Biological Invasions (2006) 8: 241–254
DOI 10.1007/s10530-004-5573-8

Boats, pathways, and aquatic biological invasions: estimating dispersal potential with gravity models

Brian Leung[1,2,*], Jonathan M. Bossenbroek[1] & David M. Lodge[1]
[1]*Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA;* [2]*Present Address: Department of Biology & McGill School of Environment, McGill University, Montreal, QC, Canada H3A 1B1; *Author for correspondence (e-mail: brian.leung2@mcgill.ca)*

Received 23 March 2004; accepted in revised form 28 October 2004

Key words: dispersal, exotic, invasive species, gravity model, non-indigenous, risk

Abstract

Biological invaders can have dramatic effects on the environment and the economy. To most effectively manage these invaders, we should consider entire pathways, because multiple species are dispersed through the same vectors. In this paper, we use production-constrained gravity models to describe movement of recreational boaters between lakes – potentially the most important pathway of overland dispersal for many aquatic organisms. These models are advantageous because they require relatively easily acquired data, hence are relatively easy to build. We compare linear and non-linear gravity models and show that, despite their simplicity, they are able to capture important characteristics of the recreational boater pathway. To assess our model, we compared observed data based on creel surveys and mailed surveys of recreation boaters to the model output. Specifically, we evaluate four metrics of pathway characteristics: boater traffic to individual lakes, distances traveled to reach these lakes, Great Lakes usage and movement from the Great Lakes to inland waters. These factors will influence the propagule pressure (hence the probability of establishment of invasive populations) and the rate of spread across a landscape. The Great Lakes are of particular importance because they are a major entry point of non-indigenous species from other continents, hence will act as the origin for further spread across states. The non-linear model had the best fit between model output and empirical observations with $r^2 = 0.80$, $r^2 = 0.35$, $r^2 = 0.57$, and $r^2 = 0.36$ for the four metrics, respectively. For the distances traveled to individual lakes, r^2 improved from 0.35 to 0.76 after removal of an outlier. Our results suggest that we were able to capture distances traveled to most but not all lakes. Thus, we demonstrate that production-constrained gravity models will be generally useful for modeling invasion pathways between non-contiguous locations.

Introduction

Invasions by non-indigenous species are a major driver of global environmental change. Invasive species have the potential to reduce global biodiversity, affect entire ecosystems (Sala et al. 2000; Lodge 2001) and impose high economic damages (Pimentel et al. 2000). Thus, it is critical to forecast where invasions will take place and to identify areas that act as major sources for further invasions to focus our management efforts and to reduce the probability of invasion or the rate of spread.

To identify areas to focus management, we need models of species dispersal – one of the major components of the invasion process. Arguably, we should focus on pathways. While analyses of specific single species are important and necessary, analyses of pathways allow prediction of the spread of multiple species and facilitate forecasting potential spread before the non-indigenous species has actually arrived. An ideal pathway model will have modest data requirements, capture the most important features of the pathway, and be applicable across a large heterogeneous geographical range.

In this paper, we use gravity models as our framework for estimating dispersal pathways. Gravity models, or spatial interaction models, have been used by geographers to predict human movement behavior (Thomas and Hugget 1980), and thus may be an ideal method for modeling patterns of human movement and the organisms that incidentally travel with them. Here we focus on the spread of zebra mussels via recreational boaters, but the same framework could be applied to non-indigenous species associated with botanical and aquaria suppliers or the live seafood trade. Recreational boating has been highlighted as perhaps the greatest pathway for aquatic non-indigenous species to be transported to inland waterways from the Great Lakes and other invaded lakes (Johnson et al. 2001). Unfortunately, there are too many lakes and too many boaters to directly measure movement patterns. For instance, in Michigan alone there are over 1500 inland lakes greater than 25 ha and over one million registered boaters. To collect data on even a few dozen lakes requires intensive sampling. Further, we are interested in not only which lakes boaters visit, but also how often they do so. The spread of non-indigenous species depends upon both the occurrence and the frequency of introductions into new areas (Kolar and Lodge 2001). Thus, models are needed that can describe the recreational boater pathway with modest data requirements.

Gravity models have been argued to provide advantages over other dispersal models such as diffusion models because gravity models supposedly capture the mechanism of spread (Bossenbroek et al. 2001). However, past research have only used gravity models to describe the distribution of invasive species such as zebra mussels (Schneider et al. 1998; Bossenbroek et al. 2001) and the spiny water flea (MacIsaac et al. in press), which are transported primarily by the recreational boater pathway (Johnson et al. 2001). The efficacy of gravity models as an estimate of the actual mechanism – the movement pattern of recreational boaters – has not been explicitly examined, which is evident when the gravity model of Bossenbroek et al. (2001) is examined more closely. The parameters developed in their model predicted that 75% of all boaters in Wisconsin travel to the Great Lakes (Bossenbroek, pers. comm.), which is in contrast to the 11% estimated by a Wisconsin boater survey (Penaloza 1991). Here, we use empirical data based on creel surveys and mail surveys of registered boaters to assess gravity models and evaluate their ability to capture relevant features of the recreational boater pathway.

To reduce the scope of analysis to a manageable level, we focus our analysis on lakes within the State of Michigan. Michigan is adjacent to four Great Lakes and Lake St. Clair. The Great Lakes are major destinations for shipping traffic, hence are an important entry point for aquatic non-indigenous species from other nations. Zebra mussels, the spiny water flea and the round goby (*Neogobius melanostomus*), to name a few, were all initially introduced into North America through the Great Lakes. Subsequent spread to inland lakes from the Great Lakes via recreational boaters can and has occurred. For instance, zebra mussels first entered Lake St. Clair in 1986 (Hebert et al. 1989) and have spread to more inland lakes in Michigan to a far greater extent than in any other state (Kraft and Johnson 2000).

We focus on one class of gravity models, termed production-constrained gravity models, which have modest data requirements in comparison with other gravity models such as doubly constrained gravity models (Thomas and Hugget 1980). Lake location (latitude and longitude), lake characteristics (e.g., lake size), and boater populations in a given county may be sufficient to capture the important characteristics of recreational boating pathways. Thus, a production-constrained gravity model is an efficient way to quantify traffic to thousands of lakes and identify the relative chance that boats carry non-indigenous species, which should be proportionally related to propagule pressure. The results from gravity models can be linked to other models, such as a habitat suitability model, or information on environmental tolerances for individual non-indigenous species to forecast where invasions will potentially take place.

In this paper, we compare a linear and non-linear version of a production-constrained gravity model and use four metrics to parameterize these models and examine their ability to capture important characteristics of the recreational boater pathway. The metrics used to examine our models were (1) distances traveled by boaters, (2) boat traffic to specific lakes, (3) proportion of boaters using a Great Lake, and (4) the proportion of boaters using both the Great Lakes and inland waters. Boater traffic is a primary determinant of propagule pressure. Distances traveled determines rate of spread. As mentioned, the Great Lakes are of particular interest due to their importance as entry points of non-indigenous species into North America. We compared the results of our models to observed data from creel surveys and mailed surveys of registered boaters.

Materials and methods

Production constrained gravity models

We use a production-constrained gravity model to predict the recreational boater pathway, in order to estimate the potential spread of non-indigenous species. Gravity models are based on the interaction between distance and the attraction of a destination. Empirical research has shown that most boaters do not travel very far (Buchan and Padilla 1999). Thus, we should expect an inverse relation between boater traffic and distance to reach a lake. Further, researchers have found that more boat users visit larger lakes (Reed-Andersen et al. 2000). Thus, the 'attractiveness' should relate to lake size. Finally, logically we should expect larger boater populations to positively relate to boat traffic, but that the number of visitors should be moderated by the number of potential destinations. Gravity models provide a formal structure that incorporates these features (Equation (1)).

For our model, the probability that a boater will travel to a given destination is based on the distance from the county residence to the lake and the attractiveness of that lake or Great Lake boat ramp. Specifically, the traffic (U_j) to a destination lake or Great Lakes boat ramp (j) depends upon the number of boats (O_i) at each source location (i), the attractiveness of the lake or boat ramp (W_j), the distance between the source and destination (D_{ij}), and the other available destinations that each boater could visit instead (A_i). The number of boats that travel from i to j was estimated via:

$$U_j = \sum_{i=1}^{K} A_i O_i W_j D_{ij}^{-d} \tag{1}$$

$$A_i = 1 / \sum_{j=1}^{L} W_j D_{ij}^{-d} \tag{2}$$

where U_j is a metric of boater movement from destination i to lake j, K is the number of counties, L is the number of lakes and Great Lakes boat ramps, and d is a shape parameter describing the relation between traffic and distance. A_i, weights the expected boater movement by the inverse of the number of other potential destinations.

Also, there was no reason to expect each boater to make only a single trip in a season, nor for their behavior to be independent of other boater behavior (e.g., late boaters may avoid areas that are in high usage). Thus, we wrote our estimate of boater traffic T_j as an unspecified function of the output from Equation (1).

$$T_j = f(U_j) \tag{3}$$

Because there was no reason to expect only a single trip per boater, nor a linear relation between U_j and T_j, we compared both a linear model

$$T_j = bU_j + a \tag{4}$$

and a power model

$$T_j = aU_j^b \tag{5}$$

These models provide two alternative options for representing the patterns of boater behavior in

Michigan. Comparison of empirical patterns to model output can provide evidence that supports either of these alternative models (Hilborn and Mangel 1997).

Empirical data

To build our model, we considered the types of information that would most easily be acquired, and would allow a large number of lakes to be modeled. The three basic components of a production-constrained gravity model are the number of individuals traveling from a particular origin (O_i), the attractiveness of a destination (W_j) and the distance between the origin and destination (D_{ij}). We defined O_i as the number of boaters registered in each county in Michigan and D_{ij} as the distance (m) between destinations and source counties (as measured from centroid of the lake to the center of lake polygons within each county; Bossenbroek et al. 2001), which permitted us to include visits from boaters in a county to lakes within the same county (i.e., modeled as distance from the centroid to the center of lake polygons). Comparison of model output to the observed data will indicate the strength of evidence that our model structure, assumptions, and levels of detail are reasonable.

In our model we had two types of destinations that were included: inland lakes and Great Lakes boat ramps. We defined the attractiveness of each inland lake as the area of the lake (m^2) because previous researchers have demonstrated a strong positive relationship between boat usage and lake size (Reed-Andersen et al. 2000). Other potential factors that could have been included in our measure of attractiveness were the number of boat ramps, the quality of fishing, etc, but these data are not easy to obtain on the regional scale of our model. To estimate the attractiveness of the Great Lakes, we estimated the attractiveness of individual boat ramps. We treated each boat ramp analogously to a separate lake, with attractiveness $W_j = g$. The value for g was fitted using maximum likelihood techniques. We chose this approach because of the large size of the Great Lakes, and because arguably the location of boat ramps was a more relevant measure of distance than the lake centroid. For simplicity, we treated all boat ramps as having homogeneous attractiveness values. However, in reality, it is possible that different ramps could differ in attractiveness. In our model we included the numbers of registered boaters in each county, lake sizes and coordinates of 1589 Michigan inland lakes greater than 25 ha and coordinates of 109 Great Lakes boat ramps (Bossenbroek 1999).

Beyond the basic data included in our models there were two parameters that we estimated by comparing model outputs to empirical metrics, based on creel and mail surveys. The parameters that needed to be estimated were d (distance coefficient) and g (attractiveness of Great Lakes boat ramps, discussed above).

The observed data used to identify the best-fit parameters included creel surveys that were conducted by the Michigan Department of Natural Resources (Lockwood 2000) and mailed surveys, which we conducted. Boat traffic was calculated based on the number of angler trips reported in the creel surveys (between May and September). On an average, there were 2.5 anglers per boat (Roger Lockwood, MI Department of Natural Resources, pers comm.). Distances traveled to reach a given lake were estimated using the numbers of visitors from each county within Michigan. Although there were a few visitors from other states, we did not include these in our calculations, as our gravity model was restricted only to Michigan. We obtained 19 estimates of lake traffic from 14 lakes, across 8 counties (Appendix A). These were the only lakes for which traffic data existed for the summer months, which we could definitively identify, based on its name and county, and which we could merge with our larger database on lake locations and characteristics, presented in Bossenbroek (1999). For distances traveled, we obtained 10 estimates from 8 lakes. Multiple measures of lakes across years allowed inclusion of some aspect of temporal variability. This was not a problem statistically, as we were not interested in the traditional P-values.

For the mailed surveys, 30 counties were chosen. We basically chose every third county to obtain a representative geographical range within Michigan. For each county selected, (Appendix A) we randomly selected 100 registered boaters to receive our survey. Each boater was asked six questions, of which two were used to test our models. The specific questions asked were: (1) During last year (2002), how many times was your boat used on Lake Michigan, Lake Huron, Lake Erie, Lake Superior, or Lake St. Clair? and (2) During last year (2002), how many times was your boat used on an inland lake (i.e., NOT a Great Lake or Lake St. Clair)? See Appendix B for the full survey instrument. We obtained 730 usable responses. We excluded boaters who indicated that they did not move their boats during the season, resulting in 498 non-resident boaters (Question 1). Only non-resident boaters would be important for the transport of non-indigenous species. Finally, we excluded counties that had too small number of respondents (set at <10), to avoid the poorest estimates of boater behavior for a given county. This resulted in a final sample size of 475 boaters across 27 counties.

We believe that our survey results are reasonably accurate. However, we note four potential sources of bias. (1) Misunderstanding the question, (2) biases if people are dishonest or if the people who respond are not representative of the boater population at large, (3) sampling error, (4) list error where our sampled population does not represent the actual population. We worked with the social science centre at the University of Notre Dame to construct the questions to minimize misunderstanding. Further, we kept the questions as simple as possible, as well as leaving space for comments (see Appendix B). One question did cause problems (question 6) in that some individuals interpreted the question as how far did they travel on a lake rather than to reach a lake. Therefore, this question was invalidated and omitted from analysis. We felt that our questions were fairly innocuous and there was little reason to lie (e.g., it did not evaluate knowledge, there were no benefits associated with any answer, and we were not evaluating potentially illegal activity – such as accidentally marketing nuisance species). It is possible that people who responded were not representative of the general population (e.g., if Great Lakes users responded more than inland lake users). We can think of no reason why this would occur and the only solution to this is to have near 100% response rates, which would require other surveying techniques that may cost hundreds of thousands of dollars. Sampling error is always a potential problem. However, this would result in an unbiased estimate, so the consequence should be to simply weaken the relations observed, and should not invalidate conclusions of the model. List error could be problematic if there are many people who boat frequently but are not registered.

Using these data, we used four metrics to analyze the viability of our gravity model. First, boat traffic estimates generated from the model (U_j, Equation (1)) were compared to empirical measures of actual boat traffic to a given lake (T_j^o) from creel surveys (Lockwood 2000). To maintain tractability, we made a simplifying assumption that while the total number of boats could potentially be affected by non-linearity, the composition of sources was not (e.g., the proportion of visitors from county A versus county B did not change). Thus, the boats visiting a given lake or boat ramp (j) from county (i) was (T_{ij}):

$$T_{ij} = T_j \frac{U_{ij}}{U_j} \tag{6}$$

Second, average distances traveled to lakes ($\overline{D}_j^o$) as determined from the creel surveys (Lockwood 2000) were compared to distances generated from the model ($\overline{D}_j$). We needed to estimate the proportion of visit to a given lake from each county. The average distance $\overline{D}_j$ traveled to each lake was estimated using the distances between a county i and destination lake j multiplied by the proportion of boaters from county i visiting lake j.

$$\overline{D}_j = \sum_{i=1}^{k} D_{ij} \frac{U_{ij}}{U_j} \tag{7}$$

Third, the proportion of boat trips to the Great Lakes (G_j^o) for county (i) (the complement of this number would be the proportion visiting inland lakes) was compared between model estimates and the observed proportions as determined from our mail survey.

The predicted proportion of boat trips to the Great Lakes (G_i) was estimated from the model by first calculating the number of boats visiting each lake j that came from county i (Equation

(6)) and then comparing the number of boaters visiting Great Lakes boat ramps to the total number of boat trips.

$$R_i = \sum_{m=1}^{M} T_{im} \quad (8)$$

$$S_i = \sum_{i=1}^{L} T_{il} \quad (9)$$

$$G_i = \frac{R_i}{R_i + S_i} \quad (10)$$

where R_i summed the predicted number of boats traveling from county i to boat ramp m on Lake Michigan and S_i summed the predicted number of boats traveling to inland lake l, M was the number of Great Lakes boat ramps, and L was the number of inland lakes.

Finally, the proportion of trip pairs that visited first the Great Lakes and then an inland lake was compared. This is of particular importance since new invasions will likely begin in the Great Lakes and then spread to inland lakes via recreational boaters (Bossenbroek et al. 2001; Johnson et al. 2001). We define trip pairs as the sequence of trips. Arguably, boaters would not necessarily travel from one lake directly to another, but instead could go home first. As such, the production-constrained gravity model was appropriate, as it modeled movement from the source location to first the Great Lakes, then from the source location to destination lakes.

The model yielded a rough composite metric across all individuals within a county during the entire boating season, whereas the empirical data had information on individual boater movements, which was highly heterogeneous, both in terms of their number of trips and their proportion of trips to the Great Lakes versus inland lakes. In order to overcome these differences in measurements, we converted heterogeneous individual data into a composite measure for a county.

Specifically, for each individual surveyed we calculated the probability that they had first gone to a Great Lake followed by a trip to an inland lake. In order to do this, we calculated the number of pairs of trips as $V^o_{i,k} - 1$, where $V^o_{i,k}$ was the observed number of trips for individual k from county i. For example, if an individual took three trips, there would be two trip pairs (1,2 and 2,3). We also needed to consider the order in which lakes were visited. The number of possible combinations of trips would be $V^o_{i,k}!$. The total number of possible trip pairs across all combinations was

$$Z_{i,k} = (V^o_{i,k} - 1)V^o_{i,k}! \quad (11)$$

We made the assumption that all possible combinations were equally likely. For a given proportion of visit to the Great Lakes ($G^o_{i,k}$), the number of trip pairs where the first trip was to a Great lake, would be $G^o_{I,k}Z_{i,k}$ We then needed to consider what proportion of the second trip in a pair would be to an inland lake. This would be given by $\frac{(1-G^o_{i,k})V^o_{i,k}}{V^o_{i,k}-1}$. The numerator in this equation gives the number of visits to inland lakes; the denominator corrects for the single first trip in the pair (i.e., so we do not count trips multiple times). Thus, after algebraic manipulations, the expected proportion of total trip pairs that go first to the Great Lakes and then to an inland lake would be given by the proportion of visit to the Great Lakes multiplied by the proportion of trips to inland lakes

$$B^o_{i,k} = \frac{G^o_{i,k}(1 - G^o_{i,k})V^o_{i,k}}{V^o_{i,k} - 1} \quad (12)$$

As the number of trips becomes large, $B^o_{i,k}$ approaches $G^o_{i,k}(1 - G^o_{i,k})$. Summing across all individuals and all trip pairs,

$$P^o_i = \frac{\sum_{k=1}^{N} B^o_{i,k}(V^o_{i,k} - 1)}{\sum_{k=1}^{N} (V^o_{i,k} - 1)} \quad (13)$$

P^o_I was the empirical estimate of the total proportion of trips pairs that first traveled to a Great Lake and then to an inland lake across all N individuals for county i.

To estimate the proportion of trip pairs from the Great Lakes to inland lakes in the model

(P_i), we used the proportion of individuals going to the Great Lakes (G_i), and of those the proportion going to inland lakes next ($1-G_i$).

$$P_i = G_i(1 - G_i) \tag{14}$$

The model estimate was a simplification of reality and represented a maximum value; P_i^o will be consistently smaller than P_i due to heterogeneity between individuals.

Fitting models to empirical metrics

We used least squares to measure the goodness of fit between model predictions and the observed data. Because the metrics were in different scales (e.g., distance traveled and proportion of trips to Great Lakes), we first standardized the least squares by the variation in the observed variable to make contributions from the metrics comparable and to integrate this information into a single metric. Generally,

$$LS = \sum_i \frac{(X_i^o - X_i)^2}{(X_i^o - \overline{X}^o)^2} \tag{15}$$

where X_i^o is a generic observed value, X_i is a generic prediction from the model, and $\overline{X}^o$ is the average observed value.

For average distances traveled to lakes (metric 2) and proportion of boat trips to the Great Lakes (metric 3), we could use Equation (15) directly, as we expected a one to one relation between prediction and observation. For average distances traveled, X_i^o was replaced with the observed average distance $\overline{D}_j^o$ from the creel surveys, and X_i was replaced with the predicted distance $\overline{D}_j$ (Equation (7)). For the proportion of trips to the Great Lakes, we again compared the observed proportion from creel surveys G_i^o to the model prediction G_i (Equation (10)).

In contrast, we did not necessarily expect a linear relation for number of boaters visiting a lake (T *versus* U in Equations (1)–(3)). Thus, we examined both a linear relation ($T_j^o = bU_j + a + \varepsilon$) and a log-linear relation ($\ln(T_j^o) = \ln(a) + b\ln(U_j) + \varepsilon$). Similarly, for the comparison of trip pairs from the Great Lakes to inland lakes, we expected the value generated from the model to represent a maximum rather than a one to one relation with observed traffic (i.e., $P_i^o = bP_i + a + \varepsilon$). Thus, Equation (15) was modified,

$$\mathrm{LS} = \sum_i \frac{(X_i^o - (bX_i + a))^2}{(X_i^o - \overline{X}^o)^2} \tag{16}$$

Coefficients a and b were determined from standard regression techniques and did not need to be determined using our non-linear search. Note that $1 - \mathrm{LS}$ from Equation (16) is the coefficient of determination from linear regression, i.e., the amount of variation explained. Where we examine only a one to one relation between prediction and observation, $1 - \mathrm{LS}$ from Equation (15) is also roughly comparable to the amount of variation explained, except that the model variation may be greater than the observed variation (i.e., LS can be greater than 1; our metric is conservative). Thus, we used $1 - \mathrm{LS}$ as a rough measure of the amount of variation explained. We also presented the slope and intercepts for each metric, to examine deviations from a one to one response.

Our integrated measure of fit was simply the sum of the least squares across all metrics, $\mathrm{LS_{tot}} = \mathrm{LS_T} + \mathrm{LS_D} + \mathrm{LS_G} + \mathrm{LS_P}$. Subscripts T,D, G, and P refer to our four metrics of fit, boater traffic to lakes, average distances traveled, proportion of Great Lakes visits, and proportion of trip pairs first to the Great Lakes and then to inland lakes. We used a simplex algorithm that minimized $\mathrm{LS_{tot}}$ to determine the best-fit values of d (distance coefficient) and g (attractiveness of Great Lakes boat ramps) for our production-constrained gravity model. Our integrated measure of fit was necessary primarily for programming purposes, and represented the simplest metric that the simplex algorithm could optimize. For the purposes of interpretation, we also analyzed the metrics separately. Our approach is analogous to traditional ecological studies (e.g., using regressions to examine the evidence for a predicted *a priori* relation), except that we use a mechanistic mathematical model as our *a priori* hypothesis – that the components and structure of the gravity model can capture the characteristics of interest.

Results

The gravity model built using the power relation (non-linear model) between T_j and U_j (Equation (5)) resulted in a considerably better fit than the linear model (Equation (4)) across all metrics. LS_{tot} was 1.98 for the non-linear model and 2.83 for the linear model; Roughly speaking, the non-linear model performed 1.43 times better than the linear model. We also examined a rough measure of percent variation explained (1 − LS) as a more intuitive comparison between models. For boater traffic to a lake, the non-linear model explained 80% of the variation compared to only 35% explained using the linear model. Similarly, the non-linear model explained 56% of the variation in the proportion of trips to the Great Lakes and 36% of the variation in the proportion of trip pairs first to the Great Lakes and then to inland lakes, whereas the linear model only explained 39% and 15% of the variation, respectively. Both models explained similar amounts of variation for distances traveled (29% for the non-linear model and 28% for the linear model). Thus, for the rest of the results and discussion we focus on the non-linear model.

The best fitting set of parameters occurred when $d = 2.08$ (Equation (1)), $g = 1.72 \times 10^8$ and $a = 158.5$, $b = 0.54$ (Equation (5)) (Figure 1). Our model predicted an average distance moved of 76.2 km, and the percent of visits to the Great Lakes at 29.2% over all counties. For the 27 counties sampled in the mail survey, the Great Lakes visits estimated by the model was similar (25.4%), and was comparable with the observed Great Lakes visits in the mail survey of 22.1%. Our model explained a high proportion of the variance in trips to the Great Lakes (slope = 0.89, intercept = 0.03, $r^2 = 0.57$, Figure 1c). The fit to distances traveled overestimated the actual distance traveled

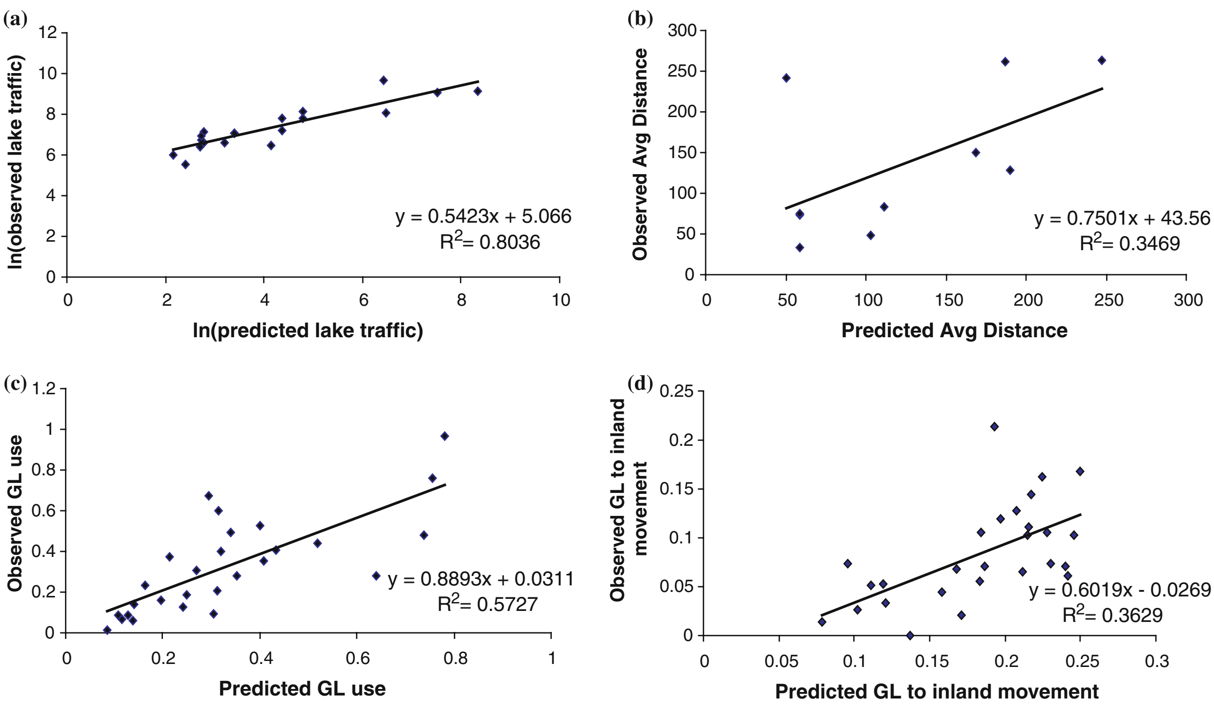

Figure 1. Comparisons of the predictions from the non-linear production constrained gravity model to observed data for four metrics, using the best fitting parameter values $d = 2.08$, $g = 1.72 \times 10^8$ m^2. (a) Number of boaters visiting specific inland lakes, (b) Average distances traveled to specific inland lakes (distance is in meters), (c) Proportion of trips to the Great Lakes, (d) Proportion of trip pairs first to the Great Lakes and then to an inland lake. Regression lines, equations, and coefficient of determinations (r^2) are also shown.

(slope = 0.75) except at short distances (intercept = 43 km) ($r^2 = 0.35$) (Figure 1b). However, this was primarily due to a single data point, with low predicted but high observed distance traveled. Removal of this outlier resulted in a stronger relationship, closer to the theoretical one to one line (slope = 1.08, intercept = −18.7, $r^2 = 0.76$). Thus, the model captured most, but not all, of the distances traveled to individual lakes. The outlying data point was Brevort Lake (lat: 45.995, long: −84.926, area: 1.73×10^7 m^2). It is possible that the discrepancy was due to the popularity of Brevort lake as a cottage resort, costing as much as $2500 USD/wk for a cottage rental during high season (www.greatrentals.com/MI/5473.html).

As mentioned, boat traffic to lakes and trip pairs to Great Lakes first and then to inland lakes were not expected to be a one to one relation between observed and model output. The power relation (and log-transformed fit) yielded a very strong fit ($r^2 = 0.80$, Figure 1a) indicating that this was a reasonable functional form. As expected, trip pairs to Great Lakes first and then to inland lakes estimated from the model was a maximum. The realized relation needed to be scaled by slope = 0.60 and intercept 0.03 ($r^2 = 0.36$; Figure 1d).

We examined the sensitivity of the results using the best fitting parameter set for the non-linear model (*d*-distance coefficient and *g*-attractiveness of the Great Lakes boat ramps), holding one parameter constant at the best fit value and varying the other. For each parameter set, *a* and *b* were fit separately using linear regression (Equation (5)). The number of boaters to a given lake (diamonds, Figures 2a, b) was much less sensitive than the other metrics. The fit to distances traveled was sensitive to variation in the distance parameter *d* (squares, Figure 2a); it was also sensitive to low values of *g* but was asymptotic and became less sensitive as *g* increased (squares, Figure 2b). Both

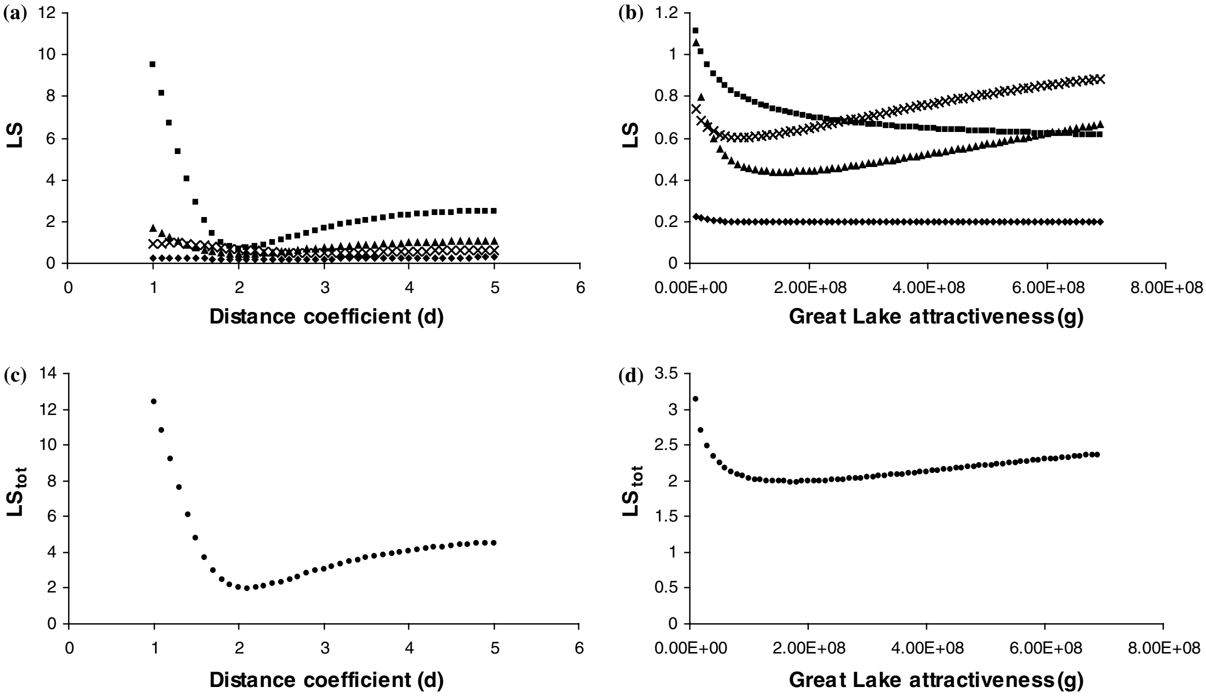

Figure 2. Sensitivity analysis of the best fit parameters; *d* (distance coefficient; panels a and c) and *g* (Great Lakes attractiveness; b and d). Model predictions were compared to observed data for four metrics (a and b) and the composite LS value (c and d) as the two parameters were varied independently. Diamonds = number of boaters visiting specific inland lakes. Squares = average distances traveled to specific inland lakes. Triangles = proportion of trips to the Great Lakes. Crosses = proportion of trip pairs first to the Great Lakes and then to an inland lakes.

the proportion of trips to the Great Lakes (triangles) and the proportion of trip pairs to the Great Lakes first and then to an inland lake (crosses) were sensitive to *g* (Figure 2b), but only marginally affected by *d* (Figure 2a). Thus, considering all empirical metrics of fit, we could be confident about the parameter values used in the non-linear model, as only a narrow range of values captured all features simultaneously.

Discussion

One of the primary goals of invasion biologists is to identify areas most at risk of invasion. Risk will depend on invasive species population dynamics, the characteristics of the receiving environment, and the pathways by which species reach new areas. We focused on pathways because they transcend individual species; in some cases, many species may be transported via the same pathway (Carlton and Geller 1993). While analysis of single species remains important, targeting prevention efforts on entire pathways enables management of multiple species simultaneously. Analysis of pathways will yield a more comprehensive estimate of the costs and benefits of prevention compared to analyses of single species. Additionally, analysis of pathways allows us to identify areas most at risk of invasion before invaders have actually arrived.

Consideration of the pathways of non-indigenous species dispersal is important both in terms of international invasions across continents as well invasions to new areas within continents. For instance, efforts to manage international pathways of invasions include regulations on ballast water (Nonindigenous Aquatic Nuisance Prevention and Control Act of 1990), the importation of fruits and vegetables, and plants (Plant Quarantine Act of 1912). Within continents, examples of pathway management include the 100th Meridian initiative to stop the westward spread of aquatic nuisance species (http://100thmeridian.org/) and the electric barrier in the Chicago shipping and sanitation canal. Analyses such as those presented here will help inform policy for prevention of invasions within continents; it will determine important sources of aquatic invaders as well as identify areas at risk of invasion so that our resources may be allocated most effectively. Gravity models can target regulation of recreational boater movement as well as education efforts.

Gravity models, if they adequately capture pathway characteristics, offer logistical advantages over other gravity models, such as doubly constrained gravity models, as discussed previously (Bossenbroek et al. 2001) and conceptual advantages over other dispersal models, such as diffusion models (MacIsaac et al. 2002). Basically, doubly constrained gravity models require a substantial amount of information and are therefore limited to analyses of a relatively small number of lakes (e.g., Schneider et al. 1998). In contrast, production-constrained gravity models can require as little information as lake size, location and numbers of registered boaters, and therefore can be applied to large systems.

Production-constrained gravity models, and gravity models in general, offer advantages over dispersal models such as diffusion models, in that gravity models reflect the actual movement process (Bossenbroek et al. 2002; MacIsaac et al. 2001). Although diffusion models have been useful, they are based on the assumption that movements are random and a result of 'atoms' bouncing off one another thus causing the individuals to spread. We believe that the disconnect between predictability of diffusion models and observed patterns of invasions such as zebra mussel infested lakes (Buchan and Padilla 1999) may be directly related to the fact that diffusion models do not model the actual movement process. Further, diffusion models are species-specific and do not consider pathways and vectors. Thus, gravity models have a foundational benefit over diffusion models, if we can demonstrate that they suitably describe the actual mechanism of movement, namely boater behavior.

Previous analyses have assumed a mechanistic link between gravity models and the boater pathway and have based parameterization purely on the fit to the observed patterns of invasions of individual species such as zebra mussels (Bossenbroek et al. 2001). Additionally, previous work has assumed a linear functional relation between model output (Equation (1)) and actual boater traffic. Here, we extended this work and parameterized our gravity model using real data on boater movement, examined functional forms, and explicitly demonstrated that gravity models were sufficient to capture the important characteristics of the recreational boater pathway. We found that a log-linear model yielded a superior fit compared to a linear functional relation between model output and actual boater traffic. The basis for this non-linear relation may be due to changes in boater behavior. Boaters may prefer larger closer lakes, but this preference is reduced when there is already high boat traffic on a lake. Thus, boaters are more evenly distributed across lakes than expected based on size and distance alone. Using the log-linear model, we could determine boater movement to individual lakes, distances traveled to reach those lakes, Great Lakes usage, and movement between locations. Thus, we can have some confidence in using the parameters and functional forms in this model and applying them to other species dispersed via the same pathway.

More specifically, our results differed dramatically from previous work (Bossenbroek et al. 2001). The previous model suggested that 82% of boaters visit the Great Lakes in Michigan (Bossenbroek, pers. comm.). Further, the attractiveness of individual boat ramps was estimated at 5.5×10^8 m^2, or roughly the size of Lake Winnebago, Wisconsin. However, based on our mail survey, 22% of Michigan boaters visit the Great Lakes (as estimated from 27 counties). More realistically, for counties sampled, our model estimated the proportion of boaters visiting the Great Lakes at 25%, and 29% for all counties in Michigan. Further, the attractiveness of the Great Lakes was considerably smaller, at 1.72×10^8 m^2 or roughly 1/3 the attractiveness.

Our results also differed from a previous survey of Wisconsin boaters, which suggested that the visitation to the Great Lakes from Wisconsin was 11% (compared to 75% Wisconsin boaters estimated by Bossenbroek's model), and that the more than 90% of boaters moved less than 50 km (Buchan and Padilla 1999). As mentioned, based on the output of our model, 29% visited the Great Lakes from Michigan. Individuals moved on an average of 76 km. This difference in boater behavior may explain the faster rate of invasions in Michigan compared to Wisconsin (e.g., zebra mussels, Kraft and Johnson 2000). However, this hypothesis represents an *a posteriori* hypothesis and would need to be explicitly tested using comparable surveys.

This work represents an important step in forecasting invasions. We used the simplest measures in our gravity models – lake size as a surrogate for attractiveness and linear distances. Explanatory power could potentially be improved by consideration of factors such as fishing opportunities, road access, and actual road distance (versus Euclidean distance). Further, such analyses should be extended to other potential pathways. Waterfowl, for instance, have been implicated in the spread of some aquatic non-indigenous species (Viviansmith and Stiles 1994; Figuerola et al. 2003). Additionally, we assumed homogeneous attractiveness for Great Lakes boat ramps. It might be possible to use some metric (such as parking spaces) to infer relative attractiveness values. Finally, while distance was treated as a primary mechanism for vector movement, for actual invasive species, distance might also relate to time out of the water which may affect survival and further influence the propagule pressure on uninfested lakes. Ultimately, we want to integrate a multitude of tools to predict invasions. We want to use the results from analyses of pathways in conjunction with species characteristics (Kolar and Lodge 2002) and environmental conditions (Ramcharan et al. 1992) to best identify invasion potential of different areas for entire suites of species that we are concerned about.

Acknowledgements

This material is based on work supported by the Integrated Systems for Invasive Species (ISIS) project (www.math.ualberta.ca/~mathbio/ISIS), funded by the National Science Foundation (DEB 02-13698) , the NOAA Illinois-Indiana Sea Grant College Program, and the US Fish and Wildlife Service.

Appendix A. Lakes used in creel survey and counties used in mail survey, used in analysis. Asterisk indicates lakes for which counties of origin for boat visitors were available, and therefore distances traveled could be calculated. Numbers in parentheses (#) indicates lakes for which boater traffic were available for multiple years. Numbers in brackets [#] indicates lakes for which distance traveled were available for multiple years. Numbers in parentheses or brackets indicate the number of years available.

Lake name	Longitude	Latitude	County	Lake size (m^2)
Lakes sampled in creel survey				
Beaver Lake *	−86.34	46.57	Alger	3,217,939.750
Bond Falls *	−89.10	46.39	Ontonagon	8,414,733.000
Brevort Lake *	−84.93	46.00	Mackinac	17,365,928.000
Chicagon Lake (2) *[3]	−88.50	46.06	Iron	4,284,049.000
Duck Lake (2) *	−89.22	46.20	Gogebic	2,467,226.250
Fletcher Pond	−83.85	44.98	Alpena	27,667,034.000
Grand Sable *	−86.04	46.63	Alger	2,662,219.250
Hagerman Lake (2) *	−88.78	46.06	Iron	2,352,329.500
Lake Gogebic	−89.58	46.49	Gogebic	52,738,424.000
Lake Pomeroy (2) *	−89.57	46.27	Gogebic	1,258,802.625
Marion Lake	−89.09	46.26	Gogebic	1,198,077.375
Mullett Lake	−84.53	45.50	Cheboygan	67,954,528.000
Silver Lake (2)	−86.50	43.67	Oceana	2,686,787.500
Tamarak Lake	−88.99	46.24	Gogebic	1,352,049.375

Counties sampled in mail survey	
Alger	Mecosta
Alpena	Menominee
Antrim	Montcalm
Baraga	Muskegon
Chippewa	Oakland
Eaton	Ogermaw
Emmet	Ontanagon
Gladwin	Oscoda
Grand traverse	Presque isle
Huron	Shiawassee
Iron	St. Joseph
Jackson	Tuscola
Kent	Van buren
Lake	Wayne
Lenawee	Wexford

Appendix B

*Note counties were noted on form before sending surveys

Michigan Boater Survey

Please circle the letter that best corresponds to your boating behavior. Please return this questionnaire in the enclosed self-addressed postage-paid envelope.

Q1: On average, how long is your boat usually kept in the water before it is removed and transported elsewhere?

a. Less than 1 week
b. 1 week to 1 month
c. More than 1 month but less than one season
d. The entire season

Q2: During last year (2002), how many times was your boat used on Lake Michigan, Lake Huron, Lake Erie, Lake Superior, or Lake St. Clair?

a. 0
b. 1
c. 2–5
e. 11–20
f. 21–30
g. 31–40

d. 6–10
h. Other: please specify___________________________

Q3: During last year (2002), how many times was your boat used on an inland lake (i.e., NOT a Great Lake or Lake St. Clair)?

a. 0	e. 11–20
b. 1	f. 21–30
c. 2–5	g. 31–40
d. 6–10	h. Other: Please Specify___________________________

Q4: During last year (2002), how many times was your boat used on the main stem of the Ohio River?

a. 0	e. 11–20
b. 1	f. 21–30
c. 2–5	g. 31–40
d. 6–10	h. Other: please specify___________________________

Q5: During last year (2002), how many times was your boat used on another river (i.e., NOT the Ohio River)?

a. 0	e. 11–20
b. 1	f. 21–30
c. 2–5	g. 31–40
d. 6–10	h. Other: please specify___________________________

Q6: Approximately, how far did you travel on your last boating trip?

a. Less than 10 miles	f. 31-40 miles
b. 11–20 miles	g. 41–50 miles
c. 21–30 miles	h. Other: please specify___________________________

References

Bossenbroek JM (1999) Modeling the dispersal of *Dreissena polymorpha* (zebra mussel) into the inland lakes of the western Great Lakes region using a gravity model. University of Wisconsin – Green Bay, Green Bay, WI

Bossenbroek JM, Kraft CE and Nekola JC (2001) Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. Ecological Applications 11: 1778–1788

Buchan LAJ and Padilla DK (1999) Estimating the probability of long-distance overland dispersal of invading aquatic species. Ecological Applications 9: 254–265

Carlton JT and Geller JB (1993) Ecological roulette – the global transport of nonindigenous marine organisms. Science 261: 78–82

Figuerola J, Green AJ and Santamaria L (2003) Passive internal transport of aquatic organisms by waterfowl in Donana, south-west Spain. Global Ecology and Biogeography 12: 427–436

Hebert PDN, Muncaster BW and Mackie GL (1989) Ecological and genetic-studies on *Dreissena polymorpha* (Pallas) – a new mollusk in the Great-Lakes. Canadian Journal of Fisheries and Aquatic Sciences 46: 1587–1591

Hilborn R and Mangel M (1997) The Ecological Detective: Confronting Models With Data. Princeton University Press, Princeton, NJ

Johnson LE, Ricciardi A and Carlton JT (2001) Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. Ecological Applications 11: 1789–1799

Kolar CS and Lodge DM (2001) Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution 16: 199–204

Kraft CE and Johnson LE (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

Lockwood RN (2000) Sportfishing angler surveys on Michigan inland waters: Fisheries Technical Report 2000-3. Michigan Department of Natural Resources, Lansing, MI

Lodge DM (2001) Lakes. In: Chapin FS III, Sala OE and Huber-Sannwald E (eds) Global Biodiversity in a Changing Environment: Scenarios for the 21st Century, pp. 277–313, Springer-Verlag, New York

MacIsaac HJ, Borbely J, Muirhead J and Graniero P (in press) Backcasting and forecasting biological invasions of inland lakes. Ecological Applications

MacIsaac HJ, Robbins TC and Lewis MA (2002) 'Biological invasions of aquatic habitats in Europe and the Great Lakes'. Modeling ships' ballast water as invasion threats to the Great Lakes. Canadian Journal of Fisheries and Aquatic Sciences 59: 1245–1256

Penaloza LJ (1991) Boating pressure on Wisconsin's lakes and rivers: Results of the 1989–1990 Wisconsin recreational boating study, Phase 1. Wisconsin Department of Natural Resources, Madison, WI

Pimentel D, Lach L, Zuniga R and Morrison D (2000) Environmental and economic costs of nonindigenous species in the United States. Bioscience 50: 53–65

Ramcharan CW, Padilla DK and Dodson SI (1992) Models to Predict Potential Occurrence and Density of the Zebra

Mussel, Dreissena-Polymorpha. Canadian Journal of Fisheries and Aquatic Sciences 49: 2611–2620

Reed-Andersen T, Bennett EM, Jorgensen BS, Lauster G, Lewis DB, Nowacek D, Riera JL, Sanderson BL and Stedman R (2000) Distribution of recreational boating across lakes: do landscape variables affect recreational use? Freshwater Biology 43: 439–448

Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, Poff NL, Sykes MT, Walker BH, Walker M and Wall DH (2000) Biodiversity – global biodiversity scenarios for the year 2100. Science 287: 1770–1774

Schneider DW, Ellis CD and Cummings KS. (1998) A transportation model assessment of the risk to native mussel communities from zebra mussel spread. Conservation Biology 12: 788–800

Thomas RW and Hugget RJ. (1980) Modeling in Geography. Barnes Owa, New Jersey.

Viviansmith G and Stiles EW (1994) Dispersal of salt-marsh seeds on the feet and feathers of waterfowl. Wetlands 14: 316–319