

Predicting invasions: alternative models of human-mediated dispersal and interactions between dispersal network structure and Allee effects

Corey Chivers^{1*} and Brian Leung^{1,2}

¹Department of Biology, McGill University, Montreal, Quebec, Canada; and ²School of Environment, McGill University, Montreal, Quebec, Canada

Summary

1. Human-mediated dispersal has been shown to be the most important vector for the spread of invasive species, yet there has been little evaluation of alternative models of dispersal in terms of differences in their predictions of invasion patterns. Moreover, no analyses have been attempted to elucidate the potential interaction between alternative models of human-mediated dispersal and population dynamical characteristics, such as Allee effects, which are central to the probability of an invasion.

2. Two prominent models in the literature which have previously been employed to predict human movement patterns are explored: (i) gravity models, which use the attractiveness of and distance to a location to predict travel patterns, and (ii) random utility models, which assume that individuals decide where to travel by maximizing the benefits that they receive according to some partially observable function of individual and site characteristics.

3. While distinction is often drawn between them in the literature, we demonstrate that these two approaches can be reduced to alternative functional forms describing the trip-taking decisions of individuals.

4. Each model was empirically parameterized using a survey of recreational boaters in Ontario, Canada. Within each model, both boater- and site-specific characteristics were important and the functional form provided by the gravity model was significantly better at capturing the behaviour of recreational boaters.

5. *Synthesis and applications.* The dispersal and establishment of species into novel habitats are central components of the invasion process and of quantitative risk assessments. However, predictions are dependent on the estimated spatial structure of the dispersal network and its potential interactions with species characteristics. This study demonstrates that Allee effects can interact with dispersal network structure to significantly alter predicted spread rates and that the consequences of these interactions manifest differently at the system and site levels. This modelling framework can be used to inform management interventions aimed at modifying human-mediated dispersal to reduce the spread of invasive species.

Key-words: Allee effects, dispersal, exotic, gravity model, invasions, networks, non-indigenous, random utility model, risk, spread

Introduction

Invasive species can cause ecological (Parker *et al.* 1999; Pejchar & Mooney 2009) and economic impacts (Aukema *et al.* 2011). To prevent or limit the spread of potentially

harmful species, management efforts must be informed by reliable estimates of where and when we expect new invasions to occur. The overland dispersal (lake to lake spread) of aquatic invasive species has been shown to be driven primarily by inadvertent human-mediated transportation. Several species have been observed ‘hitchhiking’ on the hulls and in the ballasts of recreational water vessels, which are transported on trailers from lake to lake (Johnson, Ricciardi & Carlton 2001; Kraft *et al.*

*Correspondence author: Department of Biology, McGill University, 1205 ave Docteur Penfield, Montreal, Quebec, Canada H3A 1B1. E-mail: corey.chivers@mail.mcgill.ca

2002). A recent study by Gertzen & Leung (2011) comparing human-mediated and fluvial dispersal found that human-mediated dispersal of an invasive species accounted for almost all of the propagules contributing to establishment probability. Understanding where species are likely to spread via this key human-mediated pathway is therefore an important step toward implementing mitigating measures.

There have been two general classes of modelling frameworks developed in the literature to characterize the movement of individuals across a landscape of discrete sites. Gravity models (GM) have been used extensively to characterize human movement patterns and have been applied successfully in several studies to the spread of invasive species (Leung, Drake & Lodge 2004; Potapov *et al.* 2011; Muirhead & MacIsaac 2011). They work by an analogy to Newtonian gravity, where individuals are attracted to locations proportionally to their *mass* (which can be any set of measures of desirability of the site) and inversely to the distance between an individual's current location, and the site (Schneider, Ellis & Cummings 1998; Leung, Drake & Lodge 2004; Leung, Bossenbroek & Lodge 2006). An alternative specification, developed in the field of recreational demand econometrics, is the discrete choice random utility model (RUM) (Smith & Kaoru 1986; Smirnov & Egan 2010). In this framework, individuals choose a destination from a suite of alternatives by maximizing a utility function based on any set of desirable traits, only part of which is known to the analyst. This model has been used in several recreational demand studies (e.g. Smith & Kaoru 1986; Parsons 2000), but has only recently been applied to the study of spread of invasive species (Macpherson, Moore & Provencher 2008; Timar & Phaneuf 2009). We show that these models are quite similar and that they can be reduced to simply alternative functional forms to describe an individual's trip decisions.

While it is clear that human vectors are central to the invasion process, the ramifications of employing alternative models of this vector on predicting spread is less clear. Moreover, although it has not been previously examined, one might expect that the consequences of dispersal models may interact with, and be determined by, the specific population dynamics of invaders. In particular, stochasticity and Allee effects, which are both well-known population level factors affecting invasion dynamics (Clark *et al.* 2003; Drake *et al.* 2006).

In this study, we address the following three questions: (i) Do the alternative human vector modelling frameworks (gravity and random utility models) differ in their ability to capture actual human behaviour, and therefore characterize dispersal vectors of invasive species? (ii) How do these alternative models interact with the population dynamics of invaders? (iii) What are the implications of alternative dispersal model specifications on our predictions of invasion risk across space and time (i.e. spread)?

We analysed the predictive ability of competing models of human-mediated dispersal by surveying recreational boaters and examining the ability of each model to recapture the observed trip outcomes. We recognized that differences in model fit are most important if the alternative model formulations lead to human-mediated dispersal networks that yield quantitative differences in our predictions of the spread of invasive species. Given the potential for ecological and economic harm posed by invasive species, predictions of spread across a landscape, as well as invasion risk at specific sites, are vital components of informed management policies (Landis 2004). Thus, we conducted a series of simulation experiments to examine the potential implications of each human-mediated dispersal model for risk assessments, taking into account their interaction with the population dynamics of invaders. We describe how the entropy or evenness of the predicted connectivity distribution of the dispersal network can interact with population dynamics to hinder spread. Taken together, this work provides new insights into how models of human behaviour affect the predicted structure of discrete dispersal networks and how the structure of dispersal networks interact with population level processes to influence the spatial spread of invasive species.

Materials and methods

SURVEY

We conducted a survey of recreational boaters in Ontario, Canada. We mailed 5000 invitations to participate in the survey to individuals with registered recreational licences (boating/fishing) issued by the Ontario Ministry of Natural Resources. Individual names were selected using a spatially stratified random sampling scheme. Approximately 100 invitations were sent to randomly selected individuals in each of 47 major geographical regions of Ontario as defined by the first two digits of their postal code. We developed an online survey instrument using the design approach of Dillman (2000). We employed an interactive map through which participants could quickly and easily identify the lakes that they visited. The advantage of this approach was that we were able to precisely identify lakes that may have been ambiguous because of multiple naming conventions. In this way, we were able to collect more in depth information in a visually intuitive manner. While our survey instrument was only able to capture individuals with access to the internet, 81% of households in Ontario had access to the internet as of 2009 (Statistics Canada, (<http://www.statcan.gc.ca/tables-tableaux/sum-som/101/cst01/comm36g-eng.htm>)). We have no reason to believe that those without internet access would behave differently *vis a vis* boating behaviour than those with online access. We asked participants to catalogue all of the boating trips that they took and to indicate the primary location where they kept their boat during the 2010 boating season.

Our survey response rate was 11%, with 30% of respondents indicating that they had visited multiple lakes during 2011. Given that we are interested in the behaviour of boaters who transport their boat from lake to lake during the boating season, we retained only those trip outcomes made by multi-lake boaters.

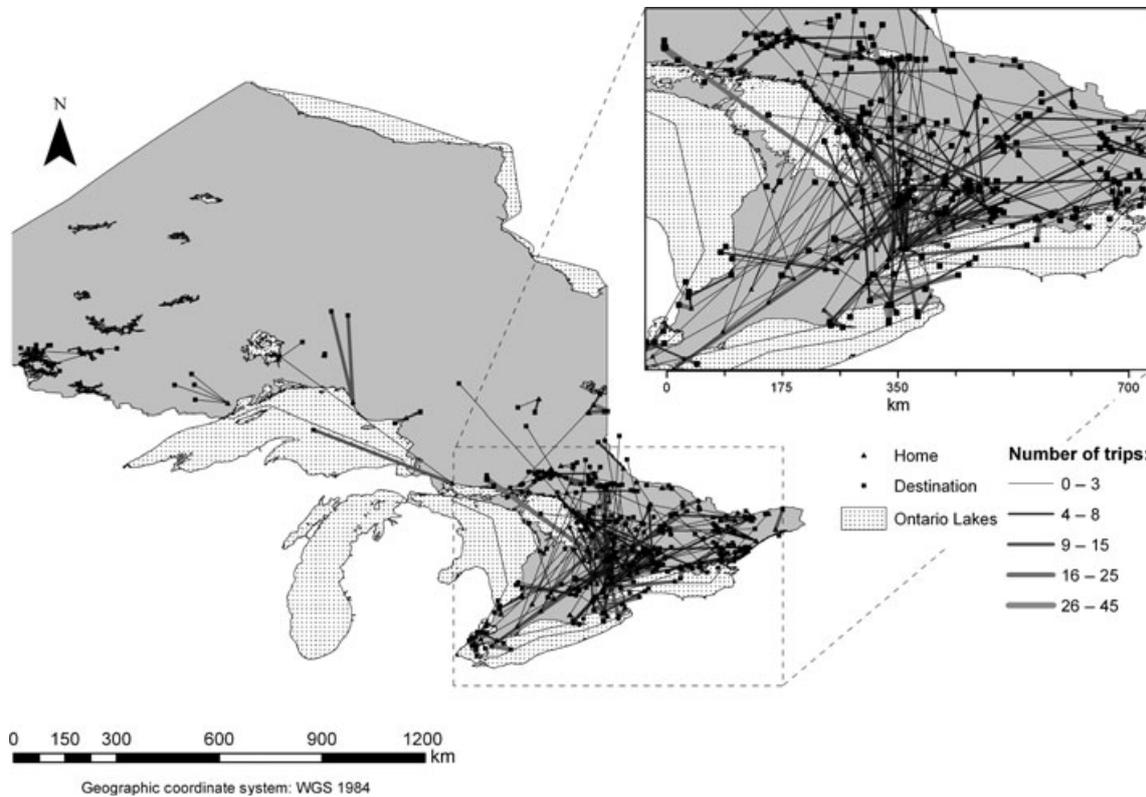


Fig. 1. Trips reported by survey respondents. The location where boaters stored their boat during the boating season is indicated with triangles, destinations are indicated by squares. The thickness of the line between home location and the destination lake is proportional to the number of trips taken. The right panel shows a zoomed in section of Southern Ontario for better visualization.

This left us with relevant observed source/destination outcomes for 146 individual boaters across Ontario making a total of 2354 boating trips (Fig. 1).

GRAVITY MODEL SPECIFICATION

Gravity models employ an analogy to Newtonian gravity, where the ‘pull’ of a given site is proportional to some function of desirable lake characteristics [termed ‘attractiveness’, e.g., size, (Bossenbroek, Kraft & Nekola 2001; MacIsaac *et al.* 2004; Leung, Bossenbroek & Lodge 2006)] and inversely related to the distance between a source location and the site. A boater chooses a site to visit according to the degree to which they are ‘pulled’ to that site, relative to the degree to which they are pulled by all other possible sites. While there are many possible formulations of gravity models, recent comparisons have found that the production constrained gravity model provides the best estimate of human-mediated dispersal of aquatic invasives (Muirhead 2007; Muirhead & MacIsaac 2011). In the production constrained formulation, it is assumed that a boater travels from their home location (primary location where they keep their boat) to a destination lake and then returns to their home location before visiting another lake. Further, the production constrained gravity model has modest data requirements compared with its alternatives (Muirhead & MacIsaac 2011), making it an accessible choice for resource managers. Because we wish to compare models in terms of their ability to capture individual level behaviour, we present a disaggregated formulation of the production constrained gravity model, in which each individual makes trip desti-

nation decisions according to a probability distribution described by the model. The site selection probability distribution, $P(T_n)$ for an individual boater n is given as:

$$P^{GM}(T_{nj}) = A_n W_j^e D_{nj}^{-d}, n = 1, \dots, n, j = 1, \dots, J. \quad \text{eqn 1}$$

Where W_j is the attractiveness of lake j , and D_{nj} is the distance between lake j and the home location (where they keep their boat) of individual n . Some authors suggest the use of least cost road networks to calculate the effective distance between source and destination (Drake & Mandrak 2010); however, for simplicity, here we use the euclidian distance between boater home location and lake centroid. The free parameters d and e describe the shape of the relationship and are fitted to the data (see Fitting and model selection). A_n is the ‘pull’ of all lakes, given by:

$$A_n = 1 / \sum_{k=1}^J W_k^e D_{nk}^{-d}. \quad \text{eqn 2}$$

Such that the probability of a boater n visiting lake j is proportional to the gravitational ‘pull’ of that lake compared to that of all other lakes. As a simple proxy for lake attractiveness, we used lake surface area in hectares. While other lake characteristics may alter the attractiveness, lake area is most readily available and has been shown to be predictive in previous studies (Leung, Drake & Lodge 2004; Leung, Bossenbroek & Lodge 2006; Muirhead & MacIsaac 2011; Gertzen & Leung 2011).

Furthermore, our survey tool provided us with additional boater level information, which we were able to incorporate into the model. Respondents identified which type of boat they owned

and we categorized them as large motor boat (> 14'), small motor boat (< 14') or other. We assumed that boater type would modulate the relationship between lake size (W) and probability of visitation. As such, we incorporated this additional information using dummy variables ($B1$ and $B2$) in the exponent of W :

$$P^{GM}(T_{nj}) = A_n(W_j^{e+\beta_1 B1_n+\beta_2 B2_n} D_{nj}^{-d}), n = 1, \dots, n, j = 1, \dots, J. \quad \text{eqn 3}$$

$$A_n = 1 / \sum_{k=1}^J W_k^{e+\beta_1 B1_n+\beta_2 B2_n} D_{nk}^{-d} \quad \text{eqn 4}$$

Where $B1$ and $B2$ equal 0 for large motor boat, $B1$ equals 0 and $B2$ equals 1 for small motor boat, and $B1$ equals 1 and $B2$ equals 0 for other. In this way, boat type determines the rate at which each additional hectare of lake area increases the attractiveness of a given lake.

RANDOM UTILITY MODEL SPECIFICATION

The RUM is a discrete choice model used extensively in the econometrics literature to predict the behaviour of recreationalists (Parsons 2003). This formulation has recently been applied to predicting the spread of invasive zebra mussels in Wisconsin (Timar & Phaneuf 2009) and in a simulation study of the spread, and management of Eurasian watermilfoil (Macpherson, Moore & Provencher 2008). In this model, boaters are assumed to behave as rational actors, maximizing their utility. For a given trip, a boater chooses the site that maximizes their utility function U , which is only partially observable by the analyst. We can separate the utility function, therefore, into two parts. The utility that boater n would derive from visiting lake j can be then be written as the sum of the observable part V_{nj} , and an error term ϵ_{nj} .

$$U_{nj} = V_{nj} + \epsilon_{nj}, n = 1, \dots, N, j = 1, \dots, J \quad \text{eqn 5}$$

Where V_{nj} is any linear function of the attributes of boater n and site j .

$$V_{nj} = \vec{\beta} \vec{X}_{nj} \quad \text{eqn 4}$$

We can then re-write the utility that would be derived by boater n by visiting each site in terms of the probability that they will choose that site over all other alternatives.

$$\begin{aligned} P^{RUM}(T_{nj}) &= \Pr(U_{nj} \geq U_{nk} \forall k \neq j) \\ &= \Pr(V_{nj} + \epsilon_{nj} \geq V_{nk} + \epsilon_{nk} \forall k \neq j) \\ &= \Pr(\epsilon_{nk} - \epsilon_{nj} \leq V_{nj} - V_{nk} \forall k \neq j), n = 1, \dots, N, j = 1, \dots, J \end{aligned} \quad \text{eqn 7}$$

If we model the error term ϵ using the type I extreme value distribution as is most commonly done, the model reduces to a simple logit, and the distribution describing the probability that boater n will choose to visit lake j is given by:

$$P^{RUM}(T_{nj}) = \frac{\exp(V_{nj})}{\sum_{k=1}^J \exp(V_{nk})}, n = 1, \dots, N, j = 1, \dots, J \quad \text{eqn 8}$$

For further details of this model, see Parsons (2003). The parameters (β) are easily fit given the observed trip outcomes using maximum likelihood (see Fitting and model selection).

As with the gravity model, we incorporated the additional boater level predictor of boat type into the utility function of the RUM. We did this by adding two dummy variables to describe the three categories of boat type, with the same definitions as in the gravity model. Our full (observable) utility function is therefore formulated as

$$V_{nj} = \beta_1 W_j + \beta_2 D_{nj} + (\beta_3 B1_n + \beta_4 B2_n) W_j, n = 1, \dots, N, j = 1, \dots, J \quad \text{eqn 9}$$

As boat type is a boater level variable, we do not include it into the main effect part of the utility function, as it cancels out when summing across the entire choice set of a given boater. Instead, we model the interaction between boat type and lake size (W).

In both the RUM and GM model formulations, we have made two key assumptions: 1) boater behaviour is constant across time and 2) boater trips are distributed independently and identically according to each model.

FITTING AND MODEL SELECTION

The parameters (θ) of each model can be fit using maximum likelihood. Our survey data provide us with observations of the number of trip outcomes S_{nj} , for each boater n , to a given lake j . From these observations, we can write the log-likelihood for model M as:

$$LL^M(\theta|D) = \sum_{n=1}^N \sum_{j=1}^J S_{nj} \ln[P^M(T_{nj}|\theta)] \quad \text{eqn 10}$$

We fit the parameters of each model, including reduced models using maximum likelihood implemented in the R statistical programming environment (R Development Core Team 2008). Reduced models were those in which we removed the boater level parameters pertaining to boat type. Each model was then compared in terms of its relative performance using two separate metrics. The first metric of model selection we used was the Akaike Information Criterion (AIC) (Burnham & Anderson 2002). The second metric we used was the simple coefficient of determination (R^2) between the predicted and observed total number of trips taken to each lake in our study system. From this metric, we could compare the relative proportions of total variation in the number of visits across all lakes explained by each model.

Spread Simulations: Examining theoretical model behaviour and interactions with population demographics

Ultimately, we are constructing our models of human movement patterns between discrete patches to use in making predictions about the spread of species that are being dispersed across this network of patches. While spread is a stochastic process, where introductions lead to viable population establishments at a given site in a probabilistic manner, we can use repeated simulations to characterize the expected trajectory of a given invasion process (Peck 2004). By simulating the spread process under each of our competing models, we can compare the predicted trajectories to make inferences about the consequences of model specification on spread prediction. Differences in predicted spread rates, as well as predicted invasion risk at the individual site level may

have an effect on management decisions regarding mitigation and control.

To conduct these simulations, we followed the procedure outlined in Leung & Delaney (2006). We model the stochastic spread process as

$$G(Q_{jt}) = 1 - e^{-(\alpha Q_{jt})^c} \quad \text{eqn 11}$$

Where the probability of invasion is given as a function of the number of propagules Q , arriving at time t to site j . The function is described by two shape parameters. The first, α , is a per propagule multiplier proportional to $-\ln(1 - p)$, where p is the per propagule probability of establishment. The additional parameter c allows us to describe an Allee effect, where the per propagule establishment probability is disproportionately lower at low propagule pressures (Dennis 2002). The strength of the Allee effect increases as $c > 1$. Non-negligible Allee effects have been observed in some aquatic invasives. This parameter has been estimated as 1.86 ($P < 0.0001$; $H_0: c = 1$) for zebra mussels using an invasion time series (Leung, Drake & Lodge 2004). Wittmann *et al.* (2011) also detected an Allee effect using a stage-structured model of the invasive zooplankton *Bythotrephes*.

To calculate the number of propagules Q arriving at site j , we sum across the probability distribution of each boater having visited an invaded lake before arriving at site j . To do this, we first calculate the proportion of boaters at each source location that have visited an invaded lake as:

$$X_{i,t} = O_i \sum_{h=1}^H P^M(T_{ih}) \quad \text{eqn 12}$$

Where O_i is the number of boaters at source location i , and $P^M(T_{ih})$ is the probability of a boater at source location i visiting lake h as given by the model M under which simulations are being carried out. $X_{i,t}$ is the number of boaters in source location i having visited an invaded lake in time step t . We derived O_i from data obtained from the Ontario Ministry Natural Resources on the number of registered boaters in Ontario in each of 526 postal regions identified by the first three postal code digits. The next step is to calculate the propagule pressure Q arriving at lake j in time t as:

$$Q_{j,t} = \sum_{i=1}^K P^M(T_{ij}) X_{i,t} \quad \text{eqn 13}$$

Which is the total boater traffic from all invaded sources to lake j in time step t . For more details, see Gertzen & Leung (2011). While each human vector model predicts a unique trip distribution matrix, the total number of boater trips taken, or the overall magnitude of traffic flow in the system as a whole, is constant across both models. Any difference in the observed rates of spread in our simulations therefore is a result of the dispersal network structure, and not the absolute magnitude of between-lake movement.

While there are roughly 250 000 lakes and rivers in Ontario, to render our simulations computationally feasible, we simulate spread across only those lakes with a surface area larger than 10 hectares. Additionally, we removed lakes above 52° latitude, as these lakes are not accessible by any roadways connecting them to the southern lakes. This left us with 781 lakes in our simula-

tion set. Each independent simulation began with a seed invasion in Lake Ontario and was run forward 30 years. By seeding the invasion in Lake Ontario, we recreate the most likely invasion scenario for Ontario inland lakes. As of 2006, the great lakes are known to have been invaded by at least 182 species (Ricciardi 2006), making it the most likely source location of a novel species spread to inland lakes.

To analyse potential interactions between population dynamics and the human vector model, we examined the effect of population establishment parameters and we ran repeated simulations across a range of parameter values of both α (7.5e05, 1.0e-04, 1.25e-04, 1.5e-04) and c (1, 1.5, 2, 2.5). For each simulation, we used either the best fitting GM or RUM of boater behaviour. As our metrics of invasion progress, for each run, we retained the cumulative number of lakes invaded. An example realization of our simulated spread procedure can be seen in Fig. 2. Additionally, we compared the relative invasion risk at each of three specific selected sites. Lakes Simcoe, Nipissing and Nipigon were selected because of their large size, making them more at risk to invasion, as well as because of their relative distances from the source location of invasion. While these lakes by no means represent a random sample, they provide a convenient gradient of baseline risk along which to observe the rate at which deviations between models occur. For these lakes, we retained the time to invasion across every simulation for every parameter combination. We calculated the risk to a given lake as the proportion of simulation realizations in which the site became invaded before the end of the 30-year time horizon.

Results

MODEL FITTING AND MODEL SELECTION

Formal model selection identified, the GM as the most likely, given the data. The GM provided superior fit to the RUM with a ΔAIC value of 3229 between the full GM and the full RUM, for the observed pattern of boater trips. Table 1 provides the ΔAIC for each model, sorted in increasing order (decreasing order of goodness-of-fit). Maximum likelihood parameter estimates and their 95% confidence intervals for each full model are given in Table 2. All fitted parameter values have direction and magnitude that we would expect. In the gravity model, $0 < e < 1$, indicating a diminishing marginal effect of each additional hectare of lake surface area. The boater-specific dummy variables ($\beta_1 > 0$, $\beta_2 < 0$) indicate that the marginal effect of each additional hectare decreases fastest for small motor boats. That $d > 0$ indicates that closer lakes are more attractive than more distant ones. Similarly, in the RUM $\beta_1 > 0$ and $\beta_2 < 0$ indicate a positive relationship between lake area and utility, and a negative relationship between distance and utility, respectively. As with the gravity model $\beta_3 > 0$ and $\beta_4 < 0$ indicate that the marginal utility of an additional hectare diminishes fastest for small motor boats. The full GM was able to account for 58% of the variation in trip outcomes across all boaters, compared to only 42% for the full RUM (Fig. 3). As a check for bias, we fit a linear regression to the predicted-observed points. The equations of fit were

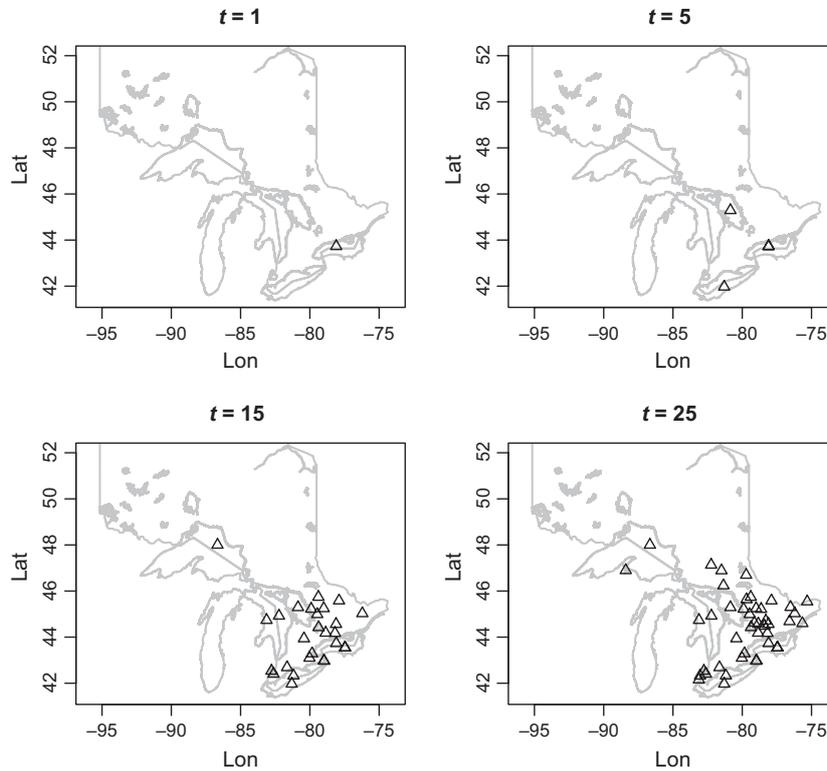


Fig. 2. Map of an example outcome of a spread simulation. Triangles indicate lakes that have become invaded as of time t . Shown is a single realization of the spread process under the gravity model with parameters $\alpha = 1.25e - 04$, $c = 2$.

Table 1. Model comparison by delta Akaike information criterion (ΔAIC)

Model	ΔAIC
<i>Gravity models (GM)</i>	
$P_{nj}^{GM} = A_n(W_j^{e+\beta_1 B_1 + \beta_2 B_2}) D_{nj}^{-d}$	0
$P_{nj}^{GM} = A_n(W_j^e D_{nj}^{-d})$	62
<i>Random utility model (RUM)</i>	
$U_{nj} = \beta_1 W_j + \beta_2 D_{nj} + (\beta_3 B_{1n} + \beta_4 B_{2n}) W_j + \epsilon^\dagger$	3139
$U_{nj} = \beta_1 W_j + \beta_2 D_{nj} + \epsilon^\dagger$	3187
<i>Null</i>	
$P_j = 1/J$	14732

* A_n follows the same form as the parenthetical part shown in the table.

†Shown are only the utility (both the observable and random) components of the random utility model. See eqns 5–9 for full specification.

$y = 0.00018[\pm 0.00016] + 0.86[\pm 0.034] x_{GM}$ and $y = -0.00031[\pm 0.00018] + 1.24[\pm 0.058] x_{RUM}$. While neither intercepts deviate significantly from zero, the slope of

the GM is less than one, indicating a tendency to overestimate the traffic to high frequency lakes, while the RUM tends to underestimate traffic to high frequency lakes.

There are three main components of the differences between the dispersal networks predicted by each model. (i) Each model generally predicted higher traffic to large lakes that are close to dense population sources, as expected. However, the rank ordering of individual lakes can differ substantially within this broader pattern (Fig. 4a). (ii) The average predicted distance travelled by boaters was higher in the GM (190 km) compared to the RUM (140 km) and (iii) How evenly, or unevenly, the predicted traffic was spread across different sites differed between models. To quantitatively evaluate this

Table 2. Maximum likelihood parameter values and 95% confidence intervals for each human vector model

Parameter	MLE ($\hat{\theta}$)	95% CI
<i>Gravity models (GM)</i>		
e	0.51	[0.486, 0.525]
β_1	0.14	[0.083, 0.203]
β_2	-0.13	[-0.179, -0.0883]
d	1.86	[1.82, 1.89]
<i>Random utility model (RUM)</i>		
β_1	0.0011	[0.00106, 0.0012]
β_2	-1.40	[-1.447, -1.344]
β_3	0.00043	[0.000273, 0.000578]
β_4	-0.00044	[-0.000614, -0.000276]

characteristic, we calculated the Shannon entropy of the traffic distributions predicted by the two models. Entropy can be thought of as a measure of evenness (Hill 1973). Probability distributions with higher entropy are more evenly dispersed. As entropy decreases, the distribution becomes more uneven, or more sharply peaked, such that more of the mass of the distribution is concentrated in fewer sites. We calculate the Shannon entropy of the predicted distributions P for model M as:

$$H(P^M) = - \sum_{i=1}^I P_i^M \log(P_i^M) \quad \text{eqn 14}$$

As $H(P^M) \rightarrow 0$, boater traffic is concentrated entirely in one lake. The maximum entropy distribution would be that which assigns $P_i = 1/n$ to all lakes in the system. Comparing the predictions of the two models, we find that the gravity model [$H(P^{GM}) = 5.018$] represents a more uneven predicted distribution than that of the RUM [$H(P^{RUM}) = 5.36$]. The stronger concentration of traffic predicted by the GM, compared with the RUM can be seen by comparison of the rank-ordered distributions in Fig. 4b. The consequences of these differences are analysed in the following section.

IMPLICATIONS FOR SPREAD AND RISK ASSESSMENT

When the per propagule probability of establishment is low (low α) and there is no Allee effect present, both dispersal models predict similar rates during the early phases of invasion (Fig. 5). The deviations between early rates of spread under the alternative dispersal models increases drastically, however, as the strength of the Allee effect increases. At the extreme end of invasiveness and Allee effect ($\alpha = 1.5e-04$ and $c = 2.5$), we observe an over ten-fold increase in the cumulative total number of sites invaded by the end of the 30-year time horizon. The degree of deviation induced by increased Allee strength is

also modulated by the independent population growth, or per propagule invasiveness parameter α . This can be seen by observing the magnitude of deviation at each row of Fig. 5, which increases as the parameter α increases.

In the absence of an Allee effect, the deviations in late-stage rates of spread can be accounted for by the differences in predicted mean distance travelled under each model. Spread in Southern Ontario occurs at similar rates under each model because of high population density, where the distances between population sources and lakes are short. However, as the invasion progresses northward into the more sparsely populated regions, spread under the RUM is slowed substantially because of the increased distances required to reach additional lakes. When an Allee effect is present, the difference in spread rate is apparent throughout the entire time series. The relative entropies of the dispersal distributions (i.e. the variance in total inbound propagules arriving across all uninvaded sites) can account for this further deviation. The expected rate at which the proportion of previously uninvaded sites become invaded can be written as $R = E[G(Q_{\bullet})]$, where $G(Q_{\bullet})$ is a vector of invasion probabilities for all uninvaded sites and is given by Eqn. 11. When an Allee effect is present, the function $G(Q_{\bullet})$ is concave over part of its range. By Jensen's inequality, we know that for a convex function $E[G(Q_{\bullet})] \geq G[E(Q_{\bullet})]$. From this we can see that as the variance of inbound propagules increases over the concave range as a result of a more uneven dispersal distribution, the rate of new invasions increases as well.

Spread at the landscape level may be of interest to regional managers; however, the risk of invasion posed at specific sites will inform management decisions made at the lake level. To see the differences in the site-level invasion risk predicted by our alternative dispersal models, we also looked at three specific inland lakes (Lakes Simcoe, Nipissing and Nipigon). These three lakes occur at increasing distances from our source location (Lake

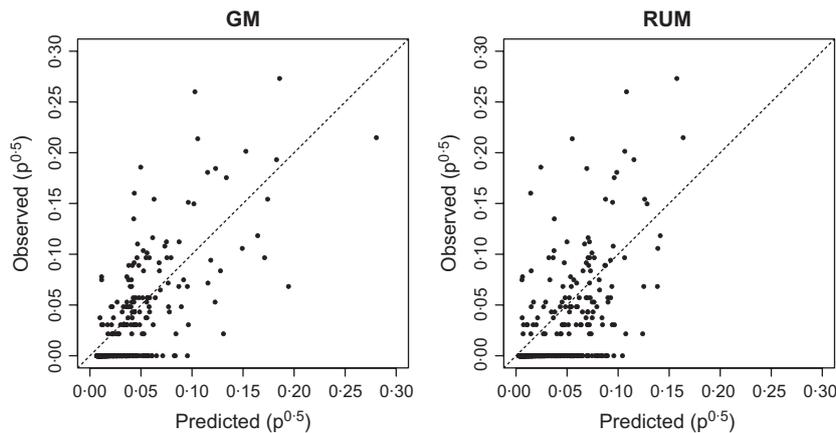


Fig. 3. Predicted versus observed relative lake visitation frequency (p) for both models. Both axes have been square-root transformed to better visualize low values. Coefficients of correlation (R^2) are 0.58 and 0.42 for the gravity (GM) and random utility (RUM) models, respectively. The dotted line is the 1:1 line.

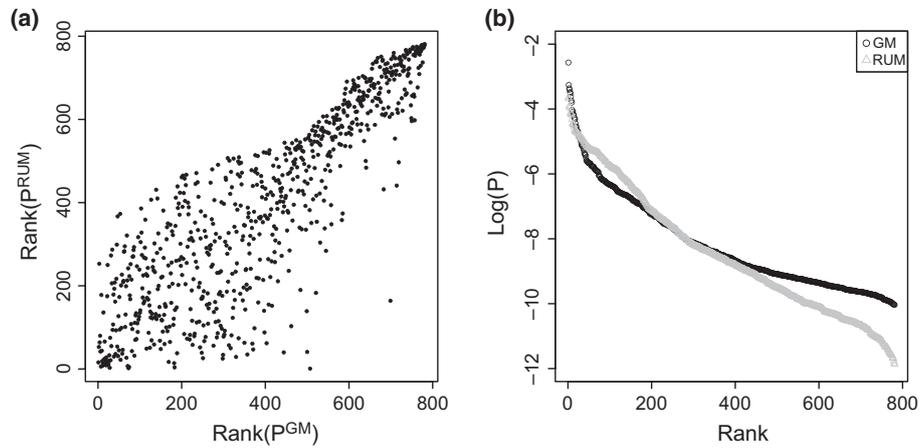


Fig. 4. Panel (a): Comparison of the ranking of predicted traffic to individual lakes for each model. Panel (b): Ranked predicted boater traffic distributions for the gravity (black circles) and random utility (grey triangles) models.

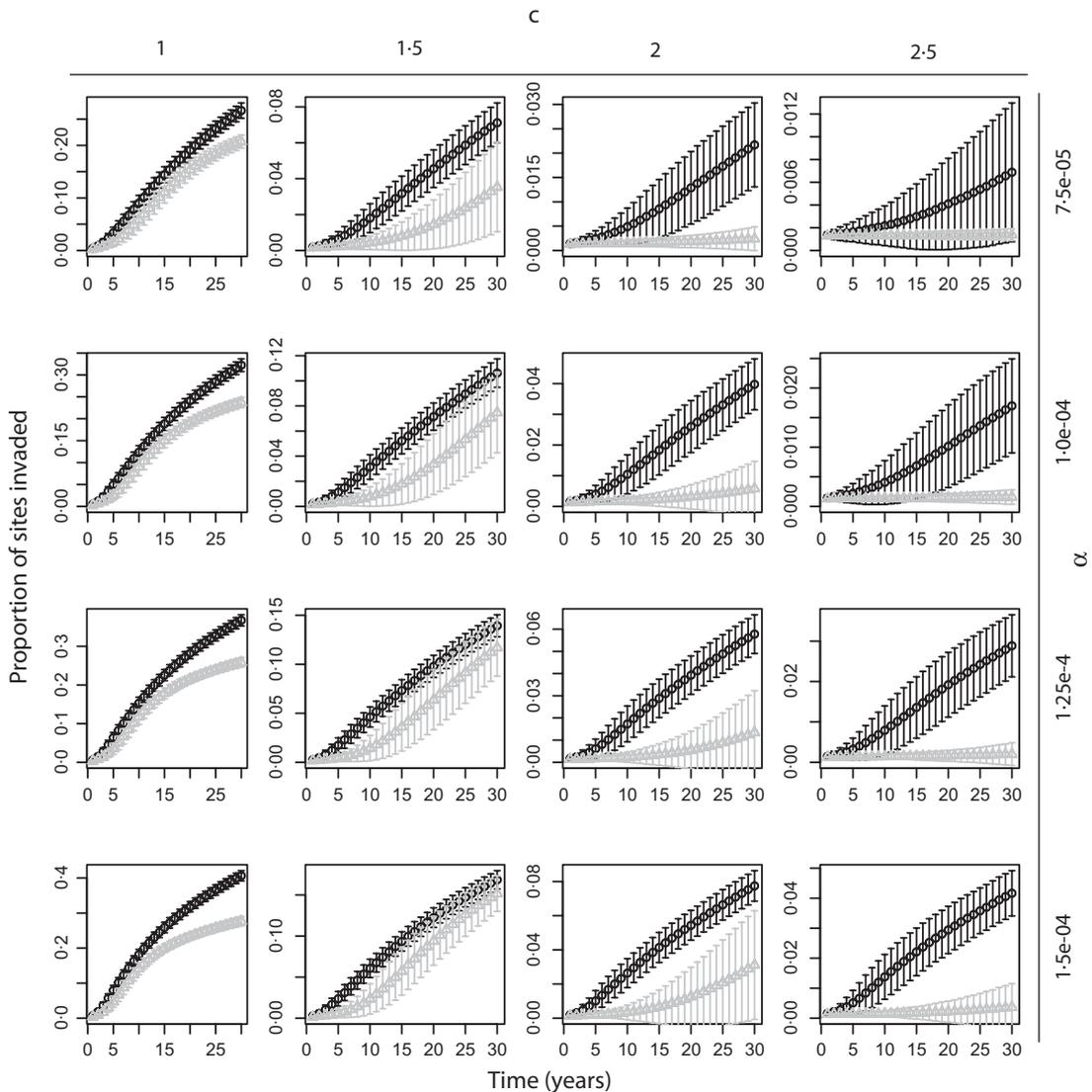


Fig. 5. Invasion trajectories (proportion of total number of lakes invaded) predicted using the gravity model (black circles) and the random utility model (grey triangles). Panels show factorial parameter combinations of $\alpha = 7.5 \cdot 10^{-5}, 1.0 \cdot 10^{-4}, 1.25 \cdot 10^{-4}, 1.5 \cdot 10^{-4}$ for no Allee effect ($c = 1$), and increasing Allee effects of $c = 1.5, 2, 2.5$. Each model and parameter set was run for 1000 replicates. Error bars show the range encompassed by 95% of invasion simulations.

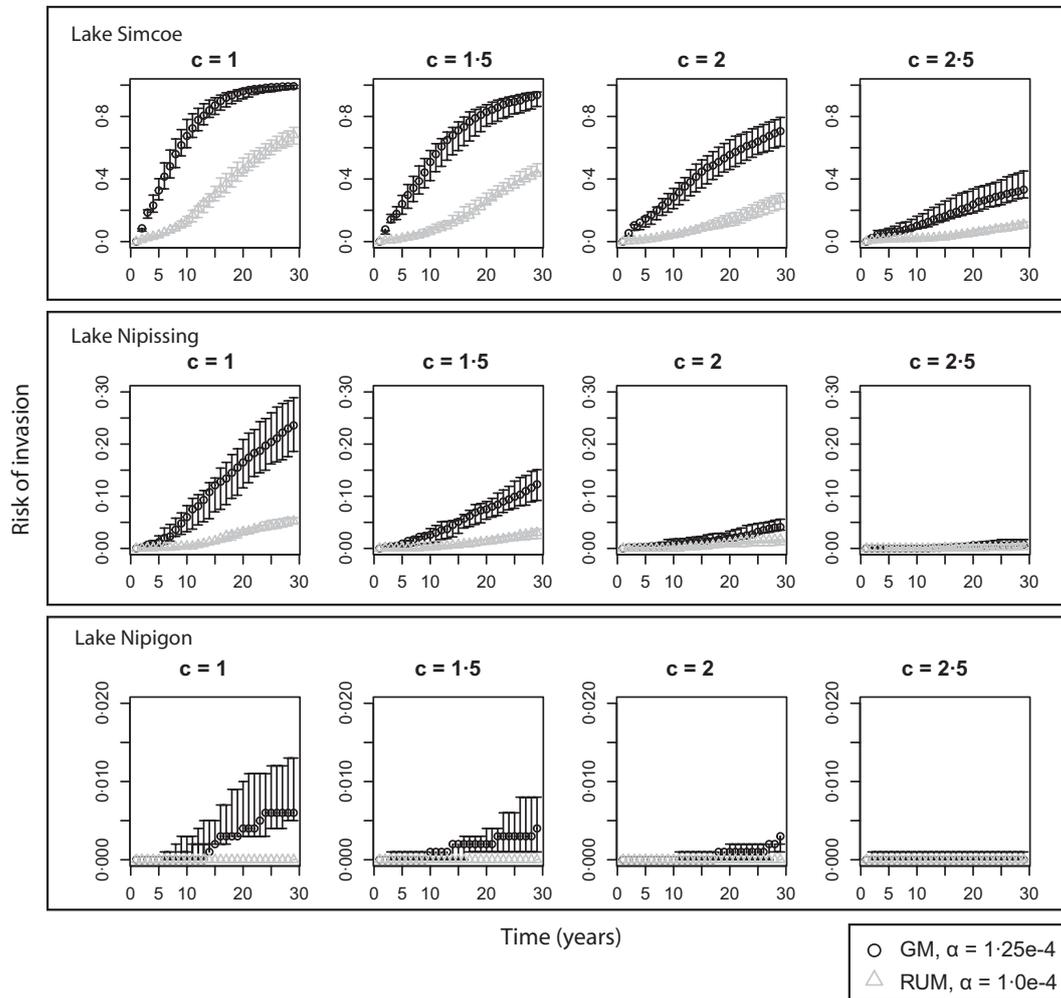


Fig. 6. Invasion risk to selected locations (Lakes Simcoe, Nipissing and Nipigon). Plots show the invasion risk (proportion of times invaded across 1000 spread simulations) across time with 95% confidence intervals. Panel columns show Allee effect increasing to the right. Alternative human vector models are shown in black circles (GM) and grey triangles (RUM).

Ontario), respectively. By observing the differences in invasion risk predicted at these sites, it is possible to see how uncertainty and deviations between model predictions increase as we move further from the source of invasion. Figure 6 shows probability of invasion (risk) as a function of time at each of the three sites across our range of population parameters. The risk of invasion posed at each of these three sites is always higher under the GM, with the exception Lakes Nipissing and Nipigon under strong Allee effect where projected risk is very near zero and indistinguishable between models.

Discussion

In this study, we have shown that a GM can better capture the behaviour of individual boaters in Ontario than an RUM. Ultimately, these two alternative models can be represented simply as different functional forms which we can use to describe a boater's trip-taking probability distribution. In the case of our sample of Ontario boat-

ers, the functional form of the GM provides a better representation of the probabilistic process through which boaters select which lakes to visit from a suite of alternatives.

Both of the behavioural models considered in this study were built using only the distance between the boater source location, lake size (surface area in hectares) and boat type as explanatory factors. We recognize that there may be a suite of additional variables that may add further explanatory power. Previous work has incorporated additional lake predictors, as well as additional interactions between individual level and lake variables. These have included lake clarity (measured as secci depth), cost of access and whether or not a given boater is an angler (Parsons 2000; Timar & Phaneuf 2009). Here, we have used lake size, distance from boater's home location, and boat type as these are the most readily available data with which to build a model of boater behaviour for the purposes of assessing invasion risk. Both the gravity and RUMs can easily be extended to incorporate any number

of additional lake and boater-specific variables. In the RUM, one would need simply to include additional linear, or higher order, predictors (β) in V_{nj} (see Eqn 9). Within a GM, lake level predictors could be modelled by the expansion of W_j into:

$$W_j^{e+\beta_1 X_{1j}+\dots+\beta_i X_{ij}} \quad \text{eqn 15}$$

With additional explanatory variables of lake attractiveness each requiring the fitting of an additional free parameter (see Fitting and model selection).

We have also demonstrated that the choice of modelling framework used to characterize the human-mediated vector can have important consequences for predicting the future spread of invasive species. The deviation between spread predictions under the two frameworks analysed here interact with population level factors of the invading species. In the presence of a strong Allee effect, boaters behaving according to RUM do not act in such a way as to generate propagule pressures high enough to overcome the demographical barriers to establishment. The inability to overcome these barriers is a consequence of the evenness of the predicted trip distribution of boaters under the RUM, as measured using the Shannon entropy of the predicted trip distribution. Without the centralized 'hub' lakes (those highly connected lakes with very high visitation frequency) predicted by the GM, a situation can arise where there is not a sufficiently concentrated flow of individuals from invaded lakes to uninvaded lakes. The data suggests rather that individuals do act in such a way as to concentrate traffic to a small number of 'hub' lakes, as predicted by a GM, and that this level of concentration is sufficient to overcome even very strong demographical barriers. The existence of such hub lakes and their importance to the spread of aquatic invasive species has been noted in the literature (MacIsaac *et al.* 2004; Muirhead & MacIsaac 2005). A misspecified behavioural model of human-mediated dispersal may underestimate the importance of these sites, leading to potentially overoptimistic projections of lake to lake spread.

We also know, however, that just predicting the rate of spread may not be the most relevant metric of interest to a resource manager who is making decisions at the local level. A more relevant measure at the lake level is the risk of invasion posed at particular sites. We analysed how our dispersal models affect site-level predictions of risk by pulling out three of the larger, more important sites in Ontario, Lakes Simcoe, Nipissing and Nipigon. For these specific lakes, the predicted probability of establishment over time differed dramatically between the GMs and RUMs, even in the absence of Allee effects. Indeed, these are three of the largest inland lakes in Ontario, all of which are probably receiving sufficient propagule pressure rather early on to overcome the demographical barrier of an Allee effect. From this result, we can see that the way in which the underlying behavioural model interacts with

the population dynamics of the invading species manifests differently at the site level, than at the landscape level.

When making policy decisions regarding invasive species, managers need informed estimates of invasion risk across space and time (Epanchin-Niell & Hastings 2010). This study suggests that for boaters in Ontario, a GM of individual behaviour most accurately characterizes this single most important vector of overland invasive spread and that alternative formulations of human vector dispersal models can interact with the population dynamics of the invading species to produce large deviations in predicted spread. These deviations manifest differently at the system level than at the level of individual lakes. While managers of inland freshwater resources should be aware of how these interactions impact assessments of risk, our results are general and hold for any species spreading across a network of discrete patches.

Future work could look at the utility of implementing GMs in the context of management interventions. Were managers to implement policies aimed at limiting the spread of an aquatic invasive species by levying a launching fee, or requiring hull sanitation procedures at either at risk or currently infested lakes, boaters may change their behaviour. Changes in boater behaviours resulting from management interventions could potentially alter the structure of the human-mediated dispersal network. This, as we have shown, will have consequences for our predictions of spread. Previous work has employed RUMs to incorporate these behavioural feedbacks (Macpherson, Moore & Provencher 2008; Timar & Phaneuf 2009); however, in the light of our current results, it may be appropriate to include these behaviours directly in a GM formulation.

While we have shown here that the interaction between intra-patch dispersal connectivity and stochastic population dynamics within patches interact to determine rates of population spread across a landscape, there will no doubt be effects of other factors, such as spatial and temporal environmental heterogeneity (Melbourne *et al.* 2007), biotic interactions (Hunt & Behrensamada 2003), as well as temporal variation in the dispersal network structure itself which should be considered in future studies.

Acknowledgements

The authors would like to thank all of our survey respondents, as well as Jason Cologna at the Ontario Ministry of Natural Resources for help registered boater mailing list data. Funding for this study was provided by the Canadian Aquatic Invasive Species Network.

References

- Aukema, J.E., Leung, B., Kovacs, K., Chivers, C., Britton, K.O., Englin, J., Frankel, S.J., Haight, R.G., Holmes, T.P., Liebhold, A.M., McCullough, D.G. & Von Holle, B. (2011) Economic impacts of non-native forest insects in the continental United States. *PLoS One*, **6**, e24587.
- Bossenbroek, J.M., Kraft, C.E. & Nekola, J.C. (2001) Prediction of long-distance dispersal using gravity models: Zebra mu invasion of inland lakes. *Ecological Applications*, **11**, 1778–1788.

- Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer, New York.
- Clark, J.S., Lewis, M., McLachlan, J.S. & HilleRisLambers, J. (2003) Estimating population spread: what can we forecast and how well? *Ecology*, **84**, 1979–1988.
- Dennis, B. (2002) Allee effects in stochastic populations. *Oikos*, **96**, 389–401.
- Dillman, D.A. (2000) *Mail and Internet Surveys: The Tailored Design Method*, 2nd edn. John Wiley & Sons, New York, NY.
- Drake, D.A.R. & Mandrak, N.E. (2010) Least-cost transportation networks predict spatial interaction of invasion vectors. *Ecological Applications*, **20**, 2286–2299.
- Drake, J.M., Drury, K.L.S., Lodge, D.M., Blukacz, A., Yan, N.D. & Dwyer, G. (2006) Demographic stochasticity, environmental variability, and windows of invasion risk for bythotrepes longimanus in North America. *Biological Invasions*, **8**, 843–861.
- Epanchin-Niell, R.S. & Hastings, A. (2010) Controlling established invaders: integrating economics and spread dynamics to determine optimal management. *Ecology letters*, **13**, 528–541.
- Gertzen, E.L. & Leung, B. (2011) Predicting the spread of invasive species in an uncertain world: accommodating multiple vectors and gaps in temporal and spatial data for Bythotrepes longimanus. *Biological Invasions*, **13**, 2433–2444.
- Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**, 427–432.
- Hunt, C.E. & Behrensyamada, S. (2003) Biotic resistance experienced by an invasive crustacean in a temperate estuary. *Biological Invasions*, **5**, 33–43.
- Johnson, L.E., Ricciardi, A. & Carlton, J.T. (2001) Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. *Ecological Applications*, **11**, 1789–1799.
- Kraft, C.E., Sullivan, P.J., Karatayev, A.Y., Burlakova, L.E., Nekola, J. C., Johnson, L.E. & Padilla, D.K. (2002) Landscape patterns of an aquatic invader: assessing dispersal extent from spatial distributions. *Ecological Applications*, **12**, 749–759.
- Landis, W.G. (2004) Ecological risk assessment conceptual model formulation for nonindigenous species. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, **24**, 847–858.
- Leung, B., Bossenbroek, J.M. & Lodge, D.M. (2006) Boats, pathways, and aquatic biological invasions: estimating dispersal potential with gravity models. *Biological Invasions*, **8**, 241–254.
- Leung, B. & Delaney, D.G. (2006) Managing sparse data in biological invasions: a simulation study. *Ecological Modelling*, **198**, 229–239.
- Leung, B., Drake, J.M. & Lodge, D.M. (2004) Predicting invasions: propagule pressure and the gravity of Allee effects. *Ecology*, **85**, 1651–1660.
- MacIsaac, H.J., Borbely, J.V.M., Muirhead, J.R. & Graniero, Pa. (2004) Backcasting and forecasting biological invasions of Inland lakes. *Ecological Applications*, **14**, 773–783.
- Macpherson, A.J., Moore, R. & Provencher, B. (2008) A dynamic principal-agent model of human-mediated aquatic species invasions. *Agricultural and Resource Economics Review*, **35**, 144–154.
- Melbourne, B.A., Cornell, H.V., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A.L., Hall, R.J., Harrison, S., Hastings, A., Holland, M., Holyoak, M., Lambrinos, J., Moore, K. & Yokomizo, H. (2007) Invasion in a heterogeneous world: resistance, coexistence or hostile takeover? *Ecology letters*, **10**, 77–94.
- Muirhead, J.R. (2007) *Forecasting dispersal of nonindigenous species*. PhD thesis, University of Windsor.
- Muirhead, J.R. & Macisaac, H.J. (2005) Development of inland lakes as hubs in an invasion network. *Journal of Applied Ecology*, **42**, 80–90.
- Muirhead, J.R. & MacIsaac, H.J. (2011) Evaluation of stochastic gravity model selection for use in estimating non-indigenous species dispersal and establishment. *Biological Invasions*, **13**, 2445–2458.
- Parker, I.M., Simberloff, D., Lonsdale, W.M., Goodell, K., Wonham, M., Kareiva, P.M., Williamson, M.H., Von Holle, B., Moyle, P.B., Byers, J. E. & Goldwasser, L. (1999) Impact: toward a framework for understanding the ecological effects of invaders. *Biological Invasions*, **1**, 3–19.
- Parsons, G.R. (2000) Familiar and favorite sites in a random utility model of beach recreation. *Marine Resource Economics*, **14**, 299–315.
- Parsons, G.R. (2003) The travel cost model. *Primer on Nonmarket Valuation* (eds P.A. Champ, K.J. Boyle & T.C. Brown), pp. 269–330. Kluwer Academic Publishers, Boston.
- Peck, S.L. (2004) Simulation as experiment: a philosophical reassessment for biological modeling. *Trends in Ecology & Evolution*, **19**, 530–534.
- Pejchar, L. & Mooney, H.A. (2009) Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution*, **24**, 497–504.
- Potapov, A., Muirhead, J.R., Lele, S.R. & Lewis, Ma. (2011) Stochastic gravity models for modeling lake invasions. *Ecological Modelling*, **222**, 964–972.
- R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ricciardi, A. (2006) Patterns of invasion in the Laurentian Great Lakes in relation to changes in vector activity. *Diversity and Distributions*, **12**, 425–433.
- Schneider, D.W., Ellis, C.D. & Cummings, K.S. (1998) Model native mussel assessment Zebra of the risk to communities from spread. *Conservation Biology*, **12**, 788–800.
- Smirnov, O.A. & Egan, K.J. (2010) Spatial random utility model with an application to recreation demand. *Economic Modelling*, **29**, 72–78.
- Smith, V.K. & Kaoru, Y. (1986) Modeling recreation demand within a random utility framework. *Economics Letters*, **22**, 395–399.
- Timar, L. & Phaneuf, D.J. (2009) Modeling the human-induced spread of an aquatic invasive: the case of the zebra mussel. *Ecological Economics*, **68**, 3060–3071.
- Wittmann, M., Lewis, M., Young, J.D. & Yan, N. (2011) Temperature-dependent Allee effects in a stage-structured model for Bythotrepes establishment. *Biological Invasions*, **13**, 2477–2497.

Received 14 January 2012; accepted 28 June 2012
Handling Editor: Jason Matthiopoulos