Available online at www.sciencedirect.com
SCIENCE DIRECT®
Fisheries Research 70 (2004) 141–159
FISHERIES RESEARCH
www.elsevier.com/locate/fishres

Standardizing catch and effort data: a review of recent approaches

Mark N. Maunder[a,*], André E. Punt[b]

[a] Inter-American Tropical Tuna Commission, 8604 La Jolla Shores Drive, La Jolla, CA 92037-1508, USA
[b] School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

Abstract

The primary indices of abundance for many of the world's most valuable species (e.g. tunas) and vulnerable species (e.g. sharks) are based on catch and effort data collected from commercial and recreational fishers. These indices can, however, be misleading because changes over time in catch rates can occur because of factors other than changes in abundance. Catch-effort standardization is used to attempt to remove the impact of these factors. This paper reviews the current state of the art in the methods for standardizing catch and effort data. It outlines the major estimation approaches being applied, the methods for dealing with zero observations, how to identify and select appropriate explanatory variables, and how standardized catch rate data can be used when conducting stock assessments.

Keywords: Abundance; Catch; CPUE; Effort; GAM; GLM; GLMM

1. Introduction

Scientific advice on fisheries management is generally based on the results of the application of some form of stock assessment technique (Hilborn and Walters, 1992). Stock assessment usually involves estimating the parameters of some form of population dynamics model by fitting it to research and monitoring data and using the results of the fitting process to estimate quantities (such as the current abundance) that are of interest to the decision makers. A variety of data types can be used when fitting stock assessment models. However, the data generally must include information on at least the removals due to harvesting and an index of relative abundance. Although the index of abundance should, ideally, be based on fishery-independent data collection methods such as surveys, fishery-independent data are often extremely costly or difficult to collect, in which case it must be based on fishery-dependent data. Therefore, assessments of many stocks (e.g. sharks and tunas) are based solely on fishery-dependent data. The most common (and easily collected) source of fishery-dependent data is catch and effort information from commercial or recreational fishers, usually summarized in the form of catch-per-unit-of-effort (CPUE) or catch rate.

* Corresponding author. Tel.: +1 858 546 7027; fax: +1 858 546 7133.
E-mail address: mmaunder@iattc.org (M.N. Maunder).

0165-7836/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2004.08.002

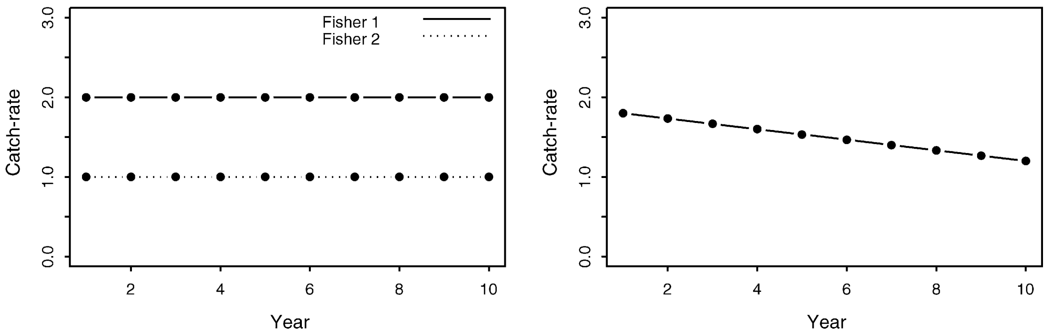

Fig. 1. Catch rate time series for two hypothetical fishers and the trend in raw catch rate (total catch divided by total effort) if the effort expended by fisher 1 decreases from 80% of the total to 20% over years 1–10.

The use of catch rate as an index of abundance assumes that, at small spatial scales, catch is proportional to the product of fishing effort and density:

$$C = qEN \tag{1}$$

where E is the fishing effort expended, N the density, and q the fraction of the abundance that is captured by one unit of effort (often referred to as the catchability coefficient).

Re-arranging Eq. (1) leads to the fundamental relationship between catch rate and density:

$$\frac{C}{E} = qN \tag{2}$$

Eq. (2) can be generalized from a small patch fished by a single fisher to the entire population fished by a large fishing fleet (in which case N is the population size rather than density) as long as q is a constant (independent of time, space, and fishing vessel). However, q may not be constant, but may change spatially and temporally due to changes in the composition of the fishing fleet, where fishing occurs, and when fishing occurs (e.g. Cooke and Beddington, 1984; Cooke, 1985; Hilborn and Walters, 1992). As an example, consider the case in which there are two fishers, and abundance is constant over time, with the result that the catch rate for each of the fishers is constant over time (Fig. 1, left panel). If the catch rate for fisher 1 is twice that for fisher 2 and the fraction the total effort expended by fisher 1 decreases from 80% of the total effort in year 1 to 20% in year 10, there is a marked decline in the 'raw' catch rate (total catch divided by total effort) over time (Fig. 1, right panel), even though there is actually no change in the true abundance of the resource.

The ability to use catch rate data as an index of abundance therefore depends on being able to adjust for (i.e. remove) the impact on catch rates of changes over time of factors other than abundance. This process is often referred to as 'catch-effort standardization'. The dangers associated with basing stock assessments on 'raw' catch rates have been known for many years, and various methods for standardizing catch and effort data have been developed (e.g. Gulland, 1956; Beverton and Holt, 1957; Robson, 1966; Honma, 1973), all of which define the efficiency of a fishing vessel as its 'fishing power' relative to that of a standard (and perhaps even hypothetical) fishing vessel. The most commonly applied method prior to the use of generalized linear modeling approaches was that developed by Beverton and Holt (1957). This method involves selecting a 'standard vessel' and determining the relative fishing power of all other vessels by

$$\mathrm{RFP}_i = \frac{C_i/E_i}{C_\mathrm{S}/E_\mathrm{S}} \tag{3}$$

where RFP_i is the relative fishing power for vessel i, C_i the total catch by vessel i during the period in which both the standard vessel and vessel i were in the fishery, C_S the total catch by the standard vessel during the period in which both the standard vessel and vessel i were in the fishery, E_i the total days fished (or whatever measure of fishing effort is chosen) by vessel i during the period in which both the standard vessel and vessel i were in the fishery, and E_S the total days fished by the standard vessel during the period in which both the standard vessel and vessel i were in the fishery.

The standardized catch rate for year t, I_t, is then defined as

$$I_t = \frac{\sum_i C_{t,i}}{\sum_i (\mathrm{RFP}_i\, E_{t,i})} \tag{4}$$

where $C_{t,i}$ is the catch by vessel i in year t, and $E_{t,i}$ the number of days fished by vessel i in year t.

Although straightforward to apply, the approach of Beverton and Holt (1957) does not generalize easily to deal with multiple factors such as month and area, and when there is no fishing vessel that has been in the fishery for many years and can be used as the standard vessel. Finally, it is not straightforward to determine the precision of the standardized catch rate estimates; this information is, however, needed when applying many of the methods of stock assessment.

More recent methods for standardizing catch and effort data involve fitting statistical models to the catch and effort data. The first examples of these methods were by Gavaris (1980) and Kimura (1981). However, the last two decades have seen a proliferation of new methods to standardize catch and effort data, most of which extend these methods to some extent. The choice among these methods should be based on an evaluation of the underlying assumptions of the models and use of appropriate statistical tests and diagnostics. Understanding of the fishery being modeled may also provide insight into which method should be used. For example, many fishery systems are inherently nonlinear, and methods that can handle nonlinear relationships between catch rate and potential variables that capture changes over time and space in catchability may be more appropriate.

This paper reviews many (but, by no means, all) of the decisions that must be made when standardizing catch and effort data. The following sections highlight various issues related to the choice of the model on which to base the analyses and the data set to be considered, focusing, in particular, on model selection and how to deal with records for which the effort is non-zero, but the catch is zero.

Most of the catch-effort standardizations on which actual assessments are based are documented in the gray literature. We have attempted in this review, as much as possible, to restrict the examples to cases in which the basic documents are fairly readily available.

2. Basic methods

2.1. Generalized linear models

Generalized linear models (GLMs; Nelder and Wedderburn, 1972) are the most common method for standardizing catch and effort data. Gavaris (1980) appears to have been the first to have used a GLM approach to standardizing catch and effort data when he extended the use of multiplicative models for this purpose (Robson, 1966) by explicitly assuming log-normal errors. Gavaris (1980) applied an analysis of variance (ANOVA) model (only categorical explanatory variables) to the natural logarithm of CPUE, assuming Gaussian error with a constant variance (equivalent to least-squares estimation) and independence among the observations. Kimura (1981) extended this approach by including both categorical and continuous explanatory variables.

GLMs are defined by the statistical distribution for the response variable (usually, but not always, catch rate) and how some linear combination of a set of explanatory variables relate to the expected value of the response variable. The key assumption of a GLM is therefore that the relationship between some function of the expected value of the response variable and the explanatory variables is linear:

$$g(\mu_i) = \mathbf{x}_i^{\mathrm{T}}\beta \tag{5}$$

where g is the differentiable and monotonic link function, $\mu_i = E(Y_i)$, $\mathbf{x}_i$ the vector of size m that specifies the explanatory variables for the ith value of the response variable,[1] β is a vector (of size m) of the parameters, and Y_i the ith random variable.

GLMs are a very general and powerful statistical technique that include, as special cases, Gaussian linear models (ANOVA, regression), log-linear models for frequency data, logistic regression models, and several others. In order to apply a GLM, it is necessary to: (a) choose the response variable, (b) select a sampling distribution for the response variable from the exponential family (e.g. normal, exponential, Poisson, binomial, gamma), (c) chose a link function appropriate to the distribution, and (d) select a set of explanatory variables. Year must be one of the explanatory variables because the primary objective of standardizing

[1] Includes the intercept by setting $x_1 = 1$ for all i.

catch and effort data is to detect trends over time in abundance. This cannot be achieved unless year is included in the model (regardless of whether or not it is statistically significant).

The choice of a statistical distribution for the response variable (often also referred to as the 'error model') should take account of the nature of the process that generated the data being modeled. For example, a discrete distribution, such as the Poisson or the negative binomial, may be the most appropriate distribution if the catch is recorded in individuals. In this case, the catch, rather than the catch rate, should be the response variable, and the fishing effort should be included as an offset (i.e. effort is added to the outcome from the linear predictor in Eq. (5)) or perhaps as an explanatory variable. However, a continuous distribution may be more appropriate if the catch is in weight, catch rate is modeled, or a large number of individuals are usually caught by each unit of effort (e.g. if many units of effort are combined into a single datum).

Natural (canonical) link functions exist for each of the distributions in the exponential family. For example, the default link function for the normal distribution is the identity function, while that for the binomial distribution is the logit function:

$$\ln\left(\frac{\mu_i}{1-\mu_i}\right) = \mathbf{x}_i^{\mathrm{T}}\beta \tag{6}$$

Nonlinearity in the relationship between the dependent and explanatory variables can be included in GLMs through the link function, through interaction terms, and by transforming the explanatory variables. For example, nonlinearity can be taken into account by including explanatory variables raised to various powers:

$$g(\mu_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,1}^2 + \beta_3 x_{i,1}^3 + \cdots \tag{7}$$

where $x_{i,1}$ is a continuous explanatory variable that might be related in a nonlinear way to the response variable.

Raising covariates to various powers should, however, be used with care and, in general, high-order terms should be avoided unless absolutely necessary. We prefer the alternative of discretizing continuous explanatory variables and treating them as categorical variables, to raising them to various powers. However, both approaches have been used in practice. For example, Ayers (2003) modeled the effects of the day of the year on the catch rate of gemfish (*Rexea solandri*) off New Zealand, using a third-order polynomial, whereas Punt et al. (2001a) modeled the impact of week on catch rates of blue grenadier (*Macruronus novaezelandiae*) off southern Australia, treating each week of the year as a categorical explanatory variable.

There are several reasons that GLMs remain the most commonly applied method for standardizing catch and effort data. These include the availability of well-tested and user-friendly software to perform the calculations. Also, some special cases of the GLM approach (e.g., multiple linear regression) have a long history in quantitative fisheries science.

2.2. GAMs and GLMMs

Although the great majority of the catch-effort standardizations have been (and still is) based on the application of the GLM approach, two other estimation frameworks have also been applied.

Generalized additive models (GAMs; Hastie et al., 2001) are extensions of generalized linear models that involve generalizing Eq. (5) by replacing the linear predictor by an additive predictor:

$$g(\mu_i) = \mu + \sum_{j=1}^{p} f_j(\mathbf{x}_i) \tag{8}$$

where f_j is a smooth function (such as a spline or a loess smoother).

The degree of smoothness achieved is balanced against the deviance by a tuning constant, often chosen by cross-validation, so that estimation is by the method of maximum *penalized* likelihood rather than of maximum likelihood. This gives GAMs a partially non-parametric aspect.

Bigelow et al. (1999) used a GAM approach to model the catch rate of swordfish (*Xiphias gladius*) and blue shark (*Prionace glauca*) in the North Pacific, and found highly-nonlinear relationships between, for example, latitude and longitude and catch rate. Rodríguez-Marín et al. (2003) used a GAM approach to determine whether the catch rates of bluefin tuna (*Thunnus thynnus*) in the Bay of Biscay varied with longitude and latitude.

Generalized linear mixed models (GLMMs; Pinheiro and Bates, 2000) extend the GLM approach by allowing some of the parameters in the linear predictor to be treated as random variables. Several recent analyses of catch and effort data (e.g. Chang, 2003; Miyabe and Takeuchi, 2003; Rodríguez-Marín et al., 2003; Brandão et al., 2004; Ortiz and Arocha, 2004) treated some of the parameters as random effects. In general, random effects have been introduced into models to deal with interactions between year and other categorical variables, such as area (see Section 5 for further details).

A full discussion of GLMs, GAMs, and GLMMs, and their advantages and disadvantages, is beyond the scope of this paper. Interested readers can consult standard textbooks (e.g., McCullagh and Nelder, 1989; Dobson, 1991; Hastie et al., 2001), the review by Guisan et al. (2002), and a special issue of *Ecological Modelling* (Vol. 157, Issues 2 and 3) for more information about generalized linear and generalized additive models, their uses, and applications.

2.3. Extracting the year effect and determining its precision

Most methods used to standardize catch and effort data estimate a year effect on which an index of abundance can be based. The year effect usually reflects changes in annual abundance, but there is no reason that a finer, or coarser, temporal resolution cannot be considered when modeling catch and effort data. For example, Maunder and Harley (2003) estimate quarterly standardized catch rate indices for the use in tuna assessments for the eastern Pacific Ocean.

The year effect must be extracted from the model. This is straightforward for models based on the log-transformed dependent variable and a normal distribution error model, since the year effect is simply $\exp(\hat{\alpha}_t + \hat{\sigma}_t^2/2)$, where $\hat{\alpha}_t$ is the estimate of the year factor for year t and $\hat{\sigma}_t$ the standard error of α_t. However, extracting the year effect is more complicated for some of the other models. For example, the year effect for year t from the delta-log-normal approach (see Section 3.3) is the prediction of the probability of a non-zero catch during year t multiplied by the prediction of the catch rate for year t, given that the catch is non-zero (Lo et al., 1992; Vignaux, 1994). The value of the year effect depends on the values of some of the explanatory variables for some error models. For example, when applying a delta approach, the probability of a non-zero catch is

$$p_t = \frac{\exp\left(\beta_0 + \alpha_t + \sum \beta_i x_{i,t}\right)}{1 + \exp\left(\beta_0 + \alpha_t + \sum \beta_i x_{i,t}\right)} \tag{9}$$

where p_t is the probability of a non-zero catch during year t, β_0 the intercept, β_i the parameter for the ith explanatory variable, and $x_{i,t}$ the ith explanatory variable for year t.

A potential problem with Eq. (9) arises in that each explanatory variable will have multiple levels (for categorical variables) and values (for continuous variables) during year t, making specification of $x_{i,t}$ a requirement for extracting the year effect. Common ways to overcome this problem include setting continuous variables to their means (or medians) over the year concerned (or over the entire data set). For categorical variables, the value used when applying Eq. (9) can be any value, but is usually the most common value in the data set (e.g. Punt et al., 2000a), or an average over all values (weighted by the relative frequency of each value in the data set). Alternatively, Eq. (9) can be applied setting the values for the explanatory variables to their highest and lowest values to determine the sensitivity of the results to the choice of these values.

The precision of the year effect can be calculated analytically (for some models) or using the delta method. However, it is not straightforward to compute the precision of the index for delta models (see Section 3.3). Vignaux (1994) used a bootstrapping approach to determine the precision of the year effect from a delta-log-normal model, while Ralston and Dick (2003) used a jackknife method. Both of these methods can be highly computationally intensive, owing to the need to repeat the GLM analysis many times.

Categorical variables are over-parameterized in GLMs if an intercept is estimated. The most common approach to overcome this problem is to fix the value of one of the parameters (usually to zero) and estimate the values for remaining parameters given the fixed value. The year effect is usually treated as a categorical variable with the consequence that one of the year effects (that for the 'base year') is set to zero and hence has no associated variance estimate. The variance estimate for a 'non-base year' therefore relates to the difference between the value of the year effect for the 'non base year' and that for 'base year'.

2.4. Summary of decisions

The previous three sections overviewed the basic estimation frameworks. However, use of these methods to standardize catch and effort data necessitates making several decisions (e.g. how to handle zero catches, which explanatory variables to consider and which to include in the final model, and diagnostic statistics). The following sections deal with these issues.

3. Dealing with zero catches

Catch and effort databases often include high proportions of records in which the catch is zero, even though effort is recorded to be non-zero (records in which effort is recorded to be zero must be either trivial if they have zero catch as well, or in error and this should be resolved in some way (e.g. discarded) prior to any analyses being conducted). This is particularly the case for less abundant species and for bycatch species. Unfortunately, these species are often those for which a standardized catch rate index is the most important (or the only) source of data on the changes in abundance (e.g. Ortiz and Arocha, 2004). The presence of many zeros can invalidate the assumptions of the analysis and jeopardize the integrity of the inferences if not properly modeled (Lambert, 1992). The zeros can also lead to computational difficulties. For example, zero catches cause computational problems for the standard log-linear approach because the natural logarithm of zero is undefined.

Zero observations can arise for many reasons:

(a) sampling a rare species or a species with low vulnerability to the gear;
(b) defining effort so that the capture of an individual is a rare event (e.g. by defining each unit of recreational effort to be a single hang);
(c) sampling a species that schools or aggregates (in this case, there may be a high probability of a zero catch, but, when a school is encountered, the catch rate may be very high);
(d) malfunctions of the gear; and
(e) recording zeros (fishers may record only the main target species if many species are caught, if the reporting form has a limited number of entries, or simply if they do not consider recording all species as being important).

It is desirable to attempt to account for the process that caused the zeros when standardizing catch and effort data. For example, zero observations caused by gear malfunctions can be detected in data by the catch for all species being zero; such records should be discarded prior to the analysis.

The simplest way to deal with zero observations is to ignore them. This might be appropriate if zero catches occur only because of gear failure or because fishing occurred in circumstances under which it would be impossible to catch the species of interest. Zero catches can also be eliminated or reduced by combining records from the same stratum (e.g. by aggregating hauls within a day to daily records or daily records into monthly records). However, this may not be sufficient to eliminate all the zero catches. Furthermore, grouping observations will result in a loss of information, and may bias the analysis. For example, if each record is associated with a different value of some continuous explanatory variable, aggregating data across records will necessitate some form of averaging over this variable (or, worse still, ignoring it).

The simplest alternative to ignoring zero catches or aggregating data is to replace zero catches by a small number, either by direct substitution or by adding a small constant to each catch. Butterworth (1996) suggests choosing this constant to give the most normal-like distribution of residuals, usually by selecting for zero skewness. Cooke and Lankester (1996) note that adding a constant to each observation leads to the series of standardized catch rates depending on the levels of the explanatory variables chosen as the standard ones. They suggest that the constant should be added to both the observed values and the predicted values when using a log-transformed dependent variable and a normal error model. Alternatively, Butterworth (1996) suggests that if an analysis is conducted in which a constant is added to all of the catch rate observations, the abundance index should be based on selecting a standard set of values for the explanatory variables and exponentiating the year effect plus the parameters multiplied by these standard values, and then removing the assumed constant from the exponentiated value, i.e. if a model of the form $\ln(\mathrm{CPUE}_{t,i} + \delta) = \alpha_t + \beta x_i + \varepsilon_{t,i}$ is fitted to a set of data, the year factor should be $I_t = \exp(\alpha_t + \beta \bar{x}_t) - \delta$ (which can, of course, be negative).

Adding a constant to each catch value may not necessarily be the most appropriate way to model the data because the results of a catch-effort standardization may be sensitive to the value of the constant. Despite the availability of objective methods for determining the size of the constant (Berry, 1987; Porch and Scott, 1994), the value of the constant is usually chosen arbitrarily. Other simple methods, such as ignoring all zero observations, may also have undesirable consequences such as positively biasing the standardized catch rates, but to a different extent each year. Fortunately, there are more appropriate methods to deal with zero catches. These methods fall into three categories: (a) statistical distributions that allow for zero observations; (b) methods that inflate the expected numbers of zeros; and (c) methods that predict the proportions of positive catches and model the catch rate when the catch is non-zero separately (the delta approach). Zero-deflated distributions are also possible, but seldom arise in practice (Ridout et al., 1998).

Historically, ignoring zero observations or replacing them by a constant was the most common approach. Probability distributions that allow for zero observations have been used in some cases (NRC, 1994) but, currently, the most popular way to deal with zeros is through the delta approach (e.g., Lo et al., 1992; Vignaux, 1994; Ayers, 2003).

3.1. Models for count data

Catch data are often recorded as counts of individuals, for which several statistical distributions are appropriate. These distributions explicitly allow for zero counts and model integer values. Catch in weight or catch rate can be converted into an appropriate form by rounding the data to the nearest integer, but the use of continuous distributions may be more appropriate for these types of data. Mullahy (1986) notes that the interest in modeling count data explicitly is due to the recognition that the use of continuous distributions to model integer outcomes might produce inconsistent parameter estimates.

The standard distribution for modeling count data is the Poisson distribution. This distribution assumes that the encounter rate of individuals is constant, with the variance being equal to the mean. However, actual count data are often overdispersed relative to the Poisson distribution (e.g. Bannerot and Austin, 1983; Punt et al., 2000a). In this case, it may be more appropriate to model the data using the negative binomial distribution, which allows for a quadratic relationship between the mean and the variance (i.e. $\mathrm{var}(Y) = \mu + \mu^2/k$, where k is a parameter to be estimated; e.g. Punt et al., 2000a).

3.2. Zero-inflated models

The proportion of zeros in the Poisson and negative binomial distributions is related to the distribution for the non-zero values (i.e. for a given distribution of non-zero observations there is only a single possible proportion of zeros). However, if the processes that lead to zero observations are not the same as those that lead to non-zero catches (e.g. gear malfunction, whether the species under consideration is being targeted), zero-inflated distributions may be more appropriate (Lambert, 1992; Hall, 2000). These distributions are a mixture of two distributions, a degenerate component that is zero with certainty and a second component that includes zeros and positive values (e.g., the Poisson distribution). The general form of these distributions is

$$Pr(Y = y) = \begin{cases} w + (1 - w) f(0), & y = 0, \\ (1 - w) f(y) & \text{otherwise} \end{cases} \qquad (10)$$

where w is the probability that an observation comes from the degenerate component.

The parameters to be modeled as functions of the explanatory variables are then the probability of a zero observation, w, and (usually) the mean of the second distribution defined by $f(y)$. In principle, the processes causing the zero catches may be the same as those leading to the distribution of positive values, so that the values for the parameters of these two models may be functionally connected (Lambert, 1992). However, it is more likely that the processes causing the additional zeros are different, and should be modeled separately. Two commonly used zero-inflated distributions are the zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB).

3.3. Delta approaches

An alternative to using zero-inflated models is to model the probability of obtaining a zero catch and the catch rate, given that the catch is non-zero, separately. These models have also been termed hurdle models

(Cragg, 1971), because a hurdle must be overcome before a positive observation occurs. If the realization is positive, the hurdle is crossed. The conditional model of the positives is governed by a standard distribution that is defined for positive values. For example, one minus the probability of a zero catch could be considered to be the probability of encountering a school, while the distribution of the positive values is the probability distribution of the school size. The general form of the delta model is

$$\Pr(Y = y) = \begin{cases} w, & y = 0, \\ (1-w)f(y) & \text{otherwise} \end{cases} \tag{11a}$$

Note that the two distributions, implied by Eqs. (10) and (11a), are formally identical in that, if

$$w^* = 1 - (1-w)(1-f(0)) \quad \text{and}$$
$$f^*(y) = \begin{cases} 0, & y = 0, \\ \frac{f(y)}{1-f(0)}, & y > 0 \end{cases} \tag{11b}$$

then Eq. (10) reduces to Eq. (11a), with w replaced by $w*$ and $f(y)$ replaced by $f^*(y)$. What are different are the parameters to be modeled: in the latter, w is the probability of a zero observation, whereas, in the former, it is the probability of an 'extra' zero.

A binary random variable is zero for a zero observation and unity otherwise, and has, by definition, a Bernoulli distribution with probability parameter, w. The probability of obtaining a zero observation is therefore usually modeled using the binomial distribution (e.g. Vignaux, 1994; Stefánsson, 1996; Punt et al., 2000a; Rodríguez-Marín et al., 2003). A variety of distributions could be used to model the catch rate given that it is non-zero. The most commonly selected distribution is the log-normal (Aitchison, 1955), as in Vignaux (1994) and Porter et al. (2003). The use of this distribution has, however, been criticized for a lack of robustness (Myers and Pepin, 1990; Syrjala, 2000). Other distributions considered when applying the delta approach are the gamma distribution (Cooke and Lankester, 1996; Punt et al., 2000a), the Poisson distribution (e.g. Ortiz and Arocha, 2004), and the negative binomial distribution (Punt et al., 2000a).

4. Selecting explanatory variables

The main goal in standardizing catch and effort data is to explain the variation in catch rate that is not a consequence of changes in population size by identifying explanatory variables that reduce the unexplained variability in the response variable. Both qualitative and quantitative variables can be included as explanatory variables in most methods. Qualitative variables are treated as factors while quantitative variables can be treated either as ordered values and used in functions, or discretized and treated as factors. In principle, the fraction of the variability explained can increase substantially by including more and more explanatory variables. However, adding explanatory variables will generally reduce bias but increase the variance of the index of abundance. This section outlines the steps commonly followed to decide on which explanatory variables should be included when standardizing catch and effort data.

4.1. Choosing explanatory variables to consider

The first step is to determine which explanatory variables are available and which of these should be considered. Explanatory variables should, however, be considered in an analysis only if there is an a priori reason that they may influence catchability. The year effect should, of course, always be included in the model, even if not statistically significant, because it is the quantity of interest. There are often many explanatory variables. For example, Horn (2003) considered 23 possible explanatory variables when standardizing the catch and effort data for ling (*Genypterus balcodes*) off New Zealand. These variables included the type of trawl gear, the time of day when the trawl was used, the vessel call sign, the characteristics of the vessel (length, breadth, etc.), and an environmental factor (the southern oscillation index). Other variables considered routinely when standardizing catch and effort data include area, month, and the catch (or catch rate) of species other than those under consideration. Including the catches (or catch rates) of other species (e.g. Punt et al., 2001a) is meant to be a way of including the impact of fishers targeting species other than that under consideration in multi-species fisheries. However, if the other species are closely related to the species of interest and are being fished down at the same time, the inclusion of these other species as explanatory variables may remove time trends in catch rate which should be attributed to the year effect.

In some cases, the decision about which variables to include may be quite subtle. For example, should each vessel be included as a categorical variable (e.g. Punt et al., 2000a; Battaile and Quinn, 2004) or should the characteristics that define a vessel be included as explanatory variables (e.g. Vignaux, 1996). This choice depends on how the vessel is thought to influence catch rate and what information is available. For example, if the captain is thought to be the main influence on catchability through differences in skill and targeting practices, then a vessel effect may be more appropriate (Punt et al., 2000a).

Inclusion of explanatory variables that are themselves correlated, the so-called problem of 'collinearity', should be avoided. This can make the model fitting process numerically unstable or lead to problems similar to those of over-fitting. For example, the length of a vessel is almost invariably highly correlated with its breadth, and including both of these quantities as explanatory variables will not contribute very much more to the predictive ability of the model than just one of them, or some simple function of the two. One way to avoid the problem of using correlated explanatory variables in an analysis is to examine the explanatory variables prior to conducting the actual analysis and to identify a set of explanatory variables that are not highly correlated. However, the presence of collinearity is less of a problem if the goal of the analysis is to generate an index of relative abundance for use in a stock assessment model rather than to determine the variables that explain variation in CPUE. This is because additional variables that are correlated tend not to add much explanatory power beyond the first variable selected, so that they are rejected in a stepwise selection procedure.

If catch is used as the dependent variable rather than catch rate, the measure of effort should be included in the analysis as an explanatory variable or as an offset. In fact, multiple effort measures can be included as explanatory variables, thereby allowing the data to choose the most appropriate measure of effort.

4.2. Model selection

There are two general categories of methods to determine which explanatory variables should be included in an analysis: (a) methods based on the fit to the data that include a penalty based on the number of parameters estimated; and (b) methods that compare model predictions to observations. The difference between these methods is that the first category uses all of the data set to estimate the model parameters, while the second group generally uses only part of the data set to estimate the parameters and the remaining part to test predictions.

Standard hypothesis testing methods (e.g. F-tests, likelihood ratio tests, and score tests) can be used to compare a more complicated model (i.e. more parameters) to less complicated models (McCullagh and Nelder, 1989; Hilborn and Mangel, 1997). These methods can be automated to select the 'best' model. However, they can be applied straightforwardly only to nested models. In contrast, information-theoretic methods, such as the Akaike information criterion (AIC; Akaike, 1973; Burnham and Anderson, 2002) and the Bayesian information criterion (BIC; Schwarz, 1978), are methods that can also be applied to non-nested models:

$$\begin{aligned} \mathrm{AIC} &= -2\ln \ell + 2p, \\ \mathrm{BIC} &= -2\ln \ell + p\ln(n) \end{aligned} \tag{12}$$

where ℓ is the likelihood function evaluated at its maximum, p the number of parameters, and n the number of observations.

As the number of data points increases, it becomes more difficult to accept additional parameters under the BIC than under the AIC. For categorical variables, it is possible to test a single level in a categorical variable (one parameter) or the categorical variable as a whole (the number of parameters is equal to the number of levels in the categorical variable minus 1).

One 'problem' with both standard hypothesis testing methods and information-theoretic approaches is that most sets of catch and effort data consist of thousands (or tens of thousands) of points. One consequence of this is that AIC or BIC may identify a model with an enormous number of explanatory variables. A commonly-applied (but ad hoc) way to deal with this (e.g. Horn, 2003) is to add explanatory variables only

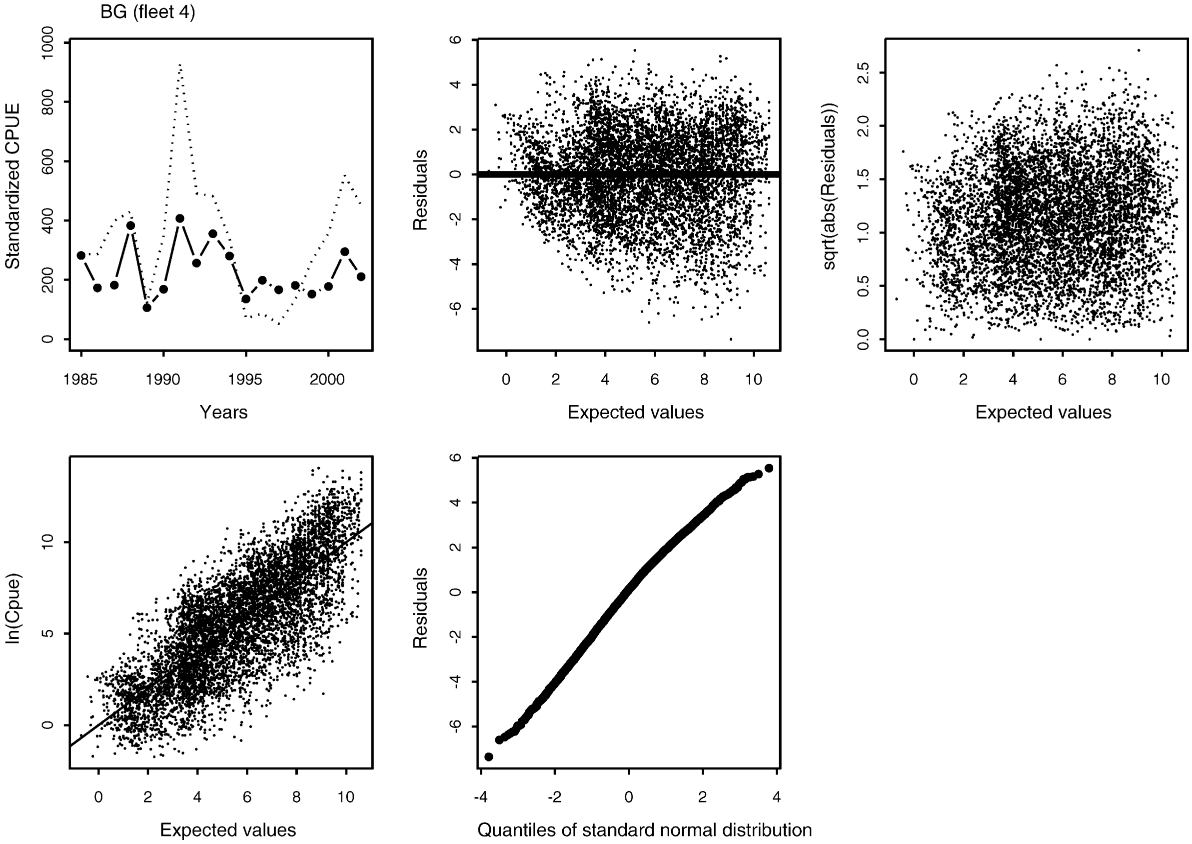

Fig. 2. Diagnostic plots for the goodness-of-fit of a log-linear model to the catch and effort data for the blue grenadier (*M. novaezelandiae*) off the west coast of Tasmania, Australia. The dotted line in the upper left panel is the geometric mean catch rate, and the solid line is the standardized catch rate index.

if the deviance is reduced (or R^2 is increased) by a pre-specified percentage (e.g. 0.5 or 2%).

Standard regression diagnostic statistics (e.g. Fig. 2) should also be examined to attempt to identify model misspecification and heteroscedascity (McCullagh and Nelder, 1989; Ortiz and Arocha, 2004). Fig. 2 shows a variety of diagnostics, including standardized residuals versus the fitted values (to assess whether model misspecification is occurring), the square root of the absolute values of the standardized residuals versus the fitted values (to assess whether the variance changes as a function of the predicted value—it should not in this case), the observed versus the predicted values (to assess qualitatively whether the explanatory variables are indeed able to reduce the variance in the data), and quantile–quantile plots. Fig. 2 shows only a small subset of the possible diagnostic statistics. Other statistics, such as those to identify outliers, should also be examined routinely when analyzing catch and effort data.

Cross-validation is also a commonly used method to select explanatory variables (Hastie et al., 2001). This involves using a 'training data set' (a subset of the total data set) to estimate the parameters of the model and using the resulting model to predict the remaining data (the 'test data set'). The ability of the model to predict the test data set is used to select the explanatory variables to include in the model. If too many explanatory variables are selected, the model fits the training data set very well, but is fitting noise rather than signal, and therefore cannot predict the test data set well. If too few explanatory variables are used, the model does not adequately mimic the data, and fits both the test and training data sets poorly. There are several versions of cross-validation, but a popular one that makes more use of the data than simple cross-validation, is k-fold validation (Hastie et al., 2001). The data are divided into k equal parts and the model is run k times, each time rotating through each of the k subsets as the test data set and using the remaining data as the training data set.

One problem with cross-validation is that it does not necessarily parallel the likelihood inference that is often used to estimate the model parameters. This is because a test criterion is required, and simple least squares are often used (Hastie et al., 2001). However, the likelihood function may differ from the least-squares criterion.

4.3. Error structure assumptions

Often, relatively little attention is paid to determining whether the distributional assumptions are actually valid (e.g. the residuals from a log-linear regression are normally distributed). Selection among statistical models can be carried out by examining the relationship between the average catch rate and the variance in catch rate (Dong and Resterpo, 1996; Punt et al., 2000a) but note that Dick (2004) finds this a weak method for selecting among statistical models. A linear relationship supports an overdispersed Poisson error model, and variance in catch rate proportional to the square of the average catch rate suggests the log-normal and gamma error models. The negative binomial error model implies that the variance in catch rate is a function of both the average catch rate and the square of the average catch rate. The choice of the error model is often dependent on the data points at high catch rates, which may be few in number (e.g. Punt et al., 2000a). In addition, the catch rates often have downward trends over time, and this trend may inflate the variance in catch rates (Punt et al., 2000a). Methods such as quantile–quantile plots (Fig. 2) can also be used following the application of a model to determine whether the residuals are consistent with the assumed error model. Dick (2004) shows that AIC can be used to select among error structure assumptions.

It is possible to allow the variance of the residuals to depend on explanatory variables. For example, Butterworth (1996) modeled the variance of the residuals about a log-linear regression as $(\alpha + \beta/E)^2$, where the values for α and β were obtained using iterative re-weighting.

Standard regression diagnostics often identify potential outliers. However, it is our experience that the presence of a small number of outliers will not affect the trends in catch rate substantially, primarily because catch-effort data sets are commonly so large that a few points can have only a limited impact on the final outcomes. In cases where outliers are considered problematic, robust regression can be used to reduce the influence of the outliers (Rousseeuw and Leroy, 1987).

5. Dealing with interactions

Interactions among factors occur fairly regularly when standardizing catch and effort data. The most common interactions are among year, month/week/day and area. Discovering significant (and substantial) interaction terms can raise some interesting hypotheses. For example, discovering a year × vessel interaction implies that the relative abundance has changed differently as seen by different fishers. However, explaining interactions can be difficult, and there is often no rational explanation for some interactions. A year × vessel interaction can be caused by several factors: different vessels targeting different size classes of fish or fishing in different areas, and some fishers having upgraded their equipment and others not having done so. In the latter case, it may be possible to eliminate the interaction by including additional explanatory variables (such as the equipment used on a vessel; see Robins et al. (1998), Bishop et al. (2000) and Rodríguez-Marín et al. (2003) for examples of catch-effort standardizations that included factors for skipper experience and the use of aids such as the Global Positioning System and plotters).

Identifying significant interactions with year means that it is no longer straightforward to use the year factor as the basis to develop an index of abundance. There are a variety of approaches for dealing with interactions when constructing an index of abundance. The first approach is to explicitly ignore any interactions between year and other explanatory variables (e.g. Vignaux, 1994). This approach avoids consideration of the problem, but may lead to a biased index of abundance if substantial interactions with year are present.

If an interaction between year and a factor (say month) is found, an appropriate way to develop the index of abundance (considering here the simplest case of a log-linear regression approach) is to average the year × month interaction terms over the year:

$$I_t = \sum_m z_m I_{t,m} \tag{13}$$

where I_t is the relative abundance index for year t, $I_{t,m}$ the index of abundance for year t and month m (calculated from the relevant year × month interaction term), and z_m the weighting factor for month m (where $\sum_m z_m = 1$).

An appropriate choice for z_m could be 1/12 if the fishery occurred over the entire year (D.S. Butterworth, pers. commun.).

Dealing with year × area interactions is not as straightforward as dealing with year × month interactions because it is difficult to define the 'area' which would be needed to apply Eq. (13). One way to do this is to define a 'habitat area' for each area included in the analysis and use that in the weighting scheme (e.g. Quinn et al., 1982; Punt et al., 2001b; Campbell, 2004). Punt et al. (2001b) weighed the year × area interactions by the physical area between 20 and 80 m depth of each area in the analysis when standardizing catch and effort data for gummy shark (*Mustelus antarcticus*) off southern Australia.

Another approach to dealing with year × area interactions is to recognize that these interactions imply different trends in abundance in different areas, which, in turn, implies some form of spatial structuring of the population. Punt et al. (2000a) standardized the catch and effort data for the school shark (*Galeorhinus galeus*) off southern Australia, and noted different trends in standardized catch rate in different areas (Fig. 3). Rather than attempting to pool the indices across areas in some way, Punt et al. (2000b) assessed this population using a population dynamics model that was spatially structured. Developing spatially-structured population dynamics models is not possible in all cases because these models rely on information, say, on movement.

If a year × area interaction is assumed to have arisen because of the random changes in the distribution of the population (unlike those in Fig. 3), it is possible to assume that the year × area interactions are random effects and use a generalized linear mixed model to standardize the data (e.g., Chang, 2003; Miyabe and Takeuchi, 2003).

The use of Eq. (13) implies that year × area interaction factors are available for all combinations of year and area. However, this is seldom the case (e.g., if the fishery developed in one area and expanded spatially thereafter). The solution to this problem (Campbell, 1998, 2004; Punt et al., 2000a; Walters, 2003) is to develop algorithms for specifying the missing year × area interactions. The algorithms should use information for the year × area combinations with data to interpolate (and perhaps extrapolate) to the remaining combinations. Unfortunately, the resulting index of abundance (and hence the results of any subsequent stock assessment) may be highly sensitive to the algorithm chosen for interpolation and extrapolation (Campbell, 1998; Butterworth et al., 2003).

6. Selecting data points

The bulk of the world's marine fish species are caught in fisheries that involve multiple target species. This is particularly true for species caught in trawl fisheries and those caught recreationally. Given the requirement to standardize the catch and effort data for a species that is caught in a multi-species fishery, it seems desirable to use only the effort that was directed at that species. Unfortunately, this is much easier said than done, even when fishers claim to record their target species in logbooks. In many cases, the fisher may just record the most prevalent species as the target species.

The most common way to deal with this problem is to base the catch-effort standardization on the records for those fishers who appear to target the species under consideration. Fishers can be chosen using criteria selected by experts (e.g., assessment scientists and fishers). These criteria can include a minimum number of years in the fishery, a minimum number of records, and a minimum average (or median/total) catch. For example, Taylor (2003) standardized the catch and effort data for orange roughy (*Hoplostethus atlanticus*) based on data for vessels that had 20 or more positive catches in at least 4 years, while Punt et al. (2001b) based their standardization of the catch and effort data for gummy shark on vessels that satisfied four criteria (in the fishery for at least 5 years, a median annual catch (all sharks) of at least 10 metric tons (t), a median annual catch (gummy shark) of 5 t, and gummy shark constituting more than 60% of the total shark catch). Both Punt et al. (2001b) and Taylor (2003) examined the sensitivity of their results to the choice of criteria for selecting vessels.

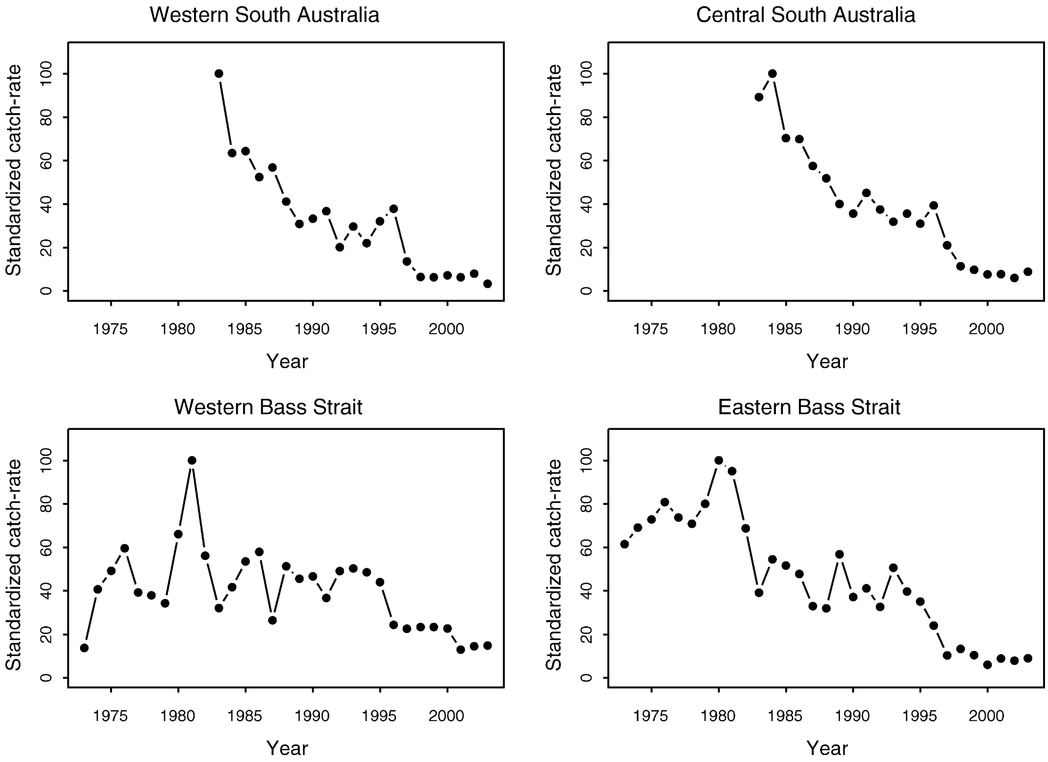

Fig. 3. Standardized catch rate indices for school shark in four regions off southern Australia.

An alternative method to overcome the problem of defining target effort is to define the characteristics of target effort. These characteristics could include the depth or area fished, or the other species caught. For example, Stephens and MacCall (2004) used this method to develop a catch rate index for bocaccio rockfish (*Sebastes paucispinis*) from data for recreational party boat fishing trips off California. These data do not contain information on where fishing occurred. Furthermore, this recreational fishery is directed at a wide range of species, including some, such as pelagic tunas, that occur where it would be impossible to catch bocaccio rockfish. Stephens and MacCall (2004) developed a method that associate the presence of bocaccio rockfish in the catch of a trip with the other species caught during the trip and use this to select which trips to use in a catch-effort standardization.

The preceding discussion has focused on excluding records that are likely to have been directed at species other than that under consideration. Catch-effort records should also be rejected for use in an analysis if they contain obvious errors. Typical errors include: (a) missing values (e.g. on location, catch, etc.), (b) very long (or short) tow duration (or large number of sets), (c) unrealistically long nets, and (d) unrealistically high catches or catch rates. Although most catch-effort data sets are sufficiently large that the occasional outliers should not affect the final estimates, it is necessary that the analyst check the data for obvious errors and remove them. If a large number of records are rejected because they are missing data for a single explanatory variable, it may be better to retain the data and ignore the explanatory variable, particularly if it explains only a small proportion of the total variability.

7. Using indices of abundance in stock assessment models

The primary reason for standardizing catch and effort data is to develop an index of relative abundance. This can be used as the basis for management advice directly, but is typically used when fitting a stock assessment model. To use an index of relative abundance estimated from catch and effort data in a stock assessment, the index of abundance must first be extracted from the standardization analysis (see Section 2.3) and then an appropriate fitting method, usually a likelihood function, must be selected.

There are two approaches to include an index of relative abundance in a stock assessment:

(a) the index can be assumed to be proportional to abundance (e.g. Butterworth and Andrew, 1984; Maunder and Starr, 2003) and the difference between the observed and model-predicted indices attributed to observation error; or
(b) the index can be used to derive 'standardized effort' (essentially by dividing the catch by the index) and the 'standardized effort' included as a model input to predict the annual catch (e.g. Schnute, 1977; Fournier and Archibald, 1982; Deriso et al., 1985; Fournier et al., 1998; Dichmont et al., 2003).

The first is used far more frequently than the second. Specifically, most age- and size-based approaches to fisheries stock assessment (e.g. Methot, 1993, 2000; Punt and Kennedy, 1997; Hilborn et al., 2000) treat the indices of relative abundance this way. The most common assumption about observation error is that it is log-normal:

$$I_t = qN_t \exp(\varepsilon_t), \quad \varepsilon_t \sim N(0, \sigma_t^2) \tag{14}$$

where q is the catchability coefficient, N_t the model estimate of the abundance during year t, and σ_t the standard deviation of the observation error for year t.

The definition of the abundance to which the index relates is the key to an appropriate use of a standardized catch rate index in a stock assessment. This is because different components of the population may exhibit different trends over time. Abundance can be defined in terms of biomass or numbers, depending on whether catch is measured in numbers or in weight, and the abundance is usually population numbers (or biomass) modified by the age- or size-specific selectivity of the gear used in the fishery.

The likelihood function for the index of abundance[2] is therefore:

$$L(\underline{I}|\underline{\theta}) = \prod_t L(I_t|\underline{\theta}) = \prod_t \frac{1}{\sqrt{2\pi}\sigma_t I_t} \times \exp\left(\frac{[\ln I_t - \ln(qN_t)]^2}{2\sigma_t^2}\right) \tag{15}$$

where $\underline{\theta}$ is the vector of the parameters of the stock assessment model.

There are several ways to specify the observation error standard deviations:

(a) assume that σ_t is independent of time, i.e., $\sigma_t = \sigma$ (e.g. Butterworth and Andrew, 1984);
(b) assume that σ_t is the product of a year-specific standard deviation, w_t, and an overall variance scaling parameter, σ, i.e., $\sigma_t = w_t\sigma$ (e.g. Cope et al., 2003; Maunder and Starr, 2003); and
(c) assume that σ_t^2 is the sum of a year-specific variance, w_t^2, and an additional variance term, σ^2, i.e., $\sigma_t = \sqrt{w_t^2 + \sigma^2}$ (e.g. Francis et al., 2003; Punt and Butterworth, 2003).

The parameter σ is either pre-specified based on a priori considerations (e.g. Francis et al., 2002) or is treated as a parameter of the model to be estimated. The values of w_t need to be pre-specified based, say, on the results of the catch-effort standardization (but see Francis, 1999; Maunder and Starr, 2003 for the correct interpretation of σ_t^2 from a GLM standardization). In general, the variability about the year factors is considerably less than that between the index of abundance and the model predictions (e.g. Cope et al., 2003), so that the value of σ is usually much greater than unity when using method (b) and much greater than w_t when using method (c). The reasons for this include that catchability may vary among years (so all the catch rates for a year may be affected by the same factor) and that the standardization may have ignored a key factor. Francis et al. (2003) analyzed a large number of

[2] In this case, the I_t in the denominator of the first term is the Jacobian from the transformation of variables (Casella and Berger, 1990, p. 50, Theorem 2.1.2). This term is often omitted from the log-normal likelihood function presented in the fisheries literature (e.g., Hilborn and Mangel, 1997, p. 248; Maunder and Starr, 2003). However, because I_t is a constant, its inclusion or otherwise has no effect on the estimates of the parameters of the model.

CPUE and research survey data sets and found that the coefficient of variation for the combined effects of observation error and annual variation in catchability was ∼0.15–0.2. Although not commonly done, Eq. (15) can be extended to allow for the correlation among the year factors that results from all the outputs being from the same model (Myers and Cadigan, 1995; Cooke and Lankester, 1996; Francis, 1999).

Methods (b) and (c) should not be used if the year-specific weights, w_t^2, are based on the standard errors of the year effects, and the year effect for the base year is set equal to zero (and so has no variance) (Maunder and Starr, 2003). A simple way to avoid this problem is apply method (b), where w_t is some function of the sample size for year *t*. Francis (1999) suggests a more elegant solution to this problem by dividing the annual indices by the geometric mean index over all years, $I_y^0 = I_y\left(\prod_{y'} I_{y'}\right)^{1/n}$, where *n* is the number of years for which indices are available. The main advantage of this 'canonical' form is that the standard error for each year can be calculated, and is independent of the year that is chosen as the base year (R.I.C.C. Francis, NIWA, pers. commun.).

The value of σ_t^2 determines, in addition to the level of uncertainty, the amount of information provided by the CPUE index relative to any other sets of data when there is more than one source of information in a stock assessment. Therefore, in some cases, it may be appropriate to fix σ at a level that reflects the quality of the data or the validity of the assumptions underlying its use (e.g. standardized CPUE is proportional to abundance). This is particularly important when a CPUE index is available for many years, and an index of abundance from survey data is only available for a few years. The greater number of data points for the CPUE index can drive the analysis even if σ is fixed at a relatively high level.

In principle, the stock assessment and the catch-effort standardization can be conducted simultaneously (e.g., Maunder, 2001; Maunder and Langley, 2004). In such an 'integrated analysis', the parameters of the population dynamics model and those related to catch-effort standardization are estimated simultaneously by optimizing an objective function for all sources of data available to the stock assessment model (e.g. catch-at-age data). Maunder (2001) showed that this integrated analysis produced much narrower confidence intervals that included the true value more often than the two-step approach of first standardizing the catch and effort data and then including its results in an assessment model. One advantage of integrating the catch-effort standardization with the fitting of the stock assessment model is that the uncertainty associated with the catch-effort standardization, including temporal correlation, is automatically accounted for, when the uncertainty associated with any model outputs is computed. Unfortunately, integrating the catch-effort standardization with the stock assessment can be very computationally demanding. This has restricted its application to date.

8. Discussion

Standardization of catch and effort data to develop an index of the relative abundance of a fish population assumes that the explanatory variables available are sufficient to remove (or explain) most of the variation in the data that is not attributable to changes in abundance. However, even if catch and effort data are standardized to remove the impact of all known factors, there is still no guarantee that the resultant index of abundance is linearly proportional to abundance (as is assumed in Eqs. (14) and (15)). Cooke and Beddington (1984) and Cooke (1985) described various scenarios in which catch rate is unlikely to be linearly related to abundance. Cooke and Beddington (1984) highlighted the possibility that catch rates may decline more slowly than abundance ('hyperstability'). Based on a meta-analysis of 297 CPUE data series, and extending the work of Dunn et al. (2000), Harley et al. (2001) found strong evidence that CPUE was hyperstable (i.e. CPUE remains high while abundance declines). However, the opposite problem ('hyperdepletion') can also occur (e.g., Prince and Hilborn, 1998).

The goal of the standardization is to remove most of the annual variation in the data not attributable to changes in abundance. However, the fraction of the overall variation in the data explained by a catch-effort standardization can be disappointingly low. For example, when standardizing catch and effort data for orange roughy east of New Zealand using the delta-lognormal approach, Anderson (2003) was able to explain only 14.5% of the deviance associated with the non-zero catches and 7.5% of the deviance associated with whether a catch is positive, even though information was available on 12 (potential) explanatory variables. Not surprisingly, the standardized index of abundance was almost identical to the annual geometric means of the catch rates. However, there is an interaction between the level of variation explained and the level of data aggregation (Paul Starr, pers. commun.). When the data are highly disaggregated (e.g. 'tow-by-tow' data), the explanatory power is generally low and can be 'increased' by aggregating the data. It may therefore not be appropriate to compare the level of variation explained among different analyses, and analysts should not base their perceptions about the reliability of their index of abundance on the extent of the variation explained.

This review has focused on the methods used most frequently to standardize catch and effort data, specifically those that can be implemented using such popular statistics packages as SPlus and SAS. However, this is a rapidly developing field. For example, Bishop et al. (2000) used generalized estimating equations (GEEs; Liang and Zeger, 1986; Zeger and Liang, 1986) to standardize catch and effort data, Maunder and Harley (2003) used neural networks and Watters and Deriso (2000) used regression trees. The latter two approaches allow for non-linear relationships between the response variable and the explanatory variables, and hence allow the data to identify the relationship(s) between catch rate and the explanatory variables.

Hinton and Nakano (1996) used a priori expectations about the distribution of the habitat, the fishing gear, and the species of interest to standardize catch rates for blue marlin (*Makaira nigricans*) using a deterministic model. However, due to differences in the spatial and temporal scales of the data for habitat distribution and the species habitat preference (and other factors), this method often gives misleading results, and a statistical version of this approach that estimates habitat preference directly from the data is now considered more appropriate (Hinton and Maunder, 2003). Hinton and Maunder (2003) describe a test based on the consistency of the index of abundance with the assumptions made in the population dynamics model and the other data used in the stock assessment. Methods used in other disciplines, but not yet used for standardizing catch and effort data could also be appropriate (e.g. ridge regression, Hastie et al., 2001).

Although many methods are now available to standardize catch and effort data, little effort has been directed toward identifying the most appropriate methods for specific instances. Some simulation work (e.g. Porch and Scott, 1994; Maunder, 2001; Campbell, 2004) has been undertaken, but additional work along these lines is clearly a high priority for the future.

Finally, although this paper has focused on standardization of fishery-dependent data, there is no reason that the methods outlined above could not be applied to fishery-independent data. For example, some of the methods described above are used to standardize fishery-independent dive survey density estimates for paua (*Halitiotis iris*) in New Zealand for the effects of diver and time of the year (e.g. Breen and Kim, 2003), while Cope et al. (2003) standardized the indices of juvenile abundance from power plant impingement rates for the effects of power station and season.

Acknowledgements

AEP was supported by NMFS grant NA07FE0473. William Bayliff, Cleridy Lennert, and Bill Venables commented on the manuscript. Paul Starr and two anonymous reviewers reviewed the manuscript. In particular, we thank Paul Starr for his thorough review and numerous suggestions for its improvement.

References

Aitchison, J., 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. J. Am. Statist. Assoc. 50, 901–908.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Proceedings of the 2nd International Symposium on Information Theory. Publishing House of the Hungarian Academy of Sciences, Budapest, pp. 268–281.

Anderson, O.F., 2003. CPUE analysis and stock assessment of the East Cape hills (ORD 2A North) orange roughy fishery for 2003. NZ Fisheries Association Report No. 2003/24.

Ayers, D., 2003. Standardized CPUE analysis for the northern gemfish (*Rexea solandri*) fisheries in SKI 1 and SKI 2, 1989–2002. NZ Fisheries Association Research Document No. 2003/34.

Bannerot, S.P., Austin, C.B., 1983. Using frequency distributions of catch per unit effort to measure fish-stock abundance. Trans. Am. Fish. Soc. 112, 608–617.

Battaile, B.C., Quinn II, T.J., 2004. Catch per unit effort standardization of the eastern Bering Sea walleye Pol-

lock (*Theragra chalcogramma*) fleet. Fish. Res. 70, 161–177.
Berry, D.A., 1987. Logarithmic transformations in ANOVA. Biometrics 43, 439–456.
Beverton, R.J.H., Holt, S.J., 1957. On the Dynamics of Exploited Fish Populations. HMSO, London.
Bigelow, K.A., Boggs, C.H., He, X., 1999. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. Fish. Oceanogr. 8, 178–198.
Bishop, J., Die, D., Wang, Y.-G., 2000. A generalized estimating equations approach for analysis of the impact of new technology on a trawl fishery. Aust. NZ J. Statist. 42, 159–177.
Brandão, A., Butterworth, D.S., Johnston, S.J., Glazer, J.P., 2004. Using a GLMM to estimate the somatic growth rate for male South African west coast rock lobster *Jasus lalandi*. Fish. Res. 70, 335–345.
Breen, P.A., Kim, S.W., 2003. The 2003 stock assessment of paua (*Haliotis iris*) in PAU 7. NZ Fisheries Association Research Document No. 2003/35.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd ed. Springer-Verlag, New York.
Butterworth, D.S., 1996. A possible alternative approach for generalized linear model analysis of tuna CPUE data. ICCAT Col. Vol. Sci. Pap. 45, 123–124.
Butterworth, D.S., Andrew, P.A., 1984. Dynamic catch-effort models for the hake stocks in ICSEAF Divisions 1.3 to 2.2. Col. Sci. Pap. Int. Commun. SE Atl. Fish. 11, 29–58.
Butterworth, D.S., Ianelli, J.N., Hilborn, R., 2003. A statistical model for stock assessment of southern bluefin tuna with temporal changes in selectivity. Afr. J. Mar. Sci. 25, 331–361.
Campbell, R., 1998. Catch and effort analysis with uncertain stock and effort dynamics: southern bluefin tuna longline. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.F., Sullivan, P.J., Zhang, C.-I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01. University of Alaska Fairbanks, Fairbanks, pp. 75–97.
Campbell, R., 2004. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70, 209–227.
Casella, G., Berger, R.L., 1990. Statistical Inference. Duxbury Press, Belmont.
Chang, S.-K., 2003. Analysis of Taiwanese white marlin catch data and standardization of catch rates. ICCAT Col. Vol. Sci. Pap. 55, 453–466.
Cooke, J.G., 1985. On the relationships between catch per unit effort and whale abundance. Rep. Int. Whal. Commun. 35, 511–519.
Cooke, J.G., Beddington, J.R., 1984. The relationship between catch rates and abundance in fisheries. IMA J. Math. Appl. Med. Biol. 1, 291–405.
Cooke, J.G., Lankester, K., 1996. Consideration of statistical models for catch-effort indices for use in tuning VPAs. ICCAT Col. Vol. Sci. Pap. 45, 125–131.
Cope, J.M., Piner, K., Minte-Vera, C., Punt, A.E., 2003. Status and future prospects for the cabezon (*Scorpaenichthys marmoratus*) as assessed in 2003. Pacific Fishery Management Council, Portland, OR.
Cragg, J.G., 1971. Some statistical models for limited dependent variables with applications to the demand for durable goods. Econometrica 39, 829–844.
Deriso, R.B., Quinn II, T.J., Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42, 815–824.
Dichmont, C.M., Punt, A.E., Deng, A., Dell, Q., Venables, W., 2003. Application of a weekly delay-difference model to commercial catch and effort data in Australia's Northern Prawn Fishery. Fish. Res. 65, 335–350.
Dick, E.J., 2004. Beyond 'lognormal vs. gamma': discrimination among error distributions for generalized linear models. Fish. Res. 70, 347–362.
Dobson, A.J., 1991. An Introduction to Generalized Linear Models. Chapman & Hall, London.
Dong, Q., Resterpo, V.R., 1996. Notes on the Poisson error assumption made to estimate relative abundance of West Atlantic bluefin tuna. ICCAT Col. Vol. Sci. Pap. 45, 158–161.
Dunn, A., Harley, S.J., Doonan, I.J., Bull, B., 2000. Calculation and interpretation of catch-per-unit-effort (CPUE) indices. NZ Fisheries Association Report No. 2000/1.
Fournier, D., Archibald, C.P., 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39, 1195–1207.
Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55, 2105–2116.
Francis, R.I.C.C., 1999. The impact of correlations in standardized CPUE indices. NZ Fisheries Association Research Document No. 99/42.
Francis, R.I.C.C., Cordue, P.L., Haist, V., 2002. Review of the 2001 hoki stock assessment. NZ Fisheries Association Report No. 2002/42.
Francis, R.I.C.C., Hurst, R.J., Renwick, J.A., 2003. Quantifying annual variation in catchability for commercial and research fishing. Fish. Bull., US 101, 293–304.
Gavaris, S., 1980. Use of a multiplicative model to estimate catch rate and effort from commercial data. Can. J. Fish. Aquat. Sci. 37, 2272–2275.
Guisan, A., Edwards Jr., T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Mod. 157, 89–100.
Gulland, J.A., 1956. On the fishing effort in English demersal trawl fisheries. Invest. Minst. Agric. Fish. Food UK (Ser. 2) 20, 1–41.
Hall, D.B., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56, 1030–1039.
Harley, S.J., Myers, R.A., Dunn, A., 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. 58, 1760–1772.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.
Hilborn, R., Mangel, M., 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton.
Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment. Chapman & Hall, London.

Hilborn, R., Maunder, M.N., Parma, A., Ernst, B., Paynes, J., Starr, P.J., 2000. Documentation for a general age-structured Bayesian stock assessment model: code named Coleraine. Fisheries Research Institute, University of Washington, FRI/UW 00/01.

Hinton, M.G., Maunder, M.N., 2003. Methods for standardizing CPUE and how to select among them. ICCAT Document No. SCRS/2003/034.

Hinton, M.G., Nakano, H., 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from the Japanese longline fisheries in the Pacific. Bull. Int. Am. Trop. Tuna Commun. 21, 171–200.

Honma, M., 1973. Estimation of overall effective fishing intensity of tuna longline fishery. Bull. Far Seas Fish. Res. Lab. 10, 63–85.

Horn, P.L., 2003. CPUE from commercial fisheries for ling (*Genypterus balcodes*) around the North Island, New Zealand: an evaluation of series for LIN 1, LIN 2, and Cook Strait. NZ Fisheries Association Report No. 2003/13.

Kimura, D.K., 1981. Standardized measures of relative abundance based on modelling log(c.p.u.e.), and the application to Pacific ocean perch (*Sebastes alutus*). J. Cons. Int. Explor. Mer. 39, 211–218.

Lambert, D., 1992. Zero-inflated Poisson regression, with application to random defects in manufacturing. Technometrics 34, 1–14.

Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 12–22.

Lo, N.C., Jacobson, L.D., Squire, J.L., 1992. Indices of relative abundance for fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. Sci. 49, 2515–2526.

Maunder, M.N., 2001. A general framework for integrating the standardization of catch per unit of effort into stock assessment models. Can. J. Fish. Aquat. Sci. 58, 795–803.

Maunder, M.N., Harley, S.J., 2003. Methodological improvements to the EPO tuna stock assessments. SCTB16 Working Paper No. MWG-2. Available from the Inter-American Tropical Tuna Commission.

Maunder, M.N., Langley, A.D., 2004. Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. Fish. Res. 70, 385–391.

Maunder, M.N., Starr, P.J., 2003. Fitting fisheries models to standardised CPUE abundance indices. Fish. Res. 63, 43–50.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall, London.

Methot, R.D., 1993. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Bull. Int. North Pac. Fish. Commun. 50, 259–277.

Methot, R.D., 2000. Technical description of the stock synthesis assessment program. NOAA Technical Mimeograph No. NMFS-NWFSC-43, 56 pp.

Miyabe, N., Takeuchi, Y., 2003. Standardized bluefin CPUE from the Japanese longline fishery in the Atlantic including those for mixing studies. ICCAT Col. Vol. Sci. Pap. 55, 1190–1207.

Mullahy, J., 1986. Specification and testing of some modified count data models. J. Economet. 33, 341–365.

Myers, R.A., Cadigan, N.G., 1995. Statistical analysis of catch-at-age data with correlated errors. Can. J. Fish. Aquat. Sci. 52, 1265–1273.

Myers, R.A., Pepin, P., 1990. The robustness of lognormally-based estimators of abundance. Biometrics 46, 1185–1192.

Nelder, J.A., Wedderburn, R.W.M., 1972. Generalised linear models. J. R. Statist. Soc. A 137, 370–384.

NRC (National Research Council), 1994. An Assessment of Atlantic Bluefin Tuna. National Academy Press, Washington, DC.

Ortiz, M., Arocha, F., 2004. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. Fish. Res. 70, 275–294.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-effects Models in S and S-plus. Springer, New York.

Porch, C.E., Scott, G.P., 1994. A numerical evaluation of GLM methods for estimating indices of abundance from West Atlantic bluefin tuna catch per trip data when a high proportion of the trips are unsuccessful. ICCAT Col. Vol. Sci. Pap. 42, 241–245.

Porter, J.M., Oritz, M., Paul, S.D., 2003. Updated standardized CPUE indices for Canadian bluefin tuna fisheries based on commercial catch rates. ICCAT Col. Vol. Sci. Pap. 55 (3), 1005–1018.

Prince, J.D., Hilborn, R., 1998. Concentration profiles and invertebrate fisheries management. In: Jamieson, G.S., Campbell, A. (Eds.), Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management. Can. Spec. Publ. Fish. Aquat. Sci., pp. 187–196.

Punt, A.E., Butterworth, D.S., 2003. Specifications and clarifications regarding the ADAPT VPA assessment/projection computations carried out during the September 2000 ICCAT west Atlantic bluefin tuna stock assessment session. ICCAT Col. Vol. Sci. Pap. 55, 1041–1054.

Punt, A.E., Kennedy, R.B., 1997. Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. Mar. Freshw. Res. 48, 967–980.

Punt, A.E., Walker, T.I., Taylor, B.L., Pribac, F., 2000a. Standardization of catch and effort data in a spatially-structured shark fishery. Fish. Res. 45, 129–145.

Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., Prince, J.D., 2000b. Stock assessment of school shark *Galeorhinus galeus* based on a spatially-explicit population dynamics model. Mar. Freshw. Res. 51, 205–220.

Punt, A.E., Smith, D.C., Thomson, R.B., Haddon, M., He, X., Lyle, J.M., 2001a. Stock assessment of the blue grenadier *Macruronus novaezelandiae* resource off southeastern Australia. Mar. Freshw. Res. 52, 701–717.

Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., 2001b. Population modeling and harvest strategy evaluation for school and gummy shark. Report of FRDC Project No. 99/102.

Quinn II, T.J., Hoag, S.H., Southward, G.M., 1982. Comparison of two methods of combining catch-per-unit-effort from geographic regions. Can. J. Fish. Aquat. Sci. 39, 837–846.

Ralston, S., Dick, E.J., 2003. The status of black rockfish (*Sebastes melanops*) off Oregon and northern Washington in 2003. Stock Assessment and Fishery Evaluation, Pacific Fishery Management Council, Portland, OR, v + 70 pp.

Ridout, M., Demetrio, C.G.B., Hinde, J., 1998. Models for count data with many zeros. In: Proceedings of the XIX International Biometric Conference, Cape Town, South Africa, 14–18 December 1998. International Biometric Society, pp. 179–192.

Robins, C.M., Wang, Y.-G., Die, D., 1998. The impact of global positioning systems and plotters on fishing power in the Northern Prawn Fishery, Australia. Can. J. Fish. Aquat. Sci. 55, 1645–1651.

Robson, D.S., 1966. Estimation of the relative fishing power of individual ships. ICNAF Res. Bull. 3, 5–14.

Rodríguez-Marín, E., Arrizabalaga, H., Ortiz, M., Rodríguez-Cabello, C., Moreno, G., Kell, L.T., 2003. Standardization of bluefin tuna, *Thunnus thynnus*, catch per unit effort in the baitboat fishery of the Bay of Biscay (Eastern Atlantic). ICES J. Mar. Sci. 60, 1216–1231.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley/Interscience, New York.

Schnute, J., 1977. Improved estimates from the Schaefer production model: theoretical considerations. J. Fish. Res. Bd. Can. 34, 583–603.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.

Stefánsson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J. Mar. Sci. 53, 577–588.

Stephens, A., MacCall, A., 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fish. Res. 70, 295–306.

Syrjala, S.E., 2000. Critique on the use of the delta distribution for the analysis of trawl survey data. ICES J. Mar. Sci. 57, 831–842.

Taylor, P.R., 2003. Standardized CPUE for the northwest Chatham Rise orange roughy fishery. NZ Fisheries Association Report No. 2003/32.

Vignaux, M., 1994. Catch per unit effort (CPUE) analysis of west coast South Island Cook Strait spawning hoki fisheries, 1987–93. NZ Fisheries Association Research Document No. 94/11.

Vignaux, M., 1996. Analysis of spatial structure in fish distribution using commercial catch and effort data from the New Zealand hoki fishery. Can. J. Fish. Aquat. Sci. 53, 963–973.

Walters, C., 2003. Folly and fantasy in the analysis of spatial catch rate data. Can. J. Fish. Aquat. Sci. 60, 1433–1436.

Watters, G., Deriso, R., 2000. Catch per unit of effort of bigeye tuna: a new analysis with regression trees and simulated annealing. Bull. Int. Am. Trop. Tuna Commun. 21 (8), 527–571.

Zeger, S.L., Liang, K.-Y., 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42, 121–130.