ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Fisheries Research 84 (2007) 210–221

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

Modeling shark bycatch: The zero-inflated negative binomial regression model with smoothing

M. Minami [a,*], C.E. Lennert-Cody [b], W. Gao [c], M. Román-Verdesoto [b]

[a] *The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo 106-8569, Japan*
[b] *Inter-American Tropical Tuna Commission, 8604 La Jolla Shores Drive, La Jolla, CA 92037-1508, USA*
[c] *School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin, 130024, China*

Received 28 April 2006; received in revised form 24 October 2006; accepted 27 October 2006

Abstract

The zero-inflated negative binomial (ZINB) regression model with smoothing is introduced for modeling count data with many zero-valued observations, and its use is illustrated with shark bycatch data from the eastern Pacific Ocean tuna purse-seine fishery for 1994–2004. Based on the generalized information criterion, the ZINB regression model provided a better fit to the data than either Poisson, negative binomial or zero-inflated Poisson regression models. To demonstrate the utility of the ZINB regression model for the standardization of catch data, standardized temporal trends in bycatch rates estimated with the ZINB regression model are computed and compared to those obtained from fits of the other three types of models to the same data. With the exception of the negative binomial, estimated temporal trends were more similar among models than would have been inferred from an analysis of model fit. Comparison of trends among models suggests that the negative binomial regression model may overestimate model coefficients when fitted to data with many zero-valued observations.
© 2006 Elsevier B.V. All rights reserved.

Keywords: CPUE; Shark; EM algorithm; GAM; GLM; Spline smoothing; Negative binomial; Zero-inflated

1. Introduction

Count data on the catch of non-target species, and some target species, can have many zero-valued observations, but also include large values when aggregations of animals are caught. This is often true for species such as sharks (e.g., Bigelow et al., 1999; Ward and Myers, 2005). Modeling these data is essential to the estimation of trends in catch rates and for understanding processes that lead to increased, or decreased levels of catch. However, the true stochastic processes that generated the data are usually not known. Although such data have been modeled with a Poisson distribution (e.g., Walsh and Kleiber, 2001) or aggregated by fishing effort and modeled with a lognormal distribution (e.g., Simpfendorfer et al., 2002), the existence of data with an excess of zero-valued observations (i.e., more than expected from a Poisson process) has led to the development of models that expressly relate covariates to the occurrence of excess zeros (e.g., Welsh et al., 1996; Barry and Welsh, 2002; O'Neill and Faddy, 2003; Lemos and Gomes, 2004).

There are two commonly used types of mixture distributions for unaggregated count data that contain an excess of zeros: (1) models that treat the presence/absence of catch separately from positive catch ("delta-F" models), and (2) models that treat zero catch separately from events in which catch could occur ("zero-inflated" models). [1] These models differ somewhat in both their formulation and interpretation. delta-F models are two-part models that describe the probability of no catch separately from the probability of positive catch. The probability of no catch is typically assumed to follow a logistic model, and positive catches are typically assumed to follow a log-linear model based on either a truncated Poisson or truncated negative binomial distribution (e.g., Welsh et al., 1996; Barry and Welsh, 2002; O'Neill and Faddy, 2003). In terms of their interpretation, delta-F models make a distinction between covariates associated with no catch and those associated with non-zero catch.

[1] The negative binomial distribution (e.g., Lawless, 1987), which can be viewed as a mixture model that extends the Poisson distribution (e.g., McCullagh and Nelder, 1991), has been used for count data (e.g., Ward and Myers, 2005). However, it does not expressly relate covariates to the occurrence of excess zeros.

* Corresponding author. Tel.: +81 3 5421 8729; fax: +81 3 5421 8729.
E-mail address: mminami@ism.ac.jp (M. Minami).

0165-7836/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2006.10.019

delta-F models are also referred to as hurdle models (Simonoff, 2003).

Zero-inflated models are also expressed in two parts: the probability of being in a ‘perfect-state’ (e.g., no catch), and the probability of being in an ‘imperfect-state’ where positive events (e.g., catch) may occur, but are not certain. That is, the imperfect-state includes both zero and non-zero values. Zero-inflated models have been used in other areas of research (e.g., Lambert, 1992; Greene, 1994; Hall, 2000; Agarwal et al., 2002; Siminoff, 2003), but appear to be only rarely used in the analysis of fisheries data (Welsh et al., 1996). The perfect-state is typically modeled with a logistic, and a complete Poisson or complete negative binomial distribution is assumed for the imperfect state. These models are referred to as zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models, respectively. In terms of their interpretation, zero-inflated models make a distinction between covariates associated with the perfect state (no catch) and covariates associated with the imperfect state in which catch can occur, but is not certain.

Conceptually, zero-inflated models may be more appropriate for catch data of infrequently encountered species, particularly when processes leading to catch of these species are poorly understood. For example, it may be plausible to assume that a species has a preferred habitat. Samples collected in the preferred habitat would be expected to yield animals, while samples collected in other habitats would be expected to yield no animals. However, in practice, even when samples are collected in the preferred habitat, sampling variability would be expected to occasionally produce samples with no animals. Perhaps more importantly, in practice, the exact characteristics of preferred habitat are not often known, further complicating the interpretation of zero-valued observations. In such a situation, a delta-F model fitted to the data would incorrectly pool zero-valued observations that potentially result from separate processes. By contrast, zero-inflated models can utilize information in covariates to determine to which process zero-valued observations might belong. ZIP models might be considered appropriate for species that are caught infrequently, but when present occur in small groups, whereas ZINB models may be more appropriate for the data of species that when present, can occur in large aggregations.

In this manuscript we introduce the ZINB regression model with smoothing. The ZINB regression model with smoothing is an extension of the classical generalized additive model (GAM; Hastie and Tibshirani, 1991). GAMs are one of several tools frequently used to standardize catch per unit effort (CPUE) data, (e.g., Maunder and Punt, 2004). To fit the ZINB model, we employ thin plate regression splines (Wood, 2003), a variant of smoothing splines that avoids complications associated with the treatment of ‘knots’. To illustrate the use of the ZINB regression model as a tool for CPUE standardization, we estimate temporal trends in the bycatch per set of silky sharks in the eastern Pacific Ocean (EPO) purse-seine fishery for tunas associated with floating objects. We compare characteristics of ZINB regression models fitted to these data, with and without smoothing, to characteristics of ZIP, negative binomial and Poisson regression models fitted to the same data. Partial dependence plots (Hastie et al., 2001) are used to summarize temporal trends in bycatch per set for each of the models, taking into consideration the average effects of other predictors. Comparison of temporal trends among models illustrates important differences in the way in which the negative binomial and the ZINB fit highly skewed count data.

2. Data

Data on the incidental mortality of silky sharks (nominally *Carcharhinus falciformis*; Román-Verdesoto and Orozco-Zöller, 2005; Minami et al., 2006) collected by IATTC observers onboard large tuna vessels of the international purse-seine fleet between 1994 and 2004 were used to demonstrate the ZINB model. Observers go to sea aboard the largest size category of fishing vessels (> 363 metric tons fish-carrying capacity) in order to collect data on the incidental mortality of dolphins and details of fishing operations. Additionally, these observers collect data on the local environment, the amounts and species of tuna caught, and, since 1993, the bycatches of non-mammal species. The term bycatch will be used herein in place of ‘catch’ to refer to the incidental mortalities of any non-target species. Target species for this fishery are yellowfin tuna (*Thunnus albacares*), skipjack tuna (*Katsuwonus pelamis*), and bigeye tuna (*Thunnus obesus*).

Purse-seine sets are categorized into three types according to the intent of the fishermen. Fishermen may target tunas associated with marine mammals, tunas associated with floating objects, or unassociated schools of tunas. Floating objects include both fish-aggregating devices (FADs) and flotsam, although since 1996, more than 80% of the objects used have been estimated to be FADs (IATTC, 2005). FADs are typically equipped with some form of relocation equipment, such as a radio beacon or a satellite transmitter. We demonstrate the use of the ZINB model with data from purse-seine sets on tunas associated with floating objects (hereafter referred to as ‘floating object’ sets). In the last decade, floating object sets were largely made within two longitudinal bands north and south of the equator, extending from the coast to as far offshore as approximately 170°W (Watters, 1999). Sampling coverage for data on non-mammal bycatch in floating object sets by IATTC observers over this 11-year period was generally greater than 64% annually (IATTC, 2006). After processing, data on 32,148 floating object sets made between 1994 and 2004 were available for analysis. Details of data processing can be found in Minami et al. (2006).

The silky shark bycatch data are characterized by many zero-valued observations and a long right tail (Fig. 1). Annually, the percentage of sets with no reported silky shark bycatch has increased from approximately 40% between 1994 and 1998 to over 60% since 2001. Overall, 51% of sets had no bycatch of silky sharks (Fig. 1). When silky shark bycatch did occur, sets involving up to about 20 animals were relatively common (Fig. 1). The large percentage of zero-bycatch sets, combined with the fact that occasional sets had bycatches of 10–100 s of animals, do not lend the analysis of these data to simple models that are sometimes used for count data (e.g., Poisson).

Download English Version:

https://daneshyari.com/en/article/4544581

Download Persian Version:

https://daneshyari.com/article/4544581

Daneshyari.com