CCAMLR Science, Vol. 15 (2008): 115–138

ESTIMATION OF EFFECTIVE SAMPLE SIZE FOR CATCH-AT-AGE AND CATCH-AT-LENGTH DATA USING SIMULATED DATA FROM THE DIRICHLET-MULTINOMIAL DISTRIBUTION

S.G. Candy
Department of the Environment,
Water, Heritage and the Arts
Australian Antarctic Division
203 Channel Highway, Kingston 7050
Tasmania, Australia
Email – steve.candy@aad.gov.au

Abstract

The incorporation of 'effective sample size' (ESS) in integrated assessments is an approximate but simple way of modelling the distribution of catch-at-age or catch-at-length frequencies using a multinomial likelihood when there is extra-multinomial heterogeneity. Accurate estimation of ESS for catch-frequency data for each fishery and fishing year is important for such assessments, and this issue is studied using simulation. Between-haul heterogeneity within fishing year was simulated using samples from the Dirichlet-multinomial (D-M) distribution, with marginal class probabilities generated using a simple age-structured model incorporating fishing selectivity. Four methods of estimation of effective sample size were compared using this simulation model and its variants. One of the methods is based on the lack-of-fit of predictions of class probabilities using aggregate year-level frequencies. The other three estimators use the haul-level frequencies, including a method based on an approximate profile maximum likelihood estimate (PMLE) of the D-M dispersion parameter. The remaining two estimators based on haul-level frequencies are derived from models for the empirical coefficient of variation (CV) in the proportions, with one being based on an existing CV model used for CCAMLR fisheries while the other is a new method. The methods that use haul-level frequencies gave accurate estimators of an ESS that is appropriate for haul-level heterogeneity with increasing accuracy in the following order: (i) the estimator based on the existing CV model; (ii) that based on the new CV model; and (iii) that based on the PMLE. The year-level method gave very inaccurate estimates of this ESS with relative mean square error two orders of magnitude worse than the best haul-level method.

To account for process error in the calculation of the ESS, the lack of fit of the age-structured model in predicting class/bin by year frequencies is used to obtain a single, across-years, over-dispersion parameter. The ESS is then rescaled by dividing by the over-dispersion parameter, and the model refitted, giving a two-step iterative procedure. The ESS will be over-corrected if there is a systematic component to the lack of fit. A simple generic model of systematic lack-of-fit (SLOF) is presented, and its performance, in terms of providing unbiased estimates of ESS when SLOF is either present or absent, is studied using perturbations of the age-structured model. These perturbations consisted of either systematic or random variation across years in one of the selectivity function parameters and similarly for the mortality rate parameter when combined with systematic or random variation in recruitment. The SLOF model substantially reduced the bias when SLOF was present and is useful when its source is not clear or cannot be rectified by changing the underlying age-structured assessment model.

Résumé

L'intégration de la "taille effective d'un échantillon" (ESS pour effective sample size) dans les évaluations intégrées est un moyen approximatif mais simple de modéliser la distribution des fréquences d'âges ou de longueurs dans la capture selon une vraisemblance multinomiale en présence d'une hétérogénéité extra-multinomiale. L'estimation exacte de l'ESS relative aux données de fréquence dans la capture pour chaque pêcherie et année de pêche étant importante pour ces évaluations, cette question est étudiée par simulation. L'hétérogénéité entre traits en une année de pêche est simulée à l'aide d'échantillons de la distribution multinomiale de Dirichlet (D-M), les probabilités marginales des classes d'âge étant générées au moyen d'un modèle simple structuré selon l'âge et tenant compte de la sélectivité de la pêche. Quatre méthodes d'estimation de la taille effective d'un échantillon sont comparées par ce modèle de simulation et ses variantes. L'une d'elles est basée sur le défaut d'ajustement des prédictions des probabilités liées aux classes d'âge en utilisant des fréquences agrégées au niveau de l'année. Les trois autres estimateurs utilisent les fréquences au niveau du trait, y compris une méthode fondée sur un profil d'estimation de maximum de vraisemblance (PMLE pour profile maximum likelihood estimate) du paramètre de dispersion de D-M. Les deux autres estimateurs basés sur les fréquences au niveau du trait sont tirés de modèles du coefficient de variation (CV) empirique dans les proportions, l'un étant basé sur un modèle de existant de CV, utilisé pour les pêcheries de la CCAMLR, alors que l'autre est une nouvelle méthode. Les méthodes utilisant les fréquences au niveau du trait ont donné des estimateurs exacts d'une ESS adéquate pour l'hétérogénéité au niveau du trait avec une précision croissante dans l'ordre suivant : i) l'estimateur basé sur le modèle de CV existant ; ii) celui basé sur le nouveau modèle de CV ; et iii) celui basé sur le PMLE. La méthode au niveau de l'année a donné des estimations très inexactes de cette ESS ; en effet, l'erreur quadratique moyenne relative est pire que celle de la meilleure méthode au niveau du trait, de l'ordre d'un facteur 100.

Pour tenir compte de l'erreur de traitement dans le calcul de l'ESS, on utilise le défaut d'ajustement du modèle structuré selon l'âge dans la prévision des fréquences des classes/lots par année pour obtenir un paramètre de surdispersion unique, sur l'ensemble des années. L'échelle de l'ESS est ensuite modifiée en divisant l'ESS par le paramètre de surdispersion, puis le modèle est réajusté, donnant une procédure itérative à deux étapes. L'ESS sera surcorrigée si le défaut d'ajustement montre un élément systématique. Un modèle générique simple de défaut d'ajustement systématique (SLOF pour systematic lack-of-fit) est présenté ; au moyen des perturbations du modèle structuré selon l'âge, on étudie sa performance, lorsqu'il s'agit de fournir des estimations d'ESS non biaisées en présence ou en l'absence du SLOF. Ces perturbations consistent en une variation systématique ou aléatoire sur plusieurs années de l'un des paramètres de la fonction de sélectivité, et il en est de même pour le paramètre du taux de mortalité lorsqu'il est combiné à une variation systématique ou aléatoire du recrutement. Le modèle du SLOF réduit considérablement le biais en présence du SLOF et il est utile lorsque sa source n'est pas évidente ou ne peut être rectifiée en modifiant le modèle d'évaluation de base structuré selon l'âge.

Резюме

Включение «эффективного размера выборки» (ESS) в комплексные оценки является приблизительным, но простым способом моделирования частотного распределения возрастов и длин в уловах с использованием мультиномиального правдоподобия при наличии дополнительной мультиномиальной гетерогенности. При проведении таких оценок важно точно определить ESS в случае данных о частотном распределении уловов для каждого промысла и промыслового года; этот вопрос изучается посредством моделирования. Гетерогенность уловов в течение промыслового года моделировалась с использованием выборок из мультиномиального распределения Дирихле (D-M), где пределы вероятности классов, получены с помощью простой возрастной модели, включающей промысловую селективность. На основе этой имитационной модели и ее вариантов было проведено сравнение четырех методов определения эффективного размера выборки. Один из этих методов основывается на неадекватности расчетов вероятности классов с использованием агрегированных частот на годовом уровне. Три других метода используют частоты на уровне улова, включая метод, основанный на приблизительной оценке профиля максимального правдоподобия (PMLE) параметра дисперсии D-M. Остальные два метода оценки на основе частот на уровне улова получены при помощи моделей эмпирического коэффициента вариации (CV) в соотношениях, где один метод берет за основу существующую модель CV, используемую для промыслов АНТКОМа, а другой метод является новым. Методы, использующие частоты на уровне улова, дают точную оценку ESS, которая соответствует гетерогенности на уровне улова, причем точность возрастает в следующем порядке: (i) оценка на основе существующей модели CV; (ii) оценка на основе новой модели CV; и (iii) оценка на основе PMLE. Метод на уровне года дал очень неточные оценки этого ESS, где относительная средняя квадратичная ошибка была на два порядка хуже, чем у наилучшего метода на уровне улова.

Для того чтобы при расчете ESS учесть ошибку обработки, неадекватность возрастной модели при прогнозировании частот классов/интервалов по годам используется для получения одного параметра избыточной дисперсии за все годы. Затем ESS пересчитывается путем деления на параметр избыточной дисперсии, и вновь подбирается модель, что представляет собой двухступенчатую итеративную процедуру. При наличии систематической составляющей несоответствия корректировка ESS будет чрезмерной. Представлена простая типовая модель систематического несоответствия (SLOF) и, с использованием отклонений возрастной модели, рассматриваются ее результаты в плане получения несмещенных оценок ESS при наличии или в отсутствие SLOF. Эти отклонения включали систематические или случайные изменения одного из параметров функции селективности по годам и, аналогичным образом, параметра коэффициента смертности в сочетании с систематическими или случайными изменениями пополнения. Модель SLOF значительно уменьшила смещение при наличии SLOF; она может использоваться, когда источник SLOF неизвестен или не может быть устранен путем изменения исходной возрастной модели оценки.

Resumen

La incorporación del "tamaño efectivo de la muestra" (ESS) en las evaluaciones integradas es una forma aproximada, pero sencilla, de modelar la distribución de la frecuencia de edades o tallas de la captura mediante una función de probabilidad multinomial cuando la heterogeneidad multinomial es mayor. La estimación precisa del ESS a partir de los datos de frecuencia de edad o talla de la captura para cada pesquería y año de pesca es importante para estas evaluaciones, y este problema se estudia mediante simulaciones. La heterogeneidad entre los lances efectuados en un año de pesca fue simulada utilizando muestras de la distribución multinomial de Dirichlet (D-M), donde la probabilidad marginal de las clases de edad fue calculada mediante un modelo simple estructurado por edades que incorpora la selectividad por pesca. Se compararon cuatro métodos para estimar el tamaño efectivo de la muestra mediante este método de simulación y sus variantes. Uno de los métodos se basa en la falla del ajuste de las predicciones de la probabilidad de las clases de edad y utiliza frecuencias agregadas de las clases anuales. Los otros tres estimadores utilizan frecuencias a nivel de lance, incluido un método basado en una estimación aproximada del perfil de máxima verosimilitud (PMLE) del parámetro de dispersión D-M. Los dos estimadores restantes basados en la frecuencia a nivel de lance se derivan de los modelos para simular el coeficiente de variación empírico (CV) en las proporciones, estando uno basado en el modelo existente del CV utilizado para las pesquerías de la CCRVMA, mientras que el otro es un nuevo método. Los métodos que utilizan las frecuencias a nivel de lance produjeron estimaciones exactas de un ESS apropiado para la heterogeneidad entre lances con precisión creciente en el siguiente orden: (i) el estimador basado en el modelo existente del CV; (ii) el estimador basado en el nuevo modelo del CV; y (iii) el estimador basado en el PMLE. El método que utiliza frecuencias agregadas de clases anuales dio estimaciones muy imprecisas de este ESS con un error cuadrático medio relativo peor -en dos órdenes de magnitud- que el mejor método basado en las frecuencias a nivel de lances.

Para tomar en cuenta el error de tratamiento en el cálculo del ESS, se utiliza la falla del ajuste del modelo estructurado por edades en la predicción de la frecuencia anual de clases o intervalos para obtener un solo parámetro de sobredispersión para todos los años. El ESS se reajusta entonces dividiéndolo por el parámetro de sobredispersión, y se vuelve a reajustar lo que resulta en un método iterativo de dos fases. El ESS será corregido en exceso si la falla del ajuste incluye un componente sistemático. Se presenta un modelo general simple de falla sistemática del ajuste (SLOF), y se estudia sus resultados, en términos de proveer estimaciones de ESS sin sesgos en presencia y ausencia de SLOF a partir de las perturbaciones del modelo estructurado por edades. Estas perturbaciones consistieron de variaciones ya sea sistemáticas o aleatorias a través de los años en uno de los parámetros de la función de selectividad y de manera similar, para el parámetro de la tasa de mortalidad cuando se combina con una variación sistemática o aleatoria en el reclutamiento. El modelo SLOF redujo substancialmente el sesgo cuando existe una falla sistemática del ajuste y resulta útil cuando el origen de esta falla es incierta o no puede ser rectificada cambiando el modelo subyacente de evaluación basado en la edad.

Keywords: Dirichlet-multinomial distribution, integrated assessment, effective sample size, profile maximum likelihood, process error, model lack-of-fit, CCAMLR

Introduction

Integrated assessments that estimate model parameters using diverse datasets implicitly assign weights to each dataset via their respective definitions of the log-likelihood. To provide the most accurate parameter estimates, the likelihood defined for each dataset should reflect as faithfully as possible the systematic and stochastic variation in the data, including both population-level variation in biological attributes and variation resulting from the sampling processes used to obtain the data. Catch-at-age or catch-at-length data consist of sample frequencies for each age class or length bin usually obtained from a large number of fish sampled randomly from hauls in each year of fishing. The CASAL assessment software (Bull et al., 2005) currently used by CCAMLR allows a multinomial likelihood to be used for the frequencies in these classes/bins aggregated across hauls for each year of fishing. CASAL does not allow haul-level frequencies to be used as data and, correspondingly, cannot incorporate a mixed-effects multinomial log-likelihood to account for multi-level sampling using units such as vessels or hauls within vessel and year.

Further, CASAL requires the catch-at-age or catch-at-length data to be input as the total yearly frequency in the form of class proportions combined with the corresponding multinomial sample size. Similar integrated assessment software of MULTIFAN-CL (Fournier et al., 1998) and Stock Synthesis Program (Methot, 2000, 2005) also limit input of catch-at-length or catch-at-age frequencies to values aggregated across hauls. In the absence of haul-level data input, an approximate but simple way of accounting for random variations in class proportions due to the two levels of sampling (i.e. fish within hauls for 'level 1' sampling and hauls within year and fishery for 'level 2' sampling) is to replace the total sample size of fish measured in the year by the effective sample size (ESS) as the nominal multinomial sample size. To demonstrate this, consider a single fishery and let the number of fish sampled from age class a for year y, aggregated across all hauls ($j = 1,\ldots,h_y$) in year y, given by $n_{ya} = \sum_{j=1}^{h_y} n_{yja}$, be multinomially distributed conditional on total sample size n_y (where $n_y = \sum_a n_{ya}$). Note that in this notation a missing subscript for a frequency implies that the variable is the value totalled over the range of the missing subscript. The variance of the observed class proportions, $o_{ya} = n_{ya} / n_y$, conditional on n_y can be specified as

$$Var\left(O_{ya} = o_{ya} \middle| p_{ya}, n_y\right) = p_{ya}\left(1 - p_{ya}\right) / n_y \qquad (1)$$

where $p_{ya} = E\left(N_{ya} / N_y \middle| N_y\right)$ (where random variables are denoted using upper case while their corresponding sample realisations by lower case).

However, if the true variance is given by

$$Var_{\phi}\left(O_{ya} = o_{ya} \middle| p_{ya}, n_y\right) = \left(\phi_y / n_y\right) p_{ya}\left(1 - p_{ya}\right) \qquad (2)$$

where ϕ_y is an over-dispersion parameter then ESS, n'_y, is given by $n'_y = n_y / \phi_y$. Therefore if the ϕ_y were known or could be estimated then n'_y could be provided to CASAL as though this n'_y-size sample of 'fish' can be assumed to be independent samples from a nominal multinomial distribution. This assumed 'distribution' has a variance relationship that corresponds marginally (i.e. averaged across hauls) to a particular mixed-effect multinomial distribution model, called the Dirichlet-multinomial (D-M) distribution (Johnson and Kotz, 1969; Polacheck et al., 2006) given that $n_y = h_y m_y$ where h_y is the number of hauls and m_y is the number of fish measured per haul in year y (in what follows, the assumption, that m_y is constant across hauls within each year, is relaxed).

A key issue is whether either of n'_y or ϕ_y can be reliably estimated from catch-at-age or catch-at-length frequencies at either the haul level or the year level (i.e. frequencies aggregated across hauls). This paper examines four methods of estimation using simulated data generated from a D-M distribution for haul-level frequencies.

A possible criticism of the use of the D-M distribution in this context derives from the study of actual catch-at-length data from surveys of Atlantic cod (*Gadus morhua*) by Hrafnkelsson and Stefánsson (2004). They examined the empirical correlation in frequency between length bins and were able to model the observed moderate to strong positive correlations between length bins for small lags (e.g. adjacent bins have lag 1) (see also Miller and Skalski, 2006) using a Gaussian-multinomial (G-M) model whereas the D-M model is constrained to estimate only a single negative and, generally, small value for across-bin correlations. Note however, that considering the sample of fish measured for a haul as a cluster sample, the D-M model estimates the correlation between any two fish in the cluster in terms of their class/bin membership as a positive value determined by the intra-cluster correlation coefficient (Zhu, 2002). Unlike the D-M model, explicit formulae for the marginal variance and covariances for the G-M model are not available and therefore ESS due to between-haul heterogeneity cannot be defined for this model. Instead, Hrafnkelsson and Stefánsson (2004) used numerical Bayesian techniques involving Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distribution of the parameters defining the G-M distribution. Although they condition on the sample sizes, *m*, which can vary across stations (i.e. hauls), the posterior estimates of variance/covariance do not explicitly account for different *m*'s as is possible using the D-M distribution. In practical terms, the G-M model although more flexible in modelling correlations, comes at the cost of considerably more parameters and is much more numerically intensive to estimate via MCMC sampling than the method described here for estimation of effective sample size under a D-M model. In order to incorporate catch-at-length data with an appropriate likelihood into integrated assessments, Hrafnkelsson and Stefánsson (2004) suggested for the G-M model that the MCMC mean vector of the posterior sample of class/bin probabilities be used as 'data' in an approximate log-likelihood. Maximising this log-likelihood is equivalent to a generalised least-squares minimisation using the empirical variance-covariance matrix of the posterior sample as the prior precision matrix for the 'data'. This approach is currently not possible in CASAL.

This study focuses on numerically efficient approaches for determining an ESS to apply within existing, commonly available, integrated assessment software. Given the catch-at-age or catch-at-length data is generated from a D-M distribution at the haul level, the performance of a set of four alternative estimators of ESS is examined.

The estimation methods examined were: (i) maximum likelihood estimation of ϕ_y; (ii) the method of McAllister and Ianelli (1997) that used aggregate year-level frequencies; (iii) the method of Dunn and Hanchet (2007); and (iv) a new method that has recently been implemented for the CASAL-based assessment for Patagonian toothfish (*Dissostichus eleginoides*) fisheries around Heard and McDonald Islands (Candy and Constable, 2008) and based in part on the method described by Constable et al. (2006). These last two methods also use haul-level catch-at-age or catch-at-length frequencies.

These four estimation methods are examined using a simulation model that uses the D-M distribution to draw catch-at-age samples from a simple age-structured model with constant recruitment and instantaneous mortality rate parameter, *M*. A model of age-dependent fishing selectivity was applied. To generate catch-at-length frequencies, the catch-at-age data were converted to catch-at-length data using an assumed von Bertalanffy growth model for length given age with lognormal distribution of length about its expected value.

Further, to account for process error in integrated assessments, an appropriate distributional model that incorporates year-to-year process error in addition to haul-level heterogeneity while giving an explicit marginal variance relationship, which allows an ESS to be appropriately defined, does not appear to be available. Therefore, in order to estimate the ESS when significant between-year process error is assumed to occur, a more heuristic approach is adopted based on Finney's heterogeneity factor (Finney, 1971; McCullagh and Nelder, 1989, p. 128) whereby the lack-of-fit in class/bin by year frequencies is used to obtain a single, across-years, over-dispersion parameter. The ESS is then rescaled by dividing by the over-dispersion parameter and the assessment model refitted in a two-step iterative procedure that has some similarities to the procedure described by Hillary et al. (2006) and Dunn and Hanchet (2007). However, the rescaled ESS will be over-corrected if there is a systematic component to the lack-of-fit. Therefore, a simple generic model of systematic lack-of-fit (SLOF) is presented and its performance in terms of providing unbiased estimates of ESS when SLOF is either present or absent, is studied using perturbations of the age-structured model. These perturbations were either systematic or random variation across years in one of the selectivity function parameters and similarly for the mortality rate parameter when combined with systematic or random variation in recruitment. The SLOF model is useful when the source of SLOF is not clear or cannot be rectified by changing the underlying age-structured assessment model. For example, different datasets may 'pull' age-structured model parameters in different directions resulting in SLOF for one or more datasets depending on the weight they receive in estimation (Candy and Constable, 2008).

First, the D-M distribution is described along with the methods used to simulate this distribution in the context of catch-at-age or catch-at-length frequency data. The estimation methods are then given for dealing with haul-level heterogeneity in class proportions as simulated by the D-M distribution. Next, methods of dealing with the addition of systematic or process error are described, followed by details of how the simulations were structured. This is followed by results of the simulation study when process error and systematic error in class proportions were set to zero. This study compares the accuracy of the alternative estimators of ESS using the theoretically derived value based on the parameter values used in the simulations.

Finally, the performance of the proposed method of accounting for year-to-year process error and systematic error is then studied when these errors were incorporated in the simulation model. Since the D-M distribution does not extend to this case, formal comparisons with a theoretical value of ESS were no longer possible.

The Dirichlet-multinomial model

Consider initially a single year, a sample of h hauls, and A age classes from age 1 to age A. The Dirichlet distribution for a set of proportions $\underline{\pi} = (\pi_1,\ldots,\pi_A)$ that sum to 1 is given by Gelman et al. (2004)

$$\Pr\left(\underline{\pi}|\underline{\theta}\right) = \frac{\Gamma(\omega)}{\prod_{a=1}^{A}\Gamma(\omega p_a)}\prod_{a=1}^{A}\pi_a^{\omega p_a - 1} \qquad (3)$$

where the parameter set is given by $\underline{\theta} = (\omega,p_1,\ldots,p_A)$, $0 < p_j < 1$ for all j, $\sum_{a=1}^{A} p_a = 1$, and $\Gamma(.)$ is the gamma function. Random values of $\underline{\pi}$ can be drawn using random draws from a set of A independent gamma distributed variables, $X = (X_1,\ldots,X_A)$, each with scale parameter 1 and shape parameter $\alpha_a = \omega p_a$ $(a = 1,\ldots,A)$, where $\omega = \sum_{a=1}^{A}\alpha_a$, to give a random haul value of $\underline{\pi} = \underline{X} / \sum_{a=1}^{A} X_a$ where $\Pr(X_a = x) = x^{\alpha_a - 1}e^{-x} / \Gamma(\alpha_a)$. The expected value and variance of the π_a are given by p_a and $p_a(1 - p_a)/(1+\omega)$ respectively, while the covariance between π_a and $\pi_{a'}$ is given by $-p_a p_{a'}/(1 + \omega)$. Given a particular value of $\underline{\pi}$ of $\underline{\pi}_j$ which represents the class probabilities for the local aggregation of fish that the jth haul samples with a size-m_j random sample of fish, then the number of fish in each class, $\underline{N}_j$, with realisation $\underline{n}_j = (n_{j,1},\ldots,n_{j,A})$, conditional on m_j and $\underline{\pi}_j$ can be simply assumed to have the multinomial distribution given by

$$\Pr\left(\underline{N}_j = \underline{n}_j \left| \underline{\pi}, m_j\right.\right) = \frac{m_j!}{\prod_{a=1}^{A} n_{ja}!}\prod_{a=1}^{A}\pi_{ja}^{n_{ja}} \qquad (4)$$

where $m_j = \sum_{a=1}^{A} n_{ja}$ and $\sum_{a=1}^{A}\pi_{ja} = 1$ (for $j = 1,\ldots,h$). More rigorously, the size-m_j sample is a random sub-sample of the total number of fish caught in the haul, say m_j', so a multivariate hypergeometric distribution (Johnson and Kotz, 1969) for $\underline{N}_j$ conditional on both $M = m_j$ and $M' = m_j'$ would be more appropriate. However, given m_j' is typically estimated, sometimes with poor accuracy at the haul-level, estimation error in M_j' adds a much greater degree of complexity and will not be considered further.

The distribution of $\underline{N}_j$ expected for a random haul is given by the marginal distribution (i.e. calculated as the expectation across random values from the distribution of $\underline{\pi}$ given by equation 3) and is given by Mosimann (1962, equation 7) and Johnson and Kotz (1969, equation 102)

$$\begin{aligned}\Pr\left(\underline{N}_j = \underline{n}_j \left| m_j\right.\right) &= \frac{m_j!}{\prod_{a=1}^{A} n_{ja}!} E\left(\prod_{a=1}^{A}\pi_{ja}^{n_{ja}}\right) \\ &= \frac{\Gamma(\omega)}{\Gamma(m_j + \omega)}\frac{m_j!}{\prod_{a=1}^{A} n_{ja}!} \\ &\prod_{a=1}^{A}\frac{\Gamma(n_{ja} + \omega p_a)}{\Gamma(\omega p_a)} \qquad (5).\end{aligned}$$

From this marginal distribution the expected value and variance of N_{ja}, conditional on m_j, are given by $m_j p_a$ and $\phi_j m_j p_a(1-p_a)$ respectively, while the covariance between N_{ja} and $N_{ja'}$ is given by $-\phi_j m_j p_a p_{a'}$ where ϕ_j is a dispersion parameter which depends on m_j obtained from the relation $\phi_j = (m_j + \omega)/(1+\omega)$ (Johnson and Kotz, 1969, equations 105, 106; Hrafnkelsson and Stefánsson, 2004). If $\bar{\phi}$ is defined as the average value of the ϕ_j then $\bar{\phi} = (\bar{m} + \omega)/(1 + \omega)$, where $\bar{m} = \frac{1}{h}\sum_{j=1}^{h} m_j$, then a single parameter corresponding to ϕ can be defined as $\bar{\phi}$ (for the following the bar notation will be dropped). The intra-cluster correlation coefficient for the D-M distribution, ρ, is given by $\rho = 1/(1 + \omega)$ (Zhu, 2002) and this correlation is therefore positive. The greater variability in the $\underline{\pi}_j$ due to lower values of ω results in higher values of ρ.

The catch-at-age or catch-at-length data that CASAL requires are the frequencies aggregated across hauls given by the sum of h random haul frequency sets to give $n_a = \sum_{j=1}^{h} n_{ja}$. If $\underline{N}_j$ has the D-M distribution (equation 5), then the sum $\underline{N} = \sum_{j=1}^{h}\underline{N}_j$ with sample realisation $\underline{n} = (n_1,\ldots,n_A)$ has a complex distribution determined by the h-fold convolution of density function (5) on itself (Feller, 1968). However, the mean and variance of $\underline{N}$ are easily calculated with corresponding ath components of $h\bar{m}p_a$ and $\varphi h\bar{m} p_a(1 - p_a)$ respectively, and (a,a')th covariance component of $-\varphi h\bar{m} p_a p_{a'}$

where $\varphi = \left(\frac{1}{h\bar{m}} \sum_{j=1}^{h} m_j^2 + \omega \right) (1+\omega)^{-1}$ (6).

This follows simply from the independence of the haul-level samples and is a special case of that given by Brown and Payne (1986). Note that if $m_j = m$ for all hauls, then $\phi_j = \phi = \varphi$.

In terms of CASAL's requirements, the theoretical ESS determined as $n' = h\bar{m} / \varphi$ can be used since the multinomial approximation to the true distribution of $\underline{N}$ is correct up to second-order moments in that the mean and (co)variance are the same as the theoretical values for the h-fold convolution of the D-M distribution.

In terms of maximum likelihood estimation of ω the distribution of $\underline{N}$ is not required since this estimation uses the haul-level frequencies with marginal likelihood (based on equation 5).

Estimation methods

In order to handle multiple years of data collection, the y subscript is introduced to represent the year. The first method estimates φ_y so that given the known values of m_{yj} the estimate of n'_y can be obtained. This estimator of φ_y assumes that, for each year considered independently, the haul-level class frequencies are random realisations of a D-M distribution and use an approximate profile likelihood to determine ω_y from haul-level class frequency data and thus determine φ_y from equation (6) in order to determine the variance of the aggregate frequency across hauls for the year y.

Maximum likelihood estimation requires the log-likelihood based on equation (5), ℓ, to be maximised with respect to the parameter set $\theta = (\omega, p_1, \ldots, p_{A-1})$ where

$$\ell = \sum_{j=1}^{h} \log_e \left\{ \Pr\left(\underline{N}_j = \underline{n}_j \middle| m_j, \theta \right) \right\} \quad (7).$$

However, in the simulation study A is large (49 age classes, 19 length bins) so in order to simplify estimation of ω, since the p's are effectively nuisance parameters, ℓ was maximised by profiling ℓ across values of ω while fixing the value of the p's to their sample estimates given by $\tilde{p}_a = n_a / n$ for $a = 1, \ldots, A$. If any of the classes have n_a of zero and thus $\tilde{p}_a = 0$ the value of $\tilde{p}_a$ in this case, for the purposes of determining ℓ from equations (5) and (7), was set to a small arbitrary constant of 1.0e^{-8}. The size of this arbitrary constant does not affect estimation since it can be seen from equation (5) that the contribution to ℓ in this case is zero. The estimate of ω, $\hat{\omega}(\underline{\tilde{p}})$, obtained in this way is called the profile maximum likelihood estimate (PMLE) but strictly it is only an approximation to the true PMLE since it is conditioned on $\underline{\tilde{p}}$ and not the maximum likelihood estimate of $\underline{p}$ obtained given a value of ω. The merits or otherwise of this approximation is left to the discussion. Suppressing the dependence on $\underline{\tilde{p}}$, this estimate of ω is used to obtain the approximate PMLE of φ, $\hat{\varphi}(\hat{\omega})$, using the relation between φ and ω that is given by equation (6). The estimate of effective sample size given by $\hat{n}'(\hat{\varphi}) = h\bar{m} / \hat{\varphi}(\hat{\omega})$ corresponds to estimation Method 1 (PML). The theoretical (i.e. true) effective sample size is given by $n'(\varphi) = h\bar{m} / \varphi(\omega)$. When catch-at-age frequencies for a series of years are given these formulae are simply replicated for each year, so the profile maximum likelihood estimator is given by

$$\hat{n}'_y = h_y \bar{m}_y / \hat{\varphi}_y (\hat{\omega}_y) \quad (8).$$

Method 2 (MI) is that described in Appendix 2 of McAllister and Ianelli (1997). McAllister and Ianelli (1997) used the year-level frequencies and so can only provide estimates of n'_y if catch-at-age frequencies for a series of years are available. They estimate n'_y directly based on the empirical estimates of the observed class proportions, at the year-level, $o_{ya} = n_{ya}/n_y$ and the relationship between their variance and expected value as a function of predicted class proportions, p_{ya}, assuming a multinomial distribution of n'_y independently sampled fish. The predicted class proportions were obtained from the fit of an integrated assessment model using an age-structured population model fitted to a commercial catch series, a catch-at-age series, and two abundance series obtained from trawl surveys. To provide initial estimates of p_{ya}, a starting constant value for effective sample size, n' (i.e. $n'_y = n'$), is used in the integrated assessment, then n'_y is estimated externally via the above method and averaged across years to give an updated estimate of n'. This two-step procedure is iterated until n' converges. Although the McAllister and Ianelli (1997) estimation method is not implemented here within a full simulated integrated assessment, the simulation environment used is a simple and ideal way to investigate the principle of their method, which is estimation of n'_y from the year-to-year 'lack-of-fit' (LOF) in p_{ya} as a predictor of o_{ya}. Since the four estimation methods are compared using the simulation model with no process error or systematic error, a simple mean of the o_{ya} over the set of years is the appropriate value to use for p_{ya}.

Therefore the equivalent method to McAllister and Ianelli (1997, Appendix 2, equation 2.5) used here is to estimate n'_y directly by

$$\hat{n}'_y = \frac{\sum_{a=1}^{A} \bar{p}_a \left(1-\bar{p}_a\right)}{\sum_{a=1}^{A} \left(o_{ya} - \bar{p}_a\right)^2} \tag{9}$$

where $o_{ya} = \tilde{p}_{ya}$ and where $\tilde{p}_{ya} = n_{ya} / n_y$ and $\bar{p}_a = \frac{1}{R}\sum_{y=1}^{R} \tilde{p}_{ya}$ where R is the number of years in the series. Note that this method of calculating ESS was also used by Methot (2000, equation 37).

The third method (Dunn) is that of Dunn and Hanchet (2007) who used the individual haul-level frequencies to estimate n'_y directly based on a non-linear regression of ln(c_{ya}) on $\ln\left(\tilde{p}_{ya}\right)$ where the c_{ya} are the empirical coefficients of variation (CVs) of the mean of the observed class proportions across hauls, where the haul-level proportions are $o_{yja} = n_{yja}/n_{yj}$ where $n_{yj} = \sum_a n_{yja}$, so that $c_{ya} = \left[Var\left(o_{yja}\right)/\left(h_y \tilde{p}_{ya}^2\right)\right]^{\frac{1}{2}}$. Dunn and Hanchet (2007) solve the following equation for n'_y

$$\ln\left(c_{ya}\right) = \ln\left(\frac{\left[n'_y\left\{1-\tilde{p}_{ya}\right\}\right]^{\frac{1}{2}}}{n'_y \tilde{p}_{ya}^{\frac{1}{2}}}\right) \tag{10}$$

using nonlinear least squares (A. Dunn, pers. comm.). The least-squares estimate of n'_y can also be obtained from the fit of the following linear regression, separately for each year, $\ln\left(c_{ya}\right) = a_y - \frac{1}{2}\ln\left(\tilde{p}_{ya} / \left\{1-\tilde{p}_{ya}\right\}\right)$ where the single parameter a_y is estimated and the fixed term $-\frac{1}{2}\ln\left(\tilde{p}_{ya} / \left\{1-\tilde{p}_{ya}\right\}\right)$ is included as an offset. In this case the ESS, n'_y, is then estimated as exp($-2\hat{a}_y$). This linear regression with a single estimated parameter was used to apply the Dunn and Hanchet (2007) method.

An alternative to the Dunn and Hanchet (2007) regression is to fit c_{ya}^2 as a linear regression, again separately for each year, through the origin, on $\left(1-\tilde{p}_{ya}\right)/\tilde{p}_{ya}$ so that the estimate of n'_y is given by the inverse of the regression slope estimate. The fourth and final method (glm) examined is that of estimating n'_y in the above way after fitting this simple linear regression as a generalised linear model (GLM) with gamma error structure and a linear 'link' function (McCullagh and Nelder, 1989) with again the regression constant omitted. The difference between Methods 3 and 4 is the assumption of a lognormal distribution for c_{ya} for Method 3 compared to a gamma distribution assumed for c_{ya}^2, for Method 4. For both the Dunn and Hanchet (2007) regression and gamma GLM methods it is necessary to exclude values of c_{ya} from the regression for which $\tilde{p}_{ya}$ is zero since in this case c_{ya} cannot be calculated. Also, to avoid extreme values of c_{ya}, those values of c_{ya} for which $h_y c_{ya}^2$ was either zero or greater than 15 were excluded.

Estimates of $\hat{n}'$ are compared to the theoretical value given earlier by $n' = h\bar{m}/\varphi$ using summary statistics for percent relative error, $100\left(\hat{n}' - n'\right)/n'$, and the relative mean square error (RMSE) given by $\sum_{r=1}^{S}\left(\hat{n}'_r - n'_r\right)^2 / \left(\frac{S}{S-1}\sum_{r=1}^{S}\left(n'_r - \bar{n}'\right)^2\right)$ where S is the total sample size of estimates (i.e. 2000 in the simulation studies) and $\bar{n}'$ is the mean of the simulated theoretical values of ESS. Smaller values of RMSE indicate better accuracy and in the limit a perfect estimator has a 1:1 relationship with the theoretical ESS, n', and thus an RMSE of zero.

Process error and systematic lack-of-fit

Assessment methods for the Ross Sea (Dunn and Hanchet, 2007) and South Georgia toothfish fisheries (Hillary et al., 2006) incorporate estimates of process error using an iterative procedure of fitting CASAL and then using lack-of-fit statistics to determine process error. These estimates of process error are then applied with an updated CASAL run and this two-step procedure repeated until there are only small changes in the estimates of process error (A. Dunn, pers. comm.). For catch-at-length or catch-at-age data, incorporation of process error involved modifying the ESS values derived from Method 3 above. However, for process errors to be random deviations of model-fitted values from observed values, then any SLOF, either across age classes/length bins or across years, should first be removed.

Fitting a generic SLOF model was achieved here by fitting a simple parametric model to the deviations between observed and fitted values in proportions by age class or length bin, where the fitted values are the simple means across years as described for Method 2. The parametric model incorporated both linear and quadratic terms in the continuous values of age (or length bin mid-point) and year number, including the interaction between years and age for these terms (i.e. lin_{bin} x lin_{yr}, lin_{bin} x $quad_{yr}$, $quad_{bin}$ x lin_{yr}, $quad_{bin}$ x $quad_{yr}$). A Poisson GLM was fitted to the nominal counts obtained as the ESS values multiplied by proportions by age class or length bin to give numbers by class/bin and year, an offset of log of predicted number, the above parametric SLOF model, and a main effect of year (as a factor) in order for the Poisson model to be constrained to give a log-likelihood equivalent to that of the multinomial. Class or bin number was also included as a factor. This model is described by

$$\log_e\left\{E\left(Z_{ya}\right)\right\} = \log_e\left(\hat{n}'_y\widehat{p}_{ya}\right) + \beta_1 + \sum_{h=2}^{Y}\beta_h I_{yh} + \sum_{r=2}^{A}\beta_{Y+r-1}J_{ar} + \beta_{Y+A}y'a' + \beta_{Y+A+1}y'^2a' + \beta_{Y+A+2}y'a'^2 + \beta_{Y+A+3}y'^2a'^2 \quad (11)$$

where a Poisson distribution is assumed for the response variable $Z_{ya} = \hat{n}'_y o_{ya}$, y' is the centred year value obtained from the integer values of year, y, so that for a 20-year simulation period $y' = y - 10$, similarly a' is the centred value obtained from the integer values of age or the mid-point of the length bins, I_{yh} is a set of dummy variables specifying years as a 'factor' so that $I_{yh} = 1$ if $h = y$ and zero otherwise, similarly J_{ar} is a set of dummy variables specifying classes/bins as a 'factor' so that $J_{ar} = 1$ if $r = a$ and zero otherwise, and the β's are parameters to be estimated. Note that since year is included as a factor, the Poisson deviance for this model is the same as that for Z_{ya} considered as multinomial conditional on the $\hat{n}'_y$ (McCullagh and Nelder, 1989, p. 212) and, additionally, quasi-likelihood theory (McCullagh and Nelder, 1989, p. 323) allows this response variable to be non-integer as long as it takes positive (i.e. including zero) values.

The adjustment to the estimated ESS due to process error was obtained simply by dividing $\hat{n}'_y$ by the dispersion parameter which was estimated as the residual mean deviance (McCullagh and Nelder, 1989) from the fit of the SLOF model. Therefore if the dispersion parameter estimate, $\hat{\Phi}$, is obtained as the residual mean deviance, then the ESS adjusted for process error, after removing SLOF, is given by $\hat{n}''_y = \hat{n}'_y / \hat{\Phi}$. This adjustment is based on Finney's heterogeneity factor (Finney, 1971; McCullagh and Nelder, 1989, p. 128) and is consistent with the approach of scaling the actual sample size given in the introduction where in this case the year-specific estimated ESSs are considered as 'actual' sample sizes. Note that a single scale parameter, $\hat{\Phi}$, is used across all years. This approach of simply scaling by $\hat{\Phi}$ is a heuristic approximation in the absence of a theoretical statistical model that can incorporate process error in addition to between-haul heterogeneity and at the same time allow these random processes to be accounted for by the use of an ESS. In practice in order to prevent over-fitting, the full SLOF model above is only fitted if the dispersion parameter for the minimal SLOF model (i.e. only including the offset and year plus class/bin factors in the linear predictor) is significantly greater than 1. It is worth noting that a parametric model is required to model the age class/length bin by year interaction since, if the non-parametric term consisting of the interaction of these two variables considered as factors was included in the GLM, then the residual deviance would be zero since the model would be 'saturated' with parameters. Also, in the simulation studies described below, the simple averages across years were used for predictions, so that $\widehat{p}_{ya} = \widehat{p}_a$, whereas for a general integrated assessment the predictions $\widehat{p}_{ya}$ would vary by year.

The simulation model

A simulation model consisted of simulating 100 replications of a series of 20 fishing years of either catch-at-age or catch-at-length frequencies.

First, a population age-structure was generated by calculating the probability density for each of 1 001 values of age, a', taken uniformly between age 0 and age 50 (i.e. consecutive age values 50/1 001 years apart) using the exponential density function $\Pr(Age = a') = P(a') = Me^{-Ma'}$ where M is the instantaneous mortality rate parameter. The raw density values, $P(a')$, were scaled by dividing by their sum to give $P'\left(a'_i\right) = P\left(a'_i\right) / \sum_{i=1}^{1001} P\left(a'_i\right)$. Using the 50 integer age classes given by {(0,1],…,(49,50]}, denoting the vector $\underline{a}$ as the upper limit of each class, the class membership of each of the 1 001 ages was determined and used to determine class probabilities. This numerical integration method of determining the age structure was used in preference to the method that uses the simple analytical integral, since the former allows more accurate calculation of class probabilities when fishing selectivity is incorporated as described next.

If fishing selectivity was to be imposed, then $P'(a)$ was multiplied by $S(a)$ to give $P''(a)$ where $S(a)$ was determined from the double-normal selectivity function given by

$$S(a) = 2^{-\left[(a-\lambda)/\sigma_L\right]^2} \quad ; \quad a \le \lambda$$
$$= 2^{-\left[(a-\lambda)/\sigma_U\right]^2} \quad ; \quad a > \lambda \qquad (12)$$

where λ is a cut-point parameter corresponding to the age at which $S(a)$ is 1, and σ_L and σ_U are parameters denoting the standard deviations of the scaled normal density functions specifying the lower and upper arms of the function respectively. For the simulations, the parameters σ_L and σ_U were always set to 1 and 25 respectively. For simulation Model 1 the cut-point parameter λ was set to 7 years. The integer-age population-class probabilities were obtained by accumulating the sum of the $P''(a)$ using their integer-age class membership to give P_a where $a = 1,\ldots,50$.

The line in Figure 1 shows the values of P_a for $S(a)$ obtained from equation (12) with λ set to 7 years (Model 1).

Simulated catch-at-length data were generated by converting P_a to be a function of length using a von Bertalanffy growth model for length as a function of age with expected value $l(a)$ and lognormal errors with a CV of 0.1. Therefore the equivalent marginal function to P_a in terms of length is given by

$$P'_k = \sum_{a=1}^{50} \Pr\left(L = l \in B_k \middle| a\right) P_a$$

where B_k represents the kth length bin where 19 bins from 300 to 2 200 mm were used giving a bin width of 100 mm and with the first bin having a range of 1 to 300 mm giving 20 bins in total. The line in Figure 2 shows P'_k versus length bin mid-points generated using the age distribution and age-dependent selectivity function used to generate Figure 1.

The second step of the simulation involved generating random catch-at-age data from simulated sampling of the population age structure. For each of the 20 years in a replicate, catch-at-age frequencies were generated for a sample of 100 hauls where across hauls the expected number of fish sampled, m, was 150. Alternative simulations where m was reduced to 100 and 50 were also carried out to determine the effect of expected sample size on the accuracy of the alternative estimators. The expected value of the total sample size for a year, n_y, is therefore 15 000, 10 000 or 5 000 corresponding to expected haul-level sample sizes of 150, 100 and 50 respectively. For each haul, random values of $\underline{\pi}$ were drawn from the Dirichlet distribution with expected value of the typical element of P_a using the method described above in 'The Dirichlet-multinomial model' where for each replicate and year (but not each haul) a separate random value of ϕ was drawn from a Poisson distribution with a specified expected value. Extra-Poisson variability in sample draws of ϕ was applied by multiplying the Poisson-expected value by the exponential of a normal error with expected value of zero and standard deviation of c_ϕ. For all simulation models c_ϕ was set to 0.1. The corresponding value of the Dirichlet parameter ω was obtained from equation (6) assuming $\phi=\varphi$. Given ω, $\underline{\pi}$ and m, multinomial samples $\underline{n}_j$ for a random haul conditional on $\underline{\pi}_j$ were drawn by first drawing a random sample size $M_j = m_j$ for the haul using a Poisson distribution given by $\Pr\left(M_j = m_j\right) = e^{-m} m^{m_j} / m_j!$ where $E(M_j) = m$. Given the value of m_j the multinomial frequencies $\underline{n}_j$ (where $\sum_{a=1}^{A} n_{ja} = m_j$) were generated using a set of sequential conditional binomial samples commonly known as continuation ratios (Fienberg, 1980) using the R-function *rmultinom*() (R Development Core Team, 2006). The same method was used to simulate catch-at-length frequencies with P'_k replacing P_a.

For simulation Model 2, to obtain the P_a for each year, the selectivity parameter λ was varied randomly as a lognormal variate with expected value 7 and CV (c_λ) of 0.2. For simulation Model 3, the selectivity parameter λ was varied linearly from 5.88 for the first year of the 20-year simulation period to 8.12 for the last year with a value at year 10.5 of 7.

For simulation Model 4 to obtain the P_a for each year, annual recruitment was varied as a lognormal variate about a mean of 100 000 with CV (c_R) of 0.6 with an initialisation period of 50 years (i.e. the 20-year simulation period for evaluating the SLOF model started in year 51). For Model 4, the mortality rate, M, was varied also as a lognormal variable about a mean of 0.13 with CV (c_M) of 0.2 from the beginning of the initialisation period. For Model 5, the simulation was the same as Model 1 for the initialisation period, but after this period M was varied linearly starting at 0.1055 for year 1 and increasing to 0.1544 by year 20 with a value of 0.13 at year 10.5 and recruitment was also varied linearly, but in this case, declining from 196 939 for year 1 to 3 061 for year 20 with a value of 100 000 at year 10.5.

Therefore, a simulation model for a given number of replicates and years, is defined by the parameter values for A (number of integer-age classes or length bins), M, ϕ, σ_L, σ_U, λ, c_λ, c_ϕ, c_R, c_M and m (note that ϕ here means its expected value not the randomly drawn value from the Poisson distribution). For all simulations ϕ was set to 10. The

Table 1: Comparison of accuracy of estimation methods for simulation model 1[1] and catch-at-length data. RMSE – relative mean square error; IQ – inter quartile; PML – profile maximum likelihood; MI – McAllister and Ianelli (1997); Dunn – Dunn and Hanchet (2007).

Estimation method (number)	Expected sample size[2], m	Statistics for %relative error of estimate $\left[100(\hat{n}'-n')/n'\right]$				RMSE
		IQ1	Mean	IQ3	SD	
PML (1)[3]	50	–2.02	1.10	4.03	4.66	0.0610
	100	–2.22	1.27	4.44	4.90	0.0132
	150	–1.79	1.41	4.59	4.78	0.0126
MI (2)[4]	50	–21.12	39.58	71.65	96.49	6.6013
	100	–18.77	45.00	75.59	100.81	5.3729
	150	–16.66	43.99	74.32	94.00	5.1519
Dunn (3)[5]	50	1.71	6.88	11.44	7.71	0.0674
	100	1.51	6.82	11.45	7.88	0.0502
	150	1.87	7.17	11.80	7.73	0.0514
glm (4)[6]	50	–0.24	4.49	8.93	6.70	0.0447
	100	–0.65	4.11	8.64	7.12	0.0315
	150	–0.48	4.34	8.62	6.85	0.0308

1 No process error and no systematic error.
2 Sample sizes across hauls generated as a Poisson random variable with given expected value. Realised maximum and minimum sample sizes (min,max) for expected values of 50, 100, 150 were (35,63), (68,145), and (116,177) respectively.
3 Based on approximate PMLE of ω.
4 McAllister and Ianelli (1997)
5 Dunn and Hanchet (2007)
6 New estimation method based on fit of a gamma GLM.

simulation models were coded in the R-package (R Development Core Team, 2006) in a way that allows all these parameters to be varied.

Results

The points in Figure 1 show the probabilities, $\tilde{p}_a = n_a / n$, for one year's set of randomly generated catch-at-age frequencies where these values were obtained with selectivity given by equation (12) with λ set to 7, a value of ϕ of 10 and m = 150. Figure 2 shows the corresponding catch-at-length bin probabilities.

The key feature of the first simulation model tested (Model 1) was that selectivity was set using equation (12) with λ set to 7 and each of c_λ, c_R and c_M set to zero thus giving zero year-to-year process error and systematic error. Table 1 gives summary statistics for percent relative error and the RMSE for each estimation method for the catch-at-length data and sample sizes of 50, 100 and 150 fish per haul. Figure 3 shows the mean of the estimates of ESS versus mean theoretical values of ESS for a sample size of 150 fish where means were calculated over the 20 years giving 100 mean values in each case. The corresponding results for Model 1 simulations of catch-at-age data are not given, since they were very similar to those for the catch-at-length data.

The performance of the full SLOF model is compared using simulation Models 2 to 5. For simulation Models 2 and 4 where only random process error is involved, the estimate $\hat{\Phi}$ for the full SLOF model should be substantially greater than 1 and only slightly less than the estimate obtained from the fit of the minimal SLOF model. For simulation Models 3 and 5 where only systematic error is involved, the estimate $\hat{\Phi}$ for the full SLOF model, if successful in removing SLOF, should be close to 1 and substantially less than the estimate obtained from the fit of the minimal SLOF model.

Table 2 gives summary statistics for $\hat{\Phi}$ for the 100 replicates of the 20-year simulation period. The residual degrees of freedom for the full and minimal SLOF models were 357 and 361 respectively.

Corresponding results for simulations of catch-at-age data were similar, with the performance of the generic SLOF model slightly better than that for the catch-at-length data for simulation Models 2, 4

Table 2: Comparison of dispersion parameter estimates from fit of minimal and SLOF models to simulation model catch-at-length frequencies. IQ – inter quartile.

Simulation model	Expected sample size, *m* (min,max)	Full SLOF model			Minimal SLOF model			Ratio[1] SLOF/Min
		IQ1	Median	IQ3	IQ1	Median	IQ3	
2[2]	50 (39,67)	3.453	3.902	4.783	3.727	4.238	5.131	0.921
	100 (75,123)	6.197	7.427	8.805	6.504	7.842	9.589	0.947
	150 (128,185)	8.877	11.060	12.920	10.010	12.090	14.100	0.915
3[3]	50 (35,65)	0.878	0.917	0.979	1.670	1.829	1.935	0.501
	100 (78,123)	1.149	1.225	1.295	2.781	2.965	3.176	0.413
	150 (124,180)	1.354	1.450	1.556	3.757	4.053	4.527	0.358
4[4]	50 (28,68)	1.665	1.976	2.546	1.910	2.334	2.712	0.847
	100 (79,122)	2.546	3.360	4.286	3.097	3.824	4.913	0.879
	150 (127,184)	4.151	5.115	6.676	4.516	5.613	6.927	0.911
5[5]	50 (32,78)	0.880	0.921	0.990	1.7585	1.905	2.066	0.483
	100 (70,121)	1.181	1.265	1.347	3.016	3.290	3.547	0.384
	150 (114,185)	1.418	1.534	1.670	4.283	4.557	4.926	0.337

1 Ratio of median dispersion estimate for full SLOF model to the corresponding median for the minimal model.
2 Random year-to-year variation in selectivity.
3 Systematic year-to-year variation (linear) in selectivity.
4 Random year-to-year variation in recruitment and mortality rate.
5 Systematic (linear) year-to-year variation in recruitment and mortality rate.

and 5 but slightly worse for simulation Model 3. For simulation Model 1, the median $\hat{\Phi}$ for both the minimal and full SLOF models was close to 0.7 for all three values of *m* which indicates over-fitting. When the class/bin factor was dropped from these models, the median estimate was close to 1 for both models and for all values of *m* as expected for this simulation model. However, when this factor was dropped from both minimal and full SLOF models for simulation Models 2 to 5, the performance of the SLOF model was not as good as that shown in Table 2. Note that Table 2 shows that $\hat{\Phi}$ increases as *m* increases which, as mentioned above, was not the case for simulation Model 1. This suggests that process and systematic error, in terms of their determination of $\hat{\Phi}$, are scaled by expected haul-level sample size, *m*.

Figure 4 shows the observed and predicted (i.e. the average across years) proportion by length bin for a single simulation of Model 3 and a sample size of 100 fish per haul. Figure 5 shows the predicted SLOF trends (i.e. constructed using only the terms in continuous variables and excluding the offset) corresponding to observed and predicted proportions shown in Figure 4.

Figures 6 and 7 show the corresponding results for a single simulation of Model 5.

Discussion

It is clear from the results of the simulation studies that when between-haul (within-year) heterogeneity in class proportions is simulated using the D-M model, and year-to-year process error is set to zero, that within-year, haul-by-haul class frequencies are required in order to accurately estimate the appropriate effective sample size. The methods that use haul-by-haul data gave acceptable accuracy. Method 1 was better than Methods 3 and 4 (Table 1;

Figure 3) for catch-at-length data with very similar results obtained for catch-at-age data (results not given). The haul-level profile maximum likelihood method (Method 1, 'PML') of estimating ESS introduced here gave substantial improvement in accuracy to that described by Dunn and Hanchet (2007) (Method 3, Dunn) under the assumption of a D-M distribution (Table 1; Figure 3). The gamma GLM (Method 4, glm) is slightly superior to Method 3, under the D-M assumption. Method 1 is reasonably precise in estimating ESS with a standard deviation in percent relative error of only 5%. The effect of varying haul-level sample size on accuracy of estimates is very slight (Table 1).

Method 1 is a simple approximate implementation of the PMLE which handles zero-class probabilities very conveniently since, as described earlier, they drop out of the log-likelihood when the zeros are replaced by an arbitrarily small number. Also, this method accounts directly for the variability in the haul-by-haul sample size, m_j. In order to estimate the D-M dispersion parameter, ω, the likelihood was conditioned on the sample values for the class probabilities where these are sufficient statistics when the haul-level data is assumed to be multinomial rather than D-M. This is an approximation but it appears adequate for the purpose of estimating ESS, although Table 1 and Figure 3 show that there is a slight positive bias in the estimates but this is minimal compared to the alternative estimators. If full maximum likelihood estimation is required for all parameters in the set $\theta = (\omega, p_1, \ldots, p_{A-1})$ simultaneously under a D-M distribution, then the exact Fisher information matrix is given by Paul et al. (2005). Full maximum likelihood estimation, with the appropriate constraint of $\sum_{a=1}^{A-1} \hat{p}_a \leq 1$, could be used to potentially improve estimation over the PML method used here but at considerable extra computational cost.

Clearly the method of McAllister and Ianelli (1997) (Method 2, MI) which relies on model lack-of-fit to estimate ESS from year-level data gives poor estimates of the appropriate ESS when heterogeneity is between hauls (Table 1; Figure 3).

When substantial year-to-year process error was introduced into the simulations using Models 2 and 4, the accuracy of Methods 1, 3 and 4 were unaffected in terms of the appropriate ESS for between-haul heterogeneity (results not given) which is not surprising since these models condition on the specific values of $\tilde{p}_{ya}$ for each year. The method of McAllister and Ianelli (1997) obtains n' by averaging year-to-year 'lack-of-fit' in the $\tilde{p}_{ya}$. Assuming for simplicity that the errors are Gaussian, then year-to-year 'lack-of-fit' in the $\tilde{p}_{ya}$ would be the weighted sum of the three components of: (i) mean squares of systematic year-to-year deviations of o_{ya} from $\hat{p}_{ya}$; (ii) variance of any additional random year-to-year deviations (i.e. process error); and (iii) variance of within-year (i.e. between-haul) deviations respectively. For mixed-effect multinomial distributions these components do not partition so neatly into additive components but the principle is the same. This study has concentrated on estimating the effective sample size to account for the last of these components, however, the case for including year-to-year process error in calculating n'_y was addressed in order to clear up some of the confusion in these issues. For example, in the justification for the need to scale down the actual sample size of fish measured for age, n_y, McAllister and Ianelli (1997, Appendix 2) mention only the effect of the cluster sampling of fish by level-2 sampling units (i.e. hauls for trawlers or sets for longliners) in order to allow a multinomial likelihood to be used when the true marginal variance is given by equation (2). However, it is clear that the sources of variation in o_{ya} in addition to (iii) above, that are involved in determining n' using their method, is that due to at least (ii) and possibly both (i) and (ii).

In contrast to the McAllister and Ianelli (1997) method, the approach of Dunn and Hanchet (2007) and the other two approaches introduced here, first estimate an ESS due to between-haul heterogeneity using haul-level data and then scale this ESS to account for possible process error. The extra step recommended here is that systematic lack-of-fit should first be examined and removed before calculating the effect of process error on the final ESS. Unfortunately, since an appropriate distributional model that incorporates process error in addition to haul-level heterogeneity while giving a marginal variance relationship which allows an ESS to be appropriately defined does not appear to be available, it is not possible to compare these two methods of adjusting ESS for process error to a theoretical value as was the case when process error was not included. Nevertheless, the results given in Table 2 show that the full SLOF has close to the required properties since when only process error was simulated the over-dispersion parameter estimate was only slightly reduced to a value of about 90% of that of the minimal SLOF model and when only SLOF was present the over-dispersion parameter estimated was substantially reduced to close to 1. This suggested that the full SLOF model can be fitted routinely since it gives a reasonably close approximation to the appropriate over-dispersion parameter estimate whether or not process errors or systematic errors are present in catch-at-length predicted frequencies. However, it is recommended that ESS values should only be scaled by $\hat{\Phi}$ when the estimate of $\hat{\Phi}$ is greater than 1, in case the minimal SLOF model is over-parameterised, particularly due to the presence in the model of the class/bin factor. For simulation Models 3 and 5, depending on the sample size of fish, failure to adjust for SLOF using the full model would have resulted in reducing the ESS to between approximately one half and one third of the appropriate value (i.e. corresponding to a bias ranging from 50% to 66%).

Figures 5 and 7 demonstrate that the generic SLOF model has been able to model the SLOF in catch-at-length proportions displayed in Figures 4 and 6 respectively. The possible exception is for years 1 and 2 in Figures 6 and 7 where the SLOF trend appears stronger in Figure 7 than that expected from Figure 6. For years 10 and 11, which straddle the time point for which the parameters assumed by the age-structured model are correct compared to the simulation, the trend line has close to zero slope across length bins and falls on the zero deviation line in both Figure 5 and 7 as expected.

The use here of the interaction of linear and quadratic terms to remove SLOF is only a crude empirical approximation to any possible true SLOF and was not derived from any particular mechanism that generates SLOF such as a model of the effect of the linear shift in selectivity parameter λ on age-class/length-bin probabilities corresponding to simulation Model 3. However, the main message from the simulation of SLOF and process error is that it is incorrect to routinely attribute all LOF to process error. If this were done, the ESS would be substantially over-corrected (Table 2). If SLOF is substantial (e.g. as in Figure 4), then steps should be taken to remove or reduce this problem by restructuring the underlying model that generates predicted age-class or length-bin probabilities. In this particular case, if the source of the SLOF were known, then this could be accommodated in CASAL using the facility to fit the selectivity function (equation 12) with a linear shift in parameter λ. However, the SLOF model is useful when the source of SLOF is not clear or cannot be rectified by changing the underlying age-structured assessment model. For example, different datasets may ‘pull’ age-structured model parameters in different directions resulting in SLOF for one or more datasets depending on the weight they receive in estimation (Candy and Constable, 2008). Nevertheless, the SLOF model should be examined to determine if all obvious SLOF have been removed. Graphical examination of Pearson or deviance residuals (McCullagh and Nelder, 1989) from the SLOF model for any trends is one way to facilitate this.

Based on the above results and discussion, the following five-step approach to estimating the ESS for catch-at-age and/or catch-at-length data in integrated assessments is recommended:

1. Estimate the ESS for each year using haul-level data as either the PML or that estimate obtained from the fit of the gamma GLM to the empirical CVs.

2. Fit overall population dynamic/fishery model using CASAL (or other software) employing all datasets using an appropriately defined log-likelihood in each case.

3. Remove SLOF from the CASAL fit to the catch-at-age and/or catch-at-length data using the simple parametric SLOF model across years and age classes/length bins.

4. Scale the ESS obtained in Step 1 for process error by dividing by the residual mean deviance for the SLOF Poisson GLM model and incorporate process error estimated for other datasets appropriately into log-likelihoods (e.g. Candy and Constable, 2008).

5. Iterate Steps 2 to 4 until Step 2 parameter estimates converge. Note that the likelihoods defined in Step 2 may be inappropriate, in terms of systematic trends in residuals, if practically significant SLOF remains after Steps 3 and 4 of the previous iteration.

The results of this study are predicated on the assumption that the D-M distribution and the corresponding (i.e. 2nd order approximation) multinomial with estimated ESS applied in integrated assessment software such as CASAL is realistic enough in terms of actual catch-at-length or catch-at-age frequencies that conclusions are robust to departures from these assumptions. Hrafnkelsson and Stefánsson (2004) and Miller and Skalski (2006) indicate that positive correlations between classes/bins, which cannot be modelled by the D-M distribution, are prevalent and reasonably strong in actual data. However, in neither of these studies were spatial analyses presented that could indicate if there were any systematic spatial trends involving aggregations of similar age or length classes that could explain these correlations with deterministic rather than stochastic model terms.

Clearly further work is required in modelling actual and simulated data using appropriate population biology and statistical methods, and further development of integrated assessment software that can more adequately model such data within practical limitations of numerical methods is required. For example, incorporation of year and year-by-age-class/length-bin random effects terms is feasible whereas incorporation of haul-level random effects for large datasets is probably not numerically feasible for the near future.

Conclusions

An existing method of estimating ESS for catch-at-age or catch-at-length data in integrated assessments that uses aggregate year-level data is shown to have poor accuracy when heterogeneity in class proportions is due to between-haul within-year variation. Three alternative methods of estimation based on haul-level data are shown to greatly improve estimation. Two of these methods, derived from models for the empirical CV in the proportions, are used in CCAMLR fisheries and these are shown to give reasonable accuracy in estimating ESS. The approximate PMLE can also be used and gives improved accuracy. The issue of process error is examined and it is shown that when systematic 'lack-of-fit' across years is mis-specified as process error, then methods which attempt to account for process error give inappropriately low values of ESS. This has important implications for determining the appropriate implicit weight given to commercial catch length or age frequencies versus other datasets via their negative log-likelihood contribution to the objective function in integrated assessment software. A five-step approach to estimating the ESS for catch-at-age and/or catch-at-length data in integrated assessments is recommended.

Acknowledgements

Thanks are due to Mr John McKinlay and Dr Andrew Constable for helpful comments on this paper. Thanks are due also to Mr Alistair Dunn for his advice on these methods. Finally the comments of Prof. André Punt and Dr Timothy Miller were most helpful in revising an earlier draft of this paper.

References

Brown, P.J. and C.D. Payne. 1986. Aggregate data, ecological regression, and voting transitions. *J. Am. Stat. Assoc.*, 81 (394): 452–460.

Bull, B., R.I.C.C. Francis, A. Dunn, A. McKenzie, D.J. Gilbert and M.H. Smith. 2005. CASAL (C++ algorithmic stock assessment laboratory): CASAL User Manual v2.07-2005/06/23. *NIWA Technical Report*, 126: 268 pp.

Candy, S.G and A.J. Constable. 2008. An integrated stock assessment for the Patagonian toothfish (*Dissostichus eleginoides*) in Division 58.5.2 using CASAL. *CCAMLR Science*, this volume.

Constable, A.J., S.G. Candy and I. Ball. 2006. An investigation of integrated stock assessment methods for the Patagonian toothfish (*Dissostichus eleginoides*) in Division 58.5.2 using CASAL. Document *WG-FSA-SAM-06/14*. CCAMLR, Hobart, Australia.

Dunn, A. and S.M. Hanchet. 2007. Assessment models for Antarctic toothfish (*Dissostichus mawsoni*) in the Ross Sea including data from the 2006/07 season. Document *WG-FSA-07/37*. CCAMLR, Hobart, Australia.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*. Third Edition, John Wiley and Sons, New York.

Fienberg, S.E. 1980. *The Analysis of Cross-classified Categorical Data*. 2nd Edition. M.I.T. University Press, Cambridge, Mass.

Finney, D.J. 1971. *Probit Analysis*. 3rd Edition. Cambridge University Press, London.

Fournier, D.A., J. Hampton and J.R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Can. J. Fish. Aquat. Sci.*, 55 (9): 2105–2116.

Gelman, A., J.B. Carlin, H.S. Stern and D.A. Rubin. 2004. *Bayesian Data Analysis*. Second Edition. Chapman and Hall/CRC, Boca Raton.

Hillary, R.M., G.P. Kirkwood and D.J. Agnew. 2006. An assessment of toothfish in Subarea 48.3 using CASAL. *CCAMLR Science*, 13: 65–95.

Hrafnkelsson, B. and G. Stefánsson. 2004. A model for categorical length data from groundfish surveys. *Can. J. Fish. Aquat. Sci.*, 61 (7): 1135–1142.

Johnson, N.L. and S. Kotz. 1969. *Distributions in Statistics: Discrete Distributions*. Houghton-Mifflin, Boston.

McAllister, M.K. and J.N. Ianelli. 1997. Bayesian stock assessment using catch-at-age data and the sampling-importance resampling algorithm. *Can. J. Fish. Aquat. Sci.*, 54 (2): 284–300.

McCullagh, P. and J.A. Nelder. 1989. Generalized linear models. 2nd Edition. *Monogr. Stat. Appl. Prob.*, 37. Chapman and Hall, London: 511 pp.

Methot, R.D. 2000. Technical description of the stock synthesis assessment program. US Dept. Commerce. *NOAA-NWFSC Tech. Memo-43*.

Methot, R.D. 2005. *Technical Description of the Stock Synthesis II Assessment Program*, Version 1.17. Available from Richard Methot, NOAA Fisheries, Seattle, WA: 54 pp.

Miller, T.J. and J.R. Skalski. 2006. Integrating design- and model-based inference to estimate length and age composition in North Pacific longline catches. *Can. J. Fish. Aquat. Sci.*, 63 (5): 1092–1114.

Mosimann, J.E. 1962. On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika*, 49: 65–82.

Paul, S.R., U. Balasooriya and T. Banerjee. 2005. Fisher information matrix of the Dirichlet-multinomial distribution. *Biometrical J.*, 47 (2): 230–236.

Polacheck, T., J. Paige Eveson, G.M. Laslett, K.H. Pollock and W.S. Hearn. 2006. Integrating catch-at-age and multiyear tagging data: a combined Brownie and Petersen estimation approach in a fishery context. *Can. J. Fish. Aquat. Sci*, 63 (3): 534–548.

R Development Core Team. 2006. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0: www.R-project.org.

Zhu, Y. 2002. Correlated multinomial data. In: A.H. El-Shaarawi and W.W. Piegorsch (Eds). *Encyclopedia of Environmetrics*, Vol. 1. John Wiley and Sons, Chichester, UK: 479–484.

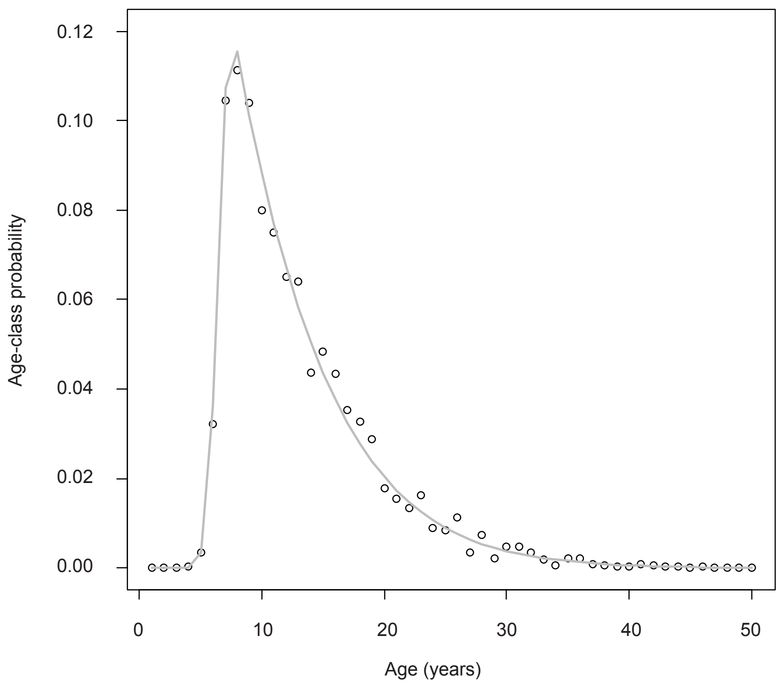

Figure 1: True population values (line) and simulated sample values for a single year (points) of integer-age class probabilities versus age, with fishing selectivity from equation (12) with λ, ϕ and *m* set to 7, 10 and 150 respectively .

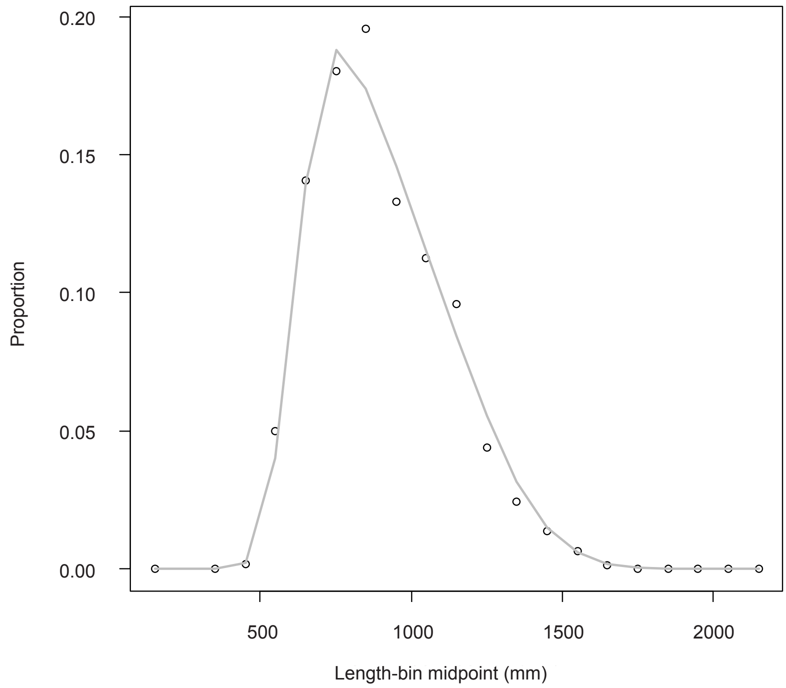

Figure 2: True population values (line) and simulated sample values for a single year (points) of length-bin probabilities versus length-bin midpoint, with fishing selectivity from equation (12) with λ, ϕ and m set to 7, 10 and 150 respectively.

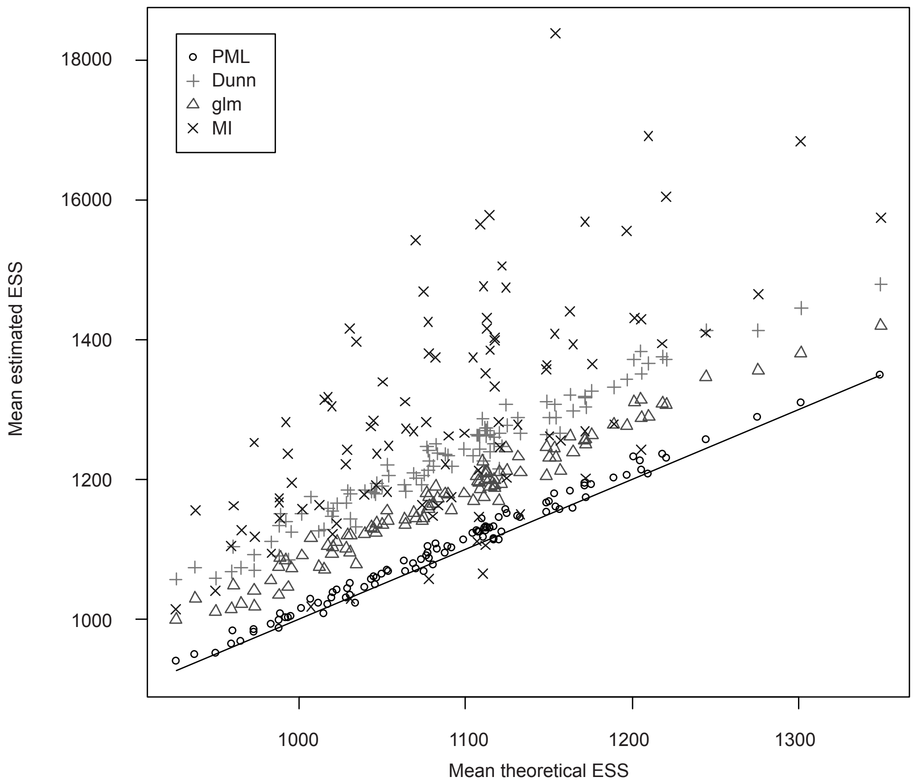

Figure 3: Mean estimated versus mean theoretical values of ESS for simulation Model 1 and catch-at-age data using Methods 1 (‘PML’), 2 (‘MI’), 3 (‘Dunn’) and 4 (‘glm’). The solid line is the 1:1 line.

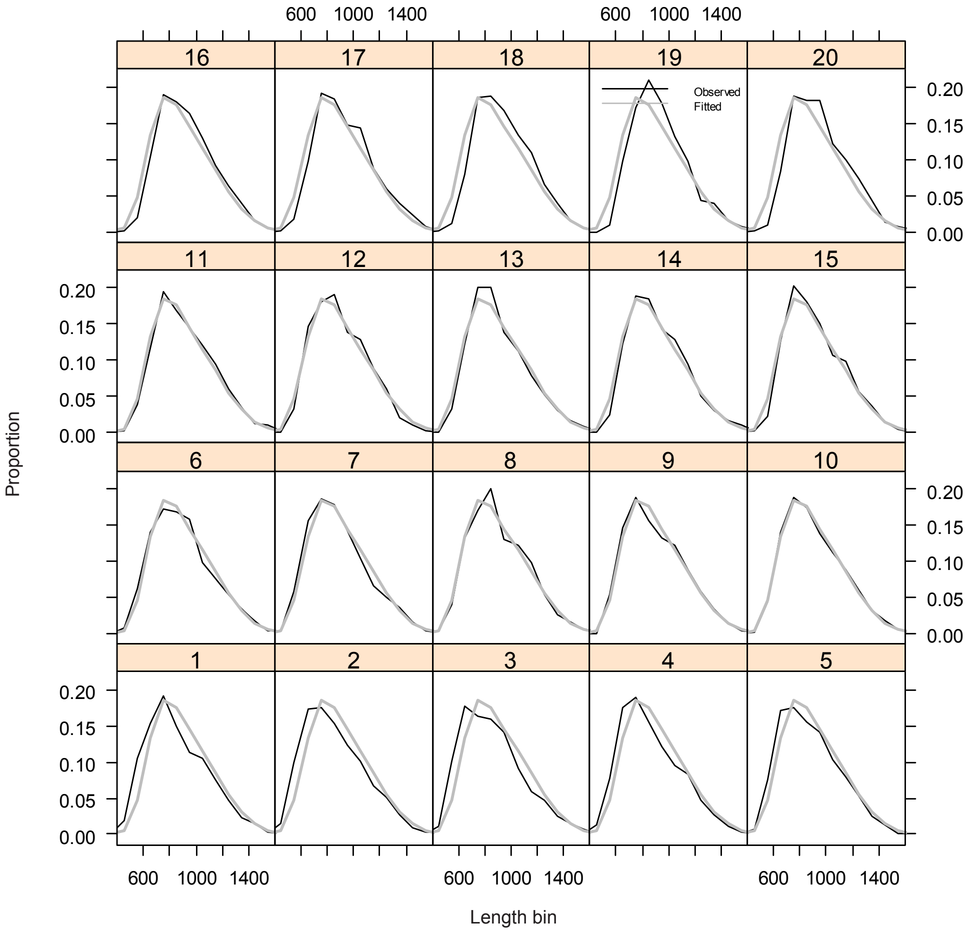

Figure 4: Observed and predicted proportions, by length bins, with predictions from an age-structured model assuming no year-to-year variation in selectivity for simulation Model 3 (Model 1 plus systematic variation in selectivity) and one replicate of a 20-year time series of catch-at-length data.

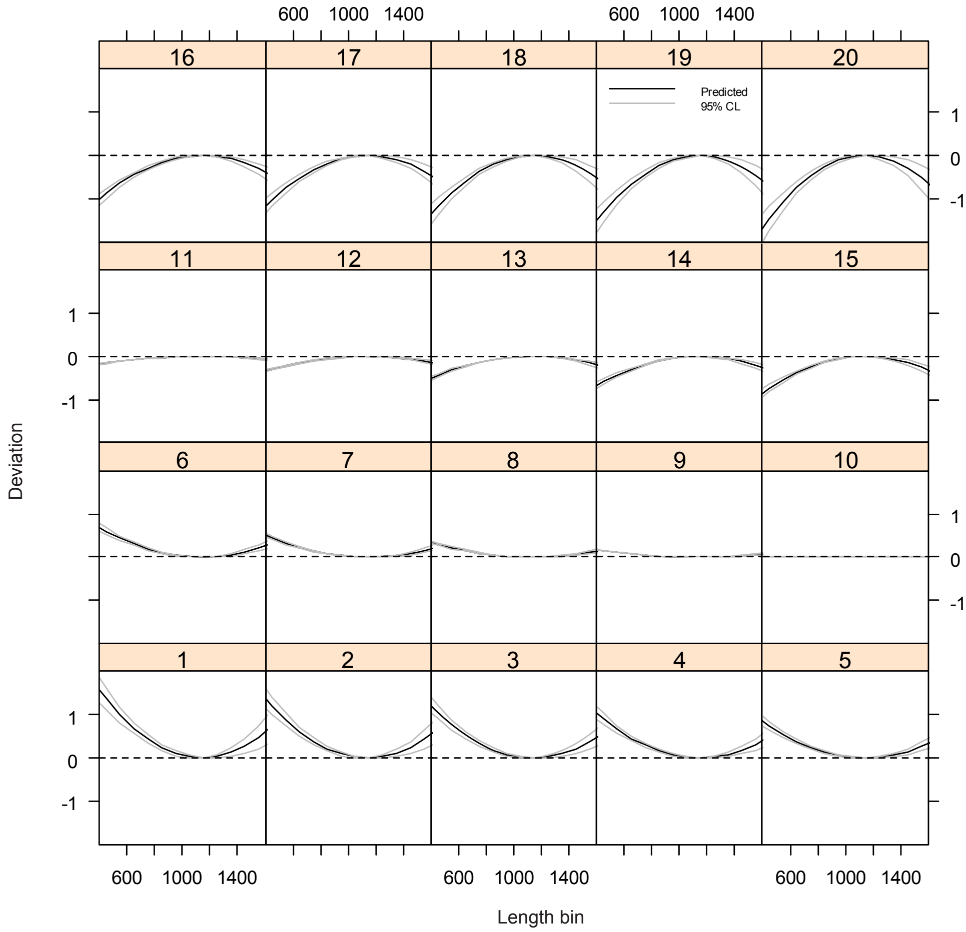

Figure 5: Predicted SLOF mean trends as deviations from an age-structured model assuming no year-to-year variation in selectivity for simulation Model 3 (Model 1 plus systematic variation in selectivity) and one replicate of a 20-year time series of catch-at-length data.

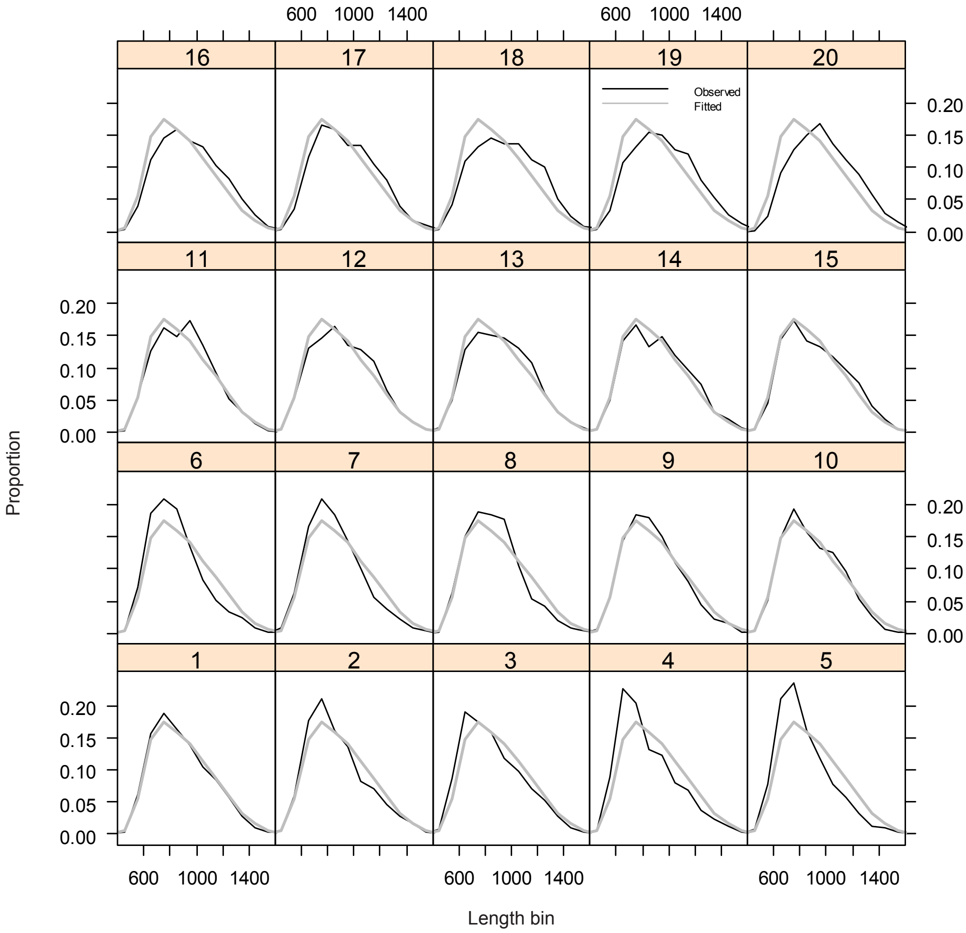

Figure 6: Observed and predicted proportions, by length bins, with predictions from an age-structured model assuming no year-to-year variation in recruitment or mortality for simulation Model 5 (Model 1 plus systematic year-to-year variation in recruitment and mortality) and one replicate of a 20-year time series of catch-at-length data.

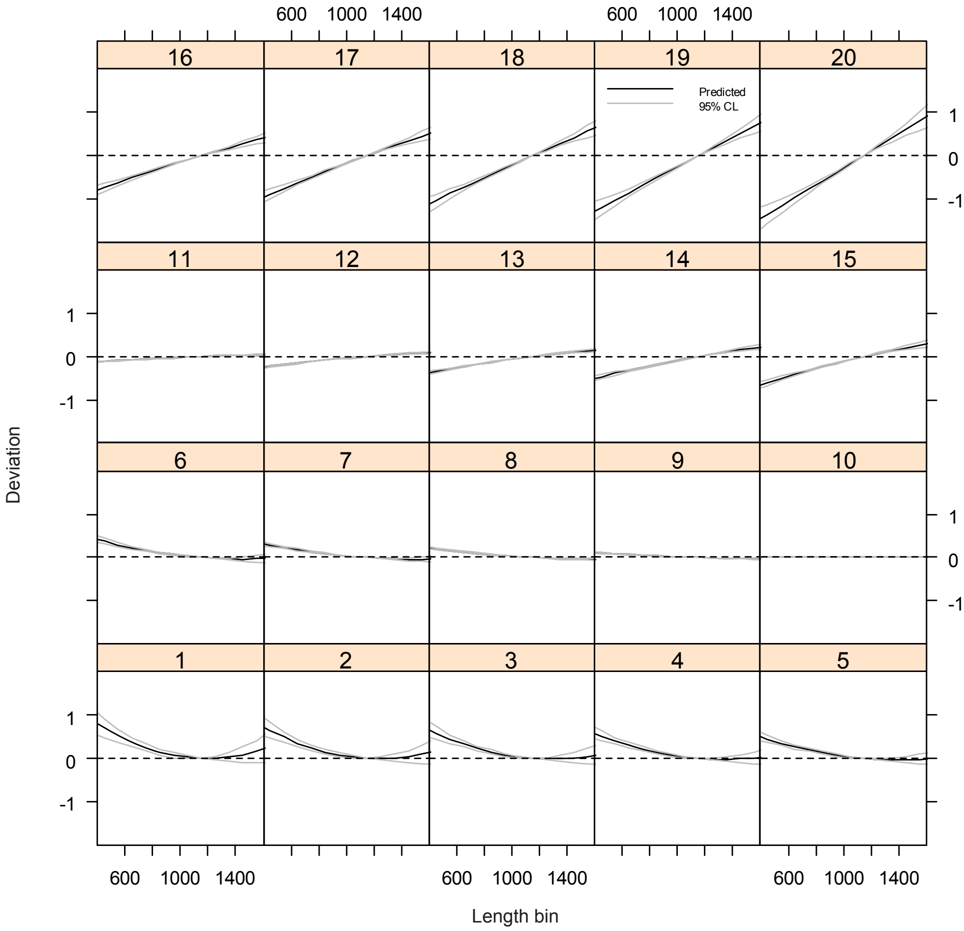

Figure 7: Predicted SLOF mean trends as deviations from an age-structured model assuming no year-to-year variation in recruitment or mortality for simulation Model 5 (Model 1 plus systematic year-to-year variation in recruitment and mortality) and one replicate of a 20-year time series of catch-at-length data.

Liste des tableaux

Liste des figures

Список таблиц

Список рисунков

Lista de las tablas

Lista de las figuras

Figura 4: Proporciones observadas y previstas, por intervalo de talla, con las predicciones efectuadas a partir de un modelo estructurado por edades suponiendo que no hay variación interanual en la selectividad para la simulación con el modelo 3 (Modelo 1 y variación sistemática de la selectividad) y una repetición de una serie cronológica de datos de captura y esfuerzo de 20 años.

Figura 5: Tendencias del promedio previstas con SLOF como desviaciones de un modelo estructurado por edades suponiendo que no existe variación interanual de la selectividad para la simulación con el modelo 3 (Modelo 1 y variación sistemática de la selectividad) y una repetición de una serie cronológica de datos de captura y esfuerzo de 20 años.

Figura 6: Proporciones observadas y previstas, por intervalo de talla, con predicciones efectuadas a partir de un modelo estructurado por edades suponiendo que no hay variación interanual en el reclutamiento o mortalidad para la simulación con el modelo 5 (Modelo 1 y variación sistemática en el reclutamiento y la mortalidad) y una repetición de una serie cronológica de datos de captura y esfuerzo de 20 años.

Figura 7: Tendencias del promedio previstas con SLOF como desviaciones de un modelo estructurado por edades suponiendo que no existe variación interanual del reclutamiento o mortalidad para la simulación con el modelo 5 (Modelo 1 y variación sistemática en el reclutamiento y la mortalidad) y una repetición de una serie cronológica de datos de captura y esfuerzo de 20 años.