# Predicting smallmouth bass (*Micropterus dolomieu*) occurrence across North America under climate change: a comparison of statistical approaches

**Sapna Sharma and Donald A. Jackson**

**Abstract:** Smallmouth bass (*Micropterus dolomieu*) is a warm-water fish species that is native to central and eastern North America. Climate change scenarios predict further extension northward of suitable habitat for smallmouth bass, which may negatively affect native fish species. We developed and compared predictive models of the distribution of bass in North America using four statistical approaches: logistic regression, classification tree, discriminant analysis, and artificial neural networks. We collected 4181 geo-referenced records of smallmouth bass occurrence and matched them with climate data. Artificial neural networks performed the best with the highest sensitivity (correctly predicting species presence) and specificity (correctly predicting absence), followed by discriminant analysis. Artificial neural networks indicated that winter air temperatures were the most important predictors of smallmouth bass occurrence, whereas the other approaches indicated that summer air temperatures were the best predictors of bass occurrence. Logistic regression and classification tree exhibited very low sensitivity, but very high specificity as a result of the large proportion of absences within the data set. Business-as-usual climate change scenarios suggest that smallmouth bass are expected to have suitable thermal habitat throughout most of Canada and the continental United States by 2100.

**Résumé :** L'achigan à petite bouche (*Micropterus dolomieu*) est une espèce de poisson d'eau chaude indigène dans le centre et l'est de l'Amérique du Nord. Les scénarios de changement climatique prédisent une extension additionnelle vers le nord de la répartition des habitats adéquats pour l'achigan à petite bouche, ce qui pourrait affecter négativement les espèces indigènes de poissons. Nous avons mis au point et comparé des modèles prédictifs de la répartition de l'achigan à petite bouche en Amérique du Nord d'après quatre méthodologies statistiques, soit la régression logistique, l'arbre de classification, l'analyse discriminante et les réseaux neuronaux artificiels. Nous avons assemblé 4181 points géoréférencés de la présence d'achigans à petite bouche et les avons associés à des données climatiques. Les réseaux neuronaux donnent le meilleur résultat avec la sensibilité (prédiction exacte de la présence de l'espèce) et la spécificité (prédiction exacte de l'absence) les plus grandes; vient ensuite l'analyse discriminante. Les réseaux neuronaux artificiels indiquent que les températures de l'air en hiver sont les variables explicatives les plus importantes de la présence des achigans à petite bouche, alors que les autres méthodes désignent les températures de l'air en été comme les meilleures variables explicatives de la présence des achigans. La régression logistique et l'arbre de classification possèdent une sensibilité très basse, mais une spécificité très élevée, le résultat d'un fort pourcentage d'absences dans la banque de données. Les scénarios de changement climatique de type « laisser-aller » indiquent que, vers l'an 2100, les achigans à petite bouche devraient pouvoir trouver des habitats thermiques adéquats sur presque l'ensemble des territoires du Canada et des États-Unis continentaux.

[Traduit par la Rédaction]

## Introduction

Smallmouth bass (*Micropterus dolomieu*) is a warm-water fish species native to central and eastern North America (Scott and Crossman 1973). Since the mid-1800s, the range of the smallmouth bass has expanded across North America and throughout Europe, Russia, and Africa as a result of introductions to provide angling opportunities (Scott and Crossman 1973). Lake stocking, unauthorized introduction by anglers, and dispersal through drainage networks has facilitated the range expansion of the smallmouth bass (Jackson 2002; Vander Zanden et al. 2004*a*; Dunlop and Shuter 2006). Although smallmouth bass is a popular angling species, the introduction of smallmouth bass has had negative consequences on native fish communities. For example, the presence of smallmouth bass has been associated with re-

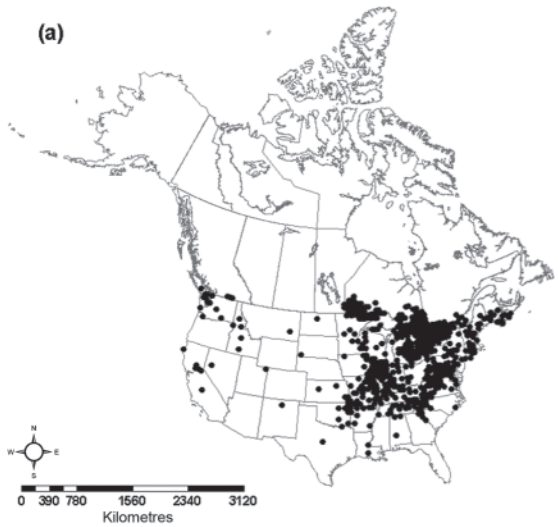
Received 12 January 2007. Accepted 24 August 2007. Published on the NRC Research Press Web site at cjfas.nrc.ca on 14 February 2008.
J19758

**S. Sharma[1] and D.A. Jackson.** Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3G5, Canada.

[1]Corresponding author (e-mail: ssharma@zoo.utoronto.ca).

 doi:10.1139/F07-178 

**Fig. 1.** (*a*) Locations of smallmouth bass (*Micropterus dolomieu*) occurrence in Canada and the United States collected for this study. (*b*) Training and validation smallmouth bass occurrence data sets used in the statistical analyses. Only smallmouth bass presence is indicated on the map. Training data and validation data are represented by grey and black circles, respectively.




duced numbers of cyprinids (MacRae and Jackson 2001; Jackson 2002). The presence of smallmouth bass has been shown to cause lake trout (*Salvelinus namaycush*) to experience reduced growth, survival, and fecundity because of a diet shift from cyprinids to energetically inefficient prey items such as zooplankton in the absence of pelagic forage fish (Vander Zanden et al. 1999).

Temperature plays a crucial role for different life stages of smallmouth bass, including likelihood, timing, and success of spawning (Shuter et al. 1980; Rejwan et al. 1999); activity levels and growth in young of the year; and overwintering survival (Shuter et al. 1980). Year-class strength is strongly determined by growth and survival rates of smallmouth bass broods (Rejwan et al. 1997; Shuter and Ridgway 2002), which are strongly correlated to thermal conditions (Casselman 2002; Casselman et al. 2002; Shuter and Ridgway 2002). Smallmouth bass growth, particularly for young individuals, had the strongest relationship with air temperature from a variety of climatic variables (Dunlop and Shuter 2006).

Climate change has the potential to greatly influence smallmouth bass populations. Increases in water temperatures are predicted to have a dramatic impact on water quality and availability of suitable fish habitat (Magnuson et al. 1990; Magnuson 2002). As temperature increases and the current thermal regime is shifted northwards, smallmouth bass populations are also expected to follow the northerly shift in thermal habitat (Jackson and Mandrak 2002; Shuter et al. 2002; Sharma et al. 2007). Currently, smallmouth bass are restricted from these northerly areas because of cool thermal habitat that prevents successful overwintering of the young of the year (Shuter et al. 2002).

The objective of our study was to predict smallmouth bass incidence based on climate data using four statistical approaches. Statistical approaches that are used with binary data were compared to evaluate their ability at predicting a single, binary response variable based on a number of predictor variables. The four statistical methods compared were multiple logistic regression, classification tree, linear discriminant analysis, and artificial neural networks. Comparison of predictive statistical approaches is an important consideration because the efficacy of a statistical approach is dependent upon the characteristics of the data set (e.g., model distribution, parametric assumptions, and interactions between predictor variables). The comparison will determine which statistical approach is the most appropriate for the data, and comparisons across methods can provide additional insight into underlying mechanisms and relationships. Based on the most appropriate statistical approach and business-as-usual climate change scenarios, we predicted smallmouth bass distribution in 2100.

## Materials and methods

### Data acquisition

We collected 4181 geo-referenced records of smallmouth bass incidence in North America (Fig. 1*a*) from a variety of sources, including FishBase (www.fishbase.org/home.htm), Fishnet, Florida Natural History Museum, University of Michigan Museum of Zoology, Illinois Natural History Survey, Ontario Fish and Habitat Inventory Index, refereed publications, and dissertations.

We compiled air temperature data for each month covering Canada and the continental United States. Climate data were obtained from the Intergovernmental Panel on Climate Change Data Distribution Centre as 1961–1990 averages. The data were interpolated from meteorological stations using thin-plate splines and summarized on a 0.5° latitude by 0.5° longitude grid.

The climate change model that was selected for this study was the Canadian General Circulation Model Version 2 (CGCM2), a conservative climate change model. We chose the "business-as-usual scenario" (IS92a) and obtained monthly air temperature data for the year 2100 from the Canadian Centre for Modelling and Analysis. Future projections in air temperature were summarized at a 3.75° × 3.75° grid level.

Using the smallmouth bass and climate data, we summarized the information of smallmouth bass incidence (categorized as either present or absent) on a 0.5° × 0.5° grid for all regions of Canada and the continental United States. Based on this grid, the data set consisted of 13 719 sites; of these, 567 sites contained at least one incidence of smallmouth bass and were then designated to contain smallmouth bass. The data set was randomly divided into training and validation data sets with the same large-scale geographic coverage in both data sets. The training data set consisted of 9602 sites, with smallmouth bass present at 396 sites. The independent validation data was composed of 4117 sites, with smallmouth bass present at 171 sites (Fig. 1*b*).

### Data analyses

To better understand the relationship between the climate variables and smallmouth bass incidence, we calculated the incidence rate as a function of monthly mean air temperature. This was used to construct frequency plots and to determine the lower and upper temperature ranges of smallmouth bass incidence. As temperature regimes for each month at any given location will be altered because of climate change, one can get an indication of the likelihood of future smallmouth bass incidence based on the frequency plots.

The relationship between smallmouth bass occurrence and climatic variables (monthly mean air temperatures) was evaluated using stepwise multiple logistic regression, classification tree, stepwise linear discriminant analysis, and artificial neural networks. Statistical analyses were conducted in SAS® (SAS Institute Inc., Cary North Carolina), with the exception of artificial neural networks, which were performed in STATISTICA® (StatSoft Inc., Tulsa, Oklahoma). The frequency of smallmouth bass incidence approximates a normal distribution with respect to the predictor variables. Mean annual air temperature was simply the average of the monthly mean air temperatures and therefore not included as a predictor. As expected with climatic variables, monthly mean air temperatures did exhibit high levels of multicollinearity, and this will impact parameter estimates.

Following the comparison of statistical approaches, we used the statistical approach providing the best predictions to estimate future smallmouth bass distribution in Canada and the continental United States using monthly mean air temperatures from 2100 based on the CGCM2–IS92a scenario. The CGCM2–IS92a is the business-as-usual climate change scenario. It uses observed greenhouse gases from 1900 to 1990, which then increase at a rate of 1% per year until the year 2100 up to 1422 parts per million by volume. This scenario includes the direct effect of sulphate aerosols (Canadian Centre for Modelling and Analysis; Canadian Institute for Climate Studies).

#### *Statistical approaches to modeling binary data*

Stepwise multiple logistic regression was used to predict smallmouth bass incidence. Significance values for predictor variables were set at a value of 0.05 to enter and remain in the model. Species presence was designated by the traditional decision threshold of 0.50 (e.g., Olden et al. 2006). In a logistic regression, response variables are subject to a logit transformation, whereas predictor variables are based on a linear combination using maximum likelihood (Olden and Jackson 2002*a*).

Classification trees are generated by dividing the data into two groups, with the division based on the predictor that best divides the group of observations such that they are as mutually exclusive and homogenous as possible (De'ath and Fabricus 2000; Olden and Jackson 2002*a*). Each resulting group is then evaluated as to whether it should be subdivided based on one of the predictor variables, and this process is repeated until some end point is reached, such as minimum group size or depth of tree. The algorithm used in classification trees aims to minimize misclassification rates when dividing the data at each split (Olden and Jackson 2002*a*). The classification tree performed in this study was based on the $\chi^2$ distance. Its significance was set at $p < 0.05$, and the maximum depth of the tree was set by identifying the number of leaves required to minimize the proportion of misclassifications. We used a cross-validation approach to identify the maximum depth of the classification tree by using the cost complexity and reduced error pruning tools available in SAS® with the training and validation data sets. The assessment plot was used to identify the number of leaves required to significantly reduce the proportions of misclassifications.

Linear discriminant analysis finds the linear combination of predictor variables that best discriminates between the two groups and capitalizes on the covariation between the predictors. In this study, groups are defined as the two groups of locations based on their respective presence or absence of smallmouth bass (Legendre and Legendre 1998).

Artificial neural networks were based on a single-layer, feedforward, back-propagation procedure in STATISTICA®. The number of hidden neurons was evaluated ranging from 1 up to matching the number of input neurons (i.e., predictors). The choice of model was based on the best model maximizing the area under the receiver operator characteristic (ROC) plot for the training data. Further details on artificial networks are available in Olden and Jackson (2002*a*, 2002*b*).

We used a large, independent data set to validate the selected models based on the training data set. The use of independent data sets has been rare in ecological studies, but necessary to evaluate the model and determine its generality (Olden and Jackson 2000; Pearce and Ferrier 2000*b*; Ozesmi et al. 2006). Fortunately its acceptance as a general procedure in evaluating models is becoming more common. Generally, predictive power is highest on training data sets, followed by bootstrapping, and then validation on an independent data set (see Olden et al. 2002 for details). Without proper validation, the model overestimates the predictive capability (Olden et al. 2002).

The relative importance of predictor variables was ranked within each of the four statistical approaches. The utility of each statistical approach used in this study was assessed by calculating model sensitivity (correctly predicting presence of smallmouth bass), model specificity (correctly predicting absence of smallmouth bass), and overall classification rate (correctly predicting presence and absence of smallmouth bass) on the independent validation data set. False presence (bass predicted but not observed) and false absence (bass not predicted but observed) were plotted spatially using ArcGIS (ESRI Corporation, Redlands, California) to determine the specific locations of the two types of misclassification and whether particular models failed in specific regions. Since a goal of the model comparison was to determine how well each method performed, rather than simply show which variables were statistically significant and their associated coefficients, colinearity does not present the same concern that it would in determining probabilities. However, as variables are correlated with one another, this may influence their order of importance within any given model.

Similarity analyses using the phi coefficient were conducted on the training and validation data sets to determine how similar the four statistical approaches were at predicting smallmouth bass incidence. The phi coefficient is a measure of association for binary data and is not influenced by variable frequency of occurrence, as are methods such as the Jaccard coefficient (Jackson et al. 1989). This analysis compared each modeling approach's prediction for each location rather than simply summarizing overall success as measured by total correct classification or the confusion matrices (Fielding and Bell 1997).

## Results

Smallmouth bass incidence rates show that smallmouth bass can be present in locations with January air temperatures as cold as –21 °C and in regions with summer average air temperatures as warm as 29 °C (Figs. 2*a*–2*l*); however, the rates clearly identify thermal regions having greater likelihood of bass occurrence. The analysis was used to determine the lower and upper air temperature limits of smallmouth bass incidence for each month of the year.

Each of the four statistical methods revealed differences in the relative importance of climatic variables in predicting smallmouth bass incidence. Generally, summer air temperatures (June or August) were the most important variables in predicting smallmouth bass incidence, followed by winter air temperatures (such as February and November). Artificial neural networks predicted that winter air temperatures were the most important predictors of smallmouth bass incidence (Table 1).

The four statistical approaches used in this study differed in their performance at predicting smallmouth bass presence and absence in locations across Canada and the United States (Table 2). Multiple logistic regression exhibited very low sensitivity, but very high specificity and yielded a 96.1% overall classification rate when tested with the independent validation data set. Multiple logistic regression predicted that most sites would not contain smallmouth bass. Thus, the majority of misclassification tended to be false absences throughout the native and introduced range of the smallmouth bass (Table 2; Fig. 3*a*).

Similar to multiple logistic regression results, the classification tree model exhibited very low sensitivity, but very high specificity and yielded a 95.9% overall classification rate when tested with the independent validation data set. The classification tree predicted the absence of smallmouth bass from most sites. The majority of misclassification of smallmouth bass incidence tended to be false absences throughout the native and introduced range of the species (Table 2; Fig. 3*b*).

Linear discriminant analysis exhibited very high sensitivity, high specificity, and an 81.2% overall classification rate when tested with the independent validation data set. Linear discriminant analysis tended to predict false absences outside the native and introduced ranges (Table 2; Fig. 3*c*). However, linear discriminant analysis yielded the greatest error of the statistical approaches in predicting smallmouth bass absence considered in our study. This approach falsely predicted the occurrence of smallmouth bass at the lower tips of the United States up to the Canadian Arctic. Smallmouth bass were even predicted to be present in the Rocky Mountain range and the island of Newfoundland (Fig. 3*c*).

Artificial neural networks exhibited very high sensitivity and specificity and an overall classification rate of 90.8% when tested with the validation data set (Table 2). The majority of misclassifications were false absences, and a small minority of misclassifications were false presences. The majority of misclassifications occurred just outside of the native and introduced ranges (Fig. 3*d*).

Assessing the overall match of the various results was done using the phi coefficient for the training and validation data sets (Table 3). The two data sets exhibited similar trends across the various predictions, with the training data set showing consistently higher coefficients of association. All values were positive, indicating that there is a greater match than mismatch between the predictions from all methods and the observed data. The observed data was most highly associated with artificial neural networks, further supporting that it was the best statistical approach to use on the data set, followed by linear discriminant analysis. Results from multiple logistic regression and classification tree were not strongly associated with the observed data, suggesting that they were not suitable methods for developing accurate predictions. Predictions from the linear discriminant analysis and artificial neural networks had a high association. Predictions made with multiple logistic regression and classification tree were also associated, but were not strongly associated with the observed data.

Model comparison suggested that the artificial neural network was the most appropriate approach to use for the data set. Therefore, we used the model generated by artificial neural networks in conjunction with the CGCM2–IS92a climate change scenario data to predict smallmouth bass distribution in 2100. We found that the majority of aquatic systems in Canada and the continental United States are predicted to be suitable for smallmouth bass by 2100 based on future air temperatures (Fig. 4).

## Discussion

### Statistical methodology

Comparison of a variety of predictive modeling approaches can provide valuable insight into the determination of which statistical approach is the most appropriate to the data (Guisan and Zimmermann 2000). We compared four statistical approaches to determine which climatic variables were the most important in determining smallmouth bass in-

**Fig. 2.** Smallmouth bass (*Micropterus dolomieu*) incidence rates as a function of monthly mean air temperature. The upper left-hand corner of each plot specifies the lower and upper mean air temperatures required for the occurrence of at least one smallmouth bass.

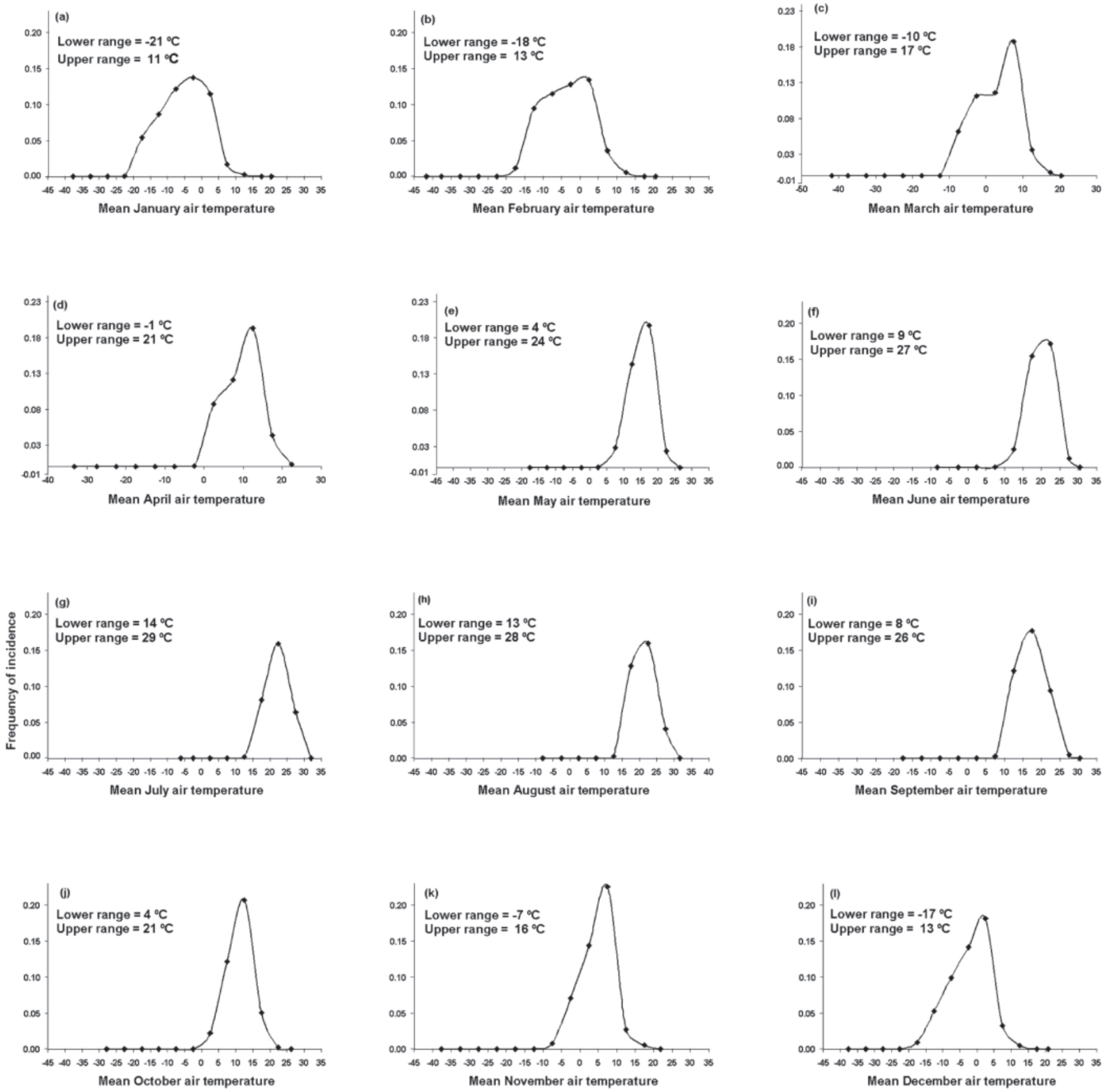



**Table 1.** Relative importance of monthly mean air temperature predictor variables in predicting smallmouth bass (*Micropterus dolomieu*) incidence.

| Method | Ranked importance of predictor variables |
|---|---|
| Multiple logistic regression | June, Feb., Nov., Oct., July, May, Dec., Jan., Aug., Mar. |
| Classification tree | Aug., Feb., Nov., May |
| Linear discriminant analysis | June, Nov., Feb., Oct., Mar., Dec., Jan., July, Aug, May, Apr., Sept. |
| Artificial neural networks | Nov., Feb., Jan., Dec., July, Sept., May, Aug., June, Mar., Apr. |

cidence. The data set was composed of binary data (presence or absence) that contained a high frequency of absences. Logistic regression and linear discriminant analysis are traditionally the most popular statistical approaches to use with binary data. However, if the data fails to meet statistical assumptions such as multicollinearity, as appeared to be the

**Table 2.** Model sensitivity, model specificity, and overall classification performance (all in percent) of the four statistical approaches on the validation smallmouth bass (*Micropterus dolomieu*) incidence data set.

| Method | Model sensitivity | Model specificity | Overall classification |
|---|---|---|---|
| Multiple logistic regression | 9.9 | 99.8 | 96.1 |
| Classification tree | 4.1 | 99.9 | 95.9 |
| Linear discriminant analysis | 92.4 | 80.7 | 81.2 |
| Artificial neural networks | 91.9 | 90.7 | 90.8 |

**Note:** The validation data set was composed of 4117 sites, which contained occurrence of smallmouth bass at 171 sites.

case with our data set, other statistical approaches (e.g., artificial neural networks or classification trees) may be more appropriate (Olden and Jackson 2002*a*).

The occurrence data were obtained through museum sources, online databases, refereed publications, and postgraduate dissertations. Data from natural history museums can be incomplete, sparse, and spatially and temporally biased (Araujo and Guisan 2006; Rondinini et al. 2006), and we supplemented the data with numerous records from the literature, as this is a well-sampled species because of its recreational and economic importance. We attained large numbers of occurrence data for smallmouth bass that reflect the current North American range of the species. Although some of the absences may reflect lack of adequate sampling information, they undoubtedly include areas suitable for smallmouth bass to exist, but where bass have not been able to colonize because of historical and current barriers to dispersal, such as drainage patterns of rivers.

Overall classification rate, model sensitivity, model specificity, and mapping of misclassification rates were used to assess the predictive models generated by the four statistical approaches on an independent data set. Overall classification rate is generally used as either the only method or one of a series of methods to assess predictive models; however, it does not provide detailed information as to the true predictive ability of the model in correctly predicting both the presence and absence of a species (Olden and Jackson 2001). If we simply limited our assessment to overall classification rate, we would conclude that multiple logistic regression and classification trees were the best predictive models for the study. Upon closer inspection of the ability to correctly predict species presence and the ability to correctly predict absence (Pearce and Ferrier 2000*b*), we find that these two models have extremely high predictive power in assessing smallmouth bass absence, but extremely low predictive power in assessing smallmouth bass presence. The artificial neural network model was the best statistical approach to use as demonstrated by the resemblance analyses and exhibited very high model sensitivity and specificity, although not the highest overall classification rate. Calculating model sensitivity and model specificity, in addition to overall classification rate, can provide a more accurate representation of model performance (Olden and Jackson 2001). Generally, as frequency of species occurrence increases, model sensitivity increases and model specificity decreases. Conversely, model sensitivity decreases as the incidence rate decreases (Olden et al. 2002). This could result in increased difficulty at predicting occurrences of rare species where conservation and management are most critical (Olden and Jackson 2001). However, there are situations in which overall classification rate is not a useful measure of accuracy, and sensitivity or specificity must be considered. For example, sensitivity may be considered to be a much more important metric when modeling rare species or in determining which areas are susceptible to an invasive species, and both specificity and total classification rate may be relatively unimportant.

Mapping misclassification rates can provide additional information when evaluating models because it spatially identifies false presences and absences. Mapping of misclassification rates based on artificial neural networks suggested that the majority of misclassifications occurred just outside the native and introduced ranges. This suggests that air temperatures may be currently suitable to permit the establishment of smallmouth bass. Artificial neural networks tended to predict false absences within the native and introduced ranges, indicating that these sites may contain suitable environmental conditions for smallmouth bass to establish, but that the model failed to adequately capture these conditions. However, linear discriminant analysis erroneously classified false presences throughout North America. Mapping the misclassification for both the multiple logistic regression and classification tree showed that the majority of misclassification tended to be false absences throughout the native and introduced ranges of the smallmouth bass. These methods did not accurately predict smallmouth bass presence. However, the false presences that were attained using these approaches tended to occur in regions that may have suitable environmental conditions for smallmouth bass to persist given their proximity to the current distribution.

Artificial neural networks was the best statistical approach in predicting smallmouth bass incidence across Canada and the United States, because it exhibited very high model sensitivity and specificity, high resemblance to the observed data set, and misclassifications just outside the native and introduced ranges of the species. Artificial neural networks present many advantages compared with traditional statistical approaches, such as fewer assumptions regarding data distributions and relationships when modeling with artificial neural networks (Olden et al. 2006; Ozesmi et al. 2006), robustness to nonlinear response relationships, modeling of varying data types such as continuous and discrete data, and can simultaneously model multiple predictor variables and their interactions without a priori knowledge and specification (Brosse et al. 1999; Olden and Jackson 2001; Olden et al. 2006). Artificial neural networks are appropriate when data follow complex or unknown distributions (Pearce and Ferrier 2000*a*). This may explain the superior ability of artificial neural networks with our data set in which the predictor variables exhibited high multicollinearity. However, artificial neural networks can be intensive computationally, may encounter problems of overtraining, has been perceived as a “black box” (but see Olden and Jackson 2002*b*), and is

**Fig. 3.** Misclassification (false absence and false presence) of smallmouth bass (*Micropterus dolomieu*) presence or absence using four statistical approaches: (*a*) multiple logistic regression, (*b*) classification tree, (*c*) linear discriminant analysis, and (*d*) artificial neural networks. False absence is represented by gray circles. False presence is represented by black triangles.

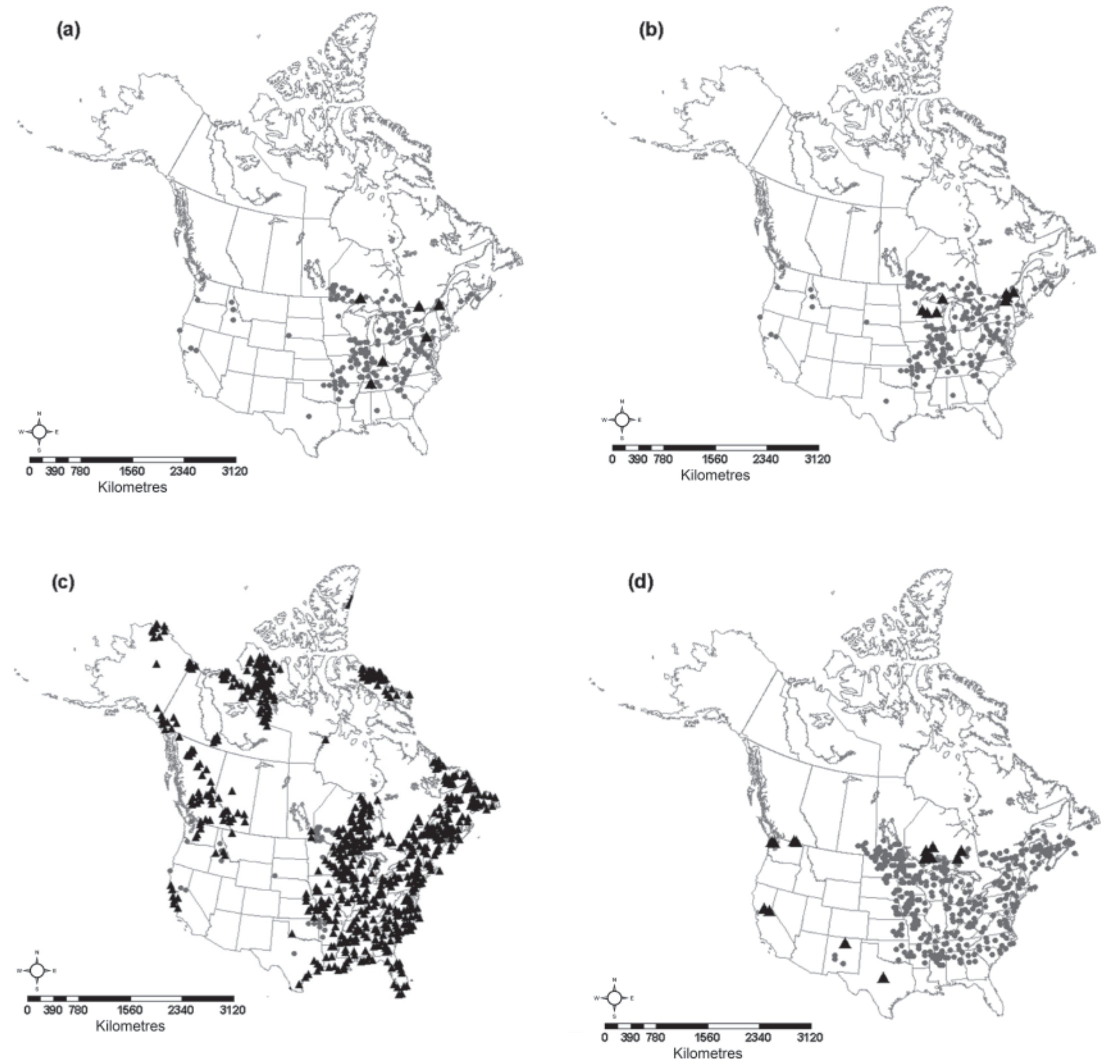


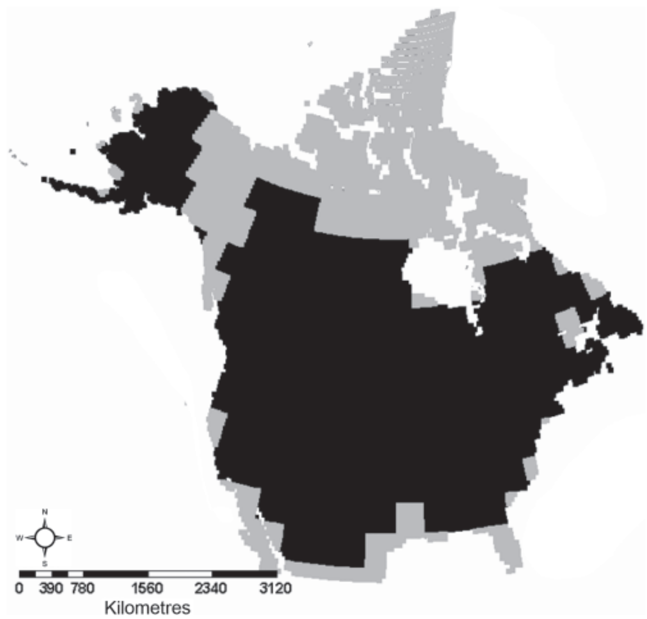

**Table 3.** Similarity index based on the phi coefficient for the training data set and validation data set.

| | Observed | MLR | CT | LDA | ANN |
|---|---|---|---|---|---|
| Observed | 1 | **0.262** | **0.157** | **0.298** | **0.449** |
| MLR | 0.236 | 1 | **0.351** | **0.151** | **0.242** |
| CT | 0.268 | 0.337 | 1 | **0.13** | **0.227** |
| LDA | 0.35 | 0.197 | 0.202 | 1 | **0.5** |
| ANN | 0.504 | 0.245 | 0.288 | 0.564 | 1 |

**Note:** Results for the validation data set are presented in bold. MLR, multiple logistic regression; CT, classification tree; LDA, linear discriminant analysis; ANN, artificial neural networks.

sensitive to structure of the training data set (Ozesmi et al. 2006). Relative to the other statistical approaches, it is more difficult to ascertain the relationships between predictor variables and their relative strength with the response variable when using artificial neural networks (Brosse et al. 1999; Olden and Jackson 2002*b*). However, there are methodologies that can be used to identify the relationships between variables in artificial neural networks such as the neural interpretation diagram (Ozesmi and Ozesmi 1999), Garson's algorithm (Garson 1991; Goh 1995); Lek's algorithm (Lek et al. 1996), a randomization test (Olden and Jackson 2002*b*), and partial derivatives (Gevrey et al. 2006). When data follow normal distributions and variables are uncorrelated, comparable results should be attained by artificial neural networks, multiple logistic regression, or linear discriminant analysis (Olden and Jackson 2001, 2002*a*). In these cases, a more time-efficient and parsimonious model may be more appropriate (Ozesmi et al. 2006), and the simpler approaches may be superior in minimizing overfitting of the data. In our case, the artificial neural networks solution provided a superior predictive model. The artificial neural network model ranked the winter air temperatures as the most important climatic variable for predicting smallmouth bass occurrence, in contrast with the other modeling methods. This discrepancy may be due to differences in the relative emphasis placed in determining whether northern versus

**Fig. 4.** Prediction of potential distribution of suitable smallmouth bass (*Micropterus dolomieu*) habitat in 2100 based on the artificial neural networks model and predicted air temperatures from the Canadian General Circulation Model 2 – Scenario IS92a. Smallmouth bass presence is indicated in black; smallmouth bass absence is shaded in gray.




southern distributions are being captured adequately, but it may also represent that the variable interactions for the artificial neural network model captured the data relationships in a fundamentally different way than the logistic regression, discriminant analysis, and classification trees.

Linear discriminant analysis exhibited the highest model sensitivity, but the lowest model specificity in predicting smallmouth bass incidence across Canada and the United States. Linear discriminant analysis tends to work well with data that have covariation among the predictors and are linearly related to the response variable (Reichard and Hamilton 1997; Olden and Jackson 2002*a*) — this was the case with our data set. Although linear discriminant analysis classified false absences just outside the range of the species, it classified false presences erroneously throughout North America from the high Arctic to the southern tips of Florida.

Multiple logistic regression exhibited the highest model specificity and very low model sensitivity. The approach tended to predict that most values were absences. Although logistic regression is widely used in ecology (Pearce and Ferrier 2000*b*), it works best when the data cleanly follow a logistic curve distribution (Pearce and Ferrier 2000*a*) and may not perform well when predictor variables are correlated. Analyses of our data set indicated that predictor variables exhibited high correlations, thereby violating one of the underlying assumptions for logistic regression. The violation of collinearity in our data set may explain the poor performance of logistic regression. However, a goal of the model comparison was to determine how well each method performed, rather than simply which variables are statistically significant and their associated coefficients; collinearity does not present the same concern that it would in determining probabilities. However as variables may be correlated with one another, this may influence their order of importance within any given model.

Classification trees exhibited the lowest model sensitivity and highest model specificity. The approach tended to predict that almost all values were absences. Generally, classification trees work well with complex ecological data with missing values, nonlinear relationships, and interactions among predictor variables (De'ath and Fabricius 2000; De'ath 2002). However, classification trees can produce discontinuities in its predictions in cases where a continuous function may be more appropriate (Austin 2002).

### Smallmouth bass distribution

Air temperature has been found to be a strong predictor of smallmouth bass distribution (Shuter et al. 1980; Dunlop and Shuter 2006). In addition to temperature, survival and abundance of smallmouth bass are dependent on factors such as habitat availability, abundance of predators and forage food (Olson et al. 2003), water levels, wind, nest desertion, predation, and angling pressure (Scott and Crossman 1973). Of the climatic variables used in the study, summer and winter air temperatures are particularly good predictors of smallmouth bass distribution in Canada and the United States. The incidence rate of smallmouth bass as a function of air temperatures indicated that smallmouth bass could persist in locations with January air temperatures as cold as –21 °C and summer air temperatures as warm as 29 °C. This method summarizes the temperature limit of smallmouth bass and provides an idea of the geographic extent of smallmouth bass with respect to air temperature under climate change. The true incidence rates will be underestimated in most cases because of the inclusion of all North American regions within the calculations (i.e., areas of western North America were also included). Bass may well be able to survive in many areas, but other factors (e.g., historical barriers) have prevented their establishment. Therefore, we provide these relationships as general patterns between incidence and temperatures and not as accurate estimates of incidence rates relative to temperature. We have identified monthly lower and upper temperature ranges of smallmouth bass incidence. The upper lethal temperature for the species is known to be 35 °C (Scott and Crossman 1973), and feeding has been observed to stop below 5 °C (Shuter et al. 1980). Laboratory and field studies showed that low winter water temperatures (i.e., <10 °C) could cause decreased activity levels in young-of-year smallmouth bass at the northern range. At temperatures below 7 °C, smallmouth bass begin to search for shelter, arrest feeding behaviour, and become inactive. As the inactivity time increases, young-of-the-year mortality increases and is size-related, beginning with the smallest fish (Shuter et al. 1980).

We found that summer and winter air temperatures were the most important predictors of smallmouth bass distribution. A positive association has been found between smallmouth bass growth and mean summer air temperature (Shuter et al. 1980; Casselman et al. 2002; Shuter and Ridgway 2002). Shuter et al. (1980) found that average July air temperatures influenced the growth rate and therefore the size of young of the year as they entered their first winter and the possibility of winter starvation. Size of young fish is extremely important, as larger young are able to survive starvation and winter better than smaller fish (Shuter et al. 1980; Wismer et al. 1985). Casselman et al. (2002) found that July–August water temperatures were positively correlated to year-class strength of smallmouth bass. The largest year class was produced after a very warm year, and the smallest year class was produced after the coolest year (Casselman et al. 2002). Temperatures in March have been correlated directly with increased probability of success and survival as warm, early springs may result in faster growth and more males maturing and spawning (Casselman et al. 2002).

As demonstrated in our study, smallmouth bass are viable in regions with a wide range of air temperatures. Business-as-usual climate change scenarios suggest that smallmouth bass are expected to have suitable thermal habitats through most of North America. Smallmouth bass are not expected to be found in alpine regions and the high Arctic. We have shown that as temperatures increase with climate change, the range of suitable habitat will extend northwards. It is expected that the current distribution of smallmouth bass will shift to the north because of the considerable extensions to their native range that have occurred during the past century. In accordance with such a climatic shift, temperatures in the far southern United States will be too warm to support viable smallmouth bass populations. By 2100, temperatures in many additional regions of Canada will be warm enough for viable smallmouth bass populations to exist (Chu et al. 2005; Sharma et al. 2007). The accelerated expansion northwards of smallmouth bass will be possible because of the northward direction of flow for aquatic systems north of the continental divide. At the southern extent of the range, winter air temperatures may be too high and inhibit gametogenesis (sensu Lukšienė et al. 2000). Whitledge et al. (2006) found that summer stream temperatures exceeding 27 °C impeded the growth of smallmouth bass in the Ozark streams in Missouri. Temperatures greater than 22 °C resulted in the displacement of smallmouth bass by largemouth bass in Ozark streams (Zweifel et al. 1999). Furthermore, at the southern extent of the range, summer water temperatures may exceed 35 °C by 2100, the upper lethal temperature for the species, thereby reducing the number of viable populations in the southern United States.

The implications of climate warming are not limited to distributional changes. Smallmouth bass are expected to experience greater growth rates that in turn will increase the likelihood of survival in concert with epilimnetic warming at the northern extent of its range (King et al. 1999; Jackson and Mandrak 2002). Year-class strength of smallmouth bass can be increased by two–five times with a 1 °C increase in temperature and six times with a 2 °C increase in temperature at the current northern extent of smallmouth bass distribution (Casselman et al. 2002). Additionally, as the duration of the ice-free period is reduced, the probability of bass winterkill attributed to low dissolved oxygen levels will also be reduced (Jackson and Mandrak 2002).

The northerly shift in the distribution of the non-native smallmouth bass will have substantial implications on native aquatic communities. Many factors will inevitably contribute to regional losses of biota, including the homogenization of the fish fauna as native cyprinids are lost (Jackson and Mandrak 2002), greater abundance of filamentous algae because of reduction in small fishes and benthic organisms that graze on algae (Power et al. 1985), and decreased growth and reproduction of native lake trout populations (Vander Zanden et al. 1999).

In conclusion, the development of predictive models is important to understand the factors that may be contributing to the current and potential future distribution of the species using the best available statistical approach. Artificial neural networks provided the most appropriate statistical approach to use for this data set and identified winter and summer air temperatures as the most important climatic variables for predicting smallmouth bass occurrence across North America. Mapping the false absences predicted by artificial neural networks identified areas where smallmouth bass may be found in the near future (Vander Zanden et al. 2004*b*). Currently, the range expansion of smallmouth bass has been facilitated by stocking by governmental agencies, unauthorized and accidental introduction by anglers, and dispersal through drainage networks (Jackson 2002; Vander Zanden et al. 2004*a*). Business-as-usual climate change data used in the model generated by artificial neural networks predicts that the majority of Canadian and American aquatic systems will contain suitable thermal conditions for smallmouth bass by 2100. Increases in air temperature will have indirect but detrimental effects on a large number of aquatic systems by the northward range expansion of smallmouth bass. The spread of smallmouth bass can be reduced by intensifying public education and regulations, which will help to limit the potential consequences of smallmouth bass on native aquatic communities in the future.

## Acknowledgements

We thank the numerous scientists who published their data in scientific journals, theses, or in online databases. We are grateful to Brian Shuter, Ann Zimmerman, Mark Poos, and Andrew Drake for comments on earlier versions of the manuscript. We thank two anonymous reviewers for helpful comments and suggestions on the manuscript. Funding for this research was provided by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to DAJ and fellowships by the University of Toronto to SS.

## References

Araujo, M.B., and Guisan, A. 2006. Five (or so) challenges for species distribution modeling. J. Biogeogr. **33**: 1677–1688.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecol. Model. **157**: 101–118.

Brosse, S., Guegan, J.F., Tourenq, J.N., and Lek, S. 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Model. **120**: 299–311.

Casselman, J.M. 2002. Effects of temperature, global extremes, and climate change on year-class production of warmwater, coolwater, and coldwater fishes in the Great Lakes basin. *In* Fisheries in a Changing Climate. American Fisheries Society Symposium 32. *Edited by* N.A. McGinn. American Fisheries Society, Bethesda, Md. pp. 39–60.

Casselman, J.M., Brown, D.M., Hoyle, J.A., and Eckert, T.H. 2002. Effects of climate and global warming on year-class strength and relative abundance of smallmouth bass in eastern Lake Ontario. *In*

Black Bass: Ecology, Conservation and Management. American Fisheries Society Symposium 31. *Edited by* D.P. Philipp and M.S. Ridgway. American Fisheries Society, Bethesda, Md. pp. 73–90.

Chu, C., Mandrak, N.E., and Minns, C.K. 2005. Potential impacts of climate change on the distributions of several common and rare freshwater fishes in Canada. Divers. Distrib. **11**: 299–310.

De'ath, G. 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. Ecology, **83**: 1105–1117.

De'ath, G., and Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, **81**: 3178–3192.

Dunlop, E.S., and Shuter, B.J. 2006. Native and introduced populations of smallmouth bass differ in concordance between climate and somatic growth. Trans. Am. Fish. Soc. **135**: 1175–1190.

Fielding, A.H., and Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. **24**: 38–49.

Garson, G.D. 1991. Interpreting neural-network connection weights. Artif. Intell. Expert, **6**: 46–51.

Gevrey, M., Dimopoulos, I., and Lek, S. 2006. Two-way interaction of input variables in the sensitivity analysis of neural network models. Ecol. Model. **195**: 43–50.

Goh, A.T.C. 1995. Back-propagation neural networks for modeling complex systems. Artif. Intell. Eng. **9**: 143–151.

Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135**: 147–186.

Jackson, D.A. 2002. Ecological effects of *Micropterus* introductions: the dark side of black bass. *In* Black Bass: Ecology, Conservation and Management. American Fisheries Society Symposium 31. *Edited by* D.P. Philipp and M.S. Ridgway. American Fisheries Society, Bethesda, Md. pp. 221–232.

Jackson, D.A., and Mandrak, N.E. 2002. Changing fish biodiversity: predicting the loss of cyprinid biodiversity due to global climate change. *In* Fisheries in a Changing Climate. American Fisheries Society Symposium 32. *Edited by* N.A. McGinn. American Fisheries Society, Bethesda, Md. pp. 89–98.

Jackson, D.A., Somers, K.M., and Harvey, H.H. 1989. Similarity coefficients: measures of co-occurrence and association or simply measures of occurrence? Am. Nat. **133**: 436–453.

King, J.R., Shuter, B.J., and Zimmerman, A.P. 1999. Empirical links between thermal habitat, fish growth and climate change. Trans. Am. Fish. Soc. **128**: 656–665.

Legendre, P., and Legendre, L. 1998. Numerical ecology. Elsevier Science B.V., Amsterdam, The Netherlands.

Lek, S., Delacoste, M., Baran, P., Dimopolous, I., Lauga, J., and Aulagnier, S. 1996. Application of neural networks to modeling nonlinear relationships in ecology. Ecol. Model. **90**: 39–52.

Lukšienė, D., Sandström, O., Lounasheimo, L., and Andersson, J. 2000. The effects of thermal effluent exposure on the gametogenesis of female fish. J. Fish. Biol. **56**: 37–50.

MacRae, P.S.D., and Jackson, D.A. 2001. The influence of smallmouth bass (*Micropterus dolomieu*) predation and habitat complexity on the structure of littoral zone fish assemblages. Can. J. Fish. Aquat. Sci. **58**: 342–351.

Magnuson, J.J. 2002. A future of adapting to climate change and variability. *In* Fisheries in a Changing Climate. American Fisheries Society Symposium 32. *Edited by* N.A. McGinn. American Fisheries Society, Bethesda, Md. pp. 273–282.

Magnuson, J.J., Meisner, J.D., and Hill, D.K. 1990. Potential changes in the thermal habitat of Great Lakes fish after global climate warming. Trans. Am. Fish. Soc. **119**: 254–264.

Olden, J.D., and Jackson, D.A. 2000. Torturing the data for the sake of generality: how valid are our regression models? Ecoscience, **7**: 501–510.

Olden, J.D., and Jackson, D.A. 2001. Fish–habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Trans. Am. Fish. Soc. **130**: 878–897.

Olden, J.D., and Jackson, D.A. 2002*a*. A comparison of statistical approaches for modelling fish species distributions. Freshw. Biol. **10**: 1976–1995.

Olden, J.D., and Jackson, D.A. 2002*b*. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. **154**: 135–150.

Olden, J.D., Jackson, D.A., and Peres-Neto, P.R. 2002. Predictive models of fish species distributions: a note on proper validation and chance predictions. Trans. Am. Fish. Soc. **131**: 329–336.

Olden, J.D., Joy, M.K., and De'ath, R.G. 2006. Rediscovering the species in community-wide predictive modeling. Ecol. Appl. **16**: 1449–1460.

Olson, M.H., Young, B.P., and Blinkoff, K.D. 2003. Mechanisms underlying habitat use of juvenile largemouth and smallmouth bass. Trans. Am. Fish. Soc. **132**: 398–405.

Ozesmi, S.L., and Ozesmi, U. 1999. An artificial neural network approach to spatial habitat modeling with interspecific interaction. Ecol. Model. **116**: 15 – 31.

Ozesmi, S.L., Tan, C.O., and Ozesmi, U. 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Model. **195**: 83–93.

Pearce, J., and Ferrier, S. 2000*a*. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecol. Model. **128**: 127–147.

Pearce, J., and Ferrier, S. 2000*b*. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. **133**: 225–245.

Power, M.E., Matthews, W.J., and Stewart, A.J. 1985. Grazing minnows, piscivorous bass, and stream algae — dynamics of a strong interaction. Ecology, **66**: 1448–1456.

Reichard, S.H., and Hamilton, C.W. 1997. Predicting invasions of woody plants introduced into North America. Conserv. Biol. **11**: 193–203.

Rejwan, C., Shuter, B.J., Ridgway, M.S., and Collins, N.C. 1997. Spatial and temporal distributions of smallmouth bass (*Micropterus dolomieu*) nests in Lake Opeongo, Ontario. Can. J. Fish. Aquat. Sci. **54**: 2007–2013.

Rejwan, C., Collins, N.C., Brunner, L.J., Shuter, B.J., and Ridgway, M.S. 1999. Tree regression analysis on the nesting habitat of smallmouth bass. Ecology, **80**: 341–348.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., and Possingham, H.P. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol. Lett. **9**: 1136–1145.

Scott, W.B., and Crossman, E.J. 1973. Freshwater fishes of Canada. Fish. Res. Board Can. Bull. 184.

Sharma, S., Jackson, D.A., Minns, C.K., and Shuter, B.J. 2007. Will northern fish populations be in hot water because of climate change? Global Change Biol. **13**: 2052–2064.

Shuter, B.J., and Ridgway, M.S. 2002. Bass in time and space: operational definitions of risk. *In* Black Bass: Ecology, Conservation and Management. American Fisheries Society Symposium 31. *Edited by* D.P. Philipp and M.S. Ridgway. American Fisheries Society, Bethesda, Md. pp. 235–249.

Shuter, B.J., MacLean, J.A., Fry, F.E.J., and Regier, H.A. 1980. Stochastic simulation of temperature effects on first-year survival of smallmouth bass. Trans. Am. Fish. Soc. **109**: 1–34.

Shuter, B.J., Minns, C.K., and Lester, N.P. 2002. Climate change, freshwater fish, and fisheries: case studies from Ontario and their use in assessing potential impacts. *In* Fisheries in a Changing Climate. American Fisheries Society Symposium 32. *Edited by* N.A. McGinn. American Fisheries Society, Bethesda, Md. pp. 77–88.

Vander Zanden, M.J., Casselman, J.M., and Rasmussen, J.B. 1999. Stable isotope evidence for the food web consequences of species invasions in lakes. Nature (London), **401**: 464–467.

Vander Zanden, M.J., Wilson, K.A., Casselman, J.M., and Yan, N.D. 2004*a*. Species introductions and their impacts in North American Shield lakes. *In* Boreal shield watersheds: lake trout ecosystems in a changing environment. *Edited by* J.M. Gunn, R.J. Steedman, and R.A. Ryder. Lewis Publishers, Boca Raton, Fla. pp. 239–263.

Vander Zanden, M.J., Olden, J.D., Thorne, J.H., and Mandrak, N.E. 2004*b*. Predicting occurrences and impacts of smallmouth bass introduction in north temperate lakes. Ecol. Appl. **14**: 132–148.

Whitledge, G.W., Rabeni, C.F., Annis, G., and Sowa, S.P. 2006. Riparian shading and groundwater enhance growth potential for smallmouth bass in Ozark streams. Ecol. Appl. **16**: 1461–1473.

Wismer, D.A., DeAngelis, D.L., and Shuter, B.J. 1985. An empirical model of size distributions of smallmouth bass. Trans. Am. Fish. Soc. **114**: 737–742.

Zweifel, R.D., Hayward, R.S., and Rabeni, C.F. 1999. Bioenergetics insight into black bass distribution shifts in Ozark border region streams. N. Am. J. Fish. Manag. **19**: 192–197.