Fisheries Research 206 (2018) 65–78

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

When are estimates of spawning stock biomass for small pelagic fishes improved by taking spatial structure into account?

André E. Punt[a,*], Daniel K. Okamoto[a,b], Alec D. MacCall[c], Andrew O. Shelton[d], Derek R. Armitage[e], Jaclyn S. Cleary[f], Ian P. Davies[g], Sherri C. Dressel[h], Tessa B. Francis[i], Phillip S. Levin[g,j], R. Russ Jones[k], Harvey Kitka[l], Lynn Chi Lee[m], Jim A. McIsaac[n], Melissa R. Poe[d,o], Steve Reifenstuhl[p], Jennifer J. Silver[q], Jörn O. Schmidt[r], Thomas F. Thornton[s], Rudiger Voss[r], John Woodruff[t]

[a] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195, USA
[b] Department of Biological Sciences, Florida State University, USA
[c] Farallon Institute for Advanced Ecosystem Research, 101 H. Street, Suite Q., Petaluma, CA, 94952, USA
[d] Northwest Fisheries Science Center, National Marine Fisheries Service, 2725 Montlake Boulevard East, Seattle, WA 98112, USA
[e] School of Environment, Resources and Sustainability, University of Waterloo, Waterloo, ON N2L 3G, Canada
[f] Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada
[g] School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195, USA
[h] Division of Commercial Fisheries, Alaska Department of Fish and Game, Juneau, AK 99811, USA
[i] Puget Sound Institute, University of Washington Tacoma, Tacoma, WA, 98421, USA
[j] Nature Conservancy in Washington, 74 Wall St., Seattle, WA, 98121, USA
[k] Haida Oceans Technical Team, Council of the Haida Nation, PO Box 98, Queen Charlotte, BC V0T 1S1, Canada
[l] Sitka Tribe of Alaska, Sitka, AK 99835, USA
[m] Gwaii Haanas National Park Reserve, National Marine Conservation Area Reserve, and Haida Heritage Site, Canada
[n] T. Buck Suzuki Foundation, Victoria, BC V8Z 4B8, Canada
[o] Washington Sea Grant, University of Washington, Seattle WA 98105, USA
[p] Northern Southeast Regional Aquaculture Association, Sitka, AK 99835, USA
[q] Department of Geography, University of Guelph, 50 Stone Road East, Guelph, ON, N1G 2W1, Canada
[r] Sustainable Fisheries, Department of Economics, Kiel University, 24118 Kiel, Germany
[s] Environmental Change Institute, School of Geography & the Environment, University of Oxford, Oxford OX1 3QY, UK
[t] Icicle Seafoods Inc., 4019 21st Ave. W, Seattle, WA 98199, USA

ARTICLE INFO

Handled by George A. Rose

Keywords:
Age-structured stock assessment methods
Simulation
Spatial structure
Spawning stock biomass
Time-varying natural mortality

ABSTRACT

A simulation-estimation approach is used to evaluate the efficacy of stock assessment methods that incorporate various levels of spatial complexity. The evaluated methods estimate historical and future biomass for a situation that roughly mimics Pacific herring *Clupea pallasii* at Haida Gwaii, British Columbia, Canada. The baseline operating model theorizes ten areas arranged such that there is post-recruitment dispersal among all areas. Simulated data (catches, catch age-composition, estimates of spawning stock biomass and its associated age structure) generated for each area are analyzed using estimation methods that range in complexity from ignoring spatial structure to explicitly modelling ten areas. Estimation methods that matched the operating model in terms of spatial structure performed best for hindcast performance and short-term forecasting, i.e., adding spatial structure to assessments improved estimation performance. Even with similar time trajectories among sub-stocks, accounting for spatial structure when conducting the assessment leads to improved estimates of spawning stock biomass. In contrast, assuming spatial variation in productivity when conducting assessments did not appreciably improve estimation performance, even when productivity actually varied spatially. Estimates of forecast biomass and of spawning stock biomass relative to the unfished level were poorer than estimates of biomass for years with data, i.e., hindcasts. Overall, the results of this study further support efforts to base stock assessments for small pelagic fishes on spatially-structured population dynamics models when there is a reasonable likelihood of identifying the sub-stocks that should form the basis for the assessment.

* Corresponding author.
E-mail address: aepunt@uw.edu (A.E. Punt).

https://doi.org/10.1016/j.fishres.2018.04.017
Received 9 August 2017; Received in revised form 24 March 2018; Accepted 22 April 2018

1. Introduction

Management strategies for many of the world's major fisheries are based on model-based harvest control rules (HCRs), which use the outputs from stock assessments that fit population dynamics models to available monitoring data (e.g., IWC, 2012). Population dynamics models that underlie these stock assessments range from those that consider only sex- and age-aggregated measures of biomass (e.g., ASPIC, Prager, 1992, 1994, 2002) to those that consider the sex, age, stage and spatial structure of the fished population (e.g., Stock Synthesis, Methot and Wetzell, 2013; MULTIFAN, Fournier et al., 1998). The type of model used for a stock assessment depends, *inter alia*, on the model outputs needed to apply the HCR, and on the available data, especially regarding the age and sex structure of the population.

Few stock assessments are currently based on population dynamics models that attempt to capture the spatial structure of fish or invertebrate populations, and those that do seldom involve a large number of spatial areas (2–3 is most common; Punt, in press). The main reason for this is that including a large number of areas in a population dynamics model can increase the complexity of the model and hence the number of estimable parameters. Most assessment analysts follow the principle of parsimony, and thus select simple models with few parameters to minimize the perceived variance of the estimates of the model outputs. Another oft-mentioned reason for not adopting spatially-structured stock assessments is lack of tagging data that would provide information about movement rates (A.E. Punt, pers. obs). Unfortunately, it is well known that ignoring spatial structure or assuming the incorrect spatial structure when applying a spatially-structured stock assessment can lead to biased (and often very imprecise) estimates of key model outputs, including estimates of spawning stock biomass, fishing mortality and recruitment (in absolute terms and relative to biological reference points) (e.g., Punt and Methot, 2004; Fu and Fanning, 2004; Cope and Punt, 2011; Garrison et al., 2011; Dougherty et al., 2013; Guan et al., 2013; Martien et al., 2013; Benson et al., 2015; Goethal et al., 2015; McGilliard et al., 2015; Punt et al., 2015). Furthermore, HCRs based on biased or imprecise stock assessments can result in unintended ecological, economic, and social consequences (Punt et al., 2016).

Of the few spatially-structured stock assessments that have been developed, most have been applied to relatively long-lived species such as groundfish and tunas (but see Dichmont et al., 2006; O'Neil et al., 2014; De Moor and Butterworth, 2015). Small pelagic species (e.g., sardines, anchovies, herrings) form the basis for some of the world's largest fisheries. However, with the exception of Cunningham et al. (2007) and de Moor and Butterworth (2015), pelagic species have not been assessed using spatially-structured methods of stock assessment. Ignoring spatial structure in management decision making for Pacific herring *Clupea pallasii* has been a concern among local and traditional knowledge holders in North Pacific communities where concentrated commercial fishing takes place on increasingly condensed spawning stocks (e.g., of herring for roe) considered critical for subsistence, trade, and other uses (Jones, 2000; Powell, 2012; Thornton et al., 2010; Thornton and Kitka, 2015; Levin et al., 2016). Further, commercial fishers and shoreworkers have identified several imperatives for Pacific herring management, including the need to build a collaborative understanding of the state of herring in its shared ecosystem,[1] with the hope that this will better acknowledge the livelihoods and fishing communities that depend on the herring fishery. This understanding could include appropriately accounting for spatial structure in stock assessments.

Spatial structure in small pelagic fishes exists at both large and small spatial scales. For example, the range of the northern subpopulation of Pacific sardine (*Sardinops sagax*) changes as a function of biomass and/or environmental conditions (e.g., Clark and Janssen, 1945); if ignored, this has been shown to lead to biased estimates of management-related quantities, including biomass (Hurtado-Ferro et al., 2014). Our paper focuses on relatively small-scale (10–100 s of km rather than 100 s–1000 s of km) spatial structure, with a focus on Pacific herring in British Columbia, Canada. The distribution and abundance of Pacific herring has varied substantially even within conventional 'stock areas' during the era of modern fisheries management. Nevertheless, current British Columbia herring stock assessments are based on the assumption that it is valid to pool data into five major and two minor stocks (Benson et al., 2015). Beyond the performance of stock assessment models, mismatches between the scales of observed or perceived population structure and aggregations used in assessment models can have consequences throughout the social-ecological system, including loss of trust in management bodies and conflict, in part because of the fine spatial scale at which traditional herring harvest practices occur (Levin et al., 2016). Problems of mismatch or fit among institutions of governance and social-ecological contexts are recognized more broadly as an enduring problem in the resource management literature (e.g., Epstein et al., 2015).

This paper uses a simulation-estimation approach to evaluate the consequences, in terms of the bias and precision of estimates of historical and projected spawning stock biomass, of various approaches to the assessment of stocks of short-lived fishes that exhibit spatial structure, based on the biological characteristics of Pacific herring at Haida Gwaii, British Columbia, Canada. Management for Pacific herring at Haida Gwaii is based on biomass estimates projected beyond the last year with data. Thus, the quantities used to evaluate estimation performance in this study include estimates of historical spawning stock biomass ('hindcast estimates') and projected biomass.

This paper aims to improve the basis for conducting stock assessments for small pelagic species, such as Pacific herring. Consequently, the key questions the analyses address are: (a) Are estimates of spawning stock biomass unbiased and precise (i.e., on average do the estimates equal the true values and is there little variation in estimates among replicate simulations) if the structure of the estimation method matches that of the spatially-explicit operating model? (b) How poor are the estimates of spawning stock biomass if the spatial structure of the operating model and estimation method differ? (c) How much spatial structure in the estimation method is sufficient to overcome any bias? and (d) How robust are the conclusions to key assumptions of the operating model, including the sample sizes for the data available for assessment purposes?

2. Materials and methods

2.1. Overview

An operating model is used to generate simulated data sets based on various specifications for the underlying system being assessed, including the number of 'sub-stocks'[2] (Table 1). It is spatially-structured and roughly mimics the population dynamics and fishery for Pacific herring (e.g., post-recruitment dispersal among sub-stocks, a fishery directed toward spawning fish only, and the possibility that an entire sub-stock skips spawning in a particular year). It includes multiple sub-stocks that are linked through dispersal. The generated data sets are analyzed using stock assessment methods (estimation methods) that range in the degree to which the assumptions of population structure match those of the operating model, from matching exactly to being

[1] See for example the open letter from the United Fisherman and Allied Workers Union to the Nuu-chah-nulth Tribal Council and British Columbia commercial herring fishermen (https://www.hashilthsa.com/news/2015-01-08/support-united-fisherman-and-allied-workers-union-herring-fishery).

[2] The term "sub-stock" is used here as these populations are neither demographically nor genetically distinguished.

Table 1
The population dynamic equations underlying the operating models and estimation methods. Table 2 provides the definitions for the symbols. The values for the first of annual deviations (i.e. for year −20) are sampled from their stationary distributions.

Eqn No	Equation	Description
T1.1	$N^1_{m,p,y,a} = \begin{cases} N_{m,p,y,a}\, e^{-M_{p,y,a}} & \text{if } a = 0 \\ \sum_j N_{m,j,y,a}\, e^{-M_{j,y,a}}\, \Omega_{j,p} & \text{otherwise} \end{cases}$	Numbers after 1+ diffusion and natural mortality
T1.2.	$N^2_{\text{mat},p,y,a} = (N^1_{\text{mat},p,y,a} + \widetilde{f}_a\, N^1_{\text{imat},p,y,a})e^{-F_{p,y}}$ $N^2_{\text{imat},p,y,a} = (1 - \widetilde{f}_a)N^1_{\text{imat},p,y,a}$	Numbers after maturation and the fishery
T1.3a	$D_{p,y} = \sum_a g_a N^2_{\text{mat},p,y,a}$	Egg production
T1.3b	$\widetilde{D}_{p,y} = \sum_a w_a N^2_{\text{mat},p,y,a}$	Spawning stock biomass
T1.4a	$N_{\text{imat},p,y,0} = \alpha_p\, D_{p,y}\, e^{-\beta_p D_{p,y}}\, \lambda_{p,y}$ if $a = 0$ $N_{\text{imat},p,y,a} = N^2_{\text{imat},p,y-1,a-1}$ otherwise	Immature numbers-at-age and recruitment
T1.4b	$N_{\text{mat},p,y,0} = 0$ if $a = 0$ $N_{\text{mat},p,y,a} = N^2_{\text{mat},p,y-1,a-1}$ if $1 \le a < x$ $N_{\text{mat},p,y,a} = N^2_{\text{mat},p,y-1,x-1} + N^2_{\text{mat},p,y-1,x}$ if $a = x$	Mature numbers-at-age
T1.5	$C_{y,p,a} = (N^1_{\text{mat},p,y,a} + \widetilde{f}_a\, N^1_{\text{imat},p,y,a})(1 - e^{-F_{p,y}})$	Catch-at-age
T1.6	$C_{p,y} = \sum_a w_a C_{p,y,a}$	Catch-in-weight
T1.7	$\lambda_{p,y} = e^{\varepsilon_{p,y} - \sigma_R^2/2}\,;\ \underline{\varepsilon}_y \sim N(\underline{0}; \Sigma)$	Recruitment deviations
T1.8	$M_{p,y,a} = \overline{M} e^{\eta_y - \sigma^2_{M1}/2} e^{\eta_{p,y} - \sigma^2_{M2}/2} e^{\omega_{p,y,a} - \sigma^2_{M3}/2}$	Natural mortality
T1.9a	$\eta_y = \rho_{M1}\eta_{y-1} + \sqrt{1 - \rho^2_{M1}}\,\tau_y \ ;\ \tau_y \sim N(0\,;\,\sigma^2_{M1})$	Annual natural mortality deviations
T1.9b	$\eta_{p,y} = \rho_{M2}\eta_{p,y-1} + \sqrt{1 - \rho^2_{M2}}\,\tau_{p,y}\,;\ \tau_{p,y} \sim N(0\,;\,\sigma^2_{M2})$	Annual sub-stock-specific mortality deviations
T1.9c	$\omega_{p,y,a} \sim N(0\,;\,\sigma^2_{M3})$	Age-, sub-stock, and year-deviations in natural mortality
T1.10	$\Omega_{j,p} = e^{-d^2_{j,p}/\sigma^2_\Omega} / \sum_{p'} e^{-d^2_{j,p'}/\sigma^2_\Omega}$	Diffusion
T1.11a	$\Sigma = \sigma^2_R \begin{pmatrix} 1 & q_{1,2} & \cdot \\ \cdot & \cdot & \cdot \\ q_{2,1} & \cdot & 1 \end{pmatrix}$	Recruitment variance-covariance matrix
T1.11b	$q_{i,j} = e^{-2\sin(\pi\delta d_{i,j}/10)}$	

based on a spatially-aggregated population dynamics model (see Fig. 1 for an overview of the simulation process).

2.2. The operating model

The operating model includes a number of sub-stocks (10 for most of the analyses for this paper in which case "area" is synonymous with "sub-stock") because several studies (e.g., Hay et al., 1989; Stick et al., 2014; Siple and Francis, 2016) have identified sub-stocks (or sections) throughout the range of Pacific herring. Each of the sub-stocks is represented using an age-structured model that keeps track of whether an animal is mature or not. The sub-stocks are arranged such that sub-stock P (P = 9 for the baseline operating model) is next to sub-stock 0 (Fig. 1), with the consequence that there is diffusion between all sub-stocks proportional to the "distance" between them. The sequential steps that occur each year to update the population dynamics (Eqns T1.1–T1.4) are: (a) removal by natural mortality, (b) diffusion of animals aged 1 and older among sub-stocks, (c) maturation, (d) fishing mortality, and (e) spawning and recruitment. The fishery is primarily for roe so selectivity is assumed for these simulations to be equivalent to being mature, and the fishery is assumed to occur on each sub-stock separately, prior to spawning (as well as any density-dependent response; Eqn T1.4a). The model treats mature and immature animals as separate groups, with maturation modeled explicitly (Eqn T1.2). The catches during year y are consequently only of animals that were mature at the start of year y or that matured during year y (Eqn T1.5). Natural mortality can change over time as a random walk (Eqn T1.8-9) and be subject to random deviations (Table 2).

The data generated by the operating model are: catches by area (assumed to be measured without error), indices of spawning stock biomass by area (with log-normally distributed observation errors), sampled age composition of catches by area, and sampled age composition of the spawning stock biomass by area. Age-composition data by area are multinomial samples from the true catches-at-age and spawning biomass age-structure, with effective sample sizes proportional to the catch-in-weight (Eqn T1.6) for the catch age-compositions and independent of area for the spawning biomass age-structure.

The operating model is initialized with an arbitrary age structure and projected forward until it reaches unfished equilibrium. It is then projected forward for 20 years (denoted as years −20 to 0) under pre-specified fishing mortality (in expectation F_{MSY}, the fishing mortality rate at which MSY is achieved given all sub-stocks are exploited at the same rate and diffusion and recruitment are not stochastic; DFO, 2017) and stochastic recruitment (and under some scenarios, stochastic natural mortality) so that the age structure of the operating model populations at the start of the first year of the assessment is neither in equilibrium nor in an unfished state in expectation. The operating model is then projected forward a further 55 years (years 1–55) with annual fishing mortality rates that are constant over space (for both the base-case operating model and most of the sensitivity scenarios, see Section 2.5 below) and set based on achieving a pre-specified proportion of the global F_{MSY} (e.g., Fig. 2a). The remaining biological parameters of the operating model (weight-at-age, maturity-at-age, and fecundity-at-age) are summarized in Fig. 2b–d.

When an estimation method is applied to spatially-aggregated data, the generated data are pooled over areas without weighting by catch or spawning stock biomass (although the approach used to generate the age-composition essentially weights by abundance). The value for the extent of variation in the recruitment, σ_R, is determined by projecting the operating model forward for 10,000 years with no catches and

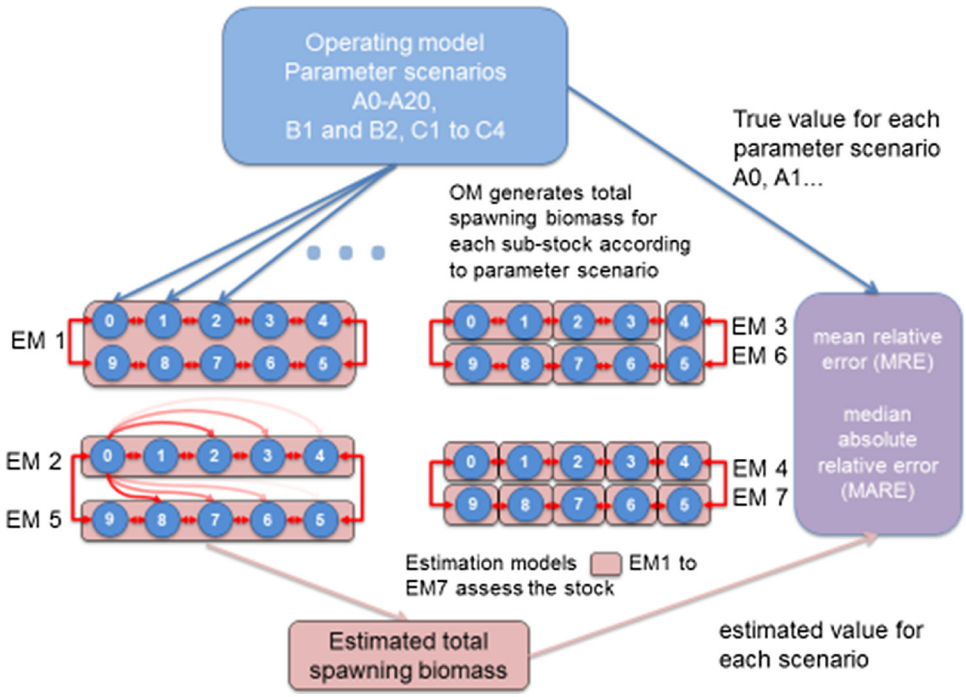

Fig. 1. Overview of the structure of a single simulation, highlighting spatial structure of the baseline operating model (sub-stocks = circles; note that diffusion can occur between non-adjacent sub-stocks) and the sub-stock structure underlying estimation models (EMs) 1–7 (boxes). The red arrows indicate the major routes of migration (although animals can move to more than adjacent sub-stocks). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Descriptions of the symbols included in the specification of the operating model and estimation methods.

Symbol	Description
$C_{p,y}$	Catch-in-weight for sub-stock p and year y
$C_{p,y,a}$	Catch of animals of age a in sub-stock p during year y
$D_{p,y}$	Egg production by sub-stock p during year y
$\widetilde{D}_{p,y}$	Spawning stock biomass of sub-stock p during year y
$F_{p,y}$	Fishing mortality on mature animals for sub-stock p during year y
$\overline{M}$	Mean rate of natural mortality (set to 0.58 yr^{-1})
$M_{p,y,a}$	Rate of natural mortality by sub-stock, year and age
$N_{m,p,y,a}$	Number of animals of maturity stage m (mature/immature) in sub-stock p that are of age a at the start of year y
$N^1_{m,p,y,a}$	Number of animals of maturity stage m in sub-stock p that are of age a after natural mortality and dispersal during year y
$N^2_{m,p,y,a}$	Number of animals of maturity stage m in sub-stock p that are of age a after maturation and fishing mortality during year y
$d_{j,p}$	"distance" between sub-stocks j and p, nominally $\lvert j\text{-}p\rvert$
f_a	Proportion of animals of age a that are mature at age a (Fig. 2b)
$\widetilde{f}_a$	Proportion of immature animals of age a that mature; $\widetilde{f}_a = (f_a - f_{a-1})/(1 - f_{a-1})$
g_a	Fecundity as a function of age (Fig. 2c)
w_a	Weight as a function of age (Fig. 2d)
x	Maximum (lumped) age-class (set to 10 years)
α_p, β_p	Stock-recruitment parameters for sub-stock p (β_p is assumed to be a constant over sub-stocks and is set to 100)
$\varepsilon_{p,y}$	The recruitment residual for sub-stock p and year y;
δ	Parameter that determines the spatial auto-correlation in recruitment
$\lambda_{p,y}$	Deviations in recruitment about the stock-recruitment relationship (by sub-stock)
$\Omega_{j,p}$	Expected proportion of 1+ animals from sub-stock j that diffuse to sub-stock p; the realized diffusion rate; $\Omega_{j,p,y}$ is a Dirichlet sample about $\Omega_{j,p}$
Σ	Variance-covariance matrix for the recruitment deviations
ρ_{M1}	Extent of auto-correlation in natural mortality
ρ_{M2}	Extent of sub-stock-specific auto-correlation in natural mortality
σ_{M1}	Standard error of the annual deviations in natural mortality
σ_{M2}	Standard error of the sub-stock-specific deviations in natural mortality
σ_{M2}	Standard error of the age-specific deviations in natural mortality
σ_R	Standard error of the annual deviations in recruitment
σ_Ω	Parameter that determines diffusion – selected to achieve a pre-specified proportion of animals leaving the sub-stock

selecting σ_R so that the coefficient of variation of the sum of recruitment over sub-stock equals the pre-specified extent of variation.

2.3. The estimation methods

The estimation methods mimic the operating model in terms of basic population dynamics (Eqns T1.1–T1.9), except that natural mortality is assumed to be constant, and independent of age, sub-stock and time. Seven variants of the estimation method are considered:

- EM1. The region to be managed is assessed ignoring sub-stock-structure.
- EM2. Sub-stocks 0–4 and 5–9 are assessed as separate units (i.e., two assessment sub-stocks); a single recruitment parameter α (as defined below) is estimated.
- EM3. Sub-stocks 0–1, 2–3, etc. are assessed as separate units (i.e., five assessment sub-stocks); a single α parameter is estimated.
- EM4. Each sub-stock is assessed separately (i.e., 10 assessment sub-stocks); a single α parameter is estimated.
- EM5. Sub-stocks 0–4 and 5–9 are assessed as separate units (i.e., two assessment sub-stocks); α_p is estimated for each assessment sub-stock.
- EM6. Sub-stocks 0–1, 2–3, etc. are assessed as separate units (i.e., five assessment sub-stocks); α_p is estimated for each assessment sub-stock.
- EM7. Each sub-stock is assessed separately (i.e., 10 assessment sub-stocks); α_p is estimated for each assessment sub-stock.

The parameters estimated are: the parameters of the stock-recruitment relationship (the slope of the stock-recruitment relationship at the origin, α_p – which determines the productivity of the population – and unfished recruitment), the diffusion rate (see Eqn T1.10), the deviations in recruitment about the stock-recruitment relationship for each year with catches, and the annual deviations in recruitment for the first year of the projection period. The model is implemented using AD Model Builder (Fournier et al., 2012). Natural mortality is set to 0.58 yr^{-1} for all years following DFO (2015), and is assumed known for the base-case operating model and most of the sensitivity scenarios. The annual fishing mortality rates are not treated as estimable parameters, but are

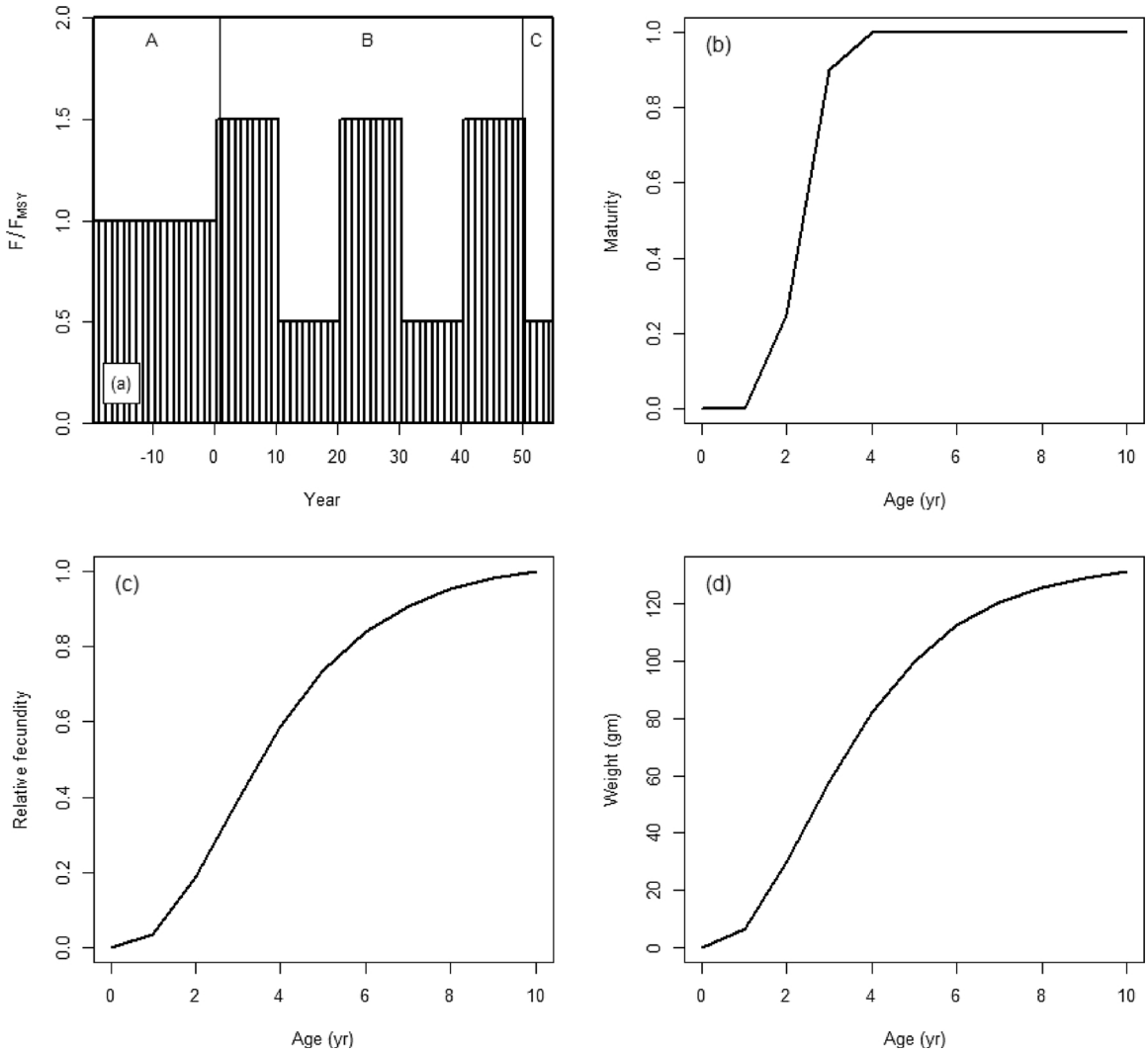

Fig. 2. Operating model parameters: (a) fishing mortality on mature animals (expressed relative to F_{MSY}), (b) proportion maturity-at age, (c) relative fecundity-at-age, and (d) weight-at-age.

rather calculated using the 'hybrid' method (Methot and Wetzell, 2013). The estimation model is fitted ignoring that the deviations in recruitment may be correlated spatially. The estimation methods are provided with sampling CVs for the index of abundance and effective sample sizes for the age-composition data. The CVs for the index are always taken to be that for a single sub-stock, while the effective sample sizes for the age-composition data are set to the numbers of animals aged. The assumed CVs and sample sizes are correct for EMs 4 and 7, but the CV will somewhat overestimate variation and the effective sample sizes will underestimate variation for the estimation methods that pool data spatially.

2.4. Performance metrics

The results of the simulations are summarized by the ability to estimate total (over all sub-stocks) spawning stock biomass (SSB; $\widetilde{D}$) over time. Particular focus is on the spawning stock biomass at the start of the second year with catches ($\widetilde{D}_2$),[3] the spawning stock biomass in the last year with monitoring data ($\widetilde{D}_{\mathrm{last}}$; where year 'last' is year 50), and the ratio of $\widetilde{D}_{\mathrm{last}}$ to $\widetilde{D}_2$. The relative errors between the true and estimated model outputs are summarized by the mean relative error (MRE; i.e. bias) and the median absolute relative error (MARE, which accounts for bias and variability):

$$\mathrm{MRE}=\underset{\mathrm{i}}{\mathrm{Mean}}\frac{x_i^{\mathrm{Est}}-x_i^{\mathrm{True}}}{x_i^{\mathrm{True}}}\quad \mathrm{MARE}=\underset{\mathrm{i}}{\mathrm{Median}}\left|\frac{x_i^{\mathrm{Est}}-x_i^{\mathrm{True}}}{x_i^{\mathrm{True}}}\right| \tag{1}$$

where x_i^{True} is the true value of the quality x for the ith replicate simulation, and x_i^{Est} is the estimated value of the quality x for the ith replicate simulation.

The aforementioned outputs $\widetilde{D}_2$, $\widetilde{D}_{\mathrm{last}}$, and $\widetilde{D}_{\mathrm{last}}/\widetilde{D}_2$ evaluate the estimation model's hindcast ability. However, management advice is often based on forecasts. Therefore, the operating model is used to project spawning stock or sub-stock biomass ahead an additional five years, with only catch data available to the estimation methods for years 51–55. The ability to estimate the spawning stock biomass in years 51–55 is evaluated by computing MREs and MAREs for $\widetilde{D}_{\mathrm{last+5}}$. Management advice for herring in Canada is based on spawning stock biomass relative to unfished spawning stock biomass (Benson et al., 2015; DFO, 2016). Hence, performance metrics are reported for $\widetilde{D}_{\mathrm{last}}/\widetilde{D}_{-\infty}$ where $\widetilde{D}_{-\infty}$ is the unfished spawning stock biomass (the spawning stock biomass in equilibrium when the only source of mortality is due to natural causes). Actual management advice is based on one year projections so Supplementary Table 1 provides MAREs for $\widetilde{D}_{\mathrm{last+1}}$, which are intermediate between those for $\widetilde{D}_{\mathrm{last}}$ and $\widetilde{D}_{\mathrm{last+5}}$.

Erroneously estimating spawning stock biomass has additional

[3] The first year is impacted by assumptions regarding initial conditions so year 2 is more reflective of estimation performance for the initial years of the assessment.

Table 3
Specifications of operating model scenarios. The specifications match those of the base-case scenario, unless otherwise specified.

Abbreviation/description	σ_{M1}; σ_{M2}; σ_{M3}; ρ_{M1}; ρ_{M2}	Global recruitment variation, σ_R	Correlation in recruitment between adjacent sub-stocks (determines $q_{i,j}$)	CV of Fishing mortality about Fig. 2a	Adult diffusion (determines, $d_{i,j}$)	Slope at the origin of the SR relationship (α_p)	Skipped Spawning	Stochastic diffusion	Fishing mortality (expected)	True sub-stocks
Base case	0; 0; 0; 0; 0	0.6	0.7071[a]	0.2	0.20	73.57 ∀ p	No	No	Fig. 2a	10
A0. With productivity variation						*LN* (73.57, 0.5)				
A1. Low recruitment variation		0.4								
A2. High recruitment variation		1.0								
A3. M variation-1	0.2; 0; 0; 0; 0									
A4. M variation-2	0.2; 0.2; 0; 0; 0									
A5. M variation-3	0.2; 0.2; 0; 0.707; 0									
A6. M variation-4	0.2; 0.2; 0.2; 0; 0									
A7. M variation-5	0.2; 0.2; 0.2; 0; 0.707									
A8. Higher diffusion					0.40					
A9. Lower diffusion					0.01					
A10. Independent recruitment			0							
A11. Higher variation in F				0.4						
A12. Lower variation in F				0						
A13. F is proportional to spawning biomass										
A14. Variation in F is spatial as well as temporal				0.2[b]						
A15. Skipped spawning (number sub-stocks not spawning each year)							Yes (3)			
A16.Stochastic diffusion (Dirichlet effective sample size)								Yes (5)		
A17: Constant expected fishing mortality									Constant	
A18: Increasing then decreasing fishing mortality									Increasing then decreasing	
A19: One true sub-stock										1
A20: Five true sub-stocks										5

[a] Assumes that half of the variation in recruitment is due to auto-correlation.
[b] There is spatial and temporal log-normal variation in fishing mortality for this sensitivity scenario.

implications for management, such as depletion of the target species, local depletion of spawning sites, and impacts on the local community, which depend on herring to satisfy traditional use, including subsistence requirements. As an approximate way to evaluate the biological implications of errors in estimating spawning stock biomass, the probability of the estimate of spawning stock biomass from the assessment being more than 40% larger than the true spawning stock biomass at the sub-stock level is reported. Undesirable levels of resource depletion would be expected if management measures such as catch limits were based on such substantially overestimated estimates of spawning stock biomass. The probability of estimates that are 40% or more smaller than the true values are also reported, as such errors could lead to underutiltization. The value of 40% is somewhat arbitrary, but reflects the average extent of error associated with data-rich stock assessments of groundfish and coastal pelagic species off the US west coast (Ralston et al., 2011). This metric is easy to calculate for the 10 sub-stock estimation methods (EMs 4 and 7) because there is a 1-1 match between sub-stocks in the operating model and the assessment sub-stocks in the underlying estimation method. However, this is not the case for other estimation methods. Consequently, the estimate of spawning stock biomass for a sub-stock that is assessed in combination with other sub-stocks is the estimate of spawning stock biomass for the entire assessed area divided by the number of sub-stocks in that area, i.e. for EM5 under the baseline operating model, the estimate of spawning stock biomass for sub-stock 0 is the estimate for sub-stocks 0 and 1 combined divided by 2.

2.5. Scenarios

The base-case scenario (Table 3) sets the values for key parameters to reference levels. It reflects a scenario that is deliberately parameterized so that spatially-aggregated estimation methods may perform adequately (e.g., by assuming that productivity does not vary spatially and that fishing mortality is the same for each sub-stock). Sensitivity analyses were conducted, by varying several of the features of the base-case scenario in turn, to explore the consequences of factors that may impact estimation performance (Table 3). The sensitivity scenarios consider sensitivity to spatial variation in recruitment, natural mortality, and productivity as well as the possibility that mean natural mortality varies over time. Most of the sensitivity scenarios involve changing a parameter value, but some are more complex. The scenario with fishing mortality proportional to biomass (A13) scales total fishing mortality each year by sub-stock proportional to the spawning stock biomass by sub-stock, keeping the total fishing mortality the same. The skipped spawning scenario (A15) involves selecting three of the sub-stocks at random each year and assuming that they do not spawn that year (so that there is neither catch nor spawning from the sub-stock – and hence no data on spawning stock biomass) – the fishing mortality rates for the other sub-stocks are set to the base-case values for this sensitivity scenario. The skipped spawning scenario is an extreme form of the observation (co-author HK pers. comm.) that some proportion of animals that would spawn at some spawning sites do not. The extent of diffusion in sensitivity scenario A16 is based on a Dirichlet sample size of 5, which implies that for an expected diffusion rate of 0.2, 90% of the actual diffusion proportions will range between ~0 and 0.86. For each combination of operating model and estimation method, 250 simulations are conducted to ensure that results are sufficiently precise to draw inferences.

The fishing mortality pattern in Fig. 2a is quite informative. Consequently, sensitivity scenarios A17 and A18 explore the consequences of different patterns of fishing mortality. Sensitivity scenario A17 explores the case in which fishing mortality is 0.7 in expectation from years −20 to 55, while sensitivity scenario A18 explores the case in which expected fishing mortality increases by 50% over the years 1–25 and reduces to its year-0 value in year 50 (and is constant thereafter).

The base-case scenario is based on 10 sub-stocks so all of the estimation methods assume either correct number of sub-stocks or a fewer number. Sensitivity scenarios A19 and A20 consider situations in which there is only one sub-stock (A19) or five sub-stocks (A20). These sensitivity scenarios involve using the areas in operating model (of which there are 10) and that one sub-stock may be found in multiple areas, and hence replacing Equation T1.2 by:

$$N^2_{\text{mat},p,y,a} = \frac{1}{n_s}\sum_q \left(N^1_{\text{mat},q,y,a} + \tilde{f}_a N^1_{\text{imat},q,y,a}\right)e^{-F_{p,y}}$$
$$N^2_{\text{imat},p,y,a} = \frac{1}{n_s}\sum_q \left(1-\tilde{f}_a\right)N^1_{\text{imat},q,y,a} \quad (2)$$

where q denotes the set of areas in which sub-stock p is found.

Table 4 lists the scenarios related to values for the coefficient of variation (CV) for the index of spawning biomass and the effective sample size for the age-composition data. Tables 3 and 4 involve single changes to the base-case operating model, but the changes may interact. Table 5 therefore lists four sensitivity scenarios that involve changing multiple features of the operating model at once, which initial analyses suggested may be consequential individually.

Table 4
Specifications of operating model scenarios related to data.

Abbreviation/ description	CV of abundance indices	Effective sample size for the survey age-composition data (by sub-stock)	Effective sample size for the catch age-composition data (in total)
A0. Base case	0.3	100	200
B1. High uncertainty	0.5	50	100
B2. Low uncertainty	0.2	200	500

Table 5
Specifications for the scenarios with multiple changes from the base-case scenario.

Abbreviation/description	Changes from the base-case scenario
C1	Productivity variation & high diffusion
C2	Productivity variation & stochastic diffusion
C3	Productivity variation. $\sigma_{M1} = 0.2$ & $\rho_{M1} = 0.7071$
C4	Productivity variation & low diffusion

3. Results

3.1. Base-case results

It is illustrative to examine the behavior of the estimation methods for a single replicate simulation before examining the results of multiple simulations. Fig. 3 shows the time-trajectories of fishing mortality, recruitment, spawning stock biomass and catch for a single replicate simulation. Results are shown by sub-stock, although there is only one time-trajectory for fishing mortality because it does not vary spatially for the base-case operating model (Table 3). The base-case scenario is a "highly informative case" in that fishing mortality is spatially constant, but varies substantially over time (Fig. 3a), leading to considerable contrast in spawning stock biomass (Fig. 3c). Estimation methods would be expected to perform better for this situation than for one with little temporal variation in fishing mortality and hence biomass (Punt, 1995, 1997; Magnusson and Hilborn, 2007). Recruitment varies considerably among years and sub-stocks (Fig. 3b), even though there is spatial auto-correlation in the recruitment deviations. The extent of variation in spawning stock biomass over time (Fig. 3c) is much smaller than in recruitment (Fig. 3b) because (a) spawning stock biomass consists of multiple age classes, which dampens the effects of

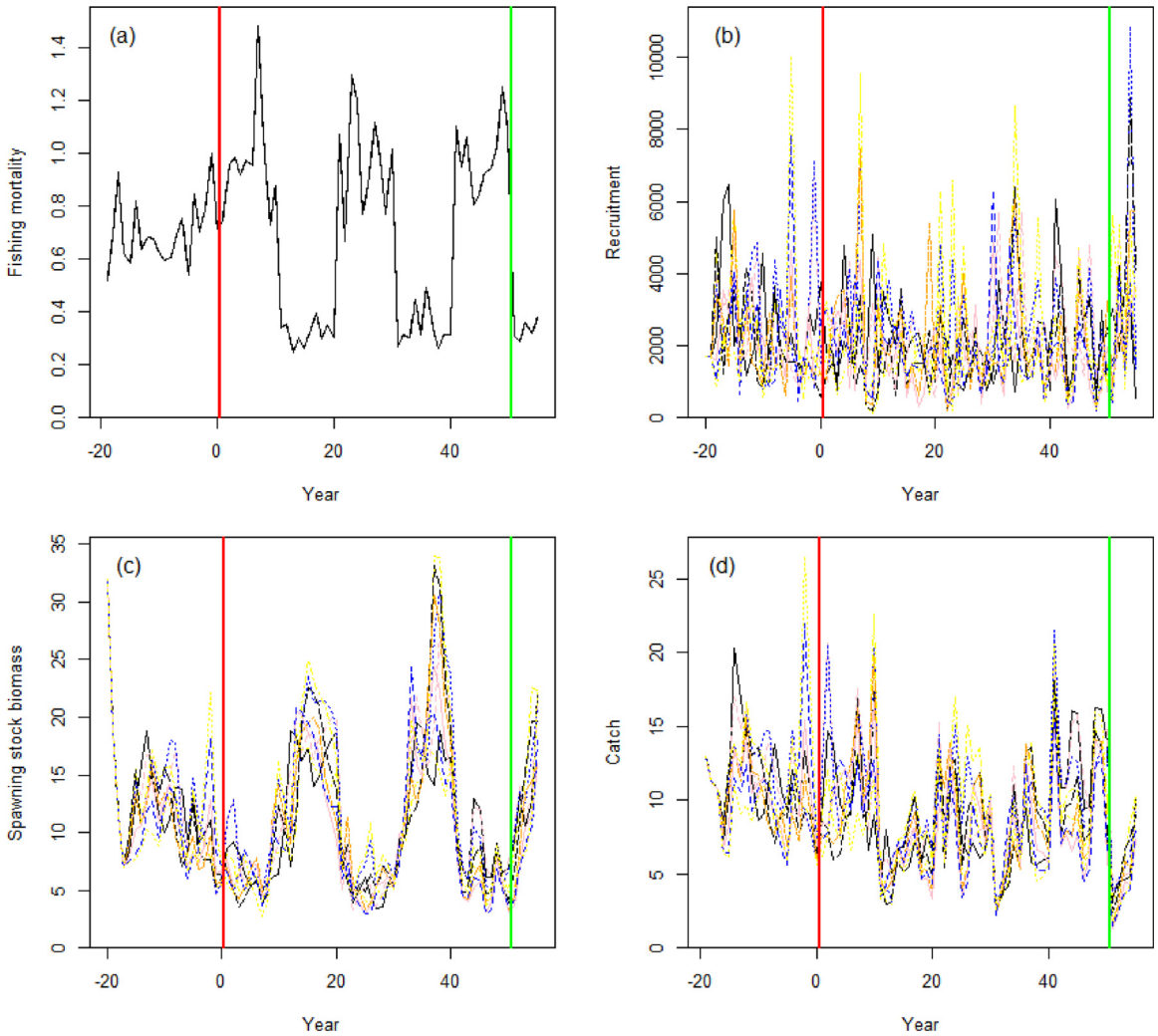

Fig. 3. Time-trajectories of operating model fishing mortality, recruitment, spawning stock biomass, and catch for a single replicate simulation based on the base-case specifications. Panels (b)–(d) show results for each of the ten sub-stocks. The red vertical lines denote when the first data used by the estimation methods are available, and the green vertical lines denote the start of the projection period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

recruitment variation, (b) productivity (in the form of the value of the slope of the stock-recruitment relationship at the origin, α_p in Eqn T1.4a) is the same for each sub-stock, and (c) there is diffusion among sub-stocks (20% of the 1+ biomass migrates from each sub-stock each year). The synchrony in operating model biomass suggests that EM1, which pools data spatially, may perform adequately for the base-case operating model.

The estimation methods are generally able to capture the broad trends in spawning stock biomass when biomass is scaled to the level of an average sub-stock, across all estimation methods (Fig. 4b, c) The estimates of spawning stock biomass are not sensitive to whether the estimation method estimates a single α_p for all sub-stocks or whether α_p is estimated by sub-stock (contrast the results for EMs 2, 3, and 4 in Fig. 4b with those for EMs 5, 6 and 7 in Fig. 4c). The estimates from EMs 1, 2 and 5 (EMs that consider a single or two sub-stocks when conducting assessments; Fig. 1) are similar. These EMs over-estimate spawning stock biomass between years 12 and 15 and under-estimate peak abundance for years 15–20 and 33–40. In contrast, EMs 3, 4, 6 and 7 are able to mimic operating model biomass better than the simpler estimation methods, although EMs 3 and 6 perform poorer than EMs 4 and 7 for some years (e.g. 8–15).

EMs 3, 4, 6 and 7 (and to a lesser extent 2 and 5) are close to unbiased (in median terms) over the historical period, while EM1 and EM2 lead to the widest distributions of relative error (Fig. 5). There is little benefit to assessing the region as two sub-stocks rather than as one sub-stock (contrast the results for EM1 and EM2), although EM2 is less biased than EM1. Including five sub-stocks in the assessment (EM3) leads to narrower relative error distributions, but EM4 performs best in this regard. Similarly, the performance metrics indicate close to unbiased estimates for $\widetilde{D}_2$, $\widetilde{D}_{\text{last}}$, and $\widetilde{D}_{\text{last}}/\widetilde{D}_2$, but with markedly lower MAREs for EMs 4, and 7 than EMs 1, 2 and 5 (Table 6). Performance metrics suggest there is little benefit to moving from a single sub-stock to a two sub-stock assessment, though EMs 4 and 7 outperform EMs 3 and 6 quite substantially for the three model outputs, demonstrating potential benefits of moving to a 10 sub-stock estimation method.

The relative error distributions broaden towards the end of the hindcast period because the ability to estimate the size of a year class depends on how often it is monitored in fisheries and surveys; the cohorts that were spawned most recently are not as well monitored as those that were spawned prior to this. In addition, and as expected, the forecast ability of the estimation model is markedly poorer than its hindcast ability, with the relative error distributions broadening substantially after year 50 (Fig. 5): MARE increases markedly from year 50 to year 55 (Supplementary Table 1). The benefits of conducting

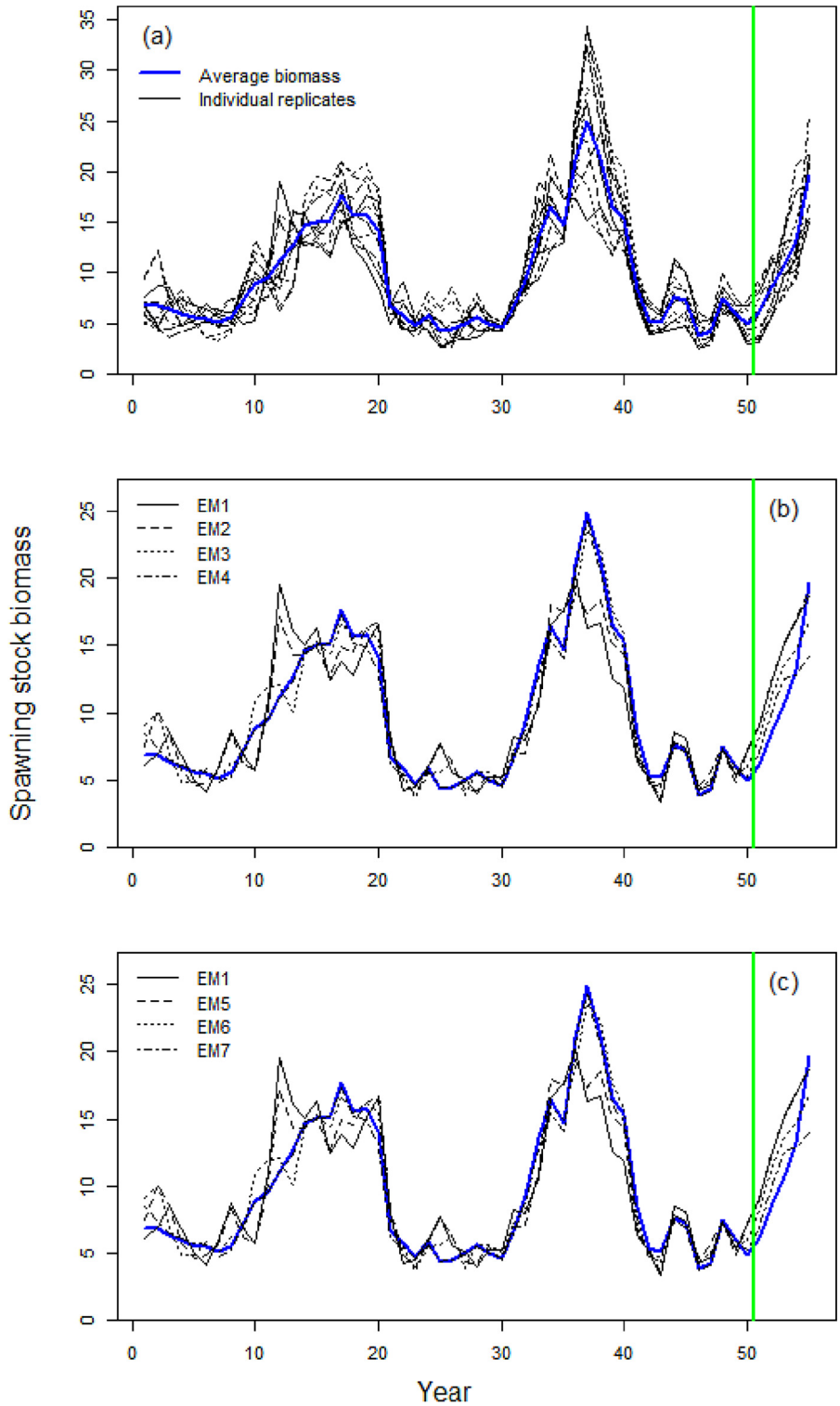

Fig. 4. Time-trajectories of estimated spawning stock biomass by sub-stock for a single replicate based on the base-case specifications (black lines) and average aggregate spawning stock biomass (blue line) (a), along with the estimates of spawning stock biomass (expected spawning stock biomass by sub-stock) from the seven estimation methods (black lines) (b,c). The green vertical lines denote the start of the projection period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

assessments with the correct number of sub-stocks are reduced when the aim is to estimate future spawning stock biomass, given only a time-series of catches. This is evident based on the MAREs for the $\widetilde{D}_{\text{last+5}}$, which are much higher than those for $\widetilde{D}_2$ and $\widetilde{D}_{\text{last}}$, with almost no improvement in estimation performance for the most complex estimation methods compared to the simpler estimation methods for beyond two years into the future (Supplementary Table 1).

In relation to the ability to estimate spawning stock biomass relative to the unfished spawning stock biomass, except for EMs 4 and 7, the

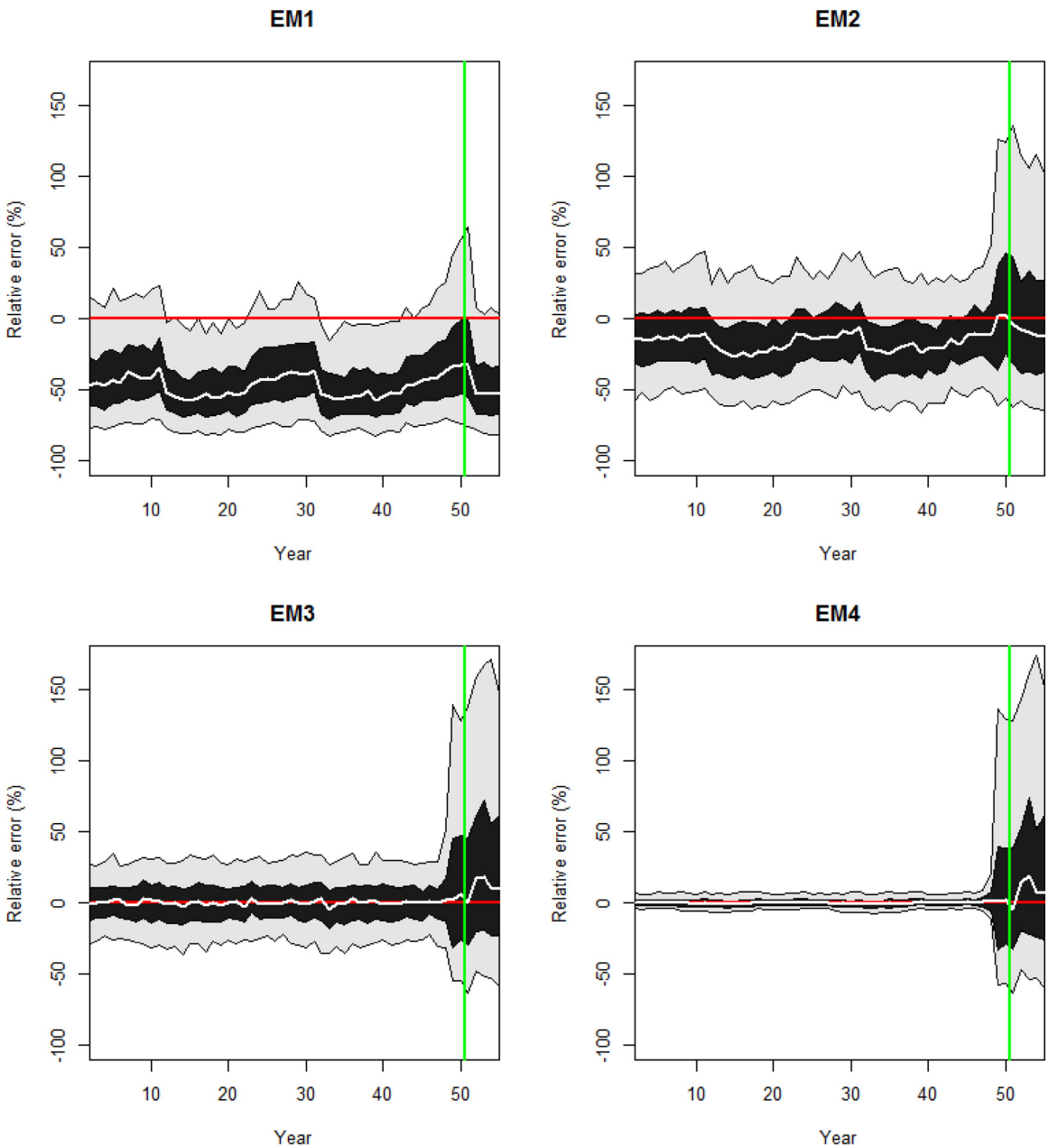

Fig. 5. Time-trajectories of percentage relative error of spawning stock biomass for EMs 1–4 for the base-case specifications (the results for EMs 5–7 are visually identical to those for EMs 2–4). The light shading encompasses 90% of the distributions and the dark shading 50%, while the white line denotes the time-trajectory of median relative errors. The green vertical lines denote the start of the projection or forecast period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

estimates of $\widetilde{D}_{\text{last}}/\widetilde{D}_{-\infty}$ are close to unbiased (MRE < 3%). The estimation variability is such that the MAREs for $\widetilde{D}_{\text{last}}/\widetilde{D}_{-\infty}$ are very similar among the EMs. It is sometimes argued that ratios of biomass are estimated better than biomass in absolute terms (e.g., Punt et al., 2002; Magnusson and Hilborn 2007), but that argument is not supported here, with the MAREs for $\widetilde{D}_{\text{last}}/\widetilde{D}_2$ and $\widetilde{D}_{\text{last}}/\widetilde{D}_{-\infty}$ larger than (or of similar magnitude to) those for $\widetilde{D}_{\text{last}}$ (Table 6).

The probabilities of over-estimating spawning stock biomass by substock are very similar for EMs 1, 2 and 5 at about 20% (increasing to 30% for forecasted biomass), while applying EMs 3 and 6, and to an even greater degree EMs 4 and 7, leads to lower probabilities (Table 6). The probability of substantially underestimating biomass is much lower than overestimating it, particular in years 2 and 50, and particularly for EMs 4 and 7. Again, however, when forecast ability is considered, there are no appreciable differences amongst any of the seven estimation methods (Table 6, column Year 55 "probability error" > 40%).

3.2. Sensitivity analyses

3.2.1. Single factor sensitivity scenarios

Allowing for spatial variation in productivity (log-normal variation in the slope at the origin of the stock-recruitment relationship with a CV of 0.5; scenario A0) leads to slightly poorer estimation performance for most EMs compared to the base case (Fig. 6b; Supplementary Table 1). The probabilities of estimation errors > 40% and < −40% also increase for almost all except EMs 4 and 7, with spatial variation in productivity. The effects of the changes in estimation performance are particularly evident for EMs 1, 2, 3, 5 and 6, but are largest for EM1. Estimating spatial variation in productivity (EMs 5, 6 and 7) improves estimation performance, but not by much (Supplementary Table 2).

Table 6
Values for the performance metrics for the simulations using the base-case specifications for the operating model.

	$\widetilde{D}_2$		$\widetilde{D}_{last}$		$\widetilde{D}_{last}/\widetilde{D}_2$		$\widetilde{D}_{last+5}$		$\widetilde{D}_{last}/\widetilde{D}_{-\infty}$		Probability error > 40%			Probability error < −40%		
	MRE	MARE	MRE	MARE	MRE	MARE	MRE	MARE	MRE	MARE	Year 2	Year 50	Year 55	Year 2	Year 50	Year 55
EM1	7.6	16.7	2.9	12.3	8.5	22.1	2.2	40.2	0.2	15.1	0.22	0.20	0.30	0.03	0.03	0.23
EM2	8.3	16.9	4.0	13.1	15.2	24.0	−0.8	37.7	0.5	15.4	0.23	0.20	0.29	0.05	0.03	0.25
EM3	8.6	13.5	4.9	10.9	12.6	16.2	−1.7	35.5	0.1	11.9	0.21	0.15	0.30	0.05	0.03	0.26
EM4	2.7	3.7	−1.0	7.2	2.5	7.8	−4.7	36.6	−11.1	11.8	0.03	0.07	0.28	0.01	0.04	0.28
EM5	7.2	16.6	4.0	13.4	15.4	23.9	0.9	38.0	0.9	15.4	0.23	0.20	0.29	0.05	0.03	0.26
EM6	8.2	13.0	5.6	11.2	12.5	16.6	−0.2	35.8	2.0	13.1	0.21	0.15	0.30	0.05	0.03	0.27
EM7	2.4	3.6	−0.9	6.8	1.6	7.9	−5.4	34.9	−10.2	11.9	0.02	0.07	0.27	0.00	0.04	0.29

The ability to estimate model outputs is sensitive to the extent of recruitment variation, with higher MAREs and probability of errors > 40% and < −40% compared to the base case for higher recruitment variation, and vice versa (sensitivity scenarios A1 and A2; Supplementary Table 2). Forecast performance deteriorates more than the hindcast performance when recruitment variation exceeds the base-case level. This is not surprising given the estimation method is not provided with data to estimate future recruitment.

Estimation performance also deteriorates when natural mortality varies over time (sensitivity scenarios A3-A7), particularly when there is serial auto-correlation in natural mortality (sensitivity scenarios A5 and A7; Supplementary Table 2 and Fig. 6c). The MAREs for the previously best performing estimation methods (EMs 4 and 7) deteriorate the most compared to the base-case scenario when natural mortality is time-varying because these EMs are now mis-specified, but they still perform best among the EMs for the scenarios with time-varying natural mortality. The effect of time-varying natural mortality is most marked for the forecasts.

Assuming that recruitment is spatially-independent (sensitivity scenario A10; Supplementary Table 2 and Fig. 6d) leads to poorer estimation performance compared to the base-case scenario, for the estimation methods where population dynamics models differ from the underlying spatially-structured operating model (i.e., EMs 1, 2, 3, 5 and 6).

Inclusion of skipped spawning in the operating model (sensitivity scenario A15) leads to poor estimation performance for all estimation methods (Supplementary Table 2 and Fig. 6f), but particularly those based on population dynamics models that do not match the operating model (EMs 1, 2, 3, 5 and 6). The high MAREs for these estimation methods are due in large part to negative bias in the estimates of spawning stock biomass (Supplementary Table 2). Negative bias arises for the estimation methods that pool data spatially because skipped spawning leads to an index of spawning stock biomass that is negatively biased owing to the exclusion of information for some sub-stocks from the index. In contrast, EMs 4 and 7 treat the lack of data for sub-stocks in some years as missing information, which does not lead to bias, but does increase estimation variation owing to the lack of data relative to that available for the base-case scenario. The probability of errors > 40% is much lower for EMs 1, 2, 3, 5 and 6 than for the base-case scenario for years 50 and 55, but the probability of errors < −40% or > 40% is correspondingly higher (Fig. 6f). This is due to the negative bias associated with the estimates of spawning stock biomass.

Stochastic diffusion (sensitivity scenario A16; Supplementary Table 2 and Fig. 6g) leads to higher MAREs for the estimation methods that are correctly specified in expectation (EMs 4 and 7). However, this scenario also leads to the lowest MAREs, relative to the other scenarios, for the forecast biomass. However, the probabilities of errors > 40% at the sub-stock level are higher relative to the other scenarios, for all estimation methods, including EMs 4 and 7, the performances of which are markedly poorly for sensitivity scenario A16 than for the base-case scenario (Fig. 6g vs a).

As expected, less informative time-trajectories of fishing mortality lead to higher MAREs, particularly in the forecast years, although the relative ranking of the seven EMs remains unchanged (sensitivity scenarios A17 and A18; Supplementary Tables 1 and 2). As expected, the differences among the estimation methods in terms of MAREs is lower when there are only five true sub-stocks (sensitivity scenario A20; Fig. 6h) and particularly when there is only one true sub-stock (sensitivity scenario A19; Fig. 6i). The MAREs are essentially identical among the seven EMs when there is only one true sub-stock (generally < 0.2 for all model outputs). EMs 4 and 7 are outperformed by the other EMs in terms of the probability of errors > 40% in years 2 and 50 for this sensitivity scenario (Supplementary Table 2).

The results are not sensitive to the extent of diffusion (sensitivity scenarios A8 and A9) or spatial and temporal variation in fishing mortality (sensitivity scenarios A11, A12, A13, and A14).

The ability to estimate the model outputs depends predictably on the quantity of the data, i.e., the MAREs and the probability of errors > 40% or < −40% increase when there are less informative data than for the base-case scenario (sensitivity scenarios B1 and B2; Supplementary Table 2).

3.2.2. *Multi-factor sensitivity scenarios*

Combining high diffusion with spatial variation in productivity (sensitivity scenario C1; Supplementary Table 2 and Fig. 6j) leads to very similar results to sensitivity scenario A0, which had 20% diffusion, although the probability of errors > 40% is lower for the mis-specified estimation methods (1, 2, 3, 5 and 6), and particularly for EMs 4 and 6.

Combining spatial variation in productivity with either stochastic diffusion (sensitivity scenario C2; Supplementary Table 2 and Fig. 6k) or time-varying natural mortality (sensitivity scenario C3; Supplementary Table 2 and Fig. 6l) generally leads to the largest MAREs and probabilities of errors > 40%. As was the case for sensitivity scenario A16, the performances of EMs 4 and 7, which are still better than those of the remaining EMs, are markedly poorer for these sensitivity scenarios than the remaining cases, with sensitivity scenario C3 leading to poorest performances overall.

4. Discussion and conclusions

An underlying principle of statistical modelling is to select a model structure that balances realism (which results in more complex models with additional parameters) with parsimony (which reduces estimation variance). Most marine species exhibit spatial structuring, but almost all stock assessments are based on population dynamics models that ignore spatial structure. The aim of this paper was to identify when simple and more complex estimation methods will lead to estimates of spawning stock biomass that have the least error for situations that are representative of small pelagic fishes, such as Pacific herring. The measures of error considered include the MARE, which combines bias and variability, for estimates of spawning stock biomass, and the probability that error in estimating spawning stock biomass by sub-

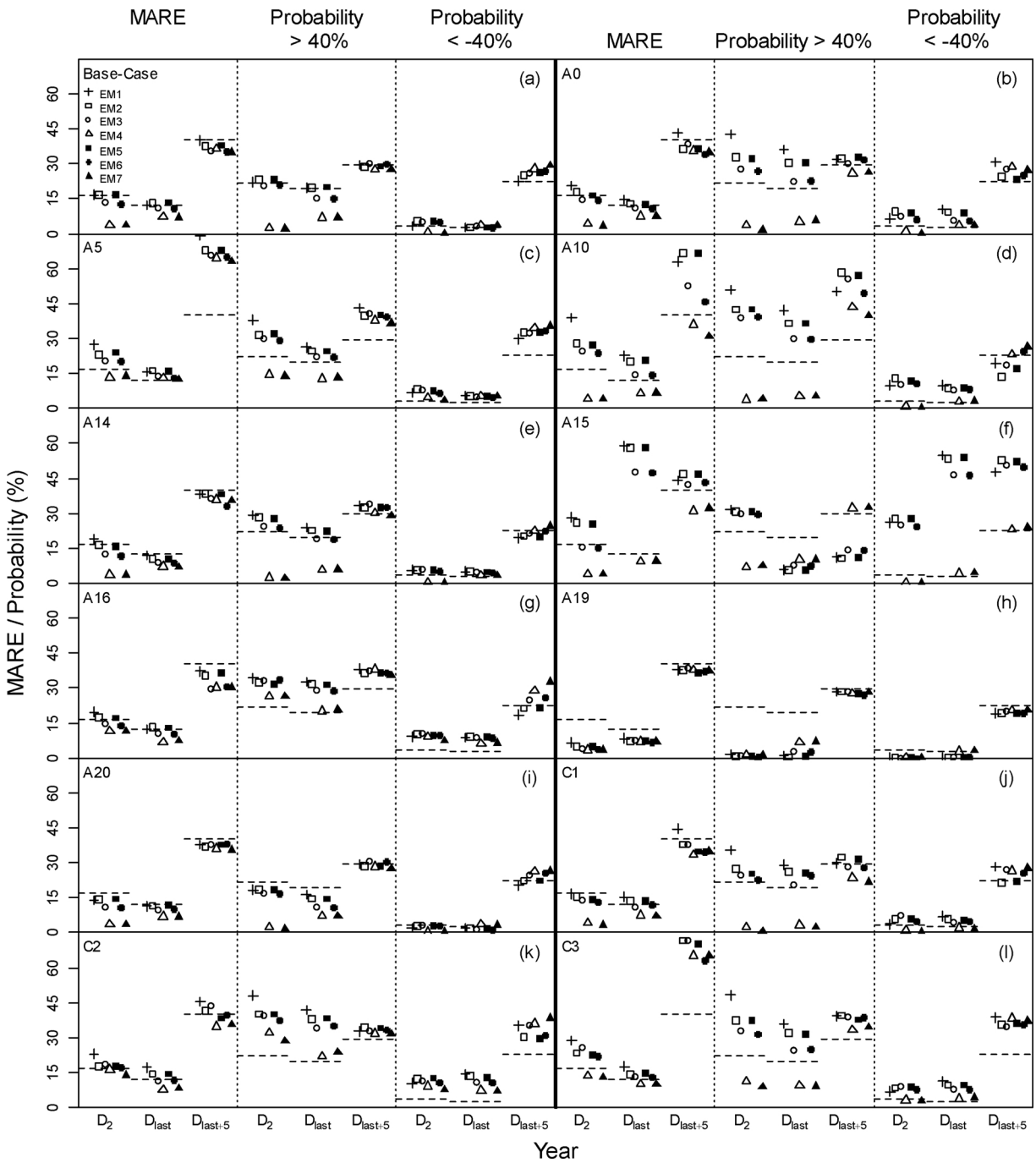

Fig. 6. MAREs for, and, the probability that the estimate of spawning stock biomass by sub-stock in years 2, 50 and 55 exceeds the true value by 40% or more, the probability that the estimate of spawning stock biomass by sub-stock in years 2, 50 and 55 is less than the true value by 40% or more for the base-case scenario and selected sensitivity scenarios. The horizontal lines for each output are the results for EM 1 for the base-case scenario. The symbols in the upper left corner of each panel denote the base-case scenario (BC) and the sensitivity scenarios (See Tables 3–5).

stock exceeds 40% in absolute terms.

In the present study, the simulated sub-stocks have similar time-trajectories of spawning stock biomass for many of the scenarios because there is diffusion and because recruitment and fishing mortality are spatially correlated. However, even with similar time-trajectories among sub-stocks, accounting for spatial structure when conducting the assessment leads to improved estimates of spawning stock biomass for the hindcast period and for short-term (1–2 year) forecasts. Predictably, the estimation methods that match the operating model perform best. This conclusion was true across all the sensitivity scenarios considered, even when all of the estimation methods were mis-specified to some extent. However, estimating sub-stock-specific values for the slope of the origin did not improve estimation performance, likely because the effects of variation in this slope are reduced owing the effects of both high variability in recruitment about the stock-recruitment relationship as well as diffusion. Thus, of the seven estimation methods, EM 4 would be considered best for hindcasts and short-term forecasts given the balance between realism and complexity/number of estimable parameters.

The simpler estimation methods did not outperform the more complex methods when the true number of sub-stocks was less than 10. Overall, if the likelihood that there are 10 or 1 sub-stock is about equal, the benefits of assuming too many sub-stocks exceeds those associated with assuming

too few sub-stocks. The study did not examine the case where the assumed boundaries between sub-stocks are incorrect, which should lead to poorer estimation precision because of increased mis-specification between the population model on which the estimation method is based and the true underlying population structure, and should be the focus of future research. In addition, the paper only considers diffusion as the mechanism linking sub-stocks for animals aged 1 and older. Future work should consider more complex spatial arrangements (e.g. following those examined by McGilliard et al., 2015).

The absolute values of the MAREs should be considered with caution because the estimation methods had information that would not be available in reality, including the correct values for biological parameters such as maturation rates and weight-at-age, the correct forms for diffusion as a function of distance, and the form and parameters of the sampling distributions for the data. However, the relative values for the MAREs (and the probabilities of errors > ± 40%) are informative about the relative performance of the estimation methods and what is estimable. EM 1 and EM 2, which pool data spatially performed poorest for almost all of management quantities and scenarios. Adding some spatial structure by dividing the region to be assessed into two assessment sub-stocks (i.e. EMs 2 and 5) did not substantially improve estimation performance. Only when five (or ten) sub-stocks were considered in the assessment did performance improve markedly. These results suggest that assessments that aim to estimate hindcast biomass and need short-term projections should more routinely be based on population dynamics models that include spatial structure.

Relatively few studies have examined estimation performance of spatially-explicit stock assessments for small pelagic fishes. Kell et al. (2009) explored the implications of uncertainty in stock structure on estimation performance of stock assessment methods for herring in the eastern North Atlantic, where the assessments are based on Virtual Population Analysis rather than integrated analysis, as examined here. The simulations explored scenarios in which the assessment boundaries matched the stock boundaries and in which there were differences. Kell et al. (2009) considered situations in which there was diffusion among sub-stocks. In common with this study, simulated assessments were found to be biased when data from multiple sub-stocks were pooled. Kell et al. (2009) also noted that pooling data spatially meant that the ability to detect overexploitation of stocks was reduced (their Fig. 4), a result also inferred by the present study. Benson et al. (2015) conducted closed loop simulation (aka Management Strategy Evaluation) to examine the management-related implications of uncertainty about abundance and dynamics of Pacific herring. Benson et al. (2015) also considered implementation error caused by fleet spatial distribution patterns. Counter to expectations, they found that these effects did not always lead to increased risk to the resource under the fleet-scenarios they considered.

For species such as Pacific herring that may or may not exhibit clearly-defined spawning aggregations, the challenge remains to develop guidelines for cases where the sub-stock structure is unclear and unidentifiable using genetic or non-genetic methods. The benefits of adopting a spatially-structured approach to stock assessment are reduced, but not eliminated, when data quality is reduced, at least within the range of data quality and precision considered in this paper. The challenge remains to develop guidelines for cases where the sub-stock structure is unclear and unidentifiable using genetic or non-genetic methods.

The factors that had the most marked impacts on estimation performance were stochasticity in diffusion rates, time-varying natural mortality in reality, but not in the estimation method, and skipped spawning. Little is known of the process of connectivity in species such as Pacific herring, but large changes in spawning biomass by spawning site are common, at least in the modern era (Hay et al., 2008, 2009), which could be attributed in part to stochastic diffusion. In contrast, some assessments of Pacific herring in Canada already include time-varying natural mortality, indicating its likely relevance for population dynamics. However, this was not included in the estimation methods examined in this paper owing to computational demands and to keep the number of results at a reasonable level. Skip spawning is thought to be underestimated across a wide suite of species, including herring (Rideout and Tomkiewicz, 2011; Engelhard and Heino, 2005, but see Kennedy et al., 2011). Given the influence of these dynamics on estimation performance, they warrant further investigation.

As expected from other simulation studies, spawning stock biomass was most precisely and accurately estimated for the years where data on age composition and estimates of spawning stock biomass are available. It might have been expected that spawning stock biomass expressed relative to unfished spawning stock biomass would be estimated fairly robustly, but that was not the case. This can be attributed to there being no data for years when fishing intensity was sufficiently low that the stock was close to its unfished level, $\widetilde{D}_{-\infty}$, for several years, a common situation for species such as herring that have been the subject of intensive harvesting for hundreds of years (Jones, 2000). Improved estimates of $\widetilde{D}_{\text{last}}/\widetilde{D}_{-\infty}$ in simulations in which data are assumed to be available from the start of the fishery (results not shown) confirm this conclusion.

Forecast biomass is estimated with much greater error than for the years with data, with the extent of error increasing monotonically with the time since the last monitoring data are available (Supplementary Table 1). In fact, the advantage of correctly knowing the spatial structure is lost when the aim is to estimate biomass 2–3 years beyond the end of the time-series of data on spawning stock biomass and catch age-composition. The consequences of this error can be evaluated using Management Strategy Evaluation (Punt et al., 2016), which we intend to do in future work.

Overall, the results of this study further support efforts to develop spatially-structured population dynamics models for small pelagic fishes, particularly when there is a reasonable likelihood of identifying species sub-stocks for inclusion in the assessment. However, the performance of estimation methods become poorer when 2–5 year forecasts are conducted, and when there is interest in spawning stock biomass relative to the unfished level. This is unfortunate because management advice for some pelagic fishes are currently based on forecasted biomass and spawning stock biomass relative to the unfished level.

The study was not designed to identify the reasons for the better performance of the more complex estimation methods, but it is clear that this study (even for the most data-poor sensitivity scenario) was based on an operating model in which there was sufficient data by sub-stock to support applying models at this level. Thus, analysts wishing to consider whether it is appropriate to move to a more spatially-structured assessment should consider the amount and quality of data by assessment sub-stock. Nevertheless, it is clear that there is value in examining alternative models in which there is sub-stock structure to determine whether the results are qualitatively as well as quantitatively different from those from standard spatially-aggregated approaches. A key future direction is to conduct further research to define guidelines for when assessments should be based on spatially-structured assessment methods given such assessment methods are now available and are computationally feasible. Further, research to consider impacts of fishing and drivers of change with explicit reference to spatial scale is needed to better address multiple fisheries management objectives (social, economic and ecological). Improved insight on these issues should help the agencies who are tasked to conduct assessments make decisions regarding data collection and allocation of resources.

Acknowledgements

Support for this project was provided by the David and Lucile Packard Foundation and by the Pew Charitable Trusts through their funding of the University of Washington's Ocean Modeling Forum. Funding for AP and PL was provided by the Packard Foundation; funding for TF was provided by the Packard Foundation and the Pew Charitable Trusts. DO was supported by a Strategic Grant from the National Sciences and Engineering Research Council of Canada to Simon Fraser University. Carryn De Moor (University of Cape Town) is

thanked for her thorough review of an earlier version of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fishres.2018.04.017.

References

Benson, A.J., Cox, S.P., Cleary, J.S., 2015. Evaluating the conservation risks of aggregate harvest management in a spatially-structured herring fishery. Fish. Res. 167, 101–113.

Clark, F.N., Janssen, J.F., 1945. Movements and abundance of the sardine as measured by tag returns. California Department of Fish and Game. Fish. Bull. 61, 7–42.

Cope, J.C., Punt, A.E., 2011. Reconciling stock assessment and management scales under conditions of spatially-varying catch histories. Fish. Res. 107, 22–38.

Cunningham, C.L., Reid, S.G., McAllister, M.K., Kirkwood, G.P., Darby, C.D., 2007. A Bayesian state-space model for mixed-stock migrations, with application to Northeast Atlantic mackerel *Scomber scombrus*. Afr. J. Mar. Sci. 29, 347–367.

DFO, 2015. Stock Assessment and Management Advice for BC Pacific Herring 2015 Status and 2016 Forecast. DFO (Can. Sci. Advis. Sec. Sci. Resp. 2015/038 56pp). http://www.dfo-mpo.gc.ca/csas-sccs/publications/scr-rs/2015/2015_038-eng.html.

DFO, 2016. Stock Assessment and Management Advice for BC Pacific Herring 2016 Status and Forecast. DFO (Can. Sci. Advis. Sec. Sci. Resp. 2016/052. 56pp.). http://publications.gc.ca/collections/collection_2017/mpo-dfo/Fs70-7-2016-052-eng.pdf.

DFO, 2017. The Selection and Role of Limit Reference Points for Pacific Herring (*clupea pallasii*) in British Columbia, Canada. DFO (Can. Sci. Advis. Sec. Sci. Resp. 2017/030. 24pp.). http://www.dfo-mpo.gc.ca/csas-sccs/Publications/SAR-AS/2017/2017_030-eng.html.

De Moor, C.L., Butterworth, D.S., 2015. Assessing the South African sardine resource: two stocks rather than one? Afr. J. Mar. Sci. 37, 41–51.

Dichmont, C.M., Deng, A.R., Punt, A.E., Venables, W., Haddon, M., 2006. Management strategies for short-lived species: the case of Australia's Northern Prawn Fishery 1. Accounting for multiple species, spatial structure and implementation uncertainty when evaluating risk. Fish. Res. 82, 204–220.

Dougherty, D.T., Hilborn, R., Punt, A.E., Stewart, I.J., 2013. Modeling co-occurring species: a simulation study on the effects of the spatial scale for setting management targets. Can. J. Fish. Aquat. Sci. 70, 49–56.

Engelhard, G.H., Heino, M., 2005. Scale analysis suggests frequent skipping of the second reproductive season in Atlantic herring. Biol. Lett. 2, 172–175.

Epstein, G., Pittman, J., Alexander, S., Berdej, S., Dyck, T., Kreitmair, U., Rathwell, K., Villamayor-Tomas, S., Vogt, J., Armitage, D., 2015. Institutional fit and the sustainability of social–ecological systems. Curr. Opt. Environ. Sustain. 14, 34–40.

Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55, 2105–2116.

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

Fu, C.H., Fanning, L.P., 2004. Spatial considerations in the management of Atlantic cod off Nova Scotia, Canada. N. Am. J. Fish. Manage. 24, 775–784.

Garrison, T.M., Hamel, O.S., Punt, A.E., 2011. Can data collected from marine protected areas improve estimates of life-history parameters? Can. J. Fish. Aquat. Sci. 68, 1761–1777.

Goethal, D.R., Legault, C.M., Cadrin, S.X., 2015. Testing the performance of a spatially explicit tag-integrated stock assessment model of yellowtail flounder (*Limanda ferruginea*) through simulation analysis. Can. J. Fish. Aquat. Sci. 72, 582–601.

Guan, W., Cao, J., Chen, Y., Cieri, M., 2013. Impacts of population and spatial structures on fishery stock assessment. Can. J. Fish. Aquat. Sci. 70, 1178–1189.

Hay, D.E., McCarter, P.B., Kronlund, R., Roy, C., 1989. Spawning areas of British Columbia herring: a review, geographical analysis and classification. Can. MS Rep. Fish. Aquat. Sci. 2019.

Hay, D.E., McCarter, P.B., Daniel, K.S., 2008. Potential impacts of the British Columbia herring roe fishery on the spatial and temporal distribution of herring spawn: examination of the serial depletion hypothesis. Can. Sci. Advis. Secr. Res. Document 2007 (86), 97.

Hay, D.E., McCarter, P.B., Daniel, K.S., Schweigert, J.F., 2009. Spatial diversity of Pacific herring (*Clupea pallasi*) spawning areas. ICES J. Mar. Sci. 66, 1662–1666.

Hurtado-Ferro, F., Punt, A.E., Hill, K.T., 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fish. Res. 158, 102–115.

International Whaling Commission, 2012. The revised management procedure (RMP) for Baleen Whales. J. Cet. Res. Manage. 13 (Suppl), 485–494.

Jones, R., 2000. The herring fishery of Haida Gwaii – an ethical analysis. In: Coward, H., Ommer, R., Pitcher, T. (Eds.), Just Fish – Ethics and Canadian Marine Fisheries. Memorial University of Newfoundland, pp. 201–224 (ISER Social and Economic Papers No. 23).

Kell, L.T., Dickey-Colas, M., Hintzen, N.T., Nash, R.D.M., Pilling, G.M., Roel, B.A., 2009. Lumpers or splitters? Evaluating recovery and management plans for metapopulations of herring. ICES J. Mar. Sci. 66, 1776–1783.

Kennedy, J., Skjæraasen, J.E., Nash, R.D.M., Slott, A., Geffen, A.J., Kjesbu, O.S., 2011. Evaluation of the frequency of skipped spawning in Norwegian spring-spawning herring. J. Sea Res. 65, 327–332.

Levin, P.S., Francis, T.B., Taylor, N.G., 2016. Thirty-two essential questions for understanding the social-ecological system of forage fish: the case of Pacific Herring. Crit. Rev. 2 (4), e01213. http://dx.doi.org/10.1002/ehs2.1213.

Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish Fish. 8, 337–358.

Martien, K.K., Gregovich, D.P., Punt, A.E., 2013. Defining the appropriate 'Unit-To-Conserve' under the international whaling commission's revised management procedure. J. Cet. Res. Manage. 12, 31–38.

McGilliard, C.R., Punt, A.E., Methot, R.D., Hilborn, R., 2015. Accounting for marine reserves using spatial stock assessments. Can. J. Fish. Aquat. Sci. 72, 262–280.

Methot, R.D., Wetzell, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99.

O'Neil, M., Leigh, G.M., Wang, Y.-G., Braccini, J.M., Ives, M.C., 2014. Linking spatial stock dynamics and economics: evaluation of indicators and fishery management for the travelling eastern king prawn (*Melicertus plebejus*). ICES J. Mar. Sci. 71, 1818–1834.

Powell, M., 2012. Divided waters: heiltsuk spatial management of herring fisheries and the politics of native sovereignty. West. Hist. Q. 43, 489–510.

Prager, M.H., 1992. ASPIC: a surplus-production model incorporating covariates. Col. Vol. Sci. Pap. ICCAT 28, 218–229.

Prager, M.H., 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92, 374–389.

Prager, M.H., 2002. Comparison of logistic and generalized surplus-production models applied to swordfish Xiphias gladius, in the north Atlantic Ocean. Fish. Res. 58, 41–57.

Punt, A.E., Methot, R.D., 2004. Effects of Marine Protected Areas on the assessment of marine fisheries. Am. Fish. Soc. Symp. 42, 133–154.

Punt, A.E., Smith, A.D.M., Cui, G., 2002. Evaluation of management tools for Australia's South East Fishery 2: how well can management quantities be estimated? Mar. Freshw. Res. 53, 631–644.

Punt, A.E., Haddon, M., 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.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. Fish Fish. 17, 303–334.

Punt, A.E., 1995. The performance of a production-model management procedure. Fish. Res. 21, 349–374.

Punt, A.E., 1997. The performance of VPA-based management. Fish. Res. 29, 217–243.

Punt, A.E., 2016. Modelling recruitment in a spatial context: a review of current approaches and suggestions for best practices. Fish. Res (in press).

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J., Conser, R.J., 2011. An approach to quantifying scientific uncertainty in stock assessment. Fish. Bull. 109, 217–231.

Rideout, R.M., Tomkiewicz, J., 2011. Skipped spawning in fishes: more common than you might think. Mar. Coat. Fish. 3, 176–189.

Siple, M.C., Francis, T.B., 2016. Population diversity in Pacific herring of the Puget Sound, USA. Oecologia 180, 111–125.

Stick, K.C., Lindquist, A., Lowry, D., 2014. 2012 Washington State Herring Stock Status Report. Fish Program Technical Report No. FPA 14-09. Washington Department of Fish and Wildlife, Olympia, Washington. http://wdfw.wa.gov/publications/01628/.

Thornton, T.F., Kitka, H., 2015. An indigenous model of a contested Pacific herring fishery in Sitka, Alaska. Int. J. Appl. Geospatial. Res. 6, 94–117.

Thornton, T.F., Butler, V.L., Funk, F., Moss, M.L., Hebert, J., Elder, J.T., Craig, R., Hamada, S., Maciejewski Scheer, A., 2010. Herring Synthesis: Documenting and Modeling Herring Spawning Areas Within Socio-ecological Systems over Time in the Southeastern Gulf of Alaska. Final Report, North Pacific Research Board Project #728. Portland State University, Portland, OR. http://herringsynthesis.research.pdx.edu/.