Estimating dispersal from patterns of spread: Spatial and local control of invasion by *Daphnia lumholtzi* in Missouri lakes

John E. Havel

Jonathan B. Shurin

John R. Jones

1—J.E. Havel, Department of Biology, Southwest Missouri State University, 901 S. National Avenue, Springfield, MO 65804, USA

2—J.B. Shurin, National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, CA 93101, USA

3—J.R. Jones, School of Natural Resources, University of Missouri, Columbia, MO 65201, USA
Abstract

The spread of exotic species can be limited by dispersal or by constraints imposed by the local environment. Using data collected from 153 Missouri lakes over seven years, we asked whether models based on dispersal or local-scale processes best predicted invasion by the exotic cladoceran *Daphnia lumholtzi*. We used multiple logistic regression to test the relative importance of 10 local physicochemical features and proximity to all known potential source populations for predicting which lakes became invaded. The decline in invasion likelihood with distance to source populations was used to estimate the shape of the dispersal kernel. Between 1992 and 1998, the cumulative prevalence of *D. lumholtzi* increased from 6 to 34% of lakes sampled, with frequent appearances of populations in new watersheds. Spatial position and physical factors were both important for predicting the new colonization events. The probability of colonization increased with lake surface area and epilimnetic temperature, declined with increasing conductivity, and was unaffected by variation in lake fertility. Invasion likelihood declined sharply as a non-linear function of distance to source populations up to around 30km, and was relatively constant at greater distances. The results suggest that dispersal and local abiotic constraints jointly limit the spread of *D. lumholtzi*. This approach illustrates how range expansion can be used to estimate dispersal rates at broad spatial scales.

Key words and phrases: biogeography, biological invasions, colonization, dispersal, dispersal kernel, *Daphnia lumholtzi*, exotic species, invasibility, rate of spread, reservoirs, zooplankton
Introduction

Humans have accelerated the rate of invasion by alien species in virtually every type of habitat on earth. Exotic species are a primary threat to biotic diversity and the integrity of natural communities and have altered rates of key ecosystem processes (Drake et al. 1989, Vitousek and Walker 1989, D'Antonio and Vitousek 1992, Sala et al 2000). Studies of the causes of species invasions generally take one of two broad perspectives. Rate-of-spread studies consider dispersal as the primary factor limiting distributions of invaders and spatial proximity to source populations as the major determinant of which habitats are invaded (Skellam 1951, Johnson and Padilla 1996, Kot et al. 1996, Veit and Lewis 1996, Clark et al. 1998, Buchan and Padilla 1999). This approach often focuses on the role of humans in facilitating dispersal of invasive organisms (Carlton and Geller 1993, Ricciardi and MacIssac 2000). Local-scale studies, by contrast, deal with aspects of communities that influence their susceptibility to invasion (Elton 1958, Moyle and Light 1996, Wiser et al. 1998, Levine and D'Antonio 1999, Levine 2000, Shurin 2000, 2001). These characteristics include resident species composition, diversity, trophic structure, and abiotic factors, such as temperature, habitat size, and chemical composition. Clearly, in order to invade new habitats, a species must both arrive there by dispersal and proliferate in the local environment. Dispersal and local-scale processes therefore jointly influence the incidence and abundance of exotic species. However, the relative importance of dispersal and local interactions in limiting the distributions of invaders are poorly understood and remain a subject of active debate (Wiser et al. 1998).

We studied the invasion of Missouri (USA) lakes by the exotic cladoceran *Daphnia lumholtzi* Sars to ask whether spatial position or local habitat features best predicted the spread of this species. *D. lumholtzi* is a native to Africa, Asia, and Australia, and was first reported in
North America from a Texas lake sampled in 1991 (Sorensen and Sterner 1992). Over the next two years, this species appeared across the south-eastern United States (Havel and Hebert 1993) and in 11 lakes in Missouri (Havel et al. 1995). Since that time, the species continued to expand its range through the continental United States (Stoeckel et al. 1996, Havel et al. 2000, Muzinic 2000) and invade many more lakes in Missouri (Fig. 1). The success of *D. lumholtzi* in North America has been attributed to its tolerance of high summer temperatures (Work and Gophen 1999a, 1999b) and resistance to fish predation due to defensive morphology (Swaffar and O'Brien 1996). Like other zooplankton, *D. lumholtzi* relies on passive dispersal to invade new habitats.

In the current study, we used patterns of invasion by *D. lumholtzi* in Missouri lakes to examine the roles of local abiotic features and spatial location as predictors of lake invasion. Zooplankton and 10 physicochemical features of 153 Missouri reservoirs were sampled over a seven-year period, during which time *D. lumholtzi* expanded its range considerably. To examine whether abiotic features limit the invasion, we used multiple logistic regression to predict which lakes became invaded based on measured habitat variables (Table 1). If *D. lumholtzi*’s distribution is constrained by particular physical and chemical tolerance limits, then variance in invasion success among lakes should be related to these variables. To assess the role of dispersal, we generated a predictor variable that summarized the position of each lake relative to all known potential sources of colonists. For all susceptible lakes (ones where *D. lumholtzi* was not found in any previous year), we included a term in the multiple logistic regression model that described the potential “load” of colonists as a function of distance to all known invaded lakes. We used likelihood profile techniques (Hilborn and Mangel 1997) to estimate the shape of the dispersal kernels. Dispersal kernels describe the decline in probability of invasion with
increasing distance from a source at geographic scales that are too large for direct measures. If dispersal plays a dominant role in limiting the spread of the invasion then the model with only the spatial term should fit the data better than one with only local variables. If \textit{D. lumholtzi}'s range is limited by both dispersal and local constraints, then we expect a model with both local and spatial terms to provide the best fit to the data.

**Methods**

**Study lakes.** We sampled 153 lakes in Missouri, USA, annually over seven years (Fig. 1). Most (137) of these lakes are reservoirs, ranging in size from main-stem impoundments to small tributary storage reservoirs used for water supply and recreation. These reservoirs occur in all physiographic regions of the state and show a wide range in size, lake fertility, and concentrations of dissolved and suspended solids (Jones and Knowlton 1993). The other 16 lakes were in the floodplain of the Missouri River, and included four oxbow lakes and 12 scour basins, formed from levee breaks during the flood of 1993 (Galat et al. 1998). These floodplain lakes tend to be high in nutrients and exhibit dynamic effects from flooding (Knowlton and Jones 1997). For convenience, we refer to both floodplain lakes and reservoirs as “lakes”.

**Sampling.** The reservoirs were sampled during 1992-98, with 83% of these lakes sampled four or more years. The 16 floodplain lakes were sampled during 1994 and 1995, with two lakes sampled again in 1996. All samples were collected in daylight hours during July and August, the period when \textit{D. lumholtzi} is most abundant in this region (Havel et al. 1995). Smaller lakes were sampled in mid-channel near the dam (or levee), whereas larger lakes were usually sampled from an up-lake site, where \textit{D. lumholtzi} tends to be most abundant (J. Havel and E.M. Eisenbacher, pers. obs.). Samples were also collected from the 12 floodplain lakes in
1994-1995. Because complete water chemistry sampling was not done on these sites, they are not included as dependent variables in the logistic regression models. They are, however, included as potential sources of colonists in the dispersal term for all models after 1995.

Zooplankton sampling and analysis. Zooplankton samples were collected with two or more vertical tows (total length 20m), using a 25cm diameter zooplankton net (mesh = 200 µm). The sample volume (982 L) should detect densities of *D. lumholtzi* over one per m³. The tows were pooled into one sample, anaesthetized with carbonated water, and preserved with buffered sugar-formalin. In order to avoid transmitting zooplankton between lakes, nets were thoroughly rinsed at each site. The entire sample from each site was later screened at 30X for *D. lumholtzi*, using the characteristics illustrated in Havel and Hebert (1993).

Limnological measures. In order to examine the effect of local features on invasion, 10 limnological characteristics were measured in each lake (Table 1). Reservoir surface areas (at conservation pool) were obtained from USGS topographic maps or from US Army Corps of Engineers brochures. Areas of the floodplain lakes were measured in 1994 (B. Thomas, Natural Resources Conservation Service, pers. com.). Depth profiles of temperature and oxygen were measured with a YSI Model 50B oxygen meter, conductivity with a YSI model 33 meter, and transparency with a Secchi disk (Lind 1985). Samples for water chemistry analysis were taken by combining four 1-L grab samples, each taken at ca. 0.5 m depth from the same site as the zooplankton sample. Whole water samples for total nitrogen and total phosphorus were placed on ice and stored frozen at −10°C. Total suspended solids samples were collected by filtering duplicate 250-1000 mL samples on tared Whatman GFC filters, with volume depending on the turbidity of the samples. For chlorophyll, duplicate 250 mL samples were filtered through GFC filters, placed in desiccant on ice, and later stored frozen with desiccant.
Analytical methods were by APHA (1985) for non-volatile suspended solids (NVSS), volatile suspended solids (VSS), and total phosphorus (TP). Chlorophyll (Chl) was extracted in heated ethanol (Sartory and Grobbelaar 1984) and analyzed by fluorometry (Knowlton 1984). Total nitrogen (TN) was analyzed by second derivative spectroscopy following persulfate oxidation (Crumption et al. 1992).

Ten abiotic features of the lakes (Table 1) were used as independent variables in the local term for the regression models. These features include indicators of primary productivity (total nitrogen, total phosphorus, total chlorophyll), transparency (Secchi depth, volatile and non-volatile suspended solids, turbidity), and lake physical features (surface area, temperature). All variables were measured on the day the lake was sampled in every year except temperature (not measured in 1998), conductivity (not measured in 1993), and turbidity (not measured in 1993). For these instances, we used the mean of all available samples as an estimate of the value for the year in which the data were missing. For each variable where mean values from other years were used, the “among years” variability was considerably less than the “among lakes” variability. Thus, the mean value from other years should be a good reflection of the actual conditions in the lake on the date sampled.

Spatial position and distance. For mapping purposes, the position of the centroid value of each lake in Missouri was calculated using Geographic Information Systems (ArcView v. 3.2. 1999. ESRI, Redlands, CA, USA), with a coverage provided by the U.S. Geological Survey (www.usgs.gov). The point locations for the lakes were recorded in the Universal Transverse Mercator (UTM) projection units. Since UTMs express lake position in units of meters, the Euclidian distance between all pairs of lakes could be readily calculated. For our dataset, inter-lake distances varied from 0.7 to 546 km (mean 211 km). The distance measures were used in
the statistical models to determine probabilities of non-invaded lakes receiving propagules from source lakes.

**Data analyses and modeling the invasion.** We illustrated the progress of the *D. lumholtzi* invasion with two descriptive measures. Prevalence is the number of lakes where *D. lumholtzi* was detected divided by the number of lakes surveyed in each year. Cumulative prevalence is the total number of lakes in which *D. lumholtzi* was ever detected divided by the cumulative number of lakes sampled.

For each year of sampling after 1992, our first year of zooplankton data, we modeled the probability that a susceptible lake became invaded as a function of measured local abiotic features and proximity to all known potential sources of colonists. A susceptible lake is defined as one that had been previously sampled where *D. lumholtzi* was never found. General linear models were used to test the roles of local habitat features and the potential supply of colonists. Invasion probability was assumed to be a linear function of the local abiotic environment and a non-linear function of proximity to potential sources of colonists. A non-linear function was used for the spatial term because the distribution of propagules for many organisms often drops off as a leptokurtic function of distance to a source (Wallace 1966, Willson 1993, Kot et al. 1996, Clark et al. 1998). To test for unimodal effects of abiotic variables, we included second-order terms for each predictor in the local model. The probability that a susceptible site $i$ becomes invaded ($p_i$) was therefore modeled using logistic regression (logit function) by \textbf{eq.1} and \textbf{eq.2} below:

\begin{equation}
\text{eq. 1) } p_i = \frac{\exp(\lambda_i)}{\exp(\lambda_i) + 1}
\end{equation}

where $\lambda_i$ is related to local and spatial factors as:
\[ \lambda_i = \beta_0 + \sum_{j=1}^{10} \beta_j (local) + \sum_{j=1}^{10} \chi_j (local)^2 + \beta_j \sum_{k=1}^{inf} \exp(-\frac{d_{ik}}{a}) \]

\( \beta_0 \) is the constant, while \( \beta_j \) and \( \beta_s \) are the scaling coefficients for the abiotic factors and the spatial terms, respectively, and \( \chi_j \) is the scaling coefficient for the quadratic of the abiotic term. The 10 abiotic variables are represented by \( j \) in the first two summation terms, \( k \) refers to the known potential source lakes in the survey, and \( inf \) is the total number of infected lakes. The distance between each potential source lake \( k \) and target lake \( i \) is \( d_{ik} \). The shape of the dispersal-by-distance function (the dispersal kernel) depends on the parameter \( a \) (described below, see Fig. 2). We tested two forms of the dispersal kernel, exponential (eq. 2 above) and Gaussian. For the Gaussian model, we replaced \( \exp(-d_{ik}/a) \) with \( \exp(-d_{ik}^2/a^2) \). The second derivative of the exponential function is continuously positive, while the Gaussian function has an inflection point. Exponential and Gaussian kernels had slightly different shapes at short distances and approached similar lower limits at greater distances. Since the general results from the Gaussian models were similar to those from the exponential models, we present only the results from the exponential models below.

In order to select the abiotic features to include in the model of local control, \( (i, \text{eq. 2}) \), we first performed simple logistic regression of the probability that a susceptible lake became invaded on each of the 10 abiotic features. Statistical analyses were done using S-Plus 2000 (MathSoft, Inc.). All variables were tested for normality using Liliefore’s test, and those that were non-normal were loge transformed (Trexler and Travis 1993, p. 1632). All variables were standardized to the same scale by subtracting the mean and dividing by the standard deviation after transformation. All independent variables and second order terms that showed significant univariate relationships with invasion were then entered together into a multiple stepwise logistic regression on invasion, using a backwards elimination procedure (Trexler and Travis 1993, Sokal...
This method was used to select the best model for local abiotic control of invasion success by *D. lumholtzi* (without the spatial term in eq. 2).

The role of spatial location was represented by a non-linear term (the third summation term in eq. 2) describing the position of each lake relative to all known potential source populations. The dispersal term represents the “load” of colonists received by a susceptible lake as a function of its proximity to all known invaded lakes in the survey under a given set of assumptions about how dispersal probability varies with distance. The approach to estimating the dispersal term in eq. 2 is shown schematically in Fig. 2. The potential load of colonists experienced by each susceptible lake was estimated by summing the dispersal kernels experienced by each target lake over all possible source populations (Fig. 2A).

In order to assess the contribution of the spatial term in eq. 2, we must first estimate the values of the unknown parameters \( a \) and \( \beta \). The dispersal parameter \( (a) \) in eq. 2 determines the shape of the dispersal kernel while the intercept \( (\beta) \) scales the kernel to the actual probability of invasion. Low values of \( a \) indicate that dispersal is highly localized around source lakes, while large values mean that propagules are broadly dispersed (Fig. 2B). The best fit values of \( a \) were estimated by the likelihood profile method (Hilborn and Mangel 1997). The full model shown in eq. 2 was evaluated for all integer values of \( a \) between 1 and 100 for the exponential dispersal function. The coefficients \( (\beta) \) in eq. 2 were estimated by likelihood iterations using the logistic regression procedure in S-Plus 2000 (MathSoft, Inc.). We assumed that the value of \( a \) that minimized the negative log likelihood of the model presents the best estimate of \( a \). That is, the model that minimizes the negative log likelihood explains the greatest portion of the variance in invasion success and therefore offers the most likely shape for the dispersal function. We compared models including the abiotic and spatial terms alone, and with both together, using the
Akaike information criterion AIC (Hilborn and Mangel 1997). The model with the lowest AIC value offered the best prediction of invasion probability. AIC was used because we were comparing non-nested models (i.e., abiotic vs. spatial, Hilborn and Mangel 1997). The 95% confidence limits on the estimate of the parameter $a$ were determined using a critical Chi-Square value of 3.84. That is, values of $a$ for which $-2 \times \log$ likelihood is 3.84 greater than the minimum value are outside the 95% confidence interval (Hilborn and Mangel 1997, p. 174).

The modeling approach we used makes four important assumptions. The first is that lakes where *D. lumholtzi* was not found in any previous year’s samples were susceptible to invasion. The range for 1992 (Fig. 1) includes all populations known from a survey of 112 reservoirs in Missouri (Havel et al. 1995). Surveys of 43 reservoirs during 1980-86 found no populations (W.R. Mabee, Missouri Department of Conservation, pers. com.). The second assumption was that lakes where *D. lumholtzi* was recorded in one year were considered invaded in all subsequent years, and could serve as potential sources of colonists (dispersal term in eq. 2). This assumption was made for two reasons. First, the volume of water from which zooplankton were sampled was very small relative to the volume of the lakes (ca. 1 m$^3$ vs. 20,000 m$^3$ for the smallest lake). It is therefore likely that we would fail to detect populations that were present at low densities. Second, *D. lumholtzi* forms resting eggs (J. Havel, pers. obs.) which, in other *Daphnia*, remain viable in lake sediments for many years (Cáceres 1998). The egg bank may re-establish the planktonic population in lakes where these plankton had become “extinct”.

Furthermore, because resting eggs are tolerant of desiccation, they provide the likely propagules for dispersal to other lakes. Therefore, all lakes where *D. lumholtzi* was found in any previous year were considered to be potential sources of colonists in the spatial term of the model (eq. 2). The third assumption was that invaded lakes were not potential sources of propagules in the first
year that *D. lumholtzi* was found in them. This assumption was made because the population
must first develop parthenogenetically before producing resting eggs. The pool of potential
source lakes in a given year therefore consisted of all lakes where *D. lumholtzi* was found in any
previous year. Finally, we assumed that the size of the source lakes did not affect the probability
of invasion. We analyzed the 1995 data (the year with the most invasions) after weighting the
dispersal term in equation 2 by the surface area of the source lake and found no improvement in
the fit of the model. Thus all remaining analyses used the un-weighted data.

The logistic regression analyses were performed for each year individually except 1997
and 1998. Only two new populations were detected in each of these years; therefore the models
had very little statistical power and tended to produce unstable parameter estimates. The
analyses were also performed after combining data for all susceptible lakes from multiple years,
including 1997 and 1998. Analyzing the data across years increased the power to discern the
shape of the dispersal function and allowed us to detect general patterns that were independent of
inter-annual variation.

**Results**

**Prevalence and the progress of the invasion.** *D. lumholtzi* dramatically expanded
its range during the study period. Each year, new lakes and watersheds were invaded (Fig. 1).
Prevalence showed a continuous increase from six to 26% from 1992-95 and, with the exception
of 1997, maintained a high prevalence (≥ 23%) each year thereafter (Fig. 3). Once lakes were
invaded, most populations were detected in subsequent years. For instance, 70% of the lakes
invaded through 1994 had detectable populations in 1995. The cumulative prevalence function
appeared to be leveling off by 1996, suggesting that the rate of new invasions had slowed down. By 1998, about 35% lakes of the sampled lakes had been invaded (Fig. 3).

Local features of invaded and non-invaded lakes. The invaded lakes were generally similar to the non-invaded lakes in terms of physical and chemical conditions, and many of these features showed large variation within each group (Table 1). Invaded lakes had significantly larger surface areas and warmer epilimnetic temperatures than non-invaded lakes, while the two categories were indistinguishable in terms of lake fertility, water clarity, and suspended solids.

The logistic regression models from individual years revealed that invasion probability generally increased with lake surface area and epilimnetic temperature (Table 2). However, these effects varied among years. Surface area contributed significantly to the fit of the model in 1993 and 1994, but not in 1995 or 1996. When data from multiple years were analyzed simultaneously, both area and temperature had significant positive effects on invasion probability (Table 3). The response to temperature was unimodal in 1995 and 1996 (Table 2) and for pooled data for 1993-1998 (excluding 1996, Table 3). That is, both the first and second order terms for temperature were significant in those years, with a negative coefficient for the second order term. The regression models also suggest that invasion probability was depressed by higher conductivity in 1995 and from the multiple years, when a non-linear effect was detected (Tables 2 and 3). Although invasion probability declined as a non-linear function of increasing conductivity, the likelihood never reached a minimum over the range of conductivity present in the survey. Measures of lake fertility (total nitrogen, total phosphorus, and chlorophyll) and water clarity (Secchi depth transparency, turbidity, and volatile and non-volatile suspended solids) showed no relationships with invasion probability in any of the years or in the combined models (Tables 2 and 3).
Modeling both local and spatial features. Including a term describing spatial position relative to all known invaded lakes (the dispersal term in eq. 2, Fig. 2) significantly improved the fit of the model in 1995 and 1996, but not in 1993 or 1994 (Table 2). The model with both abiotic and spatial terms had the lowest AIC in 1995 and 1996, while the abiotic-only model had the lowest value in the other two years (Table 2). The model with only the spatial term always had the highest AIC and therefore provided the worst fit to the data (Table 2). When data from multiple years were analyzed together, the lowest AIC was always found in the model with both spatial and abiotic terms (Table 3). The likelihood profiles for the non-linear parameter $a$ showed clear minima in every year (Fig. 4). The value of $a$ which produced the lowest negative log likelihood ($a_{min}$) is the optimal value for $a$ for inclusion in the models for the dispersal kernels. The value of $a_{min}$ ranged from 0.6-25 (mean=18.5) for the different years. All values of $a$ between 1 and 100 fell within the 95% confidence limits of the minimum value in every individual year (Fig. 4), as did all but the lowest values of $a$ for the pooled data (Fig. 5). The wide confidence intervals indicate that, although the likelihood profiles showed clear minima, no values of $a$ between 1 and 100 offered significantly improved predictions of invasion except when the data were pooled across years.

The likelihood profile for 1996 showed a substantially different pattern from the other years (Fig. 4). Because the likelihood profile did not reach a minimum for any value of $a$ between 1 and 100, we explored values between 0.1 and 4 (inset graph in Fig. 4), revealing $a_{min}$ at a value of 0.6. This low estimate can be compared with the estimates of $a_{min}$ during the period 1993-95, when *D. lumholtzi* was expanding at the greatest rate (Fig. 3). During that period, $a_{min}$ ranged from 13-25; therefore $a_{min}$ was substantially lower in 1996 than in the previous three years. Furthermore, the logistic regression model for the 1996 data had negative coefficients for
the spatial term (Table 2), suggesting that invasion likelihood increased at greater distances from potential sources.

Analysis of the pooled data revealed similar patterns to the individual years. When all years except 1996 were analyzed together, the profile had a clear minimum at $a=19$ and closely resembled the profiles from 1993-1995 (compare Fig. 5 top and bottom). In contrast, when data from all seven years were included, the minimum was considerably larger ($a_{\text{min}} = 98$).

The dispersal kernels corresponding to the spatial term in the logistic regression models are shown in Fig. 6. In every case except 1996, the probability that a lake became invaded decreased sharply up to distances of around 20-40 km from a source of colonists, and approached a lower asymptote by 80-100km. Based on the dispersal kernels from 1993 to 1995, an average of 82% of the colonization events within 100 km of a source took place within the nearest 30 km and, for the pooled data (excluding 1996), 80% of invasions were within 30 km. The shapes of the dispersal kernels suggest that the spread of $D. \text{lumholtzi}$ is localized on a scale of 10s of kilometers and decreases at distances greater than around 30km.

**Discussion**

The present study indicates that both local factors and dispersal among lakes are important for limiting the spread of $D. \text{lumholtzi}$. Invasion likelihood was greatest in larger lakes and those with warmer epilimnetic temperatures. In addition, lakes that were geographically closer to potential source populations tended to be invaded more often than more isolated lakes. The method we used for analyzing the effects of spatial position illustrates one way of considering potential dispersal effects in statistical models of the spread of invaders. This approach allows us to incorporate information about all known potential sources of colonists in
assessing the likelihood of invasion for each susceptible lake. In addition, this approach allows us to estimate the shape of dispersal functions at large spatial scales (100s of km). Our results have implications for interpreting the invasion biology of *D. lumholtzi*.

**Local factors.** Although quite variable within each group, invaded lakes tended to be larger than non-invaded lakes. This effect was detected in the first two years of our study, as well as in the data from the pooled years (Tables 2 and 3). The strong effect of area is reasonable for two reasons. First, surface area may increase the likelihood a lake receives colonists. Larger lakes offer larger targets for propagules dispersed by wind and may be visited more often by waterfowl or boaters. Recent evidence from boater surveys suggests that recreational boats are capable of dispersing zooplankton among lakes (Havel and Stelzleni-Schwent 2001). The largest reservoirs are also often located downstream from smaller reservoirs, indicating that they also may receive colonists through surface water from other lakes (Shurin and Havel, unpublished). Second, aspects of the local environment associated with surface area may also increase the likelihood for invasion success once propagules have arrived. Larger lakes have greater habitat heterogeneity and may support larger local populations that are less susceptible to stochastic extinction (Angermeier and Schlosser 1989). For instance, large reservoirs show broader horizontal gradients in productivity than small reservoirs (Thornton et al. 1990). Since the timing of *D. lumholtzi* population peaks depends on location in the reservoir (Havel and Eisenbacher, unpublished data), heterogeneity within a lake may be important for population persistence. Larger lakes may therefore be invaded more often both because they receive more propagules and because the local environment favors successful invasions. Because of their relevance to both biogeography (Angermeier and Schlosser 1989) and
metapopulation theory (Hanski 1994), it is not surprising that such area effects are important also to the ecology of invading populations.

The epilimnetic temperatures of invaded lakes tended to be slightly warmer than those of non-invaded lakes. This effect is evident from the regression models for 1995 and 1996 and from data pooled from all years (Tables 2 and 3). *D. lumholtzi* typically shows brief peaks of maximum abundance during midsummer (Havel et al. 1995), a period when native *Daphnia* usually decline (Threlkeld 1986, Havel and Eisenbacher, unpublished). Recent experiments suggest that *D. lumholtzi* has both a high thermal tolerance (30°C, Work and Gophen 1999b) and a warm thermal optimum (*r*$_{max}$ at 25°C, Lennon 1999). These data plus the tropical native range for this species suggest that *D. lumholtzi* is well adapted to living in warmer lakes and may thus be occupying a vacant niche. Nevertheless, the fact that this species has invaded Lake Erie (Muzinic 2000) suggests its thermal niche is broad.

Lakes invaded by *D. lumholtzi* tended to have a lower conductivity than those not invaded (Table 3). Nevertheless, there was no minimum over the range of conductivities observed in the current study. In Missouri lakes, conductivity is linked to region of the state, with the highest values in the Osage Plains, where many of the lakes are small (Jones 1977). Hence the conductivity effect may be an artifact of landscape position.

The current study detected no effect of lake fertility on invasibility, despite the considerable variation among study lakes in nutrient concentrations (Table 1, Jones and Knowlton 1993). This result contrasts with some previous work. In an earlier study of Missouri lakes, Havel et al. 1995 found higher levels of total nitrogen in invaded lakes than in non-invaded lakes. In contrast, in a recent study of 35 reservoirs in Kansas, invaded lakes tended to have lower levels of nitrogen, phosphorus and chlorophyll-a than non-invaded lakes (Dzialowski
et al. 2000). Similarly, in a recent mesocosm experiment (Lennon 1999), total nitrogen concentrations in invaded tanks were lower than those in reference tanks, although no significant differences were evident for total or soluble reactive phosphorus. Overall, these patterns suggest that *D. lumholtzi* is not often excluded from lakes due to variation in lake fertility and primary productivity.

We lacked data on other potentially important local features, and may therefore have underestimated the extent of local control over invasion by *D. lumholtzi*. For instance, experimental evidence (Shurin 2000, 2001) has shown that species interactions such as predation and competition are important for generating invasion resistance in zooplankton communities. Composition of the local biota, such as density of planktivorous fish, may have played a major role in determining which lakes were invaded by *D. lumholtzi*. However, we lacked information on the biotic communities of the lakes in our survey and so were unable to examine biotic effects. In addition, lake age may have influenced colonization. The reservoirs in the study were constructed between the 1920s and 1970s. *D. lumholtzi* has only invaded this system during the past 12 years. Although most of the differences among these reservoirs in fertility and conductivity are due to landscape position and not lake age (Jones and Knowlton 1993), other reservoir aging effects, such as decreasing depth and subsequent increases in temperature, could lead to conditions more conducive to invasion. Lake age may also influence structure of the biotic community in ways that affect invasibility. Despite the limited available data on local features of the lakes, our analysis indicated a major role for local abiotic control of invasion by *D. lumholtzi*.

**The effect of spatial position.** Lakes in close proximity to source lakes generally had a higher likelihood of colonization than from more remote lakes. Incorporating the term for spatial
position in eq. 2 improved the fit of the model (relative to the abiotic model) in 1995 and 1996 and when data from multiple years were analyzed together (Tables 2-3). In 1993 and 1994, the abiotic model offered the best prediction of invasion. The model with only the spatial term never provided the best fit to the data. The lack of significant spatial effects in two of the years may be a result of low statistical power. Even though the number of lakes sampled was large, the number of newly invaded lakes in most years was small (range 2-12). Nevertheless, we were able to detect effects of both local factors and spatial location.

The shapes of the dispersal kernels for *D. lumholtzi* suggest that dispersal is most likely within 30km of source lakes, but that dispersal at longer distances still occurs at a measurable rate. Several long-distance hops are apparent in the distribution maps for Missouri (Fig. 1). For instance, in 1995, new populations appeared in the southeastern and northern regions of the state that were quite remote from previous populations. Furthermore, the appearance of new populations in the western US (Arizona--J. Elser, pers. com.; Utah--C. Leucke, pers. com.) points to the importance of long-distance colonization. The likelihood of long-distance dispersal (100km) is 8-33% that of short-distance dispersal (within 1km of invaded lakes, Fig. 6). This result indicates that very remote lakes still had an appreciable probability of invasion in any given year. The overall rate of invasion among all the lakes (indicated by the horizontal line in Fig. 6) is above the dispersal kernel. This is because the kernel represents the probability of invasion from an *individual* lake, whereas lakes experience colonization from multiple potential sources.

The dispersal kernels generated by the models of invasion in 1996 were qualitatively different from the other years (Fig. 6). The kernels that resulted from the likelihood profiles had negative coefficients, implying that lakes that were *farther* from potential sources were more
likely to be colonized. The reason for the anomalous results are unknown, although the map of the invasion (Fig. 1) reveals a shift in distribution from the middle of the state in 1995 toward the southwestern region in 1996. However, the results from the other years and from the combined data were remarkably consistent (Fig. 6).

The way in which dispersal was incorporated in our model may have influenced our ability to detect spatial effects for several reasons. First, we assumed that Euclidean distance is the relevant metric of spatial position for the dispersal of *D. lumholtzi*. If the species is dispersed primarily by stream connections or boat traffic, the shortest distance between two lakes may be less important than distance via rivers or highways. A related study (Shurin and Havel, unpublished) examined the role of stream connections to source populations in the spread of the invasion. Although lakes with upstream source populations were invaded at a rate of 27% per year, lakes with no upstream lakes were still invaded at a rate of 7%. Stream connections may therefore provide one means of transport of *D. lumholtzi*, although overland dispersal occurs frequently as well. Our second assumption was that all populations exerted equivalent colonization pressure on neighboring lakes. We would expect lakes with larger populations to produce more emigrants. Although we had no information on local density, we assumed that population size is proportional to lake area. Weighting the dispersal term in equation 2 by the area of the source lakes did not improve the fit of the model. Finally, our current analysis is restricted to lakes in Missouri, whereas *D. lumholtzi* propagules could have come from source lakes in neighboring states. To correct for this possibility, we re-analyzed the data for the year of maximum invasions (1995) after omitting susceptible lakes within 30km of the border. This analysis indicated no qualitative effects on any of our results, suggesting that edge effects were of minor importance for producing the observed patterns in our study. Although the
representation of dispersal in our statistical model was simplified, including complexities such as stream connections, the size of source lakes, and edge effects had only minor impacts on our results.

The approach used in the current study shares similarities to metapopulation studies that estimate colonization and extinction rates as a function of patch size and isolation (Hanski 1994, Moilanen and Hanski 1998, Moilanen et al. 1998, Moilanen 1999, Harrison et al. 2000). In contrast to analyses of static distributions such as incidence function approaches (Hanski 1994, Moilanen 1999), we modeled colonization events over time, a dynamic approach shared with other recent studies (Wiser et al. 1998, Moilanen et al. 1998, Harrison et al. 2000). Few metapopulation studies have incorporated the role of local habitat features (e.g., Moilanen and Hanski 1998). In contrast, our model included extensive data on local habitat features along with spatial location and indicated a major role for local control relative to dispersal for limiting the spread of *D. lumholtzi*. This conclusion would have been obscured if local features of the lakes were not included in the model. Adding local biotic factors into models of spread is an important next step toward studying the constraints on species’ geographic ranges.

**The rate of invasion over a broad geographic scale.** Over the period of our study, the prevalence of *D. lumholtzi* in Missouri lakes increased from 6-34% and numerous new watersheds were invaded. During the same period, this species also expanded from the southeast and south-central US to the upper mid-west and western states (Havel, in prep). These long-distance hops are characteristic of “stratified diffusion”, such as observed in the post-glacial dispersal of oak trees in Great Britain (Hengeveld 1989). The rapid range expansion of *D. lumholtzi* contrasts with the North American invasion by another exotic crustacean zooplankter. The Eurasian *Bythotrephes cederstroemi* (and its close relative *Bythotrephes longimanus*) is
another easily-recognized cladoceran and was first discovered in the Lake Huron in 1984. Over a period of 10 years, this species invaded the adjoining Great Lakes plus 16 lakes in Minnesota and Ontario (Hall and Yan 1997). *B. cederstroemi* is primarily confined to large, oligotrophic lakes (MacIsaac et al. 2000) and has to date maintained a limited geographic extent in North America.

The rapid long-distance dispersal of *D. lumholtzi* suggests that this species is capable of taking advantage of dispersal vectors that operate over broad spatial scales (100s of km). Studies of zooplankton dispersal have concentrated on the roles of birds (Proctor and Malone 1965) and wind (Brendonck and Riddoch 1999, Jenkins and Underwood 1998) over short distances, but provided no evidence on dispersal at broader scales. Another potential vector is the live well of recreational boats. In a recent study, Havel and Stelzleni-Schwent (2001) found that boats move rapidly between lakes and that *D. lumholtzi* can survive in the live wells during summer. Such movements could readily transport this species over long distances. We do not yet know if *D. lumholtzi* has a greater tolerance of the conditions in live wells than other zooplankton. However, species having lower temperature tolerances would likely fail to move with boats during the summer. Clearly, the broad tolerance of a wide range of freshwater environments by *D. lumholtzi* implies that many lakes and wetlands are invasible by this species.

The roles of dispersal and local interactions in invasion biology have been a subject of much debate (Moyle and Light 1996, Wiser et al. 1998, Ricciardi and MacIssac 2000, Shurin 2000). Our study indicates that range expansion by *D. lumholtzi* is limited both by dispersal among lakes and by local (within lake) constraints. In addition, the approach presented here illustrates how range expansion by exotic species can be used to obtain estimates of dispersal kernels at spatial scales that are too large for direct measures. A number of studies (Hengeveld
1989, Kot et al. 1996, Veit and Lewis 1996, Clark et al. 1998) have used empirical estimates of dispersal kernels (e.g., seed shadows from trees) to predict rates of geographic spread of populations. The approach presented here works in the opposite direction by using spread to derive dispersal kernels, and has the advantage of incorporating spatial position relative to all known potential sources, as well as information about the local environment. Estimating dispersal rates at broad scales remains a difficult challenge in invasion biology because long-distance dispersal events are infrequent and difficult to observe directly. Exotic species can act as biological tracers that provide opportunities to examine dispersal among widely separated sites across the landscape.

Acknowledgements

We thank the numerous students who assisted with the fieldwork, especially Casey Meek, Matt Eisenbacher, John Calfee, and John Kubisiak. Expert technical assistance was provided by Bruce Perkins, Dan Obrecht, Brian Fredrick, and Jeanne Stephens. Bill Mabee provided the 1998 distribution data and Brian Thomas and Matt Knowlton provided the physical and chemical data on the Missouri River floodplain lakes. The profile likelihood method used in our study builds on techniques developed as part of the “Spatial Ecology of Infectious Disease” working group at NCEAS. We thank Jonathan Swinton and Ottar Bjornstad for access to unpublished work. Our work also benefited from discussions with Perry de Valpine, William Morris, Shane Richards, and Stanley Dodson, and comments on the manuscript by David Jenkins, Carol Lee, William Morris, Tim Wootton, and an anonymous reviewer. Financial support was provided by grants from the National Science Foundation (DEB 93-17869) and
Missouri Department of Conservation to JEH and the Missouri Department of Conservation and Missouri Department of Natural Resources to JRJ. Postdoctoral support for JBS was provided by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (DEB-0072909) and the University of California.

References


Figure legends

Fig. 1. Cumulative distribution of *Daphnia lumholtzi* in Missouri lakes, 1992-98. Open circles represent the position of lakes where *D. lumholtzi* was not detected in the year indicated or any previous year. Closed circles represent lakes where *D. lumholtzi* was present in either that year or in samples from earlier years. A total of 153 lakes were sampled over this period, with 80 of the same lakes sampled every year except 1997-98, when fewer lakes were sampled. The number of lakes sampled each year is indicated in figure 3. Number of new populations detected each year: 1992—7, 1993—6, 1994—5, 1995—20, 1996—12, 1997—1, 1998—2.

Fig. 2. Schematic diagram of the method for estimating the dispersal kernels. (A) Hypothetical scenario of a target lake (T) within the exponential dispersal kernels of two potential source lakes (1 and 2). The interlake distances (d₁,T and d₂,T) are the Euclidean distances between the centroids of the lakes. The potential load of colonists [exp(-d/a)] provided by each source is represented by the curved lines. The dispersal load experienced by the target lake is given by summing the loads from each source. The load is scaled to the actual probability by the coefficient (β) in the logistic regression model (eq. 2 in text). (B) Different dispersal kernels corresponding to four values of a using the exponential dispersal function. We varied the value of a from 1 to 100 in integer units and assessed the likelihood of each resulting model.

Fig. 3. Prevalence (solid symbols) and cumulative number of populations detected (open symbols) of *Daphnia lumholtzi* during the study period. Numbers of lakes sampled in each year are indicated above each data point. A total of 153 lakes were sampled during the study.
Fig. 4. Likelihood profiles for the models to predict invasion including abiotic and spatial terms with different values of $a$ (eq. 2) between 1 and 100. The y-axis shows $-2$ times the log likelihood for the model including abiotic and spatial terms for each value of $a$. The value of $a$ that minimizes the negative log likelihood of the model is indicated by $a_{\text{min}}$. The inset on the 1996 graph shows the likelihood profile for values of $a$ between 0.1 and 4 in increments of 0.1.

Fig. 5. Likelihood profiles for analyses combining data from multiple years. Solid symbols indicate values of $a$ that result in models within the 95% confidence limits of the minimum value ($a_{\text{min}}$). Open symbols indicate values of $a$ that are outside of the 95% confidence interval for the minimum.

Fig. 6. Dispersal kernels by year. The graphs show the change in probability of invasion ($p_i$) as a function of distance to a source population for the models based on data from different years. The horizontal lines show the proportion of susceptible lakes that became invaded in the year indicated.
\[ \Sigma \text{(dispersal)} = \exp(-d_{1,T}/a) + \exp(-d_{2,T}/a) \]
-2 log likelihood

-2 log likelihood vs. a for different periods:

1993-1995:
- $a_{\text{min}} = 23$

1993-1998:
- $a_{\text{min}} = 98$

1993-1998, no 1996:
- $a_{\text{min}} = 19$
Table 1. Characteristics of lakes invaded or not invaded by *D. lumholtzi* over seven years of sampling. Values shown are medians (among lakes), with the range indicated in parentheses. Data from each lake were first averaged over all available samples (years). The 16 floodplain lakes sampled in 1994-5 are not included. Range is indicated in parentheses and variables with asterisks were significantly different groups (t-tests, p < .01). Variable codes: TN—total nitrogen, TP—total phosphorus, chl-a—chlorophyll-a, NVSS—nonvolatile suspended solids, VSS—volatile suspended solids.

| Variable          | Not invaded (n= 96) | Invaded (n= 40) |
|-------------------|---------------------|----------------|-----------------|
| Area (ha) *       | 33.2                | 130.7          |                 |
|                   | (3.2 – 5819.5)      | (7.3 - 24,087.4) |
| Epilimnetic       | 24.5                | 24.8           |                 |
| temperature (°C) *| (10.3 – 27.7)       | (22.5 – 27.0)  |
| Secchi depth (m)  | 0.94                | 1.0            |                 |
|                   | (0.1 – 4.6)         | (0.22 – 3.3)   |
| TN (µg L⁻¹)       | 742.9               | 747.8          |                 |
|                   | (189.4 – 2,537.5)   | (244 – 1,446.4)|                 |
| TP (µg L⁻¹)       | 4.8                 | 39.6           |                 |
|                   | (6.6 – 277.6)       | (11.1 – 221.9) |
| Chl-a (µg L⁻¹)    | 17.4                | 16.4           |                 |
|                   | (1.5 – 178.6)       | (3.1 – 59.5)   |
| NVSS (mg L⁻¹)     | 4.7                 | 4.7            |                 |
|                   | (0.6 – 41.7)        | (0.7 – 46.0)   |
| VSS (mg L⁻¹)      | 3.2                 | 2.8            |                 |
|                   | (0.4 – 24.0)        | (0.9 – 10.9)   |
| Conductivity (µS) | 196.1               | 210.1          |                 |
|                   | (42.5 – 749.6)      | (87.1 – 382.6) |
| Turbidity (NTU)   | 5.8                 | 5.3            |                 |
|                   | (0.6 – 76.1)        | (1.4 – 27.4)   |
Table 2. Models to predict invasion by *D. lumholtzi*. The abiotic models are based on local physical characteristics of the lakes. Independent variables and second-order terms that contributed significantly to the fit of the model (*P*<0.05) are shown here. Variable codes: A—area (ha), T—surface temperature (C), C—conductivity (µS), d—distance (km). The “spatial” and “spatial + abiotic” models include the dispersal term in eq. 2 representing the potential load of colonists. The value of the parameter *a* shown minimizes the negative log likelihood of the model based on the likelihood profile (Fig. 4). The log likelihood (LL), number of parameters (K), and AIC are shown for each model. AIC is calculated as -2*LL + 2*K. The model with the lowest AIC is shown in bold. The *P* value for the “spatial + abiotic” models is the significance for inclusion of the dispersal term in the model.

<table>
<thead>
<tr>
<th>Year</th>
<th>Model</th>
<th>Regression equation (<em>λ</em>)</th>
<th>LL</th>
<th>K</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>abiotic</td>
<td>-3.92 + 1.51*A</td>
<td>-14.73</td>
<td>2</td>
<td>33.46</td>
</tr>
<tr>
<td></td>
<td>spatial</td>
<td>-3.28 + 1.35*exp(-d/38)</td>
<td>-22.76</td>
<td>3</td>
<td>51.51</td>
</tr>
<tr>
<td></td>
<td>spatial +</td>
<td>-4.21 + 1.55<em>A + 1.56</em>exp(-d/25), <em>P</em>=0.54</td>
<td>-14.49</td>
<td>4</td>
<td>36.98</td>
</tr>
<tr>
<td></td>
<td>abiotic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>abiotic</td>
<td>-4.28 + 1.70*A</td>
<td>-8.88</td>
<td>2</td>
<td>21.76</td>
</tr>
<tr>
<td></td>
<td>spatial</td>
<td>-3.35 - 234628*exp(-d/1)</td>
<td>-13.12</td>
<td>3</td>
<td>32.24</td>
</tr>
<tr>
<td></td>
<td>spatial +</td>
<td>-4.65 + 1.83<em>A + 2.96</em>exp(-d/13), <em>P</em>=0.40</td>
<td>-8.70</td>
<td>4</td>
<td>25.40</td>
</tr>
<tr>
<td></td>
<td>abiotic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>abiotic</td>
<td>-1.10 + 2.11<em>T – 1.43</em>T^2 – 0.99*C^2</td>
<td>-26.06</td>
<td>4</td>
<td>60.13</td>
</tr>
<tr>
<td></td>
<td>spatial</td>
<td>-2.24 + 0.97*exp(-d/39)</td>
<td>-39.48</td>
<td>3</td>
<td>84.97</td>
</tr>
<tr>
<td></td>
<td>spatial +</td>
<td>-1.62 + 2.31<em>T – 2.12</em>T^2 – 1.09*C^2 +</td>
<td>-23.90</td>
<td>6</td>
<td>59.80</td>
</tr>
<tr>
<td></td>
<td>abiotic</td>
<td>3.15*exp(-d/17), <em>P</em>=0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>abiotic</td>
<td>-2.50 + 3.56<em>T – 1.85</em>T^2</td>
<td>-20.70</td>
<td>3</td>
<td>47.40</td>
</tr>
<tr>
<td>Spatial</td>
<td>-1.95 – 8689702*exp(-d/0.6)</td>
<td>-24.23</td>
<td>3</td>
<td>54.47</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------</td>
<td>--------</td>
<td>---</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Spatial +</td>
<td>-2.19 + 3.47T – 1.93<em>T^2 – 2683881</em>exp(-d/0.6),</td>
<td>-19.55</td>
<td>5</td>
<td>47.09</td>
<td></td>
</tr>
<tr>
<td>Abiotic</td>
<td>P=0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Logistic regression analyses combining data from multiple years. Variable codes are as in Table 2.

<table>
<thead>
<tr>
<th>Year</th>
<th>Model</th>
<th>Regression equation ($\lambda$)</th>
<th>LL</th>
<th>K</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993-1995</td>
<td>abiotic</td>
<td>$-2.38 + 0.93<em>A + 0.61</em>T - 0.79*C^2$</td>
<td>-60.60</td>
<td>4</td>
<td>129.20</td>
</tr>
<tr>
<td></td>
<td>spatial</td>
<td>$-3.07 + 1.77*\exp(-d/28)$</td>
<td>-69.79</td>
<td>3</td>
<td>145.57</td>
</tr>
<tr>
<td></td>
<td>spatial +</td>
<td>$-2.90 + 1.01<em>A + 0.53</em>T - 0.88<em>C^2 + 2.00</em>\exp(-d/23)$, $P=0.006$</td>
<td>-57.21</td>
<td>6</td>
<td>126.42</td>
</tr>
<tr>
<td></td>
<td>abiotic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993-1998</td>
<td>abiotic</td>
<td>$-2.38 + 0.88<em>A + 1.27</em>T - 0.75*T^2$</td>
<td>-98.63</td>
<td>4</td>
<td>205.26</td>
</tr>
<tr>
<td></td>
<td>spatial</td>
<td>$-2.68 + 0.36*\exp(-d/27)$</td>
<td>-114.42</td>
<td>3</td>
<td>234.83</td>
</tr>
<tr>
<td></td>
<td>spatial +</td>
<td>$-2.94 + 0.95<em>A + 1.38</em>T - 0.74*T^2 + $</td>
<td>-96.38</td>
<td>6</td>
<td>204.75</td>
</tr>
<tr>
<td></td>
<td>abiotic</td>
<td>$0.12*\exp(-d/97.5)$, $P=0.03$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993-1998,</td>
<td>abiotic</td>
<td>$-2.15 + 0.91<em>A + 1.03</em>T - 0.82<em>T^2 - 0.45</em>C^2$</td>
<td>-73.69</td>
<td>5</td>
<td>157.37</td>
</tr>
<tr>
<td>no 1996</td>
<td>spatial</td>
<td>$-2.83 + 1.33*\exp(-d/16)$</td>
<td>-85.87</td>
<td>3</td>
<td>177.75</td>
</tr>
<tr>
<td></td>
<td>spatial +</td>
<td>$-2.54 + 0.99<em>A + 1.00</em>T - 0.79<em>T^2 - 0.44</em>C^2 + $</td>
<td>-70.02</td>
<td>7</td>
<td>154.04</td>
</tr>
<tr>
<td></td>
<td>abiotic</td>
<td>$1.16*\exp(-d/19)$, $P=0.002$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>