Fisheries Research 155 (2014) 149–159

Contents lists available at ScienceDirect

Fisheries Research

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

Proof of concept for a novel procedure to standardize multispecies catch and effort data

Henning Winker[a,*], Sven E. Kerwath[b,a], Colin G. Attwood[a]

[a] Marine Research Institute, Department of Biological Sciences, University of Cape Town, Private Bag, Rondebosch 7701, South Africa
[b] Department of Agriculture Forestry and Fisheries, Private Bag X2, Roggebaai 8012, South Africa

ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form 13 February 2014
Accepted 15 February 2014
Handling Editor A.E. Punt
Available online 27 March 2014

Keywords:
CPUE standardization
Simulation-testing
Multispecies targeting
Tweedie distribution
GAM

ABSTRACT

The effect of variability in targeting needs to be removed from catch-per-unit-effort (CPUE) data to estimate reliable abundance indices for multispecies fisheries. We test a Generalized Additive Model (GAM) that includes principal component scores (PCs) derived from the species composition in the catch, called the 'Direct Principal Component' (DPC) procedure, for its ability to remove the effect of variable targeting. Biomass trends are simulated for two multispecies, multi-habitat fishery scenarios: (i) four species distributed differentially across two habitats and (ii) ten species distributed differentially across four habitats. Tweedie distributed CPUE records are generated from the biomass trends for a fishery with constant targeting (control scenarios) and time-varying targeting (test scenarios). The DPC procedure is simulation-tested for its ability to estimate the underlying biomass trends for all species relative to the non-standardized CPUE index. The DPC procedure proved to be more accurate compared to nominal CPUE trends in the test scenarios. Even in the control scenarios, the DPC procedure offers greater accuracy for the estimated year effect by removing substantial variation from the data, with a small penalty on the accuracy of the underlying abundance trend. However, caution is advised if the DPC-derived index diverges noticeably from alternative models despite no indications for shifts in targeting. A selection procedure based on eigenvalues of the PCs is suitable to identifying the best-performing number of PCs to include in the GAM. The DPC procedure should be applicable for a variety of multispecies fisheries.

1. Introduction

The standardization of catch-per-unit-effort (CPUE) is now widely regarded as a prerequisite for the use of CPUE as abundance index in stock assessment models (Maunder and Punt, 2004; Maunder et al., 2006). The nominal CPUE index, derived from yearly means of the raw CPUE data, can be severely biased due to non-random allocation of fishing effort over time (Harley et al., 2001; Maunder et al., 2006; Carruthers et al., 2010). The most commonly used standardization procedures entail the application of Generalized Linear Models (GLMs) or Generalized Additive Models (GAMs), which aim to isolate temporal abundance trends from total variation in the CPUE data by adjusting for confounding effects on the estimated abundance trends (Guisan et al., 2002; Maunder and Punt, 2004). Influences on the CPUE other than abundance are remarkably diverse and typically include time-variant changes in spatial and seasonal effort distribution, gear, fishing power and fishing behavior (Punt et al., 2000; Maunder and Punt, 2004; Carruthers et al., 2010). The problem of estimating reliable abundance indices is exacerbated in multispecies fisheries for which the available CPUE records may reflect a number of fishing strategies, each associated with a particular choice of fishing-ground, habitat-type, and fishing-technique; even within the same fishing trip (Pelletier and Ferraris, 2000; Palmer et al., 2009; Winker et al., 2013).

An important consideration for the standardization of multispecies CPUE data is that the choice of fishing tactic allocates effort toward a particular target species or species complex and away from others, where the term 'fishing tactic' is defined as a sequence of choices of fishing strategies made by the skipper during a fishing trip (Pelletier and Ferraris, 2000; Winker et al., 2013). Temporal variations in fishing tactics inevitably violate the fundamental assumption that CPUE is equal to the product of abundance and a constant catchability (i.e. fraction of biomass caught per unit effort), because the latter will strongly depend on the choice of fishing tactic (Pelletier and Ferraris, 2000; Carvalho et al., 2010; Winker et al., 2013).

* Corresponding author. Tel.: +27 021 650 3459; fax: +27 021 650 3301.
E-mail address: henning.winker@gmail.com (H. Winker).

http://dx.doi.org/10.1016/j.fishres.2014.02.016
0165-7836/

Conventional multispecies standardization models often include the catch rates of alternative target or bycatch species as covariates to correct for the effort directed away from the target species or species under consideration (Glazer and Butterworth, 2002; Maunder and Punt, 2004; Su et al., 2008). Importantly, the alternative species should not co-occur with the target species. For example, if two species were to co-occur in the catches and would be fished down simultaneously, the use of the catch rate of the one species as a negative predictor of the CPUE of the other may result in an erroneous removal of the underlying year-effect for the species of interest (Glazer and Butterworth, 2002; Maunder and Punt, 2004). An additional challenge in situations where a large number of species are caught by the fishery is the objective selection of species-specific catch rates to be included as covariates in the standardization model.

Stephens and MacCall (2004) proposed an approach to subset catch and effort records that uses the species composition from fishing trips to discriminate between catch records from habitats where the species assessment is common and catch records from habitats where the species under is unlikely to be encountered. The idea is that the species composition (excluding the species under assessment) from a fishing trip provides information that can be used to make predictions whether the fishing trip included at least some effort expended in the target species' habitat. However, this approach does not make any inference about the extent of effort that is allocated to a habitat.

An alternative approach is based on clustering fishing trips according to their similarity in catch composition (He et al., 1997; Pelletier and Ferraris, 2000; Carvalho et al., 2010). The identified clusters are assumed to be a representation of fishing tactics, which may be treated as categorical variables in the standardization model to adjust for differences in catchability associated with each cluster (Pelletier and Ferraris, 2000; Carvalho et al., 2010; Winker et al., 2013). This approach typically requires the implementation of a rather complex analytical framework based on a sequence of ordination and clustering techniques and involves several subjective steps (Pelletier and Ferraris, 2000; Deporte et al., 2012; Winker et al., 2013).

A more direct method for the standardization of multispecies CPUE records was recently proposed by Winker et al. (2013). This 'Direct Principal Component' procedure (DPC) uses continuous principal component scores (PCs), derived from a Principal Component Analysis (PCA) of the catch composition data, as nonlinear predictor variables in a GAM to adjust for the effect of temporal variations in fishing tactics. The DPC procedure is based on the common assumption that information on the direction and extent of targeted effort can be found in the species composition of the catch (Pelletier and Ferraris, 2000; Carvalho et al., 2010).

Although the species composition does not hold direct information about the magnitude of the catch, it is arguably of concern that the information contained in the predictor variables derived from the catch composition is not entirely independent from the response CPUE and may have unpredictable impacts on the standardized CPUE trends. The standardization procedure would fail if variation in abundance of a particular species is falsely attributed to variation in targeting. Common model selection procedures, such as analysis of deviance, Akaike's information criterion (AIC) or cross-validation methods only evaluate the model based on how well it fits the data, but may fail to identify the model that provides least biased representation of the true abundance pattern (Carruthers et al., 2010).

The aim of this study was therefore to use simulation testing to evaluate if the DPC method is able to accurately track 'true' abundance trends. We consequently simulate multispecies catch data from individual fishing trips that exhibit variation in effort allocation across alternative fishing habitats over a time series of 20 years. These scenarios broadly resemble the habitat associations and catch rates of several common target species in the multispecies hand-line fishery off the South African south coast. The specific objectives were: (i) to test the efficacy of the DPC method in eliminating the effect of time-varying trends in fishing tactics on nominal CPUE, (ii) to evaluate the risk associated with the DPC method to introduce bias in terms of systematic departures from the simulated abundance trend and (iii) to evaluate alternative selection criteria for including the optimal number of PC covariates in the DPC standardization model.

2. Materials and methods

Simulation testing is a powerful tool to evaluate the performance of CPUE standardization methods (Carruthers et al., 2010; Lynch et al., 2012; Thorson et al., 2012a). The advantage of this approach is that the simulated abundance trends are known, so that the standardization method can be tested in terms of how well it predicts 'true' abundance trends. We developed a simulation framework to generate non-standardized nominal CPUE records per trip for two scenarios: (i) a simple multispecies-fishery scenario, comprising four target species that are unevenly distributed across two habitats; and (ii) a more complex multispecies-fishery scenario, comprising ten target species that are unevenly distributed across four habitats. As a study system, we use the example of the South African multispecies hand-line fishery, which represented the initial case study for the DPC procedure (Winker et al., 2013).

As is common practice, the use of CPUE as an index of abundance was based on the concept that catch rate is equal to the product of catchability and biomass: CPUE = qB, where q is the catchability representing the fraction of biomass caught by expending one standard unit of effort and B is the biomass (Maunder and Punt, 2004). This relationship only holds if q is constant, which is almost certainly violated in multispecies-fisheries that employ a variety of fishing tactics. To simulate this effect, we assumed that the choice of targeting tactic is reflected by the choices of up to two target habitats h during a fishing trip t and that each habitat is associated with a catchability for species i, $q_{i,h}$, based on the conceptual considerations outlined in Winker et al. (2013). All simulations were conducted within the statistical environment R (R Development Core Team, 2011).

2.1. Basic dynamics

The 'true' underlying abundance trends were simulated in the form of biomass trajectories for each species i over a period of 20 years:

$$B_{i,y} = B_{i,1}e^{(r_i(y-1))} \quad y = 1, 2, \ldots, 20. \tag{1}$$

where $B_{i,1}$ is the biomass of species i at start of the time-series and r_i is the rate of increase (or decrease) for species i. Nominal CPUE records from fishing trip t for each species i in year y were assumed to be Tweedie distributed with expectation:

$$\mathrm{CPUE}_{t,i,y} = \sum_h q_{i,h} B_{i,y} f_{h,t} \tag{2}$$

where $q_{i,h}$ is the catchability of species i in habitat h and $f_{h,t}$ is the fraction of effort allocated to habitat h during a fishing trip t. The Tweedie distribution belongs to the family of exponential dispersion models and is characterized by a two-parameter power mean-variance function of the form Var(Y) = $\phi\mu^p$, where ϕ is the dispersion parameter, μ is the mean and p is the power parameter. Depending on the power parameter, the Tweedie model includes the four well-known distributions: normal (p = 0), Poisson (p = 1),