Data Quality, Data Quantity, and its Effect on an Applied Stock Assessment of Cisco in Thunder Bay, Ontario

Nicholas C. Fisch[1*] and James R. Bence[1]

1. Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824-1101

*Corresponding author: nfisch@ufl.edu

Abstract

Stock assessments, or population models developed to support fishery management decisions, require informative data to produce reliable estimates. However, resources available to collect these data are limited. Thus, information relating the effects of different data collection schema on stock assessment performance should be of interest to fishery managers. We used an existing dataset on the Thunder Bay Cisco stock to simulate various degrees of reduction in available data. We considered both cluster sub-sampling of biological data from the commercial fishery harvest (which determine the observed harvest age-composition) and reductions in the frequency of hydroacoustic surveys, in order to examine their effect on fits of an age-structured stock assessment model for the Cisco stock. Our results indicate that reductions in the frequency of hydroacoustic surveys would have a greater effect on applied stock assessment performance for Thunder Bay Cisco than would reductions in biological sampling to randomly selected temporal clusters of the fishery harvest. Reduction in the frequency of the hydroacoustic survey resulted in different point estimates and larger estimated uncertainty for spawning biomass and M compared to the original assessment model. This was likely largely driven by increases in lag between the final year of the survey and the current year of the assessment. The lower influence of reduced biological sampling may be due to highly variable nature of Cisco recruitment, where large or “boom” year classes were still evident in the reduced biological samples, combined with information from survey age compositions. We suggest a priority be placed on performing hydroacoustic surveys with some regularity, such that when they are performed, they are done extensively to minimize uncertainty

(measurement error). The data subsampling approach used here could be used in many assessments to determine if a reduction in sampling of various types could be implemented without materially changing assessment results.

Introduction

Stock assessment models are important tools used in fisheries research and management. They generally use a variety of data sources from a given fish stock to develop a population model and subsequently estimate managerial and ecological quantities of interest such as spawning biomass. Where assessment models can differ in the amount and type of data used, they all require informative data on a stock of interest to produce accurate or reliable estimates (Magnusson and Hilborn 2007). Uncertainty and bias in stock assessments result from a variety of factors, including model structure and assumptions, but among these perhaps the most basal factor is the quantity and quality of data available for an assessment. Without informative data, the importance of model structure and assumptions is reduced. Management agencies possess, however, a finite amount of monies for data collection programs. Thus, there is a need to determine how to efficiently allocate resources used for data collection, such that sufficient data of each needed type are collected in a robust (statistically sound), practical, and/or efficient way.

Most stock assessments done in the United States are based on age-structured population assessment methods (Punt et al., 2017). When statistically fit, these models can be referred to as statistical catch-at-age assessment (SCAA) models, and are a form of integrated stock assessment. Such assessments rely on both indices of relative abundance (or less commonly estimates of absolute abundance), and information on the magnitude and composition of the harvest. While these data sources tend to inform on different parameters, there is overlap, and the exact influence of different types of data can be complex as this is influenced by model structure (Francis et al. 2011, Lee et al. 2014, Maunder and Piner 2015 & 2017). A very common data source utilized in age-structured stock assessments is the observed age composition of the fishery harvest. These data provide critical information within SCAAs on the relative strength of different cohorts, the fishery selectivity of a species, and the natural mortality rate (M) of a species (Lee et al., 2011; Maunder and Piner 2015). The observed age composition of the catch is generally estimated from samples of the fishery harvest and thus its accuracy depends on both the quantity and the quality of the samples. For example, as the number of samples increases, they approach a population census (in this case the population is the fishery harvest). Whereas the quality of a sample depends on how representative it is of the fishery harvest, which can depend on how different observations are spread out in time, how they are spread out by fishing trip, etc. The highest quality sample may be a truly random sample of the harvested population, or a stratified random sample, however this is nearly impossible to carry out in practice. In reality, we don’t have a final pool of the harvested population at the end of the fishing season that we can randomly sample. Instead, management agencies must determine which days, which ports, and which vessels to sample. We refer to the sampling of these nested groups of fish (select ports, select boats, etc.) as cluster sampling.

Due to the correlation among observations within clusters, in terms of space and time (i.e., characteristics of fish sampled from within clusters are not independent), a cluster sample is expected to contain less information on the biological characteristics, such as age composition, of the harvested population than would the same number of fish in a simple random sample from the entire harvested population. Some have referred to use of such data without accounting for the non-independence as pseudoreplication (Hurlbert 1984; Millar and Anderson, 2004, Murie et al., 2012). For a practical example, take port sampling, where a biologist or technician travels to a port to collect fish from the harvest and obtain information on their biological characteristics such as length or age. The port could be the only or primary landing location for the fishery or there could be many such locations. At the port on a given sampling date only a fraction of the boats that land fish at that port will be available for sampling, and often only a subset of those will be sampled. The individual boats sampled at a port on a given day are likely to have similar catch composition characteristics relative to the overall fishery catch (e.g., they fish close to the port, or closer on some days, or on the same schools of fish), resulting in observed fish characteristics that are correlated in space and time. Further, the catch composition for a specific fishing trip is likely to differ from other fishing trips landed at the same port on the same day, in ways that cannot be explained by simple random sampling of fish from a common statistical population (e.g., the specific locations fished by each vessel could have differences in age compositions).

Within assessment models, one can weight composition datasets according to their perceived quality using an effective sample size that is lower than the actual sample size (Maunder, 2011). This will ultimately affect model performance. In this study we were curious as to how cluster sampling of biological data from the fishery harvest not only affects this effective sample size of compositions (which can be calculated in numerous ways, see Francis 2011, Truesdell et al. 2017), but also how it subsequently affects stock assessment performance. Given the importance of age composition data to age-structured stock assessments, understanding how the quantity and the quality of biological samples from the fishery harvest (through its effect on the estimated observed age composition) ultimately affects stock assessment performance can provide useful direction to management agencies on how to allocate their biological sampling programs.

Indices of relative abundance or absolute estimates of abundance are critical to integrated assessments as they provide direct information on how abundance is changing over time (Francis 2011). Fishery-independent data have long been thought to be, and in some cases shown to be, important to stock assessment performance and accuracy (Chen et al., 2003; Magnusson and Hilborn, 2007; Ono et al., 2015). Fishery-independent indices of abundance can be of critical importance in stock assessments, to supplement often uninformative fishery-dependent indices of abundance used in assessment models which may not be proportional to actual stock abundance due to a variety of factors (Harley et al., 2001; Hilborn and Walters, 1992, Ono et al., 2015). Fishery-independent age composition data can also provide valuable information to assessment models in the sense that they can have a different selectivity than that of the fishery and are often able to catch smaller or younger fish, providing the model additional information on recruitment and M (Fisch et al. 2019). The downside is that fisheries-independent survey data are very expensive to collect, as contrary to fishery dependent sampling, fishery-independent surveys require additional monies for field sampling (boat time, man hours, etc.) to collect fish that would otherwise not be available. Thus determining how the frequency of fishery-independent surveys impacts stock assessment performance can provide useful information to management agencies.

In the Thunder Bay commercial Cisco (*Coregonus artedi*) fishery, the Ontario Ministry of Natural Resources and Forestry (OMNRF) samples the first 10 Cisco from each gillnet set in the fishery. This results in an extensive dataset containing biological information not only from each day that harvest occurs but at an even finer scale from each gillnet that catches fish. Although not a truly random sample of the harvested population, this is substantially more intensive and spread out sampling than is typical for most cluster sampling of biological data from a fishery's harvest. Additionally, since 2005 the Thunder Bay Cisco stock has been surveyed annually using hydroacoustic gear, to provide an estimate of spawning stock size. In 2018, a SCAA model was developed for Thunder Bay Cisco, which was informed by each of these data sources in addition to the aggregate harvest of the fishery (Fisch et al., 2019). This extensive dataset on biological samples from the commercial fishery, together with fishery-independent surveys of spawning abundance, offers a valuable opportunity to simulate both cluster sampling of biological data from the fishery and reductions in the frequency of hydroacoustic surveys, and to observe how the reductions influence the stock assessment results. We focus on cluster sampling as previous analyses indicated that simply reducing number of ages by simple random subsampling had little influence on the information content of the composition data for the Thunder Bay fishery (Fisch and Bence, 2018).

Data and its effect on stock assessment modeling is not a new subject, as many studies have examined the effect of different types and amounts of data on assessment model performance (Chen et al., 2003; He et al., 2016; Hulson et al., 2017; Magnusson and Hilborn, 2007; Muradian et al., 2019; Ono et al., 2015; Wetzel and Punt 2011). These studies have focused on the effect of leaving entire data sources out (Chen et al., 2003; Magnusson and Hilborn, 2007; Muradian et al., 2019), collecting certain data sources less frequently (e.g., every other year, second half of fishing history; Ono et al., 2015; He e al., 2016), or the amount of data collected in a given year (Ono et al., 2015; He et al., 2016; Hulson et al., 2017; Wetzel and Punt, 2011). Fewer studies have directly examined the effect of both the amount of data collected and specifically how they were collected in relation to assessment model performance.

In this study, we compare the performance of an applied stock assessment on Thunder Bay Cisco under different data collection scenarios. Our objectives for this analysis were twofold: 1) determine how cluster sampling of biological data from the fishery, through its effect on the observed harvest age composition, affects stock assessment performance, and 2) determine how the frequency of hydroacoustic surveys affect stock assessment performance. While focused on the Thunder Bay Cisco fishery, our results shed light on sampling strategies for other fisheries with some similar characteristics, and provide an example approach for evaluating how changes in sampling due to reductions in sampling effort could influence assessment results.

Methods

Thunder Bay Cisco

Cisco are a pelagic planktivore native to the Laurentian Great Lakes. They form annual spawning aggregations during the month of November in nearshore bays and areas of western Lake Superior, where contemporary spawning stocks are primarily located (Stockwell et al., 2009). In Thunder Bay (Management Areas 1-4, Figure 1 Fisch et al., 2019) the commercial Cisco fishery is largely a seasonal roe fishery, with most harvest occurring during the month of November using suspended gillnets (Ebener et al., 2008). Current management involves a limited entry fishery with aggregate quotas calculated as 10% of the estimated spawning biomass from hydroacoustic surveys.

Model

The original SCAA model developed in Fisch et al., (2019) is age- and sex-structured, beginning at age 2 and forming a plus group at age 15. The model runs from 1999 to 2015, to obtain estimates of quantities through the start of 2016. The SCAA is informed by four main sources of data; the total harvest, the age composition of the harvest, hydroacoustic surveys of spawning abundance, and the age composition of Cisco caught in additional gear deployed during the hydroacoustic surveys (mid-water trawls and multi-mesh gillnets; see Table 1 for specific years each data source was available for the original model). The model estimates M for males and females separately, treats hydroacoustic estimates of spawning stock size as absolute indices of abundance, and estimates recruitment through lognormal deviations about a median value (deviations are penalized in the likelihood). Variances of “abundance” data (i.e., hydroacoustic estimates) along with recruitment deviations were set relative to the variance of the harvest so that resulting variances for these data sources were compatible with prior expectations, consistent with recommendations from Francis (2011). The variance of the harvest was fixed at the median of its posterior distribution (-2.4 in log space) estimated in the original SCAA (Fisch et al., 2019), so as to be able to make comparisons across models. The model weights age composition data sources by iteratively reweighting effective sample sizes using method T3.4-TA1.8 of Francis (2011).

For more details on specifics of the assessment model, see Fisch et al. (2019). The only structural difference (other than fixing harvest variance) from the SCAA models used herein and the SCAA from Fisch et al., (2019) is the omission of aging error estimation within the current model, given Fisch et al., (2019) determined that (a) there was little aging error, and (b) the parameters determining aging error were very well determined.

Overview

Our overall approach was to fit the SCAA model to reduced datasets and assess model performance as the quantity and quality of data was reduced. Our main set of analyses consisted of fits to datasets for 15 dataset configurations. Herein, we use the term **dataset configuration** to denote a combination of data scenarios. We use the term **data scenario** to denote a data collection scheme for a specific type of data (e.g., sampling every other year of the hydroacoustic survey, or a version of cluster sampling biological data), and the term **dataset** to denote the actual data for a given data source used in a dataset configuration (this can change due to replicates, which differ from one another due to random data selection). The main dataset configurations included the full SCAA dataset (from Fisch et al., 2019), in addition to 14 reduced dataset configurations. We produced datasets for these 14 configurations by simulating cluster sampling of biological data from the fishery harvest in addition to leaving out select years of the hydroacoustic survey. Whenever the reduced dataset configuration involved a reduction in the number of fish providing biological data from the fishery harvest, this involved an element of randomness. Hence, for dataset configurations with cluster subsampling of the biological data, the SCAA model was fit to three replicate datasets with different random draws of the cluster subsampling, and hence different age composition data. There were three different age composition scenarios. There was the full (original) age composition and scenarios with reduced age composition datasets produced by simulating two different levels of cluster sampling of biological data from the fishery harvest. This fishery can be viewed as having its landings come via a single port, with the majority of the harvest occurring in November. Thus, we broke the November fishery into quarters and simulated cluster sampling by randomly selecting one or two quarters of the fishing season each year. For hydroacoustic data, we had the full (original) hydroacoustic dataset and produced four reduced hydroacoustic datasets by leaving out select years of the survey, resulting in a total of five main hydroacoustic scenarios. For the hydroacoustic scenarios there is no distinction between the scenario and dataset as no randomness was involved in producing the reduced datasets. Thus, we ultimately had a total of 15 different "main" dataset configurations after the three age composition and 5 hydroacoustic scenarios. There were a total of 35 SCAA fits to different datasets because the 10 configurations that involved cluster subsampling of the biological data were fit using each of the associated triplicate age compositions.

In addition to the main set of hydroacoustic scenarios, we considered three additional scenarios where the final year of the hydroacoustic survey data was always included in the dataset. These additional scenarios were intended to isolate the effect of sampling frequency from how recently a hydroacoustic survey had been conducted, and these scenarios were only paired with the full age composition (no cluster subsampling), adding three dataset configurations and datasets to which the SCAA model was fit. Thus, in total we considered 18 dataset configurations (15 "main" + 3 additional), and 38 model fits.

Cluster Sampling

In order to simulate cluster sampling of biological data from the fishery harvest, we split the number of days sampled in November each year into four parts of approximately equal duration, or quarters. For example, if harvest occurred (and thus, sampling) during every day in November in one year (30 days), then a quarter would last ~7 days for that year. Alternatively, if harvest occurred for only half of November in another year, then quarters would only span ~3 days each for that year. The total number of days was calculated from the first day sampled in November to the last day sampled in November (there could be days with no harvest in between). For one cluster sampling scenario, we randomly sampled two quarters in November each year and kept all the data collected within them. For another cluster sampling scenario, we randomly sampled only one quarter in November each year and kept all data collected within it. These two cluster sampling scenarios were termed Cluster Sample 2 (CS2) and Cluster Sample 1 (CS1), respectively. Our approach was to keep the data collected in the selected quarter or quarters, and remove the data collected outside of it. Thus, the random sampling is solely choosing which time periods (quarters) of data to keep, rather than the biological samples themselves within the quarters. For each cluster sampling scenario, in addition to removing data from non-sampled quarters (clusters), all data collected outside the month of November was omitted (not used in the model).

For each cluster sampling dataset, processing akin to the original processing for the SCAA was performed, i.e., aging data were pooled by management area each year and sex-specific age-length keys were developed to estimate the observed age composition each year. Since we are randomly removing fish from the biological database (through our simulation of cluster sampling), some of these fish will have been aged, thus we are reducing both the number of fish sampled and the number of fish aged each year. The number of fish sampled and aged in each year for each cluster sampling scenario (for the one replicate we present in the main text) can be found in Table 2. In many years the OMNRF aged a subsample of biological samples from the fishery harvest in order to develop an age-length key, aiming for 10 fish aged per 10mm length bin per management area per sex, while in some years all fish sampled were aged. Given our simulations of cluster sampling the fishery harvest randomly sample quarters (or clusters) in November each year, there is a possibility that results may be anomalous due to randomly picking particularly informative or conversely particularly uninformative samples. For this reason we produced three replicates of each cluster sampling scenario (using different random number seeds), resulting in six age-composition datasets from cluster sampling of biological data (3x CS2 & 3x CS1), and seven age composition datasets altogether including the full age composition.

Hydroacoustic Datasets

Reduced hydroacoustic datasets were produced with two questions in mind. One, what happens to assessment performance as we reduce the frequency of hydroacoustic surveys to every other year, every third year, every fourth, etc.? Two, how does assessment performance decline as the final year of the hydroacoustic survey moves further and further away from the current year of the assessment? For our main set of scenarios we produced four different reduced datasets based on the hydroacoustic survey (Table 3): sampling every other year ending in 2014, every third year ending in 2013, every fourth year ending in 2012, and every fifth year ending in 2010. These scenarios were termed AC1 – AC4, respectively. We also developed an alternative to AC4 that used data in the year 2011 instead of 2010. In an attempt to isolate the effect of the lag between the final year of the survey and the assessment year with the effect of the frequency of the survey, we developed three additional hydroacoustic datasets corresponding to scenarios that all included year 2015 of the survey. One sampled every other year, the other every third year, and the third every fifth year, termed HA2, HA3, and HA5 (Hydroacoustic-Alternate). The datasets for these additional hydroacoustic scenarios were only paired with the full composition dataset.

Given that fishery independent age composition data used in the model were calculated from biological sampling that occurred during the hydroacoustic survey, when a year of the hydroacoustic survey is removed, we also left out the fishery-independent age composition for that year. Specifics on how hydroacoustic data were collected and processed can be found in Fisch (2018).

Model Running

Models were first fit using penalized maximum likelihood to perform iterative reweighting of effective sample sizes (ESS) for age composition datasets (commercial fishery and fishery-independent age compositions) using method T3.4-TA1.8 of Francis (2011). Once effective sample sizes converged, Bayesian posteriors were generated with the ESSs fixed. MCMC chains were run for 20 million iterations, saving every 500th and burning in 2500 iterations from the final chain. Convergence was assessed based on chains of the model estimated parameters using Geweke's diagnostic at an alpha level of 0.01. Priors for model parameters can be found in Table 2 of Fisch et al., (2019).

Comparison

We compared the SCAA fit to the full Thunder Bay dataset to model fits to reduced datasets by examining changes in point estimates and estimated uncertainty for quantities such as spawning biomass and *M*. Our metrics of estimated uncertainty included 95% highest posterior density (HPD) intervals and CVs of posterior distributions. We calculated two different CV metrics for spawning biomass; the mean CV of the posterior distributions of spawning biomass over the full time series and the CV of the posterior distribution for spawning biomass in 2015. We chose to focus on spawning biomass as this value may be used to calculate quotas in the future and thus is of management interest, and *M* as this is a parameter of ecological interest. We compared estimates of spawning biomass in 2015 instead of 2016 because the 2016 spawning biomass estimate from the model does not include age 2s (model recruits), of which ~30% are generally mature in Thunder Bay. In addition, quotas are currently set based on hydroacoustic estimates of spawning stock size from the previous November, and this value most closely relates to the 2015 estimate. We believe results and conclusions would be similar had we used spawning biomass estimated in 2016.

Results

Results did not differ greatly across model fits to replicates of the cluster sampling scenarios, thus for simplicity we present results for a single replicate herein. Figures related to cluster sampling replicates are in the supplemental files. For one specific data configuration, namely AC1-CS1, the MCMC chains for the SCAA model would not converge on a stable distribution at its reweighted effective sample size for 1 out of the 3 datasets (replicates of cluster sampling scenarios).

Effective sample sizes for fishery age composition data decreased as the information content (i.e., number of fish sampled, aged, and the quality of the sample) was reduced through cluster sampling (Table 4). This result occurred across all hydroacoustic data scenarios. Effective samples sizes for fishery-independent compositions were variable, however, they generally decreased from the full hydroacoustic dataset as select years of fishery-independent compositions were removed.

Relative differences between point estimates of spawning biomass in 2015 for the original model and model subsets were variable (Table 4). Overall, the largest differences were attributed to reductions of hydroacoustic data rather than cluster sampling of biological data (Figures 1 & 2). Specifically, for the full hydroacoustic dataset, cluster sampling biological data from the fishery harvest did not change the point estimate of spawning biomass in 2015 by more than 2%. Similarly, for other hydroacoustic scenarios, the maximum change in spawning biomass estimates for 2015 was 7%. In contrast, large changes in estimates were attributable to the hydroacoustic scenario. For AC1 combined with each age composition scenario, point estimates of spawning biomass in 2015 were underestimated compared to the model fit to the full dataset by about 35%. For AC2, relative differences were modest once again, with no combination of AC2 and a given age composition scenario producing a difference greater than 4%. Combinations of AC3 and different age composition scenarios resulted in higher estimates of spawning biomass in 2015 compared to the fit to the full dataset, with the greatest difference being 25% (for AC3-CS1 model). Differences for model fits utilizing the AC4 dataset were again modest, but were the most variable within a hydroacoustic scenario (ranging from -4% to +3%), with the largest relative difference of -4% resulting from a combination with the full composition scenario (i.e., AC4-Full Comp). The model fit to the alternative AC4 scenario, which was only combined with the full age composition and used observed data in 2011 instead of 2010 from the hydroacoustic survey, produced significant differences in spawning biomass throughout the time series and in 2015 compared to the model fit to the full dataset (2015 RD = 200%, Supplemental Figure 3). In addition, this model resulted in substantial increases in estimated uncertainty for spawning biomass (in terms of 95% HPDs) compared to the model fit to the full dataset.

There did seem to be some interactive effect between the level of harvest age-composition sampling and the frequency of surveys on assessment model results during the earlier years covered by the assessment. In particular, while estimates of spawning biomass were similar in the final year for different age composition scenarios, estimates in the early years tended to vary more among the age composition scenarios for survey scenarios with less frequent sampling (Figure 1).

Estimated uncertainty as indicated by the width of 95% HPD intervals generally increased as models were fit to hydroacoustic survey from fewer years. This result was ubiquitous across cluster sampling scenarios (Figure 2). As the information content of the age composition decreased due to cluster sampling biological data from the harvest, estimated uncertainty increased marginally for most hydroacoustic datasets, although it actually decreased for the AC4 dataset (Figure 2 right panel). Another metric of estimated uncertainty, posterior distribution CVs, displayed similar results. Mean CVs for the posterior distributions of spawning biomass for the full time series increased as hydroacoustic survey became less frequent, and as the last year of survey data became further from the current assessment year (Figure 3, Table 4). Posterior CVs for spawning biomass, just for 2015, also increased as the hydroacoustic survey became less frequent and the last year of survey became further away from the current year of the assessment (Figure 3). Each CV metric (mean over time series and 2015 estimate) increased as the information content of the composition data decreased through cluster sampling of the biological data compared to the full biological dataset. However, the CV metrics did not generally increase from the CS2 to CS1 scenarios for the full acoustic dataset and the AC4 dataset (Table 4, Figure 3). The increase in CV metrics from the fits using the full biological dataset to those using cluster sampling datasets was not as pronounced as the increase in CV metrics as hydroacoustic survey frequency was reduced.

For the alternate hydroacoustic scenarios (HA2, HA3, HA5), which were attempts to isolate the effect of reductions in frequency of the survey and the effect of lag between the last year of survey and the current year of the assessment, estimated uncertainty in terms of 95% HPD intervals for spawning biomass in 2015 increased from the model fit to the full dataset to HA2 and further to HA3, then decreased from HA3 to HA5 (Figure 4 right panel). Point estimates of spawning biomass in 2015 varied with the largest difference (compared to the full model) attributed to the HA2 scenario. Mean CVs for full time series of spawning biomass were 0.28, 0.35, and 0.36, for the HA2, HA3, and HA5 scenarios, respectively. CVs for spawning biomass in the final year were 0.25, 0.29, and 0.33. For context, CV metrics (mean CV and CV in the final year) for spawning biomass for the model fit using the full dataset were 0.24 and 0.20.

Every model fit to reduced datasets estimated a higher *M* for males than for females, consistent with the fit to the full dataset. Point estimates of *M* were between 0.27-0.36 for males and 0.21-0.31 for females. These estimates for each sex varied little as the information content within the age composition data was reduced through cluster sampling (Figure 5). More variability in point estimates of *M* was attributed to reductions in the frequency of the hydroacoustic survey, with the largest difference compared to the fit to the full dataset occurring when fitting to the hydroacoustic scenario AC1. Estimated uncertainty in terms of the width of HPD intervals increased as hydroacoustic survey frequency was decreased and in most cases also increased as the information content of the composition data was decreased. Estimated uncertainty in terms of the CV of the posterior distribution of *M* generally increased as years of the hydroacoustic survey were removed (Figure 5). Results for fits for the same hydroacoustic scenario, as the information content within the age composition of the fishery harvest was reduced through cluster sampling, were more variable although most often the CV of each *M* estimate increased as the information content of the composition data was decreased (exceptions being AC2-CS2 female estimate, AC3-CS2 female estimate, and AC4-CS1 female estimate).

Discussion

Overall, the effect of reduced frequency of the hydroacoustic survey on SCAA performance was much greater than reduced biological sampling through cluster sampling the fishery harvest. This is not necessarily an unexpected result, given hydroacoustic estimates of spawning biomass are treated as absolute estimates of abundance in the assessment models. Thus including or excluding select data points of this survey will tend to "pull" the trend lines of spawning biomass up or down relative to the full assessment (based on which data points are left). However, what is surprising is how little reduced biological sampling of the fishery harvest affected assessment model performance. Although our simulations of cluster sampling biological data from the fishery harvest all produced substantial decreases in the perceived information content in the age composition datasets (based on re-weighted ESS), all of the models converged and produced similar point estimates in spawning biomass and M with only marginal increases in estimated uncertainty compared to models informed by the full age composition (conditional on hydroacoustic dataset). This may be a result of the highly variable, boom-or-bust recruitment patterns that have occurred in Cisco populations in western Lake Superior, where a large year class is produced followed by many years of effectively no recruitment (see Figure 6 in Fisch et al., 2019). For the Thunder Bay SCAA, there were 3-4 boom year classes evident over the 17 year time series. These boom year classes effectively drive the entire SCAA model in terms of selectivity, M, and recruitment (while the periodic abundance indices give an estimate of absolute scale). What may be occurring is that even if sampling is reduced or clustered (or both), given there are so few individuals in non-boom year classes, low information content age compositions may be sufficient. An additional reason why low information content in the harvest age composition data may be sufficient is because information on age compositions was also provided by the survey. Of course, some harvest composition data is necessary to fit a SCAA, as these data inform on model parameters (e.g., fishery selectivity) that other data sources provide little information. In addition, there is some threshold of ESS below which model failure will occur. By model failure we mean the model will not converge without making alterations such as fixing M, or just assuming that age compositions are more informative than they really are (artificially increasing ESS). Having to make such changes (i.e., questionable assumptions) would mean that the resulting assessment was much less reliable. Thus, it is important to maintain a biological sampling program for the harvest that results in a sufficient effective sample size for age composition data to produce a reliable assessment model.

This result, that some cluster sampling of biological data from the fishery harvest has a marginal effect on assessment performance, may not be widely generalizable across different species. Assessment model performance for a species with a different life history could be more affected by cluster sampling than Cisco. For example, He et al., (2016) found that rapidly growing species with clear signs of strong cohorts (referring to Bocaccio, *Sebastes paucispinis*), are likely to see less improvement in assessment results with increased age data than more slow growing species, or species for which recruitment is less variable. Our results are similar to those found in He et al., (2016) for Bocaccio, that decreases in the number of fish sampled from the commercial fishery have little effect on stock assessment performance, and quite likely for the same reasons as Cisco are also a fast growing species that show clear signs of strong cohorts. Similarly, Wetzel and Punt (2011) found that where the inclusion of length composition data dramatically improved assessment performance, increasing the amount of composition data only resulted in minimal improvements in performance. They went on to imply that their results may be different than those of a slow-growing, long-lived fish such as rockfish (*Sebastes* spp.). It may be that for species without as variable a recruitment pattern as western Lake Superior Cisco or slower growth, cluster sampling may indeed have a significant effect on assessment model performance and should be avoided or minimized if possible. Although as a counter example, Hulson et al., (2017) found that age composition sample size had a greater impact on SCAA uncertainty for species with higher recruitment variability compared to those with lower recruitment variability. In addition, Ono et al., (2015) found that estimation performance for species examined across three life history types (cod-, flatfish-, and sardine-like species) did not qualitatively differ from the base case when the sample size of age and length composition data was reduced throughout the entire time series. Contradictions in results may be a function of the different magnitudes of recruitment variabilities in the different studies. Sardine (fast-growing species) in Ono et al., (2015) and Walleye Pollock (*Gadus chalcogrammus*) in Hulson et al., (2017) did not have as large a recruitment SD as those of Bocaccio in He et al., (2016) or Cisco in our study (σ_r; Sardine = 0.73, Walleye Pollock = 0.70, Boccacio = 1, Cisco full model estimate = 4.5). Results and conclusions may also differ based the spatial scale or the timing of the fishery. Compared to many marine fisheries, the spatial scale of the Thunder Bay Cisco fishery is small, and it is effectively a one-month seasonal fishery. Cluster sampling may have a greater effect when the area of a fishery is larger, and for fisheries spread out over a larger time period.

As far as data collection decisions, in western Lake Superior it seems that priority should be given to collection of fisheries-independent hydroacoustic surveys for Cisco, even if this necessitated some reduction in biological sampling of the fishery. Previous studies noted the importance of fisheries-independent data to stock assessment performance (Chen et al., 2003; Ono et al., 2015; Wetzel and Punt, 2011), and our results support this view. Given we also removed select years of survey age composition data along with the hydroacoustic data (in accordance with how it is collected in Thunder Bay), we cannot be certain that effects on assessment performance are more driven by one or the other. It could be that the survey age composition data are providing critical information on recruitment and *M* to supplement the fishery dependent age composition data, and thus its reduction results in poor assessment performance. We find it more plausible that the index of abundance is a larger driver of assessment performance than the survey age composition data because of its treatment within the model as an absolute abundance estimate. The general lack of sensitivity to some reduction in biological data from the harvest through cluster subsampling is encouraging given that the coverage in both time and space is less complete (than it was for Cisco in Thunder Bay) for most other stocks of Cisco and for other species of commercially harvested fish with highly variable recruitment.

Given the importance of fishery-independent surveys to assessment performance, how often should surveys be done? In our study as both the frequency of the hydroacoustic survey was reduced and the lag between the final year of the survey and the year of the assessment increased (and years of survey age composition dropped), the performance of the assessment model decreased. When we attempted to separate these two factors using the so-called alternate hydroacoustic scenarios (HA2, HA3, & HA5), the lag between the final year of the survey and the year of the assessment seemed to have a greater effect on SCAA performance than the frequency of the survey. However, if the assessment is done annually, it is unavoidable that if the frequency of the hydroacoustic survey is reduced there will be years where the last year of the survey is not the current year of the assessment. We did see some substantial idiosyncratic changes in assessment results related to the precise survey years used rather than the frequency. This highlights the role of chance or random variation in assessment results, which could pose a danger when conducting the hydroacoustic survey too infrequently. For example, our AC4 model that included hydroacoustic data in 2005 and 2011 (instead of 2010) produced substantially greater differences in spawning biomass compared to the full model than did the original AC4 model. The hydroacoustic data in 2011 was effectively ignored in the full model fit (see Figure 9 in Fisch et al., 2019), possibly due to high measurement error. With more years of data, the effect of outlier data points is reduced. Thus, if survey frequency is reduced, decreasing measurement error within years when the survey is done should become a priority. Overall, it is likely a reduction in hydroacoustic survey frequency to every other year or every third year would not have a great effect on Thunder Bay Cisco stock assessment performance. We recommend that hydroacoustic surveys be done with at least some regularity such that the assessment model has periodic data on absolute abundance to scale the population. Future studies could involve a simulation framework to assess the tradeoff in assessment performance of many surveys with a relatively high CV vs fewer surveys with a lower CV.

An important aspect to note regarding our study is that our metrics of estimated uncertainty are almost surely underestimating the true uncertainty. Confidence intervals can underestimate true uncertainty due to constraints imposed by model structure and assumptions (He et al., 2016; MacCall, 2013; Mangel et al., 2013), e.g., by fixing a key parameter. For example, in our study we treat the hydroacoustic survey as an absolute estimate of abundance, or a survey with a catchability fixed at 1. This greatly limits the model from an uncertainty perspective as it does not have a chance to explore the uncertainty in that parameter itself, which is likely large. In this study we solely manipulated the data going into the model and the weights for the age composition datasets (ESS). Thus, the model structure and assumptions remained intact such that any changes in model performance would be attributed solely as a function of the data input. We did observe increases in estimated uncertainty with more data for some of our models (specifically AC4 models), highlighting that model estimated uncertainty may not be the best metric in assessing assessment performance. This result is consistent with the general observation that estimated uncertainty can increase with more data, if those data tell a contradictory story about the targeted fishery (Schnute and Hilborn, 1993; Schnute and Richards, 2001). A better gauge in our study regarding the robustness of an assessment in the face of a reduction in data availability is how similar the replicate results were to one another and the full assessment. We believe that a similar approach of examining how stock assessment results change as data are left out of the assessment can be generally instructive regarding potential changes in sampling designs.

Finally, as a disclaimer we do not know how reliable estimates were for the stock assessment model fit to the full data set. We chose to analyze data from a real stock as opposed to generating it based on a simulated one so as to incorporate more realism into the study. The downside of this is that we do not know the true values for the system. While it is perhaps logically possible that the original assessment was highly biased or uncertain, we still think it was reasonable to focus on metrics such as point estimates and estimated uncertainty compared to the full, original Thunder Bay SCAA, as we believe those were likely the most reliable estimates among the fits we considered.

Acknowledgments

We acknowledge the Great Lakes Fishery Commission for funding this project, and thank the OMNRF for permitting use of their dataset on Thunder Bay Cisco and working with us to interpret these data. This work was supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research, the Michigan DNR though its support of the Partnership for Ecosystem Management at MSU, and the Quantitative Fisheries Center at MSU. This is QFC publication 20XX-YY.

References

Chen, Y., Chen, L., and Stergiou, K. I. 2003. Impacts of data quantity on fisheries stock assessment. Aquatic Sciences, 65(1), 92-98.

Ebener, M. P., J. D. Stockwell, D. L. Yule, O. T. Gorman, T. R. Hrabik, R. E. Kinnunen, W. P. Mattes, J. K. Oyadomari, D. R. Schreiner, and S. Geving. 2008. Status of cisco (*Coregonus artedi*) in Lake Superior during 1970-2006 and management and research considerations. Ann Arbor, Michigan: Great Lakes Fishery Commission, Lake Superior Technical Report 1.

Fisch, N. C. 2018. On the Assessment and Sustainability of Thunder Bay Cisco (Masters Thesis, Michigan State University).

Fisch, N. C., and J. R. Bence. 2018. Sampling Intensity, Aging Intensity, and its Effect on an Age-Structured Assessment of Cisco in Thunder Bay. Quantitative Fisheries Center Technical Report T2018-02

Fisch, N. C., J. R. Bence, J. T. Myers, E. K. Berglund, and D.L. Yule. 2019. A comparison of age-and size-structured assessment models applied to a stock of cisco in Thunder Bay, Ontario. Fisheries Research, 209, pp.86-100.

Francis, R. C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68(6), 1124-1138.

Harley, S. J., R. A. Myers, and A. Dunn. 2001. Is catch-per-unit-effort proportional to abundance?. Canadian Journal of Fisheries and Aquatic Sciences, 58(9), 1760-1772.

Hilborn, R., and C. J. Walters. (1992). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Reviews in Fish Biology and Fisheries, 2(2), 177-178.

He, X., J. C. Field, D. E. Pearson, and L. S. Lefebvre. 2016. Age sample sizes and their effects on growth estimation and stock assessment outputs: three case studies from US West Coast fisheries. Fisheries Research, 180, 92-102.

Hulson, P. J. F., D. H. Hanselman, and S. K. Shotwell. 2017. Investigations into the distribution of sample sizes for determining age composition of multiple species. Fishery Bulletin, 115(3).

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological monographs, 54(2), 187-211.

Lee, H. H., M. N. Maunder, K. R. Piner, and R. D. Methot. 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fisheries Research, 109(1), 89-94.

Lee, H. H., K. R. Piner, R. D. Methot Jr, and M. N. Maunder. 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: an example using blue marlin in the Pacific Ocean. Fisheries Research, 158, 138-146.

MacCall, A.D., 2013. Use of the delta method to evaluate the precision of assessments that fix parameter values. Fisheries Research. 142, 56–60.

Magnusson, A., and R. Hilborn. 2007. What makes fisheries data informative?. Fish and Fisheries, 8(4), 337-358.

Mangel, M., A. D. MacCall, J. Brodziak, E. J. Dick, R. E. Forrest, R. Pourzand, S. Ralston. 2013. A perspective on steepness, reference points, and stock assessment. Canadian Journal of Fisheries and Aquatic Sciences. 70, 930–940.

Maunder, M. N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. Fisheries Research, 109(2-3), 311-319.

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

Maunder, M. N., and K. R. Piner. 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research, 192, 16-27.

Millar, R. B., and M. J. Anderson. 2004. Remedies for pseudoreplication. Fisheries Research, 70(2-3), 397-407

Muradian, M. L., T. A. Branch, and A. E. Punt. 2019. A framework for assessing which sampling programmes provide the best trade-off between accuracy and cost of data in stock assessments. ICES Journal of Marine Science.

Murie D. J., D. C. Parkyn, and S. L. Diamond. 2012. Sampling Commercial Fisheries. Pages 921-956 *in* A. V. Zale, D. L. Parrish, and T. M. Sutton, editors. Fisheries Techniques. American Fisheries Society, Bethesda, Maryland.

Ono, K., R. Licandeo, M. L. Muradian, C. J. Cunningham, S. C. Anderson, F. Hurtado-Ferro, K. F. Johnson, C. R. McGilliard, C. C. Monnahan, C. S. Szuwalski, J. L. Valero, K. A. Vert-Pre, A. R. Whitten, and A. E. Punt. 2014. The importance of length and age composition data in statistical age-structured models for marine species. ICES Journal of Marine Science, 72(1), 31-43.

Punt, A. E., C. A. Akselrud, and L. Cronin-Fine. 2017. The effects of applying mis-specified age-and size-structured models. Fisheries research, 188, 58-73.

Schnute, J. T., and R. Hilborn. 1993. Analysis of contradictory data sources in fish stock assessment. Canadian Journal of Fisheries and Aquatic Sciences, 50(9), 1916-1923.

Schnute, J. T., and L. J. Richards. 2001. Use and abuse of fishery models. Canadian Journal of Fisheries and Aquatic Sciences, 58(1), 10-17.

Stockwell, J. D., M. P. Ebener, J. A. Black, O. T. Gorman, T. R. Hrabik, R. E. Kinnunen, W. P. Mattes, J. K. Oyadomari, S. T. Schram, D. R. Schreiner, M. J. Seider, S. P. Sitar, and D. L. Yule. 2009. A synthesis of cisco recovery in Lake Superior: implicationsfor native fish rehabilitation in the Laurentian Great Lakes. North American Journal Fisheries Management. 29, 626–652.

Truesdell, S.B., J. R. Bence, J. M. Syslo, and M. Ebener. 2017. Estimating multinomial effective sample size in catch-at-age and catch-at-size models. Fisheries Research 192: 66-83.

Wetzel, C. R., and A. E. Punt. 2011. Performance of a fisheries catch-at-age model (Stock Synthesis) in data-limited situations. *Marine and Freshwater Research*, *62*(8), 927-936.

Tables

Table 1. Data source years for the original Thunder Bay assessment. Reproduced from Fisch (2018).

Year	Hydroacoustic Survey	Fishery Harvest	Fishery Age Composition	MWT Survey Age Composition	Gillnet Survey Age Composition
1999		X	X		
2000		X	X		
2001		X	X		
2002		X	X		
2003		X	X		
2004		X	X		
2005	X	X	X	X	
2006		X	X		
2007	X	X	X	X	
2008	X	X	X	X	
2009	X	X	X	X	X
2010	X	X	X	X	
2011	X	X	X		
2012	X	X	X		
2013	X	X	X		X
2014	X	X	X		X
2015	X	X	X	X	X

Table 2. Description of sampling of the commercial Cisco fishery for the original dataset and for each cluster sampling scenario (for the selected replicate for which results are presented in the main text). Subsampled column (Column 2) identifies which years utilized an age-length key in the original dataset (i.e., were ages subsampled or was the full sample aged?). Column 3 displays the number of Cisco sampled and aged in the original OMNRF database. Columns 4 and 5 of the table denote the cluster sampling scenarios described in methods. The first number in columns 3-5 represents the number of Cisco sampled and the second represents the number aged (Number sampled – Number Aged).

Year	Subsampled	Full Biological Dataset	Cluster Sample 2	Cluster Sample 1
1999	Yes	860 – 402	432 – 196	80 – 28
2000	Yes	3241 – 594	327 – 222	169 – 101
2001	Yes	1221 – 574	305 – 140	207 – 97
2002	Yes	1147 – 676	336 – 201	120 – 64
2003	Yes	1208 – 704	361 – 200	161 – 99
2004	Yes	1091 – 527	393 – 199	247 – 136
2005	Yes	661 – 280	220 – 79	56 – 19
2006	Yes	644 – 378	157 – 103	37 – 28
2007	Yes	839 – 330	248 – 105	107 – 45
2008	No	654 – 654	220 – 220	146 – 146
2009	No	638 – 637	190 – 190	60 – 60
2010	Yes	500 – 219	299 – 124	171 – 72
2011	No	563 – 562	140 – 140	100 – 100
2012	No	478 – 477	100 – 100	79 –79
2013	No	429 – 427	159 – 157	120 – 120
2014	Yes	733 – 517	342 – 208	230 – 149
2015	Yes	705 – 457	301 – 184	216 – 106

Table 3. Hydroacoustic dataset scenarios. Hydroacoustic-Alternate (HA) scenarios were only paired with the full composition dataset to explore the effect of lag between final data year and current assessment year. An X denotes a year that contained data from the hydroacoustic survey.

Year	Full Model	AC1	AC2	AC3	AC4	HA2	HA3	HA5
2005	X	X	X	X	X	X	X	X
2006								
2007	X		X			X		
2008	X	X		X				
2009	X					X	X	
2010	X	X	X		X			X
2011	X					X		
2012	X	X		X			X	
2013	X		X			X		
2014	X	X						
2015	X					X	X	X

Table 4. Various statistics for each assessment model fit to datasets arising from the main set of configurations and the selected replicate age composition in the study. Columns 3-5 depict the effective sample size for each age composition dataset used in each fit of the assessment model (for the selected fishery age composition replicate). Column 6 depicts relative differences (RD) of the point estimate of spawning biomass in 2015 for the fit using the full dataset compared to estimates for fits to the reduced datasets: (fit to data subset-fit to full dataset)/fit to full dataset. Columns 7 and 8 show the mean CV of the posterior distributions of spawning biomass over the time series and the CV of the posterior distribution of spawning biomass in the year 2015, respectively.

Model		Fishery ESS	MWT ESS	MMGN ESS	2015 SB RD	Mean SB CV	2015 SB CV
Full AC	Full Comp	64	45	53	NA	0.24	0.20
	CS2	23	45	46	-1%	0.25	0.20
	CS1	14	40	34	-2%	0.25	0.20
AC1	Full Comp	62	29	11	-35%	0.34	0.32
	CS2	24	19	7	-37%	0.37	0.34
	CS1	15	26	11	-34%	0.37	0.36
AC2	Full Comp	62	35	70	-4%	0.35	0.40
	CS2	24	22	50	-4%	0.35	0.42
	CS1	15	29	58	-3%	0.37	0.44
AC3	Full Comp	60	25	NA	23%	0.39	0.46
	CS2	23	16	NA	22%	0.40	0.48
	CS1	14	20	NA	25%	0.43	0.51
AC4	Full Comp	60	28	NA	-4%	0.53	0.61
	CS2	23	17	NA	3%	0.48	0.69
	CS1	15	22	NA	-3%	0.49	0.66

Figures

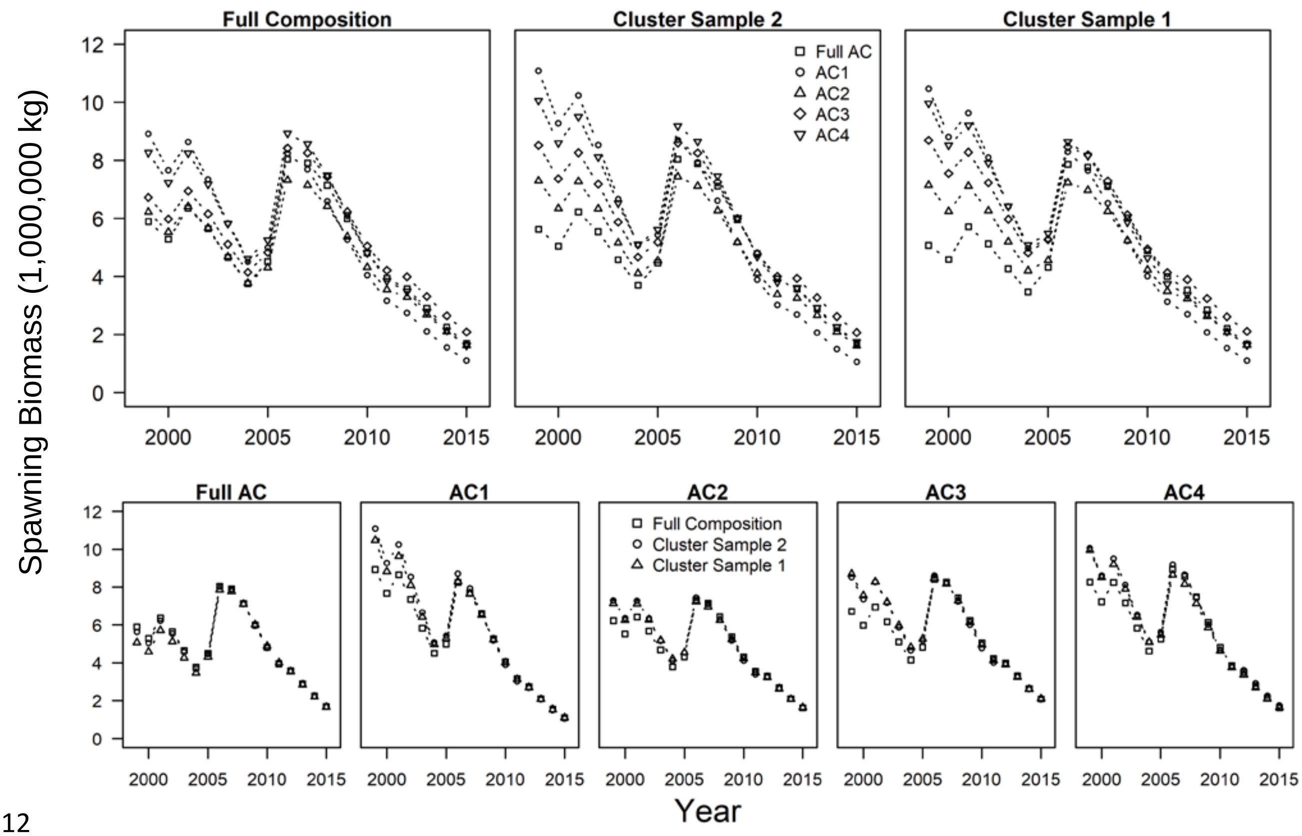

Figure 1. Point estimates (medians) of spawning biomass for the time series. Each row depicts the 15 main set of configurations with different hydroacoustic and composition scenarios. The three panels in the top row each relate to assessment models informed by a given fishery age composition dataset. Within the three panels on the top row each trend line depicts estimates from an assessment model informed by a given hydroacoustic dataset. The bottom row of plots is the opposite, where each panel depicts assessment models informed by different hydroacoustic datasets and within each panel the trend lines are estimates from assessment models informed by different fishery age composition datasets. Full AC, AC1, AC2, AC3, and AC4 refer to the hydroacoustic survey frequency scenario giving rise to that component of the data (Table 3). Full composition, Cluster sample 2, and Cluster sample 1 indicate whether the full age composition data were used or age composition datasets using just two or one randomly selected clusters per year.

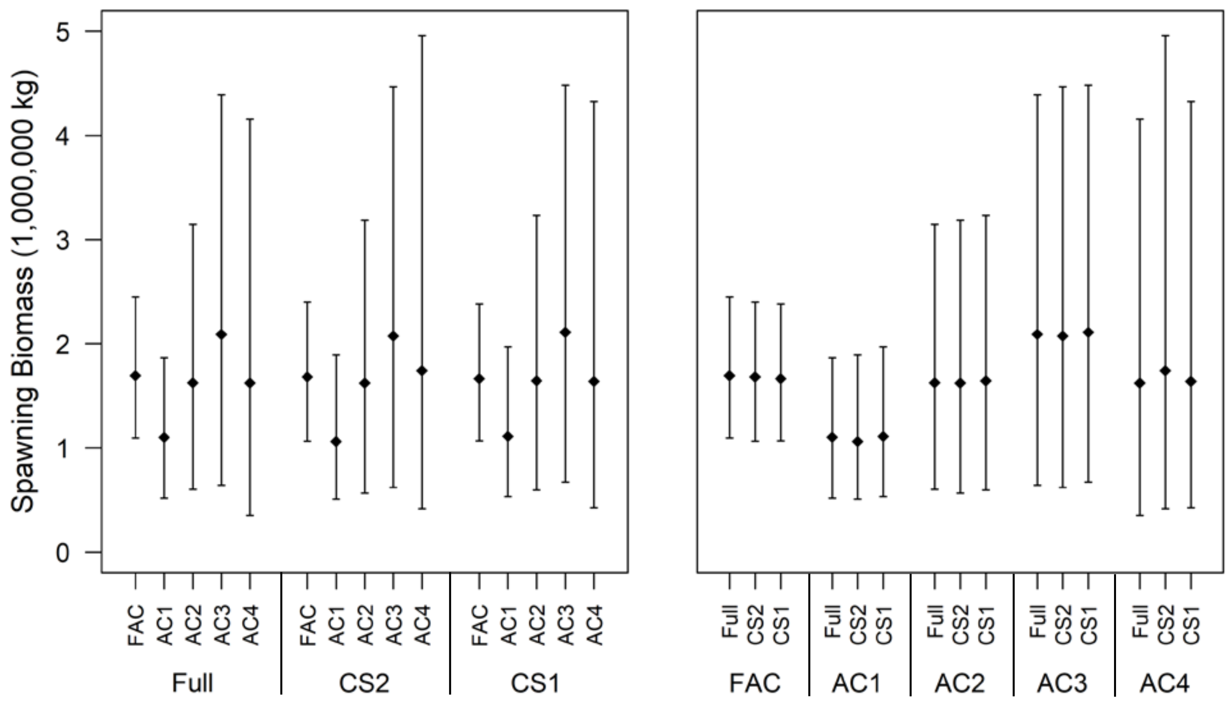

Figure 2. Spawning biomass estimates in 2015 for the main set of scenarios and selected replicate for cluster subsamples. Points denote medians of the posterior distribution and arrows denote 95% HPD intervals. Panels present the same results in a different way. The left panel examines effects of decreases in hydroacoustic survey frequency independent of fishery age composition datasets while the right panel examines the opposite. The vertical x-axis titles in the left panel depict hydroacoustic scenarios and the horizontal x-axis titles represent different fishery age composition scenarios. The opposite is true in the right panel. FAC, AC1, AC2, AC3, and AC4 all refer to hydroacoustic datasets and Full, CS2, and CS1 refer to different composition datasets (CS2 = 2 random clusters per year , CS1 = 1 random cluster per year). FAC refers to the full hydroacoustic dataset, and AC1 through AC4 to increasingly less frequent surveys (Table 3).

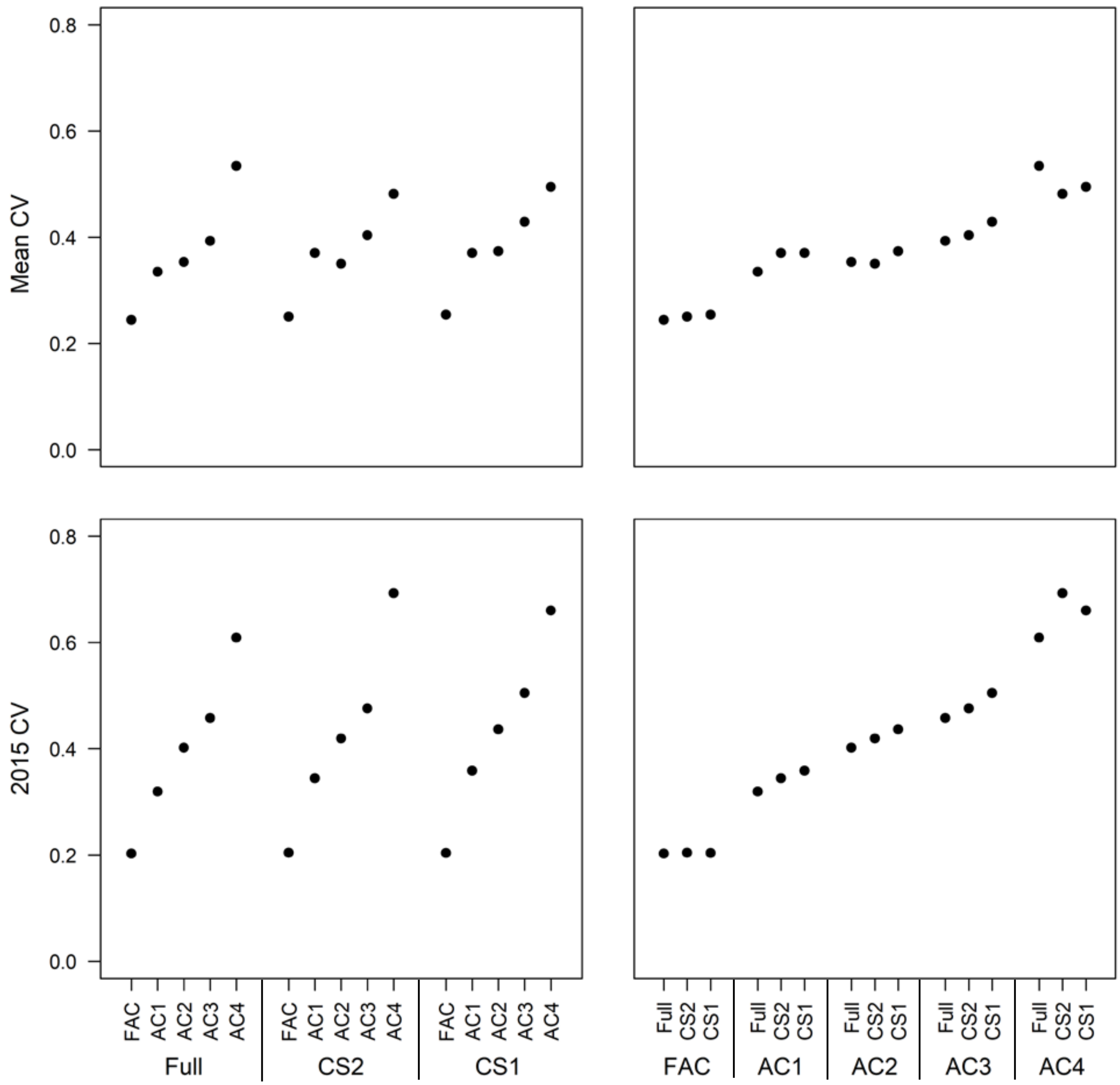

Figure 3. CVs of spawning biomass for the main set of scenarios and the selected replicate for cluster samples. Top row presents the mean CV of posterior distributions of spawning biomass over the full time series while the bottom row solely presents the CV of the posterior distribution for spawning biomass in 2015. Each column is as described in Figure 2. X axis labels as for Figure 2.

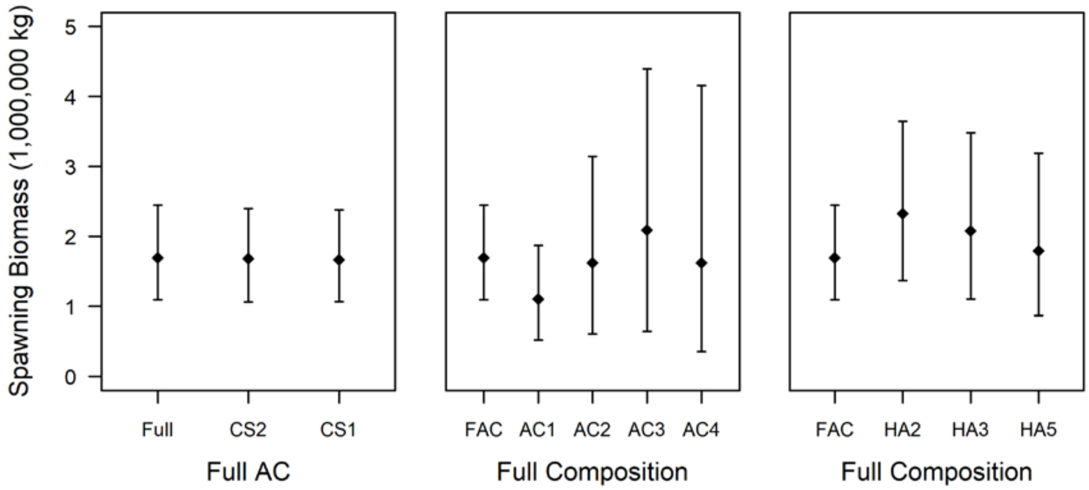

Figure 4. Shown are medians and 95% HPD intervals for spawning biomass in the year 2015. Each panel depicts a different combination of scenarios that gave rise to datasets that models were fit to. The left panel presents estimates from the assessment models informed by the full hydroacoustic dataset and three fishery age composition scenarios (and selected replicate). The middle panel presents estimates from the assessment models informed by the full fishery age composition dataset and various hydroacoustic datasets. The right panel presents estimates from the assessment models informed by the full fishery age composition dataset and various alternate hydroacoustic datasets. The alternate hydroacoustic datasets solely reduce the frequency of the hydroacoustic survey, while they maintain the final data year of 2015 (see methods). FAC refers to the full hydroacoustic dataset, whereas AC1 through AC4 and HA2 through HA5 indicate progressively less frequent surveys (Table 3).

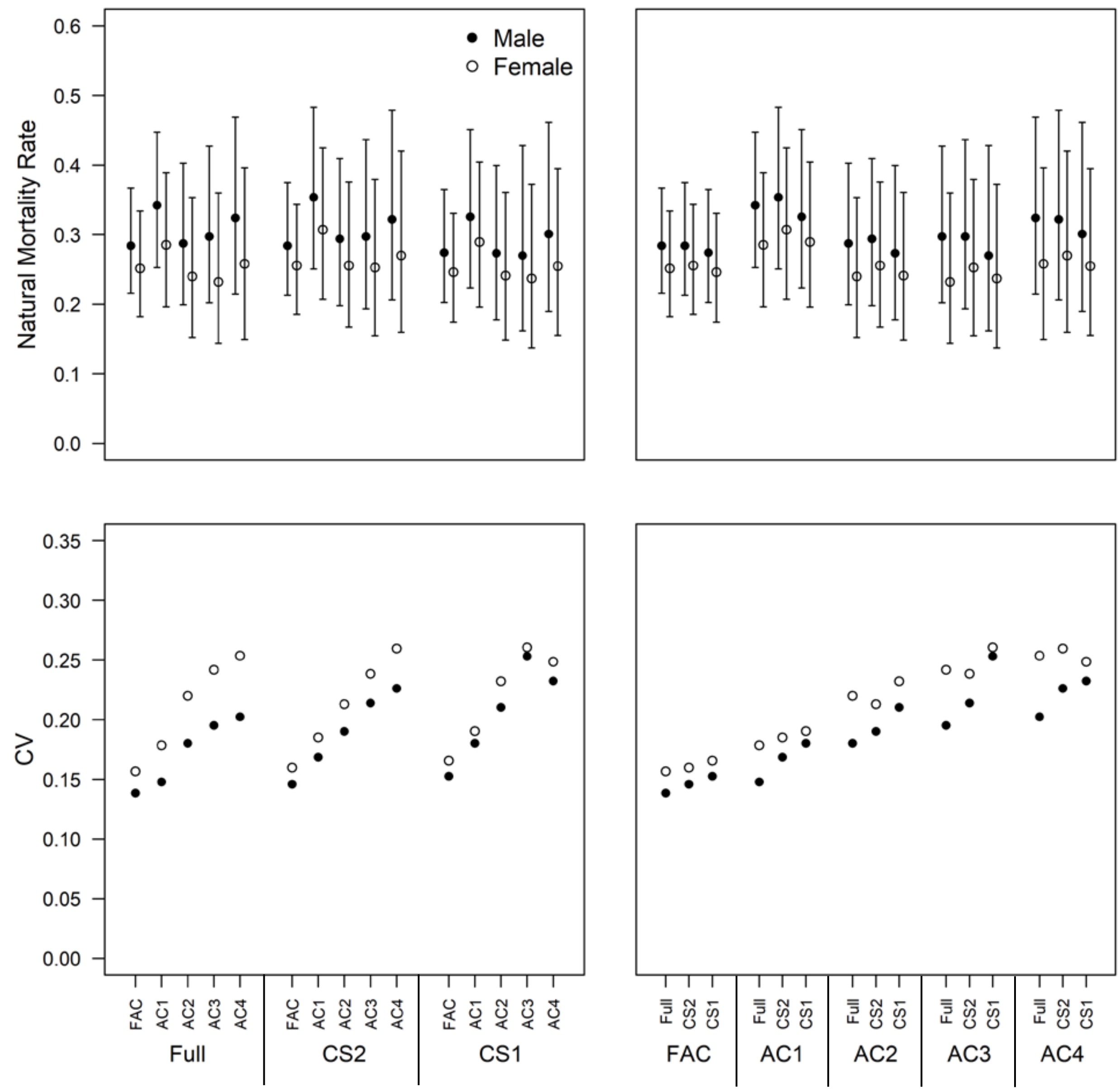

Figure 5. Estimated natural mortality rate (M) and associated estimated uncertainty within the assessment models for males and females for the main set of scenarios (and selected replicate). X axis is as defined in Figure 2. Show in the top row are medians (points) and 95% HPD intervals (arrows) for M for each sex. Shown in the bottom row are the CVs of the posterior distribution of M for each sex.