Canadian Journal of Fisheries and Aquatic Sciences
Journal canadien des sciences halieutiques et aquatiques

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

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2016-0277.R3
Manuscript Type:	Article
Date Submitted by the Author:	25-Jun-2017
Complete List of Authors:	Denson, LaTreese; University of Miami Rosenstiel School of Marine and Atmospheric Science, Marine Biology and Ecology Sampson, David; Oregon State University, Coastal Oregon Marine Experiment Station Stephens, Andi; Northwest Fisheries Science Center NOAA Fisheries
Is the invited manuscript for consideration in a Special Issue? :	AFS Spatial Processes
Keyword:	Stock Synthesis, spatial stock assessment, simulation testing, RECRUITMENT < General, spatial distribution

SCHOLARONE™
Manuscripts

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

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

LaTreese S. Denson¹, corresponding author

Department of Fisheries and Wildlife

Oregon State University

Nash Hall, 2820 SW Campus Way

Corvallis, OR 97331

Phone: (917) 605-4615

ldenson@rsmas.miami.edu

David 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

david.sampson@oregonstate.edu

Andi Stephens

Northwest Fisheries Science Center NOAA Fisheries

Newport Research Station

2032 S.E. OSU Drive

Newport, OR 97365

Andi.stephens@noaa.gov

¹ Current Affiliation

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

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, and whether an environmental index is available to inform the spatial distribution of recruitment. Stocks with limited movement of post-settlement fish may be spatially structured due to environmental forces that affect larval dispersal and recruitment distribution or from non-uniform 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 which simulates data that are used in a set of estimation models to assess bias. The following experimental factors are considered: (a) using survey data and environmental indices in the assessment; (b) using disaggregated data (two regions, as generated by the OM) or aggregated data (one region); and (c) incorporating different patterns in the OM's regional exploitation and environmentally driven recruitment distribution.

Keywords

Stock synthesis, spatial stock assessment, recruitment, spatial distribution, simulation testing

44 *Introduction*

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

58
59 Many fish populations are made up of spatially dispersed subpopulations that receive recruits
60 from a common pool but that have limited exchanges of individuals post-recruitment. These
61 subpopulations experience different catch and management histories and therefore vary in their
62 age-structure (Begg and Waldman 1999; National Research Council 1998). These spatial
63 subpopulations are often identifiable in fishery age- and size-composition data. Subpopulation
64 structure not only stems from regional differences in fishing patterns, such as those that occur at
65 the closing of an area for protection (McGilliard et al. 2015; Punt et al. 2016), but also from
66 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: 1) relatively small differences in catch histories among spatial subpopulations can create bias in estimates of stock status and 2) 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 (Methot and Wetzel 2013a), an age-structured stock assessment tool that can be configured to include spatial structure and which 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

The operating model (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 2013a; Methot and Wetzel 2013b; 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 Supplementary Data.

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 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 (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-six fish experiencing half the full rate of fishing mortality and age-eight 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 yr^{-1} in region one and 0.055 yr^{-1} in region two, which results in the expected total spawning biomass in the first year of each simulation to be 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% ($\pm 2\%$) of the unfished level when the exploitation history is constant or mixed and about 25% ($\pm 0.8\%$) 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.

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 mid-year weights-at-age. The data for the fishery catches, the environmental index, and the mid-year weights-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.

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 2013a) was used to estimate sets of parameter values for six model configurations (described below) that were fit to simulated data for each replicate. The stock synthesis 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 weights-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, stock synthesis 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 stock synthesis uses environmental indices are provided in Methot and Wetzel (2013ab). 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.

Stock synthesis 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 model configurations 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 model configuration in which the stock synthesis 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 stock synthesis were consistent.

Evaluating Convergence

Stock synthesis 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 stock synthesis 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.

Estimation Model Configurations

In addition to the two factors that control the population dynamics (*EnvirP* and *ExpRate*), the experimental design (Table 1) also evaluated: 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 model configurations (MCs, Table 2).

One model 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.

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 Operating Model to Estimation Model Outputs

To evaluate the performance of the six estimation model configurations, 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 to their true values using the relative error (RE),

$$(1) \quad RE = \frac{\hat{\theta} - \theta}{\theta}$$

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

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 model configurations 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 (Figures 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.

$$(2) \quad RE \text{ Full Model} \sim EnvirP + ExpRate + MC + EnvirP * ExpRate + \\ EnvirP * MC + ExpRate * MC + EnvirP * ExpRate * MC$$

Mini-experiment to Explore Bias Adjustments for Annual Recruitment Deviations

Previous research suggests that stock synthesis 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 stock synthesis 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 operating model, 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 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 an average 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 to their true values by calculating the

relative errors (RE), which were then compared to 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 relative error 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 model:

$$(3) \quad \Delta_{Bias} = Scenario * MC$$

Results

Convergence

All 100 replicates for each model configuration 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.

Relative Error 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 weight-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 Figures 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 model configurations. The thin grey horizontal lines in each panel connect the replicates from MC to MC. Non-parallel 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 dots 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 to the other MCs. Configurations $MC_{2,S,nE}$ and $MC_{1,S}$ show similar reduced negative bias relative to counterparts $MC_{2,nS,E}$ 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 model configurations 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 $MC_{1,nS}$ 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 model configurations to influence the bias and accuracy of the spawning biomass estimates (e.g., Figures 1, S.2 and S.3).

403 *Mini-Experiment: Recruitment Deviation Bias-Adjustment Ramp Effect on Bias*

404 Line plots illustrate the differences between relative error values when the bias-adjustment is
405 used (Ramp) or not used (No Ramp). Figure 5 shows these differences for OM scenarios $F_C:E_C$
406 and $F_M:E_A$ across the six model configurations. Similar to the previous line-plots the grey lines
407 connect replicates from estimates with the bias-adjustment to estimates without the bias-
408 adjustment. The large solid dots mark the median RE values for each treatment.

409

410 The bias-adjustment had a minimal effect on the bias in the estimation of $SSB_{Current}$, and for
411 almost all replicates the effect was to slightly accentuate the negative bias. The large solid dots
412 representing the MRE values are all either at or below the horizontal zero reference lines,
413 indicating an overall tendency to underestimate $SSB_{Current}$, regardless of whether the bias
414 adjustment ramp is used or not, and the median values from the estimates based on the bias-
415 adjustment are slightly lower than the estimates from models that did not include the bias-
416 adjustment. For scenario $F_C:E_C$ the average increase in negative bias was 1.07%; for scenario
417 $F_M:E_A$ the average increase in bias was even smaller (0.72%).

418

419 The bias-adjustment ramp had a greater effect on estimates of SSB_0 than it did on estimates of
420 $SSB_{Current}$. For scenario $F_C:E_C$ the average increase in negative bias was 1.66%; for scenario
421 $F_M:E_A$ the average increase in bias was smaller (1.47%). Again, using the bias-adjustment ramp
422 accentuated the negative bias.

423

Visually, the average differences between estimation methods seem trivial, but the factorial ANOVA results indicated significant differences between “Ramp” and “No Ramp” estimates and significant Scenario and Model Configuration effects (p value < 0.01). Model Configuration had a larger effect than Scenario on the difference in RE values.

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:

$$(4) \quad RE = \beta_0 + \beta_1 EnvirP + \beta_2 ExpRate + \beta_3 MC + \beta_4 ExpRate * MC$$

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 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 model configuration 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 “2.S.nE : 1.S - 2.nS.nE : 1.nS”. This contrast ranked 1st 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 4th for F_C and 9th 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 “2.S.E : 2.nS.E – 1.S : 1.nS”, which compares two-region models having the environmental index with one-region models (no environmental index), ranked 2nd for F_M and

16th for F_I , reducing negative bias by 18.2% and 7.8% respectively, compared to 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 model configuration with survey data. Compared to *ExpRate* the three levels for *EnvirP* had relatively little influence. None of the six contrasts between 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 to 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 2nd with exploitation history F_M and ranked 8th 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.

492
 493 Contrasts evaluating the importance for $RE(SSB_0)$ of using the survey data accounted for eight of
 494 the top 20 contrasts and occurred across all *ExpRate* levels. The contrast “2.S.nE : 1.S –
 495 2.nS.nE : 1.nS”, which gauges the importance of using survey data for either spatial assumption,
 496 ranked 1st for F_M , 3rd for F_C , and 18th for F_I producing reductions in negative bias of 8.9% ,
 497 6.7% and 3.3% respectively when survey data were used. Contrasts having a two-region model
 498 with survey biomass or an environmental index (“2.S.nE – 2.nS.nE”) ranked 6th for F_M and 12th
 499 for F_C . Survey data also reduced bias in one region models (“1.S – 1.nS”) for two of the *ExpRate*
 500 levels (F_C and F_M). The contrast evaluating the importance of survey data when we collapsed the
 501 spatial structure (“1.S – 2.nS.E”) ranked 17th under *ExpRate* F_C .

502
 503 Contrasts that evaluated the effect on $RE(SSB_0)$ of using the environmental index accounted for
 504 three of the top 20 contrasts and occurred across two *ExpRate* levels (F_M and F_I). The contrast
 505 “2.S.E : 2.nS.E – 2.S.nE : 2.nS.nE”, which examines the importance of the environmental index
 506 regardless of whether survey data are used, ranked 16th for F_M and resulted in a 3.3% reduction
 507 in negative bias.

508
 509 Among the top 20 contrasts there were seven contrasts between the *ExpRate* levels and zero of
 510 six possible contrasts between *EnvirP* levels. The contrast “ $F_I - F_M$ ” ranked 4th for the two-region
 511 model with no survey data ($MC_{2.nS.nE}$) and 5th for the one-region model with no survey data
 512 ($MC_{1.nS}$). In the first case the increasing exploitation history F_I resulted in a 5.9% reduction in
 513 negative bias relative to the mixed exploitation history F_M . The contrast “ $F_M - F_C$ ” ranked 14th
 514 for $MC_{1.nS}$ and 19th for the model configuration 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 between *EnvirP* levels indicate that they are generally of little importance to the RE values for either SSB_0 or $SSB_{Current}$ compared to 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: 1) underlying processes driving the population dynamics (regional exploitation rates and regional recruitment distribution); 2) the use of survey biomass and age-composition data as well as an environmental index for the regional recruitment distribution; and 3) whether the synthesis models assume the correct spatial structure (one-region vs. 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 .

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

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

541
542 The reduced bias in the models with survey data could be due to the survey providing improved
543 information on the abundance of younger fish. In our experiment the survey selected younger
544 fish than the fishery and thus provided more observations for gauging the strength of incoming
545 year-classes, similar to the findings of Ono et al. (2015). The improved estimates of recruitment
546 in turn led to improved estimates of spawning biomass. Our finding of reduced bias in estimates
547 of biomass when survey data were used is also consistent with results from Cope and Punt
548 (2011), who found that performance of all of their assessment models deteriorated when the
549 survey data were removed. Including an environmental index to supplement the survey data
550 provides information on the number of recruits to each region and supports the estimates of
551 subsequent spawning biomass resulting from the recruitment. Surprisingly, incorporating the
552 bias-adjustment for recruitment estimations, as suggested by Methot and Taylor (2011), did not
553 reduce bias in the estimates of biomass but instead slightly magnified the negative bias.

554
555 The different environmental patterns (abrupt, gradual, or constant) forcing the regional
556 recruitment distribution had relatively small impacts on estimation bias in $SSB_{Current}$ and SSB_0 . In
557 hindsight, this lack of a strong environmental effect is not surprising. Unlike the overall level of
558 recruitment, which has a direct but delayed impact on stock biomass, the regional distribution in
559 recruitment plays a relatively minor role in the overall population dynamics. Where recruits
560 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 model configurations would produce unbiased estimates of spawning biomass, especially configuration $MC_{2,SE}$ where the spatial structure matched that of the OM and the assessment model used all the available data. Surprisingly the $MC_{2,SE}$ 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 weights-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 $MC_{2,SE}$ 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 stock synthesis, 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 (EM) matched the configuration of the operating model. The study used stock synthesis both to generate the random observational data and also to fit the data, thereby eliminating the possibility that discrepancies between the OM and EM 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 ten 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 operating model, 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 were 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 post-settlement 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 between 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 Dr. Richard Methot and the Stock Assessment team at the NOAA Northwest Science Center (NWFSC) for sharing their expertise in stock synthesis. We also thank Drs. 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.

Supplementary Data

Supplementary data associated with this article can be found in the online version at ...

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. Fisheries Research 43(1): 35-44.
- 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. Fisheries Research 115–116: 44-59.
- Bograd, S.J., and Lynn, R.J. 2003. Long-term variability in the Southern California Current System. Deep Sea Research Part II: Topical Studies in Oceanography 50(14–16): 2355-2370.
- Bretz, F., Hothorn, T., and Westfall, P. 2011. Multiple comparisons using R. CRC Press Taylor and Francis Group, Boca Raton, FL.
- Burnham, K.P., and Anderson, D.R. 2002. Model selection and multi-model inference : A practical information-theoretic approach (2nd Edition). Springer, Secaucus, NJ, 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. Fisheries Research 107(1–3): 22-38.

- 696 Goethel, D.R., and Berger, A.M. 2017. Accounting for spatial complexities in the calculation of
697 biological reference points: effects of misdiagnosing population structure for stock status
698 indicators. *Canadian Journal of Fisheries and Aquatic Sciences* (this issue).
- 699 Green, K.M., and Starr, R.M. 2011. Movements of small adult black rockfish: implications for
700 the design of MPAs. *Marine Ecology Progress Series* 436: 219-230.
- 701 Guan, W.J., Cao, J., Chen, Y., and Cieri, M. 2013. Impacts of population and fishery spatial
702 structures on fishery stock assessment. *Canadian Journal of Fisheries and Aquatic*
703 *Sciences* 70(8): 1178-1189.
- 704 Haltuch, M.A., Punt, A.E., and Dorn, M.W. 2008. Evaluating alternative estimators of fishery
705 management reference points. *Fisheries Research* 94(3): 290-303.
- 706 Haltuch, M.A., Punt, A.E., and Dorn, M.W. 2009. Evaluating the estimation of fishery
707 management reference points in a variable environment. *Fisheries Research* 100(1): 42-
708 56.
- 709 Haltuch, M.A., Punt, A.E., and Walters, C. 2011. The promises and pitfalls of including decadal-
710 scale climate forcing of recruitment in groundfish stock assessment. *Canadian Journal of*
711 *Fisheries and Aquatic Sciences* 68(5): 912-926.
- 712 Hare, S.R., and Mantua, N.J. 2000. Empirical evidence for North Pacific regime shifts in 1977
713 and 1989. *Progress in Oceanography* 47(2): 103-145.
- 714 Hilborn, R., Parma, A., and Maunder, M. 2002. Exploitation rate reference points for west coast
715 rockfish: Are they robust and are there better alternatives? *North American Journal of*
716 *Fisheries Management* 22(1): 365-375.
- 717 Hogg, R.V., and Craig, A.T. 1970. *Introduction to Mathematical Statistics*. Macmillan
718 Publishing Co. Inc., New York.

- 719 Hulson, P.-J.F., Quinn, T.J., Hanselman, D.H., and Ianelli, J.N. 2013. Spatial modeling of Bering
720 Sea walleye pollock with integrated age-structured assessment models in a changing
721 environment. *Canadian Journal of Fisheries and Aquatic Sciences* 70(9): 1402-1416.
- 722 Hurtado-Ferro, F., Punt, A.E., and Hill, K.T. 2014. Use of multiple selectivity patterns as a proxy
723 for spatial structure. *Fisheries Research* 158: 102-115.
- 724 Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? *Fish & Fisheries*
725 8(4): 337-358.
- 726 Mangel, M., MacCall, A.D., Brodziak, J., Dick, E.J., Forrest, R.E., Pourzand, R., and Ralston, S.
727 2013. A perspective on steepness, reference points, and stock assessment. *Canadian*
728 *Journal of Fisheries and Aquatic Sciences* 70(6): 930-940.
- 729 McGilliard, C.R., Punt, A.E., Methot, R.D., and Hilborn, R. 2015. Accounting for marine
730 reserves using spatial stock assessments. *Canadian Journal of Fisheries and Aquatic*
731 *Sciences* 72(2): 262-280.
- 732 Methot, R.D., and Taylor, I.G. 2011. Adjusting for bias due to variability of estimated
733 recruitments in fishery assessment models. *Canadian Journal of Fisheries and Aquatic*
734 *Sciences* 68(10): 1744-1760.
- 735 Methot, R.D., and Wetzel, C.R. 2013a. Stock synthesis: A biological and statistical framework
736 for fish stock assessment and fishery management. *Fisheries Research* 142: 86-99.
- 737 Methot, R.D., and Wetzel, C.R. 2013b. Stock Synthesis: A biological and statistical framework
738 for fish stock assessment and fishery management Appendix A. Supplementary data.
739 *Fisheries Research* 142.
- 740 National Research Council. 1998. Improving fish stock assessments. National Academies Press.

- 741 Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F.,
742 Johnson, K.F., McGilliard, C.R., Monnahan, C.C., Szuwalski, C.S., Valero, J.L., Vert-
743 Pre, K.A., Whitten, A.R., and Punt, A.E. 2015. The importance of length and age
744 composition data in statistical age-structured models for marine species. *ICES Journal of*
745 *Marine Science* 72(1): 31-43.
- 746 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and Team, R.C. 2016. nlme : Linear and
747 Nonlinear Mixed Effects Models. *In R Packages*.
- 748 Punt, A.E., Haddon, M., Little, L.R., and Tuck, G.N. 2016. Can a spatially-structured stock
749 assessment address uncertainty due to closed areas? A case study based on pink ling in
750 Australia. *Fisheries Research* 175: 10-23.
- 751 Punt, A.E., Haddon, M., and Tuck, G.N. 2015. Which assessment configurations perform best in
752 the face of spatial heterogeneity in fishing mortality, growth and recruitment? A case
753 study based on pink ling in Australia. *Fisheries Research* 168: 85-99.
- 754 Punt, A.E., Smith, A.D.M., and Cui, G.R. 2002. Evaluation of management tools for Australia's
755 South East Fishery 1. Modelling the South East Fishery taking account of technical
756 interactions. *Marine and Freshwater Research* 53(3): 615-629.
- 757 Rahikainen, M., and Stephenson, R.L. 2004. Consequences of growth variation in northern
758 Baltic herring for assessment and management. *ICES Journal of Marine Science: Journal*
759 *du Conseil* 61(3): 338-350.
- 760 Rubinstein, R.Y., and Kroese, D.P. 2008. Simulation and the Monte Carlo method. John Wiley
761 & Sons.
- 762 Sampson, D.B., and Scott, R.D. 2011. A spatial model for fishery age-selection at the population
763 level. *Canadian Journal of Fisheries and Aquatic Sciences* 68(6): 1077-1086.

- 764 Shanks, A.L., and Eckert, G.L. 2005. Population persistence of California Current fishes and
765 benthic crustaceans: a marine drift paradox. *Ecological Monographs* 75(4): 505-524.
- 766 Taylor, I., Stewart, I., Hicks, A., Garrison, T., Punt, A., Wallace, J., Wetzel, C., Thorson, J.,
767 Takeuchi, Y., Monnahan, C., and contributors, o. 2014. r4ss: R code for Stock Synthesis.
768 R package version 1.22.1.

769

Draft

Figure Captions

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

Figure 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 (MC) described in the text. The short-dash vertical line at 0.0 indicates no bias. The long-dash vertical line represents the median relative error (MRE) across all OM scenarios and replicates for each specific MC, the MRE values are located in the top-left corner of each panel.

Figure 3. Distribution of relative error values for estimates of SSB_0 and $SSB_{Current}$ arranged by *EnvirP* pattern. Short-dash vertical lines at 0.0 indicate no bias. The long-dash vertical lines represent the median relative error (MRE) across all OM scenarios and model configurations for each *EnvirP* pattern, the MRE values are located in the top-left 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.

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

Figure 5. Line-plots of relative error values for $SSB_{Current}$ from estimation models with the bias-adjustment ramp (Ramp) and without the bias-adjustment ramp (No_Ramp). The top row displays the change in RE under OM scenario F_c : E_c and the bottom row represents the change in RE under OM scenario F_M : E_A . The effect of the ramp is also displayed across all six model configurations combined with the two OM scenarios. The two large solid circles 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 1. Description of the five experimental factors, their patterns and levels, and whether they affect the operating model or the estimation model. To simplify the analyses and displays of results, the three estimation model factors were combined into a single model configuration factor (MC), described in the text.

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

Table 2. Key components of the Operating Model (OM) compared to the different Model Configurations (MC) used to set up the stock synthesis (SS) estimation models. The six stock synthesis 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.

	Fishery Regions	Survey Regions	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

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

		Factor:					
Rank	Contrast	Level	<i>ExpRate</i>	<i>SpStru</i>	<i>EnvInd</i>	<i>SurData</i>	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}	X				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}	X		0.1158	
6	F _M – F _C	MC _{1.nS}	X		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}	X		0.0954	
11	1.S - 1.nS	F _M		A	0.0932	
12	F _C - F _I	MC _{2.nS.nE}	X		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}	X		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
----	----------------	-------	---	--------

Draft

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

Rank	Contrasts	Factor: Level	<i>ExpRate</i>	<i>SpStru</i>	<i>EnvInd</i>	<i>SurData</i>	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}	X				0.0591
5	F _I - F _M	MC _{1.nS}	X				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}	X				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	$F_C - F_I$	$MC_{2.nS.E}$	X		0.0381
14	$F_M - F_C$	$MC_{1.nS}$	X		0.0380
15	$F_I - F_M$	$MC_{2.nS.E}$	X		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	$F_M - F_C$	$MC_{1.S}$	X		0.0329
20	1.S - 1.nS	F_C		A	0.0287

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

254x190mm (96 x 96 DPI)

Figure 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 (MC) described in the text. The short-dash vertical line at 0.0 indicates no bias. The long-dash vertical line represents the median relative error (MRE) across all OM scenarios and replicates for each specific MC, the MRE values are located in the top-left corner of each panel.

254x190mm (96 x 96 DPI)

Figure 3. Distribution of relative error values for estimates of SSB_0 and $SSB_{Current}$ arranged by *EnvirP* pattern. Short-dash vertical lines at 0.0 indicate no bias. The long-dash vertical lines represent the median relative error (MRE) across all OM scenarios and model configurations for each *EnvirP* pattern, the MRE values are located in the top-left 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.

254x190mm (96 x 96 DPI)

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

254x190mm (96 x 96 DPI)

Figure 5. Line-plots of relative error values for $SSB_{Current}$ from estimation models with the bias-adjustment ramp (Ramp) and without the bias-adjustment ramp (No_Ramp). The top row displays the change in RE under OM scenario $F_c: E_c$ and the bottom row represents the change in RE under OM scenario $F_m: E_A$. The effect of the ramp is also displayed across all six model configurations combined with the two OM scenarios. The two large solid circles 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.

254x190mm (96 x 96 DPI)