Journal of Great Lakes Research 37 (2011) 325–334

Contents lists available at ScienceDirect

Journal of Great Lakes Research

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

Assessment of landed and non-landed by-catch of walleye, yellow perch and white perch from the commercial gillnet fisheries of Lake Erie, 1994–2007

Yan Li [a,*], Yan Jiao [a], Kevin Reid [b]

[a] Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0321, USA
[b] Ontario Commercial Fisheries Association, Box 2129, 45 James Street, Blenheim, Ontario, Canada N0P 1A0

ARTICLE INFO

Article history:
Received 24 August 2010
Accepted 10 February 2011
Available online 21 March 2011

Communicated by John Janssen

Index words:
Lake Erie
By-catch
Walleye
Yellow perch
White perch
Delta-AdaBoost model

ABSTRACT

Theoretical by-catch (including landed and non-landed bycatch) of walleye (*S. vitreus*), yellow perch (*Perca flavescens*), and white perch (*Morone americana*) from the Lake Erie commercial gillnet fisheries during 1994–2007, was predicted by a delta model developed on the fishery-independent survey data (Lake Erie Partnership Index Fishing Survey). The delta model consisted of one generalized additive model and one AdaBoost model. The generalized additive model was used to predict non-zero catches of the by-catch species, and the AdaBoost model was used to predict the probability of obtaining non-zero catches. Non-landed by-catch was estimated as the difference between the theoretical by-catch predicted from the delta model and the landed by-catch recorded in the commercial fishery data. The theoretical by-catch of walleye was relatively higher in the west basin in October. A higher theoretical by-catch of yellow perch occurred in the west central basin in November, and a higher theoretical by-catch of white perch occurred in the west central basin in October. We observed higher levels of non-landed by-catch of walleye in the west basin during August to September, higher levels of non-landed by-catch of yellow perch in the west central and east central basins in November, and higher levels of non-landed by-catch of white perch in the west basin in August and November. The combination of the AdaBoost model with the delta model provided an alternative model in by-catch analyses when the percentage of zero observations was high.

Introduction

By-catch, particularly non-landed by-catch (including released and discarded by-catch), is of concern to conservation biology and fisheries management because of wasted living resources, threats to endangered species, and impacts on ecosystem stability (Alverson et al., 1994; Crowder and Murawski, 1998; Hall et al., 2000; Johnson et al., 2004; Harrington et al., 2005; Kelleher, 2005). In a multi-species fishery, the major reasons for by-catch can be related to landing size limits, individual species-specific quota limits, dockside prices, and status of fish markets (Murawski, 1996; Punt et al., 2006). It is essential to characterize and assess by-catch, especially non-landed by-catch, which may have important implications for fisheries management. However, by-catch, and especially non-landed by-catch, is not always measured or reported in many fisheries (Johnson et al., 2004; Borges et al., 2005; Punt et al., 2006).

Lake Erie supports one of the world's largest freshwater commercial fisheries, where the Lake Erie Ontario commercial gillnet fisheries contribute a large proportion (Kinnunen, 2003). The Lake Erie Ontario commercial gillnet fisheries are majorly managed through quota system and restricted fishing season/location. Individual transferable quota (ITQ) species such as walleye (*Sander vitreus*), yellow perch (*Perca flavescens*) and lake whitefish (*Coregonus clupeaformis*) dominate the landings of the Lake Erie Ontario commercial gillnet fisheries (Kinnunen, 2003; Thomas and Haas, 2005). White perch (*Morone americana*), as an invasive and non-quota limited species, has imposed considerable impacts on native fish communities and lake ecosystem (Scott and Crossman, 1973; Schaeffer and Margraf, 1987; Parrish and Margraf, 1990).

Due to the overlap in biological characteristics between target and by-catch species, the overlap in spatial and temporal distributions between target and by-catch species, the relatively higher selectivity and mortality of gillnets, and the ITQ system for quota versus non-quota species, the landed and non-landed by-catch of walleye, yellow perch and white perch in the Lake Erie Ontario commercial gillnet fisheries occurred but were not recorded in the catch reporting system prior to 2010 (Scott and Crossman, 1973; Hamley, 1975; Kinnunen, 2003; Johnson et al., 2004; K. Reid, Ontario Commercial Fisheries' Association, personal communication). Failure to take landed and non-landed by-catch into account in stock assessments may increase the bias in estimating fishing mortality, population abundance and available quotas, and may conceal the impacts of by-catch on ecosystem stability (Johnson et al., 2004).

* Corresponding author. Tel.: +1 540 808 8373.
E-mail address: yanli08@vt.edu (Y. Li).

0380-1330/$ – see front matter © 2011 International Association for Great Lakes Research. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jglr.2011.02.007

In this study, the theoretical by-catch refers to all incidental captures of fish species in gillnet fishing, including landed and non-landed (released and discarded) by-catch (Crowder and Murawski, 1998; Harrington et al., 2005). The theoretical by-catch of these three species was assessed based on the commercial fishery data and the fishery-independent survey data. The commercial fishery data included only landed by-catch and there were no data about non-landed by-catch that actually occurred at lake during our study period (August–November, 1994–2007); therefore, we developed a model using the fishery-independent survey data and applied this model to the commercial fishery data to predict the theoretical by-catch and to estimate the non-landed by-catch occurring at lake. The fishery-independent survey data were collected from the Lake Erie Partnership Index Fishing Survey (PIS) conducted annually by the Ontario Ministry of Natural Resources (OMNR) and the Ontario Commercial Fisheries' Association (OCFA) since 1989.

The PIS data contained a high percentage of zero catches (75–77%). The presence of zeros may invalidate the assumption of normality we commonly use, and may cause computational difficulties. Ignoring a high proportion of zero observations may result in a loss of information about spatial or temporal distribution characteristics of fish stock. The approaches to deal with zero values in previous studies can be categorized into two types. One approach is to add a small constant to each observation of the response variable, followed by a generalized linear or additive model analysis (Ortiz et al., 2000; Maunder and Punt, 2004; Shono, 2008). However, the estimation results are sensitive to the choice of this constant (Ortiz et al., 2000; Maunder and Punt, 2004).

The other approach to deal with zeros in fishery data analyses is to apply either the delta model or the Tweedie distribution model. In the delta model, non-zero values and zero values are modeled separately by two sub-models (Lo et al., 1992; Stefansson, 1996; Ye et al., 2001; Maunder and Punt, 2004). The delta model has been widely used to analyze fishery data having a high percentage of zero observations (Lo et al., 1992; Stefansson, 1996; Ortiz et al., 2000; Ye et al., 2001). By contrast, the Tweedie distribution model can handle zero values uniformly along with non-zero values, where the Tweedie distribution is considered to be a Poisson–Gamma compound distribution when its power parameter is greater than 1 but less than 2 (Tweedie, 1984; Shono, 2008). The Tweedie distribution model was recommended in by-catch analyses when data contained a high percentage of zero observations (Shono, 2008).

To apply the delta model, we commonly use the generalized linear/additive model with an assumption of a binomial distribution as the sub-model to predict the probability of obtaining non-zero observations. Alternatively, in this study, we applied the AdaBoost model to predict the probability of obtaining non-zero catches in the delta model. The AdaBoost model was originally used for classification problems. If the response variable takes the value either 1 or −1, the algorithm used to produce the value 1 or −1 is called the classifier. The final classifier is obtained by linearly combining a sequence of classifiers fitted using reweighted data at each iteration (Freund and Schapire, 1996; Friedman et al., 2000; Hastie et al., 2001; Kawakita et al., 2005). The AdaBoost model has been employed to predict the occurrence of large silky shark (*Carcharhinus falciformis*) by-catch in a tuna (*Thunnus* spp.) purse-seine fishery, and yielded more accurate and stable predictions compared with the generalized additive models (Kawakita et al., 2005).

In this study, we aimed to (1) develop a delta model that was composed by one generalized additive model and one AdaBoost model using the PIS data; (2) predict the theoretical by-catch of walleye, yellow perch and white perch by applying the developed delta model to the commercial fishery data; (3) estimate non-landed by-catch for each species by comparing the theoretical by-catch predicted from the delta model and the landed by-catch recorded in the commercial fishery data; (4) analyze the theoretical by-catch and the percentage of non-landed by-catch by major target species, month, basin and year; and (5) highlight the implications for by-catch management regarding these three species in the Lake Erie commercial gillnet fisheries.

Methods

Data

We had two datasets available for by-catch analyses: the PIS data from 1989 to 2008 and the commercial fishery data (i.e., the commercial gillnet fishery catch and effort data) from 1994 to 2007. Both data sets contained information about fish landings, fishing effort, gear characteristics and environmental factors.

The PIS data were provided by the OCFA. In the PIS survey, experimental gillnets with fourteen mesh sizes ranging from 32 to 152 mm were set at sites randomly distributed across the Ontario waters of Lake Erie in the fall (August to November) from 1989 to 2008 (OCFA, 2007). A total of 53,662 PIS gillnet sets were available for model development. The commercial fishery data were provided by the Lake Erie Management Unit of the OMNR. In the commercial fishery data, 32,349 records were analyzed for walleye, 32,282 records for yellow perch and 32,392 records for white perch. Each record in the commercial fishery data from 1994 to 2001 represents the data from one gillnet set; each record in the commercial fishery data from 2002 to 2007 is based on the daily catch report that includes the data from 1 to 5 gillnet sets.

In this study, we analyzed those PIS and commercial fishery records in which the gillnet sets had a mesh size ranging from 51 to 140 mm, were distributed in the waters of 3 to 66 m deep, and soaked for 9 to 36 h in the fall (August to November) from 1994 to 2007; we examined those explanatory variables included in both the PIS data and the commercial fishery data, i.e., two continuous variables (site depth and soak time) and five categorical variables (basin, year, month, gear type, and mesh size). We selected these records and variables for analysis because the values of the explanatory variables in these records fell into the same range in both the PIS data and the commercial fishery data, which ensure the prediction accuracy when we applied the delta model developed on the PIS data to these records in the commercial fishery data.

Delta model based on the PIS data and the theoretical by-catch prediction

In this study, a delta model was developed to analyze the PIS data (Lo et al., 1992; Ortiz et al., 2000; Maunder and Punt, 2004; Kawakita et al., 2005; Shono, 2008). To determine the distribution of log-transformed PIS data, a small constant of 0.001 was added to each observation of response variable before log-transformation in order to avoid numerical errors caused by zero observations. Histograms of log-transformed PIS data showed that zeros were separated from non-zero values that were assumed to follow a normal distribution (Fig. 1, Stefansson, 1996; Ye et al., 2001; Murray, 2004; Damalas et al., 2007).

The delta model consisted of two sub-models, the generalized additive model (GAM) to predict non-zero catches and the AdaBoost model to predict probability of obtaining non-zero catches. The predicted theoretical by-catch ($\hat{B}$) from the delta model can be obtained by multiplying the estimates from these two sub-models (Lo et al., 1992; Pennington, 1996; Stefansson, 1996; Ortiz et al., 2000; Ye et al., 2001; Maunder and Punt, 2004; Murray, 2004):

$$\hat{B} = \hat{d} \times \hat{p},$$

where $\hat{d}$ is the predicted non-zero catches of the by-catch species, and $\hat{p}$ is the predicted probability of obtaining non-zero catches of the by-catch species.

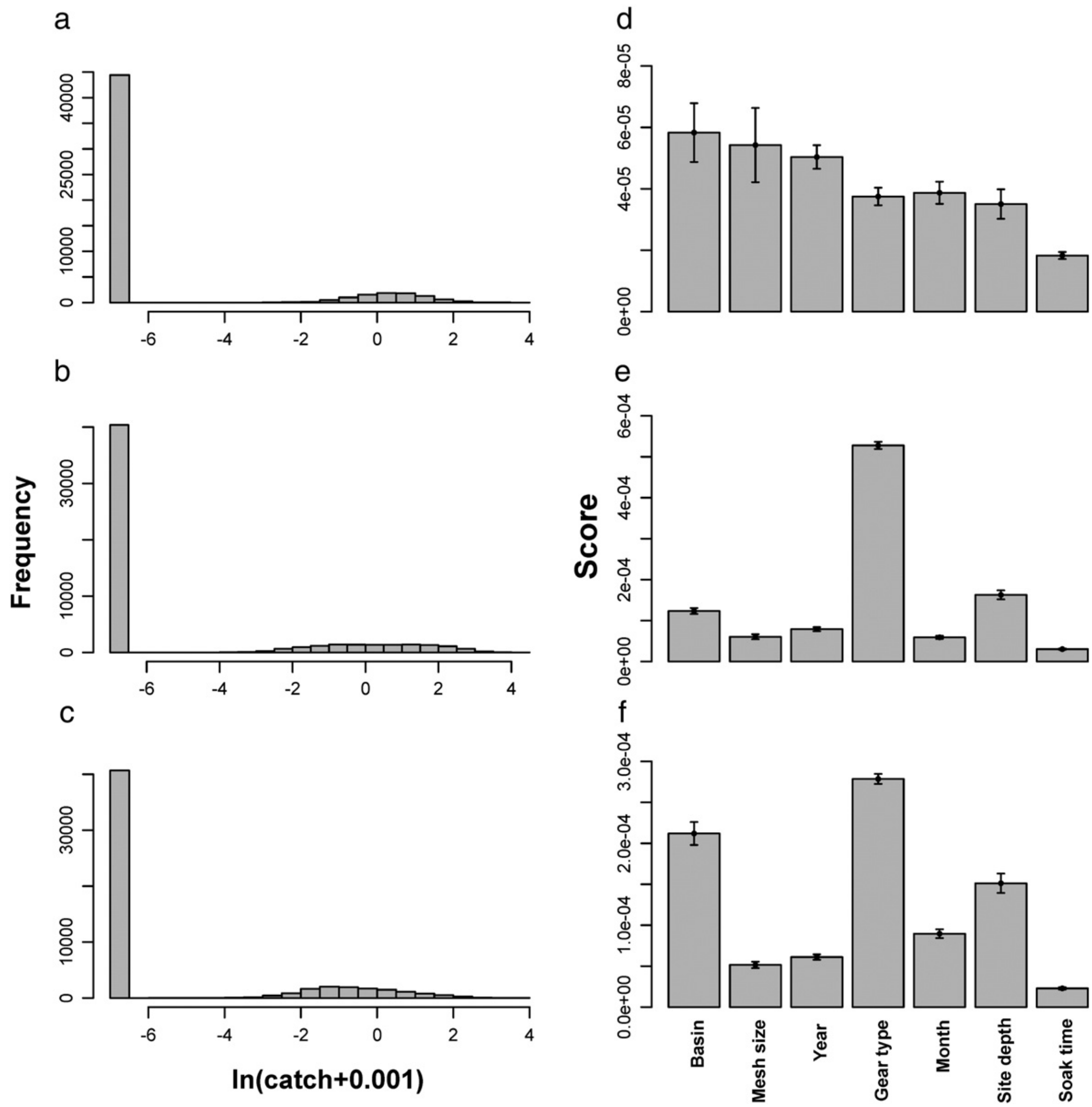

Fig. 1. Panels a–c: histograms of log-transformed catch data (kg) of walleye (a), yellow perch (b) and white perch (c), collected by the Lake Erie Partnership Index Fishing Survey (PIS), 1989–2008. Panels d–f: score plots for walleye (d), yellow perch (e) and white perch (f), derived from an AdaBoost model to predict the probability of obtaining non-zero catches. The scores are averaged over 1000 simulations. Error bars show the standard deviation.

A generalized additive model (GAM) was constructed to predict the non-zero catches $\hat{d}$. In the GAM, the effect of each variable can be modeled by a smooth function (Hastie and Tibshirani, 1990). By assuming that non-zero catches followed a lognormal distribution, the GAM can be written as:

$$ln\left(\hat{d}\right) = \beta_0 + \sum_{j=1} f_j\left(X_j\right),$$

where $\hat{d}$ is the predicted non-zero catches of the by-catch species, β_0 is the intercept, f_j is a smooth function (a spline or a loess smoother) for the jth explanatory variable X_j.

To select variables, we examined the correlation coefficients among all seven explanatory variables to detect highly correlated variables. A preliminary stepwise selection based on Akaike Information Criterion (AIC, Akaike, 1974) and statistical test was conducted to eliminate one of the highly correlated pair of explanatory variables. The variable that yielded a larger AIC value and had less significant effect on the response variable was eliminated from the correlated pair. The remaining variables were selected through a stepwise procedure based on AIC and statistical test (Akaike, 1974; Burnham and Anderson, 2002). In the stepwise procedure, at each step we evaluated the remaining variables and selected into model the one that most greatly reduced AIC value or showed the most significant effect on the response variable. We repeated this step until all significantly important variables were selected into model, i.e., all the variables left showed insignificant effects (P-value$>$0.05), and the AIC value decreased slightly (decrease in AIC$<=$5) or started to increase. We tested the interaction terms by including them in the model. Those interaction terms were either insignificant (P-value$>$0.05) or highly correlated with main factors. Therefore, we did not include interaction terms in the model (Maunder and Punt, 2004; Damalas et al., 2007).

To predict the probability of obtaining non-zero catches of the by-catch species $\hat{p}$, the AdaBoost model was applied. We denoted the vector of explanatory variables as X, and the response variable as

$Y \in \{-1, 1\}$. The value -1 represented the event of catching no fish of the species of interest (i.e., zero catches) and the value 1 represented the event of catching at least one fish of the species of interest (i.e., non-zero catches). Then the problem was treated as a two-group classification problem (Kawakita et al., 2005). In the AdaBoost model, all seven variables were included, and the final classifier $F(x)$ was constructed as follows (Freund and Schapire, 1996; Friedman et al., 2000; Hastie et al., 2001; Kawakita et al., 2005):

(1) Initialize the weights for each observation $w_i = 1/N$, $i = 1, 2, \ldots, N$, where N is the number of observations.
(2) For $t = 1$ to T, where T is the number of iterations:
  (a) Fit the classifier $g_t(x)$ using the data weighted by w_i and obtain a probability estimate $g_t(x_i) = \Pr(\hat{y}_i = 1 \mid x_i)$, i.e., the probability that the predicted value for y_i equals 1 given x_i.
  (b) Set $h_t(x_i) = \frac{1}{2} ln(g_t(x_i) / (1 - g_t(x_i)))$, which indicates the contribution of the classifier $g_t(x)$ to the final classifier $F(x)$.
  (c) Update the weights for each observation for the next iteration, $w_i = \frac{exp(-y_i h_t(x_i))}{\sum_{i=1}^{N} exp(-y_i h_t(x_i))}$.
(3) Set $H(x_i) = \sum_{t=1}^{T} h_t(x_i)$. The final classifier for the ith observation, $F(x_i) = \begin{cases} 1, & if \;\; H(x_i) > 0; \\ -1, & if \;\; H(x_i) < 0 \end{cases}$
(4) The probability of obtaining non-zero catches for the ith observation, $\hat{p}_i = \frac{e^{2H(x_i)}}{1 + e^{2H(x_i)}}$.

At iteration t, those observations that were misclassified at the previous iteration had their weights increased, whereas the weights were decreased for those classified correctly. As iterations proceeded, each classifier was forced to focus on those observations that were difficult to classify correctly. As a result of combining these classifiers, the final classifier provided accurate estimates, either the presence/absence of the fish species or the probability of obtaining non-zero catches.

Score plots generated from the AdaBoost model were utilized to detect the factors that had important influences on the response variable (i.e., the probability of catching this species in our case, Kawakita et al., 2005). In the score plots, scores for each variable were calculated as the frequency of this variable being selected to split nodes of classifiers. A higher score of an explanatory variable indicates that this variable is more frequently selected to construct classifiers and therefore more important to influence the response variable.

It is necessary to determine the optimal number of iterations T to avoid over-fitting problems, i.e., the predictive ability of the AdaBoost model may decrease as the number of iterations increases (Kawakita et al., 2005). We split the PIS data into two sub-datasets with roughly equal size when determining T. We used one sub-dataset for model building, which was called the training data. The corresponding error (i.e., the proportion of the observations misclassified among all observations) from model building was called the training error. The other sub-dataset was used for prediction, which was called the test data. When we applied the model developed on the training data to the test data, the corresponding error from the test data was called the test error. After examining the trends of training and test errors in the AdaBoost model up to 1000 interactions, we determined $T = 200$ as the optimal number of iterations in this study because both training and test errors decreased dramatically and started to be stabilized when the iteration proceeded around 200.

To confirm the superiority of the delta model composed by a generalized additive model and an AdaBoost model (Delta-AdaBoost), five candidate models were constructed for comparison. The five candidate models included three delta models and two Tweedie-distribution models. The three delta models were the delta model comprising two generalized linear models (Delta-GLM), the delta model comprising two generalized linear models with polynomial terms up to degree 3 (Delta-GLM-Poly), and the delta model composing two generalized additive models (Delta-GAM); the two Tweedie-distribution models were the generalized linear model with Tweedie distribution (GLM-Tweedie), and the generalized additive model with Tweedie distribution (GAM-Tweedie). Each of the three candidate delta models comprised two sub-models, i.e., one sub-model to estimate the non-zero catches $\hat{d}$ assuming a lognormal distribution and the other sub-model to estimate the probability of obtaining non-zero catches $\hat{p}$ assuming a binomial distribution. Each of the two candidate Tweedie-distribution models comprised one model to estimate catches that was constructed by applying the Tweedie distribution to the generalized linear or additive model (Tweedie, 1984; Shono, 2008). Each candidate model selected variables through the stepwise procedure to fit the PIS data best; therefore, the variables selected into each candidate model can be different.

Model comparison was conducted through the 5-fold cross-validation approach (Breiman et al., 1984; Tweedie, 1984; Hastie et al., 2001; Damalas et al., 2007). The PIS data were split into five sub-datasets with roughly equal size. Each sub-dataset was used as test data for prediction, and the remaining four sub-datasets were combined as training data to build the model. The Delta-AdaBoost model and each candidate model were fit using the training data, and then applied to the test data. The training error and test error from each pair of training and test datasets for each model were calculated as follows (Hastie et al., 2001; Damalas et al., 2007):

$$Training(Test) \quad error = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|,$$

where N is the number of observations, y_i is the ith observation, and $\hat{y}_i$ is the estimated value (the predicted value) by the model for the ith observation. The model providing lower training error and test error was judged as the one with better performance (Hastie et al., 2001; Damalas et al., 2007; Shono, 2008).

Non-landed by-catch estimation and bootstrap method

The delta model developed on the PIS data was applied to the commercial fishery data to predict the theoretical by-catch from the commercial gillnet fisheries. The non-landed by-catch (kg) for each species was estimated as the difference between the theoretical by-catch predicted from the delta model and the landed by-catch recorded in the commercial fishery data. For each species, we presented the results of: (1) total theoretical by-catch across all analyzed records in the commercial fishery data; (2) percentage composition of total theoretical by-catch by major target species, month, basin and year; (3) percentage of non-landed by-catch (percentage of total estimated non-landed by-catch among total theoretical landings of the species of interest; total theoretical landings were a summation of total theoretical by-catch predicted from the delta model and total target catch of the species of interest recorded in the commercial fishery data) across all the analyzed records in the commercial fishery data; and (4) percentage of non-landed by-catch by major target species, month, basin and year.

Uncertainties in the predicted theoretical by-catch and the estimated non-landed by-catch were analyzed through a nonparametric bootstrap approach. In the bootstrap approach, we re-sampled the records in the PIS data and constructed the Delta-AdaBoost model for 1000 times to obtain a joint distribution for parameters in the Delta-AdaBoost model. We applied this joint distribution of parameters to the commercial fishery data to predict the theoretical by-catch with a 95% confidence interval. This analysis was programmed in R (Version 2.9.2).

Results

Correlation analysis did not detect any high correlation among the seven explanatory variables in the PIS data for all three species. After the stepwise selection, all seven variables examined were selected into the GAM (Table 1). In the prediction of non-zero catches, the GAMs explained 20%, 31% and 22% of the deviance of PIS data for walleye, yellow perch and white perch respectively. Results from the GAM (Table 1) indicated five factors that had significant impacts on non-zero catches, i.e., mesh size, year, basin, gear type and month. Soak time was a significant factor that influenced non-zero catches of walleye and yellow perch; site depth was a significant factor that influenced non-zero catches of yellow perch and white perch. Score plots (Fig. 1) from the AdaBoost model indicated that gear type had the greatest impact on the probability of catching yellow perch and white perch; the probability of catching walleye was mostly affected by basin.

Model comparison among the Delta-AdaBoost model and five candidate models through 5-fold cross-validation confirmed that the Delta-AdaBoost model yielded the smallest training and test errors (Table 2). This provided evidence that the Delta-AdaBoost model fit the PIS data better and produced more accurate estimation and prediction compared with those five candidate models.

Yellow perch, white bass (*Morone chrysops*), lake whitefish and white perch were the major target species involved in walleye by-catch. The major target species identified for Yellow perch by-catch included white perch, walleye, white bass and lake whitefish. The major target species associated with white perch by-catch was yellow perch.

White perch had the greatest number of total theoretical by-catch across analyzed records, followed by walleye and yellow perch (Table 3). Among the total theoretical by-catch of walleye, 81% on average was observed when targeting yellow perch (Fig. 2a), 33% on average when fishing in October (Fig. 2b), 59% on average in the west basin (Fig. 2c), and 74% on average in the years before 2000 (Fig. 2d). Yellow perch by-catch occurred most often when targeting white perch (72% on average, Fig. 3a), when fishing in November (40% on average, Fig. 3b) or in the west central basin (63% on average, Fig. 3c), or when fishing during years of 2001–2005 (71% on average, Fig. 3d). Among the total theoretical by-catch of white perch, 99% on average was obtained when targeting yellow perch (Fig. 4a), 27% on average when fishing in October (Fig. 4b), 53% on average when fishing in the west central basin (Fig. 4c), and 48% on average in the years of 1994 and 2005–2006 (Fig. 4d).

White perch had the highest percentage of non-landed by-catch (44.61% on average) across analyzed records (Table 3), followed by walleye (10.89% on average) and yellow perch (0.79% on average). The highest percentage of non-landed walleye by-catch occurred when targeting yellow perch (70% on average, Fig. 2e), when fishing in August to September (12% on average, Fig. 2f) or in the west basin (19% on average, Fig. 2g), or when fishing in the years 1994 and 2003 (26% and 22% on average respectively, Fig. 2h). Higher percentage of non-landed yellow perch by-catch was observed when targeting white perch and white bass (88% and 90% on average respectively, Fig. 3e), when fishing in November (3% on average, Fig. 3f) or in the west central and east central basins (0.9% on average, Fig. 3g), or when fishing in the year 2001 (11% on average, Fig. 3h). Among total theoretical white perch landings when targeting yellow perch, 56% of white perch was discarded on average (Fig. 4e). An average of 55% and 57% of total theoretical white perch landings was discarded when fishing in August and November respectively (Fig. 4f), an average of 64% when fishing in the west basin (Fig. 4g), and an average ranging from 33% to 83% during the years 1995–2006 (Fig. 4h).

Discussion

This study is the first time that a Delta-AdaBoost model was applied in fishery data analyses. Results showed that a combination of

Table 1
A stepwise generalized additive model (GAM) building to predict non-zero catches of walleye, yellow perch and white perch. A lognormal distribution was assumed.

	Variables selected into model	df	Deviance[a]	AIC[a]	P-value (χ^2)	Deviance decrement	Cumulative% of deviance explained
Walleye							
0	Null		46,888	41,248			
1	Mesh size	13	42,881	40,448	$<2.2\times10^{-16}$	4007	9.7
2	Year	19	41,567	40,198	$<2.2\times10^{-16}$	1314	12.9
3	Basin	4	40,153	39,886	$<2.2\times10^{-16}$	1414	16.3
4	Gear type[b]	1	39,263	39,681	$<2.2\times10^{-16}$	890	18.5
5	Month	3	38,921	39,606	1.4×10^{-15}	342	19.3
6	*f* (Soak time)	1	38,815	39,583	5.1×10^{-7}	106	19.6
7	*f* (Site depth)	1	38,798	39,559	0.04	17	19.6
Yellow perch							
0	Null		368,175	81,816			
1	Mesh size	13	342,416	80,879	$<2.2\times10^{-16}$	25,759	7.0
2	Year	19	311,055	79,641	$<2.2\times10^{-16}$	31,361	15.5
3	Gear type	1	282,739	78,376	$<2.2\times10^{-16}$	28,316	23.2
4	Basin	4	258,219	77,179	$<2.2\times10^{-16}$	24,520	29.9
5	*f* (Soak time)	1	256,444	77,089	$<2.2\times10^{-16}$	1775	30.3
6	Month	3	255,430	76,917	2.2×10^{-11}	1014	30.6
7	*f* (Site depth)	1	255,138	76,887	9.7×10^{-5}	292	30.7
White perch							
0	Null		70,984	58,918			
1	Year	19	67,067	58,219	$<2.0\times10^{-16}$	3917	5.5
2	Gear type	1	64,534	57,721	$<2.2\times10^{-16}$	2533	9.1
3	Mesh size	13	60,588	56,927	$<2.2\times10^{-16}$	3946	14.6
4	Month	3	58,653	56,512	$<2.2\times10^{-16}$	1935	17.4
5	*f* (Site depth)	1	56,993	56,141	$<2.2\times10^{-16}$	1660	19.7
6	Basin	4	55,521	55,412	$<2.0\times10^{-16}$	1472	21.8
7	*f* (Soak time)	1	55,514	55,330	0.19	7	21.8

[a] The deviance and AIC values shown for each variable in the table were the deviance and AIC values for the model after selecting this variable, i.e., the model including this variable and the variables listed above it.
[b] Gear type refers to how gillnets were deployed, i.e., canned (suspended in the water column) or bottomed (set at the bottom).

Table 2
Training and test errors from the 5-fold cross-validation for the Delta-AdaBoost model and the five candidate models.

Model	Training error						Test error					
	1	2	3	4	5	Average	1	2	3	4	5	Average
Walleye												
Delta-GLM	0.476	0.474	0.480	0.482	0.476	0.478	0.482	0.485	0.475	0.468	0.481	0.478
Delta-GLM-Poly	0.472	0.469	0.475	0.477	0.471	0.473	0.478	0.480	0.471	0.463	0.476	0.473
Delta-GAM	0.472	0.469	0.475	0.477	0.471	0.473	0.478	0.480	0.471	0.464	0.476	0.474
GLM-Tweedie	0.487	0.484	0.490	0.492	0.487	0.488	0.491	0.495	0.485	0.479	0.494	0.489
GAM-Tweedie	0.482	0.480	0.485	0.487	0.482	0.483	0.487	0.489	0.481	0.474	0.489	0.484
Delta-AdaBoost	0.440	0.438	0.442	0.444	0.438	0.441	0.449	0.451	0.441	0.432	0.446	0.444
Yellow perch												
Delta-GLM	0.768	0.772	0.779	0.777	0.769	0.773	0.784	0.785	0.756	0.760	0.793	0.776
Delta-GLM-Poly	0.696	0.694	0.709	0.705	0.699	0.701	0.713	0.713	0.685	0.687	0.720	0.703
Delta-GAM	0.702	0.701	0.713	0.710	0.704	0.706	0.718	0.720	0.691	0.692	0.725	0.709
GLM-Tweedie	0.794	0.795	0.803	0.802	0.794	0.797	0.803	0.803	0.778	0.789	0.822	0.799
GAM-Tweedie	0.728	0.728	0.738	0.738	0.728	0.732	0.744	0.741	0.708	0.723	0.752	0.734
Delta-AdaBoost	0.672	0.673	0.684	0.682	0.674	0.677	0.687	0.691	0.669	0.666	0.694	0.682
White perch												
Delta-GLM	0.327	0.323	0.319	0.334	0.326	0.326	0.342	0.325	0.320	0.323	0.326	0.327
Delta-GLM-Poly	0.321	0.317	0.315	0.326	0.317	0.319	0.338	0.317	0.315	0.314	0.318	0.321
Delta-GAM	0.320	0.317	0.315	0.326	0.318	0.319	0.337	0.316	0.315	0.314	0.320	0.321
GLM-Tweedie	0.341	0.348	0.345	0.352	0.349	0.347	0.357	0.346	0.347	0.340	0.349	0.348
GAM-Tweedie	0.332	0.338	0.336	0.343	0.339	0.338	0.350	0.336	0.336	0.330	0.340	0.338
Delta-AdaBoost	0.311	0.317	0.314	0.323	0.323	0.317	0.316	0.293	0.296	0.292	0.299	0.299

a generalized additive model with an AdaBoost model increased model goodness-of-fit. Model comparison is a problem when models are less comparable through information-based criteria such as AIC (Shono, 2008). The n-fold cross-validation method was suggested in this study as a tool of model comparison in order to overcome the difficulties in comparing models that were constructed in different frameworks (Shono, 2008). The Delta-AdaBoost model was the best model outperforming the five candidate models we examined. However, the model selection could be complicated by the characteristics of the data set used to build the model and the number of candidate models constructed for comparison.

Walleye, yellow perch and white perch were easily caught incidentally in the gillnet fisheries in Lake Erie because of the overlaps in size and spatial distribution with target species (Scott and Crossman, 1973; Jester, 1977; Kinnunen, 2003; Johnson et al., 2004). Occurrence of non-landed by-catch of walleye and yellow perch can be attributed to seasonal/spatial restrictions, limited quotas and lower landed values for large walleye and yellow perch. Low dockside value and unmarketable size are the major reasons for white perch non-landed by-catch. Specific reasons for non-landed by-catch of these three species can be extrapolated by incorporating information about age and/or length structure into by-catch assessments in future studies.

Higher percentage of non-landed white perch by-catch compared with walleye and yellow perch can be attributed to fewer white perch target landings and higher theoretical white perch by-catch. The chances that fishermen targeted white perch in the commercial gillnet fisheries was low (only 1–2% records among all analyzed records targeted white perch); therefore, we had very limited white perch landings (only 1/23 of the total walleye landings and 1/18 of the total yellow perch landings) but a high chance of by-catching white perch (Table 3). Meanwhile, lower popularity of white perch leads to higher non-landed white perch by-catch even though theoretically white perch is a non-quota species and fishermen can keep all they catch. Because white perch invaded Lake Erie and has imposed considerable influences on lake ecosystem and fish communities through competition, more fishing efforts on white perch may potentially increase its fishing mortality and mitigate the competition between white perch and native species. Given the high percentage of non-landed white perch by-catch, developing a new market for white perch could be a potential market-based solution to avoid wasting its resources.

The abundance of walleye in Lake Erie kept decreasing since the late 1980s and remained at a low level during 2000–2003; the abundance of yellow perch started to increase in the late 1990s (WTG, 2010; YPTG, 2010). To keep a sustainable percid fishery, the Lake Erie Committee launched the Lake Erie Coordinated Percid Management Strategy (the Strategy) in 2001. The total allowable catch (TAC) for both walleye and yellow perch was reduced to a conservative level and this conservative TAC was kept for a minimum of three years for

Table 3
Total predicted theoretical by-catch (kg), total estimated non-landed by-catch(kg), and percentage of non-landed by-catch(%) across all analyzed records from the commercial fishery data. A 95% confidence interval is indicated in parentheses.

Species	Walleye	Yellow perch	White perch
Total analyzed records[a]	32,349	32,282	32,392
Total predicted theoretical by-catch ($\times 10^6$ kg)	1.29 (1.14 to 1.46)	0.042 (0.038 to 0.047)	1.31 (0.92 to 1.74)
Total landed by-catch ($\times 10^6$ kg)	0.54	0.008	0.60
Total estimated non-landed by-catch ($\times 10^6$ kg)	0.75 (0.60 to 0.92)	0.035 (0.030 to 0.039)	0.71 (0.31 to 1.14)
Total landed catch ($\times 10^6$ kg)	5.61	4.32	0.24
Percentage of non-landed by-catch[b] (%)	10.89 (8.86 to 13.04)	0.79 (0.70 to 0.90)	44.61 (27.24 to 57.48)

[a] In the commercial fishery data during 1994–2001, each record represents one net; in the commercial data during 2002–2007, each record represents a daily report which includes 1–5 nets.

[b] Percentage of non-landed by-catch(%) = total estimated non-landed by-catch of the species of interest/(total predicted theoretical by-catch of the species of interest + total landed catch of the species of interest).

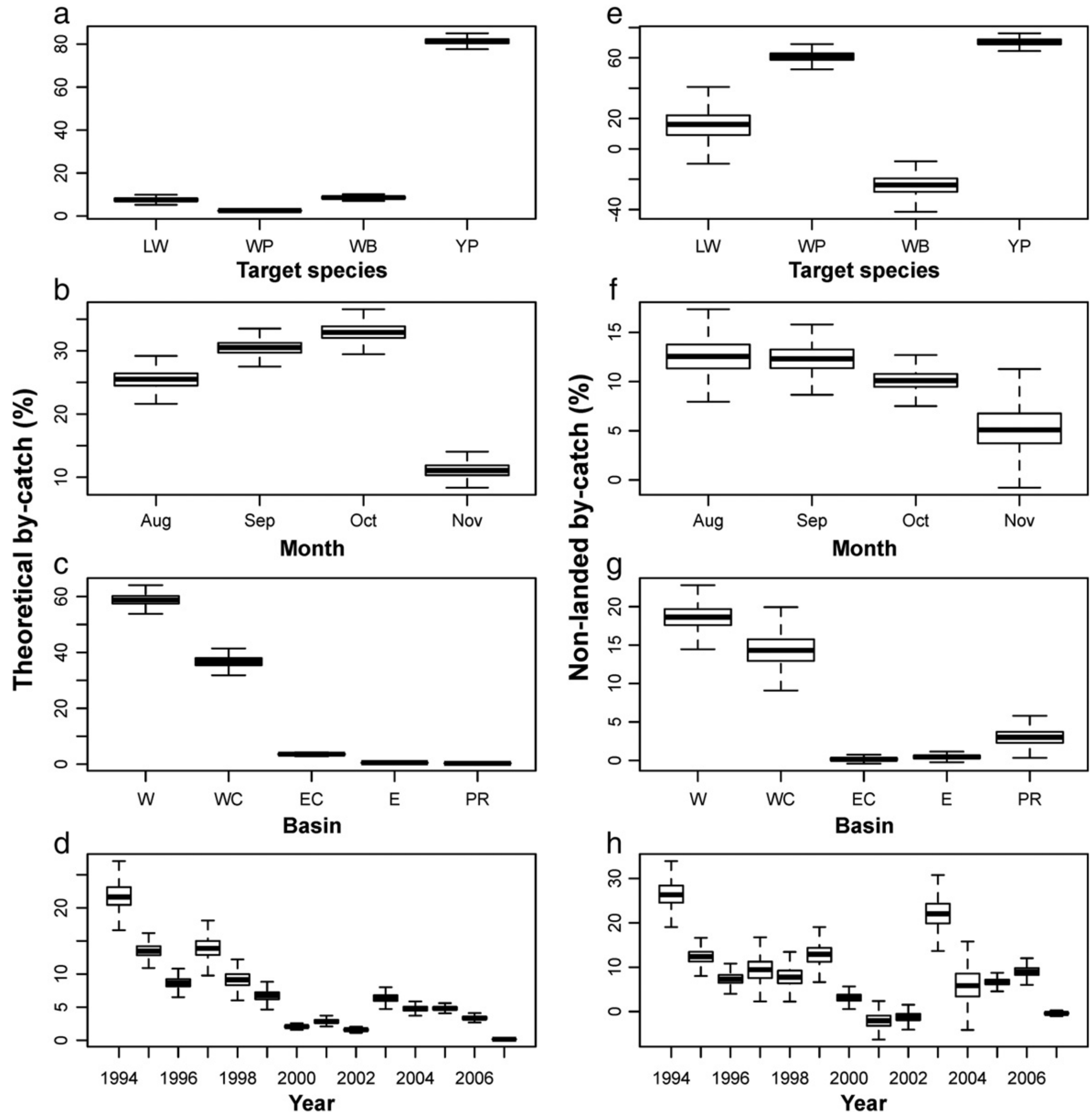

Fig. 2. Panels a–d: percentage composition of theoretical by-catch of walleye by major target species (a), month (b), basin (c) and year (d) from the Lake Erie commercial gillnet fisheries. Panels e–h: percentage of non-landed by-catch of walleye by major target species (e), month (f), basin (g) and year (h) from the Lake Erie commercial gillnet fisheries. LW — lake whitefish, WP — white perch, WB — white bass, YP — yellow perch; Aug — August, Sep — September, Oct — October, Nov — November; W — west basin, WC — west central basin, EC — East central basin, E — east basin, and PR — Pennsylvania Ridge. In each box plots, the dark line in the middle of the box represents the median, the lower and upper boundaries of the box represent the first and the third quartiles of the data, and the lower and upper bars represent the range of the data.

walleye. Low walleye abundance since 2000 may reduce the chance of by-catching walleye, which explained the lower theoretical by-catch of walleye in the years after 2000 (Fig. 2d). High abundance of yellow perch during 2001–2005 may increase the chance of by-catching yellow perch, which may lead to the higher theoretical by-catch of yellow perch in the years from 2001 to 2005 (Fig. 3d). We observed a high level of non-landed walleye by-catch in 2003 (Fig. 2h) because walleye abundance rebounded since 2003 but the TAC for walleye remained at a low level according to the Strategy. We obtained a high level of non-landed yellow perch by-catch in 2001 (Fig. 3h), which can likely be attributed to the high abundance of yellow perch and reduced TAC for yellow perch in 2001 according to the Strategy.

Methods to reduce by-catch and non-landed by-catch in the commercial fisheries in the Great Lakes have been employed, including reducing effort, modifying fishing gear, and using incentives and penalties in quota system (Johnson et al., 2004). Minimum size restrictions and closures of a fishery during spawning season or in the permanent/seasonal refuges have been applied to keep a sustainable commercial fishery in the Great Lakes (Kinnunen, 2003). This analysis highlighted the areas for higher levels of landed and non-landed by-catch of these three species in the commercial gillnet fisheries. For example, higher theoretical by-catch of walleye may be obtained in the west basin in October, higher theoretical by-catch of yellow perch in the west central basin in November, and higher theoretical by-catch of white perch in the west central basin in October (Figs. 2–4). Higher levels of non-landed by-catch of walleye may occur in the west basin of Lake Erie during August to September, higher levels of non-landed by-catch of yellow perch in the west central and east central basins in November, and higher levels of non-landed by-catch of white perch in the west basin in August and November (Figs. 2–4). If shown to be ecologically necessary, restricted fishing seasons and/or fishing locations could be applied in these areas to reduce the probability of observing non-landed by-catch, particularly for walleye and yellow perch.

Analyses on percentage of non-landed by-catch by major target species in this study may provide information for by-catch

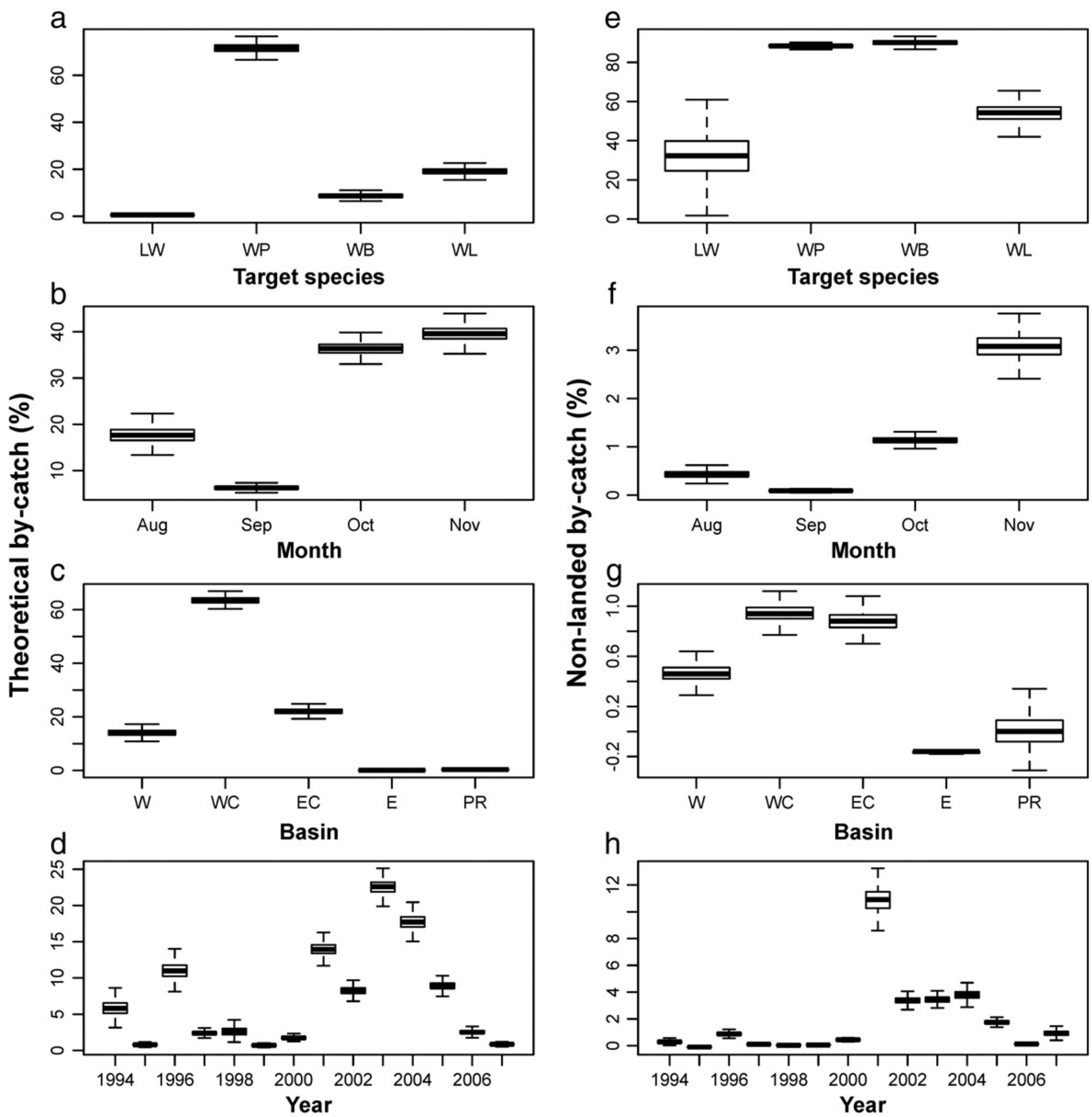

Fig. 3. Panels a–d: percentage composition of theoretical by-catch of yellow perch by major target species (a), month (b), basin (c) and year (d) from the Lake Erie commercial gillnet fisheries. Panels e–h: percentage of non-landed by-catch of yellow perch by major target species (e), month (f), basin (g) and year (h) from the Lake Erie commercial gillnet fisheries. See Fig. 2 for the explanation for abbreviations in the figure.

management. For instance, when targeting yellow perch, higher landed and non-landed by-catch of walleye (Figs. 2a and e) and white perch (Figs. 4a and e) may be observed, and targeting white perch (Figs. 3a and e) or white bass (Fig. 3e) may yield higher landed and non-landed by-catch of yellow perch. The association between target species and by-catch species in fisheries has been considered for by-catch management and a joint license framework has been currently applied in the ITQ system of the Lake Erie commercial gillnet fisheries. Quota allocation for joint-licensed species in this framework can be quantified by analyzing the ratio of target catch to by-catch, which could become an extension of this study.

The mandatory recording of all releases and discards has become a fully enforced condition for commercial fishing license on Lake Erie starting in 2010 (J. Johnson, Lake Erie Management Unit, Ontario Ministry of Natural Resources, personal communication). Nevertheless, advanced fishery data recording systems and observer programs, which can report both by-catch and release/discard information in the commercial fisheries, could provide detailed information for more accurate analyses. For example, with additional release/discard information recorded in the commercial fishery data, a model to estimate by-catch and non-landed by-catch can be developed directly based on the commercial fishery data instead of the PIS data.

The importance of fishery-independent surveys was emphasized in this study. Without the information from the Lake Erie Partnership Index Fishing Survey, it would have been difficult to assess landed and non-landed by-catch that actually occurred because this information had not been recorded in the commercial fishery data during our study period. Limited information about landed and non-landed by-catch on site is a problem commonly seen in by-catch analyses of the commercial fisheries. Even if the by-catch information was available from the commercial fishery data, additional information from a fishery-independent survey can help to calibrate the by-catch and discard estimations that are conducted directly based on the commercial fishery data. Thus, it is worthwhile to continue the fishery-independent surveys, such as PIS, for key species.

When analyzing the target species-specific, temporal and spatial variations in percentage of non-landed by-catch, especially when analyzing percentage of non-landed by-catch by major target species, we obtained negative median values in some circumstances. For

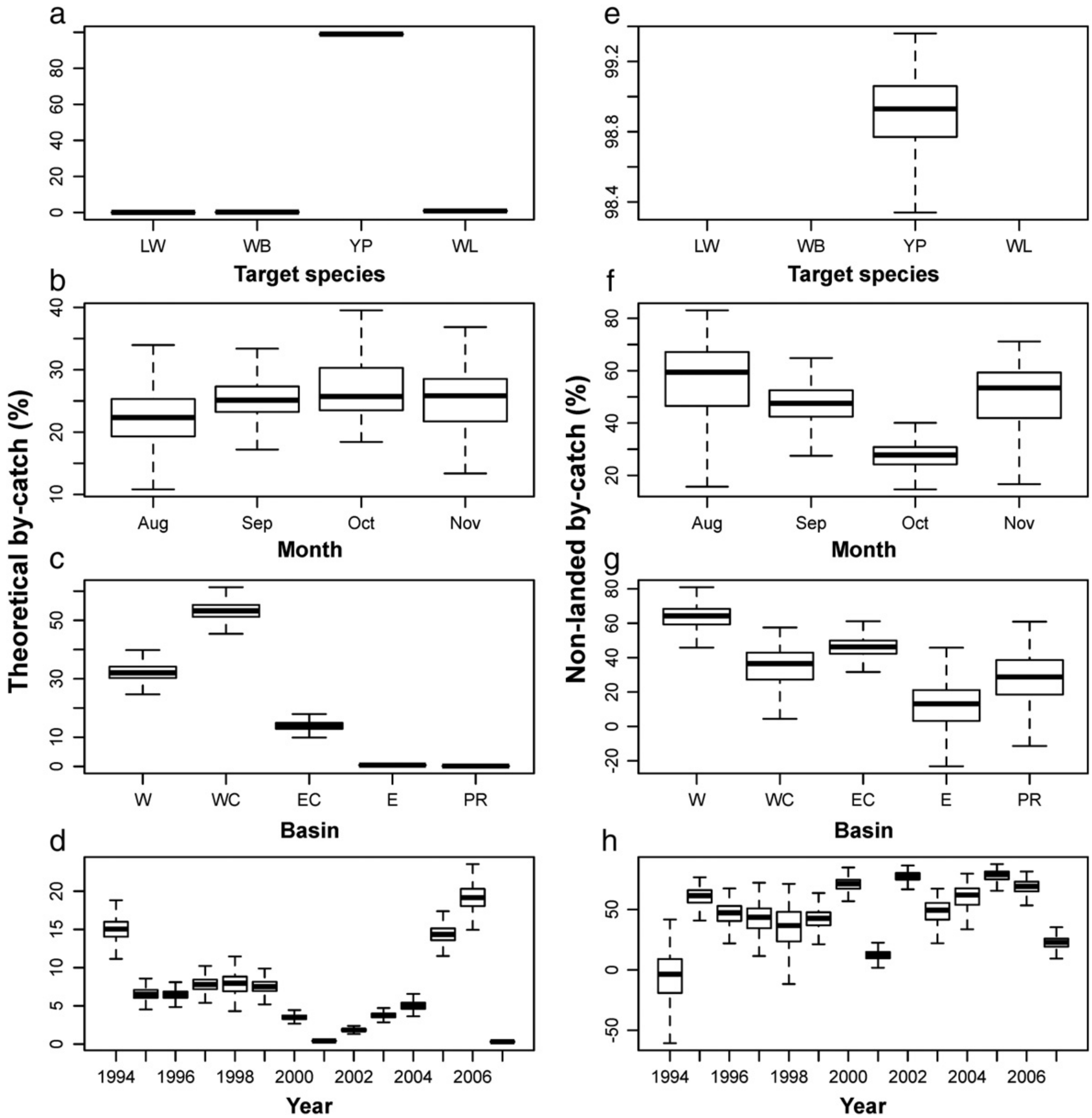

Fig. 4. Panels a–d: percentage composition of theoretical by-catch of white perch by major target species (a), month (b), basin (c) and year (d) from the Lake Erie commercial gillnet fisheries. Panels e–h: percentage of non-landed by-catch of white perch by major target species (e), month (f), basin (g) and year (h) from the Lake Erie commercial gillnet fisheries. See Fig. 2 for the explanation for abbreviations in the figure.

example, we had a negative median value for the percentage of non-landed walleye when targeting white bass (Fig. 2e). The negative median values happened when theoretical by-catch predicted for most records targeting a certain species was smaller than landed by-catch recorded in the commercial fishery data. Any of the differences between the PIS survey and the commercial fisheries that were not considered in the model developed on the PIS data could be a possible reason for the negative median values, such as net size (i.e., net length and net depth), gear set depth, fishing time, fishing location within each basin, and so on. In the PIS survey, we used the experimental gillnets which had the unified net length and net depth (30.5 m long × 1.8 m deep) whereas in the commercial fisheries, fishing gillnets had a variety of sizes; in the PIS survey, the Ontario water of Lake Erie was stratified by water depth, and the gillnets were distributed randomly within each stratus, whereas in the commercial fisheries, fishermen would go to those hot spots for fishing where the target species were aggregated and the likelihood of catching more fish was higher. Although we included basin and water depth as predictor variables in this analysis, basin represents a big range of area in terms of fishing location, i.e., there might be variable fishing locations within each basin.

The numerical values for the percentage of non-landed by-catch by major target species for walleye (Fig. 2e) and yellow perch (Fig. 3e) were much higher than those by month, basin or year. Higher values of percentage of non-landed by-catch were obtained because we assumed each trip targeted a single species at a time, i.e., when analyzing data by target species, because we targeted a certain species excluding walleye or yellow perch, the total landed catch of walleye or yellow perch was zero and we had the percentage of non-landed by-catch of walleye or yellow perch calculated as the total estimated non-landed by-catch divided by the total theoretical by-catch. By contrast, when analyzing data by month, basin or year, the total landed catch of walleye or yellow perch was a positive value for a given month, basin or year and we had the percentage of non-landed by-catch of walleye or yellow perch calculated as the total estimated non-landed by-catch divided by the sum of the total landed catch and total theoretical by-catch. We did not present the results of the percentage of non-landed by-catch of white perch when targeting lake whitefish, white bass or walleye (Fig. 4e) because the theoretical by-catch of white perch predicted from the Delta-AdaBoost model was very low (Fig. 4a) and we would not expect any considerable non-landed by-catch of white perch given such low theoretical by-catches.

In conclusion, assessment of landed and non-landed by-catch of walleye, yellow perch and white perch conducted in this study has important implications for by-catch management related to these three species in the Lake Erie commercial gillnet fisheries. The successful application of the AdaBoost model combined with a delta model indicated that the Delta-AdaBoost model can be considered as a candidate model when a high proportion of zero observations are included in the fishery data analysis.

Acknowledgments

This research was supported by grants from Virginia Polytechnic Institute and State University, the USDA Cooperative State Research, Education and Extension Service, Hatch Project #0210510, and the Ontario Commercial Fisheries' Association to Y. Jiao.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723.

Alverson, D., Freeberg, M., Pope, J., Murawski, S., 1994. A Global Assessment of Fisheries Bycatch and Discards. Food and Agriculture Organization of the United Nations, Rome.

Borges, L., Rogan, E., Officer, R., 2005. Discarding by the demersal fishery in the waters around Ireland. Fish. Res. 76, 1–13.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth Publishing Company, Belmont, CA.

Burnham, K., Anderson, D., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach2nd edition. Springer Verlag, New York.

Crowder, L., Murawski, S., 1998. Fisheries bycatch: implications for management. Fisheries 23, 8–17.

Damalas, D., Megalofonou, P., Apostolopoulou, M., 2007. Environmental, spatial, temporal and operational effects on swordfish (*Xiphias gladius*) catch rates of eastern Mediterranean Sea longline fisheries. Fish. Res. 84, 233–246.

Freund, Y., Schapire, R., 1996. Experiments with a new boosting algorithm. In: Saitta, L. (Ed.), Machine Learning: Proceedings of the Thirteenth International Conference. Citeseer, Bari, Italy.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. Ann. Stat. 28, 337–374.

Hall, M., Alverson, D., Metuzals, K., 2000. By-catch: problems and solutions. Mar. Pollut. Bull. 41, 204–219.

Hamley, J., 1975. Review of gillnet selectivity. J. Fish. Res. Board Can. 32, 1944–1969.

Harrington, J., Myers, R., Rosenberg, A., 2005. Wasted fishery resources: discarded by-catch in the USA. Fish Fish. 6, 350–361.

Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. Chapman and Hall, London, UK.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd ed. Springer, New York.

Jester, D., 1977. Effects of color, mesh size, fishing in seasonal concentrations, and baiting on catch rates of fishes in gill nets. Trans. Am. Fish. Soc. 106, 43–56.

Johnson, J.E., Jonas, J.L., Peck, J.W., 2004. Management of Commercial Fisheries Bycatch, with Emphasis on Lake Trout Fisheries of the Upper Great Lakes. Fish Res Rep. Michigan Department of Natural Resources, Lansing, MI.

Kawakita, M., Minami, M., Eguchi, S., Lennert-Cody, C., 2005. An introduction to the predictive technique AdaBoost with a comparison to generalized additive models. Fish. Res. 76, 328–343.

Kelleher, K., 2005. Discards in the World's Marine Fisheries: An Update. Food and Agriculture Organization of United States, Rome.

Kinnunen, R., 2003. Great lakes commercial fisheries. Report from Michigan Sea Grant. Michigan Sea Grant Extension, East Lansing, MI.

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., Punt, A., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141–159.

Murawski, S., 1996. Factors influencing by-catch and discard rates: analyses from multispecies/multifishery sea sampling. J. Northwest Atl. Fish. Sci. 19, 31–40.

Murray, K., 2004. Magnitude and distribution of sea turtle bycatch in the sea scallop (*Placopecten magellanicus*) dredge fishery in two areas of the Northwestern Atlantic Ocean, 2001–2002. Fish. Bull. 102, 671–681.

Ontario Commercial Fisheries' Association (OCFA), 2007. Lake Erie Partnership Index Fishing Survey. http://www.ocfa.on.ca/LakeEriePartnershipIndexFishingSurvey.htm. 2007(September 2009).

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.

Parrish, D., Margraf, F., 1990. Interactions between white perch (*Morone americana*) and yellow perch (*Perca flavescens*) in Lake Erie as determined from feeding and growth. Can. J. Fish. Aquat. Sci. 47, 1779–1787.

Pennington, M., 1996. Estimating the mean and variance from highly skewed marine data. Fish. Bull. 94, 498–505.

Punt, A., Smith, D., Tuck, G., Methot, R., 2006. Including discard data in fisheries stock assessments: two case studies from south-eastern Australia. Fish. Res. 79, 239–250.

Schaeffer, J., Margraf, F., 1987. Predation on fish eggs by white perch, *Morone americana*, in western Lake Erie. Environ. Biol. Fish. 18, 77–80.

Scott, W.B., Crossman, E.J., 1973. Freshwater fishes of Canada. Bulletin, 184. Fisheries Research Board of Canada, Ottawa, Ontario.

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., 2005. Status of yellow perch and walleye populations in Michigan waters of Lake Erie, 1999–2003. Fish Res Rep, 2082. Michigan Department of Natural Resources, Lansing, MI.

Tweedie, M., 1984. An index which distinguishes between some important exponential families. Statistics: applications and new directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Indian Statistical Institute, Calcutta.

WTG (Lake Erie Walleye Task Group), 2010. Lake Erie Walleye Task Group Annual Report. . Available: http://www.glfc.org/lakecom/lec/WTG.htm. (July 2010).

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.

YPTG (Lake Erie Yellow Perch Task Group), 2010. Lake Erie Yellow Perch Task Group Annual Report. Available: http://www.glfc.org/lakecom/lec/YPTG.htm. (July 2010).