This is to certify that the
dissertation entitled

MODEL SELECTION AND DATA WEIGHTING METHODS FOR STATISTICAL CATCH-AT-AGE ANALYSIS: APPLICATION TO 1836 TREATY WATER STOCK ASSESSMENTS

presented by

Brian C. Linton

has been accepted towards fulfillment
of the requirements for the

Doctoral	degree in	Department of Fisheries and Wildlife and Program in Ecology, Evolutionary Biology, and Behavior

James R. Bence
Major Professor's Signature

Aug 1, 2007
Date

MSU is an affirmative-action, equal-opportunity employer

MODEL SELECTION AND DATA WEIGHTING METHODS FOR STATISTICAL CATCH-AT-AGE ANALYSIS: APPLICATION TO 1836 TREATY WATER STOCK ASSESSMENTS

By

Brian C. Linton

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Fisheries and Wildlife and Program in Ecology, Evolutionary Biology, and Behavior

2007

ABSTRACT

MODEL SELECTION AND DATA WEIGHTING METHODS FOR STATISTICAL CATCH-AT-AGE ANALYSIS: APPLICATION TO 1836 TREATY WATER STOCK ASSESSMENTS

By

Brian C. Linton

Recommended harvest limits for lake trout *Salvelinus namaycush* and lake whitefish *Coregonus clupeaformis* stocks in the 1836 Treaty Waters of the Great Lakes are based on statistical catch-at-age analysis (SCAA). The assessment models and methods are similar to those used to assess fish stocks in many of the word's major fisheries. My objective was to evaluate these methods with an eye towards suggesting improvements both for 1836 treaty waters and more generally. My results provide general guidance to stock assessment scientists with regard to data weighting and selecting among alternative assessment models. As a first step, I performed an analysis of the Lake Huron lake whitefish models' sensitivity to changes in "known" inputs and model structure, selected as examples of basic type of assessment used throughout treaty waters for lake whitefish and lake trout. All of the Lake Huron lake whitefish models were sensitive to changes in the methods used to estimate recruitment and time-varying selectivity, as well as to changes in their objective functions, and this indicated that further study of these aspects of the assessment methods was warranted.

Specifically with regard to the objective function, the assessment models were sensitive to changes in pre-specified variances associated with process and observation errors, which are used to weight the different data sources. This result is consistent with concerns expressed more broadly in the literature. I evaluated alternative approaches for

estimating log catchability (process error) and log total catch (observation error) standard deviations within SCAA using Monte Carlo simulations: an *ad hoc* approach that tunes the model predicted log total catch standard deviation to match a prior value, and a Bayesian approach using either strongly or weakly informative priors for log catchability standard deviation. When process error variance is large relative to observation error (likely for many fisheries), reliable estimates of log catchability and log total catch standard deviations can be obtained in SCAA using a Bayesian approach with only a weakly informative prior on log catchability standard deviation.

The sensitivity of the Lake Huron whitefish models to the method used to model time-varying selectivity is also consistent with indications in the broader literature that SCAA assessments can be sensitive to misspecification of selectivity. I therefore evaluated four approaches for modeling time-varying selectivity within SCAA using Monte Carlo simulations: double logistic functions with one, two and all four of the function parameters varying over time, as well as age-specific selectivity parameters that all varied over time. None of these estimation methods out performed the others in all cases. In addition, I compared model selection methods to identify good (i.e., accurately matching the true fish population) estimation models. Degree of retrospectivity, the best selection method, was based on a retrospective analysis of bias in model parameter estimates as the data time series for estimation is sequentially shortened. I recommend this method of model section when considering different time-varying selectivity estimation approaches in SCAA.

ACKNOWLEDGMENTS

I would like to thank everyone who made this dissertation possible. My advisor and graduate committee chairman, Dr. James Bence, who gave me the opportunity to work on this project and taught me all about fish population dynamics and stock assessment. The rest of my graduate committee including: Dr. Michael Jones, Dr. Daniel Hayes, and Dr. Guilherme Rosa for their constructive criticism concerning my research and this dissertation. The staff and students of the Quantitative Fisheries Center provided both technical and moral support for this project. In particular, Dr. Michael Wilberg helped me get my first Monte Carlo simulation study up and running. The Modeling Subcommittee of the 1836 Treaty Waters provided the initial impetus for this work, as well as providing technical knowledge on the lake trout and lake whitefish stock assessments for the 1836 Treaty Waters. The Michigan Department of Natural Resources and U.S. Fish and Wildlife Service provided the funding for this project.

I would also like to thank my family and friends for their mental, emotional, and spiritual support over the last five years. Most notably, I would like to thank my wife, Jennifer Linton, who I had the good fortune to meet and marry during my time here. In addition, I thank my parents, Clifford and Irene Linton, and my brothers, Eric and John Linton, for supporting me in this venture, as in every undertaking I have attempted in life. Special thanks also go to all my friends at Martin Luther Chapel and the Graduate Intervarsity Christian Fellowship chapter. Lastly, and in the chief place, I thank God for blessing me with the desire and the ability needed to pursue this course. Thank you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

SENSITIVITY ANALYSIS OF LAKE WHITEFISH STOCK ASSESSMENT MODELS USED IN THE 1836 TREATY WATERS OF LAKE HURON

Introduction

In 1836, Native American Bands in the region to become the state of Michigan signed a treaty with the U.S. government which reserved their right to fish in the Michigan waters of lakes Huron, Michigan, and Superior. These fishing rights were reaffirmed by the U.S. federal courts in 1979. The federal district court later approved the fishery regulations created by the Chippewa/Ottawa Treaty Fishery Management Authority (COTFMA) in 1982, while mandating that total allowable catches (TACs) or harvest regulating guidelines (HRGs) be established for important fish species in order to prevent over-fishing. Federal, state, and tribal biologists worked together to estimate TACs for lake whitefish *Coregonus clupeaformis* during 1979-1982. During this period, the stock assessment methods used in the treaty waters were evolving and constrained by limited data. Where possible stock sizes were estimated by application of a simple age-structured model. Although there was no formal harvest policy, TACs were generally set near the estimated maximum sustainable yield if the stock size was near the associated biomass and to lower values when stock sizes were lower (e.g., AHWG 1979).

The 1985 Consent Decree laid out a 15 year agreement between federal, state and tribal agencies for the allocation of fishery harvest between the parties. The Technical Fisheries Review Committee (TFRC) was created by the decree to assess stocks of important fish species. As part of this mandate, the TFRC recommended TAC/HRGs for

lake whitefish stocks within the ceded territory to federal, state and tribal governments. Stock assessments produced for the TFRC were generally based on simple age-structured models (Clark and Smith 1984). The 1985 decree did not specify a harvest policy, but based on TTWG (1984) the TFRC adopted a policy to limit total mortality to specified levels less than 70%.

The 2000 Consent Decree was a new 20 year agreement, which set guidelines for the management of important fish species, as well as allocating fishery harvest. As part of the new decree, the Technical Fisheries Committee (TFC) was formed, which serves many of the same functions as did the TFRC under the previous decree. Also at this time, COTFMA was reorganized as the Chippewa/Ottawa Resource Authority (CORA). Unlike the previous decree, a reference mortality rate for lake whitefish of 65% was specified, which partially defines a harvest policy. New methods for conducting lake whitefish stock assessments and projecting TAC/HRGs were developed during the negotiation period for the 2000 Consent Decree by an interagency modeling group. The decree specifies that a newly formalized Modeling Subcommittee (MSC) of the TFC should build upon the work of the interagency modeling group to continue the lake whitefish stock assessment program.

The new stock assessment methods employed statistical catch-at-age models, which were created for each lake whitefish stock by the interagency modeling group and further developed by the MSC. These stock assessment models used catch-at-age and effort data from the commercial fisheries to estimate population abundances, mortality rates, fishery harvests, and other population parameters of interest. Estimated quantities from the assessment models were used to project each stock's abundance and mortality

rates into the future, and then TAC/HRGs were calculated from these projections and a reference mortality rate.

The 2000 Consent Decree established requirements governing the calculation of TAC/HRGs. The reference level of total annual mortality (65%) specified for lake whitefish plays a different role depending on whether the yield from a particular management unit is allocated entirely to the tribes (tribal unit) or partially allocated to the state (shared unit). For shared units, 65% total mortality is treated as an upper limit and TACs are established so as to allocate the yield between the parties as specified in the decree. State and tribal management agencies are responsible for separately implementing management actions (e.g., limits on entry to the fishery, gear restrictions, size limits, and trip limits) to constrain fishery yield at or below levels specified by TACs. If state or tribal fishery harvest exceeded their TAC/HRG by 25% or more, either in a single year or over the course of five years, then that party's TAC in the following year is reduced by the amount that the previous TAC was exceeded. For tribal units, 65% total mortality is viewed as an upper target level, and management actions by the tribes are intended to prevent this level from being exceeded on average.

One of the complications of applying a reference mortality rate to the results of the new age-structured assessment models is that these models account for the fact that fishing mortality varies with age. The MSC chose a conservative solution to the problem for lake whitefish by further defining the reference mortality rate. First, for the reference mortality rate, the maximum total mortality across all ages was not to exceed the specified value of 65% (for most units). In addition, the spawning stock biomass per recruit (SSBR) at this mortality schedule was required to be at least 20% of the SSBR for

the unfished stock. If the SSBR was below the 20% threshold, then the maximum total mortality was reduced until the resulting SSBR was at least 20% of the unfished SSBR.

Due to the rapid development and implementation of the stock assessment models, not all of the approaches used in the models have been fully evaluated. For example, there were numerous methods for modeling each of the biological processes represented within the models from which the MSC analysts could select. Once a particular method for modeling a process was chosen, reasonable parameter starting values and bounds on what values those parameters could take also had to be selected by the analysts. It was unknown how much these choices affected stock assessment results. Therefore, my objective was to further evaluate the stock assessment models for lake whitefish in the 1836 treaty waters of Lake Huron, with a view toward suggesting possible improvements. This objective linked to a broader goal for my work, to form the basis for advice that is broadly applicable in the field of fishery stock assessment. As a first step to achieve this objective, I performed a general analysis of the models' sensitivity to changes in "known" inputs and model structure.

Methods

The 1836 treaty waters of Lake Huron were divided into five lake whitefish management units, each thought to contain a distinct lake whitefish stock (Figure 1.1). Separate stock assessment models were developed for each of the lake whitefish management units. When the models were originally developed, it was assumed that the net movement of lake whitefish between management units was nil.

Stock Assessment Model

Here I provide an overview of the stock assessment models' general structure. Ebener et al. (2005) provides a detailed description of the models. All of the stock assessment models consisted of two basic submodels, a population submodel and an observation submodel. The population submodel described the population dynamics of the stock in terms of abundance-at-age excluding the first year and the first age in subsequent years:

$$N_{a+1,y+1} = N_{a,y}e^{-Z_{a,y}},$$

where $N_{a,y}$ was the number of fish in age a and year y and $Z_{a,y}$ was the total instantaneous mortality rate in age a and year y. Numbers-at-age in the first year were estimated as a vector of relative population variation parameters (i.e. a vector of deviations that must sum to zero). A population scaling parameter then converted these deviations to numbers-at-age. Numbers of fish in the first age of each year also were estimated as a series of scaled deviations using the same population scaling parameter, but were penalized for deviating too greatly from a Ricker stock-recruitment function:

$$N_{a_0,y} = \alpha G_{y-(a_0-1)} - e^{-\beta G_{y-a_0-1}},$$

where $N_{a_0,y}$ was the number of fish in the first age a_0 and year y, $G_{y-(a_0-1)}$ was the number of eggs produced a_0-1 years prior to year y, α was the productivity parameter, and β was the density dependent parameter. The number of eggs was calculated within the submodel, based on a constant weight-specific fecundity. The productivity and density dependent parameters were estimated within the submodel. Numbers-at-age were

converted to biomass using observed mean weight-at-age data. Total mortality consisted of four component parts:

$$Z_{a,y} = M + M_{L,a,y} + F_{G,a,y} + F_{T,a,y},$$

where M was the natural mortality rate, $M_{L,a,y}$ was the sea lamprey induced mortality rate in age a and year y, $F_{G,a,y}$ was the gill net fishing mortality rate in age a and year y, and $F_{T,a,y}$ was the trap net fishing mortality rate in age a and year y. Natural mortality was assumed to be constant for all ages and years, and was estimated as a model parameter. Pauly's equation (Pauly 1980) was used to calculate an initial value for the natural mortality parameter to provide a reasonable starting point for parameter estimation. Sea lamprey mortality was calculated externally to the model based on observed sea lamprey wounding rates. Fishing mortality was estimated by relaxing the assumptions of the fully separable fishing mortality model and allowing gear selectivity to vary with time:

$$F_{i,a,y} = S_{i,a,y} q_i E_{i,y} \zeta_{i,y},$$

where $S_{i,a,y}$ was the gear selectivity of age a fish in fishery i and year y, q_i was the catchability in fishery i, $E_{i,y}$ was the observed fishing effort in fishery i and year y, and $\zeta_{i,y}$ was the deviation in fishing mortality from direct proportionality to observed fishing effort in fishery i and year y. Selectivity was estimated with a double logistic function of age, and one of the parameters of the function was allowed to change with time according to a quadratic function. This allowed age-specific selectivity to change gradually over time. An adjustment factor was applied to the observed gill net effort in order to account for changes in the number of meshes deep that were set through time.

The observation submodel predicted catch-at-age for the gill net and trap net fisheries. Catch-at-age was predicted using Baranov's catch equation:

$$C_{i,a,y} = \frac{F_{i,a,y}}{Z_{a,y}} N_{a.y}\left(1 - e^{-Z_{a,y}}\right),$$

where $C_{i,a,y}$ was the number of age a fish caught in fishery i during year y, and all of the other parameters were estimated in the population submodel. Predicted catch-at-age was converted to a total annual catch and a proportion of catch-at-age for each fishery. An underreporting factor, representing the proportion of the actual catch that was reported, was applied to the total catch in order to account for underreporting and discards in the fisheries. The underreporting factor was obtained by comparing reported fishery landings to actual sales.

The parameter values providing the best fit were found using Bayesian methods (i.e., prior densities were assigned to all parameters). In particular, best fit parameter estimates maximized the joint posterior density, and for numerical reasons this was done by finding parameter values that minimized the weighted sum of the negative log likelihoods and the negative log prior densities. Separate likelihood components were calculated for gill net total catch, gill net proportion of catch-at-age, trap net total catch, and trap net proportion of catch-at-age. Total annual catch was assumed to follow a lognormal distribution, with the negative log likelihood (ignoring some additive constants) given by:

$$L(C_i) = \frac{1}{2\sigma_i^2} \sum_{y=1}^{n} \left[\left(\ln \frac{C_{i,y}}{\hat{C}_{i,y}} \right)^2 \right] + n \ln \sigma_i ,$$

where σ_i was the standard deviation for log-scale observed total catch in fishery i, $C_{i,y}$ was observed total numbers of fish caught in fishery i and year y, $\hat{C}_{i,y}$ was predicted total numbers of fish caught in fishery i and year y, and n was the total number of years included in the model. Observed catch was reported as weight of fish harvested, which was converted to numbers of fish using the observed mean weight of a harvested fish. Proportion of catch-at-age was assumed to follow a multinomial distribution, with the negative log likelihood (ignoring some additive constants) expressed as:

$$L(P_i) = -\sum_{y=1}^{n} N_{E,i,y} \sum_{a=1}^{m} \left(P_{i,y,a} \ln \hat{P}_{i,y,a} \right),$$

where $N_{E,i,y}$ was the effective number of fish used to calculate the age composition in fishery i and year y (Fournier and Archibald 1982), $P_{i,y,a}$ was the observed proportion of catch-at-age a in fishery i and year y, $\hat{P}_{i,y,a}$ was the predicted proportion of catch-at-age a in fishery i and year y, n was the total number of years included in the model, and m was the total number of ages included in the model. In addition to the likelihood components, the joint posterior density included terms related to prior densities for the model parameters. First, deviations of predicted recruitments from the Ricker stock-recruitment function were assumed to follow a lognormal distribution. Second, deviation of predicted natural mortality from the prior natural mortality value (i.e. the Pauly's equation value) was assumed to follow a lognormal distribution. Third, deviations in the fishing mortality from direct proportionality to observed fishing effort were assumed to follow a lognormal distribution. The log of all remaining model parameters were assigned proper uniform prior densities, which follows common practice with the intent

of being weakly informative. Therefore, prior densities of the log of the remaining parameters were constants for all parameter values.

Each likelihood component, the prior density for deviations between recruitment and the stock-recruitment function predictions, the prior density for natural mortality, and the prior density for deviations in the fishing mortality from direct proportionality to observed fishing effort were weighted by an emphasis factor as described by Methot (1990). If all likelihood components, prior densities, and their associated standard deviations or effective sample sizes were correctly specified, then the emphasis factors should all be 1.0. If there was a misspecification in the objective function, then the emphasis factors provide a simple way for analysts to adjust how closely the model attempts to fit observed and predicted data for each likelihood component.

Projection Model

Recommended yields for a reference (sometimes called target) mortality rate were then calculated using stock assessment model output in a projection model. The stock assessment model output included estimated numbers-at-age, estimated total mortality, estimated natural mortality, and assumed sea lamprey mortality, all from the last year of the model, as well as, estimated trap net and gill net mortality rates that were averaged over the last three years of the model, and estimated average recruitment (over the last ten years) of the model. Along with the stock assessment model output, observed weight-at-age in the fisheries, observed mean proportion of females in the population, observed maturity schedules represented as year and age-specific proportions, and observed time of year of spawning represented as a proportion of the year were also used in the projection model for SSBR calculations.

The projection model took the abundance-at-age estimates from the beginning of the last year of the stock assessment model, projected abundance to the beginning of the year for which recommended yields were desired, then projected yields for the trap net and gill net fisheries. Trap net and gill net fishery multiplier parameters were used to adjust age-specific fishing mortality rates by the same proportion for each age. The values of the two multipliers were set so as to achieve the reference mortality rate, while maintaining a desired allocation between trap net and gill net yield. There were two steps to determining the appropriate values for the multipliers, which corresponded to how the reference mortality was defined. First, the multipliers were adjusted so that the maximum total annual mortality for any age did not exceed the reference (typically 65%). Second, the ratio of SSBR at this mortality schedule to SSBR without fishing was calculated (hereafter the SSBR ratio). If this ratio was less than 0.2, then the multipliers were decreased until the SSBR ratio equaled 0.2.

Sensitivity Analysis

Sensitivity analysis quantifies the effect of changes made to a model's input values and underlying assumptions on the model's output (Morgan and Henrion 1990). My sensitivity analysis tested changes to the stock assessment models' input quantities and model structure (i.e., underlying model assumptions). Changes to observed input data represented possible changes in data collection (e.g., collecting more or less data), while changes to input values based on expert judgment (e.g., parameter starting values) represented a changes made by the analyst during the model fitting process. Changes in model structure were based on alternative modeling procedures suggested in the

literature. The WFH-03 management unit model was not included in the following analysis, due to a lack of convergence to a satisfactory solution.

The stock assessment models were tested for their sensitivity to changes in input values. The observed mean weight-at-age of harvested fish was varied for all ages at once by ±10% of the original values. The year- and age-specific maturity schedule was varied by reassigning maturity values from each age to the next oldest age (e.g., maturity values for age 4 fish became the maturity values for age 5 fish). Then the first age was given a maturity value of zero. Similarly, the maturity schedule was varied by reassigning maturity values from each age to the next youngest age (e.g., maturity values for age 4 fish became the maturity values for age 3 fish), and setting the maturity in the last age equal to 1.00. Fecundity was adjusted by making it a linear function of average weight-at-age at the time of spawning. The gill net adjustment factors for number of meshes deep set through time were set equal to 1.00 to test the overall effect of the adjustments. The gill net adjustment factors also were varied using the following formula, which assumed the trend in the factors over time was alternatively more and less extreme than originally thought:

$$x_y = \bar{x} + c\left(x_{0,y} - \bar{x}\right),$$

where x_y was the new adjustment factor in year y, $\bar{x}$ was the average of the original adjustment factors across all years, $x_{0,y}$ was the original adjustment factor in year y, and scalar c alternatively equaled 0.8 to represent a less extreme trend and 1.2 to represent a more extreme trend. Adjustment factors were included in the original models to account for underreporting in each year of the fisheries. The underreporting factors were set to 1.00 for one fishery at a time to test the overall effect of the adjustments. The

underreporting factors were increased and decreased by a value of 0.2 for one fishery at a time. The proportion of females in the population was set to 0.5. The proportion of females also was increased and decreased by a value of 0.2. The time of year of spawning was increased and decreased by a value of 0.2. Bounds for each model parameter, which limited the range of values a given parameter could take, were increased one at a time by decreasing the lower bound by 20% of the original value and increasing the upper bound by 20% of the original value. Bounds for each model parameter were decreased one at a time by increasing the lower bound by 20% of the original value and decreasing the upper bound by 20% of the original value. Starting values for each model parameter were increased and decreased one at a time by 20% of the original values. Natural mortality was altered by fixing the parameter to the starting value and by increasing and decreasing the starting value by 20% of the original starting value.

The stock assessment models were tested for their sensitivity to changes in model structure. Recruitment in each year was estimated as a free parameter without any penalty for deviating from stock-recruitment model predictions. Also, a Beverton-Holt stock-recruitment function, rather than a Ricker stock-recruitment function, was used to predict recruitment (Beverton and Holt 1957):

$$N_{a_0,y} = \frac{\alpha G_{y-a_0-1}}{1+\beta G_{y-a_0-1}}.$$

Rather than using deviations between observed and predicted numbers of fish caught in the objective function, deviations between observed and predicted biomass of fish caught were used. The predicted numbers of fish caught were converted to mass of fish caught, using the mass-at-age of a harvested fish, comparing them assuming a lognormal

distribution. The likelihood component emphasis factors were doubled and halved one at a time. Gamma likelihood components were substituted for all lognormal likelihood components, keeping the same coefficient of variation:

$$L(C_i) = -\phi_i \sum_{y=1}^{n} \left[\ln\left(\frac{C_{i,y}}{\hat{C}_{i,y}}\right) - \frac{C_{i,y}}{\hat{C}_{i,y}} \right],$$

where ϕ_i was the inverse of the squared coefficient of variation for observed harvest in fishery i (Cadigan and Myers 2001) and the other variables were the same as in the lognormal likelihood component. Dirichlet likelihood components were substituted for all multinomial likelihood components, with fixed parameters setting the effective sample size equal 100:

$$L(P_i) = \sum_{y=1}^{n} \left[\ln\Gamma\left(\sum_{a=1}^{m} \gamma_i \hat{P}_{i,y,a} \right) - \sum_{a=1}^{m} \ln\Gamma\left(\gamma_i \hat{P}_{i,y,a}\right) + \sum_{a=1}^{m} \left(\gamma_i \hat{P}_{i,y,a} - 1\right) \ln P_{i,y,a} \right],$$

where γ_i represented the effective sample size for fishery i, Γ was the gamma function, and the other variables were the same as in the multinomial likelihood component.

Each stock assessment model was rerun for each of the changes tested. In order to better specify the standard deviation around the stock recruitment relationship, an initial recruitment standard deviation was input into the model. The standard deviation of predicted recruitment was then calculated at the conclusion of model fitting. The predicted recruitment standard deviation then replaced the former input standard deviation, and the model was rerun leading to a new predicted recruitment standard deviation. This process was repeated 50 times with the goal of getting the ratio between input recruitment standard deviation and predicted recruitment standard variation as close to unity as possible. After the 50 runs, the model was considered to have converged to a

satisfactory solution if: 1.) the ratio of recruitment standard deviations was between 0.98 and 1.02; 2.) the maximum gradient component, which measures the maximum amount of change in parameter estimates during model fitting, was less than 1×10^{-2}; and 3.) the Hessian matrix, which is used to calculate standard deviations for the parameter estimates, was positive definite.

The sensitivity of the stock assessment models to change was monitored by tracking several of the models' output quantities. The output quantities of interest included: the estimated fully selected gill net and trap net fishing mortality rates averaged for the last three years of the assessment, estimated population biomass averaged for the last three years of the assessment, estimated SSBR of the unfished population, predicted SSBR at reference mortality levels, estimated SSBR ratio, and the estimated yield calculated for reference mortality rates for the projected population. Model sensitivity was calculated as the percent difference of the test quantity of the adjusted model from the baseline value of the test quantity of the original model (Table 1.1):

$$D_{\%} = \frac{\theta' - \theta_0}{\theta_0} \times 100,$$

where $D_{\%}$ was the percent difference, θ_0 was the baseline value of the test quantity, and θ' was the value of the test quantity from the adjusted model. I considered a model to be sensitive to a change if that change produced a 10% or greater change in one of the output quantities.

Results

All of the stock assessment models were sensitive to changes in the input values. As expected, increasing observed mean weight-at-age of a harvested fish led to an increase in the projected TAC/HRG in all of the projection models (Table 1.2). Likewise, decreasing mean weight-at-age led to a decrease in the projected TAC/HRG in all of the projection models. These effects upon projected TAC/HRGs were greater for the gill net fishery in WFH-01 and WFH-04, and were greater for the trap net fishery in WFH-02 and WFH-05. Changes in mean weight-at-age of a harvested fish in the gill net fishery had no effect upon the projected TAC/HRG in the WFH-05 model, due to the small size of the gill net fishery in that management unit.

Surprisingly, setting gill net effort adjustment factors for number of meshes deep set through time to one, increasing gill net effort adjustment factors, and decreasing gill net effort adjustment factors increased the projected TAC/HRG by 34.7% and changed the remaining test quantities to a lesser degree (0.1-4.6%), except for SSBR of the unfished population which was unaffected, in the WFH-02 model (Table 1.2). All of these changes to gill net effort adjustment factors had slight effects (0.1-3.3%) on all of the test quantities, except for SSBR of the unfished population, in all of the other models; though no clear patterns were apparent. The WFH-01 model failed to converge when gill net effort adjustment factors were set equal to one.

As anticipated, shifting the maturity schedule later by one age led to substantial decreases (21.2-48.2%) in the SSBR of the unfished population and SSBR at the reference mortality schedule, with a greater decrease in SSBR at the target schedule, because the fish were maturing later after more mortality had occurred, and mortality was

higher for the reference schedule (Table 1.3). The greater decrease in SSBR at reference mortality schedule led to a decrease in the SSBR ratio. Likewise, shifting the maturity schedule earlier by one age led to substantial increases (13.6-55.4%) in the SSBR of the unfished population, SSBR at the reference mortality schedule, and the SSBR ratio in all of the models, due to the resulting increase in spawning biomass. Unexpectedly, shifting the maturity schedule later by one age increased the projected TAC/HRG in WFH-02 and WFH-04 by 38.5% and 14.4% respectively. Changes in the maturity schedule also had modest influence on fully selected gill net and trap net mortality, biomass, and projected TAC/HRG (0.1-5.9%) in all of the models. There was some influence because maturity schedule values are used to calculate the number of eggs produced for the stock-recruitment function, and this affects the objective function. The WFH-02 model failed to converge when the maturity schedule was shifted earlier by one age.

Setting the average proportion of females in the population equal to 0.5 led to an increase (21.8-31.6%) in SSBR of the unfished population and SSBR at the reference mortality schedule in all of the models, except the WFH-04 model which failed to converge (Table 1.3). As expected, increasing the proportion of females led to an increase (47.9-52.6%) in SSBR of the unfished population and SSBR at the reference mortality schedule in all of the models, because the spawning stock was considered to be the mature females within the population. Decreasing the proportion of females led to a decrease (48.9-52.6%) in SSBR of the unfished population and SSBR at the reference mortality schedule in all of the models, due to the resulting decrease in spawning stock. All of the adjustments made to the proportion of females led to slight changes (0.1-5.4%)

in the fully selected fishing mortalities, biomass, and SSBR ratio; and larger changes (23.8-37.2%) in the projected TAC/HRG for the WFH-02 model.

As expected, increasing trap net fishery underreporting adjustment factors led to increases in fully selected trap net mortality (12.8-18.1%) in order to account for the increased trap net harvest, except for a 1.3% decrease in trap net mortality in the WFH-05 model, and decreases in fully selected gill net mortality (7.6-22.3%; Table 1.4). Decreasing trap net fishery underreporting adjustment factors led to decreases in fully selected trap net mortality (10.7-13.4%) due to the lower trap net harvest, except for a 1.1% increase in trap net mortality in the WFH-05 model, and increases in fully selected gill net mortality (5.5-22.6%). Likewise, increasing gill net fishery underreporting adjustment factors led to increases in fully selected gill net mortality (6.8-29.8%) and decreases in fully selected trap net mortality (13.4-16.7%) due to increased gill net harvest, except for a 1.8% increase in trap net mortality in the WFH-05 model. Decreasing gill net fishery underreporting adjustment factors led to decreases in fully selected gill net mortality (6.1-18.7%) and increases in fully selected trap net mortality (10.3-14.5%) due to decreased gill net harvest, except for a 1.2% decrease in trap net mortality in the WFH-05 model. The small, but unforeseen, changes in fully selected fishing mortality rates (< 2%) in the WFH-05 model appeared to be due to the small gill net fishery, which effectively makes WFH-05 a one (trap net) fishery system. It appears the WFH-05 assessment model accounted for adjustments in observed trap net harvest by making large changes to the biomass and small changes to fishing mortality. Likewise, changes in gill net harvest led to only small adjustments of the biomass and trap net fishing mortality because of the small size of the fishery. Changes in the fishery

underreporting adjustment factors also affected biomass (0.5-28.5%), SSBR at the target mortality schedule (0-4.4%), and SSBR ratio (0-4.4%), though no patterns were apparent. Changes in the fishery underreporting adjustment factors had no effect on the SSBR of the unfished population.

As anticipated, increasing the time of year of spawning led to a decrease (2.7-12.4%) in SSBR of the unfished population, SSBR at the reference mortality schedule, and SSBR ratio because fewer fish survived to spawn later in the year (Table 1.3). Increasing the time of year of spawning also led to an increase (0.3-36.2%) in the projected TAC/HRG for all of the models because the spawning stock was exposed to the fisheries for a longer period of time before spawning. Decreasing the time of spawning led to an increase (2.8-14.5%) in SSBR of the unfished population, SSBR at the reference mortality schedule, and SSBR ratio because more fish would survive to spawn earlier in the year. Decreasing the time of spawning also led to a decrease (0.4-1.1%) in the projected TAC/HRG for all of the models because the spawning stock was exposed to the fisheries for a shorter period of time, except in the WFH-02 model which had an unexpected increase in the TAC/HRG of 33.0%. Adjustments to the time of spawning led to slight changes (0-5.1%) in the fully selected fishing mortalities and biomass with no clear pattern in all of the models. These slight changes appeared because time of spawning is used to calculate the number of eggs produced for the stock-recruitment function, which influenced the objective function.

Both increasing and decreasing the parameter bounds for natural mortality led the WFH-02 model to converge to the same solution, different from the original one, where fully selected trap net mortality decreased by 0.2%, fully selected gill net mortality

increased by 1.2%, biomass increased by 4.6%, SSBR of the unfished population remained unchanged, SSBR and SSBR ratio decreased by 0.1%, and projected TAC/HRG increased by 34.7% (Table 1.5). This alternate solution was very similar to the one reached by the model for this unit when changes were made to the gill net effort adjustment factors. Increasing and decreasing the natural mortality parameter's starting value for this unit also led to the same solution described above. Decreasing natural mortality's starting value led to 0.1% changes in fully selected trap net mortality, biomass, SSBR and SSBR ratio, and projected TAC/HRG in the WFH-04 model. The WFH-01 and WFH-05 models were unaffected by changes to natural mortality.

Surprisingly, decreasing trap net catchability bounds and increasing and decreasing gill net catchability bounds in the WFH-02 model led to the same alternate solution described above, where projected TAC/HRG increases by 34.7% while all the other test quantities, except for SSBR of the unfished population, changed from 0.1-4.6% (Table 1.6). Increasing and decreasing the trap net catchability starting value and increasing the gill net starting value again led to the same alternate solution for WFH-02 with the 34.7% increase in projected TAC/HRG. None of the other models showed any sensitivity to changes in the catchability parameters.

Increasing and decreasing the population scaling parameter's bounds, and decreasing the population scaling parameter's starting value led to the state with the 34.7% increase in projected TAC/HRG in the WFH-02 model (Table 1.7). Decreasing the relative population variation parameters' bounds led to a 31.9% increase in fully selected gill net mortality, a 25.2% increase in the projected TAC/HRG, and smaller changes (2.2-8.0%) in fully selected trap net mortality, biomass, and SSBR and SSBR

ratio in the WFH-02 model. The other models were unaffected by changes to the population scaling parameter and relative population variation parameters.

Increasing the bounds of the Ricker stock-recruitment function's productivity parameter led to a 37.2% increase in projected TAC/HRG and smaller changes (0.4-5.4%) in the fully selected fishing mortalities and biomass in the WFH-02 model (Table 1.8). Decreasing the bounds of the Ricker function's productivity parameter led to changes (0.2-10.8%) in all of the test quantities, except for SSBR for the unfished population, for the WFH-01 and WFH-05 models. The WFH-02 model failed to converge when both the Ricker function's productivity parameter's bounds and starting value were decreased. Increasing the bounds of the Ricker function's density dependence parameter, increasing the starting value of the productivity parameter, and increasing and decreasing the starting value of the density dependence parameter led to the state where the projected TAC/HRG increases by 34.7% in the WFH-02 model. Increasing the bounds of the Ricker function's density dependence parameter led to a 20.6-43.4% increase in the projected TAC/HRG and smaller changes (0.3-15.2%) in the remaining test quantities, except for SSBR of the unfished population, for the WFH-01 and WFH-04 models.

Increasing the bounds of the gill net selectivity function's first inflection point and decreasing the bounds of the gill net selectivity function's first slope parameter in the WFH-02 model led to the same state noted earlier with the 34.7% increase in the projected TAC/HRG (Table 1.9). Decreasing the bounds of the gill net selectivity function's first inflection point led to changes (3.0-15.4%) in all of the test quantities, except SSBR of the unfished population, in the WFH-02 and WFH-04 models. The

WFH-02 model failed to converge when the bounds on the gill net selectivity function's second inflection point were widened. Decreasing the bounds of the trap net selectivity function's first inflection point led to changes (5.3-18.1%) in all of the other test quantities, except for SSBR of the unfished population, in the WFH-04 model (Table 1.10). The WFH-01 model failed to converge when the starting value for the gill net selectivity function's second slope parameter was increased. Increasing and decreasing the starting values for the gill net selectivity function's first and second inflection points, decreasing the starting values for the gill net selectivity function's first and second slope parameters, decreasing the starting value for the trap net selectivity function's first inflection point, increasing and decreasing the starting value for the trap net selectivity function's second inflection point, and decreasing the starting value for the trap net selectivity function's second slope parameter led to the alternate state with the 34.7% increase in the projected TAC/HRG in the WFH-02 model. The WFH-02 model failed to converge when the starting value for the trap net selectivity function's first inflection point was increased. The WFH-04 model failed to converge when the starting value for the gill net selectivity function's first inflection point was decreased. Adjustments to the starting values for the gill net selectivity function's parameters led to 0.2-131.0% changes in fully selected trap net mortality, 1.4-1,277.2% changes in fully selected gill net mortality, 1.6-50.9% changes in biomass, 0.7-14.4% changes in SSBR and SSBR ratio, 2.4-87.7% changes in projected TAC/HRG, and no change to SSBR for the unfished population for the WFH-05 model. Increasing the starting value for the trap net selectivity function's second inflection point, and increasing and decreasing the starting value for the trap net selectivity function's second slope parameter led to a 0.1% change

in fully selected trap net mortality for the WFH-01 model. Decreasing the starting value for the trap net selectivity function's second inflection point led to changes (0.1-0.4%) in all of the test quantities, except for SSBR of the unfished population, for the WFH-01 model.

Increasing and decreasing the likelihood emphasis factor for natural mortality led to the alternate solution with a 34.7% increase in the projected TAC/HRG for the WFH-02 model (Table 1.11). The WFH-01 model failed to converge when the trap net catch and age composition emphasis factors were increased. The WFH-04 model failed to converge when the trap net catch emphasis factor was increased, and when the trap net and gill net age composition emphasis factors were decreased. All the remaining adjustments to the likelihood emphasis factors led to positive and negative changes (0.1-62.0%) that showed no pattern in all of the test quantities, except SSBR of the unfished population, for all of the models.

All of the stock assessment models also were sensitive to changes in model structure. Holding natural mortality constant at its starting value in the WFH-02 model led to the state with the 34.7% increase in projected TAC/HRG (Table 1.12). Modeling fecundity as a linear function of weight led to changes (0-38.8%) in all of the test quantities, except for SSBR of the unfished population, for the WFH-01, WFH-02, and WFH-05 models, because fecundity was used to calculate the number of eggs produced (stock size) for the stock-recruitment function (Table 1.6). The WFH-04 model failed to converge when fecundity was modeled as a linear function of weight.

Estimating each year's recruitment as a free parameter led to changes (0.1-39.8%) in all of the test quantities, except for SSBR of the unfished population, for the WFH-02

and WFH-05 (Table 1.12). The WFH-01 and WFH-04 models failed to converge when recruitment was estimated as free parameters. Estimating recruitment using a Beverton-Holt stock-recruitment model led to (0.3-54.3%) changes in all of the test quantities, except for SSBR of the unfished population, in all of the models.

Fitting mass, instead of numbers, of fish caught in the objective function led to changes (3.7-39.6%) in all of the test quantities, except for SSBR of the unfished population, for the WFH-02, WFH-04, and WFH-05 models (Table 1.12). The WFH-01 model failed to converge when the mass of fish caught was used in the objective function.

The use of the gamma likelihood function in place of the lognormal likelihood function led to small changes (0.1-0.7%) in all of the test quantities, except for SSBR of the unfished population, in the WFH-01, WFH-04, and WFH-05 models (Table 1.12). The WFH-02 model failed to converge when the gamma likelihood function was used. The use of the Dirchlet likelihood function in place of the multinomial likelihood function led to changes (0-25.9%) in all of the test quantities, except for SSBR of the unfished population, for all of the models.

Most of the adjustments made to the models led to negative log-likelihood values that were the same as, or higher than, the original likelihood values, which means that the model fit was not improved. In particular, the alternate solution often arrived at by the WFH-02 model had a higher likelihood value (4,340.5) than the original model (4,337.6). There were, however, several changes that led to a decrease in the negative log-likelihood value, which means that the changes produced parameter estimates that fit the data better than the original parameter estimates. In particular a better fit was obtained after decreasing the bounds of the Ricker recruitment function's density dependence parameter

in the WFH-01 and WFH-04 models and after decreasing the starting value of the gill net selectivity function's second slope parameter in the WFH-05 model (Table 1.13). These instances of better model fit could be due to random chance given the large number of model changes explored. Likelihood values could not be directly compared to determine better model fit in cases where the model structure was changed or when the likelihood emphasis factors were adjusted, because these changes altered the objective function.

Discussion

I performed a simple sensitivity analysis of the stock assessment models for lake whitefish in the 1836 treaty waters of Lake Huron to changes in input quantities and model structure. The changes I tested could be divided into two alternate categories that affect the way in which the results are interpreted. First, changes to the observed data and model structure led to changes in the objective function (negative log-likelihood) and thus altered the optimal solution (i.e., the best-fit parameter estimates) from the optimal solution of the baseline model. In this case, changes in the output quantities represent the model seeking the new optimal solution. Second, changes to the parameter starting values and parameter bounds did not alter the optimal solution from the optimal solution of the baseline model. In this case, changes in the output quantities mean that the model has become trapped at a local minimum for the objective function or found the true global minimum for the objective function depending upon whether the likelihood value is greater than or less than, respectively, the baseline model's likelihood value.

A simple sensitivity analysis, like the one conducted here, can be useful for identifying models that are unstable and highly sensitive to change. The WFH-02 stock assessment model appears to be such a sensitive model. Thirty-five of the 111 changes

tested led the model to converge to an alternate solution, which provided a poorer fit between observed and predicted values than the original model. The alternative solution was similar to the original model's solution except for a large increase in the projected TAC/HRG, due to a change in the estimated selectivity patterns. It appears that the WFH-02 model can easily become trapped at a local minimum for the objective function, which leads to this alternate solution, rather than finding the global minimum. Besides identifying unstable models, sensitivity analysis can also provide clues for analysts as they seek the best fit for an unstable model. Sensitivity analysis can reveal to which parameters of the model important outputs are most sensitive to change. It is critical for analysts to try a wide range of starting values and bounds for those parameters in order to help ensure that the global minimum for the objective function is found each time the model is updated. Failure to find the global minimum can lead to dangerous management decisions (e.g., setting harvest limits based on an overestimated projected yield, as in WFH-02). To this end, I have created a program for the MSC to automate my sensitivity analysis, using AD Model Builder software (ADMB 2002). This program will allow analysts to more easily evaluate the sensitivity of the stock assessment models whenever the models are updated.

All of the stock assessment models were sensitive to changes in the stock-recruitment function's parameter bounds. Decreasing the density dependence parameter bounds led to better fit parameter estimates for the WFH-01 and WFH-04 models, which significantly reduced the projected TAC/HRG in both models. It appears particularly important to do sensitivity analysis using a range of starting values and bounds for recruitment parameters each time the models are updated.

Sensitivity analysis also can reveal patterns of sensitivity across models, which may point to assumptions about underlying basic model structure (i.e. the way various biological, fishery, and data producing processes are included in the models) that should receive more attention. I found that the Lake Huron stock assessment models were sensitive to my assumptions embodied in stock-recruitment functions, gear selectivity, and assumed probability distributions used to define the likelihood functions. Of particular importance, the WFH-02 and WFH-04 model test quantities underwent similar changes, which resulted in increased TAC/HRGs, when the Beverton-Holt recruitment function was employed. A number of authors have considered the consequences of assuming different stock-recruitment relationships to the management advice stemming from those assumed relationships (Myers et al. 1994; Barrowman and Myers 2000). Other authors have discussed the relative merits of estimating stock-recruitment parameters inside stock assessment models versus outside them (Maunder and Deriso 2003). The issue here is somewhat different than is addressed in that work since I was only considering short-term projections. My concern here is more on whether including stock-recruitment functions as a form of a "prior" influences and potentially improves estimates and resulting short-term management advice, given a harvest policy exists. A simulation study, where either freely estimated recruitment or priors based on different recruitment functions were used would allow for a more detailed analysis of how different approaches to estimation of recruitment fare.

I did not explicitly consider alternative approaches to estimating selectivity. However, all of the models showed sensitivity to changes in gear selectivity starting values and parameter bounds. While sometimes the changes were small, in other cases

changes were pronounced (e.g., WFH-02 and WFH-05). These results reinforce concerns that have arisen about the general suitability of the current double logistic selectivity function during the development of the stock assessment models. As a result of problems encountered during the original models' development, all the selectivity parameters are estimated in only one of the Lake Huron lake whitefish assessment models (WFH-01). In each of the other models, some of the selectivity parameters must be held constant in order for the models to even converge on a solution. Thus, issues clearly go beyond simply finding the best starting values and parameter bounds. Reduced or constrained versions of the double logistic are not the only alternatives. For example, logistic curves (Punt et al. 2001), double logistic curves (Methot 1990), gamma-type functions (Deriso et al. 1985), and polynomials (Fournier 1983) have all been used to model selectivity. Kimura (1990) and Radomski et al. (2005) found that use of an inappropriate selectivity function can greatly increase the error in modeling results. This is another area where a simulation study could be used to help evaluate the current and alternative approaches to modeling selectivity. Alternatively, an empirical selectivity experiment could be used to determine the actual gear selectivity, but this would need to be done for both gill nets and trap nets.

All of the models showed some sensitivity to changes in the negative log likelihood function, both when likelihood emphasis factors were altered and when alternate distributional assumptions were made. In theory, if the assumed standard deviations for the natural mortality, catch, and effort data and the assumed maximum effective sample sizes for the age composition data are correct, then all of the likelihood emphasis factors should be set to one. Methot (1990) warns that sensitivity to changes in

the emphasis factors indicates that there is some inconsistency between the data sources or that some process has not been modeled correctly. Sensitivity analysis of the models to the likelihood emphasis factors should be tested whenever the emphasis factors are adjusted during model updates in order to report this sensitivity along with model results. Replacing the lognormal likelihood function with the gamma likelihood function led to only small changes in the test quantities. Cadigan and Meyers (2001) found similar results when comparing the two likelihood functions, although they emphasized that the gamma likelihood function is more robust to invalid distributional assumptions than the lognormal. Williams and Quinn (2000a, 2000b) successfully used the Dirchlet likelihood function to represent age composition data for Pacific herring, where sample sizes were large. Replacing the multinomial likelihood function with the Dirchlet likelihood function led to some changes in the test quantities, particularly the TAC/HRGs, in all of the models. Again I believe a simulation study could be used to evaluate the robustness of assessments based on these alternative distributions, and to evaluate potential approaches to selecting between them.

In conclusion, sensitivity analysis provides a useful tool for analysts applying stock assessment models. Running a sensitivity analysis whenever models are updated with new data can reveal unstable models which are highly sensitive to change. Furthermore, such analysis can identify particular parameters or assumptions that generally have a large influence on outputs of interest. This can help focus attention on these aspects of the assessment models. Such attention could come in the form of using simulations to evaluate performance of alternative modeling approaches or collecting new kinds of data to distinguish among modeling choices.

References

ADMB. 2002. AD Model Builder, version 6.0.4 edition. Otter Research Ltd., Nanaimo, British Columbia.

AHWG (Ad Hoc Working Group). 1979. Second report of the Ad Hoc Working Group to assess stocks of lake whitefish, bloater chub, and lake trout in treaty waters of the upper Great Lakes—State of Michigan: Lake Huron. Memeo. Rep. 38 pp. with appendix.

Barrowman, N. J., and R. A. Myers. 2000. Still more spawner-recruitment curves: the hockey stick and its generalizations. Canadian Journal of Fisheries and Aquatic Sciences 57:665-676.

Beverton, R. J. H., and S. J. Holt. 1957. On the Dynamics of Exploited Fish Populations. Chapman and Hall, London.

Cadigan, N. G., and R. A. Myers. 2001. A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences 58:560-567.

Clark, R. D., Jr., and K. D. Smith. 1984. Quota calculations with SAP.1 (Stock Assessment Package). MDNR Institute for Fisheries Research, Ann Arbor.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences 42:815-824.

Ebener, M. P., J. R. Bence, K. Newman, and P. Schneeberger. 2005. Application of statistical catch-at-age models to assess lake whitefish stocks in the 1836 treaty-ceded waters of the upper Great Lakes. Pages 271-309 *in* L. C. Mohr and T. F. Nalepa, editors. Proceedings of a workshop on the dynamics of lake whitefish (*Coregonus clupeaformis*) and the amphipod *Diporeia* spp. in the Great Lakes. Great Lakes Fishery Commission, Technical Report 66, Ann Arbor.

Fournier, D. 1983. An analysis of the Hecate Strait Pacific cod fishery using an age-structured model incorporating density-dependent effects. Canadian Journal of Fisheries and Aquatic Sciences 40:1233-1243.

Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39:1195-1207.

Kimura, D. K. 1990. Approaches to age-structured separable sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences 47:2364-2374.

Maunder, M. N., and R. B. Deriso. 2003. Estimation of recruitment in catch-at-age models. Canadian Journal of Fisheries and Aquatic Sciences 60:1204-1216.

Methot, R. D. 1990. Synthesis model: an adaptable framework for analysisi of diverse stock assessment data. Pages 259-277 *in* L. Low, editor. Proceedings of the symposium on applications of stock assessment techniques to gadids. International North Pacific Fisheries Commission, Bulletin 50, Vancouver.

Morgan, M. G., and M. Henrion. 1990. Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge.

Myers, R. A., A. A. Rosenberg, P. M. Mace, N. J. Barrowman, and V. R. Restrepo. 1994. In search of thresholds for recruitment overfishing. ICES Journal of Marine Science 51:191-205.

Pauly, D. 1980. On the interrelationships between natural mortality, growth-parameters, and mean environmental-temperature in 175 fish stocks. Journal du Conseil 39:175-192.

Punt, A. E., D. C. Smith, R. B. Thomson, M. Haddon, X. He, and J. M. Lyle. 2001. Stock assessment of the blue grenadier Macruronus novaezelandiae resource off south-eastern Australia. Marine and Freshwater Resources 52:701-717.

Radomski, P., J. R. Bence, and T. J. Quinn II. 2005. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational kill-dominated fishery. Canadian Journal of Fisheries and Aquatic Sciences 62:436-452.

Williams, E. H., and T. J. Quinn. 2000a. Pacific herring, *Clupea pallasi*, recruitment in the Bering Sea and north-east Pacific Ocean, I: relationships between different populations. Fisheries Oceanography 9:285-299.

Williams, E. H., and T. J. Quinn. 2000b. Pacific herring, *Clupea pallasi*, recruitment in the Bering Sea and north-east Pacific Ocean, II: relationships to environmental variables and implications for forecasting. Fisheries Oceanography 9:300-315.

Table 1.1. Predicted values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass (lbs), SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, projected TAC/HRG (lbs), and the negative log-likelihood values from the unmodified lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001.

	Evaluation Quantities							
Model	TN F	GN F	Biomass	Unfished SSBR	Ref SSBR	SSBR ratio	TAC/HRG	Likelihood
WFH-01	0.07	0.37	2,397,250	0.38	0.12	0.31	374,829	6,410.34
WFH-02	0.59	0.25	2,000,810	0.63	0.28	0.44	146,597	4,337.59
WFH-04	0.24	0.35	2,341,150	0.42	0.13	0.30	333,149	5,694.84
WFH-05	0.35	0.003	4,776,130	0.34	0.17	0.49	874,957	2,987.65

Table 1.2. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when data input values were increased (+), decreased (-), and set to specific values. Some changes led the models to fail to converge (fc).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
TN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.6	-4.6
GN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.4	-15.4
GN effort adjustment = 1.0	fc		fc		fc		fc		fc		fc		fc	
GN effort adjustment	-0.1	0.1	0.3	-0.4	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	-0.3	0.3
WFH-02														
TN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.5	-16.5
GN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.5	-3.5
GN effort adjustment = 1.0	0.7		1.3		4.1		0.0		-0.3		-0.3		34.7	
GN effort adjustment	-0.4	-0.1	1.1	1.3	4.6	4.5	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.6
WFH-04														
TN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.1	-8.1
GN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0	-12.0
GN effort adjustment = 1.0	-1.4		-2.1		1.2		0.0		0.2		0.2		1.4	
GN effort adjustment	0.3	-0.3	0.5	-0.5	-0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	-0.3	0.3
WFH-05														
TN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.1	-20.0
GN weight-at-age	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN effort adjustment = 1.0	-1.2		-3.3		2.8		0.0		1.1		1.1		2.3	
GN effort adjustment	0.3	-0.3	0.8	-0.8	-0.6	0.6	0.0	0.0	-0.3	0.3	-0.3	0.3	-0.5	0.5

Table 1.3. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when data input values were increased (+), decreased (-), and set to specific values. For maturity schedule, an increase (+) means maturity values were shifted up to the next oldest age, while a decrease (-) means maturity values were shifted down to the next youngest age. Some changes led the models to fail to converge (fc).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
Maturity schedule	-0.2	0.3	-0.1	0.1	0.6	-0.9	-24.9	23.5	-48.2	53.5	-31.0	24.3	1.2	-1.8
Proportion female = 0.5	0.0		0.0		0.0		31.6		31.6		0.0		0.0	
Proportion female	0.0	0.0	0.0	0.0	0.0	0.0	52.6	-52.6	52.6	-52.6	0.0	0.0	0.0	0.0
Time of spawning	0.0	0.1	0.0	0.0	0.1	-0.2	-4.0	4.2	-11.0	13.1	-7.3	8.5	0.3	-0.4
WFH-02														
Maturity schedule	-0.5	fc	1.2	fc	5.9	fc	-21.2	fc	-37.2	fc	-20.3	fc	38.5	fc
Proportion female = 0.5	0.1		1.4		2.7		22.2		21.8		-0.4		28.9	
Proportion female	0.3	-0.4	1.3	1.1	1.2	5.4	48.9	-48.9	47.9	-48.9	-0.6	0.0	23.8	37.2
Time of spawning	-0.3	-0.1	1.1	1.3	5.1	4.0	-2.7	2.8	-7.4	8.2	-4.8	5.3	36.2	33.0
WFH-04														
Maturity schedule	-2.8	-2.1	-1.3	-0.1	2.7	3.9	-26.0	24.8	-43.2	55.4	-23.3	24.5	0.9	14.4
Proportion female = 0.5	fc		fc		fc		fc		fc		fc		fc	
Proportion female	0.0	0.0	0.0	0.0	0.0	0.0	50.0	-50.0	50.0	-50.0	0.0	0.0	0.0	0.0
Time of spawning	-0.6	0.2	-0.8	0.5	0.5	-0.1	-3.3	3.4	-12.4	14.5	-9.4	10.7	2.2	-1.1
WFH-05														
Maturity schedule	-2.2	1.9	-2.0	2.0	2.4	-2.4	-28.0	27.2	-40.5	44.5	-17.4	13.6	1.9	-1.9
Proportion female = 0.5	0.0		0.0		0.0		25.6		25.6		0.0		0.0	

Table 1.3 (cont'd).

Proportion female	0.0	0.0	0.0	0.0	0.0	0.0	50.3	-50.3	50.3	-50.3	0.0	0.0	0.0	0.0
Time of spawning	-0.7	0.6	-0.5	0.5	0.6	-0.6	-4.5	4.7	-9.9	11.3	-5.6	6.3	0.7	-0.7

Table 1.4. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when data input values were increased (+), decreased (-), and set to specific values.

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
TN harvest reporting = 1.0	-6.8		0.8		-1.4		0.0		-0.1		-0.1		-3.3	
TN harvest reporting	-11.7	15.0	7.2	-9.7	-6.1	9.7	0.0	0.0	3.2	-4.4	3.2	-4.4	-8.7	14.3
GN harvest reporting = 1.0	6.8		-1.4		-5.2		0.0		-0.4		-0.4		-4.4	
GN harvest reporting	11.9	-14.7	-7.8	8.6	-11.4	18.0	0.0	0.0	-3.6	3.8	-3.6	3.8	-8.7	14.3
WFH-02														
TN harvest reporting = 1.0	-5.4		1.9		0.5		0.0		-0.7		-0.7		31.9	
TN harvest reporting	-10.7	12.8	7.4	-7.6	-8.1	24.6	0.0	0.0	-0.6	0.5	-0.6	0.5	22.9	53.5
GN harvest reporting = 1.0	5.3		-0.2		2.4		0.0		0.4		0.4		29.7	
GN harvest reporting	10.3	-13.4	-7.2	8.7	-4.3	13.5	0.0	0.0	0.7	-0.7	0.7	-0.7	-6.5	53.3
WFH-04														
TN harvest reporting = 1.0	-4.6		2.3		-2.6		0.0		-0.5		-0.5		-2.9	
TN harvest reporting	-13.4	18.1	5.5	-7.6	-5.8	9.1	0.0	0.0	0.0	0.2	0.0	0.2	-4.7	7.6
GN harvest reporting = 1.0	5.0		-2.6		-4.2		0.0		-0.1		-0.1		-5.5	
GN harvest reporting	14.5	-16.7	-6.1	6.8	-11.8	18.5	0.0	0.0	0.0	0.2	0.0	0.2	-13.2	20.9
WFH-05														
TN harvest reporting = 1.0	1.3		8.6		-7.6		0.0		-0.4		-0.4		-7.7	
TN harvest reporting	1.1	-1.3	22.6	-22.3	-18.3	28.5	0.0	0.0	0.0	0.1	0.0	0.1	-18.6	29.0
GN harvest reporting = 1.0	-2.2		-9.9		3.4		0.0		1.0		1.0		3.1	
GN harvest reporting	-1.2	1.8	-18.7	29.8	1.2	-1.7	0.0	0.0	0.1	-0.2	0.1	-0.2	1.4	-2.1

Table 1.5. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when natural mortality parameters were increased (+) and decreased (-).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
Natural mortality bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Natural mortality initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-02														
Natural mortality bounds	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.7
Natural mortality initial value	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.5
WFH-04														
Natural mortality bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Natural mortality initial value	0.0	-0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.1
WFH-05														
Natural mortality bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Natural mortality initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1.6. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when catchability parameters were increased (+) and decreased (-).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
TN catchability bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN catchability initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN catchability bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN catchability initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-02														
TN catchability bounds	0.0	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7
TN catchability initial value	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.7
GN catchability bounds	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.7
GN catchability initial value	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7	0.0
WFH-04														
TN catchability bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN catchability initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN catchability bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN catchability initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-05														
TN catchability bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN catchability initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN catchability bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN catchability initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1.7. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when data input values were increased (+) and decreased (-).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
Population scaler bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Population scaler initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Relative population variation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-02														
Population scaler bounds	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.7
Population scaler initial value	0.0	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7
Relative population variation	0.0	8.0	0.0	31.9	0.0	0.0	0.0	0.0	0.0	2.2	0.0	2.2	0.0	25.2
WFH-04														
Population scaler bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Population scaler initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Relative population variation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-05														
Population scaler bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Population scaler initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Relative population variation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1.8. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when recruitment parameters were increased (+) and decreased (-). Some changes led the models to fail to converge (fc).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
Recruitment α bounds	0.0	0.9	0.0	0.2	0.0	-3.4	0.0	0.0	0.0	-0.6	0.0	-0.6	0.0	-9.2
Recruitment α initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Recruitment β bounds	0.0	3.0	0.0	0.8	0.0	-15.2	0.0	0.0	0.0	-2.4	0.0	-2.4	0.0	-43.4
Recruitment β initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-02														
Recruitment α bounds	-0.4	fc	1.1	fc	5.4	fc	0.0	fc	0.0	fc	0.0	fc	37.2	fc
Recruitment α initial value	-0.2	fc	1.2	fc	4.6	fc	0.0	fc	-0.1	fc	-0.1	fc	34.7	fc
Recruitment β bounds	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7	0.0
Recruitment β initial value	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.7
WFH-04														
Recruitment α bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Recruitment α initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Recruitment β bounds	0.0	-0.3	0.0	1.2	0.0	-2.1	0.0	0.0	0.0	2.4	0.0	2.4	0.0	-20.6
Recruitment β initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-05														
Recruitment α bounds	0.0	10.8	0.0	10.5	0.0	-5.4	0.0	0.0	0.0	1.0	0.0	1.0	0.0	-7.1
Recruitment α initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Recruitment β bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Recruitment β initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1.9. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when selectivity parameters were increased (+) and decreased (-). The selectivity function parameters were the first inflection point (*p1*), first slope (*p2*), second inflection point (*p3*), and second slope (*p4*). Some changes led the models to fail to converge (fc).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
GN sel. *p1* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p1* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p2* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p2* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p3* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p3* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
GN sel. *p4* initial value	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0
WFH-02														
GN sel. *p1* bounds	-0.2	6.6	1.2	3.0	4.6	-15.4	0.0	0.0	-0.1	-8.3	-0.1	-8.3	34.7	14.7
GN sel. *p1* initial value	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.7
GN sel. *p2* bounds	0.0	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7
GN sel. *p2* initial value	0.0	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7
GN sel. *p3* bounds	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0
GN sel. *p3* initial value	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.6
GN sel. *p4* initial value	0.0	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7
WFH-04														
GN sel. *p1* bounds	0.0	-4.8	0.0	-9.3	0.0	0.7	0.0	0.0	0.0	-7.9	0.0	-7.9	0.0	2.1
GN sel. *p1* initial value	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc	0.0	fc
GN sel. *p2* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p2* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1.9 (cont'd).

GN sel. *p3* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p3* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GN sel. *p4* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-05														
GN sel. *p1* bounds	not estimated as a parameter													
GN sel. *p1* initial value	-9.3	-40.5	127.1	-64.6	44.6	30.9	0.0	0.0	14.4	-5.8	14.4	-5.8	41.7	27.8
GN sel. *p2* bounds	not estimated as a parameter													
GN sel. *p2* initial value	-1.1	0.2	6.7	-8.1	4.6	-3.7	0.0	0.0	1.3	-1.2	1.3	-1.2	6.2	-5.6
GN sel. *p3* bounds	not estimated as a parameter													
GN sel. *p3* initial value	6.0	131.0	7.1	1277.2	-2.4	-50.9	0.0	0.0	1.1	10.5	1.1	10.5	-3.7	-87.7
GN sel. *p4* initial value	3.9	-4.6	1.4	1.9	-1.6	2.1	0.0	0.0	0.7	-0.9	0.7	-0.9	-2.4	3.1

Table 1.10. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when selectivity parameters were increased (+) and decreased (-). The selectivity function parameters were the first inflection point (*p1*), first slope (*p2*), second inflection point (*p3*), and second slope (*p4*). Some changes led the models to fail to converge (fc).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
TN sel. *p1* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel.*p1* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel. *p3* initial value	0.1	-0.4	0.0	-0.2	0.0	0.1	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	0.1
TN sel. *p4* initial value	0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-02														
TN sel. *p1* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel. *p1* initial value	fc	-0.2	fc	1.2	fc	4.6	fc	0.0	fc	-0.1	fc	-0.1	fc	34.7
TN sel. *p3* initial value	-0.2	-0.3	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.7
TN sel. *p4* initial value	0.0	-0.2	0.0	1.2	0.0	4.6	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	34.7
WFH-04														
TN sel. *p1* bounds	0.0	-17.0	0.0	-12.5	0.0	9.0	0.0	0.0	0.0	-5.3	0.0	-5.3	0.0	18.1
TN sel. *p1* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel. *p3* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel. *p4* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WFH-05														
TN sel. *p1* bounds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel. *p1* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel. *p3* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN sel. *p4* initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1.11. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when likelihood emphasis factors were increased (+) and decreased (-). Some changes led the models to fail to converge (fc).

	Evaluation Quantities													
	TN F		GN F		Biomass		Unfished SSBR		Ref SSBR		SSBR ratio		TAC/HRG	
Description of change	+	-	+	-	+	-	+	-	+	-	+	-	+	-
WFH-01														
Natural mortality like.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN catch like.	fc	0.2	fc	1.8	fc	-4.2	fc	0.0	fc	-0.4	fc	-0.4	fc	-6.3
TN effort like.	-7.9	13.6	-3.8	6.1	10.4	-11.7	0.0	0.0	1.0	-1.1	0.9	-1.1	17.5	-17.3
TN age comp like.	fc	0.4	fc	10.2	fc	1.6	fc	0.0	fc	0.4	fc	0.4	fc	14.3
GN catch like.	0.2	-0.4	0.0	0.2	-0.2	0.5	0.0	0.0	0.3	-0.4	0.3	-0.4	0.3	-0.5
GN effort like.	2.1	-2.7	5.8	-8.0	-1.0	1.7	0.0	0.0	1.8	-2.9	1.8	-2.9	-1.3	3.4
GN age comp like.	10.1	-5.5	6.6	-4.0	-12.6	9.0	0.0	0.0	-5.0	8.0	-5.0	8.0	-16.5	11.6
WFH-02														
Natural mortality like.	-0.2	-0.2	1.2	1.2	4.6	4.6	0.0	0.0	-0.1	-0.1	-0.1	-0.1	34.7	34.6
TN catch like.	-0.7	0.3	-0.1	3.4	6.4	1.3	0.0	0.0	0.0	-0.3	0.0	-0.3	37.2	30.3
TN effort like.	-3.0	3.0	-0.9	3.5	11.1	-2.3	0.0	0.0	1.1	-1.7	1.1	-1.7	45.9	23.1
TN age comp like.	-4.4	3.9	-8.1	17.2	6.1	0.0	0.0	0.0	-1.1	-0.1	-1.1	-0.1	37.1	27.9
GN catch like.	0.2	7.3	-0.8	34.2	-0.5	1.0	0.0	0.0	-0.2	2.5	-0.2	2.5	-0.8	25.8
GN effort like.	-0.5	0.6	2.7	-0.6	4.0	4.9	0.0	0.0	-0.1	-0.3	-0.1	-0.3	33.7	36.2
GN age comp like.	19.4	-6.5	58.3	-9.9	-16.4	15.4	0.0	0.0	0.4	0.3	0.4	0.3	-8.7	62.0
WFH-04														
Natural mortality like.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN catch like.	fc	-2.2	fc	2.3	fc	-2.4	fc	0.0	fc	0.6	fc	0.6	fc	-3.7

Table 1.11 (cont'd).

TN effort like.	-8.6	11.6	-2.9	4.4	2.3	-1.9	0.0	0.0	-2.0	3.4	-2.0	3.4	13.9	-9.8
TN age comp like.	-5.3	fc	-3.5	fc	5.2	fc	0.0	fc	5.1	fc	5.1	fc	-0.3	fc
GN catch like.	-0.9	1.1	-0.7	0.9	0.3	-0.1	0.0	0.0	-0.6	0.9	-0.6	0.9	0.7	-0.8
GN effort like.	-0.9	-1.2	0.1	-2.6	0.4	1.2	0.0	0.0	-0.4	0.2	-0.4	0.2	4.4	-0.6
GN age comp like.	13.7	fc	12.1	fc	-5.7	fc	0.0	fc	3.3	fc	3.3	fc	-11.3	fc
WFH-05														
Natural mortality like.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TN catch like.	-2.5	4.1	-3.8	6.8	4.4	-7.2	0.0	0.0	-0.7	1.2	-0.7	1.2	5.8	-9.3
TN effort like.	-9.4	18.9	-2.5	6.3	4.2	-8.4	0.0	0.0	-2.3	3.0	-2.3	3.0	9.9	-16.3
TN age comp like.	1.0	8.7	-2.1	10.4	19.7	-15.5	0.0	0.0	8.4	-7.3	8.4	-7.2	23.3	-15.4
GN catch like.	1.6	-2.3	-3.9	7.6	-2.7	4.4	0.0	0.0	-0.9	1.4	-0.9	1.4	-2.4	4.1
GN effort like.	3.7	-2.6	14.1	-10.5	-7.9	8.0	0.0	0.0	-3.0	3.0	-3.0	3.0	-7.6	7.7
GN age comp like.	19.0	-12.5	12.2	-8.8	-8.9	7.3	0.0	0.0	3.2	-2.6	3.2	-2.6	-15.6	13.1

Table 1.12. Percent difference from baseline values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass, SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, and projected TAC/HRG from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when model structure was modified. Some changes led the models to fail to converge (fc).

	Evaluation Quantities						
Description of change	TN F	GN F	Biomass	Unfished SSBR	Ref SSBR	SSBR ratio	TAC/HRG
WFH-01							
Natural mortality constant at initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fecunditiy linear function of age	0.0	0.0	0.3	0.0	0.1	0.1	0.8
Recruitment unconstrained	fc	fc	fc	fc	fc	fc	fc
Recruitment Beverton-Holt model	0.4	0.5	0.3	0.0	0.3	0.3	1.7
Fit fishery biomass	fc	fc	fc	fc	fc	fc	fc
Gamma likelihood component	0.1	-0.1	0.3	0.0	0.1	0.1	0.5
Dirichlet likelihood component	-2.9	20.1	-8.4	0.0	17.6	17.6	-16.9
WFH-02							
Natural mortality constant at initial value	-0.2	1.2	4.6	0.0	-0.1	-0.1	34.7
Fecunditiy linear function of age	-0.5	1.0	5.9	0.0	0.0	0.0	38.8
Recruitment unconstrained	-0.1	-0.1	0.2	0.0	0.0	0.0	0.6
Recruitment Beverton-Holt model	-1.0	0.7	8.0	0.0	0.4	0.4	42.9
Fit fishery biomass	8.4	39.6	6.9	0.0	5.4	5.4	3.7
Gamma likelihood component	fc	fc	fc	fc	fc	fc	fc
Dirichlet likelihood component	6.2	1.5	-0.4	0.0	-3.8	-3.8	25.9
WFH-04							
Natural mortality constant at initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fecunditiy linear function of age	fc	fc	fc	fc	fc	fc	fc
Recruitment unconstrained	fc	fc	fc	fc	fc	fc	fc
Recruitment Beverton-Holt model	-10.0	-3.5	18.6	0.0	6.9	6.9	54.3
Fit fishery biomass	-6.2	-4.9	15.7	0.0	5.0	5.0	14.6
Gamma likelihood component	0.1	-0.3	0.3	0.0	0.0	0.0	0.5

Table 1.12 (cont'd).

Dirichlet likelihood component WFH-05	-0.1	-4.7	2.2	0.0	0.0	0.0	0.5
Natural mortality constant at initial value	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fecunditiy linear function of age	1.0	0.4	-3.2	0.0	-0.6	-0.6	-5.1
Recruitment unconstrained	-3.8	-1.5	23.0	0.0	4.5	4.5	39.8
Recruitment Beverton-Holt model	-5.5	-5.7	16.7	0.0	4.3	4.3	20.8
Fit fishery biomass	-2.9	-10.2	19.3	0.0	8.2	8.2	12.8
Gamma likelihood component	-0.1	0.7	0.4	0.0	0.1	0.1	0.4
Dirichlet likelihood component	9.9	5.1	-11.6	0.0	-1.4	-1.4	-18.0

Table 1.13. Predicted values for fully selected trap-net mortality (TN F), fully selected gill-net mortality (GN F), population biomass (lbs), SSBR for unfished population, SSBR at reference mortality schedule, SSBR ratio, projected TAC/HRG (lbs), and the negative log-likelihood values from the lake whitefish stock assessment models for the 1836 treaty-ceded waters of Lake Huron in 2001, when changes improved model fit. The selectivity function parameter *p4* was the second slope. Selectivity was abbreviated sel., and decrease was abbreviated dec.

	Evaluation Quantities							
Description of change	TN F	GN F	Biomass	Unfished SSBR	Ref SSBR	SSBR ratio	TAC/HRG	Likelihood
WFH-01								
Recruitment β bounds dec.	0.07	0.38	2,032,260	0.38	0.12	0.30	212,169	6,405.53
WFH-04								
Recruitment β bounds dec.	0.23	0.36	2,292,740	0.42	0.13	0.31	264,482	5,688.16
WFH-05								
GN sel. *p4* initial value dec.	0.34	0.003	4,876,080	0.34	0.17	0.49	902,314	2,981.44

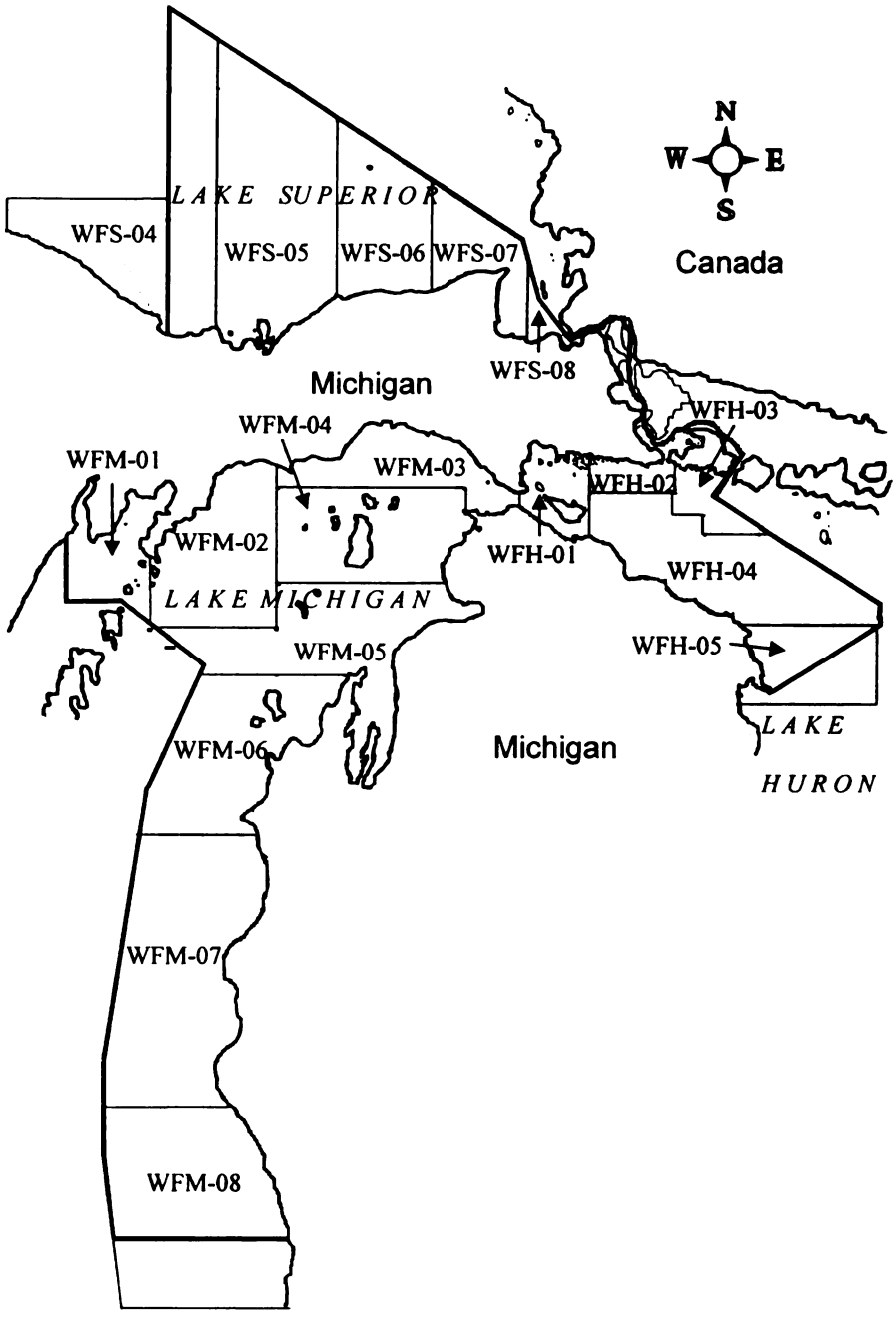


Figure 1.1. 1836 treaty-ceded waters and lake whitefish management units in lakes Huron, Michigan and Superior.

CHAPTER 2

EVALUATING METHODS FOR ESTIMATING PROCESS AND OBSERVATION ERRORS IN STATISTICAL CATCH-AT-AGE ANALYSIS

Introduction

Modern statistically-based stock assessment models allow a stock assessment analyst to explicitly account for process and observation errors. Observation errors within statistical catch-at-age analysis (SCAA) commonly take the form of differences between observed and true fishery catch or survey indices of abundance. Process errors within SCAA generally take the form of annual deviations in recruitment, catchability, or fishery selectivity. Errors within SCAA also can be combinations of observation and process error. For instance, fishing effort can be predicted within SCAA using estimates of annual fishing mortality rates on fully selected fish and fishery catchability. Some analysts implicitly treat the deviations between observed and predicted effort as observation error. In reality these deviations are due to interannual variation in catchability (process error, which will often dominate) as well as errors in observing the nominal amount of fishing effort. Similarly, deviations between model and observed values of fishery catch per unit effort (CPUE) arise from a combination of observation error and interannual variation in catchability. The variances associated with all of these error sources or the ratios of those variances are used in SCAA to weight the different data sources during the model fitting process (Fournier and Archibald 1982; Deriso et al. 1985).

It is important to understand how values for process and observation error variances are obtained because those values can affect SCAA results. Deriso et al. (1985) demonstrated that altering the assumed known ratio of catch variance to effort variance, which they used to weight fishing effort and catch data, affected estimates of fully-selected fishing mortality, surplus production and year-class strength of halibut. Chen and Paloheimo (1998) found that misspecifying the ratio of catch variance to effort variance could lead to increased estimation bias in catchability and natural mortality. The National Research Council (1998) recognized the importance of correctly weighting different data sources within stock assessments, and recommended that more research is needed to determine how those weights should be set.

Process and observation error variance values can be derived either separately from SCAA or estimated within SCAA. Derivation of error variance values separate from SCAA is the more common approach, with these estimates or their ratio then assumed known during the subsequent SCAA. A plausible estimate of observation error variance for data subsets such as observed annual catch, effort, or abundance indices often can be obtained through analysis of the raw data used to derive these quantities, taking into account the sampling designs (Law and Kelton 1982; Sitar et al. 1999). Process error variances cannot be estimated in the same way, by analysis of assessment data subsets external to the model, because by themselves these data are not informative about how model parameters such as catchability are varying. As a result, assessment scientists often rely on expert opinion to obtain estimates of this component of variance. Merritt and Quinn (2000) applied this expert opinion approach and other empirical data weighting approaches to the assessment of a recreational fishery, and judged that the

expert opinion approach produced the best model based on an analytic hierarchy process, a decision making technique. Since they were working with actual fishery data, Merritt and Quinn (2000) could not evaluate the accuracy of the variance estimates produced by expert opinion. With estimates (or educated guesses) of observation and process error variances in hand, the assessment often then proceeds assuming these values or their ratio is known. There are several potential disadvantages to such a two-step procedure. First, uncertainty surrounding the error variance estimates is ignored in the subsequent SCAA (Maunder 2001). Second, the reliability of process error variances based on expert judgment can be questioned. Francis et al. (2003) discovered that the standard values used in New Zealand stock assessments for the coefficients of variation (CV) for commercial CPUE (effectively the CV for process errors in catchability), which primarily are derived from expert opinion, typically were too low to be consistent with the resulting interannual variation in assessment model estimates of fishing mortality. In contrast, they found the prespecified trawl survey CVs were too large to be consistent with the resulting deviations between assessment model estimates of catchable stock abundance and observed survey indices.

Process and observation error variances generally are not estimated within SCAA due to difficulties in estimating the variances as parameters. This task is particularly difficult when multiple variances are being estimated. The potential advantages of estimating variances in SCAA are that 1) all of the data in the analysis can be synthesized to obtain the variance estimates, and 2) for some methods, uncertainty surrounding the variance estimates can be quantified and accounted for in the analysis. Two main statistical methods exist for estimating process and observation error variances in SCAA

(Schnute 1994). First, a Bayesian approach could be taken, in which prior information about the process and observation error variances is incorporated into the analysis to derive marginal posterior densities of the variance estimates as well as other parameters and quantities of interest. Second, a mixed model approach could be taken, in which the process errors are treated as random effects, rather than parameters, which can sometimes allow for the estimation of both the process and observation error variances as model parameters. Richards et al. (1997) suggested a third approach to estimating process and observation error variances. This method requires a prior point estimate of observation error variance. In essence the method is to repeatedly fit the assessment model, each time using a different assumed known variance ratio, and choose the ratio that produces deviations between observed data and model predictions that are most consistent with the prior estimate of observation error variance. Unlike the other two methods, this approach does not account for uncertainty in variance estimates in the analysis, and I refer to it as the *ad hoc* method, because the approach to estimating the ratio of the variances was not based on a formal statistical justification. To my knowledge, no previous attempt has been made to compare these different approaches within SCAA.

My objective was to determine whether or not process and observation error variances could be reliably estimated within SCAA. To answer this question, I evaluated two different methods for estimating the variances associated with annual variations in catchability (i.e., process error) and total catch (i.e., observation error) in SCAA. The two methods I examined were the *ad hoc* approach described by Richards et al. (1997) and a Bayesian approach. I looked at using both strongly and weakly informative priors on catchability variance for the Bayesian approach. In addition, I initially attempted to

implement a mixed model approach using the random effects module for AD Model Builder (Otter Research Limited 2005), but that estimation model failed to converge to a solution for any of the simulated data sets with which I tested it. The mixed model's failure to converge likely was due to the highly complex nature of the model which prevented the estimation of the random effects and associated variance parameters. The random effects module of AD Model Builder has not been used to estimate variance parameters in SCAA before to my knowledge. Additional experimentation with this software may produce statistical catch-at-age mixed models with better convergence properties. Monte Carlo simulations were used to investigate the performance of the different methods.

Methods

I used a simulation study to evaluate different methods of estimating process and observation error variances in SCAA. A data generating model was used to simulate data sets from a hypothetical fish population. The estimation models, each using a different error estimation method, were fit to the simulated data sets. The data generating model, *ad hoc* and Bayesian estimation models were all built using AD Model Builder software (Otter Research Limited 2004). For the following discussion, descriptions of all the symbols are given in Table 2.1, while many of the equations describing my models are given in Tables 2.2 and 2.3. I reference these equations as Equation *x.y*, where equation *y* is found in Table *x*.

Data Generating Model

I developed a data generating model to simulate the dynamics of a hypothetical fish population based on lake whitefish stocks in the upper Great Lakes. The population dynamics were described using abundance-at-age and age-specific mortality rates created by the model. A gill net fishery operating on the population produced observed total annual catch, age composition and fishing effort data.

I generated abundance-at-age using an exponential population function (Equation 2.2.1). To produce abundance-at-age in the first year (Equation 2.2.2), mortality was applied to randomly generated numbers of age-1 fish, which were drawn from a lognormal distribution (Table 2.4). The mean of the distribution was chosen by assuming the population experienced equilibrium recruitment prior to the model time series. Recruitment to the first age in subsequent years was calculated with a Ricker stock-recruitment function (Equation 2.2.3; Table 2.4). The number of female spawners was calculated as one-half of the number of fish age-3 and older, thereby assuming knife-edge maturity and a 1:1 sex ratio.

Total mortality was partitioned into natural mortality and fishing mortality sources (Equation 2.2.4). Natural mortality was a constant value for all years and ages (Table 2.4). Fishing mortality was generated using a fully separable fishing mortality model (Equation 2.2.5), where the age effect consisted of age-specific selectivity and the year effect consisted of year-specific catchability and observed fishing effort. Age-specific selectivity values were specified to create a dome-shaped selectivity curve, which is typical of gill net fisheries (Table 2.4). Catchability varied from year to year according to a lognormal white noise model (Table 2.4):

(1) $$q_y = \bar{q} e^{\varepsilon_{q,y}},$$

$$\varepsilon_{q,y} \sim N\left(0, \sigma_q^2\right).$$

The value for the standard deviation of log catchability σ_q was randomly generated from a lognormal distribution with three different means representing low, medium and high levels of catchability variation (Table 2.4). An "observed" point estimate of the log catchability standard deviation was generated, which simulated information that a stock assessment analyst might possess. The observed point estimate was drawn from a lognormal distribution:

(2) $$\sigma'_q = \sigma_q e^{\zeta_q},$$

$$\zeta_q \sim N\left(0, \sigma^2_{\sigma_q}\right).$$

The log-scale standard deviation of the log catchability standard deviation σ_{σ_q} was given the same value as was used to cause the true standard deviation of log catchability to depart from its underlying median. Therefore, I effectively assumed that the observed point estimate of the standard deviation of log catchability came from a lognormal distribution with the same median as the true standard deviation of log catchability, but with double the log-scale standard deviation as did the true standard deviation's generating distribution. Doubling the standard deviation represents the addition of observation error to the measurement of the log catchability standard deviation. Fishing effort was specified so that effort increases to a maximum in the middle of the time series and then decreases to the end of the time series (Table 2.4). This fishing effort pattern simulated a growing fishery that was regulated by effort limitations during the second

half of the time series. Total mortality Z_0 used to produce abundance-at-age in the first year (Equation 2.2.2) was generated with Equations 2.2.4 and 2.2.5 with the assumption that fishing effort in years prior to the first year of the analysis was equal to fishing effort in the first year of the analysis.

I generated observed data from a gill net fishery from simulated abundance-at-age and mortality rates. Catch-at-age was calculated using Baranov's catch equation (Equation 2.2.6). Observed total annual catch $\widetilde{C}_y$ was calculated by summing catch-at-age $C_{y,a}$ across ages for each year and incorporating observation error $\varepsilon_{C,y}$:

(3)
$$\widetilde{C}_y = \left[\sum_{a=1}^{m} C_{y,a}\right] e^{\varepsilon_{C,y}},$$

$$\varepsilon_{C,y} \sim N\left(0, \sigma_C^2\right).$$

I chose to use multiplicative lognormal errors because this is a standard assumption in SCAA (Fournier and Archibald 1982; Deriso et al. 1985). The value for the standard deviation of log total catch σ_C was randomly generated from a lognormal distribution with two different means representing low and high levels of observation error (Table 2.4). An "observed" point estimate of the log total catch standard deviation was generated, which simulated information that a stock assessment analyst likely would possess. The observed estimate σ'_C was drawn from a lognormal distribution:

(4)
$$\sigma'_C = \sigma_C e^{\zeta_C},$$

$$\zeta_C \sim N\left(0, \sigma^2_{\sigma_C}\right).$$

The log-scale standard deviation of the log total catch standard deviation σ_{σ_C} was given the same value as was used to cause the true standard deviation of log total catch to depart from its underlying median. Observed fishery age composition data was generated by drawing a random sample from a multinomial distribution with a sample size of 100, and proportions calculated from catch-at-age in the fishery (Equation 2.2.7). Any errors in measuring fishing effort were lumped with interannual variation in catchability as process error. Natural mortality was known without error.

Estimation Models

The estimation models used the same equations as the data generating model except when estimating abundance-at-age in the first year, recruitment, and selectivity. Annual recruitment was estimated as a mean recruitment parameter and a vector of annual recruitment deviation parameters (i.e., a vector of deviations that must sum to zero). Abundance-at-age in the first year was estimated as a mean abundance parameter and a vector of abundance deviation parameters (i.e., a vector of deviations that must sum to zero). Selectivity was estimated as a double logistic function of age (Equation 2.2.8). Abundance-at-age (Equation 2.2.1), total mortality (Equation 2.2.4), fishing mortality (Equation 2.2.5), catchability (Equation 2.1), catch-at-age (Equation 2.2.6), total catch (Equation 3), and proportion of catch-at-age (Equation 2.2.7) were calculated as in the data generating model. True parameter values produced by the data generating model were used as starting values for parameters in the estimation models, to expedite numerical searches during the simulations.

The estimation models differed from each other in the method used to estimate variances for process error in catchability and observation error in total catch. First, an

ad hoc approach was used in which the proportion of process error variance was set so that predicted observation error variance was consistent with an observed point estimate of observation error variance (Richards et al. 1997). This approach has Bayesian aspects because conditional on the value of the proportion of process error variance, point estimates are obtained by maximizing the posterior density (Schnute 1994). Second, a Bayesian approach with explicit priors on the variances was used in which the marginal posterior densities of the variances were estimated. I considered two variants of the Bayesian approach, one with an informative lognormal prior for log catchability variation and the second with only a weakly informative lognormal prior for this variation.

Ad Hoc *Estimation Model*

In the *ad hoc* approach, I estimated the variances using a technique developed by Richards et al. (1997). This approach requires repeated fits of the model with the proportion of total variance due to log catchability variance ρ:

$$\rho = \frac{\sigma_q^2}{\kappa^2}, \tag{5}$$

$$\kappa^2 = \sigma_q^2 + \sigma_C^2, \tag{6}$$

being varied among fits. During each fit of the model, total variance was estimated as a model parameter, and from this parameter the variances of log catchability σ_q^2 and log total catch σ_C^2 were calculated as follows:

$$\sigma_q^2 = \rho\kappa^2, \tag{7}$$

$$\sigma_C^2 = (1-\rho)\kappa^2. \tag{8}$$

I varied the proportion of log catchability variance from 0.1 to 0.95 in increments of 0.05, and refit the estimation model to a given data set for each value of ρ. I chose as best among these model fits the one where the predicted standard deviation of log total catch was closest to the "observed" point estimate of the log total catch standard deviation created by the data generating model.

For a given model fit (specific ρ) using the *ad hoc* approach, highest posterior density estimates of the parameters (a widely used approach, see Schnute 1994) were obtained by maximizing the posterior density of the parameters conditional on the observed data (Equations 2.3.1, 2.3.2a, and 2.3.3). I chose to minimize the negative log posterior density (Equation 2.3.4) for ease of computation.

The probability density of the data conditional on the parameters was separated into two components for total annual catch and proportion of catch-at-age (Equation 2.3.5). Total annual catch was assumed to follow a lognormal distribution, with the log density (ignoring some additive constants) given by Equation 2.3.6. Proportion of catch-at-age was assumed to follow a distribution that would arise if N_E fish were observed, with numbers observed at each age following a multinomial distribution, with the log density (ignoring some additive constants) given by Equation 2.3.7. Note that the probability density of the data conditional on the parameters is equivalent to the classical likelihood function. Therefore, the highest posterior density parameter estimates are equivalent to the maximum likelihood estimates.

The prior probability density of the parameters was separated into three components for the general model parameters ϕ, catchability deviations ε_q, and total

variance κ^2 (Equation 2.3.8a). Deviations in catchability were assumed to follow a lognormal distribution, with the log prior density (ignoring some additive constants) given by Equation 2.3.9. The prior densities of the log of all model parameters in ϕ and κ^2 were assigned proper uniform prior densities, which follows common practice with the intent of being weakly informative. Therefore, prior density of the log of ϕ and κ were constants for all parameter values.

Bayesian Estimation Models

In the Bayesian approaches statistical inference was made on the posterior density of the parameters conditional on the observed data (Equation 2.3.1) which was derived using a Markov Chain Monte Carlo (MCMC) method. I chose to work with the negative log posterior density for ease of computation (Equation 2.3.4). The standard deviations of log-scale catchability and total catch were included as parameters to be estimated in the model (Equation 2.3.2b). The probability density of the data conditional on the parameters was separated into two components for total annual catch and proportion of catch-at-age (Equation 2.3.5). The log densities for each of the components were the same as in the *ad hoc* estimation model (Equations 2.3.6 and 2.3.7).

The prior probability density of the parameters was separated into four components for the prior probability densities of the general model parameters ϕ, catchability deviations ε_q, log catchability standard deviation σ_q, and log total catch standard deviation σ_C (Equation 2.3.8b). Deviations in catchability were assumed to follow a lognormal distribution as in the *ad hoc* estimation model (Equation 2.3.9). In

the first version of the full Bayesian approach, hereafter referred to as the informative Bayesian approach, the standard deviations for log total catch and log catchability were assumed to follow a lognormal distribution, with log prior density (ignoring some additive constants) expressed as:

$$\ln[p(\sigma_i)] = -\frac{1}{2\sigma_{\sigma_i}^2}\left(\ln\sigma_i' - \ln\sigma_i\right)^2 - \ln\sigma_{\sigma_i}, \quad (9)$$

where i indexes the two error sources (i.e., total catch and catchability). The values for the prior standard deviations for the standard deviations of log total catch and log catchability were the same values used to create the true standard deviations in the data generating model. The prior densities of the log of all general model parameters ϕ were assigned weakly informative proper uniform prior densities. Therefore, prior density of the log of ϕ was a constant for all parameter values.

Marginal posterior densities for the standard deviations of total catch and catchability were estimated using a MCMC method. The highest posterior density parameter estimates served as starting values for the MCMC chain. A Metropolis-Hastings algorithm with a scaled multivariate normal candidate generating distribution was used to determine the marginal posterior densities (Gelman et al. 2004). The MCMC chain was run for 500,000 cycles with values being saved every 25th cycle. The first 2,000 saved cycles of the MCMC chain were dropped as a burn-in period, in order to remove the effect of the starting values (Gelman et al. 2004).

In reality, stock assessment analysts rarely have the data necessary to set such an informative prior on the standard deviation of log catchability as I did in the informative Bayesian estimation approach. Therefore, I also evaluated performance of the full

Bayesian method using a less informative prior. I refer to this as the objective Bayesian approach. This approach was identical to the informative Bayesian approach except that the prior density for the standard deviation of log catchability was assumed to follow a lognormal distribution (Equation 9) with mean (0.35) and variance (0.49) specified so that the prior density spanned all three levels of catchability variation.

Estimation Model Evaluation

My Monte Carlo simulation included six scenarios based on the three levels of catchability variation and two levels of total catch variation. Five hundred data sets were generated for each scenario for a total of 3,000 simulated data sets. Each estimation model was fit to each of the simulated data sets. Estimation model runs were dropped from the analysis if they exhibited poor convergence characteristics. After examining preliminary results, *ad hoc* estimation model convergence was judged to be poor if the maximum gradient component was greater than $1x10^{-4}$. After examining preliminary results, informative and objective Bayesian estimation model convergence was judged to be poor if the effective sample size for log catchability standard deviation, log total catch standard deviation, total abundance in the last year of analysis or highest posterior density value was less than 200. Effective sample sizes were calculated from MCMC chains using the method described by Thiebauz and Zwiers (1984) with lags out to 100 for autocorrelation calculations.

The three approaches for estimating process and observation errors were evaluated using the relative error (RE) of the standard deviations of log catchability, standard deviation of log total catch and total abundance in the last year of the analysis. The RE of the standard deviations of log catchability and log total catch indicated how

well the variances were estimated, while the RE of total abundance indicated how well the approaches estimated a key management quantity. Relative error was calculated as follows:

$$RE = \frac{\hat{X} - X}{X}, \tag{10}$$

where $\hat{X}$ is a point estimate of the quantity of interest from the estimation model, and X is the true value of the quantity of interest from the data generating model. For the Bayesian methods I used the median of the marginal posterior distribution as a point estimate, whereas for the ad hoc method the highest posterior density estimates were used. The median of the relative errors (MRE) was used to examine systematic bias in estimates from the estimation models. Median absolute relative error (MARE) , which captures elements of bias and precision, was used to compare the range of relative errors estimated by the estimation models.

Results

The following results are based on sample sizes of 500 model runs per scenario for the *ad hoc* approach, 380 to 431 model runs per scenario for the informative Bayesian approach, and 345 to 396 model runs per scenario for the objective Bayesian approach. The number of poorly converged model runs for the informative and objective Bayesian approaches is likely an artifact of my simulation study design. I had to limit the length of the MCMC chains to reduce computational times and make the study feasible. Under normal circumstances, an analyst would probably run longer MCMC chains or run multiple chains from different starting points to improve convergence properties.

The informative Bayesian approach outperformed the *ad hoc* and objective Bayesian approaches in the estimation of log total catch standard deviation. The informative Bayesian approach was less biased than *ad hoc* and objective Bayesian approaches in estimating standard deviation of log total catch (Figure 2.1). Informative Bayesian approach MRE values for all six scenarios were close to zero and ranged from -0.023 to 0.018. Objective Bayesian approach MRE values for all six scenarios exhibited positive bias and ranged from 0.021 to 0.153. *Ad hoc* approach MRE values exhibited negative bias and ranged from -0.338 to -0.019, except for the high catchability-low total catch variance scenario (0.004).

Informative and objective Bayesian approaches demonstrated higher levels of precision than the *ad hoc* approach in the estimation of log total catch standard deviation (Figure 2.1). The differences in MARE values between informative Bayesian and *ad hoc* approaches were small (-0.030 to 0.001) for medium catchability-low total catch, high catchability-low total catch, and high catchability-high total catch variance scenarios (Figure 2.2). The differences in MARE values between informative Bayesian and *ad hoc* approaches were larger (-0.249 to -0.114) for low catchability-low total catch, low catchability-high total catch, and medium catchability-high total catch variance scenarios (Figure 2.2). The differences in MARE values between objective Bayesian and *ad hoc* approaches were small (-0.074 to 0.050), except for the low catchability-high total catch variance scenario (-0.246) (Figure 2.3).

The informative Bayesian approach also out performed the *ad hoc* and objective Bayesian approaches in the estimation of the log catchability standard deviation. The informative Bayesian approach was less biased than the *ad hoc* and objective Bayesian

approaches in estimating the standard deviation of log catchability (Figure 2.4). Informative Bayesian approach MRE values for all six scenarios were close to zero, with a small positive bias, ranging from 0.002 to 0.023. Objective Bayesian approach MRE values generally were close to -1.0. The two exceptions were the medium catchability-low total catch variance and high catchability-low total catch variance scenarios for which the objective Bayesian approach MRE values were -0.048 and -0.064 respectively. *Ad hoc* MRE values were negatively biased and ranged from -0.758 to -0.377.

The informative Bayesian approach was more precise than the *ad hoc* and objective Bayesian approaches in estimating the standard deviation of log catchability (Figure 2.4), although all methods had much lower precision for estimating the standard deviation of catchability than for estimating the standard deviation of catch (note difference in scale between Figure 2.1 and Figure 2.4). The differences in MARE values between informative Bayesian and *ad hoc* approaches were substantial and ranged from -0.637 to -0.292 (Figure 2.2), where the informative Bayesian approach was more precise. The differences in MARE values between objective Bayesian and *ad hoc* approaches generally were large and ranged from 0.235 to 0.548 (Figure 2.3), where the *ad hoc* approach was more precise than the objective Bayesian approach. The two exceptions were the medium catchability-low total catch variance and high catchability-low total catch variance scenarios, -0.321 and -0.262 respectively, where the objective Bayesian approach was more precise than the *ad hoc* approach.

Differences in performance between *ad hoc*, informative and objective Bayesian approaches in the estimation of the total abundance in the last year of the analysis were less marked than for variance estimates. For all three methods, the bias of the estimates

of total abundance in the last year increased at high catchability and total catch variance levels (Figure 2.5). *Ad hoc* approach MRE values were negatively biased, and ranged from -0.271 to -0.033. Informative Bayesian approach MRE values generally were close to zero, positively biased, and ranged from 0.001 to 0.054. The one exception was the high catchability-high total catch variance scenario which was 0.107. Objective Bayesian approach MRE values generally were close to zero and ranged from -0.067 to 0.017. The one exception was the high catchability-high total catch variance scenario which was -0.124.

Precision of *ad hoc*, informative and objective Bayesian approach estimates of total abundance in the last year decreased as catchability and total catch variance levels increased (Figure 2.5). The *Ad hoc* approach was slightly less precise than the informative Bayesian approach, with differences in MARE values ranging from -0.024 to 0.011 (Figure 2.2). Differences between objective Bayesian and *ad hoc* approach MARE values were small and ranged from -0.015 to 0.059 (Figure 2.3).

Discussion

My results show that observation error variance will be more reliably estimated than process error variance in SCAA. Observation error variance is better estimated due to the availability of better prior information about observation errors. Estimates of observation error variance obtained separately from SCAA, through analysis of the raw data used to derive such quantities as observed total catch, provide a good source of prior information for estimating observation error variance in SCAA. Such prior information does not exist for process error variance because separate from SCAA the raw observed data are not informative about how model parameters such as catchability vary. This was

demonstrated in my study when the *ad hoc* and objective Bayesian approaches produced more accurate and precise estimates of log total catch standard deviation then of log catchability standard deviation. These two approaches used more weakly informative prior information or no prior information for log catchability standard deviation than for log total catch standard deviation.

Use of the Bayesian approach allows for reliable estimation of both observation and process error variances using a realistic, weakly informative prior for the process error variance, when process error variability is greater than observation error variability. Under this condition, the relatively strong informative prior for the observation error variance and the strong signal for the process errors in the observed data allow SCAA to reliably estimate the amount of total variance and successfully partition that variance between observation and process error variances. In my study, this was evident when the objective Bayesian approach was able to accurately estimate the log total catch and log catchability standard deviations in scenarios where annual variability in catchability was the dominant error source. Schnute and Richards (1995) found that, in general, their catch-at-age estimation models performed better in a Monte Carlo simulation when process error in recruitment was greater than observation error in an index of abundance. Their estimation models estimated the process and observation error variances by specifying the proportion of total variance due to process error variance, similar to the *ad hoc* approach, and obtaining maximum likelihood estimates of the variances analytically. Unfortunately, Schnute and Richards (1995) did not look specifically at how their estimation models performed at estimating the error variances. I hypothesize that process error variability likely will be greater than observation error variability, and hence the

associated variances can be estimated, in any well monitored commercial fishery, as well as most well monitored recreational fisheries. Chen and Paloheimo (1998) also have suggested that errors due to environmental variation (i.e., process errors) may be greater than observation errors for many fisheries. This finding emphasizes that another means of improving the estimation of process and observation error variances, as well as stock assessments in general, is to improve the quality of fishery monitoring data.

The *ad hoc* approach failed to reliably estimate the process and observation error variances in my study. I was not surprised by this finding since the *ad hoc* approach utilized the least amount of prior information (i.e., a single point estimate of log total catch standard deviation) to estimate both of the standard deviations. More interesting was the consistent underestimation of total variance in the system when using the *ad hoc* approach. This negative bias might in part be explained by the statistical properties of the estimator for the variances. Unlike the Bayesian approach which derived variance estimates from the median of the posterior probability density, the *ad hoc* approach simply used highest posterior density estimates to obtain variance estimates. Highest posterior density parameter estimates share many similar properties with likelihood-based parameter estimates, since highest posterior density estimates are obtained by maximizing the probability density of the data given the parameters $p(x|\theta)$ (Equation 2.3.1), which is identical to the likelihood function. Under this paradigm, the prior probability densities $p(\theta)$ could be thought of as penalty terms added to the likelihood function. The maximum likelihood estimator of variance is known to be negatively biased, thus the highest posterior density estimate of variance probably would possess the same negative bias.

The *ad hoc* approach produced unbiased estimates of the log total catch standard deviation in scenarios where catchability variation was greater than total catch variation, but this apparent success is deceptive and potentially dangerous for analysts. Even in the scenarios where catchability variation was dominant, the estimation model still underestimated the total variance as evidenced by the associated negative bias in estimates of the log catchability standard deviation. The estimation model was able to match predicted and observed log total catch standard deviation values by adjusting the proportion of total variance due to catchability variance, but the selected catchability variance proportion did not match the true proportion from the data generating model. In a real stock assessment where the true variances are unknown, such a result would lead the analyst to believe that the total variance had been well estimated when, in fact, it had been underestimated. This problem might be solved by correcting the predicted total variance by the number of parameters estimated in the model, thus producing an unbiased estimate of the total variance. Further study is needed to determine how well this total variance correction would work, but it has the potential of making the *ad hoc* approach a viable variance estimation technique.

I should point out that my study examined the ability of the *ad hoc* approach to estimate one form of process error (i.e., catchability variation). The only other published use of the *ad hoc* approach was to produce estimates of recruitment variability in a state-space age-structured model, but the approach was applied to actual fishery data and its performance was not quantified nor evaluated (Richards et al. 1997). The *ad hoc* approach can be classified with other methods that use residual model error to estimate associated variances, because the *ad hoc* approach employs the measured interannual

variation in the observed data as prior information for the stock assessment model estimates of the process and observation error variances. As another example of this class of methods, Francis et al. (2003) compared standard specified values of commercial CPUE and trawl survey CVs to resulting residual variation between observed values and stock assessment predictions of CPUE. This approach could be applied in an iterative method to obtain variance estimates. An initial variance value would be specified and the stock assessment model fit to the observed data. The resulting residual variation in model results would be used to specify a new variance value for the next model run. This process would be repeated until the specified variance value matched the resulting residual variation in model results. The assessment models used for lake whitefish in 1836 treaty waters have used a such an iterative approach to setting the variance associated with variability about an assumed stock-recruitment relationship (Ebener et al. 2005). The Francis et al. (2003) study examined actual data from New Zealand fisheries, and the whitefish assessments use actual data also, so it is unknown how accurately residual variation in stock assessment model results measures the true underlying variance. Further study of the ability of these residual-based variance estimation approaches to estimate other forms of process error variability, such as time-varying selectivity and annual recruitment variations, would be useful and informative.

The *ad hoc* and Bayesian approaches performed equally well at estimating numbers of fish in the last year of the analysis, even though the *ad hoc* approach consistently underestimated the process and observation error variances. In theory, the poor performance of the *ad hoc* approach in estimating the variances should have resulted in poorer estimates of the final number of fish. To address this issue, it is necessary to

consider when it is important to properly estimate the error variances. Methot (1990) suggested that when the different data sources used in SCAA do not trend over time, or when trends are consistent between data sources, assessment model results will be less sensitive to changes in the variances which are used to weight the different data sources. It is when trends in the different data sources are inconsistent with each other that assessment model results will be sensitive to changes in the variance values (Methot 1990). Therefore, it is most important to properly estimate the error variances when the data sources are sending mixed signals about the population dynamics to the assessment model. In my study, total catch did trend over time, but catchability did not since the catchability deviations were generated using a white noise model. If I had generated catchability so that it trended over time, then it is likely that I would have seen differences in the estimation model performances when it came to estimating the final number of fish. Such an analysis was beyond the scope of this study, since I wanted to determine if it were possible to estimate the error variances under the simplest conditions I could imagine. Actual stock assessments are generally more complex, incorporating multiple sources of observation and process error. As a result, I feel it would be informative to evaluate the Bayesian and *ad hoc* approaches when estimating more than two sources of variation.

I recommend that stock assessment analysts use the Bayesian approach when attempting to estimate process and observation error variances in SCAA. The Bayesian approach is fairly robust when existing data allow for the designation of strongly informative priors for the error variances, particularly process error variance. The Bayesian approach still can produce reliable estimates of the error variances with a

weakly informative prior for the process error variance, as long as high quality monitoring data are available. I do not recommend the use of the *ad hoc* approach based on my findings. The *ad hoc* approach consistently underestimates the error variances, which could lead to biased estimates of important management quantities when the different data sources send inconsistent signals concerning the dynamics of the population.

References

Chen, Y., and J. E. Paloheimo. 1998. Can a more realistic model error structure improve the parameter estimation in modeling the dynamics of fish populations? Fisheries Research 38: 9-17.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-at-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences 42: 815-824.

Ebener, M. P., J. R. Bence, K. Newman, and P. Schneeberger. 2005. Application of statistical catch-at-age models to assess lake whitefish stocks in the 1836 treaty-ceded waters of the upper Great Lakes. Pages 271-309 *in* L. C. Mohr and T. F. Nalepa, editors. Proceedings of a workshop on the dynamics of lake whitefish (*Coregonus clupeaformis*) and the amphipod *Diporeia* spp. in the Great Lakes. Great Lakes Fishery Commission, Technical Report 66, Ann Arbor.

Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39: 1195-1207.

Francis, R. I. C. C., R. J. Hurst, and J. A. Renwick. 2003. Quantifying annual variation in catchability for commercial and research fishing. Fishery Bulletin 101: 293-304.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian Data Analysis, 2nd edition. Chapman and Hall/CRC, Boca Raton, Florida.

Law, A. M., and W. D. Kelton. 1982. Simulation Modeling and Analysis. McGraw-Hill, New York.

Maunder, M. N. 2001. A general framework for integrating the standardization of catch per unit of effort into stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 58: 795-803.

Merritt, M. F., T. J. Quinn, II. 2000. Using perception of data accuracy and empirical weighting of information: assessment of a recreational fish population. Canadian Journal of Fisheries and Aquatic Sciences 57: 1459-1469.

Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. International North Pacific Fisheries Commission Bulletin 50: 259-277.

National Research Council. 1998. Improving Fish Stock Assessments. National Academy Press, Washington, D.C.

Otter Research Limited. 2004. An introduction to AD Model Builder version 7.0.1 for use in nonlinear modeling and statistics. Otter Research Limited, Sidney, British Columbia.

Otter Research Limited. 2005. Random effects in AD Model Builder: ADMB-RE user guide. Otter Research Limited, Sidney, British Columbia.

Quinn II, T. J., and R. B. Deriso. 1999. Quantitative Fish Dynamics. Oxford University Press, New York.

Richards, L. J., J. T. Schnute, and N. Olsen. 1997. Visualizing catch-age analysis: a case study. Canadian Journal of Fisheries and Aquatic Sciences 54: 1646-1658.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Fisheries Research Board of Canada Bulletin 191.

Schnute, J.T. 1994. A general framework for developing sequential fisheries models. Canadian Journal of Fisheries and Aquatic Sciences 51: 1676-1688.

Schnute, J. T., and L. J. Richards. 1995. The influence of error on population estimates from catch-age models. Canadian Journal of Fisheries and Aquatic Sciences 52: 2063-2077.

Sitar, S. P., J. R. Bence, J. E. Johnson, M. P. Ebener, and W. W. Taylor. 1999. Lake trout mortality and abundance in southern Lake Huron. North American Journal of Fisheries Management 19: 881-900.

Thiebauz, H. J., and F. W. Zwiers. 1984. The interpretation and estimation of effective sample size. Journal of Climate and Applied Meteorology 23: 800-811.

Table 2.1. Symbols and descriptions of variables used in data generating and estimation models.

Symbols	Description	Application
$C_{y,a}$	Number of fish caught by year and age	Both
$\widetilde{C}_y$	Observed number of fish caught by year	Both
E_y	Fishery effort by year	Both
$F_{y,a}$	Instantaneous fishing mortality by year and age	Both
M	Instantaneous natural mortality	Both
$N_{y,a}$	Abundance by year and age	Both
N_0	Mean abundance for abundance in first year	Estimation
N_E	Number of fish used to calculate age composition each year	Both
$P_{y,a}$	Proportion of catch by year and age	Both
$\widetilde{P}_{y,a}$	Observed proportion of catch by year and age	Both
R_0	Mean recruitment	Estimation
S_y	Number of female spawners by year	Generation
$Z_{y,a}$	Instantaneous total mortality by year and age	Both
$Z_{0,a}$	Instantaneous total mortality for abundance in first year by age	Generation
b_1	First inflection point of double logistic selectivity function	Estimation
b_2	First slope of double logistic selectivity funcion	Estimation
b_3	Second inflection point of double logistic selectivity function	Estimation
b_4	Second slope of double logistic selectivity funcion	Estimation
m	Total number of ages	Both
n	Total number of years	Both
$p(\theta\|x)$	Posterior probability density of parameters conditional on data	Estimation
$p(x\|\theta)$	Probability density of data conditional on parameters	Estimation
$p(\theta)$	Prior probability density of parameters	Estimation
q_y	Fishery catchability by year	Both
$\overline{q}$	Median catchability	Both
s_a	Fishery selectivity by age	Both
α	Productivity parameter of Ricker recruitment function	Generation

Table 2.1 (cont'd).

β	Density dependent parameter of Ricker recruitment function	Generation
$\varepsilon_{R,y}$	Process error in recruitment by year	Generation
$\varepsilon_{q,y}$	Process error in catchability by year	Both
$\varepsilon_{C,y}$	Observation error in number of fish caught by year	Both
ϕ	Subset of model parameters common to both estimation models	Estimation
κ^2	Total variance	Estimation
μ_N	Mean age-1 abundance for abundance in first year	Generation
θ	Set of all model parameters	Estimation
ρ	Proportion of total variance due to catchability variance	Estimation
σ_N	Standard deviation of age-1 abundance for abundance in first year	Generation
σ_R	Standard deviation of log-scale recruitment	Generation
σ_q	Standard deviation of log-scale catchability	Both
σ'_q	Observed point estimate of log catchability standard deviation	Both
σ_{σ_q}	Log-scale standard deviation of log catchability standard deviation	Generation
σ_C	Standard deviation of log-scale total catch	Both
σ'_C	Observed point estimate of log total catch standard deviation	Both
σ_{σ_C}	Log-scale standard deviation of log total catch standard deviation	Generation
υ_y	Process error in recruitment by year	Estimation
ψ_a	Process error for abundance in first year by age	Estimation
ζ_q	Observation error in log catchability standard deviation	Generation
ζ_C	Observation error in log total catch standard deviation	Generation

Table 2.2. Data generating and estimation model equations.

	Equations	Application
2.2.1	$N_{y+1,a+1} = N_{y,a}e^{-Z_{y,a}}$	Both
2.2.2	$N_{1,a} = N_{2-a,1}e^{-\sum_{j=1}^{a-1} Z_{0,j}}$; for $a > 1$	Generation
2.2.3	$N_{y,1} = \alpha S_{y-1}e^{-\beta S_{y-1}}e^{\varepsilon_{R,y}}$; $\varepsilon_{R,y} \sim N\left(0,\sigma_R^2\right)$	Generation
2.2.4	$Z_{y,a} = M + F_{y,a}$	Both
2.2.5	$F_{y,a} = s_a q_y E_y$	Both
2.2.6	$C_{y,a} = \frac{F_{y,a}}{Z_{y,a}}\left(1 - e^{-Z_{y,a}}\right)N_{y,a}$	Both
2.2.7	$P_{y,a} = \frac{C_{y,a}}{C_y}$	Both
2.2.8	$s_a = \frac{1}{1+e^{-b_2(a-b_1)}}\left(1 - \frac{1}{1+e^{-b_4(a-b_3)}}\right)$	Estimation

Table 2.3. Posterior probability density equations for estimation models.

	Equations	Application
2.3.1	$p(\theta\|x) \propto p(x\|\theta)p(\theta)$	Both
2.3.2a	$\theta = \left\{\phi, \left[\varepsilon_{q,y}\right]_{y=1}^{n}, \kappa\right\}$	Ad hoc
2.3.2b	$\theta = \left\{\phi, \left[\varepsilon_{q,y}\right]_{y=1}^{n}, \sigma_q, \sigma_C\right\}$	Bayesian
2.3.3	$\phi = \left\{N_0, \left[\psi_i\right]_{i=1}^{m-1}, R_0, \left[\upsilon_i\right]_{i=1}^{n} \bar{q}, b_1, b_2, b_3, b_4\right\}$	Both
2.3.4	$-\ln\left[p(\theta\|x)\right] \propto -\ln\left[p(x\|\theta)\right] - \ln\left[p(\theta)\right]$	Both
2.3.5	$\ln\left[p(x\|\theta)\right] = \ln\left[p(C\|\theta)\right] + \ln\left[p(P\|\theta)\right]$	Both
2.3.6	$\ln\left[p(C\|\theta)\right] = -\frac{1}{2\sigma_C^2}\sum_{y=1}^{n}\left[\left(\ln\widetilde{C}_y - \ln C_y\right)^2\right] - n\ln\sigma_C$	Both
2.3.7	$\ln\left[p(P\|\theta)\right] = \sum_{y=1}^{n} N_E \sum_{a=1}^{m}\left(\widetilde{P}_{y,a}\ln P_{y,a}\right)$	Both
2.3.8a	$\ln\left[p(\theta)\right] = \ln\left[p(\phi)\right] + \ln\left[p(\varepsilon_q)\right] + \ln\left[p(\kappa)\right]$	Ad hoc
2.3.8b	$\ln\left[p(\theta)\right] = \ln\left[p(\phi)\right] + \ln\left[p(\varepsilon_q)\right] + \ln\left[p(\sigma_q)\right] + \ln\left[p(\sigma_C)\right]$	Bayesian
2.3.9	$\ln\left[p(\varepsilon_q)\right] = -\frac{1}{2\sigma_q^2}\sum_{y=1}^{n}\left[\varepsilon_{q,y}^2\right] - n\ln\sigma_q$	Both

Table 2.4. Values of variables used in data generating model to create simulated data sets.

Variable	Level	Value
n		20
m		8
μ_N		355000
σ_N		0.4
α		10.1
β		5.10E-06
σ_R		0.4
M		0.24
E_y		0.1, 2.0, 3.0, 3.1, 3.3, 3.7, 4.4, 5.3, 6.5, 8.0, 8.0, 6.5, 5.3, 4.4, 3.7, 3.3, 3.1, 3.0, 2.0, 0.1
s_a		0.04, 0.15, 0.43, 0.85, 1.00, 0.82, 0.57, 0.37
$\bar{q}$		0.15
$\bar{\sigma}_q$	Low	0.2
	Medium	0.5
	High	0.8
σ_{σ_q}		0.2
$\bar{\sigma}_C$	Low	0.25
	High	0.75
σ_{σ_C}		0.2
N_E		100

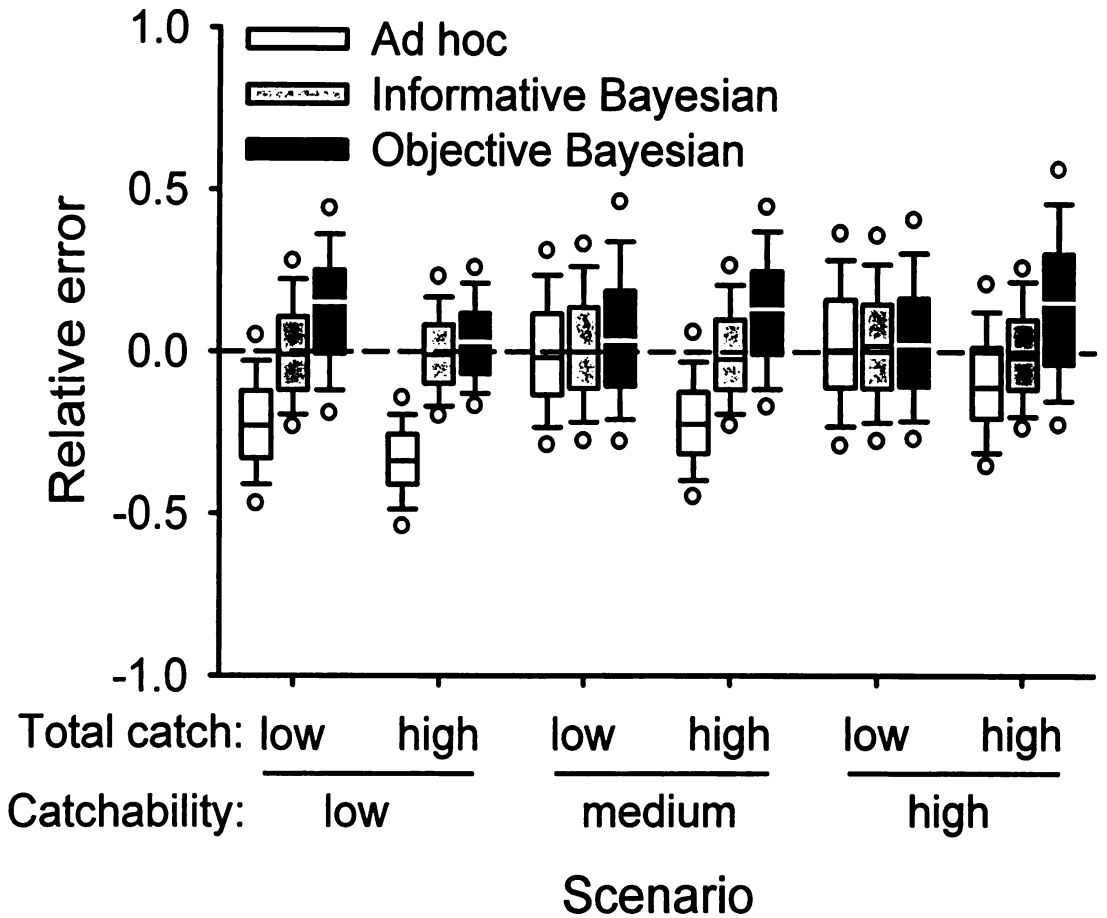


Figure 2.1. Box plots representing relative error distributions for estimates of log total catch standard deviation across different levels of catchability and total catch variance. The bars represent median relative errors. The boxes, whiskers, and circles represent 25th and 75th, 10th and 90th, and 5th and 95th percentiles of the distributions, respectively.

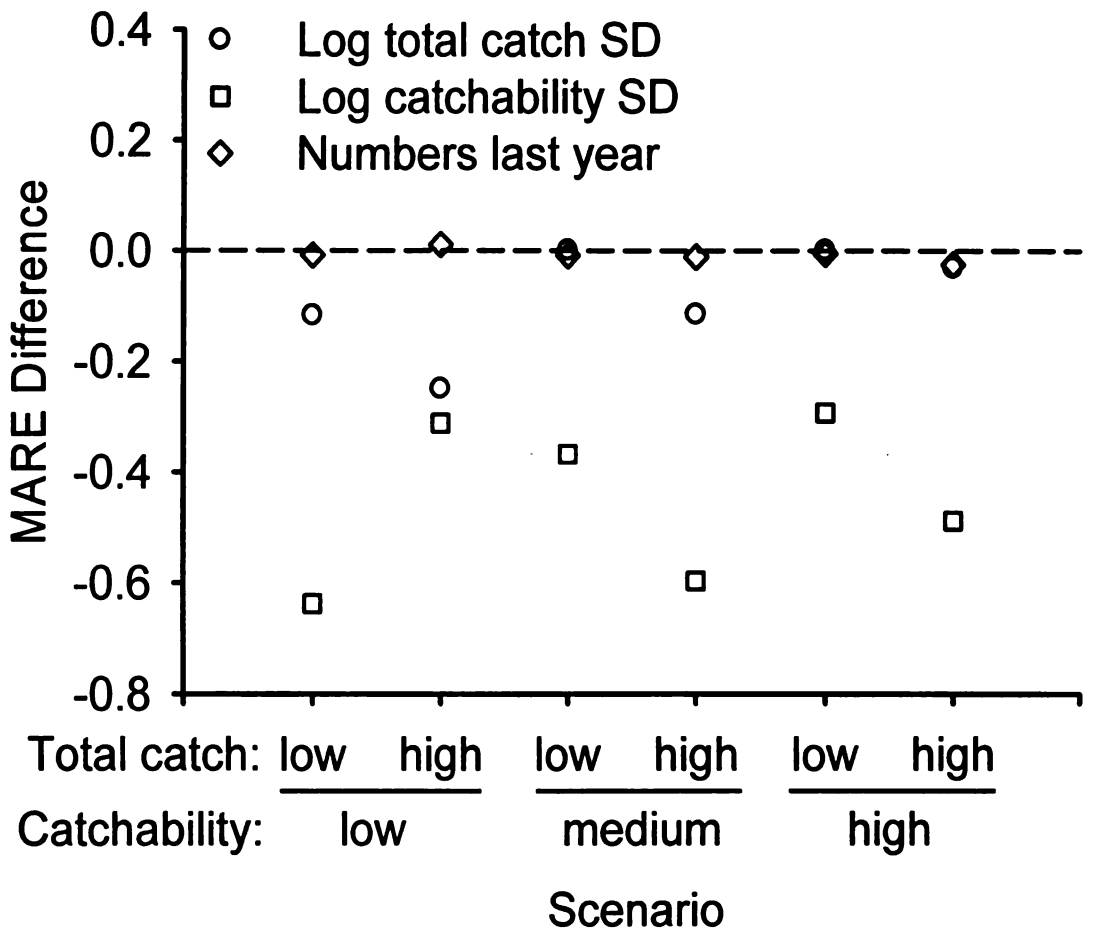


Figure 2.2. Differences in median absolute relative errors (MARE) between informative Bayesian approach and *ad hoc* approach across different levels of catchability and total catch variance. Symbols represent informative Bayesian approach MARE values minus *ad hoc* approach MARE values.

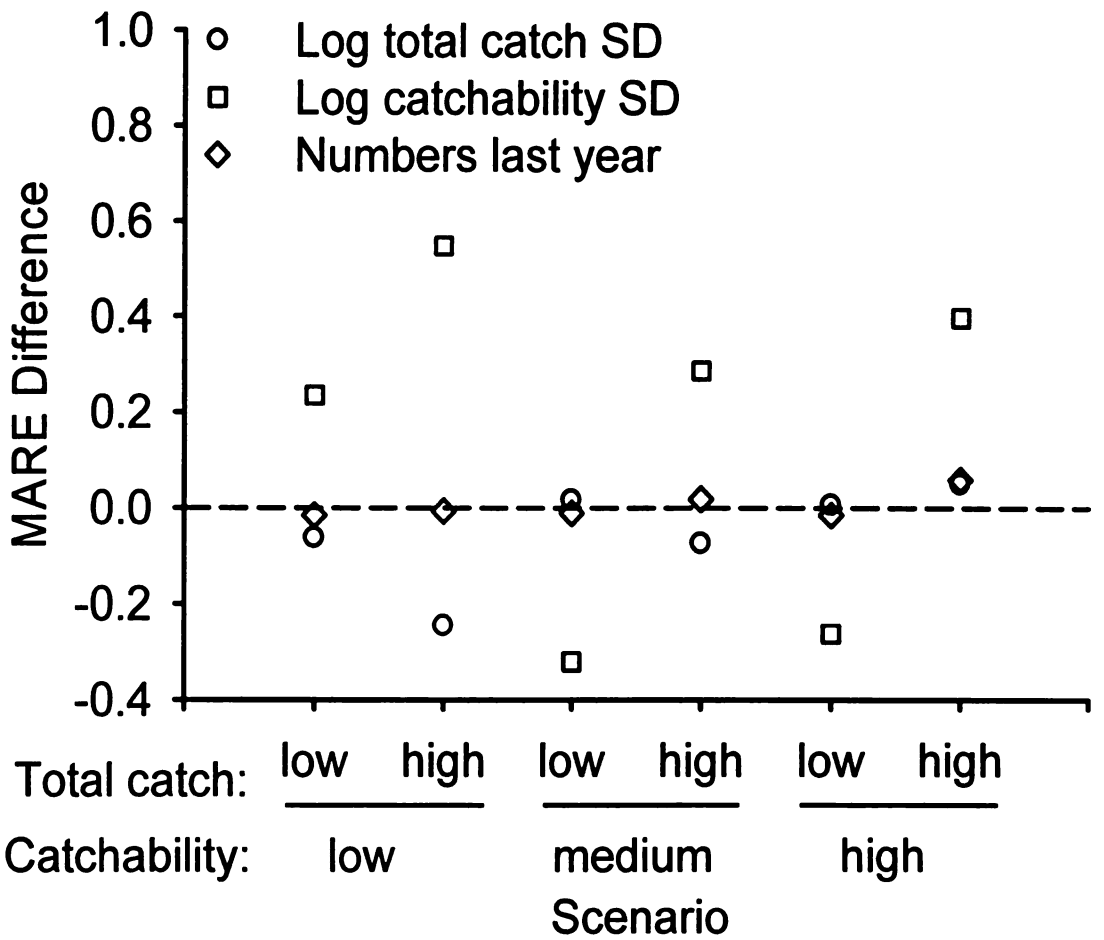


Figure 2.3. Differences in median absolute relative errors (MARE) between the objective Bayesian approach and the *ad hoc* approach across different levels of catchability and total catch variance. Symbols represent objective Bayesian approach MARE values minus *ad hoc* approach MARE values.

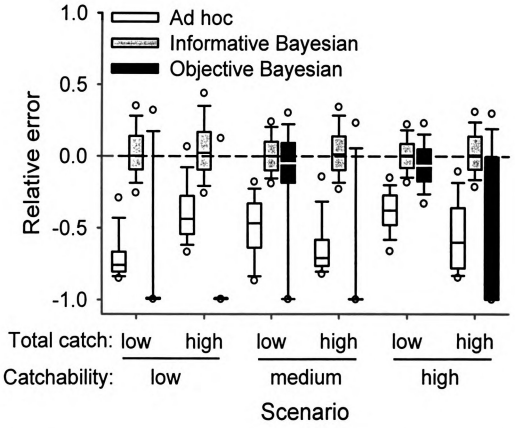


Figure 2.4. Box plots representing relative error distributions for estimates of log catchability standard deviation across different levels of catchability and total catch variance. The bars represent median relative errors. The boxes, whiskers, and circles represent 25th and 75th, 10th and 90th, and 5th and 95th percentiles of the distributions, respectively.

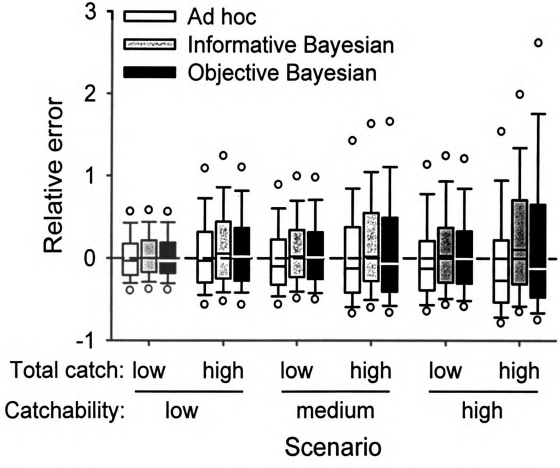


Figure 2.5. Box plots representing relative error distributions for estimates of total abundance in the last year of analysis across different levels of catchability and total catch variance. The bars represent median relative errors. The boxes, whiskers, and circles represent 25th and 75th, 10th and 90th, and 5th and 95th percentiles of the distributions, respectively.

CHAPTER 3

EVALUATING AND SELECTING METHODS FOR ESTIMATING TIME-VARYING SELECTIVITY IN STATISTICAL CATCH-AT-AGE ANALYSIS

Introduction

Statistical catch-at-age analysis (SCAA) is a common method of fisheries stock assessment. Age-structured catch data from a fishery are used to estimate quantities of interest, such as population abundance and mortality rates, using likelihood methods (Fournier and Archibald 1982). Auxiliary data that provide an index of abundance either directly or indirectly, such as survey catch-per-unit-effort (CPE) or fishery effort, are essential for reliable estimation (Deriso et al. 1985; Methot 1990). Estimated population quantities from the last year of the analysis are typically used as a starting point for short-term projections that are the basis for recommending harvest limits or targets.

In many SCAA models fishing mortality is assumed to be separable into year and age effects, with their product being the fishing mortality rate for a given year and age (Doubleday 1976). Here I refer to the year effect as fishing intensity, and to the age effect as fishery selectivity. Selectivity refers to the relative vulnerability of specific ages of fish to a fishery, so that age classes that are highly selected tend to be overrepresented in the catch in comparison to their relative abundance in the population. Selectivity is influenced by fishing gear characteristics, as well as fishing and fish behavior.

Selectivity often is modeled either as a function of age or it is allowed to vary freely among ages. The parameters of the selectivity function or the selectivity values for each age are estimated within SCAA along with other model parameters. Logistic

(Millar 1995; Punt et al. 2001), double logistic (Methot 1990; Ebener et al. 2005), exponential-logistic (Thompson 1994), normal (Millar 1995), lognormal (Millar 1995), gamma (Deriso et al. 1985; Millar 1995), and polynomials (Fournier 1983) are some of the functions used to model selectivity. Regardless of how selectivity is modeled, a restriction often must be applied to ensure a unique parameterization of the age and year effects (Doubleday 1976). Selectivity functions generally are constrained by normalizing the function to a reference age or to the age of maximum estimated selectivity. When selectivity is allowed to vary freely with age, selectivity commonly is constrained by setting selectivity at some reference age(s) equal to one.

The separability assumption can be relaxed, allowing selectivity to change over time, when there is evidence to suggest that selectivity is not constant (i.e., gear characteristics or fish behavior have changed). Separate selectivity values can be estimated for different blocks of time within SCAA (Radomski et al. 2005). Some of the selectivity function's parameters can vary over time independently from year to year (Bence and Rogers 1993), according to a polynomial in time (Ebener et al. 2005) or random walk process (Gudmundsson 1994; Ianelli 1996). Nonadditive models have been used to allow selectivity to vary with changes in fishing mortality (Myers and Quinn 2002; Radomski et al. 2005).

Statistical catch-at-age analysis has been shown to be sensitive to the choice of how selectivity is modeled. Incorrect assumptions about selectivity have been shown to generate errors in SCAA estimates of biomass (Kimura 1990), spawning biomass (Punt et al. 2002; Radomski et al. 2005), exploitation rate (Radomski et al. 2005), and the ratio of

stock biomass in the first year to the stock biomass in the final year of analysis (Yin and Sampson 2004).

Radomski et al. (2005) looked specifically at how specification of time-varying selectivity affected SCAA. They compared three methods for estimating selectivity: constant, time-blocked and nonadditive and found no one method for estimating time-selectivity performed best in all situations, but they did discover that time-varying selectivity SCAA models performed as well as constant selectivity SCAA models when selectivity was constant, and outperformed constant selectivity SCAA models when selectivity varied with time. They speculated that allowing selectivity to vary according to a random walk might improve the estimation of time-varying selectivity (Radomski et al. 2005). Radomski et al. (2005) also recommended that research was needed to determine the extent to which correct or adequate selectivity models could be identified.

The objective of my study was to compare the performance of different time-varying selectivity functions within SCAA. In addition, I strove to identify a model selection method that could allow analysts to select among alternative time-varying selectivity functions within a specific SCAA. This contrasts with an objective of determining a single "best" time-varying selectivity estimation method, which works well in most situations. Of course, one possible outcome of my work could have been that an omnibus procedure for modeling selectivity works better than selecting among alternatives. I addressed my objectives through Monte Carlo simulations, in which I evaluated different methods of both modeling time-varying selectivity within a stock assessment and selecting among the methods.

Methods

I used Monte Carlo simulations to compare four time-varying selectivity estimation methods and evaluate three model selection techniques. I used a data generating model to simulate data sets from a hypothetical fish population. The data generating model used two different approaches to simulate time-varying selectivity: 1) a double logistic function in which the first inflection point varied according to a first order autoregressive process, and 2) selectivity for each age varied independently according to a first order autoregressive process. I chose these two approaches to provide contrast in how freely selectivity varies over time. The double logistic function is constrained so that only selectivity of younger age fish changes over time. The age-specific selectivity parameters allow selectivity to vary more freely, with age-specific selectivity values changing independently of each other. I fit four estimation models, each using a different time-varying selectivity estimation method, to the simulated data sets. The selectivity estimation methods consisted of 1) a double logistic function in which the first inflection point varied according to a random walk, 2) a double logistic function in which the first and second inflection points varied according to random walks, 3) a double logistic function in which all four parameters varied according to random walks, and 4) selectivity for each age varied according to a random walk with a smoothing function across ages. I chose these estimation approaches because they represent the two general approaches for estimating selectivity in SCAA, namely modeling selectivity as a function of age and estimating age-specific selectivity parameters. In addition, these four estimation approaches form a continuum of increasing flexibility in how selectivity is allowed to vary over time. The three model selection techniques included 1) root mean

square error (RMSE), 2) degree of retrospectivity, and 3) the Deviance Information Criterion (DIC). The data generating model and four estimation models were all built using AD Model Builder software (Otter Research Limited 2004). For the following discussion, descriptions of all the symbols are given in Table 3.1, while most of the equations describing my models are given in Tables 3.2 and 3.3. I reference equations as Equation *x.y*, where equation *y* is found in Table *x*.

My Monte Carlo simulation included two scenarios based on two different methods for generating time-varying selectivity. Five hundred data sets were generated for each scenario for a total of 1,000 simulated data sets. Each of the four estimation models was fit to each of the simulated data sets. I applied the three model selection techniques to each estimation model fit to a simulated data set.

Data Generating Model

I developed a data generating model to simulate the dynamics of a hypothetical fish population based on lake whitefish stocks in the upper Great Lakes. The population dynamics were described using abundance-at-age and age-specific mortality rates created by the model. A gill net fishery operating on the population produced observed total annual catch, age composition and fishing effort data. Each simulated data set included 20 years of data for fish ages 1 to 8+, where 8+ is a plus group containing all fish age-8 and older.

I generated abundance-at-age using an exponential population equation (Equation 3.2.1). In order to produce abundance-at-age in the first year, mortality was applied to randomly generated numbers of age-1 fish (Equation 3.2.2). The number of age-1 fish was randomly drawn from a lognormal distribution (Table 3.4). I selected the mean of

the distribution by assuming the population experienced equilibrium recruitment prior to the simulated time series. I calculated recruitment to the first age in each year with a Ricker stock-recruitment function (Equation 3.2.3; Table 3.4). I calculated the number of female spawners as one-half of the number of fish age-3 and older, thereby assuming knife-edge maturity and a 1:1 sex ratio.

I partitioned total mortality into natural and fishing mortality components (Equation 3.2.4). Natural mortality was a constant value for all years and ages (Table 3.4). I modeled fishing mortality by relaxing the assumption of full separability (Equation 3.2.5). I generated year and age-specific selectivity using two different methods to create a dome-shaped selectivity curve, which is typical of gill net fisheries. I defined fishing intensity as a function of fishing effort (Equation 3.2.6). The errors associated with fishing intensity were a combination of process error due to annual variation in catchability and observation error in nominal fishing effort. I assumed that variation in catchability would outweigh observation error in fishing effort and, therefore, treat the fishing intensity errors as process error. The value for the standard deviation of log fishing intensity σ_λ was randomly generated from a lognormal distribution for each simulated data set (Table 3.4). I specified fishing effort so that effort increased to a maximum in the middle of the time series and then decreased to the end of the time series (Table 3.4). This fishing effort pattern simulated a growing fishery during the first half of the time series that was regulated by effort limits during the second half of the time series. I generated the total mortality used to produce abundance-at-age in the first year Z_0 (Equation 3.2.2) using Equations 3.2.4, 3.2.5 and 3.2.6 with the assumption that fishing effort in years prior to the first year of the analysis was equal to fishing effort in

the first year of the analysis, and selectivity in years prior to the first year of the analysis was constant at the initial values.

The two methods I chose to generate time-varying selectivity provide contrast in how selectivity changes over time. For the first method, I generated selectivity using a double logistic function of age (Methot 1990):

$$s_a = \frac{1}{1+e^{-b_2(a-b_{1,y})}}\left[1-\frac{1}{1+e^{-b_4(a-b_3)}}\right]. \quad (1)$$

I varied the first inflection point over time from an initial value according to a first order autoregressive process (Table 3.4): $\delta_y \sim N(0,\sigma_\delta^2)$

$$\log_e b_{1,y+1} = \log_e b_1' + \rho_1(\log_e b_{1,y} - \log_e b_1') + \delta_y, \quad (2)$$

$$\delta_y \sim N(0,\sigma_\delta^2).$$

I randomly drew the initial value of the first inflection point from a lognormal distribution with mean $\bar{b}_1$ and standard deviation σ_δ. The value for the standard deviation of log first inflection point σ_δ was randomly generated from a lognormal distribution for each simulated data set (Table 3.4). I normalized age-specific selectivity in a given year using the maximum generated age-specific selectivity value for that year. By allowing the first inflection point to vary over time, I was simulating a scenario in which the vulnerability of young fish to the fishery was changing over time. For the second method, I chose a more flexible approach to generating time-varying selectivity based on a method used by Butterworth et al. (2003). In this approach, age-specific selectivity varied over time from initial values according to a first order autoregressive process (Table 3.4):

(3) $$\log_e s_{y+1,a} = \log_e s'_a + \rho_2\left(\log_e s_{y,a} - \log_e s'_a\right) + \gamma_{y,a},$$

$$\gamma_{y,a} \sim N\left(0, \sigma_\gamma^2\right).$$

I used the same correlation and standard deviation parameters for all ages. I randomly drew the initial values for selectivity at each age from lognormal distributions with means $\bar{s}_a$ and standard deviation σ_γ. The value for the standard deviation of log selectivity σ_γ was randomly generated from a lognormal distribution for each simulated data set (Table 3.4). I normalized age-specific selectivity in a given year using the maximum generated age-specific selectivity value for that year. By allowing age-specific selectivity values to vary over time, I simulated a scenario in which the vulnerability of all age classes of fish to the fishery changed independently over time.

I generated observed data from a gill net fishery from simulated abundance-at-age and mortality rates. I calculated catch-at-age using Baranov's catch equation (Equation 3.2.7). I calculated observed total annual catch by summing catch-at-age across ages for each year and incorporating observation errors (Equation 3.2.8; Table 3.4). The value for the standard deviation of log total catch σ_ν was randomly generated from a lognormal distribution for each simulated data set (Table 3.4). I generated observed fishery age composition data by drawing a random sample from a multinomial distribution with a sample size of 400, and proportions calculated from catch-at-age in the fishery (Equation 3.2.9). Natural mortality and observed fishing effort were known without error (Table 3.4).

Estimation Models

The estimation models used the same equations as the data generating model except when estimating abundance-at-age in the first year, recruitment, and selectivity. I estimated annual recruitment as a mean recruitment parameter and a vector of annual recruitment deviation parameters (i.e., a vector of deviations that must sum to zero). I estimated abundance-at-age in the first year as a mean abundance parameter and a vector of abundance deviation parameters (i.e., a vector of deviations that must sum to zero). I calculated abundance-at-age (Equation 3.2.1), total mortality (Equation 3.2.4), fishing mortality (Equation 3.2.5), fishing intensity (Equation 3.2.6), catch-at-age (Equation 3.2.7), total catch (Equation 3.2.8), and proportion of catch-at-age (Equation 3.2.9) using the equations described for the data generating model. I used true parameter values produced by the data generating model as starting values for parameters in the estimation models to expedite numerical convergence during simulations.

The estimation models differed from each other in the method used to estimate time-varying selectivity for the fishery. The four methods I chose represent increasing flexibility in the estimation of time-varying selectivity. The cost associated with increased flexibility is an increase in the number of parameters that must be estimated. In the first estimation approach, I allowed the first inflection point of the double logistic function to vary over time according to a random walk:

$$s_a = \frac{1}{1+e^{-b_2\left(a-b_{1,y}\right)}}\left[1-\frac{1}{1+e^{-b_4\left(a-b_3\right)}}\right], \tag{4}$$

$$\log_e b_{1,y+1} = \log_e b_{1,y} + \eta_y,\ \eta_y \sim N\left(0,\sigma_\eta^2\right).$$

This approach is the least flexible of those I examined since it only changes the lower ages at which selectivity increases most rapidly over time. In the second estimation approach, I allowed the first and second inflection points of the double logistic function to vary over time according to random walks:

$$(5) \qquad s_a = \frac{1}{1+e^{-b_2(a-b_{1,y})}}\left[1-\frac{1}{1+e^{-b_4(a-b_{3,y})}}\right],$$

$$\log_e b_{i,y+1} = \log_e b_{i,y} + \eta_{i,y},\ \eta_{i,y} \sim N\left(0,\sigma_\eta^2\right),$$

where i indexes the inflection points of the double logistic function (i.e., $b_{1,y}$ and $b_{3,y}$). I made the simplifying assumption that the standard deviations of the two log-scale inflection points were equal. This approach of varying the two inflection points allows the ascending and descending limbs of the selectivity curve to expand and contract over the course of time. In the third estimation approach, I allowed the two inflection points and the two slopes of the double logistic function to vary over time according to random walks:

$$(6) \qquad s_a = \frac{1}{1+e^{-b_{2,y}(a-b_{1,y})}}\left[1-\frac{1}{1+e^{-b_{4,y}(a-b_{3,y})}}\right],$$

$$\log_e b_{i,y+1} = \log_e b_{i,y} + \eta_{i,y}, \eta_{i,y} \sim N\left(0,\sigma_\eta^2\right),$$

$$\log_e b_{j,y+1} = \log_e b_{j,y} + \tau_{j,y}, \tau_{j,y} \sim N\left(0,\sigma_\tau^2\right),$$

where i indexes the inflection points and j indexes the slopes (i.e., $b_{2,y}$ and $b_{4,y}$) of the double logistic function. Again, as I did for the infection points, I assumed that the standard deviations of the two log-scale slopes were equal. This approach of allowing all of the double logistic function parameters to vary over time provides maximum flexibility

in the estimation of time-varying selectivity for this functional form. In the fourth estimation approach, I allowed age specific selectivity values to vary over time according to random walks (Butterworth 2003):

$$\log_e s_{y+1,a} = \log_e s_{y,a} + \varpi_{y,a}, \tag{7}$$

$$\varpi_{y,a} \sim N\left(0, \sigma_{\varpi}^2\right).$$

I made the simplifying assumption that the year-specific standard deviations of log selectivity were equal for all years. I constrained age-specific selectivity with a curvature penalty using squared third-differences to ensure smoothness in selectivity across age classes (Butterworth 2003):

$$g\left(s_{y,a};\sigma_{\varphi}^2\right) = \sum_{y=1}^{n}\sum_{a=1}^{m-3} \frac{\left(\log_e s_{y,a+3} - 3\log_e s_{y,a+2} + 3\log_e s_{y,a+1} - \log_e s_{y,a}\right)^2}{2\sigma_{\varphi}^2}. \tag{8}$$

I made the simplifying assumption that the age-specific standard deviation of log-scale selectivity was the same for all ages. I added this curvature penalty term to the negative log posterior density. In all four time-varying selectivity estimation approaches, I normalized age-specific selectivity using the maximum estimated age-specific selectivity value. I estimated the variances associated with log total catch, log fishing intensity and log selectivity using a Bayesian approach in which the marginal posterior densities were estimated with Markov Chain Monte Carlo simulations.

I made statistical inference on the posterior density of the parameters conditional on the observed data (Equation 3.3.1) which was derived using a Markov Chain Monte Carlo (MCMC) method. More specifically, I used MCMC with the Metropolis-Hastings algorithm as it is implemented in AD Model Builder (Otter Research Limited 2004). Maximum likelihood parameter estimates were used as starting values for each MCMC

chain. I ran the MCMC chain for each model for 1,000,000 cycles, saving parameter values every 10th cycle. I dropped the first 40,000 cycles from the chain of saved MCMC values as a burn in period, which reduced the effect of chain starting values on final MCMC estimates (Gelman et al. 2004). I dropped model runs with poor convergence properties from the analysis. I judged MCMC chain convergence to be poor if the effective sample size for the highest posterior density value was less than 300. I selected the highest posterior density value because it provides an overall measure of how the MCMC chains are mixing. Effective sample sizes were calculated from MCMC chains using the method described by Thiebauz and Zwiers (1984) with lags out to 100 for autocorrelation calculations. I chose to minimize the negative log posterior density (Equation 3.3.2a) for ease of computation. For the fourth estimation approach in which age-specific selectivity values varied over time, I added the curvature penalty term (Equation 8) to my negative log posterior density (Equation 3.3.2b). The parameters estimated in the model (Equation 3.3.3) included the subset of parameters common to all of the estimation models and the subset of time-varying selectivity parameters ϕ specific to each estimation model.

The subset of parameters used to model time-varying selectivity depended upon the method used to estimate selectivity. For the first estimation approach in which the first inflection point of the double logistic function varied with time, the selectivity parameters included the first inflection point in the first year, annual deviations in the first inflection point, standard deviation of the log-scale first inflection point, and the other three parameters of the double logistic function (Equation 3.3.4a). For the second estimation approach in which both inflection points of the double logistic function varied

with time, the second inflection point was replaced by a second inflection point in the first year, annual deviations in the second inflection point, and a standard deviation of the log-scale second inflection point selectivity deviations (Equation 3.3.4b). For the third estimation approach in which all four parameters of the double logistic function varied with time, both slopes were also replaced by corresponding slopes in the first year, annual deviations for each of these parameters (Equation 3.3.4c). For the fourth estimation approach in which the age-specific selectivity values varied with time, the selectivity parameters included the age-specific selectivity values in the first year, annual deviations for each age-specific selectivity value, and standard deviations for the year and age-specific log selectivity values (Equation 3.3.4d)

I separated the probability density of the data conditional on the parameters into two components for total catch and proportion of catch-at-age (Equation 3.3.5). I assumed total annual catch followed a lognormal distribution, with the log density (ignoring some additive constants) given by Equation 3.3.6. I assumed proportion of catch-at-age followed a distribution that would arise if N_E fish were observed, with numbers observed at each age following a multinomial distribution, with the log density (ignoring some additive constants) expressed by Equation 3.3.7.

For all of the time-varying selectivity estimation approaches, I assumed the prior probability densities of the random walk deviations for selectivity parameters followed lognormal distributions, with the log prior densities (ignoring some additive constants) expressed as:

(9) $$\ln[p(\chi_i)] = -\frac{1}{2\sigma_i^2}\sum_{y=1}^{n-1}\left[\chi_{i,y}^2\right] - (n-1)\ln\sigma_i ,$$

where i indexes the time-varying selectivity parameters (e.g., first inflection point of double logistic function). I assumed the prior probability densities of the log total catch, log catchability, and log selectivity standard deviations followed lognormal distributions, with the log prior densities (ignoring some additive constants) expressed as:

$$\ln[p(\sigma_i)] = -\frac{1}{2\vartheta_i^2}\left(\ln\sigma_i' - \ln\sigma_i\right)^2 - \ln\vartheta_i, \tag{10}$$

where i indexes the error sources (e.g., observation errors in total catch). I assigned a strong informative prior density (i.e., identical to the generating distribution from the data generating model) to the log total catch standard deviation (Table 3.5). Thus, I assumed the analyst had good prior information on how observation errors in total catch were distributed, which is a reasonable assumption for a well monitored commercial fishery. I assumed the analyst would not have such strong prior information for the other standard deviations. Therefore, I assigned more weakly informative prior densities which allowed the remaining standard deviations to vary over a realistic range of values (Table 3.5). The time-varying age-specific selectivity parameter estimation model failed to converge to a solution when weakly informative prior densities were assigned to the year and age-specific log selectivity standard deviations, σ_{ϖ} and σ_{φ} respectively. As a result, I fixed the values for the year and age-specific log selectivity standard deviations at 0.15 and 0.08 respectively for all simulations. This solution followed the common practice of assuming variances to be known when they cannot be estimated in the estimation model. I assigned weakly informative uniform prior densities to the logs of all other model parameters. Therefore, prior densities for each log-scale model parameter, excluding the

selectivity random walk deviations and their associated variances, were constants for all parameter values.

I compared the performance of the four estimation models by calculating the relative error (RE) of population biomass and exploitation rate in the last year of the analysis, for each simulated data set:

$$RE = \frac{\hat{X} - X}{X}, \tag{11}$$

where $\hat{X}$ is the point estimate of the quantity of interest from the estimation model, and X is the true value of the quantity of interest from the data generating model. I used the median of the marginal posterior distribution as a point estimate. Estimated biomass and exploitation rate in the last year often play an important role when stock assessment results are used to inform management actions. In addition, I used the median of the relative errors (MRE) to examine whether there was systematic bias in estimates from the estimation models. I used the median absolute relative error (MARE), which captures elements of bias and precision, to compare the range of relative errors made when using the estimation models.

Model Selection Methods

I evaluated the performance of three model selection techniques to determine which technique(s), if any, could identify consistently the "best" time-varying selectivity estimation approach. The three model selection techniques I used to identify the best time-varying selectivity estimation approach were RMSE, degree of retrospectivity, and DIC. By best selectivity estimation approach, I mean the estimation approach which most closely predicts the true fish population as produced by the data generating model.

More specifically, for each simulated data set I measured relative performance of the different estimation models based on the RE of population biomass and exploitation rate in the last year of the analysis. I used three definitions of the best or nearly best estimation model(s) for a given simulation run: 1) the estimation model producing the lowest final population biomass or exploitation rate RE, 2) estimation models producing REs within 0.05 of the lowest RE, and 3) estimation model producing REs within 0.1 of the lowest RE. I allowed for this relaxation in the definition of best or nearly best estimation model because in a real stock assessment, where the true population characteristics are unknown, alternative estimation models which produce similar results often would be treated as equally viable. In particular, I chose the values 0.05 and 0.1 because they represented a difference in model results that most analysts would consider negligible. In addition, I used the MRE and MARE to examine bias and precision in estimates from the estimation models chosen by each selection method. Comparison of the model selection methods was made using the subset of simulation runs in which all four estimation models converged on adequate solutions to avoid problems with different convergence rates between the estimation models.

My first model selection procedure focused on proportion of catch-at-age residuals, with the selected model minimizing the RMSE for these residuals. I chose this as one possible method because I thought generally large proportion of catch-at-age residuals might occur for estimation models that incorrectly modeled selectivity patterns. These residuals were calculated from the posterior medians of predicted proportions of catch-at-age.

My second model selection method is based on retrospective analysis, which involves the comparison of successive estimates of model output quantities as additional years of data are added to the stock assessment (Parma 1993; Mohn 1999). For this selection method I selected the model that minimized the absolute value of Mohn's (1999) degree of retrospectivity statistic:

$$DR = \sum_{y=n-10}^{n} \frac{X_{(1:y),y} - X_{(1:n),y}}{X_{(1:y),y}}, \tag{12}$$

where $X_{(1:y),y}$ is the predicted value of quantity X in year y estimated from the data set spanning year 1 to year y and $X_{(1:n),y}$ is the predicted value of quantity X in year y estimated from the data set spanning year 1 to the last year of the full data set n. Here I conducted a retrospective analysis for each estimation model-simulated data set fit by dropping a year of data from the simulated data set and refitting the estimation model, repeating this process until the last 10 years of data had been sequentially removed from the analysis. Systematic retrospective patterns in model quantities can occur when time-varying processes are modeled as being constant over time (Mohn 1999). Though all of my estimation models allowed selectivity to change over time, I expected to see retrospective patterns in cases where an estimation model had difficulty tracking changes in selectivity. To make this approach practical, I used highest posterior density estimates of the parameters with the variance parameters fixed at their point estimates from the analysis of the full data set.

My final selection method was to select the model that minimized the Deviance Information Criterion (Spiegelhalter et al. 2002). Information-theoretic model selection criteria generally work by balancing model goodness of fit against model complexity

(i.e., the number of parameters in the model). The effective number of parameters in complex models, such as my SCAA models, is often less than the actual number of parameters due to various constraints placed upon those parameters. I chose to use DIC, as opposed to the more commonly used Akaike's Information Criterion (AIC; Akaike 1973) and Bayesian Information Criterion (BIC; Schwartz 1978), because DIC provides a means of estimating the effective number of parameters. Wilberg (2005) demonstrated in a different SCAA application that selection by DIC could result in estimates with lower mean square errors, than always using any particular single model.

Deviance Information Criterion is composed of two components (Spiegelhalter et al. 2002):

$$DIC = \overline{D} + p_D, \tag{13}$$

where $\overline{D}$ is the average deviance and p_D is the effective number of parameters. I estimated the average deviance as (Spiegelhalter et al. 2002):

$$\overline{D} = \frac{1}{C}\sum_{c=1}^{C} -2\log_e p(x|\theta_c), \tag{14}$$

where C is the number of MCMC cycles saved minus the burn in and $p(x|\theta_c)$ is the probability of data x conditional on parameters θ from MCMC cycle c. I estimated the effective number of parameters as (Wilberg 2005):

$$p_D = \overline{D} - D(\theta_{ML}), \tag{15}$$

where $D(\theta_{ML})$ is the deviance evaluated at the maximum likelihood parameter estimates and the other variables are described above.

For each model selection method, I calculated in what percentage of the simulation runs the method selected the best or nearly best estimation models. In addition, for each selection method I examined the distribution of REs for final population biomass and exploitation rate estimates.

I compared the performance of using estimation models selected by degree of retrospectivity to the performance of always using the same estimation model, for each of the estimation models. The objective here was to determine if this model selection technique outperformed the omnibus approach of always using the same estimation model. I used model selection by degree of retrospectivity in this evaluation because of the good performance of this model selection method (see Results). To properly make this comparison, I used degree of retrospectivity to select the best estimation based on final population biomass and exploitation rate for each simulation run, rather than for the subset of simulation runs where all four estimation models converged on solutions. Comparisons were made using MRE and MARE values for final population biomass and exploitation rate selected by degree of retrospectivity and estimated by each of the estimation models.

Results

Model runs exhibiting poor convergence characteristics were dropped from the analysis. The following results are based on sample sizes of 333 to 425 model runs per scenario (Table 3.6). All of the dropped model runs failed to converge to highest posterior density solutions, thus MCMC simulations could not be run. I suspect that with sufficient effort, which would be warranted for a real assessment, an analyst could have made adjustments in many of these cases to achieve convergence. This was not practical

in the context of this simulation study. It should be noted that the subsets of simulation runs demonstrating poor convergence characteristics generally were different for each of the estimation models (i.e., incidents of poor convergence were not due to characteristics of particular simulated data sets).

Estimation Models

There was little difference in the biases of the four estimation models' estimates of population biomass in the last year of analysis within each data generating scenario (Figure 3.1). The four estimation models produced less biased estimates of the final population biomass in the double logistic generating scenario compared to the age-specific selectivity parameters generating scenario. Median relative error values for final population biomass ranged from 0.01 to 0.13 for the double logistic function generating scenario (Table 3.6). In contrast, MRE values for final population biomass ranged from -0.23 to 0.55 for the age-specific selectivity parameters generating scenario. The estimation model using the double logistic function with four time-varying parameters produced the most biased estimates of population biomass in both data generating scenarios.

The four estimation models produced more precise estimates of population biomass in the last year of analysis when the estimation models more accurately represented the true underlying population (i.e., when selectivity estimation and data generating models were similar) (Figure 3.1). Median absolute relative error values for final population biomass varied from 0.20 to 0.26 for the three double logistic function estimation models in the double logistic function generating scenario (Table 3.6). On the other hand, the age-specific selectivity parameters estimation model had a final

population biomass MARE value of 0.54 for the double logistic function generating scenario. The age-specific selectivity parameters estimation model had final population biomass MARE value of 0.35 for the age-specific selectivity parameters generating function. In contrast, the three double logistic function estimation models had final population biomass MARE values ranging from 0.50 to 0.61 for the age-specific selectivity parameters generating scenario. The estimation model using the double logistic function with four time-varying parameters produced estimates of final population biomass that were less precise than the estimation models using double logistic functions with one and two time-varying parameters (Figure 3.1).

There was little difference in the biases of the four estimation models' estimates of exploitation rate in the last year of analysis within each data generating scenario (Figure 3.2). The four estimation models produced less biased estimates of the final exploitation rate in the double logistic generating scenario compared to the age-specific selectivity parameters generating scenario. Median relative error values for final exploitation rate ranged from -0.10 to -0.02 for all four of the estimation models in the double logistic function generating scenario (Table 3.6). In contrast, the MRE values for exploitation rate ranged from -0.36 to -0.18 for all four estimation models in the age-specific selectivity parameters generating scenario. The estimation model using the double logistic function with four time-varying parameters produced the most biased estimates of population biomass in both data generating scenarios.

The four estimation models produced more precise estimates of exploitation rate in the last year of analysis when the estimation models more accurately represented the true underlying population, though the difference was not as pronounced in the age-

specific selectivity parameters generating scenario (Figure 3.2). Median absolute relative error values for the final exploitation rate varied from 0.20 to 0.25 for the three double logistic function estimation models in the double logistic generating scenario (Table 3.6). On the other hand, the age-specific selectivity parameters estimation model had an exploitation rate MARE value of 0.58 for the double logistic generating scenario. The age-specific selectivity parameters estimation model had a final exploitation rate MARE value of 0.38 for the age-specific selectivity parameters generating function. In contrast, the three double logistic function estimation models had final exploitation rate MARE values ranging from 0.48 to 0.57 for the age-specific selectivity parameters generating scenario.

Model Selection

I compared the performance of the model selection methods by examining the subset of simulation runs where all four of the estimation models exhibited good convergence properties. All of the estimation models converged on good solutions for 438 of the 1,000 simulation runs.

Degree of retrospectivity selected the best or nearly best estimation model, based on final population biomass and exploitation rate REs, as often as or more often than DIC and RMSE (Figure 3.3). Degree of retrospectivity selected the best or nearly best model in 34-57% of the simulation runs when the best or nearly best model was chosen based on final population biomass RE, and in 33-52% of the simulation runs based on final exploitation rate RE. Deviance information criterion selected the best or nearly best model in 27-48% of the simulation runs when the best or nearly best model was chosen based on final population biomass RE, and in 27-50% of the simulation runs based on

final exploitation rate RE. Root mean square error selected the best or nearly best model in 29-49% of the simulation runs when the best or nearly best model was chosen based on final population biomass RE, and in 33-49% of the simulation runs based on final exploitation rate RE.

Selecting estimation models using degree of retrospectivity produced estimates of population biomass and exploitation rate in the last year of analysis that were as biased and precise as or less biased and more precise than estimation models selected using DIC and RMSE (Figures 3.5 and 3.6). In particular, degree of retrospectivity selected estimation models that produced final population biomass and exploitation rate estimates that were less biased and more precise than estimates selected by DIC and RMSE in the age-specific selectivity parameters generating scenario (Table 3.7).

Degree of retrospectivity performed as well as or better than the individual estimation models at estimating final population biomass and exploitation rate. Degree of retrospecitivity produced a final population biomass MRE of 0.05 and MARE of 0.24 in the double logistic generating scenario, which is comparable to the estimation performances of the three time-varying double logistic functions in that same scenario (Table 3.6). Degree of retrospecitivity produced a final population biomass MRE of -0.01 and MARE of 0.40 in the age-specific selectivity parameters generating scenario, which is less biased than any of the individual estimation models and of intermediate precision between the age-specific selectivity parameters and double logistic function estimation models in that same scenario (Table 3.6). Degree of retrospecitivity produced a final exploitation rate MRE of -0.05 and MARE of 0.24 in the double logistic generating scenario, which is comparable to the estimation performances of the three

time-varying double logistic functions in that same scenario (Table 3.6). Degree of retrospecitivity produced a final exploitation rate MRE of 0.03 and MARE of 0.42 in the age-specific selectivity parameters generating scenario, which is less biased than any of the individual estimation models and of intermediate precision between the age-specific selectivity parameters and double logistic function estimation models in that same scenario (Table 3.6).

Discussion

There was no single time-varying selectivity estimation model that outperformed the others in all situations that I examined. Rather, the estimation model(s) that produced the estimates most tightly distributed about true population biomass and exploitation rate in the last year of analysis was the one that most closely represented the true underlying population. The three estimation models that used variants of the double logistic function to model time-varying selectivity produced better estimates of final population biomass and exploitation rate than the age-specific selectivity parameters estimation model when the selectivity of the true population was generated with a double logistic function. Likewise, the age-specific selectivity parameters estimation model produced better estimates of final population biomass and exploitation rate than the three double logistic function estimation models when the selectivity of the true population was generated with age-specific selectivity parameters. This sort of result is common to simulation studies where there are similarities between data generating and estimation models (e.g., Radomski et al. 2005; Wilberg and Bence 2006).

My study suggests that if an analyst knows the underlying form that selectivity takes in a fish population, then he or she can model time-varying-selectivity reasonably

well. This begs the question, how well can an analyst know true selectivity patterns and how they vary over time? Many fishing gears such as gill nets and trap nets are size selective. If one makes the assumption that the size and age of fish are correlated and by extension selectivity and age of fish are correlated, then modeling selectivity as some function of ages, which produces a smooth selectivity curve, is a reasonable approach. It is more difficult to think of situations where age-specific selectivity values vary relatively independently of each other over time. Kimura (1990) demonstrated that estimating age-specific selectivity parameters outperformed the use of a selectivity function when the function was incorrectly specified. The approach of Butterworth et al. (2003) that I used in this study is an extension of Kimura's (1990) approach, which allows the age-specific selectivity parameters to vary over time. Further study is needed to determine whether estimation models that assume time-varying age-specific selectivity parameters outperform a time-varying selectivity function when the function is misspecified.

Model complexity is another issue that must be addressed when evaluating different time-varying selectivity models. Increased model complexity means an increased number of parameters that must be estimated, which can lead to over-parameterization of the model. An over-parameterized model can produce poor parameter estimates with high variances (Burnham and Anderson 2002). In my study, the issue of model complexity was most clearly demonstrated in the performance of the double logistic function with four time-varying parameters and two associated variances. I expected the four time-varying parameter selectivity estimation method to outperform the other double logistic function approaches when the observed data were generated using age-specific selectivity parameters, due to the increased flexibility granted by

allowing all four double logistic function parameters to vary over time. Instead, I found that the four time-varying parameter double logistic function produced more biased estimates of final population biomass than the other double logistic function estimation approaches when the observed data were generated using age-specific selectivity parameters. One of the two variances associated with the slopes and inflection points of the four time-varying parameter double logistic function was estimated as nearly zero (i.e., making it effectively the same as the two time-varying parameter double logistic function) in many of the simulation runs, which suggests that the observed data were not informative enough to estimate all of the selectivity parameters.

The performance of the four time-varying parameter double logistic function in my study could be due to my data generating model design. The observed data were generated by allowing selectivity parameters to vary over time according to a first order autoregressive process, which did not follow any trend over time, and for which the deviations among ages were not correlated. The performance of the time-varying double logistic methods, may have improved had the generating selectivity function produced correlated changes in selectivity for adjacent ages, like those that would be generated by variations in one or more parameters of a function.

I was surprised to see how well degree of retrospectivity performed as a time-varying selectivity model selection method compared to DIC and RMSE. Estimation models selected using degree of retrospectivity produced final population biomass and exploitation rate estimates that were more or equally accurate and precise compared to estimates selected by DIC and RMSE for the data generating scenarios I examined. In particular, degree of retrospectivity selected final population biomass estimates that were

much more accurate and precise than those estimates selected by DIC and RMSE when observed data sets were generated using age-specific selectivity parameters. The robustness of degree of retrospectivity as a time-varying selectivity estimation model selection method probably is due to the fact that it can detect consistent patterns in model estimates over time (i.e., as new years of data are added to the model). As I expected, these consistent or retrospective patterns do appear to be indicative of an estimation model that has difficulty estimating time-varying selectivity. Deviance information criterion and RMSE lack this ability to detect retrospective patterns since they merely evaluate the model fit to the complete time series of observed data. Parma (1993) developed an alternative metric for identifying retrospective patterns using the square root of the mean square error between the retrospective estimate of a model quantity and a corresponding reference estimate on the log scale. Mohn (1999) points out that this mean square error metric is unable to differentiate between retrospective and random patterns since it uses a mean square, rather than signed sum, in its calculation. Though I did not test Parma's (1993) metric, I suspect that it would perform similarly to my DIC and RMSE methods. I recommend that degree of retrospectivity be used to select between estimation models using different methods of estimating time-varying selectivity, based on its performance in my study.

Selecting from multiple estimation models using degree of retrospectivity worked better than choosing a single estimation model in my study. Nothing is lost in estimation performance by using degree of retrospectivity, even if an analyst is able to correctly specify time-varying selectivity. In addition, degree of retrospectivity outperforms estimation models which misspecifiy time-varying selectivity (i.e., assuming a double

logisitic function when true age-specific selectivity values vary over time). Therefore, I recommend that degree of retrospectivity be used to select between time-varying selectivity models.

I should note that my study only looked at the performance of model selection methods on an individual basis. The ability to select the best estimation model may be improved by using combinations of different selection techniques. For example, estimation models could be ranked based on their degree of retrospectivity. If multiple estimation models have equal or nearly equal degree of retrospecitvity values, then DIC or RMSE values could be used to select between those models with degree of retrospecitvity values close to zero. Further study of using such multiple model selection methods would be informative.

References

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267-281 *in* B.N. Petrov and F. Csaki, editors. Second International Symposium on Information Theory. Budapest: Akademiai Kiado.

Bence, J.R. and J.E. Hightower. 1993. Status of Bocaccio in the Conception/Monterey/Eureka INPFC areas in 1992 and recommendations for management in 1993. Appendix B in: Status of the Pacific coast groundfish fishery through 1992 and recommended acceptable biological catches for 1993. Pacific Fishery Management Council, Portland, OR.

Burnham, K.P., and D.R. Anderson. 2002. Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach. Springer-Verlag, New York.

Butterworth, D.S., J.N. Ianelli, and R. Hilborn. 2003. A statistical model for stock assessment of southern bluefin tuna with temporal changes in selectivity. African Journal of Marine Science 25: 331-361.

Deriso, R.B., T.J. Quinn II, and P.R. Neal. 1985. Catch-at-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences 42:815-824.

Doubleday, W.G. 1976. A least squares approach to analyzing catch at age data. Research Bulletin of the International Commission of Northwest Atlantic Fisheries 12:69-81.

Ebener, M. P., J. R. Bence, K. Newman, and P. Schneeberger. 2005. Application of statistical catch-at-age models to assess lake whitefish stocks in the 1836 treaty-ceded waters of the upper Great Lakes. Pages 271-309 *in* L.C. Mohr and T.F. Nalepa, editors. Proceedings of a workshop on the dynamics of lake whitefish (*Coregonus clupeaformis*) and the amphipod *Diporeia* spp. in the Great Lakes. Great Lakes Fishery Commission Technical Report 66.

Fournier, D. 1983. An analysis of the Hecate Strait Pacific cod fishery using an age-structured model incorporating density-dependent effects. Canadian Journal of Fisheries and Aquatic Sciences 40:1233-1243.

Fournier, D., and C.P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39:1195-1207.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2004. Bayesian Data Analysis, 2nd edition. Chapman and Hall, Boca Raton.

Gudmundsson, G. 1994. Time series analysis of catch-at-age observations. Applied Statistics 43:117-126.

Ianelli, J. 1996. An alternative stock assessment model of the eastern Bering Sea pollock fishery. Pages 81-104 *in* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions, November 1996. North Pacific Fisheries Management Council, Anchorage, Alaska.

Kimura, D.K. 1990. Approaches to age-structured separable sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences 47:2364-2374.

Methot, R.D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Pages 259-277 *in* L. Low, editor. Proceedings of the symposium on applications of stock assessment techniques to gadids. International North Pacific Fisheries Commission Bulletin 50.

Millar, R.B. 1995. The functional form of hook and gillnet selection curves cannot be determined from comparative catch data alone. Canadian Journal of Fisheries and Aquatic Sciences 52:883-891.

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES Journal of Marine Science 56:473-488.

Myers, R.A., and T.J. Quinn II. 2002. Estimating and testing non-additivity in fishing mortality: implications for detecting a fisheries collapse. Canadian Journal of Fisheries and Aquatic Sciences 59:597-601.

Parma, A.N. 1993. Retrospective catch-at-age analysis of pacific halibut: implications on assessment of harvesting policies. Pages 247-265 *in* G. Kruse, D.M. Eggers, C. Pautzke, R.J. Marasco, and T.J. Quinn II, editors. Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations. Alaska Sea Grant College Program.

Punt, A. E., D. C. Smith, R. B. Thomson, M. Haddon, X. He, and J. M. Lyle. 2001. Stock assessment of the blue grenadier *Macruronus novaezelandiae* resource off south-eastern Australia. Marine and Freshwater Resources 52:701-717.

Punt, A.E., A.D.M. Smith, and G. Cui. 2002. Evaluation of management tools for Australia's South East Fishery: 2. how well can management quantities be estimated? Marine and Freshwater Research 53:631-644.

Radomski, P., J.R. Bence, and T.J. Quinn II. 2005. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational kill-dominated fishery. Canadian Journal of Fisheries and Aquatic Sciences 62:436-452.

Schwartz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6: 461-464.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 63: 583-639.

Thiebauz, H. J., and F. W. Zwiers. 1984. The interpretation and estimation of effective sample size. Journal of Climate and Applied Meteorology 23: 800-811.

Thompson, G.G. 1994. Confounding of gear selectivity and the natural mortality rate in cases where the former is a nonmonotone function of age. Canadian Journal of Fisheries and Aquatic Sciences 51:2654-2664.

Wilberg, M.J. 2005. Improving statistical catch-at-age stock assessments. Doctoral Dissertation. Michigan State University, East Lansing, Michigan.

Wilberg, M.J., and J.R. Bence. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences 63: 2275-2285.

Yin, Y., and D.B. Sampson. 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. North American Journal of Fisheries Management 24:865-879.

Table 3.1. Symbols and descriptions of variables used in data generating and estimation models.

Symbol	Description	Application
$C_{y,a}$	Number of fish caught by year and age	Both
$\tilde{C}_y$	Observed number of fish caught by year	Both
E_y	Fishery effort by year	Both
$F_{y,a}$	Instantaneous fishing mortality by year and age	Both
M	Instantaneous natural mortality	Both
$N_{y,a}$	Abundance by year and age	Both
N_0	Mean abundance for abundance in first year	Estimation
N_E	Number of fish used to calculate age composition each year	Both
$P_{y,a}$	Proportion of catch by year and age	Both
$\tilde{P}_{y,a}$	Observed proportion of catch by year and age	Both
R_0	Mean recruitment	Estimation
S_y	Number of female spawners by year	Generation
$Z_{y,a}$	Instantaneous total mortality by year and age	Both
$Z_{0,a}$	Instantaneous total mortality for abundance in first year by age	Generation
$b_{1,y}$	First inflection pt. of double logistic selectivity function by year	Both
b_2	First slope of double logistic selectivity funcion	Both
b_3	Second inflection pt. of double logistic selectivity function	Both
b_4	Second slope of double logistic selectivity funcion	Both
b_1'	Mean of first inflection pt. of double logistic selectivity function	Estimation
f_y	Fishing intensity by year	Both
m	Total number of ages	Both
n	Total number of years	Both

Table 3.1 (cont'd).

$p(\theta\|x)$	Posterior probability density of parameters conditional on data	Estimation
$p(x\|\theta)$	Probability density of data conditional on parameters	Estimation
$p(\theta)$	Prior probability density of parameters	Estimation
q	Fishery catchability	Both
$s_{y,a}$	Fishery selectivity by year and age	Both
s'_a	Mean fishery selectivity by age	Generation
w_a	Mean fish weight by age	Both
α	Productivity parameter of Ricker recruitment function	Generation
β	Density dependent parameter of Ricker recruitment function	Generation
$\chi_{i,y}$	Process error in selectivity parameter *i* by year	Estimation
δ_y	Process error in first inflection point of double logistic function by year	Generation
ε_y	Process error in recruitment by year	Generation
ϕ	Subset of time-varying selectivity parameters	Estimation
$\gamma_{y,a}$	Process error in selectivity by year and age	Generation
η_y	Process error in inflection points of double logistic function by year	Estimation
λ_y	Error in fishing intensity by year	Both
μ_N	Mean number of age-1 fish for abundance in first year	Generation
θ	Set of all model parameters	Estimation
ϑ_λ	Prior standard deviation of log-scale fishing intensity standard deviation	Estimation
ϑ_η	Prior standard deviation of log-scale inflection points standard deviation	Estimation
ϑ_ν	Prior standard deviation of log-scale total catch standard deviation	Estimation
ϑ_τ	Prior standard deviation of log-scale slopes standard deviation	Estimation
ρ_1	First correlation parameter for first order autoregressive process	Generation

Table 3.1 (cont'd).

ρ_2	Second correlation parameter for first order autoregressive process	Generation
σ_N	Standard deviation of number of age-1 fish for abundance in first year	Generation
σ_δ	Standard deviation of log-scale first inflection point	Generation
σ'_δ	Generating mean of log-scale first inflection point standard deviation	Generation
σ_ε	Standard deviation of log-scale recruitment	Generation
σ_γ	Standard deviation of log-scale selectivity	Generation
σ'_γ	Generating mean of log-scale selectivity standard deviation	Generation
σ_η	Standard deviation of log-scale inflection points	Estimation
σ'_η	Prior mean of log-scale inflection points standard deviation	Estimation
σ_φ	Age-specific standard deviation of log-scale selectivity	Estimation
σ_λ	Standard deviation of log-scale fishing intensity	Both
σ'_λ	Generating and prior mean of log-scale fishing intensity standard deviation	Both
σ_τ	Standard deviation of log-scale slopes	Estimation
σ'_τ	Prior mean of log-scale slopes standard deviation	Estimation
σ_υ	Standard deviation of log-scale total catch	Both
σ'_ν	Generating and prior mean of log-scale total catch standard deviation	Both
σ_ϖ	Year-specific standard deviation of log-scale selectivity	Estimation
τ_y	Process error in slopes of double logistic function by year	Estimation
υ_y	Observation error in number of fish caught by year	Both
$\varpi_{y,a}$	Process error in selectivity by year and age	Estimation
ω_y	Process error in recruitment by year	Estimation
ψ_a	Process error for abundance in first year by age	Estimation
ζ_δ	Generating standard deviation of log-scale first inflection point standard deviation	Generation

Table 3.1 (cont'd).

ζ_γ	Generating standard deviation of log-scale selectivity standard deviation	Generation
ζ_λ	Generating standard deviation of log-scale fishing intensity standard deviation	Generation
ζ_v	Generating standard deviation of log-scale total catch standard deviation	Generation

Table 3.2. Data generating and estimation model equations.

	Equation	Application
3.2.1	$N_{y+1,a+1} = N_{y,a}e^{-Z_{y,a}}$	Both
3.2.2	$N_{1,a} = N_{2-a,1}e^{-\sum_{j=1}^{a-1} Z_{0,j}}$	Generation
3.2.3	$N_{y,1} = \alpha S_{y-1}e^{-\beta S_{y-1}}e^{\varepsilon_y}, \varepsilon_y \sim N\left(0,\sigma_\varepsilon^2\right)$	Generation
3.2.4	$Z_{y,a} = M + F_{y,a}$	Both
3.2.5	$F_{y,a} = s_{y,a}f_y$	Both
3.2.6	$f_y = qE_y e^{\lambda_y}, \lambda_y \sim N\left(0,\sigma_\lambda^2\right)$	Both
3.2.7	$C_{y,a} = \frac{F_{y,a}}{Z_{y,a}} N_{y,a}\left(1-e^{-Z_{y,a}}\right)$	Both
3.2.8	$\tilde{C}_y = \left[\sum_a C_{y,a}\right]e^{\upsilon_y}, \upsilon_y \sim N\left(0,\sigma_\upsilon^2\right)$	Both
3.2.9	$P_{y,a} = \frac{C_{y,a}}{C_y}$	Both

Table 3.3. Posterior probability density equations for estimation models.

	Equation
3.3.1	$p(\theta\|x) \propto p(x\|\theta)p(\theta)$
3.3.2a	$-\ln[p(\theta\|x)] \propto -\ln[p(x\|\theta)] - \ln[p(\theta)]$
3.3.2b	$-\ln[p(\theta\|x)] + g(s_{y,a};\sigma_\varphi^2) \propto -\ln[p(x\|\theta)] - \ln[p(\theta)] + g(s_{y,a};\sigma_\varphi^2)$
3.3.3	$\theta = \{N_0, [\psi_a]_{a=1}^m, R_0, [\omega_y]_{y=2}^n, q, \phi\}$
3.3.4a	$\phi = \{b_{1,1}, [\eta_y]_{y=1}^{n-1}, \sigma_\eta, b_2, b_3, b_4\}$
3.3.4b	$\phi = \{b_{1,1}, [\eta_{1,y}]_{y=1}^{n-1}, b_2, b_{3,1}, [\eta_{3,y}]_{y=1}^{n-1}, b_4, \sigma_\eta\}$
3.3.4c	$\phi = \{b_{1,1}, [\eta_{1,y}]_{y=1}^{n-1}, b_{2,1}, [\tau_{2,y}]_{y=1}^{n-1}, b_{3,1}, [\eta_{3,y}]_{y=1}^{n-1}, b_{4,1}, [\tau_{4,y}]_{y=1}^{n-1}, \sigma_\eta, \sigma_\tau\}$
3.3.4d	$\phi = \left\{[s_{1,a}]_{a=1}^m, \left[[\varpi_{y,a}]_{y=1}^{n-1}\right]_{a=1}^m, \sigma_\varpi, \sigma_\varphi\right\}$
3.3.5	$\ln[p(x\|\theta)] = \ln[p(C\|\theta)] + \ln[p(P\|\theta)]$
3.3.6	$\ln[p(C\|\theta)] = -\frac{1}{2\sigma_C^2}\sum_{y=1}^{n}\left[\left(\ln\tilde{C}_y - \ln C_y\right)^2\right] - n\ln\sigma_C$
3.3.7	$\ln[p(P\|\theta)] = \sum_{y=1}^{n} N_E \sum_{a=1}^{m}\left(\tilde{P}_{y,a}\ln P_{y,a}\right)$

Table 3.4. Values of quantities used in data generating model to create simulation data sets.

Quantity	Value
n	20
m	8
μ_N	355,000
σ_N	0.4
α	10.1
β	5.10E-06
σ_ε	0.4
$[w_a]_{a=1}^m$	0.20, 0.48, 0.73, 0.91, 1.32, 1.52, 1.76, 2.15
M	0.24
$[E_y]_{y=1}^n$	0.8, 1.6, 2.4, 3.2, 4.0, 4.8, 5.6, 6.4, 7.2, 8.0, 8.0, 7.2, 6.4, 5.6, 4.8, 4.0, 3.2, 2.4, 1.6, 0.8
q	0.15
σ'_λ	0.4
ζ_λ	0.1
b'_1	4.01
$[b_i]_{i=2}^4$	1.40, 3.49, 0.50
$[s'_a]_{a=1}^m$	0.04, 0.15, 0.43, 0.85, 1.00, 0.82, 0.57, 0.37
ρ_1	0.9
σ'_δ	0.2
ζ_δ	0.1
ρ_2	0.9
σ'_γ	0.15
ζ_γ	0.1
σ'_υ	0.05
ζ_ν	0.1
N_E	400

Table 3.5. Values used to define prior probability densities in estimation models.

Quantity	Value
σ'_λ	0.4
ϑ_λ	0.25
σ'_η	0.2
ϑ_η	0.42
σ'_ν	0.05
ϑ_ν	0.1
σ'_τ	0.2
ϑ_τ	0.42

Table 3.6. Median relative errors (MRE), median absolute relative errors (MARE), and number of replicates (N) for estimates of final population biomass and exploitation rate produced by the time-varying selectivity estimation models: double logistic functions with one (DL1), two (DL2), and four (DL4) time-varying parameters, and time-varying age-specific selectivity parameters (ASP).

	Population Biomass					
Estimation	DL1			ASP		
Model	MRE	MARE	N	MRE	MARE	N
DL1	0.05	0.20	414	0.23	0.50	411
DL2	0.06	0.22	361	0.33	0.57	425
DL4	0.13	0.26	333	0.55	0.61	430
ASP	0.01	0.54	382	-0.23	0.35	409
	Exploitation Rate					
DL1	-0.04	0.20	414	-0.18	0.50	411
DL2	-0.06	0.21	361	-0.24	0.51	425
DL4	-0.10	0.25	333	-0.36	0.48	430
ASP	-0.02	0.58	382	0.30	0.38	409

Table 3.7. Median relative errors (MRE), median absolute relative errors (MARE), and number of replicates (N) for estimates of final population biomass and exploitation rate chosen by the model selection methods: root mean square error (RMSE), deviance information criterion (DIC), and degree of retrospectivity (DR).

	Population Biomass					
Model	DL1			ASP		
Selection	MRE	MARE	N	MRE	MARE	N
RMSE	0.10	0.21	179	0.43	0.53	259
DIC	0.10	0.22	179	0.40	0.51	259
DR	0.07	0.22	179	-0.05	0.35	259
	Exploitation Rate					
RMSE	-0.08	0.21	179	-0.32	0.54	259
DIC	-0.09	0.22	179	-0.29	0.50	259
DR	-0.06	0.24	179	0.10	0.37	259

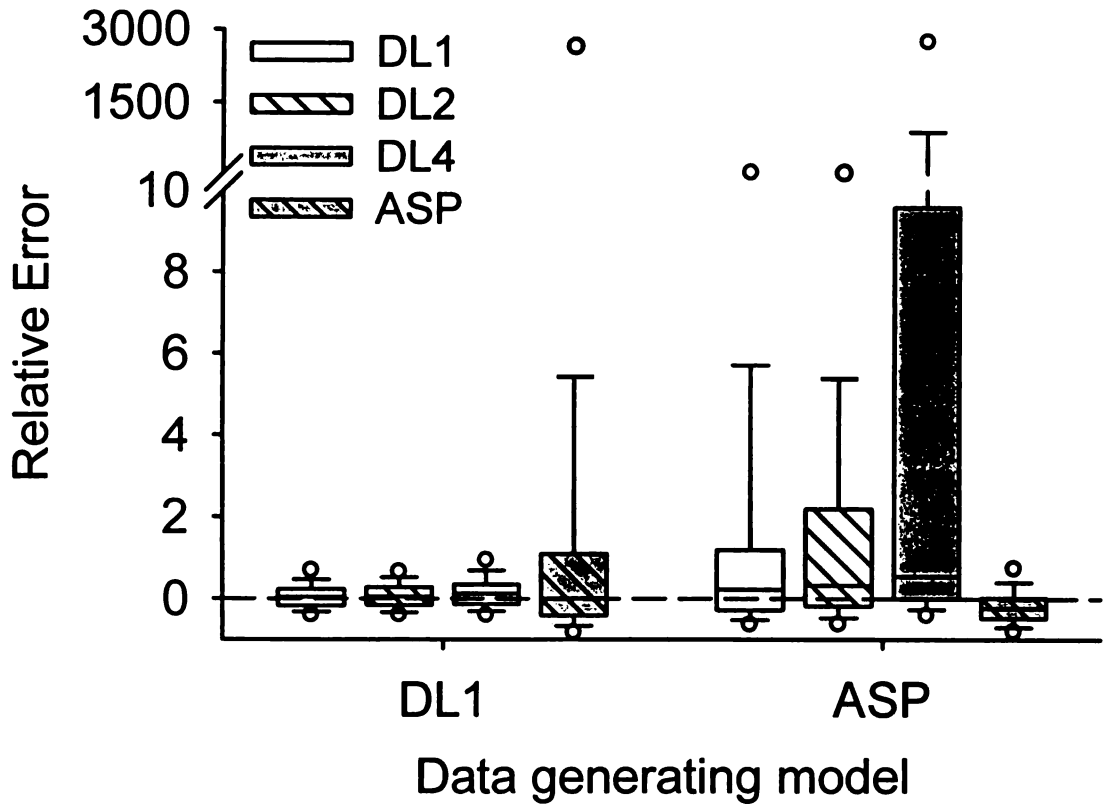


Figure 3.1. Box plots representing relative error distributions for estimates of population biomass in the last year of analysis across different data generating models. The data generating and estimation models include double logistic functions with one (DL1), two (DL2), and four (DL4) time-varying parameters, and time-varying age-specific selectivity parameters (ASP). The bars represent median relative errors. The boxes, whiskers, and circles represent 25th and 75th, 10th and 90th, and 5th and 95th percentiles of the distributions, respectively.

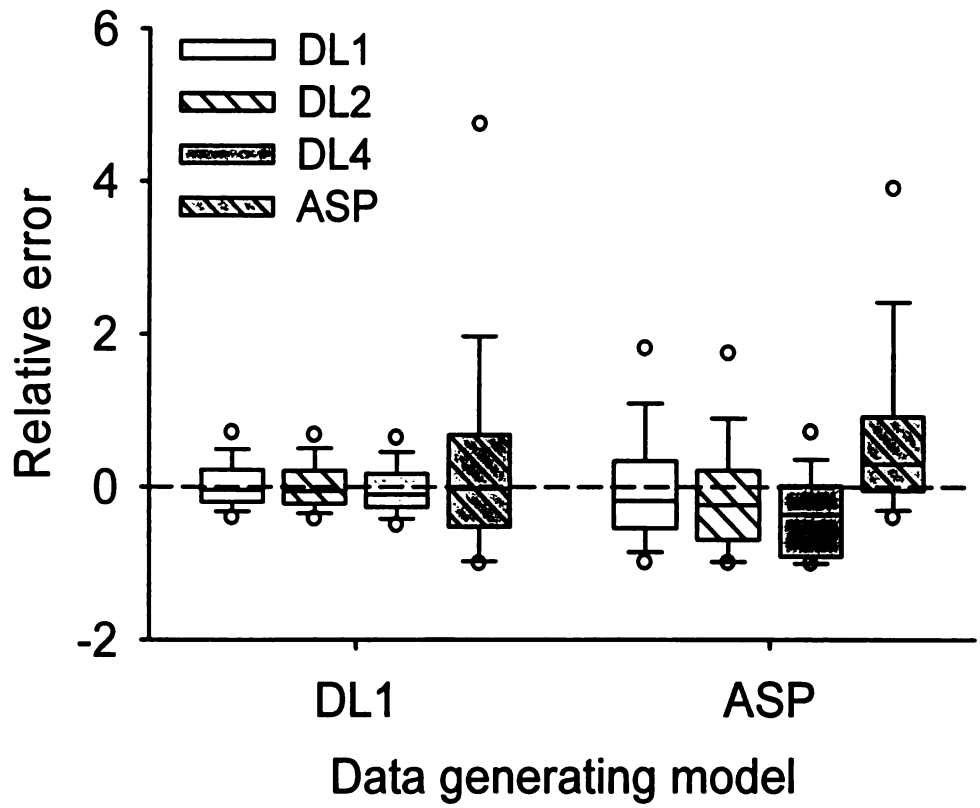
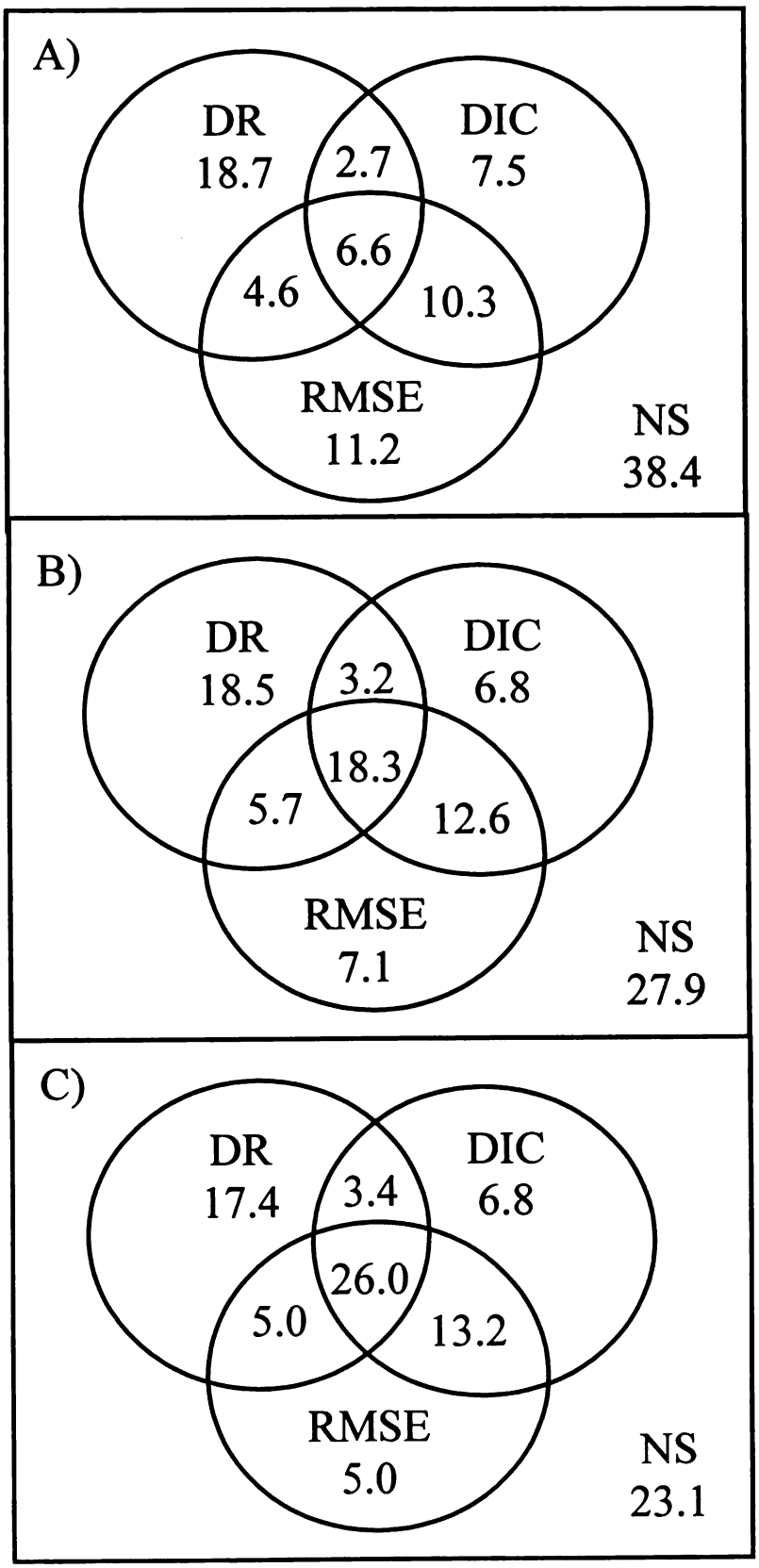


Figure 3.2. Box plots representing relative error distributions for estimates of exploitation rate in the last year of analysis across different data generating models. The data generating and estimation models include double logistic functions with one (DL1), two (DL2), and four (DL4) time-varying parameters, and time-varying age-specific selectivity parameters (ASP). The bars represent median relative errors. The boxes, whiskers, and circles represent 25th and 75th, 10th and 90th, and 5th and 95th percentiles of the distributions, respectively.

Figure 3.3. The percentage of model runs when the model selection methods chose the best or nearly best estimation model based on estimates of final population biomass. The model selection methods include root mean square error (RMSE), deviance information criterion (DIC), and degree of retrospectivity (DR). The best or nearly best estimation model(s) is defined as the model(s) producing A) the lowest final population biomass relative error, B) within 5% of the lowest final population biomass relative error, and C) within 10% of the lowest final population biomass relative error.

Figure 3.4. The percentage of model runs when the model selection methods chose the best or nearly best estimation model based on estimates of final exploitation rate. The model selection methods include root mean square error (RMSE), deviance information criterion (DIC), and degree of retrospectivity (DR). The best or nearly best estimation model(s) is defined as the model(s) producing A) the lowest final exploitation rate relative error, B) within 5% of the lowest final exploitation rate relative error, and C) within 10% of the lowest final exploitation rate relative error.


A)
DR
18.7
DIC
7.5
2.7
6.6
4.6
10.3
RMSE
11.2
NS
38.4
B)
DR
18.5
DIC
6.8
3.2
18.3
5.7
12.6
RMSE
7.1
NS
27.9
C)
DR
17.4
DIC
6.8
3.4
26.0
5.0
13.2
RMSE
5.0
NS
23.1

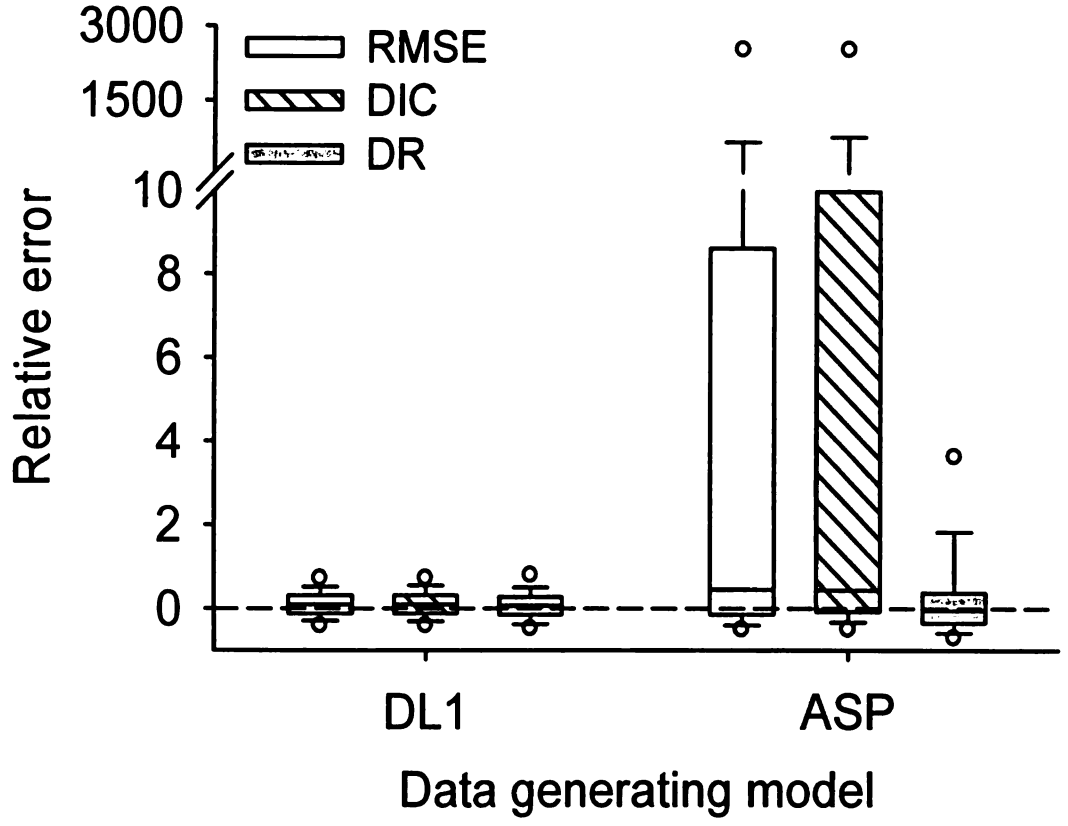


Figure 3.5. Box plots representing relative error distributions for estimates of population biomass in the last year of analysis chosen by model selection methods across different data generating models. The data generating models include double logistic functions with one time-varying parameter (DL1) and time-varying age-specific selectivity parameters (ASP). The model selection methods include root mean square error (RMSE), deviance information criterion (DIC), and degree of retrospectivity (DR). The bars represent median relative errors. The boxes, whiskers, and circles represent 25th and 75th, 10th and 90th, and 5th and 95th percentiles of the distributions, respectively.

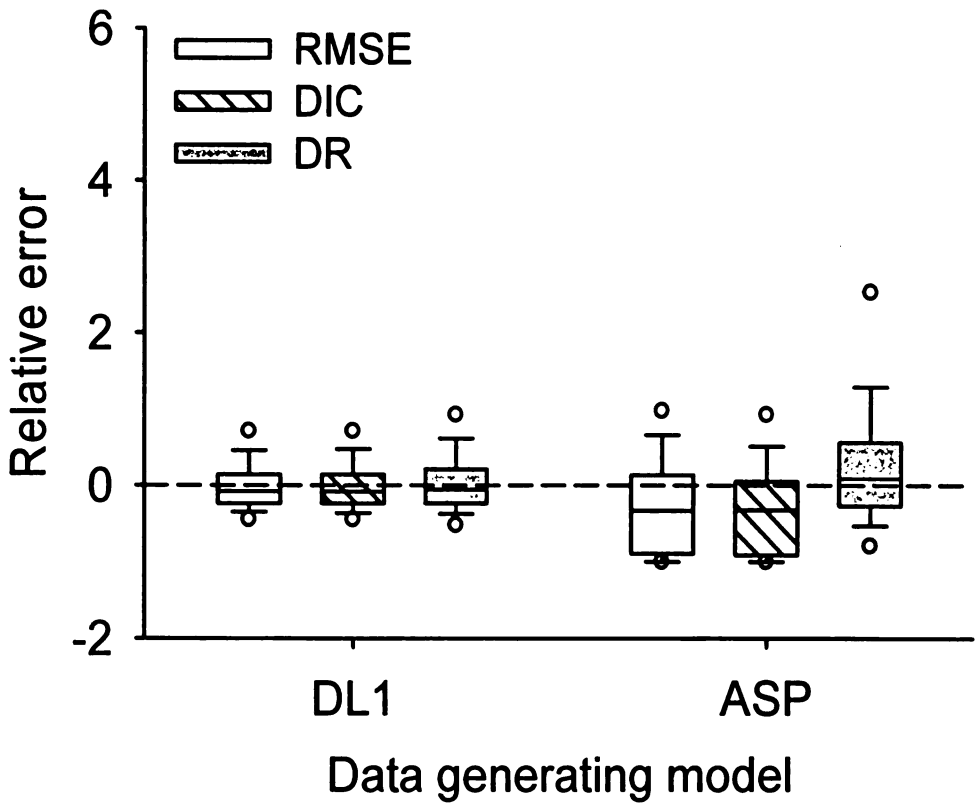


Figure 3.6. Box plots representing relative error distributions for estimates of exploitation rate in the last year of analysis chosen by model selection methods across different data generating models. The data generating models include double logistic functions with one time-varying parameter (DL1) and time-varying age-specific selectivity parameters (ASP). The model selection methods include root mean square error (RMSE), deviance information criterion (DIC), and degree of retrospectivity (DR). The bars represent median relative errors. The boxes, whiskers, and circles represent 25th and 75th, 10th and 90th, and 5th and 95th percentiles of the distributions, respectively.