Fisheries Research 238 (2021) 105885

Contents lists available at ScienceDirect

Fisheries Research

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

Hierarchical surplus production stock assessment models improve management performance in multi-species, spatially-replicated fisheries

Samuel D.N. Johnson *, Sean P. Cox

School of Resource and Environmental Management, Simon Fraser University, 8888 University Drive, BC V5K 1S6, Canada

ARTICLE INFO

Handled by A.E. Punt

Keywords:
Data-limited fisheries management
Multi-species fisheries management
Technical interactions
Management strategy evaluation
Hierarchical multi-species stock assessment
Choke species

ABSTRACT

Managers of multi-species fisheries aim to balance harvests of target and non-target species that vary in abundance, productivity, and degree of technical interactions. In this paper, we evaluated management performance of five surplus production stock assessment methods used in such a multi-species context. Production models included single-species and hierarchical multi-species models, as well as methods that pooled data across species and spatial strata. Operating models included technical interactions between species intended to produce choke effects often observed in output controlled multi-species fisheries. Average annual yield of each method under three data scenarios were compared to annual yield obtained by a simulated omniscient manager. Yield and conservation performance of hierarchical multi-species models was superior to all other methods under low, moderate, and high data quantity scenarios. Results were robust to a wide range of prior precision in assessment model biomass parameters, hierarchical prior precision for catchability and productivity, and future survey precision; however, results were sensitive to prior precision in assessment model productivity parameters under the low data scenario, where the hierarchical multi-species method had similar performance to the data pooling models and was no longer clearly the best option.

1. Introduction

Managers of multi-species fisheries aim to balance harvest of multiple interacting target and non-target species that vary in abundance and productivity. Among-species variation in productivity implies variation in single-species optimal harvest rates, and, therefore, differential responses to exploitation. Single-species optimal harvest rates (e.g., the harvest rate associated with maximum sustainable yield) typically ignore both multi-species trophic interactions that influence species' demographic rates (Gislason, 1999; Collie and Gislason, 2001), and technical interactions that make it virtually impossible to simultaneously achieve the optimal harvest rates for all species (Pikitch, 1987).

Technical interactions among species that co-occur in non-selective fishing gear are a defining characteristic of multi-species fisheries (Pikitch, 1987; Punt et al., 2002) and, therefore, play a central role in multi-species fisheries management outcomes for individual species (Ono et al., 2017; Kempf et al., 2016). Catch limits set for individual species without considering technical interactions subsequently lead to sub-optimal fishery outcomes (Ono et al., 2017; Punt et al., 2011a, 2020). For example, under-utilization of catch limits could occur when technically interacting quota species are caught at different rates (i.e., catchability) by a common gear, leading to a choke constraint in which one species quota is filled before the others (Baudron and Fernandes, 2015). Choke constraints are considered negative outcomes for multi-species fishery performance, because they reduce harvester profitability as increasingly rare quota for choke species may limit access to fishing grounds, as well as driving quota costs above the landed value of the choke species (Mortensen et al., 2018).

Setting catch limits for individual species in any fishery usually requires an estimate of species abundance, which continues to be a central challenge of fisheries stock assessment (Hilborn and Walters, 1992; Quinn, 2003; Maunder and Piner, 2015), especially when species data are of low statistical power, such as short noisy time series of observations, or uninformative catch series (Johnson and Cox, 2018). Where such low power data exists, data pooling is sometimes used to extend stock assessments to complexes of similar, interacting stocks of fish (Appeldoorn, 1996). Examples include pooling data for a single species across multiple spatial strata when finer scale data are unavailable or when fish are believed to move between areas at a sufficiently high rate (Benson et al., 2015; Punt et al., 2018), and pooling data for multiple

* Corresponding author.
E-mail addresses: samuelj@sfu.ca (S.D.N. Johnson), spcox@sfu.ca (S.P. Cox).

https://doi.org/10.1016/j.fishres.2021.105885
Received 24 June 2020; Received in revised form 14 January 2021; Accepted 16 January 2021
Available online 16 February 2021
0165-7836/

species of the same taxonomic group within an area when data are insufficient for individual species or during development of new fisheries (DeMartini, 2019). Data-pooled estimates of productivity represent means across the species complex, implying that resulting catch limits will tend to overfish unproductive species and underfish productive ones (Gaichas et al., 2012).

In multi-species and/or multi-area contexts, hierarchical stock assessment models, which treat each area/species combination as a discrete yet exchangeable replicate, may represent a compromise between single-species and data-pooling approaches. For this paper, we define a hierarchical stock assessment model as a model fit to multiple replicates (e.g. areas/species) simultaneously, using hierarchical hyperpriors on selected parameters to share information between replicates (Thorson et al., 2015). Hierarchical priors induce shrinkage effects in which parameter values are drawn towards an estimated overall mean value, thus improving model convergence for replicates with low statistical power data while still estimating replicate-specific parameters based on that data. Hierarchical methods based on data and hyper-priors stand in contrast to data-pooled methods that estimate a mean value only, or single-stock methods that usually rely on strong *a priori* assumptions about replicate specific parameters, forcing parameters to be identical among replicates, or using strongly informative priors, all of which will almost certainly increase assessment bias (Jiao et al., 2009, 2011; Punt et al., 2011b).

Although hierarchical stock assessments are expected to produce better estimates of species biomass and productivity than single-species methods in data-limited contexts, it remains unclear whether such improved statistical performance translates into better management outcomes (Johnson and Cox, 2018). Aside from some related simulations determining the benefits of manually sharing information gained when actively adaptively managing spatially replicated groundfish stocks (Collie and Walters, 1991), to our knowledge there are no evaluations of the management performance of hierarchical stock assessment models. Further, low assessment model bias and/or high precision, which are often unattainable outside of simulations, are not necessary conditions for superior management performance, because biases can, in practice, sometimes compensate for each other (e.g., negative correlation in stock size and productivity), or be offset by other parts of the management system, such as a reduction in harvest rate. A modern fisheries management oriented paradigm is more concerned with the expected performance of a fisheries management system – made up of data, assessments, and harvest rules – despite the inherent, and at some point irreducible, uncertainties in the system (de la Mare, 1998).

In this paper, we investigated whether hierarchical stock assessment models improved management performance in a simulated multi-species, spatially replicated fishery. The simulated fishery was modeled on a spatially heterogeneous complex of Dover sole (*Microstomus pacificus*), English sole (*Parophrys vetulus*), and southern Rock sole (*Lepidopsetta bilineata*) off the coast of British Columbia, Canada, fished in three spatial management areas. Closed-loop feedback simulation was used to estimate fishery outcomes when catch limits were set based on estimates of biomass from single-species, data-pooling, and hierarchical state-space surplus production models under high, moderate, and low data quantity scenarios. Assessment models were either fit to species-specific data as single-species or hierarchical multi-species models, or fit to data pooled spatially across management units, pooled across species within a spatial management unit, or totally aggregated across both species and spatial management units. Management performance of each assessment approach was measured by both the risk of overfishing, and by cumulative absolute loss in catch, defined as deviation from optimal catch trajectories generated by an omniscient manager, who could set annual effort to maximize total multi-species/multi-stock complex yield given perfect knowledge of all future recruitments (Walters, 1998; Martell et al., 2008).

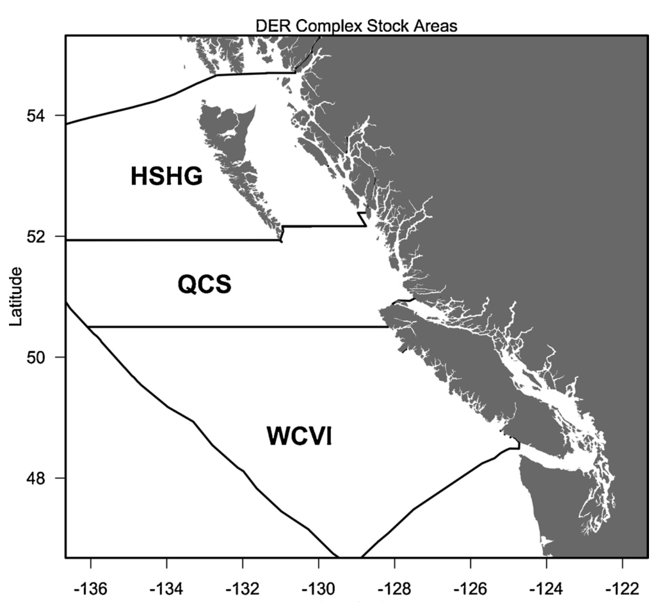

Fig. 1. Boundaries for each of the DER complex stock areas on the BC coast, showing, from north to south, Hecate Strait/Haida Gwaii (HSHG), Queen Charlotte Sound (QCS), and West Coast of Vancouver Island (WCVI).

2. Methods

2.1. British Columbia's flatfish fishery

The multi-species complex of right-eyed flounders in British Columbia (BC) is a technically interacting group of flatfishes managed over the BC coast (Fig. 1). Although there are several right-eyed flounders in BC waters, we focus on the three species, indexed by s, Dover sole ($s = 1$), English sole ($s = 2$), and southern Rock sole ($s = 3$), which we hereafter refer to as Rock sole. Taken together, these species comprise a multi-stock complex (DER complex), managed as part of the BC multi-species groundfish fishery.

The DER complex is managed in three spatially distinct stock areas, indexed by p (Fig. 1) (Fisheries and Oceans, Canada, 2015). From north to south, the first stock area – Hecate Strait/Haida Gwaii (HSHG, $p = 1$) – extends from Dixon Entrance and north of Haida Gwaii, south through Hecate Strait. The second stock area – Queen Charlotte Sound (QCS, $p = 2$) – extends from the southern tip of Haida Gwaii to the northern tip of Vancouver Island. Finally, the third area – West Coast of Vancouver Island (WCVI, $p = 3$) – extends from the northern tip of Vancouver Island south to Juan de Fuca Strait. These areas are aggregates of PFMA major statistical areas, and are primarily delineated by management breaks, although there are oceanographic features that may separate each of the stocks as defined here (e.g., gullies in the QCS area, and strong currents at the northern tip of Vancouver Island).

Each species/area combination had commercial trawl catch for the entire history (1956–2016), two commercial catch rate series, and at least one survey biomass index time series for each species (Fig. 2), for a total of four distinct fleets, indexed by f. The two commercial fishery catch-per-unit-effort (CPUE) series span 1976–2016, split between a Historical trawl fishery (1976–1995, $f = 1$) and a Modern trawl fishery (1996–2016, $f = 2$), with the split corresponding to pre- and post-implementation of an at-sea-observer programme, respectively. The fishery independent trawl survey biomass indices were the biennial Hecate Strait Assemblage survey in the HSHG area (1984–2002, $f = 3$), and the Multi-species Synoptic Trawl Survey, which operated every year but alternates between areas, so it is effectively a biennial survey in all three areas (2003–2016, $f = 4$).

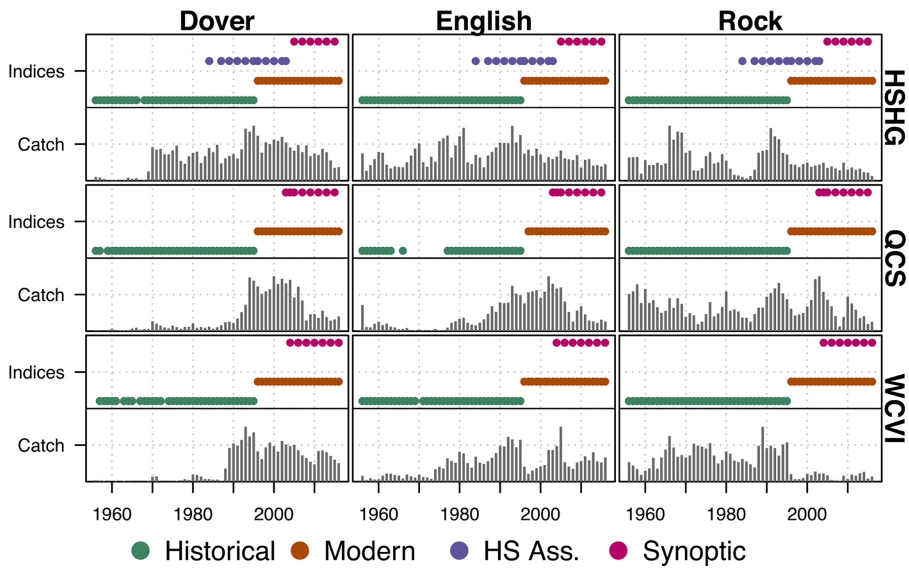

Fig. 2. A summary of DER complex data available for fitting the assessment model. Panels show species from left to right, and rows show stock areas from north to south. Each panel is split into two cells, with the top showing a filled circle when a commercial CPUE or biomass index observation exists. The bottom cell shows commercial catch (relative to the maximum catch) in grey bars.

2.2. Closed-loop feedback simulation framework

Closed-loop simulation is often used to evaluate proposed feedback management systems. In fisheries, closed-loop simulation evaluates fishery management system components, such as stock assessment models or harvest decision rules, by simulating repeated applications of these components, while propagating realistic errors in monitoring data, stock assessment model outputs, and harvest advice (de la Mare, 1998; Cox and Kronlund, 2008; Cox et al., 2013). At the end of the simulation, pre-defined performance metrics are used to determine the relative performance of the system components being tested.

Our closed-loop simulation framework included a stochastic operating model, representing stock and fishery dynamics and generating observations with uncertainty, and an assessment model component that estimated stock biomass from simulated observations and fishery catches. The operating model simulated population dynamics of a spatially stratified, multi-species flatfish complex in response to a multi-species trawl fishery in each of the three stock areas. Although total fishing effort was not restricted across the entire area, effort in each individual area was allocated such that no species-/area-specific catch exceeded the species-/area-specific total allowable catch (TAC). Within an area, fishing effort was allowed to increase until at least one species-/area-specific TAC was fully caught.

The simulation projected population dynamics for each species in each area (total nine stocks) forward in time for 32 years, or two Dover sole generations (longest generation time; Seber, 1997), with annual simulated assessments, harvest decisions, and catch removed from the population from 2017–2048. The following four steps summarise the closed-loop simulation procedure for each projection year t:

1. Update stochastic population dynamics and generate new realized catch $C_{s,p,t}(E_{p,t})$ in each area from effort (Eq. (1));
2. Generate new observation data $I_{s,p,f,t}$ (Eqs. (4)–(7))
3. Apply an assessment model (defined below) to estimate the spawning biomass for the upcoming year $\widehat{B}_{s,p,t+1}$ and an optimal harvest rate $\widehat{U}_{\mathrm{MSY},s,p}$ (Eq. (3));
4. Scale the estimated optimal harvest rate to an estimated multi-species optimal harvest rate, and use it with the biomass forecast to generate a total allowable catch $\mathrm{TAC}_{s,p,t+1}$, allocated among species/areas if using a pooled method (Eqs. (8)–(13));

Table 1
Unfished biomass B_0, single-species MSY based reference points $B_{MSY,SS}$, MSY_{SS}, and $U_{MSY,SS}$, stock status as absolute biomass in 2016 B_{2016}, depletion relative to single-species optimal biomass $B_{2016}/B_{MSY,SS}$, commercial trawl catchability scalar q^F, and multi-species reference points including technical interactions $B_{MSY,MS}$, MSY_{MS} and $U_{MSY,MS}$ for all nine DER complex stocks in 2016. Biomass quantities are given in kilotonnes, and depletion levels and harvest rates are unitless.

Stock	B_0	SS reference points			Stock status		q^F	MS reference points		
		$B_{MSY,SS}$	MSY_{SS}	$U_{MSY,SS}$	B_{2016}	$B_{2016}/B_{MSY,SS}$		$B_{MSY,MS}$	MSY_{MS}	$U_{MSY,MS}$
Dover sole										
HSHG	16.51	4.34	1.22	0.28	8.36	1.92	0.022	5.77	1.18	0.20
QCS	5.45	1.46	0.42	0.28	3.36	2.30	0.015	1.91	0.40	0.21
WCVI	13.59	3.58	1.14	0.32	8.45	2.36	0.025	3.60	1.14	0.32
English sole										
HSHG	8.60	2.21	0.87	0.39	4.85	2.20	0.026	2.31	0.87	0.38
QCS	0.57	0.15	0.06	0.39	0.41	2.78	0.016	0.16	0.06	0.36
WCVI	0.86	0.22	0.09	0.39	0.50	2.23	0.020	0.20	0.09	0.43
Rock sole										
HSHG	12.34	3.78	1.07	0.28	7.68	2.03	0.025	2.81	1.03	0.37
QCS	4.33	1.29	0.40	0.31	2.13	1.65	0.014	1.04	0.39	0.38
WCVI	1.12	0.34	0.10	0.29	0.55	1.62	0.012	0.35	0.10	0.28

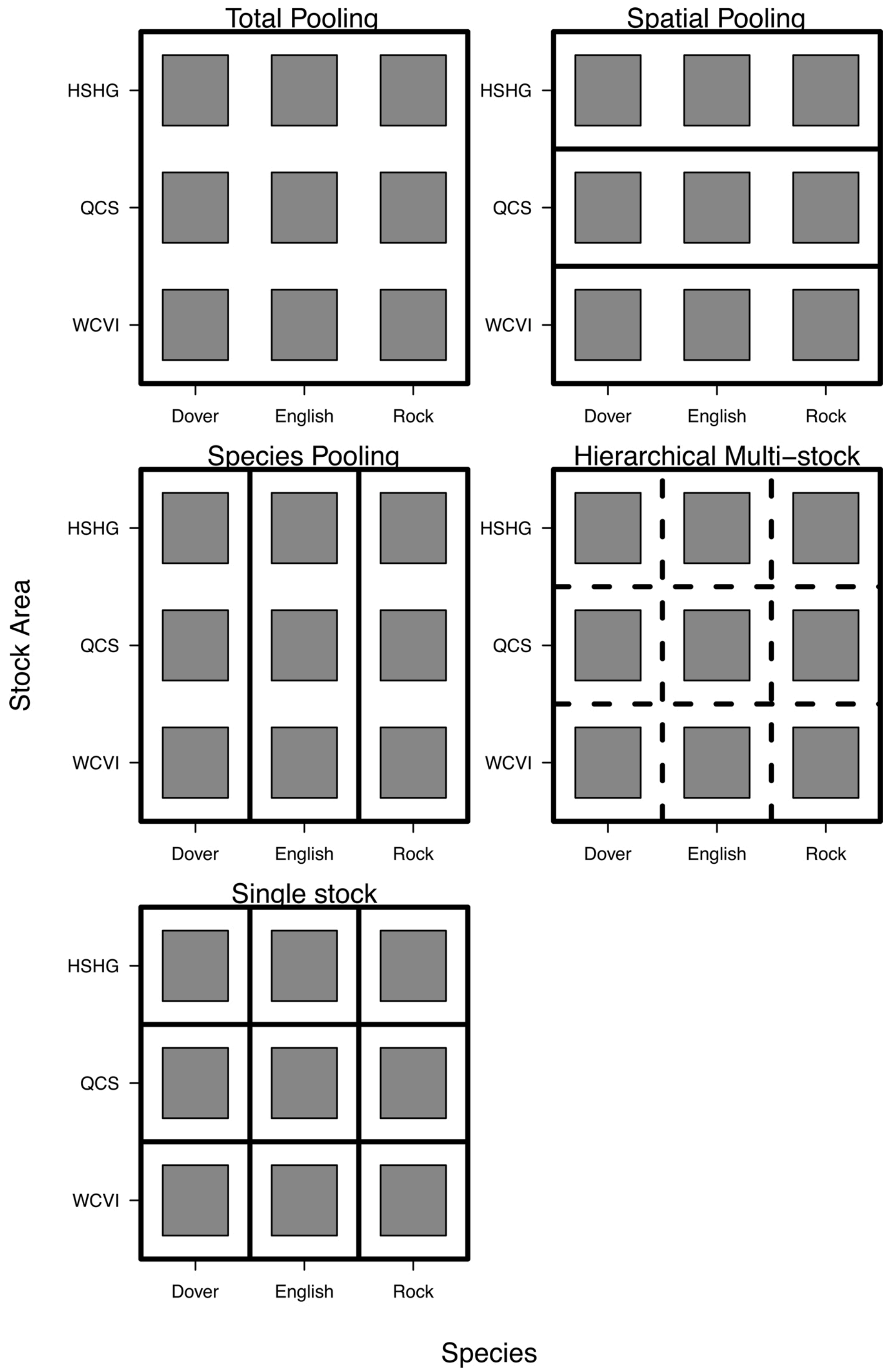

Fig. 3. Conceptual models of the five assessment model configurations. In each panel, the nine grey boxes represent each DER complex population, as indicated by the axis labels. Data are pooled for any population not separated by a black line, e.g., all nine are pooled in the total pooling case. In the hierarchical model, the broken lines indicate that data are separated, but information is shared between populations via the statistical model's hierarchical prior distributions.

5. Allocate effort $E_{p,t+1}$ to fully realise at least one TAC from step 4 in each stock area (Eq. (2)).

2.2.1. Operating model

The operating model (OM) was a multi-species, multi-stock age- and sex-structured population dynamics model (Appendix A). Population life-history parameters for the operating model were estimated by fitting a hierarchical age- and sex-structured model to data from the real DER complex.

Fishing mortality for individual stocks was scaled to commercial trawl effort via species-specific catchability parameters, i.e.,

$$F_{s,p,t} = q^F_{s,p} \cdot E_{p,t}, \tag{1}$$

where $F_{s,p,t}$ is the fishing mortality rate applied to species s in stock-area p by fleet f in year t, and $q^F_{s,p}$ is the commercial catchability coefficient scaling trawl effort $E_{p,t}$ in area p to fishing mortality (Table 1). The relationship in Eq. (1) implies a multi-species maximum yield $\mathrm{MSY}_{\mathrm{MS},p}$ within each area, from which we derived the corresponding effort $E_{\mathrm{MSY,MS},p}$, species yield $\mathrm{MSY}_{\mathrm{MS},s,p}$, the biomass $B_{\mathrm{MSY,SS},s,p}$ at which multi-species maximum yield is achieved, and the optimal harvest rate

$U_{\mathrm{MSY,MS},s,p} = \mathrm{MSY}_{\mathrm{MS},s,p}/B_{\mathrm{MSY,MS},s,p}$ that produced multi-species maximum yield, which are all given in Table 1 (Pikitch, 1987). The single-species optima, which are also in Table 1, are derived in the usual way by maximising single-species yield $\mathrm{MSY}_{\mathrm{SS},s,p}$, produced at the optimal biomass $B_{\mathrm{MSY,SS},s,p}$ by applying the optimal harvest rate $U_{\mathrm{MSY,SS},s,p} = \mathrm{MSY}_{\mathrm{SS},s,p}/B_{\mathrm{MSY,SS},s,p}$.

For simulated fishing in the projection time period, the fully realised TAC depended on the choke effects determined by relative catchabilities and absolute biomass levels of each species within an area. To simulate the choke effects, we allocated the maximum fishing effort to each area required to fully utilize the TAC of at least one species, but never exceed the TAC of any one of the three species, i.e.,

$$E_{p,t+1} = max\{E \mid C_{s,p,t+1}(E) \leq \mathrm{TAC}_{s,p,t+1} \forall s\}, \tag{2}$$

where $C_{s,p,t+1}(E)$ is the catch of species s when effort E is applied in area p in year $t+1$ (the method for determining TACs is explained below). We used maximum effort instead of an explicit effort dynamics model because the former captured choke effects present in the real fishery system, while reflecting the output controlled BC groundfish fishery, for which there is no explicit limit on the total amount of fishing effort expended by license holders.

2.2.2. Surplus production stock assessment models

At each time step t, simulated annual assessments were used to estimate the expected future spawning biomass estimate $\widehat{B}_{t+1}$ via a state-space Schaefer production model (Schaefer, 1954; Punt, 2003), modified to better approximate the biomass-yield relationship underlying the age-/sex-structured operating model (Pella and Tomlinson, 1969; Winker et al., 2018). We extended the Johnson and Cox (2018) hierarchical state-space model to fit a multi-species, spatially stratified complex, as well as fit a single-stock model to data from individual or data-pooled stocks, via the biomass Eq.

$$B_{s,p,t+1} = \left[B_{s,p,t} + U_{\mathrm{MSY},s,p} \cdot \frac{m_{s,p}}{m_{s,p}-1} \cdot B_{s,p,t} \cdot \left(1 - \frac{1}{m_{s,p}} \cdot \left(\frac{B_{s,p,t}}{B_{\mathrm{MSY},s,p}} \right)^{m_{s,p}-1} \right) - C_{s,p,t} \right] e^{\zeta_{s,p,t}}, \tag{3}$$

where the management parameters $\mathrm{MSY}_{s,p}$ (optimal single-species yield) and $U_{\mathrm{MSY},s,p}$ (optimal single-species harvest rate) are the leading model parameters, $B_{\mathrm{MSY},s,p} = \mathrm{MSY}_{s,p}/U_{\mathrm{MSY},s,p}$ is the biomass at which $\mathrm{MSY}_{s,p}$ is acheived under the optimal harvest rate, $m_{s,p}$ is the Pella-Tomlinson parameter controlling skew in the biomass/yield relationship (derived from operating model yield curves), and $\zeta_{s,p,t}$ are annual process error deviations.

In total, we defined the following five potential assessment model configurations (Fig. 3):

1. Total Pooling (1 management unit);
2. Species Pooling (3 management units, independent parameters);
3. Spatial Pooling (3 species, independent parameters);
4. Single-species (9 management units, independent parameters);
5. Hierarchical Multi-species (9 management units, hierarchical priors);

where the number of management units is shown in parentheses (i.e., delete subscripts in Eq. (7) for species or stock-area as appropriate, e.g., Spatial Pooling models have no p subscript).

Prior distributions on optimal yield $\mathrm{MSY}_{s,p}$, optimal harvest rate $U_{\mathrm{MSY},s,p}$, catchability $q_{s,p,f}$, and process error deviations $\zeta_{s,p,t}$ were defined for each assessment model, with unbiased prior means based on the true operating model values (Online supplmental, Table S.1). While unbiased prior distributions are not possible to define in practice, our research question focused on the effect of the five AM structures on management performance when setting TACs based on AM estimates of biomass and productivity. While priors certainly affect those estimates, there is considerable variation in how an analyst might choose prior mean values for biological parameters, and those choices become more uncertain as data are reduced. For example, an analyst might choose a prior mean of average catch over an agreed historical period as the MSY prior mean, and choose either a conservatively low value based on a review of stock assessments for similar species, or if it is available, the ratio of catch to an index of biomass in the same historical period to inform a prior mean U_{MSY} value. In cases where time-series of data are short, there may be no appropriate period for formulating such priors, while noisy data (e.g., commercial CPUE) lack sufficient statistical power to draw parameter estimates away from biased mean values, especially when priors may be informative in the presence of data limitations. We based our unbiased priors on the OM to keep priors as similar as possible among AMs, thereby avoiding variation in performance caused by different prior choices.

2.2.3. Data generation for assessment models

Time series of catch, commercial CPUE, and relative biomass indices were simulated in the historical and projection periods for fitting assessment models. While it is more usual to use the actual historical data in most closed loop simulations, we decided that given that pooled data must be simulated, simulating all data series removed any effect of the difference between simulated and real data on performance between pooled and non-pooled AMs. The Historical fishery CPUE and Hecate Strait Assemblage survey were simulated only for the periods shown in Fig. 2, but the Modern fishery CPUE and Synoptic trawl survey biomass index were also simulated in the projection period. All biomass indices were simulated with log-normally distributed observation error deviations with precision based on the conditioning assessment (Table A.1, Appendix A).

Biomass indices for individual stocks were all simulated as relative biomass with observation error

$$\bar{I}_{s,p,f,t} = q_{s,p,f} \cdot B_{s,p,f,t}, \tag{4}$$

where $\bar{I}_{s,p,f,t}$ is the index without observation error, $q_{s,p,f}$ is the catchability coefficient for species s and stock-area p, and $B_{s,p,f,t}$ is the biomass of species s in stock-area p vulnerable to fleet f in year t. For the Assemblage and Synoptic fishery independent surveys ($f = 3, 4$), the catchability coefficients $q_{s,p,f}$ were trawl efficiency estimates from the conditioning assessment. For commercial CPUE, catchability coefficients were estimated by the conditioning assessment by assuming that catch rates were an unbiased relative index. While the assumption that catch rates were an unbiased relative index likely increased the information contained in the simulated CPUE indices over the real data, the large CVs estimated by the conditioning assessment retain realistic variation. Moreover, simulating commercial data as relative biomass reduced differences in the variability and bias of the data simulated in the OM history and data simulated in the projection.

The method for generating pooled catch and biomass index data was analogous to the unpooled data, but combined via summation. Catch data was summed without any scaling, and biomass indices without observation error were summed after scaling by catchability (trawl efficiency in the surveys), i.e.,

$$\bar{I}^{\mathrm{pooled}}_{s,f,t} = \sum_p 1(I_{s,p,f,t} > 0) \cdot q_{s,p,f} \cdot B_{s,p,f,t} \tag{5}$$

$$\bar{I}^{\mathrm{pooled}}_{p,f,t} = \sum_s 1(I_{s,p,f,t} > 0) \cdot q_{s,p,f} \cdot B_{s,p,f,t} \tag{6}$$

$$\bar{I}^{\mathrm{pooled}}_{f,t} = \sum_{s,p} 1(I_{s,p,f,t} > 0) \cdot q_{s,p,f} \cdot B_{s,p,f,t} \tag{7}$$

where $f = 1,\ldots,4$, $\bar{I}^{\text{pooled}}_{s,f,t}$ is the spatially pooled index for species s, $\bar{I}^{\text{pooled}}_{p,f,t}$ is a species pooled index for area p, $\bar{I}^{\text{pooled}}_{f,t}$ is the totally aggregated index (all without error), and $1(I_{s,p,f,t} > 0)$ is the indicator function that takes value 1 when survey f in area p took samples in year t, and 0 otherwise.

2.2.4. Target harvest rates and total allowable catch

Simulated harvest decision rules applied a constant target harvest rate to generate TACs from one-year ahead biomass forecasts obtained from each assessment model, i.e.,

$$\text{TAC}'_{s,p,t+1} = \widehat{U}_{s,p,t} \cdot \widehat{B}_{s,p,t+1}, \tag{8}$$

where $\widehat{U}_{s,p,t}$ is the estimated target harvest rate for species s in stock-area p, and $\widehat{B}_{s,p,t+1}$ is the year $t+1$ biomass forecast from the assessment model applied in year t. The estimated target harvest $\widehat{U}_{s,p,t}$ was defined to incorporate realistic assessment errors in productivity estimates while simultaneously targeting maximum multi-species yield $\text{MSY}_{\text{MS},s,p}$. To avoid incorporating technical interactions directly into the AM dynamics, the target harvest rate scaled the AM estimate of $\widehat{U}_{\text{MSY},s,p}$ from Eq. (3) by the ratio of multi-species and single-species optimal harvest rates from the operating model, i.e.,

$$\widehat{U}_{s,p,t} = \widehat{U}_{\text{MSY},s,p} \cdot \frac{U_{\text{MSY},\text{MS},s,p}}{U_{\text{MSY},\text{SS},s,p}} \tag{9}$$

where $U_{\text{MSY},\text{MS},s,p}, U_{\text{MSY},\text{SS},s,p}$ are the optimal harvest rates maximising multi-species and single-species yield, respectively, taken from the operating model (Table 1), and $\widehat{U}_{\text{MSY},s,p}$ is the assessment model estimate of the single-species optimal harvest rate applied at time t.

Inter-annual increases in TAC were limited to 20% for all individual stocks, i.e.,

$$\text{TAC}_{s,p,t+1} = min\{\text{TAC}'_{s,p,t+1}, 1.2 \cdot \text{TAC}_{s,p,t}\}, \tag{10}$$

where $\text{TAC}'_{s,p,t+1}$ is the proposed TAC determined above, and $\text{TAC}_{s,p,t}$ is the previous year's TAC. This constraint on inter-annual TAC changes reflects a constraint on inter-annual changes in fishing effort in the objective function used for the omniscient manager solutions described below, which is meant to simulate gradual investment in additional fishing effort.

Pooled TACs were set analogously to the stock-specific case above, with pooled target harvest rates applied to biomass projections from pooled assessments. For a spatially pooled assessment of species s, we defined the operating model spatially pooled optimal harvest rates as

$$U_{\text{MSY},\text{MS},s} = \frac{\sum_p \text{MSY}_{\text{MS},s,p}}{\sum_p B_{\text{MSY},\text{MS},s,p}}, \tag{11}$$

$$U_{\text{MSY},\text{SS},s} = \frac{\sum_p \text{MSY}_{\text{SS},s,p}}{\sum_p B_{\text{MSY},\text{SS},s,p}}, \tag{12}$$

where the notation is as defined above, with species pooled and totally pooled rates defined analogously. For setting TACs under pooled assessments, the harvest rate scalar from single- to multi-species U_{MSY} in Eq. (13) used the ratio of pooled $U_{\text{MSY},\text{MS}}$ and $U_{\text{MSY},\text{SS}}$ values defined as in equations 15 and 16. Assessment model estimates of pooled optimal harvest rates were then scaled by the ratio of the pooled operating model optimal harvest rates.

Pooled TACs were split within an area or across spatial strata proportional to Synoptic trawl survey indices for the individual stocks. For example, if the TAC for area p is set by a species pooled assessment, then the proposed TAC for species s is defined as

$$\text{TAC}'_{s,p,t+1} = \frac{I_{s,p,t^*}}{\sum_{s'} I_{s',p,t^*}} \text{TAC}_{p,t+1}, \tag{13}$$

where $\bar{I}_{s,p,t^*}$ is the most recent individual biomass index from the Synoptic survey for species s in area p. The most recent index is used because the synoptic survey alternates between areas each year, so not all individual indices are present in a given year.

2.3. Simulation experiments and performance

We ran a total of 15 simulation experiments comprising five assessment models and three data quality scenarios. Simulations integrated over the stochastic processes by running a total of 200 random replicates of each combination, where each simulation was initialized with the same set of random seeds to eliminate random effects among combinations of assessments and data scenarios. Assessment convergence was defined as a positive definite Hessian matrix and a maximum gradient component less than 10^{-3} in absolute value. Replicates were considered significant when assessments were convergent in 95% of time steps, chosen to reflect that fitting models becomes more difficult as data quantity is deliberately reduced, and a simulated assessment cannot always be tuned like a real assessment performed by a real-life analyst. Results were then calculated based on the first 100 random seed values that produced significant replicates for all species and stocks considered. The operating model was run for two Dover sole generations (32 years; Seber, 1997), because this species had the longest generation time.

Operating model population dynamics were identical among replicates for each stock during the operating model historical period, except for the last few years near the end, where the operating model simulated recruitment process errors because they were not estimated in the conditioning assessment. Simulated log-normal observation and process errors in the projection were randomly drawn with the same standard deviations as the errors used in the historical period, and bias corrected so that asymptotic medians matched their expected values, i.e., for the two fishery independent surveys ($f = 3, 4$), the species/area specific biomass indices were simulated as

$$I_{s,p,f,t} = \bar{I}_{s,p,f,t} \cdot exp(\tau_{s,p,f} \cdot \delta_{s,p,f,t} - 0.5\tau^2_{s,p,f}) \tag{14}$$

where $\bar{I}_{s,p,f,t}$ is the index without error defined above, $\tau_{s,p,f}$ is the log-normal observation error standard deviation, $\delta_{s,p,f,t}$ is the annual standard normal observation error residual, and subscripts s, p, f, t are for species, stock, fleet and year, respectively. Recruitments are simulated as in Appendix A. Error is added to survey biomass indices for pooled data independently of the error added to individual indices, i.e.,

$$I_{s,f,t} = \bar{I}_{s,f,t} \cdot exp(\tau_{s,f} \cdot \delta_{s,f,t} - 0.5\tau^2_{s,f}) \tag{15}$$

$$I_{p,f,t} = \bar{I}_{p,f,t} \cdot exp(\tau_{p,f} \cdot \delta_{p,f,t} - 0.5\tau^2_{p,f}) \tag{16}$$

$$I_{f,t} = \bar{I}_{f,t} \cdot exp(\tau_f \cdot \delta_{f,t} - 0.5\tau^2_f) \tag{17}$$

where $\tau_{s,f}, \tau_{p,f}, \tau_f$ were averaged over the components of the pooled index.

2.3.1. Operating model data quantity scenarios

The three data quantity scenarios ranged from a high to a low quantity of data by successively removing the earlier CPUE index series from the full set, i.e.,

1. **High**-data: Historical CPUE (1956–1996), Modern CPUE (1996 onwards), Assemblage survey (1984–2002, biennial, HSHG only), Synoptic survey (2003 onwards, biennial);
2. **Mod**erate-Data: Modern CPUE, Assemblage survey, Synoptic survey;
3. **Low**-Data: Assemblage survey, Synoptic survey.

To improve convergence, the Hierarchical Multi-stock and Single-species assessment models were initialised later under the **Mod** and **Low** data scenarios, with the starting year of the assessments set to the

Table 2
Summary of sensitivity analyses, showing the total number of experiments, the factor being varied, the levels of that factor, and the data scenarios and AMs included in the analysis.

N	Factor	Levels	Scenarios	AMs
30	*MSY* prior CV	0.1, 0.5, 1.0	High, Low	All
30	U_{MSY} prior SD	0.1, 0.5, 1.0	High, Low	All
6	Hierarchical prior SDs τ_q, $\sigma_{U_{MSY}}$	0.1, 0.2, 0.5	High, Low	Hierarchical only
30	Synoptic survey SD τ	0.1,0.5,1.0	High, Low	All

first year with index data, which was 1984 in HSHG for both scenarios, and 1997 or 2003 for other areas under the **Mod** and **Low** scenarios, respectively.

2.3.2. Performance evaluation

2.3.2.1. Omniscient manager simulations. Assessment model performance was measured against a simulated omniscient fishery manager who was aware of all the future consequences of harvest decisions and was, therefore, able to adapt the management to meet specific quantitative objectives under any process error conditions (Walters, 1998). Omniscient manager solutions were used rather than equilibrium based metrics (Punt et al., 2016) because most stocks were in a healthy state in 2016 (i.e. above single-species B_{MSY}, Table 2) and, therefore, the time-path of fishery development was important (Walters, 1998).

The omniscient manager was implemented as an optimisation of future fishing effort by area (Appendix B), with the objective function defined as

$$\mathcal{O} = \left[\sum_{s,p} - log(\overline{C}_{s,p,\cdot})\right] + \mathcal{P}_{\text{diff}}\left(\sum_p E_{p,\cdot}\right) + \mathcal{P}_{\text{init}}\left(\sum_p E_{p,2017}\right) + \mathcal{P}_{\text{overfished}}(B_{s,p,\cdot}), \quad (18)$$

where $-log\overline{C}_{s,p,\cdot}$ is the negative log of total future catch for species s in area p over the projection period (equivalent to maximising catch). Penalty functions $\mathcal{P}$ (Eq. B.1) were applied for annual changes in total effort across all three areas being above 20% ($\mathcal{P}_{\text{diff}}$) to match the TAC smoother in stochastic experiments, differences greater than 10% between the last year of historical effort and the first year 2017 of simulated effort ($\mathcal{P}_{\text{init}}$), and a penality when biomass dropped below a lower threshold of 40% of single species $B_{\text{MSY,SS}}$ more than 5% of time-steps ($\mathcal{P}_{\text{overfished}}$). The threshold of $0.4B_{\text{MSY,SS}}$ is commonly suggested as a limit reference point for Canadian fisheries (DFO, 2006), below which the stock could experience irreparable harm and recruitment could become impaired.

An omniscient manager solution was obtained for each stochastic trajectory in the stochcastic management simulations. Each replicate was run for 80 years to produce several years free of end effects, such as transient dynamics at the beginning of the projection, or reduced penalties for overfishing at the end of the projection.

2.3.2.2. Cumulative catch loss. For each stochastic trajectory, the cumulative absolute loss in catch was calculated as (Walters, 1998):

$$L_{i,s,p} = \sum_{t=T_1}^{T_2} |C_{i,s,p,t,\text{sim}} - C_{i,s,p,t,\text{omni}}|, \quad (19)$$

where the $C_{i,s,p,t,\cdot}$ values were commercial trawl catch for replicate i,

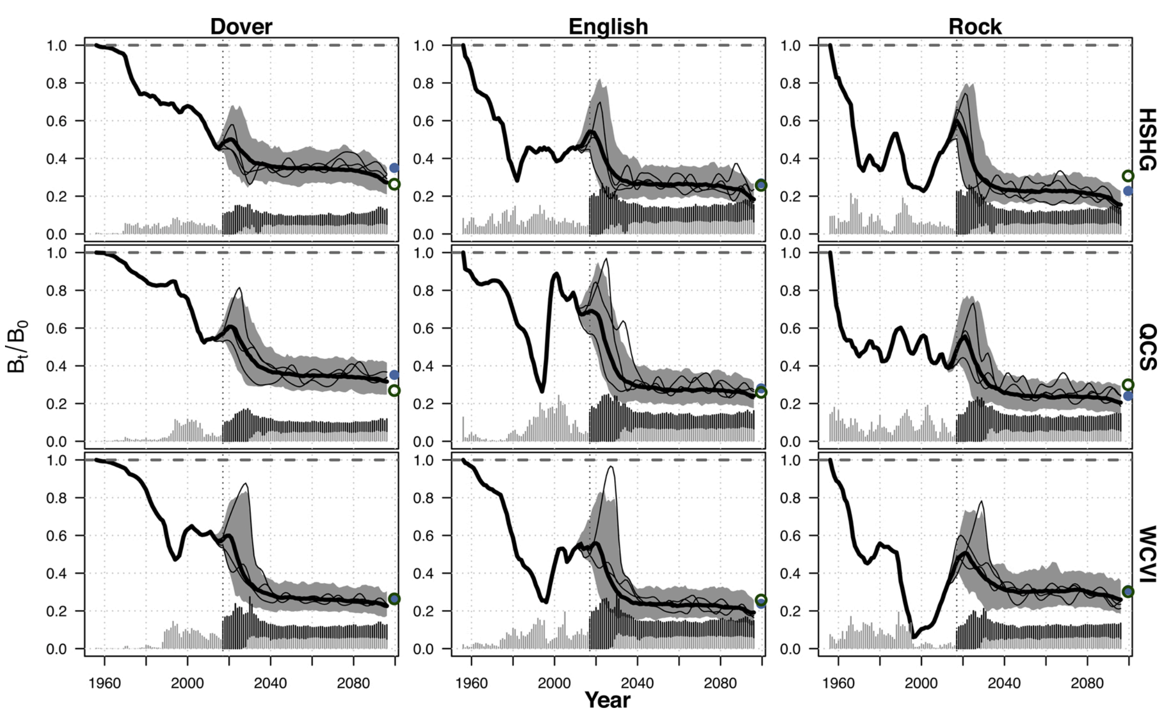

Fig. 4. Spawning biomass depletion and relative catch simulation envelopes for all nine DER complex management units from the omniscient manager simulations. Median biomass is shown by the thick black line, with the grey region showing the central 95% of the distribution of spawning biomass, and thin black lines showing three randomly selected simulation replicates. Catch is shown as grey bars in the historical period, which represent median catch in the projection, with thin vertical line segments showing the central 95% of the catch distribution. The depletion level associated with the traditional single species optimal biomass $B_{MSY,SS}$ is shown as an open green circle on the right-hand axis, while the depletion level associated with the multi-species maximum yield is shown as a blue closed circle on the right-hand axis.

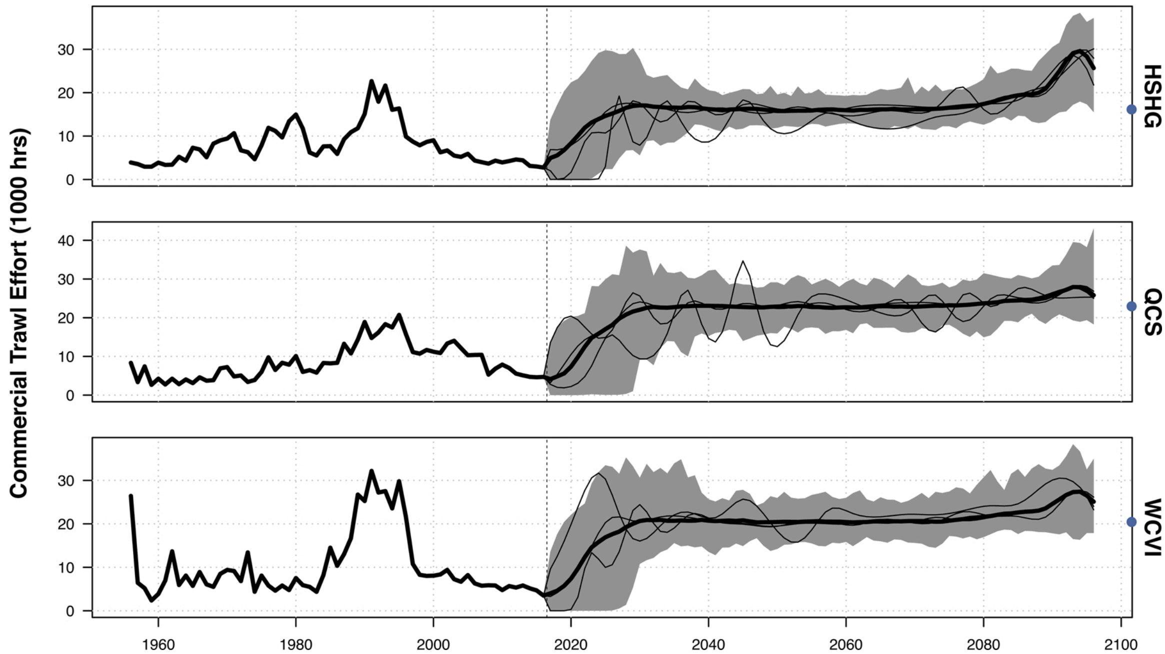

Fig. 5. Commercial fishing effort simulation envelopes for each stock area under the omniscient manager simulation. Historical and median simulated effort in the projection period are shown by a thicker black line, while the central 95% of the distribution of simulated effort in the projection period is shown as grey shaded region, with single simulation replicates shown as thinner black lines. The effort level that achieves multi-species maximum yield is shown as a closed blue circle on the right-hand axis.

species s, and stock p from stochastic simulations (sim) or the omniscient manager simulation (omni) simulation, with $C_{i,s,p,t,\text{sim}} - C_{i,s,p,t,\text{omni}}$ defined simply as catch loss, which was positive when catch was higher than the omniscient manager's, and negative when lower. When repeated over all significant random seed values, the loss functions generated a distribution of cumulative absolute catch loss, which were then used to determine relative performance of each assessment model under the three data scenarios. Cumulative absolute catch loss was calculated for the ten year period $T_1 = 2028$ to $T_2 = 2037$, chosen in the middle of the projection period because dynamics in the earlier period were dominated by the smoothers on effort and catch for the omniscient manager and TACs, respectively.

A paired analysis was used to rank AM configurations across species, stocks and replicates. Within each replicate i, each AM's cumulative absolute catch loss determined the relative rank of each AM under a species/stock/OM scenario combination, where lower loss ranked higher. Any replicates with less than 95% convergence rates for any AM on any management units were excluded from the aggregate rankings, and any species/stock combinations that failed to reach 100 replicates meeting the convergence criteria for all AMs were excluded as well, to reduce variability in ranking distributions caused by varying sample sizes and random seeds. Rankings were then pooled across remaining species and stocks within an OM/AM combination, from which the modal rank and average rank were calculated.

2.3.2.3. Biomass and overfishing risk. Biomass risk was measured by the probability of stock biomass being below both 80% and 40% of single-species operating model $B_{\text{MSY,SS},s,p}$. The threshold of 80% of $B_{\text{MSY,SS}}$ is generally considered to be the level where a fish stock transitions from an optimally fished state to an overfished state (Hilborn, 2018), while 40% of B_{MSY} is commonly suggested as a limit reference point for Canadian fisheries, below which a stock may be considered critically overfished as recruitment may become impaired and rebuilding may be required (DFO, 2006). For both biomass levels, the probability was calculated as

$$P\left(B_{s,p,t} < \lambda B_{\text{MSY,SS},s,p}\right) = \frac{1}{3200}\sum_{i=1}^{100}\sum_{t=2016}^{2047}\mathbf{1}\left(B_{i,s,p,t} < \lambda B_{\text{MSY,SS},s,p}\right)$$

where i, s, p, t are replicates meeting the 95% convergence criteria, species, stocks and time steps, respectively, $\lambda = 0.4, 0.8$, and $\mathbf{1}$ is the indicator function that takes value 1 when its argument is true, and zero otherwise. Overfishing risk was similarly calculated as the probability of fishing mortality exceeding single-species operating model $F_{\text{MSY,SS},s,p}$.

2.4. Sensitivity analyses

Parameter prior distributions area a key feature of most contemporary stock assessment models, even in data-rich contexts. Moreover, prior distributions are a defining feature of hierachical multi-species stock assessment models. Therefore, we focused most sensitivity analyses on fixed prior standard deviations for AM leading parameters MSY and U_{MSY}, and the hierarchical shrinkage prior SDs τ_q and $\sigma_{U_{\text{MSY}}}$ (Table 4). Sensitivity of performance to future increases in fishery independent survey precision was tested, where the Synoptic survey observation error variance was reduced with a linear ramp-down over the first 5 years of the projection period to simulate a gradual increase in survey effort.

3. Results

3.1. Omniscient manager performance

As expected, the omniscient manager was able to achieve the theoretical multi-species optimal yield in the presence of technical interactions during the middle of the projection period (Fig. 4, blue closed circle). Median biomass, catch, and fishing mortality reach the equilibrium after a transition period of about 20 years. During the transitionary period, effort is slowly ramped up in each area from the end of the historical period, stabilising around area-specific E_{MSY} after about 12 years (Fig. 5, blue closed circles). As is common when maximising catch over a finite time horizon, a reduction in biomass and an increase in

Table 3
Probability of being overfished and experiencing overfishing with respect to single-species reference points, and catching less than the historical miniumum during the time period 2028–2037 for all nine DER complex stocks when managed by the omniscient manager.

Stock	Prob. of being overfished		Prob. of overfishing	Prob. of low catch
	$P(B_t < .4B_{MSY,SS})$	$P(B_t < .8B_{MSY,SS})$	$P(F_t > F_{MSY,SS})$	$P(C_t < minC_{1951:2016})$
Dover sole				
HSHG	0.00	0.01	0.07	0.01
QCS	0.00	0.01	0.02	0.01
WCVI	0.00	0.07	0.54	0.01
English sole				
HSHG	0.01	0.10	0.39	0.02
QCS	0.01	0.05	0.20	0.02
WCVI	0.01	0.23	0.78	0.02
Rock sole				
HSHG	0.01	0.59	0.91	0.01
QCS	0.01	0.41	0.84	0.02
WCVI	0.01	0.08	0.42	0.02

Table 4
Summary of AM catch loss rankings and biomass risk under each scenario. The catch loss rankings show the modal and average ranks, calculated across replicates, species, and stocks. Biomass risk columns show the average probability, with range in parentheses, of stocks being in a critically overfished (i.e. below $0.4B_{MSY,SS}$) or overfished (i.e., below $0.8B_{MSY,SS}$) state, calculated across species and stocks. Within a scenario, AMs are ordered by average rank.

AM	Catch loss rankings		Biomass		Fishing mortality
	Modal rank	Average rank	$P(B_t < 0.4B_{MSY,SS})$	$P(B_t < 0.8B_{MSY,SS})$	$P(F_t > F_{MSY,SS})$
High					
Hierarchical Multi-species	1	2.08	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
Species Pooling	2	2.35	0.00 (0.00, 0.00)	0.00 (0.00, 0.01)	0.03 (0.00, 0.08)
Spatial Pooling	4	2.88	0.01 (0.00, 0.02)	0.03 (0.00, 0.09)	0.13 (0.01, 0.29)
Total Pooling	4	3.07	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.02)
Single-species	5	4.63	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
Mod					
Hierarchical Multi-species	1	2.30	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
Species Pooling	1	2.40	0.00 (0.00, 0.00)	0.01 (0.00, 0.02)	0.08 (0.02, 0.15)
Total Pooling	3	2.88	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.01 (0.00, 0.04)
Spatial Pooling	4	2.99	0.01 (0.00, 0.04)	0.04 (0.00, 0.12)	0.16 (0.02, 0.38)
Single-species	5	4.42	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
Low					
Hierarchical Multi-species	1	1.94	0.00 (0.00, 0.00)	0.02 (0.00, 0.06)	0.12 (0.00, 0.25)
Species Pooling	3	2.70	0.01 (0.00, 0.01)	0.03 (0.00, 0.06)	0.14 (0.05, 0.25)
Total Pooling	2	3.12	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.01 (0.00, 0.05)
Spatial Pooling	4	3.45	0.02 (0.00, 0.04)	0.06 (0.00, 0.15)	0.21 (0.04, 0.44)
Single-species	5	3.78	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.01)

effort was observed towards the end of the 80 year projection period, where penalties from lower biomasses and overfishing the stock relative to the multi-species optimal harvest rate are lower than the catch increases; however, the omniscient manager avoids a complete crash of the complex thanks to the penalty on avoiding the critically overfished state at 40% of single-species $B_{\mathrm{MSY,SS}}$.

Each area had similar relationships between single-species and multi-species optimal biomass levels, with one species overfished, one underfished, and one close to optimally fished, relative to single-species reference points. In HSHG and QCS, the overfished stocks were both Rock sole, fished down to 74% (HSHG) and 81% of $B_{\mathrm{MSY,SS}}$ to increase fishery access to Dover and English soles in those areas (Table 1). In WCVI, English sole was slightly overfished relative to single-species at 91% of $B_{\mathrm{MSY,SS}}$, but this would be of little concern in a real fishery if the absolute size were not so small. Despite the tendency of the omniscient manager toward overfishing at least one species in each area, very few optimal solutions risked severe overfishing below 40% of B_{MSY} (Table 3), indicating that lost yield from more intense overfishing relative to single-species optimal levels is not compensated by increased harvest from other species, which are sometimes larger populations with higher TACs. The probability of being critically overfished in the period 2028–2037 was 1% for English and Rock sole stocks, and 0% for all Dover sole stocks.

Although DER stocks begin the simulations in an overall healthy state, the omniscient manager reduced fishing effort to 0 in all areas early in the projection period in some replicates (Fig. 5, 2016–2020). In these cases, anticipatory feedback control by the omniscient manager reduced fishing effort to avoid low spawning stock biomasses, thus ensuring higher production in later time steps where recruitments were lower than average for sustained periods.

3.2. Assessment model performance

3.2.1. Rankings by catch loss

Due to low convergence rates for HSHG and QCS Rock sole and WCVI Dover sole in the High scenario, and WCVI Rock sole under the Mod scenario (Figure S1), results from those management units are excluded from catch loss rankings.

Hierarchical Multi-species assessment models ranked highest under all data quantity scenarios (Table 4). Under the Low data quantity

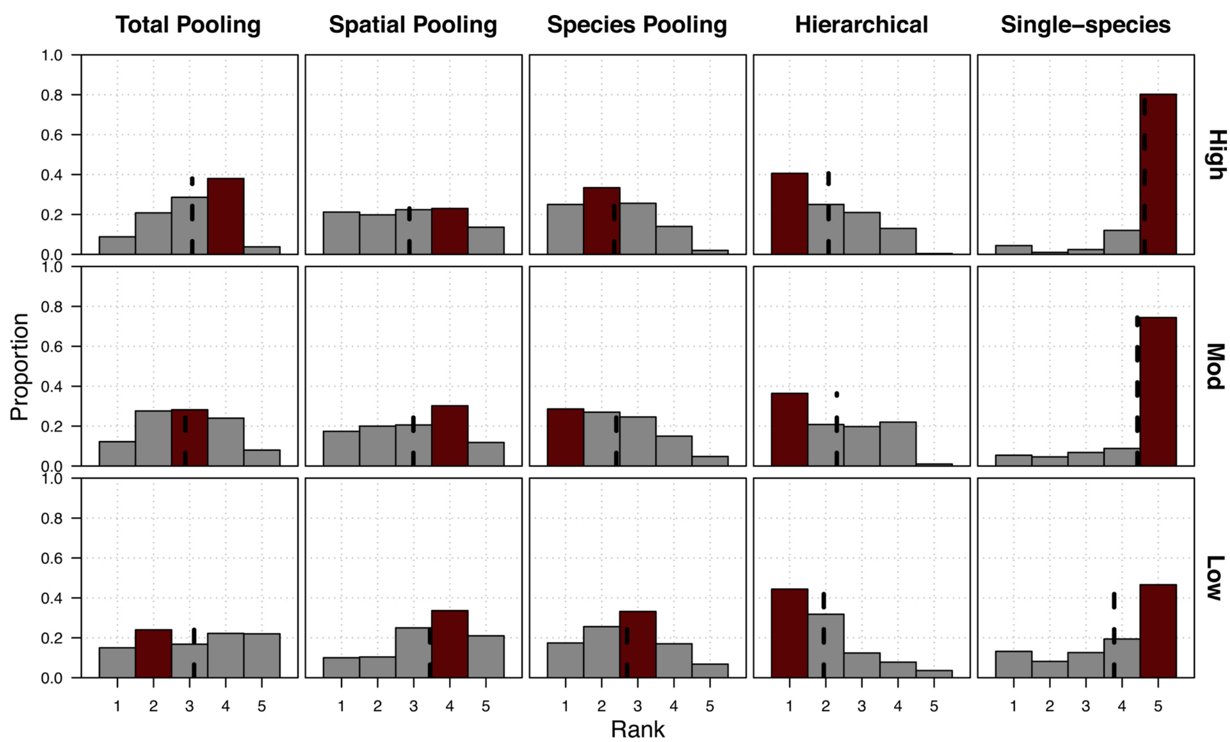

Fig. 6. Distributions of AM rankings by cumulative absolute loss in catch (kt) for the projection years 2028 to 2037 under each assessment model (columns) and OM data scenario (rows). Ranks are calculated for each species/stock combination within a replicate, then distributions of ranks are across species, stocks, and replicates that met the 95% convergent AMs criterion. The modal rank is shown asa dark red bar, and the average rank as a vertical dashed line.

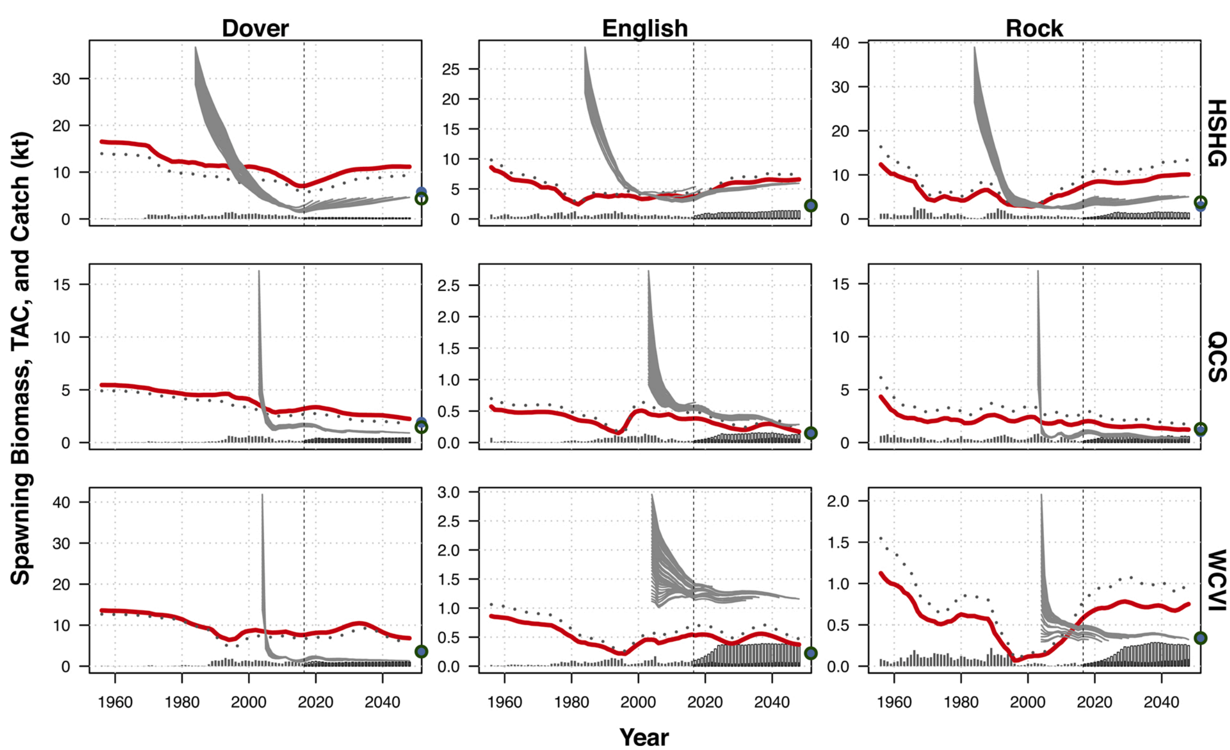

Fig. 7. Operating model spawning stock biomass (red line), commercial trawl vulnerable biomass (grey dotted line), retrospective assessment model estimates of spawning stock biomass (thin grey), and catch and TACs (grey bars) from the first simulation replicate in the Low data-quality scenario and under the Single-stock assessment model. Catch bars show realised catch in grey for the whole simulation period, and unfilled bars in the projection period show the difference between MP set TACs and realised catch. Coloured circles on the right hand vertical axis show the biomass level associated with the multi-species (closed blue circle) and single-species (open green circle) maximum sustainable yield.

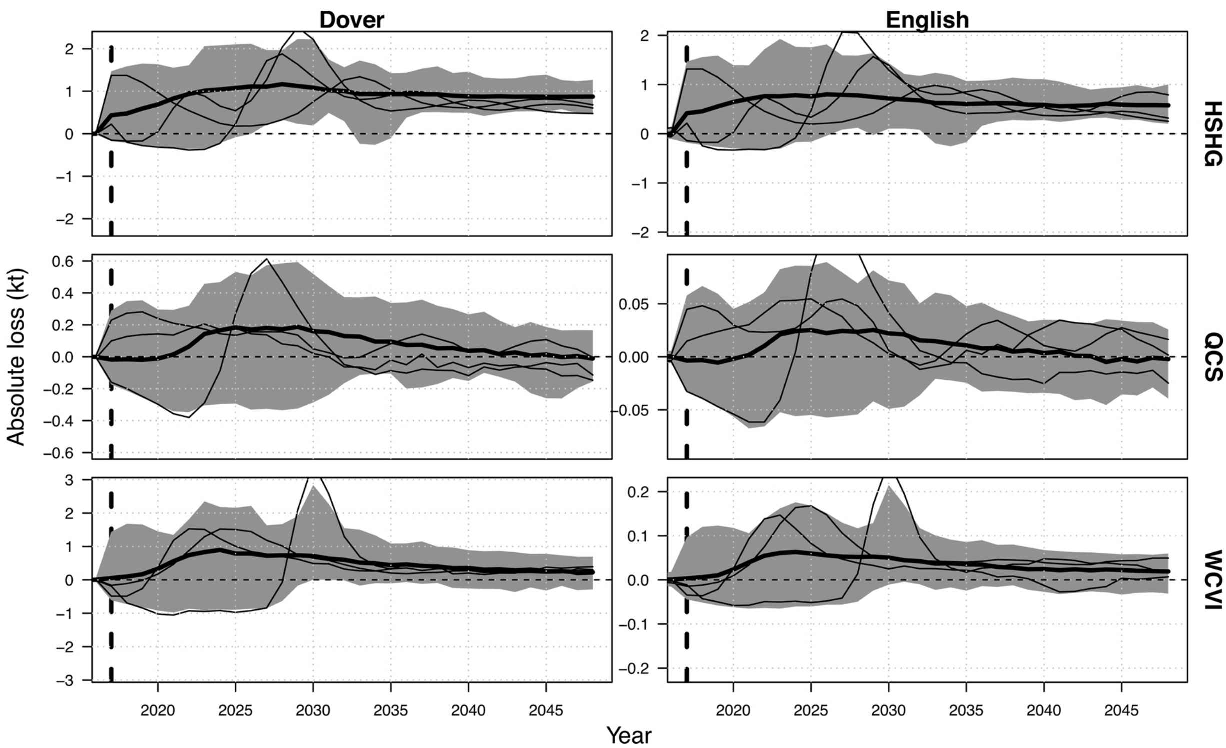

Fig. 8. Catch loss simulation envelopes for the six DER complex management units included in the catch loss ranking, when assessed using the Single-species assessment model under the Low data-quality scenario. Median catch loss is shown by the thick black line, the central 95% of the catch loss distribution is shown as the grey shaded region, and the thin black lines show three randomly selected simulation replicates. The thick vertical dashed line shows the beginning of the projection period, and the horizontal thin dashed line shows zero catch loss.

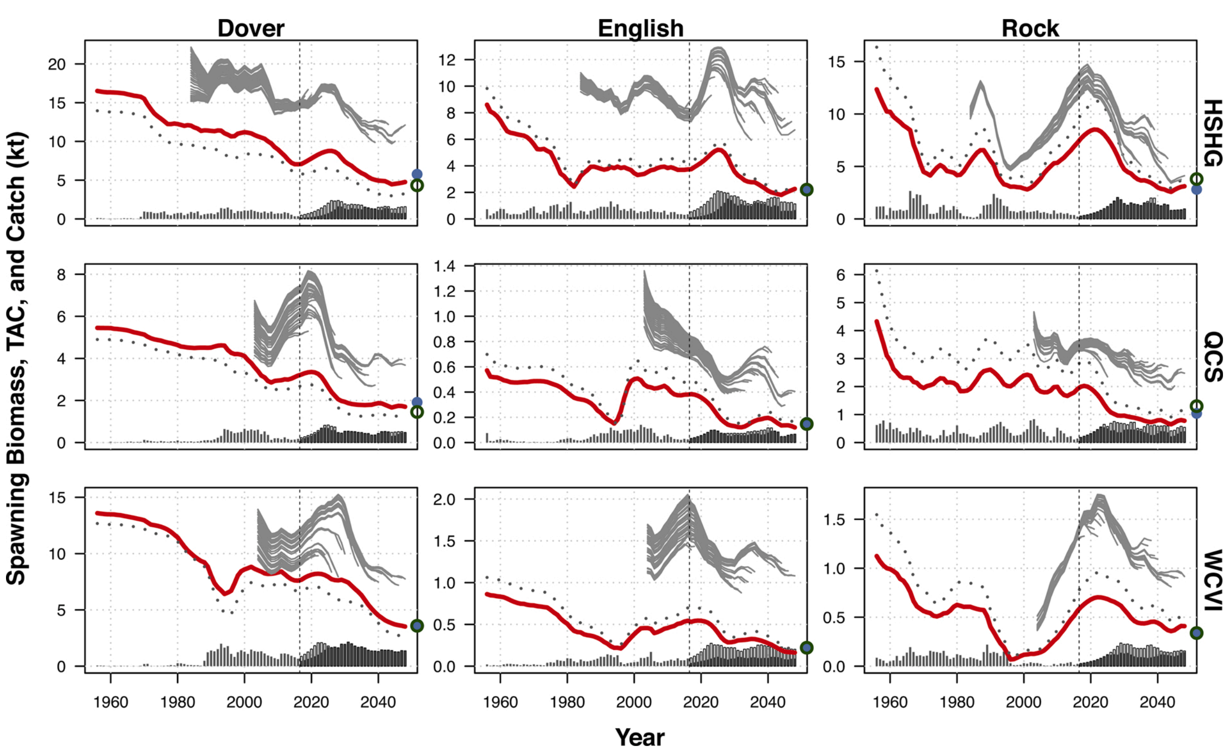

Fig. 9. Operating model spawning stock biomass (red line), commercial trawl vulnerable biomass (grey dotted line), retrospective assessment model estimates of spawning stock biomass (thin grey), and catch and TACs (grey bars) from the first simulation replicate in the Low data-quality scenario and under the Hierarchical Multi-species assessment model. Catch bars show realised catch in grey for the whole simulation period, and unfilled bars in the projection period show the difference between MP set TACs and realised catch. Coloured circles on the right hand vertical axis show the biomass level associated with the multi-species (closed blue circle) and single-species (open green circle) maximum sustainable yield.

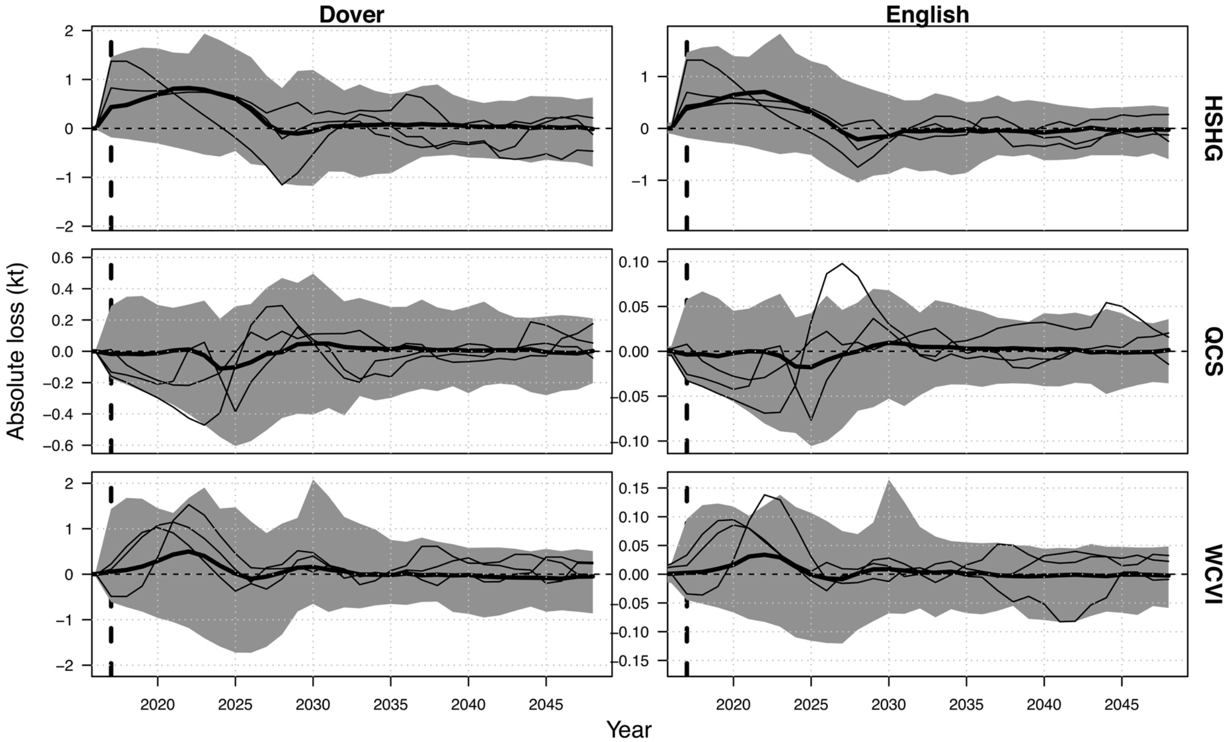

Fig. 10. Catch loss simulation envelopes for the six DER complex management units included in the catch loss ranking, when assessed using the Hierarchical Multi-species assessment model under the Low data-quality scenario. Median catch loss is shown by the thick black line, the central 95% of the catch loss distribution is shown as the grey shaded region, and the thin black lines show three randomly selected simulation replicates. The thick vertical dashed line shows the beginning of the projection period, and the horizontal thin dashed line shows zero catch loss.

scenario, the Hierarchical model had the lowest cumulative absolute catch loss in over 40% of replicate/stock/species combinations (Fig. 6), ranking highest by average and modal rank in the aggregate. When split across individual species/areas, the modal rank for the hierarchical AM remained at 1, but there was more variability in the ranking distributions in the HSHG area (e.g. Dover sole), leading to a drop in average rank (Figure S1, online supplement). As data quantity increased for the Mod scenario, the modal rank and average rank of the hierarchical model remained the highest out of all AMs in the aggregate, ranking 1st in just under 40% of cases (Fig. 6), while the variability in rankings in the HSHG area increased (Figure S1), pushing the average rank a littler lower than under the Low scenario. Results under the High scenario were similar to the Mod scenario, where the Hierarchical model was ranked first in about 40% of cases, and a slightly higher average rank than the Mod scenario, driven largely by a reduction in catch loss for WCVI English sole (Figure S1).

When ordered by average rank, Species Pooling AMs came second after the hierarchical AMs under all data quantity scenarios (Table 4). Under the High data quantity scenario, the data pooling methods ranked highest in both modal and average ranks (Table 4, High). Total and Spatial Pooling methods performed similarly under the High scenario, with only a small difference between average ranks, but quite different rank distributions, where Spatial pooling catch loss rankings were almost uniformly distributed, but Total pooling ranks had a clear mode in 4th place (Fig. 6, High scenario), which was reflected in individual stock rank distributions (Figure S1). The Total Pooling AM's 4th place rank was observed in around 40% of cases in the aggregate (Fig. 6), and was also the modal rank for most individual stock distributions (Figure S1). As data were removed for the Mod data quantity scenario, the Spatial Pooling AM dropped from 3rd to 4th place, driven largely by increased catch loss in the QCS area. The Single-species method had the worst rank under all data quantity scenarios, with a consistent modal rank of 5th across all scenarios (Fig. 6), with most consistently inferior performance observed under the High data quantity scenario, where the average rank was 4.63 (Table 4).

Choke effects both increased and reduced catch loss, depending on the assessment errors. On the one hand, there were several cases where AM underestimates of biomass and/or productivity produced TACs that produced a realised harvest rate much lower than the target associated with maximum multi-species yield. This behaviour was most prominent under Single-species AMs for the Low and Mod scenario, where the fished initialisation led the AMs to estimate a larger and less productive stock in equilibrium, estimating a steep early decline in biomass followed by apparent equilibration to a low biomass state where catch balances production (Fig. 7), producing very low TACs despite the large positive error in unfished biomass. Low TACs of more catchable species constrained the TACs of the remaining species, increasing catch loss across the whole complex (Fig. 8). On the other hand, when TACs pushed harvest rates very close to the target level, choke effects could reduce catch loss by constraining TACs that were too high compared to the target, protecting against overfishing relative to the multi-species harvest rate. These "protective" choke effects were observed under the pooling AMs and the hierarchical AM, and occured even when there were large assessment errors. For example, under the hierarchical AM in the Low scenario, AMs perceived a larger and less productive stock than the OM, similar to the Single-species AM, but shrinkage priors on productivity parameters were able to reduce the magnitude of the error in U_{MSY} so that, despite having some of the largest errors in biomass forecasts and the B_{MSY}, the AM did not perceive a stock that quickly declined from its high initial biomass (Table S2). Biomass estimates were then positively biased with respect to the operating model biomass, which were compensated by the negatively biased productivity estimates, producing more appropriately scaled TACs for all stocks, with choke effects occasionally protecting against overfishing (Fig. 9). More optimal TACs then reduced catch loss overall, with median catch loss close to zero for the 2028–2037 period (Fig. 10).

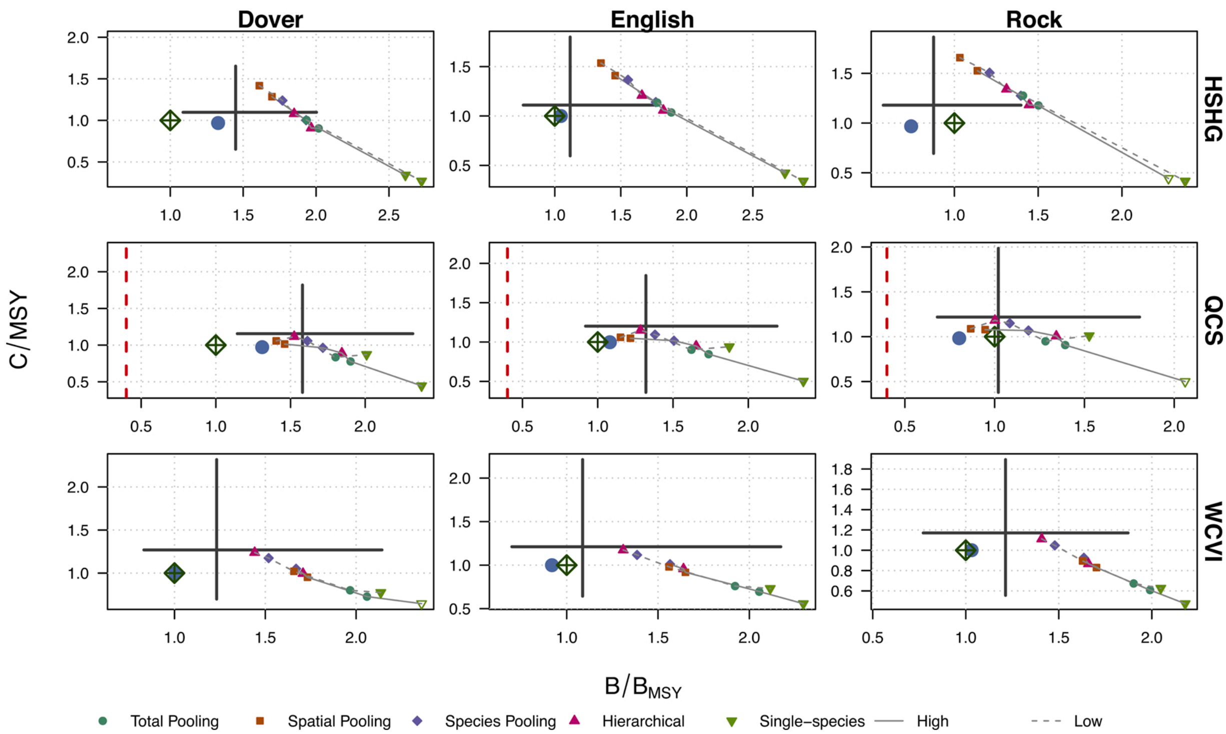

Fig. 11. Tradeoff between catch and biomass during the 2028–2037 period implied by switching between different assessment models under High and Low data-quality scenarios. Panels are gridded by species (columns) and stocks (rows), with biomass relative to $B_{MSY,MS,s,p}$ on the horizontal axis, and catch relative to $MSY_{MS,s,p}$ on the vertical axis. Distributions of biomass and catch under the omniscient manager are shown by the black crosshair, with points indicating optimal biomass and yield for single species maximum yield (open green circles) and multi-species maximum yield (closed blue circles). The biomass level at which a stock is critically overfished is shown as a vertical red dashed line. Coloured point symbols show median biomass and catch for over all replicates for different assessment models, with assessment models under the same data-quality scenario joined by a solid line (High) or dashed line (Low).

3.2.2. *Biomass and overfishing risk*

There was little chance of any stock being in an overfished state under any data scenario and AM combination (Table 4). Only thexo a critically overfished state where $B_{s,p,t} < 0.4B_{\mathrm{MSY,SS},s,p}$ under the High and Mod data scenarios with very low probability, while under the Low scenario a small probability of being critically overfished was observed under the Species Pooling (max 1%) and Spatial Pooling (max 4%) AMs. As expected, there were higher probabilities of being overfished ($B_{s,p,t} < 0.8B_{\mathrm{MSY,SS},s,p}$) under all scenarios, with the highest observed at 15% under the Spatial Pooling AM and Low scenario. The Single-species AM rarely pushed any stock into an over-fished state under any data scenario, as biomass was usually underestimated for at least one species in each area, producing a choke constraint and higher catch loss as discussed above.

As expected, there was a higher risk of overfishing relative to single-species optimal fishing mortality rates $F_{\mathrm{MSY,SS},s,p}$ when attempting to maximise multi-species catch. Under the High and Mod scenario, Pooled AMs were the only AMs with positive probabilities across the entire DER complex, with Spatial Pooling AMs overfishing more often. Unsurprisingly, overfishing was observed under all AMs for the Low scenario, reflecting the greater difficulty in estimating species productivity with low power data. Similar to the other scenarios and biomass risk above, the Spatial Pooling AM exceeded $F_{\mathrm{MSY,SS}}$ more often (max 44%), while the Single-species AM was the most conservative given that it often underestimated biomass and caused choke constraints (max 1%). Hierarchical AMs were moderate compared to the other AMs, falling in the middle of the pack and exceeding $F_{\mathrm{MSY,SS}}$ in 12% of years on average (max 25%) under the Low scenario.

3.2.3. *Catch-biomass trade-offs*

Distributions of catch and biomass relative to $\mathrm{MSY}_{\mathrm{SS}}$ and $B_{\mathrm{MSY,SS}}$, respectively, were produced for the time period $2028 \leq t \leq 2037$ for each assessment model and Scenario combination. The medians of those relative catch and biomass distributions were visually compared to each other and to the central 95% of the omniscient manager's trajectories over the same time period (Fig. 11) in order to understand the biomass and catch trade-offs between different model choices.

The Hierarchical Multi-species assessment model median catch between 2028 and 2037 came closest (i.e., smaller Euclidean distance) to the omniscient manager median catch under the Low data quantity scenario outside the HSHG area, and for HSHG Dover sole (Fig. 11, compare points to the horizontal segment in black crosshair). For HSHG English and Rock soles, while the Hierarchical model tended to take a little more catch than the omniscient manager under the Low scenario, a large biomass surplus relative to the omniscient manager was present, indicating that realised harvest rates from TACs set by the hierarchical AM were smaller than those set by the omniscient manager.

Although the range of biomass-catch trade-offs were quite broad for each stock, the majority of Scenario/AM combinations lie inside the central 95% distributions of the omniscient manager (Fig. 11, black crosshairs). Notable exceptions to this were the Single-species models for all species/areas under the High scenario and in HSHG under the Low scenario for Dover and English, the Hierarchical models in QCS and WCVI under the High scenario, and the Total Pooling method for WCVI Rock sole under the High data quantity scenario. As described above, the Single-species methods tended to under-estimate biomass and productivity under the Low scenario, so the tendency of the Single-species median biomass-catch to be towards the lower right corner of the range is expected. For the Single-species and Hierarchical methods under the High scenario, the reason is a combination of large negative assessment errors in some or all years for choke species, and the 20% limit on increases in catch, producing a low-catch/high-biomass dynamic. Finally, the Total Pooling outlier in WCVI Rock Sole is caused by a persistent small negative assessment error, and an underallocation of

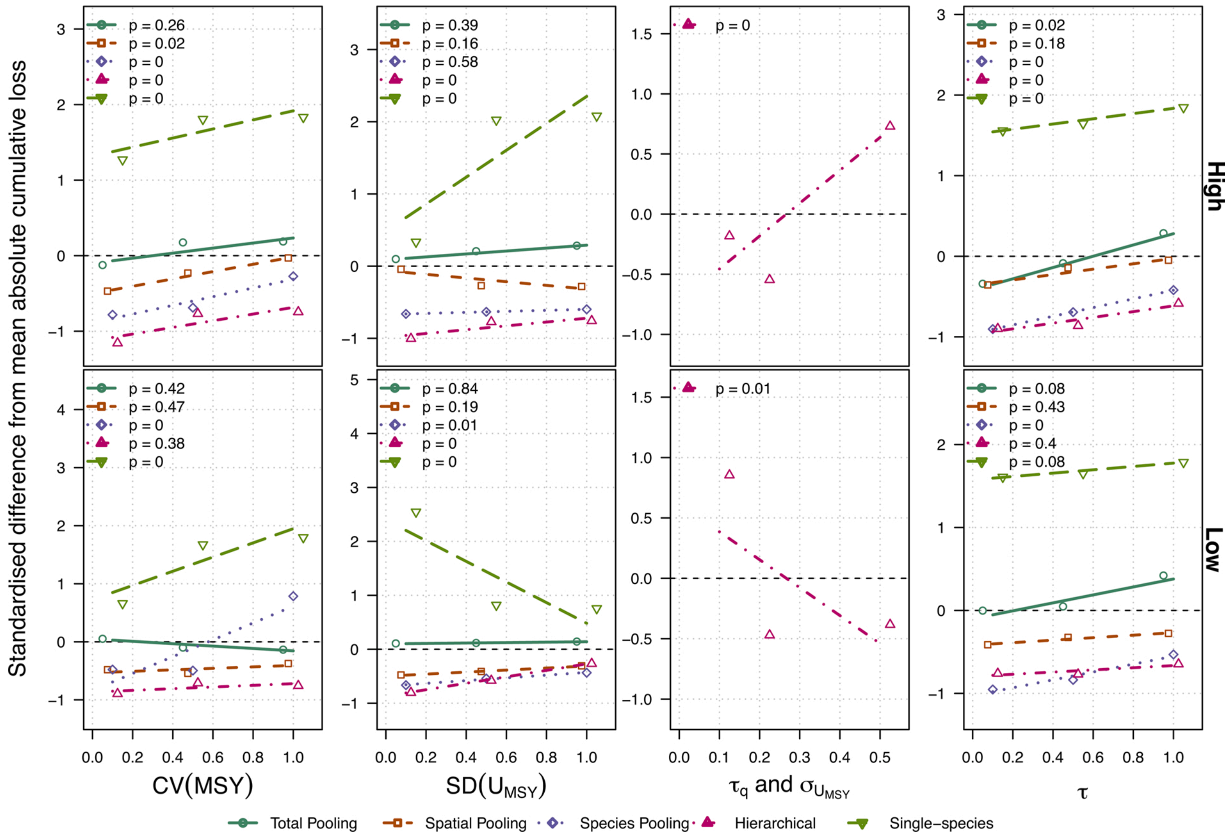

Fig. 12. Regressions showing the average sensitivity of cumulative catch loss to the prior standard deviations under each assessment model (colours, line types) for the data-rich (left column) and data-poor (right column) scenarios. The horizontal axis on each plot shows the prior standard deviation (CV for B_{MSY}), while the vertical axis shows the standardised difference between median cumulative loss for an assessment model and the mean of median cumulative loss values over AMs, stratified by species and area. See the online version of the journal for a full colour version of the plot.

the pooled TAC to WCVI Rock thanks to the differing Synoptic survey catchabilities. The Total Pooling method was also outside the central 95% of the omnisceint manager in HSHG under the High scenario, but only marginally so.

Catch-biomass trade-offs were approximately collinear under both High and Low data quantity scenarios for all stocks. All spawning stock biomasses were well above both the single-species and multi-species optimal levels at the beginning of the projection period (e.g., Table 1, Stock Status), meaning that the catch limits set by all methods were depleting a standing stock and benefited from its surplus production. Under these conditions, an increase in catch almost linearly caused a decrease in biomass as the compensatory effect of density dependence was minimal.

During the 2027–2036 period, median catch and biomass under the omniscient manager were higher than the optimal levels for both the multi-species and single-species maximum yield (green diamond cross-hair and blue circle, respectively, Fig. 11). As discussed above, this is because the transitionary period from fishery development to equilibrium dynamics is takes about 20 years under the omniscient manager. While catch is higher than MSY during the development period, this is not necessarily overfishing as the biomass also higher than B_{MSY}; however, there was overfishing relative to single-species F_{MSY} occured for all stocks under the omniscient manager, and, as expected, with higher probability for all stocks where $B_{\mathrm{MSY,MS}} \leq B_{\mathrm{MSY,SS}}$ (Table 3).

3.3. Sensitivity of results to prior standard deviations

We summarised average model sensitivities by fitting linear regressions to the distributions of median cumulative absolute catch loss (Fig. 12). To remove the effect of absolute catch scales on the regressions, median loss distributions were standardised across assessment models, stratified by species, stock-area, and data scenario. Regressions with positive slopes had increasing catch loss with increasing uncertainty, and negative or zero slopes indicated a decrease or no change in catch loss with increasing prior uncertainty. Linear model slope parameter *p*-values for each AM were also calculated to determine if an effect was significant or not, where we define significance as $p < 0.1$.

While all AMs except the Total Pooling AM were slightly sensitive to MSY prior CVs under the high scenario, there was no change in the relative rankings of the AMs (Fig. 12, first column). Under the Low scenario, only the Single-species and Species Pooling AMs were sensitive to MSY CVs, with catch loss dropping about a standard deviation as CVs decreased from 1.0 to 0.1 for both AMs, with Species Pooling AMs moving from fourth to second lowest mean catch loss, but the relative ranking of all other AMs remained the same.

Hierarchical and Single-species AMs were sensitive to the U_{MSY} prior standard deviation under both scenarios, and the Species Pooling AM was also sensitive under the Low scenario (Fig. 12, second column, $p > 0.05$). However, under the High scenario, no sensitivity led to a change in ranking for any AMs over the range of U_{MSY} prior standard deviations tested, as most AMs showed a catch loss change of less than 0.5 standard deviations, except for the Single-species AM, which was again an outlier. Under the Low scenario, the Total Pooling and Single-species AMs remained at the bottom of the rankings, despite a decrease off around 1.5 standard deviations for the Single-species AM. At the same time, Spatial Pooling, Species Pooling, and Hierarchical AM catch losses under the Low scenario appeared to converge as prior standard deviations increased from 0.1 to 1.0.

Interestingly, the Hierarchical Multi-species assessment models reacted differently to changing catchability and U_{MSY} hierarchical shrinkage priors under the High and Low scenarios (Fig. 12, third column). As $\sigma_{U_{\mathrm{MSY}}}$ and τ_q increased from 0.1 to 0.5, average catch loss increased by about 1 standard deviation under the High scenario ($p < 0.01$), mostly driven by assessment errors in the QCS and WCVI areas, leading to a widening range of catch trajectories as hierarchical prior standard deviations increased. Under the Low scenario, median absolute catch-loss dropped by about 1 standard devation, indicating that TACs were closer to the omniscient manager as hyper-prior SDs increased. The improved performance with increasing hyper-prior uncertainty was caused by a combination of subtle effects, with some assessments becoming more biased, and others less biased, which occasionally switched the choke species in an area. The combination of raised and lowered catch limits and technical interactions produced beneficial choke effects, reducing catch loss overall.

Finally, there was low sensitivity to the reduction in Synoptic survey standard errors (Fig. 12, fourth column). Under the High scenario, the all but the Spatial Pooling AM had significant improvements of around 0.3–0.5 standard deviations with decreasing standard error ($p < 0.05$), while the Spatial pooling had a non-significant change ($p = 0.18$). Under the Low scenario, only the Species Pooling AM catch loss improved significantly, dropping around 0.25 standard deviations with decreasing Synoptic survey standard error ($p < 0.01$), and switching places with the Hierarchical AM for the lowest average catch loss. For both scenarios, the decreases in catch loss appear to be caused by a slight improvement in pooled TAC allocation with increased survey precision, but choke constraints were still present.

4. Discussion

In this paper, we demonstrated that hierarchical stock assessment models may improve management performance in a spatially-replicated multi-species flatfish fishery. When available data quantity was moderate or low (indicated here by time-series length), biomass and harvest rate estimates from hierarchical stock assessment models resulted in catches that were closer to an omniscient manager's optimal reference series compared to catch limits derived from single-stock and data-pooling assessment methods. Under high data quantity scenarios, data-pooling methods outperformed hierarchical models, but the latter still outperformed single-stock assessment methods. This suggests that hierarchical assessment methods could be a better approach to making catch limit decisions than conventional single-species methods under typical fisheries data quality conditions, such as short or noisy time series of fishery-independent observations or uninformative catch series.

Ranking performance according to catch loss highlighted a fundamental trade-off between multi-species and single-species fishery objectives (Walters et al., 2005). We measured management performance via comparison to an omniscient manager simulation with the explicit objective of maximising multi-species complex yield. While targeting multi-species optima does increase complex yield overall, there are cases where the fishing mortality rates exceed single-species F_{MSY}, which is the definition of overfishing in a single-species context. An obvious question is, would the rankings continue to hold if catch limits targeted individual stocks' single-species U_{MSY}? Further, would the rankings change if ranking of AM performance was based on the risk of overfishing, or being in an overfished state, relative to single-species reference points? Targeting single-species harvest rates would likely lead to higher instances of choke effects since technical interactions are being ignored. Further, targeting single-species yields would probably not affect the rankings based on AM performance, since the estimated optimal harvest for hierarchical models had compensatory biases in biomass and harvest rates that would hold for the single-species harvest rates as well, while single-species and pooled models lacked those compensatory biases, and pooled models had issues achieving target TACs given the allocation mode. The probabilities of overfishing or being in an overfished state show that under the High and Mod scenarios the Hierarchical model would still do no worse than tie for first rank, but would fall to third rank under the Low scenario, with the single-species model ranking first. However, even at the highest probability of exceeding single-species F_{MSY} under the hierarchical AMs, there was no chance of pushing stocks into a critically overfished state, and at most 6% chance of pushing any DER complex stocks into an overfished state of less than 80% of $B_{\mathrm{MSY,SS}}$, which would be acceptable management performance in most fisheries. Furthermore, choosing single-species AMs would give up a large amount of catch. This trade-off highlights that measuring multi-species fisheries performance according to single-species objectives (i.e., reference points) may be overly conservative, and that while some overfishing may be required to achieve multi-species yield objectives, that overfishing can be sustainable. While it is certainly possible to come up with a set of commercial catchabilities that would produce critically overfished stocks while maximising complex yield, that was not the case for the DER complex simulated here, and it is unclear if such a mix would be a realistic scenario.

Our results arise from models that are necessarily a simplification of the real stock-management system. The harvest rules applied to DER complex species were relatively simple and may require more detail or complexity for practical applications. All harvest rules were constant target harvest rates, which do not include precautionary "ramping-down" of catch towards a limit biomass level (DFO, 2006; Cox et al., 2013). Including a ramped harvest rule would reduce the probability of stocks being critically overfished in some cases, probably at some further cost of choke effects, but there was very little chance of critically overfishing any stock under our simulations anyway. Second, catch limits for the simulated DER complex were set based on estimated target harvest rates that were scaled by *a priori* known scalars derived from multi-species yield curves, which may positively bias results towards lower catch loss in general. Incorporating multi-species yield curve calculations to the assessment model output into the harvest decision would be simple to do, but would require either a model of increased complexity to link fishing effort to single-species yield, or an extra assumption linking effort to surplus production model yield calculations, which would likely increase assessment model errors. Finally, the TAC allocation model for data-pooled methods was only one example from a large set of potential options. Understanding the relative risks of data-pooling would require testing alternative allocation methods, which was beyond the scope of this paper.

Replacing commercial fishery catch rates with relative biomass in the simulations increased the statistical power of commercial index data and skewed results under the High scenario. Under previous versions of the simulations where commercial indices were simulated as catch rates, but without observation error, the Hierarchical method ranked lower than the three Pooling AMs under the High scenario, but still higher than the Single-species model. The change in performance is largely caused by learning or targeting behaviour of harvesters in Historical (1956–1996) trawl fishery that caused indices to increase for some DER complex stocks while estimated biomass was decreasing in the model, indicating time-varying catchability for that fleet. Time-varying catchability was less of an issue for Pooled AMs because data pooling is intended to increase sample size and reduce variability. Replacing catch rates with relative biomass indices simulated with a constant catchability reduced the advantage of data pooling under the High scenario, allowing Hierarchical models to achieve the lowest catch loss under all three scenarios.

We only considered multi-species technical interactions, which although an important part of exploited system dynamics, are not the entire story. Although there is limited evidence for ecological interactions among DER complex species (Pikitch, 1987; Wakefield, 1984), what does exist may influence the multi-species yield relationship with fishing effort or, as with technical interactions, inhibit the ability of the management system to meet target catch levels. For example, individual survival or growth may change in response to varied fishing pressure due to unmodeled linkages (Collie and Gislason, 2001). Yet, including such ecological interactions would imply a highly data rich scenario, which is counter to our focus on surplus production models applied to multi-species fisheries. Furthermore, accounting for potential ecological interactions would require multiple OMs to test performance against a range of plausible hypotheses, since ecological uncertainties are much broader in complexity and scope than technical interactions alone. Nevertheless, future work combining technical interactions with minimum realistic models for ecological interactions could help determine the extent to which assessment approaches affect these more complex multi-species fisheries outcomes (Punt and Butterworth, 1995). For example, while diet overlap between the three species is small off the coast of Oregon, the major Rock sole prey was recently settled pleuronectiform fishes, which may include Dover and English sole young and therefore shift the complex equilibrium as fishing pressure is applied, reducing predation mortality for Dover and English sole young, and reducing prey availability for Rock sole (Wakefield, 1984; Collie and Gislason, 2001).

Our effort model applied to the DER complex was also a simplification of reality, where effort was limited only by the TACs in each area. Limiting by TACs was intended to reflect the management of the real BC groundfish fishery in which harvester decisions drive TAC utilisation among target species (via increasing catchability; Punt et al., 2011a), and non-target or choke species (via decreasing catchability; Branch and Hilborn, 2008). Changing catchability for targeting or avoidance could be simulated as a random walk in the projections, with correlation and variance based on the historical period, or perhaps simulated via some economic sub-model that accounted for ex-vessel prices and variable fishing costs. These economic factors could affect targeting and avoidance behaviour among species (Punt et al., 2011a, 2020), as well as effort allocation among stock-areas (Hilborn and Walters, 1987; Walters and Bonfil, 1999); however, it is not clear that our median results would be significantly different given the potential magnitude of assessment model errors in data-limited scenarios. Impacts of a detailed effort dynamics sub-model would probably be more important in more extreme data-limited scenarios that relied solely on fishery CPUE as an index of abundance, which we did not test here. In fact, it would be interesting to determine whether the hierarchical information-sharing approach would exacerbate assessment model errors in the (common) context where fishery CPUE is the main abundance index.

Our assessment models were all different versions of a state-space surplus production model, and rankings of AMs may vary when other model configurations with more biological realism are included. For example, the aggregate productivity parameter U_{MSY} could be separated into growth, natural mortality, and stock-recruit steepness by using a delay-difference or age-structured model formulation (Deriso, 1980; Schnute, 1985; Fournier and Archibald, 1982), which may partially offset the advantages of data-pooling and hierarchical assessment methods exhibited above. However, in the contexts where biological data are missing or of low sample size, there would be even greater reliance on strong *a-priori* assumptions for additional parameters in models of higher complexity, which we predict would lead to similar results.

Despite the limitations above, our results indicate that even in fisheries with long time series of catch and effort data, hierarchical multi-species assessment models may be preferable over typical single-species methods. The poor performance of the single-species models in all scenarios highlights the difference between data-rich (i.e., a higher quantity of data) and information-rich (i.e., data with higher statistical power) fisheries. The high data quantity scenario differed from the moderate and low scenarios by the inclusion of a historical series of fishery dependent CPUE, which was quite noisy and subject to the effects of changing harvester behaviour like targeting (variable catchability), and therefore, additional historical CPUE data had little effect on cumulative catch loss under the single-species models. In contrast, the data-pooling procedures all ranked higher than single-species and multi-species models under the data-rich scenario, as they were able to leverage additional statistical power from the historical CPUE by effectively increasing the sample size through data aggregation. The superior performance of the hierarchical model over the single-species model under the high data scenario indicate that shared priors partially compensate for low statistical power when setting TACs, but not as much as data-pooling.

Superior management performance of the hierarchical models was primarily caused by compensatory (negatively correlated) biases in biomass and productivity. Biases in biomass estimates were comparable between Single-species and Hierarchical models, but the hierarchical shrinkage prior structure defined for productivity parameters led to target harvest rate estimates that, while biased, combined with biased biomass estimates to produce TACs that were closer to the omniscient manager's. Therefore, while improved performance relative to single-species models under lower data quantity conditions is consistent with our previous study, where statistical performance of hierarchical multi-stock assessments improved with decreasing data quantity and quality (Johnson and Cox, 2018), the improved management performance of Hierarchical methods was due to a fundamentally different mechanism. This difference may be explained by a different assessment model parameterisation, increased complex size, and a different and simplified experimental design.

The benefits of parameter shrinkage induced by the hierarchical stock assessments relies on similar life histories within the the DER complex, allowing joint distributions of productivity and catchability parameters to be more precise, thereby drawing estimates close to the prior mean when data have low statistical power. Disparate life histories and catchabilities would require a decrease in the precision of the joint distribution of the hierarchically modeled parameters to preserve exchangeability, and therefore may reduce or completely eliminate the benefits of parameter shrinkage for any complex containing species with disparate life histories (Thorson et al., 2015; Gelman et al., 2014).

Hierarchical and Single-species models were sensitive to changes in prior precision for assessment model productivity parameters under the Low data quantity scenario. The Hierarchical model went from lowest median catch loss (i.e., ranked first) to highest (ranked last) as prior precision on the complex mean productivity was reduced, indicating that the compensatory biases that gave hierarchical models the advantage under low data conditions were dependent on this prior. Under the same scenario, the Single-species method improved slightly as prior precision on the stock-specific productivity parameter was reduced, allowing more compensatory bias into the harvest rate estimates. Similarly, under the same low-data scenario, Hierarchical models achieved significantly lower catch loss as hierarchical shrinkage prior precision was reduced. Lower catch loss under reduced hierarchical precision could be attributed to higher variability in catchability and productivity estimates, allowing the stock-specific estimates to achieve more optimal TACs despite the constrained complex or species mean values.

The data-pooled methods performed better under the High scenario and were generally insensitive to priors, indicating that the data were more influential than priors on the TACs and allocation. This may be because data-pooled observation errors are biased low, being simulated independently of the observation errors in the component indices, and using the average standard deviation of the components. If aggregate indices pooled errors from each component index, then the resulting observation error variance would be additive in the components, especially if those errors were positively correlated, which may be the case under a common survey or fishery.

One might expect that data-pooled methods would outperform other methods as future precision in the Synoptic survey was increased, amplifying the pooling advantage, but this was not the case. The lack of dominance by the pooled methods was caused by the allocation model and, when pooled over space, the pooling method. The allocation model allocated TAC in proportion to Synoptic survey index, which was biased away from absolute stock size by the Synoptic trawl efficiency parameter, leading to under-allocation for some species/areas and choke constraints. When data were pooled over areas, the alternating biennial observations caused a sawtooth pattern as low and high biomass areas dominated the pooled index, which inflated the observation errors independent of the survey precision.

We showed that choke effects are not a uniformly negative outcome for multi-species fisheries, and may indicate a mismatch between the target harvest rate and optimal complex yield. The usual assumption is that choke species restrict access to fishing grounds, decreasing profitability through lost yield of target species, and higher quota prices for choke species (Mortensen et al., 2018); however, we found that choke species sometimes prevented overfishing when TACs for the non-choke species were set too high, allowing harvest strategies to meet multi-species objectives despite large assessment errors for individual species in the complex. In reality, choke effects would likely be lessened by changing species catchability via harvester targeting and avoidance, creating a more complex relationship between effort and complex yield; but, the existence of a choke species would still indicate a mismatch between an individual species' TAC and the optimal exploitation rate for meeting the management objectives for the multi-species complex.

4.1. Conclusion

Hierarchical multi-species surplus production models can outperform single-species production models in meeting multi-species harvest objectives across high, moderate, and low data quantity scenarios, while avoiding states of conservation concern with high probability. While hierarchical model estimates of biomass used for setting TACs were often more biased than those produced by other methods, negatively correlated bias in biomass and productivity was better matched under hierarchical models than the other methods as data quantity was reduced, translating into better management performance across the multi-species flatfish fishery. We recommend that assessment and management of multi-species fisheries include hierarchical models that acknowledge technical interactions when designing harvest strategies and management procedures for data-limited, multi-species fisheries. Otherwise, management procedures based on single-species approaches that rely heavily on prior knowledge (inducing bias) and ignore technical interactions (making objectives impossible to achieve) may give a misleading picture of the expected management performance in multi-species, spatially heterogeneous fisheries.

Conflict of interest

The authors declare no conflict of interest.

CRediT authorship contribution statement

Samuel D.N. Johnson: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Visualization. **Sean P. Cox:** Conceptualization, Resources, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

Funding for this research was provided by a Mitacs Cluster Grantto S. P. Cox in collaboration with the Canadian Groundfish Research and Conservation Society, Wild Canadian Sablefish, and the Pacific Halibut Management Association. We thank S. Anderson and M. Surry at the Fisheries and Oceans, Canada Pacific Biological Station for fulfilling data requests. Further support for S.P.C. and S.D.N.J. was provided by an NSERC Discovery Grantto S.P. Cox. We finally thank A. E. Punt and one anonymous reviewer for their thoughtful comments that led to an improved paper.

Appendix A. The operating model

The operating model was a standard age- and sex-structured operating model, with additional structure for multi-species and multi-stock population dynamics. DER complex species and stocks were simulated assuming no ecological interactions or movement between areas. The lack of movement may be unrealistic, especially for Dover Sole given their extent, but this is how the the DER complex stocks are currently managed in practice. The lack of ecological interactions is more realistic for Dover and English soles, as although both species are benthophagus, there is evidence that they belong to different feeding guilds (Pikitch, 1987).

DER complex abundance $N_{a,x,s,p,t}$ for age a, sex x, species s and stock p at the start of year t was given by

$$N_{a,x,s,p,t} = \begin{cases} 0.5R_{s,p,t} & a = 1, \\ N_{a-1,x,s,p,t-1} \cdot e^{-Z_{a-1,x,s,p,t-1}} & 1 < a < A, \\ N_{a-1,x,s,p,t-1} \cdot e^{-Z_{a-1,x,s,p,t-1}} + N_{a,x,s,p,t-1} \cdot e^{-Z_{a,x,s,p,t-1}} & a = A^{(s)}, \end{cases}$$

where R_t is age-1 recruitment in year t, $Z_{a,x,s,p,t}$ is the instantaneous total mortality rate, and $A^{(s)}$ is the plus group age for species s.

Numbers-at-age were scaled to biomass-at-age by sex/species/area- specific weight-at-age. Weight-at-age was an allometric function of length-at-age

$$w_{a,x,s,p} = \alpha_{x,s,p} \cdot L_{a,x,s,p}^{\beta_{x,s,p}}$$

where $\alpha_{x,s,p}$ scaled between cm and kg, $\beta_{x,s,p}$ determined the rate of allometric growth, and $L_{a,x,s,p}$ was the length in cm of a fish of age a, sex x, species s and stock p. Length-at-age was given by the following Schnute formulation of the von-Bertalanffy growth curve (Schnute, 1981; Francis, 2016)

$$L_a = \overline{L}_{A_1} - (\overline{L}_{A_2} - \overline{L}_{A_1}) \cdot \left(\frac{e^{-kA_1} - e^{-k\ a})}{e^{-kA_1} - e^{-kA_2}} \right)$$

where A_1 and A_2 are well spaced reference ages, $\overline{L}_{A_1}$ and $\overline{L}_{A_2}$ are the mean lengths in cm of fish at ages A_1 and A_2, and k is the growth coefficient. Note that in the growth model we dropped the sex, species and stock subscripts for concision.

The maturity-at-age ogive was modelled as a logistic function

Table A.1
Log-normal observation error standard deviations for all DER complex biomass indices

Stock	Observation error SD			
	Historical	Modern	HS Ass.	Syn
Dover sole				
HSHG	0.549	0.548	0.527	0.349
QCS	0.530	0.534		0.355
WCVI	0.569	0.530		0.315
English sole				
HSHG	0.521	0.551	0.491	0.360
QCS	0.507	0.586		0.403
WCVI	0.520	0.626		0.336
Rock sole				
HSHG	0.509	0.571	0.524	0.324
QCS	0.512	0.620		0.360
WCVI	0.520	0.845		0.405

$$m_{a,s,p} = \left(1 + e^{-\frac{ln19(a-a^{mat}_{50,s,p})}{a^{mat}_{95,s,p}-a^{mat}_{50,x,s,p}}}\right)^{-1},$$

where $m_{a,s,p}$ was the proportion of age-a female fish of species s in stock p that were mature, and $a^{mat}_{50,s,p}$ and $a^{mat}_{95,s,p}$ are the ages at which 50% and 95% of fish of age-a, species s and stock p were mature.

Female spawning stock biomass was calculated as

$$B_{s,p,t} = \sum_a N_{a,x',s,p,t} m_{a,s,p} w_{a,x',s,p},$$

where x' denotes female fish only. Spawning stock biomass was used to calculate expected Beverton-Holt recruitment, which then had recruitment process errors applied

$$R_{s,p,t+1} = \frac{R_{s,p,0} \cdot 4h_{s,p} \cdot B_{s,p,t}}{B_{s,p,0} \cdot (1-h_{s,p}) + (5h_{s,p}-1) \cdot B_{s,p,t}} \cdot e^{\varepsilon_{s,p,t+1} - 0.5\sigma^2_{R,s,p}},$$

where $R_{s,p,0}$ is unfished equilibrium recruitment, $B_{s,p,t}$ is the spawning stock biomass at time t, $B_{s,p,0}$ is unfished spawning stock biomasss, $h_{s,p}$ is stock-recruit steepness (average proportion of $R_{s,p,0}$ produced when $B_{s,p,t} = .2B_{s,p,0}$), and $\varepsilon_{s,p,t}$ is the recruitment process error with standard deviation $\sigma_{R,s,p}$.

The operating model was initialised in 1956 at unfished equilibrium for all species s and areas p, with numbers-at-age in 1956 given by

$$N_{a,x,s,p,1956} = \begin{cases} 0.5R_{s,p,0} & a = 1, \\ N_{a-1,x,s,p,1956} \cdot e^{-M_{x,s,p}} & 1 < a < A, \\ N_{a-1,x,s,p,1956} \cdot \frac{e^{-M_{x,s,p}}}{1-e^{-M_{x,s,p}}} & a = A^{(s)}, \end{cases}$$

Fishery removals were assumed to be continuous throughout the year, with fishing mortality-at-age

$$F_{a,x,s,p,f,t} = S_{a,x,s,p,f} \cdot F_{s,p,f,t},$$

where $F_{s,p,f,t}$ is the fully selected fishing mortality rate for fleet f at time t, and $S_{a,x,s,p,f}$ is the selectivity-at-age a for sex x in species s and area p by fleet f. Selectivity-at -age was modeled as a logistic function of length-at-age

$$S_a = \left(1 + exp\left(\frac{-ln19(L_a - l^{sel}_{50})}{l^{sel}_{95} - l^{sel}_{50}}\right)\right)^{-1},$$

where L_a is length-at-age, defined above, and l^{sel}_{50} and l^{sel}_{95} are the length-at-50% and length-at-95% selectivity, respectively; stock, species and fleet subscripts are left off for concision. Catch-at-age (in biomass units) was then found via the Baranov catch equation

$$C_{a,x,s,p,f,t} = (1 - e^{-Z_{a,x,s,p,f,t}}) \cdot N_{a,x,s,p,t} w_{a,x,s,p} \frac{F_{a,x,s,p,f,t}}{Z_{a,x,s,p,f,t}},$$

where total mortality-at-age is defined as

$$Z_{a,x,s,p,f,t} = M_{x,s,p} + S_{a,x,s,p,f} \cdot F_{a,x,s,p,f,t}.$$

A.1 Observation error standard deviations

Operating model observation error standard deviations were derived from estimates from fitting a hierarchical age-structured model to DER complex data.

Appendix B. Omniscient manager optimisation

We defined penalty functions so that inside their respective desired regions the penalty was zero, and otherwise the penalty grew as a cubic function of distance from the desired region. For example, a penalty designed to keep a measurement x above a the desired region boundary ε is of the form

$$\mathcal{P}(x,\varepsilon) = \begin{cases} 0 & x \geq \varepsilon, \\ |x-\varepsilon|^3 & x < \text{epsilon}. \end{cases} \tag{B.1}$$

This form has a several advantages over simple linear penalties, or a logarithmic barrier penalty (Srinivasan et al., 2008). First, the cubic softens the boundary threshold ε, effectively allowing a crossover if doing so favours another portion of the objective function. Second, unlike lower degree polynomials, cubic functions remain closer to the x-axis when $|x-\varepsilon| < 1$. Third, zero penalty within in the desirable region stops the objective function from favouring regions far from the boundaries of penalty functions. In contrast, a logarithmic function would favour overly conservative effort series to keep biomass far from a lower depletion boundary. Finally, the cubic penalty function and its first two derivatives are continuous at every point x, allowing for fast derivative-based optimisation methods.

We used a cubic spline of effort in each area to reduce the number of free parameters in the optimisation. For each area, 9 knot points were distributed across the full 40 year projection, making them spaced by 5 years. We padded the omniscient manager simulations by an extra eight years over the stochastic simulations to avoid any possible end effects of the spline entering the performance metric calculations. Effort splines were constrained to be between 0 and 120 times the operating model $E_{\mathrm{MSY},p}$, by replacing any value outside that range with the closest value inside the range (i.e. negative values by *zero*, large values by $120E_{\mathrm{MSY}_p}$).

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fishres.2021.105885.

References

Appeldoorn, R.S., 1996. Model and method in reef fishery assessment. Reef Fisheries. Springer, pp. 219–248. https://doi.org/10.1007/978-94-015-8779-2_9.

Baudron, A.R., Fernandes, P.G., 2015. Adverse consequences of stock recovery: European hake, a new "choke" species under a discard ban? Fish Fish. 16, 563–575. https://doi.org/10.1111/faf.12079.

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. https://doi.org/10.1016/j.fishres.2015.02.003.

Branch, T.A., Hilborn, R., 2008. Matching catches to quotas in a multispecies trawl fishery: targeting and avoidance behavior under individual transferable quotas. Can. J. Fish. Aquat. Sci. 65, 1435–1446. https://doi.org/10.1139/F08-065.

Collie, J.S., Gislason, H., 2001. Biological reference points for fish stocks in a multispecies context. Can. J. Fish. Aquat. Sci. 58, 2167–2176. https://doi.org/10.1139/f01-158.

Collie, J.S., Walters, C.J., 1991. Adaptive management of spatially replicated groundfish populations. Can. J. Fish. Aquat. Sci. 48, 1273–1284. https://doi.org/10.1139/f91-153.

Cox, S.P., Kronlund, A.R., 2008. Practical stakeholder-driven harvest policies for groundfish fisheries in British Columbia, Canada. Fish. Res. 94, 224–237. https://doi.org/10.1016/j.fishres.2008.05.006.

Cox, S.P., Kronlund, A.R., Benson, A.J., 2013. The roles of biological reference points and operational control points in management procedures for the Sablefish (Anoplopoma fimbria) fishery in British Columbia, Canada. Environ. Conserv. 40, 318–328. https://doi.org/10.1017/S0376892913000271.

de la Mare, W.K., 1998. Tidier fisheries management requires a new MOP (management-oriented paradigm). Rev. Fish Biol. Fish. 8, 349–356. https://doi.org/10.1023/A:1008819416162.

DeMartini, E.E., 2019. Hazards of managing disparate species as a pooled complex: a general problem illustrated by two contrasting examples from hawaii. Fish Fish. 20, 1246–1259. https://doi.org/10.1111/faf.12404.

Deriso, R.B., 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. 37, 268–282. https://doi.org/10.1139/f80-034.

DFO, 2006. A harvest strategy compliant with the precautionary approach. DFO Can. Sci. Advis. Rep. (Technical Report 2006/023) https://www.dfo-mpo.gc.ca/csas-sccs/Publications/SAR-AS/2006/2006_023-eng.htm.

Fisheries and Oceans, Canada, 2015. Pacific Region Integrated Fisheries Management Plan: Groundfish.

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

Francis, R.C., 2016. Growth in age-structured stock assessment models. Fish. Res. 180, 77–86. https://doi.org/10.1016/j.fishres.2015.02.018.

Gaichas, S., Gamble, R., Fogarty, M., Benoît, H., Essington, T., Fu, C., Koen-Alonso, M., Link, J., 2012. Assembly rules for aggregate-species production models: simulations in support of management strategy evaluation. Mar. Ecol. Prog. Ser. 459, 275–292. https://doi.org/10.3354/meps09650.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2014. Bayesian Data Analysis, vol. 3. Taylor & Francis.

Gislason, H., 1999. Single and multispecies reference points for baltic fish stocks. ICES J. Mar. Sci. 56, 571–583. https://doi.org/10.1006/jmsc.1999.0492.

Hilborn, R., 2018. Measuring fisheries performance using the "goldilocks plot". ICES J. Mar. Sci. 76, 45–49. https://doi.org/10.1093/icesjms/fsy182.

Hilborn, R., Walters, C.J., 1987. A general model for simulation of stock and fleet dynamics in spatially heterogeneous fisheries. Can. J. Fish. Aquat. Sci. 44, 1366–1369. https://doi.org/10.1139/f87-163.

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty/Book and Disk. Springer Science & Business Media.

Jiao, Y., Cortés, E., Andrews, K., Guo, F., 2011. Poor-data and data-poor species stock assessment using a bayesian hierarchical approach. Ecol. Appl. 21, 2691–2708. https://doi.org/10.1890/10-0526.1.

Jiao, Y., Hayes, C., Cortés, E., 2009. Hierarchical Bayesian approach for population dynamics modelling of fish complexes without species-specific data. ICES J. Mar. Sci.: J. Cons. 66, 367–377. https://doi.org/10.1093/icesjms/fsn162.

Johnson, S.D.N., Cox, S.P., 2018. Evaluating the role of data quality when sharing information in hierarchical multi-stock assessment models, with an application to Dover sole. Can. J. Fish. Aquat. Sci. https://doi.org/10.1139/cjfas-2018-0048.

Kempf, A., Mumford, J., Levontin, P., Leach, A., Hoff, A., Hamon, K.G., Bartelings, H., Vinther, M., Staebler, M., Poos, J.J., et al., 2016. The MSY concept in a multi-objective fisheries environment-Lessons from the North Sea. Mar. Policy 69, 146–158. https://doi.org/10.1016/j.marpol.2016.04.012.

Martell, S.J., Walters, C.J., Hilborn, R., 2008. Retrospective analysis of harvest management performance for bristol bay and fraser river sockeye salmon (oncorhynchus nerka). Can. J. Fish. Aquat. Sci. 65, 409–424. https://doi.org/10.1139/f07-170.

Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72, 7–18. https://doi.org/10.1093/icesjms/fsu015.

Mortensen, L.O., Ulrich, C., Hansen, J., Hald, R., 2018. Identifying choke species challenges for an individual demersal trawler in the north sea, lessons from conversations and data analysis. Mar. Policy 87, 1–11. https://doi.org/10.1016/j.marpol.2017.09.031.

Ono, K., Haynie, A.C., Hollowed, A.B., Ianelli, J.N., McGilliard, C.R., Punt, A.E., 2017. Management strategy analysis for multispecies fisheries, including technical

interactions and human behavior in modelling management decisions and fishing. Can. J. Fish. Aquat. Sci. 1–18. https://doi.org/10.1139/cjfas-2017-0135.
Pella, J.J., Tomlinson, P.K., 1969. A generalized stock production model. Inter-Am. Trop. Tuna Comm. Bull. 13, 416–497. http://aquaticcommons.org/id/eprint/3536.
Pikitch, E.K., 1987. Use of a mixed-species yield-per-recruit model to explore the consequences of various management policies for the oregon flatfish fishery. Can. J. Fish. Aquat. Sci. 44, s349–s359. https://doi.org/10.1139/f87-336.
Punt, A., Butterworth, D., 1995. The effects of future consumption by the Cape fur seal on catches and catch rates of the Cape hakes. 4. Modelling the biological interaction between Cape fur seals Arctocephalus pusillus pusillus and the Cape hakes Merluccius capensis and M. paradoxus. S. Afr. J. Mar. Sci. 16, 255–285. https://doi.org/10.2989/025776195784156494.
Punt, A.E., 2003. Extending production models to include process error in the population dynamics. Can. J. Fish. Aquat. Sci. 60, 1217–1228. https://doi.org/10.1139/f03-105.
Punt, A.E., Butterworth, D.S., Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. Fish Fish. https://doi.org/10.1111/faf.12104.
Punt, A.E., Dalton, M.G., Foy, R.J., 2020. Multispecies yield and profit when exploitation rates vary spatially including the impact on mortality of ocean acidification on north pacific crab stocks. Fish. Res. 225, 105481. https://doi.org/10.1016/j.fishres.2019.105481.
Punt, A.E., Deng, R., Pascoe, S., Dichmont, C.M., Zhou, S., Plagányi, É.E., Hutton, T., Venables, W.N., Kenyon, R., Van Der Velde, T., 2011a. Calculating optimal effort and catch trajectories for multiple species modelled using a mix of size-structured, delay-difference and biomass dynamics models. Fish. Res. 109, 201–211. https://doi.org/10.1016/j.fishres.2011.02.006.
Punt, A.E., Okamoto, D.K., MacCall, A.D., Shelton, A.O., Armitage, D.R., Cleary, J.S., Davies, I.P., Dressel, S.C., Francis, T.B., Levin, P.S., et al., 2018. When are estimates of spawning stock biomass for small pelagic fishes improved by taking spatial structure into account? Fish. Res. 206, 65–78. https://doi.org/10.1016/j.fishres.2018.04.017.
Punt, A.E., Smith, A.D., Cui, G., 2002. Evaluation of management tools for australia's south east fishery. 1. Modelling the south east fishery taking account of technical interactions. Mar. Freshw. Res. 53, 615–629. https://doi.org/10.1071/MF01007.
Punt, A.E., Smith, D.C., Smith, A.D., 2011b. Among-stock comparisons for improving stock assessments of data-poor stocks: the "Robin Hood" approach. ICES J. Mar. Sci.: J. du Conseil 68, 972–981. https://doi.org/10.1093/icesjms/fsr039.
Quinn, T.J., 2003. Ruminations on the development and future of population dynamics models in fisheries. Nat. Resour. Model. 16, 341–392. https://doi.org/10.1111/j.1939-7445.2003.tb00119.x.
Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Inter-Am. Trop. Tuna Comm. Bull. 1, 23–56. http://aquaticcommons.org/id/eprint/3530.
Schnute, J., 1981. A versatile growth model with statistically stable parameters. Can. J. Fish. Aquat. Sci. 38, 1128–1140. https://doi.org/10.1139/f81-153.
Schnute, J., 1985. A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42, 414–429. https://doi.org/10.1139/f85-057.
Seber, G., 1997. Estimation of Animal Abundance, 2nd ed. Oxford University Press.
Srinivasan, B., Biegler, L.T., Bonvin, D., 2008. Tracking the necessary conditions of optimality with changing set of active constraints using a barrier-penalty function. Comput. Chem. Eng. 32, 572–579. https://doi.org/10.1016/j.compchemeng.2007.04.004.
Thorson, J.T., Cope, J.M., Kleisner, K.M., Samhouri, J.F., Shelton, A.O., Ward, E.J., 2015. Giants' shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis. Fish Fish. 16, 342–361. https://doi.org/10.1111/faf.12061.
Wakefield, W.W., 1984. Feeding Relationships Within Assemblages of Nearshore and Mid-Continental Shelf Benthic Fishes Off Oregon, Master's Thesis. Oregon State University.
Walters, C., 1998. Evaluation of quota management policies for developing fisheries. Can. J. Fish. Aquat. Sci. 55, 2691–2705. https://doi.org/10.1139/f98-172.
Walters, C.J., Bonfil, R., 1999. Multispecies spatial assessment models for the British Columbia groundfish trawl fishery. Can. J. Fish. Aquat. Sci. 56, 601–628. https://doi.org/10.1139/f98-205.
Walters, Carl J., Christensen, Villy, Martell, Steven J., Kitchell, James F., 2005. Possible ecosystem impacts of applying MSY policies from single-species assessments. ICES J. Mar. Sci. 62 (3), 558–568. https://doi.org/10.1016/j.icesjms.2004.12.005.
Winker, H., Carvalho, F., Kapur, M., 2018. JABBA: just another Bayesian biomass assessment. Fish. Res. 204, 275–288. https://doi.org/10.1016/j.fishres.2018.03.010.