Journal of Hydrology 536 (2016) 447–456

Contents lists available at ScienceDirect

Journal of Hydrology

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

Assessment of Chinese sturgeon habitat suitability in the Yangtze River (China): Comparison of generalized additive model, data-driven fuzzy logic model, and preference curve model

Yujun Yi [a,b,*], Jie Sun [a], Shanghong Zhang [c], Zhifeng Yang [a,b]

[a] State Key Laboratory of Water Environment Simulation and Pollution Control, School of Environment, Beijing Normal University, Beijing 100875, China
[b] Ministry of Education Key Laboratory of Water and Sediment Sciences, Beijing Normal University, Beijing 100875, China
[c] Renewable Energy School, North China Electric Power University, Beijing 102206, China

ARTICLE INFO

Article history:
Received 8 January 2016
Received in revised form 2 March 2016
Accepted 3 March 2016
Available online 9 March 2016
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Habitat
Preference curves
Fuzzy logic
GAM
Accuracy
Chinese sturgeon

SUMMARY

To date, a wide range of models have been applied to evaluate aquatic habitat suitability. In this study, three models, including the expert knowledge-based preference curve model (PCM), data-driven fuzzy logic model (DDFL), and generalized additive model (GAM), are used on a common data set to compare their effectiveness and accuracy. The true skill statistic (TSS) and the area under the receiver operating characteristics curve (AUC) are used to evaluate the accuracy of the three models. The results indicate that the two data-based methods (DDFL and GAM) yield better accuracy than the expert knowledge-based PCM, and the GAM yields the best accuracy. There are minor differences in the suitable ranges of the physical habitat variables obtained from the three models. The hydraulic habitat suitability index (HHSI) calculated by the PCM is the largest, followed by the DDFL and then the GAM. The results illustrate that data-based models can describe habitat suitability more objectively and accurately when there are sufficient data. When field data are lacking, combining expertise with data-based models is recommended. When field data are difficult to obtain, an expert knowledge-based model can be used as a replacement for the data-based methods.

1. Introduction

In recent decades, aquatic habitat modelling has gained increasing attention due to its relevance to river management, restoration, and conservation biology (Mouton et al., 2012). A series of aquatic habitat suitability evaluation models have been proposed in recent years. For example, the preference curve model (PCM); fuzzy logic; multivariate statistical models, such as the generalized linear model (GLM) and generalized additive model (GAM); artificial neural networks; and statistical analysis of multiple species and communities (Adriaenssens et al., 2004; Jowett and Davey, 2007; Yi et al., 2010b, 2014). Selecting appropriate models is important because inappropriate habitat suitability model may result in erroneous predictions (Lin et al., 2015). Moreover, blindly making decisions based on the result of a single model may result in suboptimal management planning decisions (Gallo and Goodchild, 2012). However, no consensus has been reached on the optimal model (Segurado and Araujo, 2004). Therefore, there is a need to compare the advantages and disadvantages of different models in a particular context.

Comparing different models under a given condition is beneficial for understanding the characteristics of each model and helpful for selecting the most suitable model (Ahmadi-Nedushan et al., 2006). A few studies have applied more than two habitat suitability evaluation models to the same aquatic species data and compared their performance. For example, the comparison of GAM and boosted regression trees for wahoo (Martínez-Rincón et al.,

Abbreviations: PCM, preference curve model; DDFL, data-driven fuzzy logic model; GAM, generalized additive model; GLM, generalized linear model; HSI, habitat suitability index, the comprehensive suitability considering several variables; HHSI, hydraulic habitat suitability index, an index to describe the integrated habitat suitability of the entire area; $CPUE_d$, catch per unit effort; TGR, Three Gorges Reservoir; SI, suitability index for single variable; TSS, true skill statistic; AUC, area under the receiver operating characteristics curve; CCI, overall accuracy index; AD, average deviation; TPR, true positive rate; TNR, true negative rate; TP, true positives; FP, false positives; FN, false negatives; TN, true negatives; MST, maximization of the sensitivity-specificity sum threshold.

* Corresponding author at: State Key Laboratory of Water Environment Simulation and Pollution Control, School of Environment, Beijing Normal University, Beijing 100875, China. Tel.: +86 18810628449; fax: +86 10 58801757.

E-mail address: yiyujun@bnu.edu.cn (Y. Yi).

http://dx.doi.org/10.1016/j.jhydrol.2016.03.005
0022-1694/

2012); fuzzy logic habitat models, univariate suitability curves, random forests, support vector machine, GAM, and ANN for brown trout (Mouton et al., 2011; Muñoz-Mas et al., 2012, 2016); data-driven species distribution models (i.e., artificial neural networks, classification and regression trees, fuzzy habitat suitability models, GAMs, GLMs, random forests and support vector machines) for spawning European grayling (Fukuda et al., 2013); PCM, expert knowledge-based and data-driven fuzzy logic models for Chinese sturgeon (Yi et al., 2014). All of these studies have provided guidelines for the selection of habitat suitability evaluation models. However, comprehensive comparative studies about using different models to evaluate aquatic habitat on a common data set (such as Fukuda et al., 2013; Muñoz-Mas et al., 2016) are limited.

A comparison of the PCM, the data-driven fuzzy logic model (DDFL), and the GAM is presented in this paper. The PCM is a conventional habitat suitability evaluation model. It determines suitability indices for different physical habitat variables (e.g., cover, substrate size, velocity, and depth). These indices are then combined to define a composite index. Arithmetic or geometric mean algorithm, product algorithm, lowest algorithm, weighted summation algorithm and weighted product algorithm are the commonly used combined algorithms. The multiplication and lowest indices are relatively conservative and are often applied in conservation planning, while the average methods are often used in the estimation of the resource amount and fishery studies (Gong et al., 2011). Although PCM has been criticized because it does not account for interactions among the habitat variables (Ahmadi-Nedushan et al., 2006; Jowett, 2003; Leclerc et al., 2003), it still is convenient analytical tool in many applications.

The DDFL was developed based on the expert knowledge-based fuzzy logic model. The fuzzy sets and fuzzy rules of DDFL are optimized based on data. This avoids subjectivity in defining fuzzy sets and fuzzy rules and emphasizes objective data compared with the knowledge-based fuzzy logic models.

The GAM, which was proposed by Hastie and Tibshirani (1990), has been widely used for habitat evaluation (Drexler and Ainsworth, 2013; Forney et al., 2012; Tao et al., 2012). The GAM is a non-parametric extension of GLMs. A GAM uses a link function to establish the relation between the mean of the response variable and a smoothed function of the predictor variables.

The PCM is an expert knowledge-based model. The DDFL and GAM are both data-based and the DDFL is an artificial intelligence model while the GAM is a statistical model (Adriaenssens et al., 2004; Fukuda et al., 2013). Although these three models have been applied to different aquatic animals to evaluate habitat suitability (Bellido et al., 2001; Mouton et al., 2009; Yi et al., 2010a), no one has, to our knowledge, used the three models on a common data set and compared their performance. The study reported here simultaneously applies the three models to assess habitat suitability for the Chinese sturgeon. The data are derived from the results of a total of 72 observations measured from 1996 to 2003. The PCM was constructed based on the research results of experiments and field surveys of biologists; the DDFL was developed using MATLAB software; and the GAM was developed using R software. The preference of the physical habitat variables for the Chinese sturgeon is studied. The three models were compared in terms of accuracy, suitable ranges of physical habitat variables (water depth and velocity), habitat maps, and hydraulic habitat suitability index (HHSI). The true skill statistic (TSS) and the area under the receiver operating characteristics curve (AUC) were used to evaluate the method's accuracy. The three models performed differently particularly with respect to accuracy and the habitat maps. The results show that the GAM outperforms the PCM and the DDFL for both the TSS and AUC. The mechanisms leading to these differences between models are discussed. The study compares the properties and effectiveness of the three models. The results can provide a framework for the choice of a habitat suitability evaluation model, and also provide a basis for habitat protection and reconstruction for the endangered Chinese sturgeon.

2. Species and study area

The Chinese sturgeon is one of the world's largest anadromous fish. Historically, it was widely distributed along China's coast and within the contiguous estuaries of large rivers. Today, the Yangtze River is the only remaining river with populations of the Chinese sturgeon (Chen, 2007; Wei et al., 1997; Yangtze Aquatic Resources Survey Group, 1988). The Yangtze is the world's third longest river and China's largest and longest river. It extends 6300 km with a basin of 1.8 million square kilometres. It flows from west to east and eventually empties into the East China Sea. Due to the impact of habitat loss, environmental pollution, overfishing, shipping and other factors, coupled with a low reproductive rate and delayed maturity, the Chinese sturgeon population in the Yangtze River has declined in recent decades (World Wide Fund for Nature, 2014). Joint monitoring by the Yangtze River Fisheries Research Institute, the Institute of Hydrobiology, and the Institute of Hydroecology reflects that no fertilized egg has been detected in the spawning grounds since 2013 (Wang, 2015). The wild Chinese sturgeon population is in danger of becoming extinct. It was listed as critically endangered in 2010 (International Union for Conservation of Nature Red List of threatened species). There is an urgent need for effective measures that strengthen the protection of this species. In addition to research on artificial reproduction, research regarding suitable spawning grounds is also very important; this would facilitate protection and reconstruction of the spawning grounds.

Prior to the construction of the Gezhouba Dam at Yichang in 1981, the historical spawning grounds covered approximately 600 km (Yangtze Aquatic Resources Survey Group, 1988). After damming, the migration route for Chinese sturgeon spawning was cut off. The area of the spawning grounds was reduced sharply, and a reach of 4.8 km below the dam became the new primary spawning area of the Chinese sturgeon (Institute of Hydrobiology, 2005). For this reason, the 5 km area below the Gezhouba Dam was chosen as the study area (Fig. 1).

3. Methods

3.1. Data sources and variable selection

The field sampling was conducted by Wei (2003) from 1996 to 2003 covering the reach from Gezhouba Dam to Miaozui. A D-shaped bottom drift net was used to capture sturgeon eggs and yolk-sac larvae from the river bottom (Wei, 2003). Most of the eggs and yolk-sac larvae were captured in the center of the river and inside the river bends. The stomachs of opportunistic predatory fishes in this area were also examined. The density of eggs was measured as the quantity of egg catch per unit of effort ($CPUE_d$). The range of the sampled $CPUE_d$ data was wide (the minimum value is 0 (no eggs collected) and the maximum is 1272.3); therefore, the $CPUE_d$ data were normalized to the range between 0 and 1 to make calculations convenient. There is much literature available on the factors that control habitat suitability for fish (Mosley, 1983). Factors related to the spawning activities of Chinese sturgeon include substrate, channel structure, area of the spawning ground, water temperature, sand concentration, water level, and so on (Wei, 2003). The Chinese sturgeon will breed when all of the hydrological parameters are within the suitable range (Yang et al., 2007). However, there is no agreement regarding the exact effect of these hydrological factors on spawning. Ban et al. (2009)