Fisheries Research 153 (2014) 83–88

Contents lists available at ScienceDirect

Fisheries Research

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

Application of finite mixture models to catch rate standardization better represents data distribution and fleet behavior

Ronan Cosgrove[a,*], Michael Sheridan[b], Cóilín Minto[b], Rick Officer[b]

[a] Bord Iascaigh Mhara (Irish Sea Fisheries Board), New Dock Road, Galway, Ireland
[b] Marine and Freshwater Research Centre, Galway Mayo Institute of Technology, Dublin Road, Galway, Ireland

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form 9 January 2014
Accepted 11 January 2014
Handling Editor A.E. Punt
Available online 15 February 2014

Keywords:
Non-normal data
Albacore tuna
Catch per unit effort
Finite mixture model
EM algorithm

ABSTRACT

Catch-per-unit-effort (CPUE) data are routinely standardized to account for factors that influence catch rates that are not related to resource abundance. Despite improvement in the methods applied to CPUE standardization, for many datasets model diagnostics can still indicate poor conformity to modeling assumptions, imprecision and unexplained fishing behaviors. In this study we examine catch rate data of an Irish mid-water pair trawl fleet targeting albacore tuna (*Thunnus alalunga*) in the North East Atlantic. A fleet strategy of searching and congregating on fish aggregations combined with negative skew in model residuals suggest that multiple components exist within the dataset. Assuming up to five components, finite mixture models are applied and compared using the Bayesian information criterion. The two component model most consistently explained observed distributions in fishing behaviors and catch rates. Finite mixture modeling markedly improved conformity to modeling assumptions, resulting in substantial improvement in the precision of specific components used in CPUE standardization and reduced inter-annual variability of the catch rate trend. These methods may facilitate investigations of technological creep but also raise questions on how best to use the results in assessment.

1. Introduction

Stock assessments for many fisheries rely on fishery-dependent catch and effort data to measure annual trends in the relative abundance of the stock (Hilborn and Walters, 1992). Ideally indices of abundance are based on fishery-independent data collected through scientific surveys which use standardized conditions to eliminate or minimise the influence on catch rates of factors other than resource abundance. However, such surveys are often not economically feasible due to the large spatial extent of many fisheries, particularly those for migratory species such as tunas. Instead catch rates calculated from fishery-dependent catch and effort data are often relied upon in stock assessment, and assumed to be proportional to underlying resource abundance (Hinton and Maunder, 2004). However, debate about the limitations of using catch per unit effort (CPUE) data questions the assumed relationship between catch rates and underlying resource abundance (Harley et al., 2001; Richards and Schnute, 1986).

Before CPUE data can be included in a stock assessment as an index of abundance, it is important to standardize the data to remove or minimise the effect that any varying factors other than resource abundance have on the catchability of a target species (Maunder and Punt, 2004). Modelling of CPUE facilitates consideration of a range of explanatory variables, or factors, that potentially affect catch rates (Campbell, 2004), whilst still achieving the primary objective of detecting trends over time in the abundance of the target species (Maunder and Punt, 2004). Generalised linear modelling (GLM) approaches are used to standardize catch and effort data assuming linearity of the model coefficients on a suitable scale (Hinton and Maunder, 2004; Maunder and Punt, 2004).

However, problems persist when carrying out fishery-dependent CPUE standardizations, generally because the proportionality between resulting indices of abundance and actual abundance remains uncertain. Campbell (2004) suggests that these problems arise mainly due to a lack of data on inexplicit factors that are likely to affect catch rates. Salthaug and Aanes (2003) question whether catch and effort data from a commercial fishing fleet, driven by economic imperatives to maintain high catch rates, can ever be used to produce an index of abundance that actually reflects stock abundance. Concentration of fishing effort in areas where target species aggregate, without coverage of the actual spatial extent of the stock, can result in catch rates displaying hyperstability and declining more slowly than underlying resource abundance (Cooke and Beddington, 1984; Erisman et al., 2011; Rose and Kulka, 1999). Harley et al. (2001) also conclude that CPUE remains hyperstable

* Corresponding author. Tel.: +353 91564318.
E-mail address: cosgrove@bim.ie (R. Cosgrove).

0165-7836/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2014.01.005

and high while actual resource abundance declines due to the behaviour of commercial fishing fleets. Nevertheless, indices of abundance that inform stock assessment for non-target species such as billfish (Ortiz and Arocha, 2004) as well as many of the world's most valuable and endangered species, such as sharks and tunas (Punt et al., 2001; Rodriguez-Marin et al., 2003), remain founded on fishery-dependent catch and effort data, due to the prohibitive cost of deploying standardized scientific surveys. It is therefore imperative that CPUE standardization methods better account for the behaviours exhibited by commercial fleets.

Albacore tuna (*Thunnus alalunga*) is a commercially important species, contributing 6% to the annual global tuna catch of 4.3 million t in 2010 (FAO, 2012). A total of 19,995 t were landed in the North Atlantic in 2011 with some 82% of landings attributed to directed French and Irish mid-water pair trawling (MWPT), Spanish bait-boat and troll, and Portuguese bait-boat fisheries operating predominantly in the greater Bay of Biscay area and south-west of Ireland (ICCAT, 2013a). Temporal and spatial availability of Atlantic albacore tuna is widely variable (Cosgrove et al., 2014; ICCAT, 2013a) obliging fleets such as the Irish MWPT to engage in strategies of searching over wide offshore areas and congregating on fish aggregations to maximise catches.

In the absence of fishery-independent surveys, assessments of Atlantic albacore rely on standardized CPUE indices. Poly-modality and skewness are common in the residuals of standardizations of these catch rate data (Andrade, 2012; Cosgrove et al., 2013; Kell et al., 2010; Kerwath et al., 2012). Potential causes include temporal (e.g. Wilberg et al., 2009) and spatial variability in catchability (e.g. Thorson et al., 2012a; Walters, 2003), gear saturation (e.g. Groeneveld et al., 2003), vessel constraints such as hold capacity or handling time (e.g. Murray et al., 2013) and vessel interactions (e.g. Gillis and Peterman, 1998). In addition we suggest that the search and congregate strategies employed by the Irish MWPT and other fleets create sub-components within catch rate data which also affect catch rate distributions.

A finite mixture model is a combination of two or more probability density functions. By combining the properties of the individual probability density functions, mixture models are capable of approximating any arbitrary distribution (McLachlan and Peel, 2004). Examples of the application of finite mixture modelling to fisheries include identification of sub-populations of recreational anglers through examination of their excursion behaviour (Provencher et al., 2002) and successfully distinguishing species and enumerating their abundances from mixed fisheries data (Fleischman and Burwen, 2003). More recently a finite mixture model was employed to improve the standardization of fishery-independent trawl survey data used to estimate abundance of shoaling Pacific rock fish (*Sebastes* spp.) (Thorson et al., 2012b). This approach has major potential to improve standardization of fishery dependent data for species such as albacore tuna by better approximating different modes of fishing and reducing variability around standardized indices. Furthermore, while yet to be applied in a fisheries context, finite mixture models have been used in climate studies to improve simulations of inter-annual variability of parameters such as rainfall (Zheng and Katz, 2008). Similar improvements to inter-annual variability of catch rate indices could maximise inclusion of operationally diverse fisheries, facilitating more thorough and comprehensive assessment of available fisheries dependant data.

In this study we test the potential benefits of applying a finite mixture modelling approach to the standardization of catch rate data by applying this approach to the Irish MWPT catch and effort series. A comparison to the extant GLM method is made to assess the relative conformity of finite mixture modelling to analytical assumptions and relative precision.

2. Materials and methods

2.1. Data sources and description

Mid-water pair trawl (MWPT) fishing for albacore (*Thunnus alalunga*) involves two vessels towing a pelagic trawl between them close to the surface at night when fish are predominantly shallow (Cosgrove et al., 2014). Detailed catch and effort data for the Irish MWPT fishery were available from mandatory logbooks compiled by the Irish Sea Fisheries Protection Authority for the years 2003 to 2012. Although Irish vessels commenced MWPT in 1998, information prior to 2003 were excluded due to major data gaps associated with diversification from drift netting to MWPT during this period. This omission is thought to at least partly negate the learning curve associated with introduction of a novel fishing technique. Available data including location, date and quantity (kg) of landed albacore, trip departure and landing dates, permitted nominal catch per day at sea to be estimated for each compiled trip.

The full dataset used in this study comprised catch information relating to 5627 days at sea carried out by 66 vessels from 2003 to 2012. Catch per day at sea ranged from 0 to 36,000 kg day^{-1} with a mean of 2674 ± 154 kg (standard error). Considerable variation and the presence of zero catches are characteristic of a fishery where vessels must search to detect schools of tuna. Very large catches occur when detections are successful, but poor or zero catches arise when schools are not located.

2.2. Standardization model

Stefansson (1996) stresses the importance of selecting a model where the zero values influence the standardized CPUE indices explicitly and non-arbitrarily. The delta-lognormal GLM advanced by Lo et al. (1992) is considered an appropriate method for handling zero catches (Maunder and Punt, 2004), and was firstly used to standardize CPUE in this study. The delta-lognormal GLM handles zero values explicitly by first modelling the probability of obtaining a zero or positive catch, assuming a binomial distribution as a function of the set of explanatory variables. It then models the positive (non-zero) catches assuming a lognormal distribution as a function of the set of explanatory variables, before combining the two models to generate standardized CPUE indices (Hinton and Maunder, 2004; Maunder and Punt, 2004). Results from the binomial model likelihood ratio tests indicated that year and vessel size category had a significant effect on the probability of a trip encountering a positive catch.

2.2.1. Positive catches

Here we focus on the standardization of positive catches (C) rather than the binomial component. Making the fewest assumptions regarding the nature of the relationship, categorical rather than quadratic or spline-form factors were employed:

- Year Y (categorical variable) $Y = \{2003, 2004, \ldots, 2012\}$,
- Quarter Q (categorical variable with 2 levels: July–September, Q3; October–November, Q4),
- Fishing zones Z (categorical variable with 2 levels: Ireland (north of 48° N), Bay of Biscay (south of 48° N)),
- Vessel size category V (m) (categorical variable with 5 levels: C1: <20; C2: 20 <25; C3: 25–<30; C4: 30–<40; C5: ≥40),
- Effort E (continuous variable of the number of days at sea).

ICES area data available from logbook data were converted to two general zones to take account of a small number of observations in some ICES areas and the general distribution of the Irish fleet between two main areas to the west and south-west of Ireland and the Bay of Biscay. Mean length of vessels involved in the fishery