Modelling spread of the invasive macrophyte *Cabomba caroliniana*

MICHAEL J. JACOBS AND HUGH J. MACISAAC
Great Lakes Institute for Environmental Research, University of Windsor, Windsor, ON, Canada N9B 3P4

SUMMARY

1. Predicting spread of non-indigenous species requires an understanding of where propagules are being transported, and whether these propagules can survive in the novel habitat and successfully integrate into the recipient community. In this study, we model potential spread of invading *Cabomba caroliniana* in Ontario, Canada, using a combination of passive and active dispersal models coupled with an environmental suitability model, thereby considering the first two stages of the invasion process.

2. Measures of propagule pressure incorporated both human-mediated dispersal via trailered boats, and advective flow from invaded to non-invaded systems, while habitat suitability was forecasted by combining native and global data sets and using boosted regression trees.

3. Risk of invasion differed depending on the combination of approaches used and the time period considered. Three lakes appear to be at greatest risk owing to a combination of high boater and water movement from invaded sources, and high environmental suitability. The best predictors of lake suitability were pH, mean lake temperature and dissolved calcium concentration. Hundreds of lakes in Ontario may be suitable for establishment of *Cabomba*, highlighting the need for vector management.

Keywords: boosted regression trees, environmental niche modelling, gravity model, non-indigenous species, propagule pressure

Introduction

The invasion process may be conceptualized as a series of barriers that an introduced species must overcome before successfully invading a community. These limitations include geographic barriers that restrict or preclude the introduction of propagules, adverse ambient environmental conditions and biological constraints that may affect integration into the new community (Richardson *et al.*, 2000; Colautti, Grigorovich & MacIsaac, 2006).

A number of recent reviews have highlighted the importance of propagule pressure to the success of non-indigenous species (NIS; e.g. Lockwood, Cassey & Blackburn, 2005; Colautti *et al.*, 2006; Hayes & Barry, 2008). Propagule pressure refers to the number of inoculation events, the number of propagules introduced per event, and the condition of introduced propagules (Williamson & Fitter, 1996; Lockwood *et al.*, 2005). Propagule pressure can be very difficult to quantify, although often it can be at least semi-quantitatively assessed (e.g. Rouget & Richardson, 2003; Drake & Lodge, 2004; Herborg *et al.*, 2007). From a management context, assessments of propagule pressure provide the first step in identification of areas vulnerable to invasion. However, introduction effort can only inform where NIS are introduced, and thus potential rather than actual distribution (Rouget & Richardson, 2003).

A complementary approach seeks to identify areas vulnerable to invasion based upon environmental suitability (e.g. Thuiller *et al.*, 2005). One increasingly popular approach by which this method is applied is through application of machine-based learning.
Predicting spread of invasive species

Dispersal of Cabomba

Gravity models have been utilized to predict spread of numerous aquatic NIS (see Muirhead & MacIsaac, 2005). Gravity models link invaded sources with non-invaded destinations frequented by human vectors. Currently, the Kasshabog Lake system is the only area where C. caroliniana occurs in Ontario. In total, Cabomba has invaded a private lake (South Lake) along with the river that connects it to Kasshabog Lake. We conducted a survey at both launch sites located on Kasshabog Lake (source) to measure boater movement to regional non-invaded lakes (destinations). We assume that boaters inadvertently disperse the species to other lakes via viable fouled plants on trailered boats (see Johnson, Ricciardi & Carlton, 2001). The survey was conducted from August to September 2006, and included 41 boaters who reported taking trailered boats from Kasshabog Lake to other lakes. An origin-specific version of a production-constrained gravity model equation was utilized to measure the potential human-mediated spread (Haynes & Fotheringham, 1984). For the purpose of this model, other potential sources of propagules, such as aquarium stores, were ignored (Cohen, Mirochnick & Leung, 2007). Potential risk of spread was assessed as follows (Haynes & Fotheringham, 1984):

$$T_j = \frac{AO_j w_j}{D(j)}$$

where \(T_j\) is the interaction between Kasshabog Lake and lake \(j\), \(A\) the balancing factor to measure the relative location of Lake Kasshabog to the destinations, \(O_j\) the propulsive power of Kasshabog Lake to lake \(j\), \(D(j)\) the distance decay function applied to lake \(j\), and \(w_j\) the destination attractiveness of lake \(j\).

Information collected from the surveys identified lakes that interacted with Kasshabog Lake, along with the strength of those interactions, effectively measuring boater movement and frequency to other lakes (\(O_j\)). It was assumed that if a boater visited many lakes in the survey, then Kasshabog Lake was visited before travelling to another lake. Destination attractiveness was based on the product of lake area and sport-fish diversity (Minns, 1990). These two factors measure...
human attractiveness for each lake (Reed-Anderson et al., 2000). Lake area was calculated by Geographic Information System (GIS) and sport-fish diversity (presence data) was obtained from the Department of Fisheries and Oceans (DFO), Burlington, ON. A distance decay function, \( D(i) \), was calculated using logarithmic regression of probability of visited lakes as a function of distance from the source, separated into three intervals. All inland lakes were given an interaction value based upon eqn 1. Lakes Ontario and Erie had the greatest outflow and distances travelled from Kasshabog Lake, respectively (Fig. 1), but were not included in the model due to our focus on spread to inland lakes only.

A second model combined the gravity model and a hydrology model for those lakes identified as at risk in the former model. This model identified additional lakes at risk owing to their down-stream location from lakes likely to be invaded via boater movements. The model was based on the ‘water virtual flow-seamless provincial dataset’ created by the Ontario Ministry of Natural Resources, and mapped using GIS. The data set is a fully connected, flow-directed, stream network with complete topological flow structure identifying connectivity and flow direction. A geometric network was created from the line layer. Next, flow directions were set on each line segment based on digitized direction. Finally, the utility network analyst extension was used to allow for downstream tracing of advective flow from a source. This procedure identified downstream lakes connected to a source.

The possibility of advective spread was explored for four lakes with the highest probability of *Cabomba* introduction from the gravity model. For downstream lakes, the distance from the closest upstream source lake was calculated along the virtual flow line layer using GIS. Water flow data was obtained from the Water Survey of Canada (Burlington, ON) for flow stations. Dispersal rate was calculated using the initial location where *Cabomba* was first reported in 1991 to where it was established by the end of 2006. An exponential cumulative distribution function was used to calculate the probability of *Cabomba* entering a non-invaded lake downstream in \( t \) years. Models were developed to project the probability of establishment after 1, 2, 5 and 10 years. Assumptions implicit to development of the hydrology model include: (i) *Cabomba* fragments move downstream at the same rate as the current (i.e. they are neutrally buoyant); (ii) flow within lakes and connecting streams is uniform; (iii) the dispersal rate (km year\(^{-1}\)) is constant; and (iv) advective flow follows the virtual dataset.

**Environmental niche model**

Environmental niche models were constructed to assess lakes in Ontario that provide suitable environmental characteristics for the establishment of *Cabomba*. Nine water chemistry parameters were used to develop the environmental niche model: dissolved oxygen (mg L\(^{-1}\)), dissolved calcium (mg L\(^{-1}\)), pH, mean surface water temperature (°C), conductivity
(μS cm⁻¹), alkalinity (mg L⁻¹), total phosphorus (μg L⁻¹), ammonia (mg L⁻¹) and nitrate (μg L⁻¹). In some cases, incomplete data were available on these parameters, although boosted regression trees (BRT; see below) allow for missing values; a minimum of four parameters were used for all lakes.

Boosted regression trees were used to develop an environmental suitability model using lake water quality parameter data (Elith et al., 2006). BRT apply an iterative method that sums the weighted contribution of a successive chain of trees that are fit to the residuals from the previous tree (Friedman, 2001). The optimum number of iterations is reached when the residuals reach zero. The models were generated using the gbm package of R software with a training factor set at 0.70 assuming a Bernoulli distribution of presence/absence of *Cabomba* (Ridgeway, 2007). Model performance was evaluated based on the area under the receiver operating characteristic (ROC) curve (AUC), and was acceptable if significantly greater than 0.5 (random). Initially, we attempted to predict the global introduced distribution (minus Ontario) for the species using only environmental data from the native range. This model was developed using data for 96 lakes in Argentina, Paraguay, Uruguay and Brazil, but fit was quite poor (AUC = 0.591, 39.3%) due to its very large surface area, followed by Rice, Scugog and Pigeon Lakes.

Once the BRT model was complete, lakes were classified as invaded if the estimated probability was greater than or equal to a threshold value based upon the shape of the ROC curve. A threshold was chosen based on the minimum distance from the upper left corner to the curve (maximum fit; Liu et al., 2005). This method has been used successfully (Pearce & Ferrier, 2000; McPherson, Jetz & Rogers, 2004), despite a recent critique of the technique (Lobo, Jiménez-Valverde & Real, 2008). The R package verification was used to calculate the ROC curve. A 2 × 2 contingency table was constructed by pooling the number of lakes where *Cabomba* was predicted present/absent to actual present/absent data to explore model fit to Ontario.

A final model combined dispersal potential (gravity and hydrology model) with environmental suitability to provide a refined assessment of invasion risk after 1, 2, 5 and 10 years. Lake vulnerability to introduction was first identified from the gravity model. Next, lakes that were considered a source were utilized in the hydrologic component. The overall introduction likelihood of each identified lake was calculated as the sum of probabilities of the gravity and hydrology models, with a maximum of 1. Probability of establishment was then determined by multiplying the probability of introduction with the probability of environmental suitability from the BRT model (Jerde & Lewis, 2007):

\[
P(\text{establishment}) = (P(\text{gravity}) + P(\text{hydrology})) \times P(\text{BRT}).
\]

### Results

Survey results indicated that 23 lakes (excluding Lakes Erie and Ontario) were the recipients of recreational, trailered boats departing from Kasshabog Lake (Table 1). Most destination lakes were <100 km (road distance) from Kasshabog Lake (Fig. 1). A logarithmic distance decay function was calculated as \(D(j) = -0.61 \ln(j) + 0.69 (r^2 = 0.99)\). Lake Simcoe had the greatest interaction score, primarily due to its vast majority of these lakes did not contain *Cabomba*, nor were they expected to. The model correctly predicted (hit rate) two of the three waterbodies that *Cabomba* has invaded in Ontario. The invaded river was incorrectly predicted as non-invaded, while 12 additional lakes were incorrectly predicted (false

positives) to be invaded. A small number of lakes and rivers in northern Ontario were identified as having high environmental suitability, while a large cluster of lakes and rivers in south-eastern Ontario have a medium-to-high probability of environmental suitability (Fig. 4).

All 28 lakes identified at risk of introduction from the combined dispersal potential model (23 gravity + 5 hydrologic) were among the 468 lakes used in the environmental niche model, thus we were able to assign each a habitat suitability probability. After combining dispersal and environmental niche models, progressively more lakes were expected to become invaded as the time scale was extended from 1 to 10 years. This combined model suggests that Rice,

Table 1 Results of the origin-specific gravity model

<table>
<thead>
<tr>
<th>Lake name</th>
<th>Surface area (km²)</th>
<th>No. of trips</th>
<th>Sport-fish diversity</th>
<th>Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simcoe</td>
<td>722</td>
<td>16</td>
<td>8</td>
<td>56.70</td>
</tr>
<tr>
<td>Rice</td>
<td>92</td>
<td>29</td>
<td>8</td>
<td>14.40</td>
</tr>
<tr>
<td>Scugog</td>
<td>34</td>
<td>18</td>
<td>17</td>
<td>6.54</td>
</tr>
<tr>
<td>Pigeon</td>
<td>52</td>
<td>9</td>
<td>20</td>
<td>6.52</td>
</tr>
<tr>
<td>Chemong</td>
<td>25</td>
<td>9</td>
<td>16</td>
<td>2.57</td>
</tr>
<tr>
<td>Stony</td>
<td>35</td>
<td>5</td>
<td>16</td>
<td>2.35</td>
</tr>
<tr>
<td>French River</td>
<td>73</td>
<td>6</td>
<td>10</td>
<td>2.34</td>
</tr>
<tr>
<td>Buckhorn</td>
<td>32</td>
<td>7</td>
<td>13</td>
<td>2.04</td>
</tr>
<tr>
<td>Sturgeon</td>
<td>45</td>
<td>3</td>
<td>19</td>
<td>1.73</td>
</tr>
<tr>
<td>Round</td>
<td>6</td>
<td>19</td>
<td>12</td>
<td>1.10</td>
</tr>
<tr>
<td>Balsam</td>
<td>48</td>
<td>2</td>
<td>15</td>
<td>0.93</td>
</tr>
<tr>
<td>Chandos</td>
<td>16</td>
<td>4</td>
<td>11</td>
<td>0.55</td>
</tr>
<tr>
<td>Head</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>0.42</td>
</tr>
<tr>
<td>Jack</td>
<td>14</td>
<td>6</td>
<td>6</td>
<td>0.40</td>
</tr>
<tr>
<td>Mississauga</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>0.28</td>
</tr>
<tr>
<td>Belmont</td>
<td>8</td>
<td>3</td>
<td>13</td>
<td>0.27</td>
</tr>
<tr>
<td>Oak</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>0.26</td>
</tr>
<tr>
<td>Sparrow</td>
<td>11</td>
<td>2</td>
<td>16</td>
<td>0.21</td>
</tr>
<tr>
<td>Weslemkoon</td>
<td>20</td>
<td>1</td>
<td>12</td>
<td>0.15</td>
</tr>
<tr>
<td>Sandy</td>
<td>4</td>
<td>5</td>
<td>11</td>
<td>0.15</td>
</tr>
<tr>
<td>Clear</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>0.04</td>
</tr>
<tr>
<td>Katchewanooka</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0.02</td>
</tr>
<tr>
<td>Little</td>
<td>&lt;1</td>
<td>2</td>
<td>10</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

The interaction score (%) between Kasshabog Lake and a destination lake measures boater movement (possible human-mediated dispersal) of Cabomba. Surface area (calculated by GIS) and sport-fish diversity are measures of lake attraction and trips resemble the number of boaters leaving Kasshabog Lake to another lake. The interaction score was determined by combining these variables using eqn (1).

The interaction score (%) between Kasshabog Lake and a destination lake measures boater movement (possible human-mediated dispersal) of Cabomba. Surface area (calculated by GIS) and sport-fish diversity are measures of lake attraction and trips resemble the number of boaters leaving Kasshabog Lake to another lake. The interaction score was determined by combining these variables using eqn (1).

Table 2 Relative influence of environmental variables measured from the boosted regression trees model combining the native and global datasets to predict Cabomba presence in Ontario

<table>
<thead>
<tr>
<th>Variable</th>
<th>Relative influence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>39.9</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>18.3</td>
</tr>
<tr>
<td>Dissolved calcium (mg L⁻¹)</td>
<td>12.7</td>
</tr>
<tr>
<td>Conductivity (µS cm⁻¹)</td>
<td>9.6</td>
</tr>
<tr>
<td>Total phosphorus (µg L⁻¹)</td>
<td>7.9</td>
</tr>
<tr>
<td>Dissolved oxygen (mg L⁻¹)</td>
<td>6.1</td>
</tr>
<tr>
<td>Alkalinity (mg L⁻¹)</td>
<td>4.0</td>
</tr>
<tr>
<td>Ammonia (mg L⁻¹)</td>
<td>1.9</td>
</tr>
<tr>
<td>Nitrate (µg L⁻¹)</td>
<td>0.5</td>
</tr>
</tbody>
</table>

All 28 lakes identified at risk of introduction from the combined dispersal potential model (23 gravity + 5 hydrologic) were among the 468 lakes used in the environmental niche model, thus we were able to assign each a habitat suitability probability. After combining dispersal and environmental niche models, progressively more lakes were expected to become invaded as the time scale was extended from 1 to 10 years. This combined model suggests that Rice,
Scugog, Round and Crowe lakes have the greatest invasion risk in future years.

**Discussion**

Identifying which habitats are most vulnerable to biological invasion has preoccupied ecologists and managers for many years (Peterson, 2003; Richardson & Rejmánek, 2004; Theoharides & Dukes, 2007). Identification of relative invasion risk patterns should allow managers to prioritize management strategies to specific areas and/or to the vectors that transmit NIS to these areas (Basse & Plank, 2008). One approach that seems to warrant attention is the marriage of vector-based and environmental suitability models, as they appear to provide refined estimates of invasion risk (Herborg *et al.*, 2007; Jerde & Lewis, 2007). In this study, we utilized this approach by combining two vector-based models with an environmental suitability model to project spread of the macrophyte *C. caroliniana*.

The gravity model, which utilized boater surveys to assess potential human-mediated transport, identified four lakes at high invasion risk. These lakes (Simcoe, Rice, Scugog and Pigeon) ranked the highest when considering lake area, sport-fish diversity, distance and boater movement from Kasshabog Lake (Table 1). Lake Simcoe was about average with respect to trailered boat movement from Kasshabog Lake, although its overall gravity score (and hence risk) was very high because of its very large surface area (722 km²). On the other hand, Round Lake had greater boater inflow and is much closer to Kasshabog Lake than Lake Simcoe, but was ranked tenth owing to its smaller surface area (6 km²) (Table 4). Our analysis of boater movement from Kasshabog Lake excluded Lakes Erie and Ontario. The current *Cabomba* distribution in the U.S.A. borders both of these lakes. Because these lakes had the highest boater inflow from Kasshabog Lake, they may be vulnerable to *Cabomba* introduction. Although we believe that the survey provides an accurate representation of overall traffic out of Kasshabog Lake, more extensive sampling might pick up additional lakes placed at risk by outbound boaters. The survey was responsible for identifying lakes at risk; as a result, misclassification of destination lakes would also influence invasion risk of those located farther downstream (i.e. underestimate true risk).

To better measure introduction effort, we applied a hydrology model to gauge the invasion risk associated with passive movement of viable *Cabomba* fragments among connected lakes (Boylen *et al.*, 2006). The combination of gravity and hydrology models recognized five additional lakes at risk of introduction. The most at-risk lakes based upon the combined dispersal model were Simcoe, Rice, Scugog, Pigeon, Round, Buckhorn and Sturgeon. The first four lakes (all source lakes) had the highest gravity model scores, while the final three had the shortest downstream distance from an invaded or source lake (Fig. 2). *Cabomba* has already invaded the North River and South Lake, both of which are downstream from Kasshabog Lake.
The combined dispersal model differs from the advection only model in that a greater number of lakes were designated an introduction risk, accounting for human-mediated dispersal potential. Our study considered only boater movement and advection to predict risk of *Cabomba* spread, although...
Cohen et al. (2007) determined that the plant is sold extensively in pet stores, which would serve as another vector of introduction to lakes.

Our environmental niche model identified 12 lakes and rivers with high habitat suitability, suggesting that these systems are vulnerable to establishment of the species should it be introduced. If the threshold is altered (i.e. reduced) to that of the North River, which has a Cabomba population, then an additional 207 waterbodies were identified as suitable habitats. Thus, Cabomba may find numerous suitable habitats for establishment in Ontario unless introduction is prevented.

pH, temperature and dissolved calcium were identified as the best predictors of Cabomba presence in the boosted regression tree model (Table 2). Two rivers in Northern Ontario (51° and 52°N latitude) were predicted as suitable habitat, illustrating that Cabomba is not limited to tropical areas. This is consistent with the occurrence of the species in the Loosdrecht lakes, the Netherlands, which share similar latitude with Northern Ontario, even though climate differs between the areas (Schooler, Cabrera & Julien, 2008). These patterns indicate that temperature alone may not provide an accurate reflection of actual or potential occurrence of Cabomba (van der Heide et al., 2006).

Ecological integration into the recipient community is the final consideration in the stage-specific approach to identification of invasion risk (Colautti et al., 2006). Charles Elton (1958) identified a number of biological factors, notably competition and predation, which can hinder the invasion success of NIS. It is not clear whether competition is likely to impede spread of Cabomba in lakes with substantial introduction effort and high environmental suitability, as Capers et al. (2007) demonstrated that the species is capable of invading systems already populated by other macrophyte species.

Forecasting invasions is an important yet imprecise science (Guisan & Thuiller, 2005). The emergence of propagule pressure and climatic suitability as key predictors of invasion success hold promise for advances in predicting vulnerability of sites to NIS invasions (Richardson & Rejmanek, 2004; Lockwood et al., 2005). By combining models that embrace propagule pressure (Cohen et al., 2007) with those that encompass niche-based modelling (Thuiller et al., 2005), predictive power should be increased (Crossman & Bass, 2008). Other studies have coupled predictive models (Herborg et al., 2007; Jerde & Lewis, 2007), although the current study is the first to utilize both active (human-mediated) and passive (advective flow) movement of propagules with analyses of environmental suitability. Our study illustrated that different combinations of lakes were deemed most vulnerable to invasion when vector-based and environmental niche models were utilized. A final model that incorporated elements of both vector- and niche-based models was most similar to the results of vector-based model that incorporated both passive and active dispersal, and highly dissimilar to the predictions from the environmental niche model. With so few lakes presently invaded in Ontario, it will take some time before the accuracy of the different models developed here can be validated.

Acknowledgments

We are grateful to the following people for providing information used in constructing models:
References


(Manuscript accepted 10 August 2008)