

Predicting the spread of invasive species in an uncertain world: accommodating multiple vectors and gaps in temporal and spatial data for *Bythotrephes longimanus*

Erin L. Gertzen · Brian Leung

Received: 30 March 2010 / Accepted: 6 January 2011
© Springer Science+Business Media B.V. 2011

Abstract Real-world uncertainties and data limitations make it difficult to predict how, when and where non-indigenous species (NIS) will spread. Typically only a small fraction of sites are sampled during only a few time intervals, such that we know neither the full spatial extent nor the true temporal progress of invasion. Yet, these unsampled locations might affect the invasion dynamics. We extend propagule pressure models to incorporate both human-mediated and natural fluvial dispersal vectors, and develop techniques to incorporate missing spatial and temporal data on invasions. We apply our model to *Bythotrephes longimanus*, a high-risk aquatic NIS, using a regional-scale 311-lake survey in a popular watershed in Ontario and extending our analysis to 1,300 unsampled lakes. Of 100 model runs with different random subsets of 50 sampled lakes reserved for validation, we were able to obtain an average area under the curve value of 0.89. Human-mediated dispersal accounted for 99.75% of the contribution of propagules to probability of establishment. Although the discovery rate is accelerating, our results suggest the annual rate of lake invasions is decelerating over

time. Management efforts controlling recreational boating traffic out of the largest lakes in the system will be the most effective way of slowing the spread of *B. longimanus* in lakes within this system.

Keywords Gravity model · Probability of establishment · Uncertainty · Invasive species · *Bythotrephes longimanus*

Introduction

Our ability to prevent and control the secondary spread of non-indigenous species (NIS) is hampered because we cannot forecast exactly how invasions will spread in space and time (Johnson and Padilla 1996; Lodge et al. 2006). Past efforts to predict spread have focused on trait-based, niche-based and propagule pressure based methods. Trait-based methods use species characteristics to predict which NIS are high risk while niche-based methods use habitat characteristics in both the invaded and native ranges to predict the full potential of spread in the invaded range (e.g., Ricciardi and Rasmussen 1998; MacIsaac et al. 2000; Kolar and Lodge 2002). Propagule pressure based approaches relate introduction effort (number and frequency of propagule introductions) to the strengths of vectors of spread to estimate invasion risk for different locations (e.g., Schneider et al. 1998; Bossenbroek et al. 2001; Hastings et al. 2005).

E. L. Gertzen (✉) · B. Leung
Department of Biology, McGill University, Montreal,
QC H3A 1B1, Canada
e-mail: erin.gertzen@mail.mcgill.ca

B. Leung
McGill School of Environment, McGill University,
Montreal, QC H3A 1B1, Canada

While trait and habitat based approaches are useful and indicate which locations are ultimately likely to become invaded by a given species, propagule pressure has been argued to be the most consistent predictor of invasion success (Lockwood et al. 2005). Moreover, management strategies are typically aimed at reducing propagule pressure by controlling vectors of spread (Lodge et al. 2006), making propagule pressure based approaches most directly linked to management.

In this manuscript, we focus on estimating probability of establishment due to propagule pressure, which provides biological relevance and potentially allows quantification of the benefits of management efforts that reduce propagule pressure (Leung et al. 2004). Despite its key advantages, linking propagule pressure to probability of establishment remains an open problem and is unmeasured for many systems, given that data are typically limited (but see Leung et al. 2004; Drake and Lodge 2006; Jerde and Lewis 2007). Here, we develop a framework to address common limitations and uncertainties which are typically encountered when building models to predict secondary spread: (1) we cannot directly measure propagule pressure from different vectors of spread (Johnson and Padilla 1996), (2) we do not know the relative contribution of propagules from different vectors, (3) we cannot sample all locations (Leung and Delaney 2006), and (4) we cannot sample locations frequently (Solow and Costello 2004).

First, it is usually impossible to measure individual introduction events; however we can use surrogates of propagule pressure. For example, in gravity models, surrogate measures of propagule pressure link lake attractiveness characteristics such as lake size and accessibility (Reed-Andersen et al. 2000) to recreational boating movement patterns (Buchan and Padilla 1999; Bossenbroek et al. 2007). To date, most modeling efforts have focused on only one vector at a regional scale (Bossenbroek et al. 2001; Leung et al. 2004; but see Muirhead 2007). Yet, for any given NIS, multiple vectors may be crucial for spread. Still, we might expect that multiple vectors provide different contributions to spread and that the viability of propagules from these vectors differs. Our framework allows us to assess the contribution of each vector to establishment.

The third and fourth limitations addressed are a lack of complete data, both in space and time. Because the entire system is interconnected, unsampled locations can act as both sources and alternative

destinations for propagules. However, researchers typically ignore unsampled locations (Schneider et al. 1998; Muirhead and Macisaac 2005) or treat them as uninvaded (Anderson and Martinez-Meyer 2004), which may change the behavior of the system and result in misleading predictions (Leung and Delaney 2006). Instead, we should incorporate these locations and develop techniques to deal with uncertainty about their invasion status. Such techniques have been developed but have only been tested theoretically (Leung and Delaney 2006); we apply these techniques to a real-world dataset for the first time.

Additionally, we are often also lacking information on the exact date of invasions for sampled sites. Most monitored locations are sampled infrequently; an invasion could exist for many years before it is detected. Further, typically it is assumed that the discovery date is equivalent to the year in which the NIS became established (Solow and Costello 2004). This can potentially result in misleading predictions on the number of lakes at risk in a given time interval, as well as on the rate of the overall invasion expansion. Here, we develop techniques to incorporate available temporal information and relate invasions to probability of establishment over time. In doing so, we provide a means to estimate when invasions most likely occurred and, thus, infer the invasion rate (number of new invasions per year) in the system.

In this paper, we develop novel approaches to predict the secondary spread of NIS in the context of limited information. We apply these approaches to the high-profile NIS, *Bythotrephes longimanus* (henceforth *Bythotrephes*), the spiny waterflea, a predatory cladoceran that invaded the Great Lakes region from the Palearctic in the early 1980s (Bur et al. 1986). Specifically, we aim to (1) assess the relative risk of anthropogenic and natural vectors of spread, (2) estimate the actual invasion rate in comparison to the discovery rate, and (3) predict which unsampled lakes are most likely invaded or at risk of becoming invaded, and make suggestions on where to focus management efforts.

Methods

Study system

We used a 311-lake survey, conducted in conjunction with the Canadian Aquatic Invasive Species

Network, to develop our approaches (Cairns et al. 2007; Weisz and Yan 2010). The survey was conducted in 2005 and 2006 in the 2 EB watershed, located in south-central Ontario on the Canadian Shield, 200 km north of Toronto. *Bythotrephes* was first observed in the popular cottage and boating watershed in 1989 (Yan et al. 1992). The 311-lake survey represents a substantial effort and generated more information than is commonly available for many invasive species; but still, it only provided us with a snapshot of information on 17% of the lakes in the watershed, and very limited temporal information on the progress of invasions. Other data on *Bythotrephes* absence, presence and discovery dates came from previously published works and personal communications, while data on several new invasions came from a re-sampling of a subset of lakes from the 311-lake survey conducted in 2007 (pers. comm. Arnott 2007; Ontario Federation of Anglers and Hunters; Yan et al. 1992; Yan and Pawson 1997; Boudreau and Yan 2004; Cairns et al. 2007). In total, we had presence/absence data for *Bythotrephes* in 336 of the 1,664 lakes (20%), of which 50 were known to be invaded.

Model development

To predict the secondary spread of *Bythotrephes*, we (1) calculated propagule pressure to each lake, (2) estimated probability of establishment from these propagule pressure estimates, (3) calibrated the model to subsets of available empirical data, and (4) validated our predictions on different subsets of data. Importantly, we re-ran the model calibration on 100 random subsets of the available empirical data and validated our predictions using 100 random subsets of lakes not included in the model calibration. This provides a truer examination of the reliability of model predictions. The majority of previous propagule pressure based modeling efforts have either not employed any validation or not used external validation data (but see Leung et al. 2004).

We expect boating traffic to be an important vector of spread because *Bythotrephes*' long caudal appendage makes it particularly susceptible to fouling boating and fishing equipment (Jacobs and MacIsaac 2007; Weisz and Yan 2010). Additionally, stream dispersal may be a source of propagules. Although the role of streams in the spread of *Bythotrephes* has

not been investigated, Johnson et al. (2006) found that 30% of downstream invasions of *Dreissena polymorpha*, zebra mussels, could be potentially attributed to stream dispersal; *Bythotrephes* putatively can be transported passively to downstream uninvaded lakes in a similar way to zebra mussel veligers.

Estimating propagule pressure due to boating traffic

We used a gravity model (Bossenbroek et al. 2001; 2007; Leung et al. 2004) to estimate relative propagule pressure to uninvaded lakes due to recreational boating traffic. The model follows a two-step process, where boaters travel to a lake, return home, and subsequently travel to another (or the same) lake (Bossenbroek et al. 2001). Specifically, we used a production-constrained gravity model which requires a modest amount of data that are reasonably easy to access. Gravity models are constructed on gravity pulls of boaters to each lake, where the boating traffic (U_j) to each destination lake j is related to the attractiveness (W_j) of that lake, the number of boaters (O_i) at each source location i , the distance along a road (D_{ij}) between source i and destination j . A_i is a scalar that takes into account the fact that boaters may visit alternative destinations (invaded/uninvaded lakes). Calculations of U_j and A_i provide the basis of the model.

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

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

where U_j provides a metric of the total boating vector traffic to each destination lake. While U_j is constant, not all boats visiting lake j will have previously visited an invaded lake. $X_{i,t}$ gives the proportion of boat traffic sources i that have been exposed to invaded lakes in year t .

$$X_{i,t} = \frac{\sum_{m=1}^M U_{im}}{\sum_{l=1}^L U_{il}} \tag{3}$$

$$Q_{j,t} = \sum_{i=1}^K U_{ij} X_{i,t} \tag{4}$$

$X_{i,t}$ is the sum of boater traffic to invaded lakes M in the current time step divided by traffic to all

lakes L . Propagule pressure $Q_{j,t}$ is the sum of vector traffic from all sources to lake j in year t that had previously visited an invaded lake (Bossenbroek et al. 2001). This is important as it allows the proportion of traffic originating from invaded locations to change as more locations become invaded over time.

Boaters O_i are the number of houses on each block. House blocks were designated as single typical street block polylines segments in a Geographic Information Systems (GIS). There were 4,496 house blocks used in the model. The value of boaters on the block is equivalent to the number of buildings nearest to the segment (data from National Topographic Database; Natural Resources Canada; <http://geogratis.cgdi.gc.ca/geogratis/en/product/search.do?id=8147>).

D_{ij} describes the road distance between each source house block i and each lake j in the watershed. The functional relation D_{ij}^{-d} was chosen because studies have shown that there exists an exponentially decreasing probability of travel with increased distance (Buchan and Padilla 1999; Leung et al. 2004). Using distance between houses and lakes is more realistic than using distances between lake pairs because people are likely to travel home between lake trips. Here, because our study system is located in cottage country, we defined ‘home’ as a residence or cottage within the watershed. If we had exclusively used permanent residence locations as ‘homes,’ our distances would have been heavily skewed toward the metropolitan Toronto area (200 km away), and would have resulted in overestimating actual distances traveled between lake visits (pers. obs.; Buchan and Padilla 1999). People very likely stay in the 2 EB area when boating, rather than traveling to and from Toronto and the 2 EB area on consecutive days. Further, for D_{ij} , we used road distances because roads deviate; thus, using road distances provides more realistic estimates of travel distance than Euclidean distances, which have been employed by most past models (Bossenbroek et al. 2001; Leung et al. 2004; but see Muirhead 2007 and Prasad et al. 2010). The road that terminated closest to each lake, as determined in the GIS, was used as the end point location when calculating the distance along a road between each house and that lake.

We used lake size (surface area in hectares) as a surrogate for attractiveness W_j (Bossenbroek et al. 2001; Leung et al. 2006). Because of the proximity of

the watershed to Georgian Bay on Lake Huron, which is invaded, we included boat launches along Georgian Bay within the boundaries of 2 EB as ‘lakes’ in our system. In Eq. 1, W_j for these locations was replaced with the parameter g , which was fitted to our model. g was used because these boat launches do not have surface area comparable to the inland lakes (see Leung et al. 2006 for details on g). Data on boat launch locations were digitized into GIS from backcountry maps (Marleau 2006). Data on lake locations and sizes and road networks were obtained from the National Topographic Database GIS data.

Estimating propagule pressure due to stream connections

While previous studies have looked at increased risk due to stream connections secondarily for a subset of known invaded lakes (e.g., Bobeldyk et al. 2005; Johnson et al. 2006; Jacobs and MacIsaac 2009), to our knowledge, this is the first attempt at a system-wide analysis of propagule pressure due to stream connections. The metric measure of propagule pressure S to each lake j in each year t is described by:

$$S_{j,t} = \sum_{n=1}^N Z_n L_{nj}^{-r} \quad (5)$$

where $S_{j,t}$ results from the sum of connections from all N upstream lakes. Propagule pressure should depend on the total water flow from the upstream lake to the downstream lake, as well as the distance along the stream propagules have to travel. We assume that propagule pressure is inversely related to the distance L_{nj} from each upstream invaded lake n to lake j because as distance increases, the probability of survival during transport decreases exponentially (Bobeldyk et al. 2005); r is a fitted coefficient describing the rate of decay in probability of survival. We used watershed area Z_n of the upstream lake n as a surrogate of water flow to downstream lake j , assuming precipitation is similar across the entire watershed. If lake j is hydrologically isolated or not connected to any invaded lakes, it will have $S_{j,t} = 0$.

Data on stream connections and elevation were obtained from the National Topographic Database and the Ontario Ministry of Natural Resources. In order to obtain the distance of the shortest stream

connection between all connected lake pairs, we developed a recursion algorithm. Elevation was used to calculate the watershed area of each lake.

Linking propagule pressure to probability of establishment

Next, we linked propagule pressure estimates in probability of establishment models. We used gravity and flow scores for each lake in each year to calculate probability of establishment over time. We compared a random null model, a boater only model, a streams only model, and a combined boater/stream model.

Generally, the probability of establishment (E) can be described mathematically as the complement of the probability of all propagules failing to establish. To begin, we consider probability of establishment for a null model ($E(null)$):

$$E(null) = 1 - e^{-\gamma^c} \tag{6}$$

where γ^c is a fitted parameter that determines the base rate probability of establishment, where invasions occur randomly throughout the system, with equal probability. We use this functional form to most clearly illustrate the relation to the three more complex models below (Eqs. 7–9).

The probability of establishment for the boat model is given by:

$$E(Q_{j,t}) = 1 - e^{-(\alpha Q_{j,t} + \gamma)^c} \tag{7}$$

where $Q_{j,t}$ is the propagule pressure due to boating traffic to lake j in year t (Eq. 4). α , γ and c are fitted coefficients. c is used to model Allee effects, which result in a disproportionately low probability of establishment at low propagule pressure (Dennis 2002; Leung et al. 2004). While Allee effects have not been studied extensively in *Bythotrephes*, they are plausible because *Bythotrephes* need to find mates during sexual reproduction. Further, Drake (2004) found some evidence for Allee effects in *Bythotrephes*; the functional form of his model is the same as that used here.

The stream model can be described in terms equivalent to the boater model:

$$E(S_{j,t}) = 1 - e^{-(\beta S_{j,t} + \gamma)^c} \tag{8}$$

where propagule pressure is due to the cumulative flow from invaded upstream lakes is described by $S_{j,t}$ (Eq. 5). β , γ and c are fitted coefficients.

The probability of establishment for the combination model is:

$$E(Q_{j,t}, S_{j,t}) = 1 - e^{-(\alpha Q_{j,t} + \beta S_{j,t} + \gamma)^c} \tag{9}$$

where coefficients are fitted simultaneously. We can directly assess the relative contribution of each vector by converting and scaling gravity scores ($Q_{j,t}$) and flow scores ($S_{j,t}$) into inputs for probability of establishment estimates ($\alpha Q_{j,t}$ and $\beta S_{j,t}$) in the combined model. Comparison of model predictions (Eqs. 6–9) to invasion data provides evidence for the relative importance of each pathway: boater and stream vectors.

Parameterizing models: managing incomplete spatial and temporal data

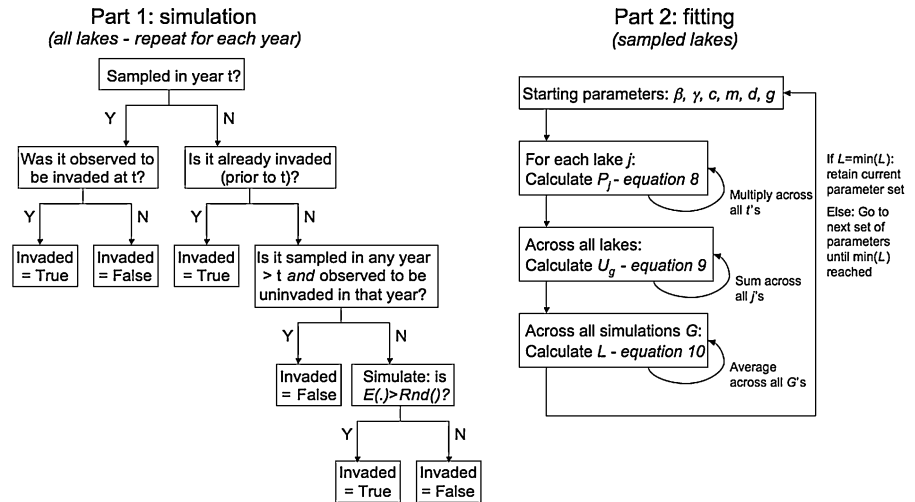
The next step is to use the available data on *Bythotrephes* invasions to fit each probability of establishment model (Eqs. 6–9). We extend theoretical techniques developed by Leung and Delaney (2006) to manage incomplete spatial invasion data and extend these techniques to accommodate incomplete temporal data. Sensitivity analyses were conducted in Leung and Delaney (2006).

The approach itself is straightforward and is comprised of a two part recipe (Fig. 1). In Part 1, for all locations and times where we do not have information, we simulate the invasion process (Eqs. 1–9; Fig. 1a). For locations and time intervals where we have empirical data on the invasion status (invaded or uninvaded), rather than simulating the process, we constrain those lakes to adopt their empirical invasion status in order to use the available information to the maximum extent possible (Leung and Delaney 2006). In Part 2, using only the sampled lakes, we derive the appropriate probability model and use maximum likelihood to find the best set of parameter values that fit the empirical data on *Bythotrephes* invasions (Fig. 1b). Specifically:

Part 1: recipe for simulating invasion patterns (Fig. 1a)

Step (1) let unsampled locations become invaded if $E(\cdot)_{j,t} > Rnd$ (where $E(\cdot)_{j,t}$ is the probability of establishment for lake j and time t for each alternative model— $E(null)$, $E(Q_{j,t})$, $E(S_{j,t})$ and $E(Q_{j,t}, S_{j,t})$;

Fig. 1 Schematic diagram describing the recipe for estimating spatial and temporal patterns of invasion in a system with limited data. In Part 1, we simulate invasions to all lakes following the rules laid out. In Part 2, we calibrate the model on sampled lakes using maximum likelihood techniques to find the best fitting parameter values. Steps, variable names and equations are described in greater detail in the main text



Eqs. 6–9, and Rnd is a random number between 0 and 1). Once invasion occurs, assume that recovery does not occur (i.e., invaded lake remains invaded).

Step (2) in year t , let lakes be uninvaded for sampled lakes that are known to be uninvaded in year t or any years after year t . Likewise, let lakes be invaded for sampled lakes that are known to be invaded in or before year t . For sampled lakes with incomplete temporal data on invasion status, apply step 1 in years for which no data. If the lake becomes invaded, let it remain invaded for all subsequent time steps.

Step (3) repeat steps 1 and 2 over all years.

Part 2: recipe for choosing the best fitting model (Fig. 1b)

We use maximum likelihood techniques to determine what parameter values ($\alpha, \beta, \gamma, c, m, d, g$, as appropriate for each model—Eqs. 1–9) best fit our observed data:

Step (1) for each sampled lake j , calculate P_j , the likelihood of observing the temporal and spatial pattern of lakes becoming invaded:

$$P_j = \left[\prod_{t=1}^{T'} (1 - E(\cdot)_{j,t}) \right] * \left[1 - \prod_{t=T'+1}^T (1 - E(\cdot)_{j,t}) \right] \tag{10}$$

where $E(\cdot)_{j,t}$ is defined above in Part 1, T is the year in which lake j was observed to be invaded (discovery year) and T' is the last year in which lake j was

observed to be uninvaded (counting from the initial invasion date in the system—1989 for *Bythotrephes* in the 2 EB watershed). For lakes that had never been sampled prior to being observed to be invaded, set $T' = 0$ (i.e., $P_j = 1 - \prod_{t=1}^T (1 - E(\cdot)_{j,t})$). For lakes that are sampled yearly (i.e., discovery date = invasion date), set $T' = T - 1$. For lakes that remain uninvaded, set $T' = T =$ last year of the study (i.e., $P_j = \prod_{t=1}^{T'} (1 - E(\cdot)_{j,t})$). The likelihood of observing the pattern of invasion for lake j (P_j) is simply the probability of lake j remaining uninvaded up until year T' (equal to the probability of failing to establish) and becoming invaded some unknown time between the intervals $T' + 1$ to T (the complement of the probability of failing to establish anytime during this interval). Equation 10 best incorporates the maximum amount of information available, properly modeling incomplete temporal data and incomplete spatial data, because $E(\cdot)_{j,t}$ includes propagule pressure in all time intervals from all sampled lakes, as well as all unsampled lakes simulated in Part 1.

Step (2) across all sampled lakes, calculate the system-wide log-likelihood of obtaining the pattern of invasion (U_g), for a given set of parameter values:

$$U_g = \sum_{j=1}^J \ln(P_j) \tag{11}$$

Step (3) for each set of parameter values, repeat Part 1 and Part 2: steps 1 and 2 G times to take into account stochasticity. Calculate the average negative

log-likelihood (L) across G simulations. In our study, we used $G = 1,000$:

$$L = -\frac{1}{G} \sum_{g=1}^G U_g \quad (12)$$

Step (4) apply maximum likelihood using Eq. 12, to find the best fitting parameter sets for each model (i.e., the lowest minimum negative log-likelihood score ($\min(L)$)).

Analyses

We randomly separated our dataset into 2 sections. We used 286 lakes to build the models and withheld 50 lakes for validation. These 50 lakes are analogous to unsampled lakes and were not used in the fitting procedure (Part 2 above). We repeated this procedure for 100 different calibration and validation sets. Such repetition was done to provide us with confidence that our model fit and predictions held. We assessed our ability to correctly predict invasions on these validation sets by calculating the area under the Receiver Operating Characteristic curve (AUC), a technique which is increasing in popularity for use in validating models of biological invasions (e.g., Jacobs and MacIsaac 2009). An AUC value significantly greater than 0.5 means that the model is more predictive than random.

Next, we examined the relative importance of the two vectors of spread. We compared the predictive ability of each model (null, boater, streams, combined) using the Akaike Information Criterion (AIC). In the combination model, we compared the relative importance of each vector to invasion risk by assessing the average amount of propagule pressure from each vector ($\alpha Q_{j,t}$ and $\beta S_{j,t}$) contributing to the probability of establishment across all lakes.

Finally, we compared the rate of new invasions occurring in the model results to the empirical discovery rate. The discovery rate was calculated as the number of discovered invasions in each year for sampled locations. The invasion rate was calculated using the number of invasions in each year averaged across G simulations, for the best fitting model. We considered the invasion rate of all lakes in the system (sampled or not).

Results

The AIC for the boater model alone was 119.67, 123.70 for the combination model, 173.05 for the

streams model and 168.39 for the null model. The streams model alone was unable to predict invasions better than the null model. In the combination model, the average relative contribution of boating traffic and stream connections to probability of establishment is 99.75 to 0.25%, respectively. Given the AIC scores, the boater model was the best fitting model. Thus, the results discussed below are for the boater model only.

Mean parameter values across the 100 model runs were $\alpha = 0.00105$, $\gamma = 4.89 \times 10^{-5}$, $c = 1.48$, $d = 1.27$ and $g = 955$. The average AUC for the 100 validation sets was 0.89 (Fig. 2). This corresponds to 65% of detection of invasions within only 10 sites at risk. We obtained an AUC value greater than 0.80 91% of the time, and an AUC value >0.90 52% of the time. The knowledge of invasions varied in each validation set—the average number of invaded lakes was 8, with a range of 2–15.

The model identified the patterns of relative risk of invasion to all unsampled lakes in the watershed (example from one validation set in Fig. 3). Lakes that became invaded more frequently, out of 1,000 simulations of the stochastic invasion process, were lakes that were larger, and close to the five largest lakes in the watershed and highly populated areas. Because the dominant lakes were all among the first invaded in the watershed, the patterns of propagule flow, probability of establishment and invasion over time ($E(Q_{j,t})$) did not change significantly once the largest lakes had become invaded (Fig. 4). The probabilities of invasion differed substantially between lakes, ranging from 0.00006 to 1. Most had low probabilities (e.g., 1,427 lakes had less than 0.3% chance of establishment) and a handful had moderate to high probabilities (e.g., 45 lakes had greater than 3.5% chance of establishment).

The inferred rate of invasion differed from the discovery rate, with more lakes being invaded earlier and a logistic shape describing the number of lakes invaded over time (Fig. 5). This logistic curve depicts a peak in the invasion rate followed by a deceleration. On the other hand, subsequent to an initial period of gradual increase in discoveries occurring during 1989–2003, the discovery rate showed an exponential increase in the number of invasions over time after 2003.

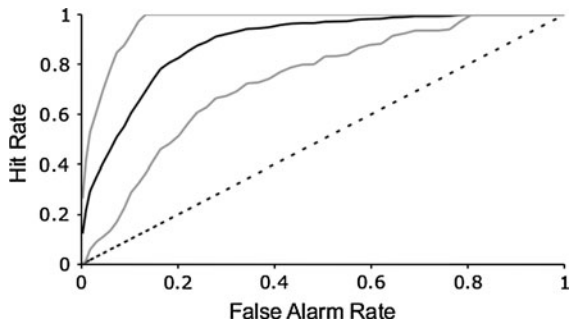


Fig. 2 Receiver operating characteristic curve showing the average curve across 100 validation sets. 95% CI are shown. The average AUC value is 0.89. The *dashed line* shows the expected relation under a random scenario

Discussion

Predicting and validating patterns of invasion

Predicting how invaders will spread in space and time is a central goal of management strategies aimed at controlling the spread and impact of NIS (Lodge et al. 2006). We showed that our model was highly predictive and consistent across different validation sets of lakes that were withheld from model calibration (average AUC = 0.89). Although other gravity models have been able to provide a good fit to the data (Bossenbroek et al. 2001; 2007; Muirhead 2007), the accuracy of those studies may be overestimated as they only reported fits to the data used to generate the model rather than a separate validation set (Fielding and Bell 1997; Leung et al. 2004). AUC is becoming a standard measure and allows comparability with other work. An AUC of 0.89 is respectable; this corresponds to an average invasion detection of 65% within 10 sites at risk, or 4 times better than random ($8/50 = 0.16\%$). A 10 site analysis is stringent, representing a relatively small sample, predicted from a model built with incomplete knowledge, which does not take into account the fact that more than 10 sites may be at risk.

With our model, managers can begin prioritizing lakes and pathways for management. First, our model suggested that boater movement was far more likely than stream connections to explain the *Bythotrephes* invasion pattern. The contribution to probability of establishment from streams was only 0.25%. In contrast, past studies on zebra mussels have found that stream connections increase invasion risk to

downstream lakes substantially (Bobeldyk et al. 2005; Johnson et al. 2006); however, invaded downstream lakes were very close to invaded upstream lakes (<1 km). Additionally, it is possible that propagule transport by boaters confounded these results (Bobeldyk et al. 2005). Therefore, stream connections may either be unimportant for *Bythotrephes* spread or increase risk to only a small subset of very proximal lakes. In many ways, this is fortuitous, as the boater pathway can arguably more easily be managed, for example, by restricting boater traffic or offering boat cleaning services on appropriate lakes.

Second, probability of establishment differed greatly between lakes, ranging from 0.00006 to 1, allowing lakes to be prioritized for management. Lakes observed with high probability (e.g., $P_j = 0.05$) were almost certain to have become invaded over a 20 year invasion process. Still, it is relevant to consider even lakes with relatively low probabilities of invasion, as the risk of invasion can become substantial over time. For instance, lakes with even a 3.5% chance of invasion in a single year (45 lakes in year 20), have a 50% chance of having become invaded over 20 years. In contrast, the majority of lakes (1,427 in year 20), which had <0.3% chance of invasion annually, would be still unlikely to be invaded (<6%). Thus, even small absolute differences in annual probability can become meaningful over time.

Our model is also useful for estimating the invasion rate. Despite the fact that one might predict an accelerating rate of *Bythotrephes* invasion in the watershed based on discovery rates, our model actually suggests that the invasion rate is decelerating. Invasions to lakes with high propagule inputs likely occurred and saturated quickly, but subsequently, invasion rates decelerated as lakes with moderate propagule inputs became invaded more slowly. This trend subscribes to Johnson et al. (2006)'s observation of a decelerating rate of new invasions occurring for zebra mussels. The trend also should caution us in directly interpreting invasion rate from discovery rate, an interpretation which is commonly done (Solow and Costello 2004; Ricciardi 2006). Discovery rate can be confounded by other variables such as sampling effort (Costello and Solow 2003), and while corrections do exist (e.g., catch-per-unit-effort approach; Loo et al. 2007), they are rarely done because obtaining data on actual effort across

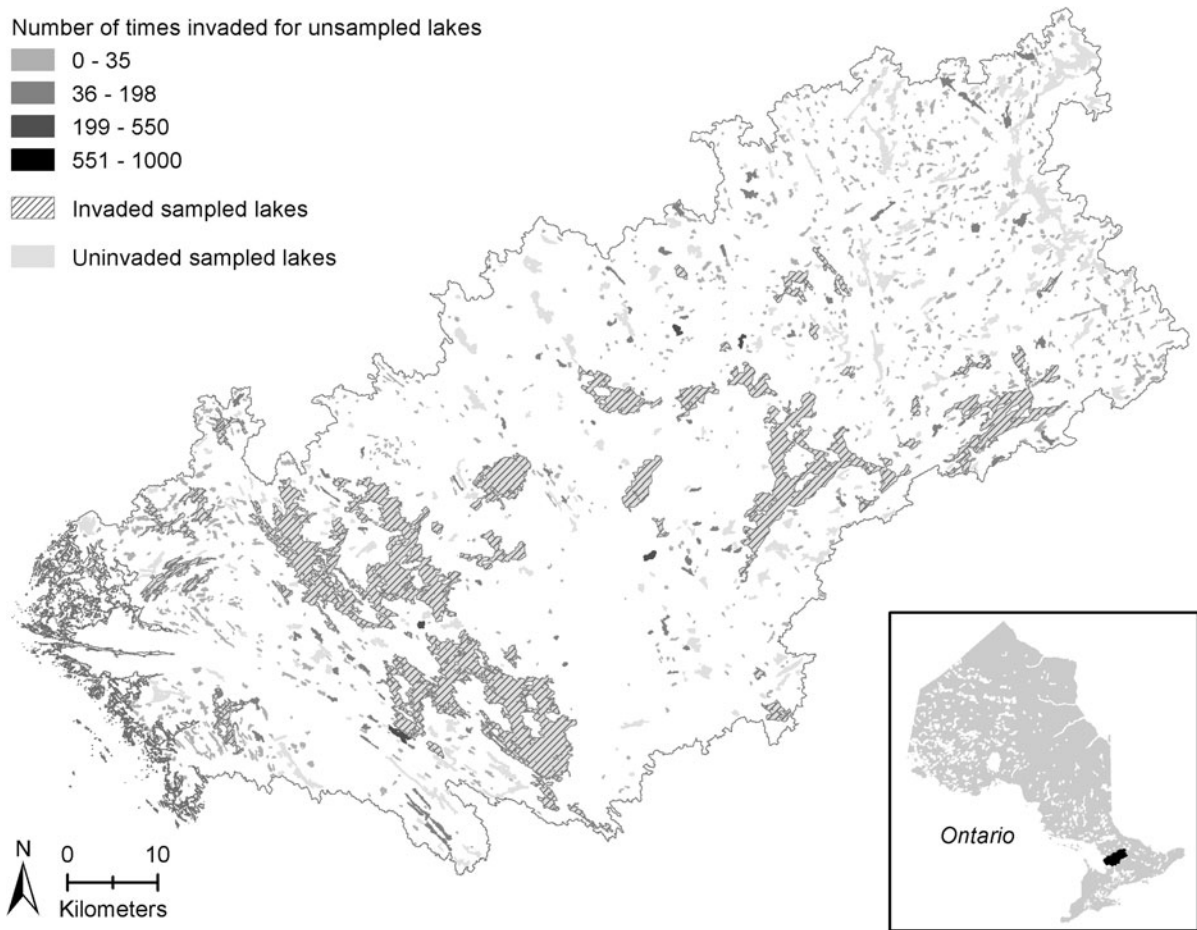


Fig. 3 Map of lakes in the 2 EB watershed with *inset map* showing the location of the watershed in Ontario. *Solid/stripes* denote status of the lake (uninvaded/invaded) and the gradient

scale represents the number of times out of 1,000 simulations for one model run that unsampled lakes in the system became invaded by the final year of the invasion process

years is often difficult (e.g., unpublished data or surveys conducted by disparate groups).

Contributions of model framework

We developed a model framework that incorporates several levels of uncertainty—unsampled locations and time intervals, and multiple vectors. We showed that we can incorporate missing locations and predict their invasion status, and incorporate unsampled time intervals, and hindcast and predict the most probable invasion dates, as discussed above. Finally, we demonstrated how multiple vectors can be incorporated and assessed. Past propagule pressure models have generally only focused on one vector of spread

at a system-wide scale (e.g., Bossenbroek et al. 2001; Leung et al. 2004; Muirhead and Macisaac 2005; but see Muirhead 2007), in static time (Bossenbroek et al. 2001; Jacobs and MacIsaac 2009; but see Leung et al. 2004), and with full knowledge or ‘full knowledge’ that excludes unsampled locations (Bossenbroek et al. 2001; Leung et al. 2004; Muirhead 2007; but see Leung and Delaney 2006 for theoretical techniques). The development of techniques that incorporate these levels of uncertainty into a single spread model should be highly useful to future risk assessments for NIS.

In the application to *Bythotrephes*, we examined recreational boating traffic and streams; however, the model framework allows for the inclusion of other

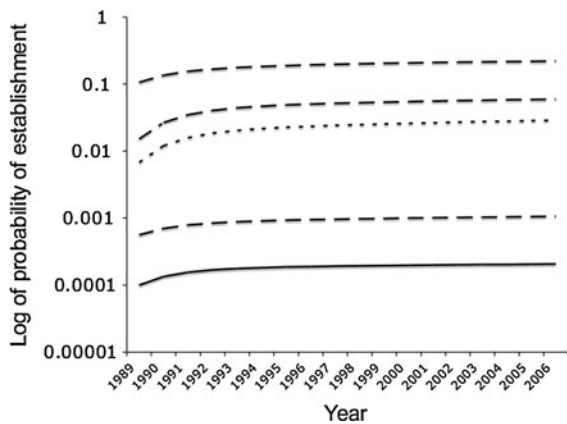


Fig. 4 Changes in the probability of establishment ($E(Q_{j,t})$) over time in the 2 EB system. Data are \log_{10} transformed. Five lakes, represented by different line styles, are shown for simplicity. Lakes shown have probability of establishment in 2006 of 0.62, 0.22, 0.06, 0.028 and 0.001—showing lakes across a range of risk—high, high, moderate, low, low, respectively. In the first few years of the invasion process (1989–1994) when the five largest lakes in the system became invaded, the probability of establishment to all lakes increased, albeit at different rates due to differences in size and location in the watershed. Subsequently, the probabilities of establishment for all lakes did not change very much with respect to the initial pulse

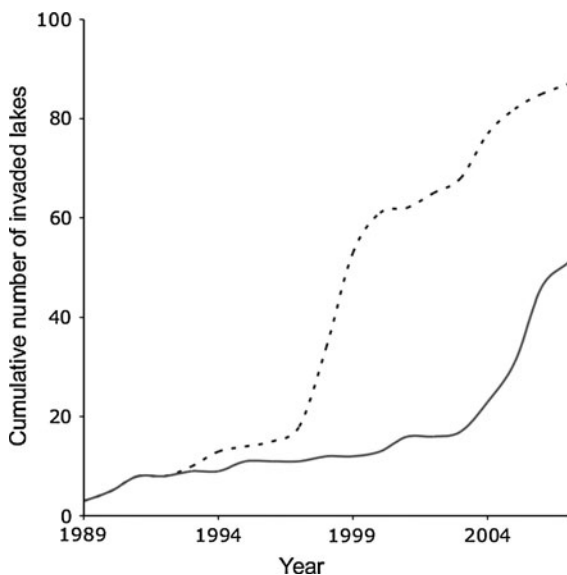


Fig. 5 The discovery rate of invasions (solid line) for sampled lakes and the invasion rate (dotted line) for all lakes predicted to be invaded in the watershed by our model

possible vectors, such as wind, waterfowl and other modes of human transport. As a first step, one needs to develop propagule pressure flows to all locations

based on metrics of movement—for example, how waterfowl move between invaded and uninvaded lakes. These flows can then be incorporated into the probability of establishment equations (Eqs. 6–9) in an additive manner with the use of scalars such as α or β , which relate propagule pressure to actual propagule viability and probability of establishment.

Here, we were able to analyze the relative contribution of vectors because the underlying basis of our model links metrics of propagule pressure to probability of establishment. Further, using probabilities allows us to connect more directly to risk analyses and management goals. It is crucial to understand that depending on the functional relation between propagule pressure and probability of establishment, a given reduction in propagule pressure will have different outcomes. Importantly, the results can be integrated with quantitative risk analyses (e.g., Hyder et al. 2008) to formally quantify the marginal gains of management actions that reduce propagule pressure.

Future studies could develop Bayesian extensions to model the uncertainty distribution in predictions of risk, or develop a joint propagule pressure–environment model for *Bythotrephes* (e.g., Herborg et al. 2007; Leung and Mandrak 2007). The characteristics of the recipient environment may modify the shape of the probability of establishment curve, providing additional information on the relative risks to locations (Levine et al. 2004). While we did not have access to environmental data, in systems where such data are available, techniques could be developed to interpolate relevant data from similar nearby locations to provide information on unsampled locations.

In summary, using our framework, we demonstrated that in order to predict the secondary spread of NIS with high power, we do not require extensive information. We can make the best use of the information we have and incorporate uncertainties in vectors of dispersal, individual lake invasion status and timing of invasions into predictive models. Further, we can link metrics of propagule pressure to probability of establishment in order to investigate changes in the system over time, and to estimate the marginal gains of management action. We applied these techniques to *Bythotrephes* spread in a popular watershed in Ontario, although the techniques can be applied to other NIS and systems where similar uncertainties exist. Overall, our framework enables us

to predict the secondary spread of NIS, and ultimately provide managers with information on which vectors and locations to focus management efforts.

Acknowledgments We thank C. Brisson, A. Irwin and J. Proville for assistance with data collection and processing. Thank you to N. Yan and the 300-lake survey team for lake data and the Ontario Ministry of Natural Resources for GIS data. We thank C. Chivers, P. Edwards and S. Kulhanek for helpful comments on the manuscript. This work was funded by the Canadian Aquatic Invasive Species Network, through grants awarded to Dr. B. Leung.

References

- Anderson RP, Martinez-Meyer E (2004) Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biol Conserv* 116:167–179
- Bobeldyk AM, Bossenbroek JM, Evans-White MA, Lodge DM, Lamberti GA (2005) Secondary spread of zebra mussels (*Dreissena polymorpha*) in coupled lake-stream systems. *Ecoscience* 12:339–346
- Bossenbroek JM, Kraft CE, Nekola JC (2001) Prediction of long-distance dispersal using gravity models: Zebra mussel invasion of inland lakes. *Ecol Appl* 11:1778–1788
- Bossenbroek JM, Johnson LE, Peters B, Lodge DM (2007) Forecasting the expansion of zebra mussels in the United States. *Conserv Biol* 21:800–810
- Boudreau SA, Yan ND (2004) Auditing the accuracy of a volunteer-based surveillance program for an aquatic invader *Bythotrephes*. *Environ Monit Assess* 91:17–26
- Buchan LAJ, Padilla DK (1999) Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecol Appl* 9:254–265
- Bur MT, Klarer DM, Krieger KA (1986) 1st Records of a European Cladoeran, *Bythotrephes-Cederstroemi*, in Lakes Erie and Huron. *J Great Lakes Res* 12:144–146
- Cairns A, Yan N, Weisz E, Petruniak J, Hoare J (2007) Operationalizing CAISN project 1.V technical report: the large, inland lake, *Bythotrephes* survey—limnology, database design, and presence of *Bythotrephes* in 311 south-central Ontario lakes, Canadian Aquatic Invading Species Network technical report
- Costello CJ, Solow AR (2003) On the pattern of discovery of introduced species. *Proc Natl Acad Sci USA* 100:3321–3323
- Dennis B (2002) Allee effects in stochastic populations. *Oikos* 96:389–401
- Drake JM (2004) Allee effects and the risk of biological invasion. *Risk Anal* 24:795–802
- Drake JM, Lodge DM (2006) Allee effects, propagule pressure and the probability of establishment: Risk analysis for biological invasions. *Biol Invasions* 8:365–375
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49
- Hastings A, Cuddington D, Davies KF, Dugaw CJ, Elmendorf S, Freestone A, Harrison A, Holland M, Lambrinos J, Malvadkar U, Melbourne BA, Moore K, Taylor C, Thomson D (2005) The spatial spread of invasions: new developments in theory and evidence. *Ecol Lett* 8:91–101
- Herborg LM, Jerde CL, Lodge DM, Ruiz GM, MacIsaac HJ (2007) Predicting invasion risk using measures of introduction effort and environmental niche models. *Ecol Appl* 17:663–674
- Hyder A, Leung B, Miao ZW (2008) Integrating Data, Biology, and Decision Models for Invasive Species Management: Application to Leafy Spurge (*Euphorbia esula*). *Ecol Soc* 13:16
- Jacobs MJ, MacIsaac HJ (2007) Fouling of fishing line by the waterflea *Cercopagis pengoi*: a mechanism of human-mediated dispersal of zooplankton? *Hydrobiologia* 583:119–126
- Jacobs MJ, MacIsaac HJ (2009) Modelling spread of the invasive macrophyte *Cabomba caroliniana*. *Freshwat Biol* 54:296–305
- Jerde CL, Lewis MA (2007) Waiting for invasions: a framework for the arrival of nonindigenous species. *Am Nat* 170:1–9
- Johnson LE, Padilla DK (1996) Geographic spread of exotic species: ecological lessons and opportunities from the invasion of the zebra mussel *Dreissena polymorpha*. *Biol Conserv* 78:23–33
- Johnson LE, Bossenbroek JM, Kraft CE (2006) Patterns and pathways in the post-establishment spread of non-indigenous aquatic species: the slowing invasion of North American inland lakes by the zebra mussel. *Biol Invasions* 8:475–489
- Kolar CS, Lodge DM (2002) Ecological predictions and risk assessment for alien fishes in North America. *Science* 298:1233–1236
- Leung B, Delaney DG (2006) Managing sparse data in biological invasions: a simulation study. *Ecol Model* 198:229–239
- Leung B, Mandrak NE (2007) The risk of establishment of aquatic invasive species: joining invasibility and propagule pressure. *Proc R Soc Lond [Biol]* 274:2603–2609
- Leung B, Drake JM, Lodge DM (2004) Predicting invasions: propagule pressure and the gravity of Allee effects. *Ecology* 85:1651–1660
- Leung B, Bossenbroek JM, Lodge DM (2006) Boats, pathways, and aquatic biological invasions: estimating dispersal potential with gravity models. *Biol Invasions* 8:241–254
- Levine JM, Adler PB, Yelenik SG (2004) A meta-analysis of biotic resistance to exotic plant invasions. *Ecol Lett* 7:975–989
- Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends Ecol Evol* 20:223–228
- Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A (2006) Biological invasions: recommendations for US policy and management. *Ecol Appl* 16:2035–2054
- Loo SE, Keller RP, Leung B (2007) Freshwater invasions: using historical data to analyse spread. *Divers Distrib* 13:23–32

- MacIsaac HJ, Ketelaars HAM, Grigorovich IA, Ramcharan CW, Yan ND (2000) Modeling *Bythotrephes longimanus* invasions in the Great Lakes basin based on its European distribution. *Arch Hydrobiol* 149:1–21
- Marleau J (2006) Backroad mapbook: cottage Country Ontario. Mussio Ventures Ltd., Burnaby, BC
- Muirhead JR (2007) Forecasting dispersal of nonindigenous species. University of Windsor, Windsor, ON
- Muirhead JR, Macisaac HJ (2005) Development of inland lakes as hubs in an invasion network. *J Appl Ecol* 42:80–90
- Prasad AM, Iverson LR, Peters MP, Bossenbroek JM, Matthews SN, Sydnor TD, Schwartz MW (2010) Modeling the invasive emerald ash borer risk of spread using a spatially explicit cellular model. *Landscape Ecol* 25:353–369
- Reed-Andersen T, Bennett EM, Jorgensen BS, Lauster G, Lewis DB, Nowacek D, Riera JL, Sanderson BL, Stedman R (2000) Distribution of recreational boating across lakes: do landscape variables affect recreational use? *Freshwat Biol* 43:439–448
- Ricciardi A (2006) Patterns of invasion in the Laurentian Great Lakes in relation to changes in vector activity. *Divers Distrib* 12:425–433
- Ricciardi A, Rasmussen JB (1998) Predicting the identity and impact of future biological invaders: a priority for aquatic resource management. *Can J Fish Aquat Sci* 55:1759–1765
- Schneider DW, Ellis CD, Cummings KS (1998) A transportation model assessment of the risk to native mussel communities from zebra mussel spread. *Conserv Biol* 12:788–800
- Solow AR, Costello CJ (2004) Estimating the rate of species introductions from the discovery record. *Ecology* 85:1822–1825
- Weisz EJ, Yan ND (2010) Relative value of limnological, geographic, and human use variables as predictions of the presence of *Bythotrephes longimanus* in Canadian Shield lakes. *Can J Fish Aquat Sci* 67:462–472
- Yan ND, Pawson TW (1997) Changes in the crustacean zooplankton community of Harp Lake, Canada, following invasion by *Bythotrephes cederstroemi*. *Freshwat Biol* 37:409–425
- Yan ND, Dunlop WI, Pawson TW, Mackay LE (1992) *Bythotrephes-Cederstroemi* (Schoedler) in Muskoka Lakes—1st Records of the European Invader in Inland Lakes in Canada. *Can J Fish Aquat Sci* 49:422–426