Effort optimization of bottom trawl subsampling of fish for length and age composition: tradeoffs between stock assessment input sample size and survey workforce health and efficiency -or- Reductions in sampling effort for fishery-independent age and length composition: balancing stock assessment input data uncertainty and workforce health and efficiency

Pete Hulson1,\*, Benjamin Williams1, Meaghan Bryan2, Jason Conner3, and Matthew Siskey4

1 Auke Bay Laboratories, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 17109 Point Lena Loop Rd., Juneau, AK 99801  
2 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  
3 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  
4 School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

\* Correspondence: [Pete Hulson <[pete.hulson@noaa.gov](mailto:pete.hulson@noaa.gov)>](mailto:pete.hulson@noaa.gov)

# Abstract

Unavoidable survey effort reduction has become a reality that must be accounted for in modern stock assessment models, and in addition, increasing negative consequences to survey staff health due to repetitive motion injuries are becoming a cost to managing agencies that is increasingly significant. In this study, we evaluated the consequences of reductions in age and length data collections for composition data that is used in stock assessment models in order to determine if a balance can be met between reducing sampling that does not excessively increase data uncertainty but provides for reductions in repetitive motions that can cause injury to survey staff. We find that reducing sampling to a maximum of 100 to 150 fish sampled per haul for length (either sex-specific or combined sex) provides for length composition data for which the uncertainty is not appreciably increased, and has no effect on the uncertainty in age composition data that is subsequently expanded from this sub-sampled length frequency data. We find that as the subsampling rate of total age collections is increased, the uncertainty in the age composition also increases, but not at a rate proportional to the reduction in sampling effort. While there is a direct link between the uncertainty in data inputs and the uncertainty of assessment model predictions, reductions in length frequency sampling can be made without large consequence, but there will be, at some level, consequence to model output with reductions in age collection sampling. The method employed here, and the results presented, can aid managing agencies in balancing the magnitude of data collection on fishery-independent survey platforms and subsequent consequences to assessment models.

# Introduction

Many integrated fishery stock assessments rely on estimates of fish population demographic composition information 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 (Council 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 should require the maximum number of samples per each survey observation. However, to increase the precision of age or length composition estimates it’s not just the magnitude of sampling that is influential, but also the manner in which the samples are collected (Siskey et al. 2023). Further, 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 that read otoliths to determine age), and budget costs (e.g., the monetary cost to perform a fisheries-independent survey).

It has become commonly understood that sampling to determine age and length composition on fishery-independent or fishery-dependent platforms can be influenced by intra-haul correlation, or, that samples taken from a school of fish that are very similar to each other in size and/or age may not be representative of the overall population’s size and/or age distribution (e.g., Pennington et al. 2000). In order to evaluate and identify the level of intra-haul correlation, and how this can be accounted for in stock assessment, the concept of ‘effective sample size’ has been developed, in which the effective sample size is smaller than the actual sample size and reflects the increase in uncertainty that is due to the intra-haul correlation displayed by the species that are sampled (e.g., McAllister and Ianelli 1997). A number of studies have used effective sample size to evaluate the impact on assessment results including whether effective sample size can be estimated by the assessment model (Hulson et al. 2012, Francis 2017, Thorson et al. 2017) and more recently as a tool to evaluate the implications of changes to sampling methodologies and the subsequent influence on assessment model results (Siskey et al. 2023). Further, methods have been developed that mimic the sampling design for length and age composition that estimate effective sample size from collected samples trough bootstrap techniques that are external to an assessment model (Stewart and Hamel 2014). Overall, the use of effective sample size has become a universal method to implement uncertainty caused by overdispersion of samples in length and age composition data.

Within a fishery-independent survey, thousands of samples are obtained 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) is responsible for the execution of several fisheries-independent bottom trawl surveys (Stauffer 2004), spanning most of the continental shelf in Alaskan waters 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). Within the quantitative time series, AFSC scientists have routinely collected observations of fish length and age composition, consisting of complete or random subsamples of fish within each trawl. There are a number of unintended consequences associated with intense sampling of species within a fishery-independent survey. The length measurement process is repeated over hundreds of thousands of times in a given survey each year, while collecting otoliths is repeated several thousands of times, representing a daunting amount of work for the field scientists working on research vessels. Further, a large portion of the length frequency data that is collected is subsequently sampled for sex determinations in order to support stock assessments that are sex-specific, which involves a substantial increase in handling and data collection effort. Each year, this work flow results in acute and repetitive stress injuries, some requiring medical interventions and claims to the U.S. Office of Workers’ Compensation Programs. Another consequence of the intensity of this work are unrecoverable errors in the observed data; as fatigue or injuries accumulate during the course of field work, so do data collection errors (e.g. failure to properly encode length measurements with the correct species or sex or incorrect identification of sex), despite extensive *in situ* quality assurance protocols. Besides the budgetary constraints that are directly associated with operationalizing a fishery-independent survey, there are downstream fiscal costs associated with composition collections, in particular, requiring properly trained staff to read otoliths collected across a number of different species.

While maximizing the number of samples collected for age and length composition is desirable from a statistical viewpoint, it is becoming increasingly the case that optimizing collections that balance the statistical quality of the data with the health of the workforce and budgetary constraints is more broadly desirable. And in this case, optimizing collections is more often associated with decreasing collection efforts from historical levels. From the stock assessment perspective, the primary concern when reducing the sampling of length frequency data (whether sex-specific or not) and age composition data collections is the impact on the uncertainty in subsequent age and length composition data. The use of effective sample size, particularity when evaluating sampling strategies, is a useful tool to determine the consequences of changing sampling effort. An additional consideration when evaluating the consequences of optimizing sampling effort is the impacts on a species-by-species basis to determine whether there are life-history characteristics that minimize the impacts of reduction in sampling effort in one species compared to another.

In this study, using data collected by the AFSC bottom trawl surveys across the EBS, AI, and GOA we evaluate the consequences of reductions in length frequency collections (including sex-specific length frequency) coupled with reductions in age composition sampling. Using effective sample size as the primary statistic to evaluate uncertainty in the age and length composition data that is derived from the bottom trawl surveys, we evaluated the impact of reducing sampling to answer three questions 1) what is the impact of reducing sampling in length frequency sampling (including sex determination sampling) and the subsequent expansion to age composition on the uncertainty in the composition data? 2) What is the impact of reducing sampling for age on uncertainty in age composition data? 3) Is there a point of diminishing returns as we increase the number of age and length samples? And, 4) are there life-history characteristics that mitigate, or exaggerate, the consequences of reductions in age or length composition sampling?

# Methods

## Computing length and age composition from bottom trawl survey data

Data collection for each AFSC groundfish bottom trawl survey is described in respective NOAA Technical Memorandums (EBS: Lauth et al. 2019, AI: von Szalay et al. 2017, GOA: von Szalay and Raring 2018). Fundamental methods of length sample collection are generally synchronized between these surveys, with species-specific exceptions for minimum length sub-sample size. Observations of sex-specific length distributions for designated species encountered within a catch sample are collected by 1) sorting the trawl sample by species, 2) weighing each species in aggregate, 3) obtaining a random sub-sample of target sample size, 4) sorting the sub-sample by sex (each fish is cut with a scalpel, gonads are identified and placed in a sex-specific receptacle), and 5) measuring and recording each fish length (each fish is placed on a measuring board, length is identified and the species, sex and measurement are recorded on a computer). To facilitate age estimation, individual fish are processed at sea to record sex, length and weight and to remove sagittal otoliths that are returned the AFSC Age and Growth laboratory for age determination. Survey age sampling protocols are specific by fish species and follow 1 of 2 paradigms: 1) a stratified collection that is distributed over both the spatial frame of the stratification scheme and the expected size range of a species; or 2) a small subsample (3-6 fish, depending on species) collected randomly per trawl. The protocol for some species has changed over the time series, which has followed a trend of transitioning from protocol 1) to protocol 2).

Length frequency samples collected by the AFSC bottom trawl surveys are expanded by area-swept type 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 is a common method to obtain population estimates at length from area-swept survey data (Miller and Skalski 2006, Ailloud and Hoenig 2019). 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 which have similar bathymetric characterstics, in particular, depth ranges, and population abundance-at-length with strata can also be summed to any sub-region level. 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). Currently, no stocks age composition are estimated using depth or area stratified ALKs, rather, all observations are pooled for the entirety of the survey area each survey year. The specific methods AFSC uses to expand length and age samples to abundance are shown in Hulson et al. (in review). The stocks selected for this analysis are shown in Table 1), and were selected because all are assessed by AFSC with statistical catch-at-age models and have corresponding expanded age and/or length composition estimates from the respective bottom trawl surveys. Historical data for each survey were used in this analysis in order to show results that reflect consequences of sub-sampling. To show the recent sampling magnitudes, the average length and age sample sizes from the most recent three survey years for the stocks selected by survey are shown in Table 1).

## Simulation-Bootstrap framework

To evaluate the effect of reductions in sampling length frequency and age collections we developed a bootstrap-simulation framework that 1) allows for reductions in the historical number of length frequencies and age specimen data collected and 2) conducts the first (length) and second (age) stage expansion process for each bootstrap replicate of length frequency and age specimen information to generate length and age composition. 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. Functions to run the sampling protocols were developed in a compartmentalized manner to provide for substantial flexibility in exploring desired resampling protocols. The order of operations (Figure 1 *ben - could we add in a part that sub-samples ages too to this figure?*) has the following schedule:

1. Resample hauls (wr) from the set of hauls with associated catch per unit effort (in numbers).
2. Within the resampled hauls from step 1, resample the observed lengths (wr).
3. From the resampled lengths in step 2, within each haul subset the lengths (wor) at pre-determined subsampling level.
4. With the resampled and subsampled lenth 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, for the total ages sampled across hauls subset the ages (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 that are collected on a per haul basis, as well as in the aggregated sample size for ages. In this bootstrap-simulation, the number of lengths (whether total or sex-specific) in a given haul must be less than or equal to the desired sample size *x* determined in step 2. In step 3, when the number of resampled lengths from step 2 in a haul is less than *x* then is used directly, if then a random draw of lengths is taken without replacement in step 3. We set the subsampling level for length frequency at numbers per haul in order to evaluate the consequences of reductions in length sampling at the haul level. Alternatively, to subsample ages we set the proportion of the total number of ages sampled in step 6 in order to evaluate the consequences of reductions in overall age sampling, as by design at the per haul bases there are a small number of ages sampled. The bootstrap-simulation then repeated steps 1-7 iteratively for each length and age subsample level, providing iterated population abundance-at-length and age that was then compared to the historical (the full sample without any resampling of data) population abundance-at-length and age determined by the bottom trawl surveys.

The length subsample levels that we evaluated from the historical length frequency collections were 50, 100, 150, 200, and 250 samples per haul. The age subsample levels that we evaluated from the historical age specimen collections were 25%, 50%, 75%, and 90% of the total number of ages collected in any given survey year. We also ran the bootstrap-simulation for the historical number of length frequency and age specimen collections without subsetting in order to compare the base level uncertainty to the increase in uncertainty gained through sub-sampling. We ran the bootstrap-simulation for 500 iterations, which was a level for which the variability in population abundance at length results had stabilized. 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 effective and input sample size

We used two performance metrics to quantify potential changes in uncertainty of length and age compositions (i.e., population abundance-at-length and age) owing to reduced sampling of length frequency and age specimen information. First, effective sample size was used to compare each bootstrap replicate of length and age composition to the original length and age composition calculated from the complete historical dataset. Second, input sample size was used to summarize uncertainty across bootstrap replicates as the harmonic mean of effective sample size replicates.

Effective sample size, as introduced by McAllister and Ianelli (1997), is a statistic that can evaluate the level of intra-haul correlation in composition samples that are collected on a survey (whether from age or length frequency collections). It is also a statistic that can evaluate the amount of uncertainty in an estimated composition compared to an observed composition. Effective sample size is given by:

where is the estimated proportion for category-*c* (which can be either age or length or any other arbitrary category across which proportions are computed) and is the observed proportion. It can be interpreted by a higher ESS indicates less uncertainty in the composition estimates, while lower ESS indicates more uncertainty.

In this bootstrap-simulation the underlying length and age composition derived from the historical bottom trawl surveys was treated as the observed proportions in equation (1). For each iteration of the bootstrap simulation for a determined sub-sample size or proportion we computed an estimated proportion () that was then compared to the underlying historical sex-specific length and age composition (the effective sample size for the total length and age composition, as the sum of population abundance at length, was also computed). Thus, across each iteration of the bootstrap simulation we computed an effective sample size that indicated the amount of increased uncertainty that was caused by sub-sampling length frequency and age specimen data. To summarize effective sample size across iterations we used the harmonic mean, which has been shown to reduce bias in recovering the true sample size in simulations for a multinomial distribution. Due to this reduction in bias the harmonic mean has also been recommended to determine the ‘input sample size’ that is used in stock assessment models to fit compositional data (Stewart and Hamel 2014). Herein, when we use the term ‘iterated effective sample size’ (iterated ESS) we are referring to the effective sample sizes that were computed for each iteration of the bootstrap-simulation, when we use the term ‘input sample size’ (ISS) we are referring to the harmonic mean of the iterated effective sample sizes. Here we show both results of the annual ISS estimates and the proportion of of the full sample ISS for each sub-sampling case evaluated for either age or length composition sampling. The proportion of full sample ISS was computed by pairing ISS as computed for the full dataset (with no sub-sampling) and the ISS for the dataset at the pre-determined sub-sampling level, and then divide the sub-sampled data ISS by the full dataset ISS for each survey year. This provides the proportion, or percent, of the full dataset ISS for the ISS at the sub-sampling level evaluated, which gives an indication of the amount of uncertainty that was increased by sub-sampling.

# Results

As the number of samples taken within hauls for length frequency decreased, the ISS decreased, but not dramatically (Figure 2). While we show the EBS stocks as an example, this trend of decreasing length composition ISS with decreasing haul sub-sampling level for length frequency also resulted across all regions, stocks, and sex categories evaluated. For all stocks evaluated the sex-specific length composition ISS were smaller than the total (combined sex) ISS, this remained true within each individual sub-sampling case as well. There were, however, stock-specific differences in the sex-specific ISS magnitude; if the magnitudes were different it was the case that female length composition ISS was larger than male length composition ISS. While the magnitude of length composition ISS was stock-specific and sex category-specific, across survey years the length composition ISS generally ranged from 250 to 3,000. The variability in length composition ISS across survey years and haul sub-sampling levels was also stock-specific, but the variability generally reduced as the haul sub-sampling level decreased. Across the length frequency haul level sub-sampling cases evaluated, the magnitude of age composition ISS for all stocks within each region was unaffected by length sub-sampling (Figure 3, shown for AI stocks as an example).

As the sub-sampling level within a haul for length frequency decreased the proportion of the full dataset length composition ISS decreased (Figure 4), following from the results of the decreasing trend in ISS magnitude across sub-sampling levels. This decrease in the proportion of length composition sub-sampled ISS compared to full dataset ISS was consistent across the stocks evaluated, although the magnitude of decrease was region and stock-specific. The largest decrease in the proportion of full dataset ISS across the sub-sampling levels evaluated resulted for EBS walleye pollock, where the smallest decrease resulted for GOA rex sole and southern rock sole. The largest variability across survey years in the proportion of full dataset ISS with the sub-sampling levels of 50 and 100 lengths per haul resulted for AI walleye pollock and GOA northern rockfish. For sub-sampling levels of 150 and greater the variability the the proporiton of full dataset length composition ISS was small and consistent across all the stocks evaluated.

While the magnitude of decrease in the proportion of length composition ISS was region and sex-specific, none of the stocks decreased to lower than 50% of the full dataset ISS; the majority of stocks did not decrease to lower than 60%. For all the stocks evaluated, the range in the annual proportion of full dataset ISS across the sex categories evaluated were either above or included 90% of the full dataset length composition ISS for the length frequency sub-sampling levels of 100 fish per haul; the only exception was for EBS walleye pollock female and male proportion of length composition ISS. Besides EBS walleye pollock, in which the range in the annual proportion of full dataset ISS included 90%, the remainder of the stocks evaluated resulted in proportions of full dataset ISS greater than 90% for length frequency sub-sampling levels of 150 fish per haul or more regardless of length composition sex category. Sampling at a level of 100 to 150 fish per haul for length frequency would result in 8,000 to greater than 30,000 fewer collections per-year within the surveys evaluated (Table 2).

Where the age composition ISS was unaffected by the length frequency sub-sampling level, the age composition ISS did markedly decrease as the proportion of total age collections decreased (Figure 5). In Figure 5 we use the stocks in the GOA bottom trawl survey as an example and note that this result was consistent across all regions, stocks, and sex categories evaluated. Similar to the results for length composition ISS, the sex-specific age composition ISS were smaller than the total (combined sex) iterated ISS across all stocks evaluated, and the magnitude of age composition ISS compared between sex categories (female and male) could differ by stocks. 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. The magnitude of sex-specific and total ISS were stock-specific, and ranged from 10 to 500; 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 larger range in age composition ISS across survey years than others.

The proportion of decrease from the full dataset age composition ISS was region and stock-specific, as was the variability in the decrease in age composition ISS across the survey years (Figure 6). Some of the largest decreases in age composition ISS across the total age sub-sampling levels resulted for arrowtooth flounder and Pacific cod in each survey evaluated, where the smallest decrease resulted for GOA walleye pollock. The largest variability across the survey years in the proportion of ISS across age sub-sampling levels again resulted for Pacific cod, but also included AI Atka mackerel and walleye pollock across each survey area, while the variablity was relatively small for the remaining stocks. Even though there were differences in the proportion of age composition ISS across the age sub-sampling levels evaluated, the results across stocks were generally consistent.

A notable result was that for all stocks evaluated, the proportional decrease in age composition ISS was dampened when compared to the proportional decrease in total age samples collected (Figure 7). For example, a reduction to 25% of the total age samples did not translate to a sub-sampled dataset age composition ISS reduction of 25% of the full dataset ISS, but rather resulted in a sub-sampled dataset age composition ISS of 30-40% of the full dataset ISS; this was true at each total age collection sub-sampling level evaluated. For rockfish and flatfish, the reduction in ISS compared to the reduction in total age samples were similar, both across stocks and sex categories. In comparison, for gadids and other (Atka mackerel), the reduction in age composition ISS compared to reduction in total age samples was further dampened, where the reduction in total age samples resulted in a proportionally smaller reduction in age composition ISS compared to rockfish and flatfish stocks. Regardless, for all stocks and sex categories evaluated, the reduction in ISS was less than the reduction in the total age samples collected on the surveys.

# 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 the proportional decrease in the ISS due to reductions in sampling. We found that the ISS for age composition data was more sensitive to reductions in sampling effort than the ISS for length composition. While the age composition ISS was sensitive to reductions in age sampling effort, we also show that the ISS for age composition was unaffected by the level of sub-sampling in the length frequency collections. Looking across multiple survey regions revealed that these reductions in ISS, or increase in uncertainty, of length and age composition were stock specific, but generally consistent. We find, based on the proportional decrease in ISS from sub-sampled data compared to full datasets, that there are recommendations that can be made to balance age and length collections on fishery-independent platforms with considerations of the amount of effort required to obtain these collections.

An unknown, but desirable outcome of research would be to specify how many length frequency collections are necessary to provide adequate information to stock assessment models. Gerritsen and McGrath (2007) suggested that sampling rates for length frequency be 10 fish per length category based on the mean weighted coefficient of variation statistic. In practical implementation, their suggestion would result in stock-specific sampling rates at the haul level, where larger fish would require more samples, and smaller fish would require less. While this is a result that makes biological sense, and is attractive as it is related to life-history characteristics of the stocks being sampled, it could result in logistical difficulties when implemented ‘on deck’ during a survey. We suggest, alternatively, that a more logistically feasible length frequency sampling method would be to set the same sampling rate across stocks at the haul level, recognizing that length frequencies for some stocks may be over-sampled. Here, we find that limiting sampling at 100 to 150 fish within a haul (whether for total or sex-specific length composition) would provide length composition data that have relatively similar uncertainty compared to length composition data derived from sampling at more intense rates (within 10% of the full dataset length composition ISS). We conclude, therefore, that there is little added benefit, in terms of reduced uncertainty, beyond sampling limits of 100 to 150 fish lengths per haul. This indicates as well that we are beyond a saturation point, or a point of diminishing returns in the improvements made by any increased sampling effort in these surveys for length composition, however, this is not the case for age composition sampling.

A related desirable outcome of research is to determine how many age collections are necessary to provide adequate information to stock assessments models. Based on the results of this analyses, we find the answer to this question to remain elusive. As we increased the subsampling rates for age composition there was no ‘leveling off’ of age composition ISS that would indicate a point of dimisnishing returns had been met like that which resulted for length composition ISS. In fact, as the subsampling rates were increased the age composition ISS continued to increase. In a similar study, Siskey et al. (2023) evaluated increasing age sampling beyond current sampling levels for a subset of stocks evaluated here using a bootstrap approach and also found that the age composition ISS continued to increase, even at levels of increasing current sampling levels by 67%. In order to determine whether and where age composition ISS reaches a point of diminishing returns simulation studies may be a more appropriate approach than bootstrapping historical data, as variables such as life-history or intra-haul correlation can be accounted for specifically (e.g., similar to Hulson et al. 2011, Xu et al. 2020). The results of our study, and ones like Siskey et al. (2023), suggest that the number of age samples beyond which there is no improvement to age composition information are likely beyond the current management agency capacities to sample and read by ageing laboratories. In terms of what is necessary to provide adequate information to stock assessment models, the historical sampling levels have resulted in model outcomes that seem to have provided reasonable advice for management of North Pacific fish stocks. Our recommendation, therefore, is to attempt to retain historical levels of sampling for age composition, with the recognition that there are consequences to age composition uncertainty with reductions in these rates that will have downstream effects on assessment model uncertainty.

Across the stocks evaluated, the decrease in length composition ISS as haul sub-sampling rates increased were extremely consistent, and no clear pattern emerged that seemed to be related to life-history, however, there were interesting species type patterns that emerged when considering the effect of reduction in age sampling rates and the subsequent age composition ISS. 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 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. Although, when comparing across species types we noted that the effect of reductions in age sampling efforts was relatively larger for flatfish and rockfish as compared to gadids (and Atka mackerel). That is, a reduction in age sampling effort for flatfish and rockfish has a relatively larger impact on the reduction in age composition ISS as compared to gadids. A similar result occurred in Hulson et al. (2017), where it was found that age sample size had a relatively larger effect on the uncertainty in age composition data for rockfish, then flatfish, then gadids (including Atka mackerel). In terms of assessment outcome, Siskey et al. (2023) showed that increasing or decreasing age sampling effort for the rockfish examples used had a larger subsequent effect on the uncertainty in management quantities than the gadid example, indicating that this relatively larger effect on age composition ISS translates through the assessment model as well. An extenuating circumstance that should be considered when evaluating the consequences of age sampling effort reduction for flatfish, is that all of the stocks included in this analysis are managed through the use of sex-specific assessment models. We show that, in comparison to the total (combined sex) age (and length) composition ISS the sex-specific age (and length) composition ISS is smaller, and between the two sexes there can be differences in the magnitude of ISS, often resulting in ISS that is larger for female age composition as compared to male. It must be noted that reductions in age sampling effort for flatfish will have further downstream effects when considering uncertainty in sex-specific age composition. While we encourage maintaining historical levels of age sampling effort, we additionally recommend that if reductions are to be considered in age sampling effort that decision-makers consider the unequal consequences of these reductions across species types.

Recently, a large amount of effort has been dedicated to evaluating the consequences of reductions in survey effort (ICES 2020, 2023). The results of this analysis fits with previous work… *Meaghan, Jason, and Matt - I think all of you have been involved with WKUSER, could you add some text here to compare what we’ve done with what was done in the WKUSER workshops, maybe focusing on whether any work was done focusing on comp data?*

While the focus of the analysis in this study was to evaluate the statistical consequences of effort reduction when collecting age and length samples, our stated goal was to speak to the balance between stock assessment input data uncertainty and workforce health and efficiency. We show that the length frequency sampling can be reduced from current sampling levels without major consequences to length composition uncertainty, this was true for both total (combined sex) and sex-specific data. While collecting total length frequency data is cheap, in terms of effort required to collect only the length from a fish, determining sex from a fish requires additional effort. *Jason - could you add a couple of sentences here with more detail from the ‘on-deck’ perspective?*

When collecting age samples the additional effort that is required beyond determining the sex of a fish involves removing the sagittal otoliths which is then sent for subsequent reading to an age reading laboratory. Thus, for age sampling there is additional effort beyond ‘on-deck’ sampling that is required by a managing agency. The cost for several stocks evaluated here is summarized in Lambert et al. (2017) and further evaluated in Siskey et al. (2023), which show that the monetary expense of each otolith is not inconsequential and the savings to an agency upon reducing age sampling could be in the tens of thousands of dollars (USD). However, Siskey et al. (2023) showed that there are downstream effects of increased uncertainty in assessment model estimates of the Over Fishing Level when age sampling is reduced. While buffers in catch limits due to assessment model uncertainty (e.g., Prager et al. 2003) are not yet incorporated in North Pacific fishery stock assessments, such buffers that would reduce catch limits based on increased uncertainty in assessment model quantities that are from increased uncertainty in age composition data due to reduced sampling would have direct effects on the value of the fisheries targeting these stocks. *Matt - anything to add from your paper?* *Jason - could you speak to the balance with workforce health here?*

Overall, we demonstrate a method from which a balance can be evaluated between stock assessment needs and workforce health and efficiency. Similar to what was done here, we recommend that fishery-independent survey groups collaborate with stock assessment scientists to determine if efficiency can be improved for age and length collections, here we found that improved efficiency could be obtained by reducing the amount of length frequency data (both total and sex-specific) on a haul-level basis. 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 intra-haul correlation levels.

# Acknowledgments

We thank *reviewer1* and *reviewer2* for their helpful reviews of this manuscript. We thank the remaining members of the AFSC Survey Workload Optimization Working Group for there 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.

# Citations

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 Journal of Marine Science 76(6): 1515–1523.

Council, N.R. 1998. Improving fish stock assessments. The National Academies Press, Washington, DC. doi: [10.17226/5951](https://doi.org/10.17226/5951).

Francis, R.I.C.C. 2017. Revisting data weighting in fisheries stock assessment models. Fisheries Research 192: 5–15.

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

Hulson, P.-J.F., Hanselman, D.H., and Quinn, I., 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 Journal of Marine Science 68(7): 1548–1557.

Hulson, P.-J.F., Hanselman, D.H., and Quinn II, T.J. 2012. Determining effective sample size in integrated age-structured assessment models. ICES Journal of Marine Science 69: 281–292.

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. Fishery Bulletin 115: 326–342.

Hulson, P.-J.F., Williams, B., Siskey, M., Bryan, M., and Conner, J. in review. 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.

ICES. 2020. Workshop on unavoidable survey effort reduction (WKUSER). ICES Scientific Reports.

ICES. 2023. Workshop on unavoidable survey effort reduction 2 (WKUSER). ICES Scientific Reports.

Lambert, G., Helser, T. E., Berger, A., Olsen, E., Hastie, J., O’Malley, J., Siegfried, K., Hulson, P.-J.F., McBride, R., Calay, S., Turner, S., and Miller, T. 2017. NOAA technical memorandum: Importance of age data collection for stock assessments: A US national perspective. Report from the otolith sampling size working group (OSSWG).

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: 260 p.

Maunder, M.N., and Piner, K.R. 2015. Contemporary fisheries stock assessment: Many issues still remain. ICES Journal of Marine Science 72(1): 7–18.

McAllister, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences 54(2): 284–300.

Miller, T.J., and Skalski, J.R. 2006. Integrating design- and model-based inference to estimate length and age composition in north pacific longline catches. Canadian Journal of Fisheries and Aquatic Sciences 63(5): 1092–1114.

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

Prager, M.H., Porch, C.E., Shertzer, K.W., and Caddy, J.F. 2003. Targets and limits for management of fisheries: A simple probability-based approach. North American Journal of Fisheries Management 23(2): 349–361.

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 Journal of Marine Science 70(1): 16–33.

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/>.

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. Canadian Journal of Fisheries and Aquatic Sciences 80(1): 115–131.

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: 205 p.

Stewart, I.J., and Hamel, O.S. 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. Canadian Journal of Fisheries and Aquatic Sciences 71(4): 581–588.

Thorson, J.T. 2014. Standardizing compositional data for stock assessment. ICES Journal of Marine Science 71(5): 1117–1128.

Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fisheries Research 192: 84–93.

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: 260 p.

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: 161 p.

Xu, H., Thorson, J.T., and Methot, R.D. 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. Canadian Journal of Fisheries and Aquatic Sciences 77(2): 247–263.

Zimmermann, M., Dew, C.B., and Malley, B.A. 2009. History of Alaska red king crab, Paralithodes camtschaticus, bottom trawl surveys, 1940-61. Marine Fisheries Review 71(1): 1.

# Tables

Table 1: Average length and age (shown in parentheses) samples 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 | Pleuronectes quadrituberculatus | – | 10,486 (327) | – |
| arrowtooth flounder | Atheresthes stomias | 9,868 (460) | 14,928 (702) | 36,842 (597) |
| Atka mackerel | Pleurogrammus monopterygius | 7,888 (610) | – | – |
| flathead sole | Hippoglossoides elassodon | 4,602 (0) | 22,356 (810) | 15,233 (225) |
| northern rock sole | Lepidopsetta polyxystra | 10,754 (0) | 24,698 (584) | 3,685 (197) |
| northern rockfish | Sebastes polyspinis | 14,928 (571) | – | 2,298 (447) |
| Pacific cod | Gadus macrocephalus | 5,723 (604) | 11,477 (1,449) | 3,452 (464) |
| Pacific ocean perch | Sebastes alutus | 3,2491 (1,019) | – | 23,319 (1,150) |
| rex sole | Glyptocephalus | – | – | 12,878 (412) |
| southern rock sole | Lepidopsetta billineta | – | – | 7,190 (730) |
| walleye pollock | Gadus chalcogrammus | 10,806 (791) | 49,544 (1600) | 16,772 (1,075) |
| yellowfin sole | Limanda aspera | – | 28,108 (845) | – |

Table 2: Average total number of reductions in length frequency samples when sampling 100 and 150 fish per haul by stock and region evaluated (rounded to the nearest 100s for stock-specific results, to the nearest 1,000s for total).

| Stock | AI | EBS | GOA |
| --- | --- | --- | --- |
| Alaska plaice | – | 200 - 700 | – |
| arrowtooth flounder | 400 - 1,200 | 600 - 2,000 | 2,900 - 8,300 |
| Atka mackerel | 400 - 1,700 | – | – |
| flathead sole | 200 - 600 | 1,000 - 3,500 | 500 - 2,200 |
| northern rock sole | – | 1,400 - 4,400 | 100 - 400 |
| northern rockfish | 1,600 - 4,800 | – | 100 - 500 |
| Pacific cod | 100 - 500 | 200 - 800 | 100 - 200 |
| Pacific ocean perch | 4,300 - 12,000 | – | 2,200 - 7,300 |
| rex sole | – | – | 100 - 800 |
| southern rock sole | – | – | 100 - 400 |
| walleye pollock | 100 - 2,500 | 5,300 - 14,500 | 2,600 - 4,700 |
| yellowfin sole | – | 2,000 - 6,300 | – |
| Total | 7,000 - 23,000 | 11,000 - 32,000 | 9,000 - 25,000 |

# Figures

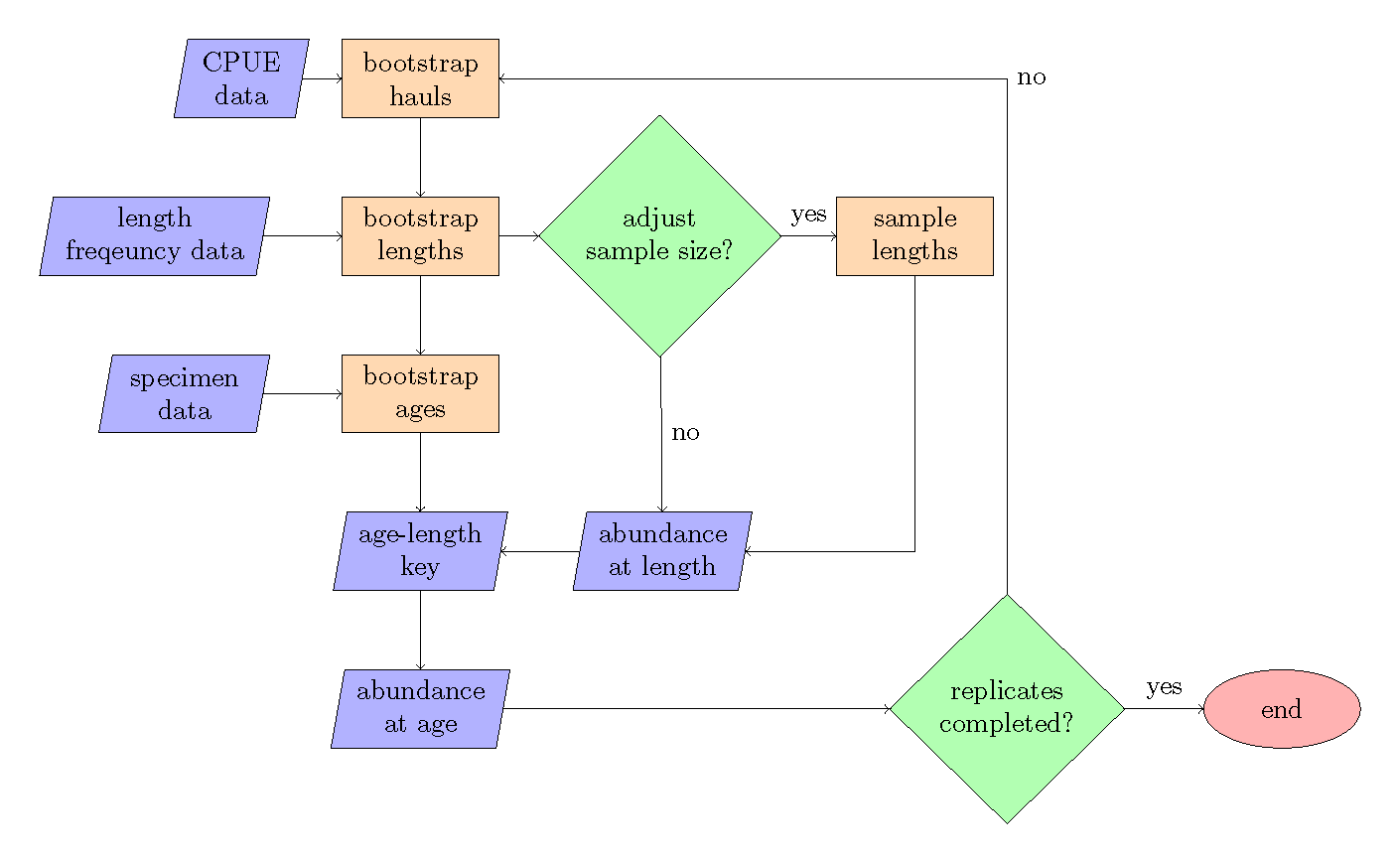


Figure 1: Bootstrap-simulation flow chart, the steps refer to the order of operations.

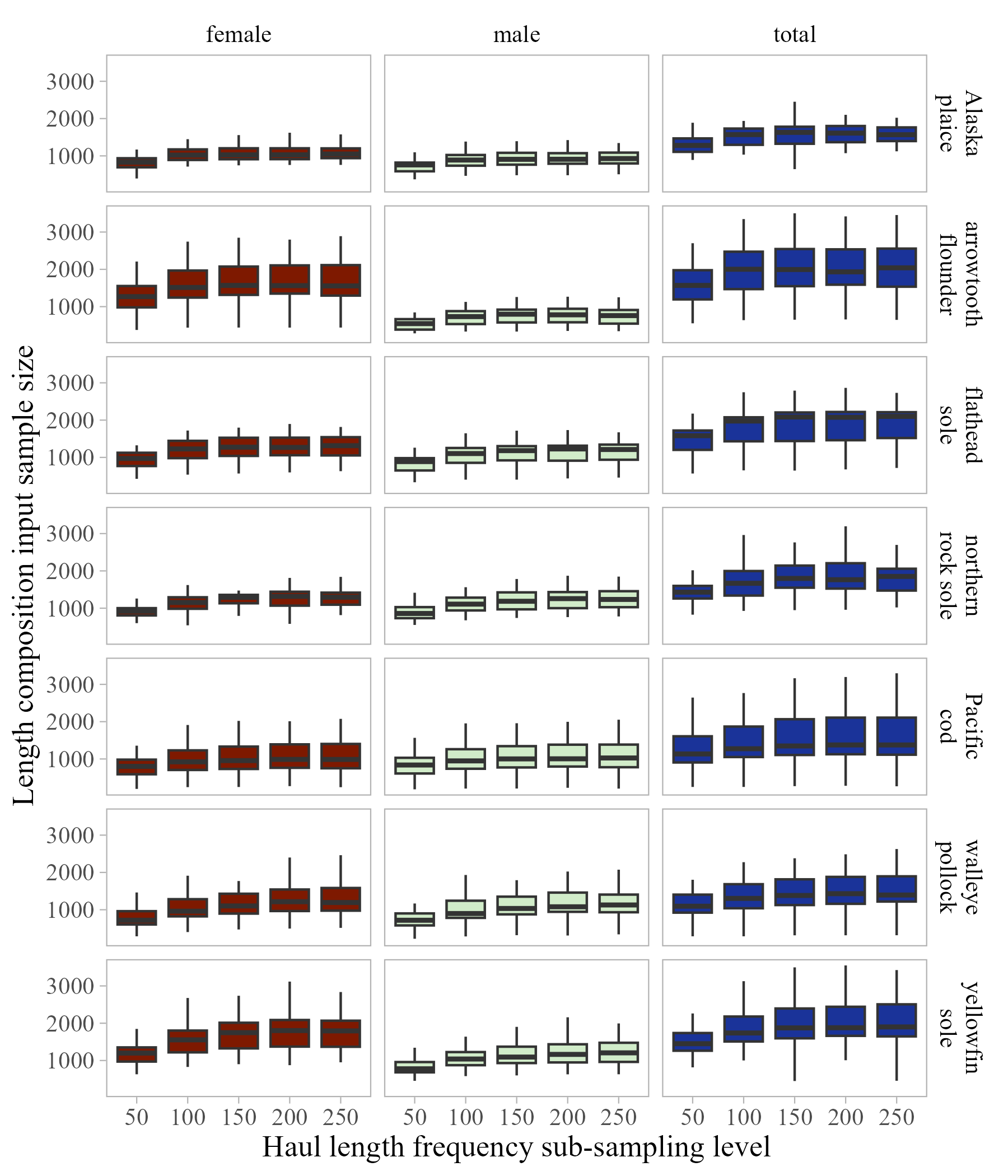


Figure 2: Boxplots of annual length composition input sample size size across length frequency sub-sampling levels evaluated for selected Eastern Bering Sea stocks.

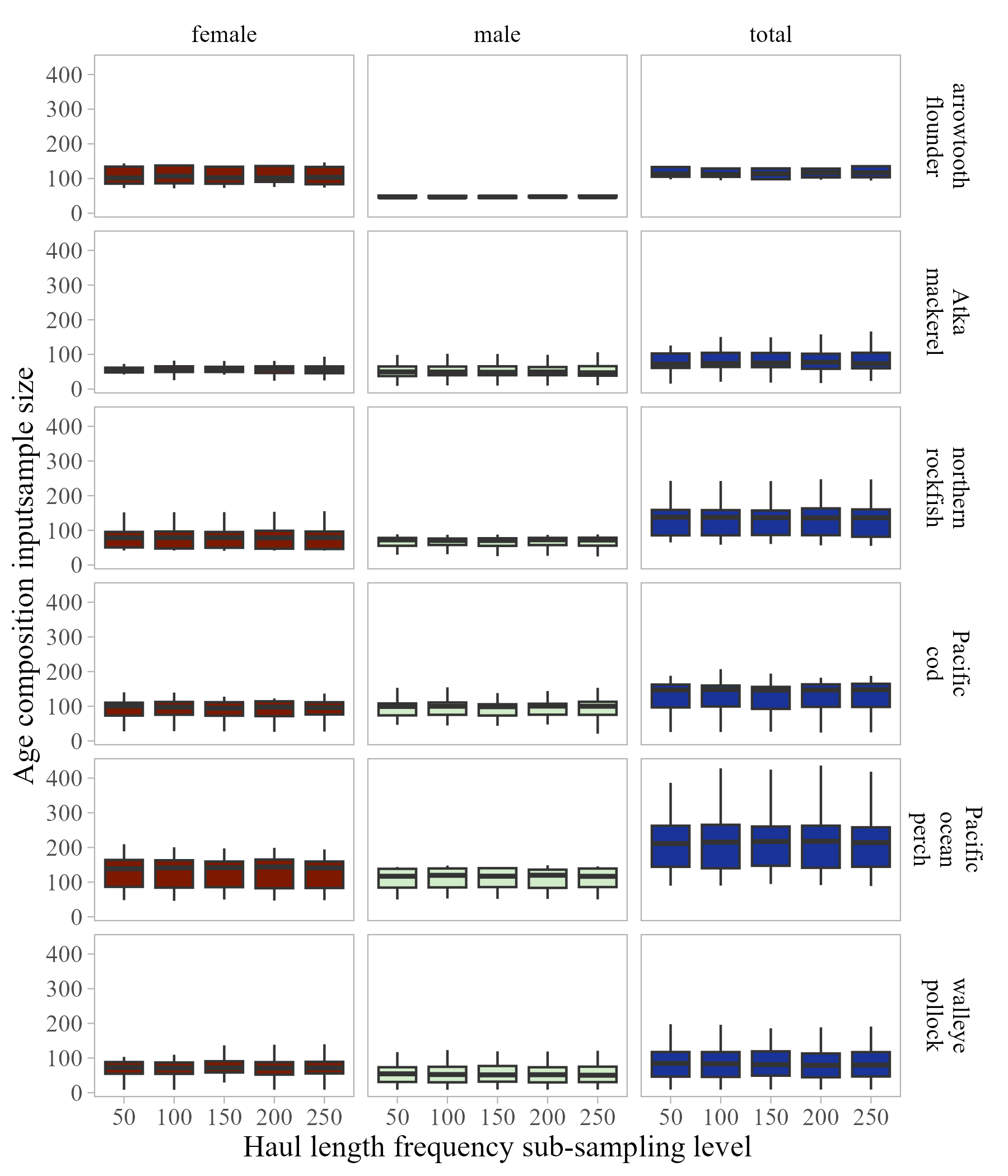


Figure 3: Boxplots of annual age composition input sample size across length frequency sub-sampling levels evaluated for selected Aleutian Islands stocks.

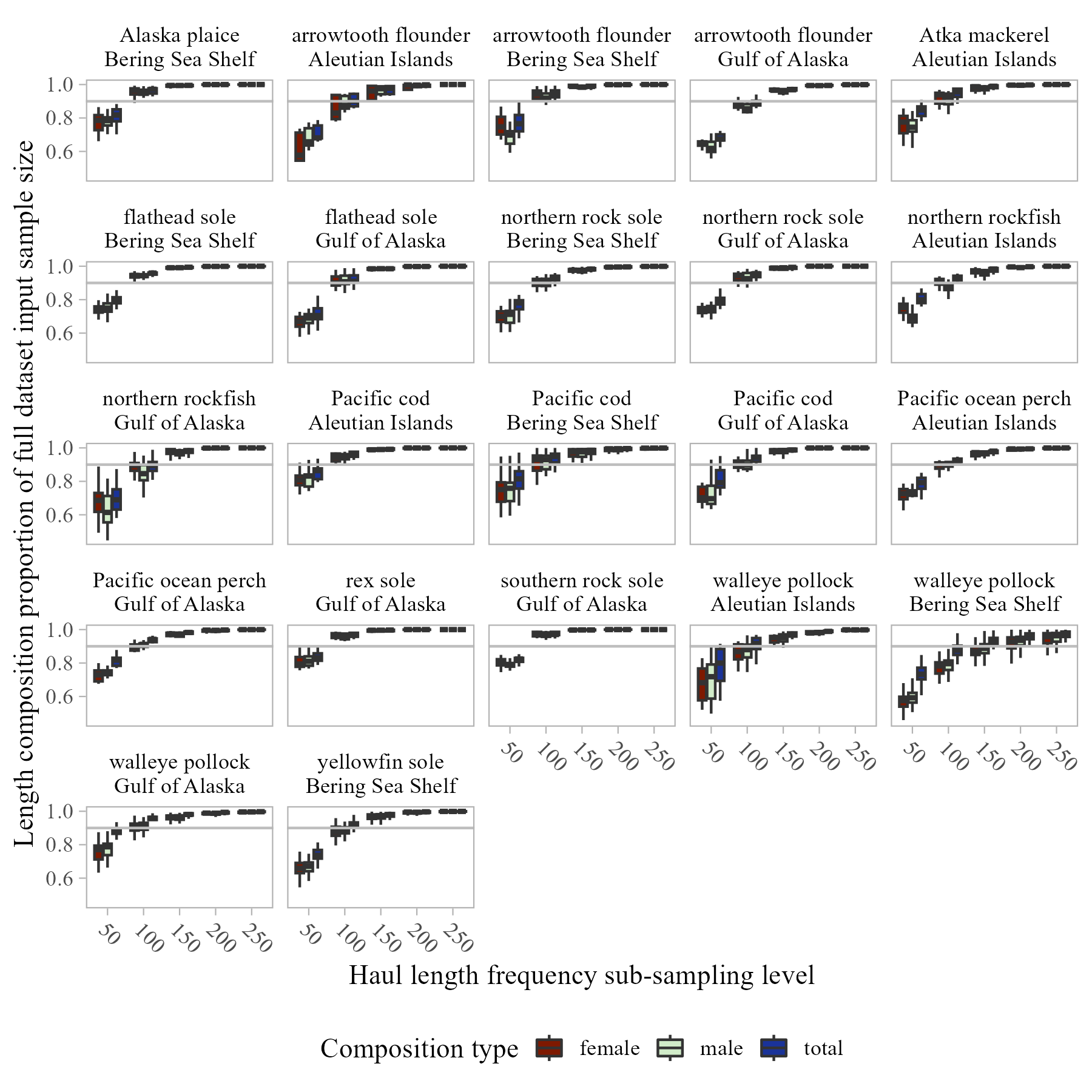


Figure 4: Proportion of full dataset length composition input sample size across length frequency sub-sampling levels evaluated (grey line at 0.9 shown for reference).

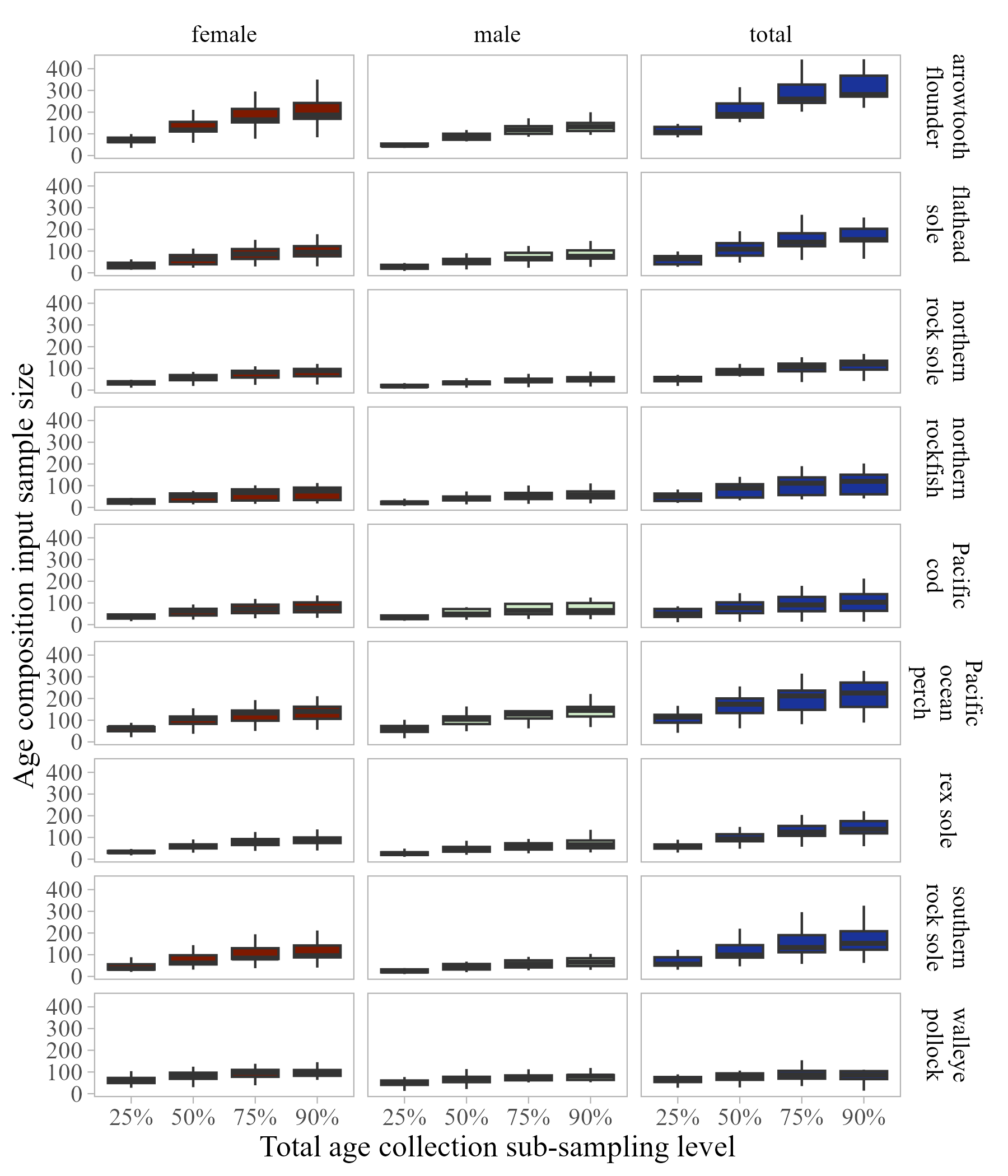


Figure 5: Boxplots of annual age composition input sample size across total age collection sub-sampling levels evaluated for selected Gulf of Alaska stocks.

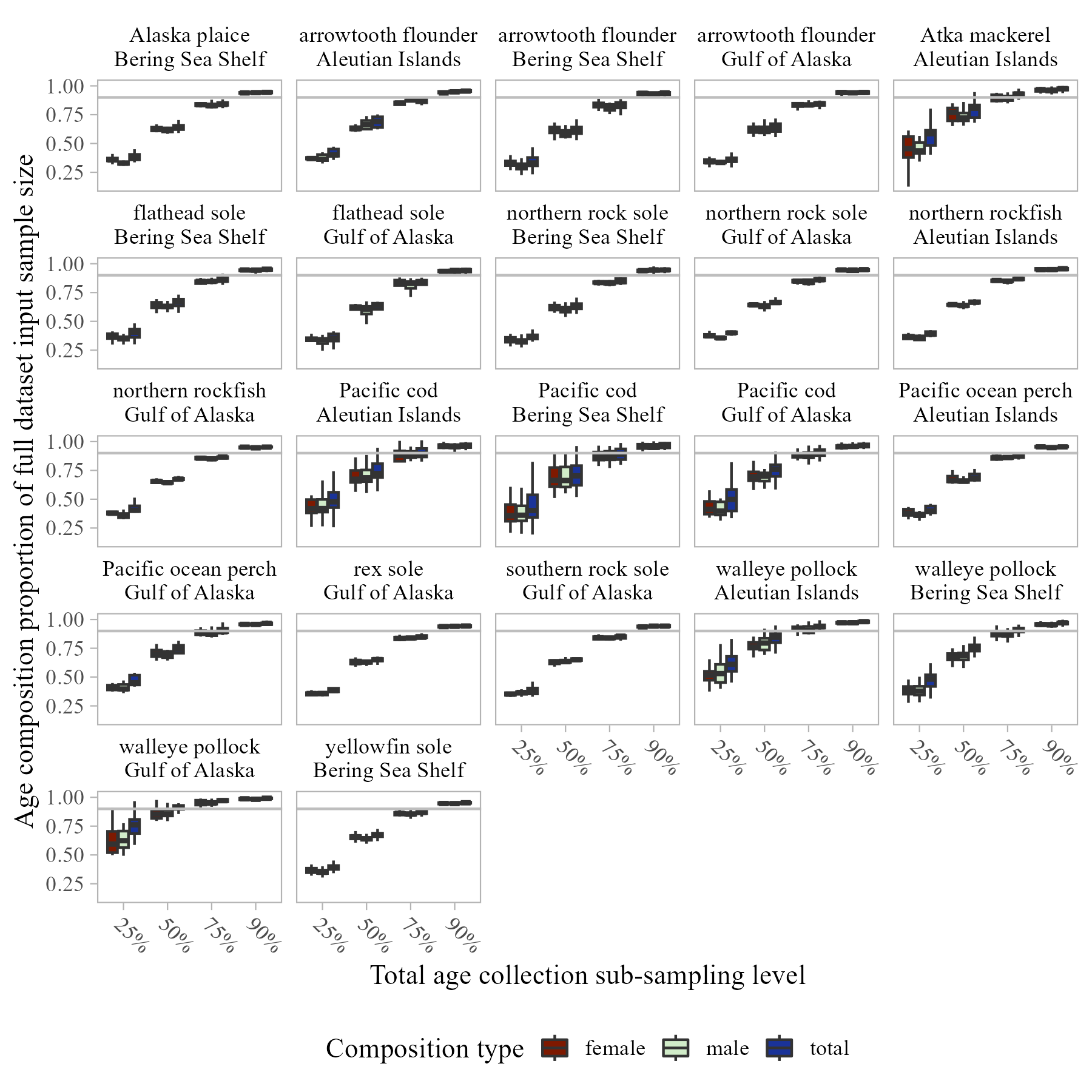


Figure 6: Proporion of reduction in input sample size compared to full dataset input sample size across total age collection sub-sampling levels evaluated (grey line at 0.9 shown for reference).

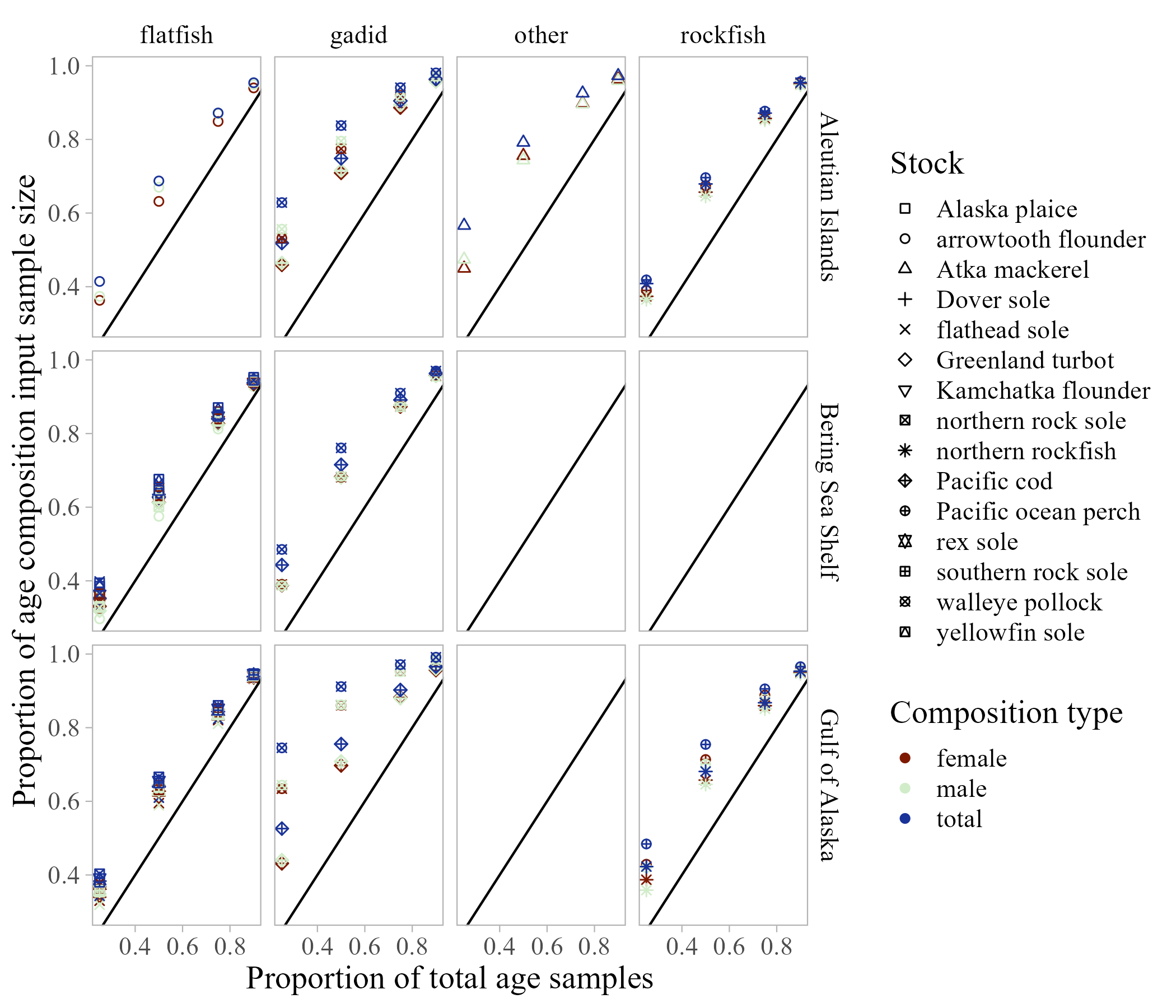


Figure 7: Average proporion of reduction in input sample size (compared to full dataset input sample size) across survey years compared to proportion of reduction in total age sample size (1-1 line shown for reference).