect Allee effects. Thus, our methods would not inflate the probability of falsely detecting Allee effects.

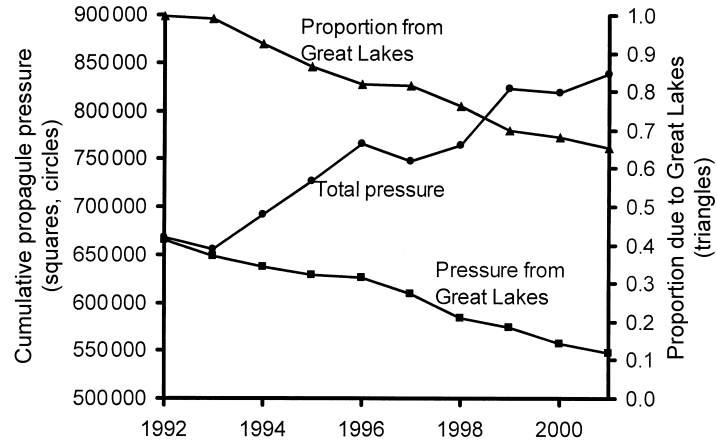
Zebra Mussel Invasions

Potential propagule pressure (Eqs. 12–14) was summed across all uninvaded lakes, keeping track of

propagules originating from the Great Lakes. Propagule pressure to uninvaded lakes increased by 25% as the invasion progressed and more inland lakes became invaded (Fig. 2, circles). The propagule pressure from Great Lakes sources decreased over time, by ~18% (Fig. 2, squares; considering only pressure to uninvaded lakes and discounting lakes once they became invaded); Great Lakes sources accounted for 99.7% of the total propagules in the beginning of the invasion (1992) and dropped to 65.3% by 2001 (Fig. 2, triangles). The shift in propagule pressure from the Great Lakes to inland lakes that occurred in the model was due to changes in invasion status over time. Specifically, inland lakes closer to the Great Lakes were invaded first; the remaining uninvaded lakes farther away from the Great Lakes were influenced less by the Great Lakes.

We found a highly significant Allee effect for our zebra mussel data set ($P < 0.0001$). We estimated the parameter values on data from 1992–1996: for the independence model, the maximum likelihood estimate was $\alpha = 1.2 \times 10^{-5}$; for the Allee model, $\alpha = 1.03 \times 10^{-4}, c = 1.86$). We then used these parameter values to examine probabilities of invasion and generate likelihood values separately for each year between 1997 and 2001 (Fig. 3). The probability of invasion during any given year was low. However, the Allee model predicted an average probability of invasion four times higher for lakes that were invaded in a given year compared to lakes that remained uninvaded. The non-Allee model predicted probabilities only twice as high. The Allee model consistently identified lakes at risk better

FIG. 2. Change in propagule pressure and importance of the Great Lakes as a source over time. Propagule pressures were summed over all uninvaded lakes. The total cumulative propagule pressure (circles), the propagule pressure from Great Lakes sources (squares), and the relative contribution to the total propagule pressure from Great Lakes sources (triangles) are shown.



than the non-Allee model and was significantly better for three out of five years (Fig. 3). Thus, the Allee model considerably outperformed the non-Allee model in our validation set.

We were able to forecast which lakes would become invaded during the years of our validation set (Fig. 4). We compared the predictive abilities of the Allee model and a null model (i.e., in the absence of lake-specific information). Ninety-eight invasions occurred during 1997–2001 out of 1065 possible previously uninvaded lakes (we did not consider lakes with inappropriate chemistry for zebra mussel establishment). Thus, the null model predicted that each lake would have the same 9.2% chance of being invaded. If we considered 10, 100, and 200 lakes in the absence of predictive models, on average, we would have correctly identified 0.92, 9.2, and 18.4 invaded lakes, respectively. In contrast, our model correctly predicted 9, 40, and 59 of the lakes that became invaded, respectively. Thus, our model provided significant improvements in predictive power over the null model.

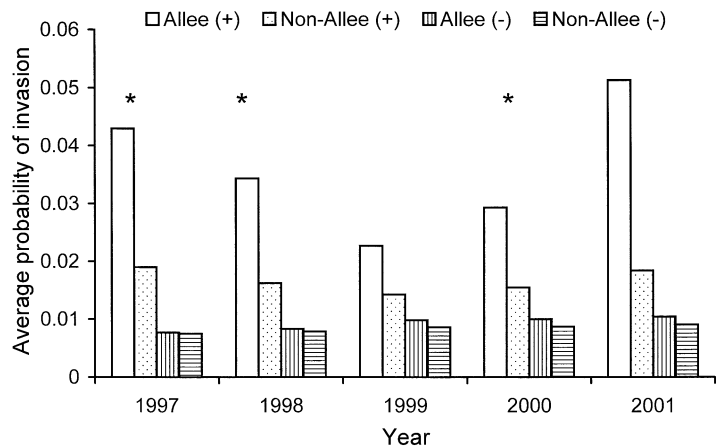
DISCUSSION

The probability of establishment

Previous studies have highlighted the importance and difficulties associated with predicting invasions

and spread of nonindigenous species. Specifically, difficulties exist in estimating vector contributions and linking them with population dynamics to determine the probability of invasion (Johnson and Padilla 1996). We demonstrated a novel method that uses presence/absence data to estimate the probability of establishment and detect the Allee effect in natural populations. Our work builds upon previous research. Specifically, earlier work examined the connectedness of lakes, but did not explicitly link them to propagule pressure and probabilities of invasion (Johnson and Padilla 1996, Buchan and Padilla 1999). Others used gravity models that linked connectedness to propagule pressure, and that suggested that propagule pressure may be predictive of future invasions (Schneider et al. 1998, Bossenbroek et al. 2001). Our efforts extend such previous work by estimating actual probabilities of invasion rather than only relative risks (Schneider et al. 1998), and examining actual lakes rather than numbers of lakes invaded per county (Bossenbroek et al. 2001). We formulated our gravity model using easily accessible data on boater registrations and lake characteristics. This extends previous work because our methods can be applied to a very large system of lakes and are not limited to only the small sample of lakes actually

FIG. 3. Relative predictive ability of future invasions. Parameter estimates were generated from data between 1992 and 1996 and validated for each year between 1997 and 2001. We show both Allee and non-Allee models and present the average probabilities associated with lakes that became invaded (+) or remained uninvaded (-) in a given year. Significant differences between Allee model and non-Allee models were indicated by an asterisk, using the likelihood ratio test.



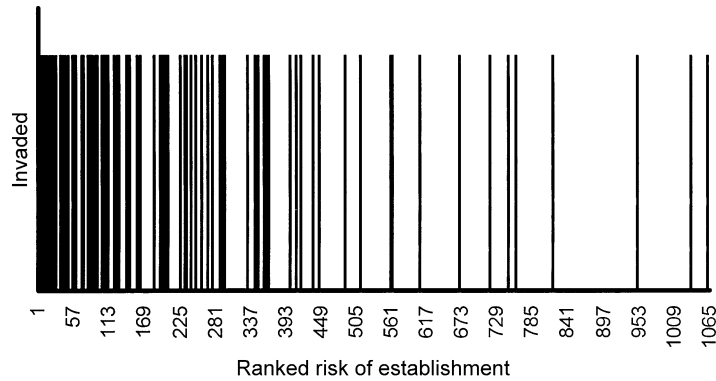


FIG. 4. Density of invaded lakes vs. model predictions. Lakes were ranked by the model prediction of probability of invasion occurring between 1997 and 2001 (1 = highest probability of invasion). Lines indicated invaded lakes.

measured in the creel survey (cf. Schneider et al. 1998), yet uses sampled lakes from creel surveys as a means of calibrating our gravity model (cf. Bossenbroek et al. 2001). Further, in previous work, assertions that gravity models could predict invasions were based purely on fitted values and were not validated. We performed such validation using invasion events that were not included in the parameterization of our model. We found, for instance, that management guided by our model could have prevented up to nine times as many invasion events as the null model (depending on the number of lakes that we chose to manage). Resource allocation could be fine-tuned further by explicitly considering the probability values generated by our model.

Previous researchers have argued for the need to consider population dynamics at early stages of invasion, especially in relation to establishment or spread (Johnson and Padilla 1996). Additional strengths of this approach compared to previous models include the demonstration of Allee effects and the demonstration of how to generate functional forms for the relation between probability of invasion and propagule pressure. These functional forms will be useful, in general, for developing quantitative probabilistic risk analyses for invasive species (e.g., Leung et al. 2002). The development of such risk analyses is a high priority in the policy and management of invasive species (National Invasive Species Council 2001, National Research Council 2002). In these ways, the approach described here advances the field of predictive ecology in general, and invasive species ecology and management in particular.

While our model possesses many useful characteristics, many potentially important factors were not included (as with any modeling effort). For instance, predictive power could possibly be improved if we considered zebra mussel populations within lakes, as this could affect the relative amount of propagules from a source lake. In the current project, we only used presence/absence data due to the availability of this information. However, surrogates for population sizes may be available (e.g., using lake chemistry [Ramcharan et al. 1992] and time since invasion as predictors). Sim-

ilarly, propagule pressure for any given year may depend in part on propagule pressures in previous years. This could be possible if propagules persisted for multiple years. These factors could be added in future versions of the model.

Additionally, there were sources of uncertainty for which we did not have data. Specifically, Great Lake boat ramps may differ in their usage. To a certain extent, heterogeneous use of these boat ramps was incorporated, since source populations and distances differed for different ramps (i.e., visitations also differed). However, the attractiveness of boat ramps to anglers also could differ. Further, only nonresident boaters transport propagules, but unfortunately the data from the creel surveys did not distinguish resident from nonresident boaters. Insofar as the proportion of boater types was equal across lakes, the results of the model are robust, because a constant proportion was incorporated in scaling parameter α (Eqs. 2 and 3). Future quantifications of differential use of boat ramps and resident vs. nonresident boaters, consideration of fishing vs. nonfishing boaters, and variation in anglers per trip might improve the predictive power of the model.

Finally, it would be of interest to consider factors related to management efforts. For instance, we might consider the consequences of early detection and time lags to detection, or of modification of human responses as an invasion progresses. As individuals become more educated about the threat of specific invasive species, they may adapt their behaviors to maximize the probability of detecting invaders and minimize the risk of transporting them. Regardless of the many potential additional considerations, our current effort provides an excellent basis for management, requiring relatively few variables that were reasonably easy to obtain. In addition, it may serve as a stepping stone for future research.

Detection of Allee effects

Our approach also advances the study of Allee effects in nature. Researchers have only rarely been able to demonstrate Allee effects in the context of population demography (i.e., establishment or extinction

probability) in natural populations (Fowler and Baker 1991, Veit and Lewis 1996, Liermann and Hilborn 1997, Groom 1998), although evidence from behavioral studies is more common (see Courchamp et al. 1999 and references therein). Most efforts to detect demographic Allee effects have incorporated natural variability into parameter uncertainty, which influences the ability to detect Allee effects (Myers et al. 1995, Veit and Lewis 1996, Caswell et al. 1999). Additionally, Allee effects are expected mostly in sparse populations, rendering accurate sampling difficult and resulting in large sampling errors. Further, in the presence of environmental heterogeneity, parameters may need to be estimated across many populations to obtain a general picture of Allee effects. These problems are compounded for invasion biology, given that invasions are rare events and failures are typically unrecorded.

Although the literature is quickly expanding, we know of only a few innovative approaches to test for Allee effects. Notably, Liermann and Hilborn (1997) analyzed fisheries time series and employed a hierarchical Bayesian method over multiple taxa in order to reduce the severe uncertainty exhibited in single populations. Veit and Lewis (1996) assayed for Allee effects by fitting a mass-action mating function in an integro-difference model for the spread of House Finch (*Carpodacus mexicanus*) in North America. Model validation was accomplished by visual comparison to observed population spread and alternative model parameterizations.

Our procedure complements these existing approaches and overcomes many of the previous problems associated with detecting Allee effects. We do not require estimation of parameters for each individual population. Instead we incorporate easily obtainable information across many populations and use information on potential and realized invasion events. We also take advantage of time series data on invasion events to improve our parameter estimates, and to provide quantitative model validation. Thus, this method provides a powerful means of detecting and estimating the strength of Allee effects. Wider application of this approach will be a useful component in forecasting future invasions to better inform management of invasive species.

ACKNOWLEDGMENTS

We thank J. M. Bossenbroek for sources of data, S. Stevens for data collection, and L. Johnson and another anonymous reviewer for their in depth and helpful comments. This work was funded by NSF Biocomplexity Incubation and IRCEB grants to D. M. Lodge.

LITERATURE CITED

- Amarasekare, P. 1998. Allee effects in metapopulation dynamics. *American Naturalist* **152**:298–302.
- Bossenbroek, J. M., J. C. Nekola, and C. E. Kraft. 2001. Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. *Ecological Applications* **11**:1778–1788.
- Buchan, L. A. J., and D. K. Padilla. 1999. Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecological Applications* **9**:254–265.
- Caswell, H., M. Fujiwara, and S. Brault. 1999. Declining survival probability threatens the North Atlantic right whale. *Proceedings of the National Academy of Sciences (USA)* **96**:3308–3313.
- Cohen, A. N., and J. T. Carlton. 1998. Accelerating invasion rate in a highly invaded estuary. *Science* **279**:555–558.
- Collette, D. 1994. *Modelling survival data in medical research*. Chapman and Hall, London, UK.
- Courchamp, F., T. Clutton-Brock, and B. Grenfell. 1999. Inverse density dependence and the Allee effect. *Trends in Ecology and Evolution* **14**:405–410.
- Dennis, B. 1989. Allee effects, population growth, critical density, and the chance of extinction. *Natural Resources Modeling* **3**:481–538.
- Dennis, B. 2002. Allee effects in stochastic populations. *Oikos* **96**:389–401.
- Etienne, R., B. Wertheim, L. Hemerik, P. Schneider, and J. Powell. 2002. The interaction between dispersal, the Allee effect and scramble competition affects population dynamics. *Ecological Modelling* **148**:153–168.
- Fotheringham, A. S., and M. E. O'Kelly. 1989. *Spatial interaction models: formulations and applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Fowler, C. W., and J. D. Baker. 1991. A review of animal population dynamics at extremely reduced population levels. *Reports of the International Whaling Commission* **41**:545–554.
- Groom, M. J. 1998. Allee effects limit population viability of an annual plant. *American Naturalist* **151**:487–496.
- Gyllenberg, M., J. Hemmini, and T. Tammaru. 1999. Allee effects can both conserve and create spatial heterogeneity in population densities. *Theoretical Population Biology* **56**:231–242.
- Hilborn, R., and M. Mangel. 1997. *The ecological detective: confronting models with data*. Princeton University Press, Princeton, New Jersey, USA.
- Hokanson, W. 2002. Zebra mussels spread to 166 inland lakes. *Michigan Riparian* **37**:13–15.
- Johnson, L. E., and D. K. Padilla. 1996. Geographic spread of exotic species: ecological lessons and opportunities from the invasion of the zebra mussel *Dreissena polymorpha*. *Biological Conservation* **78**:23–33.
- Johnson, L. E., A. Ricciardi, and J. T. Carlton. 2001. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. *Ecological Applications* **11**:1789–1799.
- Kolar, C. S., and D. M. Lodge. 2001. Progress in invasion biology: predicting invaders. *Trends in Ecology and Evolution* **19**:199–204.
- Kraft, C. E., and L. E. Johnson. 2000. Regional differences in rates and patterns of North American inland lake invasions by zebra mussels (*Dreissena polymorpha*). *Canadian Journal of Fisheries and Aquatic Sciences* **57**:993–1001.
- Lande, R. 1998. Demographic stochasticity and the Allee effect on a scale with isotropic noise. *Oikos* **83**:353–358.
- Leung, B., D. M. Lodge, D. Finnoff, J. F. Shogren, M. A. Lewis, and G. Lamberti. 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. *Proceedings: Biological Sciences* **269**:2407–2413.
- Lewis, M. A., and P. Kareiva. 1993. Allee dynamics and the spread of invading organisms. *Theoretical Population Biology* **43**:141–158.
- Liermann, M., and R. Hilborn. 1997. Depensation in fishstocks: a hierarchic Bayesian meta-analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **54**:1976–1984.
- Lockwood, R. N., D. M. Benjamin, and J. R. Bence. 1999. Estimating angling effort and catch from Michigan roving

- and access site angler survey data. Fisheries Research Report 2044. Michigan Department of Natural Resources, Fisheries Division, Lansing, Michigan, USA.
- Lodge, D. M. 2001. Lakes. Pages 277–312 in F. S. Chapin, III, O. E. Sala, and E. Huber-Sannwald, editors. Future scenarios of global biodiversity. Springer-Verlag, New York, New York, USA.
- Myers, R. A., N. J. Barrowman, J. A. Hutchings, and A. A. Rosenberg. 1995. Population-dynamics of exploited fish stocks at low population levels. *Science* **269**:1106–1108.
- National Invasive Species Counsel. 2001. Meeting the invasive species challenge: national invasive species management plan. National Invasive Species Counsel, Washington, D.C., USA.
- National Research Council. 2002. Predicting invasions of nonindigenous plants and plant pests. National Academy Press, Washington, D.C., USA.
- Pimentel, D., L. Lach, R. Zuniga, and D. Morrison. 1999. Environmental and economic costs of nonindigenous species in the United States. *BioScience* **50**:53–65.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1995. Numerical recipes in C: the art of scientific computing. Second edition. Cambridge University Press, Cambridge, UK.
- Ramcharan, C. W., D. K. Padilla, and S. I. Dodson. 1992. Models to predict the potential occurrence and density of the zebra mussel, *Dreissena polymorpha*. *Canadian Journal of Fisheries and Aquatic Sciences* **49**:2611–2629.
- Reed-Anderson, T., E. M. Bennett, B. S. Jorgensen, G. Lauster, D. B. Lewis, D. Nowacek, J. L. Riera, B. L. Sanderson, and R. Stedman. 2000. Distribution of recreational boating across lakes: do landscape variables affect recreational use? *Freshwater Biology* **43**:439–448.
- Sakai, A. K., et al. 2001. The population biology of invasive species. *Annual Review of Ecology and Systematics* **32**:305–332.
- Sala, O. E., et al. 2000. Biodiversity scenarios for the year 2100. *Science* **287**:1770–1774.
- Schneider, D. W., C. D. Ellis, and K. S. Cummings. 1998. A transportation model assessment of the risk to native mussel communities from zebra mussel spread. *Conservation Biology* **12**:788–800.
- Thomas, R. W., and R. J. Hugget. 1980. Modeling in geography. Barnes and Noble Books, Totowa, New Jersey, USA.
- Veit, R. R., and M. A. Lewis. 1996. Dispersal, population growth, and the Allee effect: dynamics of the house finch invasion of Eastern North America. *American Naturalist* **148**:255–274.