Canadian Journal of
Fisheries and
Aquatic Sciences

Article

Reductions in sampling effort for fishery-independent age and length composition: balancing sampling efficiency, data uncertainty, and workforce health

Peter-John F. Hulson[a], **Benjamin C. Williams**[a], **Meaghan D. Bryan**[b], **Jason Conner**[c], **and Matthew R. Siskey**[d]

[a]Auke Bay Laboratories, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 17109 Point Lena Loop Road, Juneau, AK 99801, USA; [b]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA 98115, USA; [c]Resource Assessment and Conservation Engineering Division, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA 98115, USA; [d]Washington Department of Fish and Wildlife, 111 Washington Street SE, Olympia, WA 98501, USA

Corresponding author: **Peter-John F. Hulson** (email: pete.hulson@noaa.gov)

Abstract

Unavoidable survey effort reduction has become a reality that must be accounted for in fisheries stock assessment. In addition, negative consequences to survey staff health due to repetitive motion injuries are becoming increasingly costly for managing agencies. In this study, we evaluated the outcomes of reductions in age and length data used in fisheries stock assessment models. The main goal was to determine whether sampling can be reduced to a level that does not excessively increase data uncertainty, yet provides a reduction in repetitive motions that can cause injury to survey staff. We found that reducing length sampling to a maximum of 100–150 fish sampled per haul (either sex-specific or combined sex) provides length composition data for which the uncertainty is not appreciably increased, and it has minimal effect on the uncertainty in age composition data that is subsequently expanded from this subsampled length frequency data. The method employed here, and the results presented, can aid management agencies to balance the magnitude of data collection and subsequent consequences to fisheries stock assessment models.

Key words: age and length sampling, fisheries stock assessment models, uncertainty

Introduction

Many integrated fishery stock assessments rely on information on fish population demographic composition in the form of length composition (Punt et al. 2013) or as a derived quantity from length composition expanded to age composition (Maunder and Piner 2015). The most reliable age and length composition estimates are based on observations from fishery-independent surveys because these platforms generally avoid the sample selection bias inherent in directed commercial fisheries (NRC 1998). Past analyses have focused on the statistical efficiency of composition estimates for data weighting and applications to likelihood functions (e.g., Thorson 2014; Francis 2017; Thorson et al. 2017). It is recognized in sampling theory that increasing the number of samples increases the precision of the estimate, and from this perspective, one would expect that the maximum number of samples per each survey observation would increase precision. Improvements in precision of age or length composition estimates are not only influenced by the magnitude of sampling, but also the way samples are collected, for example, considering the number of samples collected within a haul versus changing the number of hauls from which samples were obtained (Siskey et al. 2023). Furthermore, each fish sampled on a survey has an immediate physical labor cost (e.g., at-sea time management issues), a cumulative health cost (e.g., repetitive stress injuries), downstream labor costs (e.g., staff effort reading otoliths to determine age), and budget costs (e.g., the monetary cost to perform a fisheries-independent survey).

It is commonly understood that sampling to determine age and length composition of fishery-independent or fishery-dependent sources can be influenced by intrahaul correlation. For example, Pennington et al. (2000) demonstrated that fish sampled from hauls that are similar in size and (or) age may not be representative of the overall population's size and (or) age distribution owing to pseudoreplication. To evaluate the level of intrahaul correlation, and to account for it in fishery stock assessment, the concept of "realized sample size" has been developed (using the terminology of Stewart and Hamel 2014). Realized sample size is smaller than the number of samples collected, reflecting the increase in uncertainty due to the intrahaul correlation displayed by the sampled species (e.g., first derived in Appendix 1 of McAllister and Ianelli (1997) and also included in Stewart and Hamel (2014)). In fisheries stock assessments, this realized sample size is used within the multinomial likelihood to fit age and (or) length composition data, termed "input sample size (ISS)" to distinguish that it is being used as an input to a stock assessment model (Thorson et al. 2023). Several studies have used ISS to evaluate impacts on assessment results, including whether ISS can be estimated within the assessment model (Hulson et al. 2012; Francis 2017; Thorson et al. 2017), and more recently as a tool to evaluate the effect of changes to sampling methodologies on assessment model uncertainty (Siskey et al. 2023). Furthermore, bootstrapping methods that mimic the sampling designs for length and age composition have been developed to estimate ISS external to an assessment model (Stewart and Hamel 2014). Overall, the use of ISS has become a recommended method to account for uncertainty caused by overdispersion of samples in length and age composition data.

Each year thousands of samples are obtained from fishery-independent surveys for both age and length composition across multiple species. For example, the National Oceanic and Atmospheric Administration's (NOAA's) Alaska Fisheries Science Center (AFSC) Groundfish Assessment Program (GAP) is responsible for the execution of several fisheries-independent bottom trawl surveys (Stauffer 2004), spanning most of the continental shelf in waters off Alaska that are south of the Bering Strait, including the eastern Bering Sea (EBS), the Aleutian Islands (AI), and the Gulf of Alaska (GOA). The quantitative time series of comprehensively cataloging the biota encountered at each sampling station within these surveys began in 1982 with the adoption of standardized trawling protocols (Lauth et al. 2019), but trawl observations were made as early as 1955 (Zimmermann et al. 2009). The AFSC scientists have routinely collected observations of fish length and age composition, consisting of complete or random subsamples of fish within each trawl haul through time.

There are several unintended consequences associated with intense sampling of species within a fishery-independent survey. The length measurement process is repeated hundreds of thousands of times during an annual survey, and otolith sampling occurs several thousands of times. This represents a daunting amount of work for field scientists on research vessels and for age readers once the survey is over. Furthermore, a large portion of the fish sampled for length frequency is subsequently sampled for sex determination, which involves a substantial increase in handling and data collection effort (Link et al. 2008). Each year, this workflow results in acute and repetitive stress injuries, some requiring medical interventions and claims to the U.S. Office of Workers' Compensation Program (Stan Kotwicki, personal communication, 14 December 2021). Another consequence resulting from the intensity of this work is the introduction of unrecoverable errors in the observed data. Despite extensive in situ quality assurance protocols, as fatigue or injuries accumulate during the course of field work, so do the potential for data collection errors (e.g., failure to properly encode length measurements with the correct species or sex, or misidentification of sex).

Maximizing the number of samples collected for age and length composition is desirable from a statistical viewpoint; however, it is important to accurately define the sampling unit that needs to be optimized (e.g., the number of samples collected within a haul versus the number hauls from which samples are obtained). In addition, optimizing collections to balance the statistical quality of the data with the health of the workforce and budgetary constraints is more broadly desirable. At the AFSC, this balance has been achieved by decreasing collection efforts from historical levels. The primary concern, from a stock assessment perspective, when reducing the sampling of length frequency data and age composition data collections, is the impact on the uncertainty in subsequent age and length composition data. To this end, ISS can be used as a tool to evaluate the consequences of changing sampling effort or strategies. An additional consideration when evaluating the consequences of optimizing sampling effort is whether there are species-specific impacts (e.g., life history characteristics that minimize the impacts of a reduction in sampling effort in one species compared to another).

In this study, using data collected by the AFSC GAP bottom trawl surveys across the EBS, AI, and GOA, we evaluated the consequences of reductions in length frequency collections (including sex-specific length frequency) coupled with reductions in age composition sampling. Using ISS as the primary statistic to evaluate uncertainty in the age and length composition data derived from AFSC bottom trawl surveys, we addressed four questions: (1) What is the impact of reduced sampling of length frequency information (including sex determination sampling) and the subsequent expansion to age composition on the uncertainty in the composition data? (2) What is the impact of reduced sampling for age information on uncertainty in age composition data? (3) Are current sampling levels at a point of diminishing returns and can we decrease sampling intensities for age and length without appreciably increasing uncertainty? (4) Are there species-specific characteristics that mitigate, or exaggerate, the consequences of reductions in age or length composition sampling?

Materials and methods

Computing length and age composition from bottom trawl survey data

Data collection for each AFSC GAP groundfish bottom trawl survey is described in respective NOAA Technical Memorandums (EBS: Lauth et al. (2019); AI: von Szalay et al. (2017); and GOA: von Szalay and Raring (2018)). The fundamental methods of length sample collection are generally synchronized between these surveys. Species and sex-specific length distributions within a catch sample are collected by (1) sorting the sample by species, (2) weighing each species in aggregate, (3) obtaining a random subsample of a target sample size, (4) sorting the subsample by sex (each fish is incised to visually determine sex), and (5) measuring length and recording each fish's measurements into a database. Additionally, a subsample of these fish is processed at sea to additionally collect their weight and sagittal otoliths. These otoliths are returned to the AFSC's Age and Growth laboratory for age determination. Survey age sampling protocols are species-specific and follow 1 of 2 paradigms: (1) a stratified random collection distributed over both the spatial frame of the stratification scheme and the expected size range of a species; or (2) a small subsample (2–20 fish, species-dependent) collected randomly per trawl. The protocol for some species has changed over the time series, generally following a trend of transitioning from protocol (1) to protocol (2).

Length frequency samples collected by the AFSC GAP bottom trawl surveys are expanded by area-swept catch-per-unit-effort (CPUE) and stratum area to obtain estimates of population abundance-at-length (i.e., design-based expansion). In a design-based expansion process, this is often referred to as the "first stage expansion" and it is a common method to obtain population estimates at length from area-swept survey data (Miller and Skalski 2006; Ailloud and Hoenig 2019). To expand the species-specific length frequency samples to population-at-length, we first compute the overall population numbers within a stratum by multiplying the average CPUE within the strata (i.e., the number of fish per square kilometer averaged across the hauls performed within the strata) by the area of the strata (in square kilometers). The overall population numbers in year y within stratum s ($\widehat{N}_{s,y}$) is computed with

$$(1) \qquad \widehat{N}_{s,y} = \overline{CPUE}_{s,y} \cdot A_s$$

where A_s is the area of stratum s (in km^2), and $\overline{CPUE}_{s,y}$ is the species-specific average CPUE of numbers captured across the hauls within a strata s in year y. We then compute the relative CPUE for each haul performed within the strata. The relative CPUE for each haul ($\widehat{c}_{h,s,y}$) is computed by

$$(2) \qquad \widehat{c}_{h,s,y} = \frac{\mathrm{CPUE}_{h,s,y}}{\sum_{h=1}^{H_{s,y}} \mathrm{CPUE}_{h,s,y}}$$

where $\mathrm{CPUE}_{h,s,y}$ is the catch-per-unit-effort of numbers caught within a haul h for stratum s in year y and $H_{s,y}$ is total number of hauls for stratum s in year y (across which the denominator is summed in eq. 2). We next compute the sex-specific relative length composition for each haul ($\widehat{p}_{x,l,h,s,y}$) with

$$(3) \qquad \widehat{p}_{x,l,h,s,y} = \frac{\sum_{h=1}^{H_{s,y}} \left[N_{x,l,h,s,y}/N_{h,s,y}\right]}{\sum_{x=1}^{3}\sum_{l=1}^{L}\sum_{h=1}^{H_{s,y}} \left[N_{x,l,h,s,y}/N_{h,s,y}\right]}$$

where $N_{x,l,h,s,y}$ is the length frequency sampled, in numbers, by sex x and length l (in cm) within a haul h for stratum s in year y. We note that at the AFSC, the length bins are 1 cm; however, the formulae here are flexible to other binning structures (for example, 2 or 5 cm bins). Finally, the expanded population abundance-at-length is obtained by multiplying the overall population numbers within the strata (eq. 1), the relative CPUE of each haul (eq. (2)), and the sex-specific relative length composition (eq. 3) with

$$(4) \qquad \widehat{N}_{x,l,s,y} = \widehat{N}_{s,y} \cdot \sum_{h=1}^{H_{s,y}} \left[\widehat{c}_{h,s,y} \cdot \widehat{p}_{x,l,h,s,y}\right]$$

Population abundance-at-length are computed for three sex categories (males, females, and unsexed) at the stratum level, which are then summed across strata to obtain the population abundance-at-length for the management-scale region (i.e., EBS, AI, or GOA). Strata are defined as regions with similar bathymetric characteristics (e.g., depth ranges), and population abundance-at-length within strata can also be summed to any subregion level. We note that this formulation is equivalent to the design-based length composition expansion used in Stewart and Hamel (2014). The only difference is found in multiplying the sex-specific relative length composition (eq. 4) by the relative CPUE for each haul (eq. 2) in eq. (4) here, where Stewart and Hamel (2014) multiply the sex-specific relative length composition by the predicted number of fish in a haul.

Age-length keys (ALKs) generated from the age-length paired observations within a survey are then applied to estimated abundance-at-length to provide an estimate of abundance-at-age (e.g., Quinn and Deriso 1999), referred to as the "second stage expansion". In the second stage expansion, the sex-specific estimates of population abundance-at-length (from eq. 4) are used to estimate sex-specific population abundance-at-age. The age specimens collected during the survey, which include observations of age-at-length, are first populated into sex-specific numbers at age and length ($N_{x,a,l,y}$). Next, the sex-specific numbers-at-age and length are converted to sex-specific proportions of age-at-length (i.e., ALK) with

$$(5) \qquad \widehat{p}_{x,a,l,y} = \frac{N_{x,a,l,y}}{\sum_{a=1}^{A} N_{x,a,l,y}}$$

The proportions of age-at-length are then expanded to population abundance-at-age with

$$(6) \qquad \widehat{N}_{x,a,y} = \sum_{l=1}^{L} \widehat{p}_{x,a,l,y} \cdot \widehat{N}_{x,l,y}$$

where $\widehat{N}_{x,l,y}$ is the population abundance-at-length from eq. (4) summed across strata. Further details on the specific methods the AFSC uses to expand length and age samples to abundance are shown in Hulson et al. (2023). The stocks selected for this analysis (Table 1) are all assessed by the AFSC with statistical catch-at-age models and have corresponding expanded age and (or) length composition estimates from the respective bottom trawl surveys. The number of hauls in which these stocks were present, and the number of those hauls sampled for length frequency and age specimen data, are shown in Table S1. It was noted in Siskey et al. (2023) that, for some of these stocks, there were some age-length pairs in the AFSC age specimen data where the length was not reflected in the length frequency data. This also occurred in the

Table 1. Average length and age samples (length samples shown first, age samples are shown in parentheses) from the most recent three AFSC bottom trawl surveys by region for the stocks evaluated in the bootstrap–simulation for reduction in length and age collections.

Stock	Scientific name	AI	EBS	GOA
Alaska plaice (f)	*Pleuronectes quadrituberculatus*	–	10 486 (327)	–
Arrowtooth flounder (f)	*Atheresthes stomias*	9868 (460)	14 928 (702)	36 842 (597)
Atka mackerel (o)	*Pleurogrammus monopterygius*	7888 (610)	–	–
Flathead sole (f)	*Hippoglossoides elassodon*	4602 (0)	22 356 (810)	15 233 (225)
Northern rock sole (f)	*Lepidopsetta polyxystra*	10 754 (0)	24 698 (584)	3685 (197)
Northern rockfish (r)	*Sebastes polyspinis*	14 928 (571)	–	2298 (447)
Pacific cod (g)	*Gadus macrocephalus*	5723 (604)	11 477 (1,449)	3452 (464)
Pacific ocean perch (r)	*Sebastes alutus*	32 491 (1,019)	–	23 319 (1,150)
Rex sole (f)	*Glyptocephalus*	–	–	12 878 (412)
Southern rock sole (f)	*Lepidopsetta billineta*	–	–	7190 (730)
Walleye pollock (g)	*Gadus chalcogrammus*	10 806 (791)	49 544 (1600)	16 772 (1,075)
Yellowfin sole (f)	*Limanda aspera*	–	28 108 (845)	–

Note: Species types are shown in parentheses in the stock column (f—flatfish, g—gadid, r—rockfish, o—other). AFSC, Alaska Fisheries Science Center; AI, Aleutian Islands; EBS, eastern Bering Sea; GOA, Gulf of Alaska.

data used for this analysis; however, the age-length pairs that were dropped in the subsequent ALK and the second stage expansion represented a fraction of a percent of the total age specimen data for all of the stocks investigated (not exceeding 0.53% of the total age specimen data for any given stock). Historical data for each survey were used in this analysis to show results that reflect the consequences of subsampling. To show the recent sampling magnitudes, the average length and age sample sizes from the most recent three survey years for the stocks selected are shown in Table 1.

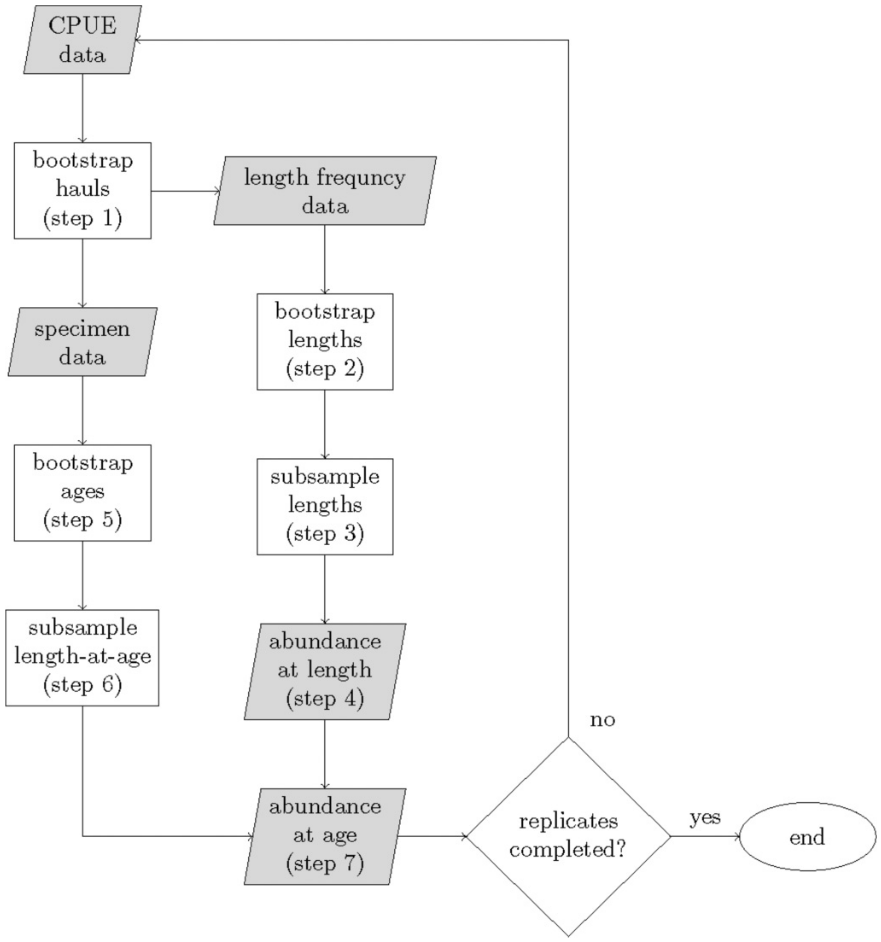
Simulation–bootstrap framework

To evaluate the effect of reductions in sampling length frequency and specimen age data, we developed a bootstrap–simulation framework that (1) allows for reductions in the historical number of length frequencies and specimen age data collected, and (2) performs first (length) and second (age) stage expansion processes for each bootstrap replicate of length frequency and specimen age data to generate length and age compositions. The bootstrap–simulation framework is composed of a suite of nested resampling protocols. Bootstrap resampling was performed either with replacement (wr) or without replacement (wor) depending on the needs of a particular protocol. With replacement was used to evaluate the uncertainty for a full sample set (for example, for the full set of hauls, lengths, and ages), while without replacement was used when subsampling from a set of samples at a pre-determined level (for example, when subsampling hauls for length frequency and age observations to investigate the consequences of reductions in sampling effort). Functions to run the sampling protocols were developed in a compartmentalized manner to provide substantial flexibility in exploring desired resampling protocols. The order of operations (Fig. 1) has the following schedule:

1. Resample hauls (wr) with associated CPUE (in numbers).
2. Within the resampled hauls from step 1, resample haul-specific observed lengths (wr).
3. From the resampled lengths in step 2, subset the haul-specific length samples (wor) at the pre-determined subsampling level (for either sex-specific or total length samples).
4. With the resampled and subsampled length frequency data in step 3, calculate sex-specific population abundance-at-length.
5. Within the resampled hauls from step 1, resample the observed ages (wr).
6. From the resampled ages in step 5, subset the total ages sampled (wor) at pre-determined subsampling level.
7. With the resampled and subsampled age data in step 6 and the sex-specific population abundance-at-length in step 4, calculate sex-specific population abundance-at-age.

The core of the bootstrap–simulation function (steps 3 and 6 above) is designed to explore reductions in the sample size of lengths collected on a per haul basis, as well as in the total sample size for ages (aggregated across hauls). In step 3 of this bootstrap–simulation, the pre-determined subsampling level for lengths (whether total or sex-specific, n_l) in a given haul must be less than or equal to the sample size x collected in the historical data for that haul. When the pre-determined subsampling level for lengths in a haul n_l is less than the historical sample size x for that haul, then a random draw of n_l from the resampled lengths in step 2 is taken without replacement; if $n_l > x$, then the resampled lengths of size x in step 2 is used directly. We evaluated the impacts of subsampling of length frequency and age specimen collections at two different levels, either at the haul level (for length frequency) or at the survey total level (for age specimen collections). We set the subsampling level for length frequency at numbers per haul to evaluate the AFSC length sampling design, as the effort expended by survey staff to measure length and determine sex of a fish is most important at the haul level. Additionally, to subsample ages, we reduced the proportion of the total number of ages sampled in step 6 to evaluate the consequences of reductions in overall age sampling. We eval-

Fig. 1. Bootstrap–simulation flowchart, with numbered steps referring to the order of operations. Shaded trapezoids are data inputs or outputs, white rectangles are processes, the diamond is a decision point, and the ellipse ends the simulation.



uated age specimen sampling at the total survey level because the number of otoliths collected per haul for age analysis is typically small and an important consideration for the level of age sampling is the total number of ages read at an age reading laboratory. The bootstrap–simulation repeated steps 1–7 iteratively for each length and age subsample level, providing iterated population abundance-at-length and age for comparison to the population abundance-at-length and age calculated from the original data.

The length subsample levels evaluated from the full length frequency collections were 50, 100, 150, 200, and 250 samples per haul. Using the GOA stocks as an example, we evaluated the impact of binning length composition data at 2 and 5 cm length bins, as well as the impact of subsampling for lengths at a subregional scale (for the western, central, and eastern GOA). The age subsample levels evaluated from the full age specimen collections were 25%, 50%, 75%, and 90% of the total number of ages collected in any given survey year. The bootstrap–simulation was run for the full number of length frequency and age specimen collections without subsetting to compare the base level uncertainty with the uncertainty observed as a result of subsampling. We ran the bootstrap–simulation for 500 iterations, the level at which variability in results did not increase (see Fig. S1). The bootstrap–simulation was developed in R (R Core Team 2022) and is available via GitHub as an R package (https://github.com/BenWilliams-NOAA/swo).

Computing ISS

Realized sample size, as introduced in Appendix 2 of McAllister and Ianelli (1997), is a statistic that can approximate the level of intrahaul correlation in age and length composition samples that can be used in several error structures when fitting composition data in stock assessments (e.g., multinomial, Dirichlet, and Dirichlet-multinomial). It is also a statistic that can be used to evaluate the amount of uncertainty in estimated composition data compared to observed composition data. Realized sample size $R_{i,y}$ is given by

$$(7) \qquad R_{i,y} = \frac{\sum_{c=1}^{C} O_{c,y}\left(1 - O_{c,y}\right)}{\sum_{c=1}^{C}\left(E_{c,i,y} - O_{c,y}\right)^2}$$

where $E_{c,i,y}$ is the estimated proportion for category c (which can be either age or length or any other arbitrary category across which proportions are computed) in iteration i and year y and $O_{c,y}$ is the observed proportion for category c in year *y*. In this bootstrap–simulation the length and age compositions derived from the historical bottom trawl surveys' full datasets without any resampling were treated as the observed proportions $O_{c,y}$ in eq. 7. For each iteration of the bootstrap simulation for a pre-specified subsampling level, we computed an estimated proportion ($E_{c,i,y}$) that was then compared to this observed sex-specific (and sex-aggregated total) length and age composition. Thus, across each iteration of the bootstrap simulation, we computed a realized sample size that indicated the change in uncertainty caused by subsampling length frequency and age specimen data. A larger realized sample size indicates higher certainty in the iterated composition estimates, while lower realized sample size indicates less certainty.

The annual realized sample sizes across iterations were summarized using the harmonic mean. This has been shown to reduce bias in recovering the true sample size in simulations for a multinomial distribution and has been recommended to determine the "ISS" used as an initial weighting in stock assessment models to fit compositional data (Stewart and Hamel 2014). Herein, we use the term "iterated realized sample size" (iterated *R*) to refer to the realized sample sizes computed for each iteration of the bootstrap–simulation from eq. (7), and we use the term "ISS" to refer to the harmonic mean of the iterated realized sample sizes.

We provide results of both the annual age and length composition ISS estimates and their relative size (termed "relative ISS") compared to the full dataset ISS. The relative ISS was computed as the ISS at a pre-determined subsampling level divided by the full dataset ISS (no subsampling) for each survey year. This provides an indication of the amount of change in uncertainty due to reductions in subsampling rates. To further explore and compare the general relationship between reduction in relative ISS due to reduction in sampling for age composition, we fit a linear model to these data by sex for each species type and survey region, while presenting the relative ISS for each stock evaluated.

To provide general results for illustration of trends, we display the ISS and relative ISS by species type (flatfish, gadid, and rockfish) and across survey regions (AI, EBS, and GOA) using box plots (which include the median, interquartile range, and 95th percentile range). Thus, when species type results are shown they are grouped across the stocks within that species type (as indicated in Table 1) and across the survey regions. Additionally, species-aggregated plots for the GOA are presented that show the influence of length composition bins and subregional scale analysis. Finally, while the primary metric of this study was ISS, as an indication of changes to uncertainty due to subsampling, we also evaluated relative bias in sex-specific mean length and the sex ratio as additional metrics. In this case, relative bias was computed as the relative bias between the full dataset mean length (or sex ratio) and the subsampled dataset mean length (or sex ratio).

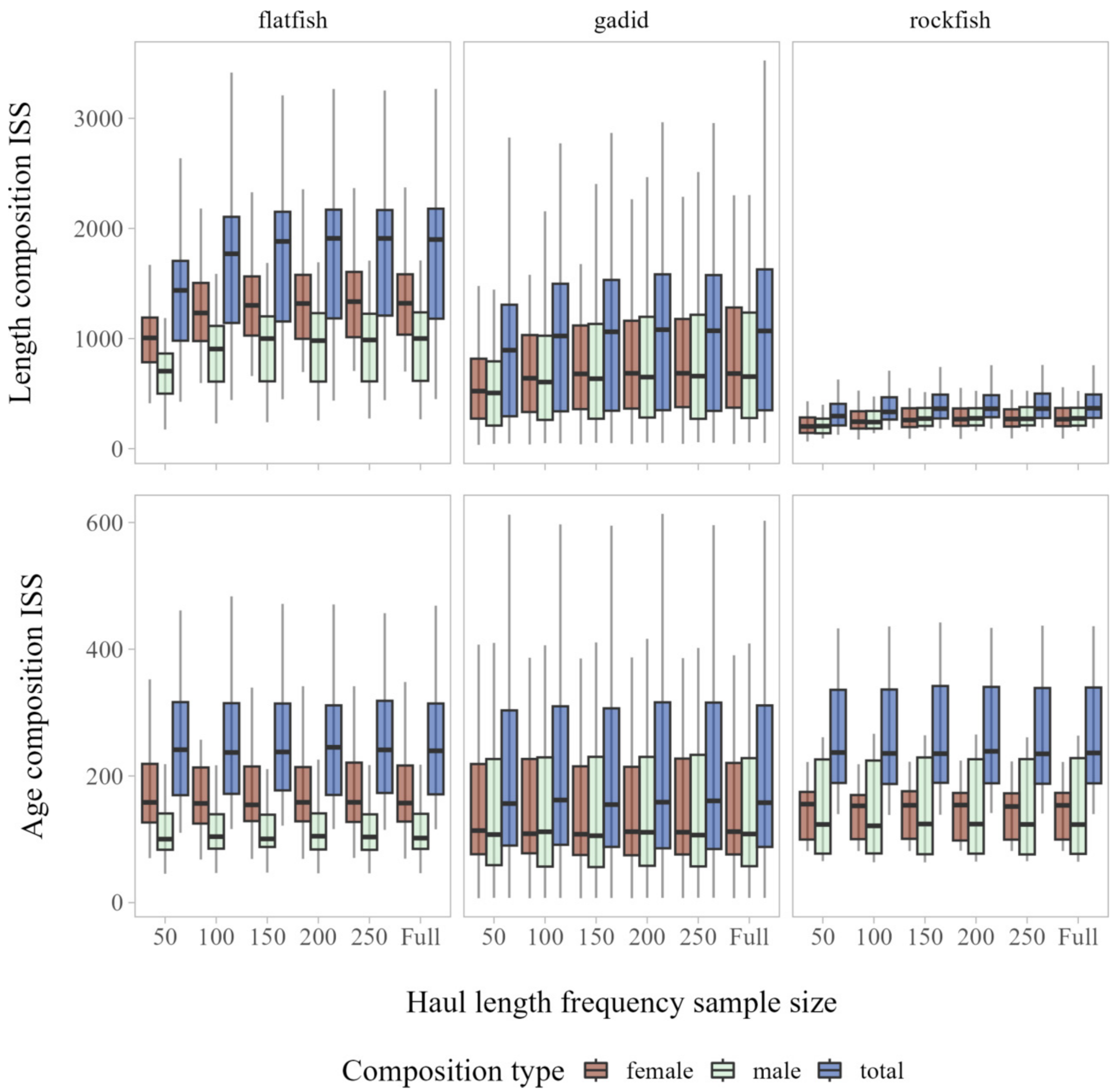
Results

As the number of length frequency subsamples taken within hauls decreased, the length composition ISS decreased, but not dramatically (e.g., moving right to left on the *x*-axis of the top panels in Fig. 2). This trend of decreasing length composition ISS associated with decreasing subsampling levels per haul for length frequency was observed across all regions, stocks, and sex categories evaluated (region- and stock-specific results are shown in Figs. S2–S4). The sex-specific length composition ISS were smaller than the total (combined sex) ISS for all stocks evaluated. This remained true for each individual subsampling case as well. However, stock-specific and species type differences in the magnitude of sex-specific ISS were observed (e.g., flatfish, including arrowtooth flounder and yellowfin sole). If the magnitudes were different, female length composition ISS was consistently larger than male length composition ISS. While the magnitude of length composition ISS was species type, stock, and sex-specific, across survey years, the length composition ISS generally ranged from 250 to 4000 for flatfish, 100 to 3500 for gadids, and 100 to 1000 for rockfish (including the full dataset case and all subsampling levels). The variability in length composition ISS across survey years and haul subsampling levels also varied between species types and stocks, but the variability generally declined as the haul subsampling level decreased. Using the full age specimen samples while changing the subsampling case for length frequency, the magnitude of age composition ISS for all stocks within each region was unaffected by length subsampling across the length frequency haul level subsampling cases evaluated (bottom panels of Fig. 2). We note that the magnitude of length composition ISS ranged from 5–10 times the magnitude of age composition ISS for flatfish and gadids, and two to three times the magnitude of age composition ISS for rockfish (comparing the top and bottom panels of Fig. 2).

As the haul level length frequency subsampling rate decreased, the length composition relative ISS also decreased (to panels of Fig. 3). This decrease in length composition relative ISS was exhibited by all stocks evaluated, and although the magnitude of decrease was region- and stock-specific, the results were consistent across regions and stocks (region- and stock-specific results are shown in Figs. S5–S7). For subsampling levels of 150 and greater lengths measured per haul, the variability in length composition relative ISS was small and consistent across all the species types and survey regions evaluated. Across the length frequency subsampling levels evaluated, the age composition relative ISS was minimally affected by the magnitude of subsampling in the length frequency data used for the first stage expansion (bottom panels of Fig. 3).

While the magnitude of decrease in the length composition relative ISS was survey region- and sex-specific, none of the stocks decreased below 50%, and most stocks did not decrease below 60% (top panels of Fig. 3). The range in the length composition annual relative ISS across the sex categories evaluated either included or was above 90% for length frequency subsampling levels greater than 100 fish per haul. EBS walleye pollock (*Gadus chalcogrammus*) was the one exception (Fig. S6) where the relative ISS was less than 90% for females and males. Besides EBS walleye pollock, in which the range in the relative ISS included 90%, the remainder of the stocks evaluated resulted in relative ISS greater than 90% for length frequency subsampling levels of 150 fish per haul or

Fig. 2. Boxplots by species type (aggregated across stocks and regions) of length composition (top panels) and age composition (bottom panels) input sample size across haul length frequency subsampling levels evaluated.



more regardless of length composition sex category (Figs. S5–S7). Sampling at a level of 100–150 fish per haul for length frequency would result in 7000–32 000 fewer collections per year for the surveys evaluated (Table 2).

Similar to the results at 1 cm bins for the length frequency data, at 2 and 5 cm bins, the length composition relative ISS decreased as the haul level length frequency subsampling rate decreased (Fig. 4, shown for GOA stocks as an example). When using 2 and 5 cm length bins, the length composition relative ISS was at or above 90% for length frequency subsampling levels greater than 100 fish per haul. We also find that the magnitude of length composition ISS decreased as the length bin increased, where for the same stock, the length composition ISS was smaller at a length bin of 5 cm compared to a length bin of 1 cm (Fig. S8). Also similar to the results of a 1 cm length composition bin, the age composition ISS for 2 and 5 cm bins was unaffected by the haul level length frequency subsampling level, and the magnitude of age composition ISS was unaffected by the size of the length composition bin (Fig. S9).

Length frequency subsampling at the subregional scale in the GOA produced results that were consistent with the broader region (Fig. 5). Regardless of the spatial scale of the length composition expansion (whether for the entire GOA or at the subregional scale of the GOA), the length composition relative ISS was at or above 90% for length frequency subsampling levels greater than 100 fish per haul. We do note that the magnitude of length composition ISS was smaller at the subregional scale compared to the larger region (Fig. S10); thus, there is an increase in uncertainty in length composition as the spatial scale is reduced. It continued to be the case at the subregional scale that the age composition ISS was unaffected by the level of length frequency subsampling (Fig. S11), and, similar to the length composition ISS, the age composition ISS was smaller at the subregional scale compared to the broader region.

Fig. 3. Relative length (top panels) and age (bottom panels) composition input sample size by species type (aggregated across stocks and regions) across haul length frequency subsampling levels evaluated (grey line at 0.9 for length composition and 1.0 for age composition shown for reference).

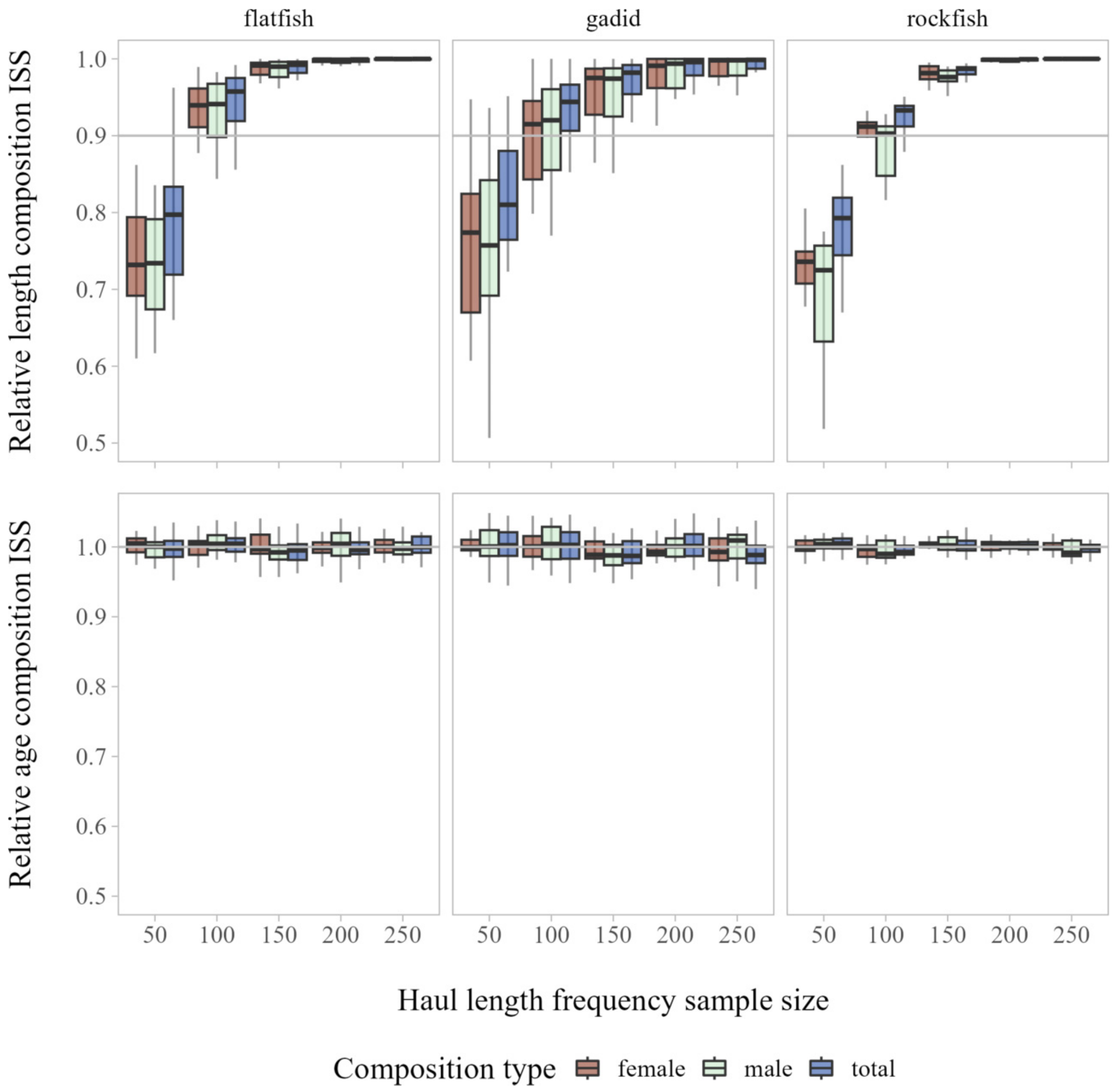
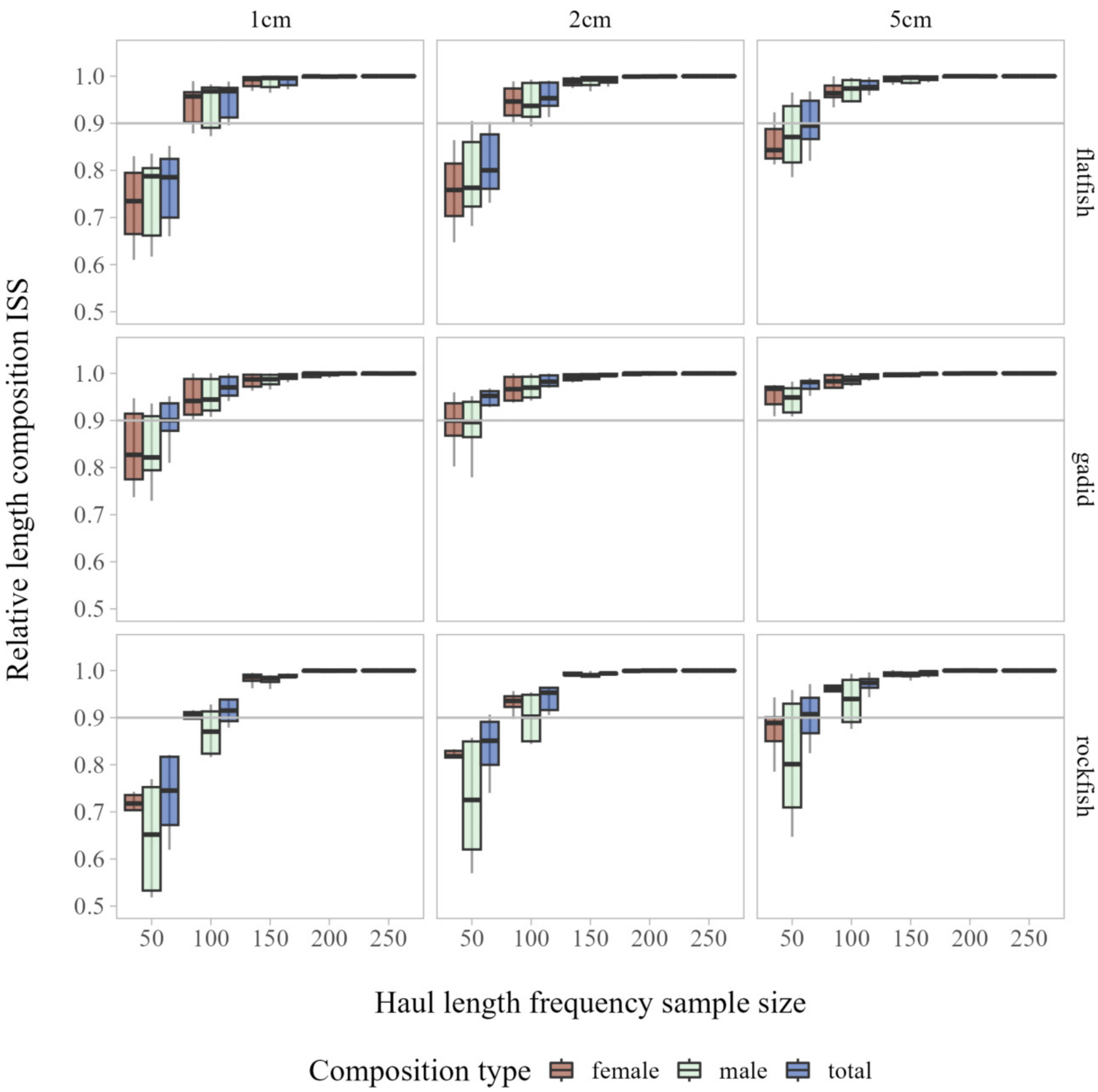


Table 2. Average reductions in total length frequency samples per survey when sampling 150 (left side) and 100 (right side) fish per haul by stock and region evaluated (rounded to the nearest 100s for stock-specific results to the nearest 1000s for total).

Stock	AI	EBS	GOA
Alaska plaice	–	200–700	–
Arrowtooth flounder	400–1200	600–2000	2900–8300
Atka mackerel	400–1700	–	–
Flathead sole	200–600	1000–3500	500–2200
Northern rock sole	–	1400–4400	100–400
Northern rockfish	1600– 4800	–	100–500
Pacific cod	100–500	200–800	100–200
Pacific ocean perch	4300–12 000	–	2200–7300
Rex sole	–	–	100–800
Southern rock sole	–	–	100–400
Walleye pollock	100 – 2500	5300–14500	2600–4700
Yellowfin sole	–	2000–6300	–
Total	7000–23000	11000–32000	9000 – 25000

Can. J. Fish. Aquat. Sci. **00:** 1–16 (2023) | dx.doi.org/10.1139/cjfas-2023-0164

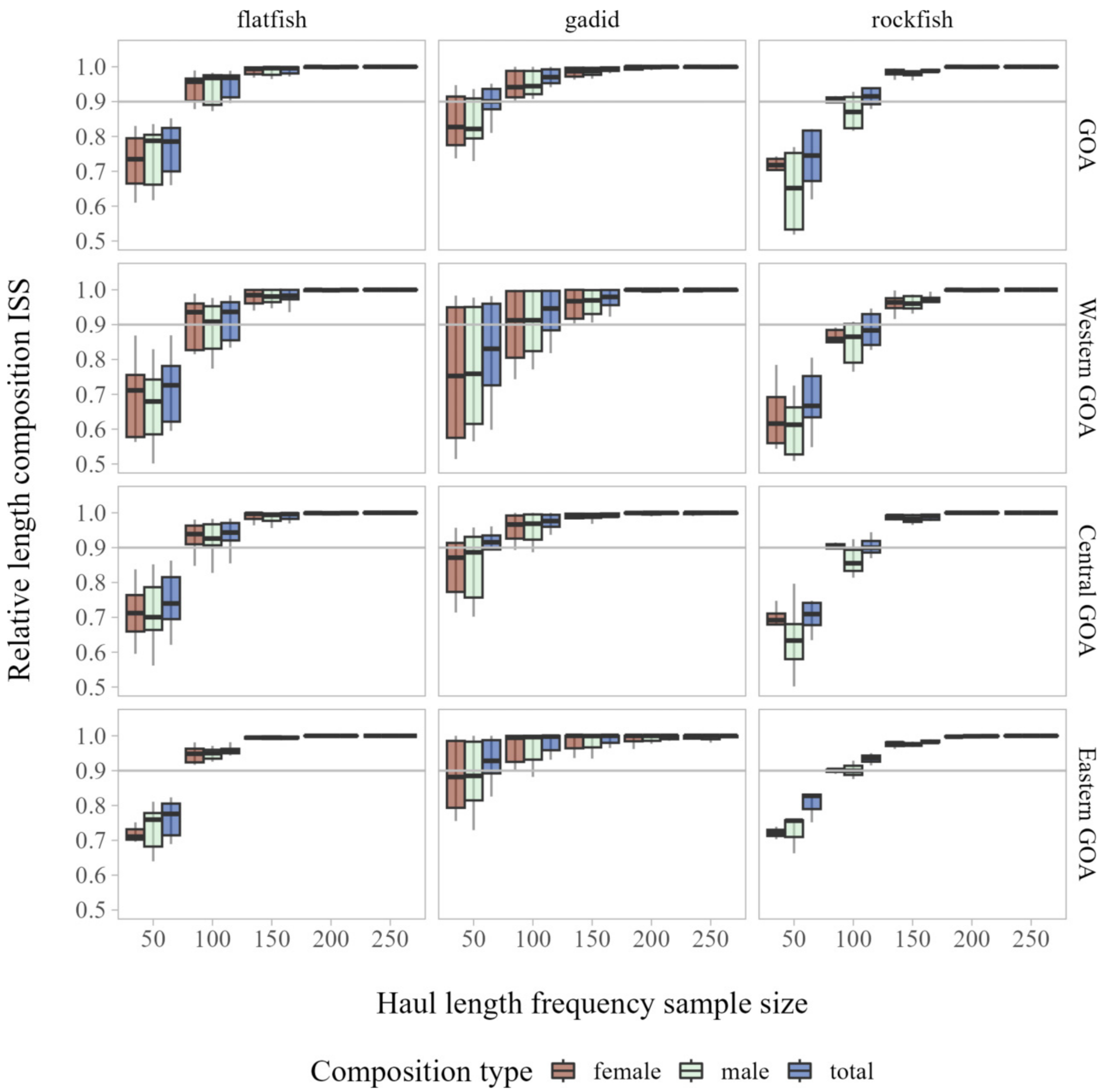
Fig. 4. Relative length composition input sample size by length bin and species type (aggregated across stocks within the Gulf of Alaska) across haul length frequency subsampling levels evaluated (grey line at 0.9 shown for reference).



While age composition ISS was unaffected by the length frequency subsampling level, the age composition ISS did markedly decrease as the magnitude of specimen age data decreased (e.g., moving right to left on the *x*-axis of the top panels of Fig. 6). We note that this result was consistent across all regions, stocks, and sex categories evaluated (Figs. S12–S14). Similar to the results for length composition ISS, the sex-specific age composition ISS were smaller than the total (combined sex) ISS across all stocks evaluated, and the magnitude of age composition ISS compared between sex categories (female and male) differed by species types. Following the results of the sex-specific length composition ISS, if the magnitudes in age composition ISS were different between the sex categories, the female age composition ISS was larger than the male age composition ISS (e.g., flatfish stocks). There could be a number of causes for the sex differences in length or age composition ISS flatfish. The smaller ISS for male flatfish indicates that there is more intrahaul correlation for males than females, or that the length composition from haul-to-haul is more variable for males than females. A number of processes could cause this, including sex-specific differences in natural mortality or spawning behavior leading to aggregations of males that are model similar in size or age, for example. Ultimately, regardless of the process, this results in availability differences to the survey between sexes. For the gadid and rockfish stocks evaluated, the magnitude of sex-specific age composition ISS was generally similar. The magnitudes of sex-specific and total ISS were stock-specific, and overall ranged from 10 to 500, which was smaller than the length composition ISS in the same year for the same stock. The variability in annual age composition ISS was also stock-specific, where some stocks displayed a larger range in age composition ISS across survey years than others.

As the age data sampling level decreased, the age composition relative ISS also decreased (bottom panels of Fig. 6). Even though there were stock- and region-specific differences in the age composition relative ISS across the age subsampling levels evaluated (Figs. S15–S17), the trends across species

Fig. 5. Relative length composition input sample size by species type and regional scale for the Gulf of Alaska (aggregated across stocks) across the evaluated haul length frequency subsampling levels (grey line at 0.9 shown for reference).



types and survey regions were generally consistent, where gadids displayed the largest range in the decrease in relative ISS within each survey region (Fig. 6). While the decrease in both length and age composition relative ISS was consistent as the subsampling level decreased, the relative length composition ISS plateaued beyond around 150 lengths per haul. The relative age composition ISS did not replicate this plateau, but rather continued to increase as the subsampling level increased.

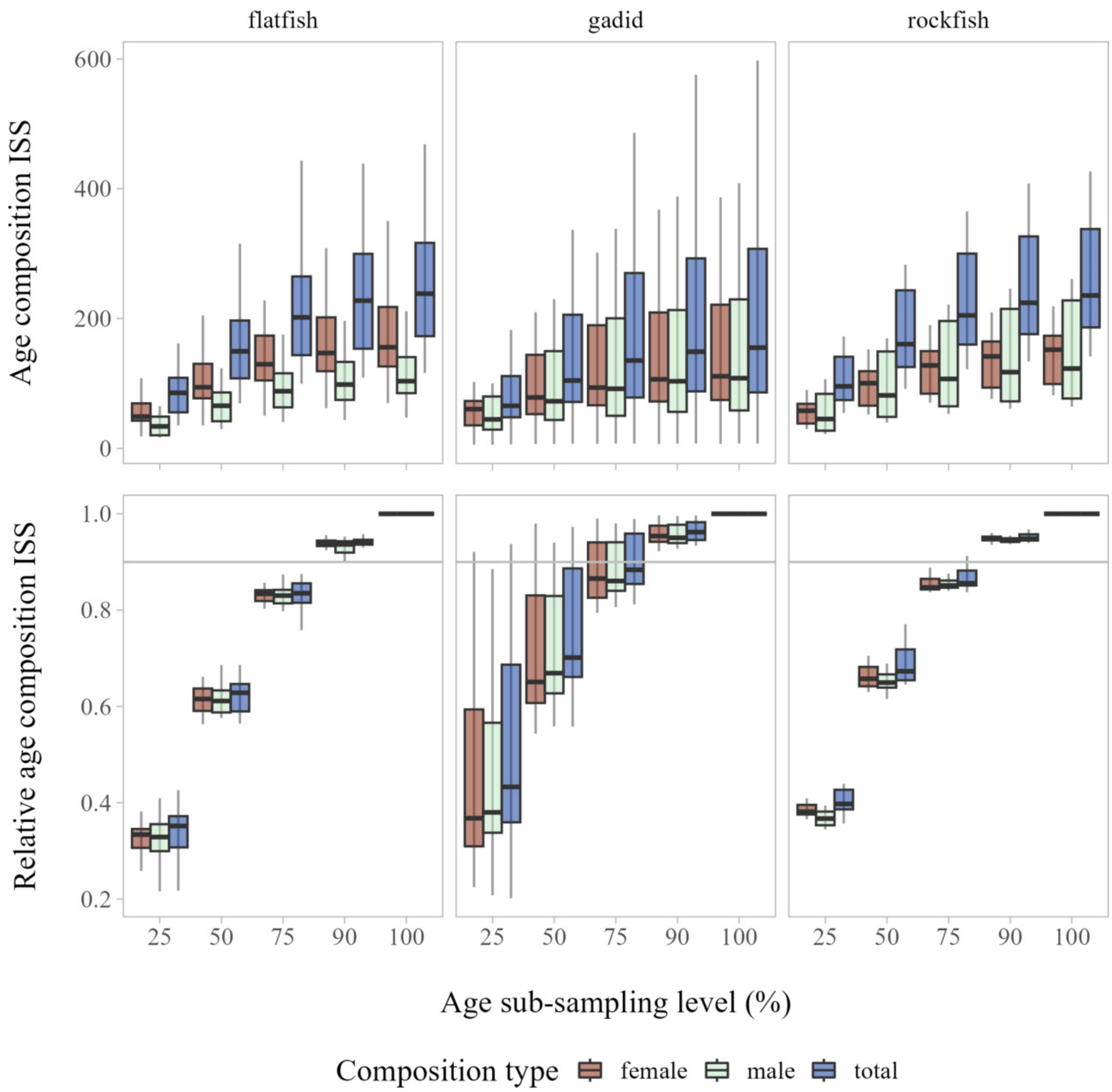
A notable result for all stocks evaluated was a dampened decrease in age composition relative ISS when compared to the specimen age subsampling level collected (Fig. 7). For example, a reduction to 25% of the specimen age sample data did not translate to an age composition relative ISS reduction to 25%, but rather it resulted in a subsampled dataset age composition relative ISS of 30%–40%; this was true at each age collection subsampling level evaluated. For rockfish and flatfish, the reduction in ISS compared to the reduction in specimen age samples was similar, both across stocks, survey regions, and sex categories. By comparison, for gadids and other (Atka mackerel), the reduction in age composition relative ISS compared to reduction in specimen age samples was further dampened, whereby the reduction in total age samples resulted in a proportionally smaller reduction in age composition ISS compared to rockfish and flatfish stocks (indicated by the slopes in the linear model fits in Fig. 7). Regardless, for all stocks and sex categories evaluated, the reduction in relative ISS was less than the reduction in the specimen age samples collected on the surveys.

At length frequency subsampling levels of 100 and 150 fish per haul, the relative bias for sex-specific mean length and sex ratio was centered at 0 (Fig. 8). This indicates that subsampling of length frequency did not cause any bias to emerge for any of these metrics.

Discussion

In this study, we developed a method to evaluate the consequences of reduced sampling effort on uncertainty in age and length composition data and demonstrated a decrease

Fig. 6. Boxplots by species type (aggregated across stocks and regions) of age composition input sample size (top panels) and relative age composition input sample size (bottom panels) across the evaluated age collection subsampling levels.



in the ISS from experimental treatments (i.e., increased uncertainty) compared to the full dataset base case due to this reduction in sampling. We found that the reductions in age sampling effort had a greater effect on the age composition ISS than reductions in length sampling effort had on the length composition ISS. Looking across multiple survey regions revealed that these reductions in ISS were stock-specific but the reductions were generally consistent. Based on the proportional decrease in ISS from subsampled data compared to the ISS generated from full datasets, we found that there are recommendations that can be made to balance the tradeoff of collecting a robust sample size of age and length compositions on fishery-independent platforms while also considering of the effort the crew must expend to obtain these samples.

One outcome of this study is guidance on how many length frequency collections are necessary to provide adequate information for stock assessment models. Gerritsen and McGrath (2007) suggested that sampling rates for length frequency be 10 fish per length category within a haul based on the mean weighted coefficient of variation statistic. In practical implementation, their suggestion results in stock-specific sampling rates at the haul level, where species that are larger (and presumably have more length bins) would require more samples than species that are smaller. While this strategy makes biological sense and is attractive as it is related to life history characteristics of the stocks being sampled, it is a more complicated sampling design compared to sampling at the same total rate for all the species caught. We also note that for a species like Pacific cod (*Gadus macrocephalus*), which can grow to upwards of 100 cm, if the sampling rate were 10 fish per 1 cm length bin, the resulting sample size (1000) would be much larger than what we have found to be necessary.

Based on the results of this study, we suggest a length frequency sampling method that sets the same sampling rate across stocks at the haul level, recognizing that length frequencies for some stocks may be oversampled. For example,

Fig. 7. Stock-specific relative age composition input sample size grouped by species type and region compared to age collection subsampling levels (1–1 line shown for reference; linear model fit by composition type for each species type also shown for reference).

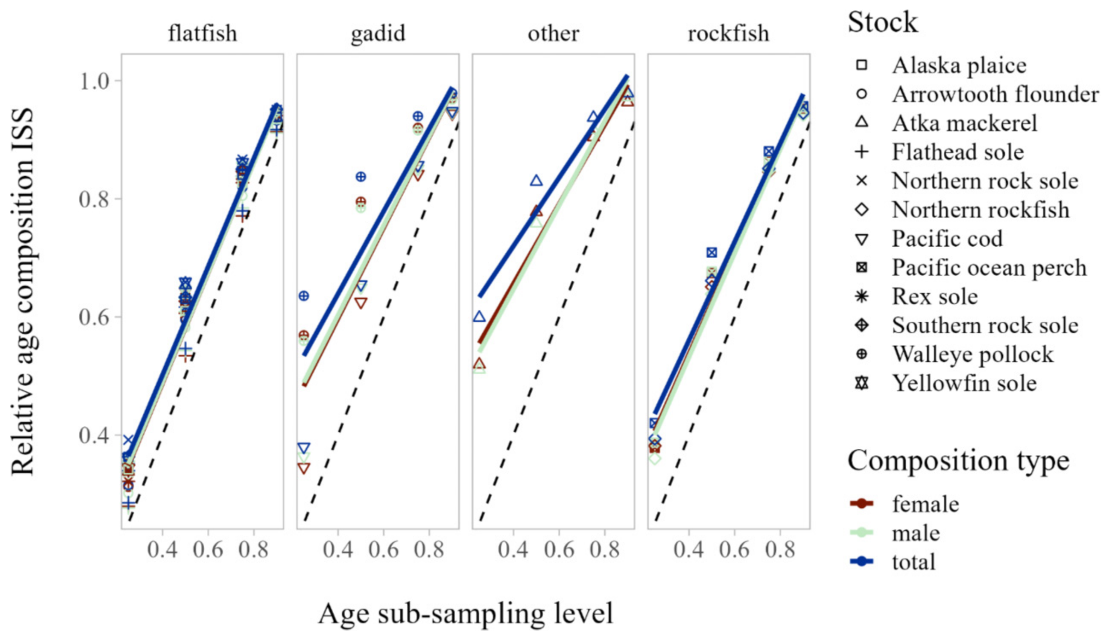
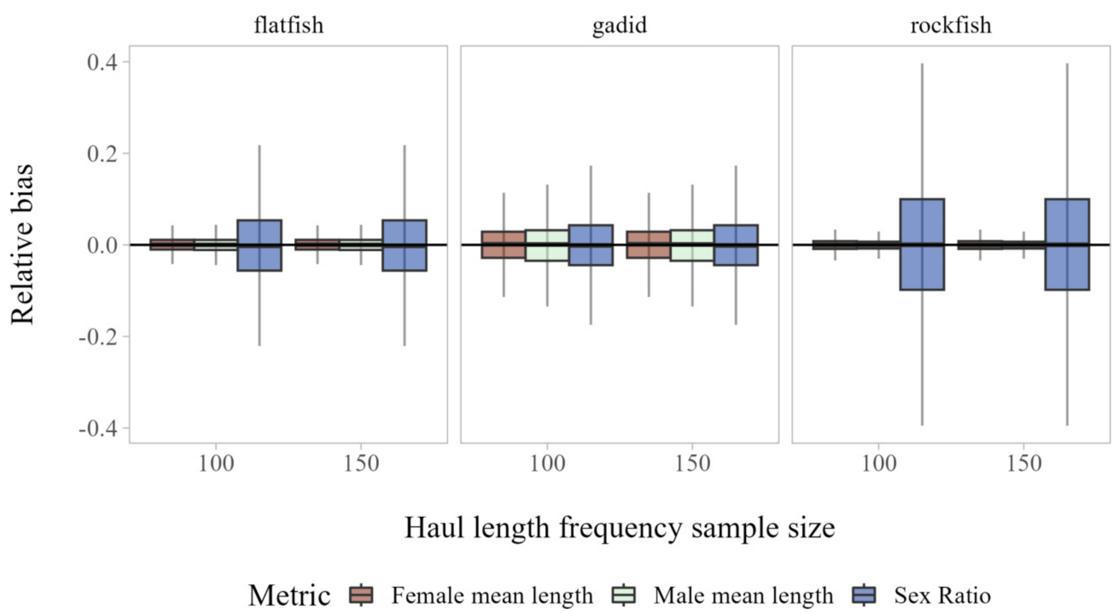


Fig. 8. Relative bias in sex-specific mean length and sex ratio by species type (aggregated across stocks and regions) for haul length frequency sample sizes of 100 and 150 fish per haul.



we found that limiting sampling at 100–150 fish per species within a haul (whether for total or sex-specific length composition) provided length composition data that had relatively similar uncertainty compared to length composition data derived from sampling at more intense rates (i.e., within 10% of the full dataset length composition ISS). Therefore, we conclude that there is little added benefit, in terms of reduced uncertainty in length composition data, beyond sampling limits of 100–150 fish lengths per haul. We also show that other metrics, such as mean length or the sex ratio of length samples, did not produce bias when subsampling 100–150 lengths per haul. Furthermore, limiting length frequency sampling at 100–150 fish produced consistent results regardless of the length bin implemented (1, 2, or 5 cm bins) or spatial survey regions. These results indicate that we are currently beyond a saturation point for length composition information on these surveys (i.e., a point of diminishing returns), whereby the current sampling rate does not offer much improvement in uncertainty over a lower sampling rate for composition data.

Can. J. Fish. Aquat. Sci. **00:** 1–16 (2023) | dx.doi.org/10.1139/cjfas-2023-0164

An interesting, and somewhat unexpected result of this analysis was that the uncertainty in age composition was largely unaffected by the amount of subsampling for length frequency. We show that the age composition ISS remained at the same magnitude for all the levels of length frequency subsampling (down to 50 lengths per haul), and that the age composition ISS was also unaffected by the bin structure in the length composition data (whether a 1, 2 cm, or 5 cm bin). We hypothesize that this result is due to the nature of the second stage expansion, where the length population estimates are converted to age population estimates through the implementation of an ALK. Within this expansion, the additional uncertainty in length composition due to subsampling is dampened by the ALK and is not translated to uncertainty in the subsequent age composition data. Thus, subsampling of length frequency data at the haul level can be achieved without consequence to the resulting age composition data.

Another application of our study is related to determining how many age collections are necessary to provide adequate information to stock assessments models. Based on the results of this analysis, we found that the answer to this question remains elusive. As we increased the subsampling rates for age composition, the age composition ISS continued to increase, indicating that a sample size producing diminishing returns had not been reached, unlike the asymptote observed for length composition ISS. In a similar study, Siskey et al. (2023) evaluated the effect of increased age sampling beyond current sampling levels for a subset of stocks and found that the age composition ISS continued to increase, up to age sampling levels 67% greater than current rates. Siskey et al. (2023) also found that, while holding the total age samples constant, improvements to age composition ISS were achieved by increasing sampling in more hauls even while decreasing the number of specimen age data collected in any given haul.

To determine whether and when age composition ISS reaches an asymptote, simulation studies might be more appropriate rather than bootstrapping historical data, as variables such as life history or intrahaul correlation can be accounted for specifically (e.g., Hulson et al. 2011; Xu et al. 2020). The results of this study and Siskey et al. (2023) suggest that the number of age samples necessary to maximize age composition information are likely beyond current management agency capacities. Our recommendation, therefore, is to attempt to retain, at a minimum, historical levels of sampling for age composition, as reductions in these sampling rates have downstream effects on assessment model uncertainty. Further research should be conducted through both simulation methods similar to those employed here and field sampling of an example species to determine whether collection of additional samples for age improves the information quality provided to an assessment. We also recommend (and reiterate the results of Siskey et al. 2023) that if sampling for ages were to be optimized, an effective approach would be to increase the number of hauls sampled while retaining the historical total number of age samples collected, effectively spreading these samples across more hauls and reducing the number of samples collected in any given haul.

Across the stocks evaluated, length composition ISS consistently increased as haul subsampling rates increased, with no clear species-specific patterns; however, there were interesting patterns within species group (e.g., flatfish, gadids, and rockfish) when examining reductions in age sampling rates. Overall, we found that the effect of decreasing age sampling rates was not one-to-one in relation to the rate of decrease in age composition ISS, but rather it had a dampened effect. For example, a 10% reduction in age sampling effort did not result in a 10% reduction in age composition ISS, but rather a 5%–8% reduction, depending on the stock. In addition, when comparing across species groups, we note that the effect of reductions in age sampling effort was relatively larger for flatfish and rockfish in comparison to gadids (and species such as Atka mackerel, *Pleurogrammus monopterygius*). That is, a reduction in age sampling effort for flatfish and rockfish results in a relatively larger reduction in age composition ISS compared to gadids. This result indicates that the intrahaul correlation in age composition for rockfish and flatfish, which can be related to patchy spatial distribution (particularly for rockfish), is large enough that a reduction in sampling effort has larger implications for uncertainty in comparison to gadids. A simpler explanation is that rockfish and flatfish are longer-lived than gadids and have more age categories to fill in age composition data, and thus, reductions in sampling have a larger impact on the resulting age composition due to missing information for a given age that was not sampled. A similar result was observed in Hulson et al. (2017), who found that age sample size had a relatively larger effect on the uncertainty in age composition data for rockfish, followed by flatfish, and then gadids (including Atka mackerel). Siskey et al. (2023) showed that increasing or decreasing age sampling effort for rockfish had a larger subsequent effect on the uncertainty in management quantities (e.g., overfishing limit) than the gadid example, indicating that this relatively larger effect on age composition ISS translates through the assessment model as well. While we encourage maintaining historical levels of age sampling effort, we recommend that if reductions must be made, decision-makers should consider the unequal consequences of these reductions across species types (i.e., make decisions on a species-specific basis), perhaps using an analytical tool like the one that was created for this study.

There were several extenuating circumstances that were revealed by these analyses, which should be considered when developing sex-specific and spatially explicit assessment models, as well as binning length composition data, as it relates to ISS for age and length composition data. First, in comparison to the total (combined sex) age and length composition ISS, we showed that the sex-specific composition ISS is smaller, and differences in the magnitude of ISS could exist between the two sexes (i.e., this resulted in an ISS that is larger for female age composition than for males for flatfish stocks). Thus, these differences in uncertainty for sex-specific age and length composition data should be considered when applying a sex-specific stock assessment model for management. Second, we showed that the magnitude of age and length composition ISS decreased as the data were split into subregions, as compared to the ISS of the entire region. This result was not surprising, but we also note that the magnitude of ISS was subregion-specific and presumably is related to the abundance and availability of the stock within the subregion. Overall, we note that there is a consequence for sex-specific and spatially explicit stock assessment models; uncertainty in age and length composition data increases for these model structures compared to assessment models that do not include these features. Finally, the results of our analysis revealed that the bin structure has implications on the magnitude of length composition ISS, where the ISS decreased (i.e., the uncertainty increased) as the bin size increased. We hypothesize that this result is a consequence of uncertainty as it relates to the coarseness of the length composition, where more uncertainty is produced for a smaller number of length bin categories.

Overall, while we have shown that the recommendation to limit subsampling for length frequency data at 100–150 fish per haul was robust to bin structure and spatial scale, we also recommend that analysts in other regions evaluate their data in a similar manner as we have to determine whether oversampling is occurring, as our results may be specific to surveys conducted by AFSC. However, we suggest that these results are general to the expansion methods employed in other regions, where nuanced differences will not have an impact to such a degree that would appreciably change these results. In terms of application to other stocks and the sample sizes that would be necessary to conduct this analysis, we would recommend that the stocks evaluated have a reasonable amount of hauls in which the length frequency sampling exceeds 100 fish. We would also recommend that stocks evaluated be sampled representatively across their distribution range (i.e., all length samples do not come from a small portion of the survey area or range of distribution for the species). For application to age composition, we suggest that whether the ages are collected randomly or stratified, expansion methods like those used here be applied (to account for different sampling strategies) when conducting analysis to determine consequences of subsampling. Finally, a universal concern when considering subsampling is the opportunity cost of not having collected samples that may be needed to address some potential future analysis. For length frequency sampling at the haul level, our results suggest that limiting sampling to 100–150 fish per haul provides adequate information to conduct current and future analyses. Our results consistently show that subsampling at these levels produces unbiased results and that no added benefit to length composition uncertainty is achieved beyond these sampling levels regardless of bin or spatial structure of the data.

While this study focused on evaluating the statistical consequences of effort reduction when collecting age and length samples on bottom trawl surveys, our stated goal was also to address the tradeoffs between stock assessment input data uncertainty and workforce health and efficiency. On fishery-independent surveys, each fish that is handled to determine sex and length represents physical efforts of filling baskets that can weigh approximately 30 kg, lifting these baskets, moving them across an unstable deck to various data collection stations multiple times, grabbing each fish to make an incision, and then handling these fish to record length measurements. We showed that the combined sex and sex-specific length frequency sampling can be reduced from current sampling levels without major consequences to length composition uncertainty. Table 2 outlines that collecting length observations at levels less than historical sampling may result in 7000–32 000 fewer collections per year for the surveys evaluated, which would amount to a substantial decrease in repetitive motions for survey teams. Because the average number of survey age collections are generally an order of magnitude less than the average number of lengths sampled on a survey (Table 1), reducing age sampling efforts may not decrease the overall repetitive motions of survey teams, although gripping a fish to cut through its cranial plate can require more force and awkward wrist movements to collect these samples. Currently, studies evaluating the monetary consequence of oversampling that causes repetitive motion injuries are non-existent, but we do note that these injuries exist in survey teams (as well as for fishery observers), and in some cases have led to Workers' Compensation claims (Stan Kotwicki, personal communication, 14 December 2021). It would be difficult, if not impossible, to predict the savings in terms of health or monetary consequences of repetitive motion injuries by subsampling at the levels we have recommended here, but we suggest that any reductions in effort would have at least some benefit to reducing these types of injuries. Of course, one should recognize that it would be impossible to completely remove these risks, as survey data need to continually be collected and there will always be some level of physical effort required to collect these samples.

In this study, we demonstrate a method for evaluating tradeoffs between stock assessment needs and workforce health and efficiency and recommend that fishery-independent survey groups collaborate with stock assessment scientists to determine how age and length collections can be optimized. Future work to evaluate the consequences of reducing, or increasing, composition data collections should include developing simulation methods that can directly evaluate variables such as life history characteristics and intrahaul correlation levels.

Acknowledgements

We thank the remaining members of the AFSC Survey Workload Optimization Working Group for their contributions to this work: William T. Stockhausen, Susanne F. McDermott, and W. Chris Long. We express our deep gratitude to all the AFSC staff who have spent time at sea and staring through microscopes to produce the data these analyses are based upon. We thank Lewis Barnett, Cecilia O'Leary, and Paul Spencer for helpful reviews of previous versions of this manuscript. We thank two anonymous reviews for their helpful and insightful comments that helped to greatly improve this manuscript.

Article information

History dates

Received: 8 June 2023
Accepted: 11 September 2023
Accepted manuscript online: 26 September 2023
Version of record online: 14 November 2023

Data availability

An example of the raw data used in this study to run exploration of the presented analysis is provided in the GitHub repository "swo" (https://github.com/BenWilliams-NOAA/swo). The output from the "swo" package that are presented in this paper are found in the "output" folder within the "swo-journal" repository (https://github.com/pete-hulson/swo-journal).

Author information

Author ORCIDs

Peter-John F. Hulson https://orcid.org/0000-0001-5651-2740
Benjamin C. Williams https://orcid.org/0000-0001-7295-2076
Meaghan D. Bryan https://orcid.org/0000-0003-1196-7930
Jason Conner https://orcid.org/0000-0001-9913-7805
Matthew R. Siskey https://orcid.org/0000-0002-7834-4749

Author contributions

Conceptualization: PFH, BCW, MDB, JC, MRS
Data curation: PFH, BCW, MDB, JC
Formal analysis: PFH, BCW, MDB
Investigation: PFH, BCW, MDB
Methodology: PFH, BCW
Project administration: PFH
Software: PFH, BCW
Supervision: PFH
Validation: PFH, BCW
Visualization: PFH, BCW
Writing – original draft: PFH, BCW, MDB, JC, MRS
Writing – review & editing: PFH, BCW, MDB, JC, MRS

Competing interests

The authors declare there are no competing interests.

Funding information

The authors declare no specific funding for this work.

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/cjfas-2023-0164.

References

Ailloud, L.E., and Hoenig, J.M. 2019. A general theory of age-length keys: combining the forward and inverse keys to estimate age composition from incomplete data. ICES J. Mar. Sci. **76**(6): 1515–1523. doi:10.1093/icesjms/fsz072.

Council, N.R. 1998. Improving fish stock assessments. The National Academies Press, Washington, DC. doi:10.17226/5951.

Francis, R.I.C.C. 2017. Revisting data weighting in fisheries stock assessment models. Fish. Res. **192**: 5–15. doi:10.1016/j.fishres.2016.06.006.

Gerritsen, H.D., and McGrath, D. 2007. Precision estimates and suggested sample sizes for length-frequency data. Fish. Bull. **105**(1): 116–120.

Hulson, P.-J.F., Hanselman, D.H., Quinn, I., and Terrance, 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**(7): 1548–1557. doi:10.1093/icesjms/fsr102.

Hulson, P.-J.F., Hanselman, D.H., and Quinn, T.J., II 2012. Determining effective sample size in integrated age-structured assessment models. ICES J. Mar. Sci. **69**: 281–292. doi:10.1093/icesjms/fsr189.

Hulson, P.-J.F., Hanselman, D.H., and Shotwell, S.K. 2017. Investigations into the distribution of sample sizes for determining age composition of multiple species. Fish. Bull. **115**: 326–342. doi:10.7755/FB.115.3.4.

Hulson, P.-J.F., Williams, B., Siskey, M., Bryan, M., and Conner, J. 2023. Bottom trawl survey age and length composition input sample sizes for stocks assessed with statistical catch-at-age assessment models at the Alaska Fisheries Science Center. U.S. Department of Commerce. NOAA Technical Memorandum NMFS-AFSC-470: 38p.

Lauth, R.R., Dawson, E.J., and Conner, J. 2019. Results of the 2017 eastern and northern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate fauna. U.S. Department of Commerce. NOAA Technical Memorandum NMFS-AFSC-396: 260p.

Link, J., Burnett, J., Kostovick, P., and Galbraith, J. 2008. Value-added sampling for fishery independent surveys: don't stop after you're done counting and measuring. Fish. Res. **93**: 229–233. doi:10.1016/j.fishres.2008.04.011.

Maunder, M.N., and Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. **72**(1): 7–18. doi:10.1093/icesjms/fsu015.

McAllister, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Can. J. Fish. Aquat. Sci. **54**(2): 284–300. doi:10.1139/f96-285.

Miller, T.J., and Skalski, J.R. 2006. Integrating design- and model-based inference to estimate length and age composition in North Pacific longline catches. Can. J. Fish. Aquat. Sci. **63**(5): 1092–1114. doi:10.1139/f06-022.

Pennington, M., Burmeister, L.M., and Hjellvik, V. 2000. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fish. Bull. **100**(1): 74–80.

Punt, A.E., Huang, T., and 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**(1): 16–33. doi:10.1093/icesjms/fss185.

Quinn, T., and Deriso, R. 1999. Quantitative fish dynamics. Oxford University Press, New York, NY.

R Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/ [accessed 14 September 2023].

Siskey, M.R., Punt, A.E., Hulson, P.-J.F., Bryan, M.D., Ianelli, J.N., and Thorson, J.T. 2023. The estimated impact of changes to otolith field-sampling and ageing effort on stock assessment inputs, outputs, and catch advice. Can. J. Fish. Aquat. Sci. **80**(1): 115–131. doi:10.1139/cjfas-2022-0050.

Stauffer, G. 2004. NOAA Protocols for Groundfish Bottom Trawl Surveys of the Nation's fishery resources. U.S. Department of Commerce. NOAA Technical Memorandum NMFS-F/SPO-65: 205p.

Stewart, I.J., and Hamel, O.S. 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. Can. J. Fish. Aquat. Sci. **71**(4): 581–588. doi:10.1139/cjfas-2013-0289.

Thorson, J.T. 2014. Standardizing compositional data for stock assessment. ICES J. Mar. Sci. **71**(5): 1117–1128. doi:10.1093/icesjms/fst224.

Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. **192**: 84–93. doi:10.1016/j.fishres.2016.06.005.

Thorson, J.T., Monahan, C.C., and Hulson, P.-J.F. 2023. Data weighting: an iterative process linking surveys, data synthesis, and population models to evaluate mis-specification. Fish. Res. **266**. doi:10.1016/j.fishres.2023.106762.

von Szalay, P.G., and Raring, N.W. 2018. Data Report: 2017 Gulf of Alaska bottom trawl survey. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-374: 260p.

von Szalay, P.G., Raring, N.W., Rooper, C.N., and A, L.E. 2017. Data Report: 2016 Aleutian Islands bottom trawl survey. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-349: 161p.

Xu, H., Thorson, J.T., and Methot, R.D. 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. Can. J. Fish. Aquat. Sci. **77**(2): 247–263. doi:10.1139/cjfas-2019-0107.

Zimmermann, M., Dew, C.B., and Malley, B.A. 2009. History of Alaska red king crab, *Paralithodes camtschaticus*, bottom trawl surveys, 1940–61. Mar. Fish. Rev. **71**(1): 1.