Journal of Great Lakes Research 40 (2014) 980–988

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

Estimating the distribution of the imperiled pugnose shiner (*Notropis anogenus*) in the St. Lawrence River using a habitat model

M.R. McCusker [a,*], N.E. Mandrak [b], S. Doka [b], E.L. Gertzen [b], J.F. van Wieren [c], J.E. McKenna [d], D.M. Carlson [e], N.R. Lovejoy [a]

[a] Department of Biological Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario M1C-1A4, Canada
[b] Fisheries and Oceans Canada, 867 Lakeshore Road, Burlington, Ontario L7R 4A6, Canada
[c] Thousand Islands National Park, 2 County Rd. 5, Mallorytown, Ontario K0E 1R0, Canada
[d] USGS Great Lakes Science Center, Tunison Laboratory of Aquatic Science, 3075 Gracie Road, Cortland, NY 13045-9457, USA
[e] New York State Department of Environmental Conservation, 317 Washington St., Watertown, NY 13601, USA

ARTICLE INFO

Article history:
Received 25 April 2014
Accepted 16 August 2014
Available online 14 October 2014

Communicated by Ed Rutherford

Index words:
Habitat suitability
MaxEnt
Freshwater fish
Species at risk

ABSTRACT

The pugnose shiner, *Notropis anogenus*, is a small minnow that occurs in the Great Lakes basin and Upper Mississippi River basin. It was listed as 'Endangered' under the Canadian Species at Risk Act (SARA) and by the State of New York, largely due to its rarity, fragmented distribution, and degraded habitat. Our objective was to use species distribution modeling to better understand the spatial extent of suitable habitat for the pugnose shiner in the upper St. Lawrence River and to guide protection of habitat for this endangered species. We performed our analyses with MaxEnt, a species distribution modeling method based on maximum entropy. The pugnose shiner is associated with shallow areas of lakes or slow-moving rivers with abundant submerged aquatic vegetation. Therefore, we created a model based on depth, water velocity, and aquatic vegetation. For water depth and velocity, we used results of a two-dimensional hydrodynamic model from the Upper St. Lawrence River generated and calibrated by Environment Canada. Aquatic vegetation was estimated from a simple algorithm based on water depth and velocity. To minimize the effect of sampling bias in the analysis (i.e., sampling occurred predominantly in shallow waters), we also used restricted depth ranges in model generation. Our model produced highly significant results when depth was not restricted, and significant results for analyses with restricted depth ranges. Our results suggest an abundance of potentially suitable habitat for the pugnose shiner in the Upper St. Lawrence River, which exceeded the minimum area for viable population estimates for this species.

Introduction

Freshwater fishes are facing a biodiversity crisis in North America and worldwide. Loss of aquatic habitat is considered a primary threat to global biodiversity (Abell, 2002; Dudgeon et al., 2006; Jelks et al., 2008). The Great Lakes region has undergone a substantial change in its fish fauna over the past 200 years related primarily to habitat alterations, invasive species, and exploitation (Mandrak and Cudmore, 2010). The pugnose shiner, *Notropis anogenus*, occurs within the Great Lakes basin and Upper Mississippi River basin. For its entire range, Jelks et al. (2008) recommended it be considered as Threatened, the IUCN considers it Near Threatened (IUCN, 2014), and NatureServe (2014) considers the species Globally Vulnerable. Although the species is widely considered to be at risk, its conservation status varies substantially among regions. Within the United States, the species is considered Extirpated in Ohio, Critically Imperiled in Illinois, Indiana, Iowa, New York, and North Dakota, Imperiled in Wisconsin, and Vulnerable in Michigan and Minnesota (NatureServe, 2014). Within Canada, the species was listed federally as Endangered by the Canadian Species at Risk Act (SARA) in 2005 although the Committee on the Status of Endangered Wildlife in Canada now considers it Threatened (COSEWIC, 2013). Within Canada, the species is only found in Ontario, where habitat degradation and loss are considered the greatest threats; and habitat protection and conservation represent the highest priorities for species recovery (DFO, 2010).

The pugnose shiner is typically found in lakes or slow-moving rivers with low turbidity and abundant, submerged aquatic vegetation (Doeringsfeld, 1993; Holm and Mandrak, 2002). The species likely spawns in waters less than 2–3 m deep in the summer, and may move to slightly deeper water during winter (DFO, 2010). It relies heavily on submerged aquatic vegetation, such as *Valisneria*, *Potamogeton*, and *Chara*, which it uses for food, shelter, and reproduction (DFO, 2010). Like other species that occupy the nearshore environment, the pugnose

* Corresponding author. Tel.: +1 416 208 4823; fax: +1 416 287 7676.
E-mail addresses: mmccusker@utsc.utoronto.ca (M.R. McCusker), Nicholas.Mandrak@dfo-mpo.gc.ca (N.E. Mandrak), Josh.VanWieren@pc.gc.ca (J.F. van Wieren).

http://dx.doi.org/10.1016/j.jglr.2014.09.014
0380-1330/

shiner is also highly susceptible to flow regime changes, which can strongly affect vegetation and habitat availability (Doka et al., 2006).

Species habitats and distributions are increasingly modeled within a Geographic Information System (GIS) framework, sometimes on very fine scales (Knouft et al., 2011). For this study, we implement MaxEnt, a program that uses a maximum-entropy model to identify suitable habitat characteristics using presence-only locality data (Phillips et al., 2006). This method was designed, in part, to take advantage of data within natural history museums or herbaria, as well as recent advances in the availability of fine-scale environmental data. We selected MaxEnt because it has been shown to perform well at both defining habitat characteristics and predicting suitable habitat in novel areas compared with a wide variety of other presence-only methods (Elith et al., 2006; Tittensor et al., 2009). MaxEnt has consistently shown strong performance compared to other methods particularly with small sample sizes of occurrence data (Hernandez et al., 2006; Pearson et al., 2007; Wisz et al., 2008). The maximum entropy model in MaxEnt can handle a high degree of model complexity and is known to be relatively robust to correlated environmental variables (Elith et al., 2011). MaxEnt's relative insensitivity to small sample sizes, its ability to model complex relationships among environmental variables, and its use of regularization to avoid over-fitting (Hernandez et al., 2006; Wisz et al., 2008; Li and Wang, 2013) make it an ideal choice for our dataset.

Here, we used MaxEnt to model mesoscale habitat suitability of the pugnose shiner in the upper St. Lawrence River. The St. Lawrence River likely represents the best habitat currently available in Ontario (DFO, 2010) and the species is relatively widely distributed throughout this region. The pugnose shiner occurs in shallow depths with low flow and abundant submerged aquatic vegetation, and we incorporated all three of these environmental variables into our model. Although the pugnose shiner is also considered to be highly susceptible to turbidity, we did not expect turbidity to vary widely within our study area (Hudon, 2000), and turbidity was not included in our analysis. We tested whether or not the species distribution could be effectively modeled with our three environmental variables, and which variables, or combinations of variables, were most effective at predicting the species distribution. Our research addresses a major knowledge gap identified in the Canadian recovery strategy for the pugnose shiner (DFO, 2010) by determining the distribution and availability of suitable habitat.

Methods

Environmental data

We assembled environmental data using ArcGIS 10 for the upper St. Lawrence River (from approximately Kingston to Cornwall, Ontario (Fig. 1), above the Moses–Saunders Dam at Cornwall). Depth and water velocity data were provided by Environment Canada and taken from a two-dimensional model of hydrologic and hydraulic data, RMA2 (Thompson and Moin, 2003). The RMA2 model simulates depth-averaged velocity and water levels for ice-free flow. The model is based on bathymetry data from the Canadian Department of Fisheries and Oceans' Canadian Hydrographic Service (CHS) for Canadian waters and from the National Oceanic and Atmospheric Administration (NOAA) and the United States Army Corps of Engineers (USACE) for American waters. Two parameters, eddy viscosity and Manning's *n* coefficient (to estimate bed roughness), were used to characterize the resistance of the river. The model generates over 82,000 nodes (data points) from Kingston to Cornwall, with spacing among nodes of approximately 150–300 m. The RMA2 model was intended to capture the complexity of the shoreline and major islands without creating an excessive number of data points. Model calibration and verification were performed using water levels at gauging stations and discharge observations from the Moses–Saunders Dam. Strong agreement was found between model predictions and empirical data, with r^2 values of 0.99 and 0.97, respectively (Thompson and Moin, 2003).

The hydrodynamic model RMA2 simulates water levels and flows for the upper St. Lawrence River based on inputs of water levels at Lake Ontario and discharge values from the Moses–Saunders Dam. For this study, the RMA2 simulation was run using the long-term average water level at the outlet of Lake Ontario from May to November (74.9 m above sea level, ASL) measured at Kingston, ON and the long-term average outflow at the Moses–Saunders dam (7580 m^3/s) for the period of May to November from 1960 to 2003. Species distribution

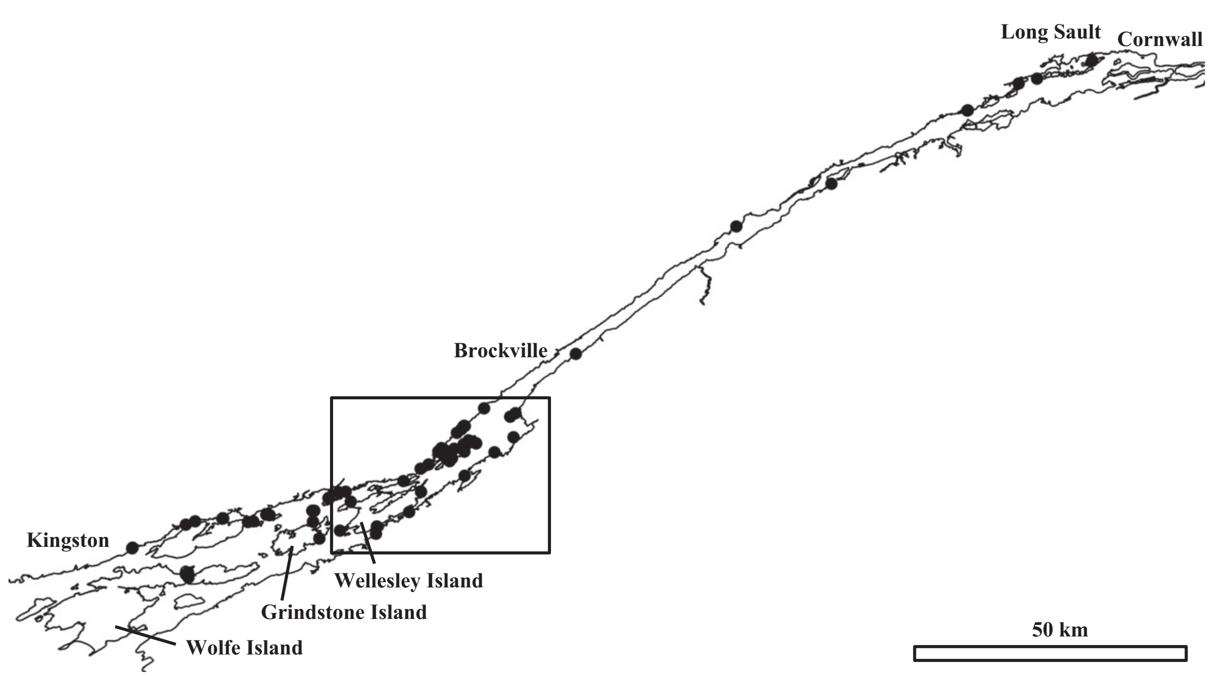

Fig. 1. The upper St. Lawrence River from Lake Ontario west of Wolfe Island to Cornwall, Ontario, illustrating the geographic extent of the study and the pugnose shiner locality data (black circles). The black square outlines the study region shown in Figs. 3 and 4.

modeling requires that environmental data are reflective of conditions during species sampling, the majority of which occurred during the past decade. Although we did not have access to outflow data from May to September at Moses–Saunders Dam during the pugnose shiner sampling period to verify this, the annual outflow over the past decade (7165 m^3/s, from 2002 to 2010) was only slightly lower than the long-term annual outflow from 1960 to 2005 (7345 m^3/s; Thompson et al., 2012). The long-term average water level (74.9 m ASL) was also comparable to the average water level during the pugnose shiner sampling period (74.8 m ASL from May to November, 2004–2011) measured at Cape Vincent, NY. Therefore, we concluded that the long-term averages were a good approximation for conditions during the sampling period and were appropriate for our purposes.

We estimated aquatic vegetation cover with a simple algorithm based on water depth and velocity values generated from the hydrodynamic model. Aquatic vegetation density was derived from depth-by-density relationships based on data from the Bay of Quinte on Lake Ontario (Minns et al., 2005) with a 5-m upper limit, which corresponds with field survey data on the average water clarity and presence/absence data for vegetation (Champoux et al., 2002; Leisti et al., 2006). A water velocity upper limit of 20 cm/s was chosen by Environment Canada, Quebec based on their field samples. Our vegetation model was intended to produce an estimate of aquatic vegetation density and is best viewed as a model of *potential* vegetation growth. The model is generic with respect to aquatic vegetation and is not intended to represent any particular plant species. The pugnose shiner is known to occupy a variety of vegetation types (DFO, 2010); however, a generic model aimed at predicting general vegetation growth was expected to be adequate for our needs. Although the pugnose shiner is predominantly associated with submerged aquatic vegetation, it can occur with both submerged and emergent communities (DFO, 2010); and our model did not distinguish between the two.

Point data for all our environmental variables (water depth, water velocity, vegetation) were interpolated to a grid using inverse distance weighting (IDW) with a 25-m resolution and the 'natural neighbor' algorithm in ArcGIS was used to examine the potential effect of the interpolation method chosen. The wetted-area grid bound by the shoreline of the upper St. Lawrence River (Minns et al., 2005) was used to define the spatial extent for all environmental layers in the GIS. Our vegetation model was generated based on field data from Lake Ontario but has not yet been fully validated in the St. Lawrence River. We evaluated whether the model was a good predictor of our field observations from the upper St. Lawrence River using all available data from Fisheries and Oceans Canada and Parks Canada. Field observations were qualitative assessments of percent vegetation cover. The aquatic vegetation model was evaluated by correlation analysis between model predictions and field observations as well as χ^2 tests to examine the performance of the model in distinguishing heavy vs. sparse cover.

Locality data

We compiled pugnose shiner locality data from government agencies in Canada and the United States including Fisheries and Oceans Canada (2004–2011), Parks Canada and Muskies Canada(2006–2010), and the New York State Department of Environmental Conservation and the United States Geological Survey (NYSDEC, USGS for 1989–2011) (Fig. 1). Although most samples were almost certainly mature adults (overall average TL ~45 mm), some sites, particularly towards the eastern end of the river, may have included young-of-the-year fish (19–27 mm). Sampling methods were similar across the three datasets. Sampling was conducted mainly by seining on foot using 30–50' bag seines with 1/4 in–1/8 in mesh size although some sites were sampled by boat with trawl nets, seine nets, or electrofishing. Much of the sampling targeted sites that were likely to yield pugnose shiner samples (i.e., slow-moving water with some vegetation) although at least one survey (USGS) randomly selected sampling sites along the American side of the river. The Parks Canada study focused on the Thousand Islands region where the pugnose shiner is generally found in greater abundance and which represents the largest amount of nearshore habitat in our study area. However, both the DFO and USGS conducted sampling throughout the river from Lake Ontario to the Moses–Saunders Dam in an effort to determine the extent of the distribution of the species.

Some presence localities had to be edited to conform to the GIS environmental layers of our study. Some presence localities that appeared on land were moved to be within the extent defined for the GIS. We edited locality data conservatively and only moved those data points that were within 100 m (usually much less) of the GIS grid. The presence localities were moved to the closest grid cell without knowledge of environmental information (i.e., high or low values of depth, velocity, vegetation) to avoid bias in the subsequent statistical analysis. In addition, 13 sites were deleted from the analysis as these were sampled from Picton Island in the Thousand Islands region that was not represented by the shoreline used. The DFO sampling records collected in 2004–2011 provided approximately 30 presence localities in the river. Sampling records from Parks Canada and Muskies Canada within the Thousand Islands portion of the St. Lawrence River provided approximately 40 presence localities, and NYSDEC and USGS provided approximately 20 presence localities. After considering redundancies across years and among government agencies, we compiled 79 distinct localities for use in our study.

MaxEnt analysis

Choice of method and assumptions

We modeled the distribution of the pugnose shiner using MaxEnt (Maximum Entropy Species Distributional Modeling, Version 3.3.3k), a presence-only, machine-learning method based on maximum entropy (Phillips et al., 2006). Although our locality data were taken mainly from surveys conducted by government agencies that provided some records of 'absences', we chose to use a presence-only method for a variety of reasons. Absence data have an inherently higher degree of uncertainty than presence data, which is especially true when detection rates are poor (MacKenzie and Royle, 2005; Lobo et al., 2010). The pugnose shiner is relatively difficult to detect and the sampling effort required to assess 'absence' with 99% confidence has been estimated to be seven seine hauls with a 61 m net (Dextrase et al., 2014). Very few (if any) sites in our surveys were sampled with that level of effort. Moreover, given low detection rates, many of our absence localities were highly interspersed with presence localities. If we were to include the absence data, deciding which data to include would not be straightforward and our selection would undoubtedly affect our results (Lobo et al., 2010). We realize that by using a presence-only approach, we have not entirely avoided the issue of low detectability in our data, as the background area is treated as pseudo-absence (see Elith et al., 2011). However, we do avoid the subjectivity of selecting which absence records to include.

Like all species distribution modeling methods, MaxEnt assumes that all environmental variables have ecological importance to the species and that the species' distribution is limited by the environmental variables rather than dispersal, both of which are satisfied in our study. Another assumption is that environmental variables are not highly correlated with one another although MaxEnt is known to be fairly robust to correlations among environmental variables (Elith et al., 2011). We evaluated correlations among variables at our presence localities and found that vegetation and depth values were highly correlated at presence sites ($r > 0.8$). Therefore, we were careful to compare results based on all three variables to those based on only two variables that did not show strong correlations (i.e., depth and velocity; vegetation and velocity). In addition, MaxEnt assumes that sampling effort is relatively unbiased with respect to the study area. As the inclusion of deeper regions of the river had the potential to artificially inflate

Table 1
Chi-square results showing a significant association between model predictions of aquatic vegetation and observations from field surveys. We divided the both the model predictions and the observed data into two categories: 'open' (<50% cover) and 'vegetated' (≥50% cover).

Observed	Field observations			
		<50%	≥50%	
Model predictions	<50%	16	72	88
	≥50%	10	127	137
		26	199	225

Expected	Field observations		
		<50%	≥50%
Model predictions	<50%	10	78
	≥50%	16	121
		p-value	0.013

statistical significance, we minimized sampling bias by eliminating deeper parts of the study area that were not well sampled (as recommended by Elith et al., 2011). We chose three restricted depth ranges to evaluate model performance (see below).

Running the model

We performed four separate analyses with MaxEnt: the whole river analysis that included all depths, and three additional analyses in which we progressively eliminated depth-based sections of the river. First, we chose a 0–5-m depth range as 5 m was identified as the upper limit for vegetation growth based on observational data in similar systems (Champoux et al., 2002; Leisti et al., 2006). We then used the more restrictive depth range of 0–3 m, as pugnose shiner adults are generally found in habitat <3-m in depth (DFO, 2011). Finally, we chose an even more restrictive range of 0–2 m as much of the sampling was conducted on foot and occurred at depths of <2 m. However, restricting the depth range entailed a loss of presence localities for analysis (from 79 localities in the whole river analysis to 71 localities in the 0–5-m depth range; 57 localities in the 0–3-m depth range; and 45 localities in the 0–2-m depth range). Analyses were performed with default settings in MaxEnt, using 75% of the data for training, and 25% for testing, with 10 replicates.

Evaluation of model performance

We evaluated model performance in several ways. First, we assessed the area under the receiver operating curve (AUC), a threshold-independent means of evaluating statistical support (Phillips et al., 2006). The AUC represents the fit of the model and represents the ability of the model to distinguish between locations where the species is present and locations where the species is absent. As MaxEnt is a presence-only method, the program compares presence sites with background sites rather than by comparing presence sites with absence sites (Phillips et al., 2006). AUC varies from 0 to 1, with 0.5 being no better than random. Although AUC has been criticized, in part, due to its dependence on the particular background area selected (Lobo et al., 2008), it remains a valid means of comparing models within a particular study area (Wisz et al., 2008). Models with an AUC of 0.9 are often considered outstanding and those of 0.7 are acceptable (Hosmer and Lemeshow, 2000). Elith et al. (2006) suggested that an AUC value of 0.75 indicates a useful model for understanding species distributions with MaxEnt. Second, we evaluated gain, which evaluates how much higher model predictions are at presence localities compared with randomly chosen sites (Phillips et al., 2006). Third, we evaluated the 1-tailed binomial probability that our model predicted the test data no better than random using the 'maximum test sensitivity plus specificity' threshold. This threshold has been found to perform well compared with other threshold approaches (Liu et al., 2005).

We examined the role of each environmental variable within our models in several ways. We visually examined response curves of all variables and evaluated the percent contribution and percent importance of each variable in the combined model. Percent contribution is defined heuristically and depends on the path chosen by MaxEnt rather than the final result. The percent importance is a more accurate assessment of each variable as it depends on the final model rather than the path taken. Percent importance is calculated by randomly permuting values for a particular variable among sites and assessing the decrease in the model's performance. We also conducted jackknife tests to evaluate model performance with subsets of variables. Jacknife tests assess performance of models (e.g., test AUC, gain) using each variable individually and after excluding each variable in turn. In addition, we evaluated model consistency by spatially comparing model predictions between those based on three environmental variables and those based on two variables (vegetation and velocity; depth and velocity). We assessed consistency across model predictions by comparing which cells were categorized as above or below a given threshold (arbitrarily chosen as 0.5) within a particular depth range (see Hernandez et al., 2006). Although this method does not measure model accuracy, per se, it does provide an estimate of model stability.

Finally, we used the results of our models to estimate the total area of suitable habitat in the study area. Presence-only methods produce a ranking of habitat suitability within the study area, but they cannot provide reliable estimates of species presence or absence because they lack information on prevalence (frequency of species occurrence within the study area) (Elith et al., 2011). Although results should be interpreted with caution, we estimated the amount of suitable habitat in our study area by applying threshold values to convert the continuous relative probability data to a binary estimate of suitable and unsuitable habitat. Phillips et al. (2006) argued that the choice of threshold should depend, in part, on the overall purpose of the assessment. If the intention is to identify areas for future sampling, a lower threshold may be useful; however, if the intention is to quantify suitable habitat for the protection of endangered species, a higher threshold value may be more appropriate. We used a range of threshold values to approach this question.

Results

Validation of the vegetation model

Our ability to statistically assess the vegetation model was limited due to the abundance of vegetation in most field observations. This may reflect the tendency of field crews to select the largest vegetation patches within a particular site for sampling. In any case, the correlation analysis between model predictions and field observations was not significant. However, a χ^2 test comparing broad categories within the field data (e.g., predominantly open vs. predominantly vegetated) with broad categories of vegetation cover predicted by the model (<50% cover vs. >50% cover) was significant ($p < 0.05$) demonstrating qualified support for the vegetation model (Table 1).

MaxEnt analysis

Species distribution models based on all three environmental variables (vegetation, velocity, depth) performed significantly better than random across all depth ranges. AUC values were highest for the 'whole river' analysis (AUC ~0.90 ± 0.04) and declined slightly as the depth range decreased (0.80 ± 0.06 for 0–5-m depth range; 0.75 ± 0.10 for 0–3-m depth range; and, 0.74 ± 0.08 for 0–2-m depth range), with very similar results for models based on two variables only (Table 2). Test gain was also positive across all analyses, with the highest values ranging from 0.30–1.39 across depth ranges (Table 2).

Table 2
MaxEnt jackknife results. AUC and gain statistics for 'test' data across all possible models (i.e., those based on all variables, all combinations of two variables, and all single variables) and depth ranges.

	n	All variables	Vegetation + velocity	Vegetation + depth	Velocity + depth	Vegetation only	Depth only	Velocity only
		AUC details						
		Whole river						
Ave	79	0.90	0.90	0.91	0.89	0.90	0.89	0.76
Stdev		0.04	0.06	0.06	0.06	0.07	0.06	0.09
		0–5 m						
Ave	71	0.80	0.78	0.80	0.73	0.78	0.72	0.67
Stdev		0.06	0.12	0.09	0.16	0.11	0.13	0.14
		0–3 m						
Ave	57	0.75	0.75	0.74	0.73	0.74	0.72	0.66
Stdev		0.10	0.11	0.07	0.12	0.07	0.09	0.15
		0–2 m						
Ave	45	0.74	0.74	0.67	0.74	0.67	0.65	0.70
Stdev		0.08	0.16	0.15	0.15	0.17	0.14	0.12
		Gain details						
		Whole river						
Ave	79	1.39	1.34	1.33	1.26	1.28	1.19	0.44
Stdev		0.67	0.71	0.65	0.62	0.70	0.56	0.37
		0–5 m						
Ave	71	0.60	0.50	0.58	0.37	0.48	0.31	0.13
Stdev		0.72	0.66	0.52	0.69	0.50	0.48	0.37
		0–3 m						
Ave	57	0.39	0.43	0.32	0.38	0.35	0.28	0.16
Stdev		0.42	0.42	0.23	0.39	0.22	0.24	0.30
		0–2 m						
Ave	45	0.28	0.28	0.08	0.30	0.08	0.10	0.23
Stdev		0.54	0.56	0.37	0.48	0.37	0.23	0.31

Given the mathematical interpretation of gain ($e^{(gain)}$), these values suggest that the predicted likelihood of suitable habitat at presence localities was approximately four times higher than at background sites for 'whole river', and 1.35 times higher for the most limited extent (0–2-m depth range). The threshold-dependent binomial tests were

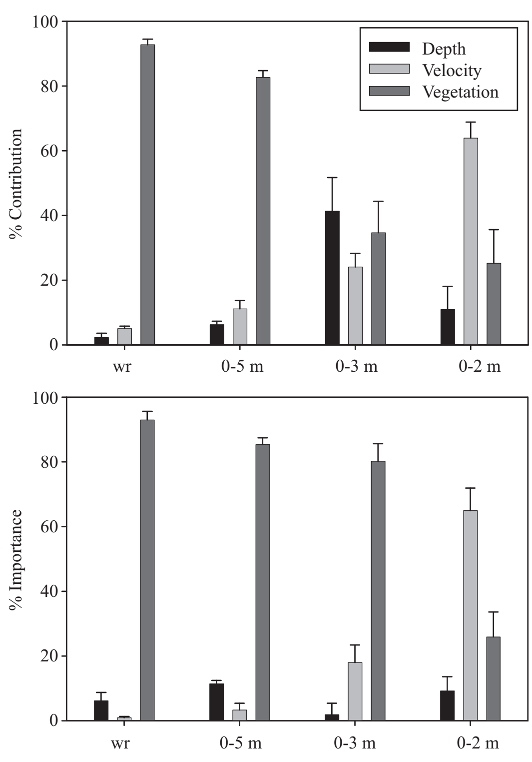

Fig. 2. Percent contribution (top) and percent importance (bottom) of depth, velocity, and vegetation in the species distribution models from MaxEnt across the four different analyses: the whole river (wr), and the three restricted depth ranges: 0–5 m, 0–3 m, and 0–2 m.

also significant across all depth ranges using the maximum test sensitivity plus specificity logistic threshold. The logistic threshold chosen by the program ranged from 0.34 for the 'whole river' ($p < 0.001$) to 0.49 in the 0–2-m depth range ($p < 0.05$). Binomial test results were also significant for models based on vegetation and velocity only (not shown). These results were based on the inverse distance weighting interpolation in ArcGIS (using default settings); similar results were found for the natural neighbor interpolation (Electronic Supplementary material Table S1).

With the exception of the 0–2-m depth range, vegetation was the single most important variable, producing the highest AUC and gain values when each variable was analyzed alone (Table 2) and the highest importance values within the model based on all three variables (Fig. 2). However, when the background extent was restricted to the 0–2-m depth range, vegetation was outperformed by velocity (Table 2). The results were the same on a percent contribution or percent importance basis (Fig. 2). As the depth range of the background extent contracted, velocity became an increasingly important variable in predicting pugnose shiner presence in shallow vegetated areas. Response curves indicated that habitat suitability increased as depth decreased, vegetation increased, and water velocity decreased, as expected. Complex interactions among our environmental variables were not expected, nor were they found. Response curves remained consistent across models and depth ranges, with the exception of depth, which occasionally showed unexpected patterns. However, this only occurred within three-variable models when depth had low contribution and low importance and consequently did not have much of an impact on the model. Within two-variable models (i.e., with depth and velocity), depth showed both greater importance and greater consistency.

We created species distribution maps for each analysis (i.e., 'whole river', and the three depth ranges) using predictions from the models based on all three variables and using all locality data (rather than withholding 25% for testing, although this made little difference) (Fig. 3). We also explored model uncertainty by creating species distribution maps for the four most highly supported models in the 0–3 m depth range (see Table 2). Visual inspection demonstrated that species distribution maps were very similar across models (Fig. 4), which was confirmed

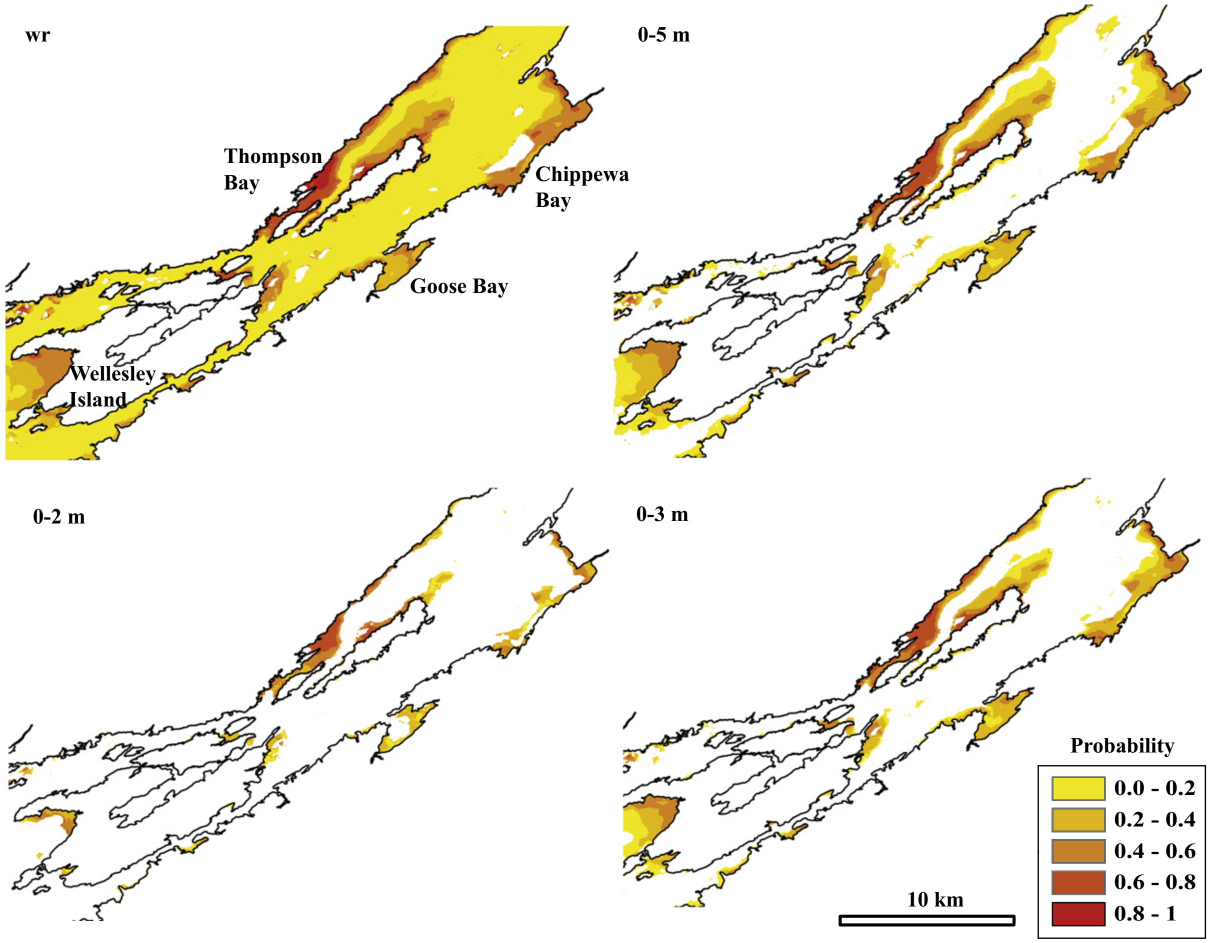

Fig. 3. MaxEnt predicted distribution of the pugnose shiner across four different depth ranges (e.g. whole river, 0–5 m, 0–3 m, and 0–2 m) based on all environmental variables. Probability of habitat suitability is shown with darkest shading indicating higher probability and lighter shading indicating lower probability.

by quantitative analysis. When we quantified consistency (i.e., whether cells were categorized above or below a threshold value of 0.5), model predictions based on only vegetation and velocity showed 95–99% similarity to those based on all three environmental variables across depth ranges. Similarly, model predictions based on depth and velocity showed 94–97% similarity to those based on all three environmental variables across depth ranges.

Finally, we quantified suitable habitat for the pugnose shiner based on the most highly supported species distribution models using a range of threshold values. We focused on the depth-restricted analyses (0–2-m, 0–3-m, 0–5-m depth ranges) to reduce the risk of overestimation. We chose 0.5 as the lower threshold as this was generally similar to the 'maximum training sensitivity plus specificity' threshold identified by MaxEnt which varied from ~0.4–0.5 across models. As this threshold may be too low (see Discussion section), we contrasted this with more conservative thresholds of 0.6 and 0.7. Based on these threshold values, the quantity of suitable habitat varied from ~2 km^2 (using a threshold of 0.7 and the 0–2-m depth range) to ~30 km^2 (using a threshold of 0.5 for the 0–5-m depth range).

Discussion

Our results provided strong support for the power of depth, velocity, and vegetation models to predict the distribution of the pugnose shiner in the upper St. Lawrence River. We found strong support for models based on all three environmental variables as well as models based on two variables. Vegetation was the single most important variable in the study in all analyses with the exception of the most restricted depth range (0–2 m). As depth range was contracted, the performance and the importance of the vegetation model declined. The weaker performance of the vegetation model at the 0–2-m depth range may have reflected limited variation for this variable at this scale. Nevertheless, even at this depth range, the model was still significant because water velocity was still a determinant of pugnose shiner distribution. Our results suggested that the hydrological model as well as the vegetation model used in this study provided valuable tools for the identification of suitable habitat.

We provided multiple models for predicting the distribution of the pugnose shiner in the upper St. Lawrence River based on different depth ranges. The 0-2-m depth range may be instructive as an indication of model effectiveness, but this analysis used fewer data points and may be unnecessarily restrictive, particularly for the level of resolution (and potential error) associated with our dataset. Therefore, we suggest that the 0–3-m or 0–5-m depth ranges may provide the most appropriate balance between restricting the geographic range considered and retaining locality data.

Models based on different subsets of environmental data were generally similar to one another and the best model was not significantly better than other models. To determine which model was the most accurate estimate of suitable habitat for the pugnose shiner, additional sampling is warranted in areas where the model predictions differ. Specifically, areas, such as Goose Bay, bays around Wolfe Island, and island

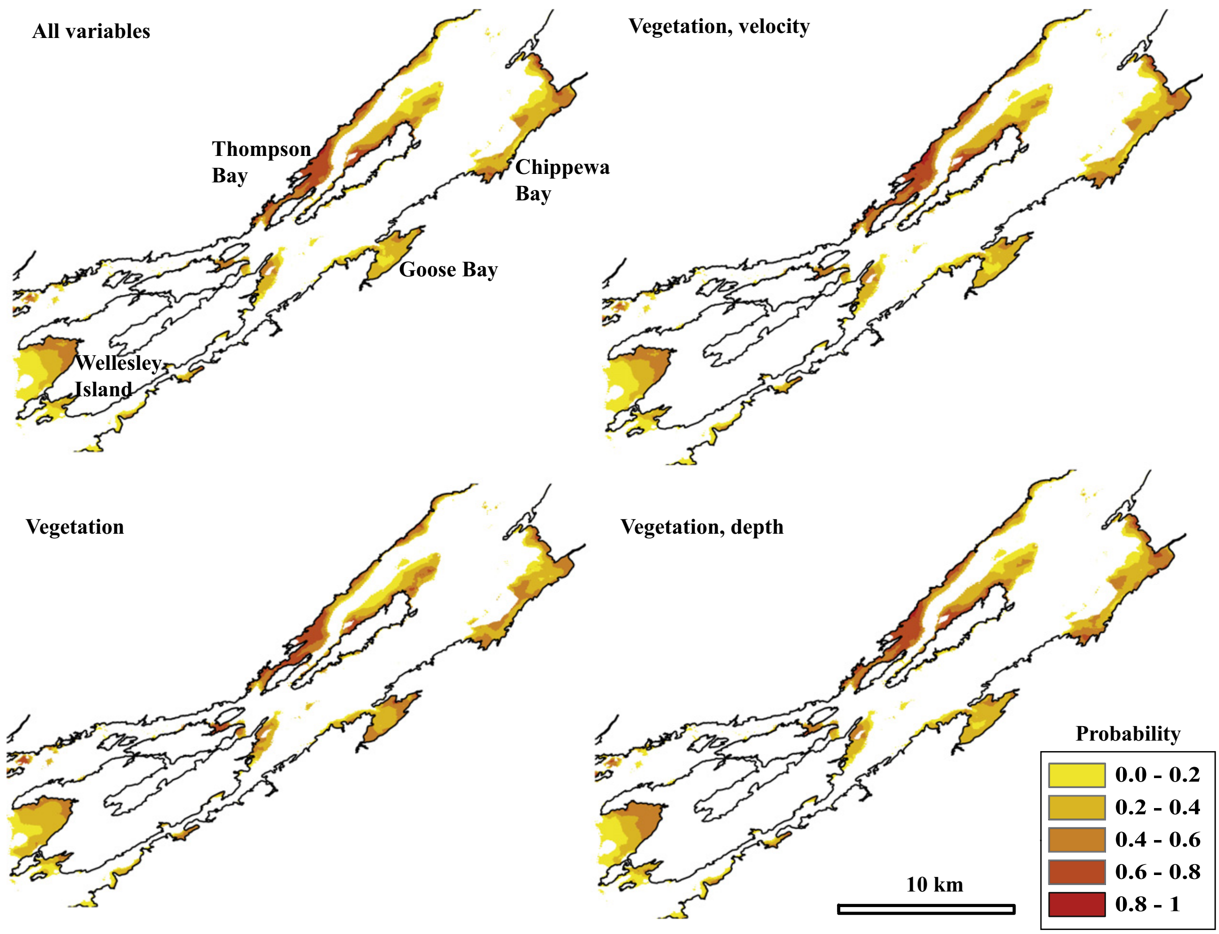

Fig. 4. MaxEnt predicted distribution of the pugnose shiner across four of the best performing models (e.g. based on all variables, vegetation and velocity, vegetation and depth, and vegetation only) in the 0-3 m depth range. Probability of habitat suitability is shown with darker shading indicating higher probability and lighter shading indicating lower probability.

habitat under the Long Sault Parkway, warrant further investigation to help gauge the true extent of the species distribution. Interestingly, models consistently showed a high probability of suitable habitat in certain areas that were likely not extensively sampled. Further sampling is warranted in several of these areas, including Eel Bay on western Wellesley Island, Flynn Bay on Grindstone Island, and many of the bays around Wolfe Island. These sites should be targeted to help refine models and perhaps extend the known distribution of the species in the upper St. Lawrence River. The identification of new populations could substantially influence conservation status and management practices in endangered species protection.

We used our estimates of habitat suitability to provide a first approximation of the total amount of suitable habitat in the upper St. Lawrence River. Our results indicate an abundance of suitable habitat within the river (~2 km^2 to 30 km^2), at least with respect to the minimum viable area for this species. The minimum area required for a viable population (MAVP) for the pugnose shiner was estimated to be 0.015 km^2 in rivers and 0.05 km^2 in lakes (DFO, 2010; Venturelli et al., 2010). However, we caution that thresholds based on ROC plots (such as 'maximum test sensitivity plus specificity') are often too low when species are rare, resulting in an overestimate of the amount of suitable habitat (Manel et al., 2001). As rarity of the species likely increases with increasing depth ranges, the amount of suitable habitat within the 0–5-m depth range may have been overestimated. A second caveat is that our model of potential vegetation growth assumes a relatively even distribution of vegetation. If vegetation growth is very patchy, our model will overestimate the amount of suitable habitat. Surveys are needed to estimate the proportion of vegetation cover at suitable sites within our study area to better estimate habitat availability.

The number of populations in the river should also be considered when comparing estimates of habitat suitability to minimum viable area for a population. Previous population genetic analyses showed significant differentiation between the Thousand Islands region and downstream regions (separated by ~50 km) (McCusker et al., 2014). Therefore, we also provide separate estimates for the amount of habitat available in the Thousand Islands region (upstream of Brockville) and that in the downstream region (from Brockville to Cornwall). Using the most conservative estimates of suitable habitat (0–2 m depth range, threshold of 0.7), our results indicate ~1.5 km^2 of suitable habitat in the Thousand Islands region, but a smaller amount of suitable habitat in the downstream region (0.4 km^2). If additional population differentiation exists in the river, suitable habitat availability per population would decline further. Although it appears that habitat is not limiting in the St. Lawrence River, this conclusion depends on assumptions such as a uniform distribution of vegetation in suitable areas and limited population structure.

Our study provided several predictions of suitable habitat for the pugnose shiner in the upper St. Lawrence River, however, our models could be further refined in a number of important ways. First, finer-scale surveys of the upper St. Lawrence River, which may be conducted in the future, could improve model predictions. Second, a well-structured and standardized vegetation field survey could provide validation and further refinement of the vegetation model. It would also help to quantify patchiness of vegetation and clarify the actual amount of suitable habitat. Third, the inclusion of substrate data (when available) could also help to refine the vegetation model by potentially providing improved predictions of quantity and quality (type) of aquatic vegetation. Fourth, we recommend additional sampling, particularly targeted sampling of younger fish and novel sampling techniques to determine whether the pugnose shiner utilizes separate overwintering habitat, which could be incorporated into future ecological niche assessments. Finally, interactions among species have long been recognized as an important factor in determining realized distributions (Newbold, 2010); and, in future, efforts could be made to incorporate presence/absence of sympatric species as well. Our model predicted the mesoscale distribution of the pugnose shiner, but future refinements have the potential to produce greater precision and finer-scale resolution of the species' distribution.

The upper St. Lawrence River is thought to provide some of the best habitat for the pugnose shiner within Ontario and possibly across its range. The region is, therefore, highly important for future survival of the species as a whole. Habitat loss and degradation represent the largest threat to pugnose shiner populations (DFO, 2010), and knowledge of a species' distribution represents an essential first step towards assessing risk of habitat perturbation and focusing recovery actions (Peterson, 2006). General population viability analyses that quantify extinction risk based on demographic parameters and habitat size have been conducted for the pugnose shiner (Venturelli et al., 2010). However, these population viability analyses have not yet been made spatially explicit by linking them to the species' distribution and its associated habitat (see Hayes et al., 2009; Minns et al., 2005). Our study can help bridge this gap, which would facilitate further understanding of habitat-related threats, including changes to hydrology, increased sedimentation and turbidity, aquatic vegetation removal, and climate change (Doka et al., 2006; Gertzen et al., 2012). Spatially explicit population viability analysis could also lead to a better understanding of critical habitat for this species, which is protected under federal law in Canada, but has not yet been comprehensively identified for specific protection. An improved understanding of the distribution and extent of suitable habitat for the pugnose shiner in the upper St. Lawrence River should help focus recovery actions that will lead to the long-term viability of this endangered species.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jglr.2014.09.014.

Acknowledgments

Funding for this study was provided by the Species at Risk program of Fisheries and Oceans Canada (DFO), and an NSERC Discovery Grant (RGPIN 238506) to NRL. M.R. McCusker was supported by the NSERC Visiting Fellowship Program. We are grateful to the many people who helped with the sampling in the St. Lawrence River whose efforts made this study possible and to anonymous reviewers whose comments improved the manuscript.

References

Abell, R., 2002. Conservation biology for the biodiversity crisis: a freshwater follow-up. Cons. Biol. 16, 1435–1437.

Champoux, O., Morin, J., Jean, M., 2002. Modèlisation de la végéetation émergente du lac Saint-Pierre. Rapport Scientifique SMC-Hydrologie RS-103 Environement Canada, (Sainte-Foy, 43 pp.).

COSEWIC, 2013. COSEWIC Assessment and Update Status Report on the Pugnose Shiner *Notropis anogenus* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa (Vi + 15 pp.).

Dextrase, A.J., Mandrak, N.E., Barnucz, J., Bouvier, L.D., Gaspardy, R., Reid, S.M., 2014. Sampling effort required to detect fishes at risk in Ontario. Can. Manuscr. Rep. Fish. Aquat. Sci. 3024 (v + 50 p.).

DFO, 2010. Recovery potential assessment of the pugnose shiner (*Notropis anogenus*) in Canada. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 1–13.

Doeringsfeld, M.R., 1993. Notes on the Life History of the Pugnose Shiner *Notropis anogenus* and Current Status in Minnesota. Gustavus Adolphus College, St. Peter, MN, pp. 1–20.

Doka, S., Bakelaar, C., Bouvier, L., 2006. Coastal wetland fish community assessment of climate change in the lower Great Lakes. In: Mortsch, L., Ingram, J., Hebb, A., Doka, S. (Eds.), Great Lakes Wetland Communities: Vulnerability to Climate Change and Response to Adaptation Strategies. Environment Canada and the Department of Fisheries and Oceans, Toronto, Ontario, pp. 101–128.

Dudgeon, D., Arthington, A.H., Gesser, M.O., Kawabata, Z.-I., Knowler, D.J., Leveque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. 81, 163–182.

Elith, J., Graham, C.H., Anderson, R.P., Dudı́k,, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton,, J.McC., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón,, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.

Gertzen, E.L., Doka, S.E., Minns, C.K., Moore, J.E., Bakelaar, C., 2012. Effects of water levels and water level regulation on the supply of suitable spawning habitat for eight fish guilds in the Bay of Quinte, Lake Ontario. Aquat. Ecosyst. Health Manag. 15, 397–409.

Hayes, D., Jones, M., Lester, N., Chu, C., Doka, S., Netto, J., Stockwell, J., Thompson, B., Minns, C.K., Shuter, B., Collins, N., 2009. Linking fish population dynamics to habitat conditions: insights from the application of a process-oriented approach to several Great Lakes species. Rev. Fish Biol. Fish. 19, 295–312.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, M.D., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29, 773–785.

Holm, E., Mandrak, N.E., 2002. Update COSEWIC Status Report on the Pugnose Shiner *Notropis anogenus* in Canada in COSEWIC Assessment and Update Status Report on the Pugnose Shiner *Notropis anogenus* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa (1–15 pp.).

Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression, 2nd ed. Wiley, New York; Toronto.

Hudon, C., 2000. Phytoplankton assemblages in the St. Lawrence River, downstream of its confluence with the Ottawa River, Quebec, Canada. CJFAS 57 (Suppl. 1), 16–30.

IUCN, 2014. The IUCN Red List of Threatened Species. Version 2014.2http://www.iucnredlist.org, Downloaded on 29 July 2014.

Jelks, H.L., Walsh, S.J., Burkhead, N.M., Contreras-Balderas, S., Diaz-Pardo, E., Hendrickson, D.A., Lyons, J., Mandrak, N., McCormick, F., Platania, S.P., Porter, B.A., Renaud, C.B., Schmitter-Soto, J.J., Taylor, E.B., Warren, M.L.J., 2008. Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries 33, 372–407.

Knouft, J., Caruso, N., Dupre, P., Anderson, K., Trumbo, D., Puccinelli, J., 2011. Using fine-scale GIS data to assess the relationship between intra-annual environmental niche variability and population density in a local stream fish assemblage. Meth. Ecol. Evol. 2, 303–311.

Leisti, K.E., Millard, E.S., Minns, C.K., 2006. Assessment of submergent macrophytes in the Bay of Quinte, Lake Ontario, August 2004, including historical context. Can. MS Rep. Fish. Aquat. Sci. 2762 (ix + 84p.).

Li, X., Wang, Y., 2013. Applying various algorithms for species distribution modelling. Int. Zool. 33, 124–135.

Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28, 385–393.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151.

Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography 33, 103–114.

Mackenzie, D.I., Royle, A., 2005. Designing occupancy studies: general advice and allocating survey effort. J. Appl. Ecol. 42, 1105–1114.

Mandrak, N., Cudmore, B., 2010. The fall of native fishes and rise of non-native fishes in the Great Lakes basin. Aquat Ecosys Health & Man. 13, 255–268.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. App. Ecol. 38, 921–931.

McCusker, M.R., Mandrak, N.E., Egeh, B., Lovejoy, N.R., 2014. Population structure and conservation genetic assessment of the endangered Pugnose Shiner, *Notropis anogenus*. Conserv. Genet. 15, 343–353.

Minns, C.K., Doka, S., Bakelaar, C., Chu, C., Leisti, K., Moore, J.E., 2005. Lake Ontario–St. Lawrence Water Regulation Study: Year 4 Final report for Burlington Fish Habitat & Modelling Group. International Joint Commission, MOU-IJC-4-M009.

NatureServe, 2014. NatureServe Explorer: an online encyclopedia of life [web application]. Version 7.1. NatureServe, Arlington, VirginiaAvailable http://explorer.natureserve.org (Accessed: July 29, 2014).

Newbold, T., 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Prog. Phys. Geog. 34, 3–22.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Peterson, A.T., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34, 102–117.

Peterson, A.T., 2006. Uses and requirements of ecological niche models and related distributional models. Biodivers. Inf. 3, 59–72.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.

Thompson, A.F., Moin, S.M., 2003. Hydrodynamic modelling of the Upper St. Lawrence River, Kingston/Cape Vincent to Cornwall Reach. Proceedings, 16th Canadian Hydrotechnical Conference, CSCE. The Canadian Society for Civil Engineering, Montreal, Burlington, Ontario (October 22–24, 2003).

Thompson, A., Stevenson, D., Morin, J., Bouchard, A., Champoux, O., 2012. The Upper St. Lawrence River: Existing Data, Past work, and Future Modelling Applications. Environment Canada. Boundary Water Issues. MSC Ontario, Hydrology and Ecohydraulic Section, MSC Quebec, pp. 1–19.

Tittensor, D.P., Baco, A.R., Brewin, P.E., Clark, M.R., Consalvey, M., Hall-Spencer, J., Rowden, A.A., Schlacher, R., Stocks, K.I., Rogers, A.D., 2009. Predicting global habitat suitability for stony corals on seamounts. J. Biogeogr. 36, 1111–1128.

Venturelli, P.A., Velez-Espino, L.A., Koops, M.A., 2010. Recovery potential modelling of pugnose shiner (*Notropis anogenus*) in Canada. DFO Can. Sci. Advis. Sec. Res. Doc. (2010/007. iv + 22 pp.).

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. Div. Dist. 14, 763–773.