Fisheries Research 195 (2017) 169–179

Contents lists available at ScienceDirect

Fisheries Research

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

Why do lake whitefish move long distances in Lake Huron? Bayesian variable selection of factors explaining fish movement distance

Yang Li[a,*], James R. Bence[a], Zhen Zhang[b], Mark P. Ebener[c]

[a] Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA
[b] Enabling Capabilities Technology Center, Dow AgroSciences, Indianapolis, IN 46268, USA
[c] Inter-Tribal Fisheries and Assessment Program, Chippewa Ottawa Resource Authority, Sault Ste. Marie, MI 49783, USA

ARTICLE INFO

Handled by A.E. Punt

Keywords:
Fish movement
Variable selection
Tag recovery
Lake whitefish
Bayesian

ABSTRACT

Understanding fish movement patterns is vital for stock assessment and fishery management. We used a variable selection procedure in a Bayesian framework to understand what factors most likely affect the net movement distance of individual fish based on a conventional tag-recovery study of lake whitefish populations in Lake Huron during 2003–2011, where fish of this species with spawning site fidelity were tagged during the spawning season and recovered throughout the year. We found that fish with greater total length, and those that were tagged and released from tagging sites near Cheboygan and Alpena, Michigan, moved longer net distances than fish from other tagging sites. Habitat conditions also had a profound effect on net movement distance. We found that shorter movement distances by lake whitefish can be expected if the relative density of the benthic amphipod *Diporeia* spp. was higher near the tagging site during the recovery year. We also found evidence that lake whitefish may start their annual spawning migration runs earlier during warmer years. More generally, our Bayesian framework for analysis of conventional tagging data has potential for wide applicability, and model details and our code are provided to facilitate this.

1. Introduction

Many fish species move for long distances at various times during their life cycle, and movements made by individuals vary from regular and predictable migration to less-predictable resource driven nomadism (Runge et al., 2014). Most previous research that evaluated changes in fish spatial locations focused on either the triggering factors or distance between initial and final fish location (e.g., Albanese et al., 2004; Radinger and Wolter, 2014), or on estimating net movement/migration rates of populations (Polacheck et al., 2006; Vandergoot and Brenden, 2014).

Fish movement is essential from both conservation and management perspectives. Movement behavior can influence how fish are distributed, whether their populations persist in the face of ecosystem changes, and how stocks are assessed. Fish movement can further influence ecological interactions and evolution (Lidicker and Stenseth, 1992). Management problems such as inaccurate assessment results, or inappropriate catch limits, can occur when actual fish movements do not agree with the spatial assumptions made in stock assessments and management decisions, which can result in local population depletion and population collapse (Fu and Fanning, 2004; Hutchings, 1996; Li et al., 2015; Mitchell and Beauchamp, 1988; Rothschild, 2007).

Despite its ecological and management importance, understanding of fish movement patterns in time and space, and how movements are related to environmental variables, is still limited. Moreover, most previous research that focused on the triggering factors (i.e., factors causing the initiation of movement) and net fish movement distance were limited to stream fish, given the easy calculation of net distance moved from conventional tagging data. Much less is known about movement of fish that live in large water areas. Most of which is known has been derived from electronic tagging data, although there are many long-term conventional tagging programs. While technological advances make the use of acoustic or pop-up tags increasingly useful, conventional tags are still more widely used for estimating population size, mortality, and tracking individual growth, given their lower price. Conventional tagging data can also provide information on the location at tag release and tag recovery, which could be used for the estimation of movement route and intensity (e.g., net fish movement distance) (e.g., Albanese et al., 2004; Gilliam and Fraser, 2001).

The goal of this study was to develop a model framework for analysis of how factors impact the distance fish move from when they are tagged until they are recovered ('net fish movement distance' hereafter) in a larger water body, based on conventional tag-recovery results. We

* Corresponding author.
E-mail addresses: liyang11@msu.edu, yangyang1008li@gmail.com (Y. Li).

http://dx.doi.org/10.1016/j.fishres.2017.07.014
Received 12 April 2017; Received in revised form 12 July 2017; Accepted 13 July 2017

based our research on several lake whitefish (*Coregonus clupeaformis)* spawning stocks in Lake Huron of the Laurentian Great Lakes of North America. As an ecological and economically important fish species in the Great Lakes, lake whitefish have been found to move freely among multiple management units during the non-spawning period, but show a high degree of natal homing, so nearly all mature fish return to spawn at the same location each year (Ebener et al., 2010b). Previous research on lake whitefish movement patterns provides a useful platform for us to derive a priori hypotheses about the potential factors that influence movement. Since the establishment of dreissenid mussels in the early 1990s, the ecosystem of four of the five Great Lakes have changed substantially, including an overall decrease in the density of lake whitefish's preferred food- *Diporeia* spp. (Barbiero et al., 2011; McNickle et al., 2006; Mohr and Nalepa, 2005). In this context, Rennie et al. (2012) evaluated the relationship between lake whitefish migration distance and growth rate, and found that the least mobile population of lake whitefish was supported by a remnant *Diporeia* spp. population. Ebener et al. (2010b) found that stock identity and season of recapture affected net movement distance most strongly, while the influence of variables such as sex, year, fish total length, and time at large was weaker. Although the role of temperature has not been directly implicated in explaining patterns in the fish movement, the association between lake whitefish harvest and surface water temperature suggested that such a connection may exist (Price et al., 2003).

The pioneering studies of net movement distance used either a regression-tree based approach or ANOVA models to test whether net movement distance varied significantly in association with the factors they evaluated (e.g., Albanese et al., 2004; Ebener et al., 2010b; Radinger and Wolter, 2014). Because some studies estimated the effects of different factors as additive (i.e., causing a given distance change rather than a percentage change in net movement distances), it is hard to generalize the results from studies with different spatial and temporal scales. When jointly considering multiple factors and continuous covariates, the ANOVA approach can provide only a rough picture of the continuous relationship between net movement distance and explanatory factors. Thus, a more thorough regression analysis is needed. The regression-tree based approach seeks to approximate nonlinearity and interactions in the relationships between the net movement distances and multiple factors by recursively partitioning the data points according to the categorization of the factors (Ebener et al., 2010b). Such partitioning may have difficulty in interpreting the effects, if the observations from the same tag or recovery area happen to be separated into different branches of the tree. Some regression-tree applications have partitioned data by site (i.e., different sites on different branches), and this can make it difficult to develop a general understanding of movement (Ebener et al., 2010b). In addition, although it is possible for regression-tree based approaches to rank or select variables based on variable importance measures, they do not provide any further insight of the uncertainty associated with their rankings or selections. Also information criteria, such as Akaike's information criterion and the Bayesian information criterion, commonly used as penalization terms for the number of parameters in model, are not applicable for nonparametric tree-based models (Claeskens and Hjort, 2008).

We therefore considered a global linear regression model that accounts for joint effects of multiple factors and the heterogeneity among sites, to study the relationship between the net movement distance and individual factors. We further conducted a variable selection procedure under a Bayesian framework to explore the plausibility of alternative regression models that include various explanatory variables, and assess the associated uncertainty. Bayesian variable selection treats the regression model itself as random among all possible models with different sets of variables. Thus, it accounts for model uncertainty in the overall assessment of uncertainty by making inferences on how probable alternative models are after consideration of the data. The implementation of Bayesian variable selection via the reversible jump Markov chain Monte Carlo (rjMCMC) (Green, 1995) procedure is substantially more efficient in exploring the model space than the traditional approaches such as all-subsets-regression (Woznicki et al., 2016). While we believe our approach has substantial advantages over regression-tree approaches, it could miss some nonlinear effects that could be identified by regression-trees. Thus, as a check on robustness we compared our results with those from regression-tree methods.

We considered how net distance moved from tagging to recapture locations changed monthly and over years, and how this net movement pattern depended upon tagging location. In addition, we considered how life history traits, namely total length, and sex, and habitat features, namely *Diporeia* spp. density and water temperature, played a role in these net movement patterns. Thus, the variables we considered as potential explanatory factors in this study were tagging year, recovery year, recovery month, year(s) between tag and recovery, fish total length, sex, tagging (spawning) site, and the habitat variables based on *Diporeia* spp. density and growing degree days.

Our goal was to provide not only insight on how those factors influenced lake whitefish movement in Lake Huron, but also a model framework for analyzing movement mechanism based on conventional tagging data. Although Bayesian variable selection in linear regression is a long-established approach (Mitchell and Beauchamp, 1988), it was rarely used in ecology or more specifically for uncovering explanations for movements (Drouineau et al., 2017; Ethier et al., 2017). Drouineau et al. (2017) used a Bayesian state-space model to analyze the effects of different environmental factors in triggering migration of silver eel in fragmented rivers. Ethier et al. (2017) used Bayesian models and variable selection to evaluate how environmental variables influenced regional variation in population trends of Bobolink. Both studies used a mixture distribution of priors (i.e., normal plus zero-inflation), which were estimated using a Gibbs sampler. However, their variable selection procedure did not introduce a penalty such as BIC for increasing number of selected variables. Also the Gibbs sampler usually involves scanning all variables at each iteration, which could be computational expensive, especially when the number of candidate variables is large.

To the best of our knowledge, this study is the first to apply the Bayesian variable selection approach to compare the effects of various factors on fish net movement distance by introducing an explicit prior penalty on model complexity, and the most comprehensive to date in terms of the range of factors affecting whitefish movement. To avoid sampling all indicators within a Gibbs sampler circle as in Drouineau et al. (2017) and Ethier et al. (2017), we adopt the reversible jump MCMC algorithm for model exploration that mimics stepwise selection and subsets regression technique, which is more computationally efficient. Thus our research introduces an approach to fish movement studies, which has the potential to be much more effectively interrogate a large number of predictor variables. To facilitate usage of our approach, we provide the open-source code for MATLAB program which is online available at to implement the method.

2. Methods

2.1. Data collection, selection, and calculation of net-movement distance

Lake whitefish were tagged and released in a study coordinated by one of us (Mark P. Ebener) at 21 individual tagging sites from nine spawning stocks in Lake Huron from late October through December (i.e., spawning season) of 2003–2006. Total length (mm) of all 35,285 tagged fish were measured before release, spatial coordinates of the tagging and release location, and date of release were recorded for each fish. Lake whitefish were tagged on or very near the spawning grounds and subsequently killed when recovered by the commercial or recreational fishery. The commercial fishing season for lake whitefish is not closed in Ontario waters during the spawning season, but it is closed in Michigan waters. Thus, fish tagged and released at Detour, Cheboygan, Alpena, and Saginaw Bay (Fig. 1) were extant 1–4 weeks before being subjected to fishing and tag recovery. At Burnt Island, the Fishing Islands, and Sarnia fish were also tagged during the spawning season, but