Fisheries Research 95 (2009) 254–259

Contents lists available at ScienceDirect

Fisheries Research

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

A simple statistical method for catch comparison studies

René Holst[a,1], Andrew Revill[b,*,1]

[a] National Institute of Aquatic Resources, Technical University of Denmark (DTU-Aqua), Box 101, DK-9850 Hirstshals, Denmark
[b] Centre for Environment Fisheries & Aquaculture Science (CEFAS), Pakefield Road, Lowestoft, NR33 0HT, UK

ARTICLE INFO

Article history:
Received 1 May 2008
Received in revised form 5 September 2008
Accepted 16 September 2008

Keywords:
Catch comparison
GLMM
Polynomial
Chebychev
Eliminator
Cod
North Sea

ABSTRACT

For analysing catch comparison data, we propose a simple method based on Generalised Linear Mixed Models (GLMM) and use polynomial approximations to fit the proportions caught in the test codend. The method provides comparisons of fish catch at length by the two gears through a continuous curve with a realistic confidence band. We demonstrate the versatility of this method, on field data obtained from the first known testing in European waters of the Rhode Island (USA) 'Eliminator' trawl. These data are interesting as they include a range of species with different selective patterns.

1. Introduction

The catch comparison experimental method has some advantages, which makes it preferred to selectivity trials for many purposes. This includes handling of the gears, commercial-like performance of the gears and conveying and interpretation of results. The same technique is also used in calibration experiments for trawl surveys.

The analysis of data from catch comparison experiments is however far less developed than that of selectivity experiments. See Millar and Fryer (1999) for a thorough review of the main achievements in statistical methods for the analysis of data from selectivity experiments. This imbalance may likely be explained by the fact that data from catch comparison experiments do not permit a model-based analysis in the sense of estimating the underlying selective properties by which the data were generated.

The more traditional approach for analysing catch comparison data uses paired tests by length classes. Paired *t*-tests have often been used in such studies, regardless of the data being discrete and potentially non-symmetric, thus violating the assumptions that underpin this test. The Wilcoxon test provides an alternative that does not rely on any distributional assumptions, but is disadvantaged by reduced power. Irrespective of the particular test, continuous curves along the length scale are more useful in describing the properties of interest, i.e. the catch efficiency at length of one gear relative to that of another gear.

Besides the intuitive appeal of making inference in terms of smooth curves with confidence bands rather than a succession of test conclusions, the approach proposed here works on the unscaled count data. In the case of sub-sampled catches the test-based methods use scaled up counts and thereby underestimate the variances and potentially produce erroneous conclusions. Our method does not suffer from this deficiency but uses the sub-sampling ratios as known offsets in the estimation and thereby obtain more realistic variance estimates.

A more recent curve fitting method/approach was proposed by Fryer et al. (2003). They used a two-stage mixed effects model similar to that commonly used for selectivity analysis (Millar and Fryer, 1999). Smooth curves were fitted for individual hauls using generalised additive models (Hastie and Tibshirani, 1990). For each length class in turn, the fitted values were input to Fryer's model of between-haul variation, giving a mean estimate over hauls. Joining up the mean estimates then gives a smooth mean curve. Imputations are used to fill in "empty" length classes. They justify their use of non-parametric smoothers by the complicated relationships of the observed proportions. We argue that the expected relationships derived from common assumptions about the selective properties makes low-order polynomials feasible approximations, with the order depending on the relative position and distance between the selective properties of the two gears. There is no requirement on

* Corresponding author. Tel.: +44 1502 562244.
E-mail address: andrew.revill@cefas.co.uk (A. Revill).
[1] Both the authors contributed equally.

0165-7836/$ – see front matter. Crown Copyright © 2008 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2008.09.027

similar coverage of length classes across hauls. The model can be fitted using standard statistical software which have routines for generalised linear mixed models.

2. Method

The GLMM method: catch comparison experiments generate binomial data and it is thus of interest to estimate the expected proportions at length of the total catch caught in the test codend. The absence of a non-selective control means that it is not possible to estimate the absolute selectivities of the two codends, by which these proportions are determined. The logit of the expected proportions at length may adequately be approximated by a low-order polynomial in length. For a single haul the curve may be fitted by a GLM. For a cruise, i.e. a collection of related hauls, the mean curve may exhibit random variation between hauls in addition to that variation accounted for by fixed effect covariates (Fryer, 1991). For such data a Generalised Linear Mixed Models (GLMM) can be used to obtain a pragmatic and reliable curve for the expected proportions-at-length and obtain realistic variance estimates by modelling the sampling structure.

Tools for GLM analyses are implemented in most general statistical packages today, whereas tools for GLMM analyses are less standardised and less widely available. This work used the `glmmPQL` function in MASS package of the R statistical software. It implements the penalised quasi likelihood function (Breslow and Clayton, 1993). Insignificant terms were removed based on the Wald's test.

Repeat GLMM analyses were also made on the same source data using the SAS PROC NLMIXED routine (Wolfinger, 1999). This routine implements a proper likelihood method for mixed models. The key syntax required for GLMM analyses for both R and SAS software packages is provided as an appendix to this work.

The polynomial GLMM method proposed here is justified by considering a model for the process by which data were generated. For ease of presentation we consider a single haul and therefore suppress reference to haul indices in the following. The two gears being compared are indexed t (test) and c (control). It is assumed that their selection properties can adequately be described by logistic curves. The conditional probability that a length ℓ fish is retained by codend i given it has entered it, is thus given by

$$r_i(\ell) = \frac{\exp(\alpha_{i,0} + \alpha_{i,1} \cdot \ell)}{1 + \exp(\alpha_{i,0} + \alpha_{i,1} \cdot \ell)} = \frac{\exp(\eta_{i,\ell})}{1 + \exp(\eta_{i,\ell})}, \quad i = t, c.$$

The split parameter, i.e. the probability that a fish "chooses" the test codend given it chooses one them, is denoted by π. The sub-sampling ratios q_t and q_c are the proportions taken out for measurements from the catch bulk of the test and control codend respectively. The logit of the expected proportion ϕ of the total catch caught in the test codend is then given by

$$\begin{aligned} \log\text{it}(\phi(\ell;\boldsymbol{\theta})) &= \log\left(\frac{\pi \cdot q_t \cdot r_t(\ell)}{(1-\pi) \cdot q_c \cdot r_c(\ell)}\right) \\ &= \log\text{it}(\pi) + \log\left(\frac{q_t}{q_c}\right) + \log\left(\frac{r_t(\ell)}{r_c(\ell)}\right) \\ &= \log\text{it}(\pi) + \log\left(\frac{q_t}{q_c}\right) + \eta_{t,\ell} - \eta_{c,\ell} \\ &\quad - \log\left(\frac{1+\exp(\eta_{t,\ell})}{1+\exp(\eta_{c,\ell})}\right), \end{aligned}$$

where $\boldsymbol{\theta} = (\alpha_{t,0}, \alpha_{t,1}, \alpha_{c,0}, \alpha_{c,1}, \pi)^T$.

This shows that when logit($\phi(\ell;\,\boldsymbol{\theta})$) is approximated by some k'th order polynomial in ℓ:

$$\log\text{it}(\phi(\ell;\boldsymbol{\theta})) \approx p_k(\ell;\boldsymbol{\beta}) = \log\left(\frac{q_t}{q_c}\right) + \beta_0 + \beta_1 \cdot \ell + \cdots + \beta_k \cdot \ell^k,$$

the intercept term β_0 absorbs the split and the intercepts $\alpha_{t,0}$ and $\alpha_{c,0}$. Inference about the efficiency of the test gear relative to that of the control gear is therefore subject to an assumed value of the split, say $\pi = 50\%$. Polynomial terms of order higher than 1 account for the non-linear component $\log(1+\exp(\eta_{t,\ell})/1+\exp(\eta_{c,\ell}))$ of logit($\phi(\ell;\,\boldsymbol{\theta})$).

When data are collected during say H hauls, a mixed effects model approach may be used to account for the variability between the hauls. For a random intercept model the polynomial associated with haul h becomes

$$p_k^{(h)}(\ell;\boldsymbol{\beta}) = \log\left(\frac{q_t^{(h)}}{q_c^{(h)}}\right) + \beta_0 + \beta_1 \cdot \ell + \cdots + \beta_k \cdot \ell^k + b_h,$$

with $q_t^{(h)}$ and $q_c^{(h)}$ being sub-sampling ratios for haul h for test and control respectively and where $b_h \sim N(0, \sigma^2)$. The model is readily extended to more complex variance-component structures (Longford, 1994).

Demonstrating the flexibility of polynomials: low-order polynomials may well approximate to the range of data likely encountered in a wide range of catch comparison scenarios. A polynomial fitted to data deviates from the "true" curve partly by the bias introduced by being an approximation and partly by the binomial errors. The bias depends on the 'wiggliness' of logit(ϕ) (as determined by the parameter $\boldsymbol{\theta}$) and the degree of the polynomial. The use of polynomial approximations is only useful if the bias can be made negligible by polynomials of relatively low-order. For known values of the parameter vector $\boldsymbol{\theta}$ and a fixed k, the bias from the assumed model curve logit(ϕ) may be quantified by the maximum absolute difference between the curve and the polynomial of degree k, which fits it the best; i.e. the polynomial of degree k with minimal maximum absolute difference to the curve it is approximating.

Since these so called minimax polynomials (Denman, 1966) are very difficult to obtain, Chebychev polynomials of the first kind are used as alternatives (Abramowitz and Stegun, 1972). These are virtually the same, but much easier to compute (Press et al., 1992). Thus, the potential of the method by finding the minimal required polynomial orders for a broad range of possible scenarios is possible. An exhaustive coverage is outside the scope of this article and instead we demonstrate the validation by three different settings. The three scenarios all describe the comparison of catches from fishing gears with widely differing L50's and selection ranges (SR) as might be encountered in most field experiments. Low-order polynomials can efficiently describe a wide range of likely catch comparison data (Fig. 1, Table 1). The maximum absolute differences between true catch comparison curves and various approximating polynomials under the three scenarios are detailed in Table 1, while the corresponding curves are plotted in Fig. 1.

3. Demonstration data

The GLMM method is demonstrated here using newly obtained unpublished data collected from a catch comparison study undertaken in the North Sea (Location 54-55°N and 001°W-001°E) during 2–8 December 2007. The catches of fish (measured to 1 cm below) from both the Experimental trawl (Eliminator trawl) (Fig. 2) and from the Control trawl (a typical North Sea demersal fish trawl) are compared. The parallel hauls method was used to collect the data, whereby two vessels towed the trawls along parallel tracks, keeping as close together as was practicable and safe (usually within 0.5