Fisheries Research 147 (2013) 186–195

Contents lists available at SciVerse ScienceDirect

Fisheries Research

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

Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study

Hao Yu[a,*], Yan Jiao[a], Laurence W. Carstensen[b]

[a] Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[b] Department of Geography, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

ARTICLE INFO

Article history:
Received 9 April 2012
Received in revised form 10 June 2013
Accepted 10 June 2013

Keywords:
Relative abundance index
Generalized linear models
Generalized additive models
Spatial interpolation

ABSTRACT

Generalized linear models (GLMs) and generalized additive models (GAMs) are commonly used to standardize catch rates as relative abundance indices in fisheries stock assessments. Spatial interpolation (SI) is an alternative way to estimate relative abundance indices but there have been no comparisons of the effectiveness of the two types of approaches. In the present study, the performances of GLMs, GAMs and SI were compared through a simulation study based on fishery independent surveys of yellow perch in Lake Erie in 1990, 1991, 1992, 2000, 2001, and 2003. Simulated scenarios were tested with sample sizes of 60, 120 and 180 drawn randomly from the survey data, and random errors variances of 0.5, 1 and 2 × the "true" estimate variances. For each combination of sample size and error, 100 simulations were calculated to estimate correlation between the "true" abundance and the estimated relative abundance indices from GLMs, GAMs and SI. The performances of all three methods improved with increasing sample sizes, but worsened with increasing magnitude of the simulated errors. SI performed better than GLMs and GAMs when the simulated errors were low, but SI was more sensitive than GLMs and GAMs to the magnitude of the simulated random errors. When simulated sampling covered the survey area incompletely, GLMs and GAMs performed better than SI.

1. Introduction

Standardized catch rate is commonly used as an index of relative abundance in fisheries studies. Catch rates are related to (and in most cases proportional to) population abundance in stock assessment models (Quinn and Deriso, 1999; Maunder and Langley, 2004; Maunder and Punt, 2004), but without standardization they may not correctly reflect abundance variation and therefore lead to biased stock assessment results. Catch rate standardization removes the effects of all factors other than population abundance variation (Maunder and Punt, 2004). This has been applied to population dynamics models in numerous research efforts (Gavaris, 1980; Lo et al., 1992; Harley et al., 2001; Walsh and Kleiber, 2001; Bishop et al., 2004; Maunder and Punt, 2004; Shono, 2008). Generalized linear models (GLMs) and generalized additive models (GAMs) are commonly used to standardize catch rates (O'Brien and Mayo, 1988; Punt et al., 2000; Ye et al., 2001; Campbell, 2004; Nishida and Chen, 2004; Damalas et al., 2007). Spatial interpolation (SI) is another way to estimate relative abundance indices (Rivoirard et al., 2000). The efficiency comparison of SI and catch rate standardization methods is important but less studied than the application of GLMs or GAMs.

GLMs and GAMs have been used to estimate abundance indices with their own advantages and limitations. GLMs were first introduced in the 1970s (Nelder and Wedderburn, 1972) and have been used to standardize catch rates since the 1980s (Gavaris, 1980). GLMs assume a linear relationship between a link function (e.g., identity, logistic, or log) of the expected response variable and the explanatory variables (Maunder and Punt, 2004). GAMs are extensions of GLMs but replace the explanatory variables with smooth functions, and they are often used to deal with nonlinear relationships between the response variable and explanatory variables (Hastie and Tibshirani, 1990; Guisan et al., 2002). Nonlinear relationships are common between fish densities and environmental factors, and therefore GAMs are also widely used in catch rate standardization (Walsh and Kleiber, 2001; Denis et al., 2002). However, both GLMs and GAMs have disadvantages when standardizing catch rates, which include: (1) requirement of a number of explanatory variables, (2) error structure assumptions, (3) model selection uncertainties, (4) dealing with high percentage of zero catches, and

* Corresponding author. Current address: Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, 146 Williams Street, Solomons, MD 20688, USA. Tel.: +1 540 231 3599; fax: +1 540 231 7580.
E-mail address: haoyu@vt.edu (H. Yu).

0165-7836/$ – see front matter. Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2013.06.002

(5) dealing with interaction terms (Maunder and Punt, 2004). When there is a high proportion of zero observations, zero-inflated models or delta models are often used to standardize catch rates (Punt et al., 2000; Martin et al., 2005). When the residuals are spatially autocorrelated, spatial-GLMs are often applied (Nishida and Chen, 2004; Yu et al., 2011). However, these new models often need a large number of observations of environmental factors besides catch data.

GLMs and GAMs are used to account for the effects of other factors (e.g., environmental factors and spatial autocorrelation) on population abundances. However, SI, in contrast, uses spatial autocorrelation to estimate the values in un-sampled areas. Geographic information systems (GIS) are widely used to display and analyze spatial characteristics in fisheries data (Rahel, 2004). SI is one of the applications of GIS, and it has been applied in estimating aquatic species densities since the 1990s (Simard et al., 1992; Maynou et al., 1996; Rivoirard et al., 2000; Wyatt, 2003). The densities of most fish species within their distribution ranges are spatially correlated because the environmental factors are more similar when the distances are closer, and therefore SI can be applied for estimating abundance indices. Kriging is one commonly used SI method (Cressie, 1993; Schabenberger and Gotway, 2005). The fundamental idea of kriging is to estimate the value of some quantity at an unknown point by using the combination of weights and the values at known local points. Therefore, it only requires sampled data and the coordinates of spatial locations. However, the disadvantages of kriging are also apparent: (1) marked sensitivity to measurement errors; (2) requirement for the coordinates of survey locations that were used to estimate the spatial distances among observations. According to different assumptions, kriging can be divided into many types (e.g., simple kriging, ordinary kriging, universal kriging, etc.). In the present study, ordinary kriging is used because it is the most commonly used SI method (Schabenberger and Gotway, 2005).

In this study, yellow perch catch rate data were used as an example to compare the performance of GLMs, GAMs, and SI. Yellow perch (*Perca flavescens*) is one of the most important commercial and recreational fish species in Lake Erie (Baldwin and Saalfeld, 1962; Regier and Hartman, 1973). Yellow perch abundance varies dramatically over time, and the catch rate data are spatially autocorrelated (YPTG, 2008). There is not a generally accepted method to estimate yellow perch relative abundance indices in Lake Erie, and the arithmetic mean (AM) of catch rates is currently used (YPTG, 2008). Therefore in this study AM is also included in the performance comparison. This is the first study that compares the performance of AM, GLMs, GAMs, and SI together in estimating relative abundance indices.

2. Materials and methods

2.1. Study area and survey method

Data used in this study are from the fishery-independent surveys conducted by the Ontario Commercial Fisheries' Association and the Ministry of Natural Resources Lake Erie Fisheries Management Unit in 1991–1993, 2000, 2001, and 2003 within the Canadian side of Lake Erie. These surveys include catch data as well as the information on gear and environmental factors (Table 1). The study area can be divided into western, central, and eastern basins and each basin was divided into 2 × 2 min cells using the ArcGIS software package (version 9.2, 2007, ESRI Inc., USA). The sample size is about 120 in each year and the sampling design is stratified random sampling. The catch rates and environmental factors of the un-sampled cells were interpolated in each year.

Table 1
Summary of data collected in the 1990–2003 yellow perch fishery-independent survey in Lake Erie.

Variables	Unit	Remarks
Catch number	Individual	Per species
Longitude	°	Converted to NAD 1983 UTM 17 N
Latitude	°	Converted to NAD 1983 UTM 17 N
Set duration	h	Standing time of gillnet in water
Bottom depth	m	Per sampling site
Gear depth	m	Depth to bottom of the gillnet
Transparency	m	Secchi depth
Water temperature	°C	At surface
Gear temperature	°C	At gear depth
Dissolved oxygen (DO)	mg/L	At gear depth

2.2. Generalized linear model

A basic GLM can be written as

$$g(\mu) = X^T\beta \tag{1}$$

where g is the link function, μ is the expectation of the observation, X is the vector of explanatory variables, and β is the vector of the regression coefficients (Montgomery et al., 2006). The log-transformation has been widely used in fisheries and has been found to be appropriate in many situations (Quinn and Deriso, 1999). Since there were zero observations in the survey data, delta-lognormal GLMs were used to generate "true" abundance data for yellow perch in the Canadian side of Lake Erie. The general form of delta-lognormal GLMs can be written as:

$$Pr(Y = y) = \begin{cases} w, & y = 0 \\ (1 - w)f(y) & \text{otherwise} \end{cases} \tag{2}$$

where w is the probability of a zero observation, and $f(y)$ is the probability function of the lognormal distribution. In the simulation procedure, we treated μ as the expectation of the log-transformed observation of catch rates, and spatial-GLMs (hereafter GLMs) were used to standardize catch rates (Nishida and Chen, 2004; Yu et al., 2011). The residuals in GLMs were assumed spatially correlated, and the covariance, $Cov(\varepsilon_i, \varepsilon_j)$ is the function of the distance d_{ij} between sample locations i and j and the range θ (the maximum distance over which the significant autocorrelation occurs):

$$Cov(\varepsilon_i, \varepsilon_j) = \sigma^2 f(d_{ij}, \theta) \tag{3}$$

The spherical covariance model was used in this study:

$$f(d_{ij}, \theta) = \frac{3d_{ij}}{2\theta} - \frac{{d_{ij}}^3}{2\theta^3} \tag{4}$$

A backward stepwise selection procedure was used to choose the best combination of explanatory variables based on Akaike's information criterion (AIC) (Akaike, 1973). The GLM catch rate standardization was conducted using the "gls" function in the R software package (Version 2.9.1, 2009, USA). The year effect was calculated and regarded as the relative abundance index after exponential transformation because it was initially log-transformed:

$$I_t = exp\left(\frac{b_t + \sigma_t^2}{2}\right) \tag{5}$$

where I_t represents the estimated catch rate in year t, b_t is the estimated year effect for year t and σ_t is the standard error of b_t.

2.3. Generalized additive model

Generalized additive models (GAMs) are nonparametric generalizations of GLMs in which linear predictors are replaced by