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

Peter-John F. Hulson1,\*, Benjamin C. Williams1, Meaghan D. Bryan2, Jason Conner3, and Matthew R. 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 Washington Department of Fish and Wildlife, 111 Washington St. SE, Olympia, WA, 98501

\* Correspondence: [Peter-John F. 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. 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 are used in stock assessment models. The main goal was to determine if 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 to 150 fish sampled per haul (either sex-specific or combined sex) provides 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 also found 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 consequences to model predictions. Conversely, there will be, at some level, consequences to model output with reductions in age collection sampling. The method employed here, and the results presented, can aid managing agencies to balance the magnitude of data collection on fishery-independent survey platforms and subsequent consequences to assessment models.

# 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 (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. Improvements in precision of age or length composition estimates are not only influenced by the magnitude of sampling, but also the manner in which the samples are collected [i.e., changing the number of samples collected within a haul versus changing the number of hauls from which samples were obtained; 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 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 intra-haul correlation. For example, Pennington et al. (2000) demonstrated that fish sampled from a haul 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 intra-haul correlation, and account for it in fishery stock assessment, the concept of ‘effective sample size’ has been developed. Effective sample size is smaller than the actual sample size reflecting the increase in uncertainty due to the intra-haul correlation displayed by the sampled species (e.g., McAllister and Ianelli 1997). A number of studies have used effective sample size to evaluate impacts on assessment results, including whether effective sample size 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). Further, methods have been developed that mimic the sampling designs for length and age composition to estimate effective sample size using bootstrap techniques external to an assessment model (Stewart and Hamel 2014). Overall, the use of effective sample size has become a recommended method to account for uncertainty caused by overdispersion of samples in length and age composition data.

Thousands of samples are obtained from a fishery-independent survey 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). AFSC scientists have routinely collected observations of fish length and age composition, consisting of complete or random subsamples of fish within each trawl haul throughout the time series.

There are a number of 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 is repeated several thousands of times. This represents a daunting amount of work for the field scientists on research vessels. Further, a large portion of the fish sampeld for length frequency are subsequently sampled for sex determinations, which involves a substantial increase in handling and data collection effort. 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 Programs. Another consequence resulting from the intensity of this work is the introduction of unrecoverable errors to 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). Aside from the budgetary constraints that are directly associated with operationalizing a fishery-independent survey, there are downstream financial costs associated with composition collections, in particular, requiring properly trained staff to accurately read otoliths collected across many species.

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 that 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 is more often 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, effective sample size 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 the impacts on a species specific 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 GAP 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 derived from AFSC bottom trawl surveys, we evaluated the impact of reduced sampling to answer four 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 species specific 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 GAP 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). The fundamental methods of length sample collection are generally synchronized between these surveys, with species-specific exceptions for minimum length sub-sample size. 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 sub-sample of a target sample size, 4) sorting the sub-sample 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 fish are processed at sea to collect their sex, length, weight, and sagittal otoliths. These otoliths are returned to the AFSC Age and Growth laboratory for age determination. Survey age sampling protocols are species specific and follow 1 of 2 paradigms: 1) a stratified 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 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 with similar bathymetric characteristics (e.g., depth ranges), and population abundance-at-length within 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). 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 the 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 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. 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 (Figure 1) has the following schedule:

1. Resample hauls (wr) with associated catch per unit effort (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.
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, ) 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 is less than the historical sample size *x* for that haul then a random draw of from the resampled lengths in step 2 is taken without replacement, if then the resampled lengths of size *x* in step 2 is used directly. We set the subsampling level for length frequency at numbers per haul in order to evaluate the AFSC length sampling design. Additionally, to subsample ages we changed the proportion of the total number of ages sampled in step 6 to evaluate the consequences of reductions in overall age sampling, because the number of otoliths collected per haul for age analysis is typically small. The bootstrap-simulation 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 population abundance-at-length and age calculated from the original data.

The length subsample levels that we evaluated from the full length frequency collections were 50, 100, 150, 200, and 250 samples per haul. The age subsample levels that we 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. We also ran the bootstrap-simulation for the full number of length frequency and age specimen collections without subsetting to compare the base level uncertainty to the uncertainty observed as a result of sub-sampling. We ran the bootstrap-simulation for 500 iterations, the level at which variability in results did not increase. 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 input sample size

Relative sample size, as introduced by McAllister and Ianelli (1997), is a statistic that can evaluate the level of intra-haul correlation in age and length composition samples. It is also a statistic that can evaluate the amount of uncertainty in estimated composition data compared to observed composition data. Relative 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) in interation-*i* and year-*y* and 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 in equation (1). For each iteration of the bootstrap simulation for a pre-specified sub-sampling level we computed an estimated proportion () 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 relative sample size that indicated the change in uncertainty caused by sub-sampling length frequency and age specimen data. A larger relative sample size indicates higher certainty in the iterated composition estimates, while lower relative sample size indicates less certainty.

The harmonic mean was used to summarize these relative sample sizes across iterations. 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 ‘input sample size’ used as an initial weighting in stock assessment models to fit compositional data (Stewart and Hamel 2014). Herein, we use the term ‘iterated relative sample size’ (iterated *R*) to refer to the relative sample sizes computed for each iteration of the bootstrap-simulation from (1), and we use the term ‘input sample size’ (ISS) to refer to the harmonic mean of the iterated relative 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 sub-sampling level divided by the full dataset ISS (no sub-sampling) for each survey year. This provides an indication of the amount of change in uncertainty due to reductions in sub-sampling rates. To provide general results for illustration of trends we display the ISS and relative ISS by species type (flatfish, gadid, and rockfish) and at times survey region (AI, EBS, and GOA) using box-plots (which include the median, inter-quartile 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 when survey region results are shown they are grouped across the stocks in that region. 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.

# Results

As the number of length frequency sub-samples taken within hauls decreased, the ISS decreased, but not dramatically (e.g., moving right to left on the *x*-axis of the top panels in Figure 2). This trend of decreasing length composition ISS associated with decreasing sub-sampling levels per haul for length frequency was observed across all regions, stocks, and sex categories evaluated (Supplementary Material Figures S1–S3). The sex-specific length composition ISS were smaller than the total (combined sex) ISS for all stocks evaluated. This remained true for each individual sub-sampling case as well. However, stock-specific and species type differences in the magnitude of sex-specific ISS were observed, such as with flatfish, including arrowtooth flounder and yellowfin sole. If the magnitudes were different, female length composition ISS was consistently larger than male length composition ISS; this same result was not replicated for the gadid and rockfish stocks, where the sex-specific ISS were relatively similar. 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 4,000 for flatfish, 100 to 3,500 for gadids, and 100 to 1,000 for rockfish (including the full dataset case and all sub-sampling levels). The variability in length composition ISS across survey years and haul sub-sampling levels was also species type and stock-specific, but the variability generally declined 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 (bottom panel of Figure 2).

As the haul level length frequency sub-sampling level decreased the relative ISS also decreased (Figure 3). This decrease in relative ISS was exhibited by all stocks evaluated, although the magnitude of decrease was region and stock-specific (Supplementary Material Figures S4–S6). In terms of stock-specific results, the largest decrease in the relative ISS across the length frequency sub-sampling levels evaluated resulted for EBS walleye pollock (Figure S5) and the smallest decrease resulted for GOA rex sole and southern rock sole (Figure S6). The largest variability across survey years in the relative ISS was observed for sub-sampling levels of 50 and 100 lengths per haul for AI walleye pollock (Figure S4) and GOA northern rockfish (Figure S6). However, in terms of species type results across the survey regions (Figure 3), the decrease in relative ISS as the haul length frequency sub-sampling level decreased was very similar, while the variability could be slightly different in some cases (e.g., variability at the haul level sub-sampling rate of 50 as compared across species types and survey regions). For sub-sampling levels of 150 and greater the variability in length composition relative ISS was small and consistent across all the species types and survey regions evaluated.

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 the majority of stocks did not decrease below 60% (Figure 3). The range in the length composition annual relative ISS across the sex categories evaluated either included or was above 90% for length frequency sub-sampling levels greater than 100 fish per haul. EBS walleye pollock was the one exception (Figure S5), 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 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 7,000 - 32,000 fewer collections per-year within the surveys evaluated (Table 2).

While age composition ISS was unaffected by the length frequency sub-sampling level, the age composition ISS did markedly decrease as the proportion of total specimen age data decreased (e.g., moving right to left on the *x*-axis of Figure 4). We note that this result was consistent across all regions, stocks, and sex categories evaluated (Supplementary Material Figures S7–S9). 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, which resulted for the flatfish stocks evaluated. For the gadid and rockfish stocks evaluated the magnitude of sex-specific age composition ISS was generally similar. The magnitude 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 range in ISS was generally of a similar magnitude across the species types (Figure 4), but could vary depending on the survey region. For example, the AI survey had smaller age composition ISS for flatfish stocks than the EBS and GOA, although, there are fewer stocks in the AI than the other survey regions. Additionally, the age composition ISS for gadid stocks were of similar magnitude an range in the AI and GOA, but were larger and had a greater range in the EBS. 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.

As the specimen age data sampling level decreased the age composition relative ISS also decreased (Figure 5). 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 region evaluated, while the smallest decrease resulted for GOA walleye pollock (see Supplementary Material Figures S10–S12). The largest variability across the survey years in the relative ISS across age sub-sampling levels again resulted for Pacific cod, but also included AI Atka mackerel and walleye pollock across each survey area, with relatively small variability observed for the remaining stocks. Even though there were differences in the age composition relative ISS across the age sub-sampling levels evaluated, the trends across species types and survey regions were generally consistent, where gadids displayed the largest range in the decrease in relative ISS within each survey region (Figure 5). While the decrease in both length and age composition relative ISS was consistent as the sub-sampling level decreased, while 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 sub-sampling level increased.

A notable result was that for all stocks evaluated, the decrease in age composition relative ISS was dampened when compared to the specimen age sub-sampling level collected (Figure 6). 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 resulted in a sub-sampled dataset age composition relative ISS of 30-40%; this was true at each age collection sub-sampling level evaluated. For rockfish and flatfish, the reduction in ISS compared to the reduction in specimen age samples were similar, both across stocks, survey regions, and sex categories. In comparison, for gadids and other (Atka mackerel), the reduction in age composition relative ISS compared to reduction in specimen 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 (this can be seen by comparing the slopes in the linear model fits in Figure 6). 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.

# 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 in the ISS from experimental treatments compared to the full dataset base case 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. We also show that while the age composition ISS was sensitive to reductions in age sample effort, 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, which represent increases in uncertainty, were stock specific but the reductions were generally consistent. Based on the proportional decrease in ISS from sub-sampled data compared to the ISS generated from full datasets, we found that there are potential 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 amount of 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 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 strategy 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 during a survey.

Based on the results of this study we suggest 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. For example, we found that limiting sampling at 100 to 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, beyond sampling limits of 100 to 150 fish lengths per haul. This also indicates 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. However, this is not the case for age composition sampling.

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 the answer to this question to remain elusive. As we increased the subsampling rates for age composition the age composition ISS continued to increase indicating that a sample size that produces 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 also 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 found by increasing sampling in more hauls even while decreasing the number of specimen age data collected in any given haul.

To determine if and when age composition ISS reaches an asymptote, simulation studies might be a more appropriate rather 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 necessary to maximize to age composition information are likely beyond the current management agency capacities. 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, as reductions in these sampling rates have downstream effects on assessment model uncertainty. We also recommend (and iterate the results of Siskey et al. (2023)) that if sampling for ages were to be optimized for effort required by a survey team ISS can be increased by spreading these samples across more hauls while retaining the historical total number of age samples collected.

Across the stocks evaluated, the increase in length composition ISS as haul sub-sampling rates increased were extremely consistent, with no clear species-specific patterns. There were interesting species group (e.g., flatfish, gadids, and rockfish) patterns when examining reductions in age sampling rates, however. 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. In addition, when comparing across species groups 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 was observed in Hulson et al. (2017), where they 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 rockfish 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 using sex-specific assessment models. In comparison to the total (combined sex) age and length composition ISS, we show that the sex-specific composition ISS is smaller and differences in the magnitude of ISS exist between the two sexes. This often resulted in an ISS that is larger for female age composition than for males. It must be noted that reductions in age sampling effort for flatfish will affect uncertainty in sex-specific age composition. While we encourage maintaining historical levels of age sampling effort, we additionally recommend that if reductions must be made that decision-makers 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 we created for this study.

Recently, a large amount of effort has been dedicated to evaluating the consequences of reductions in survey effort and the best methods to use to conduct these analyses (ICES 2020, 2023). The Workshop on Unavoidable Survey Effort Reduction (WKUSER) identified empirical resampling analyses to obtain estimates of uncertainty in abundance indices and composition data as critical when sudden changes in sampling are needed. Another recommendation from WKUSER was to take a holistic view of multispecies surveys and identify trade-offs among species rather than consider the impacts on individual species in isolation. The results of this analysis here addresses both recommendations and equips survey managers with the information to retain the data quality for a majority of species.

While the focus of this study was to evaluate the statistical consequences of effort reduction when collecting age and length samples on bottom trawl surveys, our stated goal was to address the tradeoffs between stock assessment input data uncertainty and workforce health and efficiency. 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. While collecting total length frequency data is cheap in terms of effort required, determining sex from a fish requires additional effort. Collecting age samples requires additional effort beyond determining the sex of a fish, including extraction of the sagittal otoliths which are then prepared and read in an ageing laboratory. The cost of age reading for several stocks evaluated here is summarized in Lambert et al. (2017) and further evaluated in Siskey et al. (2023). In short, the monetary expense of each otolith is not inconsequential and the savings to an agency as a result of reduced 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 limit when age sampling is reduced. Buffers in catch limits based on estimates of uncertainty in derived assessment quantities (e.g., Prager et al. 2003) are not used in the North Pacific. In areas where these types of buffers are used, a reduction in age composition sampling would lead to greater assessment uncertainty and directly impact the value of fisheries through a reduction in catch. However, from a survey effort perspective when collecting otoliths, it may be more efficient to collect less otoliths from any given haul while at the same time increase the number of hauls from which otoliths are collected.

Overall, we demonstrate a method for evaluating tradeoffs 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 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 intra-haul correlation levels.

# Acknowledgments

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 *reviewers*.

# 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 samples (length samples shown first, age samples 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. Species types are shown in parentheses in the stock column (f - flatfish, g - gadid, r - rockfish, o - other)

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

Table 2: Average total number of reductions in length frequency samples when sampling 100 (left side) and 150 (right side) 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

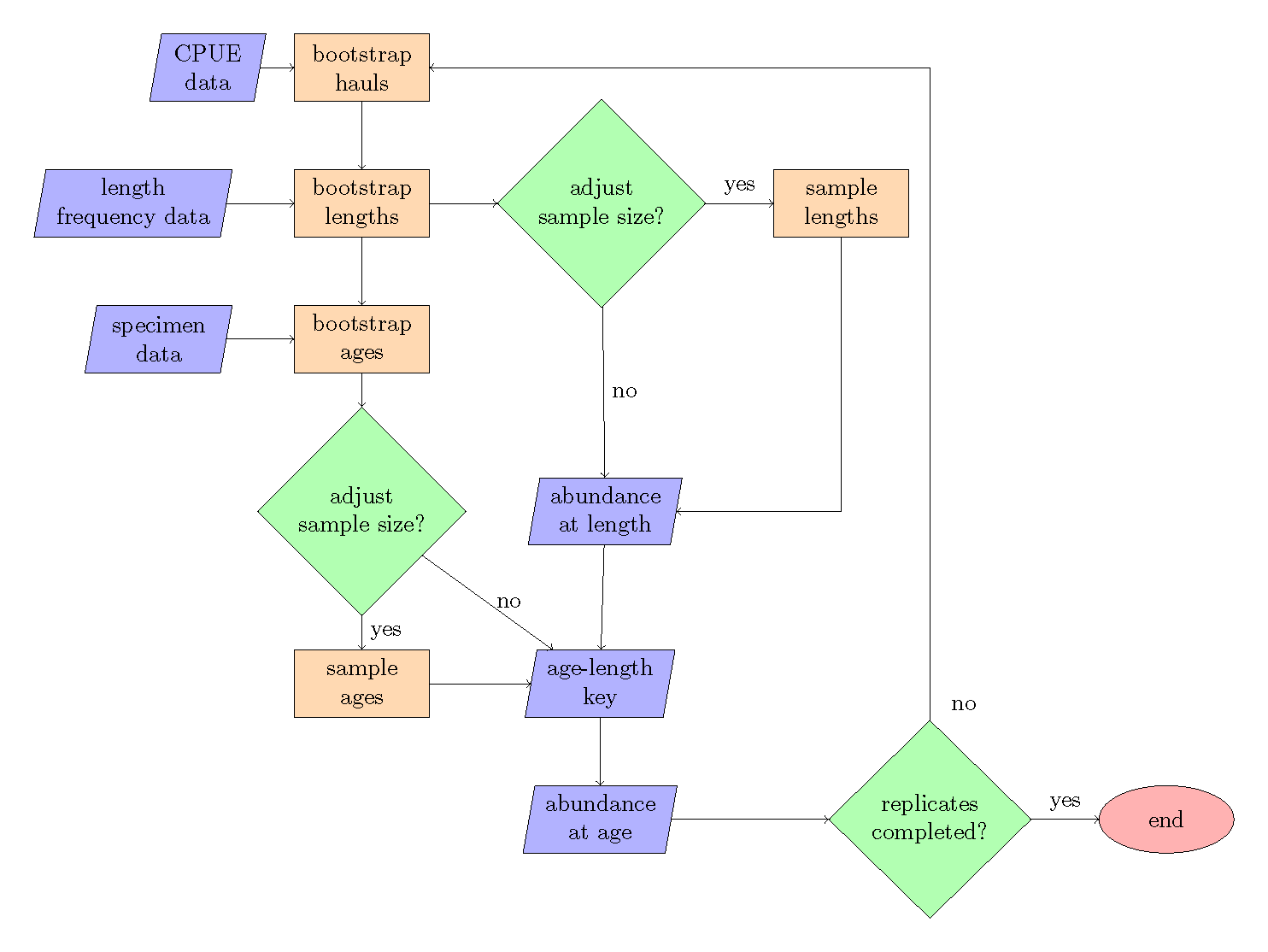


Figure 1: Bootstrap-simulation flow chart, with steps referring to the order of operations.

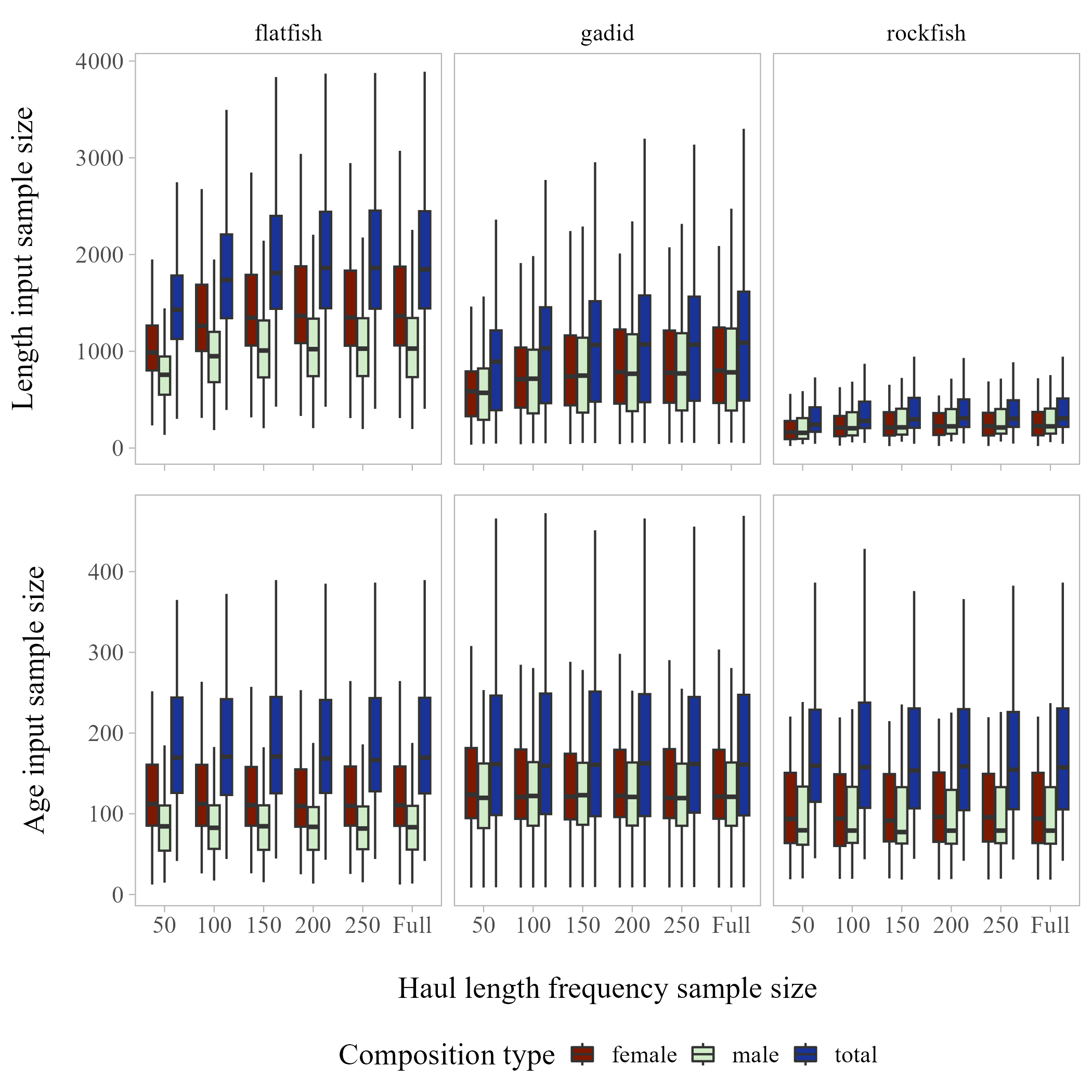


Figure 2: Boxplots aggregated by species type and regions of annual length composition (top panel) and age composition (bottom panel) input sample size across haul length frequency sub-sampling levels evaluated.

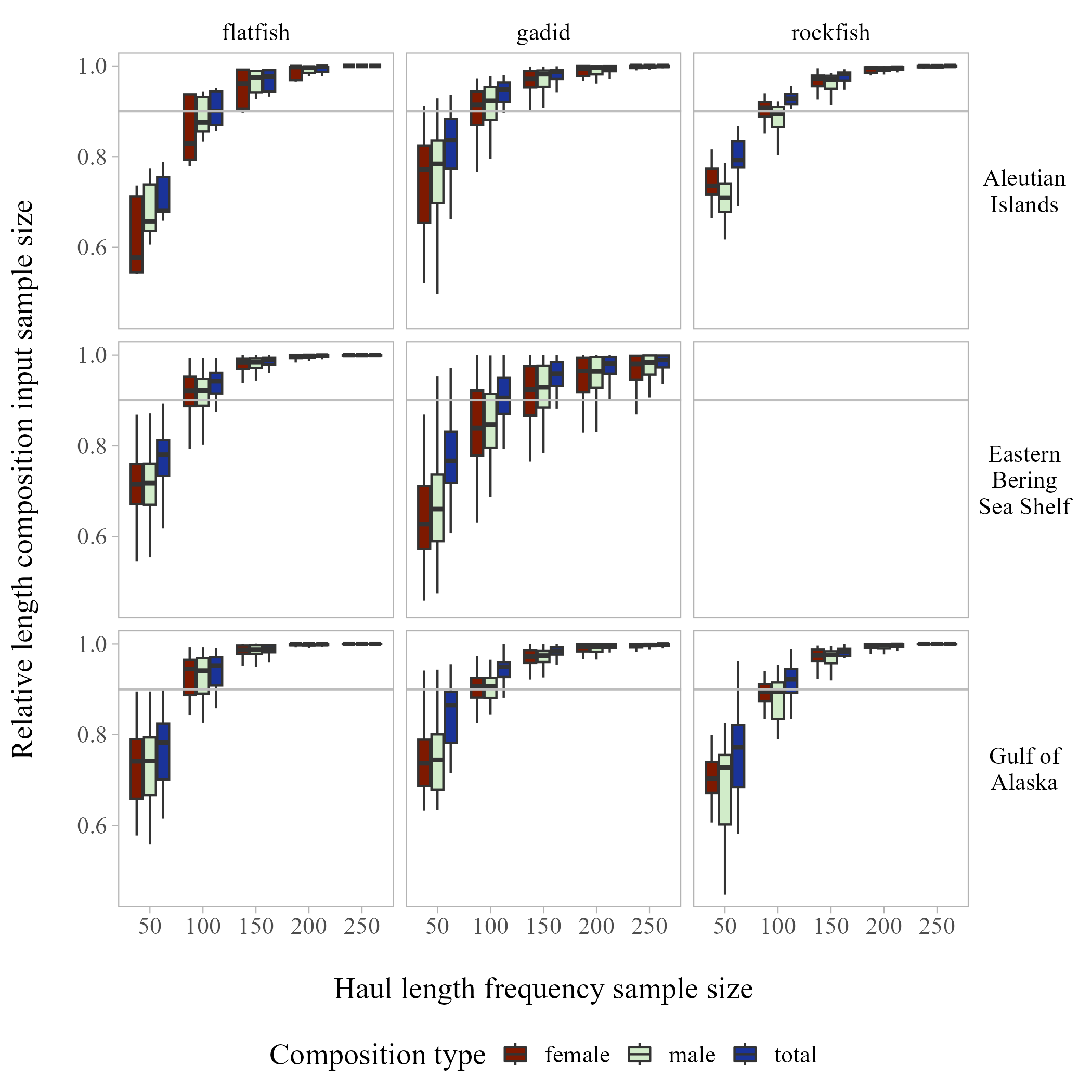


Figure 3: Annual relative length composition input sample size aggregated by species type for each region across haul length frequency sub-sampling levels evaluated (grey line at 0.9 shown for reference).

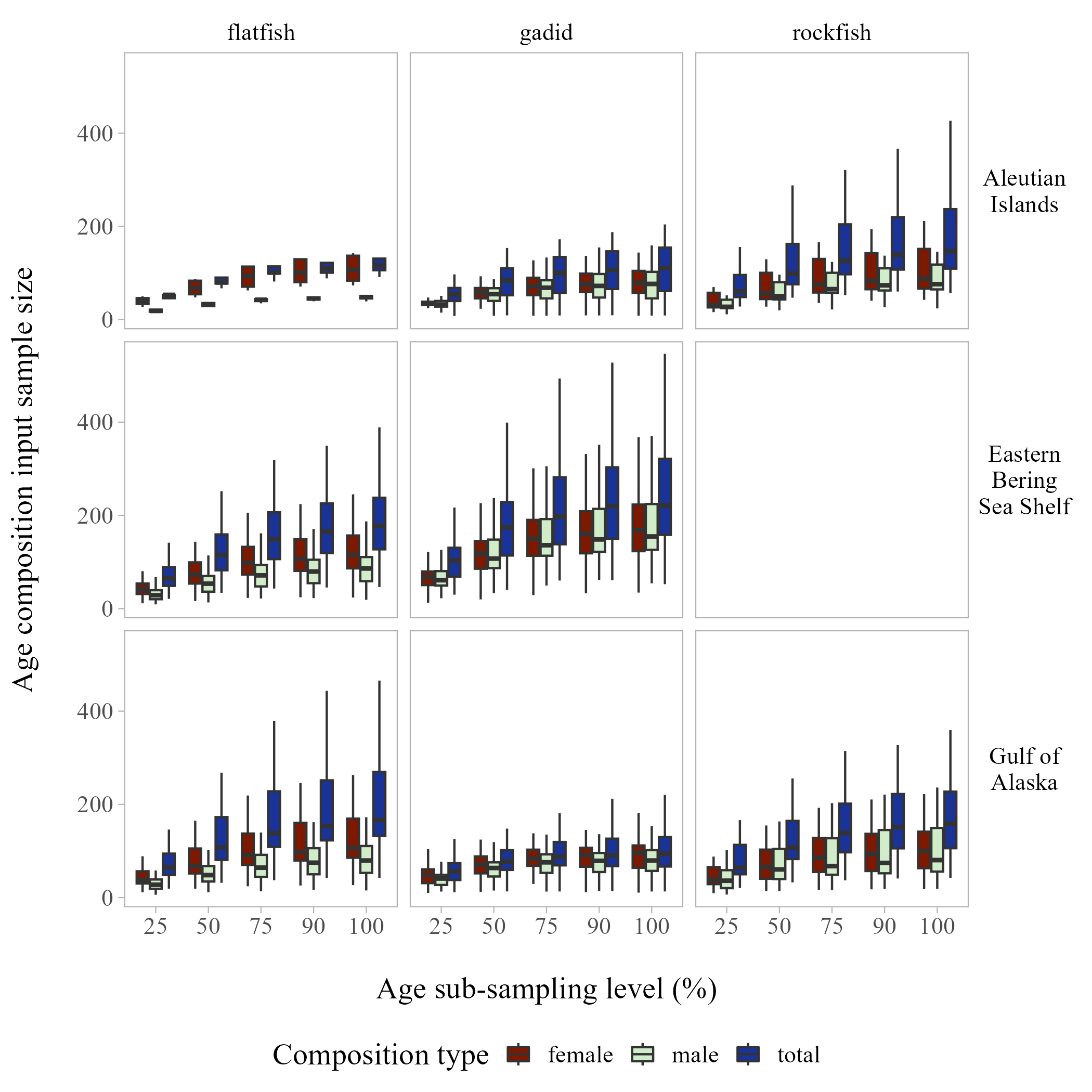


Figure 4: Boxplots aggregated by species type of annual age composition input sample size by region across age collection sub-sampling levels evaluated.

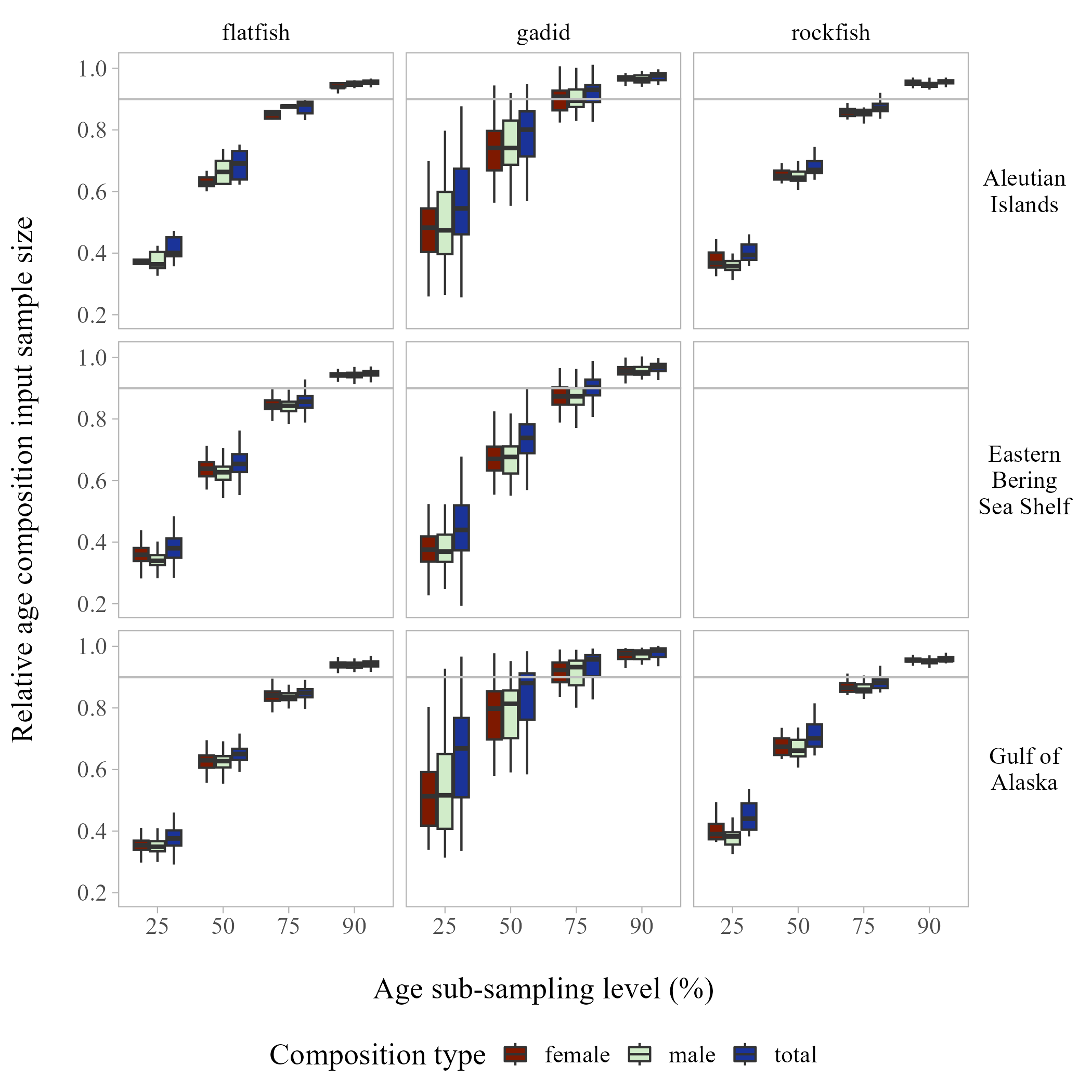


Figure 5: Annual relative age composition input sample size aggregated by species type for each region across age collection sub-sampling levels evaluated (grey line at 0.9 shown for reference).

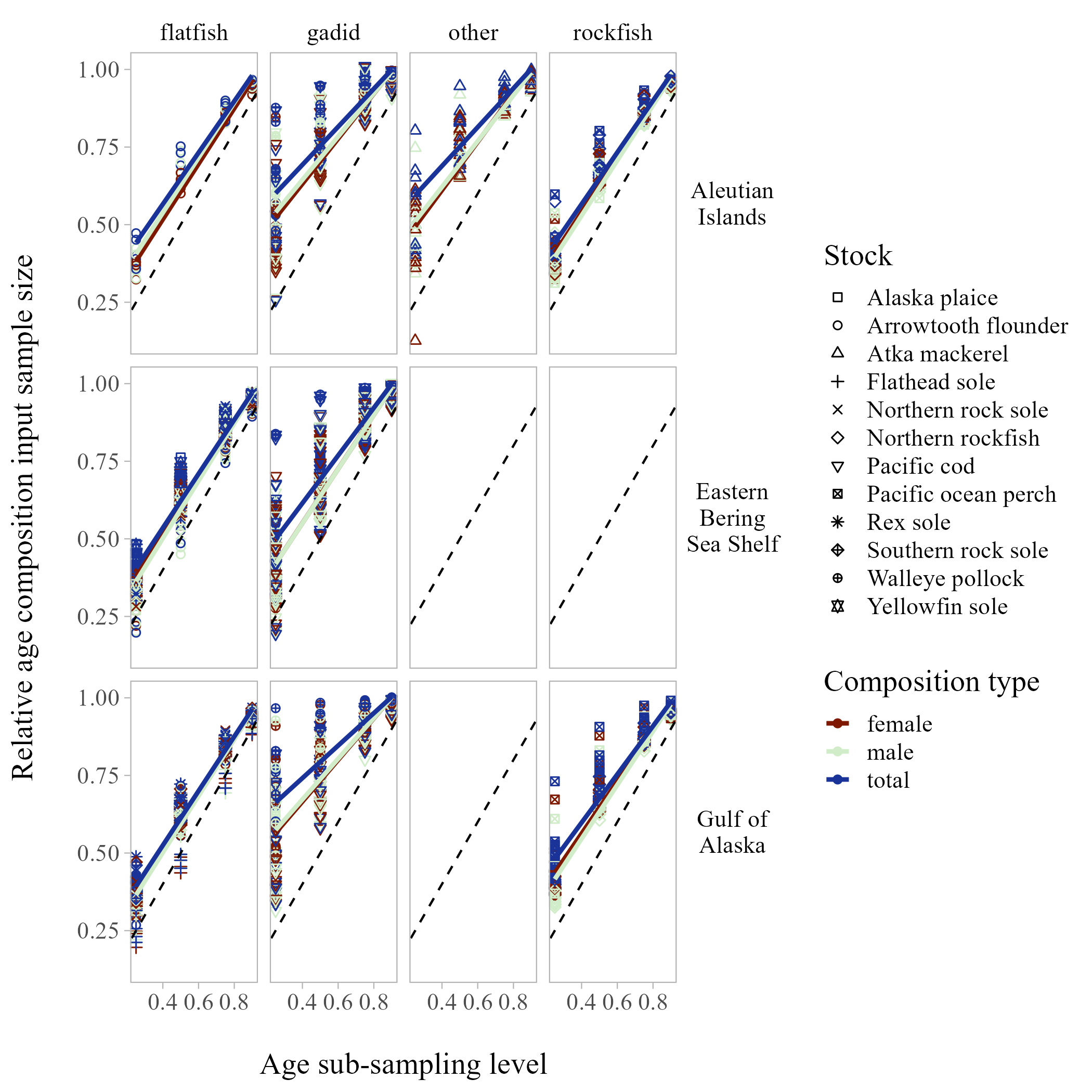


Figure 6: Stock-specific annual relative age composition input sample size grouped by species type and region compared to age collection sub-sampling levels (1-1 line shown for reference, linear model fit by composition type for each species type also shown for reference).

# Supplementary material

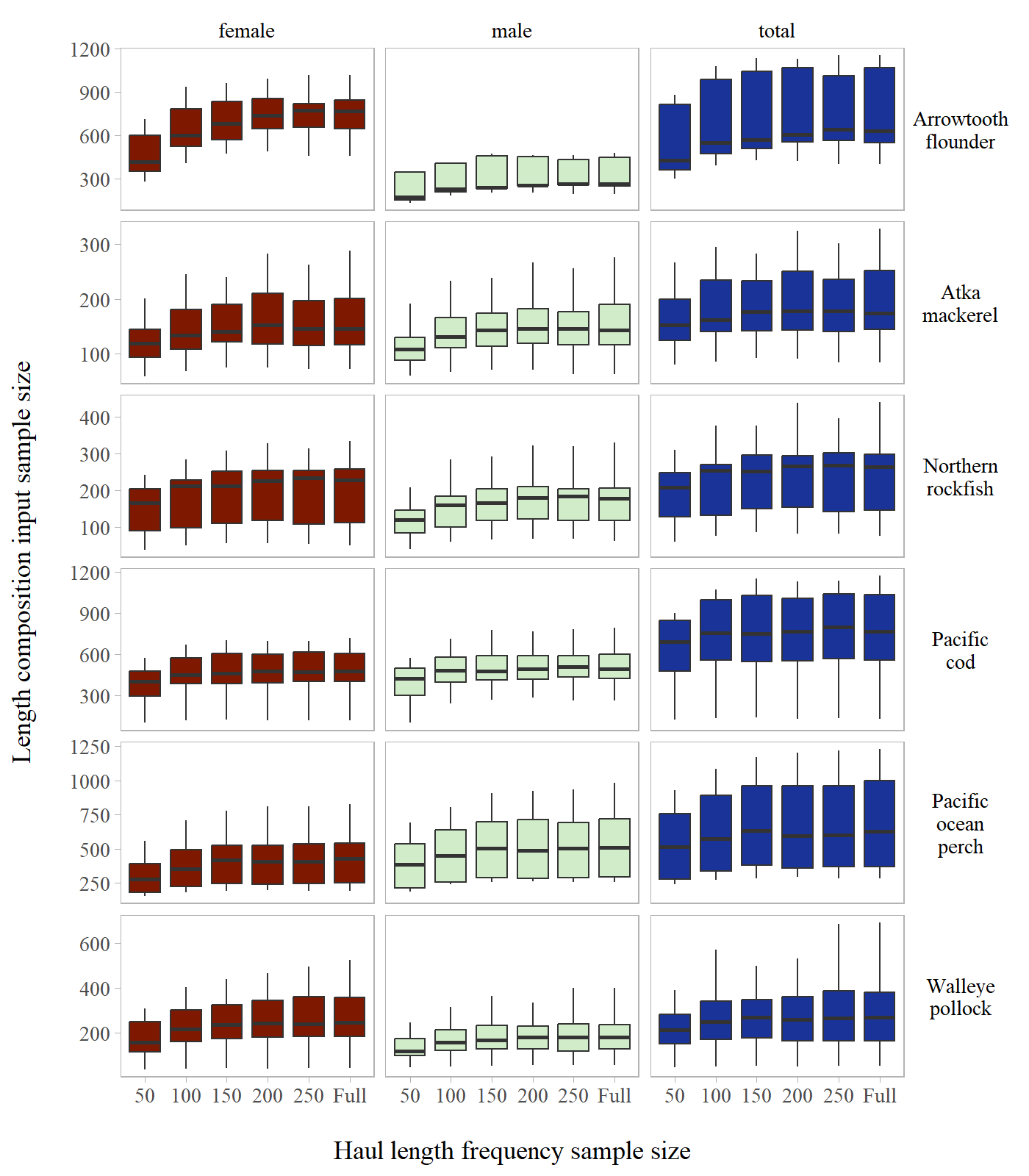


Figure S1: Annual length composition input sample size by stock across haul length frequency sub-sampling levels evaluated for the Aleutian Islands.

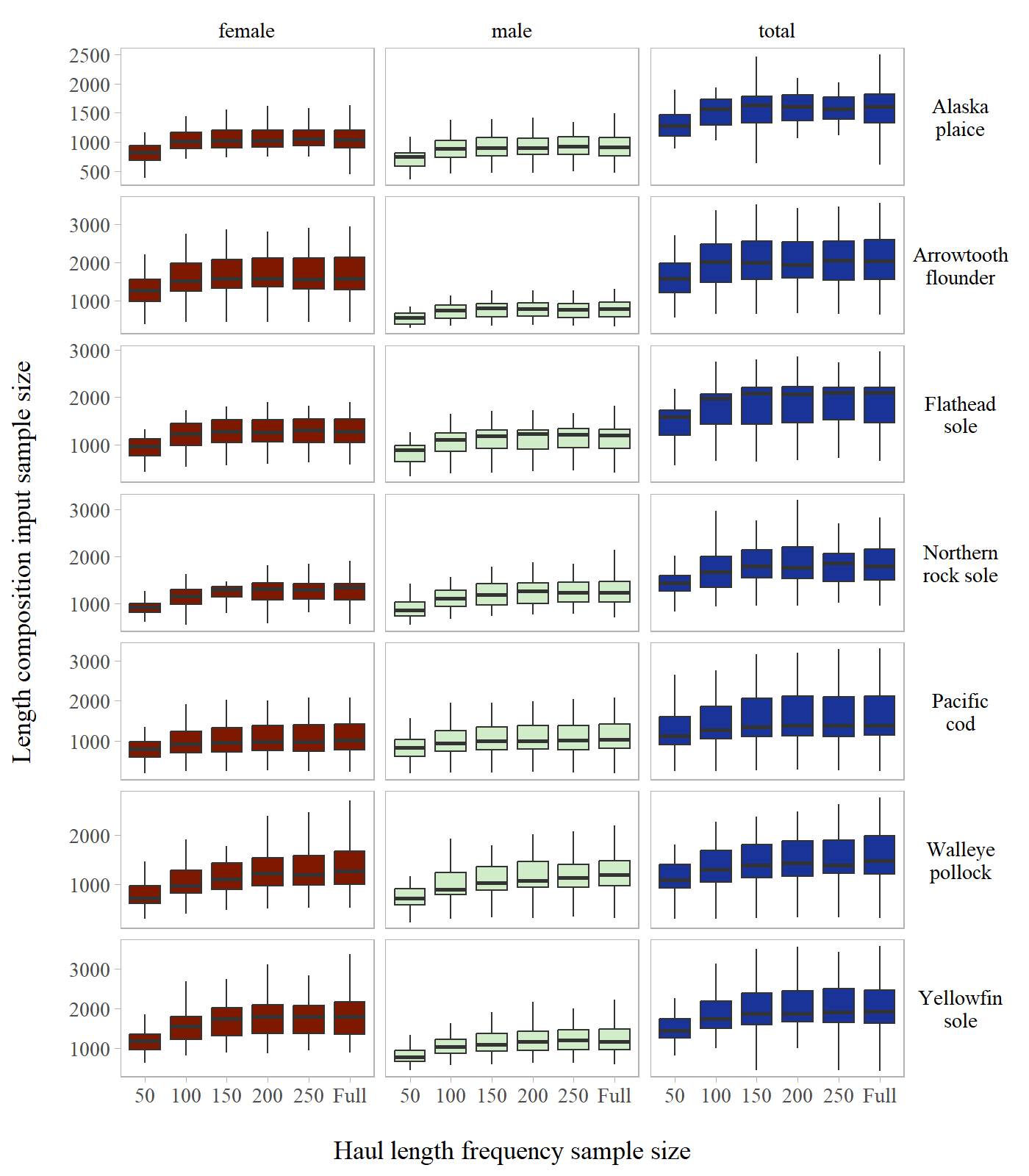


Figure S2: Annual length composition input sample size by stock across haul length frequency sub-sampling levels evaluated for the Eastern Bering Sea Shelf.

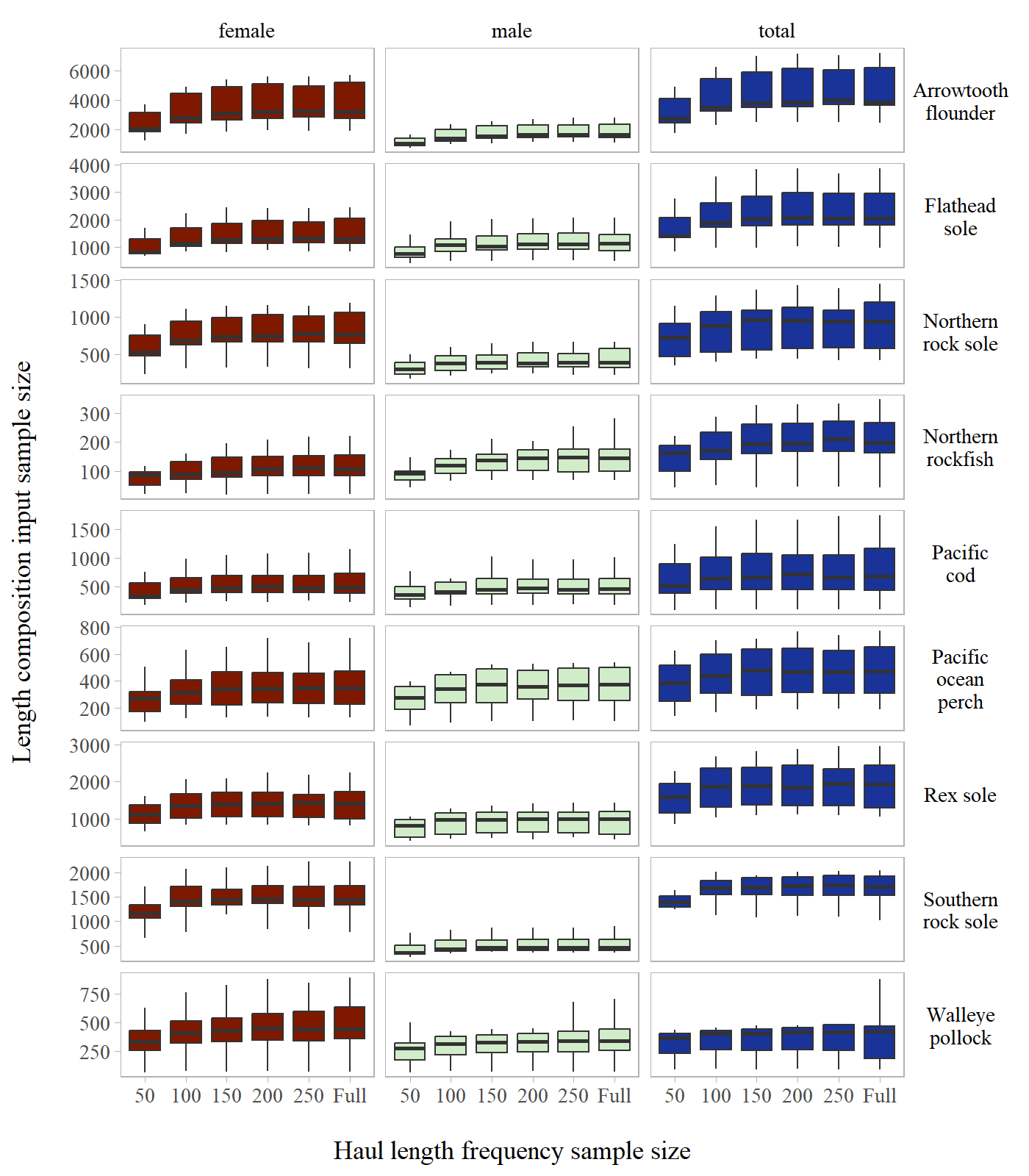


Figure S3: Annual length composition input sample size by stock across haul length frequency sub-sampling levels evaluated for the Gulf of Alaska.

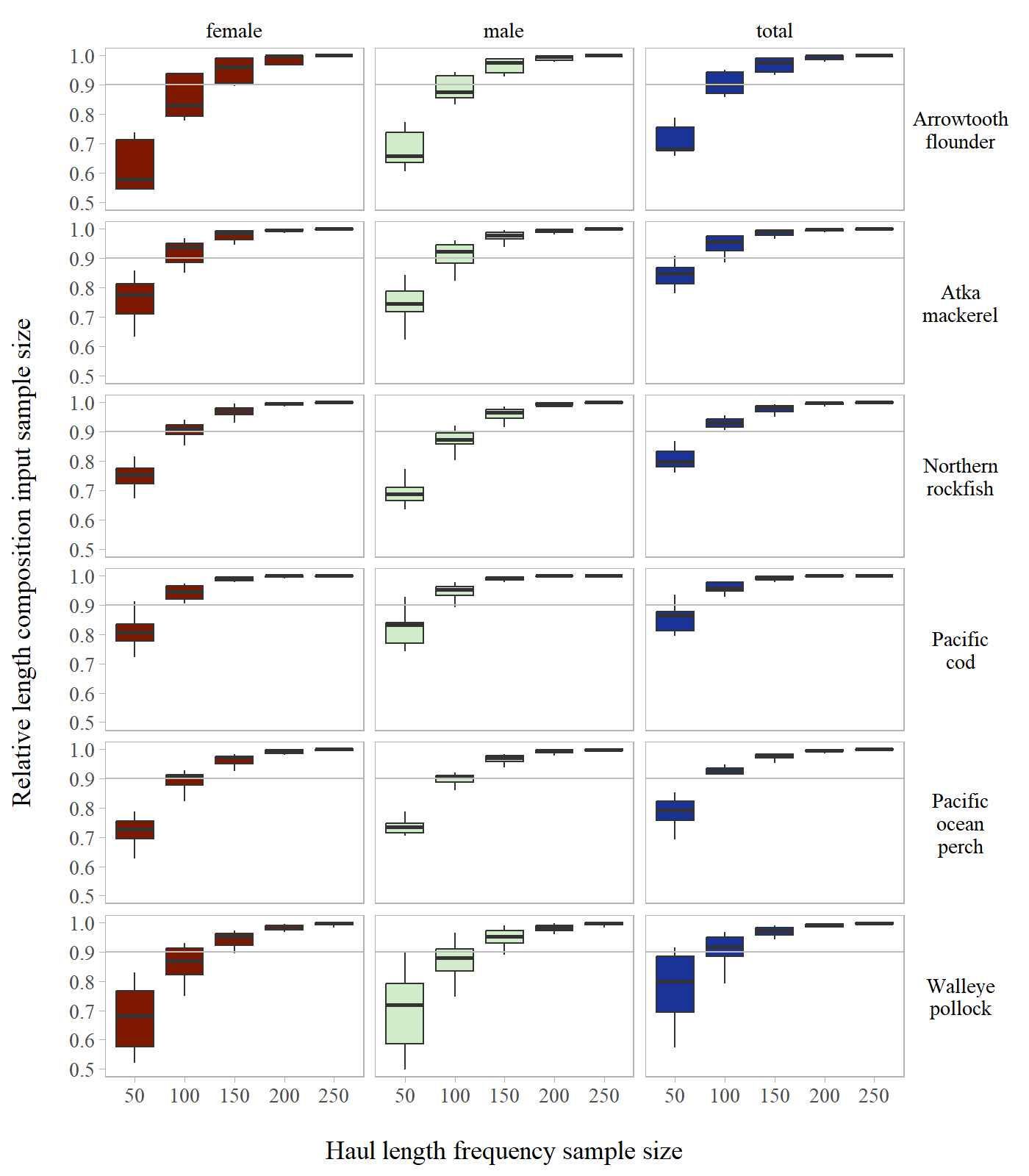


Figure S4: Annual length composition relative input sample size by stock across haul length frequency sub-sampling levels evaluated for the Aleutian Islands.

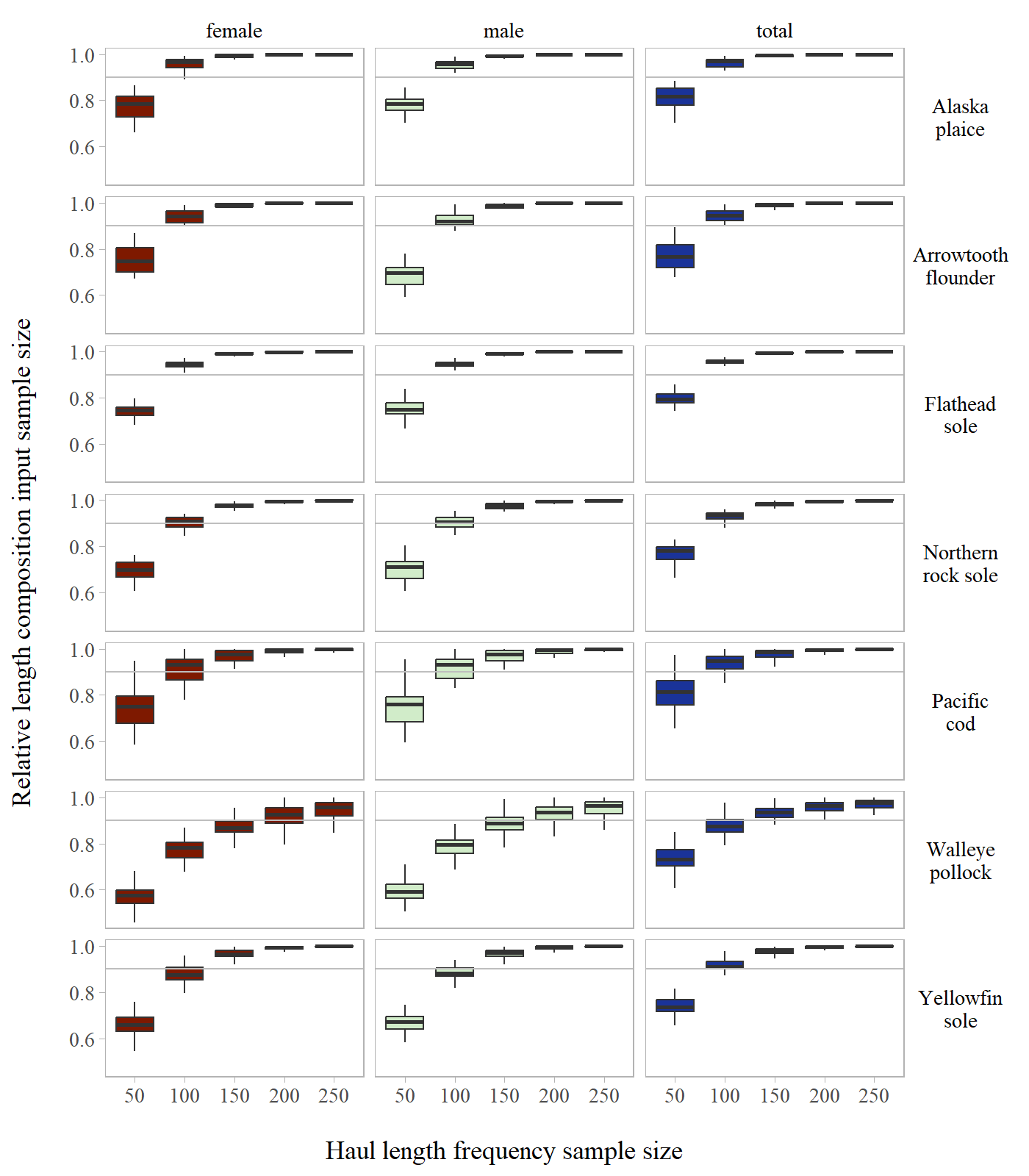


Figure S5: Annual length composition relative input sample size by stock across haul length frequency sub-sampling levels evaluated for the Eastern Bering Sea Shelf.

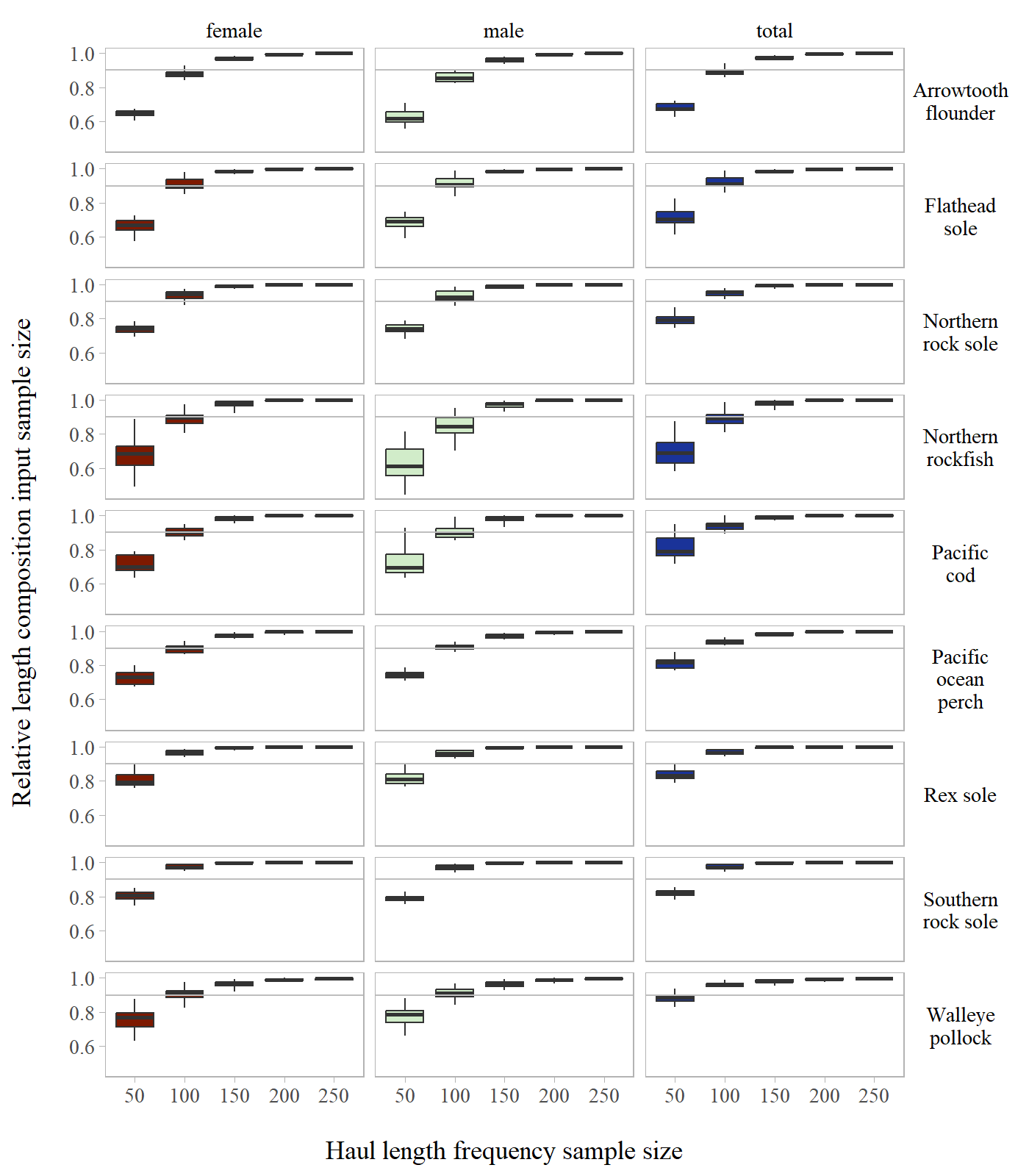


Figure S6: Annual length composition relative input sample size by stock across haul length frequency sub-sampling levels evaluated for the Gulf of Alaska.

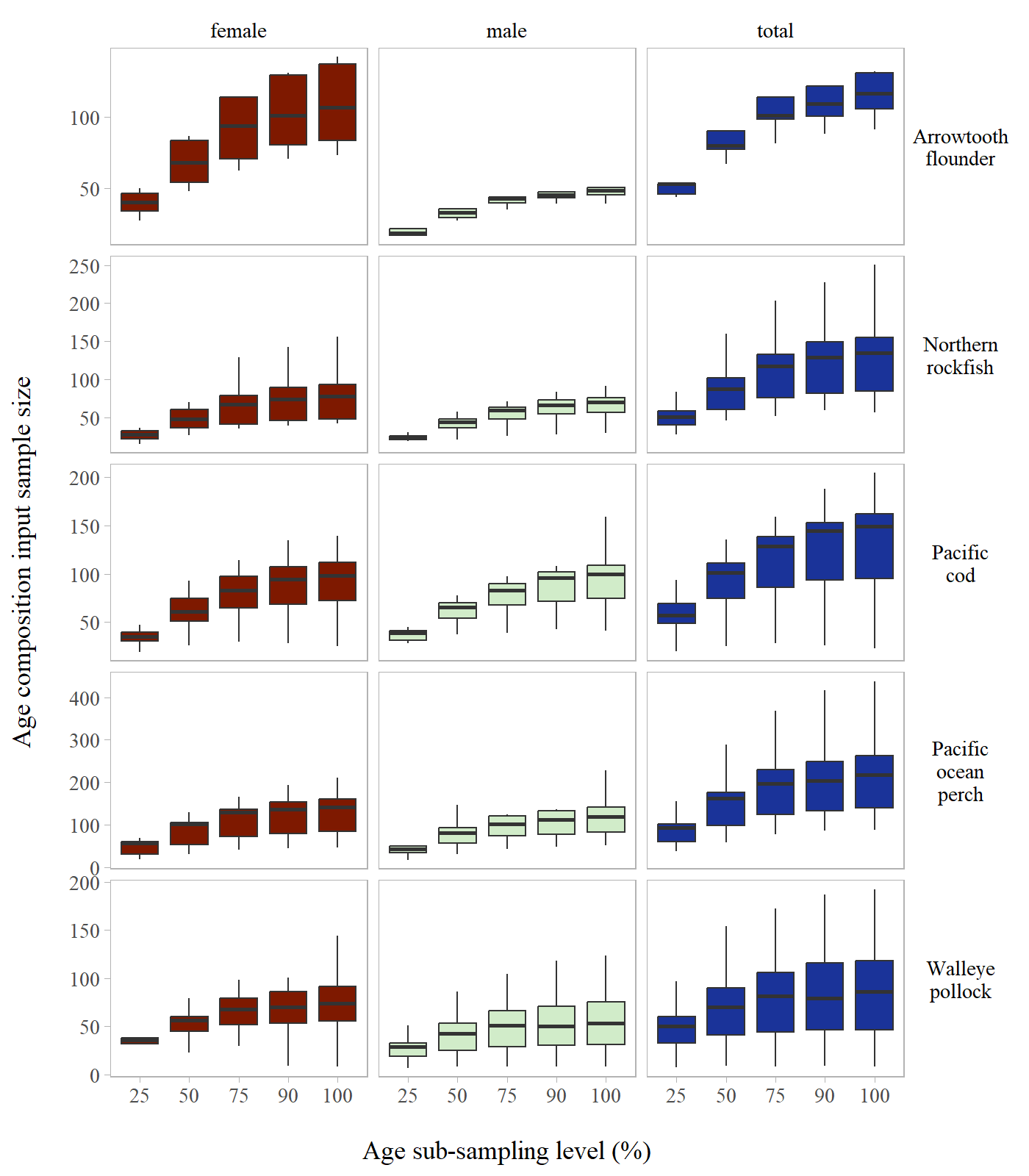


Figure S7: Annual age composition input sample size by stock across age sub-sampling levels evaluated for the Aleutian Islands.

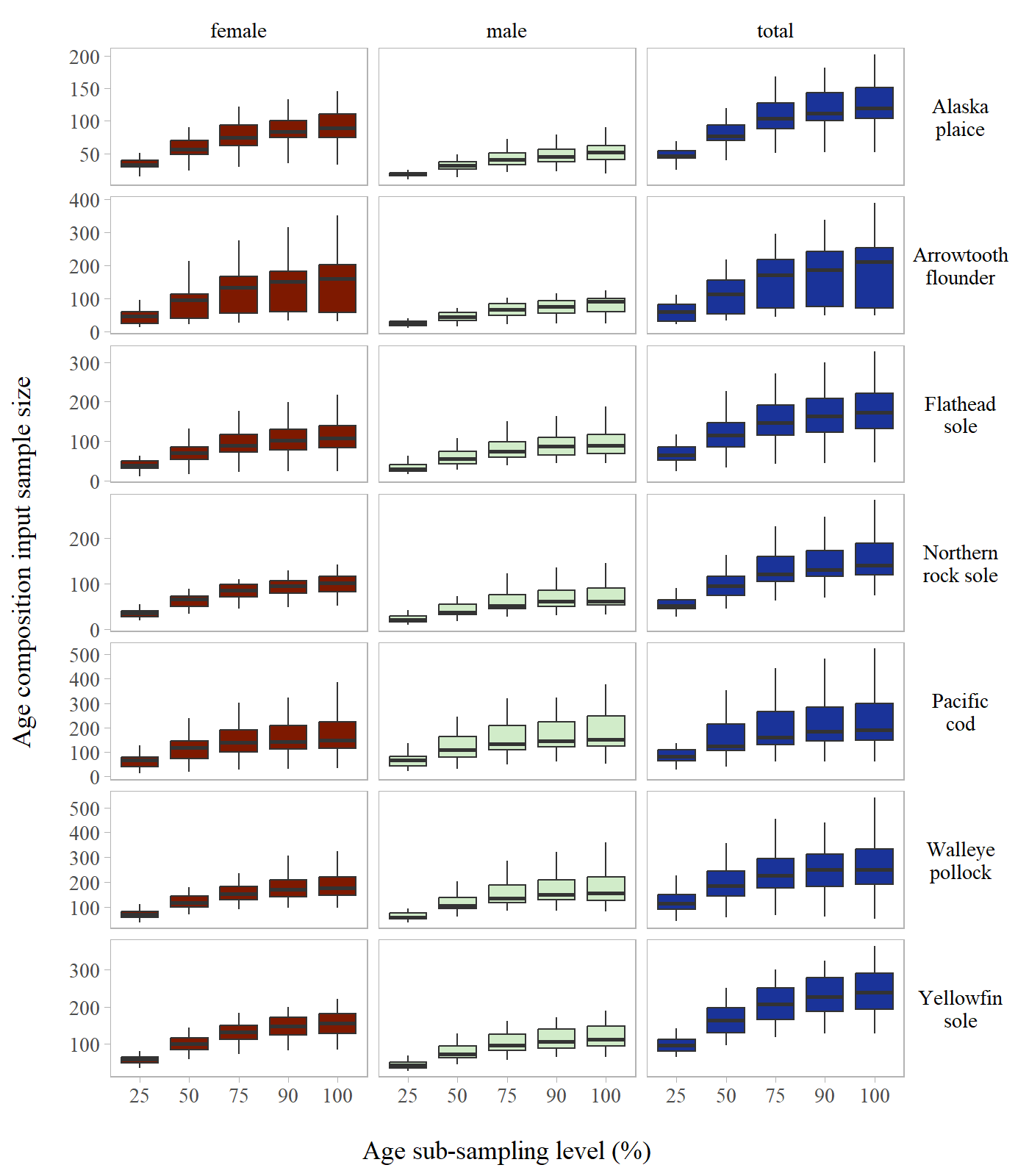


Figure S8: Annual age composition input sample size by stock across age sub-sampling levels evaluated for the Eastern Bering Sea Shelf.

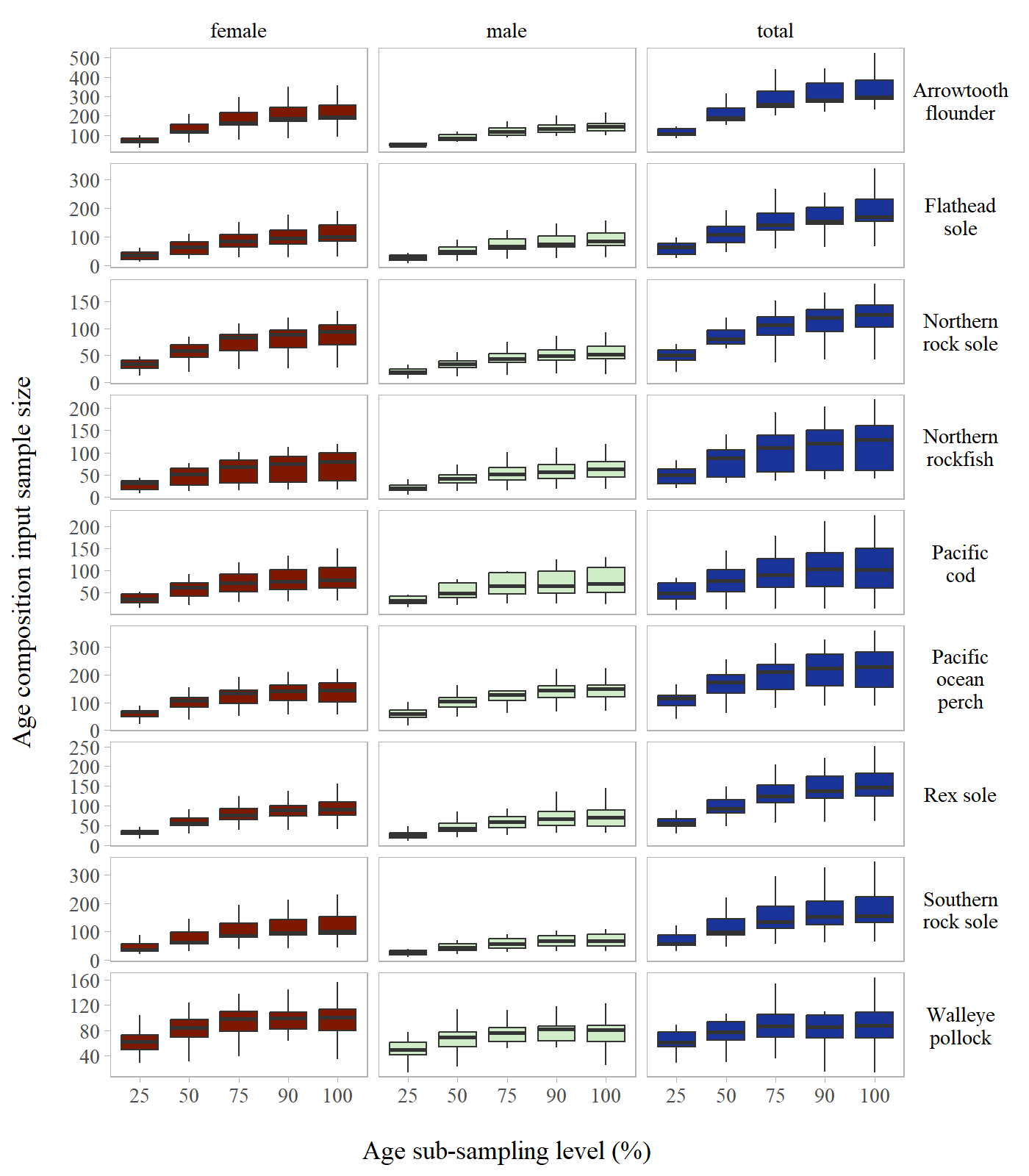


Figure S9: Annual age composition input sample size by stock across age sub-sampling levels evaluated for the Gulf of Alaska.

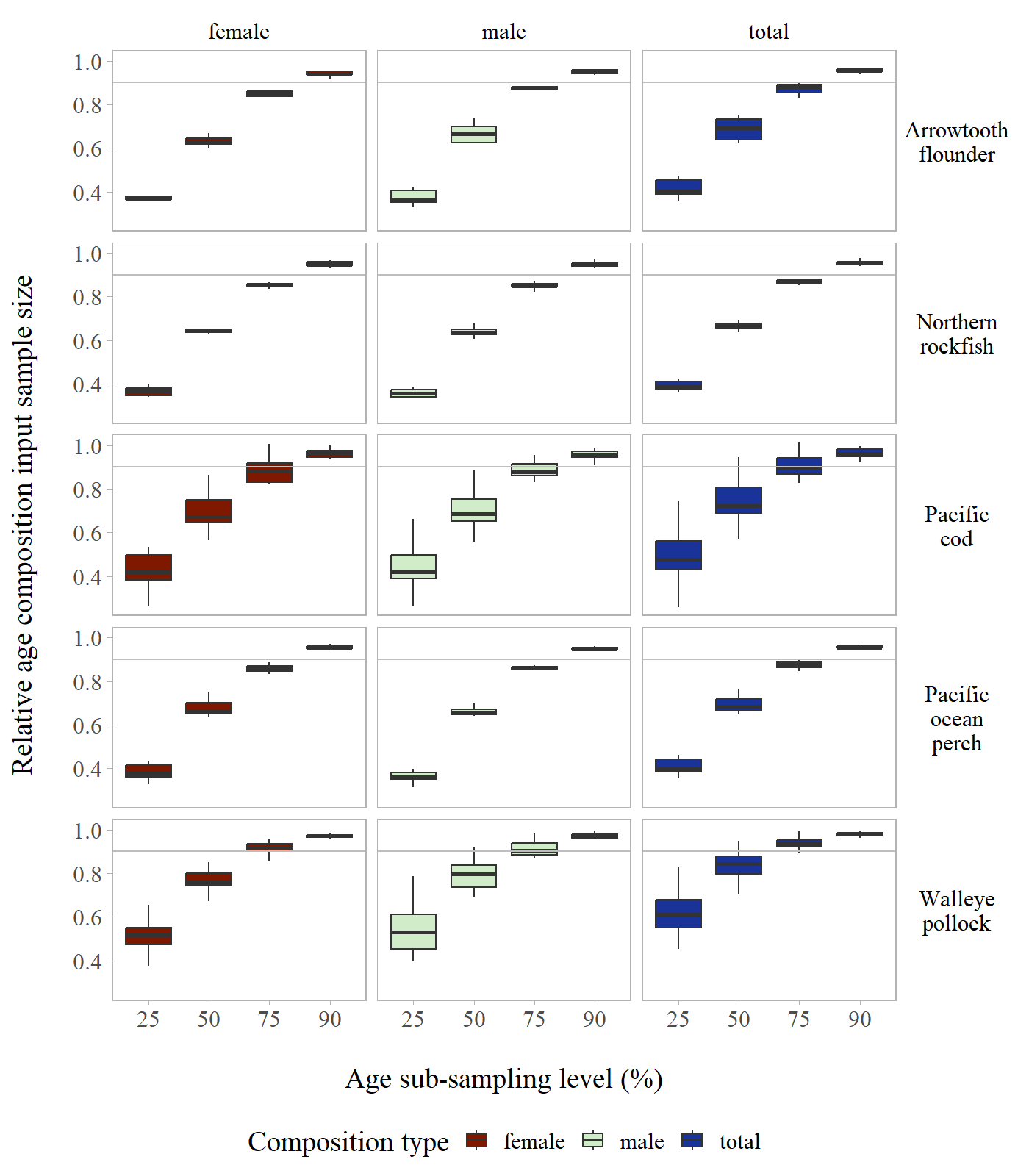


Figure S10: Annual age composition relative input sample size by stock across age sub-sampling levels evaluated for the Aleutian Islands.

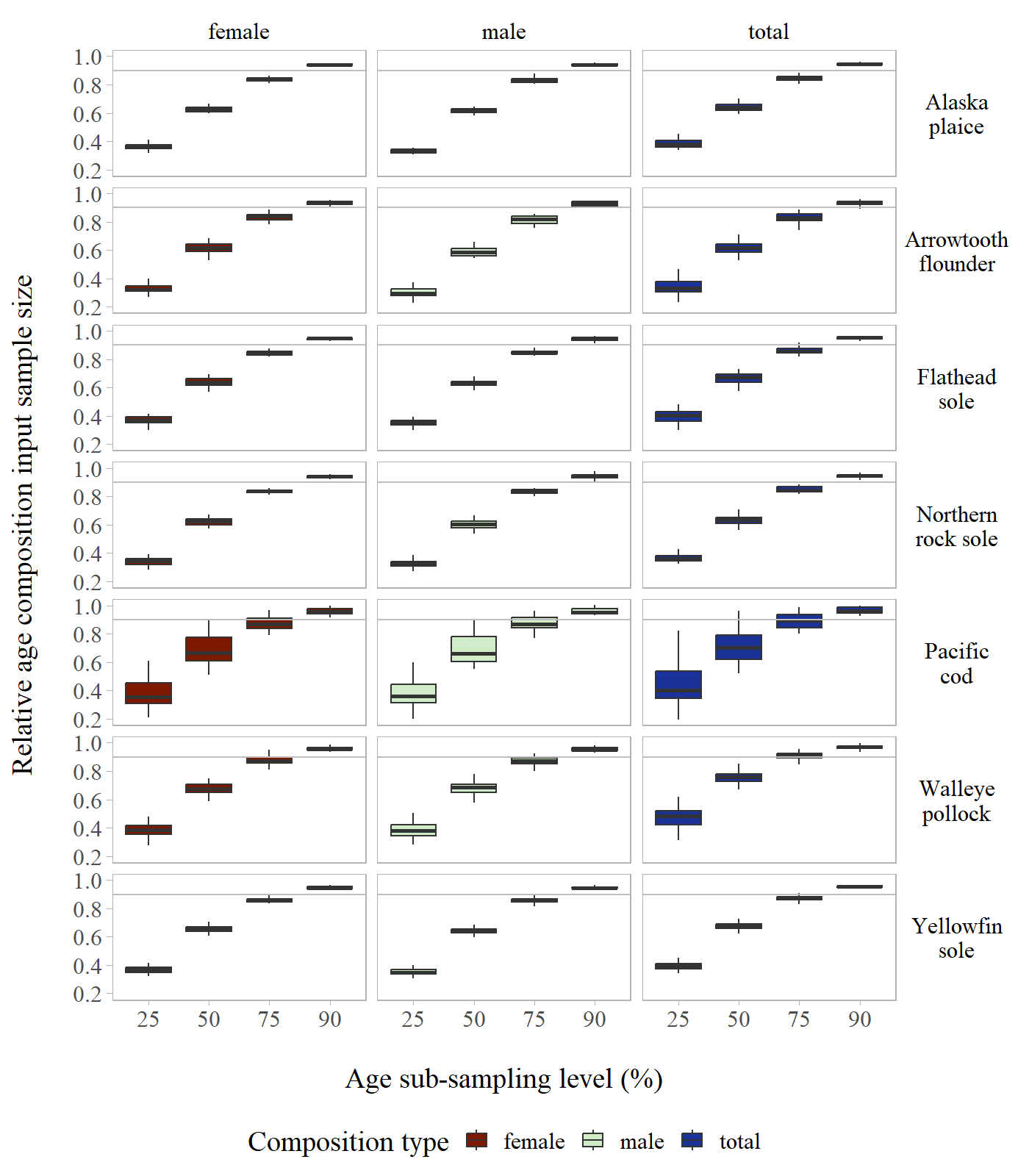


Figure S11: Annual age composition relative input sample size by stock across age sub-sampling levels evaluated for the Eastern Bering Sea Shelf.

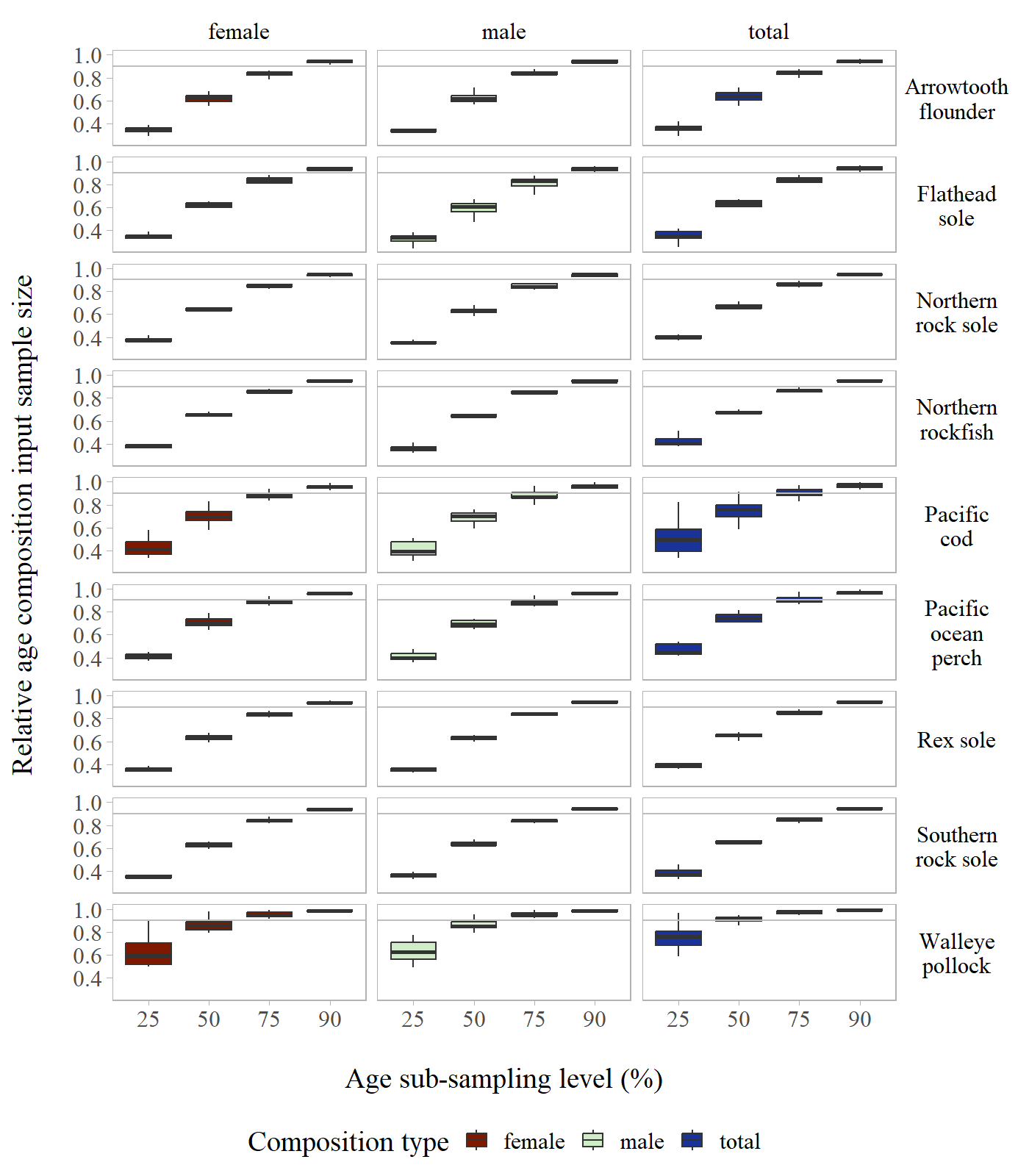


Figure S12: Annual age composition relative input sample size by stock across age sub-sampling levels evaluated for the Gulf of Alaska.