Fisheries Research 161 (2015) 109–130

Contents lists available at ScienceDirect

Fisheries Research

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

Constructing stock abundance indices from catch and effort data: Some nuts and bolts

Robert A. Campbell*

Wealth from Oceans Flagship, CSIRO Marine Research and Atmospheric Research, Private bag No. 1, Aspendale, Victoria, 3195, Australia

ARTICLE INFO

Article history:
Received 21 February 2014
Received in revised form 1 July 2014
Accepted 2 July 2014
Handling Editor A.E. Punt
Available online 31 July 2014

Keywords:
CPUE standardisation
Stock abundance indices
Simulation-testing
General linear mixed models

ABSTRACT

Construction of annual indices of stock abundance based on the standardisation of catch and effort data remains central to many fisheries assessments. However, while the use of advanced statistical methods has helped catch rates to be standardised accounting for many explanatory variables, some of the more routine aspects of constructing abundance indices, such as how the index is constructed from the model parameters, receives little explanation in many analyses. This has lead to a lack of understanding as to how the indices are constructed and in some instances incorrect techniques have been applied. This lack of understanding can be a particular issue when interactions are used in the standardising model, especially those which contain interactions with the temporal effects over which the time-series of the abundance index is required. Other issues include the use of weighted model fits, the influence of anomalous data values, how best to impute missing values when required, the consequences of model mis-specification, and the use of random-effects. In this study, the basic approach for constructing abundance indices is outlined and a worked example using simulated data is presented to explore the nature of these issues and several techniques are suggested for dealing with them and to overcome potential biases. Finally, the methods presently used by ICCAT scientists for constructing abundance indices are reviewed.

1. Introduction

The construction of relative indices of stock abundance from commercial catch-per-unit-effort (CPUE) data continues to be a routine and important aspect of many stock assessments worldwide. For example, this issue was the primary concern of 82 (23%) of the 351 papers published in the last three general issues of the Collective Volume of Scientific Papers for the International Commission for the Conservation of Atlantic Tunas (ICCAT). However, the relationship between nominal CPUE and abundance may be weak because CPUE reflects changes of catchability as well as population abundance (Clark, 1985; Harley et al., 2001, Bishop, 2006; Ye and Dennis, 2009). In an attempt to 'standardise' CPUE time series there has been a steady publication of papers on statistical techniques which may be considered appropriate for such purposes (see review by Maunder and Punt, 2004; Tascheri et al., 2010; Cao et al., 2011; Lynch et al., 2012; Brodziak and Walsh, 2013). These techniques cover a range of methods including general linear model (GLMs), general additive models (GAMs) and general linear mixed models (GLMMs) and regression trees and there has also been considerable discussion on which of the many error distributions to chose from (e.g. normal, gamma, Poisson, negative binomial and Tweedie distributions) may be the most appropriate to use with these methods. These issues all remain relevant to ongoing research in this area.

However, while this focus on statistical techniques continues, some of the perhaps less interesting aspects of constructing abundance indices appear to have been routinely ignored. Some of these problems may stem from the increasing use of 'off-the-shelf' statistical routines with little consideration given to the nuances of the data being analysed or how best to understand the outputs of these routines, especially when internal routines are relied on to construct the desired index of abundance (Goodyear and Ortiz, 2007).

Walters (2003) noted that two serious mistakes are commonly made in the analysis of spatial catch rate data. These mistakes relate to the use of the CPUE ratio estimator at an inappropriate spatial scale (the folly) and the fact that unfished strata are usually ignored in the construction of abundance indices (the fantasy). While several techniques for dealing with these issues have been proposed (e.g. Campbell, 1998, 2004; Walters, 2003) it remains uncertain whether these issues continue to be ignored. For

* Tel.: +61 3 9239 4681; fax: +61 3 9239 4444.
E-mail address: Robert.Campbell@csiro.au

http://dx.doi.org/10.1016/j.fishres.2014.07.004
0165-7836/

example, Carruthers et al. (2010) noted that it is common for abundance indices to be presented without an explicit description of how they are calculated.

Another potential problem is that little effort has been put into how best to construct the desired abundance index using the parameter estimates obtained from standardisation models. A review of the literature indicates that a range of techniques are being used for this task. However, some appear to be poorly chosen or inappropriate, resulting in the index being incorrectly constructed. This can be a particular issue when interactions are included in the standardising model, especially those which contain interactions with temporal effects (e.g. year, quarter) over which the time-series of the index is required. Other issues include the use of weighted model fits, the influence of anomalous data values, how best to impute missing values when required, the consequences of model mis-specification, and the use of random versus fixed effects.

In this study, after first outlining the basic approach for constructing abundance indices, a worked example using simulated data is presented to explore the nature of these issues and several techniques for dealing with them are suggested. Finally, the methods presently used by ICCAT scientists for constructing abundance indices are reviewed.

2. Basic equations

A number of statistical approaches have been adopted for the standardisation of CPUE data and these, together with the basic equations, have been widely disseminated in the fisheries literature (e.g. Maunder and Punt, 2004; Campbell, 2004). As such a review of these methods will not be repeated here. However, as the use of the two step delta-General Linear (Mixed) Models (delta-GLM/GLMM) approaches (Lo et al., 1992; Stefansson, 1996) appears to have increased in recent years, and has been identified as 'best-practice' for many situations (Lynch et al., 2012), I will adopt this approach here. Furthermore, as these models can be seen to generalise the simpler single-step GLM/GLMM models (i.e. by removal of the binomial component) the results can also be used to understand these simpler models. Note, in the following the notation GLM will be used to designate both GLM and GLMM methods.

The delta-GLM approach is often used to standardise CPUE data as the distribution of catch rates are usually skewed to the right with a spike at the origin. This suggests that such data might be modelled in two stages; one stage being concerned with the pattern of occurrence of successful catches, and the other with the size of the successful catch. Furthermore, for both components we can model the means as linear combinations of the factors likely to influence the probability of a successful catch and the size of the catch.

An example helps illustrate this approach. Consider a region for which there are N catch rate observations, R_i. The average catch rate, μ, can be expressed as follows:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} R_i = \frac{1}{N_S + N_F}\sum_{i=1}^{N_S} R_i = \frac{N_S}{N_S + N_F}\,\frac{1}{N_S}\sum_{i=1}^{N_S} R_i = p_s \mu_s$$

where N_S is the number of successful catches ($R_i > 0$), N_F is the number of failed catches ($R_i = 0$), p_s is the proportion of successful catches and μ_S is the average of the successful catch rates. This formulation shows that the overall mean catch rate can be expressed as the combination of the parameters from the distributions used to model the probability of a successful catch and that used to model the non-zero catch rates.

The achievement of either a successful or failed catch is a Bernoulli process and the probability of obtaining a successful catch is usually modelled as a binomial distribution. On the other hand, a number of different distributions can be used to model the size of the successful catch (or catch rates).

2.1. Stage 1: Probability of successful catch

The binominal distribution is used to model the probability of a non-zero catch where each observation is modelled as either a success ($C_i > 0$) of a failure ($C_i = 0$), with the probability of either expressed as follows:

$$Pr(C_i > 0) = p_S \quad \text{and} \quad Pr(C_i = 0) = 1 - p_S$$

Associated with each observation is a vector of covariates or explanatory variables X_j thought likely to influence the probability of a positive catch. Furthermore, the dependence of p_s is assumed to occur through a linear combination $\eta = \sum \alpha_j X_j$ of the explanatory variables X_j with unknown coefficients α_j. However, unless restrictions are placed on these coefficients $-\infty < \eta < \infty$. Therefore, in order to ensure that $0 \le p_s \le 1$ the logit link function is generally used which takes the following form:

$$\eta = \log\left(\frac{p_S}{1 - p_S}\right)$$

The inverse of this relation gives the probability of obtaining a positive catch as a function of the explanatory variables:

$$p_S = \frac{e^{\eta}}{1 + e^{\eta}} = \frac{\exp(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots)}{1 + \exp(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots)} = \mathrm{InvLogit}(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots) \quad (1)$$

This equation shows that while the effect of a unit change in X_j is to increase the linear predictor of success by an amount α_j, the effect on the probability scale is more complex.

2.2. Stage 2: Mean size of positive catch rate

Having fitted the above model to the probability of obtaining a successful catch, a separate model is fitted to the distribution of observed positive catch rates, c_S. A number of distributions (e.g. gamma, Tweedie) can be used to model the associated error distribution and the expected value of c_S is related to the linear vector of covariates or explanatory variables Y_j by a log link, i.e. $\eta = \Sigma \beta_j Y_j$. Note, however, a model with log link and a Gaussian error structure (i.e. with constant variance) is usually considered inappropriate. The size of the positive catch rate for each observation is then given by the following equation:

$$c_S = \exp(\beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \cdots) \quad (2)$$

If fitting the model to the observed catch, instead of catch rate, the negative binomial distribution, which models integer values, may be a good candidate distribution as it provides a general form of the assumed variance function ($\mu + k\mu^2$). Also, when fitting a normal linear model to log-transformed responses then there is a need to correct for bias when back transforming the predictions to obtain estimates of the mean on the response scale (Newman, 1993).

A check of the assumed variance-mean relationship (which determines the appropriate member of the exponential family for the assumed error distribution) should be undertaken using a diagnostic plot of the square-root of the absolute values of the deviance residuals against the fitted values (or monotonic transforms of the latter).

2.3. Stage 3: Construction of abundance index

To illustrate the procedure for the construction of the required abundance index, assume that the linear combination of explanatory variables in each modelled stage takes the following form (shown here using the notation of stage 2):

$$\eta = \beta_0 + \sum_{i=1}^{\mathrm{Ny}} \beta_{y,i} Y_i + \sum_{j=1}^{\mathrm{Nq}} \beta_{q,j} Q_j + \sum_{k=1}^{\mathrm{Nr}} \beta_{r,k} R_k + \sum_{m=1}^{\mathrm{Nd}} \beta_{d,m} D_m + \sum_{n=1}^{\mathrm{Ne}} \beta_{e,n} E_n \tag{3}$$

where β_0 is the intercept; $\beta_{y,i}$ is the effect of the i-th year Y_i; $\beta_{q,j}$ is the effect of the j-th quarter Q_j (or another seasonal measure); $\beta_{r,k}$ is the effect of the k-th fishing region R_k; $\beta_{d,m}$ is the effect of the m-th gear deployment setting D_m; and $\beta_{e,n}$ is the effect of the n-th environmental condition E_n. Ny, Nq, Nr, Nd and Ne are the number of levels for each of the respective effects, and due to the fact that each of the explanatory effects is fitted as a categorical variable each has the form $X_i = \delta_i^x$ where $\delta_i^x = 1$ where $i = x$ and $\delta_i^x = 0$ where $i \neq x$. Note that the effects D_m and E_n can be thought of as standardising catchability (i.e. fishing power) and availability (i.e. fish distribution) respectively, while the Y_i, Q_j and R_k effects together specify the density within each spatial–temporal stratum (Campbell, 2004). In particular, the Y_i, provide an annual measure of the density of fish in the region and quarter chosen as the standards for these effects for the model given by Eq. (3), where no interactions are included, while the R_k provide a measure of the distribution of the density of available fish across each region. The Q_j effects then allow for relative changes in these densities across each quarter, but the overall interpretation depends on whether the fitted model allows for interactions with this effect. The values of Q_j need to be interpreted as implying the same change in the density of available fish across all regions simultaneously if there are no interactions (i.e. availability cannot go down in one region and up in another simultaneously as this would imply an interaction). Differential changes in densities across regions between quarters occurs when there is an interaction between quarter and region, and these changes can be interpreted as implying some form of movement among the regions. More on the inclusion of interaction terms in the model is provided in the next section. The incorporation of continuous variables in the linear combination of explanatory variables will be considered in the worked example below.

The probability of obtaining a successful catch is also likely to be related to the amount of fishing effort (i.e. number of hooks deployed, H). An extra term therefore needs to be added to the binominal model. One option is to include a measure of fishing effort as an offset (i.e. a regression variable with a constant coefficient of 1 for each observation). This serves to normalise the probability of obtaining a catch to a per-hook basis. In order to retain the relationship between the two variables the variable Logit $H = \log(h/(1-h))$ where $h = H/M$ where $M >$ maximum(H) can be used. Some function of the fishing effort (e.g. a polynomial) can be added to the linear combination of explanatory variables if the regression coefficient for this variable is not to be fixed to 1, or it is believed that the relationship between the probability of success and effort is not linear. It is suggested that the measure of effort (e.g. effort = $[H - \mu(H)]/\sigma(H)$) be standardised first or that the variable Logit H be used. Fitting a polynomial of order greater than three should be done with care as this may introduce unnecessary non-linearity into the assumed relationship, especially if the data coverage is poor near the ends of the observed range of fishing efforts.

Walters (2003) explained that the construction of the abundance index is the same as constructing a very large table, with a row for each time period and column for each spatial strata fished. The time period should correspond to the smallest temporal period used in the standardising model, quarter in the example used here, while the range of spatial strata should correspond to the total number of distinct spatial strata fished within the period of time over which the index is to be constructed (i.e. each distinct region in the above model whether or not it was fished during all time periods).

The expected value of the standardised values of both p_S and c_S in the i-th year, j-th quarter and k-th region can be found by selecting a standardising level for each of the catchability and availability effects. In practice that level for which the related parameter is zero (e.g. $\beta_{d,s} = 0$ and $\beta_{e,s} = 0$ for the standardising level s for each effect) is selected. The choice of the standardising level for each effect is arbitrary but good practice may be to choose the category level which corresponds to the mode of the observations across all levels (i.e. the most common gear and environmental conditions in the fishery, e.g. Punt et al., 2000). In fitting the GLM this level can be set to the base level against which all others are compared and for which the corresponding parameter value $\beta_{\cdot,s}$ is usually automatically set to zero (due to the over-parameterisation inherent in the model). For continuous variables it is good practice, if not done internally by the GLM routine, to first normalise the variable so that the mean value is zero in the fitted model.

Using this approach, the expected value of the standardised catch rate in the i-th year, j-th quarter and k-th region is given by:

$$\begin{aligned} \mathrm{std}CPU_{ijk} &= p_S(i,j,k).c_S(i,j,k) \\ &= \mathrm{InvLogit}(\alpha_0 + y_{\alpha,i} + q_{\alpha,i} + r_{\alpha,k}) \exp(\beta_0 + y_{\beta,i} + q_{\beta,i} + r_{\beta,k}) \end{aligned} \tag{4}$$

where $y_{\alpha,i} = \alpha_{y,i} Y_i$ and $q_{\beta,j} = \beta_{q,j} Q_j$, etc. An index of abundance, $B(i,j,k)$, for the size of the fish population (or biomass) in the i-th year, j-th quarter and k-th region can then be obtained by multiplying this standardised catch rate (which is taken as a proxy for the density of fish) by the spatial size, A_k, of the region fished (or equivalently, the proportion of the total area of all regions represented by the k-th region), i.e. $B(i,j,k) = A_k.\mathrm{stdCPUE}_{ijk}$. A total index of abundance within a season $B(i,j)$ is then obtained by summing over all regions of the fishery.

$$B(i,j) = \sum_{k=1}^{\mathrm{Nr}} A_k.\mathrm{stdCPUE}_{ijk} \tag{5}$$

If required, an annual index of abundance for the i-th year, $B(i)$ can be obtained by taking the average over all seasons in that year. Either the arithmetic or geometric mean can be used, the latter being scale invariant and less influenced by outliers:

$$B(i) = \frac{1}{\mathrm{Nq}} \sum_{j=1}^{\mathrm{Nq}} \left[\sum_{k=1}^{\mathrm{Nr}} A_k.\mathrm{stdCPUE}_{ijk} \right] \tag{6a}$$

$$B(i) = \sqrt[\mathrm{Nq}]{\prod_{j=1}^{\mathrm{Nq}} \left[\sum_{k=1}^{\mathrm{Nr}} A_k.\mathrm{stdCPUE}_{ijk} \right]} \tag{6b}$$

Finally, a relative index which relates the average abundance in the i-th year to some reference year (or the mean of $B(i)$ over some reference period) can be calculated as:

$$I(i,\mathrm{ref}) = \frac{B(i)}{B(\mathrm{ref})} \tag{7}$$

For an analysis based on Eq. (5), and using Eq. (6a), the relative abundance index is given by:

$$I(i,\mathrm{ref})=\frac{\exp(y_{\beta,i})\sum_{j=1}^{\mathrm{Nq}}\left[\exp(q_{\beta,j}).\sum_{k=1}^{\mathrm{Nr}}A_k\,\mathrm{InvLogit}(\alpha_0+y_{\alpha,i}+q_{\alpha,j}+r_{\alpha,k})\,\exp(r_{\beta,k})\right]}{\exp(y_{\beta,\mathrm{ref}})\sum_{j=1}^{\mathrm{Nq}}\left[\exp(q_{\beta,j}).\sum_{k=1}^{\mathrm{Nr}}A_k\,\mathrm{InvLogit}(\alpha_0+y_{\alpha,\mathrm{ref}}+q_{\alpha,j}+r_{\alpha,k})\,\exp(r_{\beta,k})\right]} \tag{8}$$

From Eq. (8) it can be seen that even for the simple 'main-effects' model such as that given by Eq. (4) the quarter or regional parameters cannot be removed and there is no alternative but to first evaluate the standardised CPUE for each individual year, quarter and region strata (i.e. Walter's large table). However, in those situations where only a single step GLM is used, and dispensing with the subscripts used to distinguish between the parameters for each stage, Eq. (6a) can be written as:

$$\begin{aligned}B(i)&=\frac{1}{\mathrm{Nq}}\sum_{j=1}^{\mathrm{Nq}}\left[\sum_{k=1}^{\mathrm{Nr}}A_k\,\exp(\beta_0+y_i+q_j+r_k)\right]\\&=\frac{\exp(\beta_0+y_i)}{\mathrm{Nq}}\sum_{j=1}^{\mathrm{Nq}}\left[\exp(q_j).\sum_{k=1}^{\mathrm{Nr}}A_k\,\exp(r_k)\right]\end{aligned} \tag{9}$$

When the size of the regions fished remain constant across years the term in the square brackets is constant for all years allowing the following simple form of the value of the reference index to be obtained:

$$I(i,\mathrm{ref})=\exp(y_i-y_{\mathrm{ref}})=\exp(y_i)\quad \text{if}\quad y_{\mathrm{ref}}=0 \tag{10}$$

In this situation, the exponent of the year effect alone can be used as the relative index of abundance.

3. Inclusion of interactions

Interaction terms can be included in a standardising model when different population trends are assumed (or found) within effects included in the model. For example, inclusion of a year-by-region (Y^*R) interaction implies different annual trends in abundance in different regions, which, in turn, implies some form of spatial structuring of the population. Where appropriate, this situation can be dealt with by undertaking the analysis within each separate region. However, where interaction terms are included in the standardising model these also need to the added to the equations used to derive the annual abundance index. Where only interactions between the three main spatial and temporal effects are considered, Eq. (4) can be generalised as follows.

$$\eta=\beta_0+f_\alpha(Y_i,Q_j,R_k,\beta_{ijk})+\sum_{m=1}^{\mathrm{Nd}}\beta_{d,m}D_m+\sum_{n=1}^{\mathrm{Ne}}\beta_{e,n}E_n \tag{11}$$

where $f_\alpha(Y_i,Q_j,R_k,\beta_{ijk})$ is a function of the fitted year, quarter and region effects and related unknown parameters β_{ijk} and, again, $\beta_{d,m}$ and $\beta_{e,n}$ are the parameters related to the gear effects, D_m, and environmental conditions, E_n, respectively. For example:

(i) Full 3-way interaction:

$$f(Y,Q,R;\beta)=\sum_{i=1}^{\mathrm{Ny}}\sum_{j=1}^{\mathrm{Nq}}\sum_{k=1}^{\mathrm{Nr}}\beta_{ijk}(Y*Q*R)_{ijk}$$

where $(Y^*Q^*R)_{ijk}$ parameterises the interaction between the i-th year, j-th quarter and k-th region and for which there are NyNqNr − 1 parameters β_{ijk} to be estimated, or:

(ii) Single Y^*Q interaction:

$$f(Y,Q,R;\beta)=\sum_{i=1}^{\mathrm{Ny}}\sum_{j=1}^{\mathrm{Nq}}\beta_{ij}(Y*[(Q)])_{ij}+\sum_{k=1}^{\mathrm{Nr}}\beta_kR_k$$

where $(Y^*Q)_{ij}$ parameterises the interaction between the i-th year and j-th quarter and for which there are NyNq − 1 parameters β_{ij} and Nr − 1 parameters β_k to be estimated. Note, it is often preferential to fit a single two-way interaction between effects A and B in the form A^*B instead of the equivalent form $A+B+A^*B$ as this avoids the often overlooked interpretation that the parameter estimates obtained for effect A alone in the latter equation are relative to only one level of effect B and vice versa.

Given a standardising model of the form of Eq. (11) the abundance index (6a) becomes:

$$B(i)=\frac{1}{\mathrm{Nq}}\sum_{j=1}^{\mathrm{Nq}}\left[\sum_{k=1}^{\mathrm{Nr}}A_k\,\mathrm{InvLogit}(\alpha_0+f_\alpha(y_{\alpha,i},q_{\alpha,j},r_{\alpha,k}))\exp(\beta_0+f_\beta(y_{\beta,i},q_{\beta,j},r_{\beta,k}))\right] \tag{12}$$

The derivation of the relative annual index is still dependent upon calculation of the standardised CPUE for each of the individual year, quarter and region strata. However, in those situations where only a single step GLM is used, we have the following two cases:

(i) Quarter by region interaction:

With the inclusion of the interaction parameters $\mathrm{qr}_{j,k}$ related to the interaction term Q*R in the standardising model, Eq. (12) becomes:

$$I(i)=\frac{\exp(\beta_0+y_i)}{\mathrm{Nq}}\sum_{j=1}^{\mathrm{Nq}}\left[\exp(q_j).\sum_{k=1}^{\mathrm{Nr}}A_k\,\exp(r_k+qr_{j,k})\right]$$

Again as all the terms in the square brackets remain constant for each year, the relative annual index remains given by Eq. (10).

(ii) Year interactions:

With the inclusion of either year–quarter (Y^*Q) or year–region (Y^*R) interaction effects then we have:

$$I(i,\mathrm{ref})=\exp(y_i-y_{\mathrm{ref}})\frac{\sum_{j=1}^{\mathrm{Nq}}\left[\exp(q_j+\mathrm{yq}_{i,j})\right]}{\sum_{j=1}^{\mathrm{Nq}}\left[\exp(q_j+\mathrm{yq}_{\mathrm{ref},j})\right]}$$

or

$$I(i,\mathrm{ref})=\exp(y_i-y_{\mathrm{ref}})\frac{\sum_{k=1}^{Nr}\left[\exp(r_k+yr_{i,k})\right]}{\sum_{k=1}^{Nr}\left[\exp(r_k+yr_{\mathrm{ref},k})\right]}$$

In each situation there is a need to calculate the abundance index at either the quarter or regional level, and where both or more interactions are included in the standardising model the index has to be calculated at both the quarter and regional strata level. Except in the situation where only a Q^*R interaction is included, it is perhaps good practice in all situations where temporal interactions are included to calculate the standardised CPUE for each year, quarter and region stratum and then use these to next calculate the annual indices of abundance (and the associated relative indices) using Eq. (12) above.

4. Imputation of missing values

Given the need to use an equation of the form of Eq. (12) when a delta-GLM approach is employed there is also a need to be able to calculate a standardised CPUE for each combination of the year, quarter and region strata included in the standardising model. While this is usually easily achieved for models without interaction terms (c.f. Eq. (9)), the model may not provide an estimate of the standardised CPUE for all strata when interactions are included. For example, when a full Y^*Q^*R interaction term is included in the standardising model it will be possible to estimate the related parameter β_{ijk} for each model stratum only if there are observations for all three-way combinations of the year, quarter and region strata. There will be a need to impute a value of the standardised CPUE in those instances where strata remain unobserved. Where a full three-way interaction is not included in the model, the need for observations in all strata is not so onerous. For example, observations have to exist in only one region for all two-way year-quarter combinations when only a Y^*Q interaction term is included. Similarly, observations have to exist in only one quarter for all two-way year–region combinations with the inclusion of a Y^*R interaction. This will be demonstrated in the worked example in Section 7 below.

There is no standard procedure for imputing the value of the standard CPUE in each unobserved stratum and a number of ad hoc methods have been used in the past. Here I distinguish between two types of methods. First, the standardised CPUE estimated by the model for adjacent strata can be used to interpolate a value within each missing stratum (Walters, 2003; Campbell, 2004; Carruthers et al., 2010, 2011). Simple rule based procedures or more sophisticated Bayesian (Zang and Holmes, 2010) or geostatistical (Pereira et al., 2012) algorithms may be used for this purpose. Alternatively, the values obtained after fitting a simpler standardising model which provides values of the standardised CPUE for each unobserved stratum can be used. For example, if the data indicates the appropriateness of fitting a full three-way interaction but there are one or more unobserved strata, then a simpler model (e.g. a main effects only model) can also be fitted and the results from this latter model used to estimate the standardised CPUE in the unobserved strata (c.f. Eq. (5)). Both methods are illustrated in the simulated example provided in Section 7 below.

While imputation is a common practice, most ad hoc methods usually adopted rely on a number of, usually uncertain, assumptions. Furthermore, the resulting index of abundance (and hence the results of any subsequent stock assessment) may be highly sensitive to the imputation algorithm chosen. In these situations, several algorithms should be used to explore different assumptions about the undertaking stock and effort dynamics in the fishery (Campbell, 1998, 2004; Butterworth et al., 2003).

5. Weighting

The relative indices of annual abundance based on the parameter estimates obtained from a GLM may be biased when the number of observations in each spatial–temporal stratum varies (i.e. the data set is unbalanced) (Campbell, 2004). This is due to the fact that equal weight is given to each observation in the estimation procedure instead of giving equal weight to each region within each year. Thus, the annual indices based on the least-squares fit will be biased to favour those regions with the most number of observations. In practice this issue affects all analyses of commercial CPUE data as the distribution of effort across the fishery is usually highly heterogeneous owing to non-random (very targeted) commercial fishing strategies.

The weighting needed to be given to each observation to achieve an unbiased annual abundance index is not unique. Indeed, any weights that satisfy the condition that the sum of the individual weights given to each observation in each spatial–temporal stratum is the same for all similar strata type will ensure that all strata are treated equally. For the observations within each stratum, the weight assigned to each observation will itself ensure the importance of that observation.

Using the notation from previous sections, in the situation where there are n_{ijk} observations in the stratum for year i, quarter j and region k a simple weight for each observation in this stratum would be:

$$\text{weight}_{i,j,k} = \frac{N_{\text{obs}}}{N_{\text{strata}}} \cdot \frac{1}{n_{ijk}} \tag{13a}$$

where N_{obs} is the total number of observations and N_{strata} = NyNqNr is the total number of strata in the model. The first ratio is included to ensure the mean scaling across all observations is maintained. For example, where $n_{ijk} = n$ for all i,j,k th weight for each observation is equivalent to 1. An alternative weighting factor for each observation may be based on the corresponding effort of that observation (e_{ijk}), again scaled by the average effort across all strata:

$$\text{weight}_{i,j,k} = \frac{E_{\text{obs}}}{N_{\text{strata}}} \cdot \frac{1}{e_{ijk}} \tag{13b}$$

This would be most appropriate in situations where aggregated catch and effort observations (e.g. monthly CPUE) are used in the analysis. Punsley (1987) used a similar approach but recommended taking the log of the effort for each observation.

The inclusion of prior weights changes the relative influence of each observation in the fitted GLM. While the resulting parameter estimates depend only on the relative size of the individual weights, the standard errors of these estimates depends on their absolute sizes. It is therefore good practice to use weights that sum to N_{obs}, as suggested above, as these retain the scale of the default unit weights. The use of prior weights will also alter the distribution of the residuals and outliers. While further work to clarify the nature of these changes would be useful, nevertheless standard techniques that are useful for comparing or assessing model fit remain appropriate (Hardin and Hilbe, 2012). It can also be noted that similar parameter estimates are obtained when a full three-way interaction (i.e. a Y^*Q^*R term) is fitted to the data, and each observation is either appropriately weighted or not weighted. In this situation, the number of parameters equals the number of spatio-temporal strata and parameter estimates can be found which are independent of the relative number of observations in each region (Campbell, 2004).

6. Standard errors

Following the procedure described for the construction of the temporal abundance index, an associated standard error can be constructed in a similar manner. For example, the expected value of the positive CPUE in the i-th year, j-th quarter and k-th region (c_{ijk}) on the log scale is given by:

$$E[\log(c_{ijk})] = \mu_{\beta,ijk} = \beta_0 + y_{\beta,i} + q_{\beta,j} + r_{\beta,k}$$

and the associated variance is:

$$\mathrm{var}[\log(c_{ijk})] = \sigma^2_{\beta,ijk} = \mathrm{var}(\beta_0 + y_{\beta,i} + q_{\beta,j} + r_{\beta,k})$$

After taking the exponent, the corresponding 95-th percentile confidence limits for c_{ijk} can be written as:

$$\mathrm{CL}_{\beta,ijk} = [\exp(\mu_{\beta,ijk} - 1.96\sigma_{\beta,ijk}), \quad \exp(\mu_{\beta,ijk} + 1.96\sigma_{\beta,ijk})]$$

Similarly, from the binomial model the corresponding confidence limits of the expected value of the probability of a successful catch, p_{ijk}, being obtained in the same stratum can be written as:

$$\mathrm{CL}_{\alpha,ijk} = [\mathrm{InvLogit}(\mu_{\alpha,ijk} - 1.96\sigma_{\alpha,ijk}), \ \mathrm{InvLogit}(\mu_{\alpha,ijk} + 1.96\sigma_{\alpha,ijk})]$$

The ratio $d/3.92$, where d is the difference between the upper and lower values of these confidence limits, can then be used to approximate value of the standard errors associated with each expected values p_{ijk} and c_{ijk}. Furthermore, assuming that p_{ijk} and c_{ijk} are independent, and using the relation that for two independent random variables X and Y that $\mathrm{var}(XY) = \mathrm{var}(X)\cdot\mathrm{var}(y) + \mathrm{var}(X)\cdot E(Y)^2 + \mathrm{var}(Y)\cdot E(X)^2$ (Goodman, 1960), a variance can be determined for the standardised CPUE in each stratum, $\mathrm{stdCPUE}_{ijk} = p_{ijk}\cdot c_{ijk}$, (c.f. Eq. (4)). The standard errors associated with the abundance index Eq. (5) is then given by:

$$\mathrm{se}[B(i,j)] = \sqrt{\sum_{k=1}^{\mathrm{Nr}} \mathrm{A}_k^2 var(\mathrm{stdCPUE}_{ijk})}$$

The standard errors associated with the annual index, Eq. (6), can likewise be calculated. For those strata where a standardised CPUE needs to be imputed, a corresponding variance will also need to be inferred.

The approach outlined here for estimating standard errors for the abundance indices based on use of a delta-type two step GLMs is more direct than the first order Taylor series expansion approximation described by Shono (2008) or the alterative non-parametric bootstrap approach more recently adopted by other analysts (e.g., Winker et al., 2013).

7. Simulated example

In order to demonstrate both the inclusion of interaction terms in a standardising model and the biases which may result if (i) the annual abundance index is not correctly calculated, (ii) the standardising model is mis-specified, or (iii) an appropriate weighting is not used for each observation fitted to the model, a number of weighted and non-weighted delta-GLM models were fitted to simulated data and the results were used to calculate the associated abundance index.

7.1. Data

The simulated data were based on those obtained from the logbook data pertaining to the effort deployed per set and the associated catch of broadbill swordfish (*Xiphias gladius*) in the Australian pelagic longline fishery operating off eastern Australia. This fishery, which commenced in the mid-1980s, primarily targeted only yellowfin and bigeye tunas up until 1995 when effort reached around 3.8 million hooks. However, the catch of swordfish increased almost 10-fold to 632 mt in 1996 when the direct targeting of swordfish commenced. The fishery continued to expand further offshore given apparent high availability, and by 1999 effort and the catch of swordfish had increased to 10.3 million hooks and 2823 mt respectively. While effort continued to increase, reaching a peak of 12.8 million hooks in 2003, an apparent inshore depletion saw the swordfish catch decline to 2026 mt in 2000 and 1176 mt in 2006 (Campbell and Hobday, 2003). Effort has stabilised at around 6.8 million hooks with the catch of swordfish averaging around 1150 mt in recent years after a restructure of the fishery in 2008. There has also been a number of changes in operational practices given these changes, together with changes in the number of vessels participating in the fishery. While only a single longline set is generally deployed per day, the average number of hooks deployed has increased from around 700 hooks in 1995 to 1440 hooks in 2012 (when the average length of the mainline was around 52 km). The average number of days fished by vessels in a year has also increased from 50 days in 1996 to a peak of 115 days in 2009 and was around 101 days in 2012. The spatial extent of the fishery has also changed, initially increasing from 156 one-degree squares in 1995 to a peak of 264 squares in 2003 then declining to less than 150 squares in recent years (Campbell, 2013).

For the purposes of this study, the data were limited to the period from July 1997 (when a new logbook was introduced and began collecting much of the information on gear settings used for standardising effort) to December 2003 (when the fishery reached its maximum spatial extent) and the core spatial zone where swordfish are caught bounded by 23°S and 34°S and from the coast to 165°E. Within this zone, the catch and effort data was limited to the 73 one-degree squares where the total catch of broadbill swordfish was greater than 500 fish. Three additional squares were included to make the zone contiguous. Each observation within this zone was then assigned to one of five regions based on the distance from the coast, the proximity to inshore or offshore seamounts, and a simple spatial stratification of the mean nominal catch rates of all observations within each one-degree square. Finally, the data were further limited to those vessels which fished for four or more years and deployed more than 200 sets. The data set consisted of the catch and effort observations for 34,170 individual longline sets after removing some observations having anomalous values.

Observations were available for 129 of the 130 possible year–qtr–region spatial–temporal strata (6.5 years, four quarters and five regions). In this sense we can say that the data is highly structurally balanced but as the number of data points within each stratum varies widely (between 2 and 932) the data set can be said to be distributionally unbalanced. The missing stratum relates to the stratum $(Y, Q, R) = (1997,3,5)$. This means that a value of the standardised CPUE in this missing stratum needed to be imputed for models which included a full three-way interaction between these three effects. However, no values needed to be imputed for models which include only two-way interactions as data exists for all such two-way combinations of these effects. This can be seen in Fig. 1 which plots the nominal CPUE for each combination of spatial–temporal effects.

A delta-GLM was fitted to the data with the linear effects model for each stage based on the following linear combination of explanatory variables:

$$\begin{aligned}\eta = \beta_0 &+ \sum_{i=1}^{\mathrm{Ny}}\sum_{j=1}^{\mathrm{Nq}}\sum_{k=1}^{\mathrm{Nr}}\beta_{ijk}Y*Q*R_{ijk} + \sum_{a=1}^{\mathrm{Nbait}}\beta_a\mathrm{Bait}_a \\ &+ \sum_{b=1}^{\mathrm{Nstart}}\beta_b\mathrm{Start}_b + \sum_{c=1}^{\mathrm{Nlights}}\beta_c\mathrm{Lights}_c + \sum_{d=1}^{\mathrm{Nsoi}}\beta_d\mathrm{SOI}_d \\ &+ \sum_{e=1}^{\mathrm{Nsst}}\beta_e\mathrm{SST}_e + \beta_m.\mathrm{Moon}\end{aligned} \quad (14)$$

where: β_{ijk} is the effect of the three-way interaction between the i-th-year, the j-th quarter and the k-th region, β_a is the effect of the a-th category of bait-type used (Bait), β_b is the effect of the b-th category of start-time (Start), β_c is the effect of the c-th category of light-stick usage (Lights), β_m is the effect of the m-th category

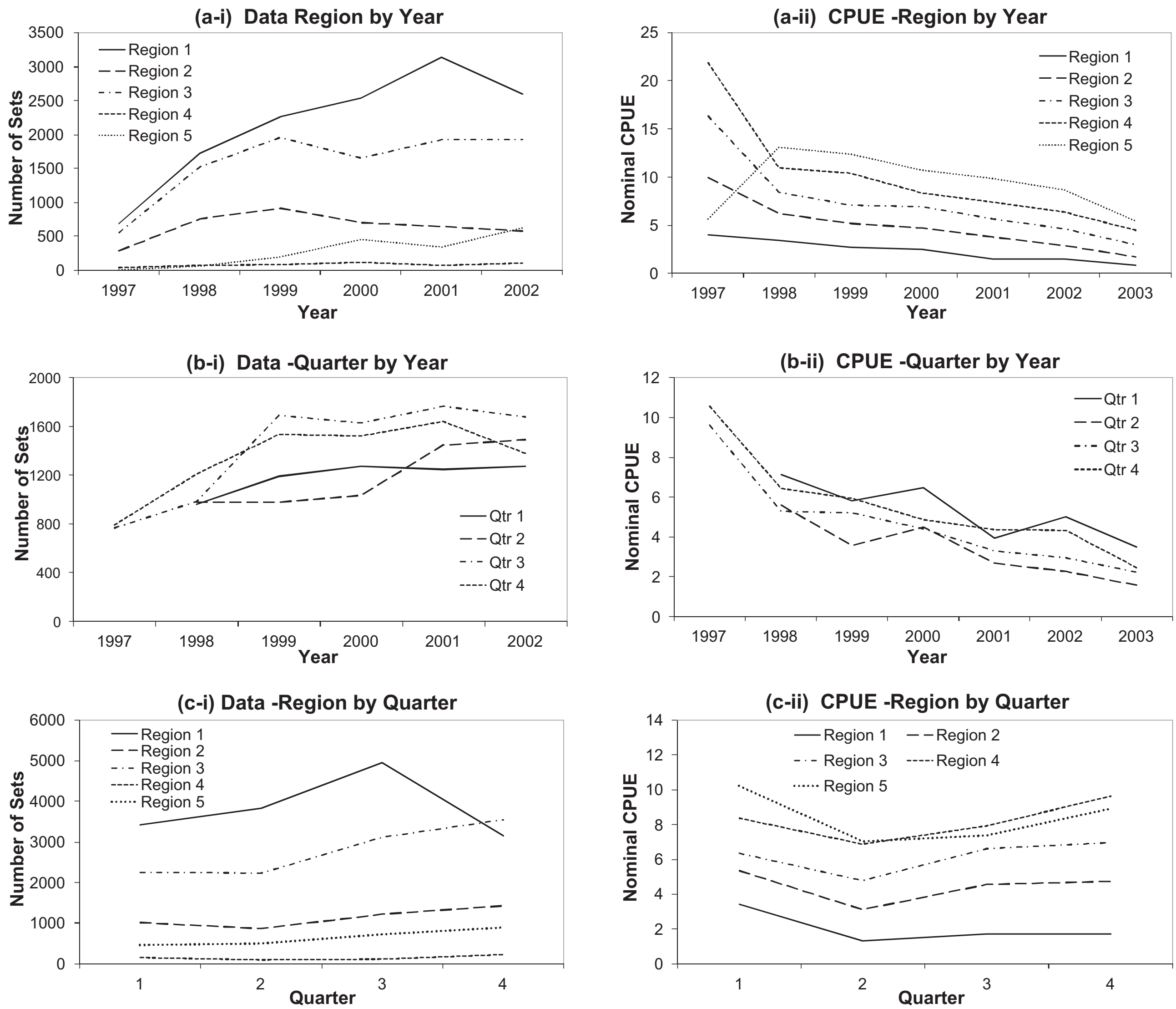

Fig. 1. (i) Number of observations (longline sets) and (ii) nominal CPUE within each combination of (a) Year and Quarter, (b) Year and Region and (c) Quarter and Region.

of the southern oscillation index (SOI), β_n is the effect of the n-th category of the sea-surface temperature (SST), and β_m is the effect of the daily moon phase (Moon).

All effects (except Moon which was fitted as a continuous variable) were fitted as categorical variables with the levels of each effect shown in Table 1. For the binominal model, the variable Logit H was fitted as a quadratic and added to the linear combination of explanatory variables while the gamma distribution with a log-link was used to model the size of the positive catch. For both models, each observation was weighted using Eq. (13a) and each model was fitted using the SAS GENMOD procedure (SAS Institute Inc., 2008). (Note that the effects included in the model given by Eq. (14) are a subset used for the purposes of this worked example of the full set of effects fitted in the analysis of the data for this fishery described in Campbell, 2012.)

After fitting each model to the data, the resulting model parameters were stored and used to generate an alternative data set. While the series of zero and positive catches were retained for each observation in the original data, where a positive catch was observed the size of the catch was modelled using the linear predictor given by Eq. (14). This generated a fully deterministic set of data based on a known set of parameters, and while the assumed random error component in the observed data has been removed it nevertheless preserves the realism of the original data set making it well suited for the following analyses.

7.2. Fitted models

Eq. (14) was fitted to the simulated data but the three-way interaction between the year, quarter and region effects was replaced by one of five functional forms $f(Y,Q,R)$ of these three effects as shown in Table 2. The non spatial–temporal effects, known henceforth as the auxiliary effects, were included in all fitted models except model 5. Two sets of model fits were obtained where: (i) each observation was weighted using Eq. (13a) and (ii) each observation was unweighted. Estimates of the standardised value of the dependent variable within each stratum and the associated confidence limits were calculated using the ESTIMATE statement in the GENMOD procedure (SAS Institute Inc., 2008).

7.3. Construction of biomass indices

The relative index of abundance for each quarter was calculated using Eq. (5) for each fitted model and compared. The size of each region, A_k, was determined as the number of one-degree squares (or parts thereof) assigned to that region. Furthermore,

Table 1
Variables included in the fitted delta-GLM models.

No.	Standardising variable	Model parameter	Category levels	Category definition
Statistical effects				
1	Year	Y	1–7	1997–2003
2	Quarter	Q	1	Jan–Mar
			2	Apr–Jun
			3	Jul–Sep
			4	Oct–Dec
3	Region fished	R	1–5	Refer to Campbell (2012)
Fishing strategy effects				
4	Percentage of hooks deployed with a lightstick	Lights	1	0%
			2	1–19%
			3	20–39%
			4	40–59%
			5	60–79%
			6	80–99%
			7	100%
5	Bait type used	Bait	1	Squid, dead
			2	Yellowftail scad, alive
			3	Pilchard, dead
			4	Other, dead
			5	Other, alive
			6	Mixed species, dead
			7	Mixed species, alive and dead
6	Start time of set	Start	1	Before 4am
			2	4am–8am
			3	8am–noon
			4	noon–4pm
			5	4pm–8pm
			6	8pm–midnight
Environmental/oceanographic effects				
7	Southern-oscillation Index	SOI	1	Normalised SOI < −1.0
			2	Normalised SOI between −1.0 and −0.3
			3	Normalised SOI between −0.3 and 0.3
			4	Normalised SOI between 0.3 and 1
			5	Normalised SOI > 1.0
8	Sea-surface temperature	SST	1–5	As for SOI
9	Daily moon phase	Moon	Continuous	ABS[cos(π*phase/29)][#]

[#] ABS = absolute value.

due to there being a missing stratum in the data, for those models which included a three-way spatial–temporal interaction (i.e. models 5 and 6) several alternative imputed values for this stratum were generated using the following two sets of procedures:

First, the following ad hoc infill method was used where $V(y,q,r)$ refers to the standardised CPUE in year y, quarter q and region r. For each year ($i \neq 1997$) the mean of the ratio of the value in region 5 to the value in each other region was calculated and then the overall mean, M, across all years was determined, i.e.

$$M = \frac{1}{6}\sum_{i=1998}^{2003}\left[\frac{1}{4}\sum_{k=1}^{4}\frac{V(i,3,5)}{V(i,3,k)}\right]$$

Table 2
Specification of the various models fitted to the simulated data, together with the methods employed to calculate the resulting abundance index and impute the value of the standardised CPUE in missing strata.

Model specifications				Index specifications	
Model	f(Y,Q,R)	Auxiliary effects	Number of parameters	Abundance index	Imputation method
1	$Y+Q+R$	Included	49	Eq. (12)	na*
1-simple		Included	49	Eq. (15)	na
2	$Y+Q^*R$	Included	60	Eq. (12)	na
2-simple		Included	60	Eq. (15)	na
3	Y^*Q+Q^*R	Included	79	Eq. (12)	na
4	$Y^*Q+Y^*R+Q^*R$	Included	114	Eq. (12)	na
5	Y^*Q^*R	No	130	Eq. (12)	na
6-a	Y^*Q^*R	Included	162	Eq. (12)	Infill
6-b		Included	162	Eq. (12)	Model 1
6-c		Included	162	Eq. (12)	Model 2
6-d		Included	162	Eq. (12)	Model 3
6-ref		Included	162	Eq. (12)	Model 4

* na = not applicable.

The value of $V(1997, 3, 5)$ was then determined by solving this equation for the year 1997 alone, i.e.

$$V(1997,3,5)=\frac{4M}{\sum_{k=1}^{4}\frac{1}{V(1997,3,k)}}$$

Second, the standardised CPUE value for the missing stratum was taken from the value estimated for this stratum based on fitting several simpler standardising models. The following four GLM models were used for this purpose:

Model 1:	$f(Y,Q,R)=Y+Q+R$
Model 2:	$f(Y,Q,R)=Y+Q^*R$
Model 3:	$f(Y,Q,R)=Y^*Q+Q^*R$
Model 4:	$f(Y,Q,R)=Y^*Q+Y^*R+Q^*R$

After construction of the quarterly biomass indices, an annual biomass index was then constructed based on Eq. (6b). The calculation of this index was limited to the years from 1998 to 2003 due to the lack of data for the first two quarters in 1997. There are differences in what terms need to be included in the calculation of the relative abundance index due to the differences in the manner that the year, quarter and region effects are included in the different models listed in Table 2. It was noted earlier that the exponent of the year effect alone can be used as the relative abundance index when a single step GLM is used, and the fitted model includes no interactions. If this simple result is extended to the delta-GLM models, then from Eq. (8) the relative index would reduce to:

$$I(i,\text{ref})=\frac{\exp(y_{\beta,i})\,\text{InvLogit}(\alpha_0+y_{\alpha,i})}{\exp(y_{\beta,\text{ref}})\,\text{InvLogit}(\alpha_0+y_{\alpha,\text{ref}})} \tag{15}$$

The relative abundance index was calculated using both Eq. (8) (generalised using Eq. (12) to include interactions) and Eq. (15) to ascertain the size of the error which may result if this simpler but incorrect form of Eq. (8) is used to calculate the abundance index, for those models which included only a year effect and no interactions with this effect (i.e. Models 1 and 2). The indices based on Eq. (15) are known as the I_{simple} indices.

7.4. Biases

To ascertain the bias resulting from fitting either: (i) a simpler model than that used to generate the data, or (ii) an unweighted model instead of a weighted model, or (iii) the use of the simpler but incorrect equation to calculate the annual index, the following measure of bias was calculated:

$$\text{Bias}(i)=\left(\frac{I(i)}{T(i)}-1\right)*100\% \tag{16}$$

where $I(i)$ is the value of abundance index in time step i for the selected model and $T(i)$ is the value of the index for the true model. A mean bias was also calculated across all time-steps.

$$\text{Mean bias}=\frac{1}{\text{Ny}}\sum_{i=1}^{\text{Ny}}\text{abs}\left(\frac{I(i)}{T(i)}-1\right)*100\% \tag{17}$$

Both quarterly and annual biases were calculated over the period 1998–2003. All indices were scaled so that the average value across this period was equal to 1 (i.e. $B(\text{ref})=\sum_{i=1}^{\text{Ny}}B(i)/\text{Ny}$ in Eq. (7)) as the index calculated for each model can be relative to a different set of standardising effects, to aid comparison across models.

Similar comparisons of biases based on the influence of individual effects can also be ascertained between each model and the true model. As before, a mean of the absolute values of the bias for each region can be calculated across all years.

Finally, for each covariate included in the model a direct comparison of the influence of each level can be made between any selected model M and the true model T. For this purpose the ratio between the respective parameter values for covariate c and level l was calculated:

$$\text{Bias}(M,\quad c,\quad l)=\left[\frac{\exp(\beta_{M,c,l})}{\exp(\beta_{T,c,l})}-1\right]*100\%$$

where $\beta_{M,c,l}$ is the associated parameter value in the linear model. A mean bias across all levels was then calculated for each covariate and model.

8. Results

The weighted fit of model 6 (i.e. the model including the full three-way interaction between year, quarter and region and all other auxiliary effects, Table 2), as expected, was found to be the only fitted model for which the estimated parameters where the same as those used to generate the data. This model will henceforth be known as the 'true' model. However, although there is only a single 'true' model, it was not possible to calculate a single 'true' abundance index covering all quarters in the data due to the need to impute a value of the standardised CPUE in the single unobserved strata. Instead, the index which uses the estimate from fitting model 4 to the data (i.e. the model most similar to the true model) to impute this missing value was taken as the 'reference' index (and used in place of the true index referred to in the previous section).

8.1. Quarterly abundance indices

Plots of the quarterly abundance indices and associated bias for each of the six main models listed in Table 2 are shown in Fig. 2 while the mean of the absolute bias for each quarter over this period is shown in Table 3. Each index is scaled so that the mean over the 24 quarters from 1998 to 2003 is equal to 1.

A number of features can be noted. First, the difference between the fitted model and the reference model generally decreases as the complexity of the fitted model increases (i.e. the number of model parameters increases) for the models which do not include the full three-way interaction. For example, the mean bias is greater than 10% for weighted models 1 and 2, decreasing to 4–5% for models 3 and 4. Second, there is a temporal trend (from positive to negative) in the quarterly bias for the simpler models. Fitting a trend linear line $y=ax+b$ through the weighted indices indicates that $a=-0.68$ for model 1, $a=-0.64$ for model 2 and $a=-0.47$ for model 3 but is close to zero (0.03) for model 4. A consequence of this trend in the bias for these simpler models is that the estimated decline in abundance is greater than for the reference model (and the assumed underlying reality). Third, there is considerable bias for model 5 (full interaction term but no auxiliary effects), indicating the importance of incorporating the auxiliary effects in the overall model. However, the bias for model 5 is the same for both the weighted and unweighted fits illustrating the point that the result for both types of model fits will be the same when a full three-way interaction is fitted to the data. In this instance the weighting becomes redundant. Fourth, for model 6 (full interaction plus auxiliary effects) there is a small (1.87%) difference between the unweighted and the weighted reference models. The number of model parameters is now greater than the number of strata resulting in the more complex relation between the model parameters and the observations in each stratum. Finally, the difference between the weighted and unweighted model fits is generally small with neither consistently giving a better fit for the first four models.

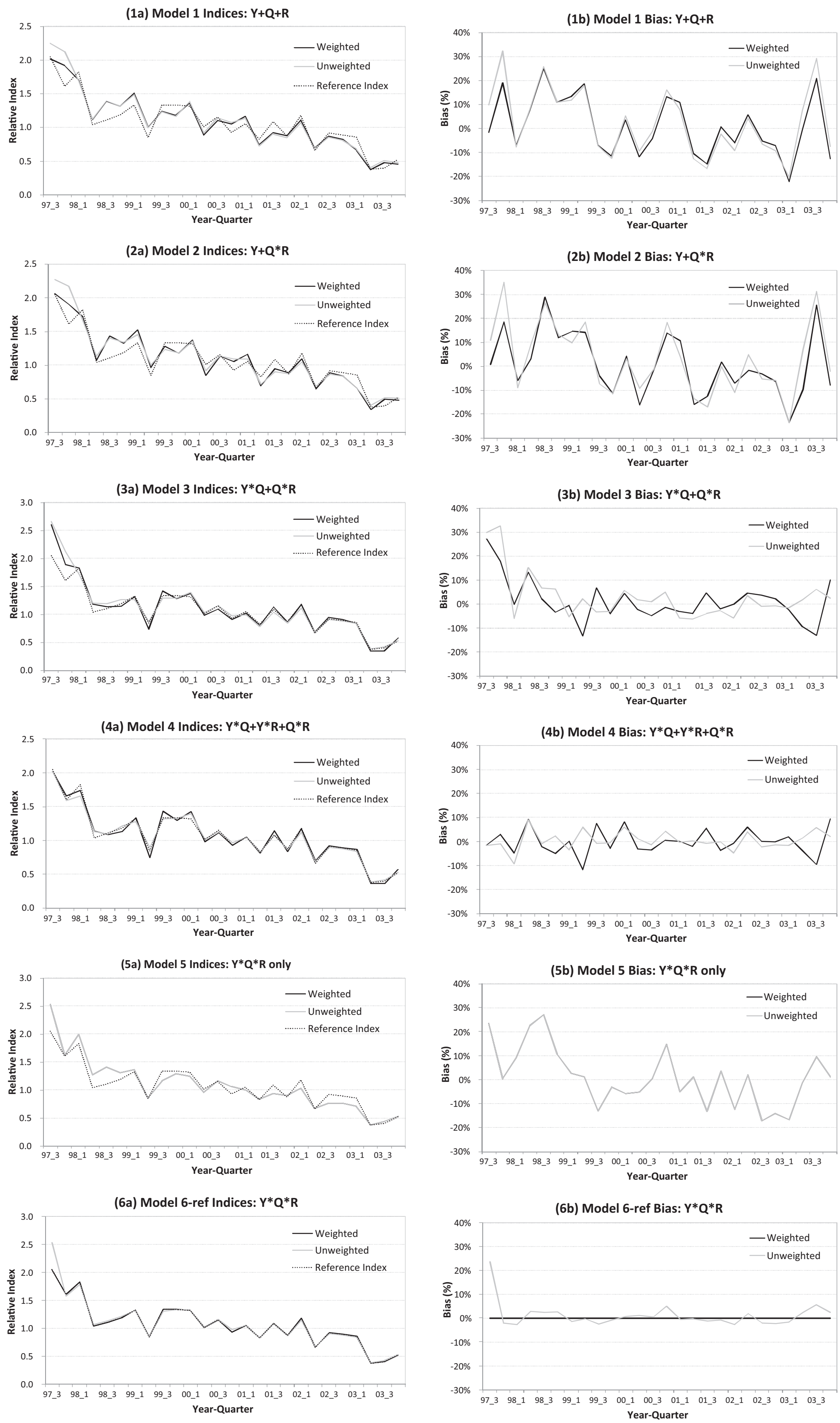

Fig. 2. Quarterly time-series of (a) the abundance index, and (b) the bias relative to the reference model for each fitted model. Results for both weighted and unweighted GLMs are shown. Each abundance index is scaled so the mean value between 1998 and 2003 is equal to 1.

Table 3
Mean of the absolute bias in the abundance index for each quarter for the six main models listed in Table 2 using model 6-ref as the reference model. The mean of the absolute bias for each year over this period is also shown.

Model	Mean quarterly bias		Mean annual bias	
	Weighted (%)	Unweighted (%)	Weighted (%)	Unweighted (%)
1	10.47	11.14	4.12	3.93
1-simple			17.00	18.23
2	10.66	10.92	4.57	3.98
2-simple			18.23	18.88
3	4.85	4.33	2.35	3.17
4	4.21	2.92	0.88	1.05
5	8.92	8.92	6.50	6.50
6-ref	0.00	1.87	0.00	1.34

8.2. Precision

The calculated coefficient of variation (CV) associated with the estimated abundance index for each quarter is shown for the weighted models 1–6 in Fig. 3. For those models where the dependent variable was imputed for strata (1997,3,5) the associated variance was set equal to the variance estimated for the adjacent spatial strata (1997,3,4). It was also found that the variances of the estimates for the Binomial stage of models 4–6 could not be evaluated as the negative of the Hessian was not positive definite. This occurred due to the presence of strata where the observed probability of obtaining a positive catch was 1. Removal of these observations allowed the variances to be calculated for the remaining strata and the variance for each removed strata (although having an observed variance of zero) was set equal to the variance estimated for that strata using the last preceding model where such a value was estimable.

There is an overall increase in the CV of the related abundance indices as the number of parameters in the fitted model increases. Mean CVs increase from 1.21% for model 1 with 49 parameters to 2.59% for model 6 with 162 parameters, and the relationship between mean CV and the number of model parameters has an R^2 of 0.97 for all models including the auxiliary effects. Despite the increase in model parameters, the CV for model 5 which does not include the auxiliary effects, was found to be the third lowest, indicating that the precision associated with the estimates of the fixed spatial–temporal effects is greater than that for the auxiliary effects. This loss of precision may be associated with the possible multicollinearity amongst some of the auxiliary variables (e.g. SST and SOI) which tends to increase the standard errors of the affected coefficients (Hocking, 1976). In such situations the test of the hypothesis that the coefficient is equal to zero may lead to a failure to reject a false null hypothesis of no effect of the explanatory effect, a type 2 error (Belsley et al., 1980).

8.3. Imputed values

For model 6, a comparison of the differences in the index for the third quarter of 1997 using the different methods for imputing the standardised CPUE for the missing strata in this quarter is shown in Fig. 4a. Compared to the reference model, the index using the ad hoc infill method has an associated difference around 28%, while the indices using the imputed value from the two simplest GLMs each have a smaller associated difference of around 20%. On the other hand, the difference associated with the index using the imputed value from model 3 is the largest, at around 31%. This last result is surprising, as the index based on the imputation model having the closest match to the reference model may have been expected to have performed better than the others, i.e. been the most similar to the reference index.

Several data issues are driving this result. First, the nominal CPUE for region 5 during 1997 is anomalously low (c.f. Fig. 1a) and

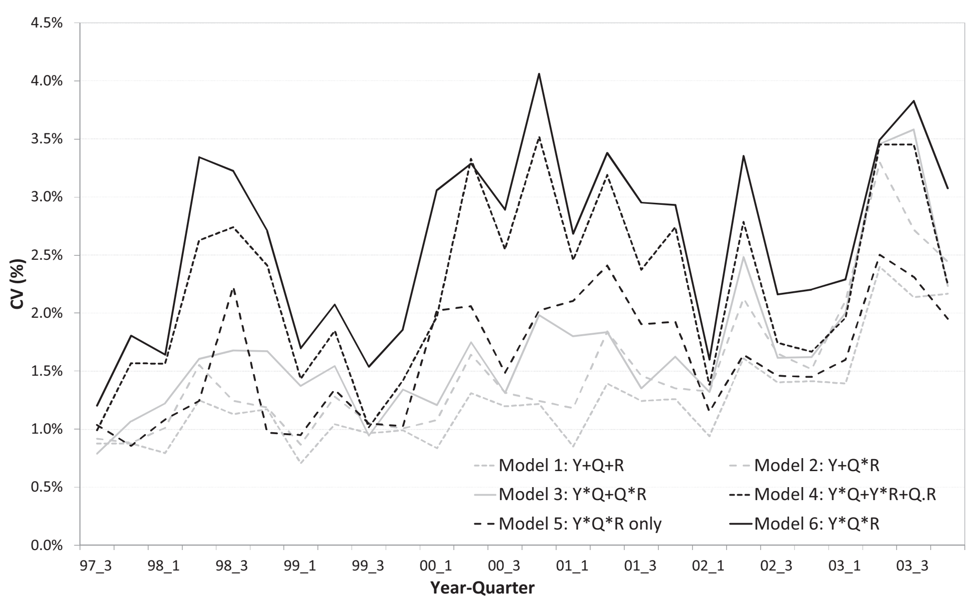

Fig. 3. Estimated coefficients of variation associated with the quarterly abundance index for each fitted model.

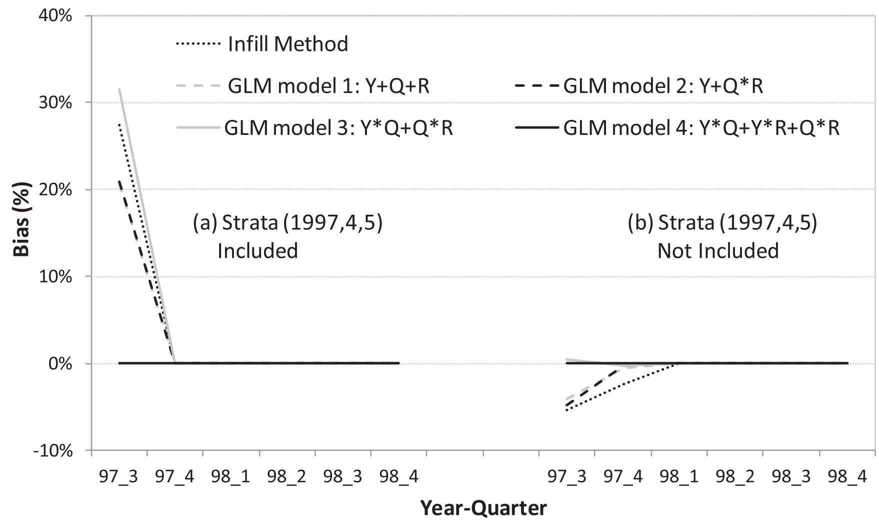

Fig. 4. Comparison of the differences in the abundance index for the third quarter of 1997 using the five methods for imputing the standardised CPUE for the missing strata in this quarter. Results are shown for when there are one or two missing strata.

is not consistent with the trends seen in the other four regions. This low value decreases the mean CPUE over all regions for this year and the associated model index. On the other hand, anomalous trends are not seen in either of the year–quarter and quarter–region CPUE values (c.f. Fig. 1b and c). As noted previously, there are no data for quarter 3 for region 5 during 1997and only two observations for quarter 4 so the amount of data used to estimate the CPUE-based density in this region in 1997 is very poor. Second, the inclusion of the Y^*R interaction in model 4 results in the low value of the CPUE for region 5 in 1997 being included in the calculation of the abundance index giving the somewhat anomalous observations in this stratum a high leverage on the result. However, this leverage is constrained when this interaction term is excluded from the fitted GLM. This explains why the index for the third quarter in 1997 for the reference model (based on an imputing model which includes the Y^*R interaction) is substantially different to the indices based on the other models.

This result raises the question as to whether the two observations for region 5 (which represents 26% of the total assessed area) during the fourth quarter of 1997 are truly representative of abundance in this stratum and, at least, suggests that a sensitivity analysis be undertaken where the previous analyses are repeated with these two observations removed from the data set. In this case there is a need to impute the standardised CPUE in region 5 for both quarters in 1997. Again, a comparison of the differences in the resulting index during 1997 for the different imputation methods is shown in Fig. 4b and all differences are seen to be appreciably smaller (<6%). Furthermore, the difference pertaining to the index which incorporates the two imputed CPUE values based on model 3 (i.e. the model closest to the true model) is now the smallest (<0.5%) for both quarters. This result appears then to be more consistent both with the overall trends in the data and the expected convergence of indices as the model used to impute the missing standardised CPUE values converges on the true model.

The decision to include or exclude the two observations for region 5 in 1997 also impacts on the quarterly index deemed appropriate for indexing the change in biomass during this early period of the fishery. The decline in the abundance index during 1997 based on the model which includes these data is appreciably lower than that based on the model which excludes them (Fig. 5) and would consequently influence the stock status inferred from the assessment using these indices. This example highlights that careful note needs to be taken of outliers and unusual features in the distribution of the observations in the data fitted to the GLM as these may have a marked impact on the resulting index.

8.4. Annual abundance indices

The bias in the annual index for the years 1998–2003 relative to the reference model is shown in Fig. 6 for each of the six models listed in Table 2 while the mean of the absolute bias for each year over this period is also listed in Table 3. The indices have been scaled so that the mean annual value over the period 1998–2003 is equal to 1 and again results are shown for both the weighted and unweighted GLM analyses.

All models show some degree of bias for most years, though the overall difference in the bias between the weighted and unweighted models is not large, with the largest differences of around 5% seen in 2003 for models 1 and 2. This result concurs with that found for the quarterly indices. The bias again decreases as model complexity increases for the weighted models, with the range and mean annual bias being around 14% and 4.5% respectively for models 1 and 2, 6.9% and 2.4% respectively for model 3, and 3.3% and 0.9% respectively for model 4 (Table 3). However, while there does not appear to be an annual trend in the bias for models 3 and 4, the trend in the bias for the simpler models 1 and 2 (i.e. with a fitted trend line giving $a = -2.51\%$ and -2.77% per year respectively)

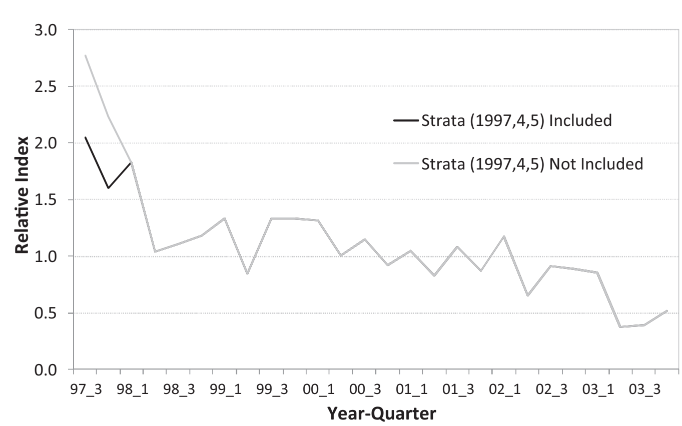

Fig. 5. Comparison of the quarterly time-series of the abundance index for the reference model fitted to the data with either one or two missing strata.

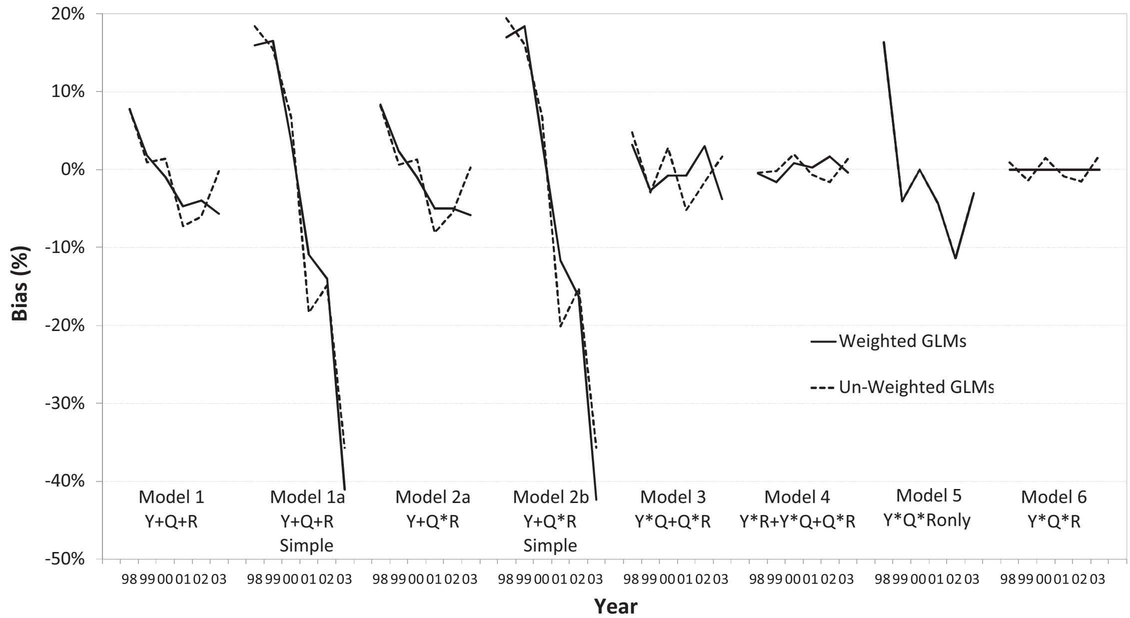

Fig. 6. Annual time-series of the bias in the abundance index for each model fitted to the data. Results for both weighted and unweighted GLMs are shown.

implies that the abundance is over-estimated in the initial years and under-estimated in the latter years. This would result in the estimated decline in overall abundance over the six years (~65%) being greater than the underlying reality (~60%). The larger range and trend in the bias for the simpler models is likely due to the non-inclusion of the interaction terms (especially the Y^*R interaction mentioned previously) in these models. These terms help account for trends and changes in the distribution of the stock over time.

The bias for the two I_{simple} (Eq. (15)) abundance indices for models 1 and 2 is appreciably higher, varying from nearly 20% in the early years to around 40% in the last year. There is also a very large overall annual trend in these biases (with a fitted trend line giving $a = -11.2\%$ and -11.9% respectively) which results in a large 20% over-estimation of the decline in the underlying biomass. This result provides some measure of the error that can occur when a simple (but incorrect) equation is used to calculate the annual abundance index. The mean annual bias for model 5 (which excludes the auxiliary effects) is also high at 6.5% and again the large trend in the annual bias would result in an over-estimation of the decline in biomass. Again, this result underlies the importance of including auxiliary covariates in the fitted model.

8.5. Bias related to individual effects

A comparison of the annual bias in the biomass estimated in each region for each of the weighted GLMs is shown in Fig. 7 while the mean bias across all years for each model is shown in Table 4. Both sets of results are based on fitting the models to the data excluding the two observations in region 5 in 1997 and use model 6-ref as the reference model. Consistent with the previous results, the bias decreases as the fitted model more closely approximates the true model. However, the bias for a given model varies considerably across each of the regions with, for example, the mean annual absolute bias for the biomass estimates generally being the highest in region 1 and lowest in regions 2 and 3. The reasons for these patterns in the bias remain uncertain, but may be related to the distribution of the number of observations and catch rates across each of the strata. For example, the CV of nominal CPUE across the four quarters is highest for region 1.

For each fitted GLM a comparison of the bias for each level for all included auxiliary effects is shown in Fig. 8 while the mean bias across all levels for each effect and model is shown in Table 5. The influence of each level for each auxiliary effect is relative to the level chosen to standardise the CPUE against and in Fig. 8 this chosen level can be identified where the bias between the chosen model and the true model is zero for all models. Several general results can be inferred. First, and as expected, the bias for the true model (6) is zero. Second, the bias for all auxiliary effects is relatively small for both the weighted and unweighted GLMs, with a mean absolute bias ranging between 1.85% and 2.33% for models 1 and 2 and between 1.26% and 0.42% for models 3 and 4. Third, the difference between the absolute bias for the weighted and unweighted GLMs is also small, with the mean difference of 0.43% (and range −0.42% to 1.64%) over all auxiliary effects for models 1–4. Except for three levels of the bait-type effect, the bias is higher for the weighted GLM. Fourth, the bias across each of the levels for most effects is not homogeneous with, for example, the bias mostly being negative for lower levels of light-stick usage and generally positive for higher levels of light-stick usage. There are also trends seen in the bias for the SOI and SST effects. The reasons for many of the above results remains uncertain but again may have to do with the relationship between the distribution of observations within each strata and the fitted parameters and/or confounding between the parameters of these effects which may be correlated.

Table 4
Mean absolute bias across all years in the abundance index estimated in each region for each of the weighted GLMs (models 1–6). Results are based on fitting models to the data excluding the two observations for region 5 in 1997 and use model 6-ref as the reference model.

Model		Region				
		1 (%)	2 (%)	3 (%)	4 (%)	5 (%)
1	$Y+Q+R$	15.68	4.23	5.81	4.96	8.79
2	$Y+Q^*R$	15.79	3.99	6.07	4.93	8.98
3	Y^*Q+Q^*R	13.90	4.51	4.76	7.03	5.73
4	$Y^*Q+Y^*R+Q^*R$	1.77	0.93	0.61	2.90	1.28
5	Y^*Q^*R only	11.02	5.76	4.94	6.43	9.78
6-ref	Y^*Q^*R	0.00	0.00	0.00	0.00	0.00

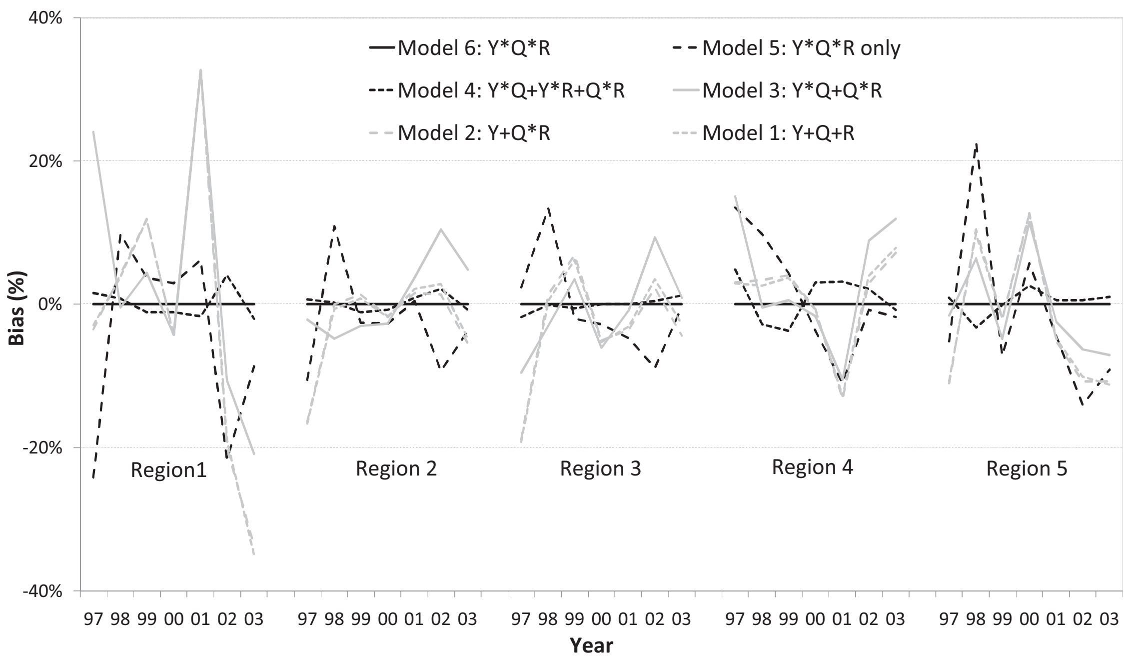

Fig. 7. Annual time-series of the bias in the estimated abundance in each region for each model fitted to the data with two missing strata and using model 6-ref as the reference model.

9. Structurally unbalanced datasets

The data set analysed in the above simulations was highly structurally balanced in that observational data were missing for only one of the 130 Year–Qtr–Region strata, though the data was highly distributionally unbalanced, with the number of observations within each of these strata varying between 2 and 932. The former property of the data lent itself to fitting a model which included a full three-way interaction term, as there was only a single stratum for which a standardised CPUE had to be imputed. However, the leverage of these latter data was found to be high as a consequence of this missing stratum, combined with the low number of observations in the only other strata within region 5 during 1997, and the modelled results in some instances were possibly misleading given the anomalous nature of these observations. A sensitivity analysis was conducted by repeating the analyses with these latter two data points removed from the data.

Potential biases in the parameter estimates can be overcome by an appropriate weighting when the data are distributionally unbalanced. On the other hand, the question of whether or not to include interactions in the fitted model when the data are structurally unbalanced remains pertinent. There is a need to impute a standardised CPUE for each missing stratum if a model with a full three-way interaction is fitted. In such instances, the results presented above suggest that a less complex model can be used to impute these missing values, noting, however, the need to check for anomalous values in each stratum. While imputing such values for a high number of strata can be achieved in practice, there is likely to be a point beyond which this can no longer be considered ‘best practice’ and it is necessary to employ a simpler form for the fitted model.

To investigate this question further, three additional data sets were generated where in each case the observations within a given number of Year–Qtr–Region strata were removed randomly from the above data to achieve a greater level of structural unbalance. Limiting the data to the years 1998–2003, in these three data sets a total of 6 (~5%), 22 (~18%) and 40 (~33%) of the possible 120 strata remained unobserved. The weighted models 1–4 and 6 were then fitted to these data and the resulting annual indices of abundance compared to a reference index based on fitting model 6 to the data set where observations existed for all strata. The standardised CPUE in each missing strata was estimated by fitting the next most

Table 5
Mean bias across all levels for each auxiliary effect for each standardising model (1–6).

Effect	Model 1 $Y+Q+R$	Model 2 $Y+Q^*R$	Model 3 Y^*Q+Q^*R	Model 4 $Y^*Q+Y^*R+Q^*R$	Model 6 Y^*Q^*R
(a) Weighted					
Light-stick usage	0.84%	0.88%	0.65%	0.69%	0.00%
Bait-type	1.97%	2.05%	1.67%	1.07%	0.00%
Set-start time	1.62%	1.27%	1.11%	1.32%	0.00%
SOI	4.34%	4.47%	1.04%	0.96%	0.00%
SST	2.94%	2.70%	1.05%	0.97%	0.00%
Moon phase	2.25%	1.45%	2.08%	1.47%	0.00%
(a) Unweighted					
Light-stick usage	0.59%	0.45%	0.34%	0.26%	0.00%
Bait-type	2.35%	2.46%	1.76%	0.87%	0.00%
Set-start time	0.28%	0.46%	0.12%	0.34%	0.00%
SOI	4.27%	3.95%	0.15%	0.30%	0.00%
SST	2.19%	2.59%	0.25%	0.56%	0.00%
Moon phase	1.60%	1.19%	0.44%	0.17%	0.00%

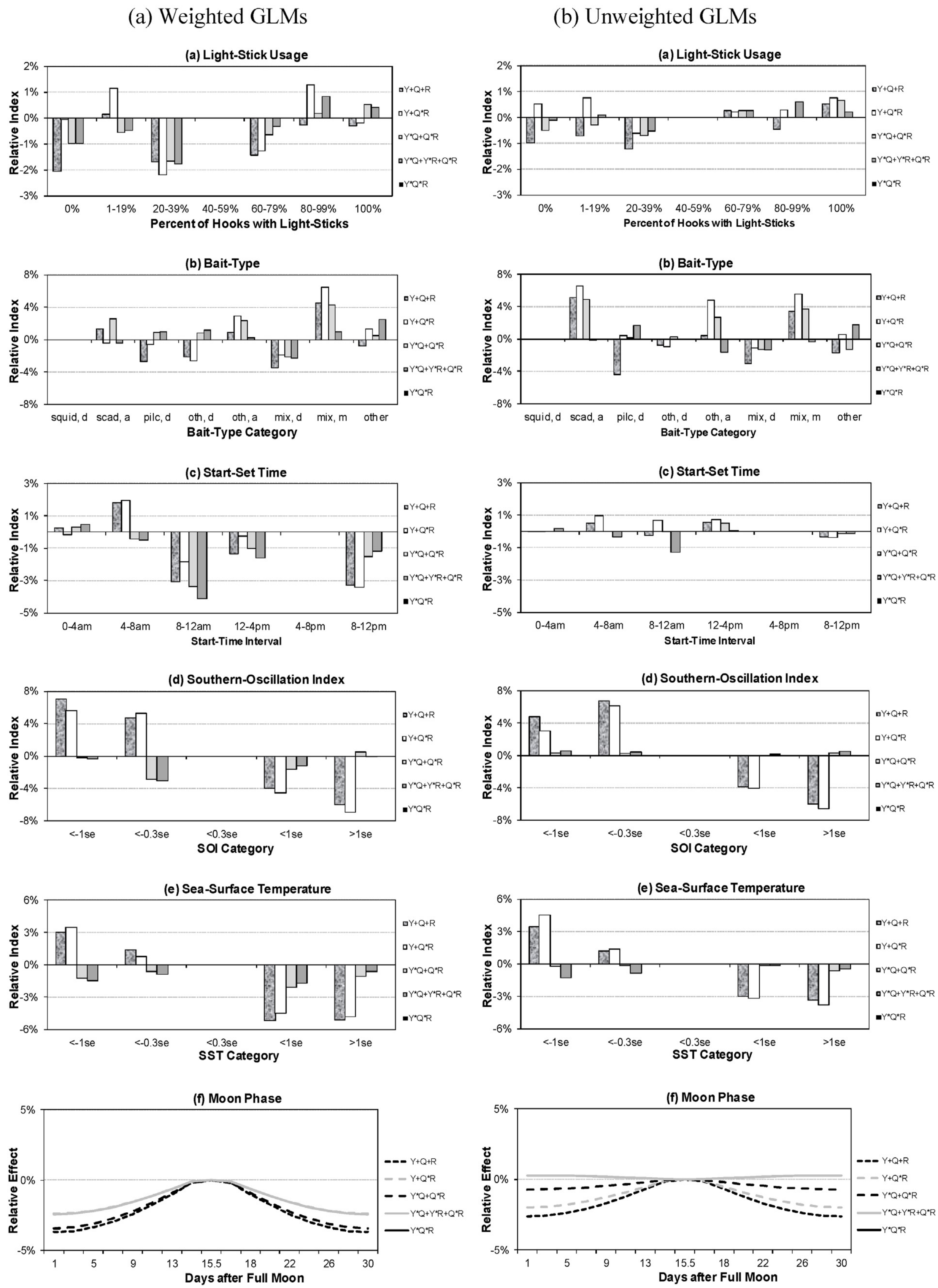

Fig. 8. Bias for each level of all covariates included in each of the models fitted to the data with two missing strata and using model 6-ref as the reference model. Results for both weighted and unweighted GLMs are shown.

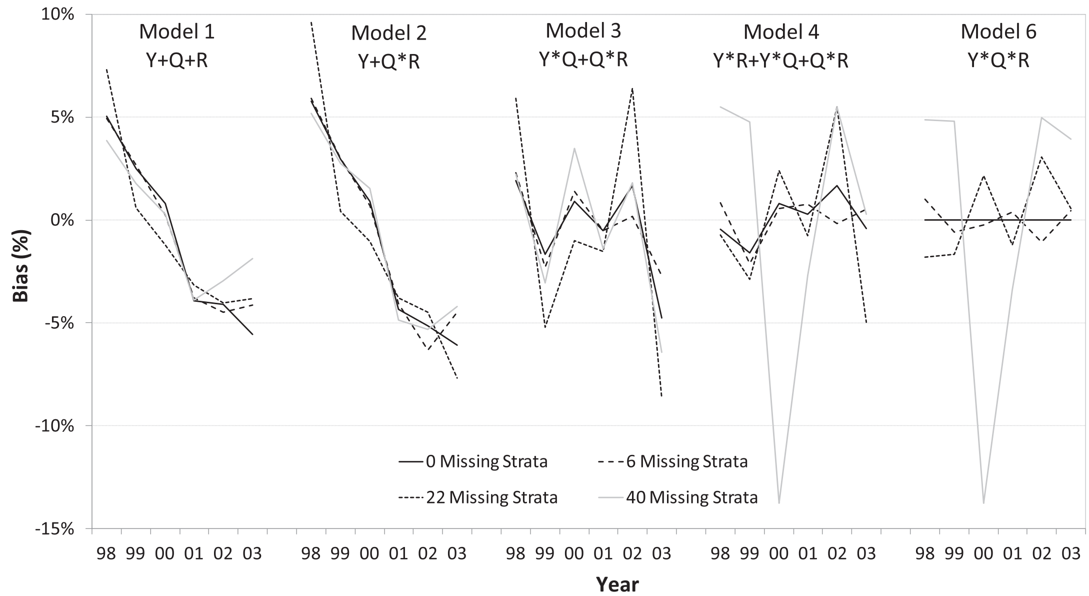

Fig. 9. Bias between a reference index and other abundance indices based on fitting the standardising model to four different sets of data each with observations removed from the shown number of *Y*–*Q*–*R* strata. The reference index uses the data with zero missing strata.

complex model which provided an estimate for that strata (in all instances this was model 4). The annual bias for each index against the reference index is shown in Fig. 9 while the mean, range and linear trend of the bias over all years are listed in Table 6.

The abundance indices for each model fitted to the data where the observations for 5% of the Year–Qtr–Region strata were removed are very similar to the reference indices, and there is a general decrease in the bias as the complexity of the model increases except for the simplest model (i.e. main effects only). The model including the three-way interaction again provides the least biased index, having the minimum values of the mean, range and linear trend of the bias over all six years. This suggests that this model may be the most optimal to use in the related GLM when only a small number of strata remain unobserved (i.e. the data set remains highly structurally balanced). However, the relative size and variability in the annual bias increases considerably for all models when the number of unobserved strata increases to 20% (Table 6). While the mean annual bias (1.7%) and range (4.9%) again remain smallest for the model including the three-way interaction, the trend in the annual bias (0.64%) is slightly greater than for the second most complex model. Finally, the results change substantially for the models fitted to the data with 33% of strata unobserved. The simplest model has both the smallest mean annual bias (2.4%) and smallest annual range (7.8%) in this situation. However, the annual trend for this model is the second highest, with the model including the three-way interaction again having the smallest trend. These results indicate that there is an obvious trade-off between the unbalanced nature of the data being analysed and the selection of the most appropriate fitted model. There is a greater need to fit simpler, less parameterised models as the data set becomes more structurally unbalanced. While this analysis has not investigated alternative options of imputing CPUE values for the missing strata (Campbell, 2004; Carruthers et al., 2010, 2011) further work on this issue is warranted.

10. Fixed versus random effects

All explanatory variables have been fitted as fixed effects in the models analysed in the previous sections. However, there has

Table 6
Mean, range and linear trend of the annual bias in the resultant abundance index between 1998 and 2003 for each standardising model. Results are shown for four sets of data each having the listed number of unobserved strata.

Bias	Model		Number of unobserved strata			
			0	6	22	40
Mean overall years	Model 1	$Y+Q+R$	4.5%	4.4%	4.1%	3.6%
	Model 2	$Y+Q^*R$	4.8%	4.7%	4.8%	4.7%
	Model 3	Y^*Q+Q^*R	2.2%	1.9%	5.3%	2.8%
	Model 4	$Y^*Q+Y^*R+Q^*R$	0.8%	1.0%	2.8%	5.2%
	Model 6	Y^*Q^*R	0.0%	0.7%	1.7%	5.7%
Range overall years	Model 1	$Y+Q+R$	13.6%	12.7%	13.6%	9.5%
	Model 2	$Y+Q^*R$	14.2%	14.4%	18.2%	10.9%
	Model 3	Y^*Q+Q^*R	8.4%	7.0%	15.9%	8.9%
	Model 4	$Y^*Q+Y^*R+Q^*R$	3.5%	4.4%	11.1%	20.0%
	Model 6	Y^*Q^*R	0.0%	2.2%	5.0%	19.6%
Trend overall years	Model 1	$Y+Q+R$	−2.78%	−2.66%	−2.47%	−1.83%
	Model 2	$Y+Q^*R$	−2.97%	−2.87%	−3.15%	−2.45%
	Model 3	Y^*Q+Q^*R	−0.92%	−0.75%	−1.22%	−0.95%
	Model 4	$Y^*Q+Y^*R+Q^*R$	0.16%	0.02%	−0.11%	−0.27%
	Model 6	Y^*Q^*R	0.00%	−0.11%	0.62%	0.28%

been interest in fitting several factors included in CPUE analyses as random effects with the introduction of statistical packages over the past decade which allow the fitting of mixed effects (i.e. the inclusion of both fixed and random effects) (Venables and Dichmont, 2004; Helser et al., 2004; Hsu, 2011; Babcock, 2013; Cortes, 2013). While the need to fit a factor as a fixed or random effect will be dependent on the question being addressed, classically fixed effect model terms have been used when means (or their analogues) are of primary interest and random effect models when the focus is on the variance components. According to Littell et al. (2002), generally a fixed effect can be thought of as treatment levels that have been selected for inclusion in an analysis and which are the only levels of the treatment in which we have an interest. Within a CPUE analysis such treatments would typically include year and region effects. Furthermore, there is usually interest in comparing the means among the different levels of these treatments (such as the year effect). On the other hand, a treatment might be modelled as a random effect if the levels that are included in the analysis can be thought of as a sample drawn from a larger (conceptual) population of levels that could (in principle) have been selected. The interest is not in the specific differences in means from one level of the treatment to another—but in the extent to which the random factor accounts for variance in the dependent variable, because we want to control for this (Littell et al., 2002). An example would be a vessel effect in a CPUE analysis, especially if not all vessels in the fleet were included in the analysis.

The standard GLM includes the following two components (i) a linear component defined by $\eta = X\beta$ where β is a vector of unknown fixed-effects parameters with known design matrix X, and (ii) a monotonic differentiable link function g which describes how the expected value of the observations y is related to the linear predictor, $g(y) = \eta = X\beta$. Things are similar with a GLMM except an additional component consisting of a vector of unknown random-effects parameters, θ, and known design matrix, Z, is added to the linear component which can now as be expressed as $\eta = X\beta + Z\theta$. Additionally, the random-effects parameters are often assumed to be normally distributed with zero mean. Predicted values of the dependent variable are then obtained by using the rows from X and Z and the estimated values of the two sets of parameters, i.e. $X\hat{\beta} + Z\hat{\theta}$. The construction of the abundance index remains the same as before, except the linear predictor used previously (c.f. Eq. (3)) now consists of both the fixed-effects and random-effect components. A difference, however, is that where there are no observations for levels of an interaction effect (e.g. some missing Year–Region strata), instead of having to impute a value of the standardised CPUE for these strata if all explanatory variables are fitted as fixed effects, the posterior mean of the assumed normal distribution of the random effects (estimated within the model) can be used together with the parameter values for the fixed effects to determine these values.

While the issue of whether it is appropriate to fit explanatory variables as fixed or random effects is an interesting one, it will not be pursued further here except for the proposal by Cooke (1997) to treat interactions including the year effect as random effects. Cooke notes that significant Y^*Q or Y^*R interactions can arise due to a number of reasons such as environmental differences between years influencing the timing or pattern of fish movements and that when such interactions are included in a purely fixed-effects GLM the resulting estimate of an annual CPUE index is no longer unique. He states that this problem can be overcome by treating these interactions as random effects. While the reason for the non-uniqueness of the index is not made clear it can be assumed that he was referring to the fact that in such situations different indices can be constructed for each region (or quarter). However, this overlooks the fact that a single index over all regions and seasons can, and should, be constructed as outlined previously.

Of course, there may be other reasons to treat interactions including the year effect as random effects. For example, it does make sense to model Y^*R effects as random because there is often very unbalanced and/or poor data coverage at this level. When this occurs (i) the resulting inferences will be unnecessarily imprecise, and (ii) standard errors may be biased low. An extreme case of (i) is when there is no data at all in an interaction. As seen in the previous example, in most CPUE analyses there are usually a few levels of one or more interactions for which the data coverage is poor, unrepresentative or missing, and in such cases there is a need to impute values for these levels if the interaction is fitted as fixed effects. This can be seen as adding an *ad-hoc* element to the analysis. While treating such interactions as random effects imposes a prior assumption of stationarity on these factors, nevertheless, from a practical point it deals elegantly with unbalanced information and the inferences about the main items of interest (e.g. the year effects) will likely be better (i.e. have lower prediction error) if you model the items that are generally not of interest as random effects (M. Bravington, pers. comm.).

In the sense that Cooke (1997) saw significant Y^*R interactions as due to random variations in environmental conditions then such interactions could be seen as a "sample drawn from a larger (conceptual) population of levels that could (in principle) have been selected" and therefore could be treated as random effects. On the other hand, and again as indicated by Cooke (1997), if interactions between year and other effects are not fully explained as random effects, for example because they show a significant trend, then such interactions should not be treated as random effects. Furthermore, if changes in movement and/or distributions pattern of the fish will be accounted for in a spatially disaggregated population dynamics model then again interactions involving year and region should not be treated as random effects.

11. Methods presently used by ICCAT

Given the issues raised in previous sections, it is useful to review some of the techniques presently being used for constructing abundance indices based on analysis of CPUE data. For this purpose, the 22 papers published in the latest Collective Volume of Scientific Papers for the International Commission for the Conservation of Atlantic Tunas (ICCAT, 2013) dealing with the standardisation of CPUE and construction of indices of abundance were examined. Whilst each paper provides a description, if somewhat briefly in some instances, of the method used to standardise the CPUE, for most papers little, if any, description is provided as to how the abundance indices were constructed using the parameter values from the fitted models. Where information was provided the following general methods were often used:

(i) The annual abundance index is the product of the estimates based on the year factor for the selected binomial and lognormal models. If included, interactions including the year effect fitted as fixed effects.
(ii) Same as (i) with interactions including the year effect fitted as random effects.
(iii) The annual abundance index is the product of the year effect least squares means (LS-means) from the binomial and lognormal components. If included, interactions including the year effect fitted as fixed effects. Note, the LS-means are within-group means appropriately adjusted for the other effects in the model, i.e. they estimate the marginal means for a balanced population (as opposed to the unbalanced design).
(iv) Same as (ii) with interactions including the year effect fitted as random effects.

As explained in the previously, the first method can only be used if the analysis is based on a simple single stage model (i.e. having no binomial stage) with no interactions including the year effect (c.f. Eq. (10)). Unfortunately this was not always the case. The other methods either fit interactions which include the year effect as random effects and/or use the LS-means to calculate the annual index. However, there are problems with the use of the LS-means approach which can be demonstrated by use of a simple example.

Consider a single-stage GLM (i.e. having no binomial stage) where the distribution of catch rates is related to a vector of Ny year and Nr region effects using a log link. The LS-means of the i-th year effect is given by:

$$LS\text{-means}(i) = \beta_0 + y_i + (r_1 + \cdots + r_{\mathrm{Nr}})/\mathrm{Nr}$$

Taking the exponent, and after some algebra, the related year index can be written:

$$B_{LS\text{-means}}(i) = \sqrt[\mathrm{Nr}]{\prod_{k=1}^{\mathrm{Nr}} \exp(\beta_o + y_i + r_k)}$$

On the other hand, from Eqs. (4) and (5) we have:

$$B(i) = \sum_{k=1}^{\mathrm{Nr}} A_k \exp(\beta_0 + y_i + r_k)$$

The desired annual abundance index relates to the total number of fish across all regions included in the standardising model and is expressed, correctly, by the last equation as the sum of the number of fish in each region (as given by the product of the density of fish in each region and the size of each region). On the other hand, the LS-means incorrectly expresses the annual abundance index as the geometric mean of the density of fish (as given by the standardised CPUE) in each region. By doing so it also does not take into account the size of each region. Even in those situations where all regions are the same size (and so cancel out when calculating a relative index) the problem persists as the ratio of the geometric mean of the densities across each region in different years is different from the ratio of the arithmetic mean of these densities. The problems with the use of the LS-means method are made worse when both stages of the delta-GLM approach are combined due to the incorporation of the non-linear logit link function. A further problem is that when fitting a model which incorporates both fixed and random effects the LS-means are based on the fixed-effects only.

In order to investigate the above approaches further a set of models were fitted to the simulated data used previously for the years 1998–2003 where:

$$f(Y, Q, R; \beta) = \sum_{i=1}^{\mathrm{Ny}} \beta_i Y_i + \sum_{j=1}^{\mathrm{Nq}} \beta_j Q_j + \sum_{k=1}^{\mathrm{Nr}} \beta_k R_k + \sum_{i=1}^{\mathrm{Ny}} \sum_{j=1}^{\mathrm{Nq}} \beta_{ij} (Y * Q)_{ij} + \sum_{i=1}^{\mathrm{Ny}} \sum_{k=1}^{\mathrm{Nr}} \beta_{ik} (Y * R)_{ik}$$

The SAS GLMMIX and SAS MIXED procedures were used for the stage 1 (Binomial) and stage 2 (log-gamma) analyses respectively. All explanatory variables were fitted as fixed effects except for the last interaction term including the year effect which was fitted as either a fixed-effect or a random-effect (known as the FE and RE models respectively). For each model the abundance index was then calculated based on:

(i) Eq. (12) using the predicted values based on all fixed and random effects included in the model, i.e. $X\hat{\beta} + Z\hat{\theta}$ (subsequently called the full index);
(ii) The transformed LS-means of the year factors (subsequently called the LS-means index), or;
(iii) Eq. (12) using the predicted values based the fixed effects only, i.e. $X\hat{\beta}$ (only calculated for the RE-models and subsequently called the FE-means index).

The analysis was repeated using the four data sets used previously having observations missing for the different numbers of the Year–Qtr–Region strata. Despite these missing strata, there were observations for all Year–Qtr and Year–Region strata in each data set. Finally, taking the full index based on the FE model fitted to the data with no missing strata as the reference index, a measure of the difference between each index and this reference index was determined using Eq. (16). The results are shown in Fig. 10.

There are small differences between the full indices for the FE model and the RE model based on the data with no missing strata. These differences will be due to the different parameterisation (i.e. treatment of the Y^*R interaction) of the two models. Further, the LS-means and FE-means indices for the RE model are also seen to be similar, due to the fact that both sets of indices are based on the same set of parameter estimates for the fixed-effects only. However, both indices differ substantially from the corresponding full index for the FE model. Finally, the LS-means index for the FE model displays the greatest difference because it incorporates parameter estimates for random Y^*R interaction effects. This pattern of results is generally repeated for the indices based on data with missing strata, with the full indices for both model types generally being similar, though the difference between these two indices increases as the number of missing strata increases. The mean over all years of the absolute difference between the annual values of these indices is 1.16%, 1.23%, 1.55% and 3.58% for the four data sets respectively. While there is some merit on further discussing whether the Y^*R interaction should be fitted as either fixed or random effects, these results indicate that the use of either the LS-means or the FE-means for determination of the desired abundance index is inappropriate and should not be used.

Another feature of the use of the LS-means index is the lack of inclusion of the size of the regional effects included in the model. This can be a particular issue when there are large differences in the relative sizes of these regions. To explore this further, the indices based on the previous analyses were re-calculated using the following different sets of sizes for each region:

Observed sizes:	$A_k = (13.1, 7.9, 28.9, 3.6, 19.0)$
Equal sizes:	$A_k = (10, 10, 10, 10, 10)$
Large differences in sizes:	$A_k = (100, 50, 25, 12, 6)$

where the first set of sizes represent the actual sizes of the fished regions and were used in all previous analyses. The results for the FE model are shown in Fig. 11 though similar results were obtained for the RE model. As expected, the full index for each set of regional sizes is different due to the fact that while the density of fish in each region is the same for each index, the number of fish and the relative proportion of the total population in each region will be different. As a consequence, the relative annual change in the total population will be different for each set of regional sizes. On the other hand, and as explained previously, the LS-means do not take the relative size of each region into account as they are based only on the density of fish in each region (which remains the same for each set of regional sizes). These results indicate that distribution of regional sizes used in the standardising models can be highly influential on the resulting abundance indices and again highlights the need to base the calculation of these indices on the correct equation. Given this result, it is somewhat disconcerting that few if any of the ICCAT papers reviewed above appeared to take into account the differential sizes of the regional effects used in the standardisation when constructing the various abundance indices.

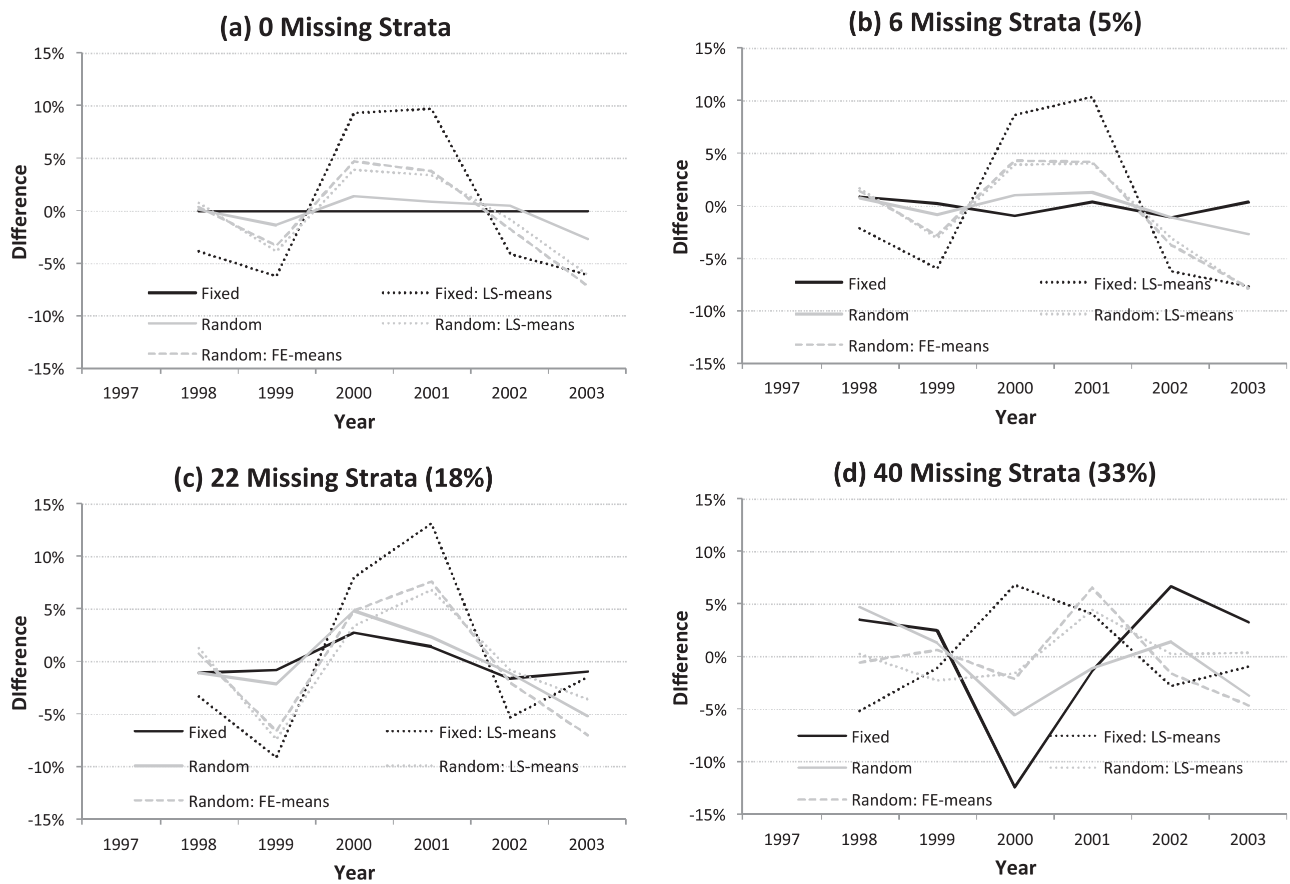

Fig. 10. Differences between a reference index and other abundance indices using different methods of calculating the index described in the text after fitting to a standardising model where (a) all effects are fitted as fixed effects, or where (b) the *Y*R* interaction term is fitted as a random effect. The reference index is based on Eq. (12) and fitted to the model with all fixed effects. Results are shown for four sets of data with observations removed from the shown number of *Y–Q–R* strata.

12. Discussion

The construction of abundance indices based on the results of statistical models used to standardise the CPUE data has become a routine but important component of many stock assessments. Given that abundance indices act as a 'backbone' for many fishery assessments, and that fishery-independent abundance indices are not available for many assessments, it is important that due care is taken to both ensure that the indices based on CPUE are correctly constructed and any potential bias is minimised so that the resulting indices reflect the underlying stock condition as accurately as possible. However, despite the importance of this task, the description of the methods and models used to standardise CPUE is often relegated to the margins of the stock assessment report with little or no description given to how the actual abundance indices are constructed. Given the importance of the resulting index to the stock assessment this should be considered an unsatisfactory state of affairs. While it can be hoped that the correct approach is being used this cannot always be guaranteed and in some instances it can be shown that an incorrect method has been used. This can be a particular issue when interactions are used in the standardising model, especially those which contain interactions with the temporal effects (e.g. year, quarter) over which the time-series of the index is required.

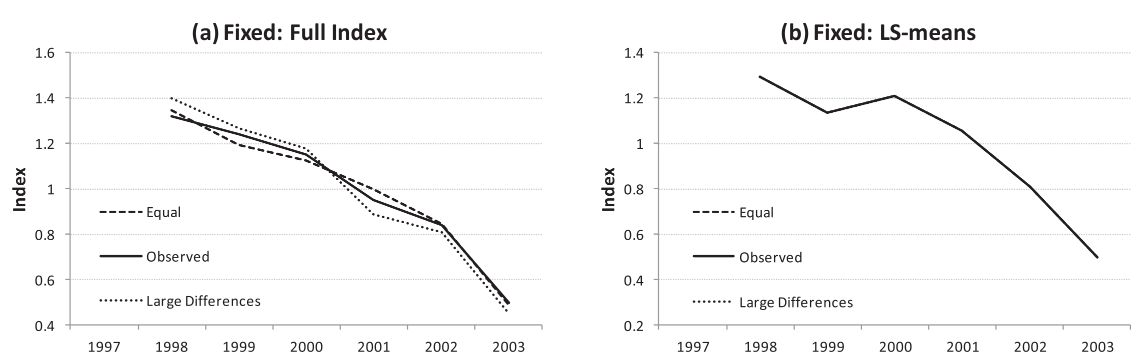

Fig. 11. Annual abundance indices calculated using (a) Eq. (12) and (b) the LS-means for the fixed-effect model and three different sets of sizes used for the regional effects in the standardising model.

While the examples presented in this study relate to pelagic longline fisheries, the methodological approaches outlined are applicable to all fisheries. Furthermore, while the construction of abundance indices should be straight-forward in practice there are several issues of which the analyst should be aware. First, the standardised CPUE should be interpreted as a measure of the density of fish in a region and not as an abundance measure (Campbell, 2004). Second, the formulation of the annual abundance index given by Eq. (6) highlights the fact that the index is the product of the density of fish within several spatial regions and the size of those regions. The common practice of reducing the index to a function of the year effect alone runs the risk of ignoring information on the spatial dynamics of the fishery which may be relevant to the underlying dynamics of the stock and the correct interpretation of the catch and effort data. Additionally, Eq. (6) will give indices of total stock abundance only if the spatial extent of the fishery coincides with, or is greater than, the spatial extent of the stock. Otherwise, the index of abundance pertains only to that portion of the stock which is found on the fishing grounds. Uncertainty will remain as to the size of the stock beyond the region fished. Third, each observation needs to be appropriately weighted to overcome potential biases due to the distributionally unbalanced nature of the data fitted to the GLM. While only small differences were found in the abundance indices between the weighted and unweighted models in this study, possibly due to the regional structure chosen appropriately stratifying the spatial distribution of catch rates, further work is required to understand the circumstances in which these differences may become more pronounced. Fourth, there is a need to take careful note of unusual features in the distribution of the observations in the data set fitted to the GLM as these may have a marked impact on the resulting index.

Fifth, with the use of the delta-GLM method it is no longer possible to just use the exponent of the year effect parameters alone as the relative index of abundance. This result holds whether only a main-effects models is used or the model includes interactions. Instead, it is necessary to calculate the standardised CPUE in each spatial–temporal strata (i.e. fill in Walter's large table) and then aggregate these over each region (and quarter) to determine the appropriate quarterly (or annual) index. Sixth, the annual index of abundance can be based on the exponent of the year effect parameters alone when standardising the CPUE using a single step GLM, but only if the fitted model includes no interactions of the temporal effects (e.g. year and quarter or month effects) and the size of each region used in the standardising model remains constant over all years. It is best practice to first calculate the standardised CPUE in each spatial–temporal strata when temporal interactions are included. If the annual abundance index is based on the parameters of the year effects only, then not only will the index be incorrect but it may be severely biased. Given these issues it is best practice to first calculate the standardised CPUE in each stratum associated with the fitted model and then aggregate across the spatial and temporal strata to obtain the desired abundance index.

Finally, analysts should be aware that the abundance index will also be influenced by the choice of standardising level for each fitted effect. This is due to the fact that in the delta-GLM approach the standardised CPUE (c.f. Eq. (4)) contains a term due to the use of the inverse logit link function in the binomial component which does not allow the cancellation of these standardising variables when the relative index is calculated. While the choice of standardising level for each fitted effect is arbitrary, in this study it has been suggested that the category levels be chosen which corresponds to the most common gear and environmental conditions observed in the fishery. It can also be argued that the same base values should be used for both components of the delta-GLM. While further work on this topic is required, this approach differs substantially from the use of the LS-means where the year effects are calculated across all levels of the fitted effects with each level given equal weight.

The examples also help to illustrate issues related to the potential errors and biases in the abundance indices that may result when the fitted model is mis-specified in comparison to the 'true' underlying model. While there will always be potential for errors in the calculated abundance index when the true model is not correctly specified (if indeed it is ever possible to do so), perhaps of more particularly concern is whether these errors add a temporal bias to the resulting index. For the data and models investigated here, both larger and temporal biases were found for the simpler models while such biases were found to be small or negligible for the more complex models (i.e. those including more interaction terms). While these particular results pertain only to the data set investigated they are not unexpected, as more complex models which include spatial–temporal interactions have a greater ability to more accurately track changes in the spatial–temporal distribution of CPUE over time. This result therefore suggests that analysts should not be reticent about attempting to fit a more complex model to the data (even if this results in the need to impute the standardised CPUE within a few strata) to avoid possibly biases in the calculated indices. If such a model is not necessary (i.e. a simpler model better represents the underlying true state of nature) then hopefully the simpler and more representative model will be identified through the systematic process of discarding terms in the model not found to be significant. Even better is to fit several models with different levels of complexity as a sensitivity analysis. Alternatively, several different models can be fitted and the resulting abundance indices compared. Occam's razor can then applied to select the most parsimonious model, i.e. proceed to the simpler model until simplicity can be traded for greater explanatory power.

As the example used in this study indicates, however, there is an associated need when fitting highly parameterised models that the data underlying the estimation of each parameter is representative of true underlying abundance associated with that parameter. The potential for significant bias in the resulting abundance index can increase where data are unrepresentative. A systematic check of the number of data points and the corresponding nominal CPUE in each stratum helps to check for anomalous observations. It may be prudent to remove such data then re-fit the model as a sensitivity test with the added requirement that it may be necessary to impute the standardised CPUE in those strata where the data were removed. Again, analysts should not be reticent about exploring the potential sources of bias in the abundance indices.

It may be possible to avoid some of the issues noted with incorporating complex interaction terms when fitting a single GLM with multiple regional effects if instead separate models are fitted to each regional set of data. Indeed this would be appropriate when a regional explicit stock assessment model is used and there is a need for a separate abundance index for each region. Alternatively, there may be evidence that the selectivity of the species differs by region or there may be quite different parameterisation for some of the effects (e.g. SST) that are better considered independently rather than as interactions with region. However, several issues need to be considered when adopting such an approach. First, it is necessary to ensure that each regional index is standardised against the same level for each fitted effect when a single index is required across all regions. Second, when using a regional explicit stock assessment model, it is important that the regions chosen attempt to stratify important features of the fishery that account for any spatial heterogeneity in the distribution of catch rates. This may be difficult when the regions chosen are relatively large such as when conducting ocean side stock assessments. Third, and as the example above illustrates, there may be significant spatial–temporal heterogeneity on a relatively small spatial scale which is useful to explore with interactions (even within sub-regional models).

This study has explored how abundance indices are constructed using the parameters estimated from fitting GLMs to the associated CPUE data and the potential for bias in these indices when the model is either mis-specified or the data is distributionally or structurally unbalanced (which unfortunately is the norm for most fishery-dependent data sets). A number of techniques (e.g. weighting, exploration of anomalous data values) have been suggested to help overcome potential biases generally, but without further work of this type it is perhaps too early to know how broadly applicable some of the other techniques (e.g. imputation methods, selection of the most parsimonious model) are across different data sets. While it is hoped that the work presented here has made a contribution to this task, it re-iterates the problem identified by Maunder and Punt (2004) that little effort has been directed toward identifying the most appropriate methods for specific instances and that additional work along these lines is clearly a high priority for the future.

Finally, this study has ignored commenting on a number of issues which are still pertinent in the standardisation of CPUE. Whilst mention was made at the start on the need for further work on identifying the most appropriate methods and associated distributions to be used, a number of other issues also need to be addressed.

First is how to propagate the uncertainty from the delta-GLM parameter estimates to error bounds on the final time series of abundance. While this paper has suggested an approach for the estimation of confidence levels for the delta-GLM model, there remain issues about the likely under-estimation of uncertainty in these models. This is a particular issue when using fine-scale or shot-by-shot data as the observations are not independent (Nishida and Chen, 2004). This will result in the degrees of freedom in the model being over-estimated resulting in the standard errors on the parameter estimates being smaller than should be otherwise. However, the standard errors do provide an indication of the relative reliability of the CPUE indices over the time-series and for this reason it may be appropriate to incorporate the relative precision of the CPUE indices within the assessment model. The under-estimation of the associated errors also has consequences for the method used when selecting which explanatory effects to include in the final model, as the standard model selection criteria based on the use of these standard errors in *F*-tests are again likely to over-estimate the influence of terms, leading to over-parameterised models. While some ad hoc approaches for dealing with this issue have been adopted over the years (such as limiting the inclusion of effects in the model to those that reduce the R^2 or deviance by more than a given amount, e.g. 1%, Cortes, 2013; Cass-Calay and Walter, 2013) further work is required on this topic (such as the inclusion of random-effects) together with other issues such as the use and effectiveness of AIC/BIC selection criteria for identifying the best model.

Second is the inclusion of environmental effects within a model used to standardise CPUE. While the worked example used in this study included such effects, this was done simply for completeness and to illustrate how such effects can be included if warranted. However, analysts need to be careful of possible confounding between environmental effects included in a model (e.g. SST and SOI) and other model effects, especially the year effect. For this purpose it would be useful to examine the covariance matrix Furthermore, unlike the use of different fishing gears, the environment can influence abundance in a number of different ways, for which there is usually no clear understanding and/or competing hypotheses. For example, a change in the thermocline structure of the ocean may result in a change in the vertical distribution of fish in the water column and influence their availability to the fishing gear. This is consistent with the comment made earlier that inclusion of environmental effects can be seen as standardising availability (within a given strata) for changes in environmental conditions. On the other hand, changes in oceanographic features under the influence of forcing mechanisms such as El Nino-Southern Oscillation may result in changes in the horizontal distribution of fish (i.e. movement to other regions) such as observed with skipjack in the equatorial Pacific (Lehodey et al., 1997). However, such changes can also be accounted for by including appropriately scaled spatial–temporal interactions in the standardising model and including both statistical and process related terms in the standardising model may lead to some degree of parameter confounding. Changes in environmental conditions can also impact on resource abundance more directly, such as influencing life-history parameters such as larval survival. In such cases it is inappropriate to standardise for such influences as it is the abundance that is of interest not the abundance which might have resulted under a standard set of environmental conditions.

While it is generally implicit when environmental effects are included in a standardisation model that the influence being modelled is indirect (i.e. not influencing the abundance directly), it is usually not explicitly mentioned whether such effects standardise catchability or availability (though it is usually assumed to be former). The assumption used in this study that environmental effects influence availability has been guided by the distinction between abundance and apparent abundance, with the latter being the abundance as affected by availability or the number of fish accessible to the fishery (Marr, 1951; Campbell, 2004). In seeking to explain the variance in observed CPUE it is also necessary to be mindful that different environmental metrics (e.g. SST and the SOI) may be correlated leading to the confounding of parameters and a loss of precision (Hocking, 1976). This may have been the situation in the worked example in this study. While this may not have an impact on the resultant abundance index it may lead to a misunderstanding of the manner in which these effects influence CPUE. In order to avoid such a situation, and given the different means by which environmental conditions can influence abundance, it is perhaps incumbent on analysts to explicitly state what they assume the environmental effect to be and provide, where possible, supporting evidence. Again, further work in this important issue is warranted.

Acknowledgements

Thanks are extended to Dale Kolody, Miriana Sporcic, Adam Langley and Simon Hoyle for useful comments on an earlier draft of this manuscript and to Mark Bravington and Bill Venables for providing insightful comments related to several statistical issues. Andre Punt and two anonymous reviewers are also thanked for their constructive comments which helped improve the manuscript. This work was partially supported under a contract with the Australian Fisheries Management Authority.

References

Babcock, E.A., 2013. Updated index of abundance for shortfin mako sharks from the U.S. marine recreational fisheries statistical survey. Collective Volume of Scientific Papers ICCAT, 69, pp. 1639–1646.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, New Jersey.

Bishop, J., 2006. Standardizing fishery-dependent catch and effort data in complex fisheries with technology change. Rev. Fish Biol. Fish. 16, 21–38.

Brodziak, J., Walsh, W.A., 2013. Model selection and multimodel inference for standardizing catch rates of bycatch species: a case study of oceanic whitetip shark in the Hawaii-based longline fishery. Can. J. Fish. Aquat. Sci. 70, 1723–1740.

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.A., 1998. Analysis of catch and effort data in a fishery with uncertain stock and effort dynamics with application to the longline fishery for southern bluefin tuna. In: Funk, F., Quinn, I.I.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, pp. 75–97.

Campbell, R.A., 2004. CPUE standardization and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70, 209–227.

Campbell, R.A., 2012. Aggregate and size-based standardised CPUE indices for longline target species caught in the south-west Pacific. In: Information Paper SA-IP-13 Presented to the Eighth meeting of the Scientific Committee for the Western Central Pacific Fisheries Commission, 7–15 August 2012, Busan, Korea http://www.wcpfc.int/meetings/8th-regular-session-scientific-committee

Campbell, R.A., 2013. Summary of catch and effort information pertaining to the Australian longline fishing operations in the Eastern Tuna and Billfish Fishery. In: Information Paper to the Seventh Meeting of the Tropical Tuna Resource Assessment Group, 24–25 July, Mooloolaba, Australia.

Campbell, R.A., Hobday, A.J., 2003. Swordfish–environment–seamount–fishery interactions off eastern Australia. In: Working Paper BBRG-3 Presented to the 16th Meeting of the Standing Committee on Tunas and Billfish, July 9–17, Mooloolaba, Australia http://www.spc.int/DigitalLibrary/Doc/FAME/Meetings/SCTB/16/BBRG_3.pdf

Cao, J., Chen, X., Chen, Y., Liu, B., Ma, J., Li, S., 2011. Generalised linear Bayesian models for standardising CPUE: an application to a squid-jigging fishery in the northwest Pacific Ocean. Sci. Mar. 75, 679–689.

Carruthers, T.R., McAllister, M.K., Ahrens, R.N.M., 2010. Simulating spatial dynamics to evaluate methods of deriving abundance indices for tropical tunas. Can. J. Fish. Aquat. Sci. 67, 1409–1427.

Carruthers, T.R., Ahrens, R.N.M., McAllister, M.K., Walters, C.J., 2011. Integrating imputation and standardisation of catch rate data in the calculation of relative abundance indices. Fish. Res. 109, 157–167.

Cass-Calay, S.L., Walter, J.F., 2013. Standardized catch rates of large bluefin tuna (*Thunnus thynnus*) from the U.S. pelagic longline fishery in the Gulf of Mexico during 1987-2-10. In: Collective Volume of Scientific Papers of ICCAT, vol. 69, pp. 992–1004.

Clark, C.W., 1985. Bioeconomic Modelling of Fisheries Management. John Wiley and Sons, New York.

Cooke, J.G., 1997. A procedure for using catch-effort indices in bluefin tuna assessments. In: Collective Volume of Scientific Papers of ICCAT, vol. 46, pp. 228–232.

Cortes, E., 2013. Standardised catch rates of mako sharks from the U.S. pelagic longline logbook and observer programs using a generalized linear mixed model. In: Collective Volume of Scientific Papers of ICCAT, vol. 69, pp. 1578–1590.

Goodman, L.A., 1960. On the exact variance of products. J. Am. Stat. Assoc. 55, 708–713.

Goodyear, C.P., Ortiz, M., 2007. GLM CPUE standardisations: interpretation risks for model fits from alternative software packages. In: Collective Volume of Scientific Papers of ICCAT, vol. 60, pp. 1622–1629.

Hardin, J.W., Hilbe, J.M., 2012. Generalized Linear Models and Extensions, third edition. Strata Press, Texas.

Harley, S.J., Myers, R., Dunn, A., 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. 58, 1760–1772.

Helser, T.E., Punt, A.E., Methot, R.D., 2004. A generalised linear mixed model analysis of a multi-vessel fishery resource survey. Fish. Res. 70, 251–264.

Hocking, R.R., 1976. The analysis and selection of variables in linear regression. Biometrics 32, 1–49.

Hsu, C.C., 2011. Verification of catch-effort data and standardisation of abundance index of bigeye tuna by Taiwanese longline fishery in the Atlantic Ocean. In: Collective Volume of Scientific Papers of ICCAT, vol. 66, pp. 368–386.

ICCAT, 2013. Collective Volume of Scientific Papers for the International Commission for the Conservation of Atlantic Tunas, vol. 69 , pp. 1–5.

Lehodey, P., Bertignac, M., Hampton, J., Lewis, A., Picaut, J., 1997. El Nino southern oscillation and tuna in the western Pacific. Nature 389, 715–718.

Littell, R.C., Stroup, W.W., Freud, R.J., 2002. SAS for Linear Models, fourth edition. SAS Institute Inc., Cary, NC.

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

Lynch, P.D., Shertzer, K.W., Latour, R.J., 2012. Performance of methods to estimate indices of abundance for highly migratory species. Fish. Res. 125–126, 27–39.

Marr, J.C., 1951. On the use of the terms abundance, availability and apparent abundance in fishery biology. Copeia 2, 163–169.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141–159.

Newman, M.C., 1993. Regression analysis of log-transformed data: statistical bias and its correction. Environ. Toxicol. Chem. 12, 1129–1133.

Nishida, T., Chen, D.-G., 2004. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacores*) longline CPUE data. Fish. Res. 70., 265–274.

Pereira, C., Leandro, R.A., Petrere Jr., M., Nishida, T., 2012. Comparison between univariate and bivariate geostatistical models for estimating catch per unit effort (cpue): a simulation study. Fish. Res. 121–122, 115–125.

Punsley, R.G., 1987. Estimation of the relative abundance of yellowfin tuna *Thunnus albacares*, in the Eastern Pacific Ocean during 1970–1985. Inter-Am. Trop. Tuna Comm. Bull. 19, 98–131.

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

SAS Institute Inc, 2008. SAS/STAT 9.2 User's Guide: The GENMOD Procedure, Cary, North Carolina, USA.

Shono, H., 2008. Confidence interval estimation of CPUE year trend in delta-type two step model. Fish. Sci. 74, 712–717.

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

Tascheri, R., Saavedra-Nievas, C., Roa-Ureta, R., 2010. Statistical models to standardize catch rates in the multi-species trawl fishery for Patagonian grenadier (*Macruronus magellanicus*) off southern Chile. Fish. Res. 105, 200–214.

Venables, W.N., Dichmont, C.M., 2004. GLMs, GAMs and GLMMs: an overview of theory and applications in fisheries research. Fish. Res. 70, 319–337.

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

Winker, H., Kerwath, S.E., Attwood, C.G., 2013. Comparison of two approaches to standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line fishery. Fish. Res. 139, 118–131.

Ye, Y., Dennis, D., 2009. How reliable are the abundance indices derived from commercial catch-effort standardization? Can. J. Fish. Aquat. Sci. 66, 1169–1178.

Zang, Z., Holmes, J., 2010. Use of generalized linear Bayesian model to mitigate the impact of spatial contraction in fishing pattern on the estimation of relative abundance. Fish. Res. 106, 413–419.