An evaluation of alternative assessment approaches for intermixing fish populations: a case study with Great Lakes lake whitefish

Yang Li*, James R. Bence, and Travis O. Brenden
Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA
*Corresponding author: tel: +1 5173550126; e-mail: liyang11@msu.edu

We used simulation modeling to explore how three statistical catch-at-age approaches for assessing intermixed fisheries performed in terms of assessment accuracy and management performance, under differing productivity, mixing, and harvest levels. Simulations were based on intermixing lake whitefish (Coregonus clupeaformis) populations in the upper Laurentian Great Lakes of North America. We found that with intermixing, the “separate” assessment approach, which ignored intermixing and treated mixed populations as unit stocks, produced biased estimates of spawning stock biomass (SSB); however, the “pooled” assessment approach, which lumped populations and assessed them as a single stock, was nearly unbiased in estimating SSB. The “overlap” assessment approach, which estimated the populations in one combined assessment model by incorporating actual mixing rates, was most strongly biased in estimating SSB in the absence of mixing, with bias decreasing as mixing levels increased. With high mixing levels, the overlap method had difficulty converging on unique solutions. The pooled approach provided better management performance than the separate approach with intermixing. When the overlap method could be applied, it provided the greatest SSB with little reductions in yield and the lowest inter-annual variation in yield. Relative performances of the assessment approaches were robust to assumed harvest levels.

Keywords: intermixing, management strategy evaluation, statistical catch-at-age, stochastic simulation, stock assessment.

Introduction

Stock-based assessment of exploited fish populations has long been used to estimate demographic rates (e.g. age-specific mortality) and population abundances, which in turn have been used to establish appropriate levels of harvest. Typically, assessments are conducted on spatial management units that have been established for a stock believed to be associated with a single, rather than multiple, spawning populations. When fisheries exploit mixtures of fish from multiple spawning populations, stock-based assessment can lead to poor estimation of abundance and unintended depletion of local subunits (Hutchings, 1996, 2000; Stephenson, 1999; Frank and Brickman, 2000; Molton et al., 2012, 2013). A common response to population intermixing has been to modify the spatial scale of fishery assessments, such as pooling harvest among areas and assessing as a single stock (Powers and Porch, 2004; Kell et al., 2009; Cope and Punt, 2011; Ying et al., 2011; Guan et al., 2013; Hart et al., 2013). Alternatively, more complex approaches that allow for movement among areas can be used, although successful application of such approaches may necessitate additional population-specific information.

Exploitation of intermixed fisheries is increasingly being recognized as a common attribute of marine and freshwater systems, and that ignoring such mixing in assessments can lead to inappropriate management advice (Hutchings, 1996; Stephenson, 1999; Booth, 2000; Fu and Fanning, 2004; Rothschild, 2007; Ying et al., 2011). The collapse of Atlantic cod (Gadus morhua) populations is believed to have partly resulted from population intermixing and failures to account for this in assessment and management (Hutchings, 1996; Fu and Fanning, 2004; Rothschild, 2007; Hutchinson, 2008). Many Pacific salmon populations are also believed to have been overexploited as a result of ocean, mixture fisheries (Morishima and Henry, 1999). Booth (2000) theorized that incorporating spatial structure into an assessment model for sparrow Pterogumnus laniarius would improve management policies for this species off the coast of South Africa.

The purpose of our research was to compare fishery management and estimation performances among three different statistical catch-at-age (SCAA) approaches for assessing intermixed fisheries and to evaluate how performance was affected by mixing and harvest levels. The type of mixing addressed in our research was...
referred to by Porch et al. (2001) as “population overlap”, whereby spawning populations have complete spawning fidelity but mix during the fishing season. We refer to the three SCAA approaches we evaluated as “separate”, “pooled”, and “overlap”. The “separate” approach assumed all stocks were assessed separately and independently, despite the occurrence of mixing. The “pooled” approach lumped all regions and fisheries into a single unit stock for assessment purposes. The “overlap” approach involved fitting a combined assessment model to the stocks by incorporating actual population mixing rates in the model. Although some studies have compared alternative assessment approaches for intermixed fisheries, they were either based on alternative assumptions regarding mixing (Ying et al., 2011), assumed no post-recruitment movements (Cope and Punt, 2011), or did not include an approach based on correctly modelling movement (Kell et al., 2009; Guan et al., 2013).

We used a management strategy evaluation (MSE) framework (Kirkwood, 1992, 1996; Kell et al., 2005a; Punt, 2008) to evaluate the performance of the different assessment approaches. Most MSE applications compare fishery control rules (e.g. constant fishing rate or constant escapement) or different parameters (e.g. harvest level) for the rules, but keep the assessment approach constant (Kell et al., 2005b). This study was somewhat unique in that we used MSE methods to compare different assessment approaches.

Great Lakes lake whitefish as a test species

Although our goal was to provide broadly relevant results as to the performance of SCAA assessment approaches, we nevertheless believed it was important to base our simulations on fish stocks that do intermix to provide a basis of realism for our evaluations. We based our simulations on lake whitefish (Coregonus clupeaformis) populations in the upper Laurentian Great Lakes of North America. Lake whitefish are ecologically and economically important in the Great Lakes and have historically been managed as nominally distinct stocks based on an assumption that populations do not mix (Ebener et al., 2005; Caroffino and Lenart, 2012). Recent evidence, however, has indicated that many Great Lakes lake whitefish populations intermix during non-spawning periods (Ebener et al., 2010). Despite this intermixing, strong genetic structuring of spawning populations has been maintained (Stott et al., 2010, 2012), likely as a consequence of lake whitefish exhibiting strong spawning site fidelity (Ebener et al., 2010).

Despite evidence indicating widespread mixing, lake whitefish stocks in the Great Lakes continue to be managed largely as discrete independent units. The SCAA assessment models used to estimate year- and age-specific abundances and mortality rates generally do not allow for the possibility of mixing between assessment areas. In northern Lake Huron, however, stock assessment scientists have responded to population intermixing by combining four formerly distinct assessment areas into one larger area and estimating a pooled total allowable catch (TAC) for lake whitefish (Caroffino and Lenart, 2011). Despite this adjustment in assessment spatial scale, an assumption is still made that fish within the aggregated area do not move to other areas and that within the pooled area fish are homogeneously mixed.

Deroba and Bence (2012) and Molton et al. (2012, 2013) previously have used simulations to evaluate the status-quo constant mortality rate policy (65% total annual mortality) for lake whitefish in the Great Lakes. Deroba and Bence (2012) concluded that the status-quo mortality rate was reasonable, but they assumed that there was no intermixing among the assessed populations. Molton et al. (2012, 2013) concluded that the 65% mortality rate could put some lower productivity stocks at risk when populations mixed, but their analysis assumed that management units would continue to be assessed separately and without accounting for mixing. Our work builds upon these previous studies by exploring both the influence of mixing and how that mixing is accounted for in stock assessment models.

Methods

In our MSE simulations, we modelled both “true” and “perceived” systems (Figure 1). The “true” system represented actual stock and fishery dynamics from which observations were collected annually. The most recent 20 years of the available time-series of data were incorporated in the assessment models that estimated stock status (i.e. the perceived system) and provided a basis for establishing target yields based on constant fishing mortality control rules and differing harvest levels. The target yield (i.e. TAC) was fed back into the “true” system, thereby influencing actual yield and fishery performance.

Simulations were conducted in AD Model Builder (Fournier et al., 2012). All equations for the simulation framework (except two core equations listed below) are given in Supplementary Table S1. Assumed life history values and stochastic modelling parameters are listed in Table 1.

The “true” system

The model for our “true” system (Figure 1) was based on that of Molton et al. (2012). We used an age-structured, forward-projection model, consisting of four hypothetical populations with a simulation length of 100 years. Only the last 25 years were summarized to evaluate performance, as the intent was to evaluate long-term performance of the approaches, independent of starting conditions. Recruits were added annually to the youngest age class based on spawning biomass and an assumed stock–recruit model (see below). During the fishing season, fish from each population mixed with fish from other areas (Figure 2), but returned to spawning areas for instantaneous spawning.

Although our use of instantaneous spawning and mixing and absolute spawning-site fidelity are simplifying assumptions, we believe they are reasonable for Great Lakes lake whitefish populations based on taggging results. Ebener et al. (2010) found that of 31 tagged lake whitefish that were recovered during the spawning season, 30 were recovered on their original spawning grounds. Ebener et al. (2010) additionally found that tagged lake whitefish routinely dispersed more than 60 km within 1 month of tagging.

Mortality was made up of both natural and fishing mortality components. Fishing mortality was determined in the perceived system based on implemented TACs and abundances on the fishing grounds during the fishing season (see more details in the perceived system). Thus, fishing mortality varied both spatially and temporally. Space- and time-invariant natural mortalities were assumed for all populations.

The four modelled spawning populations had different productivity levels as a consequence of different recruitment function coefficients (Table 2). We used a stochastic Ricker stock–recruit model because of strong evidence for over-compensation in lake whitefish populations (Healey, 1978; Henderson et al., 1983; Kratzer, 2006). Stochastic errors were multiplicative and lognormal. Given information on life history (e.g. maturity, weight-at-age), Ricker stock–recruit models can be parameterized in terms of steepness (Mangel et al., 2010) and unfished spawning stock size. We chose
Alternative assessment approaches for intermixing fish populations

Figure 1. Conceptual overview of the simulation framework and the “flow” of data and estimates between the true and perceived systems. During an individual 100-year simulation, observed data were collected annually and the assessment model was refit based on the most recent 20 years in the time-series of data. Based on the assessment results and the target mortality rate, a target amount of harvest (TAC) was set and actual harvest was then determined (accounting for implementation error). Because the actual amount of harvest was based on the assessment method and influenced future population dynamics, the different assessment methods had to be applied in separate simulations. One thousand simulations were done for each scenario (defined by mixing rate, target mortality, and assessment method).

Table 1. Parameter values assumed as part of the MSE simulation model for evaluating approaches for assessing intermixed fish populations.

<table>
<thead>
<tr>
<th>Parameters/inputs</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life history parameters</td>
<td></td>
</tr>
<tr>
<td>Maximum age</td>
<td>12 year</td>
</tr>
<tr>
<td>Instantaneous nature mortality ($M$)</td>
<td>0.25 year$^{-1}$</td>
</tr>
<tr>
<td>Growth parameters</td>
<td></td>
</tr>
<tr>
<td>$L_{\infty}$</td>
<td>60.9 cm</td>
</tr>
<tr>
<td>$K$</td>
<td>0.1689 year$^{-1}$</td>
</tr>
<tr>
<td>$t_0$</td>
<td>0 year</td>
</tr>
<tr>
<td>Weight-at-age parameters</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$8.06 \times 10^{-5}$ kg cm$^{-\phi}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>2.45</td>
</tr>
<tr>
<td>Maturity at length parameters</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.315 cm$^{-1}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>37.86 cm</td>
</tr>
<tr>
<td>Selectivity-at-age parameters</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.26 year$^{-1}$</td>
</tr>
<tr>
<td>$\eta$</td>
<td>13.074 cm</td>
</tr>
<tr>
<td>Fraction of females in the populations</td>
<td>0.5</td>
</tr>
<tr>
<td>Stochastic modelling parameters [coefficient of variation (CV)]</td>
<td></td>
</tr>
<tr>
<td>CV of process error in S–R function</td>
<td>0.66</td>
</tr>
<tr>
<td>CV of implementation error in total allowable catch</td>
<td>0.1</td>
</tr>
<tr>
<td>CV of observation error in generating observed catch</td>
<td>0.1</td>
</tr>
<tr>
<td>CV of observation error in generating observed effort</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Four steepness values for our simulated populations that were plausible for lake whitefish and produced meaningful differences in equilibrium yield. Unfished equilibrium stock size for all spawning populations was assumed to be equal ($S_0 = 519.70$ t), so that differences between populations were due only to steepnesses. Thus, the range of steepnesses we used translated to differences in productivity among the populations. The steepness and equilibrium spawning stock parameterization were then converted to the standard Ricker parameterization (Table 2; Cerviño et al., 2013). All equilibrium calculations assumed unit stocks (no movement) and followed methods outlined by Quinn and Deriso (1999). (Note: the assumption of unit stocks had no affect under unfished conditions; for more details on equilibrium calculations, see Supplementary Table S1.)
other areas during the non-spawning season, were calculated as:

\[ \text{Movement rates} = \text{a greater proportion of fish moved to that area.} \]

We assumed that a greater SR indicated a higher quality habit, and thus a greater proportion of fish moved to that area. Movement rates were determined via the stock assessment procedure described in the perceived system section (see below). True catch was set equal to the TAC multiplied by a lognormal implementation error. The fully selected fishing mortality rate necessary to produce the true catch in each area was solved using a Newton–Raphson algorithm applied to the Baranov catch equation. Age-specific fishing mortalities [i.e. the \( F_{j,a} \) in Equation (2)] were the product of the fully selected fishing mortality rate multiplied by age-specific selectivities, which were assumed equal for all four stocks.

Constant values for natural mortality, fraction of the populations that were female, and catchability were assumed. Selectivity-, length-, weight-at-age, and annual spawning-stock biomass (SSB) were determined by deterministic functions. Selectivity was as assumed to be 0 for age 2 and younger lake whitefish because these ages of lake whitefish are rarely harvested in the Great Lakes. Selectivity for other age classes was simulated by a \( \gamma \) density function scaled, so that age 10 fish had a selectivity of 1.0. A \( \gamma \) function was used to emulate the dome-shaped selectivity pattern observed for actual Great Lakes lake whitefish fisheries. Length-at-age was assumed to follow a von Bertalanffy growth model, weight-at-age assuming fraction of the population that was female. All four populations shared the same selectivity-, length-, and weight-at-age patterns. The assumption that these values were the same among the populations is clearly a simplifying assumption. We believe that this is reasonable because these life history characteristics are generally well estimated from biological data, and at least for Great Lakes lake whitefish, are not tightly linked to stock–recruit relationships (Deroba and Bence, 2012).

The initial (i.e. year 1) state for each simulation was based on equilibrium recruitment (at target mortality rates) and target mortality rates, and calculated based on deterministic models with no mixing. The actual harvest policy and management process required a 20-year time-series of data, which was not available during the initial 20-year period of each simulation. During this “burn-in” period, population abundances at age were assumed to be known exactly before applying the harvest control rule and thus the assessment model step could be skipped. The target mortality rate was applied during the burn-in period. The intent here was merely to move the system closer to steady-state conditions during this fraction of those fish residing in each destination area \( j \), and \( k \) all areas except the fishing grounds surrounding the spawning area of population \( i \). This representation of movement differs from that of Molton et al. (2012), who assumed that fish that moved were equally likely to each of the other areas. Although we had no evidence supporting our alternative mixing assumption, we believed that it was a reasonable assumption that factors influencing fish to stay in an area would result in other fish moving to that area.

Post-recruitment abundances at age were projected using an exponential population model:

\[ N_{j,y+1,a+1} = N_{j,y,a} \sum_j \theta_j \exp \left(-\left(M + F_{j,y,a}\right) \times 1 \text{ year} \right). \]  

where \( N_{j,y,a} \) is the abundance of fish from spawning population \( i \) of age \( a \) in year \( y \), \( M \) the instantaneous natural mortality rate, \( F_{j,y,a} \) the year- and age-specific instantaneous fishing mortality rate of fish occurring in area \( j \) during harvest season, and \( \sum_j \theta_j \exp \left(-\left(M + F_{j,y,a}\right) \times 1 \text{ year} \right) \) is a weighted average of survival for population \( i \) with the weights being the proportion of population \( i \) moving to area \( j \).

The initial (i.e. year 1) state for each simulation was based on equilibrium recruitment (at target mortality rates) and target mortality rates, and calculated based on deterministic models with no mixing. The actual harvest policy and management process required a 20-year time-series of data, which was not available during the initial 20-year period of each simulation. During this “burn-in” period, population abundances at age were assumed to be known exactly before applying the harvest control rule and thus the assessment model step could be skipped. The target mortality rate was applied during the burn-in period. The intent here was merely to move the system closer to steady-state conditions during this

![Figure 2. Illustration of stock intermixing according to the overlap model used in our simulations of the true populations. For clarity, here only arrows representing the movements for the LP population are included. Each population was spatially segregated during spawning, but subsequently intermixed during the non-spawning period when exploitation took place. The solid arrow represents the movement of fish each year to the area occupied during operation of the fishery, while all dashed arrows (including the bolder dashed arrow) represented the movement of all fish back to the spawning grounds during the spawning season, before being reallocated to fishing grounds for the next fishing year. The bolder arrows represent fish that stayed within the fishing ground surrounding their spawning area. For the case of non-mixing case, only movements associated with bold arrows occurred. In this case, for example, the LP population can only stay in the area (Non-spa

Table 2. Recruitment steepnesses for the four different productivity levels assigned to fish populations and the corresponding parameters \( \alpha \) and \( \beta \) for the standard ricker stock–recruit model formulation.

<table>
<thead>
<tr>
<th>Population index</th>
<th>Productivity level</th>
<th>Steepness ( \alpha )</th>
<th>Steepness ( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP population</td>
<td>Low 0.7</td>
<td>5.23 × 10^-4</td>
<td>1.51 × 10^-10</td>
</tr>
<tr>
<td>MLP population</td>
<td>Mid–low 1.1</td>
<td>9.19 × 10^-4</td>
<td>2.06 × 10^-10</td>
</tr>
<tr>
<td>MHP population</td>
<td>Mid–high 1.5</td>
<td>1.35 × 10^-3</td>
<td>2.43 × 10^-10</td>
</tr>
<tr>
<td>HP population</td>
<td>High 1.9</td>
<td>1.82 × 10^-3</td>
<td>2.72 × 10^-10</td>
</tr>
</tbody>
</table>

were set based on results from tagging studies (Ebener et al., 2010). We assumed that a greater SR indicated a higher quality habit, and thus a greater proportion of fish moved to that area. Movement rates were calculated as:

\[ \theta_j = (1 - \theta_j) \frac{\theta_j}{\sum_{k \neq j} \theta_k}, \]  

where \( 1 - \theta_j \) was the fraction of fish from population \( i \) that moved to other areas during the non-spawning season, \( \theta_j/\sum_{k \neq j} \theta_k \) the weight of the population in year- and age-specific instantaneous fishing mortality rate of fish occurring in area \( j \) during harvest season, and \( \sum_j \theta_j \exp \left(-\left(M + F_{j,y,a}\right) \times 1 \text{ year} \right) \) is a weighted average of survival for population \( i \) with the weights being the proportion of population \( i \) moving to area \( j \).
period. Stochastic recruitment errors were still included during the burn-in period.

The "perceived" system

In the perceived system (Figure 1), the management procedure for lake whitefish in the 1836 Treaty-ceded waters was duplicated, including data collection, stock assessment, and application of the harvest control rule. When generating observed data, we assumed that total true catch and age composition of the catch were affected by observation error. Observed age compositions were generated from multinomial distributions with probabilities equal to actual age compositions of the catch and an effective sample size of 200 for separate and overlap approaches and 800 for the pooled method, which are within the range of age composition sample sizes collected for lake whitefish in 1836 Treaty-ceded waters (ftp://glpd.fw.msu.edu/MSCFTP/Assessment_models/). Observed fishing effort was assumed to be proportional to the fully selected fishing mortality multiplied by a lognormal error with a dispersion parameter equal to the true value used in generating observed effort data (Table 1). The above data were collected annually and used as inputs for stock assessments. Stock assessments were assumed to have knowledge of the true natural mortality rate and other life history parameters listed, except for selectivity function parameters, which were estimated.

As previously indicated, three different assessment approaches were evaluated. The separate approach assumed no movement among areas and was applied to each fishing ground separately. Because the pooled approach lumped all four fishing grounds into a single unit stock, this approach resulted in combined predictions of harvest at age and abundance. The overlap approach used a single, combined assessment model for all four stocks, but correctly accounted for movements (i.e. stay and movement rates were model inputs) among the regions. Thus, the four populations were tracked according to Equation (2) and predictions accounted for movements (i.e. stay and movement rates were model inputs) among areas and was applied to each fishing ground separately. Therefore, movement among areas was included in the overlap approach.

For all three assessment approaches, recruitment, initial abundance-at-age in the first assessment year, selectivity at age, catchability, and fully selected fishing mortality rate were estimated (either area specific or pooled). Age of recruitment in the assessment models was defined as age 3 due to age 2 and younger lake whitefish not being subject to harvest. Although the ways we defined recruitment were nominally different in the "true" and the "perceived" systems, they were effectively the same because the natural mortality between age 0 and age 3 was constant and could have been incorporated into the productivity parameter "α" of the Ricker stock-recruitment model.

Best fit parameters were estimated via iterative methods so as to minimize the negative penalized log-likelihood (i.e. the objective function). The objective function consisted of three components. Two of these represent fit to observed data. The first was based on an assumed lognormal distribution of errors for annual fishery catch, and the second was based on a multinomial distribution for the harvest age compositions. The objective function also included a penalty for deviations from direct proportionality between fishing mortality and observed fishing effort, with those deviations assumed to have a lognormal distribution (Fournier and Archibald, 1982). The negative loglikelihood component for the age composition of the fishery harvest was weighted by effective sample size (800 for the pooled approach and 200 for the separate and the overlap assessment approaches). For the pooled and overlap assessment approaches, the objective function consisted of just the summation of these three components. For the separate assessment approach, the objective function was the summation of the three components for each of the four modelled areas. Models were considered to have converged on a solution when the maximum gradient of the parameters was <0.001 and the Hessian matrix was positive definite. Occasionally, the estimation algorithm did not converge because no further progress could be made in reducing the objective function or the maximum number of iterations (i.e. 10 000) was reached. For the overlap assessment approach, it was necessary to increase the maximum number of iterations to 50 000 because convergence, when it was achieved, oftentimes took more steps than for the other assessment approaches. Model runs that ended without the assessments reaching the gradient criterion were classified as run gradient failures. If the Hessian matrix of the penalized likelihood was not positive definite, we termed this a Hessian problem. The proportion of gradient failures and Hessian problems that occurred in all simulations were recorded for each assessment approach. We defined the convergence rate as the fraction of simulations which encountered neither gradient failures nor Hessian problems.

We mimicked the timing of assessments and harvest policies of lake whitefish fisheries in 1836 Treaty-ceded waters. In this area, annual assessments are conducted, based on data collected through the previous year, and are used to establish harvest targets for the following year. Thus, the stock assessment estimates abundance-at-age at the start of the year when the assessment is conducted. Projections through the assessment year were based on the exponential population model, in which fishing mortality rates were assumed to be the mean of the last 3 years’ values, and recruitment (age 3 lake whitefish) were assumed to be the mean of the last 10 years’ values. For the year in which the TAC was set, the same level of recruitment was assumed, but fishing mortality was adjusted, so the target mortality rate was achieved (assuming the assessment estimates were correct). The current harvest control rule for lake whitefish in 1836 Treaty-ceded waters, which equates to a constant fishing mortality rate policy, was used in the simulation. Such constant total mortality rules are widely used in the Great Lakes because some populations (particularly of lake trout (Salvelinus namaycush)) experience substantial temporal changes in natural mortality due to sea lamprey (Petromyzon marinus) attacks. The control rule sets the maximum total annual mortality percentage experienced by any age of fish to a specified value. The percentage used as of 2013 was 65%. The TAC was calculated from assessment-based estimates of abundance-at-age at the start of the year the TAC applies to and target fishing mortality-at-age using the Baranov catch equation. The age-specific fishing mortality rates were based on the estimated selectivity-at-age from the assessment multiplied by the target fishing mortality rates, which was the target instantaneous total mortality rate minus the natural mortality rate as assumed in the assessment model. For the separate and overlap assessment methods, a TAC for each area was calculated separately because assessment results were area specific for those methods. For the pooled population assessment, only a pooled TAC could be calculated and this was allocated to the areas according to an annually varying allocation rule. For this allocation rule, TAC allotments were proportional to the annual area-specific catch per unit effort (cpue) observed from the "true" system. To reduce annual variation in yield, we used the average cpue over the last 3 years of available data (preliminary simulations suggested this averaging outperformed alternative averaging approaches).
Experimental design

We evaluated six intermixing scenarios (Table 3). Scenarios 1–4 corresponded to cases where all populations shared the same SR: SR = 100% (scenario 1), SR = 75% (scenario 2), SR = 50% (scenario 3), and SR = 25% (scenario 4). Scenario 1 assumed no mixing, which represented a null model for evaluating the effect of intermixing. Scenario 5, which we refer to as “Po-Cor”, assumed that productivity and SR for each population were positively correlated (fish from higher productivity populations tended to stay in natal areas). Scenario 6, which we refer to as “Ne-Cor”, assumed a negative correlation between productivity and SR for each population. In the preliminary study, we tested two other scenarios: SR = 90% for each population and an unpredictable scenario which assumed no consistent relationship between productivity and SR of population. Those results were not included in later model comparison because the SR = 90% had very similar results compared with SR = 75%, and the unpredictable scenario showed an intermediate effect to that of scenarios Po-Cor and Ne-Cor. Each of these intermixing scenarios was applied using four annual total mortality harvest levels: (i) the current rate used in 1836 Treaty-ceded management of lake whitefish (65%), as well as (ii) 55%, (iii) 45%, and (iv) 35% harvest levels.

Performance statistics

For each assessment method and investigated scenario, 1000 simulations were performed with the following performance statistics calculated: average SSB for each population, the proportion of years SSB was <20% of the unfished SSB level ($B_{20\%} = 103.94 \text{ t}$), average annual total yield, interannual variation (IAV) in the total yield, IAV in yield by area, and the median relative error (MRE) of SSB estimates for the assessment methods. $B_{20\%}$ was used as a benchmark because (i) this is a widely used value, though somewhat arbitrary (e.g., Ianelli et al., 2011; Mace et al., 2013) and (ii) because based on our discussions with the MSC, this benchmark can be used to define reasonable performance measures for lake whitefish. The IAV in yield was calculated as:

$$\text{IAV} = \frac{\text{mean} | y_i - y_{i-1} |}{\bar{y}},$$

where $\bar{y}$ is the mean annual yield, $y_i$ the yield for a given year, and $y_{i-1}$ the yield in the previous year (Buijse et al., 1991). The statistics allowed us to compare the performance of alternative assessment methods under different mixing scenarios with respect to management and assessment objectives. Given that most performance statistics approached a long-term mean by year 75, performance statistics were calculated using the last 25 years of the simulations. Additional performance statistics (average annual yield by area, and median absolute relative error of estimating SSB over the last 25 years) are available for inspection in the Supplementary data. All performance statistics except MRE of SSB estimates were generated to describe the real stock and fishery (e.g., true SSB and yield in the real system).

Results

Convergence of assessment methods

In general, assessment models converged on solutions. For the pooled and separate methods, convergence was achieved in >96% of assessments across all scenarios, and thus we do not present or discuss their convergence results further. However, the overlap assessment method had more frequent convergence issues (Table 4). In particular, when the SR was 25%, the overlap assessment method never resulted in a positive definite Hessian matrix, suggesting that a unique solution could not be found. The same issue was seen in nearly 25% of the assessments when the SR and productivity were negatively correlated. As a consequence, we excluded the overlap method when comparing assessment methods for these cases (scenarios 4 and 6).

Comparison of assessment methods at the current 65% target mortality rate

Management performance

Using $B_{20\%}$ as a benchmark, the 65% mortality target was found to be overly aggressive for the LP population for most scenarios (Figures 3 and 4). Except for the overlap method in the Po-Cor Scenario, the inter-quartile range of average true SSB for the LP population overlapped $B_{20\%}$ for all other scenarios and all three assessment methods, even in the absence of mixing.

In the absence of mixing, all three assessment methods resulted in similar performance in terms of true SSB, with the largest differences in SSB among the methods occurring for the HP population, which tended to have slightly greater values for the overlap assessment method (Figure 3). While there were some differences between assessment methods when mixing occurred, observed differences were modest relative to differences among populations.

<table>
<thead>
<tr>
<th>Scenario index</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SR = 100%</td>
<td>All populations have the same high stay rate (100%)</td>
</tr>
<tr>
<td>2</td>
<td>SR = 75%</td>
<td>All populations have the same medium high stay rate (75%)</td>
</tr>
<tr>
<td>3</td>
<td>SR = 50%</td>
<td>All populations have the same medium low stay rate (50%)</td>
</tr>
<tr>
<td>4</td>
<td>SR = 25%</td>
<td>All populations have the same low stay rate (25%)</td>
</tr>
<tr>
<td>5</td>
<td>Po-Cor</td>
<td>Positive-correlated stay rates and productivity. Four stay rates (0.25, 0.5, 0.75, 0.9) matched in rank order to productivity of four populations</td>
</tr>
<tr>
<td>6</td>
<td>Ne-Cor</td>
<td>Negative-correlated stay rates and productivity. Four stay rates (0.25, 0.5, 0.75, 0.9) matched in reverse rank order to productivity of four populations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SR = 100%</th>
<th>SR = 75%</th>
<th>SR = 50%</th>
<th>SR = 25%</th>
<th>Po-Cor</th>
<th>Ne-Cor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hessian warning (%)</td>
<td>0.05</td>
<td>0.29</td>
<td>5.53</td>
<td>10.00</td>
<td>0.90</td>
<td>24.69</td>
</tr>
</tbody>
</table>

Table 3. Simulation scenarios evaluated in this research. In all scenarios, each of the four populations had different productivity, with one population taking each of the four alternative productivity levels (Table 2).

Table 4. The frequency of the lack of convergence (Hessian warning percentage) for the “overlap” assessment method under each scenario when target mortality rate was 0.65.
differing in productivity. With mixing, the pooled and overlap methods tended to result in higher SSB for each population than the separate method. For the SR = 50% and Po-Cor scenarios, the overlap method provided the highest SSB for each population, whereas for other scenarios, the pooled method resulted in the highest SSB. Overall, the pooled and separate methods provided stable results in terms of true SSB among different mixing scenarios, while SSB resulting from the overlap assessment method was sensitive to the assumed mixing rates. We also evaluated SSB in terms of the proportion of the last 25 years when true SSB was below $B_{20\%}$. 

**Figure 3.** Mean annual SSB for each population over the last 25 years of simulations done for a 65% mortality target in the true system. The x-axis of each panel represents all four simulated populations, from the LP population to the HP population. The triplets for each population are results for each of the assessment methods. The dashed line is at 20% of unfished SSB ($B_{20\%}$). The box tops and bottoms cover the inter-quartile range; the horizontal middle line represents the median value of all simulations for each scenario.

**Figure 4.** Mean of the proportion of years SSB was <20% of the unfished SSB level ($B_{20\%}$) over the last 25 years of simulations done with a 65% mortality target in the true system. The x-axis of each section represents all four simulated populations, from the LP population to the HP population. The triplets in each section are results for each of the assessment methods. The box tops and bottoms cover the inter-quartile range; the horizontal middle line represents the median value of all simulations for each scenario.
and these results qualitatively were similar to the true SSB results (Figure 4). The only noticeable difference was for MLP, MHP, and HP populations in the absence of mixing, where the pooled method had a greater (median of close to 10%) proportion of years with SSB < B_{g0}. In general, assessment method had little influence on total yield regardless of mixing scenario (Figure 5a). Here, we contrast results for SR = 100% and SR = 50% scenarios (for the full range of mixing scenarios, see Supplementary Figure S5). In the absence of mixing, the separate assessment tended to provide the highest total yield, followed by the overlap and the pooled methods; a similar pattern, though with smaller difference in total yield, was observed when mixing occurred. Across all mixing scenarios, the overlap method always provided the lowest IAV in total yield (Figure 5b). With no mixing, the pooled assessment method had the highest IAV in yield. This result was likely due to higher variability in biomass for the pooled method in this scenario, which was reflected in the more frequent low SSB values for the pooled method (Figure 4), although the mean SSB was similar to the other methods (Figure 3). However, the IAV of area-specific yield (Figure 5c) for the pooled method under SR = 50% was much lower than the separate method. Thus, the larger IAV in yield for each area for the separate assessment method largely cancelled each other out when considering aggregate yield. 

Assessment performance

In terms of assessment performance, the primary metric that was considered was bias associated with SSB estimates. The separate method was the least (median) biased method in the absence of mixing, but was positively median-biased in estimating SSB (overestimated the real total SSB) when mixing occurred (Figure 6). In the absence of mixing, the pooled method was positively median-biased, but became nearly (median) unbiased under all other scenarios. As for the overlap model, it was negatively median-biased for SSB (underestimated the real total SSB) in the absence of mixing, but became less biased as the SR decreased. In the Po-Cor scenario, the overlap model became more negatively median-biased (−0.05). Thus, with intermixing, the overlap model underestimated total SSB, and the separate method overestimated total SSB, and the pooled method provided a nearly unbiased estimator of total SSB.

Comparison of assessment methods at lower mortality targets

Given the results for 65% mortality, we explored the performance of the three assessment approaches for lower target mortality rates.
Here, we present the results for the 55% target for the scenarios SR = 100 and SR = 50%, to contrast with the results for the 65% target above. The full set of results for each target mortality rate and the full range of mixing scenarios are presented in the Supplementary data. We chose to emphasize the 55% alternative target because this led to a modest decrease in total annual yield (∼7 vs. 22% or more for lower mortality rates), while lowering the probability of SSB, B20% to near zero (Figure 7a). In addition, the relative performances of the three methods were similar at the 55, 45, and 35% target rates (Supplementary data). We present results for the SR = 100 and SR = 50% scenarios because they sufficiently contrast the methods and show how target mortality influenced the differences.

When target total mortality was decreased to 55%, the SSB increased for all four populations, with a near doubling in SSB for the LP population compared with the 65% target rate (compare Figure 7a with Figure 4 and Figure 7b with Figure 3). Differences in SSB between methods did not change in obvious ways when the target mortality rate was lowered. The relative harvest performance among assessments methods (total yield, IAV of total yield, IAV of area-specific yield) was also similar to that seen for the 65% target mortality rate (Supplementary data). Thus, the results regarding differences among methods for the 65% target mortality rate were still applicable with lower mortality target.

The results for MRE of estimating true SSB with the 55% target mortality rate were very similar to what was seen at the 65% target rate when intermixing occurred (compare Figure 8 with Figure 6). With mixing, the pooled method still had an MRE distribution centred near zero, whereas for the separate method, most average estimates of SSB were higher than the true value, indicating a positive (median) bias in estimating SSB (overestimated the real total SSB) under all scenarios (Figure 8). Only the results for the overlap model changed slightly for the lower mortality target. In the scenarios SR = 75 and SR = 50%, the overlap model produced a nearly (median) unbiased estimator of SSB under 55% target rule in contrast to negative bias for 65% rules (compared with Figure 6).

Discussion

The goal of this research was to evaluate three different assessment methods (separate, pooled, and overlap) when mixing occurs among spawning populations during the harvest season [referred to as overlap mixing (Porch et al., 2001; Goethel et al., 2011; Taylor et al., 2011)]. We found that when mixing occurred, separate assessments for each area were biased and pooled assessments were nearly unbiased and led to good management performance in terms of high SSB and low interannual variation in harvest. Although the overlap method, which explicitly accounted for movement,
performed well under some movement scenarios, in other scenarios, convergence of the assessment model was problematic.

Although the assessment methods considered in this research were similar to those considered by Cope and Punt (2011), Ying et al. (2011), and Guan et al. (2013), examinations were based on different assumptions regarding mixing. While Cope and Punt (2011) used a meta-population operating model, dispersal was assumed to occur before recruitment, with no subsequent movements of recruited fish. Thus, post-recruitment dynamics matched those of a separate population and would agree with a separate assessment assumption at the finest spatial scale. Cope and Punt (2011) considered various spatial pooling alternatives for assessment and management. Guan et al. (2013) considered one type of overlap without a direct correspondence between spawning and fishing areas. In both the Cope and Punt (2011) and Guan et al. (2013) studies, only separate and pooled assessment methods were considered and performance was evaluated in terms of estimation accuracy. Ying et al. (2011) considered three approaches to assessment using surplus production modelling in the face of intermixing that were similar to our assessment approaches. In their study, the simulated mixing scenario was a "meta-population" mixing pattern, where a proportion of resident biomass moved to and spawned in other areas. To our knowledge, the work we report here is the first simulation study comparing separate, overlap, and pooled assessment approaches for overlap mixing using SCAA assessment modelling.

With high levels of intermixing, overlapping populations approximate a pooled one, and as mixing rates decline, the populations more closely resemble unit stocks (Ying et al., 2011). Our study yielded some results that matched our prior expectations based on these limiting cases, but there were also some results that were not anticipated. In the absence of mixing, with the current 65% total mortality target, the separate method performed better than the pooled assessment method based on the performance statistics we used. However, in the face of intermixing, even with SRs as high as 75%, the pooled method outperformed the separate method in terms of SSB estimation accuracy. Thus, even a modest amount of mixing may lead to the pooled method outperforming the separate method in estimating SSB. In other systems when mixing occurred, the pooled method has also been found to be more accurate than the separate method, with the separate method resulting in positively biased estimates of stock size (Ying et al., 2011; Guan et al., 2013).

In some simulation studies, pooling has been found to lead to biased estimates, particularly when there are separate populations (at least post-recruitment) that experience substantially different fishing rates and histories (e.g. Cope and Punt, 2011; Kell et al., 2012; Hart et al., 2013). Although we observed some bias with the pooled method in the absence of mixing, the level of bias was modest. The low bias observed in this study likely reflects how TACs were allocated among areas. In our simulations, this was done based on relative cpue for the areas, leading to similarity in fishing rates among areas and stocks. In some preliminary simulations not presented herein, we allocated a constant proportion of the TAC to each area (based on population productivity in that area), and this led to more disparate fishing rates among areas and poorer management and assessment performance in the pooled approach.

We constructed an overlap assessment model that used known mixing rates and this method produced useable estimates of stock size and mortality rates except when mixing rates were high or negatively correlated with productivities. When the overlap method was included in the model comparisons, this method provided nearly unbiased estimates of SSB (median at about −1%), except for the SR = 100% and Po-Cor scenarios where results were more strongly negative biased. The convergence issues that were encountered for the overlap model with high mixing levels and the Ne-Cor scenario makes intuitive sense, given that in our simulations, all assessment data were from sampling during the fishing season. The basic issue was that recruits were too mixed among areas during the time when data were collected to allow the SCAA to separately estimate recruitment to each population, even when mixing rates were known. In particular, with an SR = 25%, area-specific abundances during the fishing season do not depend on from where fish recruit, just on total recruitment, as fish are then equally distributed among areas. Mathematically, this issue can be framed by considering the matrix of movement rates (θij, i and j varied from 1 to 4). With SRs of 25% for every population, the movement matrix is non-invertible because all elements are equal. Thus, even in the absence of sampling error, and with SRs known, there is no unique set of population abundances among areas that correspond to a given set of abundances among areas during the fishing season.

To some extent, the differences between the separate and overlap assessment models are surprising in the absence of mixing, given that dynamics assumed by the separate assessment method corresponded to the overlap assessment model without mixing. There were, however, differences in constraints that were placed on parameters. In particular, given the challenges to estimating the parameters for the overlap model, the selectivity, catchability, and standard deviation for catch data were assumed to be the same across the areas for the overlap model (which was indeed the case) but were separately estimated for the separate method (see Supplementary Table S1 for more details).

With high mixing levels, additional population-specific data, such as genetic information that would allow catches to be allocated to the population sources, could help solve the convergence issues encountered in the overlap assessment approach. Alternatively, if data could be allocated to source populations, population-specific datasets could simply be created for each population and the separate assessment approach applied to these datasets. Guan et al. (2013) discussed and evaluated such an approach, and found the estimation bias of the overall stock status was generally small when this was done (MRE of SSB <1%). Powers and Porch (2004) also discussed the potential for converting area-specific catch data to population-specific data and then applying separate assessments. An area for future research would be to evaluate how investment in population-specific data (used either to allocate catch for the separate method or to apply the overlap method) would pay off in terms of assessment and management performance relative to simply applying the pooled method in cases of high mixing among populations.

We attributed differences in management performance of the assessment methods in part to the biases associated with estimation. For example, the separate method generally overestimated SSB when mixing occurred, so the TAC set based on the assessment model results led to too high target fishing mortality rates. Unless the target rate was conservative, this would lead to poor overall management performance (e.g. lower yield and/or SSB). We also found that the estimation bias for all three models was less at the 55% mortality target than at the 65% target. It may often be the case that sensitivity to such biases (or the extent of them) would be reduced by a more conservative policy with relatively little cost in yield. These results reinforce the idea put forward by Kell et al. (2009) that precaution is warranted in the face of uncertainty associated with population intermixing.
When we reduced target mortality from 65 to 55% (or even lower levels), we found that our qualitative conclusions about the assessment model comparison did not change. Thus, we did not find evidence suggesting that a change in mortality target should necessarily lead to changes in the way spatial structure should be accounted for in assessments. Likewise, our results reinforce Molton et al. (2012, 2013) conclusion that the 65% target mortality rate may put LP lake whitefish populations at risk in the Great Lakes. The greatest benefit for reducing from a 65 to a 55% mortality target is the increase in SSB for each population, especially for LP populations, with a modest (7%) decrease in yield. Such consideration of the trade-off between precaution and realized yield should always be explicitly considered by fishery managers (McAllister et al., 1999). As Molton et al. (2012, 2013) argued, when mixing is not accounted for, LP populations can be falsely assessed to have high abundance, harvest, and productivity, by virtue of fish from other populations moving on to fishing grounds. With substantial mixing, the harvest from the area surrounding the spawning grounds of the LP population is likely to be composed largely of fish from other more productive populations. This is clearly evident in our results, where yield for the area associated with the LP population increased as mixing rates increased (see Supplementary data). To the extent to which such mixing influenced historical stock assessments and the estimations of biological reference points, this in turn tends to make the lowest productivity values we assumed even more plausible (Frank and Brickman, 2000; Ying et al., 2011).

Accurately quantifying mixing rate remains important for managing mixture fisheries. Even if we can identify strategies robust to intermixing, it is clear that appropriate mortality rates depend on population productivities. Without information on the population source of fish caught in different areas, productivity levels cannot be refined nor can it be assessed if they are changing. The current critical task for managing overlapping populations, such as the pooled management unit for northern Lake Huron, is to collect data from genetic or tagging studies to identify movement among spawning populations and how this might be associated with population productivity. If mixing rate is high, LP populations are at risk of overexploitation, which could have adverse ecological consequences due to loss of spawning components (Stephenson, 1999).

Although we incorporated considerable complexity in our simulation framework and multiple sources of uncertainty, as Irwin et al. (2008) pointed out, “not all uncertainty can be captured by any model and unexpected changes could occur”. Similar to other simulation studies (Irwin et al., 2008; Ying et al., 2011; Deroba and Bence, 2012; Guan et al., 2013), we made a number of simplifying assumptions and choices, pertaining to factors such as mixing patterns and levels of mixings, population productivities, and how fisheries operate. Likewise, we ignored temporal variation in parameters such as catchability and natural mortality. These assumptions and choices may influence the final results. In general, we based our assumptions on data available for lake whitefish and on previous studies, suggesting that qualitative results would be robust to the simplifications. While the assumptions and simplifications deserve scrutiny, we believe that the necessary simplifications we made are of secondary importance, compared with the large uncertainties regarding productivity and mixing.

**Supplementary data**

Supplementary material is available at the ICESJMS online version of the manuscript.

**Acknowledgements**

Funding for this project was provided by Quantitative Fisheries Center contributing partners, by Michigan DNR through the PERM program, and by grants from the Great Lakes Fishery Commission Fishery Research Program and the Great Lakes Fishery Trust (Project 2012.1250). 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. We thank two anonymous reviewers and Dr Mikko Heino for commenting on earlier manuscript drafts. This is publication 2014-07 of the Michigan State University Quantitative Fisheries Center.

**References**


Healey, M. C. 1978. Fecundity changes in exploited populations of lake whitefish (Coregonus clupeaformis) and lake trout (Salvelinus namaycush). Canadian Journal of Fisheries and Aquatic Sciences, 35: 945–950.


Molton, K., Brenden, T. O., and Bence, J. R. 2013. Harvest levels that conserve spawning biomass can provide larger and more stable and sustainable yields in intermixed fisheries. Fisheries Research, 147: 264–283.


Handling editor: Mikko Heino