A Comparison of Age- and Size-Structured Assessment Models Applied to a Stock of Cisco in Thunder Bay, Ontario

Nicholas C. Fisch[1*], James R. Bence[1], Jared T. Myers[2], Eric K. Berglund[3], and Daniel L. Yule[4].

1. Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824-1101
2. U.S Fish and Wildlife Service, Ashland Fish and Wildlife Conservation Office, 2800 Lake Shore Dr. East, Ashland, WI 54806
3. Ontario Ministry of Natural Resources and Forestry, Upper Great Lakes Management Unit, 435 James Street South, Suite 221e, Thunder Bay, ON P7E 6S8, Canada
4. U.S. Geological Survey, Lake Superior Biological Station, 2800 Lake Shore Dr. East, Ashland, WI 54806

Abstract

Stock assessments are critical to modern fisheries management, supporting the calculation of key reference variables used to make informed management decisions. However, there is still considerable uncertainty as to which class of assessment models is appropriate to use under different circumstances. A common class of models used when age data are available are statistical catch-at-age assessment (SCAA) models, which track annual cohorts through time. When age data are unavailable, as is often the case in invertebrate fisheries where the lack of a bony structure such as otoliths makes aging difficult, statistical catch-at-size assessment (SCSA) models are more often employed, tracking fish or invertebrates through time by size-classes rather than ages. Do SCAA models actually perform better than SCSA models when age data are available, or is this just an assumption we make in fisheries research and management? We examined this question by evaluating the effectiveness of both SCAA and SCSA models in characterizing cisco, *Coregonus artedi*, population dynamics in Thunder Bay, Ontario. Both models were fit using an integrated framework with multiple sources of data including hydroacoustic estimates of spawning stock, fishery-dependent and -independent age/length compositions, and harvest data. Our results suggest that for cisco in Thunder Bay, data-limitations related to lack of size-composition data over the size range for which cisco growth is rapid resulted in difficulty estimating relative year-class strength within a SCSA. This led to parameter confounding and ultimately the inability to estimate natural mortality within a SCSA. This hampered the utility of a SCSA model in comparison with a SCAA model when age-composition data were available.

*Corresponding author: fischnic@msu.edu

1. Introduction

Stock assessment is a critical aspect of fisheries research and management, supporting the calculation of key quantities such as spawning biomass, abundance, exploitation rate, recruitment, and their associated uncertainties. Most assessments conducted in the United States are based on age-structured assessment methods (Punt et al., 2017), which, when statistically fit, can be referred to as statistical catch-at-age assessment (SCAA) models. SCAA models are based on the assumption that most population processes are a function of age, and they work by tracking cohorts of fish through time, using observations of catch-at-age and auxiliary information to estimate population parameters (Fournier and Archibald, 1982; Deriso et al., 1985). When catch-at-age data are unavailable for a species of interest, as is the case in many invertebrate fisheries where lack of a bony structure such as an otolith makes aging difficult, size-structured assessment methods are often employed (Punt et al., 2013). Similarly, when statistically fit these types of models can be referred to as statistical catch-at-size assessment (SCSA) models. SCSA models, contrary to SCAA counterparts, are largely based on the assumption that most population processes are a function of size rather than age. Sullivan et al. (1990) developed and applied a framework for SCSA, which differs from SCAA in that it utilizes observations of catch-at-size and tracks fish in size bins rather than age-classes through time, often making use of a growth model that determines transition probabilities of size bins in subsequent time steps. Although age-structured models can be fit using harvest size-composition data, generally by using a model to convert predicted age-compositions to size-compositions (Fournier et al., 1990, 1998), contemporarily the use of SCSA is preferred when the sole or primary harvest composition data are for sizes rather than ages (Punt et al., 2013).

Each method offers distinct advantages and disadvantages. For size-based methods, the model can directly account for the size structure of removals from a population (Punt et al., 2017), it thus can more appropriately model some fishery processes such as selectivity as size-based, and importantly, size-composition data is almost always more abundant because it is both easier and cheaper to collect. While the observation model of a SCAA can account for size-based selectivity, accounting for how size-at-age of the survivors is altered by fishing is more challenging (see Methot (2000) for one approach). SCSA models can considerably decrease the number of fish that need to be aged, as age-compositions of the catch are not required. SCSA is not without its challenges. Primary among them is the need for a growth model to determine transition probabilities through size bins for each time step, where additional aspects such as time-varying or density-dependent growth can add complexity. Although growth models are often specified outside of SCAA models to convert abundances-at-age to biomass, their derivation is not critical to model fit as they are often not used to predict data (provided yield is predicted using mean weights-at-age from harvest data). The transition of fish from one year to the next is much simpler within SCAA models, which benefits from the fact that a fish must be a year older in the next (yearly) time step; a caveat being that our ability to observe ages is not perfect, as there is measurement error involved in aging organisms, and ignoring this error can result in biased assessment output (Coggins and Quinn, 1998; Reeves 2003; Bertignac and Pontual, 2007). Although aging error is not always accounted for in SCAA models, it can be (Thompson et al., 2011; Methot and Wetzel, 2013). In addition, the effects of aging error can be minimized using quality control in aging techniques (Campana, 2001).

Perhaps due to the deterministic transition of fish through age bins, and advantages associated with this, very seldom are SCSA models developed for species when age data are available. Additionally, few studies have compared the two methods. One such study, Punt et al., (2017),

80 used simulation analysis to compare the performance of age-, size-, and age- and size-structured
81 assessment methods and concluded, based on an age- and size-structured operating model, that
82 size-structured and age- and size-structured assessment methods performed best, while age-
83 structured methods performed poorest. A key factor specified in the operating model for this study
84 was that growth was modelled using a size-transition matrix, which likely gave the size structured
85 approaches an advantage. This highlights that this study was done, as are most simulation studies,
86 based on known population dynamics pre-specified by researchers. The advantage of this approach
87 is the ability to compare assessment results to what is pre-specified in the operating model as the
88 true population dynamics of the stock. This specification of the operating model can also limit the
89 applicability of results, if the researchers' conception on the dynamics of the stock and fishery
90 (e.g., survey selectivity as age-based process in Punt et al., 2017), do not actually reflect underlying
91 processes. Fitting alternative models to empirical data can be highly useful in helping to better
92 define plausible processes and informing the direction of future simulation studies.
93   We develop and fit both integrated SCAA and SCSA models for a stock of cisco, *Coregonus*
94 *artedi*, in Thunder Bay (Lake Superior), Ontario. Our objective was to compare and contrast
95 performance of the different assessment methods when applied to an actual stock and to provide
96 recommendations on which type of model may be preferred under different scenarios. We were
97 specifically interested in the overall question: "Does the collection of age-composition data, and
98 its use in a SCAA lead to an improvement in assessment performance over what could be obtained
99 using a size-structured model, without using age composition data?" Given the expenses
100 associated with collecting age-composition data, it is important to know if as good or better results
101 can be obtained with size models, perhaps because they better model fishery processes. To our
102 knowledge, only one study has performed a comparison between age- and size-structured models
103 on a actual stock with true dynamics unknown (Akselrud et al. 2017, concluding that age-
104 structured fit data best). In a time of shrinking natural resource agency budgets, it seems these
105 comparisons could provide managers with valuable information on how they might implement
106 their overall assessment programs.

2. Methods
2.1. Study species
109   Cisco are a planktivorous species native to the Laurentian Great Lakes. They are largely
110 pelagic and form annual spawning aggregations during the month of November in nearshore bays
111 and areas of western Lake Superior, where contemporary spawning stocks are primarily located
112 (Stockwell et al., 2009). These aggregations support a lucrative commercial roe fishery, as fishers
113 generally target spawning fish during November using suspended gillnets (Ebener et al., 2008).
114 Additionally, since 2005 these aggregations have been surveyed annually using hydroacoustic
115 surveys in Thunder Bay. Current management in Thunder Bay relies on a fixed exploitation rate
116 control rule where 10% of spawning biomass estimated from the hydroacoustic surveys are
117 allocated as quota in the subsequent year to a limited number of fishers. No formal assessment
118 models have previously been developed for this or any other stock in western Lake Superior.

2.2. Stock area
120   We treated Ontario Ministry of Natural Resources and Forestry (OMNRF) Quota Management
121 Areas 1-4 (QMAs; Fig. 1) as the stock area for Thunder Bay cisco. This stock has been
122 hypothesized to be discrete because cisco in an adjacent embayment (i.e., Black Bay) have not
123 shown any sign of recovery since a collapse in the 1980s. If cisco from Thunder Bay belonged to
124 a non-discrete spawning stock that inhabited a broader geographic range, it is expected there would

125 have been some level of recovery in Black Bay over the last 30+ years (Ebener et al., 2008).
126 Additionally, this area was chosen based on coverage of the hydroacoustic surveys, which
127 generally sample over QMAs 1-4 in Thunder Bay.

128 *2.3. Data*
129   The SCAA and SCSA models made use of six main sources of observed data in the fitting
130 process (Table 1): (1) Number of cisco > 250 mm in Thunder Bay estimated from hydroacoustic
131 surveys (2005, 2007-2015), age- or size-composition of cisco caught in fisheries-independent (2)
132 mid-water trawls (2005, 2007-2010, 2015) and (3) multi-mesh gillnets (2009, 2013-2015), (4) age-
133 or size-composition of the commercial fishery catch subsamples (1999-2015), and (5) male and
134 (6) female biomass harvested by the fishery each year (1999-2015). The SCSA made use of one
135 additional source of data; (7) individual growth increments of cisco back-calculated from otolith
136 increment data. Details on how data were processed for input into assessment models can be found
137 in Fisch (2018).

138 *2.4. Process model*
139   Predicted quantities needed to compare to the observed data listed above were calculated using
140 a variety of equations describing the stock and fishery. The assessment models ran from 1999 to
141 2015, with parameters estimated using a Bayesian framework. The ages in the SCAA model ages
142 began at 2 and ended at a plus group age of 15 (denoting all cisco older than 14) while the SCSA
143 model size bins were divided in 10 mm increments beginning at 170 mm and ending at a plus
144 group of 410 mm (denoting all cisco $\geq$ 410 mm). The SCSA model starting size bin of 170mm
145 was chosen as this is effectively the minimum size for age 2 fish (Online Supplemental Fig. 1).
146 Age or size bins are referenced throughout the manuscript with subscript j. Given the fishery
147 operates primarily as a roe fishery, it captures a disproportionate number of females in Thunder
148 Bay each year (81% on average; Online Supplemental Fig. 2). For this reason, it was decided to
149 make the assessment models sex-specific, tracking the numbers of male and female fish separately
150 through time. This is presented in subsequent equations with the subscript s, denoting sex. We
151 generally utilized model comparison criteria PSIS-LOO (Pareto smoothed importance sampling
152 leave-one-out; Vehtari et al., 2017) to select between different parameterizations within fitting the
153 SCAA and SCSA. PSIS-LOO is an efficient approximation of the exact cross-validation model
154 comparison criterion, and has been shown to be asymptotically equal to the Widely Applicable
155 Information Criterion (WAIC; Watanabe, 2010). In addition, it is more robust in cases with weak
156 priors or influential observations (Vehtari et al., 2017), both of which occur in assessment
157 modeling. Online Supplemental Table 1 contains an overview of different parameterizations
158 attempted for each model and the reason some were not included in the final parameterization for
159 either the SCAA or the SCSA.
160   To initialize the SCAA model, we estimated sex-specific cisco numbers at age as individual
161 parameters for each sex/age-class combination in the first year of the assessment model. For the
162 SCSA, sex-specific cisco abundance in the first year was estimated through a combination of size
163 and abundance components; two estimated parameters of abundance, ψ_s (one for each sex),
164 multiplied by an initial size composition (non sex-specific) derived using a gamma distribution

165
$$N_{j,1999,s} = \psi_s * \int_{j*-5}^{j*+5} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \qquad (1)$$

where α and β are the estimated shape and rate of the gamma distribution, and $j*$ is the midpoint of size bin j. We fit models with alternative parameterizations including assuming a separate initial size composition for each sex and estimating one value of abundance, in addition to estimating two size-compositions and two abundance values, each of which did not outperform the chosen parameterization (identified above) in terms of model comparison criterion PSIS-LOO (Online Supplemental Table 1).

Sex-specific numbers of cisco at each age or size in each year were calculated from the exponential survival equation

$$\text{SCAA} \quad N_{j+1,y+1,s} = N_{j,y,s} e^{-(F_{j,y,s}+M_s)} \tag{2}$$

$$\text{SCSA} \quad N_{j',y+1,s} = \sum_{j=1}^{j'} P_{j,j'} N_{j,y,s} e^{-(F_{j,y,s}+M_s)} + R_{j',y+1,s} \tag{3}$$

where $N_{j,y,s}$ denotes the number of cisco in age or size bin j alive at the start of year y of sex s. $P_{j,j'}$ is the size transition matrix, denoting the probability that a fish in length bin j will grow into length bin j' in the next time step. $R_{j',y,s}$ is the sex and year specific number of fish that recruit into length bin j'. $F_{j,y,s}$ represents the sex-specific instantaneous fishing mortality during year y for a given age or size bin. M_s is sex-specific instantaneous natural mortality, and is assumed constant over time and ages or sizes. We decided to allow natural mortality to vary by sex because previous studies indicated male cisco may experience higher natural mortality than females (TeWinkel et al., 2002; Yule et al., 2008). In the interest of numerical stability, we added an informative prior on each natural mortality parameter based on the updated Hoenig linear model surrogate equation from Then et al. (2014).

We calculated the size transition matrix, with typical element $P_{j,j'}$, using growth parameters L_∞ and K to model an average growth increment and parameters a and b to model the variance in growth increment as a function of the expected growth increment

$$E(\Delta_j) = (L_\infty - j^*)(1 - e^{-K}) \tag{4}$$

$$Var(\Delta_j) = a + b * E(\Delta_j) \tag{5}$$

We assumed sex-invariant growth because preliminary analyses indicated negligible differences in growth between the sexes both as a function of age and as a function of size. We assumed that these growth increments followed a gamma distribution and for our parameterization of the gamma probability density function (pdf; eq. 1), the relationship between the expected value and the variance and variance is given by $E(\Delta_j) = \dfrac{\alpha_j}{\beta_j}$ and $Var(\Delta_j) = \dfrac{\alpha_j}{\beta_j^2}$, respectively. The probability of remaining in the same length bin in the next time step, $P_{j,j}$ (the diagonal elements of the matrix), was calculated by integrating the gamma pdf ($g(\Delta_j | \alpha_j, \beta_j)$) from 0 to 5 mm (assumes fish are at midpoint of the length bin). Assuming no negative growth, the rest of the transition matrix

199   elements, $P_{j,j'}$ ($j \neq j'$), were calculated by integrating from the growth increment required to

200   reach the lower bound of length bin j' (Δ'_1) to the growth increment required to reach the upper

201   bound of that bin (Δ'_2).

202
$$P_{j,j'} = \int_{\Delta'_1}^{\Delta'_2} g(\Delta_j \mid \alpha_j, \beta_j) d\Delta_j \tag{6}$$

203   In calculations of the probability density, α_j and β_j were first solved for using the $E(\Delta_j)$ and

204    $V(\Delta_j)$ obtained from eqs. 4 and 5. The size transition matrix was started at the 60-70mm size bin

205   (instead of model starting size of 170mm) so as to facilitate calculation of the recruitment size

206   distribution, further described below. The size transition matrix was also derived using data on fish

207   smaller than the model starting size.

208   For the SCAA, we modeled recruitment in each year as multiplicative deviations about a

209   median recruitment level (μ):

210
$$R_y = \mu \delta_y \tag{7}$$

211   The log of the deviations, $\log(\delta_y)$, was assumed to be normally distributed with mean 0 and

212   variance σ^2.

213
$$\log(\delta_y) \sim N(0, \sigma^2) \tag{8}$$

214   We assumed equal sex ratios at recruitment, apportioning 50% of the recruitment each year to

215   the model starting age of each sex.

216
$$N_{2,y,s} = 0.5 R_y \tag{9}$$

217   where $N_{2,y,s}$ denotes the number of cisco age 2 (model starting age) in year y of sex s. Note that

218   this model does not assume any prior relationship between the magnitude of recruitment and stock

219   size. We modeled recruitment in the SCSA model with an added length-based component p_j,

220   representing the proportion of recruits going into each size class.

221
$$R_{j,y} = R_y p_j \tag{10}$$

222   where R_y is modeled in the same fashion as in the SCAA. We calculated the proportion of recruits

223   going into each length bin, p_j, by specifying a size distribution of fish smaller than the model

224   starting size and growing them (within the assessment) one time step into the future according to

225   the growth transition matrix (i.e., take a distribution of fish from bins 10-170 mm, how many of

226   those fish and what model size bins would they be in (> 170 mm) if they grew for one time-step

227   according to the size transition matrix). We calculated this pre-model size distribution (size

228   distribution of the cisco less than 170 mm) using the mean and variance in length-at-age of cisco

229   ages 0-2 (Online Supplemental Fig. 3). To account for depletion in abundance, we weighted

230   different ages using a natural mortality value of 0.3yr^{-1} (i.e., age-0 = 1, age-1 = 0.74, and age-2 =

231   0.55). Within the SCSA, the pre-model size distribution is grown according to the size-transition

232   matrix for one time step and the proportion of fish in each model size bin (> 170 mm) divided by

233 the total number in size bins greater than 170 mm is p_j. Once again, we assumed equal sex ratios
234 at recruitment, and apportioned 50% of the recruitment each year to the model length bins of each
235 sex, $R_{j,y,s} = 0.5 * R_{j,y}$. Other standard approaches to estimating the recruitment size distribution
236 in size-structured assessment models include specifying the distribution using a parametric
237 distribution such as a normal or a gamma (Punt et al., 2013). We initially ran the SCSA model
238 with the recruitment size distribution specified as a gamma distribution estimated using two
239 parameters. However, the resulting recruitment distribution was implausible given a steep drop in
240 the posterior distribution from length bin 170-180mm to 180-190mm. In addition, results of
241 recruitment and spawning biomass from this model were nearly identical to those resulting from
242 the SCSA parameterization described above, suggesting our results are robust to this specification
243 of the recruitment size distribution (Online Supplemental Fig. 4).
244   We calculated instantaneous fishing mortality for both models using

$$F_{j,y,s} = s_j f_{y,s} \tag{13}$$

246 where s_j is the fishery selectivity for cisco over ages or size bins and $f_{y,s}$ is the fishing intensity
247 in a given y year for sex s. During preliminary analysis and in previous work for cisco from
248 Lake Superior (Rick Clark, Quantitative Fisheries Center, Michigan State University, East
249 Lansing, USA, pers. Com.), there was no clear relationship between hydroacoustic estimates of
250 abundance and gillnet catch or catch per unit effort (CPUE). Due to this, fishery effort was not
251 used when fitting the model. Instead we directly estimated fishing intensity in each year.
252   We modeled fishery selectivity using a two parameter gamma function as in Deriso et al.
253 (1985)

$$s_j = \frac{j^\alpha e^{-\beta j}}{s_{ref}} \tag{14}$$

255 where α and β are gear selectivity coefficients and the denominator denotes the value that would
256 be obtained for the numerator for a reference category, made age 7 and size bin 380-390 mm for
257 each respective model. We initially estimated fishery selectivity parameters independently for each
258 sex. However, in the interest of parsimony, given near identical estimates we decided to assume
259 that the same selectivity function applied to both sexes. We chose to model fishery selectivity as
260 time-invariant as preliminary analyses found that mesh sizes used and mean sizes at age of fish
261 caught in the commercial fishery have remained fairly constant since 1999.

262 *2.5. Observation model*
263   Our basic approach was to use age-composition, but not size-composition data when fitting the
264 SCAA, and size-composition, but not age-composition data when fitting the SCSA. Based on a
265 reviewer suggestion, we also attempted to fit the SCAA only using size-composition data, as a
266 way to evaluate whether differences in performance were more due to the process model or the
267 data used. This alternative model failed to converge and thus is not considered further (see
268 Appendix for additional details).
269   Predictions of data source 1, the hydroacoustic estimates of the number of cisco greater than
270 250 mm, $\hat{H}_y$, were modeled using

7

271   SCAA $$\hat{H}_y = e^\gamma \sum_s \sum_j P(Fish_j > 250mm) N_{j,y,s} e^{-Z_{j,y,s}\frac{5}{6}}$$ (15)

272   SCSA $$\hat{H}_y = e^\gamma \sum_s \sum_{j=250}^{L} N_{j,y,s} e^{-Z_{j,y,s}\frac{5}{6}}$$ (16)

273 where $P(Fish_j > 250mm)$ is the probability that a cisco in age bin j is greater than 250 mm, L

274 represents the terminal size bin, and γ is the logarithm of the hydroacoustic survey calibration

275 coefficient (Hulson et al., 2008), which when presented as e^γ, can be referred to as hydroacoustic

276 catchability. Given that selectivity of the hydroacoustic survey is assumed to be knife edged at 250

277 mm (based on a target strength cutoff), where all fish become fully selected to the gear,

278 hydroacoustic estimates of spawning stock are in theory absolute estimates of spawning stock, so

279 γ was expected to be at or very near 0. The 250 mm cutoff is used for hydroacoustic surveys in

280 Thunder Bay as cisco of this size are generally mature (Yule et al., 2006, 2008). We applied

281 mortality to numbers of fish-at-age or –at-length for the first 10/12ths of the year ($N_{j,y,s} e^{-Z_{j,y,s}\frac{5}{6}}$)

282 given the hydroacoustic surveys are performed during the spawning season in November. The

283 probability that a fish of a given age is greater than 250 mm, $P(Fish_j > 250mm)$, was calculated

284 outside of the model by characterizing the size distribution of each age of cisco using mean length

285 and variance in length at age of cisco. A full description of how this was done can be found in

286 supplemental files (Online Supplemental Fig. 5).

287   Predictions of the age or size composition of Thunder Bay cisco each year obtained from mid-

288 water trawls (data source 2) and multi-mesh gillnets (data source 3), were modeled using

289 $$\hat{P}_{j,y,s}^i = \frac{s_j^i N_{j,y,s} e^{-Z_{j,y,s}\frac{5}{6}}}{\sum_s \sum_j s_j^i N_{j,y,s} e^{-Z_{j,y,s}\frac{5}{6}}}$$ (17)

290 where s_j^i is the survey selectivity of gear i (midwater trawls or multi-mesh gillnets) for age or

291 size j. $\hat{P}_{j,y,s}^i$ is the sex-specific predicted proportion at age or size caught from each survey gear

292 in a given year. Once again we applied mortality to numbers of fish at age or length for the first

293 10/12ths of the year ($N_{j,y,s} e^{-Z_{j,y,s}\frac{5}{6}}$) given the survey bio-data are collected during the spawning

294 season in November. Survey selectivity was modeled to adequately characterize the selective

295 nature of mid-water trawl surveys and multi-mesh gillnets using a two parameter gamma function

296 identical to the formula described for the fishery, however without a denominator to standardize

297 the selectivity, given we were calculating relative values. We chose not to use the CPUE data from

298 midwater trawl or multi-mesh gillnet surveys as often the gears are deliberately set on schools of

299 fish seen from hydroacoustic gear so as to make an inference on species composition in the water

300 column for use in separating out hydroacoustic targets. Thus we did not think mid-water trawl or

301 multi-mesh gillnet CPUE would be reflective of abundance.

Predictions of data source 4, the age- or size-composition of the fishery catch, was modeled using the Baranov catch equation

$$\hat{C}_{j,y,s} = \frac{F_{j,y,s}}{F_{j,y,s} + M_s} N_{j,y,s} (1 - e^{-(F_{j,y,s}+M_s)}) \tag{18}$$

$$\hat{P}^f_{j,y,s} = \frac{\hat{C}_{j,y,s}}{\sum_s \sum_j \hat{C}_{j,y,s}} \tag{19}$$

where $\hat{C}_{j,y,s}$ is the predicted number of commercially caught cisco age or size j in year y of sex s, and $\hat{P}^f_{j,y,s}$ is the predicted sex-specific age or size composition of the fishery catch each year. Predictions of data sources 5 and 6, the yield of each sex in each year, was modeled using

$$\hat{Y}_{y,s} = \sum_j \hat{C}_{j,y,s} w_{j,s} \tag{20}$$

where $\hat{Y}_{y,s}$ is the predicted sex-specific fishery yield each year and $w_{j,s}$ is the mean sex-specific weight of a commercially caught cisco age or size j. For the SCSA, this term was obtained from a weight-length regression (Fisch, 2018). For the SCAA, to account for age-length key fixed allocation bin sampling bias in mean weight-at-age (Quinn and Deriso, 1999), we calculated the adjusted average weight of a commercially caught cisco similar to $P(Fish_j > 250mm)$, using

$$w_{j,s} = \frac{\sum_i (n_{i,j,s} * W_i)}{\sum_i n_{i,j,s}} \tag{21}$$

where $n_{i,j,s}$ is the number of fish in bin i that are age j of sex s, and W_i is the average of the length bin endpoints converted to weight using the same weight-length relationship mentioned previously. We used 10 mm bins, the same bin sizes used in the age-length key sampling procedure.

Predictions of data source 7, the individual cisco growth increments, which were only used in the SCSA, were calculated using

$$E(\Delta_l) = (L_\infty - l)(1 - e^{-K}) \tag{22}$$

$$Var(\Delta_l) = a + b * E(\Delta_l) \tag{23}$$

where l represents the starting length of a cisco (length at start of annulus) and $E(\Delta_l)$ denotes the expected growth increment of a cisco given starting length.

2.5.1. Aging Error

Aging error was included in the SCAA model by multiplying the model-predicted catch-at-age and the predicted relative catch-at-age from survey gear by an aging error matrix. The aging error

329 matrix, was estimated by characterizing the expected coded age $E(C_j)$ given true age j and the
330 coefficient of variation of coded age given true age as linear functions.

331 $$\mathrm{E}(C_j) = c + d * j \tag{24}$$

332 $$CV(C_j) = e + f * j \tag{25}$$

333 where c, d, e, and f are estimated parameters. For ease of computation, given preliminary fits
334 suggested c and f were ≈ 0, further runs of the model fixed both parameters at zero. The
335 probability that a fish would be assigned coded age i given its true age j, the $\{i, j\}$ element of
336 the aging error matrix, was computed using the difference in the cumulative distribution function
337 (CDF) of the lognormal distribution, between ages $i + 0.5$ and $i - 0.5$ based on $E(C_j)$ and
338 $CV(C_j)$. The plus group was calculated as 1 minus the CDF of 14.5. The predicted "true" catch-
339 at-age matrix (i.e., without aging error), output by the model, was multiplied by the aging error
340 matrix to obtain the predicted catch-at-age matrix used in calculations of the predicted age-
341 composition of the catch. The same was done for the predicted relative catch-at-age from the
342 survey, which was multiplied by the aging error matrix prior to calculating predicted survey age-
343 compositions.

344 2.5.2. Likelihood
345 We calculated the log likelihood components, L_i, for data sources 1, 5, and 6 (hydroacoustic
346 survey index and annual sex-specific yields) in each model through a lognormal likelihood (with
347 additive constants dropped)

348 $$L_i = -\frac{1}{2\sigma_i^2} \sum_y \left(\log(x_{i,y}) - \log(\hat{x}_{i,y}) \right)^2 - n \log(\sigma_i) \tag{26}$$

349 where σ_i is the standard deviation for likelihood component i, $x_{i,y}$ and $\hat{x}_{i,y}$ are the observed and
350 model-predicted values for year y, and n is the number of observations.
351 The log likelihood components for data sources 2, 3, and 4 for the SCAA (i.e., the age
352 compositions) assumed a robust multinomial likelihood equation as in Starr et al. (1999)

353 $$L_i = \sum_y \sum_j 0.5 \log\left((1 - p_{i,j,y}) p_{i,j,y} + \frac{0.1}{Nb} \right) - \log\left(\exp\left[\frac{-(p_{i,j,y} - \hat{p}_{i,j,y})^2}{2\left((1 - p_{i,j,y}) p_{i,j,y} + \frac{0.1}{Nb} \right) \bigg/ \tilde{N}_{i,y}} \right] + 0.01 \right) \tag{27}$$

354 where $\tilde{N}_{i,y}$ denotes the effective sample size from data source i in year y, $p_{i,j,y}$ and $\hat{p}_{i,j,y}$ are the
355 observed and predicted proportions of cisco in year y that are in age by sex category j from data
356 source i, and Nb represents the number of age by sex bins. Robust likelihoods aid in keeping a
357 small number of outlier composition data points from unduly influencing model fit (Fournier et al.
358 1990; Francis 2011). This is especially important given the nature of cisco year classes in western
359 Lake Superior, which exhibit a "boom-or-bust" pattern where there may be a very large cohort
360 moving through the time series followed by many years of almost no recruitment (Online
361 Supplemental Fig. 6). For the SCSA, given we expect less outlier composition data points as

disparity in year-class strength is reduced due to recruitment into size bins (Online Supplemental Fig. 7), we utilized a regular multinomial likelihood

$$L_i = \sum_y \tilde{N}_{i,y} \sum_j p_{i,j,y} \log(\hat{p}_{i,j,y}) \tag{28}$$

where $p_{i,j,y}$ and $\hat{p}_{i,j,y}$ are the observed and predicted proportions of cisco in year y that are in size by sex bin j from data source i. We initially attempted to fit the SCSA using the robust multinomial likelihood for composition data similar to the SCAA however found that it ultimately led to implausibly low effective sample sizes and very poor fits to the fishery-independent size composition data. For all composition data in the SCAA and SCSA models, both sexes went into one likelihood for each i, meaning only one value of $\tilde{N}_{i,y}$ was used for each data source. This results in double the number of bins (compared to just size bins or age categories) for each i, to account for both males and females.

Data source 7, the individual growth increment data, which was only used in the SCSA, were assumed to come from a gamma distribution, with a gamma log likelihood

$$L_7 = \sum_l \varsigma_l \log(\tau_l) - \log(\Gamma(\varsigma_l)) + (\varsigma_l - 1)\log(\Delta_l) - \tau_l \Delta_l \tag{29}$$

where and ς_l and τ_l are the shape and rate parameters of the gamma distribution, and Δ_l denotes an observed growth increment of a cisco with starting length l.

Log prior components for natural mortality and recruitment deviations that were not compared to data, but rather to expectations specified as informative priors also contributed to the objective function through a normal prior distribution

$$L_i = -\frac{1}{2\sigma_i^2} \sum_j (\log(\hat{x}_{i,j}))^2 - n\log(\sigma_i) \tag{30}$$

where $\hat{x}_{i,j}$ represents the model predicted deviations.

The objective function was then the negative sum of the log likelihood and log prior components

$$L = -\sum_i L_i \tag{31}$$

2.5.3. Data weighting

Standard deviations, σ_i, in likelihood equations for data sources 1, 5 and 6, and for recruitment deviations, were modeled as one estimated parameter and two assumed variance ratios, denoting what we might expect the multiplicative difference in standard deviations to be. We expected the fishery harvest (data sources 5 and 6, sharing a σ) to have the smallest standard deviation (< 0.1), based on a well-developed catch reporting system with extensive monitoring for Thunder Bay cisco. We expected the hydroacoustic standard deviation to be < 1 based on within year CVs for hydroacoustic data (variability among segments) being less than 1. Lastly, from meta-analyses of recruitment deviation suggesting sd of ≈ 0.71 for the order Salmoniformes (Thorson et al., 2014), we expected cisco recruitment variability to be greater given the substantial differences in cohort strength among years (Online Supplemental Fig. 6). We estimated the standard deviation for

fishery harvest, while assuming variance ratios (Vr) of 0.04 and 0.0004 for hydroacoustic estimates of abundance and recruitment deviations, respectively.

$$\sigma_i = \sqrt{\frac{1}{Vr}\sigma_f^2} \qquad (32)$$

where σ_f denotes the standard deviation for fishery harvests. During preliminary analyses, we adjusted variances ratios so that the variances set by ratios were reasonable consistent with prior expectations. The model results were largely insensitive to these.

Effective sample sizes, $\tilde{N}_{i,y}$, for the composition datasets were calculated using the iterative reweighting procedure T3.4 of Francis (2011):

$$\tilde{N}_{i,y} = N_{i,y}w_i \qquad (33)$$

where $N_{i,y}$ denotes the previous iterations effective sample size. w_i was calculated using TA1.8 of Francis (2011):

$$(w_i)^{-1} = Var_y\left[\frac{\left(\bar{O}_{i,y} - \bar{E}_{i,y}\right)}{\left(v_{i,y} / N_{i,y}\right)^{0.5}}\right] \qquad (34)$$

where $\bar{O}_{i,y} = \sum_j x_j * O_{i,j,y}$ and $\bar{E}_{i,y} = \sum_j x_j * E_{i,j,y}$ are the observed and expected mean ages or

lengths and $v_{i,y} = \sum_j (x_j^2 E_{i,j,y}) - \bar{E}_{i,y}^2$ is the variance of the expected distribution of age or length

calculated over sexes for each year. In these formulas, i denotes a composition data set (age or size; commercial fishery, mid-water-trawl survey (MWT), or multi-mesh gillnet survey), j denotes age or size by sex bin, and x_j is the age or length (bin midpoint) for bin j. The initial effective sample sizes were set at the total number of fish sampled in a year. We fit the models iteratively (changing effective sample sizes) using penalized maximum likelihood estimation until effective sample sizes converged on values within one unit of the previous iteration's value. Once effective sample sizes converged, the models were run using Bayesian methods. All models were run in Automatic Differentiation Model Builder (ADMB; Fournier et al., 2012). Effective sample sizes for the SCAA converged on 62, 45, and 50 for the fishery, MWT, and multi-mesh gillnet compositions, respectively. Similarly, effective sample sizes for the SCSA converged on 58, 22, and 11 for the fishery, MWT, and multi-mesh gillnet compositions, respectively.

2.6. Model fitting/calibration/troubleshooting
2.6.1. SCAA
The SCAA model was unable to converge on an estimate of γ, which denotes the logarithm of hydroacoustic catchability. Essentially this parameter scales our entire population by representing what proportion of spawning cisco the hydroacoustic survey is actually detecting. We decided to assume a conservative scenario where $\gamma = 0$, which assumes the hydroacoustic survey is an absolute index of spawner abundance. By conservative, we mean that actual catchability is likely lower and abundance is likely higher. This in turn means that quotas calculated when $\gamma = 0$

430   will likely be lower than if the target exploitation rate were applied to an abundance estimate when
431   $\gamma < 0$. The hydroacoustic surveys are generally thought to be a conservative estimate of
432   abundance as all areas of the water column are not sampled effectively with the gear (Yule et al.,
433   2012). The MCMC sampler was run for 10 million iterations each saving every 500th, dropping
434   the first 2,500 values from the saved chains as a burn-in period when summarizing posterior
435   distributions. Chain burn in was assessed visually and convergence determined using Geweke's
436   convergence diagnostic (Geweke, 1991). Long single chains or multiple chains are alternative
437   reasonable approaches to checking MCMC convergence (Cowles and Carlin, 1994). We elected
438   for a long single chain because we had the computing power to do so. For long single chains,
439   Geweke's diagnostic (Geweke, 1991) is a common statistic used to diagnose non-convergence of
440   a MCMC chain (Cowles and Carlin, 1994). In addition to results from the Geweke's diagnostic,
441   we did start chains from alternative starting points and found that our results were not sensitive to
442   this.

443   2.6.2. SCSA
444   It became clear at the start of model calibration for the SCSA that the model was going to be
445   unable to output plausible estimates of natural mortality (*M*). The model would confound estimates
446   of recruitment, selectivity, and natural mortality. It was unable to converge on plausible estimates
447   of natural mortality even when given assumed known growth parameters at levels previously
448   estimated using fixed natural mortality at prior point estimates. The model would increase natural
449   mortality to an implausibly high value, inflate recruitment, and make larger fish more selected.
450   What the model was doing was creating many fish through recruitment, killing them off at high
451   rates through natural mortality in order to have enough fish at spawning sizes to fit the
452   hydroacoustic data. Few large fish were predicted to survive, but fishery selectivity was highest
453   for the largest fish to fit the fishery composition data. We confirmed the confounding of natural
454   mortality with selectivity and recruitment by fixing *M* at a range of alternative values (with
455   hydroacoustic catchability fixed), which led to substantial changes in recruitment and selectivity
456   but little overall change in model fit when *M* was increased. Similar to the SCAA, the SCSA was
457   also unable to converge on an estimate of hydroacoustic catchability. Given these issues, we
458   decided to fix natural mortality at its prior point estimates (0.283yr⁻¹ for males, 0.256 yr⁻¹ for
459   females), fix hydroacoustic catchability at 1 ($\gamma = 0$), and estimate growth. We ran the model for
460   double the number of samples (20 million) and also doubled chain burn in (first 5,000 iterations
461   of the saved chain). Likelihood profiles over γ for each model and over M for the SCSA can be
462   found in Online Supplemental Fig. 8.

463   *2.7. Comparison*
464   Given different data used in each assessment, it was not possible to compare the final models
465   in terms of predictive accuracy/information theoretic measures such as PSIS-LOO, WAIC, or DIC.
466   Instead, final models were compared using a variety of criteria. First, we considered what
467   assumptions we had to make to fit each model. We also looked at retrospective patterns,
468   parameter/output uncertainty and model fit/residuals. We also fit an additional SCAA model with
469   fixed natural mortalities (at the same values fixed in the SCSA) so as to be able to compare
470   uncertainty and retrospective patterns for both models when natural mortality was fixed.
471   Retrospective analyses primarily focused on spawning biomass and exploitation rate. Mohn's rho
472   (Mohn, 1999) was calculated for spawning biomass and exploitation rate as the mean relative error
473   for the last year of each peel compared to the corresponding year in the last assessment.

474
$$\rho = \overline{\left(\frac{Y_{F,p} - Y_{F,ref}}{Y_{F,ref}}\right)} \qquad (35)$$

475   where Y is the assessment output quantity, either spawning biomass or exploitation rate, *ref* refers
476   to the last assessment, and F refers to the final year of a given assessment peel, p. Five years
477   were removed from the assessment. We also calculated a mean final year absolute difference for
478   the retrospective analyses, as the mean absolute value of the relative error for the last year of each
479   peel compared to the corresponding year in the full assessment. This statistic considers the
480   difference in estimates in the final year of each peel to the reference assessment as opposed to
481   whether or not there is a consistent pattern.

482
$$\lambda = \overline{\left(\frac{|Y_{F,p} - Y_{F,ref}|}{Y_{F,ref}}\right)} \qquad (35)$$

483   We refer to λ as a "precision indicator" because the last assessment is typically viewed as
484   producing more precise estimates than do the peels in their terminal years. Whereas ρ can provide
485   useful information to potentially diagnose model misspecification (Hurtado-Ferro et al., 2015), λ
486   can provide useful information on the quality of your model estimates based on how much they
487   change as data are removed, regardless of a consistent pattern. Statistics ρ and λ were calculated
488   using medians of the posterior distribution as point estimates.
489   Residuals for common data sources were compared using the standard deviation of the
490   normalized residuals (SDNR, Breen et al., 2003; Francis, 2011; Carvalho et al., 2017). These were
491   calculated as the standard deviation of the normalized residual for each data point (formulas in
492   Table B1 in Francis 2011). A relatively good model fit is characterized by smaller residuals and a
493   SDNR near 1 (Carvalho et al., 2017), although Francis (2011) notes that a value much less than 1
494   is not a cause for concern, but rather means that the data set is fitted better than was expected. Due
495   to their correlative nature, composition data points cannot be compared using this metric (Francis,
496   2011), so these were compared visually.

497   **3. Results**
498   Point estimates of quantities output from the models are reported as medians of the posterior
499   distribution, with 95% highest posterior density (HPD) intervals reported in parentheses. The
500   SCAA estimated a total of 89 parameters while the SCSA estimated 67 parameters. All parameters
501   in each model indicated convergence based on Geweke's diagnostic at an alpha level of 0.01. Two
502   parameters for the SCAA (one initial abundance and one recruitment deviation parameter)
503   compared to nine parameters in the SCSA (three fishing intensities, five recruitment deviations,
504   plus the male initial abundance scaler) produced a Geweke's diagnostic that was significant at an
505   alpha level of 0.05. Natural mortality estimates within the SCAA for males and females were
506   0.285yr^{-1} (0.212-0.371) and 0.253 yr^{-1} (0.178-0.340) respectively.

507   *3.1. Spawning biomass*
508   SCAA spawning biomass, defined as the mature female biomass (>250 mm), began at 4.99
509   (2.07-11.12) million kg, initially declined then rose to an estimate of 4.90 (2.87-8.01) million kg
510   in 2006 and ended the time series at 0.95 (0.58-1.40) million kg (Fig. 2). For the SCAA model
511   with fixed natural mortality values, point estimates and 95% HPD intervals for spawning biomass

in 1999, 2006, and 2015 were 5.26 (3.36-7.88), 5.02 (3.92-6.32), and 0.95 (0.64-1.33), respectively (Fig. 2). SCSA spawning biomass increased from 2.04 (1.26-3.43) million kg at the start of the time series to a peak of 3.01 (2.15-3.97) million kg in 2008 before decreasing to 1.45 (0.98-2.03) million kg in the final year, 2015.

3.2. Exploitation rate

Exploitation rate was defined as yield divided by the biomass of fish larger than 250 mm. Exploitation in the SCAA was modest throughout the time series, hovering around 3.5%, although in 2010 began to increase resulting in a final year estimate of 9.1% (5.5%-13.5%; Fig. 2). This resulted in fully-selected fishing mortality rate estimates of 0.08yr^{-1} (0.04-0.12) and 0.20 yr^{-1} (0.11-0.32) for males and females in 2015, respectively. For the SCSA, exploitation rate decreased from 8% (4%-13%) at the start of the time series to 3% (2%-4%) in 2007 and increased throughout the rest of the time series to a final year estimate of 6% (4%-8%). Final year fully-selected fishing mortality rates were 0.05 yr^{-1} (0.03-0.08) and 0.17 yr^{-1} (0.10-0.25) for males and females.

3.3. Recruitment

As expected, recruitment was highly variable throughout the time series, with evidence of ~4 "boom" recruitment years in the SCAA, belonging to 1998, 2003, 2005, and 2009 year classes. Estimates of recruitment (age-2 fish) for these years (2000, 2005, 2007, and 2011) were 19.27 (7.08-41.86), 36.84 (18.76-64.99), 1.87 (0.87-3.32), and 4.40 (2.33-7.17) million fish, respectively (Fig. 3). Recruitment was low for 10 of the years (i.e., estimated to be around 15,000 fish). Recruitment for the 3 remaining years were estimated at modestly low values, with estimates ranging from 0.86 (0 -1.89) million in 2004 to 0.21 (0-1.01) million fish in 1999.

Recruitment in the SCSA showed a similar trend to SCAA recruitment with 3-4 clear modes most likely attributed to the introduction of fish >170 mm from the 1998, 2003, 2005, and 2009 "boom" year classes (Fig. 3).

3.4. Abundance

SCAA abundance echoed spawning biomass results, with intermittent spikes due to "boom" recruitment years and an overall declining trend at the end of the time-series. In 1999 the model predicted there were around 12.71 (5.30-28.53) million cisco, a high of 44.25 (23.36-77.21) million estimated in 2005, and in the final year 3.90 (2.49-5.73) million (Fig. 2). SCSA-predicted abundance began the time series at 8.02 (3.92-14.42) million fish, 30.95 (21.31-41.30) million fish at its peak in 2005, followed by a decrease to around 5.60 (3.60-8.48) million fish in 2015.

3.5. Aging error

Very little aging error was estimated within the SCAA model. Approximately no bias was estimated in aging as true age increased ($d \approx 1.00$, Table 2), and the estimated CV ($e = 0.02$, Table 2) was very low.

3.6. Growth

L_∞ and K were estimated at 428 mm (419-438) and 0.28 (0.25-0.31), respectively (Fig. 4). Estimates for growth variance parameters a and b were 124.82 (32.25-252.62) and 2.68 (0.60-7.19), respectively.

3.7. Retrospective analyses

Retrospective patterns for the latter half of the time series for each model were very similar (Fig. 5). Mohn's ρ estimates for spawning biomass and exploitation rate were both ≈ 0 for the SCAA and 0.19 and -0.13 for the SCSA, respectively. All of these ρ values are within a range of values deemed "not a cause for concern" in retrospective analyses (Hurtado-Ferro et. al., 2015). Precision indicators (λ) for spawning biomass and exploitation rate were 0.16 and 0.18 for the SCAA and 0.21 and 0.16 for the SCSA, respectively. Retrospective statistics for the SCAA model fit with fixed natural mortalities were ≈ 0 and 0.02 for ρ and 0.11 and 0.14 for λ with reference to spawning biomass and exploitation rate, respectively (Fig. 5).

3.8. Model fit to data

Assessment model fits to the hydroacoustic data were very similar (Fig. 6). Both assessments treated the observed hydroacoustic spawning abundance estimate in 2011 as an outlier. Outside of that outlier data point, both models predicted a near linear decline in spawning abundance since 2005, in accordance with the observed data points. The median of the negative log-likelihood for the fit to hydroacoustic data was lower for the SCAA (Table 3). SDNR values for the hydroacoustic data were also closer to 1 for the SCAA, indicating better fit.

Fits to harvest data were nearly identical between the assessments (Online Supplemental Fig. 9). HPD intervals were slightly smaller for the SCSA model. The medians of the negative log-likelihoods for male and female harvest were lower for the SCSA (Table 3), and the SDNR values were smaller for the SCSA. Both model SDNR yield values were well below 1, indicating better model fit than expected.

Both models fit the fishery composition data points well (see Online Supplemental Figs 10-17). These two fits cannot be directly compared as they used different data.

3.9. Computational intensity

The SCSA was considerably more computationally intensive than the SCAA, requiring about 5x the run time for the same number of iterations (where the SCAA took ≈3.5 hours for 10 million, the SCSA took ≈35 hours for 20 million).

4. Discussion

Overall both models showed similar trends in outputs and modest differences in final year estimates (Fig. 2). However, the SCAA model had a larger degree of uncertainty during the first half of the time series, which decreased throughout the second half of the time series to actually end up being less than SCSA uncertainty in the final year (Fig. 2). This decrease in uncertainty and the relative change for the SCAA relative to the SCSA is not just a scaling effect due to the different stock size estimates. The coefficient of variation (CV, calculated as the MCMC-based standard deviation divided by the median estimate) for spawning biomass declined from 0.6 at the start in 1999 to values between 0.15 and 0.3 during 2010-2016 for the SCAA, whereas the SCSA had a CV of about 0.3 in 1999, and similar CVs as the SCAA during 2010-2016. This disparity in uncertainty at the beginning of the time series is likely due to differences in the initial parameterization of each model, where much more flexibility was afforded to the SCAA by estimating 26 initial abundance parameters (one for each age-sex combination above recruitment age). Conversely, the SCSA estimated only four initial abundance parameters; two for the initial size composition, and one for each sex as abundance scalars. Increased certainty in the SCSA during most of the time series may also be driven by the assumption of known natural mortality values. In fact, when we re-ran the SCAA with assumed known natural mortalities at their prior

point estimates, uncertainty in model output decreased substantially (Fig. 2), indicating that the certainty in output expressed by the SCSA is likely in some part due to assuming known natural mortality values.

In terms of model fit, the SCAA had better fit to the hydroacoustic data while the SCSA had better fit to the yield data [although the slightly better fit of the SCSA to yield may not be all that significant, as all yield SNDRs were well below 1, indicating better fit than expected by each model; Francis, 2011]. For the retrospective analysis, where ρ estimates were larger for the SCSA, upon visual inspection the patterns appear comparable between the two assessments, if not worse for the SCAA, specifically at the start of the time series (Fig. 5). The large discrepancies between the two retrospective patterns prior to 2005 are likely due once again to the flexibility afforded to the SCAA in initial abundance parameterization, resulting in highly variable initial abundances for each peel in the SCAA. This result of smaller ρ estimates for the SCAA even though patterns may appear more severe if not equal to those in the SCSA is driven by equally large deviations in the terminal years of peels for the SCAA in opposite directions (i.e., not a consistent pattern but equal numbers of over and under estimates). Given no ρ estimates are at values considered "cause for concern" (Hurtado-Ferro et al., 2015), in this study it may be more prudent to consider λ in comparing retrospective analyses, which was larger for exploitation rate and smaller for spawning biomass for the SCAA compared to the SCSA. Once again, it is likely that assuming known natural mortality values at their prior point estimates led to both a smaller λ estimate for exploitation rate and the appearance of less severe retrospective patterns within the SCSA. When we re-ran the SCAA retrospective analysis with assumed known natural mortality values at their prior point estimates, the precision indicator (λ) decreased. In fact, all retrospective statistics were lower (closer to 0 for ρ) for the SCAA with assumed known natural mortality than the SCSA counterparts (Fig. 5).

An important result of our study is the inability to estimate natural mortality within the SCSA. Given natural mortality is one of the most influential quantities in stock assessment and its estimation within an assessment can be difficult (Lee et al., 2011; Brodziak at al., 2011; Sippel et al., 2017), the ability to estimate this parameter in the SCAA certainly favors the SCAA as an assessment model choice. An interesting note is the remarkable similarity of the prior natural mortality point estimates (the fixed, assumed known SCSA M values; 0.283yr^{-1} and 0.256 yr^{-1}) to the estimated natural mortality point estimates for the SCAA. This is not an artifact of an influential prior; when we ran the SCAA model without specifying informative priors on natural mortality, a similar result occurred (Male M = 0.284 yr^{-1}, Female M = 0.252 yr^{-1}), suggesting that the age-composition data are providing crucial information on natural mortality. The similarity between assumed known natural mortality in the SCSA and estimated natural mortality in the SCAA in addition to the utilization of the same hydroacoustic and yield data likely led to similar output between the two assessments.

The inability to estimate natural mortality within the SCSA due to its confounding with estimates of recruitment and selectivity is not a new finding, as parameter confounding has been noted to be potentially more serious in size-structured assessments (Punt et al., 2013). Where parameter confounding here did not change growth parameters much (mainly influenced selectivity, recruitment, and natural mortality), its underlying cause may have been variation in individual growth, such that variation in size-at-age makes it hard for size-structured models to discern cohorts from length-composition data (Punt et al., 2013). Even when we assumed growth was known within the SCSA, the model still confounded selectivity, recruitment, and natural mortality. Another aspect that may have led to the inability to estimate natural mortality within the

SCSA is the range of vulnerability to the fishery for cisco in Thunder Bay, where by the time cisco start to show up in the fishery length-compositions they are at or very near asymptotic size (Online Supplemental Fig. 7). This results in similarity in fishery size-composition data between years making it difficult to observe strong year classes pulse through the fishery composition data. While the fishery-independent survey gear does select smaller fish and is, to some extent, able to discern cohorts from its length-composition data (likely why recruitment in SCSA for 2003 and 2005 cohorts were approximated well), our survey composition data was limited, only having started in 2005 and missing critical years in 2006 and 2011-2012. The missing survey data pre-2005 and in 2011-2012 likely resulted in recruitment of the 1998 cohort being spread over ~5 years and recruitment of the 2009 cohort entering the population in 2010 in the SCSA (as opposed to 2011, Fig. 3). Where temporal variation in growth could have caused this, we find it less plausible because for the years for which survey composition data are available length distributions given age were fairly constant over time, and the mismatch of recruitment timing occurred when the survey composition data were lacking but not when they were not (i.e., 1998 & 2009 vs 2003 & 2005 cohorts). Fishery-independent survey size-composition data throughout the full time series would have likely resulted in a better approximation of year-class strength and possibly allowed estimation of natural mortality within the SCSA. Alternatively, in the SCAA model, likely due to the boom-or-bust recruitment pattern, the model was clearly able to distinguish 3-4 large year classes moving through the fishery and estimate their associated depletion. Estimation of natural mortality within the SCAA may also have been made possible by relatively light exploitation, effectively making the major source of mortality and transition through the population-age matrix one of natural depletion.

Estimation of natural mortality within size structured assessments is possible. Punt et al. (2013) reviewed integrated size-structured assessment methods and two out of nine assessments that were reviewed in depth estimated natural mortality within the assessment. One of the assessments modeled selectivity as logistic (Fu and Mckenzie, 2010) and the other modeled it as a double normal (Breen et al., 2009; Starr et al., 2009), while both modeled recruitment as lognormal deviates entering the population through a specified size distribution. These selectivity functions are less flexible than a gamma function, which may indicate a reason they did not experience parameter confounding to the extent we did with regard to natural mortality, selectivity and recruitment. However, even if we fixed selectivity and growth at values estimated using assumed known natural mortalities, and then tried to estimate natural mortality, the SCSA model would still inflate recruitment and estimate implausibly high natural mortalities. Where some size-structured models may indeed be able to estimate natural mortality, our study indicates that this may be an even taller task than it is in SCAA models, and depends on a multitude of factors from variability/patterns in recruitment, variability in growth and size at age, and vulnerability range of organisms within size composition data. In addition, our study utilized otolith increment data to estimate the growth transition matrix, whereas in many SCSA models tagging data are used for this purpose (Punt et al., 2013, 2016). Theoretically tagging data can also be used to provide information on mortality rates, and thus may influence the ability to estimate natural mortality within the model. However, they are not often used in this fashion because of concerns over tag-loss and tag-reporting rates (Punt et al., 2013).

4.1. Conclusions and recommendations

Although more uncertain, primarily due to the ability to estimate natural mortality, we conclude that the SCAA is more appropriate for modeling population dynamics of cisco in the Lake Superior. While size-based assessment models can considerably decrease the amount of fish

688 that need to be aged, as this study shows, age-composition data can be crucial to the ability to
689 estimate natural mortality and recruitment within a model. We prefer not to rely on assumed known
690 scale parameters (e.g., hydroacoustic catchability) and natural mortalities, if this can be avoided.
691 Where the assessments both resulted in similar natural mortality estimates, in other case studies
692 this may not occur, and using a surrogate equation for natural mortality may result in biased
693 assessment output. In addition, assuming known natural mortality may artificially decrease model
694 uncertainty. We do not necessarily expect this conclusion to apply for all, or even most species. In
695 fact, this result is likely largely driven by the specific life history of cisco and data availability for
696 the Thunder Bay stock in Lake Superior. Boom-or-bust recruitment most likely facilitated
697 estimation of natural mortality within the SCAA. For the SCSA, however, the fact that most
698 growth occurs before cisco are vulnerable to the fishery, and in some years fishery-independent
699 data that do sample cisco at sizes where they exhibit rapid growth were not available, made the
700 estimation of natural mortality and relative year-class strength difficult within the model. For
701 species with less variable recruitment, less variable growth, and more size-composition data
702 available throughout the growth period of their life span, size-based assessment methods may
703 perform equally well, or better, than age-structured methods. In this specific case study, were a
704 SCSA to replace a SCAA for Thunder Bay cisco, perceived stock biomass would be higher and
705 perceived exploitation rate lower. If the SCSA model were to be used to calculate quotas, they
706 would likely be increased compared to quotas calculated using the SCAA. For stocks with no age
707 data available, if a fishery is targeting individuals at asymptotic size, an onus should be placed on
708 the collection of fishery-independent composition data that targets individuals of smaller sizes so
709 as to provide an assessment model with more informative data on natural mortality and relative
710 year-class strength. In addition, our inability to get an age-structured model to fit to size-
711 composition data may suggest that size-structured models are more appropriate when stocks do
712 not have much age data.
713   Our conclusion, that the SCAA was more appropriate than the SCSA when applied to cisco, is
714 driven by our desire to estimate natural mortality, and our concern that the SCSA recruitment
715 estimates are not reliable. While it is tempting to contrast our conclusion with other comparisons
716 of size- and age-based assessment models (Akselrud et al., 2017; Punt et al., 2017), those studies
717 did not attempt to estimate natural mortality within the assessment models. While Punt et al.,
718 (2017) concluded that age-structured methods performed poorest and Akselrud et al., (2017)
719 concluded that age-structured methods fit the data best, we believe that the conclusions of these
720 studies might depend on their assumption that natural mortality was known. Akselrud et al., (2017)
721 and Punt et al., (2017) also considered a third type of assessment model that take into account both
722 age- and size-based processes in their analyses. Where these age-size models may improve
723 assessment accuracy (Gilbert et al., 2006; McGarvey et al., 2007; Punt et al., 2017), they are also
724 very computationally intensive. We did not consider them in our analyses. It is possible that an
725 age- and size-structured model could outperform both SCAA and SCSA in application to cisco in
726 Thunder Bay. It is also possible to fit both age- and size-composition data within a SCAA (Methot
727 and Wetzel, 2013). We did not pursue this approach because we assumed adding size-composition
728 data to an age-structured model already fit to informative age-composition data would provide
729 little benefit. However, we cannot rule out that such an approach could have outperformed the
730 SCAA and SCSA defined in our analysis for Thunder Bay cisco. Additionally, while we believe
731 the comparisons we made and conclusion we reached in preferring the age-based model is valid,
732 we cannot be sure that the estimated population sizes and mortality rates are closer to true values
733 than those generated by a size-based model, given the truth is not known. Further, our analysis

734 cannot define the conditions under which the natural mortality is estimable and produces useable
735 assessment results, as we had only one data set resulting from one set of conditions. This is an
736 advantage of simulations such as those of Punt et al. (2017) over empirical comparisons of
737 alternative models as shown here. Our empirical comparisons highlighted some aspects of the
738 performance of size- and age-based models contrasted in a real world application and thus can
739 point the way for future simulations. More work is needed that directly investigates the ability to
740 estimate natural mortality and other parameters within size-structured assessment models both
741 from a simulation perspective and in empirical assessments.

Appendix

756   **Appendix**
757   We attempted to fit the SCAA to size-composition data, hereafter referred to as SCAA-CL (Catch-
758   at-Length). The changes from the original SCAA parameterization were as follows:

759   1) Size-composition data replaced age-composition data for the fishery and survey gears.
760   2) Aging error was dropped from the model.
761   3) Mean length-at-age was calculated using a Von-Bertalanffy growth function where L_∞,
762   t_o, and K were estimated as parameters.
763   4) Standard deviation of length-at-age was specified as a linear function of mean size-at-age,
764   with two estimated parameters (intercept and slope).
765   5) The size-at-age distribution was specified using a normal distribution (using mean and SD
766   described above).
767   6) Predicted catch-at-age was converted to catch-at-size and then multiplied by a weight-
768   length function to predict yield (previously catch-at-age was multiplied by mean weight-
769   at-age).
770   7) Selectivity of the hydroacoustic survey was specified as length-based (all fish over
771   250mm). Previously this was specified as the probability that a fish of a given age is greater
772   than 250mm.

773   The model ran into some severe issues, which were very similar to those encountered with the
774   SCSA. The main problem the model had was an inability to output reliable estimates of selectivity.
775   It would estimate an exponential increase past the age at which asymptotic size begins (~8) for the
776   fishery selectivity. This occurred regardless of whether natural mortality or growth was estimated
777   or fixed. The problem, similar to the SCSA, is that the model can track ages 2-8 as fish are growing
778   (only when there are fishery-independent survey compositions). However, once growth stops the
779   model has very little information to inform it for ages 9-15. In the original age-structured model,
780   the model could follow these cohorts late into senescence using information from the age-
781   compositions. However, all old ages in the SCAA-CL have nearly the same size-at-age
782   distribution, thus the model once again necessitates composition data on young fish as they are
783   growing to inform it on relative year class strength and selectivity. These problems were more
784   severe in the SCAA-CL than the SCSA likely due to reasons such as age- vs size-based selectivity,
785   or the transition through age bins (when all old fish have the same size-at-age distribution) vs
786   transitioning through size bins.
787

References

Akselrud, C.I.A., Punt, A.E., Cronin-Fine, L., 2017. Exploring model structure uncertainty using a general stock assessment framework: The case of Pacific cod in the Eastern Bering Sea. Fish. Res., 193, 104-120.

Bertignac, M., and De Pontual, H., 2007. Consequences of bias in age estimation on assessment of the northern stock of European hake (*Merluccius merluccius*) and on management advice. ICES J. Mar. Sci., 64(5), 981-988.

Breen, P.A., Kim, S.W., and Andrew, N.L., 2003. A length-based Bayesian stock assessment model for the New Zealand abalone *Haliotis iris*. Mar. Freshw. Res. 54(5), 619–634.

Breen, P.A., Haist, V., Starr, P.J., and Kendrick, T.H., 2009. The 2008 stock assessment of rock lobsters (Jasus edwardsii) in CRA 3. New Zealand Fisheries Assessment Report 2009/23.

Brodziak, J., Ianelli, J., Lorenzen, K., and Methot Jr, R.D., 2011. Estimating natural mortality in stock assessment applications. NOAA Technical Memorandum NMFS-F/SPO, 119, 38p.

Campana, S.E., 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. J. Fish. Biol., 59(2), 197-242.

Carvalho, F., Punt, A.E., Chang, Y.J., Maunder, M.N., and Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments?. Fish. Res., 192, 28-40.

Coggins, L.G., and Quinn, T.J., 1998. A simulation study of the effects of aging error and sample size on sustained yield estimates. In Fishery Stock Assessment Models, pp. 955–975. Ed. by F. Funk, T.J. Quinn, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C-I. Chang. University of Alaska, Fairbanks, Alaska Sea Grant College Program Report No. AK-SG-98-01.

Cowles, M.K. and Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. J. Am. Stat. Assoc., 91(434), 883-904.

Deriso, R.B., Quinn Ii, T.J., and Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci., 42(4), 815-824.

Ebener, M.P., Stockwell, J.D., Yule, D.L., Gorman, O.T., Hrabik, T.R., Kinnunen, R.E., Mattes, W.P., Oyadomari, J.K., Schreiner, D.R., Geving, S., Scribner, K., Schram, S.T., Seider, M.J., Sitar, S.P., 2008. Status of cisco (Coregonus artedi) in Lake Superior during 1970-2006 and management and research considerations. Gt. Lakes Fish. Comm., Lake Superior Tech. Rep, (1), 126.

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

Fournier, D., and Archibald, C.P., 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci., 39(8), 1195-1207.

Fournier, D.A., Sibert, J.R., Majkowski, J., and Hampton, J., 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (Thunnus maccoyii). Can. J. Fish. Aquat. Sci., 47(2), 301–317.

Fournier, D.A., Hampton, J., and Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, Thunnus alalunga. Can. J. Fish. Aquat. Sci., 55(9), 2105-2116.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. and Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods. Softw., 27(2), 233-249.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci., 68, 1124–1138.

Fu, D., and McKenzie, A., 2010b. The 2010 stock assessment of paua (Haliotis iris) for Milford, George, Central, and Dusky in PAU 5A. New Zealand Fisheries Assessment Report 2010/46.

Geweke, J., 1991. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (ed JM Bernado, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK

Gilbert, D.J., Davies, N.M., and McKenzie, J.R., 2006. Development of an age–length structured model of the Hauraki Gulf–Bay of Plenty snapper (Pagrus auratus) population. Mar. Freshwater. Res., 57(5), 553-568.

Hulson, P.J.F., Miller, S.E., Quinn, T.J., Marty, G.D., Moffitt, S.D., and Funk, F., 2007. Data conflicts in fishery models: incorporating hydroacoustic data into the Prince William Sound Pacific herring assessment model. ICES J. Mar. Sci., 65(1), 25-43.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci., 72(1), 99-110.

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

McGarvey, R., Feenstra, J.E., and Ye, Q., 2007. Modeling fish numbers dynamically by age and length: partitioning cohorts into" slices". Can. J. Fish. Aquat. Sci., 64(9), 1157-1173.

Methot, R.D., 2000. Technical description of the Stock Synthesis assessment program. NOAA Tech Memo. NMFS-NWFSC-43.

Methot, R.D., and Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res., 142, 86-99.

Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci., 56(4), 473-488.

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 J. Mar. Sci., 70(1), 16-33.

Punt, A.E., Haddon, M., and McGarvey, R., 2016. Estimating growth within size-structured fishery stock assessments: what is the state of the art and what does the future look like?. Fish. Res., 180, 147-160.

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

Quinn, T.J., and Deriso, R.B., 1999. *Quantitative fish dynamics*. Oxford University Press.

Reeves, S.A., 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. ICES J. Mar. Sci., 60(2), 314-328.

Sippel, T., Lee, H.H., Piner, K., and Teo, S.L., 2016. Searching for M: Is there more information about natural mortality in stock assessments than we realize? Fish. Res., 192, 135-140.

Starr, P.J., Bentley, N., and Maunder, M.N., 1999. Assessment of the NSN and NSS stocks of red rock lobster (Jasus edwardsii) for 1998. New Zealand Fisheries Assessment Research Document 99/34. Ministry of Fisheries, Wellington, New Zealand.

878 Starr, P.J., Breen, P.A., Kendrick, T.H., and Haist, V., 2009. Model and data used for the 2008
879   stock assessment of rock lobsters (Jasus edwardsii) in CRA 3. New Zealand Fisheries
880   Assessment Report 2009/22.
881 Stockwell, J.D., Ebener, M.P., Black, J.A., Gorman, O.T., Hrabik, T.R., Kinnunen, R.E., Mattes,
882   W.P., Oyadomari, J.K., Schram, S.T., Schreiner, D.R., Seider, M.J., Sitar, S.P., Yule, D.L.,
883   2009. A synthesis of cisco recovery in Lake Superior: implications for native fish rehabilitation
884   in the Laurentian Great Lakes. N. Am. J. Fish. Manag., 29, 626–652.
885 Sullivan, P.J., Lai, H.L., and Gallucci, V.F., 1990. A catch-at-length analysis that incorporates a
886   stochastic model of growth. Can. J. Fish. Aquat. Sci., 47(1), 184-198.
887 Then, A.Y., Hoenig, J.M., Hall, N.G., and Hewitt, D.A., 2014. Evaluating the predictive
888   performance of empirical estimators of natural mortality rate using information on over 200
889   fish species. ICES J. Mar. Sci., 72(1), 82-92.
890 TeWinkel, L.M., Kroeff, T., Fleischer, G.W., and Toneys, M., 2002. Population dynamics of
891   bloaters (Coregonus hoyi) in Lake Michigan, 1973–1998. Adv. Limnol., 57, 307–320.
892 Thompson, G.G., A'Mar, Z.T., Palsson, W.A., 2011. Assessment of the Pacific cod stock in the
893   Gulf of Alaska. In: Plan Team for Groundfish Fisheries of the Gulf of Alaska (compiler), Stock
894   Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska.
895   North Pacific Fishery Management Council, 605 W. 4th Avenue Suite 306, Anchorage, AK
896   99501, pp.161–306.
897 Thorson, J.T., Jensen, O.P., and Zipkin, E.F., 2014. How variable is recruitment for exploited
898   marine fishes? A hierarchical model for testing life history theory. Can. J. Fish. Aquat. Sci.,
899   71(7), 973-983.
900 Vehtari, A., Gelman, A., and Gabry, J., 2017. Practical Bayesian model evaluation using leave-
901   one-out cross-validation and WAIC. Stat. Comput., 27(5), 1413-1432.
902 Watanabe, S., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable
903   information criterion in singular learning theory. J Mach Learn Res. 11, 3571-3594.
904 Yule, D.L., Stockwell, J.D., Cholwek, G.A., Evrard, L.M., Schram, S., Seider, M., and Symbal,
905   M., 2006. Evaluation of methods to estimate lake herring spawner abundance in Lake Superior.
906   Trans. Am. Fish. Soc., 135(3), 680-694.
907 Yule, D.L., Stockwell, J.D., Black, J.A., Cullis, K.I., Cholwek, G.A., and Myers, J.T., 2008. How
908   systematic age underestimation can impede understanding of fish population dynamics:
909   lessons learned from a Lake Superior cisco stock. Trans. Am. Fish. Soc., 137, 481–495.
910 Yule, D.L., Schreiner, D.R., Addison, P.A., Seider, M.J., Evrard, L.M., Geving, S.A., and Quinlan,
911   H.R., 2012. Repeat surveys of spawning cisco (Coregonus artedi) in western Lake Superior:
912   timing, distribution and composition of spawning stocks. Adv. Limnol., 63, 65-87.
913

914 Table 1. Data source years for each assessment and input sample sizes for each year of composition
915 data. An X denotes that a data source was available for that year in each assessment. The word
916 "Composition" refers to both age and length composition data.

Year	Hydroacoustic Survey	Fishery Harvest	Fishery Composition	MWT Survey Composition	Gillnet Survey Composition
1999		X	X – 860		
2000		X	X – 3241		
2001		X	X – 1221		
2002		X	X – 1147		
2003		X	X – 1208		
2004		X	X – 1091		
2005	X	X	X – 661	X – 794	
2006		X	X – 644		
2007	X	X	X – 839	X – 1845	
2008	X	X	X – 654	X – 559	
2009	X	X	X – 638	X – 994	X – 302
2010	X	X	X – 500	X – 520	
2011	X	X	X – 563		
2012	X	X	X – 478		
2013	X	X	X – 429		X – 678
2014	X	X	X – 733		X – 135
2015	X	X	X – 705	X – 478	X – 824

917

918

919 Table 2. Posterior point estimates (medians) and standard deviations in parentheses for parameters
920 in each assessment model. Priors specified are presented in column 4. Fixed parameters are in
921 bold. NA indicates a parameter that was not used for a given model (e.g., aging error for SCSA).
922 Parameters excluded from this table include recruitment deviations, sex- and age-specific initial
923 abundance parameters for the SCAA, and fishing intensities.

| Parameter	SCAA	SCSA	Priors
Male Natural Mortality - M_m (log scale)	-1.255	**-1.261**	N(-1.26, 0.44)
Female Natural Mortality - M_f (log scale)	-1.374	**-1.362**	N(-1.36, 0.44)
Median Recruitment - μ (log scale)	10.92 (0.85)	13.52 (0.40)	U(5, 20)
Fishery Harvest SD - σ_f (log scale)	-2.40 (0.16)	-2.53 (0.15)	U(-5, 5)
Initial Size Distribution Gamma Shape - α (log scale)	NA	2.95 (0.31)	U(-10, 5)
Initial Size Distribution Gamma Rate - β (log scale)	NA	-2.34 (0.30)	U(-5, 10)
Male Initial Abundance Scalar - ψ_m (log scale)	NA	13.61 (0.47)	U(1, 25)
Female Initial Abundance Scalar - ψ_f (log scale)	NA	15.26 (0.22)	U(1, 25)
Fishery Selectivity Parameter 1 - α (log scale)	1.88 (0.07)	4.80 (0.11)	U(-10, 5)
Fishery Selectivity Parameter 2 - β	0.65 (0.07)	4.50 (0.54)	U(-5, 10)
MWT Survey Selectivity p1 - α_{MWT} (log scale)	0.56 (0.38)	2.25 (0.23)	U(-1.5, 5)
MWT Survey Selectivity p2 - β_{MWT}	0.11 (0.09)	0.50 (0.13)	U(-5, 10)
Gillnet Selectivity p1 - α_{GN} (log scale)	0.42 (0.75)	2.96 (0.54)	U(-1.5, 5)
Gillnet Selectivity p2 - β_{GN}	0.21 (0.15)	0.99 (0.46)	U(-5, 10)
L-infinity - L_∞ (log scale)	NA	6.06 (0.01)	U(5.9, 6.5)
Brody Growth Coefficient - K (log scale)	NA	-1.28 (0.05)	U(-5, 1)
Growth Variance Intercept - a (log scale)	NA	4.83 (1.10)	U(-5, 10)
Growth Variance Slope - b (log scale)	NA	0.99 (0.63)	U(-1, 5)
Aging Error Intercept - c	**0**	NA	U(0, 10)
Aging Error Slope - d	1.001 (0.01)	NA	U(0.5, 1.5)
Aging Error CV Intercept - e	0.018 (0.01)	NA	U(0, 0.5)
Aging Error CV Slope - f	**0**	NA	U(0, 10)
Hydroacoustic Calibration Coefficient - γ (log scale)	**0**	**0**	NA

924

925

926 Table 3. Negative log-likelihood (NLL) and standard deviation of normalized residuals (SDNR)
927 values for common data sources.
928

	Male Yield	Female Yield	Hydroacoustic Data
SCAA – NLL	-32.98	-33.00	0.32
SCSA – NLL	-35.17	-35.23	2.03
SCAA – SDNR	0.09	0.12	1.26
SCSA – SDNR	0.05	0.04	1.51

929

930

931
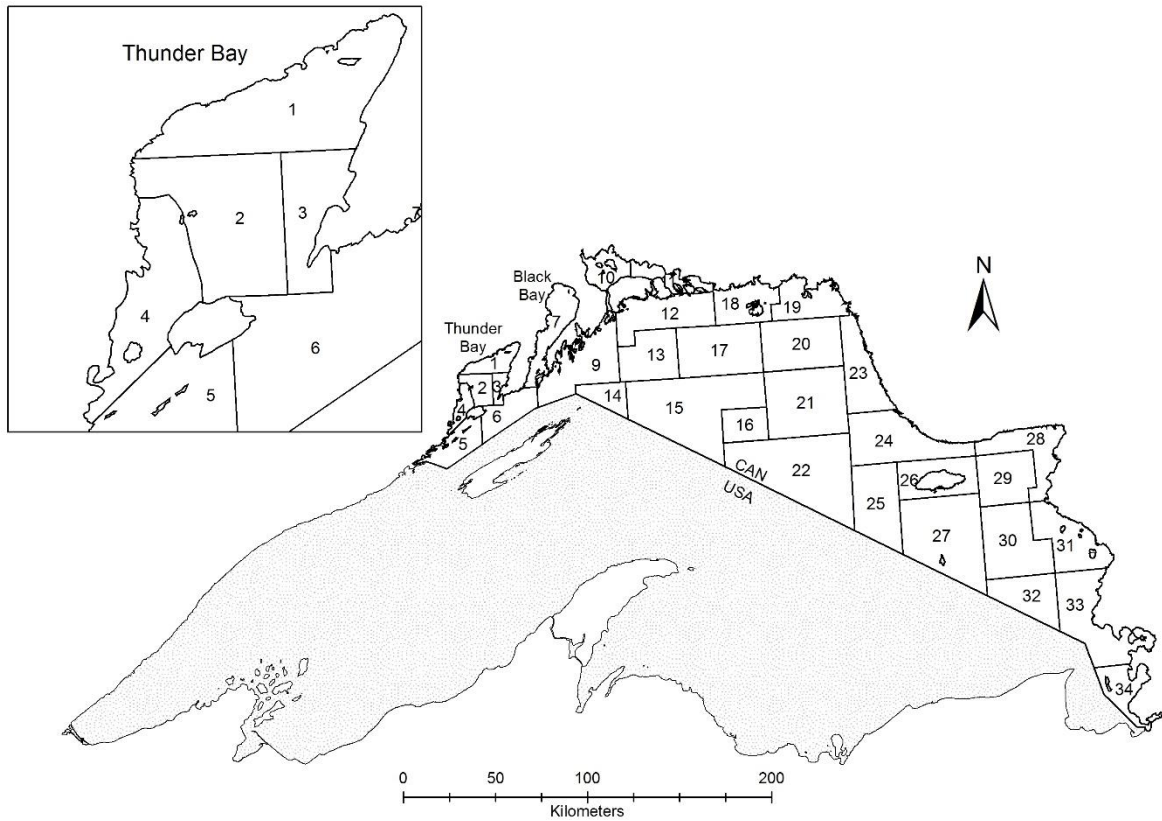
Figure 1. OMNRF Quota Management Areas (QMAs) pre-2016. We are characterizing Thunder
Bay stock area as QMAs 1-4 which are shown zoomed in at the top left.

934

935

936

937

938

939
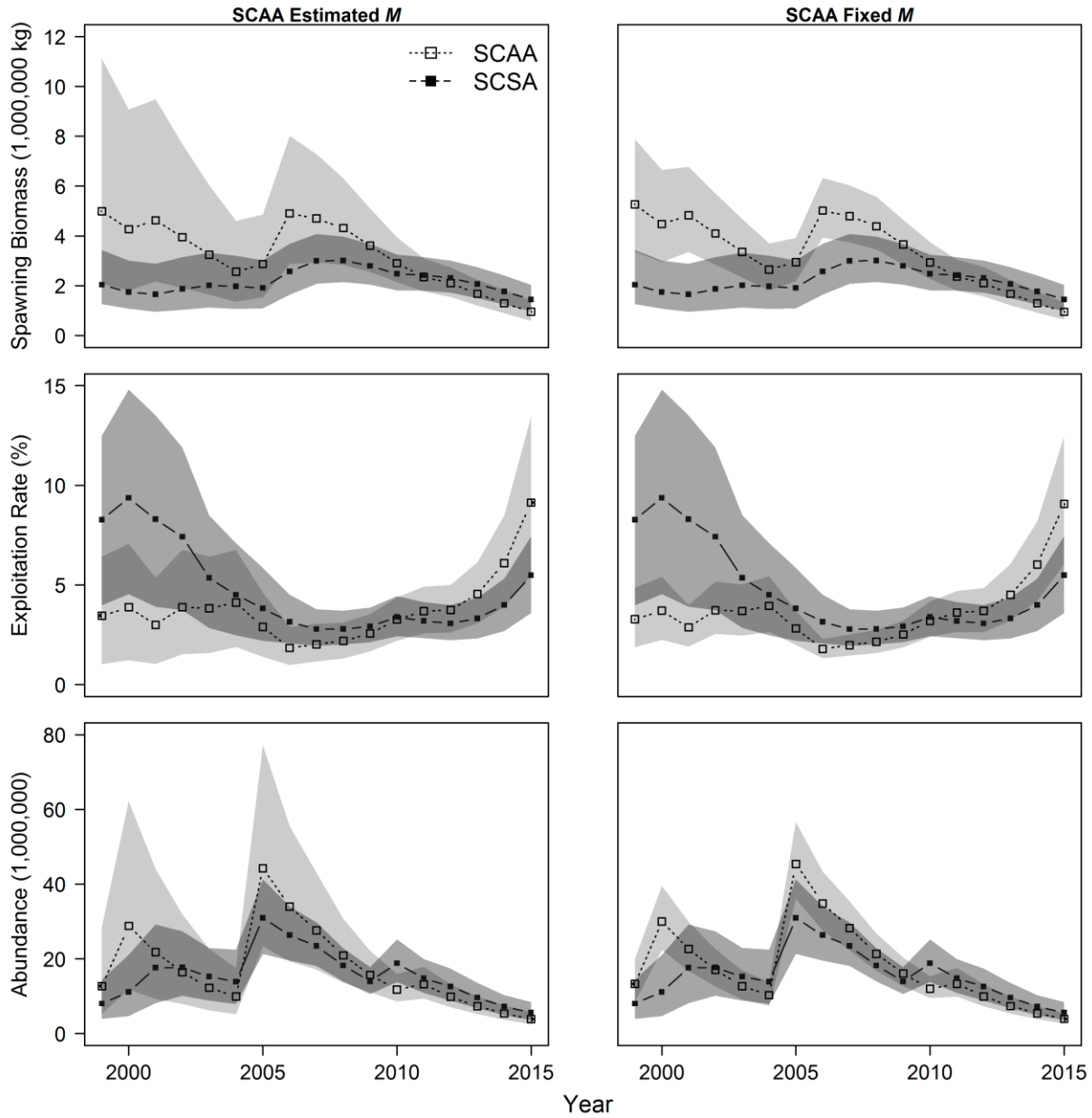
Figure 2. Spawning biomass in millions of kg of mature females (>250 mm), exploitation rate (harvest/biomass of fish > 250 mm), and abundance in millions of fish for the SCAA and the SCSA. First column represents the original assessment models where M is estimated in the SCAA (however not in the SCSA). The second column represents assessment results where M is fixed (assumed known) in both the SCAA and the SCSA. SCSA results are identical in each column. Shaded regions denote 95% HPD intervals and dashed lines or points are medians of the posterior distribution. Hollow squares denote SCAA output while filled squares denote SCSA output. Light shading denotes the HPD for the SCAA, darker shading denotes the HPD for the SCSA, and the darkest shading is where the two intervals overlap.

29

Figure 3. Recruitment for the SCAA and SCSA. Both represent the number of fish entering the model in a given year, but for the SCAA it is the number of age 2 fish entering the population and in the SCSA it is the number of fish greater than 170 mm that are entering the population. Points denote medians of the posterior distribution and error bars are the 95% HPD intervals.

958

959
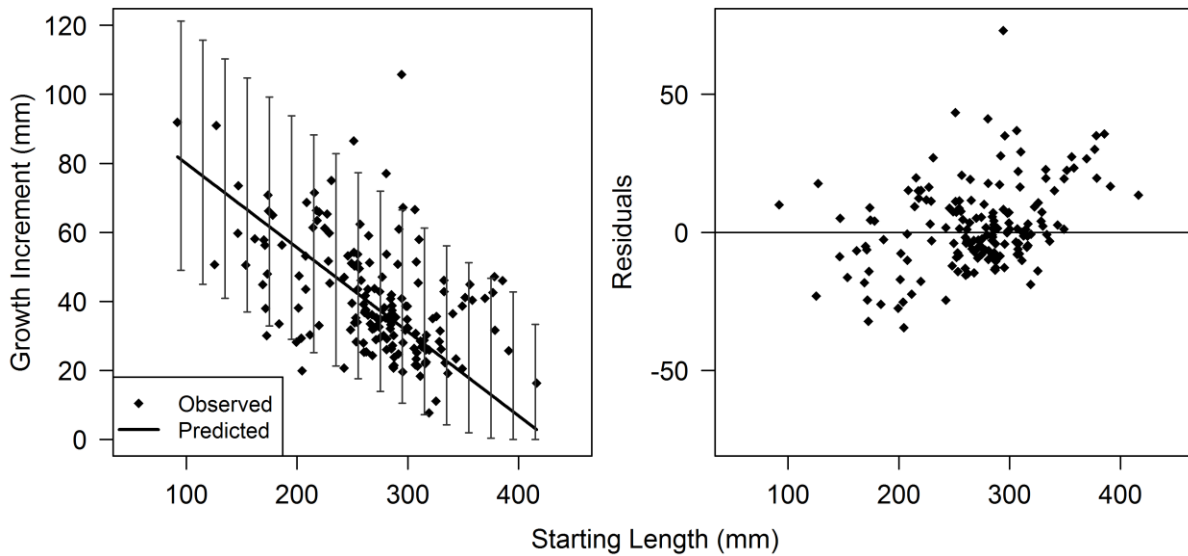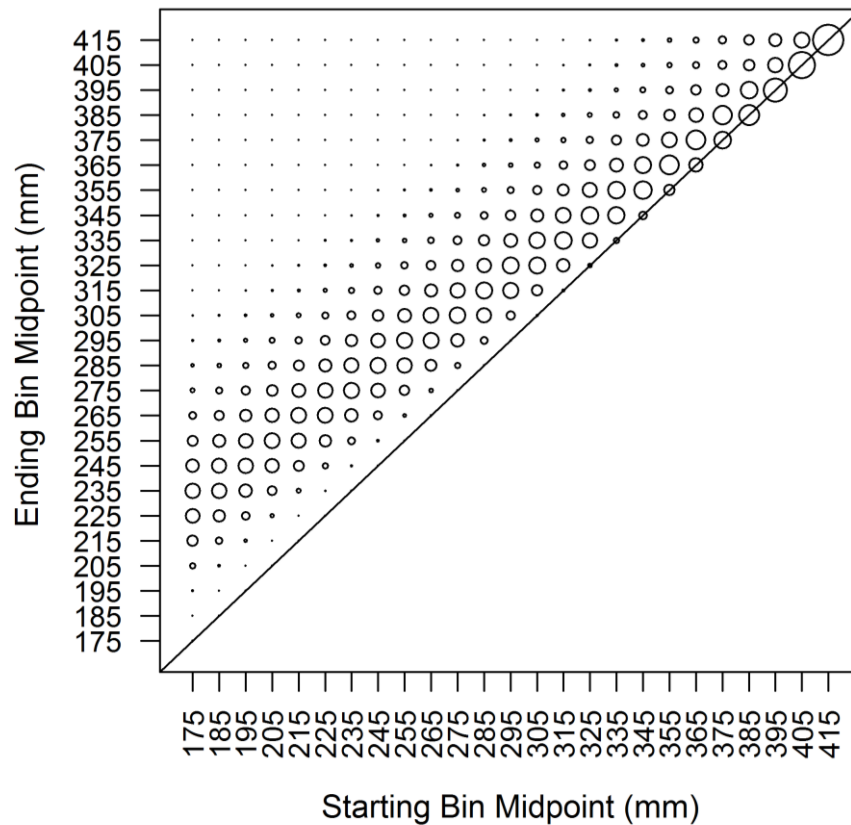
Figure 4. Upper left: Fit to growth increment data. Black line depicts the median of the posterior
distribution of the expected growth increment ($E(\Delta_l)$) and arrows represent middle 95% quantiles
of gamma distributions derived using point estimates of parameters a, b, and $E(\Delta_l)$ that define
the variance about an expected growth increment. Upper right: Residuals, medians of the posterior

964 distribution, from fit to growth increment data. Lower panel: Growth transition matrix at the
965 posterior medians for growth parameters. Note that the area of the circles represent the probability
966 of growing into a length bin given a starting length bin. Length bins are represented on axes as
967 midpoints. Plus group is length bin 410-420 mm.

968
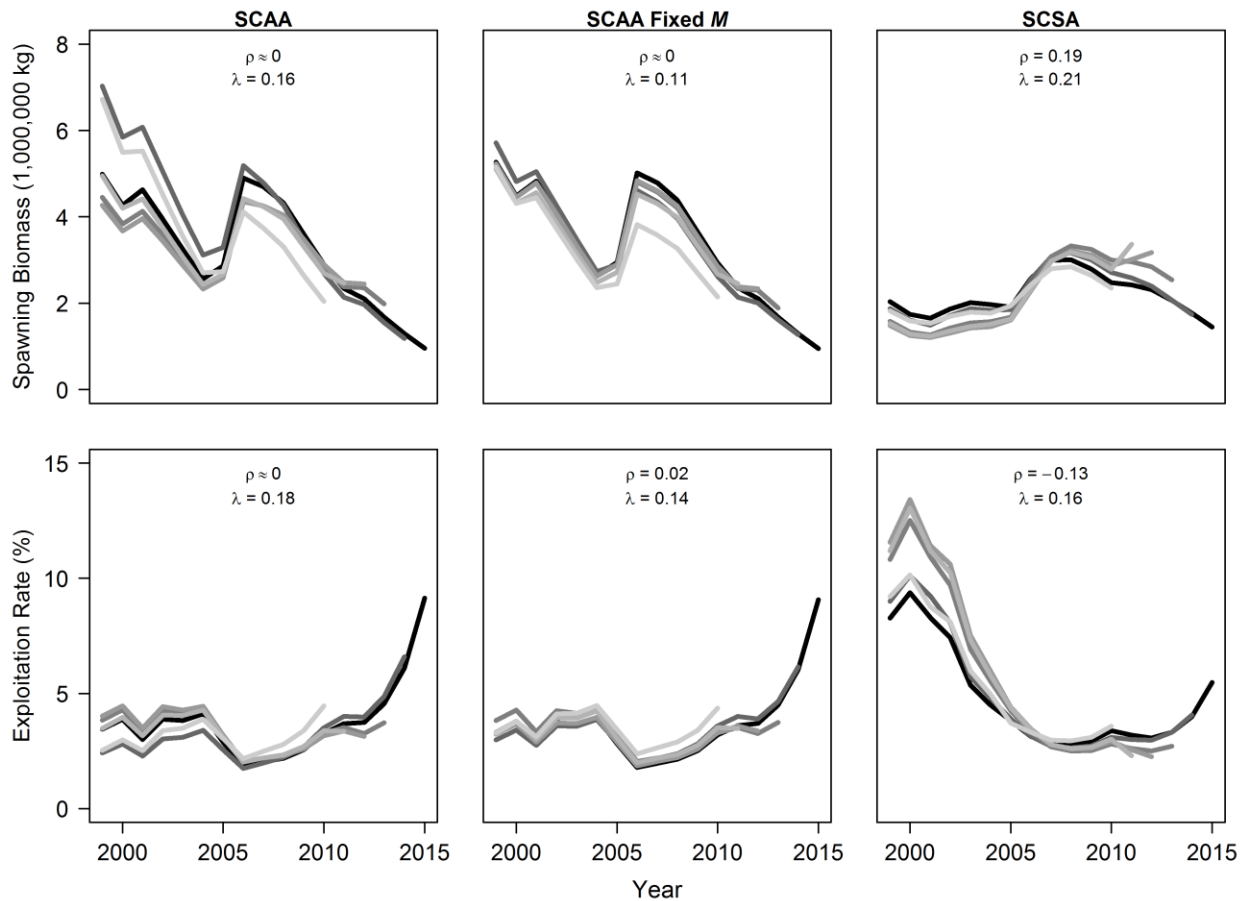
969

970 Figure 5. Retrospective analyses for spawning biomass and exploitation rate. Shown are times
971 series estimates of spawning biomass and exploitation rate when five terminal years of data are
972 sequentially dropped from each assessment. Each column depicts a different assessment model.
973 The first column depicts retrospective results from the SCAA, the second depicts the SCAA with
974 fixed, assumed known natural mortality values, and the third depicts the SCSA. Spawning biomass
975 is reported as millions of kg of mature females and exploitation rate as yield/biomass of fish > 250
976 mm. Shown on each panel is the corresponding value of Mohn's rho (ρ) and the precision
977 indicator (λ).

978

979
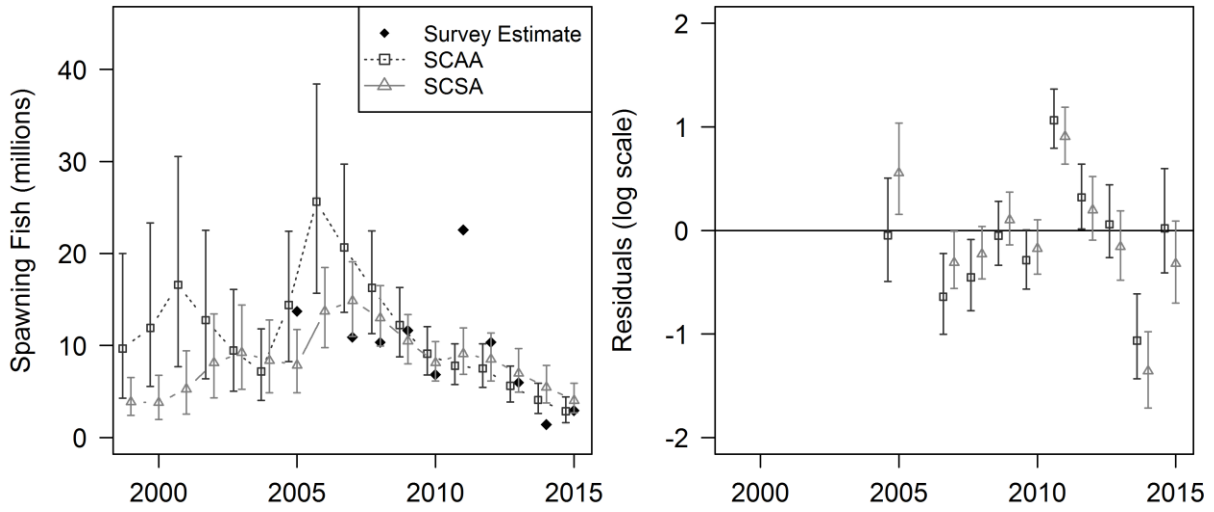
980
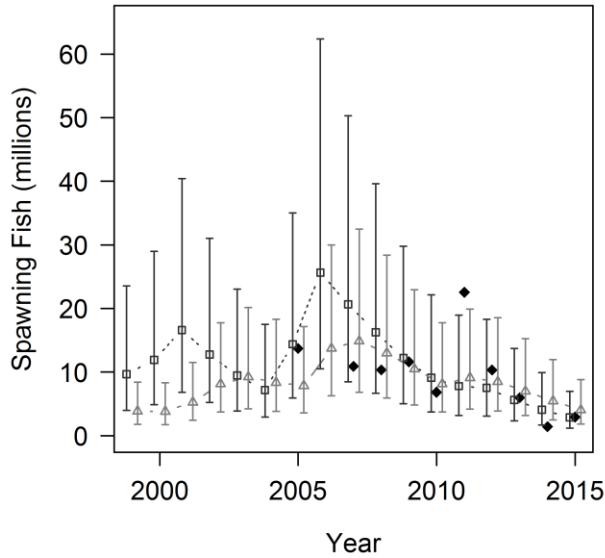
981

982 Figure 6. Fit to hydroacoustic estimates of spawning abundance. No data from 1999-2004 and in
983 2006. Spawning fish is reported as millions of fish. For the upper row, points denote medians of
984 the posterior distribution and error bars are 95% HPD intervals. For the bottom plot, points denote
985 medians of the posterior distribution and error bars are sampling intervals (95%) for the data
986 calculated at the posterior median of the standard deviation σ for hydroacoustic data. Filled
987 diamonds represent the hydroacoustic survey estimates of spawning abundance, used as observed
988 data in the models. Hollow squares and triangles represent the SCAA and SCSA fit to the
989 hydroacoustic survey data set, respectively. Note SCAA and SCSA points are offset to increase
990 visibility.

991

992