Estimating multinomial effective sample size in catch-at-age and catch-at-size models

Samuel B. Truesdell[1,*] James R. Bence[1] John M. Syslo[1] and Mark P. Ebener[2]

[1]Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, 48824, USA

[2]Chippewa/Ottawa Resource Authority, 179 W. Three Mile Road, Sault Ste. Marie, Michigan, 49783, USA

*Corresponding author:

Samuel Truesdell
Quantitative Fisheries Center
375 Wilson Rd., Room 101
East Lansing, MI 48824
e-mail: truesd16@msu.edu
ph: 517-355-0126

Abstract

Catch-at-age or catch-at-size stock assessment models require specification of an effective sample size (ESS) as a weighting component for multinomial composition data. ESS weights these data relative to other data that are fit, and is not an estimable parameter within a model that uses a multinomial likelihood. The ESS is typically less than the actual sample size (the number of fish) because of factors such as sampling groups of fish (clusters) that are caught together. A common approach for specifying ESS is to iteratively re-fit the model, estimating ESS after each fit so that the standardized residual variance is "correct," until ESS converges. We survey iterative methods for determining ESS for a multinomial likelihood and apply them to two Great Lakes whitefish stocks. We also propose an extension of such methods (the Generalized Mean Approach - GMA) for the case where ESS is based on mean age (or length) to account for correlation structures among proportions. Our extension allows for greater flexibility in the relationship between ESS and sampling intensity. Our results show that the choice of ESS estimation method can impact assessment model results. Simulations (in the absence of correlation structures) showed that all the approaches to calculating effective sample size could provide

reasonable results on average, however methods that estimated annual ESS independently across years were highly imprecise. In our simulations and application, methods that did account for correlation structure in annual proportions produced lower ESS than those that did not and suggested that these methods are adjusting for a deviation from the multinomial correlation structure. We recommend using methods that adjust for correlation structures in the proportions, and either assuming a constant ESS or, when there is substantial inter-annual variation in sampling levels, assuming ESS is related to sampling intensity and using the GMA or a similar approach to estimate that relationship.

1. Introduction

Catch-at-age and catch-at-size models are commonly used tools in stock assessment (e.g., Legault and Restrepo 1998, Methot and Wetzel 2013, Punt et al. 2013). These models use observations of cohorts through time to estimate population parameters. Because cohort size is a fundamental component, an accurate implementation of the relative abundance of age or size classes is critical to model accuracy. In a model's likelihood function, observations of the relative abundance of class size (expressed as proportions) are frequently compared to model-produced estimates during fitting using the multinomial likelihood (Francis 2014). The influence of the proportions-at-age or at-size on the fit of the likelihood function is determined by the multinomial's effective sample size parameter (ESS), which defines the expected amount of variability from a simple random sample of fish ages or sizes (Folmer and Pennington 2000, Methot and Wetzel 2013). Determining ESS is important because this weighting factor can impact the model output quantities used by managers such as population size and fishing mortality rates (Francis 2011).

The observed population composition data may be more variable than or have a correlation structure that differs from that of a multinomial sample of the observed number of fish. Two causes are the spatial behavior of the fish and the spatial grouping of the sampling method (e.g., a trawl catches many fish together). This amounts to cluster samples (Cochran 1977), which carry less information than the number of individuals actually aged or measured (McAllister and Ianelli 1997, Folmer and Pennington 2000, Stewart and Hamel 2014), so ESS is typically smaller than the number of individuals processed. A third cause, applicable to length-structured models, is the potential for large recruitment events to impact multiple adjacent length bins, producing such correlations. Further complicating the issue, age compositions are often calculated based on both a length composition and an age-length key. Due to this complex data structure, ESS cannot be determined directly from the number of fish aged or measured, although in some cases it can be estimated based on sampling theory (e.g., Crone and Sampson 1998, Pennington et al. 2002) or using an approach such as bootstrapping (e.g., Stewart and Hamel 2014); however it has been suggested that these data should not be weighted independently of an assessment model because much of the composition error may result from model process error rather than observation error (Francis 2016). ESS also cannot be included as a parameter in models that use multinomial likelihoods for composition data because it is not estimable in the multinomial likelihood function.

Various methods have been employed for fixing and estimating multinomial ESS (Francis 2011, Maunder 2011), and these include ad-hoc and iterative approaches. To recognize that the information content of the samples is less than the actual number of fish observed, ad-hoc methods may set a fixed ESS (e.g., Fournier and Archibald 1982; Fig. 1A) or treat the annual number of observations as the ESS up to a maximum value, and use this maximum when the number of observations exceeds the threshold (e.g., Fournier et al. 1998, Caroffino and Lenart 2010; Fig. 1B). These ad-hoc approaches can be based on estimation of actual variances in other fisheries if formal sampling designs permit this (Crone and Sampson 1998), informal consideration of the observed variation in age compositions relative to what would be expected from a multinomial, or other forms of professional judgement.

A variety of iterative approaches have been advanced (e.g., McAllister and Ianelli 1997, Francis 2011, Maunder 2011). Francis (2011) argued that decisions regarding weighting (variances) for other data should be made first, followed by tuning ESS using iterative approaches. Most approaches determine how variable data are about the model predictions, relative to how variable they are expected to be given the assumed ESS, and then refit the stock assessment model repeatedly, adjusting the ESS at each iteration to be consistent with the variation seen at the last iteration until ESS is stable.

These iterative methods were classified by Francis (2011) based on whether they accounted for correlation structures or not, and their assumptions about "process error" (which in this case can be viewed as over-dispersion relative to a multinomial distribution based on the number of fish aged or measured). Herein, correlation structure refers to a deviation from the weak negative correlation in proportions between all pairs of bins that arises from the multinomial distribution and the constraint that proportions sum to 1.0. Our expectation is that such structure will generally involve the strongest positive correlations in observed proportions from proximal bins (e.g., ages 5 and 6) with positive correlations weakening and eventually becoming negative between proportions in bins that are farther apart (e.g., 4 and 9). Methods that do not allow for correlation structures generally seek to set ESS to match variation in the proportions at age or length versus what would be expected from a multinomial distribution. This includes McAllister and Ianelli's (1997) commonly used approach (e.g., Wilberg et al. 2005, Campana et al. 2010, Berger et al. 2012). Methods that can account for correlation structure seek to set ESS to match variation in mean age or mean length that would be expected if the composition data arose from a multinomial distribution. As originally implemented by McAllister and Ianelli, their iterative approach calculated an ESS for each year (for a data type), and then averaged these and used the same ESS for each year in the next iteration of the assessment model. Thus they assumed that information content was constant over years and unrelated to any variation in sampling effort (Fig. 1A). Francis (2011) proposed two hypotheses that account for overdispersion, based on the idea that the adjustment of ESS from the number of samples should either be multiplicative or additive. For the multiplicative case, if a particular composition sample was based on $\tilde{N}$ observations, then its information content (ESS) is $\ddot{N} = w\tilde{N}$, where $\ddot{N}$ is the ESS and w is a multiplicative scaling factor (Fig. 1C). For the additive case, $\frac{1}{\ddot{N}} = \frac{1}{\tilde{N}} + \frac{1}{N_{MAX}}$, the information content initially increases directly with sample number but approaches an asymptote, N_{MAX} (Fig. 1D).

The hypothesized direct proportionality between ESS and sampling intensity arising from multiplicative error could apply to other measures of sampling intensity such as number of trips rather than number of fish aged or measured. Iterative methods that do not account for correlation structure and use the observed variation in proportions along with the variability expected in a multinomial sample can at least theoretically calculate an ESS for each year. Maunder (2011) suggested that in such cases rather than using these directly one could fit a statistical model relating these nominal effective sample sizes to observed sampling intensities, and use the predictions from the statistical model as the ESS in the next iteration. This allows for consideration of more general relationships between ESS and sampling intensity than those arising from multiplicative or additive error acting alone. For example one might hypothesize that information content of composition samples increases to an asymptote as a function of the number of trips sampled, rather than number of fish aged, but there would be no reason to assume an initial slope of 1.0. Even when using number of ages or lengths as the predictor, an initial slope of less than 1.0 seems possible, i.e., both multiplicative and additive error could operate together. This approach is not directly applicable to the methods that allow for correlation structures, as only one deviation between observed and expected means is available for each year. Thus, for those methods, Francis (2011) indicated that either w for the multiplicative hypothesis or N_{MAX} for the additive hypothesis is adjusted so the resulting variation is matched exactly for each iteration.

In 1836 Treaty Waters of the Great Lakes, lake whitefish catch-at-age assessments use a multinomial likelihood for age compositions (Ebener et al. 2005, Truesdell and Bence 2016). In these assessments, ESS is set to the actual number aged up to a maximum and to this maximum for higher levels of sampling (one of the ad-hoc methods described above; Fig. 1B). The maximum is set by the professional judgement of individuals conducting the assessment, taking into account factors such as typical coverage and representativeness of the biological sampling (number of trips sampled and seasonal and spatial coverage of the fishery) as well as informal examination of the regularity of age compositions. This study is partially motivated by a desire to evaluate whether these ESS are consistent with those generated by iterative approaches. Two lake whitefish stocks were chosen as examples for this study: the North Huron assessment area consolidated four previous lake whitefish assessment areas: WFH-01, WFH-02, WFH-03 and WFH-04. The WFM-04 whitefish assessment area is in the northeastern part of Lake Michigan. This is referred to here as the Lake Michigan stock area. For details on these areas see MSC (2015).

Specifically, the objectives for this study were to (1) estimate annual effective sample sizes using a range of iterative approaches for the two lake whitefish assessments, and determine how sensitive the assessments are to effective sample sizes, (2) compare results obtained from iterative approaches with those from status quo assessments, and (3) to extend existing iterative methods for determining effective sample sizes so that a more flexible asymptotic relationship between ESS and level of sampling could be estimated statistically for the case that allows for correlation structure. This last objective was motivated by the observation that methods that do not account for correlation structures may often overestimate the information content of the data (Maunder 2011, Francis 2011), but the multiplicative and additive relationships may be too restrictive to capture how information content varies with actual sampling intensity.

2. Methods

2.1 Methods for estimating effective sample size

We considered a range of iterative approaches for calculating effective sample sizes for use in age- or size-based assessment models. These models use annual age or size compositions where each age or size group (hereafter "bin") is a proportion and each set of annual proportions sum to 1. The annual proportions are assumed by the assessment model to behave as though they arose from a multinomial sample of given size (the ESS). These approaches were tested against simulated data and also applied to two example lake whitefish assessments from 1836 Treaty ceded waters of the North American Laurentian Great Lakes.

2.1.1 Iterative approaches

The basic iterative approach requires initial specification of ESS for each year and data type (e.g., type of fishery or survey) for which composition data are available. For simplicity we have dropped a subscript for data type in our equations, but in our examples we have applied them separately by data type (in these cases trap net and gillnet fishery age compositions). These initially specified effective sample sizes are identical to those used in the actual assessments and are used in the iteration 0 stock assessment. Results from the assessment then are used to evaluate how much the observed proportions (or annual summary of proportions, such as the mean age or mean length) deviate from the predictions of proportions for each bin (or predicted annual summaries). New effective sample sizes are then calculated using this comparison such that the ESS from a multinomial sample would produce the observed amount of deviation from the measured values. These generated effective sample sizes are then used in iteration 1 of the stock assessment model. The steps of (i) evaluation of deviation between observed and predicted values from the assessment, and (ii) adjustment of effective sample sizes to be consistent with this variation, are then repeated until effective sample sizes converge (Fig. 2A). For the purposes of this paper, convergence was defined as a maximum difference in estimated ESS from iteration t to iteration $t + 1$ (over years for all data types) of less than five, and iterations continued until convergence was achieved or a maximum of 25 iterations were completed.

The different iterative approaches we considered are described in detail below, classified by whether they account for the possibility of correlation structures in proportions among bins, and any assumed relationship between ESS and a measure of sampling intensity (e.g., actual number of fish aged or number of trips sampled). Francis (2011) proposed an ESS calculation based on variation in annual mean length or age, rather than the variation among individual bins, to account for correlation structure. These are the methods we refer to as accounting for correlation structure. Symbols associated with all methods are given in Table 1.

2.1.2 Methods that do not account for correlation structures.

Equations for determining the ESS used in each year ($\ddot{N}_y$) for these methods are given in Table 2. The naming conventions for the methods that follow are based on Tables 2 and 3 (see Table 2 caption for an example). Here we consider variations of three basic approaches. The first approach (A) corresponds to the method originally proposed by McAllister and Ianelli (1997), and adapted by Francis (2011) as his method TA1.1. The second (B) and third approaches (C) were presented by Francis (2011) as methods TA1.2 and TA1.3, respectively (see his Appendix Table 1). These basic approaches can be applied using several different sub-approaches: (i) unconstrained year-specific values, (ii) constant values unchanged over years based either on (a) a geometric average of the year-specific values or (b) values using equations from iii, but with input sampling intensity specified as identical in each year, (iii) values directly proportional to sampling intensity (e.g., number of fish aged or number of trips sampled), or (iv) values following an asymptotic relationship with sampling intensity, where the asymptotic model's parameters are estimated based on the relationship between the unconstrained year-specific values and sampling intensity. In this last sub-approach, the parameters are estimated based on a nonlinear regression of the log of unconstrained year-specific estimates for that iteration ($\widehat{N}_y$) versus sampling intensity ($\widetilde{N}_y$). We suggest applying the regression approach using an asymptotic function (Table 2), and use that function in our applications, but note the basic approach is more general. In preliminary work we found that some individual unconstrained year-specific values could converge on unreasonably high ESS. We therefore specified a year-specific maximum value (the actual number of fish aged for that year (for each data type) in our application) when determining the final ESS for the unconstrained year-specific approaches (Table 2). Approaches B and C, as originally put forward by Francis (2011), provided a single proportionality constant (w) between ESS and sampling intensity. Our unconstrained year-specific calculations of ESS, based on these approaches, simply applies those methods separately by year and algebraically re-expresses the result in terms of ESS instead of Francis' w. Sub-approaches ii-iv often involved estimation of initial year-specific ESS as for sub-approach i, and then further processing of these estimates to obtain the ESS used in the next iteration of the assessment model ($\ddot{N}$). In our applications, in all cases where year-specific ESS were used in calculations, the initially calculated unconstrained year-specific ESS were reduced to the actual number of fish aged if the calculated ESS exceeded the number aged in that year. We used this constraint because the effective sample size would generally not be greater than the actual number of fish sampled (i.e., realistic situations where compositional data would be under-dispersed are hard to contrive). This change was made prior to any additional processing or use in the assessment and also applied when the number of trips were used as $\widetilde{N}$. The methods B.iii and C.iii did not first involve calculation of unconstrained year-specific ESS, but instead the calculation of a single weight, w, which was then used to calculate year-specific values proportional to sampling intensity.

2.1.3 Methods that did account for correlation structures.

In cases with correlation structure, a given ESS might produce variation between observed and expected proportions that is consistent with what would be expected from a multinomial distribution with that ESS, but variation between the observed and predicted mean age or length that is inconsistent with what would be expected from that multinomial distribution. In such cases it has been argued that matching the variation in the means more properly acknowledges the information content of the data. The methods in this section are based on this principle. Each also assumes that the ESS in a given year will be function of an input value $\widetilde{N}_y$, which is a measure of sampling intensity (in our applications we use number of fish aged or the number of sampling trips contributing to the age composition sample).

Equations used to determine ESS used for each year ($\ddot{N}_y$) for these methods are given in Table 3. Each bin has an age or length (e.g., at the midpoint length for the bin) associated with it and thus a mean observed age or length can be calculated. The variance of such means can be determined based on the age or size distribution and the ESS. In particular, the variance for the mean of x (where x is age or length) for a given year is given by $v_y/\ddot{N}_y$, where $v_y = \sum_b x_b^2 E_{by} - \bar{E}_y^2$, and $\bar{E}_y = \sum_b x_b E_{by}$.

For method (D) we treat ESS as directly proportional to sampling intensity and use the estimator proposed by Francis, corresponding to his multiplicative error case. For method (E) we assume that at very low sampling levels ESS increases directly (with slope 1) with sampling intensity but eventually approaches a maximum ESS for high sampling levels, corresponding to Francis' additive error case. For our last method (F) we generalize the first two approaches and allow ESS to increase to an asymptote, but do not restrict the value for the slope at the origin to be 1, as in method E. Approaches D and E rely on the over-year variance of standardized deviations between observed and predicted values equaling 1.0. Given there is just one variance, ESS can be calculated to make the variance match this criterion exactly by either altering the slope or asymptote. In the generalized approach (F) we consider annual individual standardized deviations of mean age (or length); $(\bar{O}_y - \bar{E}_y)^2/v_y/\hat{N}_y)$ – note that asymptotically these have a standard (i.e., variance of 1.0) normal distribution (due to the central limit theorem) and that the square of a standard normal distribution is χ^2 distributed with one degree of freedom. The asymptote is estimated by minimizing the sum of the log of the χ^2 densities of the squared standardized deviations (i.e., the log likelihood). Although Francis' algebraic and our statistical approach should perform similarly, using a statistical model can reduce the impact of outliers on the asymptote estimate, as long as the errors are in fact χ^2 distributed as assumed by the model. We attempted to simultaneously estimate both the slope and the asymptote but found that the data for both gear types in both the North Huron and Lake Michigan assessments did not provide sufficient contrast to do so. Still, there need not be an expectation for an origin slope (α) of 1.0. To incorporate this observation we specified various values for α and recorded their impact on the standardized deviations for each data set in each assessment. We chose the most appropriate value for α qualitatively by considering both the variance and graphical depictions of the distribution of standardized deviations. As indicated above these should have variance of 1.0 and approximate a normal distribution. In these qualitative analyses the slope at the origin (α) for the gillnet and trap net fisheries varied together, i.e. we did not test all combinations of gillnet slope with each potential value for trap net slope.

2.2 Catch-at-age Model

In 1836 treaty waters of the North American Laurentian Great Lakes, statistical catch-at-age models are used in lake whitefish stock management. The assessments are based entirely on fishery-dependent data and typically model both a trap net and gillnet fishery. There are a total of 13 such age-structured assessments applied on a regular basis in this region. Here we illustrate two examples, one for the northern Lake Huron assessment area and the second for the WFM-04 assessment area of Lake Michigan. We started with the most recently fit models used for making harvest recommendations in October 2015. These models spanned the years 1976-2014 and 1981-2014, and recognized ages 4 and 3 (age of recruitment) to 12 and 16 (an accumulating age including that age and all older ages) for the North Huron and WFM-04 areas respectively. The models were coded in AD Model Builder (ADMB; Fournier et al. 2012). In both units we had available the number of aged fish as a measure of sampling intensity. In the North Huron area we also had access to the number of sampled trips. The fisheries and their management were described in detail by Ebener et al. (2005) and details of the assessment models were reported by Truesdell and Bence (2016). The model components directly related to ESS are described here, and additional details are reported in Appendix 1.

The proportions-at-age come directly from annual age sampling and were not inferred from an age-length key or a growth model. Observed and predicted ages and annual ESS were incorporated in the likelihood function using the multinomial log-likelihood: $L_M = \sum_{y=1}^{Y} N_{E,y} \sum_{a=1}^{A}\left[p_{y,a} \log\left(\hat{p}_{y,a}\right)\right]$ where L_M is the multinomial log-likelihood component, $N_{E,y}$ is the ESS in year y, Y is the number of years, a is the age-class index and A is the number of age classes, and $p_{y,a}$ and $\hat{p}_{y,a}$ are, respectively, the observed and predicted annual proportions-at-age.

We elected to fix the variances for each normally distributed data type or penalty in the objective function (see Appendix 1) at their final estimated values in the original assessment fit as we explored alternative approaches to estimating ESS. We followed this approach to be consistent with the suggestion of Francis (2011), who suggested they be fixed and that the weighting of age compositions occur in a second stage.

The ESS ($N_{E,y}$) in the baseline 2014 models were set to the actual number of fish aged up to a specified maximum value of 100. In years that the fisheries operated, gillnet sample size was always greater than 100 fish ages in both North Huron and the Lake Michigan area. Trap net sample size was typically greater than 100 fish ages, however in North Huron only 46 fish ages were available one year and in the Lake Michigan area there were four instances of less than 100 ages (see Table 4).

2.3 Sensitivity of model-estimated quantities to ESS

We examined the sensitivity of model-estimated quantities to ESS before moving on to our ESS estimation methods. We did this by systematically varying the maximum trap net and gillnet ESS within the assessment models in Northern Huron and Lake Michigan (i.e., the maximum in Fig. 1B). The scale of the variances for the non-multinomial likelihood components were fixed during these simulations at the values estimated in the two base models, but this was the only parameter that was fixed across these analyses. An R program (R Core Team 2015) was designed to update this maximum in the ADMB data file at each iteration, and this program was linked to the model executable. The maximum ESS was varied in each fishery between 4 and 400 at a resolution of 3 for ESS < 30 and at a resolution of 20 for ESS > 30. The comparison of results for models of varying ESS was made using the average fishing mortality for ages 10-12 over the last 10 model years. We also tabulated the sum of the negative penalized log likelihood (NPLL) for all components excluding the age compositions. We did this to illustrate how the overall model fit, exclusive of the age composition log likelihood, depended on the assigned maximum effective sample sizes.

2.4 Validating ESS estimation and performance of the methods

When data are generated from a multinomial distribution these methods should, at least on average, be able to reproduce the actual multinomial sample size. We evaluated this by considering the case where the actual proportions in each year were known. Thus this evaluation did not involve fitting a stock-assessment model, nor iterative adjustment of ESS, but instead a single application of the equations in Tables 2 and 3 to simulated data. The the E_{by} and $\bar{E}_y$ in Tables 2 and 3 were known and the equations in those tables were applied once for each simulated dataset. While knowing the proportions is not realistic, this procedure provides an upper bound on how well these methods can perform in recovering the true underlying sample sizes assuming the data are multinomial in nature.

To reflect variability in annual sampling effort, the true ESS was varied over both 25 and 100 years. The sampling intensity was assumed to come from a truncated normal distribution $\widetilde{N}^{N}_{TRUE,y} \sim N(\mu, \mathrm{cv}^2\mu^2, \tau)$ where the mean, μ, was 100, the CV was 1.8, and the minimum value, τ, was 10 to prevent unrealistically small numbers of samples. The CV used in the normal distribution was the average CV from gillnet and trap net number of trips and number of samples in North Huron. ESS was assumed to follow the asymptotic relationship to number of samples as used in the regression methods (Table 2, column iv) with asymptote of 125 and slope at the origin of 1.0. Given the truncated distribution and asymptotic relationship, the CV (among years) in true ESS was 0.39. A vector of 9 probabilities (p), summing to 1 represented the true proportions in each age (or length) bin. In each year of the simulation these probabilities were drawn randomly from the set of trap net and gillnet proportions-at-age from the North Huron data set where all age classes had proportions greater than zero. In each year a random multinomial vector of counts (O_y) was generated $O_y \sim M(p, E^N_{TRUE,y})$, and these counts were converted to observed annual proportions in each bin. The ESS was then estimated using the methods described above. The ESS estimates were then compared to the known values ($E^N_{TRUE,y}$) by subtracting the known ESS from the ESS estimates. This was repeated 1000 times. Some results were excluded from the analysis: in 12% of cases method E did not produce an estimated ESS (i.e., the ESS equation did not have a solution within the wide range of potential values we searched) and in < 2% of cases for model B.iv the nonlinear regression that determined ESS estimates did not converge.

We also assessed the performance of the methods we tested for datasets with or without correlation structure. To do this, we simulated data from the logistic-normal distribution 1000 times. To derive these values we first drew year-specific values from the multivariate normal $O^*_y \sim N(E, C)$, with age-specific elements O^*_{ay}, where N denotes the multivariate normal, C the variance-covariance matrix and E the mean. The elements of C were consistent with an AR(1) structure, with adjacent ages having the highest correlation. Each element of the observed annual proportions O_y , O_{ay}, were obtained from the multivariate normal elements by $O_{ay} = \exp(O^*_{ay}) / \sum_i \exp(O^*_{iy})$, resulting in values between 0 and 1.0. The elements of E , E_a, were set equal to the log of the elements of *p* (the expected proportions used in simulations from the multinomial distribution above), so on average the simulated proportions were close to the expected proportions used in those simulations. The variance-covariance matrix C was parameterized by the variance (assumed equal among ages and set to 0.25) and the correlation between adjacent ages (ρ set to either 0 [no correlation structure] or 0.5 [correlation structure]). With the logistic-normal distribution one can view the proportions as arising from relative abundance indices that are multivariate lognormal, with the same CV across ages (approximately 0.5 in this case). While the CV used in these simulations is somewhat arbitrary, we found qualitatively similar patterns as those we present using alternative values. See Schnute and Haigh (2007) and Francis (2014, Appendix A) for more information about the logistic-normal distribution and use of the AR(1) structure to represent age-compositions with correlation structure. The number of samples $\widetilde{N}$ is used in the ESS estimation methods, and these were generated from a truncated normal distribution, following the same procedure as in the simulations from the multinomial simulation. Because here we generated proportions at age from the same distribution each year, there was no relationship between simulated sampling intensity and the information content of the age-compositions, and this would likely disadvantage approaches to estimating ESS that assumed there was a relationship. This made estimating two parameters for model B.iv unrealistic so the slope (α) was fixed at 1.0. Each of the approaches was applied to a 25-year data set of simulated proportions-at-age to estimate ESS.

For the logistic-normal simulations, unlike for the multinomial simulations, the true effective sample size is not known, so it is not possible to formally assess bias. The methods that attempt to address correlation structure, however, are based on the idea that ESS should be set so that a multinomial distribution with a particular sample size would have the right variance in average age. When using the multivariate logistic-normal distribution with specific parameters to generate composition samples, we found the true value for this variance by simulating 10,000 age composition samples that were multivariate logistic-normal samples, calculating the average age for each sample, and then the among sample variance in these averages. The sample size that would produce this variance in mean age for multinomial samples was then determined from the analytic relationship between sample size and variance in average age (see section 2.1.3). In at least one sense this is a value the estimated ESS values should match.

2.5 Application

We next applied each of the iterative methods to both of the assessment models, and summarized results in terms of estimates of ESS and fishing mortality for ages 10-12 (fully selected or nearly fully selected in both areas for both trap nets and gillnets) in the last 10 years of the assessment. For the North Huron assessment we used both number of aged fish and number of sampled trips as our $\widetilde{N}_y$ in different trials. Our baseline evaluation used the ESS ($\ddot{N}_y$) that were assumed in the original assessments in the iteration 0 assessment. The α levels we used (for method F) were: 0.75 for both the North Huron trap net and gillnet fisheries using number of fish sampled as $\widetilde{N}_y$, 5 and 65 for the north Huron trap net and gillnet fisheries using trips as $\widetilde{N}_y$ and 1.0 for both the Lake Michigan trap net and gillnet fisheries using fish as $\widetilde{N}_y$. We performed a simple analysis to verify that different starting values produced the same estimates for ESS and in most cases they were consistent (save some instances when using methods E and F (Appendix 2).

3 Results

3.1 Sensitivity of model-estimated quantities to ESS

In the North Huron analysis, the average *F* over the last 10 model years was generally < 0.15 except in some cases where both trap net and gillnet ESS were greater than about 150 (Fig. 3). In the Lake Michigan assessment, the average *F* generally decreased with increasing gillnet maximum ESS, and increased as trap net maximum ESS increased. This inverse relationship was case-specific as the same was not true in North Huron. The most variability occurred when either trap net or gillnet maximum ESS were low.

In both assessment models the NPLL decreased with increased trap net and/or gillnet ESS (Fig. 3) because when the model weighted the age compositions heavily less relative weight was assigned to the other likelihood components. The North Huron model was more robust to combinations of ESS – in the Lake Michigan assessment the NPLL was still relatively low if either trap net or gillnet ESS was increased but when they increased together the fit to the non-composition data became poorer more quickly than in North Huron.

3.2 Validating ESS Estimation

The two quantities of interest when evaluating the performance of these methods by estimating a known ESS from simulated multinomial data are (1) bias and (2) precision. All methods produced negligible bias (Fig. 4). Methods A.ii.b and B.ii.b were biased slightly low, but the average bias (< 10) was small relative to the true mean ESS of 100 and especially relative to the noise generated by many methods; for example the 90th percentile for bias in methods A.i, B.i and C.i was > 70. Our results demonstrate that the unconstrained annual estimates are imprecise and provide little information regarding the year-specific true ESS (i.e, the variation in estimated ESS was much greater than variation in true ESS (CV of 0.83 to 1.27 rather than CV of 0.39). Thus their use may be problematic, as they might typically provide more noise than information about year-specific information in age compositions. The other potential methods that could not account for correlation structure (A.ii.a, B.ii.a, C.ii.a, A.ii.b and B.ii.b and B.iv) summarized the annual estimates in some way (e.g., via a geometric mean or a regression model). Method B.iv performed the best in terms of both accuracy and precision under these simulated conditions. All the methods that can incorporate correlation structure (D-F) performed well in terms of bias, though method D was typically more precise. Methods E and F performed similarly and had more outliers than the variants of methods A, B and C which included some kind of summarization (i.e., were not A.i, B.i or C.i). The methods that accounted for correlation structures (D, E and F) performed better when 100 years of data were used instead of 25 because there were a greater number of observations (years) available.

When data were simulated using the logistic-normal distribution with no assumed correlation, the methods that can account for correlations (D, E and F) performed similarly to the methods that cannot account for correlations in terms of their average value (Fig. 5). Methods A.i, B.i and C.i again produced some unrealistically high estimates, as did some of the estimates from methods D, E and F (though to a lesser extent). When the data were simulated using a correlation of 0.5, methods D, E and F estimated smaller effective sample sizes than the methods that cannot account for correlations. The sample size that produced variances in mean age for samples from a multinomial distribution that matched the variances in mean age for samples from the multivariate logistic-normal distributions we used were 44 and 29, for $\rho=0$ and $\rho=0.5$, respectively. We found that the methods that accounted for correlation structure changed ESS, on average, roughly in accord with these values, whereas methods that did not account for correlation structure showed no such change (Figure 5).

In summary, the results of the multinomial and logistic-normal simulations – under ideal conditions that are unlikely to be replicated in real-world scenarios – show that (1) unconstrained year-specific values are noisy and will often reflect sampling error rather than true variability in ESS; (2) all of the methods can perform well in terms of bias; (3) methods that can incorporate correlation structures have a more pronounced increase in precision with a longer time series and (4) when correlations are actually present methods that use the mean length or age (D, E and F) tend to produce smaller ESS estimates than methods that cannot account for correlations in the data.

3.3 ESS estimates

Methods A.i, B.i, and C.i resulted in annual estimates of ESS that varied widely and were unrealistically high in some years for both gear types in North Huron (Fig. 6) and Lake Michigan (Fig. 7). Computing a summary statistic among years for these methods (methods A.ii.a, B.ii.a and C.ii.a) reduced the ESS estimates to values more consistent with levels typically used in stock assessment models.

Methods that assume ESS to be proportional to annual sample size (methods A.iii, B.iii and C.iii) also reduced the range of ESS estimates relative to methods that freely estimated annual values (methods A.i, B.i and C.i). Methods A.iii where $\widetilde{N}$ is the number of samples and A.iii where $\widetilde{N}$ is the number of trips resulted in larger estimates of ESS than the corresponding methods B.iii and C.iii for the trap net fishery in the North Huron model, but resulted in similar estimates for the gillnet fishery.

Methods B.iv using fish as $\widetilde{N}$ and B.iv using trips as $\widetilde{N}$ (in North Huron) incorporated a regression of the predicted ESS from method B.i against the actual sample size used in nonlinear models to predict N_y (Fig. 8). These methods produced estimates for ESS that spanned a similar range to methods A.ii, B.ii, Cii and A.iii, B.iii, and Ciii (Figs. 6 and 7). The models for trap nets and gillnets that were based on the number of fish aged all had origin slopes that were larger than 1.0, though in North Huron these were larger by <0.1 (Table 5). The asymptotes for these models ranged from 94.6 to 324. The origin slope for the North Huron trap net fishery that used a model based on number of trips was approximately 1000, and consequently there was little relationship between estimated ESS and number of trips for North Huron trap nets, and thus essentially an average value (the asymptote, 334) was used in all years. The origin slope for the North Huron gillnet fishery was 5.70.

For both the North Huron and Lake Michigan models, approaches accounting for correlation structures (methods D, E and F) generally produced smaller ESS estimates than methods that did not, with method D generally producing slightly larger values and a larger range in North Huron (Figs. 6 and 7). Methods E and F generally produced the smallest ESS. Across all years and all methods that were tested, the median ESS for the methods that could incorporate correlation structure were always smaller than the median for the methods that could not incorporate correlation structure (Table 5).

For the generalized mean approach we chose values for α according to the variance of the standardized residuals and the appearance of the relationship between the standardized residuals and the number of fish or number of trips (e.g., Fig. 9). The variances that were nearest to 1.0 produced plots that looked closest to standard normal, though in some cases a range of αs produced plots that were nearly indistinguishable. The αs that were chosen and the subsequent maximum likelihood estimates for the asymptotes (N_{MAX}) were: 0.75 and 32 (North Huron trap net using fish), 0.75 and 132 (North Huron gillnet using fish), 5 and 96 (North Huron trap net using trips), 65 and 140 (North Huron gillnet using trips), 1 and 64 (Michigan trap net using fish), and 1 and 11 (Michigan gillnet using fish).

The most relevant findings from our applications of these methods to the two Great Lakes stocks were (1) the unconstrained year-specific approaches produced very noisy ESS estimates; (2) various methods were similarly successful at reducing the noise by summarizing the unconstrained estimates in different ways; and (3) the methods that allowed for correlation structures in the composition data produced lower estimates of ESS.

3.4 Fishing mortality estimates

The North Huron and Lake Michigan assessment models responded differently to the variability in ESS estimates. For the North Huron model, most average *F* estimates were between 0.1 and 0.11 (Fig. 10A). For the Lake Michigan model, estimates of *F* were more variable among ESS estimation methods, ranging from 0.17 to 0.36. The methods accounting for correlation structure produced estimates of average *F* > 0.3 while most estimates not accounting for correlation structure produced estimates <0.3 (Fig. 10B). Despite large differences in estimated ESS in North Huron (Fig. 6) the different methods corresponded with similar average *F* (Fig. 10A). However, in Lake Michigan the variability in ESS estimates (Fig. 7) caused measurable differences in the *F* estimates (Fig. 10B).

4 Discussion

The ESS used in a catch-at-age or catch-at-size model's objective function (what is minimized when fitting the model) can have considerable impact on the estimates for stock quantities important to sustainable management, such as fishing mortality. When ESS is high the model is forced to fit proportions-at-age (or at-size) closely; conversely when ESS is relatively low the model provides a better fit to other quantities such as the total catch or a survey CPUE. The alternative ESS weighting changes the likelihood surface and impacts the estimates, most notably in cases where the composition and other data sources are in conflict. Even if all data sources are in agreement (i.e., there is zero process error), mis-weighting the composition data may reduce the precision of stock quantity estimates, even if they remain unbiased (Maunder 2011).

The impact of ESS on model output in applications to actual fish populations can be substantially greater than is suggested by simulations where composition and other data sources are not in conflict. The Lake Huron and Lake Michigan examples had very different relationships between ESS (for both gears) and estimated fishing mortality, but both demonstrate the importance of assigning accurate weights to compositional data. Mis-specifying the ESS can clearly change estimates of quantities such as fishing mortality which will affect stock management strategies. These findings re-emphasize that the sensitivity of models to ESS should be tested as part of assessment model diagnostics (Brodziak 2002, Maunder 2003).

Year-specific unconstrained annual ESS were imprecise in our multinomial simulations (Fig. 4), and when used in our applications led to some years with unrealistically high ESS – a result alluded to by de Moore et al. (2012). Hulson et al. (2012) reported more frequent and consistent stock assessment errors in a simulation study when annual rather than mean ESS was used (estimated as parameters using the Dirichlet distribution) and suggested that this result might indicate an overparameterization issue when using annual ESS. One way to reduce the amount of noise is to combine information from different years. Using all the years of data to estimate a single ESS to apply to each year is one such approach, and we illustrated a variety of ways of doing this based on methods presented by McAllister and Ianelli (1997) and Francis (2011).

Another approach to combining information across years is to assume a relationship between ESS and sampling intensity. Maunder (2011) evaluated using a zero intercept linear regression between unconstrained annual values and sampling intensity, and suggested that this approach could be generalized by using an asymptotic relationship. We applied this more general regression approach and also assumed an asymptotic relationship in our generalized mean approach. Francis (2011) also presented non-regression approaches that assumed either direct proportionality or a specific asymptotic relationship. We found in our simulations that all these approaches did increase precision of ESS estimates relative to the unconstrained values. When information content varies, use of unconstrained estimates is not the only choice and we strongly recommend using one of the other approaches that estimate ESS as a function of annual sampling intensity.

When estimating a functional relationship between ESS and sampling intensity, we sometimes estimated high origin slopes (greater than 1.0 when using fish and greater than the average number of fish per sampling event when using trips), which has the potential to produce ESS that is consistently larger than the actual number of fish sampled in years with low sampling intensity. This could be avoided, if considered undesirable, by restricting the slope near the origin to values less than 1.0 (or some other reasonable value if sampling intensity is measured in units other than numbers of fish).

The methods we tested that allowed correlation structures (i.e., D-F), tended to estimate smaller ESS for real data in our applications, and were not biased when data were actually multinomial in simulations. In addition when we introduced positive correlation between adjacent ages in simulations, these methods estimated lower ESS, which was not the case for methods that did not account for correlation structure. These results suggest that in our applications correlation structures in the composition data were large enough to substantially influence information content for both stocks, and were likely a function of the clustered samples or model inadequacies that cause such correlations. This finding agrees with other studies: Francis (2014) considered composition data from 28 stock assessments, and found evidence for correlation structures in most of them, and other studies also suggest that such correlation structures may be common (Pennington and Vølstad 1994, Hulson 2011, Maunder 2011).

Despite widespread use of the multinomial likelihood for describing composition data in catch-at-age or catch-at-size models, the application of this distribution in these likelihood functions has two important limitations (Francis 2014). First it is not possible to estimate ESS within the model, which is why iterative approaches such as those discussed here are necessary. Second, the multinomial distribution cannot directly account for correlation structures within the composition data. An alternative to methods that account for correlation structures by using mean age or mean length is to replace the multinomial with a distribution capable of incorporating correlation structure and that has a weighting that can be estimated within the model. A number of distributions with estimable weights have been suggested and evaluated (e.g., Maunder 2011, Hulson et al. 2012). Francis (2014) argued for the consideration of the logistic-normal, which he considered promising because the data are restricted to the 0-1 range and because it can account for correlation structures. Maunder (2011) evaluated distributions with estimable weights (but that could not incorporate correlation structures) and found that precision of stock estimates was degraded due to some types of process errors that produce correlation structures.

This supports Francis' (2014) argument that the most promising distributions with internally estimated weights are those allowing for correlation structure.

A different approach to accounting for correlation structures in compositions is to model the catch-at-age using multivariate distributions (such as the multivariate lognormal) rather than treating the composition and total catch data separately (Myers and Cadigan 1995). Use of the catch-at-age data, rather than proportions and totals, underpinned early age-structured assessments but was largely dropped because of the unrealistic variance-covariance structure, although advances in statistical modeling have made the specification of more realistic variance-covariance structure feasible (Berg and Nielsen 2016). Fournier and Archibald (1982) argued for separate treatment of total catch and composition data in part because these data often arise from separate data collection efforts.

There are hurdles to overcome in generally applying the logistic-normal or other distributional approaches. One is to identify appropriate correlation structures, and another is how to handle zero values for observed proportions, which the logistic-normal does not allow. Similar issues apply when modeling catch-at-age directly. Francis (2014) suggested initial approaches to both. With regard to correlation structures he considered both AR(1) and AR(2) correlation models among bins, but identified the sex-specific case as remaining problematic. With regard to zero values he suggested compression of composition data (aggregating the bins for youngest/smallest and oldest/largest) to reduce the number of zeros, combined with adding a small constant to proportions. He took the view that zeros are more a problem to deal with than a phenomenon to model, although he did indicate that an alternative approach would be to consider compound distributions (e.g., logistic-normal combined with multinomial) to allow for zeros. Thus the use of distributions that allow for correlation structures is promising, although more work is needed to develop them and evaluate alternative approaches to their implementation. In the meantime, there is no doubt that multinomial distributions will remain a widely used approach.

Another alternative to the iterative approaches is to fix the ESS a priori. One approach to this is to follow general guidance such as use of actual number of fish contributing to the composition up to a maximum (Fournier et al. 1998) or the square-root of sample size (Thompson 1995, cited by Hulson 2012). An alternative is to base the ESS on the analysis of data outside the assessment such as by using survey design theory (e.g., Pennington et al. 2002). The adequacy of the first of these approaches depends entirely on how well the ESS from the general guidance reflects the information content in the specific composition data. The Pennington et al. (2002) survey design estimator (PSDE) has promise as a way to define ESS a priori as it is based on the actual data used to generate compositions and accounts for correlation structures by use of means. Although this estimator has largely been used in sampling design applications, Hulson et al. (2011) recognized its potential use in stock assessments. Hulson et al. (2011) used a detailed process model (e.g., incorporating age-related schooling and depth distributions, cluster sampling, and aging error) to simulate age compositions and, as in our simulations, applied ESS estimators outside a stock-assessment model. Hulson et al. (2011) found two estimators that did not account for correlation structure (the unconstrained annual estimator (A.i) and a maximum likelihood estimator using the Dirichlet distribution) produced similar estimates of ESS on average, whereas the PSDE sometimes produced quite divergent average ESS results. At least some of these differences could be due to the type of correlation structure present. For example, Hulson et al. (2011) found higher average ESS from the PSDE than other estimators when they assumed single ages schooled together, in contrast to mixed age schools, where the average from PSDE was lower than for the other estimators. This is consistent with the PSDE adjusting for correlation structure, given that single-aged schools would be expected to cause greater than anticipated negative correlations in proportions (compared to

Dirichlet/Multinomial), whereas mixed age schools would be expected to produce positive correlations between adjacent ages. However, the PSDE produces imprecise annual values (Hulson et al. 2011) so it may be that this approach should be adapted to link estimates over years (e.g., via regression or using hierarchical models). One concern with a priori estimators is they can only account for influences detectable from the sampling data, not from other process errors that influence the data but are not accounted for in the assessment (Francis 2011). An open issue is the extent to which such process errors can and should be addressed by treating them as part of the observation process (i.e., adjusting ESS) versus explicitly modeling them, perhaps using state-space approaches (e.g., Berg and Nielsen 2016, Stewart and Monnahan 2016). However, given that current practices do not approximate the full range of process error (e.g., arising from temporal variation in selectivity that may not be as smooth as assumed), it is likely that a priori specification of ESS will lead to over-fitting of the processes that are included in the assessment to the observed composition data.

We recommend our generalized mean approach for estimating annual ESS; this approach allows for correlated proportions as did approaches using means suggested by Francis (2011) but is more flexible in the relationship between ESS and sampling intensity than those methods. We think the asymptotic function that is not constrained to have a slope of 1.0 at the origin may have quite general utility because it allows consideration of different measures of sampling intensity that scale differently. We found that the slope at the origin and the asymptote for the relationship between ESS and $\widetilde{N}$ could not be uniquely estimated through optimization of an objective function for the applications we considered. However, the generalized mean approach forces consideration of what the slope near the origin should be (and if its precise value matters). In our example (Fig. 9) the standardized residuals did not change substantially between α =0.25 and α =1, however these were clearly better than smaller αs. This was a common theme across the fisheries and data types we examined; however, much of the benefit is in ruling out unrealistic αs rather than choosing the best. The reason that the standardized residuals did not change much is probably lack of contrast in the sampling. While there was annual variability there were not many instances where very few samples were taken, which is where α is most important. It follows, then, that we could not estimate α and that different values do not dramatically affect the standardized deviations. However, some fisheries may have few samples in enough years to make the slope at the origin an influential term and we think it is important to be explicit about the slope at the origin when using an asymptotic model.

The analysis we used to determine fixed values for α (e.g., Fig. 9) was not global because we did not look at all combinations of trap net α against all combinations of gillnet α (we used the same slope for both gillnets and trap nets). If α cannot be estimated, a more comprehensive (though still ad-hoc) approach would be to estimate the asymptote conditional on a set of fixed αs, in this case for both trap nets and gillnets, and evaluate all combinations. For each combination a χ^2 probability could be calculated as the sum of the individual probabilities, resulting in a table of χ^2 probabilities. This landscape of χ^2 probabilities could then be used to pick a satisfactory α to use for each gear. If the choice is not obvious it would be prudent to run a sensitivity analysis for α on the assessment model to ensure that any subjective decision does not substantially affect the results.

In our application we found that ESS results were quite similar whether we used numbers of fish or number of trips as a measure of sampling intensity. This will not always be the case when the average number of fish sampled per trip varies more substantially among years. Thus it is possible that one measure of sampling effort will be a better predictor of ESS than another, or even that ESS could be best predicted by considering multiple measures of sampling effort at the same time. The generalized mean approach could quite readily be modified to consider such options or alternatively to simply assume and estimate a constant ESS if there was little contrast in among-year sampling levels.

References

Berg, C.W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES J. of Mar. Sci. 73, 1788-1797.

Berger, A.M., Jones, M.L., Zhao, Y., Bence, J.R. 2012. Accounting for spatial population structure at scales relevant to life history improves stock assessment: the case for Lake Erie walleye *Sander vitreus*. Fish. Res. 115, 44–59.

Brodziak, J.K. 2002. An age-structured assessment model for Georges Bank winter flounder. Northeast Fish. Sci. Cent. Ref. Doc, 02–03.

Campana, S., Frisk, M.G., Martell, S.J.D., Miller, T.J., Sosebee, K. 2010. Exploring the population dynamics of winter skate (*Leucoraja ocellata*) in the Georges Bank region using a statistical catch-at-age model incorporating length, migration, and recruitment process errors. Can. J. Fish. Aquat. Sci. 67, 774–792.

Caroffino, D.C., Lenart, S.J. 2010. Statistical catch-at-age models used to describe the status of lean lake trout populations in the 1836-Treaty ceded waters of lakes Michigan, Huron, and Superior at the inception of the 2000 Consent Decree. Modeling Subcommittee, Technical Fisheries Committee.

Cochran, W.G. 1977. Sampling Techniques. John Wiley & Sons, New York.

Crone, P.R., D.B. Sampson 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. Pages 355-369 in F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, C.-I. Chang (eds), Fishery Stock Assesment Models - proceedings of the the international symposium on fishery stock assessment models for the 21st century. Publication AK-SG-98-01 of University of Alaska Sea Grant College Program.

de Moor, C.L., Butterworth, D.S., and Coetzee, J.C. 2013. Can anchovy age structure be estimated from length distribution data collected during surveys? Afr. J. Marine Sci. 35, 335–342.

Ebener, M.P., Bence, J.R., Newman, K., Schneeberger, P.J. 2005. Application of statistical catch-at-age models to assess lake whitefish stocks in the 1836 treaty-ceded waters of the upper Great Lakes. Pages 271-309 *in* L.C. Mohr, and T. F. Nalepa (eds), Proceedings of a workshop on the dynamics of lake whitefish (Coregonus clupeaformis) and the amphipod Diporeia spp. in the Great Lakes. Great Lakes Fish. Comm. Tech. Rep. 66.

Fielder, D.G., Bence, J.R. 2014. Integration of auxiliary information in statistical catch-at-age (SCA) analysis of the Saginaw Bay stock of Walleye in Lake Huron. N. Am. J. Fish. Manage. 34, 970–987.

Folmer, O., Pennington, M. 2000. A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. Fish. Res. 49, 165–178.

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

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J.R. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27, 233-249.

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

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

Francis, R.C. 2016. Revisiting data weighting in fisheries stock assessment models. <http://dx.doi.org/10.1016/j.fishres.2016.06.006 >.

Hulson, P.-J.F., Hanselman, D.H., and 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. 68, 1548–1557.

Hulson, P.-J.F., Hanselman, D.H., and Quinn, T.J. 2012. Determining effective sample size in integrated age-structured assessment models. ICES J. Mar. Sci. 69, 281–292.

Legault, C.M., Restrepo, V.R. 1998. A flexible forward age-structured assessment program. ICCAT. Col. Vol. Sci. Pap 49, 246–253.

Maunder, M.N. 2003. Paradigm shifts in fisheries stock assessment: from integrated analysis to Bayesian analysis and back again. Nat. Resour. Model. 16, 465–475.

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.

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.

Modeling Subcommittee, Technical Fisheries Committee (MSC). 2015. Technical Fisheries Committee Administrative Report 2015: Status of Lake Trout and Lake Whitefish Populations in the 1836 Treaty-Ceded Waters of Lakes Superior, Huron, and Michigan, with Recommended Yield and Effort Levels for 2015. <http://www.michigan.gov/greatlakesconsentdecree>.

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

Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. ICES J. Mar. Sci. 39: 175–192.

Pennington, M., Burmeister, L.-M., Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fish. Bull. 100, 74–80.

Pennington, M., Vølstad, 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.

Punt, A.E., Huang, T., Maunder, M.N. 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. 70, 16–33.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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(2): 218–233.

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

Stewart, I.J. and Monnahan, C.C. 2016. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fish. Res. <http://dx.doi.org/10.1016/j.fishres.2016.06.018>.

Thompson, G.G. 2011. Pacific cod. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions as Projected for 1996, pp. 2-1–2-28. Compiled by Plan Team for the Groundfish Fisheries of the Bering Sea and Aleutian Islands. North Pacific Fishery Management Council, Anchorage, AK.

Truesdell, S.B., Bence, J.R. 2016. A Review of stock assessment methods for Lake Trout and Lake Whitefish in 1836 treaty waters of Lake Huron, Lake Michigan and Lake Superior. Quantitative Fisheries Center Technical Report T2016-01. doi: 10.6084/m9.figshare.3123949.

Wilberg, M.J., Bence, J.R., Eggold, B.T., Makauskas, D., Clapp, D.F. 2005. Yellow perch dynamics in southwestern Lake Michigan during 1986–2002. N. Amer. J. Fish. Manage. 25, 1130–1152.

Acknowledgements

We acknowledge feedback from staff at the Quantitative Fisheries Center. This work is based on stock assessments and associated data prepared through the Modeling Subcommittee of the Technical Fisheries Committee for 1836 Treaty of Washington and we wish to acknowledge our appreciation for all their efforts in assessing this fishery and their support of our work. We would also like to thank two anonymous reviewers for their helpful comments. This work was supported by funding from QFC partners. This is publication 201X-YY of the Michigan State University Quantitative Fisheries Center.

Funding

This research was supported through MSU Quantitative Fisheries Center base funding that is jointly provided by Michigan State University (AgBioResearch, Extension, & General Fund College of Agriculture and Natural Resources), Michigan Department of Natural Resources, the Great Lakes Fishery Commission, and Council of Lake Committee agencies. Additional Funding was provided by the Michigan Department of Natural Resources through the Partnership for Ecosystem Management (Grant 751N6600045) and contract 751N7700014 (which was partially supported by NOAA Interjurisdictional Act Funding) and Minnesota Department of Natural Resources through a Joint Powers Agreement (which was supported by NOAA Interjurisdictional Act Funding).

Tables

Table 1. Definitions of symbols used in Tables 2 and 3.

Symbol	Description
y	Index indicating year
b	Index indicating age or length bin
$\ddot{N}_y$	Effective sample size used in assessment model and updated during each iteration.
O_{by}	The observed proportion in a bin for a year
O_y^*	Vector of observed counts in bins for a year from a multinomial distribution
E_{by}	The assessment model estimate of the probability in a bin for a year
Y	Number of years for which there are composition data
B	Number of bins for each year of composition data
$\bar{O}_y$	The observed average for age or length
$\bar{E}_y$	The assessment model estimated average for age or length
v_y	The variance in age or length in a given year based on the assessment model estimates of the composition for that year (the E_{by}).
N_y	Unconstrained effective sample size, either used as $\ddot{N}_y$ or as input to assessments when using sub-approach i, or as input for some calculations using other sub-approaches.
$\breve{N}_y$	Year-specific and pre-specified upper limit used in calculation of N_y
N	Single value for effective sample size calculated for approach ii and used as $\ddot{N}_y$ for all years.
$\widetilde{N}_y$	Measure of sampling intensity such as number of fish aged or number of trips sampled.
$\widehat{N}_y$	A prediction of effective sample size based on sampling intensity.
w	A proportionality constant relating predicted effective sample size to sampling intensity
N_{Max}	Asymptote to relationship between $\widehat{N}_y$ and $\widetilde{N}_y$
α	Slope at origin for asymptotic relationship between $\widehat{N}_y$ and $\widetilde{N}_y$

Table 2. Methods that do not allow for correlation structures among proportions. The ESS to be used in fitting the assessment model in the next iteration is given by $\ddot{N}_y = N_y$, $\ddot{N}_y = N$, or $\ddot{N}_y = \hat{N}_y$, depending on whether sub-approach i, ii, or iii and iv, respectively, are used. See Table 1 for descriptions of notational conventions and variables. The geometric mean function is indicated by gm. The maximum function, max, is taken over its arguments. $Var(x)$ is the usual sample variance. Without a subscript it is calculated over years and bins. When subscripted by year it is calculated over bins for the specified year. Naming conventions for the methods in this paper are based on this table. For instance, McAllister and Ianelli's (1997) method using a constant value based on the geometric mean is named A.ii.a.

Basic Method	(i) Unconstrained Year-specific Values	(ii) Constant value	(iii) Values proportional to sampling intensity	(iv) Statistical estimates (via regression)
A. McAllister and Ianelli	$N_y = \max\left[\frac{\sum_b E_{by}(1-E_{by})}{\sum_b (O_{by}-E_{by})^2}, \tilde{N}_y\right]$	a. $N = \text{gm}(N_y)$ or b. $N = \frac{\sum_{by} E_{by}(1-E_{by})}{\sum_{by}(O_{by}-E_{by})^2}$	$w = \text{gm}\left(\frac{N_y}{\tilde{N}_y}\right)$ $\hat{N}_y = w\tilde{N}_y$	$\hat{N}_y = \frac{N_{Max}\tilde{N}_y}{\frac{N_{Max}}{\alpha}+\tilde{N}_y}$
B. Francis TA1.2	$N_y = \max\left(1/Var_y\left[\frac{O_{by}-E_{by}}{\sqrt{E_{by}(1-E_{by})}}\right], \tilde{N}_y\right)$	a. $N = \text{gm}(N_y)$ or b. $N = 1/Var\left[\frac{O_{by}-E_{by}}{\sqrt{E_{by}(1-E_{by})}}\right]$	$w = 1/Var\left[\frac{O_{by}-E_{by}}{\sqrt{E_{by}(1-E_{by})/\tilde{N}_y}}\right]$ $\hat{N}_y = w\tilde{N}_y$	$\hat{N}_y = \frac{N_{Max}\tilde{N}_y}{\frac{N_{Max}}{\alpha}+\tilde{N}_y}$
C. Francis TA1.3	$N_y = \max\left[\frac{(B-1)}{\sum_b (O_{by}-E_{by})^2/E_{by}}, \tilde{N}_y\right]$	a. $N = \text{gm}(N_y)$ b. Use C.iii with $\tilde{N}_y = \tilde{N}$, so: $N = Y(B-1)/\sum_{by}(O_{by}-E_{by})^2/E_{by}$	$w = Y(B-1)/\sum_{by}\tilde{N}_y(O_{by}-E_{by})^2/E_{by}$ $\hat{N}_y = w\tilde{N}_y$	$\hat{N}_y = \frac{N_{Max}\tilde{N}_y}{\frac{N_{Max}}{\alpha}+\tilde{N}_y}$

Table 3. Methods that allow for correlation structures among proportions by using observed and predicted average values from compositions rather than bin-specific proportions. The ESS used in the assessment model at the next iteration is given by $\ddot{N}_y = \widehat{N}_y$. D_{χ_1} is the probability density function for a χ^2 distribution with one degree of freedom. See Table 1 for descriptions of notational conventions and variables. The maximum function is given as a function of an argument that is implicitly a function of other parameters, and the parameters the function is maximized over (α and N_{Max}) are given below the function. For approach E, N_{Max} is adjusted until the ancillary equation is satisfied. For approach F, α is (in this example) fixed a priori and N_{Max} is estimated using maximum likelihood.

Estimator	Effective N equation	Ancillary equation
D. Francis TA1.8	$\widehat{N}_y = w\widetilde{N}_y$	$w = 1/Var\left[(\bar{O}_y - \bar{E}_y)/\sqrt{v_y/\widetilde{N}_y}\right]$
E. Francis TA1.9	$\widehat{N}_y = \dfrac{N_{Max}\widetilde{N}_y}{N_{Max} + \widetilde{N}_y}$	$Var\left\{(\bar{O}_y - \bar{E}_y)/\sqrt{v_y(1/\widetilde{N}_y + 1/N_{Max})}\right\} = 1$
F. Generalized mean-based	$\widehat{N}_y = \dfrac{N_{Max}\widetilde{N}_y}{\frac{N_{Max}}{\alpha} + \widetilde{N}_y}$	$L = \max\limits_{N_{Max},\alpha} \sum_y \log\left[D_{\chi_1}\left((\bar{O}_y - \bar{E}_y)^2/v_y/\widehat{N}_y\right)\right]$

Table 4. Annual number of fish sampled and sampling events (trips sampled) for the two stock areas. The number of sampling events was not available for the Lake Michigan stock area.

	Fish Sampled			Sampling Events		
	Min	Med	Max	Min	Med	Max
North Huron						
Gillnet	415	923	2020	3	15	33
Trap net	46	929	2288	1	11	26
Lake Michigan						
Gillnet	126	343	1082	-	-	-
Trap net	30	39	658	-	-	-

Table 5: Results for the North Huron and Lake Michigan trap net and gillnet fisheries giving (1) slopes (α) and asymptotes (N_{Max}) from the final iterations of method B.iv and (2) the median ESS across all methods that (a) do not incorporate correlation structures and (b) do incorporate correlation structures. Slopes and asymptotes were obtained by nonlinear regression of $\log(N_y)$ on $\log(\widetilde{N}_y)$.

	(1) Method B.iv				(2) Median ESS	
	Fish		Trip			
	α	N_{Max}	α	N_{Max}	Methods A-C	Methods D-F
North Huron						
Trap Net	1.08	324	1000[1]	334	172	31
Gillnet	1.08	212	5.70	473	142	128
Lake Michigan						
Trap Net	1.59	100	-	-	62	50
Gillnet	1.81	94.6	-	-	70	12

[1]Such a high slope indicates essentially no relationship between ESS and number of trips for this fishery so effectively a mean ESS is used over all years.

Figures

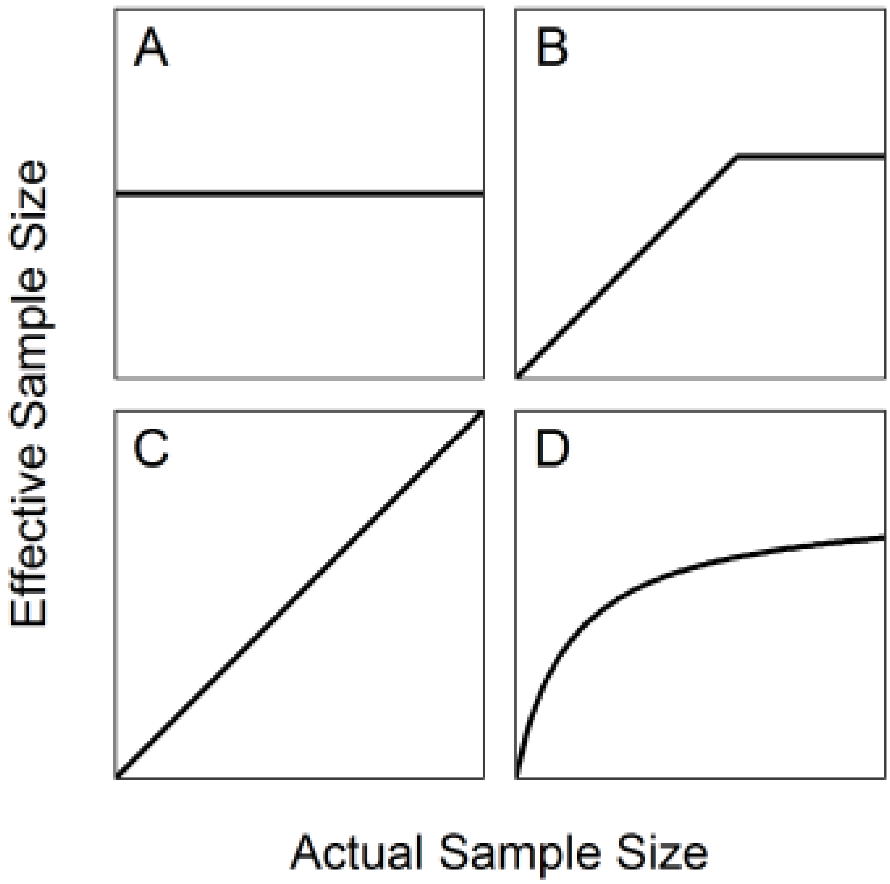

Figure 1. Options for relating ESS to sampling intensity in catch-at-age or catch-at-size models: (A) a set ESS no matter the measured sample size, (B) proportional relationship between ESS and measured sample size up to a maximum, (C) proportional relationship between ESS and measured sample size, and (D) asymptotic relationship between ESS and measured sample size. In principle, relationships between ESS and actual sample size could apply to other measures of sampling effort, such as the number of trips sampled rather than number of fish aged or measured.

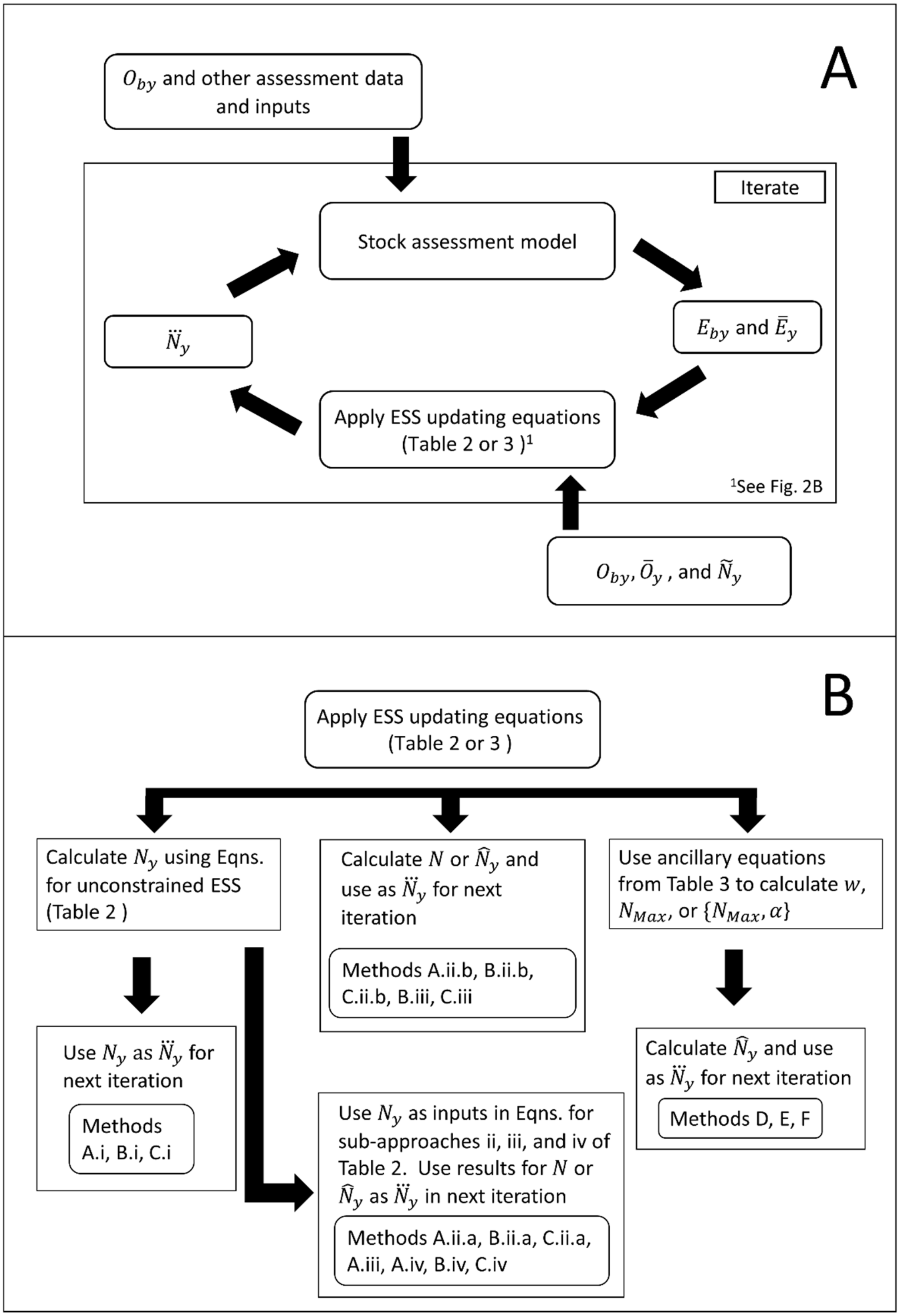

Figure 2. Process for estimating ESS using different iterative methods. A: Flowchart describing the iterative process and the data that are used at each step. B: The approaches described here for estimating ESS – details can be found in Tables 2 and 3.

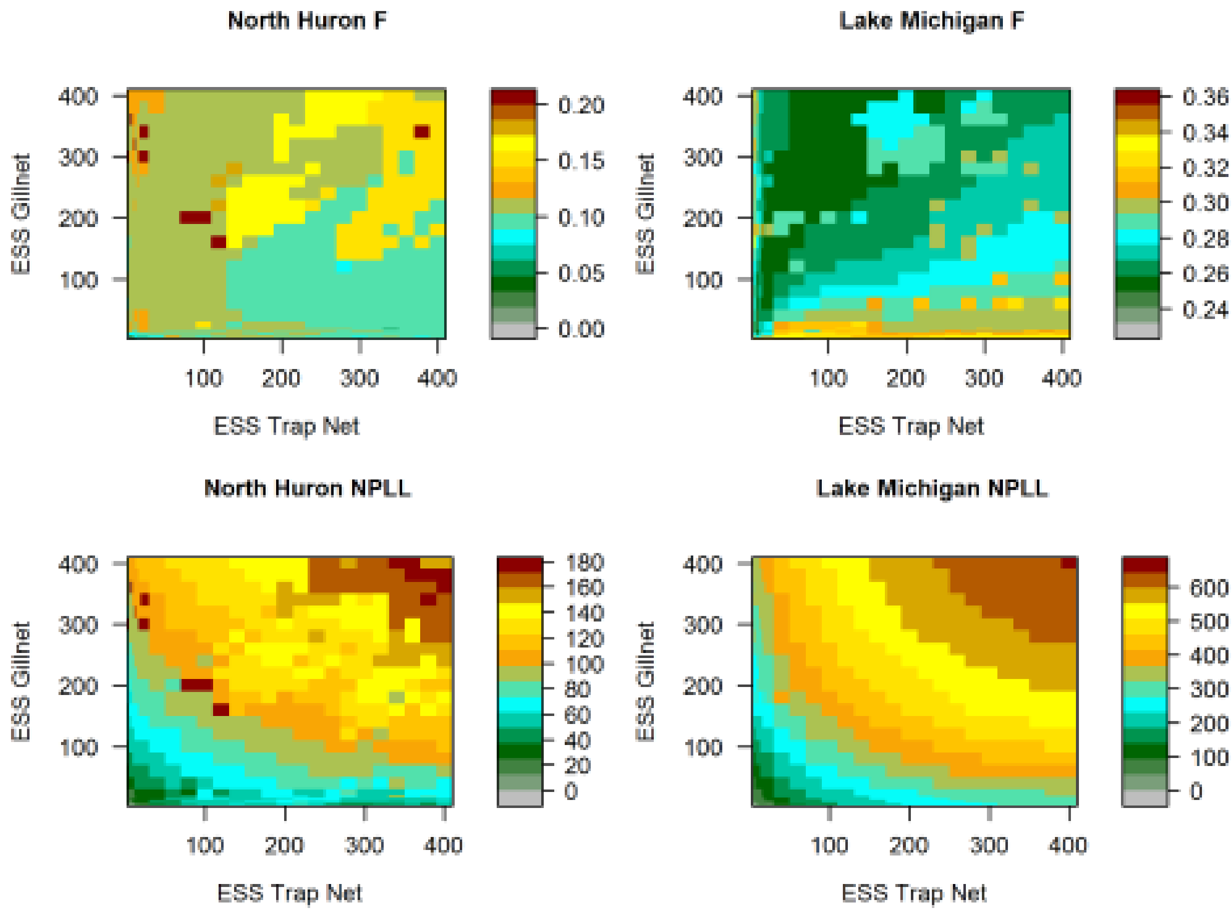

Figure 3. The top two panels are the average fishing mortality over ages 10-12 over the last 10 model years (indicated by the color scale) for varying combinations of trap net and gillnet ESS in the Northern Lake Huron and Lake Michigan stock areas. Ages 10-12 are essentially fully selected by for both trap nets and gillnets. For visualization values in North Huron > 0.2 (about 1%) were set to 0.2. The bottom panels are the sum of the negative penalized log likelihood for all likelihood components excluding the age compositions (NPLL), scaled so the smallest values are 0. For visualization NPLL values larger than 171 in North Huron (about 2%) were set to 171. Note the differences in each scale bar.

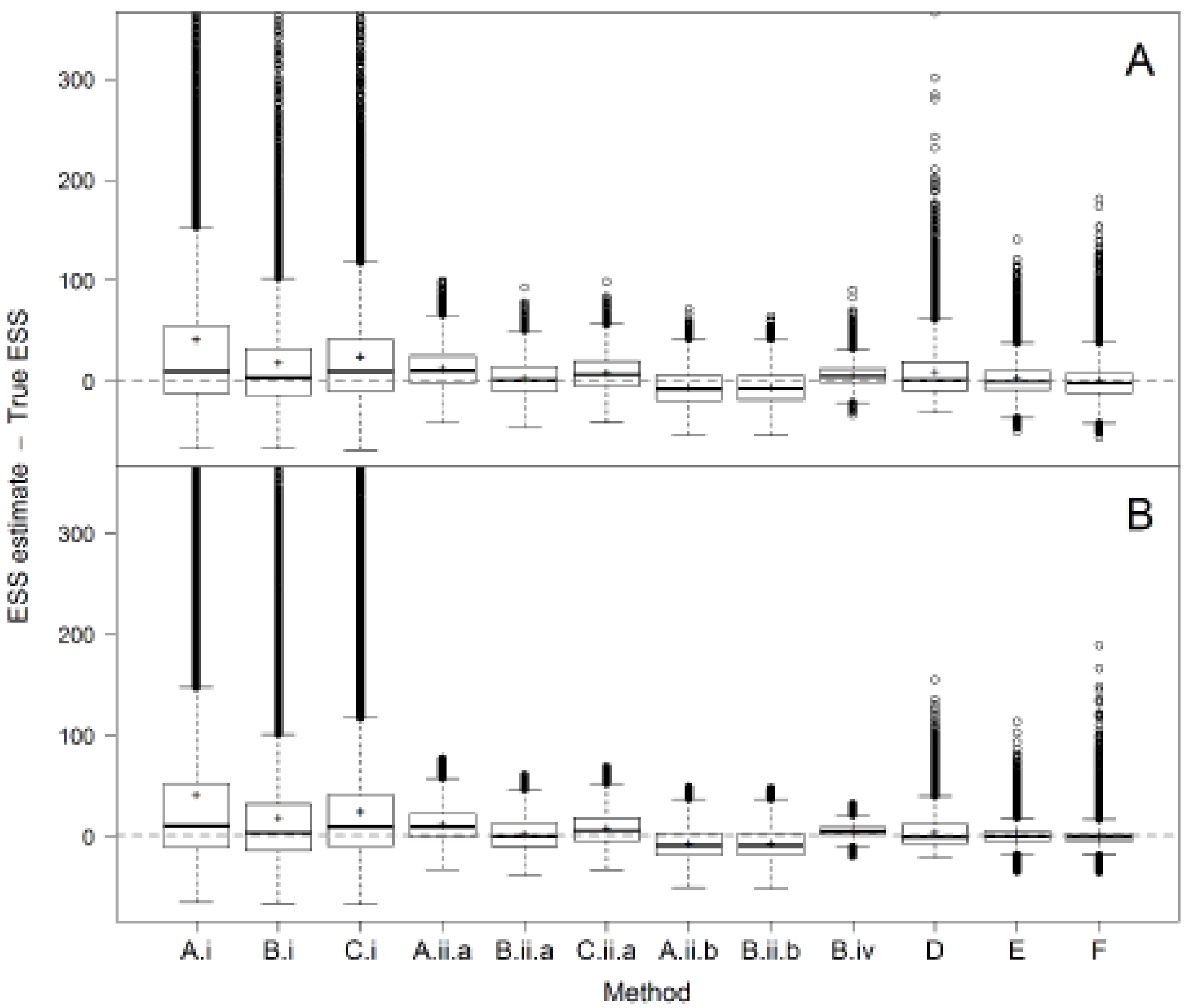

Figure 4. Box and whisker plot summaries indicating the performance of ESS estimation methods under actual multinomial sampling. The boxes indicate the interquartile range and the whiskers extend to 1.5 times this range. The horizontal lines within the boxes indicate the medians and the "+" denote the means. The specific results of this test depended on the number of multinomial categories and the annual distribution of true ESS. Plot A is over 25 years and plot B is over 100 years. The y-axis range was set so all the data are not shown: outliers in methods A.i, B.i and C.i ranged to over 4000.

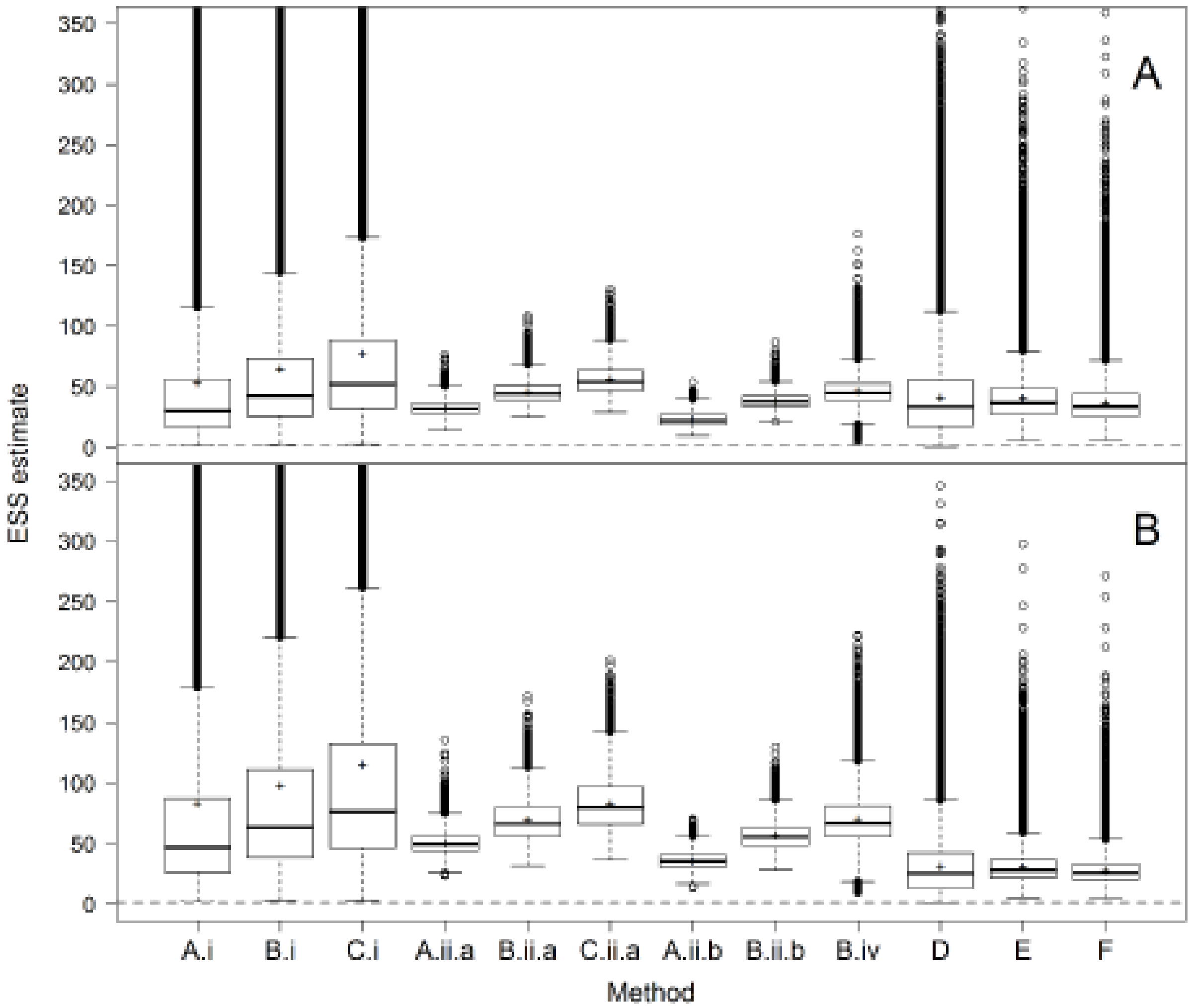

Figure 5. Box and whisker plot summaries showing the estimated ESS for the different methods when the data are simulated from a logistic-normal. Panel A assumes no correlation structure and panel B a correlation structure arising from $\rho = 0.5$. Methods D, E and F are can account for correlation structure. These methods performed similarly to the other methods (in terms of their means and medians) but estimated smaller ESS when correlation structure was present. The boxes indicate the interquartile range and the whiskers extend to 1.5 times this range. The horizontal lines within the boxes indicate the medians and the “+” denote the means. There were 25 years of sampling. The y-axis range was set so all the data are not shown: outliers in methods A.i, B.i and C.i always ranged over 5000.

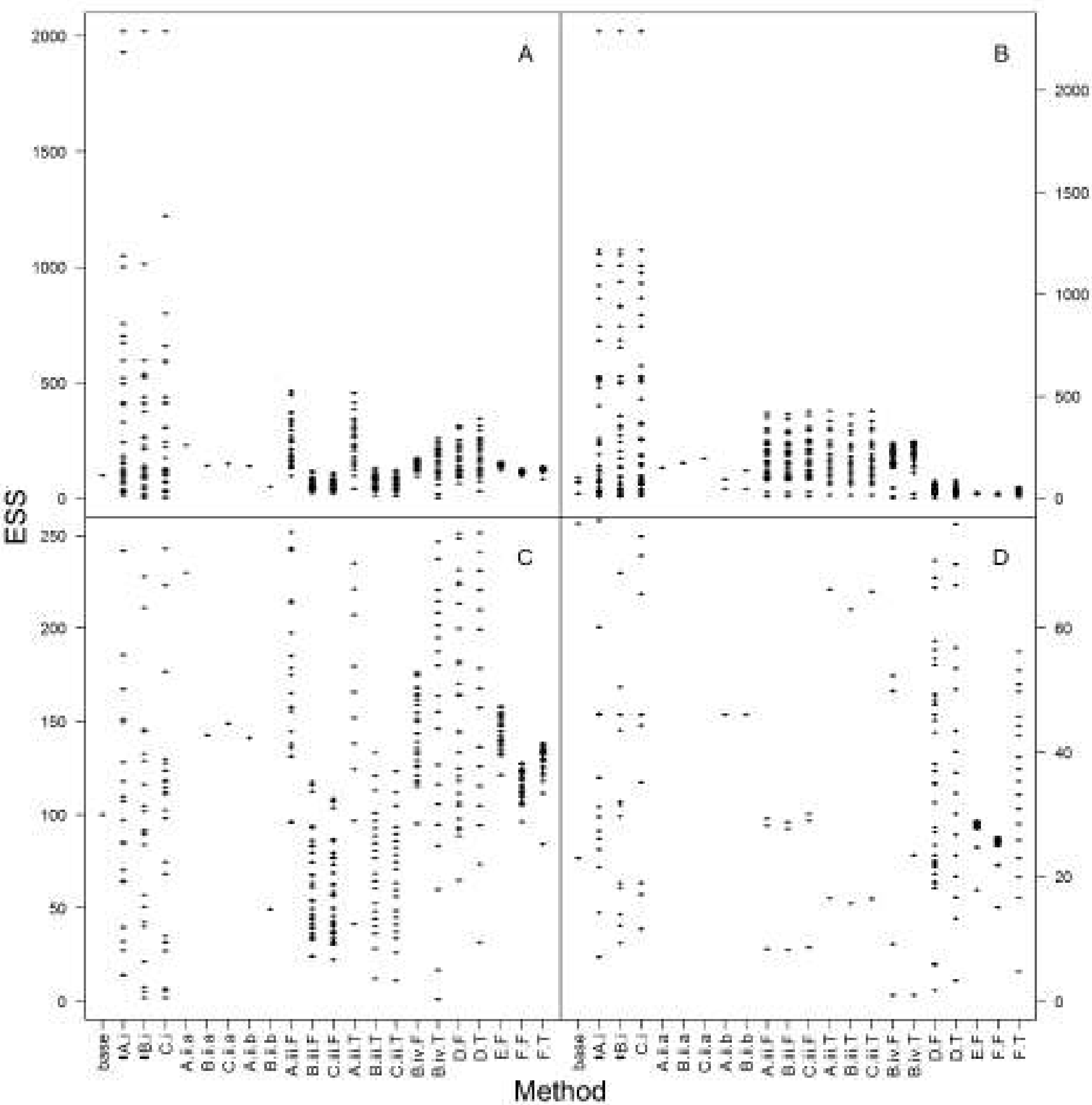

Figure 6. ESS estimates for the North Huron gillnet (A, C) and trap net (B, D) compositional data sets. Panels C and D give the same information as A and B, but the Y axis scale has changed to give better resolution within the typical range of ESS estimates that are used in practice. Underlining in the axis labels (i.e., methods A.i and B.i) indicates that the models did not converge after 25 iterations. Methods that used $\widetilde{N}_y$ end with F or T, depending on whether the measure of sampling intensity was number of fish aged (F) or number of trips sampled for ages (T). The exception is method E, which only used number of fish aged.

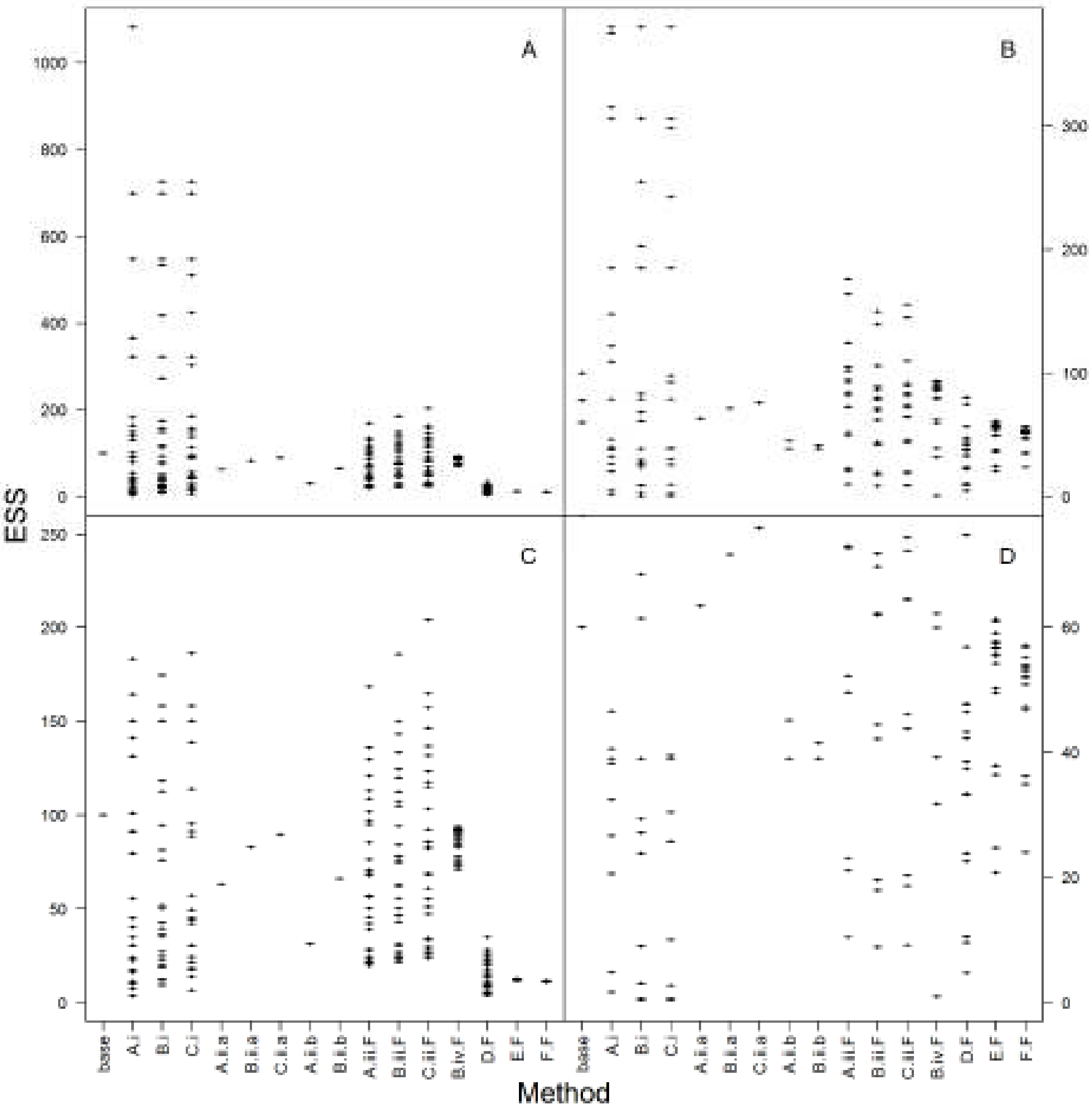

Figure 7. ESS estimates for the Lake Michigan stock area gillnet (A, C) and trap net (B, D) compositional data sets. Panels C and D give the same information as A and B, but the Y axis scale was changed to give better resolution within the typical range of ESS estimates that are used in practice. Methods that used $\widetilde{N}_y$ always used the number of fish aged (indicated by F) as the measure of sampling intensity.

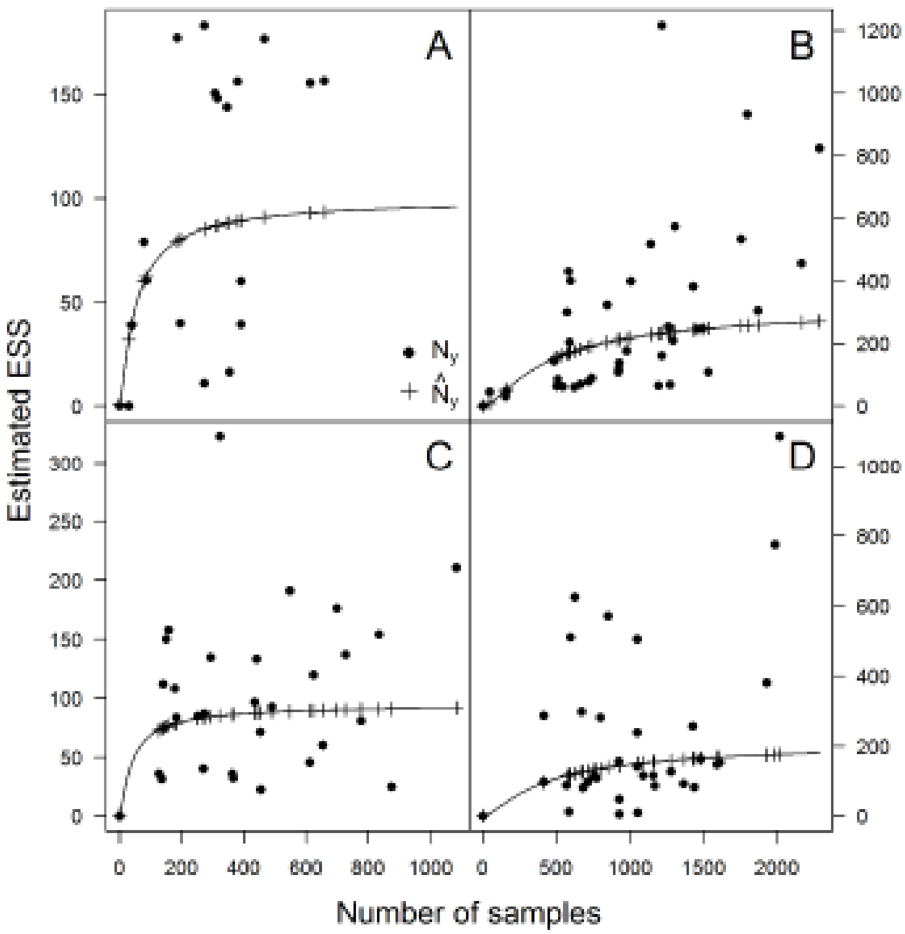

Figure 8. Relationship between the estimated ESS using method B.i and the actual number of samples in the Lake Michigan (A and B) and North Huron (C and D) assessment areas for trap nets (A and C) and gillnets (B and D). This relationship is used in method B.iv. Solid circles are unconstrained annual estimates. Curves give asymptotic relationship estimated by regression. The crosses are the regression predicted values that were used as ESS in the next model iteration. The plots given here represent the final relationship when ESS had converged.

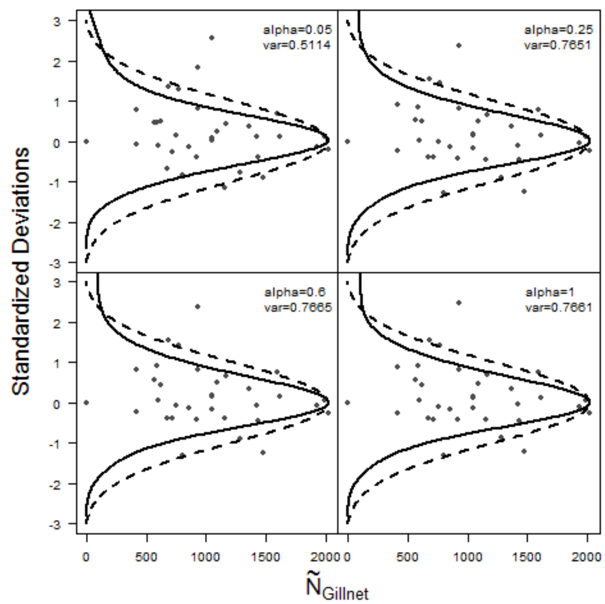

Figure 9. Example plot of $\widetilde{N}$ against standardized deviations for use in the generalized mean method (for North Huron gillnets where $\widetilde{N}$ is the number of fish sampled). The best value for α results in a variance of the standardized deviations that is closest to 1.0 with a distribution of standardized deviations that is closest to standard normal. The points are standardized deviations, the solid line is a kernel density of the standardized deviations (bandwidth = 0.5) and the dashed line is the standard normal curve. Plots such as these can be evaluated in order to determine an appropriate α level.

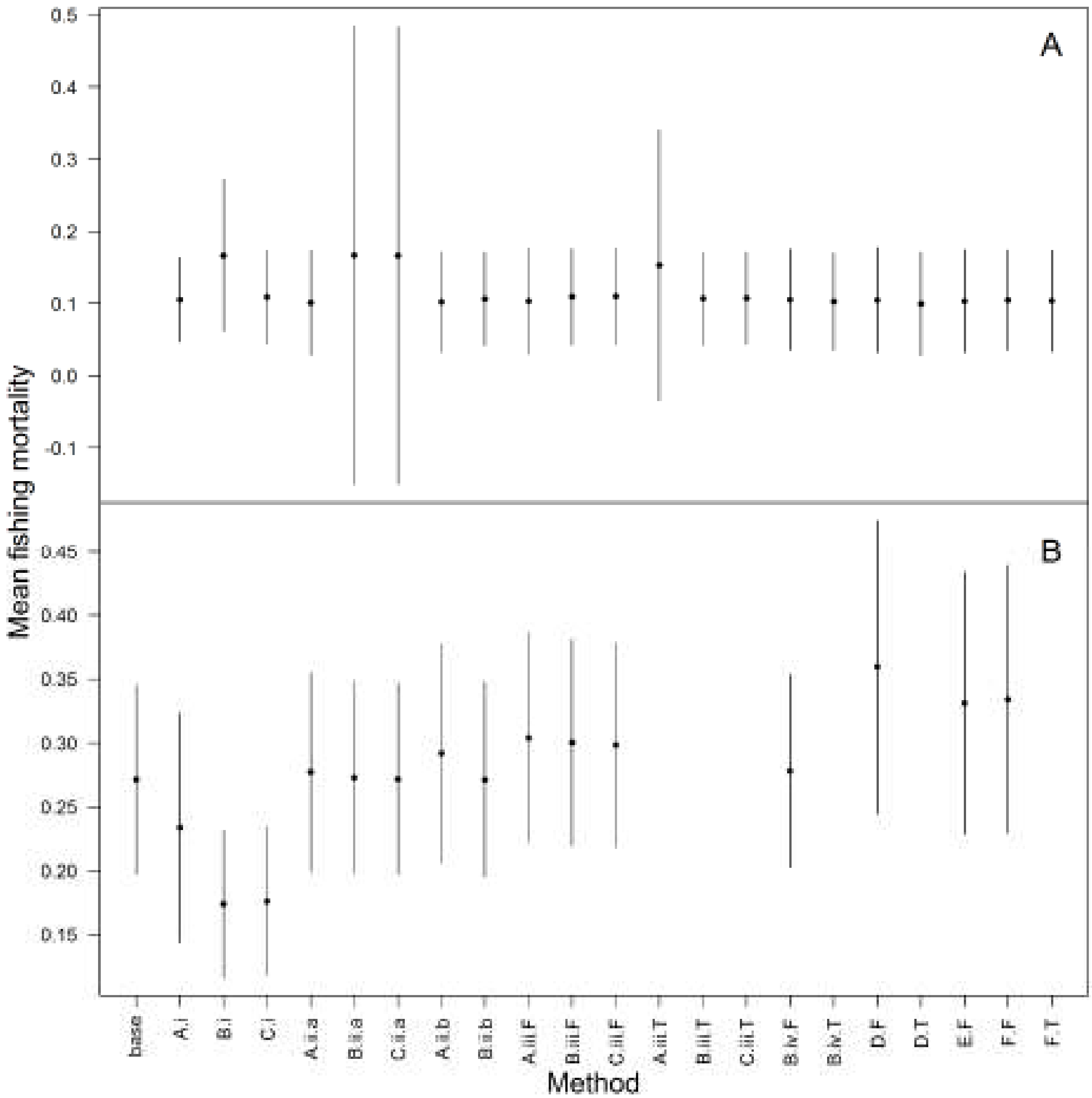

Figure 10. Average fishing mortality (ages 10-12 over the last 10 model years) for the North Huron stock area (A) and the Lake Michigan stock area (B) using the different methods to estimate ESS. Vertical lines are the standard deviations. Methods that used $\widetilde{N}_y$ end with F or T, depending on whether the measure of sampling intensity was number of fish aged (F) or number of trips sampled for ages (T). The exception is method E, which only used number of fish aged.

Appendix 1: ESS sensitivity

Table A1.1. Results of a simple sensitivity analysis to starting effective sample size. Three starting scenarios were tested: (1) the number of fish or trips actually observed (the baseline); (2) the observations multiplied by 0.5; and (3) the observations multiplied by 2. This gives the ESS estimates for each method in each model year for both assessments and both gears. The maximum difference of the three scenarios was used here. Each entry in this table represents the proportion of annual ESS where the maximum difference was less than 5. In most cases the scenarios gave very similar ESS estimates. The outliers were methods E.F for the North Huron gillnet fishery, and method F.T for both the North Huron gillnet and trap net fisheries. The reason for low correspondence for method E.F in the North Huron gillnet fishery is unknown. The reason for low correspondence for method F.T (which uses number of trips as $\widetilde{N}$) in both North Huron fisheries is probably that these sensitivity analyses lacked a thorough investigation of potential values to use for the slope at the origin (α).

	North Huron		Lake Michigan	
Method	Trap net	Gillnet	Trap net	Gillnet
A.i	1	0.97	1	1
B.i	0.82	0.85	1	1
C.i	1	1	1	1
A.ii.a	1	1	1	1
B.ii.a	1	1	1	1
C.ii.a	1	1	1	1
A.ii.b	1	1	1	1
B.ii.b	1	1	1	1
A.iii.F	1	1	1	1
B.iii.F	1	1	1	1
C.iii.F	1	1	1	1
A.iii.T	1	1	x	x
B.iii.T	1	1	x	x
C.iii.T	1	1	x	x
B.iv.F	1	1	1	1
B.iv.T	1	1	x	x
D.F	1	0.92	1	1
D.T	1	0.92	x	x
E.F	1	0.36	1	1
F.F	1	1	1	1
F.T	0.38	0.15	x	x

Appendix 2: Further information on Lake Whitefish catch-at-age models

Here we describe the age-structured stock assessment models and how they were fit for the two lake whitefish stocks reported on in the main text. Model structure and fitting approach have been developed by the Modeling Subcommittee for 1836 Treaty Waters (MSC). In this paper we use these assessments as examples, and take the assessment and basic estimation approach as given. Additional background and details on the assessment approach can be found in Truesdell and Bence (2016).

Point estimates were obtained by fitting the models to input data by maximizing the penalized log-likelihood (the objective function). During model fitting the initial abundance-at-age was freely estimated and annual recruitments were estimated including a penalty for deviations from a Ricker stock-recruit function (Table A2.1, Eqn. 1). Abundance-at-age outside the first year and recruiting age was calculated using instantaneous mortality rates composed of fishing and natural mortality (Table A2.1, Eqn. 2). In both areas fishing mortality was broken into trap net and gillnet components.

In both models, and for both fisheries, instantaneous fishing mortality rates were calculated as products of catchability, age-specific selectivity, and input annual fishing effort (Table A2.1, Eqn. 2). Catchability and selectivity depended on parameters that were estimated during model fitting. The log of catchability was time-varying and followed a random walk (Table A2.1, Eqn. 3). In the Lake Huron model selectivity was a double logistic function for both fisheries (Table A2.1, Eqns. 4 and 5), whereas in the Lake Michigan model selectivity was lognormal for the gillnet fishery (Table A2.1, Eqn. 6) and logistic (Table A2.1, Eqn. 7) for the trap net fishery. The selectivity functions were time-varying, where one parameter in each function varied via a random walk (Table A2.1, Eqns. 3, 4, 6 and 7, where $\log \beta_{2,y}$ in the logistic and double logistic and $\log \sigma_y$ in the lognormal selectivity functions followed random walks). Selectivities were a function of mean lengths-at-age (rather than direct functions of age), so selectivity-at-age could change if growth changed over time even if all selectivity parameters remained constant over years.

The data that contributed to the objective function were annual catch- and proportions-at-age for each of the two fisheries. The difference between the log annual catch for each fishery and the log predicted annual catch was modeled as normal (Table A2.1, Eqn. 8). The objective function also included penalties for the deviations in recruitment, catchability random walks, selectivity parameter random walks and the difference between the log of the estimated natural mortality rate and that produced by Pauly's temperature- and growth-based estimator (Pauly 1980). Each of these penalty components was based on the assumption that the process errors involved (typically on a log-scale) were normally distributed so they each had the same basic form as the log likelihood component for catch (Table A2.1, Eqn. 8), but with component-specific standard deviations. Observed and predicted proportions-at-age and annual ESS for each fishery were incorporated in the objective function using multinomial log-likelihood components (Table A2.1, Eqn. 9). In all calculations length-at-age, the growth and temperature parameters used in Pauly's equation, and other life-history values used to calculate spawning stock size were provided to the assessment model and were not estimated or adjusted as the model was fit.

In both lake whitefish assessments as they were originally fit by the MSC, a variance for a reference data source (for normally distributed variables) was estimated during model fitting, and ratios of this estimate to variances for other normally distributed data and process errors involved in penalties were specified (as in, for example, Fielder and Bence 2014). These ratios were adjusted during model development by the MSC so as to produce source-specific variances in accord with prior expectations. These variances are the square of the standard deviations (e.g., σ_c^2 in the equation for the catch component). In this study we elected to fix the variances for each data type or penalty at the final values that were obtained by the MSC as we explored alternative approaches to estimating ESS. We followed this approach to be consistent with the suggestion of Francis (2011), who suggested they be fixed and that the weighting of age compositions occurs in a second stage.

Equation	Description	Eqn. Number
$R_y = \alpha G_y e^{-\beta G}$	Ricker model where α and β are estimated parameters and G is the annual total calculated stock female egg weight based on the model abundance estimates. Differences between estimated recruitment and the R_y produced by Eqn. 1 contribute to a penalty term in the objective function (see Eqn. 8).	1
$F_{g,y,a} = S_{g,y,a} q_{g,y} E_{g,y}$ $Z_{y,a} = M_a + F_{G,y,a} + F_{T,y,a}$ $N_{y,a} = N_{y-1,a-1} e^{-Z_{y,a}}$ $N_{y,p} = N_{y-1,p-1} e^{-Z_{y,p}} + N_{y-1,p} e^{-Z_{y,p}}$	Equations for generating abundance-at-age outside the initial age composition and the recruiting age. Annual gear- and age-specific fishing mortality ($F_{g,y,a}$) is the product of annual age-specific selectivity ($S_{g,y,a}$), annual catchability ($q_{g,y}$) and annual effort ($E_{g,y}$). Gear types were gill net g=G, and trap net g=T. Natural mortality-at-age (M_a) is the sum of a non-age-specific background rate and an age-specific rate from sea lamprey in North Huron. Lamprey mortality was zero in the Lake Michigan area. Abundance-at-age ($N_{y,a}$) for all ages besides the plus group is the product of numbers in the previous year and annual age-specific survival ($e^{-Z_{y,a}}$). Plus-group abundance-at-age ($N_{y,p}$) is calculated in the same manner but includes surviving individuals from the plus group in the previous year.	2
$x_y = x_{y-1} + d_y$	A random walk function where x is a time series of annual values following the random walk (e.g., log catchability) and d_y is an annual change in that quantity from year $y-1$ to year y.	3
$S^*_{y,a} = \left[\frac{1}{1+\exp(-\beta_1 L_a - \beta_{2,y})}\right]\left[1 - \left(\frac{1}{1+\exp(-\beta_3 L_a - \beta_4)}\right)\right]$	The double logistic equation. The β_1 and β_3 parameters represent the slope of the increasing and decreasing logistic functions, respectively, and $\beta_{2,y}$ and β_4 represent the position of the inflection point of the increasing and decreasing functions. $S^*_{y,a}$ is the non-standardized selectivity and L_a is the average length at age a. Selectivity is standardized by dividing each annual selectivity-at-age by the selectivity calculated at a reference length (Eqn. 5). The function is applied to specific gears with gear-specific	4

	parameters, but the subscript for gear is suppressed to simplify notation.	
$S_{g,y,a} = \dfrac{S^*_{g,y,a}}{S^*_{g,y,r}}$	Selectivity standardization. Annual selectivity-at-age ($S_{g,y,a}$) for gear g is standardized using the raw selectivity-at-age function value ($S^*_{g,y,a}$) and the selectivity at a reference length in each year ($S^*_{g,y,r}$).	5
$S^*_{y,a} = \dfrac{1}{\sigma_y L_a \sqrt{2\pi}} \exp\left(-\dfrac{(\ln(L_a) - \mu)^2}{2\sigma_y^2}\right)$	The lognormal equation. σ_y is the lognormal standard deviation in year y, L_a is the average length at age a and μ is the lognormal mean. Selectivity is standardized by dividing each annual selectivity-at-age by the selectivity calculated at a reference length (Eqn. 5). The function is applied to specific gears with gear-specific parameters, but subscript for gear is suppressed to simplify notation.	6
$S^*_{y,a} = \left[\dfrac{1}{1 + \exp(-\beta_1 L_a - \beta_{2,y})}\right]$	The logistic equation. The β_1 parameter represents the slope and $\beta_{2,y}$ the (annual) position of the inflection point of the function. $S^*_{y,a}$ is the non-standardized selectivity used in the model and L_a is the average length at age a. Selectivity is standardized by dividing each annual selectivity-at-age by the selectivity calculated at a reference length (Eqn. 5). The function is applied to specific gears with gear-specific parameters, but subscript for gear is suppressed to simplify notation.	7
$L_k = -\ln \sigma_k - \dfrac{\left(\log k - \log \hat{k}\right)^2}{2\sigma_k^2}$	The normal log density for parameter k used to specify penalties. The equation given is the likelihood for a single value. Likelihoods for vectors of annual values (e.g., catch) are the sum of this function over all values. The negative of these are "penalties." Random walk parameters (e.g., catchability) are assumed to have a mean of zero (so $\log \hat{k}$ is zero).	8
$L_M = \sum_{y=1}^{Y} N_{E,y} \sum_{a=1}^{A} \left[p_{y,a} \log(\hat{p}_{y,a})\right]$	The multinomial log-likelihood for a series of annual proportions-at-age where $N_{E,y}$ is the ESS in year y, Y is the number of years, a is the age-class index, A is the number of age classes, and $p_{y,a}$ and $\hat{p}_{y,a}$ are, respectively, the observed and predicted annual proportions-at-age. This equation was	9

	applied separately to the age compositions for each fishery (i.e., ESS and proportions were fishery-specific).