Effort optimization of bottom trawl subsampling of fish by sex and size: tradeoffs between stock assessment effective sample size and survey workforce health and efficiency

Pete Hulson1,\*, Benjamin Williams1, Meaghan Bryan2, Jason Conner3, Matthew Siskey4, William Stockhausen2, Susanne Mcdermott3, and Chris Long3

1 Auke Bay Laboratories, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 17109 Point Lena Loop Rd., Juneau, AK 99801  
2 Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA 98115  
3 Resource Assessment and Conservation Engineering Division, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA 98115  
4 School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

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

# Abstract

[Pete]

# Introduction

Many integrated fishery stock assessments rely on estimates of fish population length composition either directly (Punt et al. 2013) or as a derived quantity [e.g. age composition; Maunder and Piner (2015)]. The most reliable length composition estimates are based on observations from fishery-independent surveys because these platforms generally avoid the sample selection bias inherent in directed commercial fisheries (Council 1998). Past analyses have focused on the statistical efficiency of composition estimates for data weighting and applications to likelihood functions (e.g., Thorson 2014, Francis 2017, Thorson et al. 2017). It is recognized in sampling theory that increasing the number of samples increases the precision of the estimate, and from this perspective, one should require the maximum number of samples per each survey observation. However, to increase the precision of age or length composition estimates it’s not just the magnitude of sampling that is influential, but also the manner in which the samples are collected (Siskey et al. In Review). Further, each fish sampled on a survey has both an immediate physical labor cost (e.g. at-sea time management issues) and a cumulative health cost (e.g. repetitive stress injuries).

The National Oceanic and Atmospheric Administration’s (NOAA’s) Alaska Fisheries Science Center (AFSC) has conducted effort-standardized bottom trawl surveys in the eastern Bering Sea (EBS), Gulf of Alaska (GOA) and Aleutian Islands (AI) since 1982 (Stauffer 2004). These resource surveys comprehensively catalog the biota encountered at each sampling station. Observations of sex-specific size distributions for each encountered species are collected by: (1) sorting the trawl sample by species, (2) weighing each species in aggregate, (3) obtaining a random subsample of target sample size, (4) sorting the length subsample by sex (each fish is cut with a scalpel, gonads identified and placed in a sex-specific receptacle), and (5) measuring and recording each fish length (each fish is placed on a measuring board, length is identified and the species, sex and measurement are recorded into a computer). This length measurement process is repeated over 100,000 times in a given survey each year, representing a daunting amount of work for 6 field scientists per research vessel. Each year, this work flow results in acute and repetitive stress injuries, some requiring medical interventions and claims to the U.S. Office of Workers’ Compensation Programs. Another consequence of the intensity of this work are unrecoverable errors in the observed data; as fatigue or injuries accumulate during the course of field work, so do data collection errors (e.g. failure to properly encode length measurements with the correct species or sex or incorrect identification of sex), despite extensive *in situ* quality assurance protocols.

At the AFSC, length frequency sampling from the bottom trawl surveys is used in stock assessment models in a variety of ways to inform estimates of population abundance that are subsequently used to set management quantities. The most common use of length frequency sampling is to derive estimates of the population abundance at length that are then used in an age-length key to estimate population estimates at age and the model is fit to these age composition estimates Monnahan et al. (2021). Length frequency samples are also used in many flatfish assessments in a conditional-age-at-length framework (e.g., Rudd et al. 2021) that both fits the length composition and enables estimation of growth internally to the assessment (e.g., Turnock et al. 2017). In some cases, where age data is not available, length frequency samples which have been expanded to population abundance at length estimates are used directly as composition data within the assessment (e.g., McGilliard et al. 2019). Finally, recent developments have included using length frequency samples in a model-based framework to estimate length and age composition estimates (Thorson and Haltuch 2019, Ianelli et al. 2021, Thompson et al. 2021).

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

In December of 2021 an AFSC working group was formed (the ‘Survey Workload Optimization’ working group, SWO) to evaluate the impacts of reducing sampling for sex within the length frequency data collections due to repetitive motion injuries that have been occuring on the bottom trawl surveys. From the stock assessment perspective, the primary concern when reducing the sampling for sex within the length frequency data collections was the impact on the uncertainty in subsequent sex-specific length composition data. This is of particular importance to stock assessments that are sex-specific, such as the flatfish assessments conducted by AFSC. To that end, using effective sample size as the primary statistic to evaluate uncertainty in the length composition data that is derived from the bottom trawl surveys, we evaluated the impact of reducing sampling for sex within the length frequency data collection to answer three questions: (1) what is the impact of reducing sampling for sex on the uncertainty in the sex-specific length composition? (2) Is there a point of diminishing returns as we increase the number of lengths that are sampled for sex? And, (3) is there an acceptable level of increase in uncertainty in sex-specific length composition data due to subsampling for sex?

# Methods

## Bottom trawl survey data

[Paras on survey background, sampling for length and ages, and sample sizes - Jason and Meaghan]

## Computing length and age composition

[Para on expansion methods, maybe refer to tech memos or include appendix - Pete and Matt]

## Simulation-Bootstrap framework

[Included what we had from tech memo to build on - Ben and Pete]

To evaluate the effect of sub-sampling length frequency collections for which sex is subsequently determined we developed a bootstrap-simulation framework that allowed for reductions in the number of sexed length frequencies collected. We used the historical length frequency data that were collected from the bottom trawl surveys to evaluate the impact of reduced sex-specific length frequency data. In simple terms, the simulation framework that we developed would select a pre-determined number of fish from the length frequency collections that would then be subsequently sexed, the remaining length frequency data (regardless of whether sex was actually determined in the historical data) was classified as ‘unsexed.’

The bootstrap-simulation framework is composed of a suite of nested resampling protocols. Bootstrap resampling was performed either with replacement (wr) or without replacement (wor) depending on the needs of a particular protocol. Functions to run the sampling protocols were developed in a compartmentalized manner to provide for substantial flexibility in exploring desired resampling protocols. The order of operations (Figure 1) has the following schedule, with steps 1-3 being optional switches:

1. Resample hauls (wr) from the set of hauls with associated catch per unit effort (in numbers)
2. Within the resampled hauls from step 1, resample the observed lengths (wr)
3. From the resampled lengths in step 2, subset the lengths (wor) with observed sex (either male or female) and sample these sex-length pairs at the sub-sampling level determined in step 2; equation (1)
4. Calculate sex-specific population abundance at length, using equations (??) - (??)

The core of the bootstrap-simulation function (step 3 above) is designed to explore reductions in the sample size of lengths that are then sexed on a per haul basis. In this bootstrap-simulation, the number of lengths in a given haul must be less than or equal to the desired sample size *x* determined in step 2. In step 3, when the number of resampled lengths from step 2 in a haul is less than *x* then is used directly, if then a random draw of lengths is taken without replacement in step 3.

The bootstrap-simulation then repeated steps 1-3 iteratively for each sex sub-sample size determined in step 1, providing iterated sex-specific population abundance at length that was then compared to the historical sex-specific population abundance at length determined by the bottom trawl surveys.

We applied the bootstrap-simulation to species that were most commonly captured in the EBS, AI, and GOA bottom trawl surveys (Tables 1 - ??). The sub-sample levels that we evaluated for subsequent sex determination from the length frequency collections were 50, 75, 100, 125, 150, and 175 samples. We also ran the bootstrap-simulation for the historical number of sexed length frequency collections without subsetting in order to compare the base level uncertainty to the increase in uncertainty gained through sub-sampling. We ran the bootstrap-simulation for 500 iterations, which was a level for which the variability in population abundance at length results had stabilized, and applied the bootstrap-simulation to the most recent 3 years of the respective bottom trawl surveys, which were the most indicative of the current sampling levels. The bootstrap-simulation was developed in R (R Core Team 2022) and is available via GitHub as an R package (<https://github.com/BenWilliams-NOAA/swo>).

## Computing effective and input sample size

[Including what we have in the tech memo to build on - Pete and Matt]

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

where is the estimated proportion for category-*c* (which can be either age or length or any other arbitrary category across which proportions are computed) and is the observed proportion.

In this bootstrap-simulation the underlying length composition derived from the historical bottom trawl surveys was treated as the observed proportions in equation (2). For each iteration of the bootstrap simulation for a determined sex sub-sample size we computed a sex-specific estimated proportion () that was then compared to the underlying historical sex-specific length composition (the effective sample size for the total length composition, as the sum of population abundance at length, was also computed). Thus, across each iteration of the bootstrap simulation we computed an effective sample size that indicated the amount of increased uncertainty that was caused by sub-sampling sexed length frequency data. To summarize effective sample size across iterations we used the harmonic mean, which has been shown to reduce bias in recovering the true sample size in simulations for a multinomial distribution. Due to this reduction in bias the harmonic mean has also been recommended to determine the ‘input sample size’ that is used in stock assessment models to fit compositional data (Stewart and Hamel 2014). Herein, when we use the term ‘effective sample size’ we are referring to the effective sample sizes that were computed for each iteration of the bootstrap-simulation, when we use the term ‘input sample size’ we are referring to the harmonic mean of the iterated effective sample sizes.

# Results

[length results for selected species, including samples saved, ess and iss reductions - Pete and Ben/Meaghan]

[age results, including ess and iss reductions - Pete and Ben/Meaghan]

# Discussion

[Summary para of main results]

[Cost-benefit of precision compared to survey injuries]

[Still thinking on other paras]

# Acknowledgments

We thank *reviewer1* and *reviewer2* for their helpful reviews of this manuscript.

# Citations

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.

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

Ianelli, J.I., Fissel, B., Stienessen, S., Honkalehto, T., Siddon, E., and Allen-Akselrud, C. 2021. Assessment of the walleye pollock stock in the eastern Bering Sea. In: Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. Anchorage, AK, North Pacific Fishery Management Council.

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.

McGilliard, C.R., Palsson, W., Havron, A., and Zador, S. 2019. Assessment of the Deepwater Flatfish stock complex in the Gulf of Alaska. In: Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. Anchorage, AK, North Pacific Fishery Management Council.

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. Anchorage, AK, North Pacific Fishery Management Council.

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

Punt, A.E., Huang, T., and Maunder, M.N. 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES Journal of Marine Science 70(1): 16–33.

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/>.

Rudd, M.B., Cope, J.M., Wetzel, C.R., and Hastie, J. 2021. Catch and length models in the Stock Synthesis framework: expanded application to data-moderate stocks. Frontiers in Marine Science 8.

Siskey, M.R., Punt, A.E., Hulson, P.-J.F., Bryan, M.D., Ianelli, J.N., and Thorson, J.T. In Review. 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.

Spencer, P.D., and Ianelli, J.I. 2020. Assessment of the Pacific ocean perch stock in the Bering Sea/Aleutian. In: Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. Anchorage, AK, North Pacific Fishery Management Council.

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.

Thompson, G.G., Barbeaux, S., Connor, J., Fissel, B., Hurst, T., Laurel, B., O’Leary, C.A., Rogers, L., Shotwell, S.K., Siddon, E., Spies, I., Thorson, J.T., and Tyrell, A. 2021. 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 Isalands. Anchorage, AK, North Pacific Fishery Management Council.

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

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., 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.

Turnock, B.J., McGilliard, C.R., and Palsson, W. 2017. Assessment of the flathead sole stock in the Gulf of Alaska. In: Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. Anchorage, AK, North Pacific Fishery Management Council.

# Tables

[will determine which tables to include - kept this table in as place holder]

Table 1: Total length frequency samples from the most recent three Aleutian Islands surveys for the species evaluated in the bootstrap-simulation for reduction in sexed length-frequency collections.

| Species | 2014 | 2016 | 2018 | Avg |
| --- | --- | --- | --- | --- |
| arrowtooth flounder | 10,201 | 9,166 | 10,225 | 9,864 |
| Atka mackerel | 9,999 | 6,080 | 7,301 | 7,793 |
| flathead sole | 4,945 | 3,904 | 4,941 | 4,597 |
| northern rockfish | 14,738 | 15,116 | 14,640 | 14,831 |
| Pacific cod | 4,541 | 6,500 | 6,093 | 5,711 |
| Pacific ocean perch | 29,569 | 36,065 | 30,980 | 32,205 |
| shortspine thornyhead | 3,201 | 4,241 | 2,954 | 3,465 |
| walleye pollock | 10,300 | 9,112 | 12,972 | 10,795 |

# Figures

[Will figure out what figures to include - pun intended, left length flowchart here as we will have the total flowchart included]

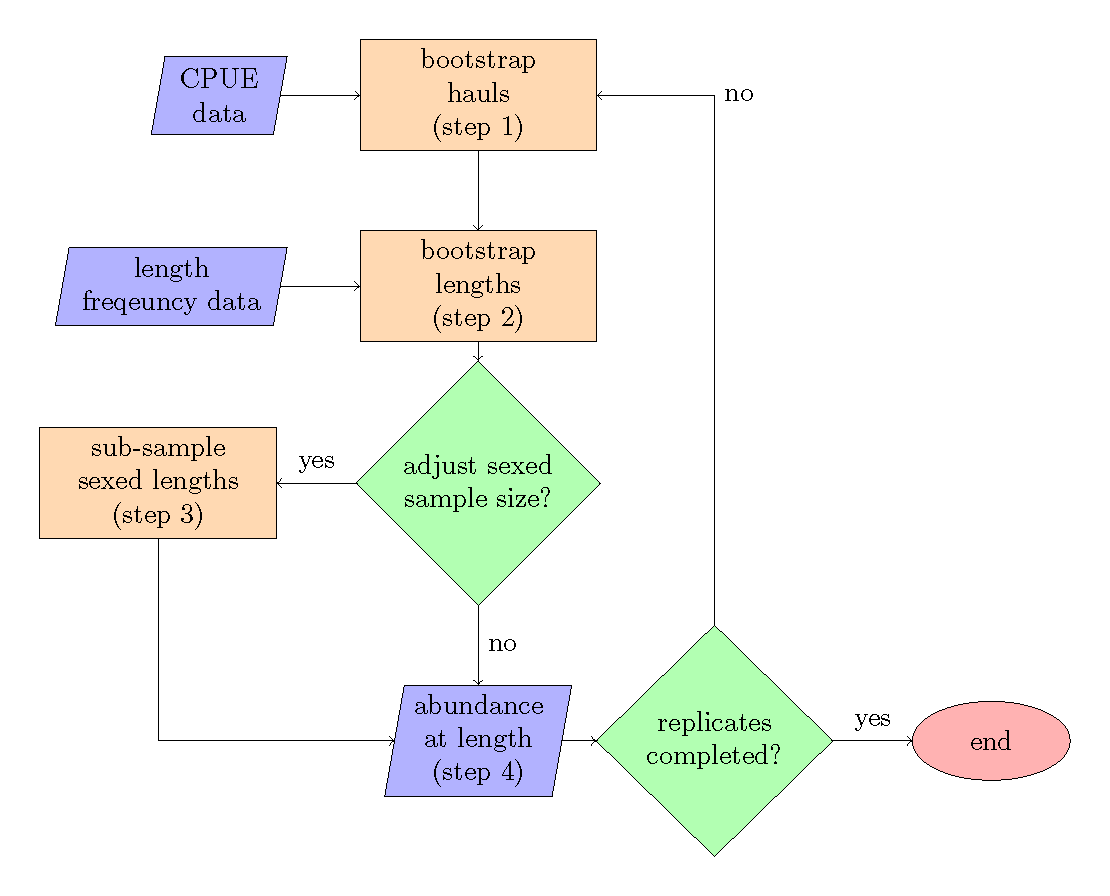


Figure 1: Bootstrap-simulation flow chart, the steps refer to the order of operations as described in the *Bootstrap-simulation framework* section.