NRC Research Press

ARTICLE

Data needs and spatial structure considerations in stock assessments with regional differences in recruitment and exploitation[1]

LaTreese S. Denson, David B. Sampson, and Andi Stephens

Abstract: This study uses a simulation experiment to demonstrate that bias in estimates of spawning biomass is influenced by the spatial configuration of a stock assessment model, whether survey data are used or not and whether an environmental index is available to inform the spatial distribution of recruitment. Stocks with limited movement of postsettlement fish may be spatially structured due to environmental forces that affect larval dispersal and recruitment distribution or from nonuniform spatial exploitation. Data are frequently aggregated across space in stock assessments, thus disregarding this complex spatial structure and possibly introducing bias into estimates of stock status. An operating model (OM) is created that simulates data that are used in a set of estimation models to assess bias. The following experimental factors are considered: (*i*) using survey data and environmental indices in the assessment; (*ii*) using disaggregated data (two regions, as generated by the OM) or aggregated data (one region); and (*iii*) incorporating different patterns in the OM's regional exploitation and environmentally driven recruitment distribution.

Résumé : L'étude fait appel à une expérience de simulation pour démontrer que le biais dans les estimations de la biomasse reproductrice est influencé par la configuration spatiale d'un modèle d'évaluation des stocks, par l'utilisation ou non de données de relevés et par la disponibilité ou non d'un indice environnemental fournissant de l'information sur la répartition spatiale du recrutement. Les stocks caractérisés par des déplacements limités des poissons après leur établissement pourraient avoir une structure spatiale en raison de forces environnementales qui influent sur la dispersion des larves et la répartition du recrutement ou d'une exploitation non uniforme dans l'espace. Les données sont fréquemment regroupées dans l'espace dans les évaluations de stocks, une pratique qui fait fi de cette structure spatiale complexe et peut introduire un biais dans les estimations de l'état des stocks. Un modèle opérationnel (MO) est créé qui simule les données qui sont utilisées dans un ensemble de modèles d'estimation pour évaluer le biais. Les facteurs expérimentaux suivants sont examinés : (*i*) l'utilisation de données de relevés et d'indices environnementaux dans l'évaluation, (*ii*) l'utilisation de données non regroupées (deux régions générées par le MO) ou regroupées (une région) et (*iii*) l'incorporation de différents motifs dans la répartition régionale dans le MO de l'exploitation et du recrutement modulé par les conditions environnementales. [Traduit par la Rédaction]

Introduction

Numerous factors can affect the accuracy of stock assessment estimates of population size, including the extent, accuracy, and precision of data, the validity of biological input parameters, the soundness of assumptions made about the data, and the underlying biological processes such as recruitment and mortality. An assessment model that considered all potential factors would be highly complex and would likely not have adequate supporting data. A common simplifying assumption is that the fish in a stock can be treated as being spatially homogeneous. An additional simplification is to assume that biological processes are time-invariant rather than being influenced by changes in the surrounding environment over time. These simplifications could lead to an overoptimistic assessment of a stock's status and productivity and subsequently to over-harvesting in less productive areas (Hilborn et al. 2002; Magnusson and Hilborn 2007; National Research Council 1998). Conversely, simplifications that lead to an overly pessimistic assessment could result in under-fishing in more productive areas and an unnecessary potential loss of revenue for the fishers and fish for consumers.

Many fish populations are made up of spatially dispersed subpopulations that receive recruits from a common pool but that have limited exchanges of individuals postrecruitment. These subpopulations experience different catch and management histories and therefore vary in their age structure (Begg and Waldman 1999; National Research Council 1998). These spatial subpopulations are often identifiable in fishery age- and size-composition data. Subpopulation structure not only stems from regional dif-

Received 29 June 2016. Accepted 12 July 2017.

L.S. Denson.* Department of Fisheries and Wildlife, Oregon State University, Nash Hall, 2820 SW Campus Way, Corvallis, OR 97331, USA.
D.B. Sampson. Coastal Oregon Marine Experiment Station and Department of Fisheries and Wildlife, Oregon State University, Hatfield Marine Science Center, 2030 SE Marine Science Drive, Newport, OR 97365, USA.
A. Stephens. Northwest Fisheries Science Center NOAA Fisheries, Newport Research Station, 2032 S.E. OSU Drive, Newport, OR 97365, USA.
Corresponding author: LaTreese S. Denson (email: ldenson@rsmas.miami.edu).
*Present address: Rosenstiel School of Marine and Atmospheric Science, Department of Marine Biology and Fisheries, University of Miami, 4600 Rickenbacker Causeway, 212 S, Grosvenor, Miami, FL 33149, USA.

[1]This article is being published as part of the special issue "Space Oddity: Recent Advances Incorporating Spatial Processes in the Fishery Stock Assessment and Management Interface" arising from a related theme session at the 145th Annual Meeting of the American Fisheries Society, Portland, Oregon, USA, August 2015.

 Published at www.nrcresearchpress.com/cjfas on 18 July 2017.

ferences in fishing patterns, such as those that occur at the closing of an area for protection (McGilliard et al. 2015; Punt et al. 2016), but also from temporal changes in the relative distribution of recruitment among regions, possibly driven by the surrounding ocean dynamics (Ainley et al. 1993; Charter and Sandknop 2000; Hurtado-Ferro et al. 2014). These environmental processes are especially important for organisms that are sedentary as adults (Shanks and Eckert 2005).

Although there have been a number of recent simulation studies evaluating the effects of spatial structure on age-structured stock assessments, as discussed in Punt et al. (2015), only a relatively small number of investigations have incorporated population spatial structure and environmental influences into assessment models. Cope and Punt's (2011) simulation study suggested that (*i*) relatively small differences in catch histories among spatial subpopulations can create bias in estimates of stock status, and (*ii*) the mechanisms driving spatial structure in the underlying population dynamics (e.g., growth, mortality, and reproduction) will influence the spatial scale needed to accurately reflect the population dynamics. Haltuch et al. (2011) explored the potential benefits of including an environmental index as a driver of annual recruitment deviations. A study by Hulson et al. (2013) investigated the performance of spatially explicit models that include tagging data and movements of fish between regions, where the movement rates were affected by environmental changes.

In the current study, we explore a stock assessment that incorporates both an environmental driver and spatial structure, combining elements of the published studies mentioned above. We use a factorial experimental design and a Monte Carlo simulation approach (Rubinstein and Kroese 2008) to evaluate the performance of different configurations of an age-structured assessment model that is applied to simulated data from a spatially structured stock. Our operating model (OM) generates spatially structured fishery and survey data, as well as an environmental index for the spatial distribution of recruitment. The OM calculates true values of the population size and various stock status metrics. We analyze the generated data using the estimation platform Stock Synthesis (SS; Methot and Wetzel 2013), an age-structured stock assessment tool that can be configured to include spatial structure and that produces estimates of the population size and stock status metrics. We also experiment with different methods for treating the data within the estimation model. The estimates are compared with the OM's calculated true values to evaluate the strength and importance of the various experimental factors on bias in the estimation model results. Our goal is to provide new insight into the bias that may result when stock assessments are applied to populations that are heterogeneously distributed and exploited.

Methods

Operating model (OM)

The OM simulates the dynamics of an age- and spatially structured population using standard equations for survival, growth, and recruitment to project the population forward 25 years. The equations are consistent with those commonly used in the stock assessment simulation literature (e.g., Haltuch et al. 2008; Methot and Wetzel 2013; Punt et al. 2002). Values for the biological parameters are similar to those used to describe the life history characteristics of black rockfish (*Sebastes melanops*), which is an important target species for nearshore commercial and recreational fisheries along the United States west coast. Details of the OM's equations and parameter values are provided in the online Supplementary Data[2].

The two regional subpopulations in the simulated stock have the same natural mortality and growth characteristics. Natural mortality (M) is held constant across ages and years for simplicity. The unfished stock size is determined by arbitrarily setting the unfished level of recruitment to 2000 fish, apportioned to the two regions at 60% and 40% under equilibrium conditions. After fish recruit to a region, they do not move between regions, which is a reasonable simplifying assumption for black rockfish, which tend to be sedentary as adults (Green and Starr 2011).

Annual recruitment in the simulated stock depends on the combined spawning biomass of the two regional subpopulations, and the expected values are calculated using a Beverton–Holt spawner–recruit relationship. The OM assumes that density dependence operates at the scale of the entire population (i.e., with both regions combined). The overall pool of recruits is distributed to the regions each year based on a recruitment distribution parameter $P_{Er}(t)$, which is driven by an environmental factor. The annual values for recruitment include lognormal random variability, and the $P_{Er}(t)$ values also include random variability.

The simulated environmental factor follows one of three patterns: constant, abrupt, or gradual. When the environmental factor (EnvirP) follows the constant pattern (denoted as E_C), there are no systematic changes in the regional distributions of the recruits. When EnvirP follows the abrupt pattern (E_A), there is an environmental change in year 12 that forces a step change (e.g., from 40% to 60%) in the expected regional distributions of the recruits in that year and all subsequent years. Similar to how the 1976–1977 regime shift is thought to have affected recruitment dynamics for a variety of North Pacific fish stocks (Bograd and Lynn 2003; Hare and Mantua 2000), we postulate that more subtle changes in hydrographic features could impact the regional distribution of recruitment without influencing the magnitude of recruitment (Bograd and Lynn 2003). Having the shift in year 12 provides time for the changed influx of recruits to influence the age composition data prior to the end of the 25-year simulation. When EnvirP follows the gradual pattern (E_G), there is a gradual but cyclical environmental change that causes wave-like patterns in the regional distributions of the recruits over the 25-year simulation (Supplementary Data[2] Fig. S.1).

The simulated population is subjected to one of three different exploitation histories (ExpRate), modeled as linear trends in the expected values for the regional fishing mortality rates with random lognormal variability. When the ExpRate factor is constant (denoted as F_C), there are no systematic changes in the expected fishing rates. With the mixed pattern (F_M), fishing mortality in one region tends to increase linearly over time, while the fishing rate in the other region tends to decrease with the same step size. If the ExpRate factor is increasing (F_I), the fishing mortality rates in both regions tend to increase linearly over time. The operating model used simulated environmental patterns and exploitation histories rather than patterns observed historically. This provided complete experimental control over the simulated time series and lessened the complexity of the OM. A visual display of the ExpRate and EnvirP patterns is available in the Supplementary Data[2] (Fig. S.1).

Although each region has its own independent fishery, the OM assumes that the fisheries use similar gear and have the same pattern of logistic selectivity, with age-6 fish experiencing half the full rate of fishing mortality and age-8 fish experiencing 98.07% of the full rate. Catches-at-age are calculated using the standard Baranov catch equation. The regional fishing mortality rates for years prior to the simulated period are fixed at 0.11 year^{-1} in region one and 0.055 year^{-1} in region two, which results in the expected total spawning biomass in the first year of each simulation to be

[2]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2016-0277.

40% of the total unfished spawning biomass when there is no variability in recruitment or fishing mortality. As part of the experimental design, regardless of the setting of the environmental factor, the total spawning biomass in the last year (year 25), if there is no variability in recruitment, recruitment fractions, or exploitation, is about 62% (±2%) of the unfished level when the exploitation history is constant or mixed and about 25% (±0.8%) when ExpRate is increasing.

The three environmental patterns and three exploitation histories provide nine OM scenarios for the simulated population dynamics. For each of these scenarios, the OM generated 100 replicate data sets that include process error due to annual random deviations in recruitment, in the regional recruitment distribution fractions, and in the regional fishing mortality rates. Details are provided in the Supplementary Data[2].

The data sets generated for analysis with stock synthesis include the following elements: regional series of annual values for the fishery catches, fishery catch age-composition observations, fishery catch-per-unit-effort (CPUE) indices of abundance, survey biomass indices and age-composition observations, an environmental index for the recruitment distribution fractions, and midyear masses-at-age. The data for the fishery catches, the environmental index, and the midyear masses-at-age are generated to be perfectly accurate, but the other elements include observation error to mimic random sampling processes. The age-composition data are generated to have multinomial error; the fishery CPUE index and the survey biomass index data include lognormal error. Details of the data generation process are provided in the Supplementary Data[2].

For each replicate, the OM also calculates the following five "true" focal quantities: (*a*) the unfished equilibrium spawning stock biomass (SSB_0), which does not change between replicates, (*b*) spawning stock biomass in the last (25th) year of the series ($SSB_{Current}$), which does change between replicates due to recruitment variability, (*c*) the maximum sustainable yield (MSY) for the entire stock (both regions), (*d*) the spawning stock biomass that produces the maximum sustainable yield (SSB_{MSY}), and (*e*) depletion in the final year of the simulated period (Depletion = $SSB_{Current}$ / SSB_0).

Estimation methods

Stock Synthesis (SS) version 3.21e (Methot and Wetzel 2013) was used to estimate sets of parameter values for six model configurations (MCs, described below) that were fit to simulated data for each replicate. The SS estimation models were provided the correct values for the natural mortality rate and the steepness of the spawner–recruit relationship, as well as the correct values for the masses-at-age (thus eliminating the need to estimate growth parameters). The models were configured to estimate values for the following parameters: deviations from equilibrium in the initial age composition, the unfished and initial equilibrium recruitment, the initial recruitment fraction to region one, the series of annual fishing mortality rates for each region, annual recruitment deviations for the simulated period, parameters for the logistic selectivity curves, and the catchability coefficients for the two regional fisheries and the surveys. Survey selectivity and catchability were specified to be the same in both regions as in the OM. For some of our experimental treatments, SS was configured to estimate a parameter that controls the strength of the linkage between the environmental index and the regional recruitment distributions. Details of how SS uses environmental indices are provided in Methot and Wetzel (2013). Other estimation models were configured to estimate annual deviations in the regional recruitment distribution. In these cases the standard deviations of the annual deviations were calculated by the operating model and used in the estimation models to constrain the regional recruitment distribution deviations.

SS derives its maximum likelihood parameter estimates by finding the set of parameter values that minimizes the total negative log-likelihood, calculated as the weighted sum of log-likelihood components from each data source. The synthesis models were provided the correct forms for each likelihood component (e.g., lognormal for the indices and multinomial for the age compositions) and the correct values for scaling each component (e.g., the same log-scale standard deviations and effective sample size values that the OM used to generate the observations). For MCs that ignored spatial structure, sample sizes for the compositional data were added together to reflect the number of aged fish in the combined regions. For survey abundance and fishery CPUE data series, the log-scale standard deviations for the combined indices were the same values as the regional log-scale standard deviations. The log-likelihood components were weighted equally in the total log-likelihood, except for one experimental treatment in which the components for the survey data were down-weighted to near zero to remove their influence. For the MC in which the SS model exactly matched the OM, we provided the synthesis models with perfect data (no observation error but process error in recruitment, recruitment fractions, and fishing mortality) to verify that the OM and SS were consistent.

Evaluating convergence

SS uses an iterative approach for locating the set of parameter values that maximize the likelihood (minimize the negative log-likelihood) function, but the search process may not converge, either because the likelihood surface does not have a unique set of parameter values that result in a single maximum likelihood value (e.g., there is a ridge) or because the surface is irregular with several sets of parameter values that each result in a local maximum likelihood value. To provide some assurance that SS found maximum likelihood estimates, we took the following approach. For each of the 100 iterations of each experimental treatment, there were at least six stock synthesis runs, each starting from parameter estimates that were randomly "jittered" from the true parameter values. Each run was tested to determine that it had produced a valid likelihood value and an invertible Hessian matrix. From the set of runs with valid likelihoods and invertible Hessians, the run with the results that produced the maximum likelihood value were selected (the one with the smallest negative log-likelihood value). While there is no guarantee that this approach always resulted in the maximum likelihood estimates, we found no indications to the contrary (e.g., outlying results). Details of the algorithm are provided in the Supplementary Data[2].

Estimation MCs

In addition to the two factors that control the population dynamics (EnvirP and ExpRate), the experimental design (Table 1) also evaluated the following: (*a*) the importance of using survey data (SurData) in the estimation model, (*b*) whether or not the estimation model used an environmental index to drive the regional distribution of recruitment (EnvInd), and (*c*) whether or not the model accounted for spatial structure (SpStru). Exploring all feasible combinations of these factors required six MCs (Table 2).

One configuration (denoted $MC_{2.S.E}$) is fully consistent with the OM and makes full use of all the available data. The synthesis model for this configuration is set up for two regions, uses the two-region spatial data, the two-region survey data, and the environmental index. Configuration $MC_{2.nS.E}$ is similar but omits the survey data. Configuration $MC_{2.S.nE}$ includes the survey data but omits the environmental index. $MC_{2.nS.nE}$ omits both the survey data and the environmental index.

The environmental index is irrelevant for the models having only one region because recruitment to this region is always 100%. Configuration $MC_{1.S}$ is a one-region model in which the fishery and survey data are included in collapsed form (aggregated to one region). The final configuration ($MC_{1.nS}$) has an incorrect spatial structure (one region) and omits the survey data.

Table 1. Description of the five experimental factors, their patterns and levels, and whether they affect the operating model or the estimation model.

Factor	Description	Level 1	Level 2	Level 3
Operating model				
EnvirP	Environmental pattern used to drive the regional recruitment distribution	Constant E_C	Abrupt E_A	Gradual E_G
ExpRate	Exploitation history patterns for the regional subpopulations	Constant F_C (increasing population)	Mixed F_M (increasing population)	Increasing F_I (decreasing population)
Estimation model				
SpStru	Assumed spatial structure	One region with combined data	Two regions with region-specific data	—
SurData	Use of survey data, age composition, and biomass	Data included	Data not included	—
EnvInd	Use of an environmental index to drive the regional recruitment distribution; not applicable in one-region model configurations	Index included	Index not included	—

Note: To simplify the analyses and displays of results, the three estimation model factors were combined into a single model configuration (MC) factor, described in the text.

Table 2. Key components of the operating model (OM) compared with the different model configurations (MCs) used to set up the stock synthesis (SS) estimation models.

	Fishery region	Survey region	Environmental index
OM	2	2	Yes
$MC_{2.S.E}$	2	2	Yes
$MC_{2.nS.E}$	2	No survey	Yes
$MC_{2.S.nE}$	2	2	No index
$MC_{2.nS.nE}$	2	No survey	No index
$MC_{1.S}$	1	1	Not applicable
$MC_{1.nS}$	1	No survey	Not applicable

Note: The six SS models were applied to data from each replicate of the OM. The numbers in the cells are the number of regional components. The one-region synthesis models combined the regional components produced by the OM.

The estimation models described above did not include any recruitment deviation bias adjustment to the likelihood, as recommended in Methot and Taylor (2011). To evaluate the potential effect of including the recruitment bias adjustment, we conducted a mini-experiment for two operating model scenarios (described below).

Comparing OM with estimation model outputs

To evaluate the performance of the six estimation MCs, for each of the nine OM scenarios and 100 replicates, the derived quantities SSB_0, $SSB_{Current}$, MSY, SSB_{MSY}, and Depletion were chosen and estimates for these quantities were compared with their true values using the relative error (RE):

$$\text{RE} = \frac{\hat{\theta} - \theta}{\theta} \tag{1}$$

where $\hat{\theta}$ is an estimated quantity, and θ is the true value calculated by the OM.

The median relative error (MRE) was used to summarize the overall bias in the estimates. The median is more robust to outliers than the mean of the relative errors. The MRE and RE values for SSB_0, $SSB_{Current}$, and the other focal quantities were examined graphically and statistically to gauge the importance of the different MCs and levels of EnvirP and ExpRate. The median absolute relative error (MARE) was used to quantify accuracy. To simplify the presentation, only the MRE results are reported in the main paper. Comparisons of the MRE and MARE values for SSB_0, $SSB_{Current}$, and Depletion are presented in the Supplementary Data[2] (Figs. S.2 and S.3).

Linear mixed effect models (LME) were fit across the RE values from the three-by-three factorial experimental design, and a series of linear hypothesis tests were used to measure the importance of various factor-level combinations based on contrasts (differences) between the LME model coefficients. The LME models were fit using the R package "nlme" (Pinheiro et al. 2016). Factors EnvirP, ExpRate, and MC were fixed effects, and MC was treated as a repeated measure. Repeated measures were needed because the analysis used each of the replicates six times, once with each of the six MCs, which leads to a lack of independence in the RE variables from the six MCs. The random effects in this mixed model were the 100 replicates for each of the nine OM scenarios. The linear hypothesis tests were executed using the "multcomp" R package (Bretz et al. 2011).

The best LME models were selected to represent the RE of SSB_0 and $SSB_{Current}$ using a backwards-stepwise model selection process and the Akaike information criterion (AIC; Burnham and Anderson 2002). The model selection process for each variable started from the full model that included all main effects and interactions (eq. 2) and iteratively removed predictors that produced the highest AIC values.

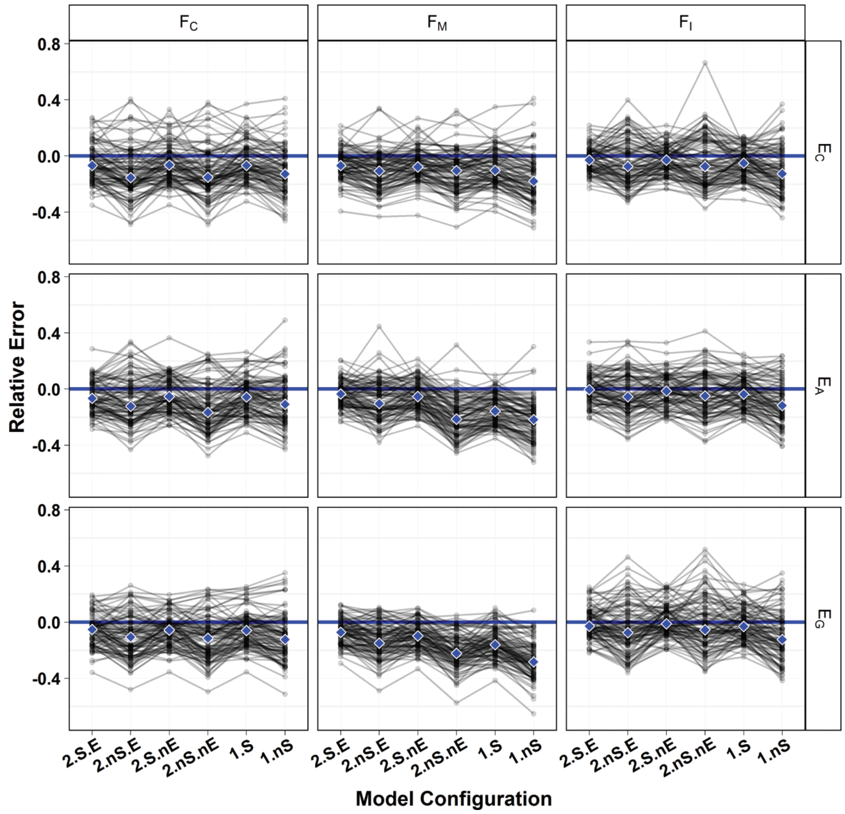
$$\begin{aligned} \text{RE Full Model} \sim\ & \text{EnvirP} + \text{ExpRate} + \text{MC} \\ & + \text{EnvirP} \times \text{ExpRate} + \text{EnvirP} \times \text{MC} \\ & + \text{ExpRate} \times \text{MC} + \text{EnvirP} \times \text{ExpRate} \times \text{MC} \end{aligned} \tag{2}$$

Mini-experiment to explore bias adjustments for annual recruitment deviations

Previous research suggests that SS tends to underestimate the true recruitment deviations (Methot and Taylor 2011). This is due to an interaction between the likelihood component that penalizes the recruitment deviations, causing them to shrink toward zero, and changes over the extent of the assessment period in the quality of the data that inform the recruitment deviation estimates. To reduce the bias in the estimation of the recruitment deviations, Methot and Taylor (2011) suggest modifying the likelihood function for recruitment to include a bias adjustment that fluctuates over time. Initially bias adjustments were not included in SS runs for this study. To evaluate the potential impact of the bias adjustment on the biomass estimates from our full experiment, the following mini-experiment was conducted.

Sets of 25 replicates were randomly drawn from the original 100 replicates produced by the OM for two scenarios chosen arbitrarily. These randomly drawn data sets were used in additional estimation models that contained the bias adjustment for recruitment deviations based on formulas given in Methot and Taylor (2011), as implemented in the r4SS package (Taylor et al. 2014). The

Fig. 1. Three-by-three line plots of relative error values for $SSB_{Current}$. Each panel represents a different operating model (OM) scenario, defined by the combination of patterns from factors EnvirP and ExpRate. For each OM scenario, the six different model configurations (MCs) were applied to each of the 100 random replicates, as indicated by the horizontal lines connecting the points across the MCs in each panel. The large solid symbol for each MC in each panel marks the median relative error for $SSB_{Current}$ produced by that specific MC and specific OM scenario. The median relative error (MRE) values below the zero line indicate underestimation of $SSB_{Current}$ (negative bias).



bias adjustment fractions were estimated for each replicate using a two-step process. The first step used five predetermined values to define the bias correction "ramp". The predetermined bias adjustment fraction increased from 0.0 in year 1975.2 to a maximum of 0.97 from 1984.4 to 2002.0 and decreased to 0.0 in 2022.0. Using these as base values, new values were estimated for each replicate in a single step. Across the 25 replicates used in the mini-experiment, the estimated bias adjustment fractions increased from 0 in 1973 (SD = 1.98 years) to a mean maximum value of 0.97 (SD = 0.0069) for the period 1985 (SD = 1.36 years) to 2020.8 (SD = 0.9211) and decreased to 0 in 2022.6 (SD = 1.12). These values were then used in the final estimation models for this mini-experiment.

To determine the influence of the bias adjustment, estimates of SSB_0 and $SSB_{Current}$ from the estimation model runs with bias adjustment were compared with their true values by calculating the REs, which were then compared with the RE values of the estimation models that did not use the bias adjustment. The comparisons were examined graphically and statistically to gauge the importance of the bias adjustment on our estimates. Having the same replicate data set for the bias-adjusted and non-bias-adjusted RE values allowed us to conduct the statistical analysis on the difference between the two RE values (Δ_{Bias}) using the following simple analysis of variance (ANOVA) model:

(3) $$\Delta_{Bias} = \text{Scenario} \times \text{MC}$$

Results

Convergence

All 100 replicates for each MC and OM scenario combination resulted in sets of parameter values, at least one of which produced an invertible Hessian, based on at least six runs that started from "jittered" initial values. The run that was accepted for each replicate produced an invertible Hessian and had the smallest negative log-likelihood.

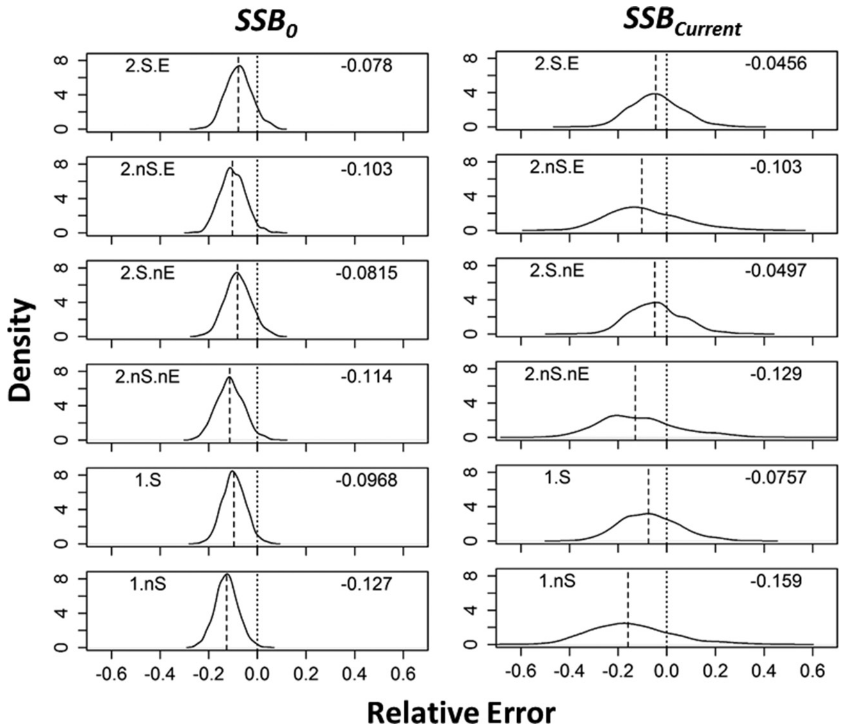
RE of the focal quantities

Many of the RE values for the five focal quantities were relatively highly correlated. Correlations among the values for MSY, SSB_{MSY}, and SSB_0 were almost perfect, suggesting that two of these variables could be eliminated with little loss of information. The high correlations reflect the fact that all three quantities are defined by similar equations that multiply the estimate of unfished recruitment with functions involving natural mortality, steepness, and mass-at-age, which were all fixed at their true values in the experiment. Mangel et al. (2013) have previously noted that the quantities MSY, SSB_{MSY}, and SSB_0 are tightly related when natural mortality and steepness are fixed. Our analyses here focus on the unfished spawning stock biomass (SSB_0) and the spawning stock biomass in the last year ($SSB_{Current}$). For simplicity, Depletion is excluded in the display and interpretation of results presented below. $SSB_{Current}$ was highly correlated with Depletion ($\rho_{X,Y}$ = 0.93), and we obtained similar results for both variables. Summary results for Depletion, $SSB_{Current}$, and SSB_0 are shown in the Supplementary Data[2] Figs. S.2 and S.3.

Factor effects on bias

Given our experimental design, the RE of an estimate can be visually summarized in a three-by-three set of line plots. Figure 1 shows the RE values for $SSB_{Current}$ by replicate for the nine OM scenarios across the six MCs. The thin grey horizontal lines in

Fig. 2. Density plots showing the distribution of relative error values for estimates of SSB_0 and $SSB_{Current}$. The plots are arranged by the six model configurations (MCs) described in the text. The short-dashed vertical line at 0.0 indicates no bias. The long-dashed vertical line represents the median relative error (MRE) across all operating model (OM) scenarios and replicates for each specific MC; the MRE values are located in the top right corner of each panel.



each panel connect the replicates from MC to MC. Nonparallel lines indicate interactions between MC and the replicates. The spread between the grey lines for a given MC indicates the variability in the estimates. The large solid symbols mark the median RE values for each experimental treatment. They all are below the horizontal zero reference lines, indicating an overall tendency to underestimate $SSB_{Current}$. All of the nine OM scenarios show the least amount of negative bias (closer to zero) when using $MC_{2.S.E}$ compared with the other MCs. Configurations $MC_{2.S.nE}$ and $MC_{1.S}$ show similar reduced negative bias relative to counterparts $MC_{2.nS.nE}$ and $MC_{1.nS}$, which have no survey data. The line plots of RE values for SSB_0 (not shown but comparable to Fig. 1) show similar patterns with less variability between replicates.

Density plots provide simpler summaries of the RE values. Figure 2 shows the RE values for SSB_0 and $SSB_{Current}$ for the six MCs pooled across the nine OM scenarios. The plots for SSB_0 indicate that most of the configurations tend to underestimate SSB_0. The least negatively biased estimate (MRE = –7.8%) was for $MC_{2.S.E}$, the configuration fully consistent with the OM (two regions and uses all the available data). The most biased estimate (MRE = –12.7%) was for $MC_{1.nS}$, the configuration that is both least consistent with the OM and least data-inclusive (data collapsed to one region; no survey data, no environmental index). $SSB_{Current}$ was also consistently underestimated. As was the case for SSB_0, configuration $MC_{2.S.E}$ produced the least negatively biased estimates of $SSB_{Current}$ (MRE = –4.6%), and $MC_{1.nS}$ was the most biased (MRE = –15.9%). The $SSB_{Current}$ RE values are more widely spread than the SSB_0 RE values, suggesting that SSB_0 can be estimated more precisely than $SSB_{Current}$.

The density plots also illustrate how the RE values for SSB_0 and $SSB_{Current}$ vary with the environmental patterns EnvirP (Fig. 3) and exploitation histories ExpRate (Fig. 4). The three environmental patterns had little effect on the median RE values. The exploitation histories had a slight effect on the median RE value for SSB_0 and a much stronger effect on the MRE for $SSB_{Current}$. Although these density plots suggest that EnvirP and ExpRate are relatively unimportant factors, the plots obscure how EnvirP and ExpRate interact with the six MCs to influence the bias and accuracy of the spawning biomass estimates (e.g., Figs. 1, S.2, and S.3[2]).

Mini-experiment: recruitment deviation bias adjustment effect on bias

Line plots illustrate the differences between RE values when the bias adjustment is used (Ramp) or not used (No Ramp). Figure 5 shows these differences for OM scenarios $F_C : E_C$ and $F_M : E_A$ across the six MCs. Similar to the previous line plots, the grey lines connect replicates from estimates with the bias adjustment to estimates without the bias adjustment. The large solid symbols mark the median RE values for each treatment.

The bias adjustment had a minimal effect on the bias in the estimation of $SSB_{Current}$, and for almost all replicates the effect was to slightly accentuate the negative bias. The large solid symbols representing the MRE values are all either at or below the horizontal zero reference lines, indicating an overall tendency to underestimate $SSB_{Current}$, regardless of whether the bias adjustment is used or not, and the median values from the estimates based on the bias adjustment are slightly lower than the estimates from models that did not include the bias adjustment. For scenario $F_C : E_C$, the mean increase in negative bias was 1.07%; for scenario $F_M : E_A$ the mean increase in bias was even smaller (0.72%).

The bias adjustment had a greater effect on estimates of SSB_0 than it did on estimates of $SSB_{Current}$. For scenario $F_C : E_C$, the mean increase in negative bias was 1.66%; for scenario $F_M : E_A$, the mean increase in bias was smaller (1.47%). Again, using the bias adjustment accentuated the negative bias.

Visually, the mean differences between estimation methods seem trivial, but the factorial ANOVA results indicated significant differences between "Ramp" and "No Ramp" estimates and significant scenario and MC effects (p value < 0.01). MC had a larger effect than scenario on the difference in RE values.

Fig. 3. Distribution of relative error (RE) values for estimates of SSB_0 and $SSB_{Current}$ arranged by EnvirP pattern. Short-dashed vertical lines at 0.0 indicate no bias. The long-dashed vertical lines represent the median relative error (MRE) across all operating model (OM) scenarios and model configurations for each EnvirP pattern; the MRE values are located in the top right corner of each panel. The RE values for SSB_0 have a smaller spread than the ones for $SSB_{Current}$. The MRE values are fairly similar for both variables and all EnvirP patterns.

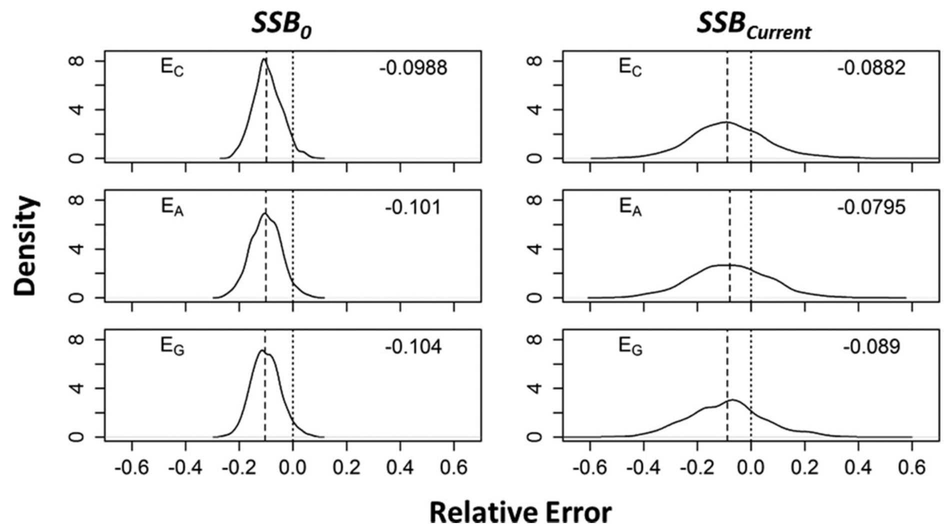
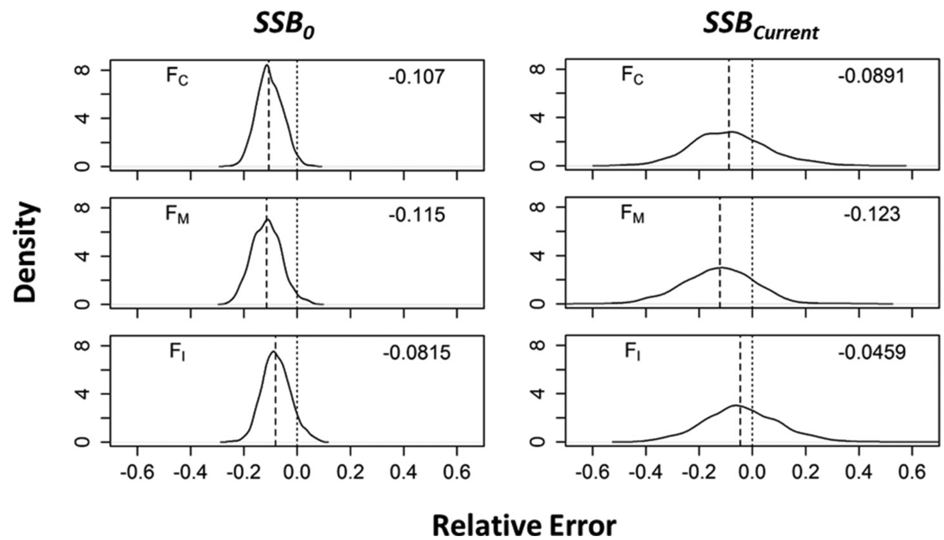


Fig. 4. Density plots for relative error (RE) values for estimates of SSB_0 and $SSB_{Current}$ by ExpRate pattern, similar to Fig. 3. The median relative error (MRE) values for SSB_0 (pooled across all MC and EnvirP patterns) are located in the top right corner of each panel.



Linear hypothesis testing

We used backwards model selection and AIC to choose parsimonious models for the RE values for SSB_0 and $SSB_{Current}$. The resulting model for both response variables was as follows:

$$\text{RE} = \beta_0 + \beta_1\text{EnvirP} + \beta_2\text{ExpRate} + \beta_3\text{MC} + \beta_4\text{ExpRate} \times \text{MC} \tag{4}$$

The model includes all three main effects and one interaction with MC. Summaries of the models we considered during the selection process are provided in Supplemental Data[2] Table S.2.

The LME models were used to calculate specific contrasts and gauge the importance of the different factors. The contrast denoted as "2.S.nE – 2.nS.nE" in Table 3, for example, measures the effect of having survey data within a two-region MC that does not include the environmental index. The contrast denoted by "2.S.nE : 1.S – 2.nS.nE : 1.nS" measures the importance of survey data in either the two-region or the one-region configuration. Because the model for RE (eq. 4) includes the interaction ExpRate × MC, there are six contrasts involving F_M and F_I, one for each of the six MC levels.

The contrasts were ranked by their coefficient values, and we report the top 20 of 54 contrasts for $SSB_{Current}$ (Table 3) and SSB_0 (Table 4). Below we highlight the influence of specific factor-level combinations on bias in the estimates of SSB_0 and $SSB_{Current}$.

Linear hypothesis tests for RE($SSB_{Current}$)

The overall effect on RE($SSB_{Current}$) of configuring the synthesis models for two regions rather than one (SpStru), combining cases where survey data were used or not, is represented by the contrast "2.S.nE : 2.nS.nE – 1.S : 1.nS" (Table 3). This contrast ranked in the top 20 only when the exploitation history (ExpRate) was mixed (F_M; rank 8) or increasing (F_I; rank 13), and in both cases negative bias in the estimate of $SSB_{Current}$ is reduced with the two-region models.

The importance of having data from a survey (SurData), regardless of the spatial configuration, was measured by the contrast

Fig. 5. Line plots of relative error values for $SSB_{Current}$ from estimation models with the bias adjustment (Ramp) and without the bias adjustment (No_Ramp). The top row displays the change in relative error (RE) under operating model (OM) scenario $F_C : E_C$, and the bottom row represents the change in RE under OM scenario FM : EA. The effect of the bias adjustment (ramp) is also displayed across all six model configurations combined with the two OM scenarios. The two large solid symbols in each box represent the median RE for $SSB_{Current}$ produced by that specific model configuration, OM scenario, and use of a bias adjustment or not.

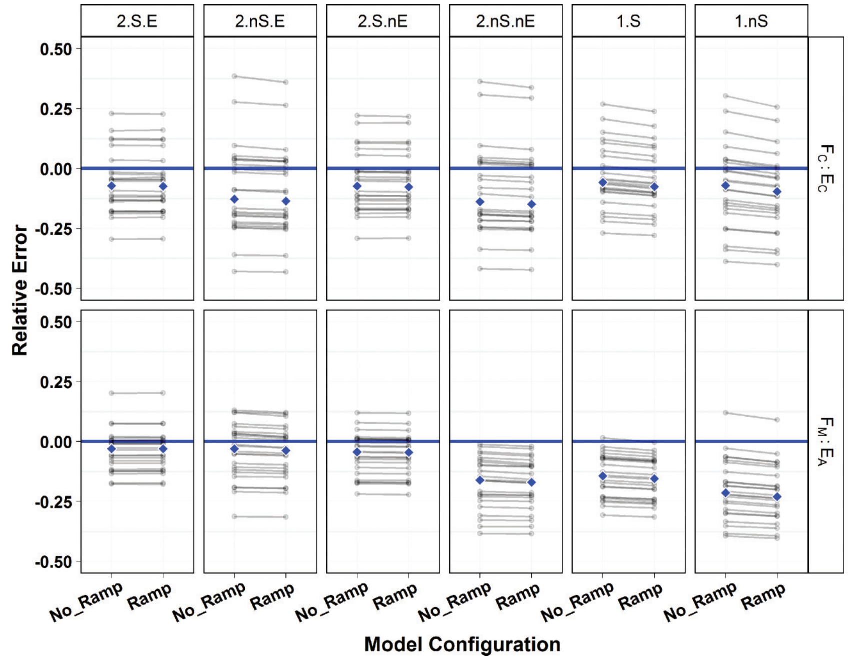


Table 3. Top 20 contrasts for model coefficients for the relative error of $SSB_{Current}$ estimates.

Rank	Contrast	Factor level	ExpRate	SpStru	EnvInd	SurData	Coefficient
1	2.S.nE : 1.S – 2.nS.nE : 1.nS	F_M				A	0.1984
2	2.S.E : 2.nS.E – 1.S : 1.nS	F_M		A	A		0.1820
3	$F_I - F_M$	$MC_{2.nS.nE}$	×				0.1394
4	2.S.nE : 1.S – 2.nS.nE : 1.nS	F_C				A	0.1348
5	$F_I - F_M$	$MC_{1.nS}$	×				0.1158
6	$F_M - F_C$	$MC_{1.nS}$	×				0.1134
7	2.S.nE – 2.nS.nE	F_M				A	0.1052
8	2.S.nE : 2.nS.nE – 1.S : 1.nS	F_M		A			0.1046
9	2.S.nE : 1.S – 2.nS.nE : 1.nS	F_I				A	0.0967
10	$F_I - F_M$	$MC_{1.nS}$	×				0.0954
11	1.S – 1.nS	F_M				A	0.0932
12	$F_C - F_I$	$MC_{2.nS.nE}$	×				0.0923
13	2.S.nE : 2.nS.nE – 1.S : 1.nS	F_I		A			0.0910
14	$F_M - F_C$	$MC_{1.S}$	×				0.0786
15	1.S – 2.nS.nE	F_C		A		R	0.0783
16	2.S.E : 2.nS.E – 1.S : 1.nS	F_I		A		A	0.0782
17	2.S.E : 2.nS.E – 2.S.nE : 2.nS.nE	F_M			A		0.0774
18	2.S.nE – 2.nS.nE	F_C				A	0.0765
19	1.S – 1.nS	F_I				A	0.0727
20	2.nS.nE – 1.nS	F_I		A			0.0699

Note: Contrasts were examined using linear hypothesis testing methods applied to the model given by eq. 4. Each contrast was evaluated for all levels of any relevant factor interacting with the contrast. Contrast "1.S – 1.nS", for example, measures the mean difference between the relative error values from model configurations $MC_{1.S}$ and $MC_{1.nS}$ and appears in the table for two of the three levels of ExpRate. The four columns (ExpRate, SpStru, EnvInd, and SurData) indicate the factor whose importance was measured by the contrast. The contrasts are ranked by the values of their coefficients, which are all positive, indicating that the first set of levels in the contrast (e.g., 2.S.nE : 1.S) caused a reduction in bias towards zero compared with the second set (e.g., 2.nS.nE : 1.nS). In the columns for SpStru, EnvInd, and SurData, which are the factors that control the model configurations in the contrast, an "A" indicates that adding data for this factor caused bias to be reduced towards zero; an "R" indicates that removing data for this factor caused bias to be reduced. The table does not include a column for EnvirP because contrasts between levels of that factor had ranks greater than 20.

"2.S.nE : 1.S – 2.nS.nE : 1.nS". This contrast ranked first when the exploitation history was at the mixed level (F_M), producing a 19.8% reduction in bias towards zero when survey data were included. The same contrast ranked fourth for F_C and ninth for F_I, in both cases resulting in reduced negative bias when survey data were added.

The overall importance of using the environmental index, combining cases where survey data were used or not, was measured with the contrast "2.S.E : 2.nS.E – 2.S.nE : 2.nS.nE". This contrast ranked 17th given the mixed exploitation history level and produced a 7.7% reduction in negative bias when the environmental data were used, but it ranked over 20 for F_C and F_I. The contrast

Table 4. Top 20 contrasts for model coefficients for the relative error of SSB_0 estimates, following a similar arrangement and format as Table 3.

Rank	Contrasts	Factor: level	ExpRate	SpStru	EnvInd	SurData	Coefficient
1	2.S.nE : 1.S – 2.nS.nE : 1.nS	F_M				A	0.0889
2	2.S.E : 2.nS.E – 1.S : 1.nS	F_M		A	A		0.0807
3	2.S.nE : 1.S – 2.nS.nE : 1.nS	F_C				A	0.0669
4	$F_I - F_M$	$MC_{2.nS.nE}$	×				0.0591
5	$F_I - F_M$	$MC_{1.nS}$	×				0.0526
6	2.S.nE – 2.nS.nE	F_M				A	0.0481
7	2.S.E : 2.nS.E – 1.S : 1.nS	F_M		A			0.0475
8	2.S.E : 2.nS.E – 1.S : 1.nS	F_I		A	A		0.0473
9	$F_C - F_I$	$MC_{2.nS.nE}$	×				0.0458
10	2.S.nE : 2.nS.nE – 1.S : 1.nS	F_I		A			0.0443
11	1.S – 1.nS	F_M				A	0.0408
12	2.S.nE – 2.nS.nE	F_C				A	0.0382
13	$F_C - F_I$	$MC_{2.nS.E}$	×				0.0381
14	$F_M - F_C$	$MC_{1.nS}$	×				0.0380
15	$F_I - F_M$	$MC_{2.nS.E}$	×				0.0338
16	2.S.E : 2.nS.E – 2.S.nE : 2.nS.nE	F_M			A		0.0333
17	1.S – 2.nS.nE	F_C		A		R	0.0333
18	2.S.nE : 1.S – 2.nS.nE : 1.nS	F_I				A	0.0332
19	$F_M - F_C$	$MC_{1.S}$	×				0.0329
20	1.S – 1.nS	F_C				A	0.0287

Note: The table does not include a column for EnvirP because contrasts between levels of that factor had ranks greater than 20.

"2.S.E : 2.nS.E – 1.S : 1.nS", which compares two-region models having the environmental index with a one-region models (no environmental index), ranked second for F_M and 16th for $F_{I,}$ reducing negative bias by 18.2% and 7.8%, respectively, compared with the one-region configurations. For the third exploitation level, this contrast did not rank among the top 20.

The three exploitation histories had relatively large influences on the RE of $SSB_{Current}$, appearing as main contrasts (e.g., "$F_I - F_M$") and as relevant factor levels in six of the top 20 contrasts (ranks 3, 5, 6, 10, 12, and 14). The first five of these rankings occurred in contrasts having no survey data ($MC_{2.nS.nE}$ and $MC_{1.nS}$); the rank 14 contrast occurred with the one-region MC with survey data. Compared with ExpRate, the three levels for EnvirP had relatively little influence. None of the six contrasts among the EnvirP levels (e.g., "$E_C - E_A$") ranked among the top 20.

Linear hypothesis tests for RE(SSB_0)

The contrast coefficients for RE(SSB_0) (Table 4) are much smaller than the coefficients for RE($SSB_{Current}$), implying that the experimental factors have much less effect on the estimation of SSB_0 than on the estimation of $SSB_{Current}$. The contrast "2.S.nE : 2.nS.nE – 1.S : 1.nS", which combines cases where survey data were used or not when the environmental index is not used, ranked 10th when the exploitation history is increasing (F_I), and the two-region models produced less-biased estimates of SSB_0 compared with the corresponding one-region versions. Contrasts involving SpStru had greater importance in association with other factors. For instance, the contrast "2.S.E : 2.nS.E – 1.S : 1.nS", which tests the spatial assumption plus the use of the environmental index, ranked second with exploitation history F_M and ranked eigth with exploitation history F_I. This contrast also had a large influence (rank 2 in Table 3) on the estimates of $SSB_{Current}$ for F_M, but was otherwise not among the top 20.

Contrasts evaluating the importance for RE(SSB_0) of using the survey data accounted for eight of the top 20 contrasts and occurred across all ExpRate levels. The contrast "2.S.nE : 1.S – 2.nS.nE : 1.nS", which gauges the importance of using survey data for either spatial assumption, ranked first for $F_{M,}$ third for $F_{C,}$ and 18th for F_I producing reductions in negative bias of 8.9%, 6.7%, and 3.3%, respectively, when survey data were used. Contrasts having a two-region model with survey biomass or an environmental index (2.S.nE – 2.nS.nE) ranked sixth for F_M and 12th for F_C. Survey data also reduced bias in one-region models (1.S – 1.nS) for two of the ExpRate levels (F_C and F_M). The contrast evaluating the importance of survey data when we collapsed the spatial structure (1.S – 2.nS.E) ranked 17th under ExpRate F_C.

Contrasts that evaluated the effect on RE(SSB_0) of using the environmental index accounted for three of the top 20 contrasts and occurred across two ExpRate levels (F_M and F_I). The contrast "2.S.E : 2.nS.E – 2.S.nE : 2.nS.nE", which examines the importance of the environmental index regardless of whether survey data are used, ranked 16th for F_M and resulted in a 3.3% reduction in negative bias.

Among the top 20 contrasts there were seven contrasts among the ExpRate levels and zero of six possible contrasts among EnvirP levels. The contrast "$F_I - F_M$" ranked fourth for the two-region model with no survey data ($MC_{2.nS.nE}$) and fifth for the one-region model with no survey data ($MC_{1.nS}$). In the first case, the increasing exploitation history F_I resulted in a 5.9% reduction in negative bias relative to the mixed exploitation history F_M. The contrast "$F_M - F_C$" ranked 14th for $MC_{1.nS}$ and 19th for the MC with survey data ($MC_{1.S}$). The contrast "$F_C - F_I$" only ranked in the top 20 contrasts for $MC_{2.nS.nE}$. The poor rankings for the contrasts among EnvirP levels indicate that they are generally of little importance to the RE values for either SSB_0 or $SSB_{Current}$ compared with the other factors that were tested in the analysis.

Discussion and conclusions

This study set out to determine the influence of the following factors on stock synthesis estimates for a spatially structured population: (*i*) underlying processes driving the population dynamics (regional exploitation rates and regional recruitment distribution); (*ii*) the use of survey biomass and age-composition data as well as an environmental index for the regional recruitment distribution; and (*iii*) whether the synthesis models assume the correct spatial structure (one region versus two regions). Results indicated that bias in the estimation of $SSB_{Current}$ and SSB_0 depends in a complex manner on the specific combinations of the experimental factors. However, some general patterns were evident:

(1) Using survey data (biomass index and age-composition observations) reduced bias in estimates of SSB_0 and $SSB_{Current}$ in all OM scenarios.
(2) Including an environmental index sometimes reduced the estimation bias, but the direction and magnitude of the effect

depended on whether or not survey data were used, and they varied among OM scenarios with no clear pattern.

(3) The exploitation history experienced by the regional subpopulations had a strong influence on estimation bias in some cases and was often stronger than the effect of the environmental driver of the regional recruitment distribution.

(4) The estimates of $SSB_{Current}$ and SSB_0 were negatively biased and for most treatments the estimates of $SSB_{Current}$ were more biased than the estimates of SSB_0.

(5) The estimates of $SSB_{Current}$ were more variable than the estimates of SSB_0.

(6) Including the bias adjustment for recruitment in the estimation model slightly magnified the negative bias in estimates of SSB_0 and $SSB_{Current}$.

The reduced bias in the models with survey data could be due to the survey providing improved information on the abundance of younger fish. In our experiment, the survey selected younger fish than the fishery and thus provided more observations for gauging the strength of incoming year classes, similar to the findings of Ono et al. (2015). The improved estimates of recruitment in turn led to improved estimates of spawning biomass. Our finding of reduced bias in estimates of biomass when survey data were used is also consistent with results from Cope and Punt (2011), who found that performance of all of their assessment models deteriorated when the survey data were removed. Including an environmental index to supplement the survey data provides information on the number of recruits to each region and supports the estimates of subsequent spawning biomass resulting from the recruitment. Surprisingly, incorporating the bias adjustment for recruitment estimations, as suggested by Methot and Taylor (2011), did not reduce bias in the estimates of biomass but instead slightly magnified the negative bias.

The different environmental patterns (abrupt, gradual, or constant) forcing the regional recruitment distribution had relatively small impacts on estimation bias in $SSB_{Current}$ and SSB_0. In hindsight, this lack of a strong environmental effect is not surprising. Unlike the overall level of recruitment, which has a direct but delayed impact on stock biomass, the regional distribution in recruitment plays a relatively minor role in the overall population dynamics. Where recruits settle becomes an important feature only if there are large regional differences in mortality. If we had used more extreme temporal changes in the fractional distribution of recruits or larger differences between regions in the exploitation histories, we probably would have seen greater sensitivity to the environmental forcing. In a one-region simulation study, Haltuch et al. (2009) also found that different patterns of environmental variability did not have a strong effect on assessment performance.

Differences in exploitation histories between subpopulations had an important impact on estimation bias, as illustrated by the generally high ranks found for the contrasts between the exploitation rate patterns for the estimates of SSB_0 and $SSB_{Current}$. This result is similar to the findings by Cope and Punt (2011) that relatively small differences in regional catch histories could create spatial structure and impact estimates of depletion if spatial structure is not correctly modeled.

We expected that some of the assessment MCs would produce unbiased estimates of spawning biomass, especially configuration $MC_{2.S.E}$, where the spatial structure matched that of the OM and the assessment model used all the available data. Surprisingly the $MC_{2.S.E}$ models underestimated SSB_0 and $SSB_{Current}$ on average for all combinations of the experimental factors, even though we fixed steepness (*h*), natural mortality (*M*), and the masses-at-age at their true values. This was also the result when we used the recruitment bias correction in the mini-experiment. In contrast, when we provided the $MC_{2.S.E}$ models with perfect data, the models were able to exactly fit the data and recover the true parameter values, indicating that the assessment models correctly reflected the underlying population dynamics. Some unknown aspect of the data variability or the assessment model caused the negative estimation bias, but uncovering its source was beyond the scope of this study.

There is no theory to support the notion that maximum likelihood estimation, such as used in SS, will in general produce unbiased parameter estimates, especially when sample sizes are relatively small (Hogg and Craig 1970). Further, other studies that simulation-tested age-structured stock assessment approaches provide evidence that stock assessment estimates and derived quantities can be biased, even when the assessment model has the same structure as the operating model and uses the same assumptions for the observation error structure and variability of the generated observations (e.g., Ono et al. 2015; Punt et al. 2015). Ono et al. (2015) reported that estimated model parameters "had low bias (MRE below 4%) for the Base case", in which the configuration of the estimation model matched the configuration of the OM. The study used SS both to generate the random observational data and also to fit the data, thereby eliminating the possibility that discrepancies between the OM and estimation model caused the bias in the base case. In the current study, the maximum MRE values across all focal quantities, OMs, and MCs are also relatively low (below 4%). Punt et al. (2015) also included a case where the estimation model matched the operating model (described as the "FULL" configuration). Although they did not report numerical values for bias, graphical results (in their figure 5, panel *g*) indicate median relative errors of about 5% during the first 10 years of the assessment period.

In studies such as these and ours, the inclusion of a "control" or "base" model provides a set of reference estimates against which experimental treatments can be compared to draw inferences about the strength of the treatments, even if the control is biased. While having slight bias in results from stock assessments is not a desirable feature, it may be unavoidable.

Our experiment, which did not manipulate any biological parameters in the OM, represented a long-lived species with relatively slow population dynamics and moderate variability in recruitment. To evaluate the generality of our results, this experiment could be repeated with the OM tuned to the biology of a shorter-lived species with faster dynamics. Haltuch et al. (2009) reported that high variability in Pacific hake (*Merluccius productus*) recruitment resulted in more variable dynamics and therefore more variable estimates of SSB_0. We hypothesize that with more variable recruitment in the OM, there would be greater contrast in the resulting spawning biomass that in turn would allow the synthesis models to better estimate recruitment and biomass.

Our study design considered two simple factors that affect spatial structure: regional recruitment distribution and exploitation history. Other simulation studies, such as Cope and Punt (2011) and Berger et al. (2012), have taken a similar approach. Alternative mechanisms could also result in spatial structure, and it seems likely that they could degrade stock assessment performance. For example, subpopulations could have regionally varying growth, as is evident from commercial fishery data for herring in the northern Baltic Sea (*Clupea harengus membras*) (Rahikainen and Stephenson 2004). The effects of this phenomenon on assessment results was explored in a simulation study based on pink ling in Australia (*Genypterus blacodes*) (Punt et al. 2015). Even more complex spatial structure would arise if there was density-dependent predation on recently settled fish, leading to natural mortality rates that varied regionally and temporally. An additional factor that could influence a population's dynamics and might influence the performance of an assessment is whether density-dependent effects on the recruitment relationship operate at the regional level rather than at the level of the population (as our OM assumed).

Another real-world phenomenon that we did not include in our operating model is the movement between regions of postsettlement recruits. This was a focus of the study by Guan et al. (2013) that assessed the spatial structure of US Atlantic herring (*Clupea harengus*) as they migrated among four management areas to locate feeding grounds and respond to environmental changes. Many fish species also exhibit complex ontogenetic movements, such as settling in shallow water and moving to deeper water as fish mature. Such a phenomenon could cause selection-at-age to change if fishing mortality is not uniform with depth (Sampson and Scott 2011) and would likely result in poor performance of stock assessment models that were not appropriately configured for such spatial structure (Goethel and Berger 2017). This is also evident in the work of Hurtado-Ferro et al. (2014), who suggest that addressing movement by using multiple fleets with different selectivity curves could address the uncertainty in spatial structure caused by movement of fish.

This study provides a starting point for investigating factors that influence the performance of spatially structured stock assessments, specifically data availability and recruitment distribution. The results indicate that in some circumstances ignoring spatial structure tends to degrade assessment performance. Whether improvements in performance are sufficient to justify the additional complexities that arise when constructing a spatially structured assessment remains an unresolved question. When considering how to include spatial structure in a stock assessment, the assessment analyst must decide how many spatial regions to include and must set spatial boundaries for partitioning data to the regions. Data on oceanographic features or bottom substrate may provide information to support the partitioning. Decisions regarding the number of regions and the placement of boundaries are likely to influence performance of the assessment. Having data available that can be appropriately structured may be a key limitation. It is relatively simple to develop spatially distinct survey age-composition, biomass data, and environmental indices, given that this information is usually recorded at a fine spatial scale. In contrast, developing a data series of spatially resolved fishery catches and age compositions can be a major or even insurmountable challenge because fishery monitoring systems generally collect information at coarse spatial scales.

Acknowledgements

We thank Richard Methot and the Stock Assessment team at the NOAA Northwest Science Center (NWFSC) for sharing their expertise in stock synthesis. We also thank Owen Hamel, Aaron Berger (NWFSC), André Punt (University of Washington), and one anonymous reviewer for their helpful feedback on earlier drafts of this manuscript. We are also grateful to the NOAA Living Marine Cooperative Science Center for providing student funding for L. Denson. This paper is based on her thesis submitted in partial fulfillment of the requirements for a Master of Science degree from Oregon State University.

References

Ainley, D.G., Sydeman, W., Parrish, R., and Lenarz, W. 1993. Oceanic factors influencing distribution of young rockfish (*Sebastes*) in central California: a predator's perspective. CalCOFI Rep. **34**: 133–139.

Begg, G.A., and Waldman, J.R. 1999. An holistic approach to fish stock identification. Fish. Res. **43**(1–3): 35–44. doi:10.1016/S0165-7836(99)00065-X.

Berger, A.M., Jones, M.L., Zhao, Y., and Bence, J.R. 2012. Accounting for spatial population structure at scales relevant to life history improves stock assessment: The case for Lake Erie walleye *Sander vitreus*. Fish. Res. **115–116**: 44–59. doi:10.1016/j.fishres.2011.11.006.

Bograd, S.J., and Lynn, R.J. 2003. Long-term variability in the Southern California Current System. Deep Sea Res. Part II Top. Stud. Oceanogr. **50**(14–16): 2355–2370. doi:10.1016/S0967-0645(03)00131-0.

Bretz, F., Hothorn, T., and Westfall, P. 2011. Multiple comparisons using R. CRC Press Taylor and Francis Group, Boca Raton, Fla.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multi-model inference: a practical information-theoretic approach. 2nd ed. Springer, Secaucus, New Jersey, USA.

Charter, S., and Sandknop, E. 2000. Abundance and distribution of rockfish (*Sebastes*) larvae in the Southern California Bight in relation to environmental conditions and fishery exploitation. Calif. Coop. Oceanic Fish. Invest. Rep. **41**: 132–147.

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.

Goethel, D.R., and Berger, A.M. 2017. Accounting for spatial complexities in the calculation of biological reference points: effects of misdiagnosing population structure for stock status indicators. Can. J. Fish. Aquat. Sci. **74**(11). [This issue.] doi:10.1139/cjfas-2016-0290.

Green, K.M., and Starr, R.M. 2011. Movements of small adult black rockfish: implications for the design of MPAs. Mar. Ecol. Progr. Ser. **436**: 219–230. doi:10.3354/meps09263.

Guan, W.J., Cao, J., Chen, Y., and Cieri, M. 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.

Haltuch, M.A., Punt, A.E., and Dorn, M.W. 2008. Evaluating alternative estimators of fishery management reference points. Fish. Res. **94**(3): 290–303. doi:10.1016/j.fishres.2008.01.008.

Haltuch, M.A., Punt, A.E., and Dorn, M.W. 2009. Evaluating the estimation of fishery management reference points in a variable environment. Fish. Res. **100**(1): 42–56. doi:10.1016/j.fishres.2009.03.001.

Haltuch, M.A., Punt, A.E., and Walters, C. 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Can. J. Fish. Aquat. Sci. **68**(5): 912–926. doi:10.1139/f2011-030.

Hare, S.R., and Mantua, N.J. 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. Progr. Oceanogr. **47**(2–4): 103–145. doi:10.1016/S0079-6611(00)00033-1.

Hilborn, R., Parma, A., and Maunder, M. 2002. Exploitation rate reference points for west coast rockfish: Are they robust and are there better alternatives? North Am. J. Fish. Manage. **22**(1): 365–375. doi:10.1577/1548-8675(2002)022<0365:ERRPFW>2.0.CO;2.

Hogg, R.V., and Craig, A.T. 1970. Introduction to mathematical statistics. Macmillan Publishing Co. Inc., New York.

Hulson, P.-J.F., Quinn, T.J., II, Hanselman, D.H., and Ianelli, J.N. 2013. Spatial modeling of Bering Sea walleye pollock with integrated age-structured assessment models in a changing environment. Can. J. Fish. Aquat. Sci. **70**(9): 1402–1416. doi:10.1139/cjfas-2013-0020.

Hurtado-Ferro, F., Punt, A.E., and Hill, K.T. 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fish. Res. **158**: 102–115. doi:10.1016/j.fishres.2013.10.001.

Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? Fish Fish. **8**(4): 337–358. doi:10.1111/j.1467-2979.2007.00258.x.

Mangel, M., MacCall, A.D., Brodziak, J., Dick, E.J., Forrest, R.E., Pourzand, R., and Ralston, S. 2013. A perspective on steepness, reference points, and stock assessment. Can. J. Fish. Aquat. Sci. **70**(6): 930–940. doi:10.1139/cjfas-2012-0372.

McGilliard, C.R., Punt, A.E., Methot, R.D., and Hilborn, R. 2015. Accounting for marine reserves using spatial stock assessments. Can. J. Fish. Aquat. Sci. **72**(2): 262–280. doi:10.1139/cjfas-2013-0364.

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.

Methot, R.D., and Wetzel, C.R. 2013. Stock Synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.

National Research Council. 1998. Improving fish stock assessments. National Academies Press.

Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F., Johnson, K.F., McGilliard, C.R., Monnahan, C.C., Szuwalski, C.S., Valero, J.L., Vert-Pre, K.A., Whitten, A.R., and Punt, A.E. 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES J. Mar. Sci. **72**(1): 31–43. doi:10.1093/icesjms/fsu007.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and the R Core Team. 2016. nlme: linear and nonlinear mixed effects models. *In* R Packages.

Punt, A.E., Smith, A.D.M., and Cui, G.R. 2002. Evaluation of management tools for Australia's South East Fishery 1. Modelling the South East Fishery taking account of technical interactions. Mar. Freshw. Res. **53**(3): 615–629.

Punt, A.E., Haddon, M., and Tuck, G.N. 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment? A case study based on pink ling in Australia. Fish. Res. **168**: 85–99. doi:10.1016/j.fishres.2015.04.002.

Punt, A.E., Haddon, M., Little, L.R., and Tuck, G.N. 2016. Can a spatially-structured stock assessment address uncertainty due to closed areas? A case study based on pink ling in Australia. Fish. Res. **175**: 10–23. doi:10.1016/j.fishres.2015.11.008.

Rahikainen, M., and Stephenson, R.L. 2004. Consequences of growth variation in northern Baltic herring for assessment and management. ICES J. Mar. Sci. **61**(3): 338–350. doi:10.1016/j.icesjms.2004.02.005.

Rubinstein, R.Y., and Kroese, D.P. 2008. Simulation and the Monte Carlo method. John Wiley & Sons.

Sampson, D.B., and Scott, R.D. 2011. A spatial model for fishery age-selection at the population level. Can. J. Fish. Aquat. Sci. **68**(6): 1077–1086. doi:10.1139/f2011-044.

Shanks, A.L., and Eckert, G.L. 2005. Population persistence of California Current fishes and benthic crustaceans: a marine drift paradox. Ecol. Monogr. **75**(4): 505–524. doi:10.1890/05-0309.

Taylor, I., Stewart, I., Hicks, A., Garrison, T., Punt, A., Wallace, J., Wetzel, C., Thorson, J., Takeuchi, Y., Monnahan, C., et al. 2014. r4ss: R code for Stock Synthesis. R package version 1.22.1.