Inclusion of ageing error and growth variability in the bootstrap estimation of age composition input sample size for fisheries stock assessment models

Peter-John F. Hulson1,\*, and Benjamin C. Williams1

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

\* Correspondence: [Peter-John F. Hulson <[pete.hulson@noaa.gov](mailto:pete.hulson@noaa.gov)>](mailto:pete.hulson@noaa.gov)

# Abstract

Statistical catch-at-age assessment models used for fisheries management integrate various sources of information that are statistically weighted in a joint likelihood framework; the relative statistical weighting between these sources of information is an important, yet often a subjective aspect of stock assessment. Input sample size (ISS) is a quantity that is used to statistically weight composition data in these types of models. Both design-based bootstrap and model-based estimators have been proposed, however, these previous methods to determine ISS do not explicitly include sources of uncertainty from ageing error and growth variability that are inherent to expanded age composition data. In this study, we evaluate the impact of including ageing error and growth variability within bootstrap methods that estimate age composition ISS. We find that for all the stocks evaluated the ISS determined from bootstrap methods decreased as these addition sources of uncertainty were included. The decrease in ISS was species type specific, but generally decreased up to 40% when ageing error was introduced, up to 50% when growth variability was included, and up to 60% when both sources of uncertainty were included. These results indicate that there is more variability within age composition data than would be accounted for with ISS estimates that do not include these sources of uncertainty. The method and results provided here allow for assessment scientists to statistically weight age composition with ISS that takes into account ageing error and growth variability that are implicit to any expanded age composition from either fishery-independent or fishery-dependent sources. This has not previously been investigated and including these sources of uncertainty improves bootstrap estimates of ISS to capture all the sources of variability in age composition and will subsequently improve stock assessment model quality.

Key Words: stock assessment, aging error, growth variability, maximum likelihood, uncertainty, input sample size

# Introduction

Compositional information on age and length are critical data products used in statistical catch-at-age assessment (SCAA) models as they facilitate the tracking of year classes and size-structure over time to facilitate our understanding of a fish stock’s population dynamics (Quinn and Deriso 1999), including the size and age based mortality processes through the selectivity of the fisheries. The two primary sources for age and length composition data used in SCAA models are fishery-independent and fishery-dependent. Fishery-independent sources typically include randomized and standardized collection of samples from hauls placed across space in a non-targeted framework. Fishery-dependent sources, on the other hand, are based upon collection of age and length samples, randomized at some level, but obtained from hauls within a trip that is targeting a specific species. A common challenge in using compositional information in SCAA models to estimate population processes is the statistical weighting in the joint likelihood, as the statistical weighting effects the performance of the model. Due to the strong influence that compositional data can have in SCAA models, the statistical weight assigned to these data products are important for providing accurate advice for management (e.g., Hulson et al. 2012, Xu et al. 2020).

Regardless of the source of composition data (whether fishery-independent for fishery-dependent), it is commonly accepted that overdispersion of the data is inherent due to intra-haul correlation (e.g., Pennington and Volstad 1994, Pennington et al. 2002). The concept of effective sample size [ESS; introduced by McAllister and Ianelli (1997)], a reduced sample size from the actual number of fish measured or aged to account for this overdispersion, can be implemented within the likelihood function to statistically weight the age or length composition data. The statistical weight assigned to annual composition data can follow a myriad of methods (e.g., fixed values as in Monnahan et al. (2021), number of samples or tows sampled upon as in Hulson et al. (2021) or Spencer and Ianelli (2022), bootstrapping compositions as in Stewart and Hamel (2014)). The primary consideration when assigning a statistical weight to composition data is to account for the potential variability and correlation in the sampling process that result in overdispersion.

Throughout the development and implementation of the ESS concept in SCAA models a variety of terms have been used, often having multiple meanings for the same term. Often ESS is a term that has been used to denote the sample size used in statistical weighting of age or length composition data (e.g., Hulson et al. 2012, Punt et al. 2021), it has also be used to denote the performance of a SCAA estimates of composition data compared to the observed data (e.g., Thorson and Haltuch 2019). Input sample size (ISS) has also been used as a term to denote the sample size used in statistical weighting of age or length composition data (e.g., Thorson and Haltuch 2019, Thorson et al. 2023). In addition, relative sample size is a term introduced when using bootstrap methodologies (Stewart and Hamel 2014). In order to provide consistency in the literature we propose the following usage of terms as it relates to this issue: \* Nominal sample size: the actual sample size obtained for age or length composition data from fishery-independent or fishery-dependent sources. \* Input sample size: the reduced sample size that accounts for overdispersion of age or length composition data used to statistically weight the composition data in SCAA models. \* Effective sample size: the statistic used to measure the difference in fit between SCAA model estimates of age or length composition data and the observed composition data. \* Relative sample size: the sample size that measures the difference between bootstrap estimates of age or length composition and the observed composition for a given bootstrap iteration. Much of this terminology follows from Thorson et al. (2023) and we reiterate and expand upon it here in an attempt to convince researchers to adhere to a uniform set of terms across the fisheries literature when studying age and length composition data used in SCAA models.

When age is capable of being determined from otoliths, there is further variability in age composition data duet o the ageing of the otolith, often called ‘ageing error’ (e.g., Punt et al. 2008). There are a number of factors that can influence the magnitude of ageing error [for example, the number of age classes or the sample size; Nesslage et al. (2022)], but inherent to obtaining ages from otoliths is variability in the age readings across the laboratory age readers. To account for this source of variability, ageing laboratories regularly evaluate precision through obtaining multiple readings of the same otolith across different age readers (Morison et al. 2005). Several methods have been developed to account for ageing error in SCAA models when fitting age composition by integrating an ageing error matrix as an additional input data source for the model (Punt et al. 2008, Candy et al. 2012). The ageing error matrix is used to ‘correct’ the numbers-at-age estimated by the assessment model by assigning a certain proportion of fish in a given age class to adjacent age-classes based on the magnitude of the ageing error within the specific age-class. Since the development and implementation of ageing error matrices a number of studies have been devoted to quantifying the effects of ageing error on assessment model estimates Liao et al. (2013). Within each of these studies, and in each application of an ageing error matrix within a SCAA model, the age composition data will be statistically weighted by an ISS. As described previously, the ISS selected to statistically weight the age composition data should reflect the variability in the sampling process, thus, it should also include the variability in the age readings themselves.

In the process of obtaining an observed age composition, it is the case that an age-length key (ALK) is employed to expand the estimated population numbers-at-length to population numbers-at-age (Quinn and Deriso 1999, Ailloud and Hoenig 2019). For either fishery-dependent or fishery-independent sources of age composition, if an expansion process is used to obtain an observed age composition it is generally the case that length frequency is expanded to some geographic area by weighting haul-level length frequency by haul-level catch-per-unit-effort (in numbers), this provides an estimated population numbers-at-length. Then, an ALK is constructed with age-length paired data and multiplied by the estimated population numbers-at-length to obtain estimated population numbers-at-age, often referred to as expanded age composition data (Siskey et al. 2023). An intrinsic component to the ALK is the variability in length for a given age. This variability in growth, or the range in lengths that are observed for a given age, is directly linked to the variability in the expanded age composition, and thus, should be reflected in the ISS selected to statistically weight the age composition data within a SCAA model.

To date, no method has been developed to integrate the variability in the ageing process when reading otoliths and in the growth process upon which age-length keys are based within the estimation of ISS used to statistically weight age composition data in SCAA models. The bootstrap method developed by Stewart and Hamel (2014) allows for resampling techniques to be employed at each level of the sampling design, and provides an objective avenue to estimate ISS that is based on the observation variability contained within the sampling process. In this study, we extend the methods of Stewart and Hamel (2014) to estimate age composition ISS that includes both ageing error and growth variability in the estimation process. We show, in a step-wise process, the added variability in age composition sample size from including ageing error and growth variability across a number of species that reflect differing life histories and levels of ageing difficulty. *Through a case study involving several North Pacific SCAA models we also show the potential consequences of not including these sources of variability in the application of ISS to weight age composition data.*

# Methods

## Data

We used historical data collected from bottom trawl surveys conducted by the Alaska Fisheries Science Center (AFSC) in the Eastern Bering Sea (EBS: Lauth et al. 2019), Aleutian Islands (AI: von Szalay et al. 2017), and Gulf of Alaska (GOA: von Szalay and Raring 2018). Within the AFSC bottom trawl surveys both length frequency data and age specimen data are collected, in addition to other survey data (e.g., catch, effort, location). Generally, a subsample of fish from each haul were processed at sea to collect their sex, length, and weight. A subsample of these fish have their sagittal otoliths collected; these otoliths were sent (with haul and specimen data) to the AFSC Age and Growth laboratory for age determination. Periodically, a subset of aged otoliths are selected for reader-tester agreement tests. These tests are used to evaluate the reproducability of an age reading when two different readers age the same fish without knowledge of the other reader’s age determination of the otolith (Kimura and Lyons 1991). The average annual bottom trawl survey age sample sizes by region, and the total number of otoliths used for reader-tester agreement tests are shown in Table 1) for the species evaluated. The species or stocks selected for this analysis all have greater than 5,000 reader-tester paired otolith readings. These stocks are all assessed using integrated SCAA models that require input sample sizes for the age composition data.

## Length and age composition expansion

Details of how the length frequency and age collections are expanded to population abundance-at-length and -age then subsequently used as compositional data in stock assessment models at AFSC are provided in Hulson et al. (in review). Here we generalize these methods to provide the reader with a broad understanding of how length and age composition are expanded in the AFSC bottom trawl surveys.

Length frequency samples collected by the AFSC bottom trawl surveys are expanded by catch and stratum area to obtain estimates of population abundance-at-length. This is often referred to as the ‘first stage expansion’ and is a common method to obtain population abundance estimates at length from area-swept survey data (Miller and Skalski 2006, e.g., Ailloud and Hoenig 2019). To expand the species-specific length frequency samples to population-at-length we first compute the overall population numbers within a stratum by multiplying the average catch per unit effort within the strata (i.e., the number of fish per square kilometer averaged across the hauls performed within the strata) by the area of the strata (in square kilometers). The overall population numbers year-*y* within stratum-*s* () is computed with

where is the area of stratum-*s* (in km2), and is the species-specific average catch per unit effort of numbers captured across the hauls within a strata in year-*y* We then compute the relative catch per unit effort for each haul performed within the strata and the sex-specific relative length composition for each haul. The relative catch per unit effort for each haul () is computed by

where is the catch per unit effort of numbers caught within a haul-*h* for stratum-*s* in year-*y*. The sex-specific relative length composition for each haul () is computed with

where is the length frequency sampled, in numbers, by sex-*x* and length-*l* (in cm) within a haul-*h* for stratum-*s* in year-*y*.

Finally, the expanded population abundance-at-length is obtained by multiplying the overall population numbers within the strata ((1)), the relative catch per unit effort of each haul ((2)), and the sex-specific relative length composition ((3)) with

and to obtain the sex-specific estimates of population abundance-at-length in a management area one would simply sum across strata.

In the second stage expansion the sex-specific estimates of population abundance-at-length are used to estimate sex-specific population abundance-at-age. The annual specimen data that are collected during the survey, which include observations of age-at-length, are first populated into sex-specific numbers at age and length (). Next, the sex-specific numbers-at-age and length are converted to sex-specific proportions of age-at-length (i.e., age-length key) with

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

where is the population abundance-at-length from (4) summed across strata.

## Simulation-Bootstrap framework

To evaluate the effect of the inclusion of ageing error and growth variability on uncertainty in age compostion datasets we modified a bootstrap-simulation framework (Hulson et al. in review) to include these additional sources of error. In simple terms, the simulation framework is a two-stage bootstrap that first resamples hauls, then resamples lengths and ages collected within the resampled hauls following from the methods in Stewart and Hamel (2014). The simulation framework was modified to account for ageing error by resampling from tester ages associated with a given reader age. Growth variability was incorporated, for each species, by resampling fom all lengths associated with a given age and sex.

The order of operations (Figure 1 has the following schedule:

1. Resample hauls 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
3. With the resampled length frequency data from step 2, calculate population abundance-at-length
4. Within the resampled hauls from step 1, resample the observed ages from the specimen data
5. For the resampled ages in step 4, resample a length from the set of lengths observed for the given age
6. For the resampled ages in step 4, resample an age from the set of tester ages for the given age
7. With the resampled age data in steps 4-6 and the population abundance-at-length in step 3, calculate the population abundance-at-age

Steps 5 and 6 were designed to explore inclusion of ageing error and growth variability. The bootstrap-simulation repeats steps 1-7 iteratively providing iterated population abundance-at-age for comparison to the historical (the full sample without any resampling of data) population abundance-at-age.

We applied the bootstrap-simulation in a step-wise manner to evaluate the consequences of adding each source of additional error to the age composition estimates across what we term ‘uncertainty scenarios’. First, we ran the bootstrap-simulation with the historical data without adding any extra error, thus, omitting steps 5 and 6 above (‘Base’ scenario). Next, we added ageing error (‘AE’ scenario) and growth variability (‘GV’ scenario) separately, thus, omitting either step 5 or 6 depending on the source of uncertainty desired. Finally, we added both ageing error and growth variability (‘AE & GV’ scenario) to the bootstrap-simulation framework. To increase reader-tester sample sizes for each species, we pooled reader-tester data across the three regions (we note that age readings for all three regions are produced in the same age reading laboratory at AFSC). Sex-specific age-length pairs were pooled across all survey years, by region, in order to evaluate the maximum influence of growth variability on the replicated sex-specific age composition estimates. The bootstrap-simulations were run for 500 iterations, a level at which the variability in population abundance-at-age 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/surveyISS>).

## Computing input sample size

A useful statistic that can quantify the variability in age composition is relative sample size, introduced by McAllister and Ianelli (1997). This statistic evaluates the amount of uncertainty in an estimated composition compared to an observed composition and is given by:

where is the estimated proportion for category-*c* (which can be age or any other arbitrary category across which proportions are computed) and is the observed proportion. Here, the underlying age composition derived from the historical bottom trawl surveys with the full and unsampled data was treated as the observed proportions in equation (1). For each iteration-*i* of the bootstrap-simulation we computed an estimated proportion () that was then compared to the underlying historical age composition () to determine the relative sample size () of the resampled age composition. Thus, across each iteration of the bootstrap-simulation we computed a relative sample size that indicated the amount of uncertainty in the resampled age composition.

To summarize relative sample size across iterations we used the harmonic mean. This has been shown to reduce bias in recovering the true sample size in simulations for a multinomial distribution and has also been recommended to determine the input sample size (ISS) that is used in stock assessment models to fit compositional data (Stewart and Hamel 2014). While we present the results of the annual ISS for each stock evaluated when incorporating ageing error, growth variability, or both, we also compute the proportion of ‘base’ ISS in order to present the relative decrease in ISS when incorporating these sources of additional uncertainty. The ‘relative ISS’ is computed by dividing the ISS as determined after incorporating ageing error, growth variability, or both, by the base ISS without these sources of uncertainty.

## Evaluating life-history relationships to consequences of added uncertainty

Trends across stocks were aggregated and evaluated using three species types: flatfish (arrowtooth flounder, flathead sole, northern rock sole, and yellowfin sole), gadids (walleye pollock and Pacific cod), and rockfish (Pacific ocean perch and northern rockfish). Two statistics were used to evaluate trends across the cases that added ageing error and growth variability. The first statistic we used to evaluate the influence of adding ageing error was the average coefficient of variation (CV) in age agreement. We computed this by computing the CV in the tester ages for a given reader age, then averaged these CVs across the reader ages. The second statistic we used to evaluate the influence of adding growth variability was the average CV in age-length data. We computed the average age-length CV by computing the CV in length for a given age, then averaging the CVs in length-at-age across age.

# Results

While the magnitude of age composition ISS was stock and region specific, the general result that was consistent across the stocks evaluated was a reduction in age composition ISS as additional sources of uncertainty were introduced in the bootstrap procedure (Figure 2). This reduction in age composition ISS resulted for both sex-specific and total (combined sex) age composition ISS. The relative magnitude of adding ageing error compared to growth variability was also stock and region specific. For example, adding ageing error to arrowtooth flounder age data resulted in smaller ISS than adding growth variability in the GOA, but larger ISS in the EBS. Variability in the age composition ISS across the scenarios evaluated was also stock and region specific, with Pacific cod and walleye pollock in the EBS having the greatest reduction in ISS and AI Atka mackerel having the least.

The age composition relative ISS across uncertainty scenarios revealed patterns across species types within each region evaluated (Figure 3). Including ageing error for rockfish had a larger proportional reduction in relative ISS than including growth variability. However, growth variability had a larger proportional reduction for flatfish and gadid relative ISS than adding ageing error. Flatfish exhibited the greatest variability in the proportional reduction in relative ISS in any given region, while gadids had the smallest relative ISS, in general. For example, after adding ageing error and growth variability the relative ISS for gadid was 85-95% in the AI, 40-50% in the EBS, and 70-80% in the GOA (Figure 3). Overall, adding both ageing error and growth variability decreased the relative ISS to 50-90% of the Base ISS for flatfish, 40-90% for gadids, and 80-95% for rockfish.

A positive relationship is observed, by species group, between the number of age samples taken per haul and the age composition ISS per sampled haul (top panel of Figure 4). However, as sources of uncertainty were added to age data this positive trend was dampened for each species type; this was particularly true for gadids as compared to flatfish and rockfish. Across species types the age composition ISS per haul was around half of the number of sampled ages per haul, less for rockfish and gadids. The age composition ISS per haul was, on average, the largest for flatfish, followed by rockfish, and was the smallest for gadids (bottom panel of Figure 4). The variability in the age composition ISS per haul across years, regions, and sex categories was also the largest for flatfish compared to gadids and rockfish.

Comparing between statistics for ageing error and growth variability resulted in different trends in the relative ISS by species types (top panels of Figure 5). The relative ISS when adding ageing error was similar when compared to the average reader-tester CV between gadids and rockfish, with lower variability for rockfish (top left panel of Figure 5). While the range in average reader-tester CV was smaller for rockfish and gadids than for flatfish, the resulting range in relative ISS was larger for gadids than rockfish or flatfish. The relative ISS when adding growth variability resulted in the smallest range in both ISS reduction and average age-length CV for rockfish, though flatfish and gadid had a larger CV spread (top right panel of Figure 5). On average, the relative ISS when adding both ageing error and growth variability was largest for rockfish, intermediate for flatfish, and smallest for gadids (bottom panel of Figure 5). The same trend resulted when evaluating the range in the relative ISS when both ageing error and growth variability were added.

# Discussion

*add a consequences of data weighting para somewhere*

In this study we find that accounting for ageing error and growth variability using bootstrap procedures decreased age composition ISS for all stocks examined. The impact of the sources of uncertainty on resulting ISS was species type specific, with ageing error being more influential for rockfish than growth variability, and growth variability more influential than ageing error for flatfish and gadids. We propose that these results are due to larger growth variability observed in gadids and flatfish compared to rockfish. Further, the effects of ageing error are not unexpected for rockfish, as they are so long-lived. However, the average CV for reader-tester agreement was much larger for some flatfish stocks relative to the rockfish stocks evaluated, thus, there was some interaction between the number of age bins and ISS in this analysis. When considering both ageing error and growth uncertainty the largest reduction in ISS magnitude was for gadids, followed by flatfish, with the least effect observed for rockfish, though results varies by stock and region.

Understanding effects of survey reduction effort has recently been the focus of a number of studies (ICES 2020, 2023), as survey reductions may be inevitable in many regions. A recent study investigated the reductions in length frequency and age collection effort, using AFSC bottom trawl survey as example (Hulson et al. in review) and found that reduction in age collections had larger effect on flatfish and rockfish as compared to gadids. Here we find that including additional sources of uncertainty has largest effect on gadids, and smallest on rockfish, and it is potentially the case that the effect of decreases in sampling effort for gadids and flatfish would be smaller given the magnitude of effect by these sources of uncertainty as compared to rockfish. However, we acknowledge that this should be evaluated in future studies to understand the specific effects on stocks and when comparing among species types.

It is well known that misspecification of ISS when fitting compositional data can lead to bias results in assessment model predictions (e.g., Stewart and Monnahan 2017, Xu et al. 2020). Here we show for gadids that there is upwards of a 50% decrease in ISS for some stocks when adding additional uncertainty as compared to the base case, an upwards of 40% decrease for flatfish, and an upwards of 20% decrease for rockfish. Without these additional sources of uncertainty taken into account, using the bootstrap procedure would result in ISS that are larger than what they should be. While we did not investigate implications to specific assessment model outcomes, it can be inferred that reductions of ISS on this scale would have downstream effects on model predictions and the associated uncertainty. We note that these sources of uncertainty would not be contained only to fishery-independent sources, like evaluated here, but would also be inherent to age collections for fishery-dependent sources as well. Thus, future investigations into the impacts of adding ageing error and growth variability into ISS estimation on assessment model results should also include fishery-dependent ISS implications as well.

A number of operational assessment models use hauls as a proxy in some form when setting ISS for age and length composition data Hulson et al. (2021). This is derived from a result found in Pennington et al. (2002) who investigated length frequency sampling and, based on the level of intra-haul correlation, determined that for the species that were investigated, on average, the ISS was one fish per haul However, we note that the conclusion made in Pennington et al. (2002) was not that the number of hauls should be used as a proxy for ISS in the assessment model data fitting procedure, but rather, that in order to potentially increase the ISS and have a better estimate of the level of intra-haul correlation that samples should be taken from an increased number of hauls. This result was further substantiated in Siskey et al. (2023), who also suggested that increasing the number of hauls while decreasing the sampling effort on any given haul was preferable to increasing the sampling effort within any given haul at the expense of sampling from a greater number of hauls. Here we find that the age composition ISS per haul was dependent on the species type, had a large range depending on the type of error included in determining ISS, and was generally larger than one fish per haul.

Previous work has investigated both the inclusion of ageing error (Punt et al. 2008, Liao et al. 2013) and growth (Taylor and Methot 2013) within stock assessment models, however, none have applied these additional sources of uncertainty in the context of estimating ISS. As it pertains to ageing error, in many current assessments an ageing error matrix is implemented (e.g., Williams et al. 2022) in order to account for ageing error that is inherent to the age composition data. Use of an ageing error matrix effectively ‘smudges’ assessment model estimates of population-at-age into adjacent age classes prior to fitting the ‘observed’ age composition input data from either fishery-independent or fishery-dependent sources. Then, in the model fitting step, an ISS is used to fit the models ‘smudged’ estimates of age composition to the ‘observed’ age composition. Here we argue that unless ageing error is accounted for in the age composition ISS used to fit these data then we are likely ‘over-fitting’ the model estimates to the ‘observed’ age composition and that the use of only an ageing error matrix half-way accounts for this source of uncertainty in the assessment modeling process. The results from this study suggest that in some cases the addition of ageing error to the bootstrap method decreases the age composition ISS by 20-40%, thus, an ISS that does not take into account this source of uncertainty can be 125-165% too large.

It is commonly the case that age composition is produced through a two-stage expansion process (Quinn and Deriso 1999), in which length frequency data is expanded to population-at-length in the first stage, and an ALK is used to expand population-at-length to population-at-age in the second stage (Ailloud and Hoenig 2019). In the second-stage of this expansion process, the ALK is produced through the use of age-length paired data that are obtained in the age sampling collection, and within the ALK the variability in length-at-age is implicitly accounted for. However, there have been no previous attempts to include this source of variability when considering the ISS that is used to fit the expanded age composition data. We find that the magnitude of effect on age composition ISS is species type dependent, in which a larger effect results for species that exhibit larger variability in growth, for example, gadids and flatfish compared to rockfish. When implementing growth variability there was a 10-50% decrease in the magnitude of bootstrapped age composition ISS, thus, if not taking this source of uncertainty into account the ISS can be up to 200% too large.

Overall, we find that expanding upon the method introduced by Stewart and Monnahan (2017) by including ageing error and growth variability into estimation of age composition ISS can have large effect in reducing the magnitude of ISS. We provide two primary recommendations from this work. First, we recommend that stock assessment scientists consider the use of bootstrap methods like this one (which was developed from Stewart and Monnahan 2017) to set and and length composition ISS. With modern computing power, for a single species using the package we built (<https://github.com/BenWilliams-NOAA/surveyISS>) it takes on the matter of a couple of hours to obtain both age and length composition bootstrap ISS for a historical survey time series (longer than 40 years in some cases); for a single year it takes a matter of minutes. Second, for all estimates of age composition ISS we recommend implementing ageing error and growth variability to more explicitly and thoroughly take these sources of uncertainty into account in stock assessment models. We note that while we used fishery-independent data here as an example, that these sources of uncertainty would also be inherent to fishery-dependent data.

# Acknowledgments

We thank *reviewer1* and *reviewer2* for their helpful reviews of this manuscript. We also thank all the AFSC survey staff who collected the data over the last 40 years used in this analysis.

# 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.

Barbeaux, S.J., Barnett, L., Connor, J., Nielson, J., Shotwell, S.K., Siddon, E., and Spies, I. 2022. Assessment of the Pacific cod stock in the Eastern Bering Sea. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, L92 Building, 4th floor, Anchorage, AK 99501.

Candy, S.G., Nowara, G.B., Welsford, D., and McKinlay, J.P. 2012. Estimating an ageing error matrix for patagonian toothfish (*dissostichus eleginoides*) otoliths using between-reader integer errors, readability scores, and continuation ratio models. Fisheries Research 115: 14–23.

Henriquez, V., Licandeo, R., Cubillos, L.A., and Cox, S.P. 2016. Interactions between ageing error and selectivity in statistical catch-at-age models: Simulations and implications for assessment of the chilean patagonian toothfish fishery. ICES Journal of Marine Science 73(4): 1074–1090.

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., Williams, B., Bryan, M., Conner, J., and Siskey, M. in review. Reductions in sampling effort for fishery-independent age and length composition: balancing stock assessment input data uncertainty and workforce health and efficiency.

Hulson, P.-J.F., Williams, B.C., Fissel, B.E., Ferriss, B.E., Hall, M., Yasumiishi, E.M., and Jones, D.T. 2021. Assessment of the Pacific ocean perch stock in the Gulf of Alaska. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, L92 Building, 4th floor, Anchorage, AK 99501.

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.

Kimura, D.K., and Lyons, J.J. 1991. Between-reader bias and variability in the age-determination process. Fishery Bulletin 89: 53–60.

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.

Liao, H., Sharov, A.F., Jones, C.M., and Nelson, G.A. 2013. Quantifying the effects of aging bias in atlantic striped bass stock assessment. Transactions of the American Fisheries Society 142(1): 193–207.

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.

Monnahan, C.C., Dorn, M.W., Deary, A.L., Ferriss, B.E., Fissel, B.E., Honkalehto, T., Jones, D.T., Levine, M., Rogers, L., Shotwell, S.K., Tyrell, A., and Zador, S. 2021. Assessment of the walleye pollock stock in the Gulf of Alaska. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, L92 Building, 4th floor, Anchorage, AK 99501.

Morison, A., Burnett, J., McCurdy, W., and Moksness, E. 2005. Quality issues in the use of otoliths for fish age estimation. Marine and Freshwater Research 56.

Nesslage, G., Schueller, A.M., Rezek, A.R., and Mroch III, R.M. 2022. Influence of sample size and number of age classes on characterization of ageing error in paired-age comparisons. Fisheries Research 249: 106255.

Pennington, M., Burmeister, L.M., and Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fishery Bulletin 100(1): 74–80.

Pennington, M., and Volstad, J.H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics 50(3): 725–732.

Punt, A.E., Smith, D.C., KrusicGolub, K., and Robertson, S. 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in australia’s southern and eastern scalefish and shark fishery. Canadian Journal of Fisheries and Aquatic Sciences 65(9): 1991–2005.

Punt, A.E., Tuck, G.N., Day, J., Burch, P., Thomson, R.B., and Bessell-Browne, P. 2021. The impact of alternative age-length sampling schemes on the performance of stock assessment methods. Fisheries Research 238.

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.

Spencer, P.D., and Ianelli, J.I. 2022. Assessment of the Pacific ocean perch stock in the Bering Sea/Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, L92 Building, 4th floor, Anchorage, AK 99501.

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.

Stewart, I.J., and Monnahan, C.C. 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fisheries Research 192: 126–134.

Taylor, I.G., and Methot, R.D. 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fisheries Research 142: 75–85.

Thorson, J.T., and Haltuch, M.A. 2019. Spatiotemporal analysis of compositional data: Increased precision and improved workflow using model-based inputs to stock assessment. Canadian Journal of Fisheries and Aquatic Sciences 76(3): 401–414.

Thorson, J.T., Monnahan, C.C., and Hulson, P.-J.F. 2023. Data weighting: An iterative process linking surveys, data synthesis, and population models to evaluate mis-specification. Fisheries Research.

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.

Williams, B.C., Hulson, P.-J.F., Lunsford, C.R., and Ferriss, B. 2022. Assessment of the Northern rockfish stock in the Gulf of Alaska. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, L92 Building, 4th floor, Anchorage, AK 99501.

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.

# Tables

Table 1: Average age samples from the AFSC bottom trawl surveys by region (rounded to the nearest 10), and total reader-tester age pairs (rounded to the nearest 100) for the species evaluated in the bootstrap-simulation.

| Stock | Scientific name | AI | EBS | GOA | R-T |
| --- | --- | --- | --- | --- | --- |
| arrowtooth flounder | Atheresthes stomias | 450 | 480 | 850 | 6,100 |
| Atka mackerel | Pleurogrammus monopterygius | 600 | – | – | 12,900 |
| flathead sole | Hippoglossoides elassodon | – | 560 | 520 | 9,400 |
| northern rock sole | Lepidopsetta polyxystra | – | 460 | 450 | 8,900 |
| northern rockfish | Sebastes polyspinis | 570 | – | 450 | 6,400 |
| Pacific cod | Gadus macrocephalus | 800 | 1070 | 650 | 21,200 |
| Pacific ocean perch | Sebastes alutus | 940 | – | 1030 | 13,500 |
| walleye pollock | Gadus chalcogrammus | 790 | 1500 | 1300 | 84,400 |
| yellowfin sole | Limanda aspera | – | 750 | – | 10,300 |

# Figures



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

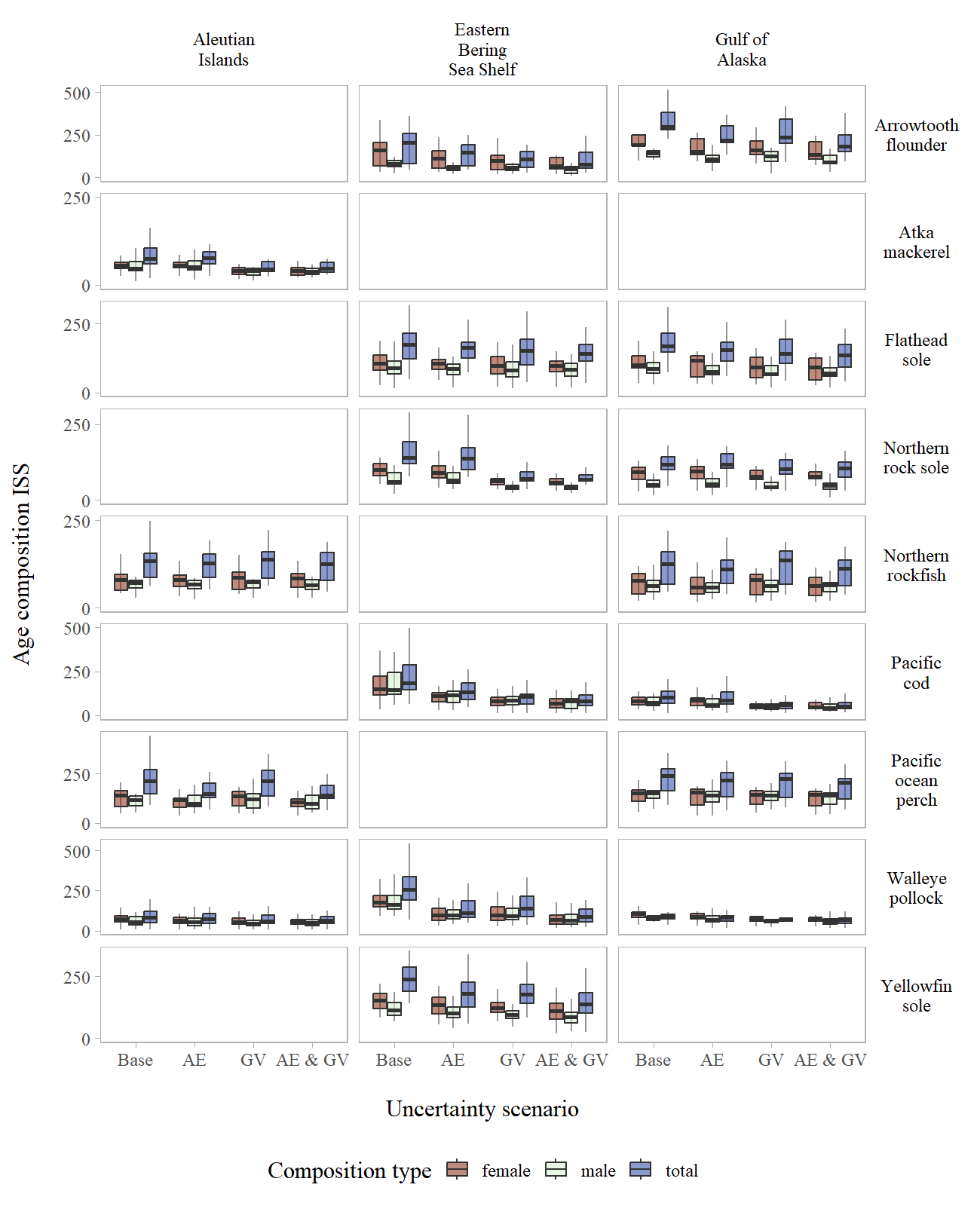


Figure 2: Age composition input sample size for the stocks and regions evaluated across uncertainty scenarios.

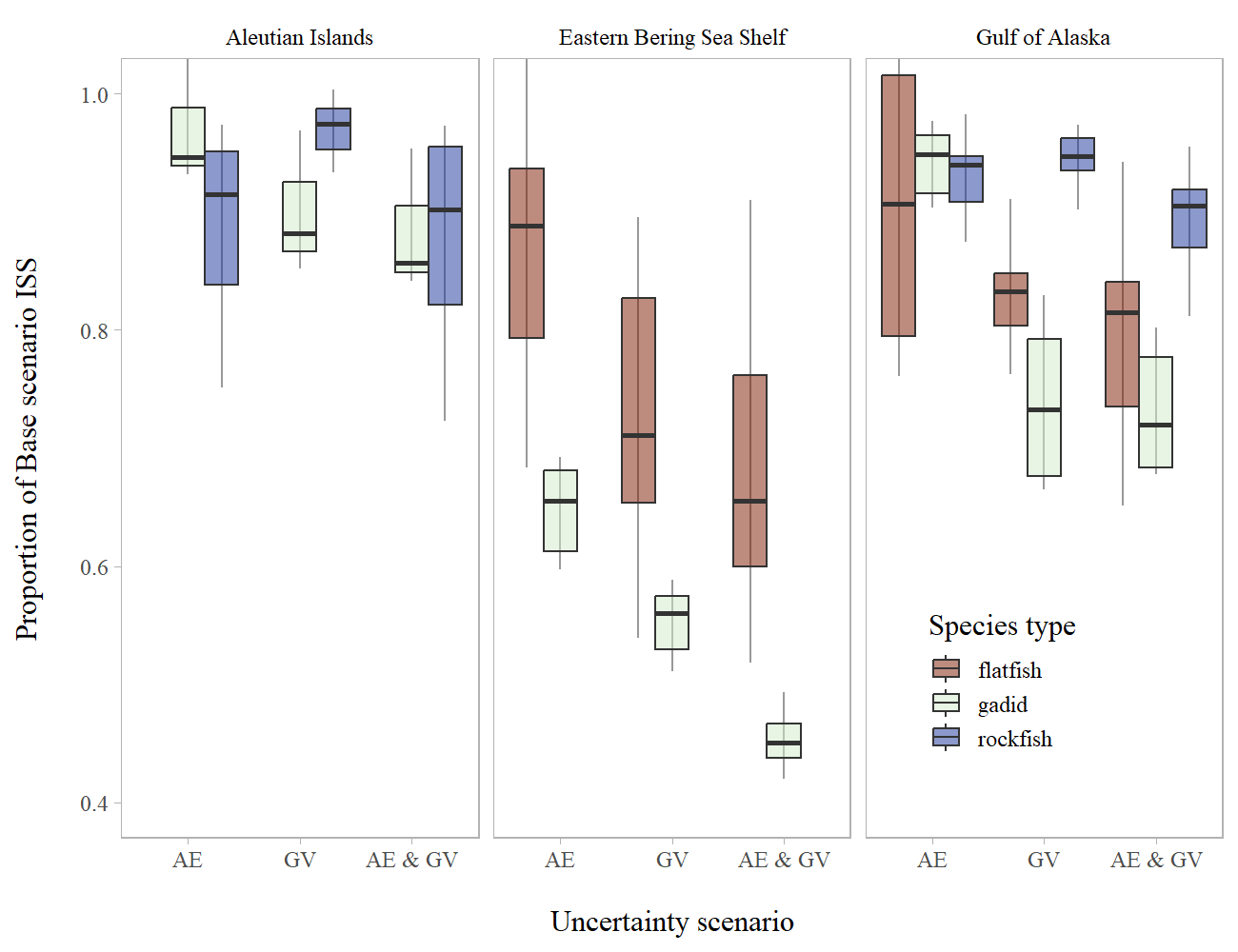


Figure 3: Age composition relative input sample size by species type across sex categories and uncertainty scenarios.

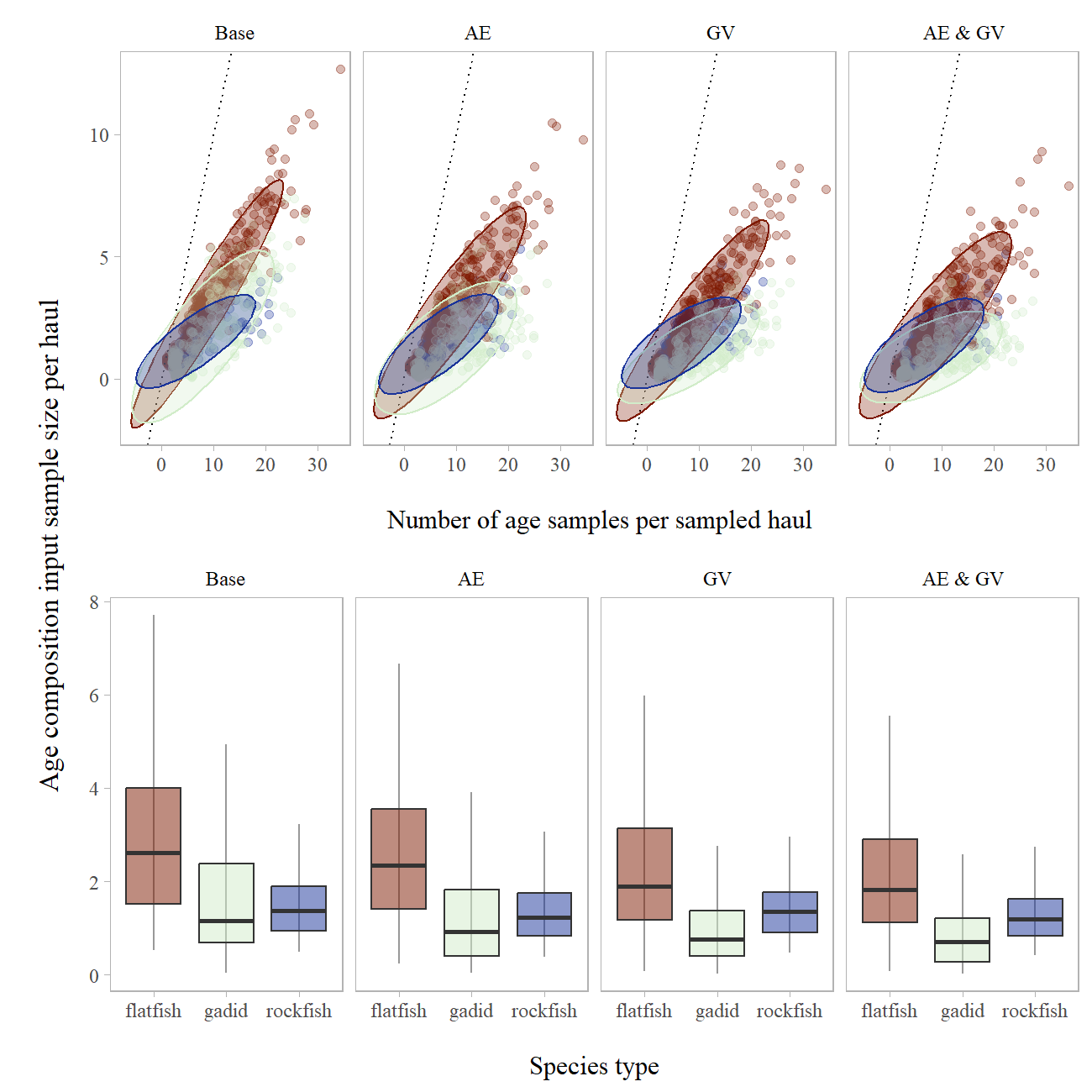


Figure 4: Age composition input sample size per sampled haul compared to number ages sampled within a haul (top panel) across uncertainty scenarios and species types (bottom panel).

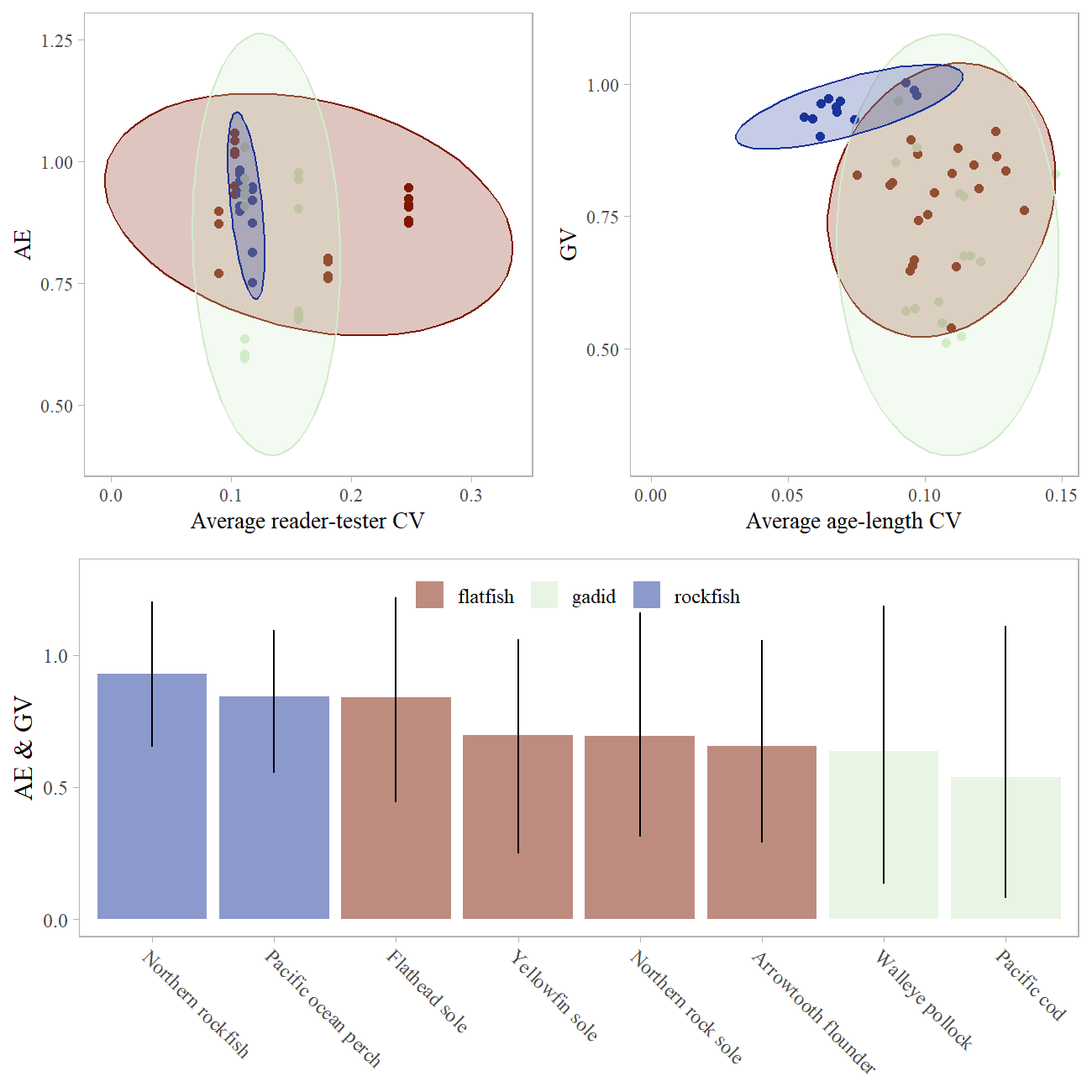


Figure 5: Age composition relative input sample size compared to age and growth statistics (top panel) and across stocks evaluated (bottom panel).