Research Article

Assessing the probable distribution of the potentially invasive Chinese mystery snail, *Cipangopaludina chinensis*, in Nova Scotia using a random forest model approach

Sarah Kingsbury[1,*], Megan Fong[1], Donald F. McAlpine[2] and Linda Campbell[1,*]

[1]Saint Mary's University, Department of Environmental Sciences, 923 Robie St., Halifax, Nova Scotia, B3H 3C3, Canada

[2]New Brunswick Museum, 277 Douglas Avenue, Saint John, New Brunswick, E2K 1ES, Canada

Author e-mails: *sarahberry231@gmail.com (SK)*, *meganfong@dal.ca (MF)*, *donald.mcalpine@nbm-mnb.ca (MFM)*, *lm.campbell@smu.ca (LC)*

Corresponding authors

Abstract

Non-native species that become invasive threaten natural biodiversity and can lead to socioeconomic impacts. Prediction of invasive species distributions is important to prevent further spread and protect vulnerable habitats and species at risk (SAR) from future invasions. The Chinese mystery snail, *Cipangopaludina chinensis*, native to Eastern Asia, is a non-native, potentially invasive, freshwater snail now widely established across North America, Belgium, and The Netherlands. This species was first reported in Nova Scotia, eastern Canada in 1955, but was not found to be established until the 1990s and now exists at high densities in several urban lakes. Nonetheless, the presence and potential distribution of this species in Nova Scotia remains unknown. Limited resources make it difficult to do a broad survey of freshwater lakes in Nova Scotia, however a species distribution probability model has the potential to direct focus to priority areas. We apply a random forest model in tandem with a combination of water quality, fish community, anthropogenic water use, and geomorphological data to predict *C. chinensis* habitat in Nova Scotia (NS), Canada. All predicted probabilities of suitable *C. chinensisi* habitats in Nova Scotia were > 50% and include Cape Breton Island, the Nova Scotia-New Brunswick border, and the Halifax Regional Municipality. Suitable habitats predicted for *C. chinensis* overlap with many SAR habitats, most notably brook floater mussel, *Alasmidonta varicosa*, and yellow lampmussel, *Lampsila cariosa*. Our results indicate that *C. chinensis* could become widespread throughout NS, appearing first in the aforementioned areas of highest probability. Further research is required to test *C. chinensis* ecological thresholds in order to improve the accuracy of future species distribution and habitat models, and to determine *C. chinensis* impacts on native freshwater mussel populations of conservation concern.

Key words: freshwater, mollusc, species distribution, species at risk, statistical model, *Bellamya*, Viviparidae

Introduction

The Chinese mystery snail, *Cipangopaludina* (= *Bellamya*) *chinensis* (Gray, 1834), is a non-native, freshwater, gastropod native to Eastern Asia. The species was introduced to North America in the 1890s via the Asian food market and has since become widespread in natural environments in Canada, the United States of America (USA), Belgium, and The

Netherlands (U.S. Fish & Wildlife Services 2011; Collas et al. 2017a; Jokinen 1892; Matthews et al. 2017; McAlpine et al. 2016; Kingsbury *personal observation*). It is suspected that *C. chinensis* is spreading through illegal aquarium releases and accidental boat transfers between infected and uninfected water bodies (Collas et al. 2017a, b; Matthews et al. 2017; Rothlisberger et al. 2010). *Cipangopaludina chinensis* adults can survive more than nine weeks of air exposure, making long range, over-land dispersal possible (Unstad et al. 2013). *Cipangopaludina chinensis* reported occurrences in continental North America suggest that this species is far more widespread than currently documented.

Once introduced, *C. chinensis* presence may lead to blocked water pipes, fouled beaches, altered nitrogen: phosphorus ratios which can elevate eutrophication, and altered food webs through changes in bacterial community composition (Chao et al. 1993; Collas et al. 2017b; Cui et al. 2012; Fang et al. 2001; Hanstein 2012; Harried et al. 2015; Xing et al. 2016). Pressure on native mollusc species may also occur through competition for resources (e.g. food, space, dissolved calcium). Female *C. chinensis* produce large broods of live young. Estimates of > 100 offspring / female/ year have been suggested (Haak 2015; Havel 2011; Stephen et al. 2013; Unstad et al. 2013). Being more resistant to predation than native molluscs due to large size and the presence of a protective operculum "trapdoor" (Haak 2015; Johnson et al. 2009; Karatayev et al. 2009; Olden et al. 2013; Plinski et al. 1978; Sura and Mahon 2011), *C. chinensis* can become established rapidly within new water bodies. Although there is recognition in some provinces that *C. chinensis* introduction to Canada should be of concern, the species has not been identified by federal agencies as potentially invasive across the country, due to lack of published data for Canada (McAlpine et al. 2016; Schroeder et al. 2014).

It is difficult to manage *C. chinensis* once a population is established because the species is resistant to typical species management techniques such as heat (Burnett et al. 2018), desiccation (Unstad et al. 2013), and chemical treatments (Haak et al. 2014). Therefore, it is important to predict future *C. chinensis* distribution trends so as to proactively prevent introductions, especially to vulnerable habitats where species at risk (SAR) are present. Species distribution models (SDMs) are based on simulations that predict a species future distribution using habitat variables from areas of known species presences and real- or pseudo-absences (Hijmans and Elith 2016; Jackson et al. 2000; Law and Kelton 1991). SDMs have become important for freshwater ecosystem management as a means to predict aquatic invasive species (AIS) introductions and spread (Rodríguez-Rey et al. 2019; Drake and Bossenbroek 2009; Love et al. 2015; Phillips et al. 2004; Robbins 2004; Václavík and Meentemeyer 2009). SDMs have been used to predict the distributions of native species, rare species, marine ecosystem assemblages, and to forecast plant dispersal (Baldwin 2009; Best et al. 2007;

DeAngelis and Grimm 2014; Evans et al. 2011; Joy and Death 2004; Merow et al. 2013; Mi et al. 2017; Phillips et al. 2004; Roberts et al. 2010). However, modeling AIS distributions can be challenging because AIS monitoring often excludes absence reports (Václavík and Meentemeyer 2009, 2012). Other AIS modeling challenges include selecting the right model for the species (Mi et al. 2017), choosing appropriate absence data (whether choosing to use true-absences or pseudo-absence or a combination of the two) (Václavík and Meentemeyer 2012), and balancing the assumptions made to fit a model to the current known species distribution without overfitting the model so that it prevents accurate prediction (DeAngelis and Grimm 2014; Hijmans and Elith 2016; Jackson et al. 2000; Merow et al. 2013; Rodríguez-Rey et al. 2019). *Cipangopaludina chinensis* is particularly difficult to accurately model because little is known about the species ecological thresholds or biological needs (Kingsbury et al. *in prep*).

Previous studies have used either generalized linear models (glm), Ecopath, or MaxEnt to predict *C. chinensis* distribution or environmental parameters important for its invasion success. However, all the previous modeling attempts for *C. chinensis* have been restricted to a single well-studied area of Wisconsin, USA (Haak 2015; Haak et al. 2017; Latzka et al. 2015; Papes et al. 2016) and may not translate to other geographic regions that are less well monitored. Moreover, model type selection such as MaxEnt and glm, are problematic when applied specifically to AIS because MaxEnt has been misapplied (i.e. used to predict species distribution when MaxEnt is a tool for assessing habitat suitability) and glms assume linear relationships in data (Phillips et al. 2004; Richmond 2019). Finally, studies have yet to combine water chemistry, physical environmental parameters, ecosystem composition, and anthropogenic aquatic ecosystem interactions into one statistically robust species distribution model. Perhaps, due to gaps in freshwater monitoring data, or the limitations of models that tend to overfit when extreme numbers of parameters are included (e.g. glm), previous modeling attempts have been unable to incorporate multiple parameter types. As demonstrated previously in conservation science (Balden et al. 2020; Evans et al. 2011; Haungnadan et al. 2020; Maguire and Mundle 2020; Mi et al. 2017; Pearman et al. 2020), random forest models (RFMs) have the potential to be more accurate than other model types currently used (e.g. TreeNet, glm, MaxEnt). Nonetheless, RFMs have yet to become widely used for freshwater species, including invasive forms. A new statistically robust predictive modeling approach using species specific data is needed for more accurate predictions of *C. chinensis* distributions.

Random forest models (RFMs) build hundreds or thousands of decision trees by random selection of variables from different database subsets in various orders. Each tree is incorporated to build a "forest" that is resilient to background noise, handles large datasets well, and allows for parameter-type (e.g. water chemistry, ecosystem composition, etc.) variation (Breiman 2001;

Koehrsen 2018). RFMs average predictions made by each tree in the forest to give one prediction encompassing as many parameter interactions as possible within that forest size (Breiman 2001; Koehrsen 2018). RFMs are a collection of many individual regression trees, they do not assume linearity and, therefore, can identify important variables for accurate predictions (Breiman 2001). This model may be suitable for *C. chinensis* SDM as it can incorporate a number of variables, leading to predictions that are more robust, reliable, and flexible than other models.

Finally, RFM will indicate which model variables are highly correlated with *C. chinensis* presence, suggesting the most important factors for *C. chinensis* spread. This study aims to (1) predict suitable *C. chinensis* habitats in Atlantic Canada using a random forest model approach to indicate future distribution, (2) determine potential overlap between areas of high-probability for *C. chinensis* invasion and SAR habitat, and more generally, (3) determine the feasibility of using RFM for freshwater AIS distribution modeling.

Materials and methods

Data collection:

Habitat parameter datasets were obtained from various sources (e.g. Atlantic Data Stream, provincial government websites/databases, Boat Launches Canada, Google Maps) for 343 freshwater bodies (i.e. lakes, rivers, ponds, harbours, and bays) from six Canadian provinces: Nova Scotia (NS) (n = 250), New Brunswick (NB) (n = 44), Prince Edward Island (PEI) (n = 3), Ontario (ON) (n = 40), Alberta (AB) (n = 1), and British Columbia (BC) (n = 5). We logged 36 parameters, including 3 that identified sites (water body name, county name/station number, and data source), 2 that recorded spatial position (latitude and longitude), 17 water quality parameters (e.g. salinity, alkalinity), 6 geomorphic features (surface area, depths, shoreline development, and water body connectedness), 3 that were reflective of human perturbation (stocking frequency indicated greater recreational fishing pressures, number of public boat launches indicate recreational use, and distance to a highway from a boat launch indicate accessibility), and 2 that were indicative of current ecosystem make-up (number of fish species recorded and number of invasive species). See Supplementary material Table S1 for a list of data sources and Figure 1 for a brief overview of the database. Surface area refers to the water body surface area (km^2). Depths reported both maximum depth (m) and mean depth (m). Water body connectedness was ascertained by counting the number of freshwater bodies directly connected to the water body in question (e.g. Lake Banook in Dartmouth, NS has two water bodies directly connected to it, Sullivan's Pond and Lake Micmac, therefore, the number of connected water bodies is 2).
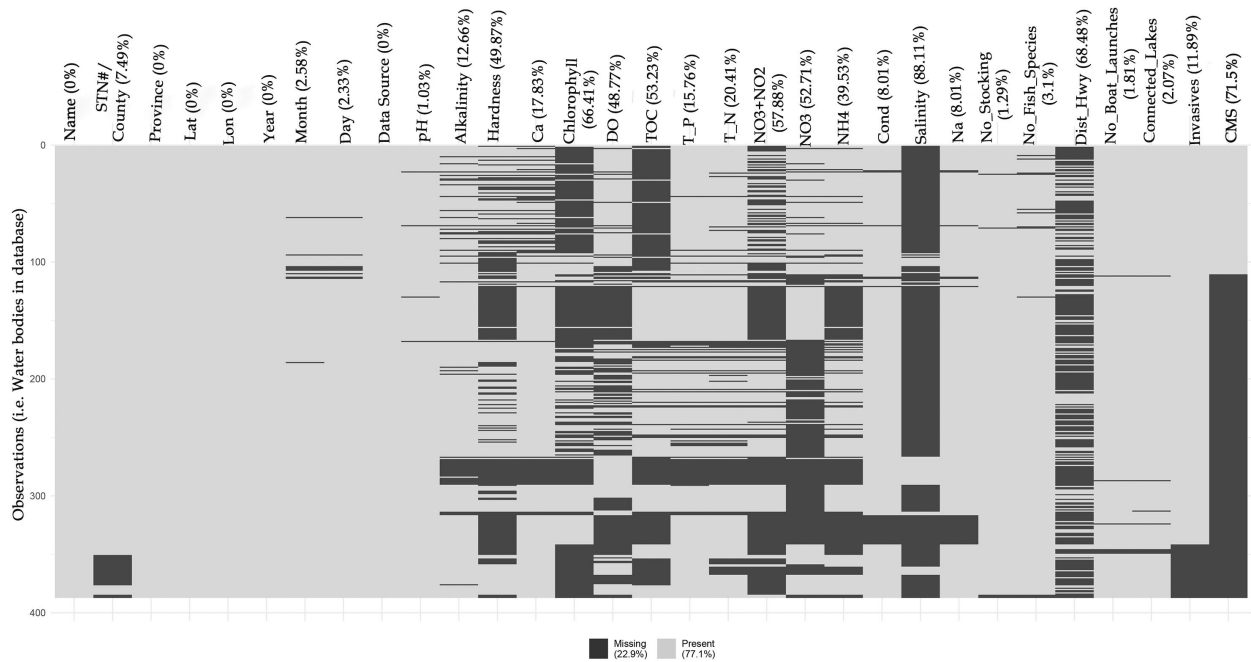
Figure 1. Data categories used to develop random forest models to predict *Cipangopaludina chinensis* distribution in Maritime Canada. Data absences are the result of a lack of standardization in freshwater parameters collected across Canadian jurisdictions. Note, the geomorphologic data is not included in this figure because the relatively large amounts of missing data lead to the exclusion of these parameters from all models. The parameters along the y-axis are: name (name of water body), STN #/County (the station number or county used to identify the water body), province, Lat (latitude in decimal degrees), Lon (longitude in decimal degrees), Year (year of water sampling), Month (month of sampling), Day (day of sampling), Data Source (the source of data, see Table S1), pH, Alkalinity, Hardness (water hardness), Ca (calcium concentration), Chlorophyll (chlorophyll-*a* concentration), DO (dissolved oxygen concentration), TOC (total organic carbon concentration), T_P (total phosphorous concentration), T_N (total nitrogen concentration), NO_3+NO_2 (nitrate plus nitrite concentration), NO_3 (nitrate concentration), NH_4 (ammonia concentration), Cond (conductivity), salinity (salinity concentration), Na (sodium concentration), No stocking (number of annual fish stocking events), No Fish_Species (number of fish species recorded), Dist Hwy (distance to nearest highway), No Boat Launches (number of public boat launches), Invasives (number of known invasive species reported), CMS (known *C. chinensis* presence/absence).

As a means of quality assurance/quality control, each water body was searched using Google Maps to ensure that the name and location data were identical to the original dataset. Water bodies were excluded from the database if the water body name and latitude/longitude data did not match, if the given latitude/longitude did not align with any water body, or if it was not possible to confirm the location and water body name. See Table S2 for a complete list of parameters recorded. To gather the data on each parameter, multiple data sources were used and combined from the year 2000 forward. If multiple sources of information were available for a single water body, the most recent data were used.

Training the model

RFMs were created to analyse water bodies with known *C. chinensis* presence/absence (or pseudo-absences) and predict *C. chinensis* probability of occurrence in water bodies where potential presence/absences were unknown (Table 1). Of the 343 water bodies, 46 had *C. chinensis* presence/absence data or pseudo-absence data which were used to train the RFMs. Based on preliminary work, we found data on the model parameters for 40 water bodies with known *C. chinensis* occurrences from ON (n = 21),

Table 1. Twelve (numbered 1 through 12) random forest model (RFM) itinerations were generated using different combinations of model formula (see footnotes) and datasets used to train and validate the models. Only RFM1 was unvalidated, all other models were validated using subsets of the training datasets. Additionally, the RFM1–8 models were trained on unnormalized data (all units retained). The models RFM9–12 were normalized, with all units converted to a scale of 0 to 1. For most models, all waterbodies that had data were included, regardless the amount of data, while for RFM7, only waterbodies that had more than 50% of the characteristics in the dataset were included. In regards to the Formulas 1–4 below, CMS is the verified presence/absence or pseudo-absence of *Cipangopaludina chinensis*; Lat is latitude; Lon is longitude; Ca is dissolved water calcium concentration; T_P is total phosphorous concentration; T_N is total nitrogen concentration; Cond is conductivity; Na is water sodium concentration; No_Stocking is number of times per year a water body is stocked with fish; No_Fish_Species is the number of reported fish species per water body; Dist_Hwy is the distance to the nearest highway from a publicly accessible boat launch; No_Boat_Launches is the number of publicly accessible boat launches per water body; Connected_Lakes is the number of freshwater bodies connected to the water body in the dataset; and Invasives is the number of invasive species reported in each water body. Note that species labelled as "invasive" are those identified as such by each respective province. Also, the number of recorded fish species varies depending on the individual provincial desire to monitor, stock, or record the number of native fish species in freshwater bodies.

Model phase	Normalized?	Datasets	Waterbodies	RFM Formula No.	RFM Itineration
Unvalidated	No	Verified Presence/Absence data from ON & NS	All waterbodies having data	1	RFM1
Validated	No	Verified Presence/Absence data from ON & NS	All waterbodies having data	1	RFM2
Validated	No	Verified Presence/Absence data from ON & NS	All waterbodies having data	2	RFM3
Validated	No	Verified Presence/Absence data from ON & NS	All waterbodies having data	3	RFM4
Validated	No	Verified Presence/Absence data from ON & NS Pseudo-absence data from ON	All waterbodies having data	1	RFM5
Validated	No	Verified Presence/Absence data from ON & NS Pseudo-absence data from ON	All waterbodies having data	2	RFM6
Validated	No	Verified Presence/Absence data from ON & NS Pseudo-absence data from ON	Waterbodies with large amounts (> 50%) of missing data removed from dataset	1	RFM7
Validated	No	Verified Presence/Absence data from ON & NS Pseudo-absence data from ON	Waterbodies with large amounts (> 50%) of missing data removed from dataset	2	RFM8
Validated	Yes	Verified Presence/Absence data from BC, AB, ON & NS Pseudo-absence data from ON	All waterbodies having data	2	RFM9
Validated	Yes	Verified Presence/Absence data from BC, AB, ON & NS Pseudo-absence data from ON	All waterbodies having data	1	RFM10
Validated	Yes	Verified Presence/Absence data from BC, AB, ON & NS Pseudo-absence data from ON	All waterbodies having data.	3	RFM11
Validated	Yes	Verified Presence/Absence data from BC, AB, ON & NS Pseudo-absence data from ON	All waterbodies having data.	4	RFM12

1. CMS ~ Lat + Lon + pH + Alkalinity + Ca + T_P + T_N + Cond + Na + No_Stocking + No_Fish_Species + Dist_Hwy + No_Boat_Launches + Connected_Lakes + Invasives
2. CMS ~ Lat + Lon + pH + Alkalinity + Ca + Cond + Na + No_Stocking + No_Fish_Species + No_Boat_Launches + Connected_Lakes + Invasives
3. CMS ~ Lat + Lon + pH + Alkalinity + Ca + T_P + Cond + Na + No_Stocking + No_Fish_Species + No_Boat_Launches + Connected_Lakes + Invasives
4. CMS ~ pH + Alkalinity + Ca + T_P + T_N + Cond + Na + No_Stocking + No_Fish_Species + Dist_Hwy + No_Boat_Launches + Connected_Lakes + Invasives

BC (n = 5), and AB (n = 1). Generated models were coded as RFM#, the number denoting the order in which each model was generated (RFM1, RFM2,....RFM12). The NS *C. chinensis* presence data (n = 13) and true absence data (n = 5) was kept separate for model validation (RFM 9–12) or was mixed with the ON data which were randomly subdivided into

training and validation sets (RFM 2–8). We followed the MaxEnt method of selecting pseudo-absences, making our selection from a geographical area that is limited to the region with the greatest number of known species occurrences (in this case, ON). This allowed us to select water bodies with estimated *C. chinensis* absences that could then be used to train our models (Baldwin 2009; Phillips et al. 2004; Richmond 2019).

We semi-randomly selected 19 pseudo-absences from ON (in the Kingston-Toronto area). The known presence points from Kingsbury et al. (*in prep.*), were mapped using ESRIs ArcGIS (ArcMap 10.7). The species presence points were aggregated (within a 35 km radius) and spatially joined to the aggregate polygon (hereafter called the TK polygon) in order to determine the area with the most reported *C. chinensis* presences. The background data points from our water body database were loaded as a layer (water bodies with unknown *C. chinensis* presence/absence). The background points that fell within the TK polygon were selected and exported into an Excel spreadsheet and randomly ranked. Points 1 through 30 were selected as potential pseudo-absence points. Points were excluded if, when cross-checked with known *C. chinensis* presence, they were located in water bodies with known populations. Hence, semi-randomness of selected pseudo-absences was used to train our model. The final training dataset, containing 27 presence points and 19 absence points (total n = 46), was built from as many water bodies as possible.

Each model used different combinations of formulae (Table 1) and datasets (i.e. different combinations of normalized or non-normalized data, Atlantic Canada + ON data or Atlantic Canada + ON + AB + BC, and with/without pseudo-absences). The error rates denoted in Table 2 are the out-of-bag errors, which represent a percentage of trees within the forest that incorrectly classified water bodies from the training or validation datasets. The number of predictions made by each model (Table 2) vary depending on the ability of that model to make predictions. Some models made fewer predictions than others due to data not being normalized on a scale of 0 to 1 (i.e. models that were confused by variations in parameter magnitude), model formulae including a significant number of parameters missing large amounts of data, or the training dataset being relatively smaller than that which was used to train models that produced a greater number of predictions.

Model validation

A variety of validation methods were used. RFM1 was not validated as all of the *C. chinensis* presence/absence data were used to train the model. Although this provided an understanding of relative parameter importance for *C. chinensis* establishment success in NS, it was not statistically reliable. For this reason, we have chosen not to use RFM1 to represent predicted

Table 2. Error rates, number of predictions and formula used to develop Random forest models used to predict *Cipangopaludina chinensis* distribution in Maritime Canada. Model 9 (RFM9) was selected for use as it was able to predict presence/absence at the greatest number of water bodies while maintaining relatively low training and validation error rates.

Model	Error Rate (Training)	Error Rate (Validation)	Number of Predictions	Formula Used
RFM1	18.92	NA	43	1
RFM2	24	25	44	1
RFM3	24	60	168	2
RFM4	24	40	142	3
RFM5	15.62	33	44	1
RFM6	21.88	22.22	162	2
RFM7	27.78	100	43	1
RFM8	27.78	42.86	141	2
RFM9	10.87	13.16	169	2
RFM10	8.7	7.69	46	1
RFM11	10.87	13.16	142	3
RFM12	28.26	7.69	46	4

probabilities in NS. RFM2–8 used the k-fold cross validation technique (Al-Mukhtar 2019; Guillaumot et al. 2019). RFM9–12 were validated on *C. chinensis* presence/absence data including BC, AB, ON, and NS data. Validation error and the number of predictions each model was able to make were considered when selecting the final model (RFM9).

Mapping predicted C. chinensis *habitat suitability, species at risk (SAR), and significant habitat*

Model prediction probabilities were mapped using ArcGIS (Figure 2). The predicted probabilities and their associated latitude/longitude were saved as a csv file, imported into ArcMap, and categorized based on four percentile groupings starting at the lowest predicted probability of 0.5 (50%). The predicted *C. chinensis* distribution was mapped separately from the SAR and Significant Habitat maps (Figure 3) to maximize clarity when comparing the geographic locations of each and to delineate the actual distribution of SAR (current context) versus the predictions (future context). SAR and significant habitat were compared to establish any areas within regions of general overlap with predicted *C. chinensis* distribution. Federally-recognized SAR in NS include *Salmo salar* (Atlantic salmon), *Thamnophis sauritus* (eastern ribbon snake), *Osmerus mordax* (rainbow smelt), *Hydrocotyle umbellata* (water pennywort), *Coregonus huntsmanii* (Atlantic whitefish), *Alasmidonta varicosa* (brook floater mussel), *Emydoidea blandingii* (Blandings turtle), and *Lampsilis cariosa* (yellow lampmussel). SAR and significant habitat polygons were created based on species recovery plans, point distribution, and geospatial files accessed through multiple government organizations (see Table S3 for details). Note that our SAR and significant habitat map only included freshwater aquatic habitats. Our study focused only on aquatic species or semi-aquatic species (e.g. *T. sauritus*) as these are the species predicted to be most impacted by *C. chinensis* presence.
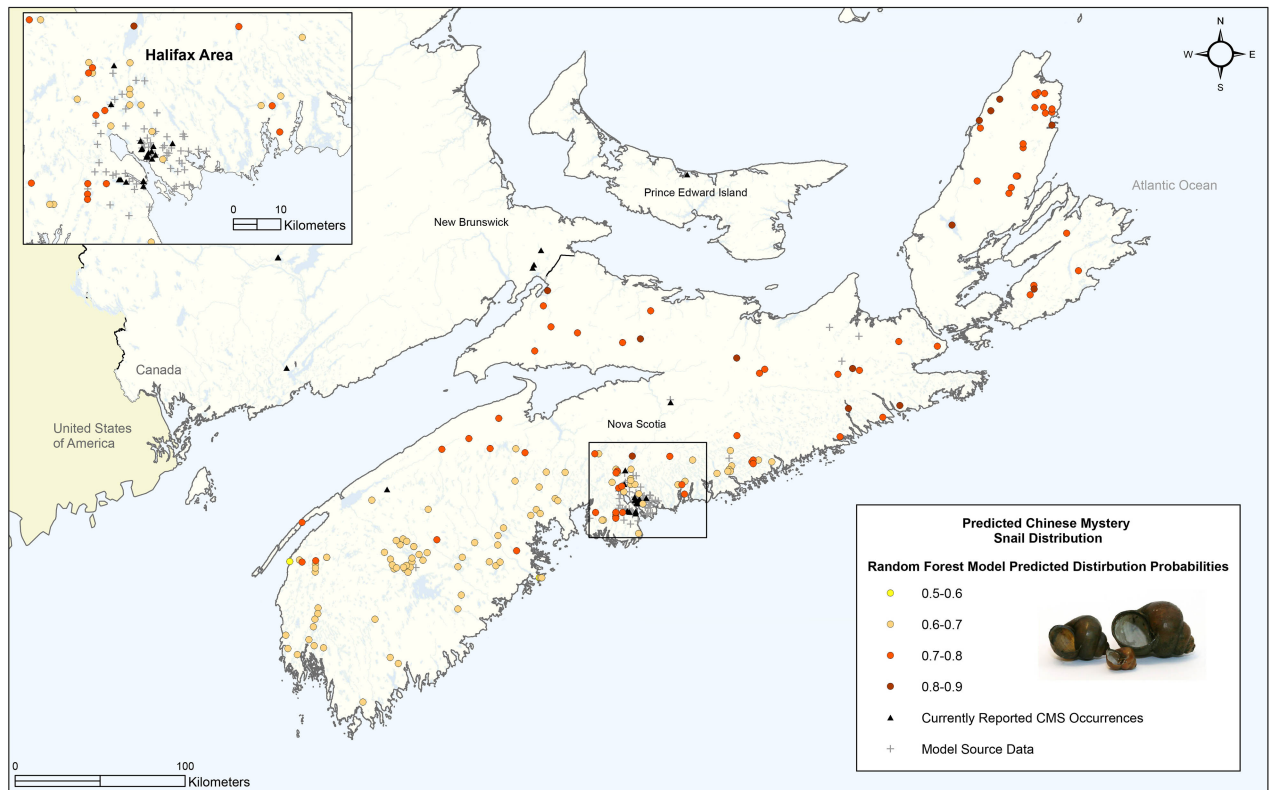
Figure 2. Known and predicted distribution of *Cipangopaludina chinensis* in the Maritimes. Currently confirmed reports for *C. chinensis* occurrences in the Maritimes are shown as black circles. Predicted *C. chinensis* distribution with probabilities are shown as red circles. Crosses mark source data for development of random forest models. Cape Breton, the northeastern island in Nova Scotia, had the high predicted probabilities and Digby, the southern tip, had the lowest.
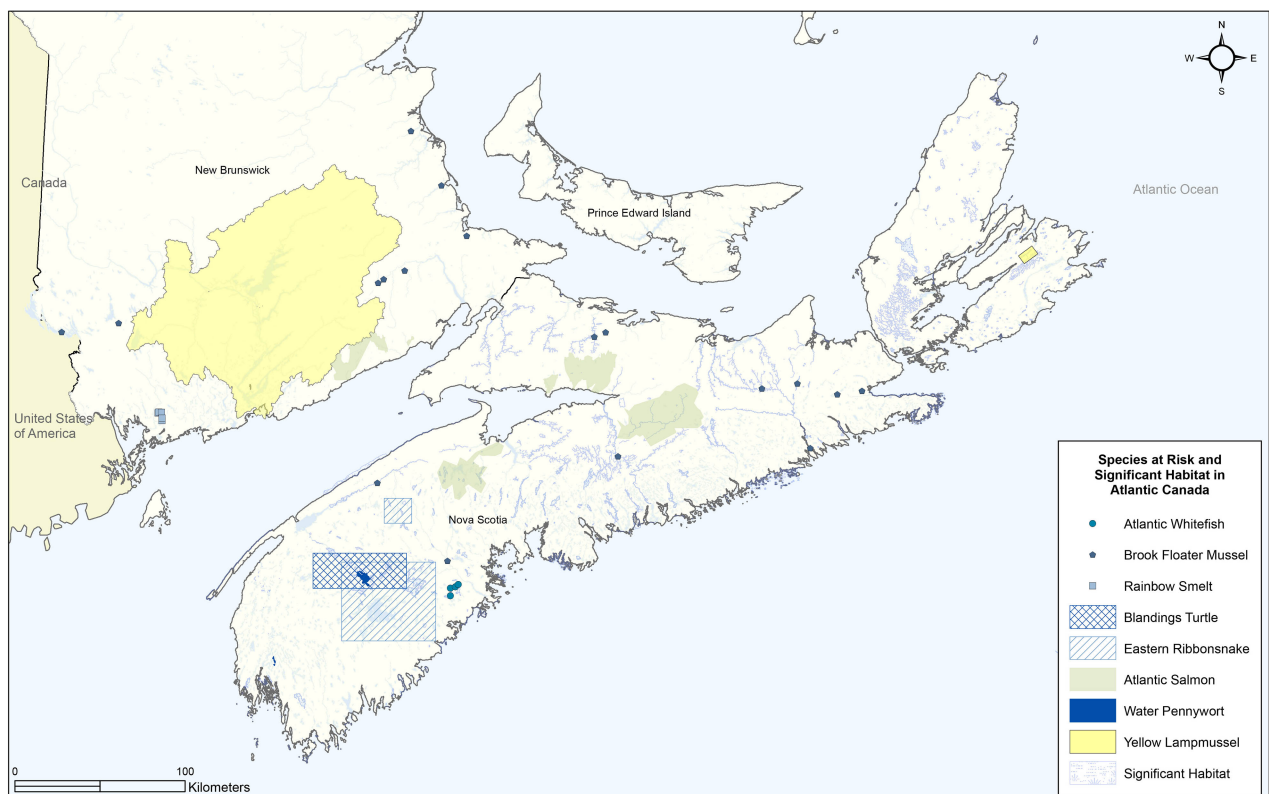

Figure 3. Distribution of selected species at risk (SAR) habitat and significant habitat in Nova Scotia and New Brunswick. See Table S3 for a list of various Government of Canada SAR conservation websites where the data for this map was obtained.
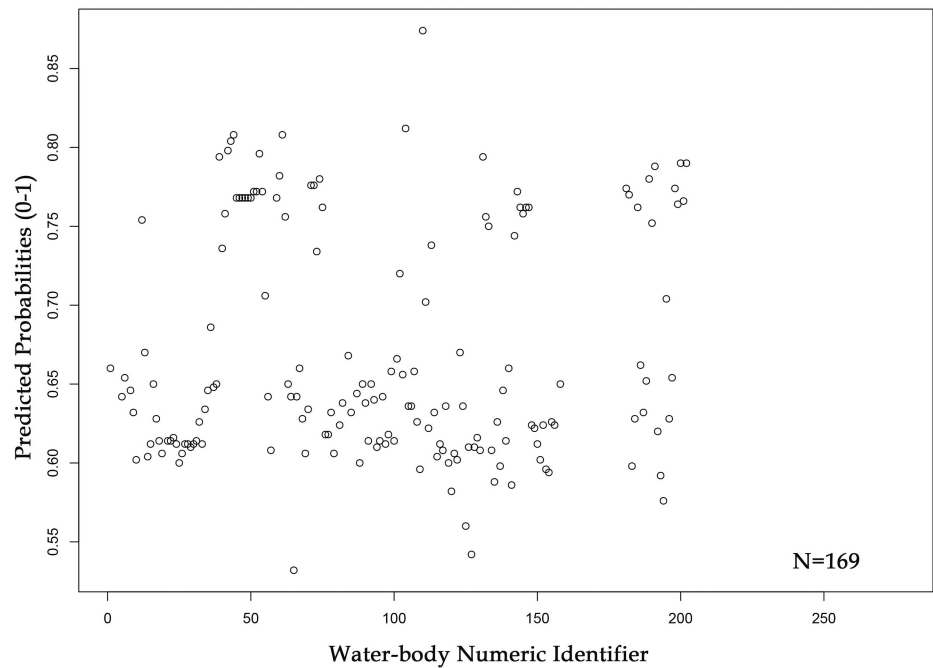
Figure 4. Predicted probability of *Cipangopaludina chinensis* in 169 Nova Scotia (NS) water bodies based on a random forest model. Each water body was given its own numeric identifier (see Supplementary Materials for list of all lakes with its numeric identifier). All predicted probabilities for NS suggest presence of *C. chinensis* with > 50% probability.

Results

The final model (RFM9) predicted 169 of 279 NS water bodies to have a > 50% likelihood of suitability for *C. chinensis* distribution (Figure 4). Among the models generated, none were able to provide predictions for water bodies in NB or PEI (NB predictions: 0/44, PEI: 0/6). The model results listed in Table 2 therefore include only predictions among the 250 NS water bodies examined. Lack of model ability to make predictions for NB or PEI, may be because parameters measured in these provinces did not adequately match parameters used to train the model in terms of data availability, or perhaps due to small sample size. This would appear to be the result of a lack of standardization in freshwater parameters reported across jurisdictions.

Three regions with the highest probability of *C. chinensis* distribution include (a) Cape Breton, NS (the north-eastern Nova Scotian island), (b) along the NS-NB border, (c) and the Halifax Regional Municipality (HRM) (Figure 2). The lowest predicted probability, at 51.6%, is located at the tip of Digby, NS (the south-western edge of NS) on the Bay of Fundy, where high salinity is expected. The highest predicted probability of *C. chinensis* invasion was 85.2% at Lake Ainslie, NS in Cape Breton. Water bodies in northern NS typically have a higher pH and alkalinity than elsewhere in NS and have a higher probability of *C. chinensis* presence, whereas water bodies in southern NS tend to be acidic and dystrophic and not support calcium-bearing invertebrates. It is important to note that all models (RFM1–12) followed similar trends (Figure 5).
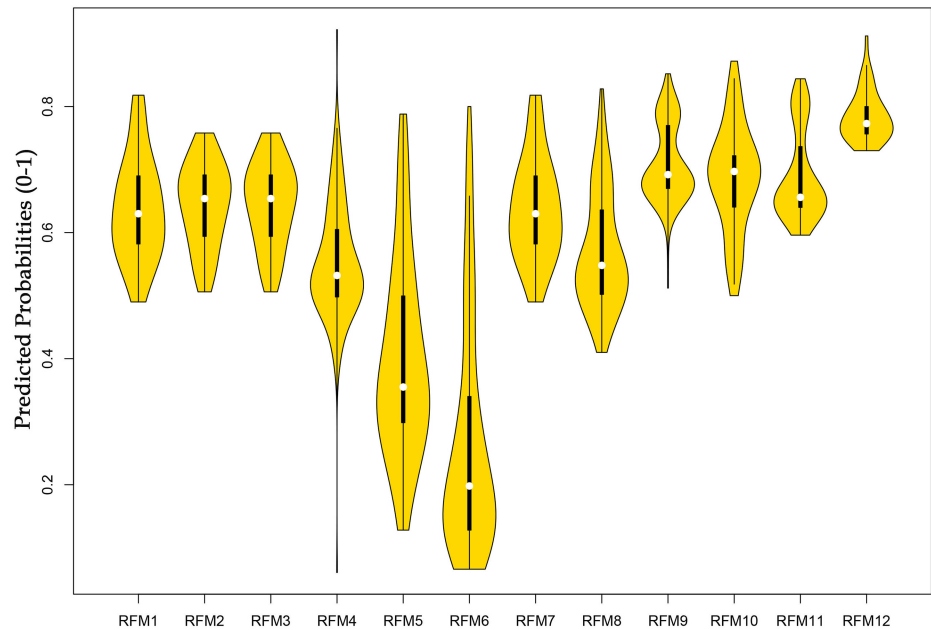
Figure 5. Violin plots comparing each of twelve random forest models (RFM) developed (RFM1–12) to predict *Cipangopaludina chinensis* distribution in Maritime Canada. The x-axis lists RFM developed along with sequential numbering (e.g. RFM1, RFM2). The y-axis shows the predicted probability (0–1) of water bodies in the dataset having suitable habitat for *C. chinensis*. See Table 1 for the list of the RFM equations and parameters. Note that RFM5 and RFM6 included pseudo-absence data and data were not normalized in these models. We selected RFM9 to represent the model predictions seen in RFM1–12 because RFM9 contained data from a national context, data were normalized, the training error and validation error were relatively low and similar, and RFM9 was able to make the largest number of predictions (n = 169) of all models developed.

The most important predictive parameters for *C. chinensis* distribution in NS are the number of fish species present, alkalinity, presence of other invasive species, number of publicly accessible boat launches, and number of connected freshwater bodies (Figure 6). Longitude and latitude are also considered relatively important parameters for the model, but we found national datasets were biased with high proportion of ON reports. Sodium concentration, dissolved calcium concentration, conductivity, pH, and stocking frequency were found to be relatively less important for predicting *C. chinensis* distribution. Multiple areas with high probability of *C. chinensis* distribution were identified in regions supporting habitat for SAR (Figure 3), especially the two freshwater mussel species (brook floater and yellow lampmussel) and Atlantic salmon. Habitats for other species (pennywort, Blanding's turtle, eastern ribbon snake, and the Petite Rivière Watershed which is inhabited by Atlantic whitefish) had low-predicted probability of suitable habitat for *C. chinensis*.

Discussion

The predicted presence of *C. chinensis* in NS was highest in areas with high alkalinity, neutral pH (7–8), high recreational fishing and boating, other invasive species already present in the watershed, and lower fish diversity.
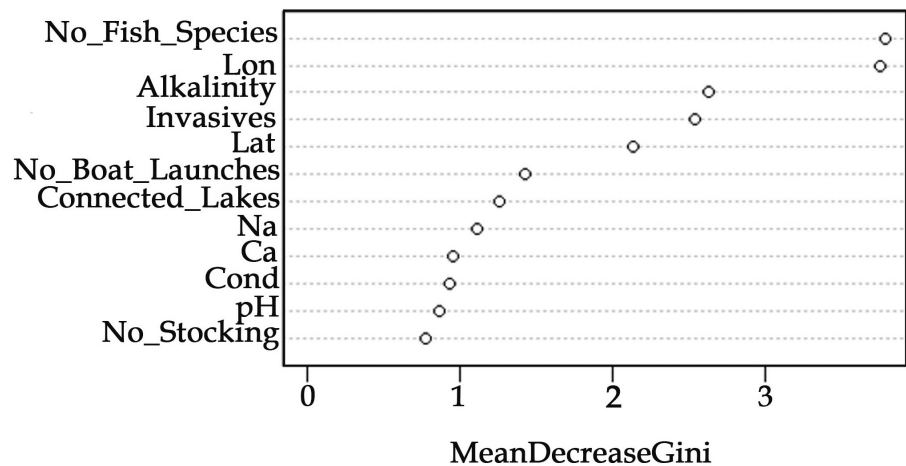
Figure 6. The importance of selected parameters to the preferred random forest model used to predict the distribution of *Cipangopaludina chinensis* in Nova Scotia. The higher mean decrease gini, the more important the variable in predicting *C. chinensis* presence in Nova Scotia. The parameters along the y-axis are: No Fish_Species = number of fish species recorded; Lon = longitude; Invasives = number of known invasive species reported; Lat = latitude; No Boat Launches = number of public boat launches; Na = sodium; Ca = calcium concentration; Cond = conductivity; No stocking = number of annual fish stocking events. Mean decrease gini stems from the gini impurity which is used in decisions trees to determine the importance of random selected variables to correctly split/subset the data. Because RFMs are multiple decision trees which makes a forest, the mean decrease gini is the average of each tree's gini impurity for each parameter. Gini impurity is the measure of probability of misclassifying *C. chinensis* absence if each variable were introduced to the prediction model formula..

This is in line with the results from social network models of *C. chinensis* in Nebraska lakes (Haak 2015; Haak et al. 2017) and classification tree and maximum entropy models of *C. chinensis* in Wisconsin, USA (Latzka et al. 2015; Papes et al. 2016). Predicted *C. chinensis* distribution for NS was high (> 50%) throughout the entire province. Previous research has indicated that *C. chinensis* can tolerate a range of pH values (Haak 2015), salinity concentrations (personal observation), dissolved calcium levels (Haak 2015; Latzka et al. 2015), water temperatures (Burnett et al. 2018), and nutrient concentrations (*personal observation*). *Cipangopaludina chinensis* is a habitat generalist and an opportunistic feeder, and is therefore adaptable to a variety of environments (Ricciardi 2013). In Atlantic Canada, most freshwater bodies are at risk of *C. chinensis* establishment, especially given ongoing anthropogenic disturbance to wetland habitat and climate change leading to range expansion in potential AIS (Bellard et al. 2013; Lassuy and Lewis 2013; Lozon and MacIsaac 1997; Pyšek et al. 2010; Ricciardi 2007; Spear et al. 2013). Many studies have predicted that NS may become a "hot-spot" for invasive species in the future, highlighting the importance of developing effective SDMs for AIS (Barbosa et al. 2013; Bellard et al. 2013; Gallardo et al. 2015; González-Moreno et al. 2015; Lowry et al. 2013; Pyšek et al. 2010; Robbins 2004; Spear et al. 2013).

The principal means of *C. chinensis* introductions to Atlantic Canada is thought to be chiefly boater transfers and aquarium releases (McAlpine et al. 2016). The distribution of *C. chinensis* across multiple locations within the Shubenacadie watershed suggests that this species may also be spreading

naturally. Our own observations of juvenile *C. chinensis* in laboratory cultures demonstrate that juveniles may be carried and spread via water currents. Even though it is likely that many of the initial *C. chinensis* introductions are the result of aquarium releases, it is reasonable to conclude that natural dispersal is now a pathway for the continued spread of *C. chinensis* within NS watersheds. Unfortunately, there is a large temporal-gap (~ 1955–1990) in *C. chinensis* occurrence data for NS, so it has been impossible to track distributional change within the province, except to some degree over the past two decades, and even then, data is patchy. *Cipangopaludina chinensis* continues to be sold in Canada in the water garden and aquarium trade. For regions where models predict high probability for future distribution, strong provincial and federal regulations need to be enacted for boater cleaning (e.g. clean-drain-dry programs). Legally preventing the sale or release of *C. chinensis* is of utmost importance. For regions with established *C. chinensis* populations, including the NS-NB border and HRM, and where our model has predicted especially high probability of continued *C. chinensis* spread, containment of existing populations must be ensured. Active programs of public education and boater awareness may help in this regard. Furthermore, additional modeling approaches may improve predictability by including distance to urban areas, where *C. chinensis* is likely to be sold, and proximity to established *C. chinensis* populations. Such data may help reveal further information about the pathways by which *C. chinensis* has spread.

C. chinensis cannot be effectively managed through culling, drawdowns (Unstad et al. 2013), or chemical treatments (Haak et al. 2014). Our best model indicates that ecosystem composition (defined by number of fish species and number of invasive species known to inhabit a water body), alkalinity, and presence of publicly accessible boat launches were relatively important in predicting *C. chinensis* presence. Hence, as already mentioned, the importance of public education and boater clean-drain-dry programs (Matthews et al. 2017; Rothlisberger et al. 2010). Increasing public involvement through citizen science programs can assist in species monitoring and can increase awareness of the threats that non-native species present to water bodies. Such approaches have been identified as essential elements in *C. chinensis* monitoring and control (Kingsbury et al. *submitted*). Previous public involvement in AIS management and prevention has slowed the spread of zebra mussels (*Dreissena* spp.) in North America, and helped prevent Asian carp (*Cyprinus carpio*), invasion of the Great Lakes (Fisheries and Oceans Canada 2019a; Meyers 2016). A better understanding of how humans impact and enable *C. chinensis* establishment in North America is required so that critical pathways of introduction can be targeted (Barbosa et al. 2013; Gallardo et al. 2015; González-Moreno et al. 2014; Lassuy and Lewis 2013; Lozon and MacIsaac 1997; Pyšek et al. 2010; Robbins 2004; Spear et al. 2013).

The predicted *C. chinensis* distribution in NS is concerning because it frequently overlaps with SAR occurrence and significant habitat for these species. In particular, the predicted habitat overlap between *C. chinensis* and yellow lampmussel is of special concern. Nova Scotia supports only three populations of yellow lampmussel, the species distribution in Canada is restricted to NS and NB, and the species is believed to be in decline (Fisheries and Oceans Canada 2010a). Although the protective measures in place at some yellow lampmussel habitats in NS appear to be robust (e.g. Pottle Lake is a municipal water supply and has many restriction on use), and may decrease the likelihood of *C. chinensis* introduction, other sites are less well protected (e.g. Blacketts Lake). The yellow lampmussel is sensitive to changes in water quality (Fisheries and Oceans Canada 2010) and sedimentation and eutrophication (Sabine et al. 2004). In the Saint John River system, NB, *C. chinensis* and yellow lampmussel have already been shown to overlap in distribution (Sabine et al. 2004; McAlpine et al. 2016), although any impacts on the latter are unknown. From our model, the most notable areas of concern are yellow lampmussel habitat in Cape Breton, brook floater habitats in northern and northwestern NS, the Shubenacadie Watershed connecting HRM and the Bay of Fundy, and the St. Mary's River in Guysborough County, where predicted *C. chinensis* distribution is high (> 70%). The SAR mussel populations are of concern because, should *C. chinensis* densities become sufficiently high, this may directly impact mussel populations through competition for food and calcium necessary for shell development. As an example, the Shubenacadie Watershed, parts of which are already occupied by *C. chinensis*, has great historical and economic importance for NS, and is at high risk of expanded *C. chinensis* occupation. Within this watershed, Shubenacadie Grand Lake is predicted to have an 84.2% likelihood of future *C. chinensis* presence.

At high densities, *C. chinensis* can be considered an ecosystem engineer as it alters habitat to *C. chinensis* benefit. *C. chinensis* prefers diatoms as food and *C. chinensis* presence encourages the growth of diatoms over other algal forms (Gonzalez et al. 2008; Olden et al. 2013). *Cipangopaludina chinensis* alters nitrogen:phosphorous ratios leading to higher concentrations of chlorophyll-*a*, which can lead to eutrophication (Bobeldyk 2009). Furthermore, filter feeding *C. chinensis* are likely to be in direct competition with mussel species for food resources (Olden et al. 2013). Other SAR are also considered to be sensitive to water quality changes, including Atlantic salmon, eastern ribbon snake, rainbow smelt, and water pennywort (COSEWIC 2014; Fisheries and Oceans Canada 2010b, 2016, 2019b; Parks Canada 2012). Alterations in water chemistry or food web structure may negatively impact other species that rely on aquatic ecosystems, including Atlantic whitefish and brook floater mussel (COSEWIC 2009; Fisheries and Oceans Canada 2018). Of the eight SAR listed and mapped in our study, only one, Blandings turtle, is at low risk of being affected by *C. chinensis* presence in NS. The Blandings turtle is apparently less sensitive to changes

in water quality or impacted by habitat perturbation than the other species we considered in our analysis (Parks Canada 2010). Conversely, Kejimkujik National Park and National Historic Site, where the NS population of Blandings turtles is concentrated, showed relatively low probabilities for future *C. chinensis* occurrence (generally < 70%), due to the acidic water chemistry (pH 4–5) of many of those dystrophic lakes. Further research is required to determine the probable consequences of *C. chinensis* presence for each SAR in NS, although we suggest that the best strategy is to prevent *C. chinensis* introductions and to protect the critical habitats of SAR.

Conclusion

The non-native and potentially invasive freshwater mollusc, *C. chinensis,* currently has a limited, although widespread, distribution throughout Atlantic Canada. Our RFM predicted, with > 50% probability, *C. chinensis* establishment throughout NS, with the highest predicted probability in northern-NS, the Halifax Regional Municipality (HRM), and near the NS-NB border. When SAR habitat is compared with predicted *C. chinensis* distribution, overlaps with SAR habitat were identified. Most concerning is potential overlap that *C. chinensis* may have with the yellow lampmussel, a mussel species of conservation concern in Canada. Further research is needed to define the ecological thresholds of *C. chinensis* and to identify the manner in which human factors influence the distribution of this species across its non-native range. Although our model was successfully applied to NS, we were not able to make any predictions for NB or PEI. This was likely due to a lack of standardization in freshwater parameters reported across jurisdictions. Future modelling exercises for AIS in Canada would benefit from greater conformity across provinces in the collection and reporting of water quality and other parameters important to SDMs. Due to the high probability of expanding C. *chinensis* range in Atlantic Canada, and the potential negative impacts this species may have on SAR, it is important that *C. chinensis* be recognized as potentially invasive in Canada and managed as such.

References

Al-Mukhtar M (2019) Random Forest, Support Vector Machine, and Neural Networks to Modelling Suspended Sediment in Tigris River-Baghdad. *Environmental Monitoring and Assessment* 191: 673, https://doi.org/10.1007/s10661-019-7821-5

Balden D, Piniewski M, Funk A, Gumpinger C, Flodl P, Hofer S, Hauer C, Heina T (2020) A multi-scale, integrative modeling framework for setting conservation priorities at the catchment scale for the Freshwater Pearl Mussel *Margaritifera margaritifera. Science of the Total Environment* 718: 137369, https://doi.org/10.1016/j.scitotenv.2020.137369

Baldwin RA (2009) Use of Maximum Entropy Modeling in Wildlife Research. *Entropy* 11: 854–866, https://doi.org/10.3390/e11040854

Barbosa AM, Pautasso M, Figueiredo D (2013) Species-people correlations and the need to account for survey effort in biodiversity analyses. *Diversity and Distributions* 19: 1188–1197, https://doi.org/10.1111/ddi.12106

Bellard C, Thuiller W, Leroy B, Genovesi P, Bakkenes M, Courchamp F (2013) Will climate change promote future invasions? *Global Change Biology* 19: 3740–3748, https://doi.org/10.1111/gcb.12344

Best BD, Halpin PN, Fujioka E, Read AJ, Qian SS, Hazen LJ, Schick RS (2007) Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecological Informatics* 2: 210–223, https://doi.org/10.1016/j.ecoinf.2007.07.007

Bobeldyk A (2009) Spread and Impact of Freshwater Invasive Invertebrates in North America. PhD Thesis, University of Notre Dame, Notre Dame, Indiana, 179 pp

Breiman L (2001) Random Forests. *Machine Learning* 45: 5–32, https://doi.org/10.1023/A:1010933404324

Burnett J L, Pope KL, Wong A, Allen CR, Haak DM, Stephen BJ, Uden DR (2018) Thermal Tolerance Limits of the Chinese Mystery Snail (*Bellamya chinensis*): Implications for Management. *American Malacological Bulletin* 36: 140–144, https://doi.org/10.4003/006.036.0106

Chao D, Wang LC, Huang TC (1993) Prevalence of larval helminths in freshwater snails of the Kinmen Islands. *Journal of Helminthology* 67: 259–264, https://doi.org/10.1017/S0022149X00013249

Collas F, Breedveld S, Matthews J, van der Velde G, Leuven R (2017a) Invasion biology and risk assessment of the recently introduced Chinese mystery snail, *Bellamya* (*Cipangopaludina*) *chinensis* (Gray, 1834), in the Rhine and Meuse River basins in Western Europe. *Aquatic Invasions* 12: 275–286, https://doi.org/10.3391/ai.2017.12.3.02

Collas FPL, De Hoop L, Van der Velde G, Leuven RSEW (2017b) Risk assessment of the alien Chinese mystery snail (*Bellamya chinensis*): An update using results of field surveys in 2017. Department of Environmental Science, Faculty of Science, Institute for Water and Wetland Research, Radboud University. Project number 626460RL2017-2

COSEWIC (2009) COSEWIC assessment and status report on the brook floater, *Alasmidonta varicosa*, in Canada. Committee on the Status of Endangered Wildlife in Canada, 79 pp, https://central.bac-lac.gc.ca/.item?id=CW69-14-580-2009E&op=pdf&app=Library

COSEWIC (2014) COSEWIC Assessment and Status Report on the Water Pennywort *Hydrocotyle umbellata* in Canada. Committee on the Status of Endangered Wildlife in Canada, 44 pp, https://www.sararegistry.gc.ca/default.asp?lang=En&n=25B43406-1#_03

Cui J, Shan B, Tang W (2012) Effect of periphyton community structure on heavy metal accumulation in mystery snail (*Cipangopaludina chinensis*): A case study of the Bai River, China. *Journal of Environmental Sciences* 24: 1723–1730, https://doi.org/10.1016/S1001-0742(11)60976-5

DeAngelis DL, Grimm V (2014) Individual-based models in ecology after four decades. *F1000Prime Reports* 6: 39, https://doi.org/10.12703/P6-39

Drake JM, Bossenbroek JM (2009) Profiling Ecosystem Vulnerability to Invasion by Zebra Mussels with Support Vector Machines. *Theoretical Ecology* 2: 189–198, https://doi.org/10.1007/s12080-009-0050-8

Evans JS, Murphy MA, Holden ZA, Cushman, SA (2011) Modeling Species Distribution and Change Using Random Forest. In: Drew C, Wiersma Y, Huettmann F (eds), Predictive Species and Habitat Modeling in Landscape Ecology. Springer New York, NY, pp 139–159, https://doi.org/10.1007/978-1-4419-7390-0_8

Fang ZQ, Cheung RY, Wong MH (2001) Heavy metal concentrations in edible bivalves and gastropods available in major markets of the Pearl River Delta. *Journal of Environmental Sciences (China)* 13(2): 210–217

Fisheries and Oceans Canada (2010a) Management Plan for the Yellow lampmussel (*Lampsilis cariosa*) in Canada [Final]. Species Act Risk Act Management Plan Series, 44 pp, https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/management-plans/yellow-lampmussel.html#ch1.5

Fisheries and Oceans Canada (2010b) Recovery Strategy for the Atlantic Salmon (*Salmo salar*), Inner Bay of Fundy Populations [Final]. Species at Risk Act Recovery Strategy Series, 58 pp, https://www.registrelep-sararegistry.gc.ca/document/doc1917f/ind_e.cfm

Fisheries and Oceans Canada (2016) Recovery strategy for the Lake Utopia rainbow smelt (*Osmerus mordax*), small-bodied population (sympatric with the large-bodied population),

in Canada. Species at Risk Act Recovery Strategy Series, 57 pp, https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/recovery-strategies/lake-utopia-rainbow-smelt-proposed.html#_5

Fisheries and Oceans Canada (2018) Amended recovery strategy for the Atlantic Whitefish (*Coregonus huntsman*) in Canada. Species at Risk Act Recovery Strategy Series, 62 pp, https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/recovery-strategies/atlantic-whitefish-amended-final.html

Fisheries and Oceans Canada (2019a) Asian Carp. https://www.dfo-mpo.gc.ca/species-especes/profiles-profils/asiancarp-carpeasiatique-eng.html (accessed 02 March 2020)

Fisheries and Oceans Canada (2019b) Action plan for the Atlantic Salmon (*Salmo salar*), inner Bay of Fundy population in Canada. Species at Risk Act Action Plan Series, 61 pp, https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/action-plans/atlantic-salmon-inner-bay-fundy-2019.html

Gallardo B, Zieritz A, Aldridge DC (2015) The Importance of the Human Footprint in Shaping the Global Distribution of Terrestrial, Freshwater and Marine Invaders. *PLoS ONE* 10: e0125801, https://doi.org/10.1371/journal.pone.0125801

Gonzalez A, Lambert A, Ricciardi A (2008) When does ecosystem engineering cause invasion and species replacement? *Oikos* 117: 1247–1257, https://doi.org/10.1111/j.0030-1299.2008.16419.x

González-Moreno P, Diez JM, Ibáñez I, Font X, Vilà M (2014) Plant invasions are context-dependent: Multiscale effects of climate, human activity and habitat. *Diversity and Distributions* 20: 720–731, https://doi.org/10.1111/ddi.12206

González-Moreno P, Diez JM, Richardson DM, Vilà M (2015) Beyond climate: Disturbance niche shifts in invasive species. *Global Ecology and Biogeography* 24: 360–370, https://doi.org/10.1111/geb.12271

Guillaumot C, Artois J, Saucède T, Demoustier L, Moreau C, Eléaume M, Agüera A, Danis B (2019) Broad-Scale Species Distribution Models Applied to Data-Poor Areas. *Progress in Oceanography* 175: 198–207, https://doi.org/10.1016/j.pocean.2019.04.007

Haak DM (2015) Bioenergetics and habitat suitability moels for the Chinese mystery snail (*Bellamya chinensis*). PhD Dissertation, University of Nebraska, Nebraska, United States of America, https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1108&context=natresdiss

Haak D, Stephen B, Kill R, Smeenk N, Allen C, Pope K (2014) Toxicity of copper sulfate and rotenone to Chinese mystery snail (*Bellamya chinensis*). *Management of Biological Invasions* 5: 371–375, https://doi.org/10.3391/mbi.2014.5.4.08

Haak DM, Fath BD, Forbes VE, Martin DR, Pope KL (2017) Coupling ecological and social network models to assess "transmission" and "contagion" of an aquatic invasive species. *Journal of Environmental Management* 190: 243–251, https://doi.org/10.1016/j.jenvman.2016.12.012

Hanstein B (2012) That's a lot of snails: 504 pounds pulled from Clearwater Lake. *Daily Bulldog*, http://www.dailybulldog.com/db/features/thats-a-lot-of-snails-504-pounds-pulled-from-clearwater-lake/

Harried B, Fischer K, Perez KE, Sandland GJ (2015) Assessing infection patterns in Chinese mystery snails from Wisconsin, USA using field and laboratory approaches. *Aquatic Invasions* 10: 169–175, https://doi.org/10.3391/ai.2015.10.2.05

Havel JE (2011) Survival of the exotic Chinese mystery snail (*Cipangopaludina chinensis malleata*) during air exposure and implications for overland dispersal by boats. *Hydrobiologia* 668: 195–202, https://doi.org/10.1007/s10750-010-0566-3

Hijmans RJ, Elith J (2016) Spatial Distribution Models [R Studio]. Spatial Data Science, https://rspatial.org/raster/sdm/index.html (accessed 11 August 2019)

Haungnadan F, Kefi S, Deter J (2020) Identifying key-conservation areas for *Posidonia oceanica* seagrass beds. *Biological Conservation* 247: 108546, https://doi.org/10.1016/j.biocon.2020.108546

Jackson LJ, Trebitz AS, Cottingham KL (2000) An Introduction to the Practice of Ecological Modeling. *BioScience* 50: 694, https://doi.org/10.1641/0006-3568(2000)050[0694:AITTPO]2.0.CO;2

Johnson PTJ, Olden JD, Solomon CT, Vander Zanden MJ (2009) Interactions among invaders: Community and ecosystem effects of multiple invasive species in an experimental aquatic system. *Oecologia* 159: 161–170, https://doi.org/10.1007/s00442-008-1176-x

Jokinen EJ (1982) *Cipangopaludina chinensis* (Gastropoda: Viviparidae) in North America, Review and Update. *Nautilus* 96(3): 89–95

Joy MK, Death RG (2004) Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. *Freshwater Biology* 49: 1036–1052, https://doi.org/10.1111/j.1365-2427.2004.01248.x

Karatayev AY, Burlakova LE, Karatayev VA, Padilla DK (2009) Introduction, distribution, spread, and impacts of exotic freshwater gastropods in Texas. *Hydrobiologia* 619: 181–194, https://doi.org/10.1007/s10750-008-9639-y

Koehrsen W (2018) An Implementation and Explanation of the Random Forest in Python. https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76 (accessed 04 February 2020)

Lassuy D, Lewis P (2013) Invasive Species: Human-Induced-Arctic biodiversity, Conservation of Arctic Flora and Fauna (CAFF). Conservation of Artic Flora and Fauna. https://www.arcticbiodiversity.is/index.php/the-report/chapters/invasive-species-human-induced (accessed 11 December 2019)

Latzka AW, Crawford JT, Koblings AS, Caldeira Y, Hilts E, Vander Zanden MJ (2015) Representing calcification in distribution models for aquatic invasive species: Surrogates perform as well as $CaCO_3$ saturation state. *Hydrobiologia* 746: 197–208, https://doi.org/10.1007/s10750-014-2001-7

Law AM, Kelton WD (1991) Simulation modeling and analysis (2nd ed). McGraw-Hill, Tuscon, Arizona, 759 pp

Love JW, Newhard JJ, Greenfield B (2015) A Geospatial Approach for Estimating Suitable Habitat and Population Size of the Invasive Northern Snakehead. Journal of Fish and *Wildlife Management* 6: 145–157, https://doi.org/10.3996/102014-JFWM-075

Lowry E, Rollinson EJ, Laybourn AJ, Scott TE, Aiello-Lammens ME, Gray SM, Mickley J, Gurevitch J (2013) Biological invasions: A field synopsis, systematic review, and database of the literature. *Ecology and Evolution* 3: 182–196, https://doi.org/10.1002/ece3.431

Lozon JD, MacIsaac HJ (1997) Biological invasions: Are they dependent on disturbance? *Environmental Reviews* 5: 131–144, https://doi.org/10.1139/a97-007

Maguire TJ, Mundle SOC (2020) Citizen Science Data Show Temperature-Driven Declines in Riverine Sentinel Invertebrates. *Environmental Science & Technology Letters* 7: 303–307, https://doi.org/10.1021/acs.estlett.0c00206

Matthews J, Collas FPL, de Hoop L, van der Velde G, Leuven RSEW (2017) Management approaches for the alien Chinese mystery snail (*Bellamya chinensis*). *Environmental Sciences* 558: 1–22, https://repository.ubn.ru.nl/handle/2066/177158

McAlpine DF, Lepitzki DAW, Schueler FW, McAlpine FJT, Hebda A, Forsyth RG, Nicolai A, Maunder JE, Noseworthy RG (2016) Occurrence of the Chinese mystery snail, *Cipangopaludina chinensis* (Gray, 1834) (Mollusca: Viviparidae) in the Saint John River system, New Brunswick, with review of status in Atlantic Canada. *BioInvasions Records* 5: 149–154, https://doi.org/10.3391/bir.2016.5.3.05

Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography* 36: 1058–1069, https://doi.org/10.1111/j.1600-0587.2013.07872.x

Meyers SL (2016) Outreach Efforts Stabilize Spread Of Zebra Mussels. Wisconsin Public Radio, https://www.wpr.org/outreach-efforts-stabilize-spread-zebra-mussels (accessed 3 February 2020)

Mi C, Huettmann F, Guo Y, Han X, Wen L (2017) Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5: e2849, https://doi.org/10.7717/peerj.2849

Nova Scotia Lands and Forestry (2013) Significant Species and Habitats Database. https://novascotia.ca/natr/wildlife/habitats/hab-data/ (accessed 02 March 2020)

Olden JD, Ray L, Mims MC, Horner-Devine MC (2013) Filtration rates of the non-native Chinese mystery snail (*Bellamya chinensis*) and potential impacts on microbial communities. *Limnetica* 32(1): 107–120

Papes M, Havel JE, Vander Zanden MJ (2016) Using maximum entropy to predict the potential distribution of an invasive freshwater snail. *Freshwater Biology* 61: 457–471, https://doi.org/10.1111/fwb.12719

Parks Canada (2010) Recovery Strategy for the Blanding's turtle, Nova Scotia population, in Canada. Species at Risk Act Recovery Strategy Series. https://sararegistry.gc.ca/document/doc2245p/tdmtoc_e.cfm

Parks Canada (2012) Recovery strategy for the eastern ribbonsnake (*Thamnophis sauritus*), Atlantic population in Canada. Parks Canada. https://central.bac-lac.gc.ca/.item?id=En3-4-81-2010-eng&op=pdf&app=Library

Pearman TRR, Robert K, Callaway A, Hall R, Lo Iacono C, Huvenne VAI (2020) Improving the predictive capability of benthic species distribution models by incorporating oceanographic data - Towards holistic ecological modelling of a submarine canyon. *Progress in Oceanography* 184: 102338, https://doi.org/10.1016/j.pocean.2020.102338

Phillips SJ, Dudık M, Schapire RE (2004) A Maximum Entropy Approach to Species Distribution Modeling. In: Proceedings of the twenty-first international conference on Machine learning (ICML '04). Association for Computing Machinery, New York, NY, USA, 83, https://doi.org/10.1145/1015330.1015412

Plinski M, Ławacz W, Stanczykowska A, Magnin E (1978) Etude quantitative et qualitative de la nourriture des *Viviparus malleatus* (Reeve) (Gastropoda, Prosobranchia) dans deux lacs de la région de Montréal. *Canadian Journal of Zoology* 56: 272–279, https://doi.org/10.1139/z78-036

Pyšek P, Jarošík V, Hulme PE, Kühn I, Wild J, Arianoutsou M, Bacher S, Chiron F, Didžiulis V, Essl F, Genovesi P, Gherardi F, Hejda M, Kark S, Lambdon PW, Desprez-Loustau M-L, Nentwig W, Pergl J, Poboljšaj K, Mooney HA (2010) Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences of the United States of America* 107: 12157–12162, https://doi.org/10.1073/pnas.1002314107

Ricciardi A (2007) Are Modern Biological Invasions an Unprecedented Form of Global Change? *Conservation Biology* 21: 329–336, https://doi.org/10.1111/j.1523-1739.2006.00615.x

Ricciardi A (2013) Invasive Species. In: Leemans R (ed), Ecological Systems. Springer, New York, NY, pp 161–178, https://doi.org/10.1007/978-1-4614-5755-8_10

Richmond S (2019) Maxent. EHelp Support. https://support.bccvl.org.au/support/solutions/articles/6000083216-maxent (accessed 15 October 2019)

Robbins P (2004) Comparing Invasive Networks: Cultural and Political Biographies of Invasive Species. *Geographical Review* 94: 139–156, https://doi.org/10.1111/j.1931-0846.2004.tb00164.x

Roberts JJ, Best BD, Dunn DC, Treml EA, Halpin PN (2010) Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software* 25: 1197–1207, https://doi.org/10.1016/j.envsoft.2010.03.029

Rodríguez-Rey M, Consuegra S, Börger L, de Leaniz CG (2019) Improving Species Distribution Modelling of freshwater invasive species for management applications. *PLoS ONE* 14: e0217896, https://doi.org/10.1371/journal.pone.0217896

Rothlisberger JD, Chadderton WL, McNulty J, Lodge DM (2010) Aquatic Invasive Species Transport via Trailered Boats: What is Being Moved, Who is Moving it, and What Can Be Done. *Fisheries* 35: 121–132, https://doi.org/10.1577/1548-8446-35.3.121

Sabine DL, Makepeace S, McAlpine DF (2004) The Yellow Lampmussel (*Lampsilis cariosa*) in New Brunswick: a population of significant conservation value. *Northeastern Naturalist* 11: 407–420, https://doi.org/10.1656/1092-6194(2004)011[0407:TYLLCI]2.0.CO;2

Schroeder B, Mandrak N, Cudmore B (2014) Application of a Freshwater Mollusc Risk Assessment to Non-Indigenous Organisms in Trade in Canada. Fisheries and Oceans Canada (DFO). DFO Canadian Science Advisory Secretariat Research Document 2013/060, 26 pp

Spear D, Foxcroft LC, Bezuidenhout H, McGeoch MA (2013) Human population density explains alien species richness in protected areas. *Biological Conservation* 159: 137–147, https://doi.org/10.1016/j.biocon.2012.11.022

Stephen BJ, Allen CR, Chaine NM, Fricke KA, Haak DM, Hellman ML, Kill RA, Nemec KT, Pope KL, Smeenk NA, Uden DR, Unstad KM, VanderHam AE, Wong A (2013) Fecundity of the Chinese mystery snail in a Nebraska reservoir. *Journal of Freshwater Ecology* 28: 439–444, https://doi.org/10.1080/02705060.2013.769127

Sura SA, Mahon HK (2011) Effects of competition and predation on the feeding rate of freshwater snails. *The American Mainland Naturalist* 166: 358–368, https://doi.org/10.1674/0003-0031-166.2.358

Unstad K, Uden D, Allen C, Chaine N, Haak D, Kill R, Pope K, Stephen B, Wong A (2013) Survival and behavior of Chinese mystery snails (*Bellamya chinensis*) in response to simulated water body drawdowns and extended air exposure. Nebraska Cooperative Fish & Wildlife Research Unit – Staff Publications. https://digitalcommons.unl.edu/ncfwrustaff/123

U.S. Fish & Wildlife Services (2011) Chinese mystery snail (*Cipangopaludina chinensis*) Ecological Risk Screening Summary. United States of America Government Species Report, 16 pp

Václavík T, Meentemeyer RK (2009) Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling* 220: 3248–3258, https://doi.org/10.1016/j.ecolmodel.2009.08.013

Václavík T, Meentemeyer RK (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions* 18: 73–83, https://doi.org/10.1111/j.1472-4642.2011.00854.x

Xing Y, Yang P, Yu J (2016) Biosorption of Pb(II) by the shell of vivipaird snail: Implications for heavy metal bioremediation. *Separation Science and Technology* 51: 2756–2761, https://doi.org/10.1080/01496395.2016.1217242

Supplementary material

The following supplementary material is available for this article:

Table S1. Data Sources by Province.

Table S2. random forest model Parameters.

Table S3. Species at risk (SAR) and Significant Habitat Data Sources.

Table S4. List of acronyms used in the manuscript text.

This material is available as part of online article from:

http://www.reabic.net/aquaticinvasions/2021/Supplements/AI_2021_Kingsbury_etal_SupplementaryTables.xlsx