This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Chrysafi, Anna; Cope, Jason M.

Testing approaches to determine relative stock abundance priors when setting catch recommendations using data-limited methods

Published in:
FISHERIES RESEARCH

DOI:
[10.1016/j.fishres.2019.105343](https://doi.org/10.1016/j.fishres.2019.105343)

Published: 01/11/2019

Document Version
Peer reviewed version

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Chrysafi, A., & Cope, J. M. (2019). Testing approaches to determine relative stock abundance priors when setting catch recommendations using data-limited methods. *FISHERIES RESEARCH*, 219, [105343].
<https://doi.org/10.1016/j.fishres.2019.105343>

Testing approaches to determine relative stock abundance priors when setting catch recommendations using data-limited methods

Anna Chrysafi^{1,2*} and Jason M. Cope³

¹Water & Development Research Group (WDRG), Aalto University, Tietotie 1E, Espoo, 02150, Finland

²Department of Environmental Sciences, University of Helsinki, Viikinkaari 2a, Helsinki, 00014, Finland

³Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, United States.

Abstract

Data-limited methods for managing stocks have expanded greatly over the last decade due to the necessity of quantitatively assessing exploited populations with limited information. A special category of such approaches is based on stock reduction analysis. These “catch-only” methods provide a way to handle low data availability, but also require as an input relative stock status (e.g., current biomass/initial biomass), a difficult to determine value that leads to large sensitivity in method output and performance. Published methods have been developed to devise informative priors for this quantity, but have not been evaluated together with the assessment methods. Here, relative stock abundance priors derived from elicited expert knowledge, vulnerability analysis and catch trends are compared to the common assumption of a stock being at $B_{40\%}$ (40% of the initial biomass). The performance of each prior source is evaluated both in the degree of bias in estimating stock status and in the estimation procedure of catches for ten data-rich stocks with six stock assessment models that require stock abundance input. The results from both performance metrics show that these alternative sources can provide more accurate priors than assuming current biomass equals $B_{40\%}$, with priors elicited from stock assessment experts performing best. Finally, based on the findings of this work and the data requirements to construct a stock abundance prior, we make recommendations on how to navigate the options for devising a relative stock status prior.

Keywords: Data-limited, Stock abundance, Stock assessment, Fisheries management, Expert knowledge

* Corresponding author: Anna Chrysafi, Tietotie 1E, Espoo, 02150, Finland

e-mail: anna.chrysafi@aalto.fi

1. Introduction

The literature on the development of data-limited stock assessment methods (DLMs) highlights the broad necessity of such approaches (Bentley, 2014; Geromont and Butterworth, 2015; Chrysafi and Kuparinen, 2015), while demonstrating the limitations and applications in real management schemes (Fulton et al., 2016; McDonald et al., 2018). The methods are diverse, ranging from simple indicators of stock status to methods incorporating more biological complexity using simplified population models. Life history parameters (Beddington and Kirkwood, 2005; Mangel, 2005) and some data (e.g., lengths, catches, or catch-per-unit-effort) can be used to derive either a stock status indicator or a catch recommendation (e.g., Froese, 2004; Cope and Punt, 2009; Kokkalis et al., 2014; Hordyk et al., 2015). The methods are often very sensitive to parameter inputs that may not be directly measured, but instead are borrowed from other species and/or derived from empirical methods (Froese and Binohlan, 2000), global databases (Fishbase; Froese and Pauly, 2016) and/or meta-analyses (Prince et al., 2015; Thorson et al., 2017).

A special type of data-limited assessments are the so called “catch-only” methods, which require prior information on relative stock abundance (expressed in the form of B_t/B_r , where B is biomass, t is current year and r is a reference year, often representing the initial condition or carrying capacity of the population). The majority of these methods are modified versions of the classic Stock Reduction Analysis (SRA, e.g., Kimura and Tagart, 1982; Kimura et al., 1984; Cope, 2013; Dick and MacCall, 2011; Martel and Froese, 2013; Froese et al., 2017; Zhou et al., 2018) but not limited to it (MacCall, 2009). These approaches have been applied for actual management decisions (e.g., Dick and MacCall (2011) in Dick and MacCall, 2010; MacCall (2009) in ICES, 2012a; Martel and Froese (2013) in Sabater and Kleiber, 2013 and Sharma, 2013; Cope (2013) in Cope et al., 2015a). The performance of “catch-only” methods has been evaluated several times using simulation (e.g., Wetzel and Punt, 2011; Wiedenmann et al., 2013; Carruthers et al., 2014), and has demonstrated that the estimated yields (e.g., maximum sustainable or optimum yield) are highly affected by under- or over-estimating the relative stock abundance input. However, relative stock abundance (hereafter referred to as “stock status”) is typically a derived stock assessment output and therefore usually not known when needed for a data-limited application.

Consequently, a common practice when applying these models is to assume that the stock is at a target biomass, typically $B_{40\%}$, (e.g., $B_t/B_r=0.4$; Dick and MacCall, 2010), which is only precautionary if the stock is not below 40% of the unfished population. In a search for alternative ways to provide more accurate stock status priors, Cope et al. (2015b) used Productivity-Susceptibility Analysis (PSA; Patrick et al., 2010) vulnerability scores that were estimated excluding susceptibility attributes informed from stock assessments (commonly done when applying PSA to non-assessed stocks), to mimic data-limited scenarios, and suggested that the PSA-derived prior improves the performance of DLMs that need stock status as an input. Another option is to explicitly obtain required information using expert judgment (Berkson and Thorson, 2015; Newman et al., 2015). Chrysafi et al. (2019) elicited stock status priors from experts by providing to them data on catch time-series, limited fishery length compositions and basic life history information, similar to the data-availability in a typical data-limited stock assessment. The elicitation experiment demonstrated that more experienced stock assessment experts may provide more accurate stock status priors than fisheries experts with no experience in stock assessment. Perälä et al. (In revision) developed a hierarchical Bayesian model for expert calibration using the data and results from Chrysafi et al. (2019), to incorporate expected human behavioral bias (Morgan, 2013), and demonstrated that calibrated expert judgments can improve stock status predictions compared to uncalibrated ones and uninformed distributions. Furthermore, in an effort to obtain accurate stock status priors for use in the “catch-only” models, Zhou et al. (2017) developed boosted regression trees (BRTs) that correlate stock status with predictors calculated from catch data in an attempt to automate the calculation of this parameter. All the above approaches aim to provide better alternatives to the default practice of either assuming

current biomass equals $B_{40\%}$ or rough estimates of stock status based on catch (e.g., in Martel and Froese, 2013).

This work explores the above three methods for constructing a stock status prior in “catch-only” models in the following ways: a) how expert elicited stock status priors compare to PSA- and BRT-derived priors, b) how “catch-only” model estimation is affected by the different prior elucidation methods, and c) constructing a decision tree-based guideline for selecting appropriate stock status prior(s) based on performance and resource availability.

2. Materials and methods

2.1 Data collection

A subset of the data from Chrysafi et al. (2019) is used in this work. Ten species managed by the US. Pacific Fishery Management Council (PFMC) with analytical stock assessments using Stock Synthesis (SS; Methot and Wetzel, 2013) are the basis of the analysis. All selected species have estimated vulnerability scores from a PSA conducted for the year 2010 (Cope et al., 2011) which are used to approximate the PSA derived stock status prior from Cope et al. (2015b). Species-specific life history, vulnerability values and model derived stock status, treated as the true stock status, in the last year of the selected time-series are described in Table 1; hereafter species are referred to by their code name as shown in Table 1.

2.2 Stock status prior study cases

Nine stock status prior options (referred to as SSPM hereafter; Table 2) were tested. Cases 1-3 are priors elicited from individuals reported in Chrysafi et al. (2019), representing experts with background in fisheries science but with differing degrees of experience conducting stock assessments. Case 4 aggregates expert opinions using a linear opinion pool approach (Stone, 1961):

$$P(\theta) = \sum_{i=1}^n w_i p_i(\theta) \quad (1)$$

where n is the number of experts, $p_i(\theta)$ represents expert i 's probability distribution for unknown θ , $P(\theta)$ represents the combined probability distribution, and the weights w_i are non-negative and sum to one. An equal weight was assigned to all experts, assuming that we have no prior information on their performance. In case 5, the posterior distributions from the best performing calibration model were used (Model 3 with hierarchical marginally uniform prior for expert bias combined with only experts' mean; Perälä et al., In revision). Case 6 uses the BRT method for catch data (Zhou et al., 2017) and its' implementation in the datalimited2 R package (Free, 2018). To describe the non-beta distributions in cases 4-6, the summary statistics (mean and standard deviation of the smoothing kernel) were used for sampling from a beta distribution. Case 7 is PSA-derived stock status prior, where the vulnerability scores for 2010 (Table 1), were used to approximate stock status from the constructed prior in Cope et al. (2015b). Case 8 is the often-used assumption of a stock being at $B_{40\%}$, while case 9 is using the true, model derived, stock status for year 2010 and was created as a reference of the "best possible" stock status input in the assessment models. Hereafter, all cases are referred to by their code name as described in Table 2.

2.3 Evaluation environment and design

The DLMtool package (Carruthers and Hordyk, 2018) in R was selected as the platform for evaluating the different SSPMs. This package is developed for evaluating management procedures (MPs; made up of the combination of a DLM and a control rule) and includes many DLMs and control rules. The DLMtool is very versatile; a single input data file can be applied, according to data availability, to several DLMs at once. Furthermore, the user can create new DLMs and MPs tailored to user-specific requirements. Here, we were interested only on estimated catches from various DLMs, testing different stock status options, and thus we did not perform closed-loop simulation analyses. The following DLMs available in the DLMtool were used to observe how each of the stock status priors affect “catch-only” DLMs catch estimates: Depletion-corrected average catch (DCAC;

MacCall, 2009), Depletion-based SRA (DBSRA; Dick and MacCall, 2011) and Catch trend Surplus Production MSY (SPMSY; Martel and Froese, 2013). In addition, the updated Monte Carlo version (CMSY; Froese et al., 2017) of Martel and Froese (2013), the Optimized Catch-Only Method (OCOM; Zhou et al., 2018) and Simple Stock Synthesis (SSS; Cope, 2013) were coded for implementation in the DLMtool. Details for each of the specific stock assessment models can be found in the original publications. One notable modified application is that SSS uses the Ricker-Power function (Punt and Cope, 2019) instead of the typical Beverton-Holt stock recruit function (The two versions of SSS and detailed documentation about the differences can be found in <https://github.com/shcaba/SSS>). The Ricker-Power stock recruit function uses inputs similar to DBSRA, and was selected to make sure productivity assumptions across models were matched as best as possible. Input requirements and model specifications for the stock status prior used in each DLM are provided in Table 3. The beta distribution was selected to describe the stock status prior as it is bounded in the $[0,1]$ space and is typically used in data-limited applications.

Further DLM-specific inputs were as follows: After exploring catch estimates from the DLMtool version of DCAC, using the average catch estimated for all the years in the catch time-series, an additional catch scenario was considered and used in our analysis: the average catch was estimated for the years between the maximum catch ($C_{\max}$), in the time-series, and year 2000 when management changes significantly impacted removals. Even though the stock statuses for the tested SSPMs are estimated for year 2010, when using average catch it is more appropriate to exclude low catches due to developing fisheries and/or catches reduced by management, otherwise, the average catch is artificially lowered, and result in catch estimates biased low. SPMSY and CMSY do not input a stock status prior, but instead use relative stock status groupings with lower and upper bounds based on some relationship with catch (if $C_{\text{cur}}/C_{\max} \geq 0.5$ then $B_t/B_r \in [0.3, 0.7]$ and if $C_{\text{cur}}/C_{\max} < 0.5$ then $B_t/B_r \in [0.01, 0.4]$, where C_{cur} is the catch in the last year in the time series). Biomass trajectories that fall within the pre-defined ranges are then accepted, based on the stock status rule and random draws on r (intrinsic growth rate) and K (population carrying capacity). We took the following steps to modify these models such as the retained trajectories represent the specific stock status priors being tested: Before running the models, the upper and lower bounds for the stock status range were set at the 1% and 99% bounds of the desired stock status distribution. After the simulations were completed and the default rules had retained the trajectories that fall within the defined range, the stock status values (and consequently the respective r - K combinations that lead to the accepted biomass levels) were sampled to retain the tested stock status prior. This process was selected on the basis of not modifying how the specific DLMs operate but rather only to add a final rule on the accepted biomass trajectories. This was a necessary change, as the SSPMs tested here provide a distribution that is more informative than bounded uniform priors. Additionally, SPMSY, CMSY and OCOM require input on stock productivity, r . In SPMSY and CMSY r was estimated based on the DLMtool's SPMSY default rule using maximum age and age-at-maturity, assuming $F_{\text{msy}} = r/2$. Alternatively, OCOM uses the same relationship to instead estimate r via $r = 2 * F_{\text{msy}}$, where F_{msy} is derived from multiplication of the F_{msy}/M and M inputs for each species.

Data objects in the DLMtool environment were created with the life history information and the catch time-series for each species (Table 1). The ratio of biomass at maximum sustainable yield (MSY) to virgin biomass (B_{msy}/B_0) was set at $E[B_{\text{msy}}/B_0] = 0.4$, $SD[B_{\text{msy}}/B_0] = 0.1$ (Thorson et al., 2012; Punt et al., 2014). The fishing mortality at MSY to natural mortality M (F_{msy}/M) ratio was set at $E[F_{\text{msy}}/M] = 0.694$, $SD[F_{\text{msy}}/M] = 0.1$ for rockfishes, $E[F_{\text{msy}}/M] = 1.16$, $SD[F_{\text{msy}}/M] = 0.154$ for flatfishes and $E[F_{\text{msy}}/M] = 0.896$, $SD[F_{\text{msy}}/M] = 0.162$ for roundfishes (Zhou et al., 2012). Furthermore, for natural mortality $SD[M] = 0.1$ was set for all species and specifically for SSS, a $CV = 0.1$ was set for length input parameters in the SSS-specific control file. All other life history input parameters were fixed to a single value. Finally, the first age at size zero (t_0) for constructing the von Bertalanffy growth curve, was set to zero for all species. All DLMs with the discussed modifications, input data

files and steps for performing the analysis are available in Github (https://github.com/chrysafi/stock_status_prior_evaluation).

2.4 Performance measures

The predicted stock status distributions from each SSPM were compared to the “true”, model derived, stock status by species shown in Table 1. The performance of the DLMs for all nine cases was evaluated with 1,000 retained runs and the catch estimates were compared to the stock assessment-based estimated catches for 2011 for each species (2010 for Cabezon). After Wetzel and Punt (2011), the following overall performance metric was used to evaluate each SSPM and DLM across species and stock status groupings (stocks $\leq B_{40\%}$ and stocks $> B_{40\%}$):

$$OPA_{c,m} = \sum_1^n |0.5 - Prob_over_{c,m}| \quad (2)$$

where Prob_over is the probability of overestimating the catch estimate relative to the stock assessment estimated catch, n is the total number of species, c is the stock status case and m is the DLM. This metric is used to summarize results across species status groupings, and it can identify which SSPM is more suitable than others overall and how robust DLMs are to stock status misspecification. The assumption here is that both overestimating and underestimating the catch is equally undesirable. In addition, assuming in a data-limited context it is less risky to be underestimating catches, the cumulative probability of overestimating the catch was also reported:

$$OPB_{c,m} = \sum_1^n (Prob_over_{c,m}) \quad (3)$$

This metric is used to distinguish whether overall bad performance is due to overestimating or underestimating catches on average.

Two stock status groups (stocks $\leq B_{40\%}$ and stocks $> B_{40\%}$) are reported to summarize results, as any more groups suffered from low sample sizes. We note that even though overall performance metrics described in equations (2) and (3) are compared relative to the official stock assessment, this work is not designed explicitly to focus on the performance of DLMs relative to a data-rich integrated stock assessment model. The official stock assessment estimated catches are used as a needed common reference point for all tested cases in the DLMs, and the reference case 9 with the unbiased prior, represents the best possible stock status input. Therefore, the performance of each individual DLM for cases 1-8 is evaluated relative to the performance of the assessment-based (“true”) stock status input (case 9).

3. Results

3.1 Stock status predictions

As identified in Chrysafi et al. (2019), elicited stock status priors from fisheries experts with the greatest experience in stock assessment (EXP1) led to less biased, but more uncertain, distributions relative to the other experts. On the other hand, fisheries experts inexperienced in stock assessment (EXP3) overestimate stock status more than any other tested SSPM (Fig. 1; Table 4). In general, the same pattern observed for fisheries experts (Chrysafi et al., 2019), is also present in the other SSPMs, where stock status tends to be overestimated for stocks with low true stock status and underestimated for stocks with higher true stock status (Table 4). The overall most biased and least accurate priors are produced by EXP3 and the default model pre-specified assumption of a stock being at $B_{40\%}$, which clearly shows this latter practice is problematic unless the stock is indeed near its’ target stock status. The expert opinion pool leads to relatively low bias yet imprecise results, which seems a natural result of averaging the biases of all experts that tend to follow opposite directions (Fig. 1). The bias correction model greatly improves predictions for EXP2 and EXP3 for some of the instances but performs poorly in others. The BRT and PSA methods produce similar and consistent stock status predictions with BRT being less biased for stocks with higher true stock status (Fig. 1; Table 4).

3.2 Catch estimation

The effect of stock status misspecification on estimated catches is obvious from Fig. 2, in conjunction with the patterns in Fig. 1. In general, as demonstrated above, the ability of the SSPMs to predict stock status is clearly reflected in the catch estimates, where an overestimation is consistent across DLMs for the stocks with the lowest status and vice versa (Fig 2; Table 5). The average estimated median catches exhibited the least variability, compared to the other cases, in the reference case when the “true” (unbiased but imprecise) stock status is used (Table 5, TS). This confirms that it is possible to achieve relatively good catch predictions overall when given an accurate prior. Among the tested SSPMs, EXP3 yields the highest estimates of catches, with cases of severe overestimations (Fig. 2; Table 5; e.g. PTSL and BCAC). Other notable behavior is observed with the SSS and DBSRA estimates of catches. For CARY, WIDOW, ARRA and STHY especially, both DLMs estimate very high catches with very wide distributions, when using stock status priors with high medians and large uncertainty, leading to the long tails in catch estimation (Figs. 1 and 2). Therefore, the average estimation of catches, especially for EXP1, is highly affected due to these outliers (Table 5). DBSRA also performs poorly (i.e., catch estimates are too high) for DVSL (Fig. 2) and this behavior could be due to the prior issues raised above combined with the specific life-history, as it is not observed for SSS. Additional runs and samples of DBSRA outputs confirmed the repeatability of these results.

3.3 Individual SSPM evaluation and DLM sensitivity

Equations (2) and (3), representing the overall performance of each SSPM (OPA) and the cumulative probability of overestimating catches (OPB) respectively, were compared separately for stocks below $B_{40\%}$ and those above $B_{40\%}$ to be able to identify specific stock status grouping performance. This categorization shows a clear pattern, that regardless of SSPM and DLM, poor performance is generally due to overestimation of catches for stocks with low stock status and underestimation of catches for stocks above $B_{40\%}$ (Fig. 3; Online Supplementary Tables S1-S4). EXP3 exhibits the worst overall performance across DLMs for stocks below $B_{40\%}$, and this is due to the very high probabilities of overestimating catches (Fig. 3a,c; Table 6; Online Supplementary Tables S1 and S3). In contrast, EXP1, OPOOL, ECLBR and EXP2 perform the best overall relative to the reference case with the true stock status. BRT, PSA and $B_{40\%}$ priors perform similarly, with $B_{40\%}$ performing the worst among the three (Fig. 3a,c; Table 6; Online Supplementary Table S1 and S3). From the applied DLMs, DCAC is more robust to stock status misspecification, there is less variation in overall performance as demonstrated by mean absolute deviation (MAD) relative to the reference case. Higher values of MAD are observed for SSS, DBSRA and SPMSY while OCOM and CMSY are affected the most by the stock status misspecification, scoring the highest MAD values (Fig. 3a; Online Supplementary Table S1, MAD).

The pattern is different when considering stocks above $B_{40\%}$. The cumulative probability of overestimating catch level for EXP3 is closer to the perfect score (Fig. 3b,d; Table 6; Online Supplementary Tables S2 and S4), despite EXP3 demonstrating worst overall performance. Overall bad performance of EXP3 is due to inconsistent stock status predictions (overestimating or underestimating), but this is masked in the cumulative probability results. Next in overall worse performance is $B_{40\%}$ and EXP2, which resulted in the greatest distance from the perfect score in OPB, compared to the other SSPMs. In practical terms, the SSPMs are on average underestimating the catch level. The best overall performance is observed for OPOOL, followed by EXP1, PSA, EBCTL and BRT (Fig. 3b,d; Table 6; Online Supplementary Tables S2 and S4). When it comes to the sensitivity of DLMs to stock status misspecification, the pattern changes to a degree. DCAC is still least sensitive to the stock status misspecification, as demonstrated by the MAD relative to TS. Higher MAD values are observed for CMSY, DBSRA and SSS while SPMSY and OCOM are affected the most by the stock status misspecification (Fig. 3b; Online Supplementary Table S2, MAD).

4. Discussion

A major challenge of implementing “catch-only” DLMs is the required input of stock status. Rather than assuming a target biomass (e.g., $B_{40\%}$) or another uninformed value, alternative methods have been developed to construct priors that are more accurate. A general pattern of overestimating stock status, and thus catches, for stocks below $B_{40\%}$ and underestimating stocks status, and thus catches, for stocks above $B_{40\%}$ was observed (Figs. 1-3; Tables 4-5). The performance patterns derived from expert elicitation are indicative of the increased difficulty in predicting stock status and estimating catches, the closer a stock is to either the limits of near extinction or unfished status owed to inherently higher bias at the boundary levels. This problem also seems to affect the non-expert elicitation approaches as well. We demonstrate that priors based on fisheries experts with experience in conducting stock assessment can provide stock status priors with performance comparable to when “true” stock status priors are used (Fig. 3; Tables 4 and 6; Online Supplementary Tables S1 and S2; EXP1 and TS).

Different expert groups may reach different conclusions (ICES, 2015) so it is critical to evaluate the implications of any discrepancies. For example, an expert panel comprised of only stock assessment experts raises little concern, though such a panel is not always available as stock assessment scientists are in short supply (Lynch et al., 2018). When experience in stock assessment is moderate or limited, the constructed priors can be very biased (Fig. 1; Table 4; EXP2 and EXP3). In such cases, individual opinions or a simple opinion pool could be misleading (Fig 2; Table 5) and therefore should be used cautiously and with consideration of other possible priors. In our example, the mixed expert opinion pool produced less biased (but very uncertain) estimates of stock status and as a result, good performance in estimating catches overall, but this was due to combined good and moderate performance of EXP1 and EXP2 respectively and the opposite bias direction between them and EXP3. We therefore recommend the additional step of calibrating expert priors with test data, especially when expertise in stock assessment is limited. Correcting for expert bias can also be useful when stock assessment experience is extensive, since the calibration model (ECLBR) can improve predictions of stock status and the overall performance of DLMs (Figs. 1 and 3; Online Supplementary Tables S1-S4).

Obtaining expert judgments might not always be a feasible (due to lack of data or available participants) or desirable way to construct a stock status prior. In such cases, the other tested methods (BRT and PSA) can be used with acknowledgement of their biases, being significant improvements over default assumptions such as $B_{40\%}$. The $B_{40\%}$ rule performed poorly in our examples (Table 6; Online Supplementary Tables S1 and S2) and failed to provide a precautionary measure of stock status when stocks were below $B_{40\%}$.

In addition, we were also interested in the sensitivity of individual DLMs to different levels of stock status misspecification. Some DLMs appear to be more robust than others (Fig. 3; Online Supplementary Tables S1 and S2; MAD), potentially owing to the additional built-in precaution for a given stock status and life history assumption. However, further evaluation using simulation is required to be able to make recommendations for appropriate DLM selection as done in Carruthers et al. (2014). Specifically, the variants of SPMSY and CMSY used here that altered the rule for accepted biomass trajectories, and the average catch choice for DCAC will likely yield different results than those observed in the original publications and previous performance evaluations thus, necessitating further exploration. In addition, the poor behavior and sensitivity of SSS and DBSRA for high expected value stock status priors, requires further attention. This behavior may be partially attributable to the beta distribution’s shape when a high value for expected stock status is combined with high uncertainty. Due to its nature, that is bounded in the $[0,1]$ space, in these instances, the constructed beta prior exhibits very high densities around the expected value and long tails to the opposite direction, leading to the very wide ranges of catch estimates. Alternative distributions may be more suitable and should be explored for actual stock assessment. In addition, the F_{msy}/M and B_{msy}/B_0 priors used in SSS and DBSRA cannot, by definition, be matched exactly to the productivity

in the actual stock assessment that is defined through steepness h of the stock-recruitment relationship (see Punt and Cope [2019] for a description of this issue). This mismatch leads to a bias in catch estimation and its individual effect can be seen in case 9, where stock status is specified correctly. Using a steepness prior instead of priors for F_{msy}/M and B_{msy}/B_0 when applying SSS may improve performance if the Beverton-Holt stock recruit relationship is appropriate for a particular stock.

In addition, model performance can also be affected by specific life-history. Wetzel and Punt (2011) observed that the output from DLMs was affected when M was misspecified for specific life-histories but only to a limited degree, a finding in agreement with Carruthers et al. (2014). The authors observed very limited interactions between DLMs performance and life-history thus, these effects are not considered in our work. These types of details for each DLM need to be fully understood when evaluating the appropriateness of applying any given method.

Although the sensitivity of specific DLMs to the misspecification of stock status has been discussed in the existing stock assessment literature (MacCall, 2009; Dick and MacCall, 2011; Froese et al., 2017; Cope et al., 2015b; ICES, 2012b, 2014, 2015) exploration of model behavior, is still limited. The tested stock status alternatives were user-specified and thus, conditioned on subjective bias (Wetzel and Punt, 2011, 2015; Cope, 2013; Carruthers et al., 2014; Arnold and Heppell, 2015). Here, we explored alternative formulations of the stock status input, which correspond to a real stock assessment situation. They extend from expert opinion (Chrysafi et al., 2019) to methods specifically developed to assist selection of priors for stock status (Cope et al., 2015b; Zhou et al., 2017). Our approach allows us to evaluate the performance of these different sources and identify their limitations and caveats. Furthermore, although stock status misspecification is an important and critical issue, as it strongly affects estimation procedure, there are no specific recommendations for overcoming the existing constraints to select an appropriate stock status prior. More specifically, in the first such approach, MacCall (2009) mentions that if nothing is known about the value of interest then, it may be appropriate to assume a value (e.g., 0.5) for precautionary purposes. Such an ad-hoc assumption has been used to guide management decisions ($B_{40\%}$ assumption; Dick and MacCall 2010, 2011) with explicit management procedures (DCAC40 and DBSRA40) included in the DLMtool (Carruthers and Hordyk, 2018). Given the challenge in specifying stocks status as an input, it is reasonable to assume that potential DLMs users could go back to past studies as a starting point. In addition, in the DLMtool description for the specific implementations of management procedures, there are no warnings regarding the assumptions of the $B_{40\%}$ rule, and users might prefer the readily available options instead of trying to define their own prior. Here we explicitly demonstrated that such an assumption a) is not precautionary for stocks below $B_{40\%}$, and b) overly precautionary for stocks above $B_{40\%}$, making clear the need for better informed stock status inputs.

General guidelines on selecting an appropriate prior have been lacking even though the existing pool of potential choices has expanded (e.g., Cope et al., 2015b; Chrysafi et al., 2019; Zhou et al., 2017). For example, the ICES WKLIFE working group, specifically intended to improve the assessment of data-limited stocks, has identified the main drawback of these models many times over the years (ICES 2012b, 2014, 2015). One report recommends that assumptions related to the stock status prior should be decided by a group of experts or, preferably, informed by independent data-limited methods. However, it is unclear what such guidelines should be and secondly, decisions about how to construct a stock status prior should be a) specific, meaning that a SSPM should be selected on the basis of available resources and management objectives and b) account for biases identified in performance evaluations, such as that presented here, and not on the assumption that experts' knowledge is inferior compared to stock status indicators such e.g., $L_{\text{mean}}/L_{\text{opt}}$ (L_{mean} represents the mean length in the catch and L_{opt} the optimum length at capture), that are suggested instead (ICES, 2015). The lack of sufficient guidance can be due to the general perception that experts and stakeholders are likely to have reasonably good knowledge about current stock status (Froese et al., 2017). As demonstrated by Chrysafi et al. (2019) experts' knowledge can be reasonably accurate at times, but inaccurate at others, and these vagaries can have significant impacts on the estimated

catches and subsequent management decisions. Therefore, it is necessary to fully explore the potential ranges of estimated catch owed to the associated bias in different sources of stock status information and to construct guidelines for appropriate use of the available methods and to account for this unavoidable characteristic.

To this end, we provide a step-by-step decision tree (Fig. 4) of the potential routes that can be taken during a stock assessment procedure to create the most appropriate stock status prior based on data availability in conjunction with the overall performance of each SSPM (demonstrated in Fig. 3, Table 6 and Online Supplementary Tables S1-S4). The decision tree does not limit the user to a specific approach, but rather recommends appropriate course(s) of action for a given situation. The reason for not creating a clear ranking for the SSPMs in the decision tree is that the appropriate choice is case-specific depending on data availability, research capacity and management objectives. For example, one could construct a prior based on experienced experts (preferably calibrated), but also combine them with BRT and/or PSA derived priors. These SSPMs exhibit consistent patterns in bias behavior that can be accounted for in decision making, thus, it is possible to apply them individually or in an opinion pool. Avoidance of specifying stock status at $B_{40\%}$ is recommended because it strongly assumes the stock is at the target level and the previous catch history represents yields taken sustainably. These recommendations focus only on the methods explored here, and do not include indicators such as $L_{\text{mean}}/L_{\text{opt}}$, $C_{\text{cur}}/C_{\text{max}}$ that provide only point estimates to create ranges of potential stock status (as in Martel and Froese, 2013; Froese et al., 2017). Provided that we have the means to create informative and more accurate priors, there is no need for such point indicators, as valuable information is lost otherwise (Zhou et al., 2017).

4.1 Caveats, future directions and conclusion

This analysis would benefit from an increased number of test stocks. Expert elicitation with calibration data is a long and timely process, thus Chrysafi et al. (2019) were limited in the number of experts examined, though results within and among experts groups were consistent. Furthermore, we only used 10 of the total 18 stocks included in Chrysafi et al. (2019) due to the lack of available PSA-derived stock status for all stocks. Nevertheless, the selected stocks ranged from severely depleted to close to virgin biomass stocks (Table 1) and thus, we were able to identify patterns in the performance of SSPMs and the sensitivity of DLMs. In addition, DBSRA and SSS consistently resulted in catch estimates higher than the other DLMs tested here due to a mismatch in the productivity assumption from the stock assessment, decreasing the overall performance for certain SSPMs. Moreover, SSS performance would have been improved if the Beverton-Holt stock recruitment relationship was used as the productivity behavior would have been consistent with the assumption used in the stock assessments. Nevertheless, the purpose of this study was not to evaluate and compare the DLMs in estimating catches relative to the official stock assessments or which SSPMs can lead to successful management (which depends on objectives) but rather to explore to what degree the SSPMs affect the estimation procedure relative to a reference case. The baseline for evaluation in this study was the best available scientific information (BASI), because expert elicited priors and PSA vulnerability scores cannot be simulated. The benefits and limitations of the BASI approach are discussed extensively in Cope et al. (2015b) as it offers a valuable alternative and/or complement to simulation testing.

The expert calibration model improved the predictions of stock status for some instances, but performed poorly in others. As discussed thoroughly in Perälä et al. (In revision) this is due to the small calibration data sample size. To overcome this limitation and to increase the applicability of such a model in data-limited fisheries stock assessment, we would need to use stock assessment outputs, as already done in Chrysafi et al. (2019) but scaled up to more stocks. A fisheries stock assessment database, such as the RAM legacy database (Ricard et al., 2012) could provide the data that can directly be uploaded to an on-line elicitation tool, such as that developed in Chrysafi et al. (2019). This approach would increase accessibility to calibration data with consistent stocks, thus

creating a reference database for calibration and allowing the performance of experts to be comparable. The accumulation of experts that complete the calibration data-sets could lead to further understanding of expert biases and even explore the value of information (Chrysafi et al. 2019), allowing the expert elicitation approach to become more widely available. Determining a stock status prior is an important and often complicated issue because it can bring into the estimation procedure a source of bias with meaningful consequences. We are hopeful the results of this work will raise questions about the implications of inappropriate stock status prior selection and that the recommended steps (Fig. 4) will be beneficial in assisting data-limited practitioners to be thoughtful about and use the best way(s) to determine this important input parameter.

Acknowledgements

This work has received funding from the University of Helsinki. We would also like to thank Owen Hamel and two anonymous reviewers for their thoughtful comments that helped improve this article.

Figure 1: Predicted stock status for the nine cases for each of the 10 selected species. Species are arranged from low stock status to higher stock status. The horizontal dashed line indicates true stock status point estimate from SS outputs.

Figure 2: Estimated catches for each species, for all cases. a) SSS, b) DBSRA, c) DCAC, d) OCOM, e) CMSY, and f) SPMSY. The vertical solid line is the stock assessment catch for 2011 and the vertical dashed line is the median estimated catch from the average catch function in DLMtool after 1,000 repetitions. X-axis upper limit was drawn at the 85% percentile of the variable with the highest values.

Figure 2 continued

Figure 1 consists of four heatmaps (a, b, c, d) showing the performance of six methods (SSS, DBSRA, DCAC, OCOM, CMSY, SPMSY) across eight scenarios (EXP1, EXP2, EXP3, OPOOL, ECLBR, BRT, PSA, B40%, TS). The heatmaps are arranged in a 2x2 grid. The top row (a, b) shows OPA (Worst to Best) and the bottom row (c, d) shows OPB (Over to Under). The columns represent the scenarios. The rows represent the methods. The color scale for OPA ranges from black (Worst) to white (Best). The color scale for OPB ranges from red (Over) to blue (Under). The heatmaps show that the methods generally perform better (higher OPA, lower OPB) in the OPOOL and ECLBR scenarios compared to the other scenarios.

Figure 4: Alternative routes to create a stock status prior. $B_{40\%}$ is presented here as it has been used in the past in lieu of any other information, but should be cautiously employed. Arrows indicate information/ impulse flows.

References

- Arnold, L.M., Heppell, S.S., 2015. Testing the robustness of data-poor assessment methods to uncertainty in catch and biology: a retrospective approach. *ICES J. Mar. Sci.* 72 (1), 243-250, <https://doi.org/10.1093/icesjms/fsu077>
- Beddington, J.R., Kirkwood, G.P., 2005. The estimation of potential yield and stock status using life-history parameters. *Phil. Trans. Roy. Soc. B.* 360, 163–170, <https://doi.org/10.1098/rstb.2004.1582>
- Bentley, N., 2014. Data and time poverty in fisheries estimation: potential approaches and solutions. *ICES J. Mar. Sci.* 72 (1), 186-193, <https://doi.org/10.1098/rstb.2004.1582>
- Berkson, J., Thorson, J. T., 2015. The determination of data-poor catch limits in the United States: is there a better way? *ICES J. Mar. Sci.* 72(1), 237-241, <https://doi.org/10.1093/icesjms/fsu085>
- Carruthers, T., Punt, A., Walters, C., MacCall, A., McAllister, M., Dick, E., Cope, J., 2014. Evaluating methods for setting catch limits in data-limited fisheries. *Fish. Res.* 153, 48–68, <https://doi.org/10.1016/j.fishres.2013.12.014>
- Carruthers, T., Hordyk, A.R., 2018. The Data-Limited Methods Toolkit (DLMtool): An R package for informing management of data-limited populations. *Meth. Ecol. Evol.* 9 (12), 2388-2395, <https://doi.org/10.1111/2041-210X.13081>
- Chrysafi, A., Kuparinen A., 2015. Assessing abundance of populations with limited data: Lessons learned from data poor fisheries stock assessment. *Env. Rev.* 24, 25-38, <https://doi.org/10.1139/er-2015-0044>
- Chrysafi, A., Cope, J.M., Kuparinen A., 2019. Eliciting expert knowledge to inform stock status for data-limited stock assessments. *Mar. Policy*, 101, 167-176, <https://doi.org/10.1016/j.marpol.2017.11.012>
- Cope, J.M., Punt, A.E., 2009. Length-Based Reference Points for Data-Limited Situations: Applications and Restrictions. *Mar. Coast. Fish.* 1(1), 169-186, <https://doi.org/10.1577/C08-025.1>
- Cope, J.M., DeVore, J., Dick, E.J., Ames, K., Budrick, J., Erickson, D.L., Grebel, J., Hanshew, G., Jones, R., Mattes, L., Niles, C., Williams S., 2011. An approach to defining stock complexes for US West Coast Groundfishes using vulnerabilities and ecological distributions. *North Am. J. Fish. Man.*, 31, 589-604, <https://doi.org/10.1080/02755947.2011.591264>
- Cope, M.J., 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. *Fish. Res.* 142, 3-14, <https://doi.org/10.1016/j.fishres.2012.03.006>
- Cope, M.J., Dick, E.J., MacCall, A., Monk, M., Soper, B., Wetzel, C., 2015a. Data-moderate stock assessments for brown, China, copper, sharpchin, striptail, and yellowtail rockfishes and English and rex soles in 2013, <https://swfsc.noaa.gov/publications/CR/2013/2013Cope.pdf>
- Cope, M.J., Thorson, J.T., Wetzel, C.R., DeVore, J., 2015b. Evaluating a prior on relative stock status using simplified age-structured models. *Fish. Res.* 171, 101-109, <https://doi.org/10.1016/j.fishres.2014.07.018>
- Dick, E.J., MacCall, A.D., 2010. Estimates of sustainable yield for 50 data-poor stock in the Pacific coast groundfish fishery management plan. NOAA Technical Memorandum, <https://swfsc.noaa.gov/publications/TM/SWFSC/NOAA-TM-NMFS-SWFSC-460.pdf>
- Dick, E.J., MacCall, A.D., 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. *Fish. Res.* 110, 331–341, <https://doi.org/10.1016/j.fishres.2011.05.007>
- Free, C.M., 2018. Datalimited2: More stock assessment methods for data-limited fisheries. R package version 0.1.0. <https://github.com/cfree14/datalimited2>
- Froese, R. and Binohlan, C., 2000. Empirical relationships to estimate asymptotic length, length at first maturity and length at maximum yield per recruit in fishes, with a simple method to evaluate length frequency data. *J. Fish Biol.* 56 (4), 758-773, <https://doi.org/10.1111/j.1095-8649.2000.tb00870.x>

- Froese, R., 2004. Keep it simple: three indicators to deal with overfishing. *Fish Fish.* 5, 86–91, <https://doi.org/10.1111/j.1467-2979.2004.00144.x>
- Froese, R., Pauly, D., Editors. 2016. FishBase. World Wide Web electronic publication, www.fishbase.org
- Froese, R., Demirel, N., Coro, G., Kleisner, K.M., Winker, H., 2017. Estimating fisheries reference points from catch and resilience. *Fish Fish.* 18, 506–526, <https://doi.org/10.1111/faf.12190>
- Fulton, E.A., Punt, A.E., Dichmont, C.M., Gorton, R., Sporcic, M., Dowling, N., Little, L.R., Haddon, M., Klaer, N., Smith, D.C., 2016. Developing risk equivalent data-rich and data-limited harvest strategies. *Fish. Res.* 183, 574–587, <https://doi.org/10.1016/j.fishres.2016.07.004>
- Geromont, H.F., Butterworth, D.S., 2015. A Review of assessment methods and the development of management procedures for data-poor fisheries. FAO Report, the Marine Resource Assessment and Management Group (MARAM), University of Cape Town, South Africa, https://www.researchgate.net/publication/297195982_FAO_Report_A_review_of_assessment_methods_and_the_development_of_management_procedures_for_data-poor_fisheries_219pp
- Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J., 2015. A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. *ICES J. Mar. Sci.* 72(1), 217–231, <https://doi.org/10.1093/icesjms/fsu004>
- ICES, 2012a. ICES Implementation of Advice for Data-limited Stocks in 2012 in its 2012 Advice, ICES DLS Guidance Report. ICES CM 2012/ACOM:68, <https://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2012/ADHOC/DLS%20Guidance%20Report%202012.pdf>
- ICES, 2012b. Report of the Workshop on the Development of Assessments based on LIFE history traits and Exploitation Characteristics (WKLIFE). ICES CM 2012/ACOM:36, <http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2012/WKLIFE/WKLIFE%202012.pdf>
- ICES, 2014. Report of the Workshop on the Development of Quantitative Assessment Methodologies Based on Life-history Traits, Exploitation Characteristics, and Other Relevant Parameters for Data-limited Stocks (WKLIFE IV). ICES CM 2014/ACOM: 54, http://ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2014/WKLIFE4/wklifeIV_2014.pdf
- ICES, 2015. Report of the Fifth Workshop on the Development of Quantitative Assessment Methodologies based on Life-history Traits, Exploitation Characteristics and other Relevant Parameters for Data-limited Stocks (WKLIFE V). ICES CM 2015/ACOM:56, http://ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2015/WKLIFEV/wklifeV_2015.pdf
- Kimura, D., Tagart, J., 1982. Stock reduction analysis, another solution to the catch equations. *Can. J. Fish. Aquat. Sci.* 39, 1467–1472, <https://doi.org/10.1139/f82-198>
- Kimura, D.K., Balsinger, J.W., Ito, D.H., 1984. Generalized stock reduction analysis. *Can. J. Fish. Aquat. Sci.* 41(9), 1325–1333, <https://doi.org/10.1139/f84-162>
- Kokkalis, A., Thygesen, U.H., Nielsen, A., Andersen, K.H., 2014. Limits to the reliability of size-based fishing status estimation for data-poor stocks. *Fish. Res.* 171, 4–11, <https://doi.org/10.1016/j.fishres.2014.10.007>
- Lynch, P. D., Methot, R. D., Link, J. S., (eds.) 2018. Implementing a Next Generation Stock Assessment Enterprise. An Update to the NOAA Fisheries Stock Assessment Improvement Plan. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/ SPO-183, 127 p. <https://doi.org/10.7755/TMSPO.183>
- MacCall, A.D., 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. *ICES J. Mar. Sci.* 66, 2267–2271, <https://doi.org/10.1093/icesjms/fsp209>

- Mangel, M., 2005. Invariant Ratios vs Dimensionless Ratios. *Science* 310 (5753), 1426-1427, <https://doi.org/10.1126/science.310.5753.1426>
- Martel, S., Froese, R., 2013. A simple method for estimating MSY from catch and resilience. *Fish. Res.* 14, 504–514, <https://doi.org/10.1111/j.1467-2979.2012.00485.x>
- McDonald, G., Campbell, S.J., Karr, K., Clemence, M., Granados-Dieseldorff, P., Jakub, R., Kartawijaya, T., Mueller, J.C., Prihatinningsih, P., Siegel, K., Syaifudin, Y., 2018. An adaptive assessment and management toolkit for data-limited fisheries. *Ocean Coast. Manag.* 152, 100–119, <https://doi.org/10.1016/j.ocecoaman.2017.11.015>
- Methot, R.D., Jr., Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.* 142, 86–89, <https://doi.org/10.1016/j.fishres.2012.10.012>
- Morgan, M.G., 2013. Use (and abuse) of expert elicitation in support of decision making for public policy. *PNAS* 111 (20), 7176-7184, <https://doi.org/10.1073/pnas.1319946111>
- Newman, D., Berkson J., L. Suatoni, 2015. Current methods for setting catch limits for data-limited fish stocks in the United States. *Fish. Res.* 164, 86-93, <https://doi.org/10.1016/j.fishres.2014.10.018>
- Patrick, W.S., Spencer, P., Link, J., Cope, J., Field, J., Kobayashi, D., Lawson, P., Gedamke, T., Cortes, E., Ormseth, O., 2010. Using productivity and susceptibility indices to assess the vulnerability of United States fish stocks to overfishing. *Fish. Bull.* 108, 305–322, <http://fishbull.noaa.gov/1083/patrick.pdf>
- Prince, J., Hordyk, A., Valencia, S.R., Loneragan, N., Sainsbury, K., 2015. Revisiting the concept of Beverton–Holt life-history invariants with the aim of informing data-poor fisheries assessment. *ICES J. Mar. Sci.* 72(1), 194-203, <https://doi.org/10.1093/icesjms/fsu011>
- Punt, A.E., Cope, J.M., 2019. Extending integrated stock assessment models to use non-depensatory three-parameter stock-recruitment relationships. *Fish. Res.* 217, 46-57, <https://doi.org/10.1016/j.fishres.2017.07.007>
- Ricard, D., Minto, C., Jensen, O.P., Baum, J.K., 2012. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish. Res.* 13 (4), 380-398, <https://doi.org/10.1111/j.1467-2979.2011.00435.x>
- Sabater, M., Kleiber, P., 2013. Improving specification of acceptable biological catches of data-poor reef fish stocks using a biomass-augmented catch MSY approach. WPRFMC, http://www.wpcouncil.org/wp-content/uploads/2014/02/9.C2-Sabater-and-Kleiber-2013-Technical-Report-Augmented-Catch-MSY-Approach_FINAL.pdf
- Sharma, R., 2013. Stock assessment of three billfish species in Indian ocean, blue, black and striped marlin using stock reduction methods. IOTC-WPB11-28, <https://www.iotc.org/documents/stock-assessment-three-billfish-species-indian-ocean-blue-black-and-striped-marlin-using>
- Stone, M., 1961. The opinion pool. *Ann. Math. Stat.* 32, 1339–1342, <https://www.jstor.org/stable/2237933>
- Thorson, J.T., Cope, J.M., Branch, T.A., Jensen, O.P., 2012. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Can. J. Fish. Aquat. Sci.* 69, 1556–1568, <https://doi.org/10.1139/f2012-077>
- Thorson, J. T., Munch, S.B., Cope, J.M. and Gao J., 2017. Predicting Life History Parameters for All Fishes Worldwide. *Ecol. App.* 27(8), 2262–76, <https://doi.org/10.1002/eap.1606>
- Wetzel, C.R., Punt, A.E., 2011. Model performance for the determination of appropriate harvest levels in the case of data-poor stocks. *Fish. Res.* 110, 342–355, <https://doi.org/10.1016/j.fishres.2011.04.024>
- Wetzel, C.R., Punt, A.E., 2015. Evaluating the performance of data-moderate and catch-only assessment methods for U.S. west coast groundfish. *Fish. Res.* 171, 170-187, <https://doi.org/10.1016/j.fishres.2015.06.005>

- Wiedenmann, J., Wilberg, M.J., Miller, T.J., 2013. An evaluation of harvest control rules for data-poor fisheries. *North Am. J. Fish. Man* 33, 845–860, <https://doi.org/10.1080/02755947.2013.811128>
- Zhou, S., Yin, S., Thorson, J.T., Smith, A.D.M., Fuller M., 2012: Linking fishing mortality reference points to life history traits: an empirical study. *Can. J. Fish. Aquat. Sci.*, 69(8), 1292-1301, <https://doi.org/10.1139/f2012-060>
- Zhou, S., Punt, A. E., Ye, Y., Ellis, N., Dichmont, C. M., Haddon, M., Smith, D. C. et al., 2017. Estimating stock depletion level from patterns of catch history. *Fish Fish.* 18 (4), 742-751, <https://doi.org/10.1111/faf.12201>
- Zhou, S., Punt, A.E., Smith, A.D.M., Ye, Y., Haddon, M., Dichmont, C.M., Smith, D.C., 2018. An optimized catch-only assessment method for data poor fisheries. *ICES J. Mar. Sci.* 73 (3), 964-976, <https://doi.org/10.1093/icesjms/fsx226>

Table 1: Species-specific information included in this analysis. Stock status in 2010 as estimated in the official stock assessment output. Species are ordered from low to high stock status.

Species	Petrale sole	Canary rockfish	Bocaccio	Darkblotched rockfish	Cabezon*	Widow rockfish	Aurora rockfish	Longspine thornyhead	Shortspine thornyhead	Dover sole
Scientific Name	<i>Eopsetta</i>	<i>Sebastes pinniger</i>	<i>Sebastes</i> <i>jordani</i>	<i>Sebastes crameri</i>	<i>Scorpaenichthys</i> <i>marmoratus</i>	<i>Sebastes entomelas</i>	<i>Sebastes aurora</i>	<i>Sebastolobus</i> <i>altivelis</i>	<i>Sebastolobus</i> <i>alascanus</i>	<i>Microstomus</i> <i>pacificus</i>
Species code	PTSL	CARY	BCAC	DRBL	CBEZ	WDOW	ARRA	LTHY	STHY	DVSL
Vulnerability ₂₀₁₀	1.54	2.01	1.93	1.92	1.68	2.05	2.1	1.8	1.96	1.54
Natural mortality M	0.15	0.076	0.15	0.05	0.25	0.12	0.035	0.11	0.05	0.116
von Bertalanffy K	0.13	0.125	0.215	0.2	0.15	0.2	0.091	0.1	0.018	0.15
L _∞ (cm)	54	60	68	43	60	48	32	28	75	48
L _{50mat} (Length at 50% maturation)	33	40.5	40	34.5	35	33	25.5	17	18	35
A _{max} (Maximum age)	40	95	45	105	20	35	125	100	100	60
Weight-Length(W- L) relationship a	2.09E-06	1.55E-05	7.36E-06	1.11E-05	9.18E-06	5.45E-06	9.93E-06	4.3E-06	4.77E-06	2.81E-06
W-L relationship b	3.473703	3.03	3.114	3.13512	3.118	3.28781	3.144807	3.352	3.264	3.345
Stock Status ₂₀₁₀	0.104	0.225	0.238	0.29	0.452	0.507	0.63	0.684	0.75	0.847

*Stock status in 2009 when the last stock assessment was conducted.

Table 2: Alternative tested options for the stock status input requirement

Cases	Case Code	Stock status prior	Distribution
Case 1	EXP1	Experienced expert	Beta
Case 2	EXP2	Novice expert	Beta
Case 3	EXP3	Inexperienced expert	Beta
Case 4	OPOOL	Opinion pool of the 3 experts	non-parametric
Case 5	ECLBR	Expert-corrected bias MCMC	non-parametric
Case 6	BRT	BRTs and catch history	Skewed normal
Case 7	PSA	PSA and vulnerability scores	Beta
Case 8	B40%	The stock is at B _{40%}	Beta with $E[x]=0.4$ and $SD[x]=0.2$
Case 9	TS	True stock status	Beta with $E[x]=TS$ and $SD[x]=0.2$

Table 3: Stock status prior modifications and input requirements for the selected DLMs.

Model	Original stock status prior	Modified stock status prior	Input requirements
SSS	Beta	-	Catch, M^1 , F_{msy}/M^2 , B_{msy}/B_0^3 , Bt/B_r^4 , $MaxAge^5$, K^6 , $t0^7$, L_∞^8 , L_{mat50}^9 , $W-L$ parameters ¹⁰
DBSRA	Beta	-	Catch, M , F_{msy}/M , B_{msy}/B_0 , Bt/B_r , K , $t0$, L_∞ , L_{mat50}
DCAC	Log-normal	Beta	Average Catch, M , F_{msy}/M , B_{msy}/B_0 , Bt/B_r
SPMSY	lower and upper bound	Beta	Catch, Bt/B_r , $MaxAge$, K , $t0$, L_∞ , L_{mat50}
CMSY	lower and upper bound	Beta	Catch, Bt/B_r , $MaxAge$, K , $t0$, L_∞ , L_{mat50}
OCOM	BRTs derived – skewed normal distribution	Beta	Catch, M , F_{msy}/M , Bt/B_r

¹ Natural mortality² Fishing mortality at maximum sustainable yield (MSY) to natural mortality ratio³ Biomass at MSY to virgin biomass ratio⁴ Current biomass to reference year r biomass ratio⁵ Maximum age attained⁶ Von Bertalanffy curve growth rate⁷ Age at length zero in Von Bertalanffy curve⁸ Asymptotic length in Von Bertalanffy curve⁹ Length at 50% maturity¹⁰ Weight-length relationship a and b parameters

Table 4: Probability of overestimating stock status for each tested method relative to TS. Dark grey shaded cells indicate Prob_over>75%, white cells indicate a 25%<Prob_over<75% and light grey shaded cells indicate Prob_over<25%.

	EXP1	EXP2	EXP3	OPOOL	EBCTL	BRT	PSA	B40%	TS
PTSL	0.72	0.98	1	0.9	0.85	0.96	0.96	0.95	0.24
CARY	0.68	0.58	1	0.75	0.58	0.52	0.65	0.78	0.41
BCAC	0.5	0.34	1	0.6	0.4	0.65	0.78	0.74	0.41
DRBL	0.47	0.48	1	0.65	0.49	0.49	0.65	0.67	0.43
CBEZ	0.64	0.58	0.84	0.68	0.98	0.84	0.57	0.4	0.49
WDOW	0.63	0.17	0.86	0.55	0.18	0.05	0.23	0.27	0.51
ARRA	0.67	0.29	0.44	0.46	0.42	0.02	0.1	0.14	0.53
LTHY	0.21	0	0.99	0.4	0.02	0.55	0.42	0.11	0.56
STHY	0.75	0.27	0.03	0.34	0.03	0.49	0.11	0.05	0.58
DVSL	0.45	0.01	0	0.15	0.07	0.32	0.17	0.01	0.68

Table 5. Official stock assessment catches for the selected species and average % of under/overestimation from the median estimated catches among DLMs for each tested SSPM. Light grey shaded cells indicate an overestimation greater than 50%.

	Catch 2011 (t)	EXP1	EXP2	EXP3	OPOOL	EBCTL	BRT	PSA	B40%	TS
PTSL	1021	78.9%	140.3%	1240.1%	152.3%	152.1%	155.1%	136.2%	142.3%	20.2%
CARY	586	229.3%	88.6%	306.3%	176.7%	75.4%	59.4%	103.3%	157.7%	35.5%
BCAC	737	183.3%	76.6%	665.1%	264.9%	93.2%	177.9%	239.6%	240.1%	91.0%
DRBL	508	-15.5%	-25.4%	103.4%	-3.2%	-31.2%	-28.9%	-2.3%	3.9%	-33.9%
Average ≤B_{40%}		119.0%	70.0%	578.7%	147.7%	72.4%	90.9%	119.2%	136.0%	28.2%
CBEZ¹¹	359	-64.1%	-68.9%	-64.2%	-66.2%	-3.3%	-53.3%	-69.1%	-76.6%	-72.7%
WDOW	4872	82.6%	-17.0%	129.7%	29.8%	-4.5%	-19.1%	-6.7%	15.5%	51.5%
ARRA	47	78.7%	-36.9%	-9.2%	-13.5%	-19.9%	-77.0%	-67.0%	-57.8%	6.0%
LTHY	3571	-71.2%	-83.0%	187.7%	-68.5%	-76.4%	4.6%	-24.8%	-57.6%	4.1%
STHY	2384	317.7%	-9.5%	-23.9%	-3.6%	-58.5%	-21.8%	-59.5%	-56.5%	81.4%
DVSL	44400	-56.5%	-78.0%	-75.9%	-73.2%	-55.5%	-45.5%	-68.2%	-75.4%	-36.0%
Average > B_{40%}		47.9%	-48.9%	24.0%	-32.5%	-36.4%	-35.3%	-49.2%	-51.4%	5.7%

¹¹ for year 2010

Table 6. Summary statistics for performance metrics (OPA and OPB) from equations (2) and (3) respectively. The performance for each SSPM is summed across DLMs for the stocks included in the two species groupings. Perfect score describes the cases that the probability of overestimating the catch level is exactly 50% and the worst score the cases that the probability is either 0% or 100%. Scores vary relative to the number of stocks in each species group.

	Grouping	EXP1	EXP2	EXP3	OPOOL	EBCTL	BRT	PSA	B40%	TS	Perfect score	Worst score
OPA	$B \leq B_{40\%}$	4.12	5.14	9.70	4.89	4.85	5.80	6.16	6.33	3.60	0	12
	$B > B_{40\%}$	10.41	11.96	13.81	9.55	10.60	10.96	10.49	12.81	9.76	0	18
OPB	$B \leq B_{40\%}$	16.07	16.14	23.15	16.99	15.67	16.59	18.74	19.03	12.83	12	0/24
	$B > B_{40\%}$	13.82	7.16	15.42	11.19	9.15	9.29	7.76	6.42	15.45	18	0/36