

The multivariate-Tweedie: a self-weighting likelihood for age and length composition data arising from hierarchical sampling designs

James T. Thorson ^{1,*}, Timothy J. Miller ² and Brian C. Stock³

¹Habitat and Ecological Processes Research Program, Alaska Fisheries Science Center, NOAA, Seattle, WA 98115, USA

²Resource Evaluation and Assessment Division, Northeast Fisheries Science Center, NOAA, Woods Hole, MA 02543, USA

³Institute of Marine Research, Nye Flødevigveien 20, 4817 His, Norway

*Corresponding author: tel: +001 206 302 1772; e-mail: james.thorson@noaa.gov

Weighting data appropriately in stock assessment models is necessary to diagnose model mis-specification, estimate uncertainty, and when combining data sets. Age- and length-composition data are often fitted using a multinomial distribution and then reweighted iteratively, and the Dirichlet-multinomial (“DM”) likelihood provides a model-based alternative that estimates an additional parameter and thereby “self-weights” data. However, the DM likelihood requires specifying an input sample size (n_{input}), which is often unavailable and results are sensitive to n_{input} . We therefore introduce the multivariate-Tweedie (MVTW) as alternative with three benefits: (1) it can identify both overdispersion (downweighting) or underdispersion (upweighting) relative to the n_{input} ; (2) proportional changes in n_{input} are exactly offset by parameters; and (3) it arises naturally when expanding data arising from a hierarchical sampling design. We use an age-structured simulation to show that the MVTW (1) can be more precise than the DM in estimating data weights, and (2) can appropriately upweight data when needed. We then use a real-world state-space assessment to show that the MVTW can easily be adapted to other software. We recommend that stock assessments explore the sensitivity to specifying DM, MVTW, and logistic-normal likelihoods, particularly when the DM estimates an effective sample size approaching n_{input} .

Keywords: data weighting, Dirichlet-multinomial, hierarchical sampling design, stock assessment, Tweedie, WHAM.

Introduction

Integrated fisheries stock assessment models combine multiple sources of data to estimate population dynamics parameters, abundance, and status through time (Maunder and Punt, 2013). The three most common data types are catch (i.e. fishery removals), abundance indices (i.e. surveys), and age/length composition (i.e. proportions of fish in age or length bin categories), although many integrated models do not simultaneously include all three of these (e.g. Rudd *et al.*, 2021). Time series of composition data are particularly useful to inform relative year class strength, that is, years of strong recruitment (Hjort, 1926). Integrated assessments involve specifying the joint likelihood of parameters given data, and the likelihood of each datum is weighted (explicitly or implicitly) by its estimated or assumed variance. The relative scale of weights strongly influences assessment results when the data sources provide conflicting information about parameters conditional upon model assumptions (Francis, 2011; Maunder and Punt, 2013), and the absolute value of weights affects the overall estimate of assessment uncertainty. Given the typical sample sizes assumed for compositional data, these data will often overwhelm other types of data in estimation of model parameters. As a result, there is a large literature regarding methods to “downweight” compositional data to avoid assigning them larger leverage than other data.

The multinomial distribution has typically been used as the composition data likelihood, which relates the observed proportions to the model expected proportions given a set of parameters (Francis, 2014). The multinomial arises from the simple assumptions that for a fixed number of fish n_{input} and num-

ber of categories C (e.g. age or length classes), the category observed for each fish is independent and the true probabilities of each category, $\{\pi_1, \dots, \pi_c, \dots, \pi_C\}$, are the same for each fish. The multinomial distribution then results in a vector of counts $\{y_1, \dots, y_c, \dots, y_C\}$ of fish in each category, where the expected proportion $\mathbb{E} \left(\frac{y_c}{n_{\text{input}}} \right) = \pi_c$. Importantly, the sampling variance of the observed proportion, $\mathbb{V} \left(\frac{y_c}{n_{\text{input}}} \right) = \frac{\pi_c(1-\pi_c)}{n_{\text{input}}}$, is completely defined by the mean and n_{input} and is a decreasing function of n_{input} , so that n_{input} determines the statistical leverage (“weight”) of composition data. We refer to this multinomial sample size as n_{input} because it must be input as data by the assessment scientist and cannot be estimated as a parameter within the model (Francis, 2014). Fisheries composition data typically exhibit more variation than expected by the multinomial distribution with sample size n_{input} defined as the number of measured animals (i.e. they are “overdispersed”), for example, because the measurement of each animal is not statistically independent as assumed in a multinomial distribution. For example, Pennington and Vølstad (1994) introduced the term “intra-haul correlation” to describe the common observation that fish from the same haul are often more similar in age and length than fish from other hauls. There are two general approaches for handling overdispersion and correlations in composition data: (1) continuing to use the multinomial distribution but with a reduced n_{input} that represents an appropriate “effective” sample size n^* , or (2) replacing the multinomial with a distribution that can reweight n_{input} to estimate the effective sample size and/or correlations.

Received: April 8, 2022. Revised: August 8, 2022. Accepted: August 11, 2022

Published by Oxford University Press on behalf of International Council for the Exploration of the Sea 2021. This work is written by (a) US Government employee(s) and is in the public domain in the US.

Researchers have developed a variety of approaches to estimate input sample size directly from field data, and/or estimate effective sample size within stock-assessment models. Crone and Sampson (1997) demonstrated an *ad hoc* method for estimating n_{input} from a two-stage sampling design typical for commercial fisheries catch age composition, with fish chosen for ageing nested within boat trip. They developed a regression model relating n_{input} to the number of boat trips sampled and showed that n_{input} was between the total number of fish aged and the number of boat trips sampled. This method has been subsequently extended using alternative bootstrap or model-based estimators (Stewart and Hamel, 2014; Thorson, 2014; Thorson and Haltuch, 2018). Another method for calculating n_{input} is to iteratively fit a stock assessment model using output of the previous run to adjust n_{input} . McAllister and Ianelli (1997) proposed to tune n_{input} to achieve “internal consistency”, that is, such that the variance between the observed and predicted proportions at age matches the multinomial variance using the predicted proportions. Pennington and Vølstad (1994) calculated n_{input} for the length composition that would produce the standard error of the mean length under independent sampling, which in one example was about half of the original sample size. Miller and Skalski (2006) showed how n_{input} can be calculated for each category by equating age or length class-specific design-based variance estimates with the variance equation for those classes under the multinomial assumption. Francis (2011) suggested an iterative procedure to estimate effective sample size in assessment models under the similar assumption that it should be standardized relative to the error in mean age/length. Although using these methods to specify or reweight n_{input} in the multinomial is better than ignoring the issue, several drawbacks remain: (1) they do not propagate uncertainty about n_{input} ; (2) model results change based on n_{input} , which hinders model development and reproducibility; and (3) the time required to re-run the model makes it difficult to conduct sensitivity runs, simulation studies, and Bayesian estimation.

The alternative to iterative reweighting algorithms is to use a distribution other than the multinomial. Several distributions can “self-weight” composition data by estimating one additional parameter and have been recommended for use in stock assessment: the Dirichlet (Maunder, 2011; Hulson *et al.*, 2012), Dirichlet-multinomial (“DM”) using either saturating (Candy, 2008) or linear parameterizations (Thorson *et al.*, 2017), and logistic-normal using either additive (Schnute and Haigh, 2007; Miller *et al.*, 2016; Stock and Miller, 2021) or multiplicative parameterizations (Cadigan, 2016). Composition data are often higher or lower than model estimates for a sequence of adjacent years and age/length bins (e.g. Francis, 2014 Figure 4), and some replacements for the multinomial estimate the magnitude of these correlations while also weighting the composition data (e.g. Berg and Nielsen, 2016). For example, the logistic-normal can also be extended with additional parameters to estimate correlation structures, such as first and second order autoregressive, AR(1) and AR(2), and autoregressive moving average (ARMA) processes (Francis, 2014; Perreault *et al.*, 2020; Fisch *et al.*, 2021).

In a simulation study, Fisch *et al.*, (2021) found that the logistic-normal and DM performed better than other likelihoods, and which was best depended on the amount of composition data and process error. One limitation of the DM is

that it still requires specifying n_{input} , which is often unavailable due to inability to access raw data (e.g. when sharing data across jurisdictions or for historical data sets) or costly for survey teams to produce (e.g. when developing and maintaining code to bootstrap raw samples), and results depend somewhat upon the specific value used. The DM also cannot upweight compositional data, that is, estimate an effective sample size greater than n_{input} . In cases when n_{input} is unavailable due to data-access or cost and is therefore fixed at an arbitrary value with little scientific basis, it might be a nuisance that the DM cannot estimate $n_{\text{effective}}$ higher than this arbitrary n_{input} value. Alternatively, when n_{input} is available and accurately represents sampling variability, the fact that the DM requires $n_{\text{effective}} < n_{\text{input}}$ provides an upper bound on the variance of time-varying parameters.

In this study, we introduce an alternative self-weighting likelihood for compositional data, the multivariate-Tweedie (MVTW) distribution. We first describe how the Tweedie distribution arises in nature from the hierarchical expansion of field measurements for age and length samples. In particular, age/length measurements typically arise from a hierarchical (a.k.a. multi-stage) sampling design with unequal sampling probability due to spatially stratified sampling and length-stratified subsampling. If the inverse-sampling probability follows a gamma distribution, we show that this gives rise to the compound-Poisson-gamma (“Tweedie”) distribution. We then use an age-structured simulation model to show that (1) the MVTW is more precise than the DM in estimating effective sample size, and (2) the MVTW can also upweight age-composition data when needed. Finally, we implement the MVTW in the Woods Hole Assessment Model software package (WHAM; Miller and Stock, 2020) and compare its performance to the DM, multinomial, and logistic-normal distributions in an application to data for southern New England–Mid-Atlantic Bight (SNE-MAB) yellowtail flounder. We find that the MVTW estimates higher effective sample sizes than (but comparable uncertainty estimates to) the DM.

Methods

We seek to develop a self-weighting likelihood that generalizes the multinomial distribution but incorporates one or more additional parameters that represent the ratio between input and effective sample size. To do so, we first discuss a useful relationship between the multinomial and Poisson distributions for parameter estimation, summarize the recently developed MVTW distribution, discuss how the Tweedie distribution arises in nature from hierarchical expansion of age and length measurements, and derive properties of that MVTW likelihood when used to fit age and length composition data.

Using the Poisson to estimate multinomial parameters

Researchers have known for decades that a generalized linear model using a log-linked Poisson distribution can be specified to give identical estimates to a multivariate-logit-linked multinomial regression (Birch, 1963; Palmgren, 1981; Baker, 1994). Cormack (1989) also showed how to take advantage of this relationship for inferences from mark-recapture studies. However, we are not aware of a derivation showing this equivalency in a stock-assessment context. In this context, a

population-dynamics model generates a predicted proportion π_c for each age/length bin c , which is compared with an observed proportion p_c that has been assigned a sample size n_{input} .

Conventionally, the discrepancy between prediction π_c and observation $p_c n_{\text{input}}$ is computed using a multinomial distribution:

$$\begin{aligned} \log \mathcal{L}(\pi; n_{\text{input}}, \mathbf{p}) &= \log(d\text{multinom}(\mathbf{p}|\pi, n_{\text{input}})) \\ &= \log\left(\frac{n_{\text{input}}!}{\prod_{c=1}^C (p_c n_{\text{input}})!} \prod_{c=1}^C \pi_c^{p_c n_{\text{input}}}\right) \\ &= \sum_{c=1}^C p_c n_{\text{input}} \log(\pi_c) + \delta_1, \end{aligned} \quad (1)$$

where $\delta_1 = \log\left(\frac{n_{\text{input}}!}{\prod_{c=1}^C (p_c n_{\text{input}})!}\right)$. However, an identical log-likelihood (up to a constant term) is obtained by using the product of independent Poisson distributions:

$$\begin{aligned} \log \mathcal{L}(\pi; n_{\text{input}}, \mathbf{p}) &= \log\left(\prod_{c=1}^C d\text{Poisson}(n_{\text{input}} p_c | n_{\text{input}}, \pi_c)\right) \\ &= \log\left(\frac{\prod_{c=1}^C (\pi_c n_{\text{input}})^{p_c n_{\text{input}}} \prod_{c=1}^C e^{-\pi_c n_{\text{input}}}}{\prod_{c=1}^C (p_c n_{\text{input}})!}\right) \\ &= \sum_{c=1}^C p_c n_{\text{input}} \log(\pi_c) + \delta_2 + \delta_3 + \delta_4, \end{aligned} \quad (2)$$

where $\delta_2 = n_{\text{input}} \log(n_{\text{input}})$ captures the component $\prod_{c=1}^C (n_{\text{input}})^{p_c n_{\text{input}}}$ from the first term in the numerator because $\sum_{c=1}^C p_c = 1$, $\delta_3 = -n_{\text{input}}$ captures the second term in the

numerator because $\sum_{c=1}^C \pi_c = 1$, and $\delta_4 = \log\left(\frac{1}{\prod_{c=1}^C (p_c n_{\text{input}})!}\right)$ captures the denominator. This derivation shows that Poisson and multinomial distributions are proportional, and therefore can be used to generate identical maximum likelihood estimates, asymptotic standard errors, likelihood profiles, and Bayesian posterior distributions. Furthermore, the two can be equivalent and therefore used to generate identical likelihoods (e.g. for use in likelihood ratio tests and the Akaike Information Criteria) if the integration constants $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ are properly incorporated. This equality is sometimes called Poissonization (Adhikari and Pitman, 2021, Chapter 7); said another way, if we have a total count N that follows a Poisson distribution with intensity μ , and a multinomial with size N and proportions \mathbf{p} , then the resulting compound distribution is identical to a series of independent Poisson distributions with intensity $\mu \mathbf{p}$.

Generalizing the mean–variance relationship

Given that the multinomial likelihood can be replaced with a vector of independent Poisson distribution, we propose to use the Tweedie distribution which provides a more flexible mean–variance relationship than the Poisson distribution. We acknowledge that there is no general closed-form expression for the sum of Tweedie random variables (as used during Poissonization), and therefore no equivalent version of the multinomial distribution that arises from a vector of Tweedie

random variables. We instead we refer to the proportions arising from normalizing a vector of Tweedie distributions a new “MVTW distribution” (MVTW):

$$\mathbf{p} \sim \text{MVTW}(\pi, \phi, \psi | n_{\text{input}}),$$

where response \mathbf{p} is a vector containing the observed proportions $p_c = \frac{y_c}{\sum_{c=1}^C y_c}$ arising from expanded count y_c for each category, π is a vector containing the predicted proportion π_c , ϕ and ψ are Tweedie variance parameters defined below, and n_{input} is an “input sample size” that is assumed at a fixed value. This approach was termed the MVTW distribution by Thorson *et al.*, (2022), where it was derived from an thinned and marked Poisson process for animal foraging. The MVTW is “multivariate” in the sense that (1) the response \mathbf{p} arises from normalizing a vector of expanded counts \mathbf{y} , and (2) its calculation involves calculating a vector of proportions π . In other context, a vector of proportions is typically calculated from a “multivariate logit” link function, and the MVTW is therefore multivariate in a similar sense as the multinomial. However, the MVTW has strong constraints on the covariance among categories $\mathbb{V}(\mathbf{p})$, and this is again similar to the multinomial distribution.

We next provide support for three claimed benefits of using the MVTW:

- (1) It supports composition data that includes zeros, as well as a continuous proportion p_c for each category.
- (2) It arises naturally from the process of generating age and length composition data, where animals are subsampled to measure age/length and then expanded to calculate proportions p_c .
- (3) It can account for mis-specification of input sample size n_{input} .

To illustrate these claims, we start by specifying a Poisson distribution, $y_c \sim \text{Poisson}(\lambda_c)$, where $y_c \equiv n_{\text{input}} p_c$ is constrained to the set of non-negative integers $y_c \in \{0, 1, \dots, n_{\text{input}}\}$. However, y_c is not in practice constrained to non-negative integers when fitted in stock-assessment models but are instead specified such that $0 \leq y_c \leq n_{\text{input}}$. This arises because p_c is typically calculated as a hierarchical expansion of age or length samples where expansion-factors are derived from subsampling rates (Crone and Sampson, 1997; Thorson, 2014), or via the action of an age-length key. In practice, stock-assessment software accommodates this by replacing all factorial functions (which are defined only for integers) with suitably modified gamma functions (where $\Gamma(X) = (X-1)!$ for integers X but is also defined for non-integer values).

We next illustrate how hierarchical expansion of age and length composition samples can be approximated as a compound-Poisson-gamma process (Foster and Bravington, 2013). A hierarchical sampling design typically results in a count of fish X_c for each age, and hierarchical expansion then computes an estimator for the true proportion p_c for each age. However, each age or length sample may have a different sampling (a.k.a. inclusion) probability, arising from (1) subsampling rates for measuring age or length of animals in a given sample, and (2) spatially stratified sampling rates; these are sometimes termed “first-stage” and “second-stage” expansion, respectively (Thorson, 2014). Each individual animal x_c is then expanded by the inverse of its sampling probability (representing the inverse product of sampling and subsampling rates), where the expansion factor for animal x_c

is defined as Z_{xc} . If expansion factor Z_{xc} follows a gamma distribution, this then results in a compound-Poisson-gamma process:

$$\begin{aligned} X_c &\sim \text{Poisson}(\lambda_c) \\ Z_{xc} &\sim \text{Gamma}(k, \theta_c) \\ y_c &= \sum_{x=1}^{X_c} Z_{xc}, \end{aligned} \quad (3)$$

where k is the shape and θ_c the scale parameters for a gamma distribution. If we express these in terms of $\mathbb{E}(y_c) = \pi_c n_{\text{input}}$ and define two variance parameters such that $\mathbb{V}(y_c) = \phi(\pi_c n_{\text{input}})^\psi$, the compound process can be evaluated efficiently as a Tweedie distribution (Kendal, 2004):

$$y_c \sim \text{Tweedie}(\pi_c n_{\text{input}}, \phi, \psi), \quad (4)$$

where the parameters of the compound-Poisson-gamma process can be calculated as $\lambda_c = \frac{(\pi_c n_{\text{input}})^{2-\psi}}{\phi(2-\psi)}$, $k = \frac{2-\psi}{\psi-1}$, and $\theta_c = \phi(\psi-1)(\pi_c n_{\text{input}})^{\psi-1}$. These Tweedie parameters $\{\pi_c, \phi, \psi\}$ result in the compound-Poisson-gamma with the constraint that k is constant across categories while θ_c varies among categories (as assumed in equation 3). We then compute the likelihood for all age/length bins using a sequence of Tweedie likelihoods,

$$\mathcal{L}_{MVTW}(\pi, \phi, \psi; \mathbf{p}, n_{\text{input}}) = \prod_{c=1}^C dtweedie(y_c | \pi_c, n_{\text{input}}, \phi, \psi)$$

where response $y_c = n_{\text{input}} p_c$.

In the following, we optimize this MVTW likelihood with respect to the following parameters: proportion π_c for each age/length bin, log-variance ϕ^* defined such that $\phi = \exp(\phi^*)$, and logit-scale power ψ^* defined such that $\psi = 1 + \frac{\exp(\psi^*)}{1 + \exp(\psi^*)}$. Variance parameters ϕ^* and ψ^* are defined to have unbounded continuous support while maintaining the Tweedie restrictions $1 < \psi < 2$ and $\phi > 0$. We assume that ϕ^* and ψ^* are constant for all age/length bins, although future research could explore the consequences of relaxing this assumption. The resulting Tweedie distribution then has non-negative continuous support for response $y_c = p_c n_{\text{input}}$.

We next define the approximate sample size n^* as the number of multinomial samples with the same variance as a given distribution for age or length composition data. We seek to calculate n^* after estimating Tweedie parameters $\{\phi, \psi\}$ by fitting them in a stock assessment, and call n^* the “effective sample size” when it is calculated in this way. We follow Thorson and Haltuch (2018) in approximating the effective sample size by calculating $\mathbb{V}(p_c) = \mathbb{V}\left(\frac{y_c}{\sum_{c'=1}^{n_c} y_{c'}}\right)$, that is, the variance of a ratio where the numerator is the expanded count y_c for a given bin and the denominator is the total count across all bins. Calculating this involves a second order Taylor series approximation to $\mathbb{V}\left(\frac{A}{A+B}\right)$ where $A = y_c$ and $B = \sum_{c'=1}^{n_c} y_{c'} - y_c$ are treated as two uncorrelated random variables. After some algebra (see Thorson and Haltuch (2018) Supplementary Material A for details), this then yields:

$$n_c^* \approx \frac{\pi_c(1-\pi_c)}{\pi_c^2 \times \left(\frac{\phi(n_{\text{input}}\pi_c)^\psi}{n_{\text{input}}^2\pi_c^2} - 2\frac{\phi(n_{\text{input}}\pi_c)^\psi}{n_{\text{input}}^2\pi_c} + \frac{\sum_{c'=1}^{n_c} \phi(n_{\text{input}}\pi_{c'})^\psi}{n_{\text{input}}^2} \right)}, \quad (5)$$

where n^* is calculated as the median of n_c^* across bins c for that vector of age or length compositions, that is, $n^* = \text{Median}_c(n_c^*)$. Numerical exploration shows that this

Table 1. Quick facts about the MVTW distribution, including how the likelihood can be evaluated numerically, how to calculate effective sample size for each category n_c^* , and how the median effective sample size can be further approximated to gain intuition about model behaviour.

Characteristics	Details
Likelihood calculation	$\mathcal{L}_{MVTW}(\pi, \phi, p; \mathbf{p}, n_{\text{input}}) = \prod_{c=1}^{n_c} dtweedie(p_c n_{\text{input}} n_{\text{input}} \pi_c, \phi, \psi)$
Effective sample size by category	$n_c^* \approx \frac{\pi_c(1-\pi_c)}{\pi_c^2 \times \left(\frac{\phi(n_{\text{input}}\pi_c)^\psi}{n_{\text{input}}^2\pi_c^2} - 2\frac{\phi(n_{\text{input}}\pi_c)^\psi}{n_{\text{input}}^2\pi_c} + \frac{\sum_{c'=1}^{n_c} \phi(n_{\text{input}}\pi_{c'})^\psi}{n_{\text{input}}^2} \right)}$
Simplified approximation to median effective sample size	Where: $n^* \approx \text{Median}_c(n_c^*)$ $n^* \approx n_{\text{input}} \times \text{Median}(\phi^{-1} e^{(\psi-1)\log(\pi_c)})$

can be further approximated as:

$$n^* \cong \text{Median}_c \left(n_{\text{input}} \times \frac{e^{(\psi-1)\log(\pi_c)}}{\phi} \right). \quad (6)$$

These properties are summarized in Table 1. We note that it is not necessary to calculate or report the approximate sample size n^* . However, n^* is a simple summary of the variance resulting from n_{input} and the weighting estimated by the MVTW distribution. It therefore could be useful to compare across different assessment models with alternative specification, to predict the likely impact of changes in n_{input} under future sampling designs (Thorson *et al.*, 2020), or when measuring the magnitude of measurement and process errors using n_{input} and n^* (Francis, 2011, Figure 2; Thorson and Haltuch, 2018, Figure 7).

In summary, we emphasize two interpretations for how the MVTW distribution arises in nature:

- (1) *Hierarchical expansion process*: First, we interpret latent variable Z_{xc} as the expansion factor associated with each animal that is measured for age or length in the raw data (e.g. from spatially stratified sampling and length-stratified subsampling). We define the mean expansion factor $w_c \equiv \mathbb{E}(Z_{xc}) = k\theta_c$ and the variance $v_c \equiv \mathbb{V}(Z_{xc}) = k\theta_c^2$ (see equation 3), such that the coefficient of variation is $CV = 1/\sqrt{k}$. In the limit that each animal in a population has equal subsampling probability, then the $CV \rightarrow 0$ such that $k \rightarrow \infty$. Plugging this into Tweedie parameters (see below equation 4), we see that $\psi \rightarrow 1$ and the Tweedie approaches a Poisson distribution in this limiting case, such that the Tweedie provides a maximum likelihood estimate that approaches a multinomial likelihood with equivalent input sample size. Alternatively, as the expansion factor Z_{xc} for each animal becomes more variable (i.e. $CV \gg 1$ and $\psi \rightarrow 2$), the effective sample size decreases relative to the limiting case of equal expansion rates (see equations 5–6). This latter property makes sense intuitively given that unequal expansion factors produce expanded age or length composition vectors \mathbf{p} that are dominated by those few measured animals with large expansion factors Z_{xc} .
- (2) *Mis-specified input sample size*: Second, the input sample size n_{input} might be mis-specified. For example, some stock assessments specify n_{input} a priori (i.e.

Table 2. A comparison of the linear parameterization of the DM and the MVTW in terms of how the distribution arises in nature (the generative process) and a closed-form expression for each.

	Generative process	Closed-form calculation
DM	A multinomial distribution for age/length composition data: $y_c \sim \text{Multinomial}(\pi^*, n_{\text{input}})$, where $y_c = \mathbf{p}_t n_{\text{input}}$, and selectivity varies randomly in each sampling effort: $\pi^* \sim \text{Dirichlet}(\pi, n_{\text{input}}\theta)$ yields a linear parameterization of the DM process $y_c \sim$ $\text{Dirichlet.Multinomial}(\pi, n_{\text{input}}\theta)$, where $y_c = \mathbf{p}_t n_{\text{input}}$, and equivalent sample size is decreased: $n^* = \frac{1}{1+\theta} + \frac{\theta}{1+\theta} n_{\text{input}}$.
MVTW	A process of counting animals of different age/length categories: $X_c \sim \text{Poisson}(\lambda_c)$, where each encountered animal is expanded by an expansion factor Z_{xc} that varies among individuals: $Z_{xc} \sim \text{Gamma}(k, \theta_c)$. $y_c = \sum_{x=1}^{X_c} Z_{xc}$... yields a MVTW distribution. $\mathbf{p}_c \sim \text{MVTW}(\pi, \phi, \psi n_{\text{input}})$, where equivalent sample size can be greater or less than input sample size: $n^* = n_{\text{input}} \times \text{Median}(\phi^{-1} e^{(\psi-1)\log(\pi_c)})$.

Table 3. Summary of differences between the DM and MVTW distributions.

Characteristics	DM	MVTW
Number of data-weighting parameters	1: $\log(\theta)$	2: $\log(\phi)$ and $1 \leq \psi \leq 2$
Can upweight data above n_{input}	No	Yes
Can estimate heteroscedasticity	No	Yes
Numerically stable for all values of parameters	Yes	No ¹
Could be extended by making data-weighting parameter(s) a function of age/length bin	No	Yes
Ratio of effective and input sample size is only a function of data-weighting parameters	Yes	No ²
Can use AIC or likelihood-ratio test to compare with multinomial distribution	Yes	No
Parameter estimates are invariant to a proportional change in input sample size	No	Yes

¹The MVTW likelihood becomes multimodal as power ψ approaches 1, and in practice we recommend constraining $1.05 < \psi < 2$ to promote numerical stability.

²The effective sample size of the MVTW depends on data-weighting parameters as well as bin proportions whenever power $\psi > 1$.

$n_{\text{input}} = 100$) or based on the number of sampled fish or survey tows. In these cases, the “weight” associated with each sample Z_{xc} will be systematically greater or less than one; intuitively, each of n_{input} animals sampled in the multinomial distribution represents w_c “effective” samples from the age/length distribution.

We further compare the Tweedie and DM models in terms of how the distribution arises in nature (Table 2) and how they differ in practice (Table 3).

Simulation experiment

To explore the performance of the new MVTW likelihood, we compare it with the existing linear-parameterization of the DM likelihood (Thorson *et al.*, 2017) within an age-structured simulation and estimation package, CCSRA (<https://github.com/James-Thorson/CCSRA>). This package has been used in numerous prior simulation studies (Thorson and Cope, 2015;

Thorson and Kristensen, 2016; Thorson *et al.*, 2018; Winker *et al.*, 2020), and includes code to simulate and estimate population and fishery dynamics. The package is implemented in the R statistical environment (R Core Team, 2021) and estimates fixed effects using maximum likelihood while integrating stochastic variation in recruitment via R-package TMB (Kristensen *et al.*, 2016). We have added the MVTW distribution to the package for purposes of this analysis.

We use package CCSRA to simulate data arising from age-structured population dynamics driven by fishery removals. The fishing mortality rate F_t follows a Gompertz effort-dynamics model (Thorson *et al.*, 2013) that results in increasing and then decreasing fishing mortality (a “two-way trip”). We specifically simulate dynamics over 20 years for ages 0–20, given von Bertalanffy growth rate $K = 0.2 \text{ year}^{-1}$ and $L_\infty = 100 \text{ cm}$, weight-at-length $W = 0.01 \times L^{3.01}$, knife-edge maturity at age 3, mortality rate $M = 0.3 \text{ year}^{-1}$, a Beverton-Holt stock-recruit function with steepness $b = 0.86$, average unfished recruitment $R_0 = 10^9$ where recruitment variability follows a lognormal distribution with log-standard deviation $\sigma_R = 0.4$, and logistic fishery selectivity with 50% selection at age-3 and logit-slope of 1.0. An abundance index is available in each year that is proportional to exploited abundance, where measurement errors follow a lognormal distribution with a log-standard deviation $\sigma_I = 0.1$, fishery catches are known without error, and age-composition samples are available from multinomial subsamples of fishery catches. These settings provide a relatively simple and informative example assessment with low process-error and precise data inputs, useful to illustrate the behaviour of compositional likelihoods under well-behaved conditions.

We explore estimation model performance using a 3-by-3 factorial cross of two factors:

- (1) *True compositional sample size*: We explore high, medium, and low sample-size scenarios for compositional data involving 100, 40, or 10 fishery age-composition samples per year.
- (2) *Mis-specified input sample size*: We also explore three scenarios where input sample size is substantially inflated relative to the true variance of age-composition samples (i.e. where n_{input} is specified as 2.4 times the true value), slightly inflated (n_{input} is 1.2 times the true value), or set too low (n_{input} is 0.6 times the true value).

For each simulated data set, we fit the age-structured estimation model using either the DM or the MVTW distribution and record their estimates of effective sample size. We then replicate each configuration 50 times, for a total of $3 \times 3 \times 2 \times 50 = 900$ total estimation model fits. In each replicate r , we record the average across years of effective sample size $n_r = \frac{1}{20} \sum_{t=1}^{20} n_{rt}^*$ and the estimate of recruitment N_{0t} , and compare these with true values. We specifically record the epsilon bias-correction estimator of recruitment to account for retransformation bias (Thorson and Kristensen, 2016). The epsilon estimator acknowledges that recruitment deviation estimates δ_t are penalized towards the Beverton-Holt function, but are lognormally distributed and hence exponentiated when calculating recruitment $N_{0t} = f(SB_t)e^{\delta_t}$. Therefore, calculating the expected recruitment $\mathbb{E}(N_{0t})$ requires accounting for the standard error of δ_t , which epsilon-correction does automatically in a way that generalizes alternative methods (Methot and Taylor, 2011; Thorson, 2019). We explore performance when estimating recruitment because (1) bias in estimated recruitment across years is informative about biased population scale, and (2) recruitment is more variable than other population-dynamics measures and has a high-experimental power for detecting performance differences.

Case-study demonstration

We also demonstrate the utility of the MVTW distribution in stock-assessment packages that are used for real-world assessments by implementing it in the R package WHAM (Miller and Stock, 2020; Stock and Miller, 2021). We then compare alternative configurations of the model fitted to the southern New England/Mid-Atlantic yellowtail flounder data set, which is used as the basic example for the package. The data set for this stock is structured with six age classes (ages 1–5, and 6+, where 6+ is an accumulator group) for years 1973–2016, with freely estimated, age-dependent natural mortality and dynamics driven by one fishery that generates age-composition data for all years. It also includes empirical weight-at-age data, as well as indices of abundance and age-composition samples from two surveys for 1973–2016. We configured selectivity and recruitment as in model 1 in the example 1 script available with WHAM, with three age-specific selectivity blocks and annual log-recruitment as independent Gaussian random effects with estimated mean.

We fit these data using four configurations which differ in how fishery and survey age-composition data are fitted:

- (1) *Multinomial*: Using the input sample size of 100 for the fishery and 50 for each of two surveys.
- (2) *DM*: Estimating three additional parameters representing the downweighting ratio for the fishery and each of two surveys.
- (3) *MVTW*: Estimating six additional parameters, two each for the fishery and each of two surveys, governing the weighting ratio for each fleet.
- (4) *Logistic-normal*: Estimating three additional parameters representing the variance for the fishery and each of the two surveys.

and see Appendix B of Stock and Miller (2021) for the likelihood for each configuration. We compare models in terms of estimating fishing mortality, recruitment, and spawning

biomass, as well as estimated effective sample size for the DM and MVTW models for the fishery and each survey.

Results

We first explore the properties of the MVTW distribution, specifically calculating the properties of the MVTW as a function of variance parameters (Figure 1). When the power parameter $\psi = 1$ (blue line in bottom row of Figure 1), the coefficient of variation of bin proportions is identical to a multinomial distribution such that effective sample size n_c^* is identical regardless of estimated proportion π_c for that age/length bin. By contrast, when $\psi > 1$ (green and yellow lines in bottom row of Figure 1), the coefficient of variation decreases slower than the multinomial distribution as a function of proportion π_c , such that effective sample size n_c^* decreases for age/length bins with higher proportional abundance. In general, n_c^* decreases proportionally for all bins with increasing parameter ϕ (Figure 1 middle row), and increases proportionally with increasing input sample size n_{input} (Figure 1 top row), such that an increasing n_{input} can be exactly offset by a proportional decrease in parameter ϕ . Finally, the effective sample size n^* can exceed input sample size n_{input} (i.e. data are upweighted relative to a multinomial distribution) whenever $\phi < \text{Median}_c(e^{(\psi-1)\log(\pi_c)})$ (see equation 6).

The simulation experiment shows that the DM and MVTW estimate different effective sample sizes n^* depending upon the rate of data inflation, which results in estimates of the different data-weighting ratio (defined as n^*/n_{input}). When the input sample size is set too low (left column, Figure 2), the DM consistently estimates a data-weighting ratio, n_{input} , approaching the upper bound of 1.0. By contrast, the MVTW upweights the data to approximately the correct value of $1.66 = 1/0.6$, and is most accurate in the medium and high sample-size scenarios. At the other extreme, when n_{input} is specified well above the true value (right column, Figure 2), the DM underestimates the data-weighting ratio in particular when true sample size is low (i.e. top-right panel). Even when the n_{input} is specified at approximately the correct level, the MVTW results in a more precise estimate of the data-weighting ratio (middle column, Figure 2), where both DM and MVTW become more precise (i.e. a tighter distribution around the true value) when the true sample size is moderate or high (i.e. $n_{\text{sample}} \geq 40$ in middle or bottom rows).

These differences in data-weighting ratio result in small differences in model performance, that is, in estimates of recruitment (Figure 3). Surprisingly, the MVTW has somewhat worse performance when n_{input} is set too low (Figure 3, left column), and superior performance when n_{input} is set much too high (Figure 3, right column). The latter presumably occurs because the DM downweights the age-composition data more than is necessary, whereas the MVTW estimates an effective sample size approaching the true value.

Finally, the alternative age composition likelihoods in an assessment model for SNE-MAB yellowtail flounder result in the modest differences in estimates of annual fishing mortality, recruitment, and spawning biomass (Figure 4). The DM and multinomial assumptions resulted in the largest and smallest SSB estimates, respectively; this pattern is inverted for the estimates of fully selected fishing mortality (as expected for these two variables), and this arises because the DM estimates somewhat higher fishery selection for ages 2–3 than the other models (and particularly relative to the multinomial distribu-

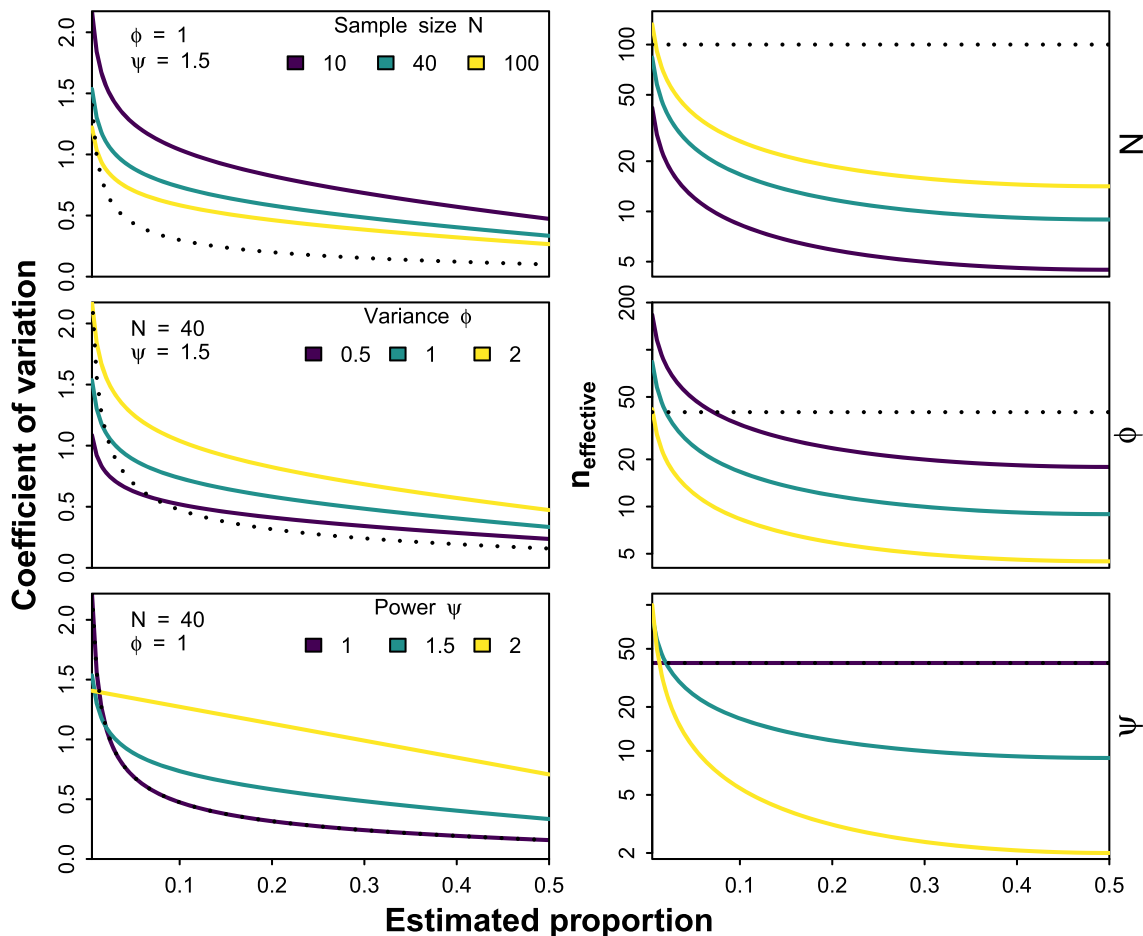


Figure 1. The coefficient of variation (left column, y-axes) and effective sample size n_c^* calculated using the approximation in equation (5) (right column, y-axes) for a category c having an estimated proportion π_c (x-axes) for different values of input sample size n_{input} (top row) and two data-weighting parameters ϕ (middle row) and ψ (bottom row), showing three values in each case (see color panel in left column for each row), and comparing these with the same values for a multinomial distribution as a reference (dashed line).

tion). The logistic-normal and MVTW assumptions resulted in estimates intermediate to the DM and multinomial assumptions. The MVTW estimates SSB 30% higher than the multinomial and 16% lower than the DM, averaged over the last 10 years (Figure 4). Consistent with the simulation-experiment, however, the MVTW estimates a somewhat higher effective sample size (23.0 for fishery, 20.5 and 18.6 for surveys) than the DM (18.6 for fishery, 17.6 and 15.2 for surveys), and hence results in estimates somewhat closer to when using a multinomial distribution using the original input sample size (100 for fishery, 50 and 50 for surveys). Meanwhile, the DM and MVTW estimate uncertainty (Appendix A, Figure S1) that is generally intermediate between multinomial model (estimating lowest uncertainty) and the logistic-normal (estimating highest uncertainty).

Discussion

We have introduced a new “MVTW” likelihood to fit age and length-composition data in stock-assessment models. This distribution allows zeros to be fitted in age and length composition, which is not possible using alternative logistic-normal, multivariate-lognormal, or Dirichlet distributions but is possible using the multinomial or DM distributions. We provide

a novel derivation of this likelihood from the hierarchical expansion of hierarchical sampling designs (as are widely used worldwide), whereby individual animals are counted and then expanded when calculating a total proportion which is then treated as data in many stock-assessment models. We specifically derive important characteristics of this new likelihood, including (1) its effective sample size as a function of estimated parameters, (2) that it can allow upweighting of data in cases when input sample size is unknown, and (3) that it can estimate heteroscedasticity, whereby age/length bins with higher proportion are estimated to have proportionally higher variance than expected under the multinomial or DM distributions. A simulation experiment shows that the new likelihood can perform better than the DM in some cases, while a case-study confirms that it is feasible to implement for real-world assessments. Both simulation and case-study examples involve models implemented using TMB, in part because the *dtweedie* likelihood function was already available in that platform, and we have added it as option to the assessment model WHAM. However, the *dtweedie* function is also being added to ADMB-13 in 2022 (J. Ancheta, personal communication), so the “MVTW” likelihood will soon be feasible to implement in other models, including Stock Synthesis (Methot and Wetzel, 2013).

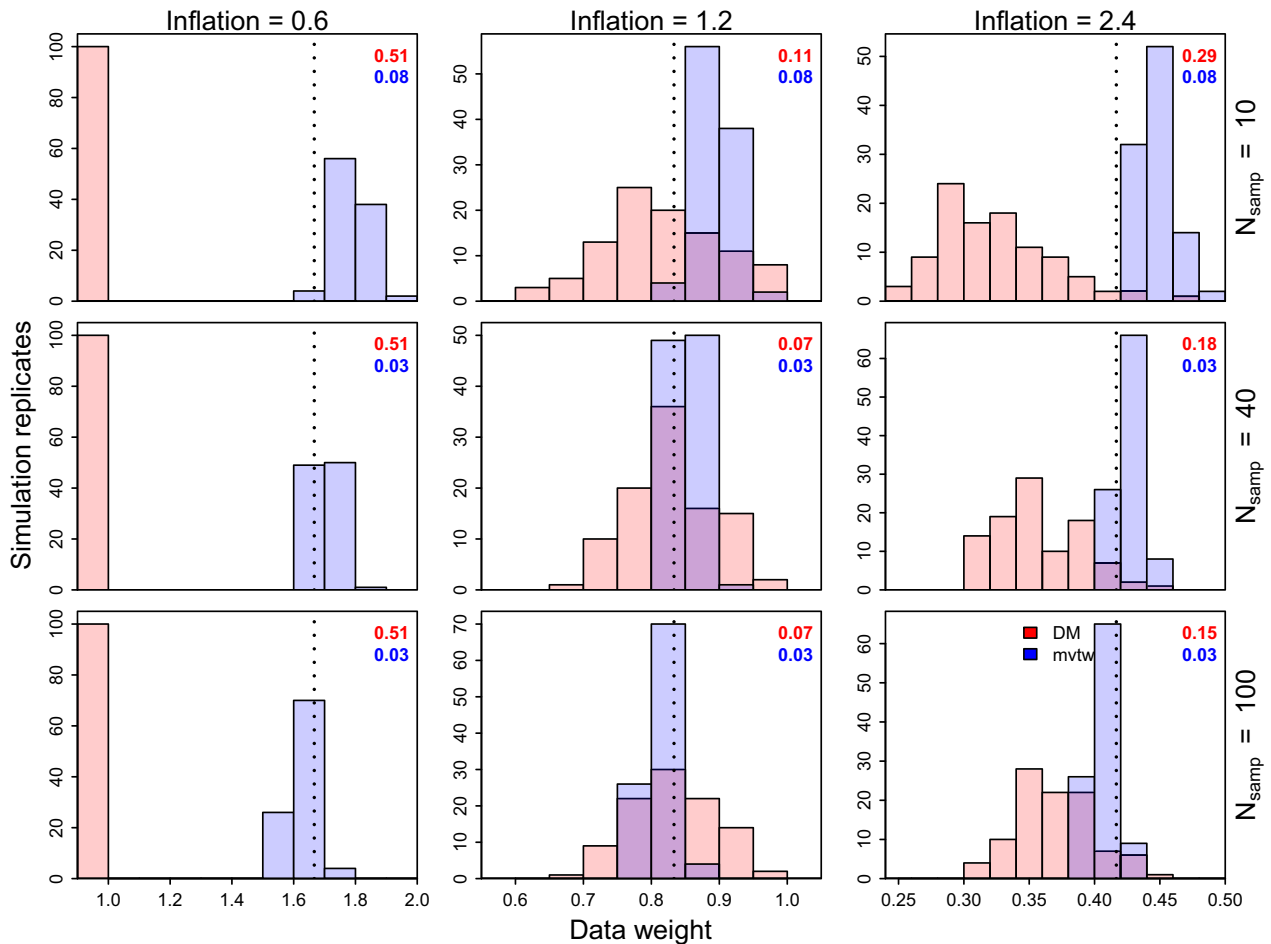


Figure 2. Estimated data weight (x-axes) for each simulation replicate (y-axes) for the DM (red) or MVTW (blue) data-weighting methods relative to the true value (black dashed line), given an input sample size that is too low (inflation = 0.6, left column), slightly too high (inflation = 1.2, middle column), or substantially too high (inflation = 2.4, right column), when true sample size is low ($N_{\text{samp}} = 10$, top row), moderate ($N_{\text{samp}} = 40$, middle row), or high ($N_{\text{samp}} = 100$, bottom row). Each panel also shows the root-mean-squared-log-ratio of estimated and true effective sample size (top-right numbers), where a value close to 0 would indicate optimal performance.

We have introduced the MVTW likelihood in part via analogy to the equivalence between properly specified Poisson and multinomial distributions. Although these distributions are equivalent in terms of maximum likelihood estimates, asymptotic standard errors, and model selection, they are not equivalent in terms of properties when simulating new replicated data sets (e.g. for use in simulation residuals or as an operating model of a simulation experiment). Using a Poisson distribution to simulate frequencies at age would result in variation in total sample sizes across simulation replicates, and this also occurs for the Tweedie distribution. However, this simulated variation in total sample size may not be realistic because the total number of age observations is often a function of cost, and the total captured in a tow is a function of the density of fish. One response is to re-scale the vector of simulated observations at age so that they add up to the input sample size, such that the total amount of data is constant across simulated observations and the outcomes for each category are still variable. However, such a transformation of the simulated MVTW observations would result in a multivariate observation that departs from the MVTW used to generate the untransformed observations. This ability to estimate but not properly simulate data is analogous to quasi-

likelihood and generalized estimating equations (GEE) methods, which can estimate parameters but not necessarily be used in simulation studies (Fieberg *et al.*, 2009). We therefore recommend further research to simulate data from a MVTW distribution while conditioning upon a fixed value for the sum across bins.

We noted that the MVTW as defined here is a vector of independent Tweedie random variables. For some types of age composition observations that are formed by applying proportions at length to age-length keys, this assumption is not met because applying the same proportions at length to each age induces correlation among the age-specific observations (Miller and Skalski, 2006). To relax this independence in the context of a Tweedie distribution, Furman and Landsman (2010) describe a “common shock model” wherein a common Tweedie distributed random variable is added to a vector of independent Tweedie random variables with different parameters to the former. This is analogous to similar developments of a “multivariate Poisson” distribution, and results in positive correlation among bins. The covariance structure of MVTW random variables was relaxed further by Jorgensen (2013). Future research could explore these extensions to the likelihood proposed here.

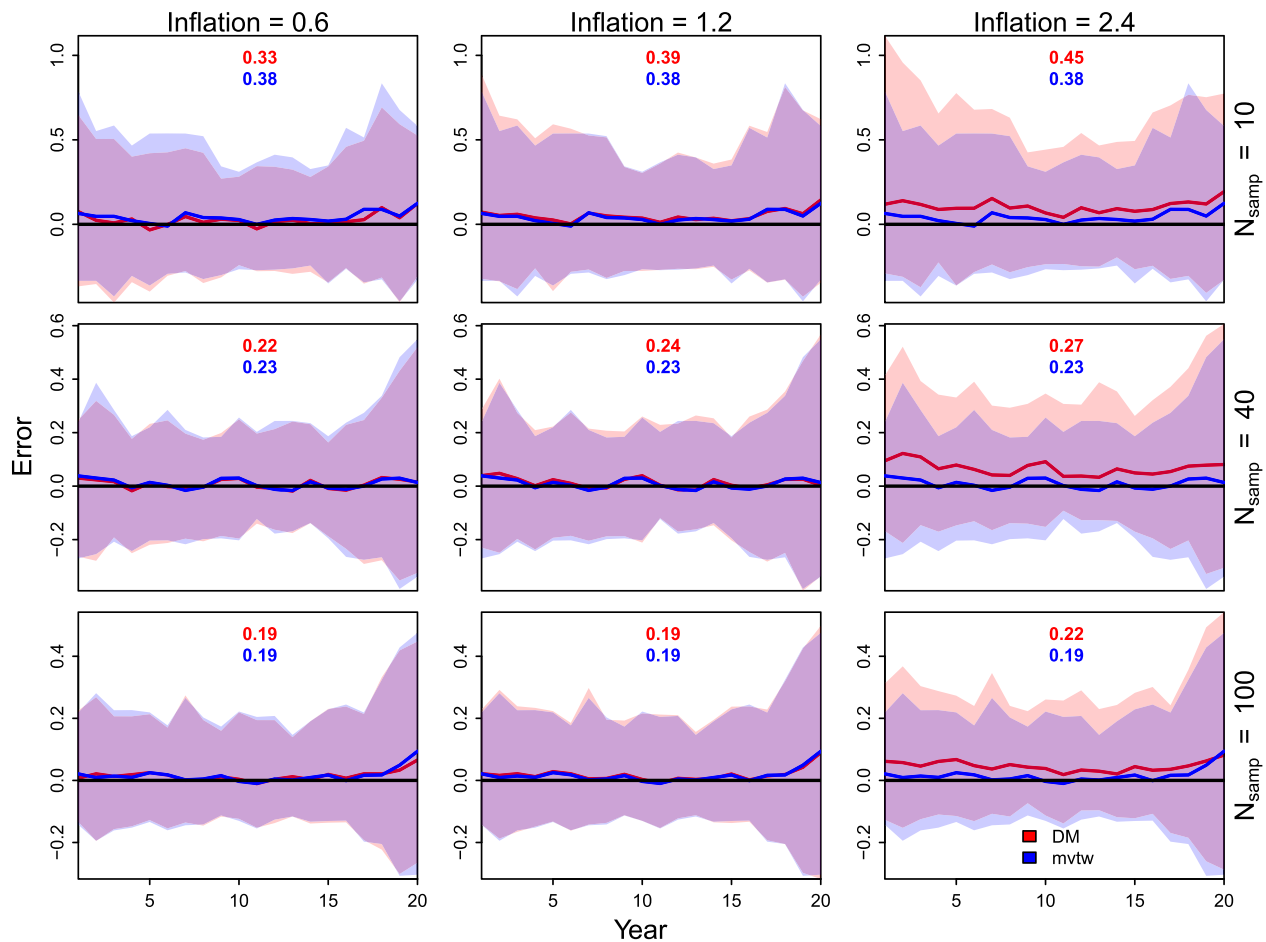


Figure 3. Error in annual recruitment (y-axis, in billions) for each of 20 years (x-axis) in replicates of a simulation experiment, specifically showing the 10% and 90% percentiles (shaded area) and median error (solid lines) for the DM (red) or MVTW (blue) models. The simulation experiment involves a 3-by-3 factorial cross of three true sample sizes (rows) and degrees of inflation (columns), where we also list the root-mean-squared error for all years and replicates (top of each panel). Note that y-axis range is identical across rows of the panel figure, and different among columns.

Similarly, residuals for composition data are often correlated for adjacent years and ages (Francis, 2014, Figure 4; Hrafnkelsson and Stefánsson, 2004, Figure 4). Some research has argued that these correlations can be implicitly addressed by using a metric that incorporates correlations when tuning composition data (Francis, 2011, 2014, 2017). However, composition data often arise from a “fleet” that also provides catch or abundance-index data, and in some cases the correlated residuals might also be informative when interpreting other data from that fleet. For example, analysts may want to condition upon correlated residuals in fishery age-composition data when assigning fishery catch to different ages, or condition upon correlated residuals in survey age-composition when interpreting an abundance index. In these cases, we argue that it is helpful to develop a state-space model that incorporates time-varying selectivity and/or growth. By doing so, random effects are conditioned-upon when interpreting age/length composition data, catches, abundance indices, or other model components, and this ensures that correlated residuals attributed to random effects are used when interpreting all data jointly. Recent research suggests that model-based data weighting can improve estimates of time-varying processes (Xu *et al.*, 2020), but there is relatively little research comparing performance among model-based

methods (although see Fisch *et al.*, (2021)). Similarly, studies show that data-weighting cannot by itself address misspecified selectivity or growth (Punt, 2017; Stewart and Monahan, 2017; Wang and Maunder, 2017), and hence attributing correlated residuals to time-varying processes (rather than the distribution for composition data) could be helpful (although also see Szuwalski (2022) for cautions). We re-iterate Francis (2011, Figure 2) in noting that the estimated magnitude of process errors can be used to diagnose model mis-specification (Thorson and Haltuch, 2018, Figure 7). We therefore recommend future research to investigate which data-weighting method has the best performance when partitioning variance into multiple measurement and time-varying processes, and hypothesize that having multiple admissible likelihoods to select from (e.g. DM, MVTW, and logistic-normal) will likely improve optimal performance in this context. Similarly, we recommend further research regarding model-selection and diagnostics to select between (or combine) results from different data-weighting methods.

There are alternative ways to model age-specific observations in assessment models than we used here. For example, SAM (Nielsen and Berg, 2014) is another state-space age-structured model that treats the numbers at age for surveys as multivariate lognormal observations rather than

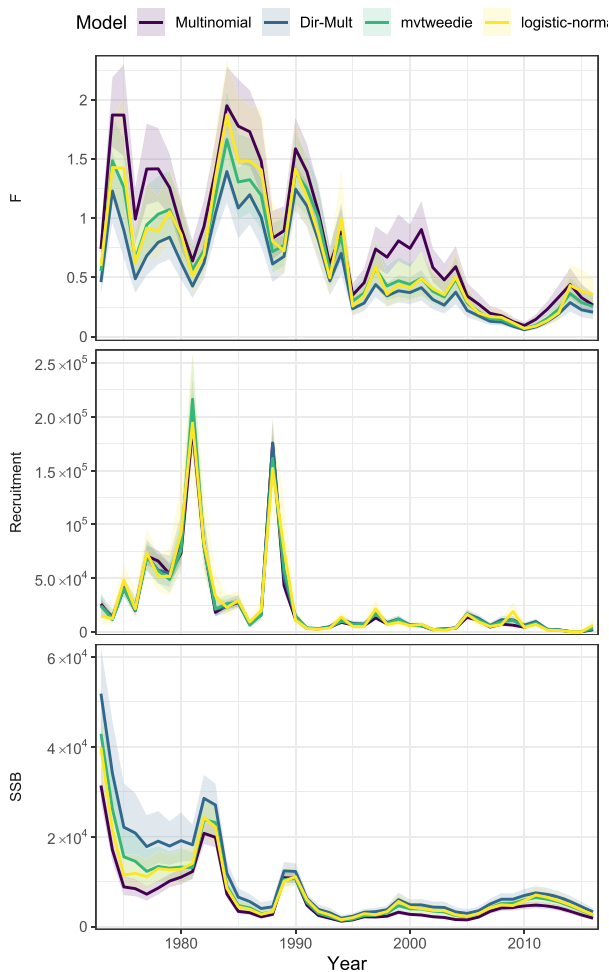


Figure 4. Comparison of estimated fishing mortality rate (top), recruitment (middle), and spawning biomass (bottom) in each year (x-axis) for the yellowtail flounder case study fitted using WHAM using a multinomial, DM, MVTW, or logistic-normal (see top legend for colors), showing the maximum likelihood estimate (solid line) and 95% confidence interval (shaded area).

separately modeling the observed aggregate catch or indices and age composition. This multivariate lognormal distribution can then be inflated relative to an estimate of the multivariate lognormal sampling error that arises from hierarchical expansion of sampling data (Berg and Nielsen, 2016). The MVTW could also be an alternative to the multivariate lognormal for these observations while naturally allowing zero observations that occur for some age classes in stocks with low sample sizes or large maximum ages.

Finally, we also note that DM and MVTW distributions both require an input-sample size n_{input} that is used as starting point during data weighting. By contrast, the alternative logistic-normal likelihood does not use information about input-sample size, which can be a strength (when this information is unavailable) or a weakness (when analysts know that sample sizes have changed over time but cannot supply this information to the model). We showed that the MVTW is insensitive to the value n_{input} , where a change in n_{input} will be exactly offset by a proportional change in variance parameter ϕ such that effective sample size remains unchanged (i.e. the MVTW distribution has the same stated strength as the logistic-normal

likelihood). However, we still emphasize the continued importance of using bootstrapping or model-based estimators to generate n_{input} for two reasons. First, n_{input} may differ substantially among years, for example, when sample sizes or survey designs change over time (O’Leary *et al.*, 2020), and both DM and MVTW would propagate information about well or poorly-sampled years, while still proportionally up- or down-weighting these data across years (thus the MVTW addresses the stated weakness of the logistic-normal likelihood). Second, this n_{input} remains a useful guidepost for model interpretation, that is, where substantial down-weighting can diagnose misspecification of selectivity or other time-varying processes and predict consequences of changing field sample sizes (Thorson *et al.*, 2020). We therefore recommend that continued research regarding age and length-composition likelihoods occur in parallel with real-world efforts to estimate input-sample-sizes directly from field sampling data.

Supplementary Data

Supplementary material is available at the ICESJMS online.

Data availability statement

The data used in the yellowtail flounder case study are distributed with the R package WHAM (<https://github.com/timjmiller/wham>), and the MVTW distribution is available within release 1.0.7. Similarly, the simulation experiment used release 1.3.0 of R package CCSRA (<https://github.com/James-Thorson/CCSRA>).

Author contributions

JT developed the MVTW distribution, derived its analytical properties, and identified its linkage to hierarchical expansion of multistage sampling designs. JT and TM refined its interpretation, parameterization relative to compound-Poisson-gamma process, and identified difficulties with simulation. JT modified package CCSRA to include the MVTW distribution, and implemented the simulation study. TM and BS developed the package WHAM, and JT and TM modified it to include the MVTW distribution. TM and BS developed the case study, and JT, TM, and BS implemented the case study. JT led writing, and all authors contributed to interpretation, writing, and editing.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Acknowledgements

We thank prior discussions with C. Monnahan, J. Sullivan, N. Cadigan, E. Liljestr and, and J. Bence, which contributed to the development of ideas explored here. We continue to thank K. Kristensen, H. Skaug, and other developers of TMB, without which this research would likely be impossible. We also thank J. Ancheta for collaborating to add *dtweedie* to ADMB-13, and I. Stewart, J. Sullivan, and two anonymous reviewers for helpful comments on a previous draft.

References

- Adhikari, A., and Pitman, J. 2021. Probability for Data Science. UC Berkeley. Berkeley, CA.
- Baker, S. G. 1994. The multinomial-poisson transformation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43: 495–504.
- Berg, C. W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES Journal of Marine Science*, 73: 1788–1797.
- Birch, M. W. 1963. Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25: 220–233.
- Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Canadian Journal of Fisheries and Aquatic Sciences*, 73: 296–308.
- Candy, S. G. 2008. Estimation of effective sample size for catch-at-age and catch-at-length data using simulated data from the Dirichlet-multinomial distribution. *CCAMLR Science*, 15: 115–138.
- Cormack, R. M. 1989. Log-Linear models for capture-recapture. *Biometrics*, 45: 395–413.
- Crone, P. R., and 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. 8–11 October.
- Fieberg, J., Rieger, R. H., Zicus, M. C., and Schildcrout, J. S. 2009. Regression modelling of correlated data in ecology: subject-specific and population averaged response patterns. *Journal of Applied Ecology*, 46: 1018–1025.
- Fisch, N., Camp, E., Shertzer, K., and Ahrens, R. 2021. Assessing likelihoods for fitting composition data within stock assessments, with emphasis on different degrees of process and observation error. *Fisheries Research*, 243: 106069.
- Foster, S. D., and Bravington, M. V. 2013. A poisson-gamma model for analysis of ecological non-negative continuous data. *Environmental and Ecological Statistics*, 20: 533–552.
- Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 68: 1124–1138.
- Francis, R. I. C. C. 2014. Replacing the multinomial in stock assessment models: a first step. *Fisheries Research*, 151: 70–84.
- Francis, R. I. C. C. 2017. Quantifying annual variation in catchability for commercial and research fishing. *Fisheries Research*, 192: 5–15.
- Furman, E., and Landsman, Z. 2010. Multivariate Tweedie distributions and some related capital-at-risk analyses. *Insurance: Mathematics and Economics*, 46: 351–361.
- Hjort, J. 1926. Fluctuations in the year classes of important food fishes. *ICES Journal of Marine Science*, 1: 5.
- Hrafnkjellsson, B., and Stefánsson, G. 2004. A model for categorical length data from groundfish surveys. *Canadian Journal of Fisheries and Aquatic Sciences*, 61: 1135–1142.
- Hulson, P. J. F., Hanselman, D. H., and Quinn, T. J. 2012. Determining effective sample size in integrated age-structured assessment models. *ICES Journal of Marine Science: Journal Du Conseil*, 69: 281–292.
- Jørgensen, B. 2013. Construction of multivariate dispersion models. *Brazilian Journal of Probability and Statistics*, 27: 285–309.
- Kendal, W. S. 2004. Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. *Ecological Complexity*, 1: 193–209.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: automatic differentiation and laplace approximation. *Journal of Statistical Software*, 70: 1–21.
- McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling: importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences*, 54: 284–300.
- Maunder, M. N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. *Fisheries Research*, 109: 311–319.
- Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142: 61–74.
- Methot, R. D., and Taylor, I. G. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 68: 1744–1760.
- Methot, R. D., and Wetzel, C. R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, 142: 86–99.
- Miller, T. J., Hare, J. A., and Alade, L. A. 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern new england yellowtail flounder. *Canadian Journal of Fisheries and Aquatic Sciences*, 73: 1261–1270.
- Miller, T. J., and Skalski, J. R. 2006. Integrating design-and model-based inference to estimate length and age composition in north pacific longline catches. *Canadian Journal of Fisheries and Aquatic Sciences*, 63: 1092–1114.
- Miller, T. J., and Stock, B. C. 2020. The Woods Hole Assessment Model (WHAM) (Version 1.0) [Computer software]. <https://timjmillergithub.io/wham/>.
- Nielsen, A., and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, 158: 96–101.
- O'Leary, C. A., Thorson, J. T., Ianelli, J. N., and Kotwicki, S. 2020. Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (*Gadus chalcogrammus*) stock assessment. *Fisheries Oceanography*, 29: 541–557.
- Palmgren, J. 1981. The fisher information matrix for log linear models arguing conditionally on observed explanatory variable. *Biometrika*, 68: 563–566.
- Pennington, M., and Volstad, J. H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics*, 50: 725–732.
- Perreault, A. M., Wheeland, L. J., Morgan, M. J., and Cadigan, N. G. 2020. A state-space stock assessment model for american plaice on the grand bank of Newfoundland. *Journal of Northwest Atlantic Fishery Science*, 51: 45.
- Punt, A. E. 2017. Some insights into data weighting in integrated stock assessments. *Fisheries Research*, 192: 52–65.
- R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rudd, M. B., Cope, J. M., Wetzel, C. R., and Hastie, J. 2021. Catch and length models in the stock synthesis framework: expanded application to data-moderate stocks. *Frontiers in Marine Science*, 8: 1119.
- Schnute, J. T., and Haigh, R. 2007. Compositional analysis of catch curve data, with an application to *Sebastes maliger*. *ICES Journal of Marine Science: Journal Du Conseil*, 64: 218–233.
- Stewart, I. J., and Hamel, O. S. 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences*, 71: 581–588.
- Stewart, I. J., and Monnahan, C. C. 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. *Fisheries Research*, 192: 126–134.
- Stock, B. C., and Miller, T. J. 2021. The woods hole assessment model (WHAM): a general state-space assessment framework that incorporates time-and age-varying processes via random effects and links to environmental covariates. *Fisheries Research*, 240: 105967.
- Szuwalski, C. 2022. Estimating time-variation in confounded processes in population dynamics modeling: a case study for snow crab in the eastern Bering Sea. *Fisheries Research*, 251: 106298.
- Thorson, J. T. 2014. Standardizing compositional data for stock assessment. *ICES Journal of Marine Science: Journal Du Conseil*, 71: 1117–1128.

- Thorson, J. T. 2019. Perspective: let's simplify stock assessment by replacing tuning algorithms with statistics. *Fisheries Research*, 217: 133–139.
- Thorson, J. T., Arimitsu, M. L., Levi, T., and Roffler, G. H. 2022. Diet analysis using generalized linear models derived from foraging processes using r package mvtweedie. *Ecology*, 103: e3637.
- Thorson, J. T., Bryan, M. D., Hulson, P.-J. F., Xu, H., and Punt, A. E. 2020. Simulation testing a new multi-stage process to measure the effect of increased sampling effort on effective sample size for age and length data. *ICES Journal of Marine Science*, 77: 1728–1737.
- Thorson, J. T., and Cope, J. M. 2015. Catch curve stock-reduction analysis: an alternative solution to the catch equations. *Fisheries Research*, 171: 33–41.
- Thorson, J. T., and Haltuch, M. A. 2018. Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 76: 401–414.
- Thorson, J. T., Johnson, K. F., Methot, R. D., and Taylor, I. G. 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. *Fisheries Research*, 192: 84–93.
- Thorson, J. T., and Kristensen, K. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fisheries Research*, 175: 66–74.
- Thorson, J. T., Minto, C., Minto-Vera, C., Kleisner, K., and Longo, K. 2013. A new role of effort dynamics in the theory of harvest populations and data-poor stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 70: 1829–1844.
- Thorson, J. T., Rudd, M. B., and Winker, H. 2018. The case for estimating recruitment variation in data-moderate and data-poor age-structured models. *Fisheries Research*, 217: 87–97.
- Wang, S.-P., and Maunder, M. N. 2017. Is down-weighting composition data adequate for dealing with model misspecification, or do we need to fix the model? *Fisheries Research*, 192: 41–51.
- Winker, H., Carvalho, E., Thorson, J. T., Kell, L. T., Parker, D., Kapur, M., Sharma, R., *et al.* 2020. JABBA-select: incorporating life history and fisheries' selectivity into surplus production models. *Fisheries Research*, 222: 105355.
- Xu, H., Thorson, J. T., and Methot, R. D. 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. *Canadian Journal of Fisheries and Aquatic Sciences*, 77: 247–263.

Handling Editor: Ernesto Jardim