Fisheries Research 181 (2016) 14–24

Contents lists available at ScienceDirect

Fisheries Research

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

Modeling spatially-varying ecological relationships using geographically weighted generalized linear model: A simulation study based on longline seabird bycatch

Yan Li[a,*,1], Yan Jiao[a], Joan A. Browder[b]

[a] Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0321, USA
[b] NOAA National Marine Fisheries Service Southeast Fisheries Science Center, 75 Virginia Beach Dr., Miami, FL 33149, USA

ARTICLE INFO

Article history:
Received 9 November 2015
Received in revised form 29 March 2016
Accepted 30 March 2016
Available online 8 April 2016

Keywords:
Spatially varying relationships
Seabird bycatch
Binomial distribution
Spatial expansion model
Geographically weighted regression

ABSTRACT

Geographically weighted regression (GWR) is a relatively new technique to explore spatially-varying relationships between biological and environmental processes. It allows parameters to vary over space and assumes data to follow a normal distribution. We extend GWR to a geographically weighted generalized linear model (GW-GLM) by incorporating statistical distributions other than the normal distribution (i.e., the binomial distribution). We demonstrate the application of GW-GLM with an empirical example, U.S. Atlantic pelagic longline seabird bycatch. Due to the high percentage of zero observations in the seabird bycatch data, we analyzed the positive catch rates (number of seabirds caught per 1000 hooks) and the probability of catching a seabird separately. Parameter estimates exhibited considerable spatial variation, especially for target catch rate when analyzing the positive catch data, and for intercept, water depth and water temperature when estimating the probability of catching seabirds. We compared model performance of GW-GLM with a global generalized linear model, a mixed effect model with a random areal effect, and a spatial expansion model that is an early technique to model spatially-varying ecological relationships by modeling each of the parameters as a function of location. The GW-GLM performed best. Simulations with hypothetical datasets having different percentages of zeros showed that, regardless of the zero percentage in the data, GW-GLM performed best on average. Applying a range of bandwidth indicated that the GW-GLM was more robust to an overestimated bandwidth than an underestimated bandwidth.

1. Introduction

Understanding key relationships between biological processes and environmental factors is important in natural resource management and conservation. For example, catch rate standardization based on relationships between catch rate and environmental factors can be used to estimate annual patterns of catch rate for fish stock assessment (e.g., Stefansson, 1996; Maunder and Punt, 2004; Damalas et al., 2007). Bycatch assessment using these relationships can enable prediction of bycatch hotspots (e.g., Murray, 2004; Li et al., 2011; Winter et al., 2011). A “global” model is commonly applied to examine the relationships between biological processes and environmental factors. The global model assumes these relationships to be stationary over the entire study area, i.e., constant and independent of location and direction. This assumption may not be well suited to spatial ecological data given the dynamic spatial interactions between biological and environmental factors in natural ecosystems. Instead, assuming spatially-varying relationships between biological and environmental factors (Brunsdon et al., 1996) is more appropriate.

One of the early techniques to model spatially-varying ecological relationships is the spatial expansion model (SEM) (Casetti, 1972; Casetti and Jones, 1992). In the SEM model, each of the parameters is itself a function of location. The nature of the function must be predetermined (e.g., linear or polynomial) based on prior knowledge. However, model results may be sensitive to the specification of the expansion function (Charlton et al., 2009). In addition, computation and interpretation may become difficult when higher order polynomials are used for the expansion function, especially in the case of limited data.

A relatively new technique to address spatial variation in ecological data is geographically weighted regression (GWR, Brunsdon

* Corresponding author.
E-mail addresses: yxl73@psu.edu, yanli9010@me.com (Y. Li), joan.browder@noaa.gov (J.A. Browder).
[1] Current address: Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University, University Park, PA 16802, USA.

http://dx.doi.org/10.1016/j.fishres.2016.03.024
0165-7836/

et al., 1996; Fotheringham et al., 1998). In GWR, neighboring observations are weighted based on the strength of spatial dependence and thus drive parameters to vary across the study area. GWR has several advantages in exploration of spatial variation in ecological relationships. First, it yields a set of estimates for each of the parameters that vary over space. The set of estimates can be mapped over the entire study area, which provides a visualization of the spatially-varying relationships between biological and environmental variables. Second, GWR extends the traditional regression approach with spatially-varying parameters and can easily be fitted using the weighted least square (WLS) method. Third, spatial relationships among observations in GWR can be investigated by testing different spatial weighting matrices that combine information on neighbor structure and spatial dependence. The original formulation of GWR (Brunsdon et al., 1996; Fotheringham et al., 1998), is limited to the normal assumption. We extend GWR to the geographically weighted generalized linear model (GW-GLM) by applying distributions other than the normal from the exponential family (i.e., the binomial and Poisson distributions). GW-GLM can be fitted using iteratively reweighted least squares (IRLS, McCullagh and Nelder, 1989). GW-GLM is also an extension of the global generalized linear model (GLM) in which spatial weights are applied in the fitting procedure (i.e., the IRLS) to consider spatial locations in parameter estimation. GWR is a special case of GW-GLM in which a normal distribution and an identity link function are applied, and spatial weights are directly applied to observations.

We demonstrate the application of GW-GLM using the example of seabird bycatch from the U.S. pelagic longline fishery in the Western North Atlantic. The incidental mortality of seabirds in longline fisheries has raised a global concern among marine conservationists because it has threatened several albatross (Diomedeidae) species and other species in the order of Procellariiformes (Brothers et al., 1999a; Tasker et al., 2000; Belda and Sanchez, 2001; Furness, 2003). Occurrence of seabird bycatch has been related to environmental variables such as wind speed (Klaer and Polacheck, 1998; Brothers et al., 1999b). Seabird bycatch is likely to be spatially-varying given the highly mobile nature of seabirds and the large heterogeneous habitat covered by longline fishing. Therefore, conservation and management of seabird bycatch will benefit from a more complete understanding of the relationships (e.g., the spatially-varying relationships) between behavioral processes and environmental factors.

We address two major questions. First, is GW-GLM better than the global GLM, the mixed effect model (MEM) where observations were assumed dependent within the same fishing zone and independent between zones, and the SEM, an alternative model for spatially-varying relationships? Second, are relationships between biological processes and environmental factors in GW-GLM significantly varying over space?

2. Methods

2.1. Geographically weighted generalized linear model (GW-GLM)

The GW-GLM is an extension of the global generalized linear model by considering locations in parameter estimation and thus allowing parameters to vary locally rather than being constant over space. A global GLM takes the form:

$$g\left(E\left[y_i\right]\right) = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}, \tag{1}$$

where $i = 1, 2, \ldots, n$ denotes the ith location, y is the observation, $E[y]$ is the expectation of y, β_0 is the intercept, β_k is parameter for the kth explanatory variable x_k and $k = 1, 2, \ldots, p$, and $g(.)$ is the link function that describes the relationship between $E[y]$ and explanatory variables. The parameters in the global GLM can be estimated using an iterative algorithm in which weighted least squares (WLS) is applied to the adjusted response variable $Z = [z_1, z_2, \ldots, z_n]^T$ at each iteration step (McCullagh and Nelder, 1989):

$$\hat{\beta}^{new} = \left(X^T W^{old} X\right)^{-1} X^T W^{old} Z, \tag{2}$$

where T denotes the transpose of a matrix, $\hat{\beta}^{new} = \left[\hat{\beta}_0^{new}, \hat{\beta}_1^{new}, \ldots, \hat{\beta}_p^{new}\right]^T$ is a vector of estimated parameters that are updated at each iteration step, and X is the design matrix defined as:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}. \tag{3}$$

The adjusted response variable z_i is defined as:

$$z_i = x_i^T \hat{\beta}^{old} + \left(y_i - E\left[y_i\right]\right)\left(G'\left(x_i^T \hat{\beta}^{old}\right)\right)^{-1}, \tag{4}$$

where $G(.)$ is the inverse function of $g(.)$ and $G'(.)$ denotes the first derivative of function G(.) with respect to parameters β. The $n \times n$ weighting matrix $W = [w_i]$ can be written as:

$$W = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{bmatrix}, \tag{5}$$

and w_i is updated accordingly at eatch iteration step:

$$w_i^{new} = \frac{G'\left(x_i^T \hat{\beta}^{new}\right)^2}{V\left(E\left[y_i\right]\right)}, \tag{6}$$

where $V(.)$ is the score function that is defined as the first derivative of the log-likelihood with respect to parameters β. The whole algorithm progresses until parameter estimates converge. In the global GLM, a unity weight is assigned to each observation as the prior weights. parameter estimation can be adjusted by controlling prior weights. Linear regression is a special case of GLM in which $G'(.) \equiv 1$ and $E[y_i] = x_i^{T\beta}$, and thus prior weights can be applied directly to the observations and no iteration is necessary.

As an extension of the global GLM with locally varying parameters, the geographically weighted generalized linear model (GW-GLM) can be written as:

$$g\left(E\left[y_i\right]\right) = \beta_0\left(u_i, v_i\right) + \sum_{k=1}^{p} \beta_k\left(u_i, v_i\right) x_{ik}, \tag{7}$$

where (u_i, v_i) denotes the coordinates of the ith location (longitude, latitude), and β_0 (u_i, v_i) and β_k (u_i, v_i) are parameters for the ith location. In common with most spatial analyses methods, GW-GLM implicitly assumes that observations nearer to location i have more influence on the parameter estimation for this location than observations farther away from location i (Cressie, 1993; Fotheringham et al., 1998). Thus, in a GW-GLM, parameters for the ith location are estimated using its neighboring observations through the IRLS method as described above, except that prior weights are assigned to its neighboring observations according to their influence on the