Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution

James T. Thorson[1], Kelli F. Johnson[2], Richard D. Methot[3], Ian G. Taylor[1]

[1]Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

[2]School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

[3]NOAA Senior Scientist for Stock Assessments, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

Abstract

Theoretical considerations and applied examples suggest that stock assessments are highly sensitive to the weighting of different data sources whenever data sources conflict regarding parameter estimates. Previous iterative reweighting approaches to weighting compositional data are generally ad hoc, do not propagate uncertainty about data-weighting when calculating uncertainty intervals, and often are not re-adjusted when conducting sensitivity or retrospective analyses. We therefore incorporate the Dirichlet-multinomial distribution into Stock Synthesis, and propose it as a model-based method for estimating effective sample size. This distribution incorporates one additional parameter per fleet (with the option of mirroring its value among fleets), and we show that this parameter governs the ratio of nominal ("input") and effective ("output") sample size. We demonstrate this approach using data for Pacific hake, where the Dirichlet-multinomial distribution and an iterative reweighting approach previously developed by McAllister and Ianelli (1997) give similar results. We also use simulation testing to explore the estimation properties of this new estimator, and show that it provides approximately unbiased estimates of variance inflation when compositional samples capture clusters of individuals with similar ages/lengths. We conclude by recommending further research to develop computationally efficient estimators of effective sample size that are based on alternative, *a priori* consideration of sampling theory and population biology.

Keywords: data weighting; Dirichlet-multinomial; integrated stock assessment model; multinomial; statistical catch-at-age; overdispersion; length composition; age composition;

1. Introduction

Stock assessment models are quantitative tools that are used to provide a scientific basis for the management of marine fishes (Walters and Martell, 2004). Assessment models increasingly incorporate biological assumptions regarding the population dynamics of fished species, and population dynamics parameters are estimated by fitting the assessment model to available data (Maunder and Punt, 2013). Fitting population models to available data is typically done using likelihood-based statistics, and the proper estimation of confidence and forecast intervals therefore generally requires accounting for heteroskedastic and correlated residuals as caused by unmodeled biological or measurement process (Thorson and Minto, 2015). Theoretical considerations and applied examples suggest that integrated statistical stock assessments are sensitive to the weighting of different data sources whenever sources conflict regarding parameter estimates. Consequently, estimates of stock status and productivity are often highly dependent upon the weighting of different data sources (Francis, 2011).

Stock assessment models frequently are fitted to sampling data that are informative about the proportion of the vulnerable population belonging to different observable categories. Common categories include the proportion of survey or fishery catch that is associated with different ages, lengths, and/or sexes. Most often, compositional sampling is assumed to follow a multinomial distribution, e.g., drawing 10 marbles with replacement from an urn that contains 15 red, 45 blue, and 40 green marbles. The multinomial distribution is derived from the assumption that a given compositional sample represents independent sampling with replacement from a fixed and known number of individuals (i.e., 10 marbles), where each individual is from one of several possible categories, and where there is a true “fixed” probability p_c associated with each category c (i.e., p_c=0.15, 0.45, and 0.40 for red, blue, and green marbles). Each sample will not perfectly represent the true distribution, e.g., a single sample of 10 marbles might yield 1 red, 4 blue, and 5 green (i.e., where the observed proportion is 0.1, 0.4, and 0.5), and another sample might yield 2 red, 3 blue, and 5 green (an observed proportion of 0.2, 0.3, and 0.5). The multinomial distribution implies that the sampling variance (i.e., variation if the sampling process was replicated) is a function of both the true probability and sample size, $Var(p_{obs+} = p(1-p \ /n$, where n is the number sampled and p is the true probability for each category. Thus, as n increases, the coefficient of variation for the proportion in each category decreases by $1/\sqrt{n}$.

In practice, compositional data for fish populations arises from a process of sampling fish (e.g., non-extractive visual samples or by capturing and measuring fishes), and this sampling process is more complicated than the process implied by a multinomial distribution. In particular, compositional data are likely to have greater variance than predicted by a multinomial distribution based on the number of individual fish that are sampled (termed “overdispersion”). In general, overdispersion arises whenever individuals within a sample are not statistically independent. This assumption of statistical independence (i.e., underlying the multinomial distribution) is often violated, e.g., when fish schooling behavior leads to a single age being over-represented in each individual sample (McAllister and Ianelli, 1997), or when juvenile or adult fish have an affinity for a particular depth range leading to proportions that vary spatially (Kristensen et al., 2014) and between sampling tows (Crone and Sampson, 1997). In practice, compositional data are processed to transform raw compositional sampling data into an aggregated estimate of the proportion in each category in a given year for the entire modeled population. The resulting estimates of the proportion in each category for each year is sometimes termed “expanded compositional data” when the process uses a simple design-based estimator, whereas we prefer the term “standardized compositional data” in recognition that the process sometimes involves complicated statistical methods to estimate input sample sizes or account for missing data (Shelton et al., 2012; Thorson, 2014). Compositional standardization results in an estimate of "input" sample size for the compositional data in a given year, where estimates of input sample size are frequently a function of both (i) the number of tows and (ii) the total number of sampled fish (Crone and Sampson, 1997; Stewart and Hamel, 2014). Compositional standardization can also estimate the covariance among categories (e.g., Miller and Skalski, 2006), although this is not always done.

The multinomial distribution is often used in the likelihood function that is maximized to estimate parameters in an integrated assessment model. In this usage, the multinomial distribution is used to approximate the probability that the standardized proportions in each category arose from the fish population given proposed values for estimated parameters. We define the "input sample size" as the sample size calculated during compositional standardization (or assumed at a fixed value *a priori*), and this input sample size is often used when evaluating the multinomial likelihood of estimated parameters. In this usage, input sample size controls the weighting of compositional data relative to other data sources included in the likelihood function. However, model misspecification may cause this input sample size to be an inappropriate measure of data weighting. As a thought experiment, imagine that all participants in a fishery falsify fish sizes in their catch. These data would have no information about the size-composition of the population, and a stock assessment model would have optimal performance if it assigned zero weight to these data. As a less extreme example, age-composition data are often obtained by laboratory examination of fish samples (otoliths or spines), and these laboratory methods sometimes mis-identify the age of a given fish. Ageing error will cause age-composition data to be a blurred measure of the true age-composition such that age-composition data are less informative than if ageing error were absent (Coggins and Quinn, 1998). However, if the stock assessment model incorporates double-reading and ageing-error methods to correct for the ageing error (Methot and Wetzel, 2013; Punt et al., 2008), these data might be more informative about population age structure.

The previous example highlights that the optimal weight of composition data depends upon the specification of the model, where model misspecification (e.g., neglecting the impact of ageing error) results in a lower optimal weight for available compositional data. This conclusion implies that compositional weighting can be informed by inspecting the goodness-of-fit between the compositional data and estimated proportions from the assessment model, and consequently decreasing the sample size for data that generally do not match. This process was suggested by McAllister and Ianelli (1997), who proposed iteratively estimating the "effective sample size" for compositional data from a given fleet via the match between predicted and observed compositional data. However, iterative reweighting approaches require the following steps: (1) fit the assessment model to available data; (2) extract estimates of compositional proportions; (3) calculate the effective sample size; (4) input the new effective sample size; (5) iterate steps 1-4 a fixed number of times, or until subsequent iterations cause little change in the estimate of effective sample size. Decreasing the effective sample size has an identical impact to multiplying the multinomial likelihood function by the same percent change (Francis, 2011), such that this process is essentially reweighting the compositional data during each iteration of the algorithm. This iterative-reweighting algorithm has several draw-backs, including that it is infeasible to repeat for every sensitivity run, it is difficult to explore when parameter estimation is slow (e.g., when using Bayesian estimation via Markov-chain Monte Carlo), it is difficult to incorporate into simulation designs, it is potentially influential when estimating likelihood profiles for stock assessment parameters, and it does not propagate uncertainty about data weighting into estimates of parameter uncertainty.

In the following, we seek to develop a method to estimate effective sample size during parameter estimation. If this were done by estimating a new parameter that governs the ratio of input and effective sample size, then uncertainty about the data-weighting parameter could be estimated using conventional methods (Magnusson et al., 2013), and its uncertainty could be propagated and evaluated during stock projections. We therefore specifically seek a method to estimate effective sample size as a model parameter. For this purpose, we implement the Dirichlet-multinomial distribution for compositional data in the likelihood function of an integrated assessment model. We show that using the Dirichlet-multinomial distribution involves estimating a new parameter, and can be parameterized such that it estimates a linear relationship between input and effective sample size. We incorporate this new distribution into the Stock Synthesis stock assessment software, which is widely used in the United States and internationally (Methot and Wetzel, 2013). The Dirichlet-multinomial is now available as a feature in Stock Synthesis when calculating the probability of age- or length-composition samples from the entire population ("marginal" age- or length-composition data), the probability of age-composition samples from a given length category ("conditional age-at-length data"), or the probability of length-composition samples from a given age-category ("conditional length-at-age data"). We then use a case study and simulation experiment to show that the Dirichlet-multinomial distribution provides estimates of effective sample size that are similar to iterative reweighting methods, but without requiring multiple iterations of running the assessment model.

2. Methods

2.1 Introducing the Dirichlet-multinomial distribution

Many stock assessment models use the multinomial distribution for fitting compositional data while calculating the likelihood of model parameters:

$$L(\boldsymbol{\pi}|\tilde{\boldsymbol{\pi}}, n+= \text{Multinomial}(\tilde{\boldsymbol{\pi}}|\boldsymbol{\pi}, n+= \frac{\Gamma(n+1+}{\prod_{a=1+}^{a_{max}} \Gamma(n\tilde{\pi}_{a+}\,1} \prod_{a=1+}^{a_{max}+} \pi_a{}^{n\tilde{\pi}_a} \qquad (1)$$

where $\tilde{\boldsymbol{\pi}}$ is the proportion at age in the available data such that $\sum_{a=1+}^{a_{max}+} \tilde{\pi}_{a+} = 1$ (we use vector-matrix notation where vectors are bold, while elements of a vector are italicized with a subscript), $\boldsymbol{\pi}$ is the estimated proportion at age (such that $\sum_{a=1}^{A} \pi_a = 1$), *n* is the total number of samples in the available data (which is restricted to any non-negative real number), a_{max} is the maximum age in available data, and $\text{Multinomial}(\tilde{\boldsymbol{\pi}}|\boldsymbol{\pi}, n+$is defined as the multinomial probability mass function (we present theory using notation for age-composition data, but note that the theory is applicable to length-composition data as well). However, using the multinomial distribution for compositional data involves the assumption that the true proportion at age $\boldsymbol{\pi}$ is constant for all age-composition samples, but schooling or spatial behaviors may in fact cause the "true" age-composition (i.e., its average if the sample was replicated at that place and time) to vary among samples. Variability in a proportion can be approximated using a Dirichlet distribution:

$$p(\boldsymbol{\pi}_i|\boldsymbol{\pi}+= \text{Dirichlet}(\boldsymbol{\pi}_i|\boldsymbol{\alpha}+ \qquad (2)$$

where $\text{Dirichlet}(\boldsymbol{\alpha}+$is the probability density function for the Dirichlet distribution and $\boldsymbol{\alpha}$ is a vector of a_{max} parameters (restricted to be positive) that govern the mean and variance of this distribution. Now imagine that, for each age-composition sample, we take a random draw $\boldsymbol{\pi}^{*} \sim \text{Dirichlet}(\boldsymbol{\alpha}+$from a Dirichlet distribution, and then take a draw from a multinomial distribution $\boldsymbol{\pi} \sim \text{Multinomial}(\boldsymbol{\pi}^{*}, n+$with mean proportion $\boldsymbol{\pi}^{*}+$from the Dirichlet draw. In this case, the observed proportion $\tilde{\boldsymbol{\pi}}$ follows a compound "Dirichlet-multinomial" distribution with a probability density function:

$$p(\tilde{\boldsymbol{\pi}}|\boldsymbol{\alpha}, n+= \int \text{Multinomial}(\tilde{\boldsymbol{\pi}}|\boldsymbol{\pi}^*, n\ \ \text{Dirichlet}(\boldsymbol{\pi}^*|\boldsymbol{\alpha}\ \ d\boldsymbol{\pi}_{i+} \qquad (3)$$

where the marginal probability density function for data $\tilde{\boldsymbol{\pi}}$ is computed via integrating across the "unobservable" average proportion $\boldsymbol{\pi}^{*+}$for that sample (Thorson and Minto, 2015).

Fortunately, the likelihood function for the Dirichlet-multinomial distribution can be computed using interpretable parameters without recourse to numerical integration:

$$L(\boldsymbol{\pi}, \beta|\tilde{\boldsymbol{\pi}}, n+=+\frac{\Gamma(n+1+}{\prod_{a=1+}^{a_{max}+}\Gamma(n\tilde{\pi}_{a+}\,1+}\frac{\Gamma(\beta+}{\Gamma(n+\beta+}\prod_{a=1+}^{a_{max}+}\frac{\Gamma(n\tilde{\pi}_{a+}\,\beta\pi_{a+}}{\Gamma(\beta\pi_{a+}} \qquad (4)$$

where β is a new parameter representing the overdispersion caused by the Dirichlet distribution. Here, we use the gamma function, rather than the conventional factorial function, so that the Dirichlet-multinomial is defined for all non-negative sample sizes *n*, such that it reduces to the conventional Dirichlet-multinomial distribution whenever input sample size is a whole number. The first term $\frac{\Gamma(n+1+}{\prod_{a=1+}^{a_{max}+}\Gamma(n\tilde{\pi}_{a+}\,1+}$ does not depend upon the parameters, but ensures that the value of the Dirichlet-multinomial function $L(\boldsymbol{\pi}, \beta|\tilde{\boldsymbol{\pi}}, n+$converges on the value of the conventional multinomial function $L(\boldsymbol{\pi}|\tilde{\boldsymbol{\pi}}, n+$as $\beta \rightarrow \infty$, such that the multinomial distribution is a special case of the Dirichlet-multinomial distribution. Similar to the multinomial, the Dirichlet-multinomial likelihood can be computed even for cases with zero observations (i.e., where $\tilde{\pi}_a =+$ 0 for some *a*), and this is not true of other proposed methods to account for overdispersion (e.g., Francis, 2014).

2.2 Computing the effective sample size:

We define the effective sample size $n_{eff+}$of a distribution g for compositional data $\mathbf{c}\sim g(\boldsymbol{\pi}+$as the sample size of a multinomial distribution $\mathbf{c}^*\sim\text{Multinomial}(\boldsymbol{\pi}, n_{eff+}$ that has the same variance on average across categories (i.e., $\sum_{a=1+}^{a_{max}+}\text{Var}(c_{a+} = \sum_{a=1+}^{a_{max}+}\text{Var}(c_a^{*+})$. The variance of a single element from a multinomial distribution is:

$$\mathrm{Var}(c_a|n,\boldsymbol{\pi}+= n\pi_a(1-\pi_{a+} \qquad (5)$$

where *n* is the sample size. Defining observed proportion $\tilde{\pi}_a = c_a/n$, we see that:

$$\mathrm{Var}(\tilde{\pi}_a|n,\boldsymbol{\pi}+=\frac{\pi_a(1-\pi_{a+}}{n+} \qquad (6)$$

i.e., variance decreases as the reciprocal of sample size.

We next return to the Dirichlet distribution, $\tilde{\boldsymbol{\pi}}\sim Dirichlet(\beta\boldsymbol{\pi}+$; where $\alpha_a = \beta\pi_{a+}$and $\pi_{a+}$is the true proportion at age. The Dirichlet distribution has variance:

$$\mathrm{Var}(\tilde{\pi}_a|\beta,\boldsymbol{\pi}+=\frac{\alpha_a(\beta-\alpha_{a+}}{\beta^2(\sum_{a=1+}^{a_{max}+}\alpha_a\ \ 1}=\frac{\beta\pi_a(\beta-\beta\pi_{a+}}{\beta^2(\beta+1+}=\frac{\pi_a(1-\pi_{a+}}{\beta\ \ 1+} \qquad (7)$$

such that $\beta+1$ is the effective sample size of the Dirichlet distribution:

Finally, the variance of the observed proportion at age for a Dirichlet-multinomial distribution is:

$$\mathrm{Var}(\tilde{\pi}_a|n,\beta,\pi+=\frac{\pi_a(1-\pi_{a+}}{n+}\left(\frac{n\ \ \beta}{1+\beta}\right)+ \qquad (8)$$

such that the variance (and also the covariance) is equal to the variance (and covariance) for the multinomial distribution multiplied by $(n+\beta+/(1+\beta+$(Eq. 15-16 in Mosimann, 1962). We therefore calculate the estimated effective sample size n_{eff} of a Dirichlet-multinomial distribution as:

$$n_{eff+}=\frac{n+n\beta+}{n\ \ \beta+} \qquad (9)$$

where this formula is similar to an approximation obtained by summing the variance of the Dirichlet and multinomial distributions (i.e., the sum of multinomial sampling variance and Dirichlet-distributed overdispersion). This formula illustrates that the Dirichlet-multinomial distribution has equal overdispersion for all bins (e.g., sizes or ages). In some cases, overdispersion may vary substantially among bins (Miller and Skalski, 2006), presumably due to spatial variation in population densities associated with each bin (Kristensen et al., 2014;

Thorson, 2014), and we suggest that future research explore the impact of varying overdispersion on the performance of assessment models using the Dirichlet-multinomial likelihood.

2.3 Two potential parameterizations

Given the Dirichlet-multinomial distribution and the closed-form computation of its effective sample size, we propose two alternative parameterizations that may be useful in practice for length- and age-composition samples in stock assessment models. These parameterizations differ in terms of the function relating input and effective sample size (Fig. 1), and correspond to different hypotheses regarding the mechanisms underlying overdispersion. Both use the input sample size to distinguish among years that have relatively more or less information about the true proportion.

2.3.1 Parameterization #1 – Linear version

As a default, we recommend a re-parameterization of the Dirichlet-multinomial distribution, wherein the variance-inflation parameter β is replaced by a linear function of input sample size *n,* i.e., $\beta = \theta n$. This results in the following probability distribution function:

$$L(\boldsymbol{\pi}, \theta | \tilde{\boldsymbol{\pi}}, n+= \frac{\Gamma(n+1+}{\prod_{a=1+}^{a_{max}} \Gamma(n\tilde{\pi}_a \ \ 1} \frac{\Gamma(\theta n+}{\Gamma(n \ \ \theta n+} \prod_{a=1+}^{a_{max}+} \frac{\Gamma(n\tilde{\pi}_{a+} \theta N \pi_{a+}}{\Gamma(\theta n \pi_{a+}} \qquad (10)$$

which has effective sample size:

$$n_{eff+}= \frac{1+\theta n+}{1 \ \ \theta+} = + \frac{1+}{1+\theta+} + n + \frac{\theta+}{1 \ \ \theta+} \qquad (11)$$

where we see that effective sample size is a linear function of input sample size with intercept $(1 + \theta +^{-1+}$and slope $\theta(1 + \theta +^{-1+}$. If θ becomes large $(\theta \gg n+$then $n_{eff+} \rightarrow n$ such that there is no variance inflation in this case, and if θ is small $(\theta \ll n+$while n is large $(n \gg 1)$ then θ is approximately the ratio of effective and input sample size $(\theta \rightarrow n_{eff}/n)$. We recommend using the "linear effective sample size" parameterization, given that previous methods for weighting compositional data have generally multiplied the likelihood of compositional data by a fixed quantity λ<*1* (Francis 2011), and this parameterization has similar behavior when sample sizes are high and samples are strongly overdispersed ($n \gg 1$ and $\theta \ll n$).

2.3.2 Parameterization #2 – Saturating version

As a potential alternative, analysts may instead use the original parameterization of the Dirichlet-multinomial distribution (Eq. 4), which has effective sample size:

$$n_{eff+} = +\frac{n+n\beta+}{n\quad\beta+} \tag{12}$$

This parameterization can revert to the multinomial distribution with sufficiently large β, i.e., $n_{eff+} = n$ when $\beta \gg n$. However, it provides an upper bound on effective sample size with lower values of $\hat{\beta}$, i.e., $n_{eff+} \rightarrow 1 + \beta$ when $n \gg \beta$. Therefore, this parameterization could be useful when analysts seek to estimate an upper bound on the effective sample size for a given year.

We have implemented both parameterizations of the Dirichlet-multinomial distribution in Stock Synthesis (version 3.30; public release planned for Aug 2016, and please contact Richard.Methot@noaa.gov for a beta version). In the following, we focus exclusively on the linear parameterization (version #1). However, we recommend future research comparing the performance of these two parameterizations using real-world data, and developing more-complicated two-parameter forms for the Dirichlet-multinomial distribution that could combine the characteristics of both versions. In particular, the saturating parameterization resembles an "additive" influence of process errors while the linear parameterization is more similar to the "multiplicative" influence of process errors (Francis, this issue), and we hypothesize that a two-parameter form could be used to distinguish between additive and multiplicative forms for process error. In the following, we also restrict ourselves to the case where the variance-inflation parameter is constant for all years, but note that future studies can estimate different levels of variance inflation for each year, or for different blocks of years.

2.4 Case study: Pacific hake

To demonstrate this new data-weighting method, we compare its performance with that of other data-weighting methods when applied to a recent stock assessment for Pacific hake, *Merluccius productus* (Taylor et al., 2015). Pacific hake is a semi-pelagic schooling species of commercial importance to fisheries off of the US West Coast and Western Canada. Recent management is conducted following procedures determined by an international agreement between the United States and Canada, and are informed by annual stock assessments implemented using Stock Synthesis. Data used in the 2015 stock assessment includes (1) catches from 1966 to 2014, (2) fishery age–composition samples from 1975-2014, (3) an index of abundance from ten acoustic surveys conducted between 1995 and 2013, (4) survey age-composition samples associated with each acoustic survey, (5) cohort-specific definitions of ageing error that specify improved ageing accuracy with larger cohorts, and (6) "empirical" weight-at-age data calculated from all fisheries and the acoustic survey for years 1975 to 2014, which are assumed to be known without error (Taylor et al., 2015).

Four assessment models were fitted to data for Pacific hake, where each model used a different approach to data-weighting for the fishery age-composition data: (i) unweighted (i.e., treating input sample size as effective sample size), (ii) tuned using an iterative approach, (iii) estimated using the Dirichlet-multinomial distribution, and (iv) weight of zero. Option (ii) is the approach commonly used in West Coast assessments, including the Pacific hake assessment (Taylor et al., 2015), and involved fitting the model to available data, computing the ratio of the harmonic mean of yearly effective sample size (as computed by Stock Synthesis) to the arithmetic mean of yearly input sample size for fishery age-composition data, multiplying this value by the "weighting factor" for the fishery age-composition data used during parameter estimation, and then inputing this value as the new weighting factor. We use the harmonic mean of effective sample sizes, rather than the arithmetic mean, following recent research (Punt, In press) and common practice for West Coast assessments (e.g., Taylor et al., 2015). This process was repeated two times and the third fit to data was used as the final estimate of parameters. The initial weighting factor was set to one and all additional weighting factors had an upper bound of one to ensure that effective sample size was never greater than the original input sample size. In the following, we refer to this as the McAllister-Ianelli iterative-reweighting method, although we note that this algorithm has evolved since its original version in McAllister and Ianelli (1997). Option (iv) specifies that the stock assessment was fitted only to abundance indices and survey age-composition data, and represents the extreme case of "zero" weight assigned to fishery compositional data. To achieve convergence in this option, we turned off parameters representing variation in fishery selectivity over time, and fixed parameters representing average fishery selectivity at their estimates from Option (ii). Fishery compositional data are the only source of information regarding age-structure prior to 1975, so we assume that this option will result in large differences in estimates during early years. Preliminary exploration showed that the input sample size is approximately equal to effective sample size for survey age-composition data (i.e., the iterative approach results in a ratio of 0.94, and the Dirichlet-multinomial results in a ratio approaching 1.00, i.e., θ increases indefinitely). We therefore chose to not re-weight the survey age-composition data (i.e., we did not estimate the Dirichlet-multinomial parameter for the survey age-composition data, nor did we tune them). We inspected model fit for the fishery age-composition samples using Pearson residuals:

$$r_{a,t+}=+\frac{\tilde{\pi}_{a,t}-\pi_{a,t+}}{\sqrt{\frac{\pi_{a,t}(1-\pi_{a,t+}}{n_{eff,t+}}}} \qquad (13)$$

where $r_{a,t+}$is the Pearson residual for age *a* and year *t*, $\tilde{\pi}_{a,t+}$is the proportion in the observed data for that age and year, $\pi_{a,t+}$is the expected proportion, and $n_{eff,t+}= (1+n_t\theta+/(1+\theta+$is the estimate of effective sample size using the linear parameterization where $n_{t+}$is the input sample size for year *t*. We expect that a well-fitted model will have (1) no consistent patterns in residuals for consecutive ages in a given year, (2) no pattern in residuals for consecutive years for a given age, and (3) no pattern in residuals among fleets.

2.5 Simulation testing

The performance of the Dirichlet-multinomial distribution implemented in Stock Synthesis was explored using simulated data. To do so, we simplified the Pacific hake estimation model in five ways: (1) changed fishery selectivity to be stationary over time (i.e., removed time-varying selectivity parameters), (2) changed all fishery age-composition sample sizes to a single fixed value per year, (3) changed all survey age-composition sample sizes to 100 samples per year, (4) changed age-specific ageing error to be stationary over time and equal to the baseline ageing-error matrix, and (5) changed to using an "explicit-F" parameterization, wherein instantaneous, fully-selected fishing mortality in each year is estimated as a fixed effect. We made changes (1) and (4) because fishery selectivity and ageing error in the original assessment are related to realized cohort size, and our simulation is randomly generating new time series of relative cohort size. We made change (5) so that the simulated fishing intensity is plausible given the simulated vector of recruitment deviations for each simulation replicate, and changes (2) and (3) to simplify interpretation of results (e.g., so that time series estimates are not influenced by annual variation in sample sizes). We then ran the modified Pacific hake assessment model on available data, extracted estimated parameters, and used these estimates as the “true” values during the simulation experiment (while confirming that estimated stock status and productivity was generally similar to that in the case study).

We then generated new, simulated data sets using the Stock Synthesis parametric bootstrap simulator. For each simulation replicate, we simulated a new vector of recruitment deviations with a standard deviation of recruitment deviations (σ_R) set at 0.9, and also simulate a new deterministic pattern for fishing mortality, where instantaneous fishing mortality F for fully-selected ages increases linearly from $F = 0.01$ in the first year (1966) to $F = 0.30$ in the final year (2013). The bootstrap simulator then calculated the population abundance-at-age resulting from the input vector of recruitment deviations and fishing mortality, and simulates an abundance index and age-composition samples from their specified distributions (i.e., using a lognormal distribution with the input log-standard deviation for the abundance index and a multinomial distribution with the input sample size for the age-composition samples).

The simulation experiment involves a factorial design with three simulation scenarios, five levels of an inflation factor, and three estimation models. For each combination, we ran 100 simulation replicates, for a total of $3 \times 5 \times 3 \times 100 = 4{,}500$ total estimation model runs. We define three simulation scenarios, where we generate age-composition samples $\mathbf{c}_t$ in each year t from a multinomial distribution i.e., $\mathbf{c}_\mathrm{t} \sim \mathrm{Multinomial}(\boldsymbol{\pi}, n_{true+}$, and where the “true” sample size varies among scenarios (n_{true} = 25, 100, or 400). Given this age-composition sample, we then provide the estimation model with an input sample size of $n_{input+}= \theta_{sim} n_{true}$, such that the “observed” age-composition sample is inflated by inflation factor θ_{sim}, with value $\theta_{sim+}=+ \{1,2,5,25,100\}$. We then use estimation methods (i), (ii), and (iii) defined in the section titled *Case study: Pacfic hake* (see above).

2.6 Simulation model evaluation

Estimation procedures were evaluated by comparing estimated parameters and derived quantities of interest to management to their true values as defined in the operating model. Estimation error was quantified using relative error ($RE = \left(\hat{P} - P\right)/P$, where $\hat{P}$ and P are estimated and true parameter values respectively). Results were recorded for converged models, where convergence was defined as obtaining a gradient less than 0.1, and we also record the proportion of non-convergence for each estimation model and simulation scenario.

3. Results

3.1 Case study application: Pacific hake

Comparing four alternative methods for weighting compositional data in the Pacific hake assessment (Fig. 2) shows that estimates of relative spawning output and fishing intensity are generally bracketed by the two naïve approaches, i.e., either treating input sample size as effective sample size ("unweighted") or removing fishery age-composition data entirely ("no fishery ages"). However, spawning output is higher for the tuned and Dirichlet-multinomial models than the unweighted model because the unweighted model estimates lower unfished recruitment. In particular, removing fishery age data results in a higher estimate of average unfished spawning output and lower spawning output estimates from the mid-1980s onward, as well as large differences in abundance trends prior to 1975. Meanwhile treating input sample size as the effective sample size results in estimates of strong year-class strength in 1980 and 1999. By contrast, the default iterative and new Dirichlet-multinomial weighting methods result in similar estimates of spawning output, with the exception of early years (prior to 1980) when the Dirichlet-multinomial estimator results in somewhat elevated estimates of spawning output relative to the iterative method. Similarly, the iterative and Dirichlet-multinomial estimates of fishing intensity are more similar than the other weighting methods, particularly for early years (prior to 1970). Inspection of Pearson residuals when using the Dirichlet-multinomial likelihood to estimate overdispersion (Fig. 3) shows little evidence for correlated residuals among ages within a year, among years within an age, or among fleets (except perhaps for the negative residual for individuals in the oldest age category). However, cohorts born during 1977, 1980, and 1984 generally have small, positive residuals. This pattern arises because the recruitment penalty (i.e., penalizing recruitment deviations towards zero) encourages less variation in cohort strength than the age-composition data suggest for these years.

3.2 Simulation experiment

Estimates of the Dirichlet-multinomial parameter are different among the different scenarios and levels of the inflation factor (Fig. 4, panel a). However, estimates of effective sample size are generally similar for all levels of the inflation factor for a given scenario (Fig. 4, panel b). In general, the estimated effective sample size closely matches the true sample size for all scenarios and levels of the inflation factor. However, we detect a small positive bias in the estimates of effective sample size when the true sample size is 400 (i.e., median effective sample size estimate is close to 450), and a negative bias when true sample size is 25 and variance inflation is high (θ_{sim}>25).

Comparison of parameter estimates from the unweighted multinomial, iterative reweighting algorithm, and the linear parameterization of the Dirichlet-multinomial distribution shows that the iterative reweighting and Dirichlet-multinomial approaches have similar precision and accuracy when estimating natural mortality and average unfished recruitment for all levels of the inflation factor (Fig. 5). By contrast, the unweighted model has substantially degraded estimates of natural mortality and unfished recruitment for any inflation factor other than 1. We note that the Dirichlet-multinomial algorithm has a small fraction (2 of 100) of replicates that do not converge for some levels of the variance inflation (θ_{sim}=100, see Fig. 5). We therefore conclude that the Dirichlet-multinomial method has similar estimation performance to the previous iterative reweighting approach.

4. Discussion

In this study, we implemented two parameterizations of the Dirichlet-multinomial distribution in the Stock Synthesis software that is widely used to conduct stock assessments in the US and internationally. We then compared the Dirichlet-multinomial distribution with a version of the McAllister-Ianelli iterative-reweighting approach that is commonly used for US West Coast groundfish stock assessments. We believe that the Dirichlet-multinomial approach is superior to this iterative-reweighting approach for several reasons.

1. *Slow or inconsistent exploration of alternative models*: Iterative reweighting methods require fitting a stock assessment model to data to calculate effective sample sizes, and then re-estimating the model with revised input sample sizes. This iterative tuning procedure either slows exploration of alternative models (due to the need for re-tuning after each model change) or causes inconsistent exploration of alternative models (where analysts neglect to re-tune for every sensitivity run, and therefore compare between runs that are not tuned in a consistent manner).
2. *Failure to account for uncertainty in data weighting*: Iterative reweighting methods provide no obvious method for propagating uncertainty about data-weighting. By contrast, the Dirichlet-multinomial approach represents data-weighting via an estimated parameter, and the uncertainty in this parameter can be captured via standard statistical methods (e.g.,

likelihood profiles, asymptotic confidence intervals, or Bayesian posteriors, (Magnusson et al., 2013)).

3. *Clear standards for convergence*: Iterative reweighting methods require subjective decisions regarding when to stop tuning the sample size, what order to tune multiple fleets, and how to combine data-weighting information from multiple fleets. These subjective decisions are rarely documented and different decisions by different analysts may cause substantial differences in ultimate estimates of stock status and productivity in assessments where data weighting is an important axis of uncertainty (e.g., US West Coast sablefish). By contrast, the Dirichlet-multinomial method allows for a single, unambiguous definition of convergence (i.e., via maximizing the model likelihood function), which can be independently replicated by different authors and does not require further documentation. If estimates of the parameter governing effective sample size using the Dirichlet-multinomial likelihood do not converge, we suggest that the analyst could perform one model run using the iterative reweighting approach (to get an initial value for the Dirichlet-multinomial parameter), and then proceed to fully estimate that parameter in a final model run.
4. *Interpretable estimates of effective sample size*: Analysts have previously suggested alternative model-based methods for estimating effective sample size. For example, an analyst might use a Dirichlet distribution, which performed relatively well in previous simulation testing (Hulson et al., 2011; Maunder, 2011), rather than the Dirichlet-multinomial distribution used here. However, the Dirichlet distribution can have effective sample size that ranges from 0 to infinity, i.e., it can exceed the input sample size (Hulson et al., 2011; Maunder, 2011; Schnute and Haigh, 2007). By contrast, the Dirichlet-multinomial distribution ensures that the effective sample size can never be greater than the input sample

size. We believe that restricting the effective sample size to be less than or equal to input sample size is useful when analysts have properly estimated the variance of standardized compositional data (Stewart and Hamel, 2014; Thorson, 2014), as we and others have recommended in general. When analysts have not estimated the input sample sizes for standardized compositional data, the Dirichlet distribution might be a suitable approach for estimating an effective sample size greater than the input sample size. We hypothesize that the Dirichlet distribution will be less numerically stable that the Dirichlet-multinomial distribution (see e.g., Maunder, 2011), because the Dirichlet distribution may lead to model estimates with implausible high weight for compositional data.

These benefits of the Dirichlet-multinomial distribution relative to iterative reweighting approaches should facilitate the development, exploration, testing, and review of stock assessment models in real-world applications.

The Dirichlet-multinomial distribution assumes a fixed, negative correlation in residuals among categories in a given year and fleet. Residuals in real-world assessments might have a more complicated pattern of correlation for two general reasons:

1. *Covariation in sampling data* – Many circumstances may cause individual samples of compositional data in natural populations to represent a disproportionately large number of juvenile or adult fishes. For example, when fishes aggregate in groups with similar age or size the age of each individual from that school will be highly correlated. This correlation also occurs when fishes partition available habitat by size or age, such that each sample will occur in a habitat preferred by a particular age or size category. Correlations among size or age measurements for each sample will cause the standardized estimate of proportions by category (inputted as data into assessment models) to also be correlated. This covariation

can be estimated by proper analysis of raw compositional data (Hrafnkelsson and Stefánsson, 2004; Miller and Skalski, 2006).

2. *Model mis-specification* – Alternatively, model residuals (i.e., the difference between compositional data and model predictions of proportions for each category) may be correlated among categories when the population dynamics model is mis-specified (e.g., by assuming the wrong value for natural mortality rate, or not accounting for error in reading fish otoliths Maunder (2011)). Unmodeled processes (e.g., spatial variation in fishing intensity) will generally result in residuals for compositional data that are correlated among categories (e.g., between age-1 and age-2 samples in a given year), years (e.g., between adjacent years for age-2 individuals), sexes (between males, females, and unsexed individuals for a given age and year), and fleets (between survey and fishery compositional data for a given age and year). For example, positive correlations among years for a given age are likely to arise whenever unmodeled processes have a similar effect on individuals of that age. Potential causes of correlated residuals for compositional data include time-varying or non-parametric fishery selectivity, time-varying growth, and time-varying rates of natural mortality.

We acknowledge that covariation arising from the process of sampling compositional data (mechanism #1 listed above) is not adequately captured by the Dirichlet-multinomial likelihood function, and that alternative functions have been developed to simultaneously model correlations and overdispersion in compositional data. One example is the logistic-normal function, which Francis (2014) proposed as a general replacement for the multinomial distribution. However, Francis (2014) only explored correlations among categories (inter-class correlation), and did not attempt to account for correlations in a given category among years or fleets. We therefore encourage further research regarding likelihood functions that can use information regarding correlations caused by sampling while still estimating a reduction in effective sample size (to account for model mis-specification).

We hypothesize that correlations arising from model mis-specification (mechanism #2 listed above) will generally include correlations among fleets, ages, years, and sexes, and are best dealt with by using adding random effects to account for important forms of model mis-specification. Mixed-effects estimation is useful to elicit the correlation among data that is induced by unobserved processes (Thorson and Minto, 2015); therefore, mixed effects are a natural tool for modeling correlations in compositional data that are caused by model mis-specification. Mixed-effect methods have already been developed for time-varying selectivity, natural mortality, and individual growth, and are increasingly feasible for age-structured population models using maximum likelihood or Bayesian estimation methods (Kristensen et al., 2014; Mäntyniemi et al., 2013; Nielsen and Berg, 2014; Thorson et al., 2015). We therefore recommend future research to explore whether accounting for these processes can adequately approximate the correlations in model residuals for compositional data, or whether it is also necessary to explicitly incorporate covariation caused by sampling.

As with any new method, we also encourage simulation testing using a variety of operating models, forms of model mis-specification, and harvest control rules (Hulson et al., 2011; Maunder, 2011; Punt, In press). Different forms of spatial structure or cohort-specific selectivity will generally result in different forms of correlation among years, categories, fleets, and sexes, and therefore will likely result in better or worse performance of the Dirichlet-multinomial distribution (given its inability to account for correlated residuals). We hope that future studies comparing the performance of the Dirichlet-multinomial likelihood relative to generalized likelihood functions that account for among-bin correlation (e.g., Francis, 2011) will include a variety of forms of model misspecification. Until these studies are conducted, we do not believe there is sufficient evidence to have a strong opinion regarding the full trade-off between either (1) modeling correlations via time-varying biological and fishery parameters vs. (2) modeling correlations via a generalized likelihood function.

5. Conclusions

In this paper, we have shown that the Dirichlet-multinomial distribution can be used to generate model-based estimates of effective sample size for age- and length-compositional data in stock assessment models. Using a real-world stock assessment for Pacific hake, we showed that the Dirichlet-multinomial distribution provides similar estimates of effective sample size to the McAllister-Ianelli approach to iterative reweighting using the harmonic mean. We also provide a simulation experiment to verify that it provides approximately unbiased estimates of effective sample size given that the model is otherwise specified correctly. We conclude that the Dirichlet-multinomial distribution is a reasonable method to estimate the magnitude of overdispersion in compositional data, and recommend future research combining it with mixed-effects estimates of time-varying selectivity and individual growth to account for correlated residuals among categories, years, and fleets.

Acknowledgements

This publication was partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. XXX, Contribution XXX. We thank Allan Hicks, Jim Hastie, Chris Francis, and an anonymous reviewer for comments on an earlier draft.

References

Coggins, L.G., Quinn, T.J., 1998. A simulation study of the effects of aging error and sample size on sustained yield estimates. Fish. Stock Assess. Models 955–975.

Crone, P.R., Sampson, D.B., 1997. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition., in: Int. Symp. on Fishery Stock Assessment Models for the 21st Century, Anchorage, Alaska, EEUU. 8Á11 Oct 1997.

Francis, R.C., 2014. Replacing the multinomial in stock assessment models: A first step. Fish. Res. 151, 70–84.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.

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

Hulson, P.J.F., Hanselman, D.H., Quinn, T.J., 2011. Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ratio and likelihood methods. ICES J. Mar. Sci. J. Cons. 68, 1548–1557.

Kristensen, K., Thygesen, U.H., Andersen, K.H., Beyer, J.E., 2014. Estimating spatio-temporal dynamics of size-structured populations. Can. J. Fish. Aquat. Sci. 71, 326–336. doi:10.1139/cjfas-2013-0151

Magnusson, A., Punt, A.E., Hilborn, R., 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish Fish. 14, 325–342.

Mäntyniemi, S., Uusitalo, L., Peltonen, H., Haapasaari, P., Kuikka, S., 2013. Integrated, age-structured, length-based stock assessment model with uncertain process variances, structural uncertainty, and environmental covariates: case of Central Baltic herring. Can. J. Fish. Aquat. Sci. 70, 1317–1326. doi:10.1139/cjfas-2012-0315

Maunder, M.N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. Fish. Res. 109, 311–319.

Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.

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

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99.

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

Mosimann, J.E., 1962. On the Compound Multinomial Distribution, the Multivariate β-Distribution, and Correlations Among Proportions. Biometrika 49, 65–82. doi:10.2307/2333468

Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158, 96–101.

Punt, A.E., In press. Some insights into data weighting in integrated stock assessments. Fish. Res.

Punt, A.E., Smith, D.C., KrusicGolub, K., Robertson, S., 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia's southern and eastern scalefish and shark fishery. Can. J. Fish. Aquat. Sci. 65, 1991–2005.

Schnute, J.T., Haigh, R., 2007. Compositional analysis of catch curve data, with an application to *Sebastes maliger*. ICES J. Mar. Sci. J. Cons. 64, 218–233.

Shelton, A.O., Dick, E.J., Pearson, D.E., Ralston, S., Mangel, M., Walters, C., 2012. Estimating species composition and quantifying uncertainty in multispecies fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data. Can. J. Fish. Aquat. Sci. 69, 231–246.

Stewart, I.J., Hamel, O.S., 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. Can. J. Fish. Aquat. Sci. 71, 581–588.

Taylor, I., Grandin, C., Hicks, A.C., Taylor, N., Cox, S., 2015. Status of the Pacific Hake (whiting) stock in US and Canadian waters in 2015. Prepared by the Joint Technical Committee of the U.S. and Canada Pacific Hake/ Whiting Agreement.

Thorson, J.T., 2014. Standardizing compositional data for stock assessment. ICES J. Mar. Sci. J. Cons. 71, 1117–1128. doi:10.1093/icesjms/fst224

Thorson, J.T., Hicks, A.C., Methot, R.D., 2015. Random effect estimation of time-varying factors in Stock Synthesis. ICES J. Mar. Sci. J. Cons. 72, 178–185. doi:10.1093/icesjms/fst211

Thorson, J.T., Minto, C., 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES J. Mar. Sci. J. Cons. 72, 1245–1256. doi:10.1093/icesjms/fsu213

Walters, C.J., Martell, S.J.D., 2004. Fisheries Ecology and Management. Princeton University Press, Princeton, New Jersey.

Table 1. Parameters used to generate simulated data sets (the "operating model") and during model fitting (the "estimation model"). A modified version of the 2015 Pacific hake assessment model with 134 estimated parameters is used as both the operating and estimation model (the model uses empirical weight-at-age techniques, and therefore does not estimate individual growth parameters). Survey and fishery selectivity values are not listed but follow the non-parametric form used in Taylor et al. (2015), but without variation over time.

	Operating model	*Estimation model*	
Name	True value	Estimated or fixed?	Number of estimated parameters
Natural mortality rate	0.217	Estimated	1
Expected recruits at unfished level (natural logarithm)	14.470	Estimated	1
Beverton-Holt steepness	0.850	Estimated	1
log-standard deviation of recruitment deviations	0.900	Fixed	-
Additional variance for accoustic survey index	0.313	Estimated	1
Accoustic survey selectivity at age	-	Estimated	4
Fishery selectivity at age	-	Estimated	5
Recruitment deviations	-	Estimated	72
Instantaneous fishing mortality rates	-	Estimated	49

Figure captions:

Fig. 1. Input sample size (x-axis) and effective sample size (N_{eff}; y-axis) for two paramaterizations of the Dirichlet-multinomial distribution across varying values for the Dirichlet-multinomial parameter specific to each parameterization. The dashed line represents the 1:1 line where input sample size is the same as N_{eff}.

Fig. 2. Comparison of spawning output relative to average unfished levels (left-left), spawning output (SPB; top-right), exploitation fraction (catch divided by estimated biomass for individuals aged 3 and older; bottom-left), and recruitment (age-0 abundance; bottom-right) for the Pacific hake assessment given four alternative methods of weighting the age-composition data: (i) weight of zero for the age-composition data (red); (ii) unweighted (green), (ii) iteratively tuned (black); or (iii) Dirichlet-multinomial distribution (blue), where for each model we show the maximum likelihood estimates (solid line) and +/- 1 standard error (shaded region).

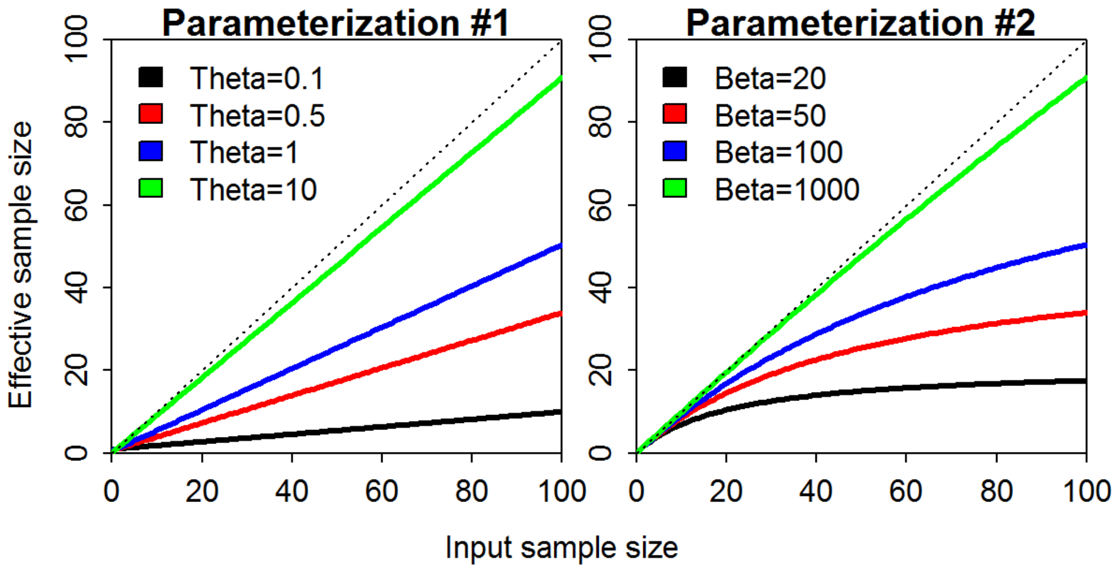
Fig. 3. Pearson residuals for age-composition data from the fishery (top panel) and survey (bottom panel) using the Dirichlet-multinomial to estimate overdispersion (and hence data weighting) for the fishery simultaneously with other model parameters, where each panel shows a circle with area proportional to the Pearson residual (see Eq. 13 for calculation), and with sign indicated by shading (grey: positive residual; white: negative residual).

Fig. 4. Estimated Dirichlet-multinomial variance inflation parameter (top row) and effective sample size (N_{eff}, bottom row) from the “linear” parameterization (parameterization #1) of the

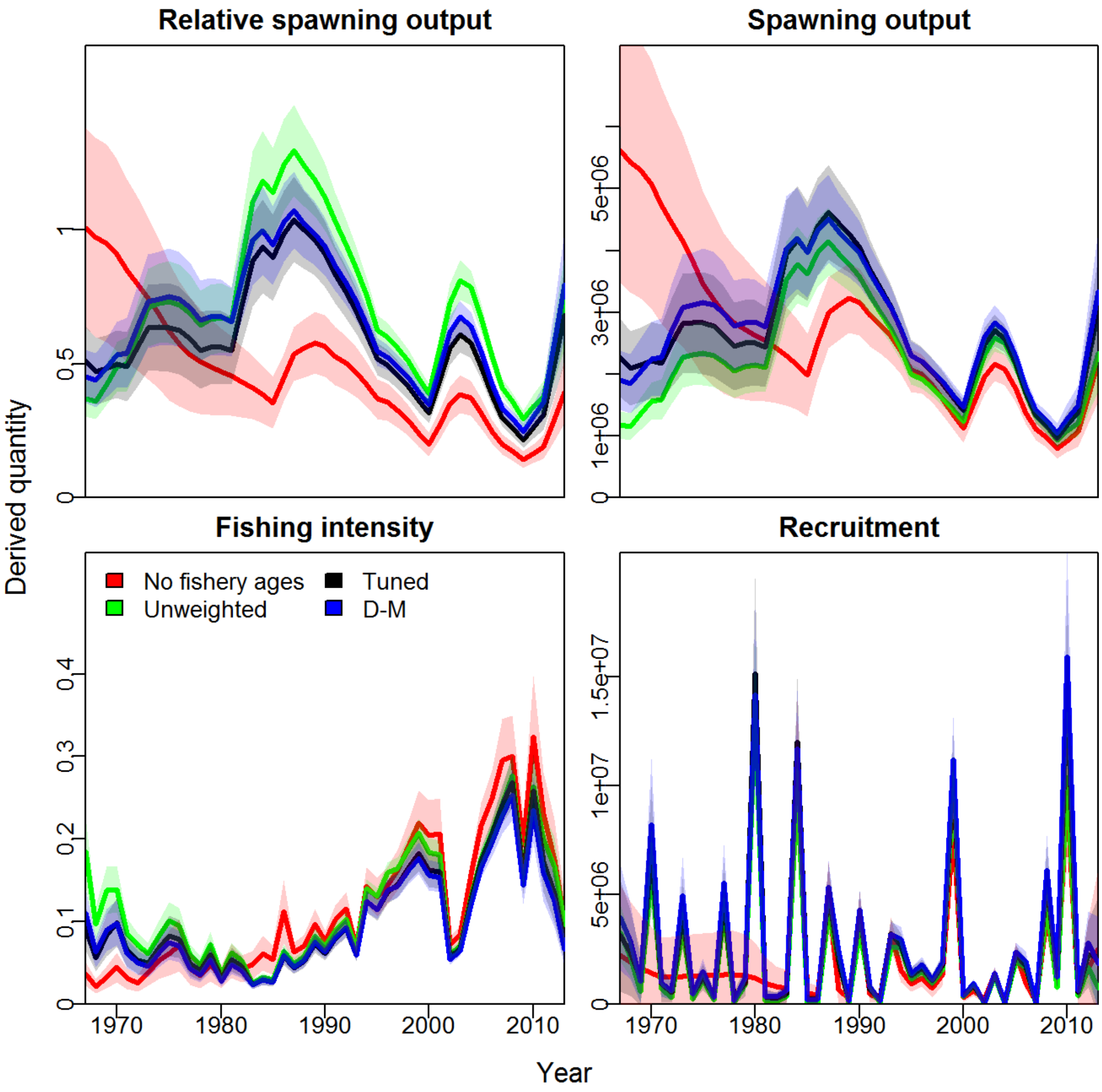
Dirichlet-Multinomial distribution implemented in Stock Synthesis shown for three "true sample sizes" (1st column: 25; 2nd column: 100; 3rd column: 400 samples per year) and four levels of variance inflation (wherein the input sample size provided to Stock Synthesis is 2, 5, 25, or 100 times the true sample size).

Fig. 5. Relative error in parameter estimates across estimation methods (rows; " tuned": using the ratio estimator of the harmonic mean to input sample size; "unweighted": conventional multinomial treating input as effective sample size; "DM": linear-parameterization of the Dirichlet-multinomial distribution) and levels of the inflation factor for the fishery age-composition data in the operating model (columns). Each panel depicts the maximum likelihood estimates of natural mortality rate (*M,* y-axis) and average unfished recruitment (*ln(R_0)*, x-axis), where colors are used to distinguish estimates. We only show results for estimation models where the maximum final gradient was <0.1 (the number of replicates across models is indicated in each panel, where 300 implies that all 100 replicates converged for each of three estimation models), and confirm that results are qualitatively similar if using a different convergence threshold. The lower left panel is not plotted because the DM estimation method was not used when the inflation factor was one.

Fig. 1


Parameterization #1
Theta=0.1
Theta=0.5
Theta=1
Theta=10
Parameterization #2
Beta=20
Beta=50
Beta=100
Beta=1000
Effective sample size
0
20
40
60
80
100
Input sample size

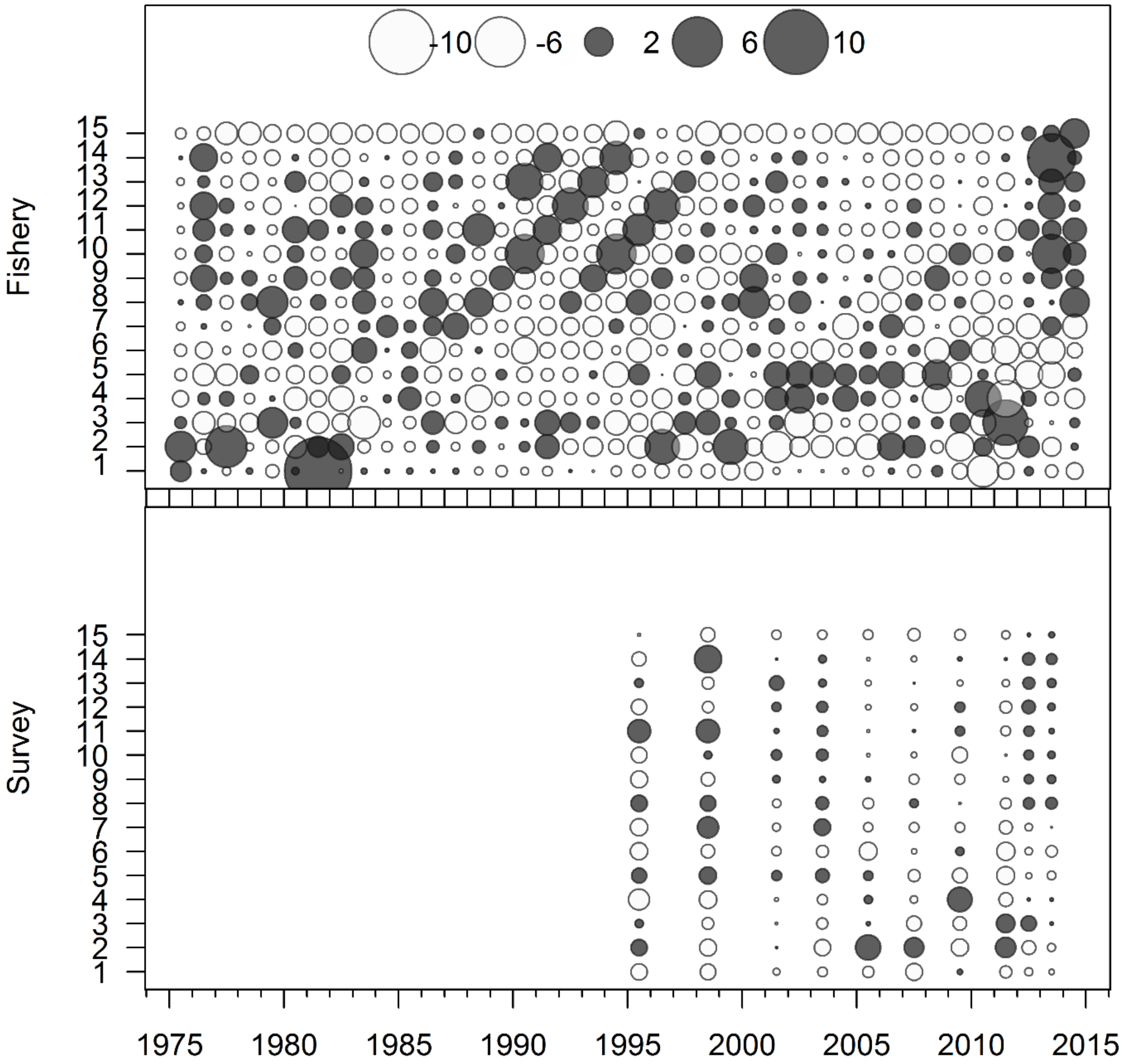
661 Fig. 2



662

663

Fig. 3



-10
-6
2
6
10
Fishery
Survey
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
1975
1980
1985
1990
1995
2000
2005
2010
2015

Fig. 4

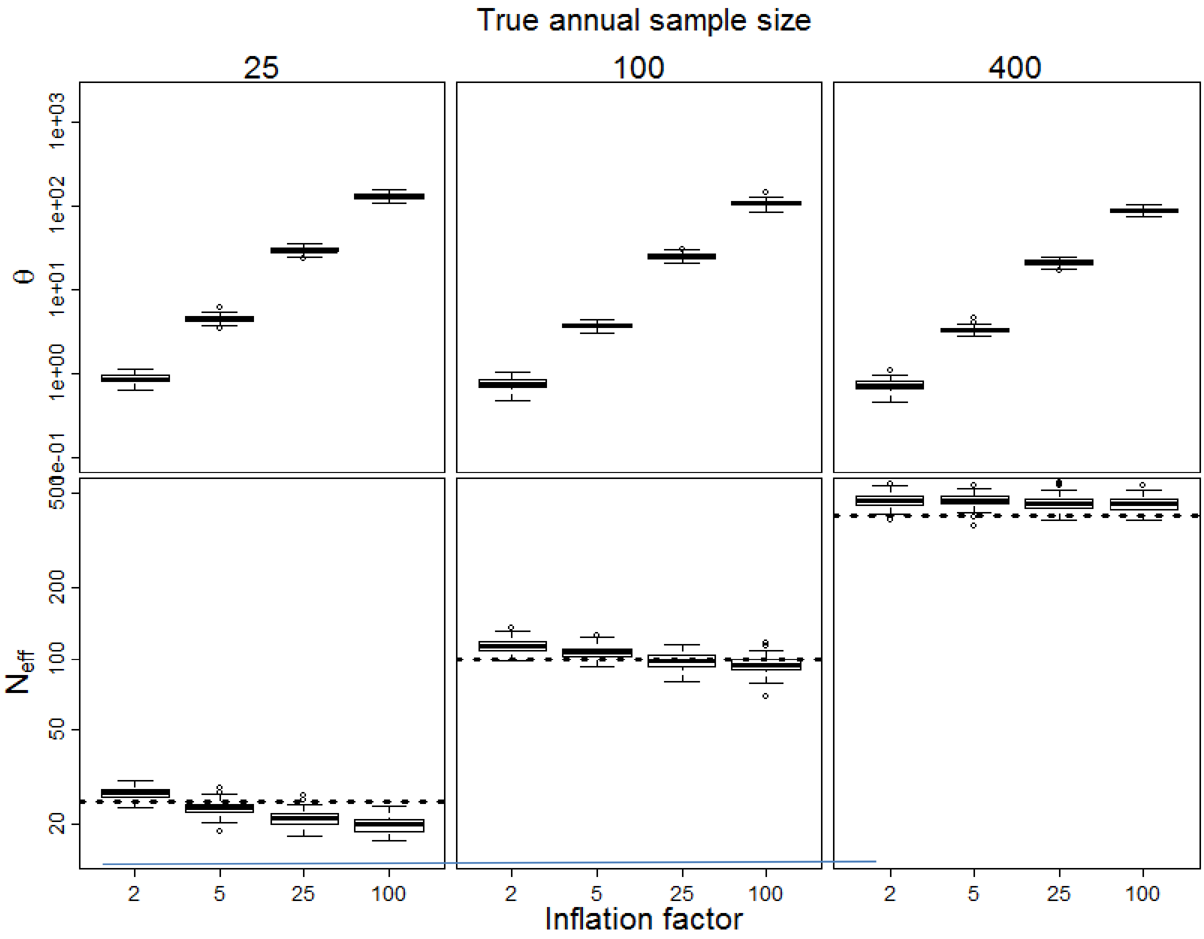
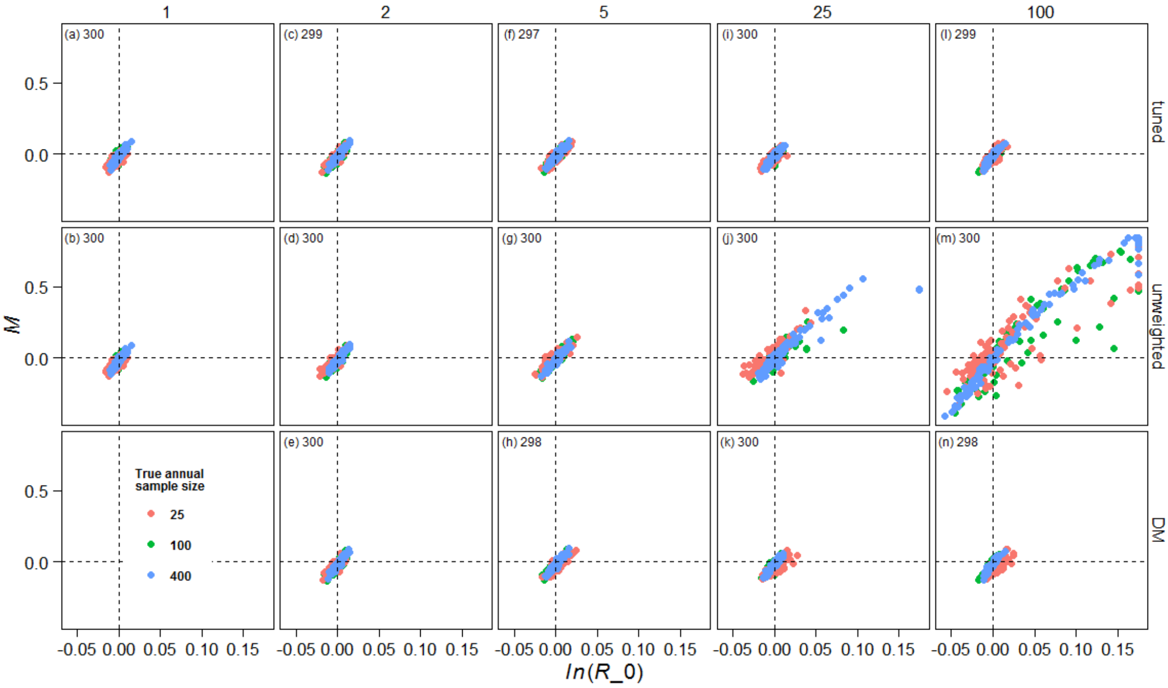

True annual sample size
25
100
400
θ
1e-01
1e+00
1e+01
1e+02
1e+03
N_eff
20
50
100
200
500
2
5
25
100
Inflation factor

Fig. 5


1
2
5
25
100
tuned
unweighted
DM
(a) 300
(c) 299
(f) 297
(i) 300
(l) 299
(b) 300
(d) 300
(g) 300
(j) 300
(m) 300
(e) 300
(h) 298
(k) 300
(n) 298
True annual sample size
25
100
400
M
0.5
0.0
-0.05 0.00 0.05 0.10 0.15
ln(R_0)