Can spawning origin information of catch or a recruitment penalty improve assessment and fishery management performance for a spatially structured stock assessment model?

Yang Li [1](email: liyang11@msu.edu)
Quantitative Fisheries Center, Department of Fisheries and Wildlife
Michigan State University, 375 Wilson Rd., UPLA Room 101
East Lansing, Michigan, USA, 48824-1101

James R. Bence (email: bence@msu.edu)
Quantitative Fisheries Center, Department of Fisheries and Wildlife,
Michigan State University, 375 Wilson Rd., UPLA Room 101
East Lansing, Michigan, USA, 48824-1101

Travis O. Brenden (email: brenden@msu.edu)
Quantitative Fisheries Center, Department of Fisheries and Wildlife
Michigan State University, 375 Wilson Rd., UPLA Room 101
East Lansing, Michigan, USA, 48824-1101

[1]Corresponding author: Yang Li (email: liyang11@msu.edu, Telephone:517-355-0003).

Abstract

We used simulations based on Lake Whitefish (*Coregonus clupeaformis*) populations to explore the benefits of using spawning origin information for parsing catch to spawning populations in stock assessments for intermixed fisheries exhibiting an overlapping movement strategy. We compared this origin-informed assessment model with a standard assessment model that did not parse catch. We additionally evaluated the influence of including annual recruitment penalties. For standard assessment models, spawning stock biomass estimates could be unstable and biased (sometimes by more than 50%), depending upon population mixing and productivity, and in some cases estimated near average zero recruitment in the terminal year. Incorporating information on population-specific harvest age composition improved spawning stock biomass estimation (e.g., by sometimes essentially removing 50% biases, and improving accuracy). Assessments with recruitment penalties produced less biased terminal recruitment estimates (sometimes a 100% bias was removed). Under status quo target mortality rates improvements in assessments did not necessarily translate to improved fishery management performance (e.g., avoiding depletion of spawning biomass), but such improvements, and overall better performance, were seen at lower target mortality rates.

Introduction

Accurate estimation of spawning stock biomass and recruitment is important for the management of fishery stocks. Biased or imprecise estimates can influence measures of population productivity and year-class strength, stock-recruitment relationships, and management decisions (e.g., harvest regulations) that depend on these assessment results. When fish from distinct spawning populations intermix on fishing grounds during harvest periods (i.e., populations exhibit spatial structuring), estimating recruitment and spawning stock biomass dynamics for each spawning population from sampling programs that only target intermixed fisheries can be challenging. Statistical catch-at-age or catch-at-size models are commonly used for the assessment of commercial harvested fish populations for estimating biomass of spawning adults and recruitment dynamics. However, one known feature of such assessment models is that recruitment in the last several assessment years cannot be reliably estimated because there is little information about recruitment levels for those years. In addition, such assessment models typically ignore spatial structure and assume harvest is from a single population (i.e., the "unit stock" assumption).

When assessment data are collected from intermixed fisheries but a single population assumption is made in the stock assessment model, population abundance can be overestimated, which can further lead to inappropriate management advice especially for low productivity populations (Hutchings 1996; Fu and Fanning 2004; Ying et al. 2011; Hintzen et al. 2015; Li et al. 2015). For example, it has been argued that some Atlantic cod (*Gadus morhua*) and Pacific salmon (*Oncorhynchus* spp.) populations were overharvested due to intermixed fisheries that did not properly account for differences in population productivities (Hutchings 1996; Morishima and Henry 1999; Fu and Fanning 2004). To facilitate management of intermixed fisheries, spatially-explicit stock assessment models can be used that either incorporate tagging data within the stock assessment framework (Eveson et al. 2009; Vincent et al. 2017), or incorporate mixing and migration rates in assessment models as fixed quantities (Guan et al. 2013; Li et al. 2015). Both approaches allow for spatially-explicit estimation of abundances, mortality components, and other dynamic rates within an integrated stock assessment model.

When accounting for spatial structure in stock assessments, two alternative movement strategies are commonly recognized: diffusion and overlap (Porch et al. 2001). The diffusion movement strategy, also known as meta-population mixing (Ying et al. 2011), assumes that the fraction of fish populations that move away from their original spawning areas become part of the spawning populations near to which they move (i.e., their spawning population identity changes according to their movement behavior). Conversely, the overlap movement strategy assumes 100% spawning site fidelity meaning that fish always move back to their original natal areas during the spawning season, and thus spawning population identity is maintained throughout a fish's lifetime. In this paper we focus on stock assessment models assuming an overlap movement strategy. While this is clearly a simplification for any given stock, it is a reasonable approximation of spatial structure for many stocks.

A known problem for assessment models, when applied to populations exhibiting spatial structuring with moderate to high levels of intermixing, is that population-specific estimates of recruitment are uncertain or not estimable, and estimates of spawning stock biomass are unstable or biased, even when mixing rates are assumed known (Ying et al.

2011; Molton et al. 2012; Li et al. 2015). Li et al. (2015) proposed an overlap stock assessment model in which an integrated statistical catch-at-age (SCAA) assessment model was fit to overlapping fish populations by incorporating actual mixing rates in the model. They found that mixing among areas caused problems in estimating population-specific annual recruitments, and this led to substantial uncertainty and bias in estimation of recruitment and biomass. They hypothesized that this problem could be resolved if additional population-specific data were provided to the assessment model, such that harvest data could be allocated to source populations. Hintzen et al. (2015) evaluated the influence of fishery-independent survey data on the performance of an integrated catch-at-age method for intermixing fish populations, in which information on the classification of the catch to their spawning origin were used to inform survey indices (i.e., the proportions of survey sample to spawning populations). However, the catch data they used in the assessment model were not reallocated back to the spawning populations because their assessment model ignored spatial structure. Thus, mismatch between spatial structures in the assessment data and in the assessment model still existed. They found that spatially-explicit survey data marginally reduced bias in estimation of biomass, but when there were errors in classification rates inaccuracies could actually increase.

The goal of our research was to evaluate the benefits of including information on catch composition for the management of intermixing fish populations. Our research extended the overlap SCAA assessment model proposed by Li et al. (2015) by including information on population-specific harvest age composition, which could arise from having genetic or some other type of discriminatory characteristic (e.g., parasite community, meristic or morphometric feature) of the populations from a biological sample collected from the intermixed fisheries. Herein, we refer to the overlap assessment model proposed by Li et al. (2015) as the “standard assessment model”, and the extended one with additional data on population source as the “origin-informed assessment model”. In both assessment models, annual recruitments were estimated as free parameters, which is the same approach used by Li et al. (2015). We further propose two alternative assessment models that are identical to these two models except that a penalty on annual recruitment residuals was incorporated in each model. Several studies conducted for single populations (no spatial structure) have shown that adding such penalties or other constraints can improve estimates of annual recruitment, particularly for terminal assessment years (Maunder and Deriso 2003; Methot et al. 2011; Korman et al. 2012). We tested how assessment and management performance of the standard and origin-informed assessment models were influenced by the magnitude of recruitment variation, assessment data quality, uncertainty regarding mixing rates, and target mortality rates.

The dynamics of our simulations were based on lake whitefish (*Coregonus clupeaformis*) populations and fisheries in the upper Laurentian Great Lakes of North America, although results should have general applicability to populations with similar life history and movement patterns given the stochastic modeling of uncertainty and the range of sensitivity analyses we report. An overlap movement strategy was assumed for the simulated lake whitefish populations, because evidence suggests that lake whitefish populations in the Laurentian Great Lakes region overlap during non-spawning seasons but move back to where they were born during the spawning season of each year (Ebener et al. 2010a; Stott et al. 2010; Li et al. 2017). Although tagging studies have suggested that considerable movement of lake whitefish in the Laurentian Great Lakes region from management units containing their spawning grounds to other management units during the non-spawning and harvest seasons (Ebener et al. 2010b; Li et al. 2017), they are still largely managed as unit stocks. To our best knowledge, our research is the first to evaluate the influence of including population-specific catch information on a spatial structured stock assessment model. Compared to Hintzen et al. (2015), we propose a different approach of using such information for the management of intermixing stocks with a focus directed towards spatially structured stock assessments.

Methods

Simulation framework

Our simulation framework followed a management strategy evaluation approach (i.e., full closed-loop feedback simulation framework to evaluate alternative management procedures, Figure 1). These at simulations were designed to determine the long-term assessment and management performance for both standard and origin-informed assessment models with or without a lognormal penalty on annual recruitment residuals (Table 1). The operating model consisted of four hypothetical lake whitefish populations with age-structure and an overlap movement strategy (i.e., 100% natal fidelity was assumed) that intermixed across four areas of harvest. Observations from the four regions of harvest were then generated for input for the stock assessment models. Assessment models were fit to the observed data, and a harvest control rule was applied each year based on the assessment results so that target harvest levels (i.e., total allowable catch in our case) could be set. The management procedure then fed back to the operating model by implementing actual harvest based on the target with implementation error in the operating model of next year. Given we were considering alternative stock assessment models and the stock assessment results influenced dynamics, separate simulations were conducted for each assessment approach, albeit using the same random number seeds. To evaluate long-term performance of each assessment model, we ran each simulation for 100 years, and summarized results for the last 25 years. All symbols of index variables and accents used in the equations of this paper are identified in Table 2.

Operating model

The operating model was stochastic and age-structured (i.e., ages 3 to 12 with the last age class an aggregate group including age-12 and older fish), operated in annual time steps, and recognized four geographic fishing grounds that were presumed to surround the four spawning areas (i.e., each spawning area was associated and located within a fishing region). Yearly time steps were considered because evidence suggested that the movement of lake whitefish populations in the upper Great Lakes generally occurred soon after spawning (i.e., between late October and early December, Li et al. 2017). Thus, we assumed that fish moved away from their spawning areas on the first day of each year, and all surviving fish returned to their original spawning areas to spawn at the end of each year.

As described in detail below, many parameters of the operating model are taken from Li et al. (2015), which were based on a review of existing Lake Whitefish stock assessments. A single set of life history (growth, maturity) parameters was used, representative of those estimated from biological data used in those stock assessments. General levels of recruitment stochasticity and productivity, and variations among populations were based on analysis of recruitment and spawning stock sizes from the existing assessments. The existing assessments are unit stock assessments, and the influence of this on perceived differences in recruitment productivity was taken into account when specifying varying productivity levels (Li et al. 2015). In real assessments, with spawning populations that differ in life history, it is likely that there would be additional advantages of biological data that is spawning population specific, which we have not evaluated here.

For each simulated population, we modeled recruitment (age-3 fish) from a Ricker stock-recruitment function with a first-order autoregressive process (AR1):

$$N_{i,y,a=3} = \alpha_i SSB_{i,y-3} e^{-\beta_i SSB_{i,y-3}} e^{\varepsilon_{R,i,y}}. \quad (1)$$

$$\alpha_i = \alpha_i' e^{-0.5\sigma^2}.$$

$$\varepsilon_{R,i,y} = \rho \times \varepsilon_{R,i,y-1} + \tau_{R,i,y}.$$

$$\tau_{R,i,y} \sim Normal\ (0, {\sigma_R}^2).$$

$$\sigma^2 = \frac{{\sigma_R}^2}{1-\rho^2}.$$

where $N_{i,y,a=3}$ is the abundance of age-3 fish from population i at the beginning of year y, $SSB_{i,y-3}$ is the spawning stock biomass of population i in year $y-3$, and α_i and β_i are Ricker stock-recruitment function parameters for population i. The parameters ρ and σ_R defined the stochastic process for deviations of recruitment from the underlying Ricker stock-recruitment function, producing temporally autocorrelated recruitment. The level of process error presented in Table 3 was used for all simulated populations in the baseline scenario. Process error parameters were varied in the sensitivity analysis for evaluating the influence of recruitment variation on modeling results. The stock-recruitment parameter α', together with β, were chosen so that the deterministic stock recruitment would produce the desired average level of recruitment given stock size. For the simulations, α' was scaled by $e^{-0.5\sigma^2}$ so that the expectation of the stochastic form of the recruitment relationship would equal the deterministic value and not depend on the assumed level of recruitment variation.

Total spawning stock biomass (SSB) for population i in year y was calculated as the product of female percentage in the population (50%), weight-, maturity-, and abundance-at-age, and weight-specific fecundity (19733/kg). All equations and parameter values used for calculating SSB are defined in Table 4, which are the same as used by Li et al. (2015).

For each population, post-recruitment (after age-3) abundances at age (*a*) at the beginning of each year were forward projected using an exponential mortality model with a constant natural mortality (*M*) of 0.25, and age-, year-, and region-specific (*j*) fishing mortality (*F*):

$$N_{i,y+1,a+1} = N_{i,y,a} \sum_j \theta_{ij} \exp\left(-M - F_{j,y,a}\right). \quad (2)$$

According to equation 2, fish from a spawning population either remained in the region surrounding their natal area during the non-spawning season or moved to one of the other harvest areas, depending on the assumed mixing rates θ_{ij}. Thus, the survival of fish in a population was a weighted average of the survival rates in each of the harvest regions, with weights equal to the proportions of fish from the population residing in the regions during the non-spawning season. In some scenarios, mixing rates varied among the populations in the operating model, but in all cases were temporally invariant for each population.

We used stay rate θ_{ii} (i.e., the proportion of fish from spawning population i that stay in the area surrounding that population's spawning area during the non-spawning season) to represent movement dynamics for population i, and assumed that a greater stay rate indicated higher-quality habitat, so that a greater proportion of fish from other population moved to that area (Table 5). Thus, mixing rates θ_{ij} (i.e., the proportion of fish from spawning population i that move to the area surrounding population j's spawning area during the non-spawning season) were calculated as (Li et al. 2015):

$$\theta_{ij} = (1 - \theta_{ii})\frac{\theta_{jj}}{\sum_{k \neq i} \theta_{kk}}. \tag{3}$$

where the summation is overall all areas k except the fishing grounds surrounding the spawning area of population i. Total allowable catch (TAC) for each harvest area was determined via the management procedure described below. Actual harvest (C) in each year was set equal to the TAC multiplied by a lognormal implementation error term with a coefficient of variation (CV) of 10%:

$$C_{j,y} = TAC_{j,y} \exp\left(\zeta_{j,y} - 0.5{\sigma_{tac}}^2\right). \tag{4}$$

$$\zeta_{j,y} \sim Normal\ (0, {\sigma_{tac}}^2).$$

where σ_{tac} is the standard deviation of TAC implementation error ζ. The fully selected fishing mortality rate f that produced the actual harvest level given age-specific abundances was solved for using a Newton-Raphson algorithm and Baranov's catch equation:

$$C_{j,y} = \frac{s_a F_{j,y}}{M + s_a F_{j,y}} \left(1 - e^{-M - s_a F_{j,y}}\right) \sum_i N_{i,y,a} \theta_{ij}. \tag{5}$$

Age-specific *F*s were set equal to the solved *f* multiplied by age-specific selectivities s_a:

$$F_{j,y,a} = s_a f_{j,y}. \tag{6}$$

Selectivities for age-3 and older ages were calculated from a gamma function that produced a dome-shape selectivity pattern with peak selectivity for age-10:

$$s_a = \frac{a^{\eta} \exp(-\tau a)}{10^{\eta} \exp(-\tau 10)}. \tag{7}$$

where selectivity parameters τ = 1.26 year^{-1}, η = 13.074 cm (from Li et al. 2015), were assumed to be the same for different populations.

We used the same approach as Li et al. (2015) to determine initialization abundances for each simulation. Specifically, initialization abundances for the populations were set to their equilibrium values based on the target mortality rate and a deterministic version of our model (equilibrium for populations at different productivity levels are shown as the intersections in Figure 2). As well, like Li et al. (2015), during the initial 20-year period of each simulation, the harvest control rule based on the target mortality rate was applied to the actual abundances at age (i.e., the assessment modeling was skipped). This was necessary as prior to year 20 the required data time series for conducting assessments was not available. We were not interested in the transient dynamics during this initial period, and we set the starting conditions at the deterministic equilibrium solely to better ensure that the final 25 years of our 100-year simulations approximated steady-state conditions.

Management Procedure

We attempted to emulate key aspects of the management procedures for lake whitefish in the 1836 Treaty-ceded waters, including data collection, stock assessment, and application of a constant total mortality harvest control rule (1836 Treaty Waters Modeling Subcommittee 2017). The underlying premises were that collected data were used to assess the populations (Figure 1), that the assessment results provided estimates of the abundance of fish present in each region, and that target harvests were set based on estimated abundances in an attempt to achieve the same target total mortality rate in each harvest region. All evaluated assessment models used an integrated SCAA assessment model that correctly accounted for movements (i.e., stay and mixing rates were model inputs and were accurately known) among the regions, with the exception of the sensitivity analyses that evaluated the consequences of uncertain mixing rates. All assessment models fit the same population dynamic model to each of their observed data sets to estimate the parameters used to summarize population status and determine target harvest. When fitting the assessment models, only the most recent 20 years of data were used. We elected to use a fixed-length time series so that the amount of information available to an assessment remained stationary during the performance evaluation period (the last 25 years of each 100-year simulation). While relatively short by assessment standards, 20 years represents more than three times the expected period between birth and production of offspring, given the assumed life history, fishery selectivity, and target mortality rate in our operating model, based on Lake Whitefish. Simulations using a 40-year assessment period for a subset of scenarios produced nearly identical results to those with the 20-year assessment period. Age range of the assessment models was the same as that of the operating model. By minimizing the negative log-likelihood (see objective function subsection below), the assessment models were considered to have converged on a solution when the maximum gradient of the parameters was less than 0.001, and the Hessian matrix was positive definite. Convergence rate is defined as the fraction of simulations that met both of the above criterions.

For the standard assessment models with or without a recruitment penalty (i.e., S and S W/Rec in Table 1), observed harvest, effort, and aggregated (across populations) harvest age composition data were collected annually for each region. For the origin-informed assessment models (i.e., O and O W/Rec in Table 1), observed harvest, effort, and population-specific harvest age composition data were collected annually for each region. Observed harvest differed from actual harvest as a result of observation error, which was modeled with a lognormal error term υ:

$$\tilde{C}_{j,y} = C_{j,y} \exp(\upsilon_y - 0.5{\sigma_c}^2). \qquad (8)$$

$$\upsilon_y \sim Normal\ (0, {\sigma_c}^2).$$

The observed fishing effort was a function of fishing mortality f, catchability q, and a lognormal observation error μ and we assumed ${\sigma^2}_F = 4\ {\sigma_c}^2$.

$$E_{j,y} = \frac{f_{j,y}}{q} exp(\mu_{j,y} - 0.5{\sigma^2}_F). \qquad (9)$$

$$\mu_{j,y} \sim Normal(0, {\sigma^2}_F).$$

In the baseline scenario, baseline level of CVs for the error terms of observed harvest and effort were used (Table 3) while different levels of CVs were explored in the sensitivity analyses for data quality.

For the standard assessment models, aggregated observed age compositions for area-specific harvests were generated from multinomial distributions with probabilities equal to the actual age composition. For the origin-informed assessment models, observed population-specific age compositions for area-specific harvests were generated from multinomial distributions with probabilities equal to the actual population-specific age compositions in each region. The effective sample size (N_{eff}) for the multinomial distribution used to generate aggregated and population-specific age compositions was assumed at its baseline level (Table 3), except for the sensitivity analyses for data quality.

Recruitment ($\widehat{N}_{i,y,a=3}$) of each assessment year, abundances at age (except age at recruitment) in the first assessment year ($\widehat{N}_{i,y=1,a>3}$), gamma function selectivity parameters ($\hat{\tau}$, $\hat{\eta}$), catchability ($\hat{q}$), the annual deviation from general level of fishing mortality($\widehat{\varepsilon F}_{j,y}$, Fournier and Archibald 1982), and the standard deviation from observed harvest ($\hat{\sigma}_c$) were estimated during assessment model fitting. Recruitments in the standard and origin-informed assessment models without recruitment penalty were estimated as free parameters. For the assessment models that included a recruitment penalty, recruitment for each population i was reparameterized as the product of average recruitment ($\widehat{R\mu_i}$) multiplied by an annual residual ($\varepsilon'_{i,y}$) that was exponentiated and bias corrected, so that the annual recruitment was assumed to come from a lognormal distribution:

$$N'_{i,y,a=3} = \widehat{R\mu_i} e^{\varepsilon'_{i,y} - 0.5\sigma'^{2}_{R}}. \quad (10)$$

$$\varepsilon'_y \sim Normal(0, \sigma'^{2}_{R}).$$

Post-recruit abundances at age in the first assessment year were estimated as free parameters. The fishing mortality in the assessment models was modeled in the same way as for the operating model, which was a product of selectivity at age and fully selected fishing mortality (same as in Equations 6 and 7, but here $\hat{\tau}$ and $\hat{\eta}$ were estimated parameters). The fully selected fishing mortality ($f'_{j,y}$) was modeled as a product of assessed catchability ($\hat{q}$), observed effort ($\tilde{E}_{j,y}$), and assessed annual deviation from general level of fishing mortality ($\widehat{\varepsilon F}_{j,y}$).

The natural mortality rates assumed in all assessment models were the same as those used for the operating model. The parameters of all assessment models were estimated in AD Model Builder (Fournier et al. 2012).

The population dynamics in all stock assessment models (i.e., S, S W/Rec, O, and O W/Rec) followed:

$$N'_{i,y+1,a+1} = N'_{i,y,a} \sum_j \theta_{ij} \exp(-M - F'_{j,y,a}). \tag{11}$$

$$C'_{j,y,i,a} = \frac{F'_{j,y,a}}{M + F'_{j,y,a}} (1 - e^{-M - F'_{j,y,a}}) N'_{i,y,a} \theta_{ij}. \tag{12}$$

$$C'_{j,y,a} = \sum_i C'_{j,y,i,a}. \tag{13}$$

For each harvest area, aggregated harvest age composition for the standard assessment models (Equation 14, Table 1), and population-specific harvest age composition for the origin-informed assessment models (Equation 15, Table 1) were:

$$p'_{j,y,a} = C'_{j,y,a} / \sum_a C'_{j,y,a}. \tag{14}$$

$$p'_{j,y,i,a} = C'_{j,y,i,a} / \sum_{i,a} C'_{j,y,i,a}. \tag{15}$$

Predicted SSB was calculated from estimated abundance at age $N'_{i,y,a}$ by using equation 1, and assuming weight, maturity at age and weight-specific fecundity were known (Table 4).

Objective function

The objective function for each assessment model was the summation of at least three negative log-likelihood and log-prior/penalty components (Table 1). All four assessment models assumed the same lognormal distributions for the log-likelihood component of total fishery annual harvest by harvest area and for the log-prior components associated with the fishing mortality-effort relationship for each harvest area.

The total negative log-likelihood component for the log of area-specific annual fishery harvest was based on a normal distribution

$$\ell_c = \sum_j (nlog_e(\hat{\sigma}_c) + \left(\frac{1}{2\hat{\sigma}_c{}^2}\right) \sum_y (log_e(\frac{\tilde{C}_{j,y}}{\hat{C}_{j,y}}))^2), \tag{16}$$

where n was the number of assessment years (i.e., 20 years). A normal distribution was also assumed for the log-prior penalty associated with the log annual deviation from the general level of fishing mortality

$$\ell_{\varepsilon F} = \sum_j (nlog_e(\sigma_F{}') + \left(\frac{1}{2\sigma_F{}'^2}\right) \sum_y (log_e(\widehat{\varepsilon F}_{j,y}))^2), \tag{17}$$

where $\sigma_F{}'^2$ was assumed to be four times greater than $\hat{\sigma}_c{}^2$, which matched what was assumed in the operating model. This penalty was equivalent to predicting effort as proportional to estimated fishing mortality and treating deviations between the log of observed and predicted fishing effort as normally distributed (Fournier and Archibald 1982).

The third likelihood component was associated with harvest age composition and was based on a multinomial distribution, but there were differences in this likelihood component for standard and origin-informed assessment models. For the standard assessment model (assessment models S and S W/Rec, Equation 18), the negative log likelihood component was for the aggregate harvest age composition for the harvest regions

$$\ell_a = -\sum_j \sum_y N_{eff} \sum_a (\tilde{p}_{j,y,a} log_e p'_{j,y,a}). \quad (18)$$

where $\tilde{p}_{j,y,a}$ and $p'_{j,y,a}$ are the observed and estimated proportions of harvest in area j by age a in year y and N_{eff} is the assumed effective sample size. For the origin-informed assessment models (assessment models O and O W/Rec, Equation 19), the negative log likelihood component was for the population-specific harvest age composition for the harvest regions

$$\ell_{pa} = -\sum_j \sum_y N_{eff} \sum_{i,a} (\tilde{p}_{j,y,i,a} log_e p'_{j,y,i,a}). \quad (19)$$

where $\tilde{p}_{j,y,i,a}$ and $p'_{j,y,i,a}$ are the observed and estimated proportions of harvest in area j by age a from population i in year y, respectively. For baseline scenarios, N_{eff} was set equal to 50 for both standard and origin-informed assessment models, but was varied in sensitivity analyses to evaluate the influence of data quality.

For standard and origin-informed assessment models that included a penalty on annual recruitment residuals (i.e., S W/Rec and O W/Rec in Table 1), the objective function included a log-penalty component that constrained the annual recruitment residuals $\varepsilon'_{i,y}$ of equation 10 based on a normal distribution with standard deviation σ'_R equal to 2.0. In other words, the log-penalty on annual recruitment residuals was modeled as

$$\ell_R = \sum_j(\sum_y log_e(\sigma'_R) + \frac{{\varepsilon'_{i,y}}^2}{2{\sigma'_R}^2}). \qquad (20)$$

Application of the harvest control rule

To mimic the timing of implementing assessments and setting harvest targets of lake whitefish fisheries in 1836 Treaty-ceded waters, we included a one-year lag between data collection and incorporation in the four stock assessment models. More specifically, an annual assessment was conducted each year of a simulation based on data collected through the previous year, to set the harvest targets for the following year. In the lag year, abundances were projected based on an exponential population model where total mortality rates were assumed to be the mean of the last three years' value, and recruitments were assumed to be the mean of the most recent 10 years. During the year of setting harvest targets (after the lag year), we used the same approach as in the lag year to project abundance at the beginning of that year. We then used Baranov's catch equation (same as in equation 12 and 13) to calculate harvest target, while the fishing mortality rates were adjusted to the target fishing mortality rates, which can be calculated based on target mortality rates, estimated selectivity-at-age, and natural mortality rate.

Simulation Scenarios

We evaluated five productivity and movement scenarios (Table 5), and six sensitivity analysis scenarios (Table 3 and 6). We also evaluated all cross-combinations of productivity/movement scenarios and sensitivity analysis, and full results are available in the supplementary material. For each evaluated scenario, 200 simulations were conducted. In the baseline scenario (Table 5), we assumed the four simulated populations had equal stay rates and productivity levels to establish a baseline for comparison of assessment and management performance results. Then we explored alternative operating model settings with different productivity and movement assumptions, to evaluate the consequences of different combinations of productivity and movement dynamics of lake whitefish populations on stock assessments. We also evaluated outcome sensitivity to different quality of assessment data, uncertain mixing rates assumptions, and recruitment variability.

Baseline scenario and alternative productivity and movement scenarios

We explored five scenarios of population-specific movement dynamics and productivity (scenario 1 is the baseline scenario) (Table 5). Overall, there were three different levels of productivity (i.e., low, baseline, and high), and three different stay rates during non-spawning season (low, medium, high). Each productivity level corresponded to a specific steepness parameter, and different productivity levels shared the same unfished equilibrium spawning stock size (Table 3). However, higher productivity levels would lead to greater fished equilibrium stock size and recruit levels (Figure 2). Target mortality rate (Target_A; annual death rate=1.0-annual survival rate) was assumed to be 0.65 as a baseline level, which is the current rate used in 1836 Treaty-ceded management of lake whitefish, although as part of sensitivity scenarios explored the effects of a lower target mortality rate.

In the baseline scenario (scenario 1), the four populations had identical "baseline" productivity and stay rates set to "medium" levels. Scenario 2 explored a case in which the four populations still had equal medium levels of movement, but two of the populations had low productivity while the other populations had high productivity. In scenarios 3 to 5, the four populations had different stay rates and either had equal productivity levels (scenario 3) or unequal productivity levels (scenario 4: positive correlation between productivity level and stay rate; scenario 5: negative correlation between productivity level and stay rate).

Sensitivity Analyses

A total of six sensitivity scenarios (Table 6) were conducted to determine whether baseline results remained consistent after modifying specific conditions of the examined scenario (e.g., poor data quality). The purpose of the sensitivity analyses was to determine the general applicability of model results.

Data Quality—The first two sensitivity scenarios considered different levels of data quality available for assessment models: low and high (relative to the baseline level), by varying effective sample size (N_{eff}) and the CVs for harvest and effort (Tables 3 and 6). The low and high levels of data quality were chosen to reflect the extreme data quality cases evaluated by Li et al. (2016) based on ranges seen in retrospective errors for actual lake whitefish stock assessments in the 1836 Treaty-ceded waters.

Uncertain Mixing Rates—In the baseline scenario, the mixing rates were consistent across populations and simulation years in the operating model, and assumed as correctly known parameters in the stock assessment model. In the third sensitivity scenario, we assumed that annual stay rates in the assessment models were still treated as known parameters, but did not match the true θ_{ii} in the operating model. The annually varying stay rates $\theta_{ii,y}$ used in the assessment model were parameterized by a 'logistic' function of re-parameterized rates (ω_y)

$$\theta'_{ii,y} = \exp(\omega'_y)/(\exp(\omega'_y) + 1). \quad (21)$$

The annual values for ω'_y were generated from a normal distribution (Table 3). Different sets of mean and variance values were assumed to ensure the annually varying stay rates used in the assessments were within 10% of the true θ_{ii}.

Recruitment Variation—For the next two sensitivity scenarios, we explored two recruitment variability levels (Table 3 and Table 6). In the high recruitment variability scenario, we kept the autocorrelation coefficient at 0.45 as in the baseline scenario but increased the stationary standard deviation in the recruitment process error to 1.5. For the second level, we removed the autocorrelation component of recruitment variation so that the recruitment variation was simply white noise, and kept the same stationary variance as for the baseline scenario.

Target mortality—For the last sensitivity scenario, a lower target mortality rate (Target_A) of 0.55 was implemented in the management procedure because this rate has been identified as sustainable for a wide range of lake whitefish populations with different productivities (Li et al. 2015).

Performance Statistics

Performance statistics for evaluating the different assessment models were average SSB, the proportion of years SSB was less than 20% of the unfished SSB level ($P(SSB<B_{20\%})$), average annual total yield and inter annual variation (IAV) in yield by area, relative error (RE) in the terminal assessment year SSB, and RE of estimating recruitment for all assessed years, over the last 25 years of the simulations. Relative error was calculated as $RE = (\bar{x} - x)/x$, where $\bar{x}$ is the predicted value based on the assessment results and x is the true value generated from the operating model. We additionally estimated the autocorrelation in RE in the terminal assessment year SSB over the last 25 years for each

simulation. This was intended to assess autocorrelation in assessment errors under stationary conditions. The autocorrelation was estimated by fitting an AR1 model to the time series of REs in terminal SSB resulting from each simulation by ordinary least squares. We used the ar.ols function from stats package in R 3.2.2 for the autocorrelation coefficient (ARC) calculation (R Core Team 2016). A large positive ARC would imply that the assessment errors tended to be similar for multiple years in a row. The distributions of the performance statistics calculated over all 200 simulations for an evaluated scenario, were summarized by the median and inter-quartile range. We choose to run 200 simulations because preliminary results of the baseline scenario suggested that results from 200 simulations were nearly identical from those based on 1000 simulations.

Results

In general, all four assessment models converged on solutions. Convergence rate of the assessments was >93% across all scenarios for the origin-informed model (O), the origin-informed model with recruitment penalty (O W/Rec), and the standard model with recruitment penalty (S W/Rec). Although the convergence rate of the standard assessment model (S) was 95% for the baseline scenario, it was less than 90% for other evaluated scenarios. Including a recruitment penalty increased the convergence rate for both standard and origin-informed models by 8.0% and 1.7% on average across all scenarios, with the largest improvement in convergence by 20.8% for the standard assessment approach under scenario 5 with low data quality (Tables 5 and 6).

Baseline scenario

Under the baseline scenario, where the simulated populations had the same stay rates and productivity levels, the expected assessment and management performance was the same across all populations, and indeed the realized performance results were nearly identical (see full results in Supplementary material). Consequently, we summarize the results for only one of the four populations (i.e., Population 1 in Table 5). Compared to the standard assessment models (i.e., S and S W/Rec), adding population-specific harvest age composition in the origin-informed assessment models (i.e., O and O W/Rec) in general resulted in less bias and more weakly autocorrelated estimates of SSB in the terminal assessment year with smaller inter-quartile ranges (Figures 3a and 3f), and less uncertainty in estimates of recruitment (based on smaller inter-quartile ranges of RE) over all assessment years except for the final two years (Figure 3b). However, the origin-informed assessment model performance did not translate into benefits in the management performance statistics, such as average true SSB and yield, with only slightly improvement in the IAV of yield (3c, 3d and 3e, and supplementary materials). When a recruitment penalty was added to both the standard and origin-informed assessment models (comparing S W/Rec and O W/Rec with S and O), this resulted in less IAV of yield (median IAV of yield decreased by 0.05 and 0.04 for standard and origin-informed models, Figure 3e), and lower bias in estimates of recruitment for the last two assessment years (Figure 3b), but slightly higher risk of SSB being lower than 20% of its unfished level (median P(SSB<$B_{20\%}$) increased by 3.8% and 7.7%, Figure 3c).

Both the standard and origin-informed assessment models without recruitment penalties had considerable difficulty in estimating recruitment levels in the terminal assessment year. In most simulations, the recruitment RE in the terminal assessment year was -100%, meaning that recruitment was being estimated at essentially 0 fish (Figure 3b and 4). However, when a recruitment penalty was included in the assessments (comparing S

W/Rec with O W/Rec), the origin-informed assessment model (i.e., O W/Rec) produced less biased estimates for the terminal assessment year recruitment (Figure 3b and Figure 4).

Alternative productivity and movement scenarios

For the alternative productivity and movement scenarios, we present results only for populations 1 and 3 because for these scenarios populations 1 and 2 and populations 3 and 4 had nearly identical results due to their same productivity and stay rates. When low and high productivity populations intermixed (Scenario 2, 4, and 5 in Figure 5), low productivity populations generally had high risk of being overfished (i.e., the interquartile ranges of average true SSB were below 20% of the unfished level) across all scenarios. Regardless of whether a penalty for annual recruitment residuals was included, the origin-informed assessment models (i.e., O and O W/Rec) substantially outperformed the standard assessment models (S and S W/Rec) in terms of estimation of SSB of the terminal assessment year for low productivity populations, but using population-specific harvest age composition data had only a slight influence on estimation of SSB for high productivity populations. More specifically, for the low productivity populations, the RE of estimated terminal assessment year SSB in year 100 was less biased, and the autocorrelation for these estimates over the last 25 years was lower for assessment models O and O W/Rec than for S and S W/Rec. Such differences in assessment performance were greater for scenarios where there was a negative correlation between stay rates and productivity. For the scenario where populations had the same productivity but different stay rates (Scenario 3), assessment performance results were similar to those of the baseline scenario.

With respect to the estimation of terminal assessment year recruitment and for management performance statistics, results for all alternative productivity and movement scenarios were similar to those found in the baseline scenario. Neither the standard or origin-informed assessment models without recruitment penalties could produce reliable estimates of recruitment in the terminal assessment year. When low productivity populations intermixed with high productivity populations (Scenario 2, 4, and 5 in Figure 5), standard and origin-informed assessment models with recruitment penalties resulted in unbiased recruitment estimates in the terminal assessment year for high productivity population, but positive bias in recruitment estimates in the terminal assessment year for low productivity populations.

Sensitivity Analyses

The assessment and management performances for all the assessment models were generally insensitive to changes in the magnitude of actual recruitment variation, target mortality, data quality, and to uncertain mixing rates assumptions (Figure 6), with patterns in performance statistics similar to those of the baseline scenario. There were only three exceptions. First, with a lower total mortality target (55%), the origin-informed assessment models both with and without recruitment penalties had better management and assessment performance than the standard assessment models, as evidenced by lower P(SSB<$B_{20\%}$) (median at 0.08 for O and at 0.12 for S), similar or even higher yield (median at 204.8 for O and at 204.2 for S), lower IAV of yield (median at 0.32 for O and at 0.35 for S), and less biased with smaller inter-quartile range (inter-quartile range [-

0.13,0.10] for O and [-0.18,0.15] for S), and less autocorrelated estimates of SSB (median at 0.37 for O and at 0.44 for S) in the terminal assessment year. Second, when recruitment variation was high, P(SSB<$B_{20\%}$) was higher, and average yields were lower for all four assessment models. In addition, for this high recruitment scenario both assessment models with recruitment penalties tended to overestimate recruitment (RecV_H in Figure 6). Finally, when assessment data quality was low (RecV_L in Figure 6), all four assessment models tended to underestimate SSB, have greater IAV of yield, and greater inter-quartile range for the RE of estimating terminal year SSB.

Discussion

Attempting to account for movement in fish stock assessment models has become increasingly common for the management of intermixed fisheries (Cope and Punt 2011; Ying et al. 2011; Molton et al. 2012; Li et al. 2015; Vincent et al. 2017). In this study, we evaluated four spatially-structured SCAA models (standard assessment, standard assessment with recruitment penalty, origin-informed assessment, origin-informed assessment with recruitment penalty) for assessing lake whitefish populations that were assumed to exhibit an overlap movement strategy. We aimed to evaluate if considering additional assessment data about classification of catch to spawning origin, and adding a penalty for annual recruitment residuals, could improve the assessment and management performance of the overlap SCAA model proposed by Li et al. (2015). We found that data allowing parsing of catch from a management area to the specific spawning population the fish came from could result in less biased and less auto-correlated estimates of spawning stock biomass (SSB) in terminal assessment years, and less uncertainty in estimates of recruitment early in the time period assessed; while including a lognormal penalty on annual recruitment residuals in assessment models substantially improved the estimation of recruitment in the terminal assessment years. With the penalty, data on population source also led to improved terminal recruitment estimates.

When we used data on the classification of catch to spawning origin in our proposed overlap assessment models, we assumed a multinomial distribution of population-age composition for each year of harvest from an area. This is an extension of what we assumed in our standard SCAA model in which a multinomial distribution was assumed, as is often done, for age composition of harvest. Use of these additional data did provide better estimation of the spawning stock biomass (SSB) in the terminal assessment year. Hintzen et al. (2015) reached a similar conclusion but with a small level of improvement when they used such data to inform survey indices for an integrated stock assessment model. This may be due to the mismatch between the spatial structures in their assessment data of catch and in the assessment model. Although spawning origin information allowed the assessment model to incorporate correct (or with uncertainty) survey indices, because their assessment model ignored spatial structure in the observed catch data such a mismatch can still lead to biased estimation of biomass and recruitment. Our results suggested that such improvements in assessment performance did not necessarily translate into improved management performance, except when we used a lower than status-quo mortality target. Under the status-quo mortality target, although the origin-informed assessment models provided better estimation of SSB than the standard overlap models, the calculated total allowable catch (TAC) based on the estimated SSB was still not sustainable. Coincidentally, because the standard assessment models tended to underestimate SSB, it resulted in a more “appropriate/conservative” TAC. This argument is evidenced by our sensitivity analysis with lower target mortality rate (Target_A=55%) in which origin-informed assessment models had better management and assessment performance than standard assessment models.

Past studies have found that when populations with different productivity levels intermix during harvest season, populations with lower productivity are generally more vulnerable to overharvest (Ricker 1958; Paulik et al. 1967; Hintzen et al. 2015; Li et al. 2015). The results from this study are consistent with those studies. We found that there was a high risk of being overfished for low productivity populations, especially when low productivity populations with high stay rates intermixed with high productivity populations with low stay rates. In such a case, for low productivity populations, standard assessment models tended to overestimate SSB, while the origin-informed assessment models provided nearly median unbiased estimation of SSB. We suspect that the standard assessment model is challenged to identify the correct age composition for low productivity populations from the aggregate sample collected from each harvest area, because they consist of mixtures of age compositions from populations with different productivity, with contributions depending on population productivities and movement rates. Conversely, information on population-specific age compositions for area-specific harvests provides sufficient information to prevent inaccuracies in SSB estimates.

Our sensitivity analysis suggested that the improvement by including population-specific age compositions for area-specific harvests was limited to scenarios without high assessment data quality. In other words, when data quality is high, standard assessment models can provide sufficiently accurate estimates of population-specific SSB when supplied with accurate mixing rates. Thus, an origin-informed assessment model may not be necessary in conditions of high data quality and accurate information on mixing. We must emphasize that our consideration of data quality was focused on precision rather than potential biases in data. We also did not consider model misspecification except for the unmatched mixing rates assumed in the operating and assessment models in the sensitivity analyses, and our stochastic assumptions regarding recruitment for the models with recruitment penalties. A formal evaluation of how model misspecification affects the performance of spatially structured stock assessment model was outside the scope of our research but we would encourage investigations on this topic. We anticipate consequences of model misspecification to be case specific. Some cases of model misspecification may change the scale of biomass assessment, and this would not change the relative performance of the four assessment models we evaluated because target F in all assessment models would be adjusted to count for bias in similar manners. In other cases, however, model misspecification may lead to too high estimation errors. In such there may not be a strong justification for collecting population-specific data because the advantages of origin-informed assessment models over the standard models may not be clear.

The other major finding from this research was that including a lognormal penalty on annual recruitment residuals in both standard and origin-informed assessment models markedly improved the estimation of recruitment at the end of the assessment period. This is consistent with what has been found in evaluations of stock assessments without spatial structure (Maunder and Deriso 2003; Methot et al. 2011; Korman et al. 2012). Although the inclusion of a recruitment penalty did not prevent recruitment from being overestimated when recruitment variation in the operating model was high, its performance was still better than when a recruitment penalty was not included. This overestimation may stem, in part, from the large standard deviation for the distribution governing the annual recruitment deviations in the assessment models with recruitment deviations. We also found that IAV of yield was lower when a recruitment penalty was incorporated. This may result from the more stable/reasonable estimation of recruitment at the end of the assessment year period. Such stabilization of recruitment estimates can lead to a more stable prediction of future abundance, and that is what the TAC calculation is based on. Also, because we included a 1-year lag between assessment data collection and assessment model implementation to mimic the real management procedure for lake whitefish in Laurentian Great Lakes region, the impact of recruitment estimation near the end of the time series is magnified, given we needed to project an additional year over what is assumed in some studies.

In summary, we found that for a spatially structured SCAA model that incorporated information on population-specific age composition of harvest resulted in less biased and less correlated estimates of spawning stock biomass (SSB) in terminal assessment years, and less uncertainty in estimating recruitment in early assessment years. Including a lognormal penalty on annual recruitment residuals in the spatial structured SCAA model substantially improved the estimation of recruitment in the terminal assessment years, which we suggest as “best practice” for spatially-structured assessment models. Despite the improved assessment performance, preventing overharvest of low productivity populations when using such assessments will still require an appropriate harvest policy, such as lower target mortality rates or precautionary reference points. Different approaches for parsing catch to contributing populations are likely to have different levels of classification accuracy. For example, genetic classification methods may be more accurate than otolith microchemistry methods if there are not strong environmental differences among spawning locations. Further research into how assessment model performance is affected by classification accuracy would be beneficial. We also recommend additional investigation of factors such as the inclusion of more complex spatial structure (e.g., seasonal movement), alternative harvest policies, model misspecification, and alternative spatial structured stock assessment models (e.g., spatially structured virtual population analysis, tag integrated assessment model) to evaluate the benefits of parsing catch to spawning populations when it comes to the management of spatially-structured populations.

Acknowledgements

Funding for this project was provided by the Great Lakes Fishery Commission (Project 2013_BRE_44025) and Great Lakes Fishery Trust (Project 2012.1250). Additional funding support was provided by supporting partners of the Quantitative Fisheries Center and Michigan Department of Natural Resources through the Partnership for Ecosystem Research and Management program. We acknowledge the support of the Michigan State University High Performance Computing Center and the Institute for Cyber-Enabled Research. We wish to thank the Modeling Subcommittee of the Technical Fisheries Committee for 1836 Treaty-ceded waters for providing input on our research. This is publication 201X-XX of the Michigan State University Quantitative Fisheries Center.

Reference

1836 Treaty Waters Modeling Subcommittee, Technical Fisheries Committee. 2017. Technical fisheries committee administrative report 2017: status of lake trout and lake whitefish populations in the 1836 treaty-ceded waters of Lakes Superior, Huron and Michigan, with recommended yield and effort levels for 2017.

doi:10.13140/RG.2.2.19581.15844.

Cope, J.M., and Punt, A.E. 2011. Reconciling stock assessment and management scales under conditions of spatially varying catch histories. Fish. Res. **107**(1–3): 22–38. doi:10.1016/j.fishres.2010.10.002.

Ebener, M.P., Brenden, T.O., and Jones, M.L. 2010a. Estimates of fishing and natural mortality rates for four lake whitefish stocks in northern Lakes Huron and Michigan. J. Great Lakes Res. **36**: 110–120. doi:10.1016/j.jglr.2009.06.003.

Ebener, M.P., Brenden, T.O., Wright, G.M., Jones, M.L., and Faisal, M. 2010b. Spatial and temporal distributions of lake whitefish spawning stocks in Northern lakes Michigan and Huron, 2003–2008. J. Great Lakes Res. **36**: 38–51. doi:10.1016/j.jglr.2010.02.002.

Eveson, J.P., Laslett, G.M., and Polacheck, T. 2009. A spatial model for estimating mortality rates, abundance and movement probabilities from fishery tag-recovery data. *In* Modeling Demographic Processes In Marked Populations. doi:10.1007/978-0-387-78151-8.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. **27**(2): 233–249. doi:10.1080/10556788.2011.597854.

Fournier, D., and Archibald, C.P. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. **39**(8): 1195–1207. doi:10.1139/f82-157.

Fu, C., and Fanning, L.P. 2004. Spatial considerations in the management of Atlantic cod off Nova Scotia, Canada. North Am. J. Fish. Manag. **24**(3): 775–784. doi:10.1577/M03-134.1.

Guan, W., Cao, J., Chen, Y., Cieri, M., and Quinn, T. 2013. Impacts of population and fishery spatial structures on fishery stock assessment. Can. J. Fish. Aquat. Sci. **70**(8): 1178–1189. doi:10.1139/cjfas-2012-0364.

Hintzen, N.T., Roel, B., Benden, D., Clarke, M., Egan, A., Nash, R.D.M., Rohlf, N., and Hatfield, E.M.C. 2015. Managing a complex population structure: exploring the importance of information from fisheries-independent sources. ICES J. Mar. Sci. **72**(2): 528–542. doi:10.1093/icesjms/fsu102.

Hutchings, J.A. 1996. Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock’s collapse. Can. J. Fish. Aquat. Sci. **53**(5): 943–962. doi:10.1139/f96-097.

Korman, J., Martell, S.J.D., Walters, C.J., Makinster, A.S., Coggins, L.G., Yard, M.D., Persons, W.R., and Quinn, T. 2012. Estimating recruitment dynamics and movement of rainbow trout (Oncorhynchus mykiss) in the Colorado River in Grand Canyon using an integrated assessment model. Can. J. Fish. Aquat. Sci. **69**(11): 1827–1849. doi:10.1139/f2012-097.

Li, Y., Bence, J.R., and Brenden, T.O. 2015. An evaluation of alternative assessment approaches for intermixing fish populations: a case study with Great Lakes lake whitefish. ICES J. Mar. Sci. **72**(1): 70–81. doi:10.1093/icesjms/fsu057.

Li, Y., Bence, J.R., and Brenden, T.O. 2016. The influence of stock assessment frequency on the achievement of fishery management objectives. North Am. J. Fish. Manag. **36**(4): 793–812. doi:10.1080/02755947.2016.1167145.

Li, Y., Bence, J.R., Zhang, Z., and Ebener, M.P. 2017. Why do lake whitefish move long distances in Lake Huron? Bayesian variable selection of factors explaining fish movement distance. Fish. Res. **195**: 169–179. doi:10.1016/j.fishres.2017.07.014.

Maunder, M.N., and Deriso, R.B. 2003. Estimation of recruitment in catch-at-age models. Can. J. Fish. Aquat. Sci. **60**(10): 1204–1216. doi:10.1139/f03-104.

Methot, R.D., and Taylor, I.G. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. **68**(10): 1744–1760. doi:10.1139/f2011-092.

Modeling Subcommittee, T.F.C. 2017. Technical fisheries committee administrative report 2017: status of lake trout and lake whitefish populations in the 1836 treaty-ceded waters of Lakes Superior, Huron and Michigan, with recommended yield and effort levels for 2017. doi:10.13140/RG.2.2.19581.15844.

Molton, K.J., Brenden, T.O., and Bence, J.R. 2012. Control rule performance for intermixing lake whitefish populations in the 1836 Treaty waters of the Great Lakes: A simulation-based evaluation. J. Great Lakes Res. **38**(4): 686–698. doi:10.1016/j.jglr.2012.09.014.

Morishima, G.S., and Henry, K.A. 1999. The history and status of pacific northwest chinook and coho salmon ocean fisheries and prospects for sustainability. *In* Sustainable Fisheries Management: Pacific Salmon. *Edited by* E.E. Knudsen, C.R. Steward, D.D. MacDonald, J.E. Williams, and D.W. Reiser. CRC Press, Boca Raton, FL. pp. 219–236.

Paulik, G.J., Hourston, A.S., and Larkin, P.A. 1967. Exploitation of Multiple Stocks by a Common Fishery. J. Fish. Res. Board Canada **24**(12): 2527–2537. doi:10.1139/f67-202.

Porch, C.E., Turner, S.C., and Powers, J.E. 2001. Virtual population analyses of atlantic bluefin tuna with alternative models of transatlantic migration: 1970-1997. Int. Comm. Conserv. Atl. Tunas Collect. Vol. Sci. Pap. **52**(3): 1022–1045.

R Core Team. 2016. R: A language and environment for statistical computing. R Found. Stat. Comput. doi:10.1007/978-3-540-74686-7.

Ricker, W.E. 1958. Maximum Sustained Yields from Fluctuating Environments and Mixed Stocks. J. Fish. Res. Board Canada **15**(5): 991–1006. doi:10.1139/f58-054.

Stott, W., VanDeHey, J.A., and Sloss, B.L. 2010. Genetic diversity of lake whitefish in lakes Michigan and Huron; sampling, standardization, and research priorities. J. Great Lakes Res. **36**(SUPPL. 1): 59–65. doi:10.1016/j.jglr.2010.01.004.

Vandergoot, C.S., and Brenden, T.O. 2014. Spatially varying population demographics and fishery characteristics of Lake Erie walleyes inferred from a long-term tag recovery study. Trans. Am. Fish. Soc. **143**(1): 188–204. doi:10.1080/00028487.2013.837095.

Vincent, M.T., Brenden, T.O., and Bence, J.R. 2017. Simulation testing the robustness of a multiregion, tag-integrated assessment model that exhibits natal homing and estimates natural mortality and reporting rate. Can. J. Fish. Aquat. Sci. **74**(11): 1930–1949. doi:10.1139/cjfas-2016-0297.

Ying, Y., Chen, Y., Lin, L., and Gao, T. 2011. Risks of ignoring fish population spatial structure in fisheries management. Can. J. Fish. Aquat. Sci. **68**(12): 2101–2120. doi:10.1139/f2011-116.

Figure 1. The full closed-loop feedback simulation framework, which followed a management strategy evaluation approach.

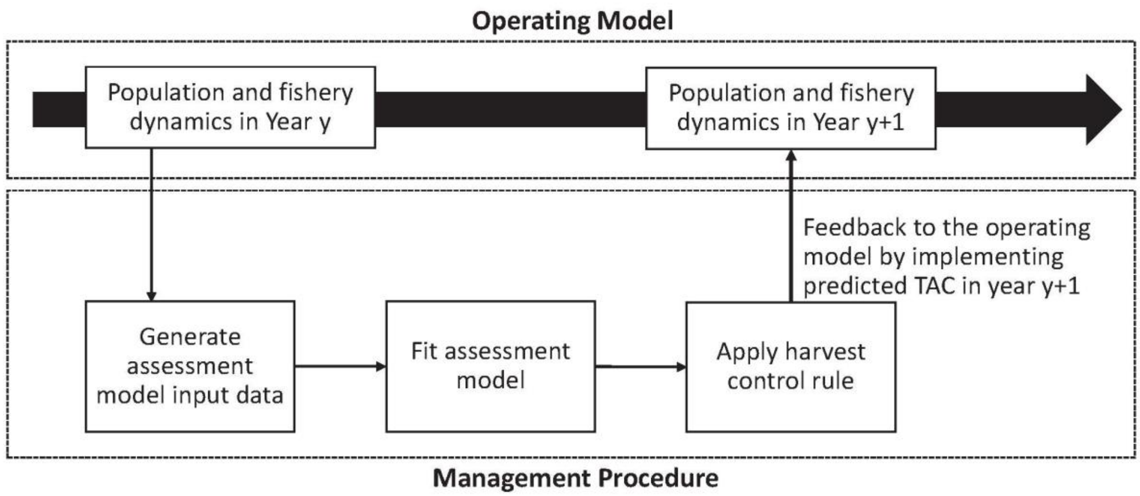

Figure 2. Ricker stock-recruitment relationships for populations with low, medium, and high level of productivity (Table 3). Two dashed lines represent the replacement lines for *F*=0 and target *F* and their intersections with stock-recruitment curves (dots) define equilibrium for low, baseline, and high productivity. Note that the target *F* is calculated based on the natural mortality rate and the status quo target total mortality (*A*=0.65).

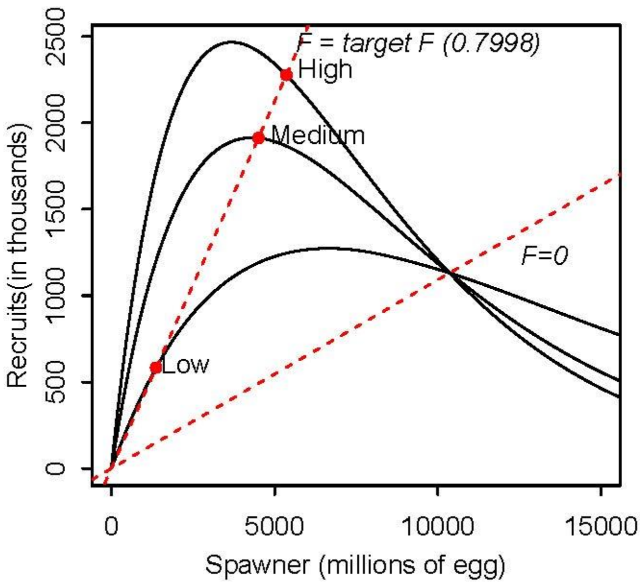

Figure 3. Simulation results (median $\pm$ interquartile range) for population 1 (Table 5) in the baseline scenario. Full model names are in Table 1. (a) Relative error of estimating terminal assessment year SSB during simulation year 91 to 100. (b) In simulation year 100, relative error of estimating recruitment of the last ten assessment years. (c) Proportion of years SSB was lower than 20% of the unfished SSB level ($B_{20\%}$) over the last 25 years of simulations. (d) Mean annual yield for the fishing area surrounding spawning grounds of Pop1 over the last 25 years of simulations. (e) Mean interannual variation (IAV) in yield over the last 25 years of simulation. (f) Estimated autocorrelation for terminal year estimates of SSB during simulation years 75 to 100.

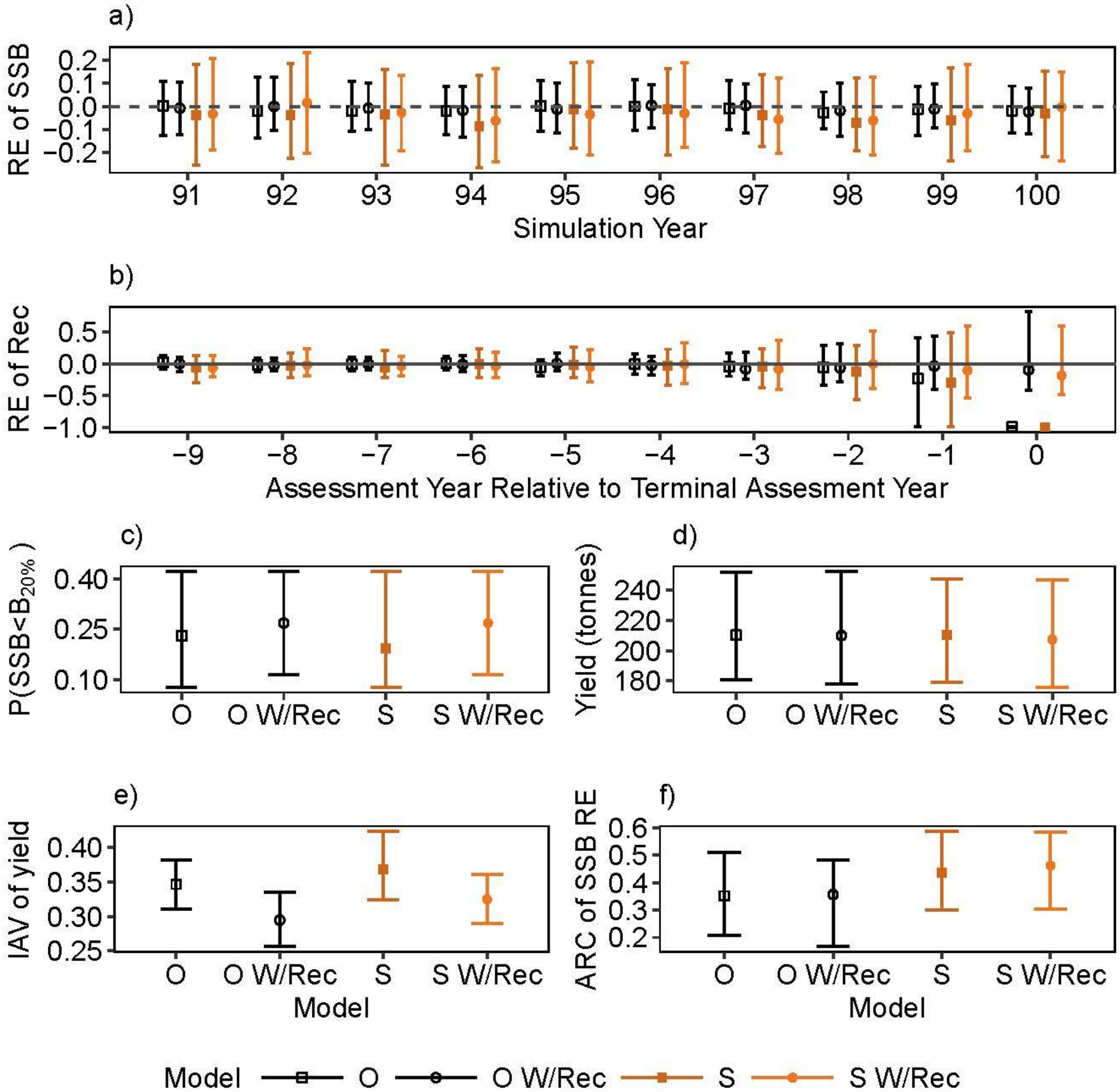
a)
RE of SSB
Simulation Year
b)
RE of Rec
Assessment Year Relative to Terminal Assesment Year
c)
P(SSB<B20%)
d)
Yield (tonnes)
e)
IAV of yield
Model
f)
ARC of SSB RE
Model
Model O O W/Rec S S W/Rec

Figure 4. Relative error in estimates of recruitment for the terminal assessment year during the simulation year 76 to 100 for an example simulation. Full model names are in Table 1.

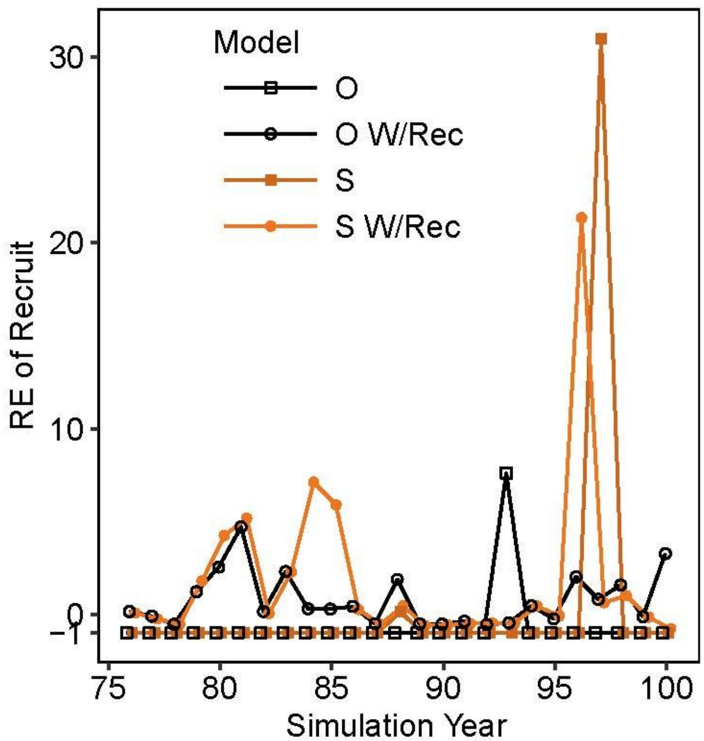

Figure 5. Simulation results (median $\pm$ interquartile range) for populations 1 and 3 under scenarios 2 to 5 (Table 5). Full model names are in Table 1. Each column represents a different productivity and movement scenario, and each row presents a different performance statistic. The x-axis of each column indicates the productivity levels (L, A, H are low, average, and high productivity levels) and stay rates associated with the two populations results are presented for. For example, L70% means low productivity population with 70% stay rate. For each such productivity level and stay rate, results are given for the four different assessment methods, distinguished by different symbols. The second, fourth, and sixth rows represents the same performance statistics as for Figure 3c, 3e, and 3d. The first and third row are relative error of estimating terminal year SSB and recruitment in simulation year 100, respectively, with a 0 dashed line. The fifth row

represents the average SSB over the last 25 years of simulation, and the dashed line is 20% of the unfished SSB.

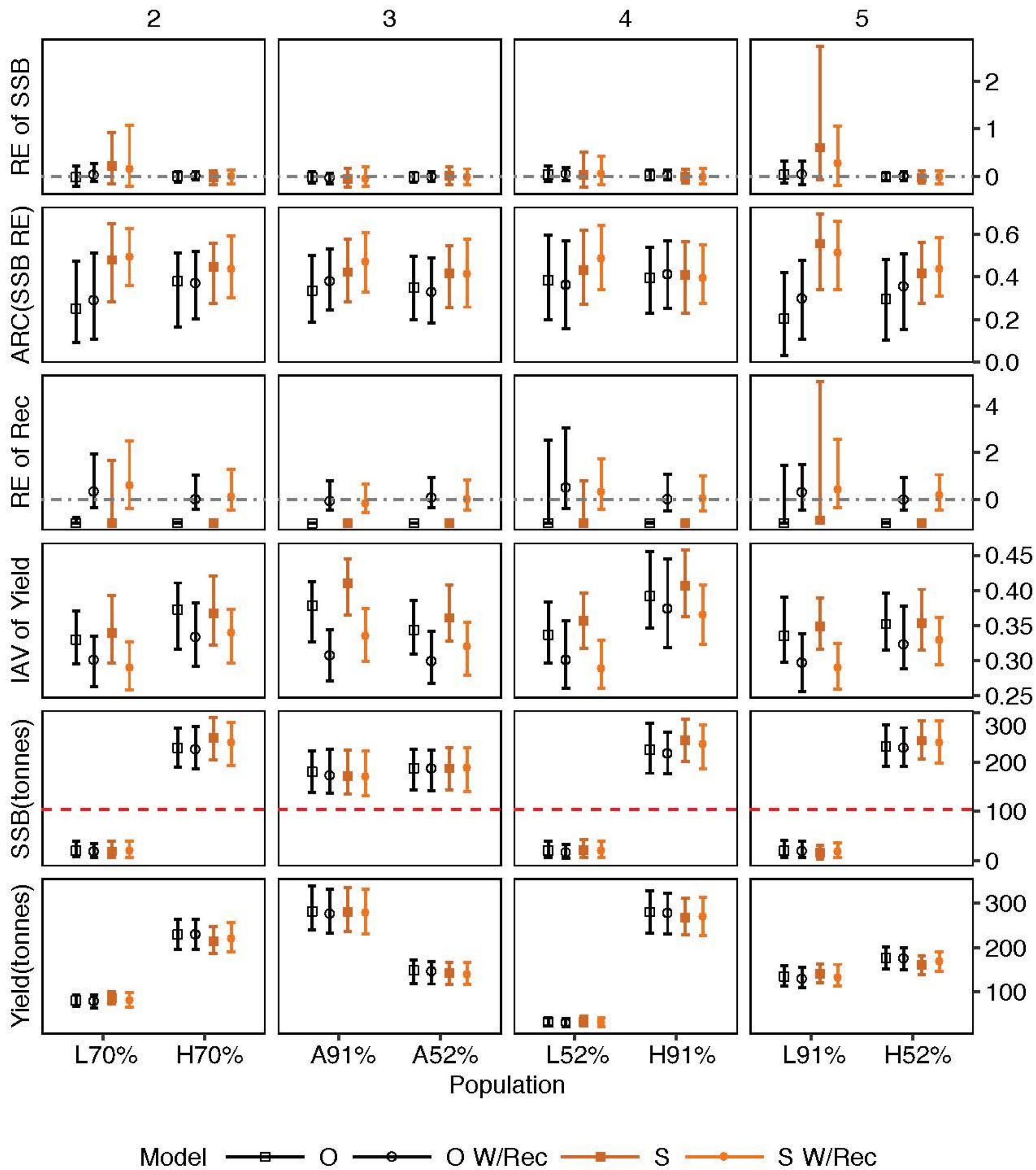

Figure 6. Simulation results (median $\pm$ interquartile range) for Pop1 (Table 5) in sensitivity analyses. Full model names are in Table 1. Each column represents a sensitivity scenario, each row represents a performance metric (as described in Figure 5), and results in each panel are for the four assessment models.

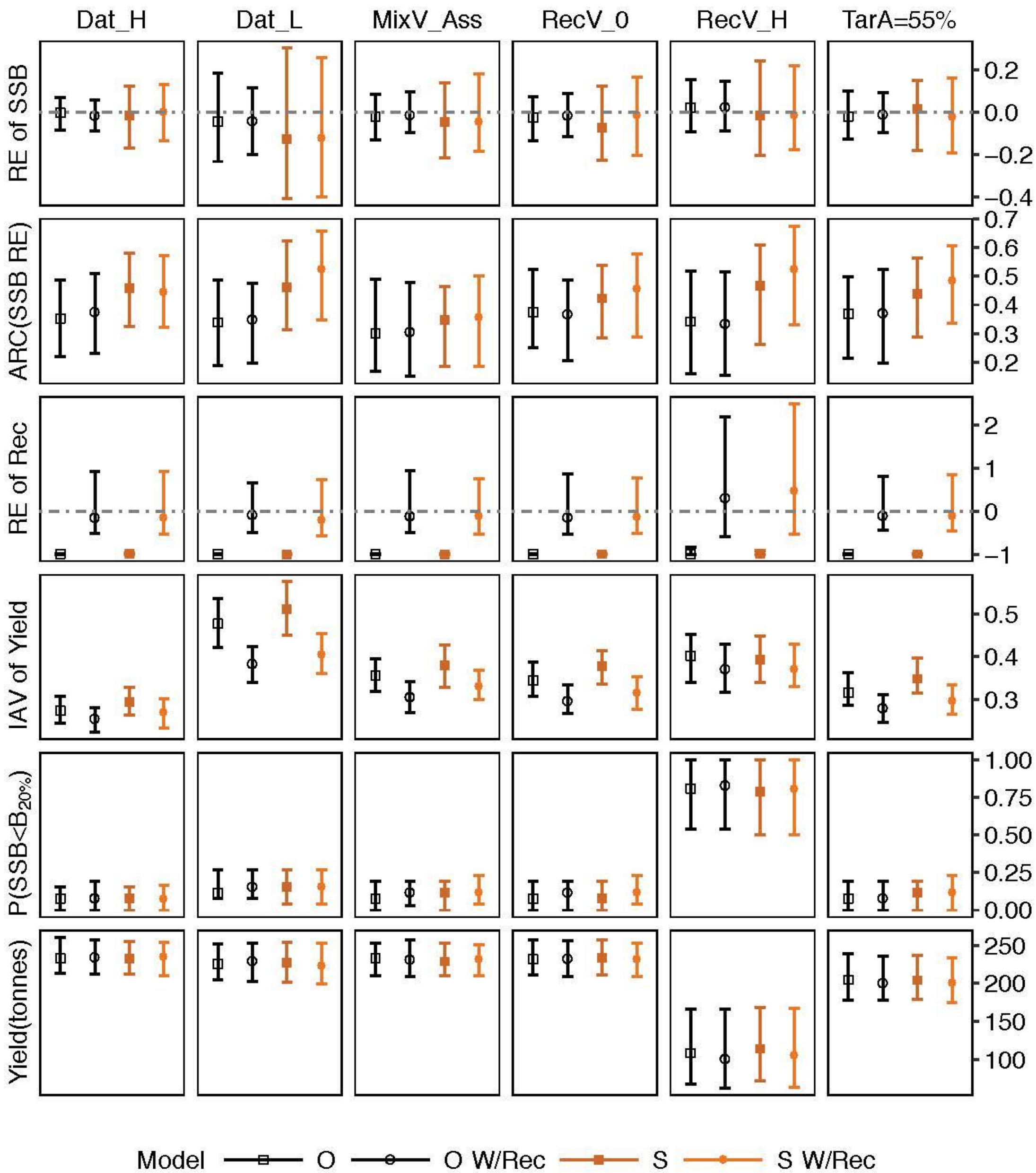

Table 1. Composition of the assessment input data and objective function for the four assessment models we evaluated.

Assessment model		Standard assessment model without a recruitment penalty (S)	Standard assessment model with a recruitment penalty (S W/Rec)	Origin-informed assessment without a recruitment penalty (O)	Origin-informed assessment with a recruitment penalty (O W/Rec)
Input data	Observed harvest	✓	✓	✓	✓
	Observed effort	✓	✓	✓	✓
	Aggregated harvest age composition	✓	✓		
	Population-specific harvest age composition			✓	✓
Objective function components (negative log likelihood or log-prior penalty for)	Area-specific fishery harvest	✓	✓	✓	✓
	Annual deviation from the general level of fishing mortality	✓	✓	✓	✓
	Aggregate harvest age composition	✓	✓		

	Population-specific harvest age composition		✓	✓
	Annual recruitment residuals	✓		✓

Table 2. Index variables and accents used in all equations.

Symbol	Definition
i	Population
j	Fishing ground
y	Year
a	Age
$\sim$	Observed variable
$\frown$	Estimated variable
$\prime$	Derived variable

Table 3. Coefficients for parameters used to generate different levels of productivity, data quality, recruitment variation, and annual-varying random generated rates in both operating and stock assessment models.

Coefficient name	Definition	Coefficient values		
Productivity levels		Low	Baseline	High
Steepness	S-R steepness	0.7	1.3668	1.9
α'	Ricker S-R parameter	0.0003169815	0.0007316319	0.001104342
β	Ricker S-R parameter	$1.511359e^{-10}$	$2.318631e^{-10}$	$2.716004e^{-10}$
Data quality levels		Low	Baseline	High
$effN$	Effective sample size	25	50	100
Harvest CV	CV for observed harvest about actual harvest	0.4	0.15	0.1
Effort CV	CV for observed harvest about actual effort	0.8	0.3	0.2
Annual-varying random generated rates		Stay rate=91%	Stay rate=70%	Stay rate=52%
μ_ω	Mean of ω_y	2.313635	0.8472979	0.08004271
${\sigma_\omega}^2$	Variance of ω_y	0.3364	0. 0625	0.21
Recruitment variation levels		No autocorrelation	Baseline	High
ρ	Autocorrelation coefficient	0	0.45	0.45
σ_R	Innovative standard dev. in rec process error	0.8734	0.78	1.3395
σ	Stationary standard dev. in rec process error	0.8734	0.8734	1.5
Target mortality levels		Low	Baseline (Status quo)	
A	Annual total mortality rate	0.55	0.65	

Table 4. Biomass calculation in the operating model.

Model name	Model equation	Equation number
Age-specific SSB	$SSB_{i,y} = \sum_a Fem W_a m_a N_{i,y,a} Fec$ where Fem=0.5 (from Li et al. 2015)	2.1
Length at age	$L_a = L_\infty(1 - \exp(-\kappa(a - t_0)))$ where L_∞=60.9 cm, κ=0.1689 year^{-1}, $t_0 = 0$ year (from Li et al. 2015)	2.2
Weight at age	$W_a = \gamma {L_a}^{\psi}$ where γ =8.06×10^{-5}, ψ= 2.45 (from Li et al. 2015)	2.3
Maturity at age	$m_a = \dfrac{m_\infty}{1 + \exp(-\vartheta(L_a - \delta))}$ where ϑ = 0.315 cm^{-1}, δ= 37.86 cm (from Li et al. 2015)	2.4

Table 5. Simulation scenarios, including the baseline scenario and other combinations of productivity levels and stay rates, for four hypothetic populations used in the simulations.

Scenario index	Scenario	Population identifier	Productivity	Stay rate
Baseline (1)	Equal mixing with baseline productivity	Pop1	Baseline	70%
		Pop2	Baseline	70%
		Pop3	Baseline	70%
		Pop4	Baseline	70%
2	Equal mixing with different productivity	Pop1	Low	70%
		Pop2	Low	70%
		Pop3	High	70%
		Pop4	High	70%
3	Unequal mixing with baseline productivity	Pop1	Baseline	91%
		Pop2	Baseline	91%
		Pop3	Baseline	52%
		Pop4	Baseline	52%
4	Unequal mixing with different productivity (Positive correlation between productivity and stay rates)	Pop1	Low	52%
		Pop2	Low	52%
		Pop3	High	91%
		Pop4	High	91%
5	Unequal mixing with different productivity (Negative correlation between productivity and stay rates)	Pop1	Low	91%
		Pop2	Low	91%
		Pop3	High	52%
		Pop4	High	52%

Table 6. Scenarios for sensitivity analyses. In each sensitivity scenario, except for the change descripted below all other parameters are at their baseline levels.

Scenario index	Description	Description of change from baseline scenario
Dat_L	Data quality levels (Table 3) all low.	Data quality
Dat_H	Data quality levels (Table 3) all high.	Data quality
MixV_Ass	Allowed mixing rates in the assessment model to vary annually about the true value assumed in the operating model.	Mixing rates in the assessment model
RecV_H	Recruitment variation levels (Table 3) all high.	Recruitment variation
RecV_0	Recruitment variation levels (Table 3) all no autocorrelation.	Recruitment variation
TarA=55%	Target mortality levels all low (Table 3).	Target mortality