



Influences of gillnet fishing on lake sturgeon bycatch in Lake Erie and implications for conservation

Yan Li*, Yan Jiao

Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0321, USA

ABSTRACT: Three candidate classification tree models were constructed to estimate the probability of obtaining lake sturgeon *Acipenser fulvescens* bycatch under specific environmental and gillnet fishing conditions in Lake Erie. This analysis was based on a fishery-independent survey, the Lake Erie Partnership Index Fishing Survey (PIS), from 1989 to 2008. The 3 classification tree models included 1 conditional-inference classification tree generated by the R-package 'party' and 2 exhaustive-search-based classification trees generated by the R-package 'tree' and 'rpart,' respectively. The discriminative performance of each tree was evaluated by the receiver operating characteristic (ROC) curve and the area under the curve (AUC) using a jackknife approach. Most of the lake sturgeon captured in the PIS were juveniles. The 3 tree models identified fishing basin and gear type as factors related to gillnet fishing that had important influences on lake sturgeon bycatch and implications for lake sturgeon conservation and management. Results indicated that the west basin of Lake Erie could be a hotspot for lake sturgeon bycatch in the commercial gillnet fisheries, and the use of bottom gillnets might increase the probability of catching lake sturgeon. A model comparison indicated that the conditional-inference tree model could provide unbiased predictor selection and better discriminative performance in predicting the probability of taking lake sturgeon as bycatch.

KEY WORDS: Gillnet fishing · Lake sturgeon · Bycatch · Lake Erie · Conservation · Classification tree

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Lake sturgeon *Acipenser fulvescens* were abundant in the Great Lakes during the late 1800s, contributing to both recreational and commercial fisheries (Auer 1999). The biological characteristics of lake sturgeon, such as long lifespan, late maturity, and intermittent spawning, make this species vulnerable to heavy exploitation and dramatic environmental changes (Noakes et al. 1999, Welsh 2004). Within the last century, lake sturgeon have been dramatically reduced or extirpated due to overfishing, habitat degradation, anthropogenic impacts on nursery and spawning areas, damming that impedes migration, and water pollution (Scott & Crossman 1973, Birstein et al. 1997, Auer 1999, Bogue 2000). Although lake sturgeon are not listed under the US Endangered Species Act of

1973, they have been listed as endangered or threatened under local legislation in 19 of the 20 states within their original range in the USA (GLLSCM 2002, 2010, Welsh 2004). In the Canadian province of Ontario, although the recreational or commercial harvest of lake sturgeon is permitted, it is considered to be a sensitive species and is regulated by a quota system with increasing concerns of fishery management and fish conservation (CESCC 2001, 2005, Welsh 2004).

Lake sturgeon migrate to rivers throughout the lake basins to spawn and spend the remaining time in the open waters of the Great Lakes. Lake Erie supports several non-spawning and few spawning populations of lake sturgeon (Welsh 2004). Some rivers off Lake Erie have been identified to be major spawning areas of lake sturgeon, including the Detroit River connecting Lake Erie and the St. Clair system (Lake St. Clair

*Email: yanli08@vt.edu

and the St. Clair River) and the Niagara River (Thomas & Haas 1999, 2002, GLLSCM 2002, Boase 2003, Caswell 2003, Welsh 2004). Lake Erie also supports large commercial gillnet fisheries of some key species, such as walleye *Sander vitreus* and yellow perch *Perca flavescens* (Kinnunen 2003, Thomas & Haas 2005). Because of the overlap in life history traits between juvenile lake sturgeon and the major target species in the gillnet fisheries, the high size selectivity and mortality of gillnets, and the management status of lake sturgeon, gillnet bycatch might cause additional threats to lake sturgeon populations in Lake Erie (Scott & Crossman 1973, Hamley 1975, GLLSCM 2002, Kinnunen 2003, Johnson et al. 2004).

Previous studies centered on the demographic characteristics and stock structure of lake sturgeon populations residing in the river systems using genetic tools (Ferguson & Duckworth 1997) and experimental surveys (Threader & Broussau 1986, Chiasson et al. 1997, McKinley et al. 1998, Noakes et al. 1999, Thomas & Haas 1999, 2002). However, few studies have addressed the impacts of gillnet fishing on the bycatch of lake sturgeon in Lake Erie. Failure to take the influences of gillnet fishing into account in lake sturgeon conservation and fishery management may ignore the potential threats from gillnet fishing, and may reduce the efficiency of recovering lake sturgeon populations.

The Ontario Ministry of Natural Resources (OMNR) and the Ontario Commercial Fisheries' Association (OCFA) have been conducting a fishery-independent survey program—the Lake Erie Partnership Index Fishing Survey (PIS)—since 1989. In the PIS, experimental gillnets are deployed at sites across the entire Ontario waters during August to November annually using commercial fishing crews and vessels (OCFA unpubl.). The PIS data contain a high percentage of 0 captures of lake sturgeon, which is typically encountered in bycatch analyses and catch analyses of rare species (Maunder & Punt 2004). The presence of zeros may invalidate the assumption of normality commonly used in fishery analysis, and may cause computational difficulties (Maunder & Punt 2004). Elimination of a considerable proportion of zeros may result in a loss of information on spatial or temporal distribution characteristics of the fish stock (Maunder & Punt 2004).

The methods employed in previous studies to deal with the fishery data containing zeros can be categorized into 2 types. One approach is to add a small constant to each observation in the generalized linear/additive model analysis (Ortiz et al. 2000, Maunder & Punt 2004, Shono 2008). The other approach is to use the delta model or Tweedie distribution model. The delta model deals with positive values and zeros using 2 sub-models separately (Lo et al. 1992, Stefansson

1996, Ye et al. 2001, Maunder & Langley 2004), whereas the Tweedie distribution model handles zeros uniformly along with the positive values (Tweedie 1984, Shono 2008). However, based on a preliminary analysis, it was difficult to quantitatively capture the complexity of the data and the underlying relationship between the lake sturgeon bycatch and the predictor variables by applying those commonly used methods because of the extremely high percentage of 0 observations (>99%) in the PIS data.

The tree-based approach was introduced by Breiman et al. (1984) from a statistical perspective and has been widely used for classification problems in clinical, genetic and ecological studies because it allows for the easy interpretation of the model, the graphical visualization of model structure, and the identification of complex interactions in the data (Ribic & Ainley 1997, Fielding 1999, Harrell 2001, Gansky 2003, Austin 2008, Cutler et al. 2009). Two types of classification tree method are commonly used: the exhaustive-search-based tree and the conditional-inference tree. The exhaustive-search-based tree selects the predictor at a node and splits this node simultaneously through an exhaustive search over all possible splits on all predictors that satisfy the splitting criterion (Hothorn et al. 2006, Cutler et al. 2009), whereas the conditional-inference tree selects the predictor and splits the node separately (Hothorn et al. 2006). The discriminative performance of these 2 types of tree method has been proven equivalent (Hothorn et al. 2006).

In this study, we aimed to (1) construct 3 candidate trees, i.e. a conditional-inference classification tree and 2 exhaustive-search-based classification trees, to estimate the probability of obtaining lake sturgeon bycatch under specific environmental and gillnet fishing conditions in Lake Erie based on the PIS data; (2) investigate factors related to gillnet fishing that have potential influences on lake sturgeon bycatch and highlight their implications for lake sturgeon conservation and management; and (3) compare the discriminative performances of 3 candidate trees and recommend the application of classification tree models in data analyses with a high percentage of 0 observations.

MATERIALS AND METHODS

Data. In the PIS data, 53 562 nets were available for analysis, with a total of 31 lake sturgeon captured. The predictors included 6 continuous variables (site depth, gear depth, Secchi depth, gear temperature, dissolved oxygen, and soak time) and 5 categorical variables (basin, year, month, gear type, mesh size). The information on predictors is summarized in Table 1.

Table 1. Predictor variables included in the classification tree models for lake sturgeon in Lake Erie. Penn. Ridge: Pennsylvania Ridge. For continuous variables, mean and range values are shown

Predictor variable	Type	Mean	Range	Category
Site depth (m)	Continuous	21.0	3.3–65.5	
Gear depth (m) ^a	Continuous	14.4	0.9–65.2	
Secchi depth (m)	Continuous	3.3	0.2–11.0	
Gear temperature (°C) ^b	Continuous	16.8	2.7–26.2	
Dissolved oxygen (mg l ⁻¹)	Continuous	8.8	0.2–21.4	
Soak time (h)	Continuous	22.2	9.5–35.9	
Basin	Categorical			West, west central, east central, east, Penn. Ridge
Year	Categorical			1989–2008
Month	Categorical			August–November
Gear type	Categorical			Canned or bottomed
Mesh size	Categorical			32–152 mm

^aWater depth at which the gillnet was set
^bWater temperature at the depth where the gillnet was set

Trees. The conditional-inference classification tree is a unified framework where the recursive partitioning of a tree is conducted on the basis of well-defined statistical tests (Hothorn et al. 2006). We denoted the response variable as $Y \in \{0,1\}$, where the value 0 represents the event of capturing no lake sturgeon and the value 1 represents the event of capturing at least 1 lake sturgeon. The m predictor variables were denoted as $X = (X_1, \dots, X_m)$. The problem was then converted to a 2-group classification problem. Our goal was to classify each observation with the label 0 or 1 given a complex of predictor variables. The procedure for growing a conditional-inference classification tree is described as follows (Breiman et al. 1984, Fielding 1999, Hothorn et al. 2006, Cutler et al. 2009):

(1) Test the global null hypothesis of independence between the response variable and any of the m predictor variables. Stop if this global null hypothesis cannot be rejected at a pre-determined significance level α based on the test statistics. Otherwise (i.e. at least 1 of the m predictor variables is significantly associated with the response variable), the association between Y and each of the m predictor variables is measured by the test statistics, and the variable X_j with the strongest association to Y , say X_j , is selected to split a node, where $j \in \{1, \dots, m\}$.

(2) Determine a set $A \subset X_j$ to split X_j into 2 disjoint sets A and \bar{A} , where $A \cap \bar{A} = \emptyset$ and $A \cup \bar{A} = X_j$. The set A was determined by measuring the discrepancy between all possible A and the corresponding \bar{A} based on the test statistic. The set A that maximizes the discrepancy is selected to determine the split value at this node. The observations at this node are assigned to the left if $X_{j,i} \in A$ or to the right if $X_{j,i} \in \bar{A}$ in order to form 2 descendant nodes, where $i = 1, \dots, n$, and n is the num-

ber of observations at this node, and for the root node n is the total number of observations.

(3) Recursively repeat steps (1) and (2) to select a predictor variable at a node and to determine the split value for splitting the node and forming the descendant nodes.

(4) When the tree stops growing, the node that cannot be split any further is called a terminal node. The observations at a terminal node are used to compute the proportion of each class that gives the probability of obtaining each class, or compute the class label based on the most frequent class label the observations have at this terminal node.

In this study, the R-package 'party' (R Version 2.9.2) was utilized to construct the conditional-inference classification tree, and this tree is referred to as the 'party' tree in the following analysis.

The 2 exhaustive-search-based trees were grown in a similar way to that described above except that steps (1) and (2) were combined (Hothorn et al. 2006). The particular split to partition a node was selected by exhaustively searching for every possible split on every predictor variable (Breiman et al. 1984, Harrell 2001, Hothorn et al. 2006). The predictor and split combination that satisfied the splitting criterion best was selected to partition the node (Breiman et al. 1984, Cutler et al. 2009). The 2 exhaustive-search-based classification trees were programmed using the R-package 'tree' and 'rpart,' and are referred to as the 'tree' tree and the 'rpart' tree in the following analysis.

Model comparison. The discriminative performance of each candidate tree was evaluated by plotting the receiver operating characteristic (ROC) curve and calculating the area under the ROC curve (AUC). The ROC curve has long been used to visualize the performance of a classification algorithm, which plots the true positive rate $\Pr(\hat{Y} = \oplus | Y = \oplus)$ against the false positive rate $\Pr(\hat{Y} = \oplus | Y = \ominus)$ (Bradley 1997, Austin 2008). In our case, the true positive rate refers to the probability that the predicted label is 1 given the observed label 1, and the false positive rate refers to the probability that the predicted label is 1 given the observed label 0. The classification algorithm is considered to have a better discriminative performance if it yields a ROC curve shooting to the top left corner of the plot and extending towards the top right corner which gives a larger AUC value (Bradley 1997, Austin 2008).

The jackknife approach was employed to conduct the uncertainty analysis (Miller 1974, Efron 1979, Meyer et al. 1986). In the jackknife approach, we left out 1 observation each time. The 1 observation left out represented the test data and the $N-1$ observations were the training data, where N is the total number of observations. At each run of jackknifing, the classification tree was developed using the training data, and the corresponding AUC value was calculated. The classification tree built based on the training data was applied to the test data for prediction. A ROC curve was generated by combining the predicted value of each test data from each run of jackknifing across the whole dataset.

RESULTS

In total, 31 lake sturgeon were observed to have taken as bycatch in the Lake Erie PIS conducted in the fall from 1989 to 2008 (Table 2). These are tabulated according to the predictor variables selected by the conditional-inference classification tree (see Fig. 2). Of the 31 lake sturgeon, 26 (83.9%) were captured in the west basin, 25 (80.6%) were captured in waters where the temperature at the gear set depth was less than 22.6°C, 30 (96.8%) were captured in waters where the dissolved oxygen level was greater than 6.9 mg l⁻¹, and 16 (51.6%) were captured in waters with a depth less than 8.5 m. The average total length of the lake sturgeon bycaught in the PIS was 570 mm (95% confidence interval: 229 to 857 mm) (Fig. 1).

The conditional-inference classification tree (Fig. 2) was grown starting with basin at the root node, and selected year, site depth, gear temperature, and dissolved oxygen at the descendant internal nodes. This indicated that basin ($p < 0.001$) was the factor most strongly associated to the presence/absence of lake sturgeon bycatch in Lake Erie, and year ($p < 0.001$), site depth ($p < 0.001$), gear temperature ($p < 0.001$), and dissolved oxygen ($p = 0.002$) also had important impacts on the lake sturgeon bycatch. Seven termi-

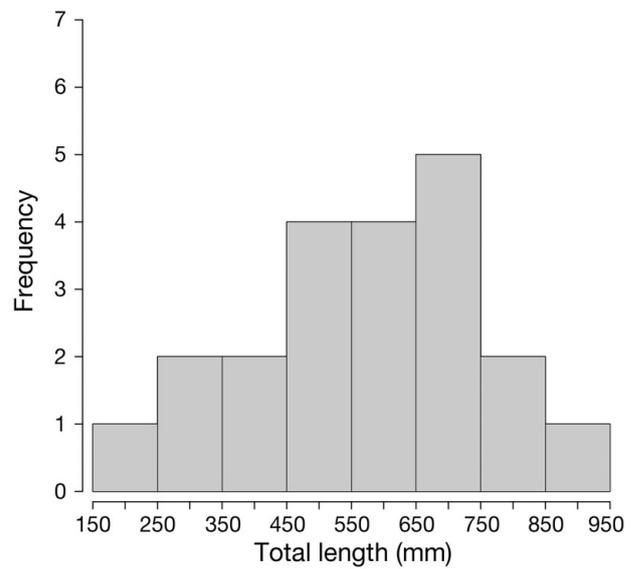


Fig. 1. *Acipenser fulvescens*. Total length (mm) of lake sturgeon taken as bycatch in the Lake Erie Partnership Index Fishing Survey (PIS), 1989–2008

nal nodes were identified in the well-constructed tree. The lake sturgeon bycatch was most likely (35.3% chance) to be observed in waters of the west basin during the years 1994, 1998, and 2007, where the water temperature at gear set depth was less than 22.6°C and the dissolved oxygen level was greater than 6.9 mg l⁻¹.

The 'tree' tree (Fig. 3) was grown with basin at the root node, and the 'rpart' tree (Fig. 4) was rooted at site depth. Both trees contained 10 terminal nodes, and selected year, mesh size, soak time, and site depth at the internal nodes. Gear type was also included in the 'tree' tree at an internal node, and basin was also included in the 'rpart' tree.

The 'tree' tree (Fig. 3) indicated a 100% chance of observing lake sturgeon bycatch when fishing using a bottomed gillnet with soak time >22 h in waters deeper than 9.1 m in the west basin during the years 1994,

Table 2. *Acipenser fulvescens*. Observed lake sturgeon bycatch (in numbers) by basin, gear temperature, dissolved oxygen, and site depth (\leq or >8.5 m). Data from the Lake Erie Partnership Index Fishing Survey (PIS), 1989–2008

Gear temperature (°C)	Dissolved oxygen (mg l ⁻¹)	West		West central		East central		East		Pennsylvania Ridge		Total
		≤ 8.5	> 8.5	≤ 8.5	> 8.5	≤ 8.5	> 8.5	≤ 8.5	> 8.5	≤ 8.5	> 8.5	
≤ 22.6	≤ 6.9	0	1	0	0	0	0	0	0	0	0	1
	> 6.9	14	5	0	3	0	2	0	0	0	0	24
> 22.6	≤ 6.9	0	0	0	0	0	0	0	0	0	0	0
	> 6.9	2	4	0	0	0	0	0	0	0	0	6
Total		16	10	0	3	0	2	0	0	0	0	31

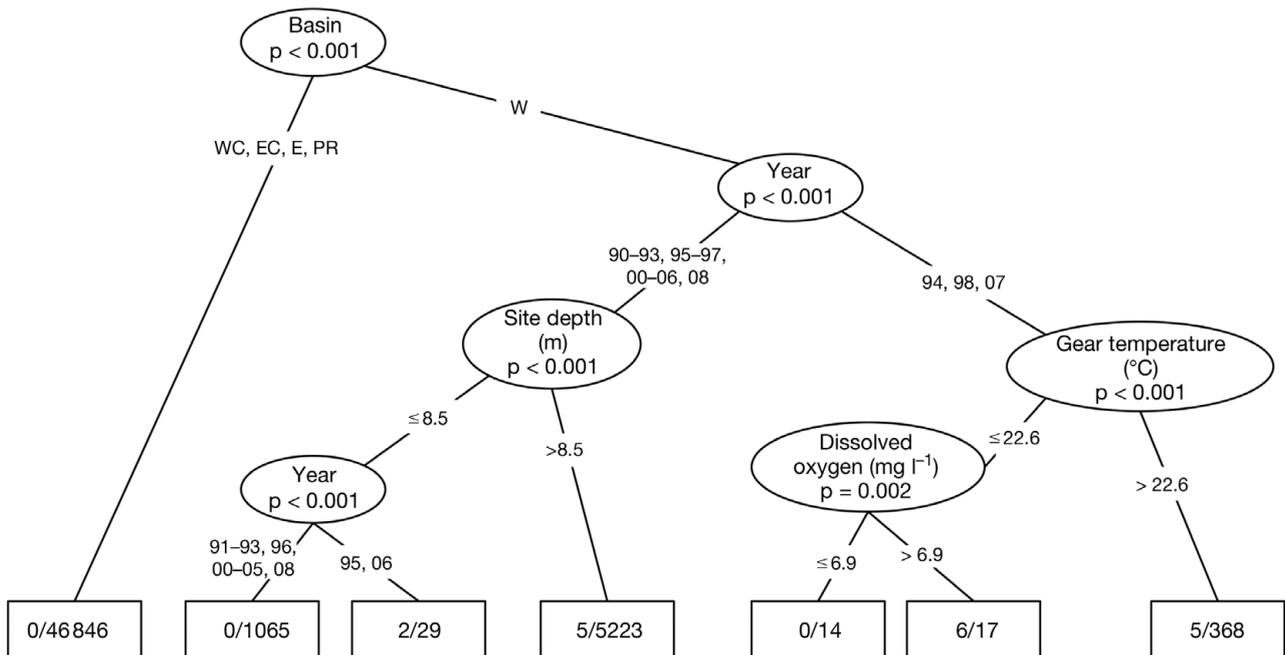


Fig. 2. Conditional-inference classification tree for lake sturgeon bycatch generated by the R-package 'party' ('party' tree in text). The internal nodes are denoted by ovals, and the terminal nodes by rectangles. At each terminal node, the number of lake sturgeon observations at this node/the number of all observations (n) falling into this node is indicated in the rectangle. 'Gear temperature' indicates the water temperature at the depth where the gillnet was set. W: west basin; WC: west central basin, EC: east central basin; E: east basin; PR: Pennsylvania Ridge

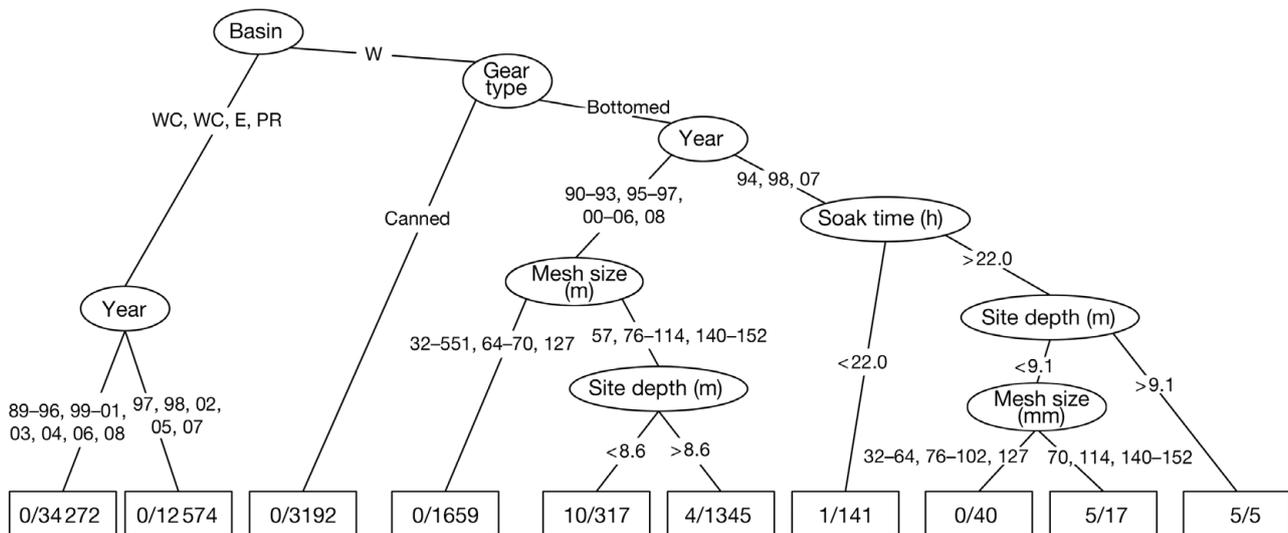


Fig. 3. Exhaustive-search-based classification tree for lake sturgeon bycatch generated by the R-package 'tree' ('tree' tree in text). See Fig. 2 for the explanations for internal and terminal nodes and basin abbreviations

1998, and 2007. If gillnet fishing took place in waters shallower than 9.1 m in the west basin during the same years, and gillnets with a mesh size of 70, 114, 140, and 152 mm were set at bottom for >22 h, there was a 29% chance of lake sturgeon bycatch.

According to the 'rpart' tree (Fig. 4), there was 100% chance of taking lake sturgeon as bycatch in waters of the west basin where the water depth was >9.1 m and when gillnets were soaked for >22.5 h during the years 1994 and 2007. When fishing using gillnets with a mesh

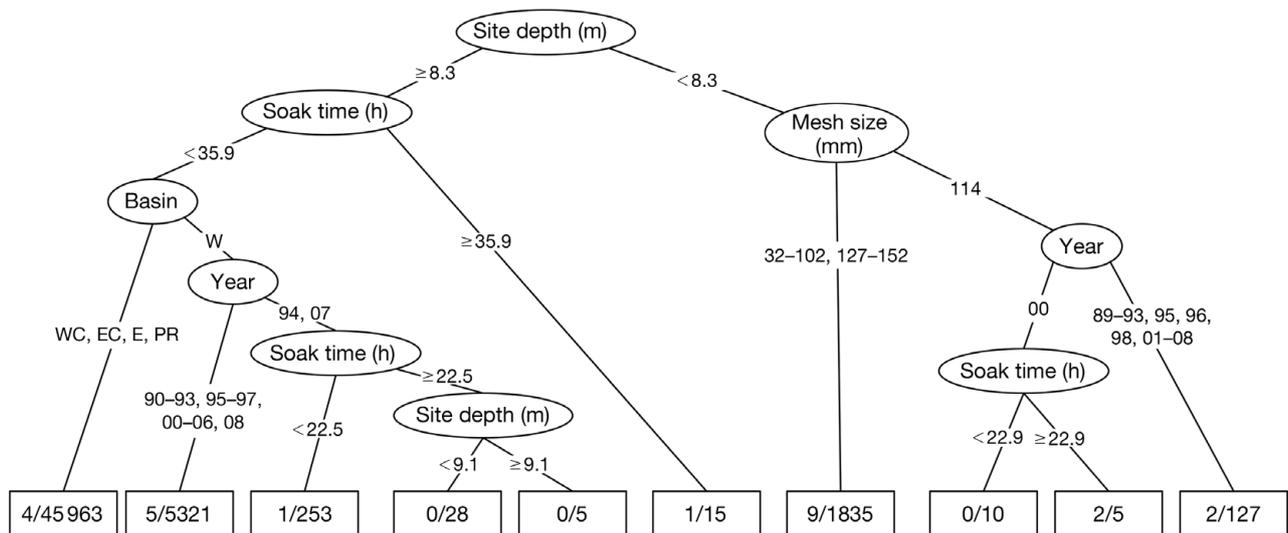


Fig. 4. Exhaustive-search-based classification tree for lake sturgeon bycatch generated by the R-package 'rpart' ('rpart' tree in text). See Fig. 2 for the explanations for internal and terminal nodes and basin abbreviations

size of 114 mm and deploying the nets for >22.9 h in waters shallower than 8.3 m during the year 2002, there was a 60 % chance of taking lake sturgeon as bycatch.

The 'tree' tree produced the largest AUC values on average (mean \pm SD: 0.9745 \pm 0.0252) on the training data, followed by the 'rpart' tree (0.9014 \pm 0.0039) and the 'party' tree (0.8908 \pm 0.0003), which suggested a better performance by the 'tree' tree on the training data

(Table 3). The ROC curves and AUC values of the test data (Fig. 5) implied that the 'party' tree (the conditional-inference classification tree) performed best (AUC = 0.77) in predicting the probability of catching lake sturgeon.

DISCUSSION

Most of the lake sturgeon caught in the PIS were juveniles (average total length: 570 mm; Fig. 1). The rare capture (31 individuals in 53 562 gillnets) of lake sturgeon indicated its low abundance in Lake Erie, and the relatively frequent captures of juveniles indicated the high mortality of juveniles caused by gillnet fishing, which might reduce recruitment in subsequent years and impede population recovery over the long term.

In this study, all 3 tree models (i.e. 'party,' 'tree' and 'rpart' trees) revealed that the west basin was a hotspot for lake sturgeon bycatch in the gillnet fisheries. Frequent incidental captures of juvenile lake sturgeon in the west basin of Lake Erie have previously been documented (GLLSCM 2002). Relatively higher abundance of juvenile lake sturgeon in the west basin of Lake Erie is likely attributed to better food availability and more suitable habitat. Availability of prey largely determines the spatial and temporal distribution of juvenile lake sturgeon (Chiasson et al. 1997). Juvenile lake sturgeon feed primarily on benthic macro-invertebrates in substrate dominated by sand and clay, such as burrowing mayflies (Ephemeroidea: *Hexagenia*) (McKinley et al. 1993, Chiasson et al. 1997, Beamish et al. 1998, McCabe et al. 2006). The west basin is the most productive basin in

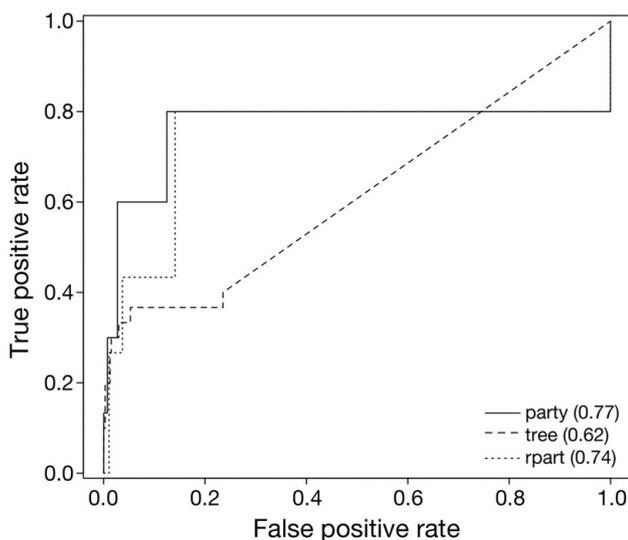


Fig. 5. Receiver operating characteristic (ROC) curves for the test data. The corresponding area under the curve (AUC) is indicated in parentheses. party: conditional-inference classification tree generated by the R-package 'party'; tree: exhaustive-search-based tree generated by the R-package 'tree'; rpart: exhaustive-search-based tree generated by the R-package 'rpart'

Lake Erie and supports a variety of benthic macro-invertebrates (Krieger et al. 1996).

Spawning migration of lake sturgeon during the summer and fall complicates their spatial distribution among Lake Huron, the St. Clair system, the Detroit River, and Lake Erie (Thomas & Haas 2002). The St. Clair system and the Detroit River are known to support relatively abundant spawning populations of lake sturgeon (Thomas & Haas 1999, 2002, GLLSCM 2002, Caswell 2003). Tagging and telemetry studies support the assumption that there are the upstream and downstream movement patterns whereby lake sturgeon migrate up to Lake Huron or down to Lake Erie from their natal spawning sites in the St. Clair system and the Detroit River (Thomas & Haas 2002, Boase 2003, Caswell 2003). Because of the spawning migration of lake sturgeon, the sources of the juveniles captured in the PIS are unclear without further research efforts.

The influence of gear type on lake sturgeon bycatch was identified by the 'tree' tree. The use of bottom gillnets increased the probability of catching lake sturgeon compared to canned gillnets, which agrees with the benthic feeding preference of lake sturgeon (Scott & Crossman 1973). Among the gillnet sets deployed in the commercial fisheries in Lake Erie from 1994 to 2007, around 63% gillnets were bottom sets (Lake Erie commercial fishery logbook data provided by the OCFA), which implies a potential threat to lake sturgeon populations from gillnet fishing.

None of the 3 classification trees detected the seasonal distribution characteristics of lake sturgeon bycatch in Lake Erie, due to limited data in our study, although the seasonal migration patterns have been well documented (Thomas & Haas 2002, Caswell 2003). A downstream migration of lake sturgeon into Lake Erie during the summer and fall was observed by Caswell (2003), suggesting high complexity in lake sturgeon migration patterns (Knights et al. 2002, Thomas & Haas 2002). Given that PIS data were collected in the fall rather than year-round, it was impractical to characterize the seasonal distribution of lake sturgeon bycatch in the present study.

Among the 3 candidate tree models (Figs. 2 to 4), the conditional-inference ('party') tree tended to select fewer variables at the internal nodes than the 2 exhaustive-search-based trees ('tree' and 'rpart'), but the split values for the common variables were similar. For example, all 3 trees split basin into west basin and other basins, and split site depth at 8.5 m ('party'), 8.6 m ('tree'), or 8.3 m ('rpart'). We suggest predicting the probability of catching lake sturgeon and making inferences on the importance of predictor variables based on the conditional-inference tree because (1)

this tree avoided over-fitting and predictor selection bias by separating the procedures for predictor selection and node splitting (Hothorn et al. 2006); (2) the smaller AUC values on the test data produced by the conditional-inference tree suggested a better predictive performance than the 2 exhaustive-search-based trees.

The tree-based model has been applied in various research fields, largely because it can capture the complex interactions in the data and the results can be interpreted easily with a visualized diagram (Ribic & Ainley 1997, Fielding 1999, Harrell 2001, Gansky 2003, Austin 2008, Cutler et al. 2009). The largest drawback of the tree-based model is the reduced accuracy compared to the commonly used generalized-linear/additive-based model approaches (Gansky 2003, Cutler et al. 2009). More accurate results can be obtained by combining a variety of suitably chosen trees, which is a method known as tree-based ensembles wherein bagging and boosting techniques can be employed (Breiman 1996, Friedman et al. 2000, Hastie et al. 2001, Cutler et al. 2009). The application of tree-based models in fishery data analyses could be extended by applying tree ensembles in future studies.

CONCLUSIONS

The conditional-inference tree model and the exhaustive-search-based tree model were applied in this study due to the high percentage of 0 observations of lake sturgeon bycatch in the PIS data. Fishing basin and gear type were identified as factors related to gillnet fishing that might have important implications for lake sturgeon conservation and management: the west basin of Lake Erie could be a hotspot for lake sturgeon bycatch in the commercial gillnet fisheries, and bottom gillnets could be regulated or modified to reduce the probability of catching lake sturgeon. Tree model comparison among the 3 candidate trees indicated that the conditional-inference tree model could provide unbiased predictor selection and better discriminative performance in predicting the probability of taking lake sturgeon as bycatch.

This study highlighted the conservation and management implications by examining the influences of gillnet fishing on the probability of obtaining lake sturgeon bycatch. Further information on overall population size, age structure, spatial and temporal distribution of lake sturgeon, gillnet fishing efforts, and gillnet set distribution in the commercial gillnet fisheries would be needed for assessing the influences of gillnet fishing on the overall lake sturgeon population in Lake Erie.

Acknowledgements. This research was supported by grants from Virginia Polytechnic Institute and State University, the USDA Cooperative State Research, Education and Extension Service, Hatch Project No. 0210510, and the Ontario Commercial Fisheries Association to Y.J.

LITERATURE CITED

- Auer NA (1999) Lake sturgeon: a unique and imperiled species in the Great Lakes. In: Taylor WW, Paola Ferreri C (eds) Great Lakes fisheries policy and management: a binational perspective. Michigan State University Press, East Lansing, MI, p 515–536
- Austin PC (2008) R and S-PLUS produced different classification trees for predicting patient mortality. *J Clin Epidemiol* 61:1222–1226
- Beamish F, Noakes D, Rossiter A (1998) Feeding ecology of juvenile lake sturgeon, *Acipenser fulvescens*, in northern Ontario. *Can Field Nat* 112:459–468
- Birstein V, Bemis W, Waldman J (1997) The threatened status of acipenseriform species: a summary. *Environ Biol Fishes* 48:427–435
- Boase JC (2003) Integrating sonic tracking and GIS to determine habitat selection and benthic prey distribution of adult lake sturgeon in Lake St. Clair. MSc thesis, University of Michigan, Ann Arbor, MI
- Bogus M (2000) Fishing the Great Lakes: an environmental history, 1783–1933. University of Wisconsin Press, Madison, WI
- Bradley A (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn* 30:1145–1159
- Breiman L (1996) Bagging predictors. *Mach Learn* 24: 123–140
- Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and regression trees. Wadsworth, Belmont, CA
- Canadian Endangered Species Conservation Council (CESCC) (2001) Wild species 2000: the general status of species in Canada. Public Works and Government Services, Ottawa
- Caswell NM (2003) Population characteristics, spawning sites, and movements of lake sturgeon (*Acipenser fulvescens*) in the Detroit River. MSc thesis, Central Michigan University, Mt. Pleasant, MI
- Chiasson W, Noakes D, Beamish F (1997) Habitat, benthic prey, and distribution of juvenile lake sturgeon (*Acipenser fulvescens*) in northern Ontario rivers. *Can J Fish Aquat Sci* 54:2866–2871
- Cutler A, Cutler D, Stevens J (2009) High-dimensional data analysis in cancer research. Springer, New York, NY
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7:1–26
- Ferguson M, Duckworth G (1997) The status and distribution of lake sturgeon, *Acipenser fulvescens*, in the Canadian provinces of Manitoba, Ontario and Quebec: a genetic perspective. *Environ Biol Fishes* 48:299–309
- Fielding AH (1999) Machine learning methods for ecological applications. Kluwer Academic Publisher, Norwell
- Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. *Ann Stat* 28: 337–374
- Gansky SA (2003) Dental data mining: potential pitfalls and practical issues. *Adv Dent Res* 17:109–114
- GLLSCM (Great Lakes Lake Sturgeon Coordination Meeting) (2002) Basin overview presentations of status and assessment activities. Sault Ste. Marie, MI
- GLLSCM (2010) Lake Sturgeon biology and population history in the Great Lakes. Available at www.fws.gov/midwest/sturgeon/biology.htm
- Hamley J (1975) Review of gillnet selectivity. *J Fish Res Board Can* 32:1944–1969
- Harrell FE (2001) Regression modeling strategies. Springer, New York, NY
- Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining, inference, and prediction. Springer Verlag, New York, NY
- Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning. *J Comput Graph Statist* 15:651–674
- Johnson JE, Jonas JL, Peck JW (2004) Management of commercial fisheries bycatch, with emphasis on lake trout fisheries of the Upper Great Lakes. Fisheries Research Report. State of Michigan Department of Natural Resources, Ann Arbor, MI
- Kinnunen R (2003) Great Lakes commercial fisheries. Report from Michigan Sea Grant, Michigan Sea Grant Extension, Marquette, MI
- Knights B, Vallazza J, Zigler S, Dewey M (2002) Habitat and movement of lake sturgeon in the upper Mississippi River system, USA. *Trans Am Fish Soc* 131:507–522
- Krieger K, Schloesser D, Manny B, Trisler C, Heady S, Ciborowski J, Muth K (1996) Recovery of burrowing mayflies (Ephemeroptera: Ephemeridae: *Hexagenia*) in western Lake Erie. *J Gt Lakes Res* 22:254–263
- Lo N, Jacobson L, Squire J (1992) Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can J Fish Aquat Sci* 49:2515–2526
- Maunder M, Langley A (2004) Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. *Fish Res* 70:389–395
- Maunder M, Punt A (2004) Standardizing catch and effort data: a review of recent approaches. *Fish Res* 70: 141–159
- McCabe D, Beekey M, Mazloff A, Marsden J (2006) Negative effect of zebra mussels on foraging and habitat use by lake sturgeon (*Acipenser fulvescens*). *Aquat Conserv* 16: 493–500
- McKinley S, Singer T, Ballantyne J, Power G (1993) Seasonal variation in plasma nonesterified fatty acids of lake sturgeon (*Acipenser fulvescens*) in the vicinity of hydroelectric facilities. *Can J Fish Aquat Sci* 50:2440–2447
- McKinley S, Van Der Kraak G, Power G (1998) Seasonal migrations and reproductive patterns in the lake sturgeon, *Acipenser fulvescens*, in the vicinity of hydroelectric stations in northern Ontario. *Environ Biol Fishes* 51:245–256
- Meyer J, Ingersoll C, McDonald L, Boyce M (1986) Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. *Ecology* 67:1156–1166
- Miller R (1974) The jackknife—a review. *Biometrika* 61:1–15
- Noakes D, Beamish F, Rossiter A (1999) Conservation implications of behaviour and growth of the lake sturgeon, *Acipenser fulvescens*, in northern Ontario. *Environ Biol Fishes* 55:135–144
- Ortiz M, Legault C, Ehrhardt N (2000) An alternative method for estimating bycatch from the US shrimp trawl fishery in the Gulf of Mexico, 1972–1995. *Fish Bull* 98:583–599
- Ribic C, Ainley D (1997) The relationships of seabird assemblages to physical habitat features in Pacific equatorial waters during spring 1984–1991. *ICES J Mar Sci* 54: 593–599
- Scott WB, Crossman EJ (1973) Freshwater fishes of Canada. *Fish Res Board Can Bull* 184. Fisheries Research Board of Canada, Ottawa

- Shono H (2008) Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fish Res* 93:154–162
- Stefansson G (1996) Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES J Mar Sci* 53:577–581
- Thomas M, Haas R (1999) Capture of lake sturgeon with set-lines in the St. Clair River, Michigan. *N Am J Fish Manag* 19:610–612
- Thomas M, Haas R (2002) Abundance, age structure, and spatial distribution of lake sturgeon, *Acipenser fulvescens*, in the St Clair System. *J Appl Ichthyol* 18:495–501
- Thomas M, Haas R (2005) Status of yellow perch and walleye populations in Michigan waters of Lake Erie, 1999–2003. *Fish Res Rep No 2082*. Michigan Department of Natural Resources, Ann Arbor, MI
- Threader RW, Broussaeu CS (1986) Biology and management of the lake sturgeon in the Moose River, Ontario. *N Am J Fish Manag* 6:383–390
- Tweedie M (1984) An index which distinguishes between some important exponential families. In: Ghosh JK, Roy J (eds) *Statistics: applications and new directions*. Proc Indian Stat Inst Golden Jubilee Int Conf. Indian Statistical Institute, Calcutta, p 579–604
- Welsh A (2004) Factors influencing the effectiveness of local versus national protection of migratory species: a case study of lake sturgeon in the Great Lakes, North America. *Environ Sci Policy* 7:315–328
- Ye Y, Al-Husaini M, Al-Baz A (2001) Use of generalized linear models to analyze catch rates having zero values: the Kuwait driftnet fishery. *Fish Res* 53:151–168

*Editorial responsibility: Dave Hodgson,
University of Exeter, Cornwall Campus, UK*

*Submitted: July 7, 2010; Accepted: December 30, 2010
Proofs received from author(s): March 9, 2011*