INTERNATIONAL REVIEW PANEL REPORT FOR THE 2018 INTERNATIONAL FISHERIES STOCK ASSESSMENT WORKSHOP
26 -30 November 2018, UCT

[Sean Cox[1], Sarah Gaichas[2], Malcolm Haddon[3], and André E Punt[4]]

Summary of general issues

The Panel recognised the very high quality of the research presented at the 2018 International Fisheries Stock Assessment Review Workshop. This included research on Southern African hake, sardine and rock lobster, as well as new methods for conducting stock assessments where the main data sources are indices of abundance and catch. The Panel thanked the workshop participants for their hard work preparing and presenting the workshop papers, for the extra analyses undertaken during the workshop, and for the informative input provided during discussions.

This report starts with observations from the Panel on some general issues for the species and analysis programmes reviewed, and then focuses on answering the questions posed by DAFF Research, providing a more detailed technical review where necessary, and finally recommending further work concerning each fishery. The recommendations are annotated by their priorities (H, M, L). Much of this report reflects responses to the questions. For ease of reading, answers to the questions that also have research components are indicated by an asterisk.

Hake

The Panel was tasked with evaluating several modifications to the models and OMP projection scenarios leading up to a proposed new OMP-2018. The bridging analysis was especially helpful in facilitating the review, and the Panel commends the analysts for the speed and quality of work conducted during the workshop in tracking down mechanisms behind certain model behaviours and changes in assessments of status (e.g. current biomass relative to B_{MSY}). However, several important issues related to model convergence and parameter stability were identified for models in the reference set that could not be fully resolved during the workshop. The Panel, therefore, was unable to address some of the key questions related to the performance of the proposed OMP-2018 and proposed metarules, aside from apparently adequate performance in some robustness tests. Panel recommendations mostly suggest ways to diagnose and improve the assessment process and robustness tests, so as to ensure that OMP simulations provide reliable indications of expected performance.

Sardine

The Panel evaluated several supporting analyses for the sardine OMP-18 revision, including methods for projecting recruitment, for approximating assessment model estimation uncertainty and for establishing risk equivalence across old and new sardine operating models. Overall, the Panel supported the intent to project recruitment with a similar distribution to that observed historically, and made recommendations to further refine how recruitment is generated for projections. The Panel supported methods proposed by the sardine analysts to evaluate risk to the stock, noting that risk measures will vary among Operating Models and trials (previous OMP evaluations for sardine had few trials so this was less of an issue in the past), and recommended a consistent approach to risk metrics. Other Panel recommendations related to sardines apply in general to all assessed species. In general, the Panel considered it

[1] School of Resource and Environmental Management, Simon Fraser University, Canada
[2] Northeast Fisheries Science Center, NOAA, Woods Hole, USA
[3] CSIRO Oceans and Atmosphere, Hobart, Australia, Australia
[4] School of Aquatic and Fishery Sciences, University of Washington, USA

reasonable to use the simplified approach to approximating estimation uncertainty from the assessment using a multivariate normal approximation centred on the variance-covariance matrix from an inverted Hessian. The Panel offered a structured approach based on this idea to streamline evaluation of estimation uncertainty when MCMC sampling is not possible given time constraints.

West Coast Rock Lobster

The analysis of trends in poaching catches has been much improved by compliance data that recently became available and that link policing effort type to confiscations (the 'new' database) (though these are somewhat limited in number). Interestingly, the trend in poaching (taken to be indexed by the ratio of confiscations per unit of policing effort) over time using the revised data and analyses is similar to that computed last year for both super-areas 3+4+5+6+7 and super-area 8+. Nevertheless, this estimate of trend in poaching is more reliable given the availability of the 'new' database. The Panel recommends that, to enable more detailed analyses, this database be improved further by recording cases in which no confiscations occurred rather than only those cases in which there was a confiscation.

The estimates of poaching are based on the compliance data and data from TRAFFIC. However, these data sources do not account for domestic sales of illegal catches. The Panel encourages efforts to obtain data on local sales to better estimate total poaching catches, and hence to better evaluate the impact on this resource.

The latest stock assessment indicates that the WCRL resource has been at low levels since about 1968 and is now down to 1.8% the estimated 1910 biomass[5]. This pattern of long-term decline implies low resilience to historical removals (legal and illegal) combined with environmental shocks and periods of low recruitment. Poaching is now at high levels irrespective of how it is estimated.

The projections used to select WCRL TACs consistent with avoiding further decline were implemented by projecting poaching at current levels and the central tendency of recent recruitment (given by the geometric mean) forward through to 2025. These projections could be improved in several ways: (a) bias-correct the geometric mean assuming log-normality, (b) use an arithmetic mean recruitment, (c) use bootstrap samples of the empirical distribution of recruitment values in the projections, or (d) preferably by re-parameterizing the 1975-2017 recruitment parameters via an estimated mean level ($\bar{R}$) multiplied by annual recruitment deviates. This last parameterization would enable projections via randomly selecting recruitment values from their estimated distribution. Even so, further potential declines are predicted without a substantial reduction in both catch and poaching.

Assuming that poaching continues at current levels, a 244 t TAC is needed to achieve a male WCRL harvestable biomass 7% above the baseline 2006 level by year 2025. The TAC for the current season has already been reduced from 1,924 t to 1,084 t, with another reduction down to 244 t proposed for the following season. While the increase in male harvestable biomass is 30% over the estimated level in 2018, the predicted harvestable biomass in 2015 will still be only 2.31% of the estimated 1910 level. Improvements in recruitment projection methods, as suggested above, may result in a lesser TAC decrease than currently indicated for the 2019/2020 season.

Linefish

The Panel evaluated the JABBA-Select assessment framework. This method is appropriate for "data moderate" scenarios in which a time-series of catches and at least an index of abundance is available. It is based on an age-aggregated model of the population dynamics. Uncertainty

[5] More specifically, the metric used is the male biomass above the lowest legal size of 75 mm carapace length.

is quantified using Bayesian methods. Unlike earlier methods, JABBA-Select can account for changes over time in selectivity, and is parameterized in terms of biological parameters. Simulations, the results of which were presented to the Panel, indicate that JABBA-Select performs at least as well as, and usually better than, alternative approaches. The Panel thus endorses JABBA-Select for use in data-moderate assessment cases, including for local linefish species.

General considerations

There is a need for greater attention to be given to error checking of assessments: consideration should be given to technical checking of an assessment before it is presented at a DAFF Scientific Working Group. Appendix 1 of this document provides a "process check list" to help such technical task teams conduct and document assessments.

A. Assessment issues common across all stocks

A.1. Reporting of assessment results

The Panel's ability to fully review some of the assessments was hampered due to a lack of standardized reporting of model outputs and common assessment model diagnostics. The Panel notes that other jurisdictions (e.g. the west coast[6] and the northeast of the US[7]) have developed Terms of Reference for stock assessments that specify model outputs and diagnostics needed for a high-quality review. These and other similar documents should be consulted and Terms of Reference for assessments developed that provide relevant information for South African fisheries. In particular, the assessment reviews would have been easier to conduct had a "bridging analysis" (e.g. MARAM/IWS/2018/Hake/P2) been available for all stocks, and also had assessment documents reported both an historical analysis[8] and a retrospective analysis[9]. The results of jittering analyses (which are now routinely conducted) should be included in the assessment document where model performance is sensitive (measured in some standard way) to initial parameter values. Further, simple summaries of data sources with year and sample size are helpful to reviewers because they identify potential data gaps and changes between assessments (see Figure 1 for an example plot, and Appendix 1 for one potential sequence of steps for conducting assessments – note also implementing a scheme such as that outlined in Appendix 1 will have both logistical and financial implications).

A.2. What to do when the Hessian matrix is not positive definite

The inverse of the Hessian matrix is an estimate of the variance-covariance matrix. The lack of a positive definite Hessian matrix is suggestive of a lack of convergence and precludes automatic application of MCMC sampling methods within ADMB. If the Hessian is not positive definite, the maximum gradient of the objective function should be examined, and the gradients for individual parameters examined to determine which parameter(s) is (are) likely leading to the lack of convergence. There are also benefits to using the ADMB "derivative checker" to check that analytical derivatives are consistent with numerical approximations. When differences are found, this may indicate a programming error.

A positive definite Hessian matrix is required to start the default MCMC algorithm in ADMB. It is possible to specify the variance-covariance matrix directly, rather than use the matrix that arises from the fitting algorithm. A generalized inverse based on a non-positive

---

[6] https://www.pcouncil.org/wp-content/uploads/2018/10/Stock_Assessment_ToR_2019-20_SEPT2018_Final.pdf

[7] https://www.nefsc.noaa.gov/saw/saw66/saw-sarc66-tor-fluke-strpbass-2018-01-16.pdf

[8] An historical analysis shows time-trajectories of spawning biomass and recruitment from the reference case model for each assessment conducted over time.

[9] A retrospective analysis involves using the model configuration on which the reference case model for the current assessment is based and removing one, then two, then three, etc. years of data.

definite Hessian matrix can be used to provide an approximate variance-covariance matrix, although the convergence of the MCMC algorithm may be worse in this instance.

A.3. Approximating MCMC sampling

Some of the models used for South African assessments are very complicated; this can make conducting many applications of the MCMC algorithm computationally prohibitive. Thus, it is desirable have an approximation to MCMC sampling available. One such approach would involve:

- Reparameterizing the model so that, for example, all parameters that have to be positive are estimated in log-space, and all parameters that have lower and upper bounds are estimated in logit-space.
- Selecting a threshold difference from the maximum posterior density (MPD – the PD is the product of the likelihood and any priors) that determines when the predictions from a parameter vector are unrealistic. One such threshold would be based on the 1st percentile of a chi-square distribution with degrees of freedom equal to the number of parameters.
- Generating parameter vectors from a multivariate normal (or t) distribution centered on the MPD estimates with the variance-covariance matrix set to that based on inverting the Hessian matrix. Any parameter vectors for which the objective function differs from that corresponding to the MPD by more than the threshold difference would be rejected in this procedure.

One way to use this procedure would be to: (a) construct a posterior using the MCMC algorithm for the reference case model, and an (approximate) posterior using the simpler approach, (b) compare the biomass and mortality trajectories from the two posterior distributions, (c) compare the performance metrics for a small number of OMP variants based on the two posterior distributions, and (d) conduct robustness tests based on the simpler approach and compare the results with those for the reference case operating model based on the simpler approach.

B. Hake

B.1. There have been a number of modifications to the basic assessment over the last 18 months, which have resulted in an appreciably better estimate of status relative to B_{MSY} for the *M. paradoxus* resource (whose status previously – fluctuation to below B_{MSY} - had been a matter of debate and concern) – are these modifications together with their consequent changes in results justified?

Several changes to the assessment have been made and these are all justified. The Panel notes that there are three major causes for the changes to the estimate of ratio of the current biomass to B_{MSY}: (a) correction of how selectivity-at-age is calculated, (b) correction of how catch-at-age is computed when calculating F_{MSY}, and (c) adoption of the natural mortality-at-age vector from the predation model (rather than the earlier somewhat arbitrary vectors). However, several of the other changes to the assessment, including the addition of the most recent data, have also increased the ratio of current biomass to B_{MSY}.

B.2 (*) The revised OMP proposed (OMP-2018) is more "aggressive" than its predecessor OMP-2014 in giving higher TACs for the same abundance (increasing the *b* control parameters in the HCR by 5%, and increasing the cap on the TAC from 150 000 to 160 000 MT. Do the results from the updated Operating Models and simulation tests justify a revision in this direction?

The Panel identified several issues with the reference set of Operating Models (OMs) that could not be fully addressed during the workshop. This made it infeasible to address this question. In particular, in some of the OMs the parameters for the female length-at-age relationship were sensitive to changes in other aspects of the assessment model (e.g. the stock-recruitment relationship) and this sensitivity was sometimes appreciable to the point of affecting convergence, as well as producing unrealistic inferences about female biomass and productivity for *M. paradoxus*. Moreover, at least one trial (RS05b) apparently converged to a local minimum, which may be related to general convergence issues.

There is a need to consider re-parameterizing the model to avoid such unrealistic behaviour. For example, the growth model should be re-parameterized to a 2-parameter linear model, where the parameters are the lengths corresponding to two reference ages for which there are reliable data on length-at-age. In addition, it may be necessary to fix some parameters (e.g. the observation error sigmas) and use analytical solutions for the others (e.g. the catchability coefficients).

The next revision of the assessment should follow a more detailed scheme such as that in Appendix 1, to ensure that certain important aspects are not missed again during checking.

B.3 (*) A particular concern arising for the revised OMP-2018 development has been the possibility of needing in the future to substitute an industry vessel for the standard research vessel (which is now old and experiencing many maintenance problems) to carry out hake abundance estimation surveys, and furthermore the possibility that funding limitations may impact the (regular) continuation of these surveys. A number of robustness tests have been conducted to evaluate the consequences, and the proposed revised OMP-2018 has been considered to have shown adequately robust performance for these. Especially in circumstances where a more "aggressive" OMP has been proposed, which will yield greater TACs than the previous OMP-2014, have the tests conducted been sufficient, and if not what further tests are suggested?

The Panel recommends that future analyses consider robustness trials in which (a) fishery catchability is increasing at a faster rate than 2% per annum, (b) fishery catchability is density-dependent (e.g. CPUE is proportional to the square root of abundance), and (c) there is a failure in recruitment. These robustness trials were selected to more fully test the implications of less frequent surveys on the performance of OMP-2018. Results were presented during the workshop that showed that OMP-2018 is robust to the impacts of (a) and (c), assuming that the reference case analysis is appropriate. However, density-dependent catchability should be examined before final conclusions are drawn.

B.4. A new metarule has been proposed for OMP-2018 which involves the specification of a threshold to indicate when extra measures may be necessary to deal with especially low *M. capensis* abundance. Are this rule and the basis used to develop an initial value for this threshold appropriate?

The proposed meta-rule seems appropriate, but further investigation of the value of such a rule should await finalization of the reference set of OMs.

B.5. What are priority needs (if any) for further robustness tests of OMP-2018? In particular, has adequate attention been paid to the possibilities of recruitment failure (currently surrogated by a decrease in *K* for both species in the future)?

See response to question B.3.

B.6. The assessments generally estimate fairly low values of B_{MSY}/K. These might be argued as leading to acceptance of recovery targets that are too low. Do these "low" values constitute a concern, or a need for alternative higher "targets", give that:

a) **they follow in large part from the stock-recruitment functions estimated (were the forms considered sufficient and appropriate?) (MARAM/IWS/2018/Hake/P3);**
b) **they are arguably a reflection of poor estimates of *K* rather than of current *B* or B_{MSY};**
c) **the hake explicit-predation model (MARAM/IWS/2018/Hake/BG7) indicates that *K* for *M. paradoxus* is "over-estimated" in the standard assessments because these ignore the predation release on this species arising when the fishery commenced concentrating on *M. capensis*;**
d) **for the great majority of Reference Set Operating models, under OMP-2018 both hake species are predicted to "stabilise" at median levels well above their B_{MSY}'s (MARAM/IWS/2018/Hake/P3); and**
e) **for economic reasons the industry needs high CPUE values (which OMP-2018 is projected to provide) (MARAM/IWS/2018/Hake/P4)?**

The Panel did not discuss this issue owing to the concerns with the reference set of OMs.

B.7 (*) Is there a need for a trawl-ID covariate in the GLMM analysis underlying the hake catch species-splitting model used?

The Panel considers that adoption of this approach is currently premature. Prior to making the decision regarding the need for a trawl-ID covariate, the following recommendations and suggestions are offered.

- Implement the approach to model selection outlined by Zuur *et al.* (2009).
- The structure of the binomial GLMM should be examined to determine to what the binomial sample size is set.
- Consideration should be given to a nested random effects structure (Trawl ID within vessel).
- There may be value in moving to a beta or beta-binomial regression framework.
- Calculate the predicted values by integrating over the random effects.

B.8 (H) Other recommendations

It is still not possible to obtain a positive definite Hessian matrix for the hake assessment. Efforts should be made to do so because this will increase confidence that the parameter estimates correspond to the true minimum of the objective function. Convergence of the minimization procedure would likely be enhanced by implementing the "hybrid" method for calculating fishing mortality rates.

C. Sardine

C.1 (*) Is the method followed to estimate a fixed σ_R's to apply in sardine projections for OMP testing appropriate?

The analysts provided an approach (MARAM/IWS/2018/Sardine/P3) aimed mainly to address the issue of how to mimic the observed distribution of recruitments. There are several ways to address this issue. The Panel recommends that the following empirical approach be adopted (for each replicate):

1. Calculate the log-deviations, ε_y, about the fitted stock-recruitment relationship.
2. Convert these to residuals by adjusting them for auto-correlation (with the extent of auto-correlation, ρ, replicate–specific, unless this leads to difficulties), $\eta_y = (\varepsilon_y - \rho\varepsilon_{y-1}) / \sqrt{1-\rho^2}$.
3. In future, project the log-deviations forward by sampling a residual, η_y^*, from step #2 and then computing the future log-deviation, i.e.: $\varepsilon_y = \rho\varepsilon_{y-1} + \sqrt{1-\rho^2}\eta_y^*$.

C.2. How might one best check whether use of the variance-covariance matrix from the Hessian to reflect stock assessment uncertainty is an acceptable alternative to the Bayesian sampling approach to develop joint-distributions for parameters in question for OMP testing.
See discussion under Section A.3 above.

C.3 (*) Is the general approach used in (MARAM/IWS/2018/Sardine/P4) appropriate for attempting to determine the reasons underlying different operating models (for sardine) indicating different levels of harvest intensity to correspond to the same level of risk (as expressed by leftward shift)? How would one best apply the approach further to uncover the underlying mechanism(s) causing such differences?
The general approach in MARAM/IWS/2018/Sardine/P4 is appropriate, and suggests that it is some aspect related to the sample of parameter values from the posterior distribution that is leading to the effect observed. This can be investigated further by studying how risk changes with different hinge points for the stock-recruitment relationship.

This investigation also highlighted that risk measures will vary among Operating Models (OMs)/trials (previous OMP evaluations for sardine had few trials so this was less of an issue in the past) and leads to several recommendations:

- The risk measure should be based on biomass values for "the current regime". Care should be taken to apply objective methods for defining "the current regime".
- The value for the β parameter should be calculated for one OM (as recommended by IWS 2017) or over a weighted set of OMs, but naturally this requires that the set of OMs and their weights be selected.
- Final selection of an OMP should consider all the performance statistics (and not, for example, just average catch).
- The values for risk measures should be considered in the context of the risk under zero catch.
- The approach should be based on setting the specifications for computing the "leftward shift" for OMP-14 in terms of the biomass component measure available from the application of the OM in use at that time that most closely corresponds to that appropriate for the current OMs (e.g. basing it on changes to the distribution for spawning rather than total biomass).
- It is better to consider absolute rather than relative changes in risk (i.e. 0.005 to 0.1 is an increase of 0.095 rather than an increase by a factor of 20), as the latter is very sensitive to the initial risk chosen.

C.4 (*) When risk is to be related to wishing to avoid dropping below a certain level of abundance, how is that risk best measured in a way that is readily interpreted, and also shows appreciable differences when the management controls are changed substantially?
Appendix 2 outlines the thresholds used in management of other small pelagic species.

MARAM/IWS/Sardine/P5 presented results for four variants of OMP CMP3 for three trials. The variants differ in terms of how the OMP reacts to values for the proportion of the TAC taken on west coast. The performances of the variants differ in terms of catch variability, but minimally in terms of the probability of west component biomass falling below 150 000t for the base case trial. The differences in the risk measure for CMP3 among all the trials are very small (maximum 1 percentage point) and the Panel would consider such differences insufficiently large that they should be the focus for much attention. Sensitivity of risk measures to changes in management controls depends on the structure of the dynamic models and stochastic processes. Lack of sensitivity should be investigated to better understand possible hidden feedbacks within the closed-loop simulation that offset impacts of the control. Further, it should be confirmed that the closed-loop simulation model structure adequately represents the risk (e.g. that deviations from the OMP have appropriate structure to impact outcomes if warranted). The Panel also recommends that:

- Correlation in the residuals about the relationship between the proportion of the TAC taken from the west coast and the ratio of the west component biomass to the total TAC should be examined. If this is found to be statistically significant, it should be included in trials (although analyses presented to the Panel suggest that this is not the case (Figure 1)), Following that, robustness trials in which there is correlation in the residuals should be developed to better capture the observed behaviour of the industry and therefore the consequences of the industry not allocating their effort spatially as in the past.
- A performance metric that is the proportion of consecutive x years that the spawning biomass of the west component is less than, say, the 2007 level may better quantify the impact of short-term changes to the total TAC on the west coast component.

D. West Coast Rock Lobster

D.1 (*) Have the Panel's 2017 recommendations about the analysis of the compliance data to estimate poaching trends been appropriately addressed?

The Panel's evaluation of the analyses in MARAM/IWS/WCRL/P1 focused on whether the compliance data had been re-analysed as recommended during IWS 2017 and not (as requested) on how the poaching scenarios used in assessments and projections were developed. The revised analysis was based on a small 'new' database that includes data on confiscations and the associated effort type, which hence addresses the recommendation from IWS 2017. The analysis in MARAM/IWS/WCRL/P1 of the 'new' database led to fairly imprecise estimates owing to low sample size. An alternative analysis that uses both datasets simultaneously (and propagates uncertainty of the estimates of effort type efficiency factors better) would be:

$$C_{y,m,t}^{new} \sim E_{y,m,t}^{new} \text{Negative\_binomial}(\exp(\mu^{new} + \alpha_m + \delta_y + \beta_t), \delta)$$
$$C_{y,m}^{old} \sim \text{Negative\_binomial}(\sum_t Q_t e^{\beta_t} E_{y,m,t}^{old} \exp(\mu^{old} + \alpha_m + \delta_y), \delta)$$

where Q_t is the factor to account for the fact that the new database does not include zero catches (pre-specified based on Table 1 of MARAM/IWS/WCRL/P1). One disadvantage of this approach is that the information in the 'new' database is included in the old database. However, the overlap is likely small given the low number of observations in the new database, and this disadvantage is likely outweighted by the benefits of estimating common month and year factors, and propagating the uncertainty in the estimates of efficiency factors.

D.2 (*) Have updated assessments of the resource been carried out appropriately?

There is only one major change to the assessment methodology since the 2010 assessment, which is the estimation of the recruitment for 2010. However, many of the data sources have been updated based on recent data. The Panel did not detect any reasons that the updated assessment could not be used for management purposes.

The Panel strongly recommends that the assessment model be re-coded in ADMB to facilitate easier model diagnostic analyses such as assessing parameter sensitivity via the gradient, likelihood profiling, parameter correlation and possibly Bayesian MCMC simulation, once model structure and parameterization is adequate to provide convergence.

D.3 (*) How could these assessments be improved, with prioritised suggestions related to data and to analysis methodology?

1. The Panel notes with concern that no length-frequency data have been collected from the fishery catches since 2008. These data are important to allow recruitments to be estimated more precisely. The Panel recommends that DAFF scientists and managers work to re-initiate collection of these data. In other regions of the world (e.g. New Zealand) length-frequency data are collected at low cost by industry with some observer validation. A similar approach in South Africa would enhance future assessments and should be discussed.
2. Possible use of a stock-recruitment relationship relies on the estimation of egg production. Calculation of egg production should account for changes over time in maturity-at-length. If there is to be an increased focus on egg production, additional tagging to estimate recent somatic growth rates for females should take place.
3. The fits to the data should be annotated with measures of uncertainty and bubble plots provided for the fits to the composition data to help assess whether the model fits the data adequately.
4. The assessment assumes that somatic growth prior to 1968 equals the average from 1968 onwards. This assumption may have been appropriate when the assessment was first developed, but should be reviewed by the DAFF WCRL SWG, and the alternative assumption that pre-1968 somatic growth equals the average over some of the first years for which estimates of somatic growth are available examined.
5. The assessment is coded in FORTRAN, which means that exploration of scenarios is slow and measures of uncertainty cannot be produced easily. The model should be recoded in ADMB – perhaps with a prior on the recruitment in 1910 (unless recruitment is re-parameterized as per below).
6. The assessment does not fit some of the average (over time) length-composition data sets, especially those for females. Alternative selectivity assumptions should be explored to assess whether the fits can be improved. However, care should be taken to avoid overly and unrealistically complex selectivity patterns (a recommendation from IWS 2010).
7. Explore whether the data currently available are sufficient to construct models that split some of the super-areas into inshore and offshore components. Such models will likely have to estimate movement between the two components, which will require tagging data.
8. Inclusion of environmental covariates in the GLM may explain some of the variance in catch rates, but such inclusions often do not change the results and should only be considered if there is there is a plausible biological mechanism for the relationship between the covariate and catchability. There may be value in including depth as a covariate in the GLM (assuming the necessary data are collected).
9. Conduct sensitivity tests in which catch-rate is assumed to be non-linearly related to abundance.

D.4 (*) Have updated resource projections under alternative future levels of catch (both legal and illegal/poaching) been carried out appropriately and how might such projections be best improved methodologically in future, including in particular taking account of stock-recruitment effects?

Key assumptions underlying the methodology used for future projections are that (a) somatic growth for males is the average of the estimates for 1989-2017, and (b) future recruitment is equal to the simple geometric mean from 1975-2010. Although assuming constant recruitment is common in some circumstances where resource abundance is relatively high, there is concern that projected recruitment and abundance will not adequately reflect impacts of fishing, especially under a constant TAC plus poaching on the very low abundance stocks in all the WCRL super areas. Figures 3 and 4 plot somatic growth rates and recruitment spatially, and all of these are clearly suggestive of variation among years.

Future projections should consider uncertainty in recruitment, somatic growth, and resource abundance.

- Recruitment processes are poorly understood for many invertebrate stocks, mainly because of uncertainty about spatial and reproductive dynamics (e.g. role of males, male-female ratios). In the absence of clear hypotheses for reproductive dynamics, the Panel recommends two parsimonious assumptions about future recruitment: (1) sampling from recruitment values – this approach makes the fewest assumptions about the form and distribution of recruitment, and simply relies on the best-estimated historical pattern though at the cost of assuming that future recruitment is independent of abundance, and (2) sampling recent recruitment rates (i.e., recruits/egg production) from the stock assessment – this approach would additionally create feedback between egg production (and hence link to legal and illegal catch) and future recruitment. As in (1), process error could be bootstrapped from residuals about the mean recruitment rate.
- The recruitment estimates for 2007 and 2010 each had a penalty to constrain their values and prevent their estimates being unrealistically large or small. The penalties should be calculated without including the recruitment values for 2007 and 2010. Projections of recruitment, which are currently based on the geometric mean, need to be adjusted to better reflect the production from the resource which is the most likely in the future. There are various ways that merit consideration to do this, which the DAFF WCRL SWG will need to evaluate: for example, bias-correct the geometric mean assuming log-normality; use an arithmetic mean; use bootstrap samples of the empirical distribution of recruitment values in the projections, although these would best be set up as recruitment residuals (deviates) from some measure of central tendency ($\bar{R}$) (see below).
- The estimation of $\bar{R}$ would permit the generation of a likelihood profile on $\bar{R}$ as a means of characterizing uncertainty around recent recruitment.
- Somatic growth has changed over time, which leads to two scenarios: (1) somatic growth remains low, and (2) somatic growth increases back to historical levels. Projections should be conducted for both scenarios, with weights assigned by the DAFF WCRL SWG.
- Previously, uncertainty in historical and projected resource abundance was dependent on uncertainty in R_{2010}, which, unfortunately, does not fully quantify uncertainty. Another way to represent resource uncertainty is to re-parameterize recruitment (after 1975) as multiplicative deviations about a mean level ($\bar{R}$). Projections could then be conducted by repeatedly selecting a value of $\bar{R}$ based on its probability from a likelihood profile. A range of options for conducting projections from these $\bar{R}$ values should be considered by the DAFF WCRL SWG, including basing these on the full set of parameter estimates corresponding to each selected $\bar{R}$ and resampling from the associated recruitment deviations.
- The results of projections should be shown for each scenario regarding somatic growth and not for the weighted result alone. In addition to showing projection envelopes, the probability of dropping below the current and the 2006 biomass should be reported to include in the basis for selecting future TACs.

The stock-recruitment plots in MARAM/IWS/WCRL/P5 were hard to interpret because (a) no account was taken of time-varying maturation schedules, and (b) the data points do not exhibit compensatory dynamics, which preclude fitting standard (e.g. hockey-stick) stock-recruitment relationships. Some of the projections lead to reductions in biomass below previously observed levels. It is plausible that recruitment would be expected to decline if this occurs, but the current data provide little quantitative basis for specifying expected recruitment at egg production levels lower than have been observed historically. Given the current stock status, any decline in egg production should be considered to reflect a large increase in risk to the resource.

D.5 In the current situation, what schedule would be appropriate to return to an OMP basis for management recommendations, and how should future data/indices on poaching best be treated in this process?

The Panel emphasizes that OMP-based management remains the preferable way to provide scientific management advice for marine renewable resources. However, the Panel does not recommend that an OMP development process be re-initiated until there is more clarity regarding the future management structure (e.g. the allocation among sectors) and also future data collection strategies. Given the likely changes in allocation, the Panel strongly recommends that the small-scale sector be monitored at least for both total (landed) catch, discards and catch-rate.

E. Linefish

E.1a (*) Comment on the *Multivariate Normal (MVN)* prior approach to account for the non-independence of key JABBA-Select input parameters, H_{MSY} and the shape parameter *m*, addressing IWS 2017

The newly implemented approach using the MVN addresses a concern raised in last year's IWS. This approach allows for correlations between log($H_{MSY,S1}$) and log(*m*), but ignores any correlation between these parameters and $H_{MSY,S2}/H_{MSY,S1}$, $H_{MSY,S3}/H_{MSY,S1}$ etc. The correlation between these parameters and log($H_{MSY,S1}$) and log(*m*) should be explored and if found to be appreciable, log(*m*), log($H_{MSY,S1}$) and all the log($H_{MSY,Sx}/H_{MSY,S1}$) parameters should be generated from a multivariate normal distribution. This request was examined in a plot provided by the analyst (Figure 5), which illustrated at least some other correlations. Use of these correlations should be included in the MVN to account for this non-independence.

E.1b. Advise on the robustness of the currently used deterministic, external approximation of the non-linear relationship between exploitable biomass (EB) and spawning biomass (SB). If possible, suggest potential options that could be explored to allow updating the functional form of EB/SB within the JABBA-Select model.

The fits of the relationship between SB and EB are good, with the only potential concern being that they are based on the means of the priors for *M* and *h*. The adequacy of the assumed relationship was tested as part of the simulations in MARAM/IWS/2018/Linefish/P2.

E.1c (*) Broadly comment on the completeness of the revised documentation of the JABBA-Select model provided, addressing IWS 2017.

The documentation of JABBA-Select (MARAM/IWS/2018/Linefish/P2) is much improved compared with the document provided to IWS 2017. JABBA-Select continues to be developed further, and it is important for the documentation to continue to be updated regularly. The Panel applauds the continued open source development and publicly accessible code repository on GitHub. Linking to the repository with current features (https://github.com/Henning-Winker/JABBAbeta) would be helpful in the updated documentation. Ultimately, there will be benefit in developing a user manual and/or further vignettes.

E.2. Evaluate the suitability of the revised simulation framework in terms of comparability between modelling frameworks and performance metrics, addressing IWS 2017.

The simulations in MARAM/IWS/2018/Linefish/P2 compare the estimation performance of a Bayesian state-space Schaefer model, JABBA-Select, a deterministic age-structured production model and a stochastic age-structured production model for an informative scenario, as well as of scenarios in which *h* and *M* are not equal to the modes of their priors. Results were shown during the workshop for a scenario in which biomass declines essentially continuously during the historical period. While additional simulations can be conducted, the analyses in MARAM/IWS/2018/Linefish/P2 suggest that the performance of JABBA-Select is at least as good as that of the alternative models, and has the best coverage probability of the methods evaluated.

E.3a. (*) Comment on the revised the posterior and prior diagnostic plots and metrics in MARAM/IWS/2018/Linefish/P2 and provide guidance regarding correct interpretation.

Several diagnostics and performance metrics are now available for JABBA-Select, including measures of whether the MCMC algorithm has converged, whether the model fits the data adequately, retrospective analyses and cross-validation of prediction ability. The existing diagnostics can be complemented by (a) reporting trace plots for multiple chains and the objective function, (b) reporting posterior predictive checks, and (c) automatic application of the runs test to the model fits. The intervals used to evaluate predictive skill should be based on observation error as well as process error (i.e. the intervals should be posterior predictive intervals). The Posterior to Prior ratio of variance calculation needs checking (currently this is based on the CV; this should be changed to the variance).

E.3.b (*) The stochastic age-structured simulator, which was suggested as a diagnostic tool for process error variance estimate, infers the process error from an unfished state. However, the simulation results in MARAM/IWS/2018/Linefish/P2 indicate that the fisheries-induced truncation of the population is likely to inflate variation in the log-biomass variation at low abundance. Provide directions for future research towards generalizing this approach to inform process error prior parameterization and process error diagnostic tests.

The simulations in MARAM/IWS/2018/Linefish/P2 confirm that the extent of process error in biomass depends on natural mortality rate, the age-at-maturity and the extent of recruitment variability. In addition, the extent of inter-annual variation in biomass (and its temporal autocorrelation) depends on the depletion of the population. The values for biomass process error can be used to guide the priors for process error in JABBA-Select. JABBA-Select can be extended by developing a relationship (similar to that between exploitable biomass and spawning biomass) between the extent of biomass process error and depletion.

E.4 Other recommendations

E.4.1 (H) JABBA-Select is based on pre-specified selectivity parameters. Selectivity is often not easy to estimate. Applications of JABBA-Select should document the parametric forms for selectivity and the associated parameter values, as well as the basis for the forms and values.

D.4.3 (H) Conduct simulations to explore the performance of JABBA-Select when selectivity is dome-shaped (e.g. double-normal, double-logistic, gamma) rather than asymptotic. Depending on the results of the simulations, it may be necessary to modify JABBA-Select to account for alternative selectivity patterns.

E.4.3 (H) The estimated process errors exhibit auto-correlation over recent years, but this may reflect incorrect choices for priors. This possibility should be examined by considering sensitivity to the assumed priors. Alternatively, auto-correlation which may reflect model mis-specification or the incorrect attribution of some dynamic process, may be occurring from 2000 – 2015, but is not explicitly included in the model.

E.4.4 (M) Assuming that catchability (q) is constant over time can lead to bias when selectivity is time-varying. Consider extending JABBA-Select to include the approaches developed for removing this bias (e.g. Anon., 2009).

F. References

Anon. 2009. Report of the Operating Model and Management Procedure Technical Meeting, 13 - 17 July 2009. Unpublished CCSBT Report. 64pp.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. and Smith, G.M. 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer 549pp.

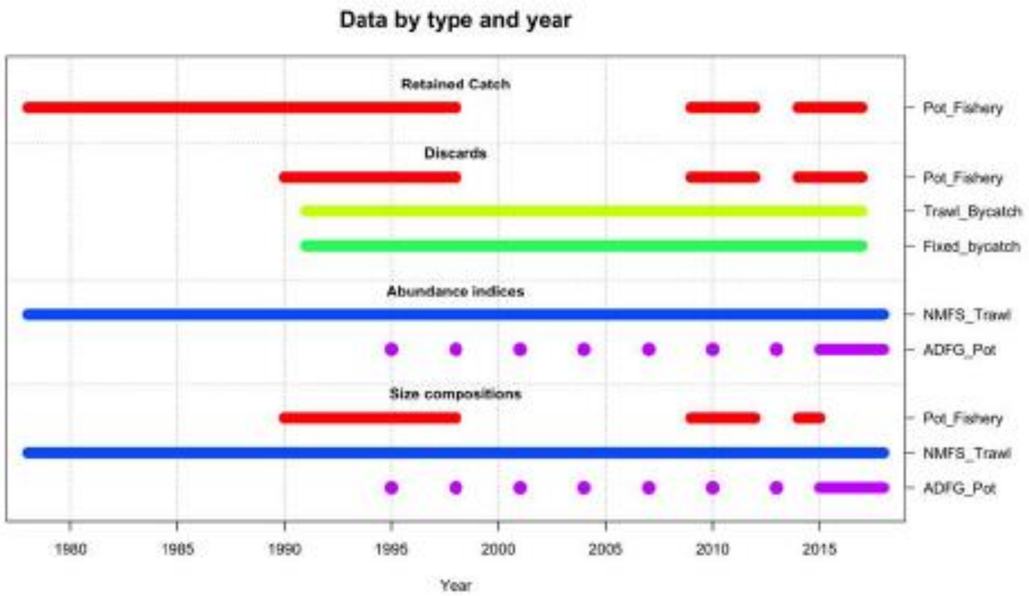

Figure 1. Example of a display of data availability by year and data source (for St Matthews Island blue king crab).

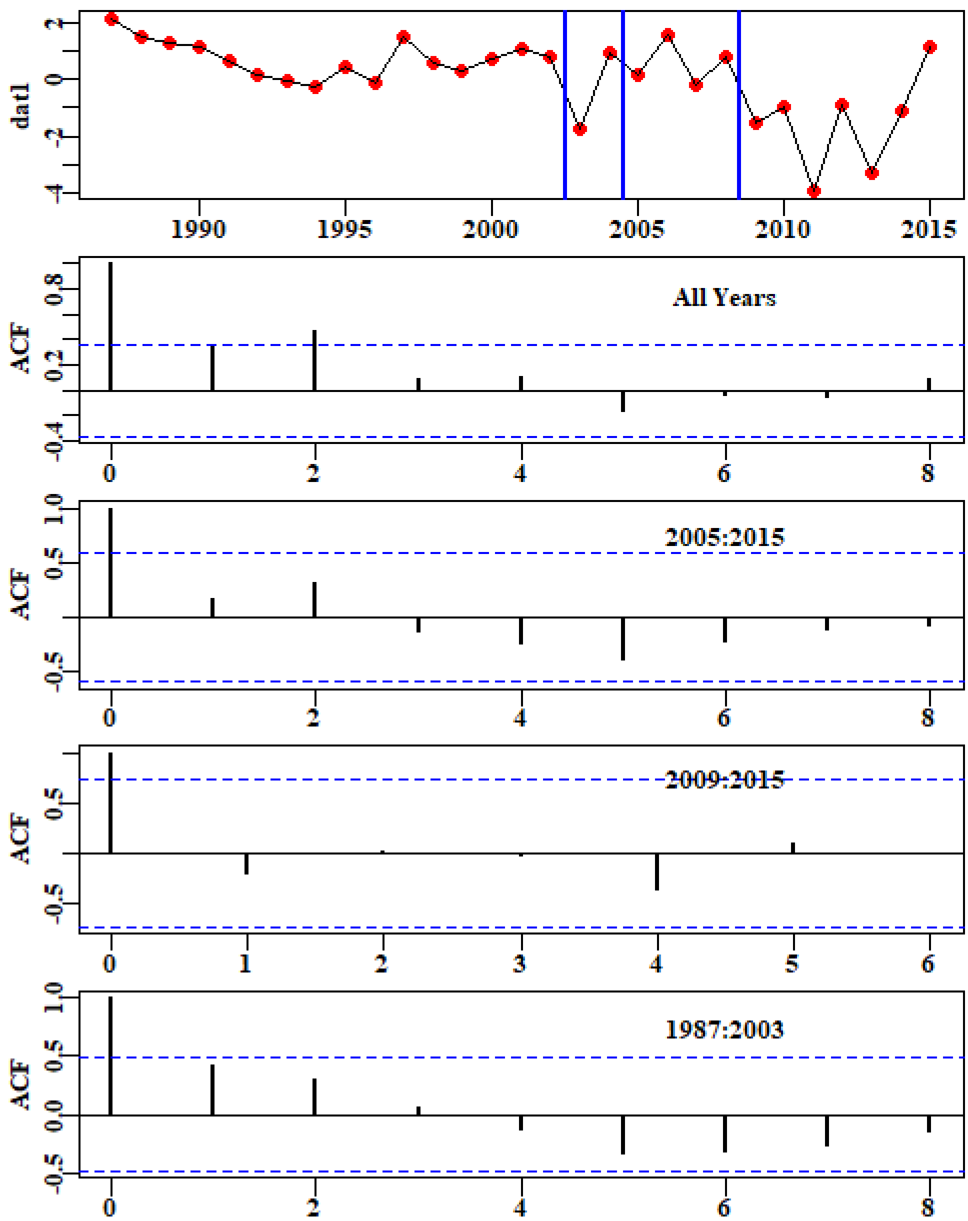

Figure 2. Residuals about the relationship between the proportion of the sardine TAC taken from the west coast and the ratio of survey estimate of spawning biomass (upper panel), and autocorrelations for various time periods (lower four panels). The dashed lines indicate 95% confidence intervals inside of which there is no significant correlation. The only significant auto-correlation had a period of 2 years when all 29 data points were considered.

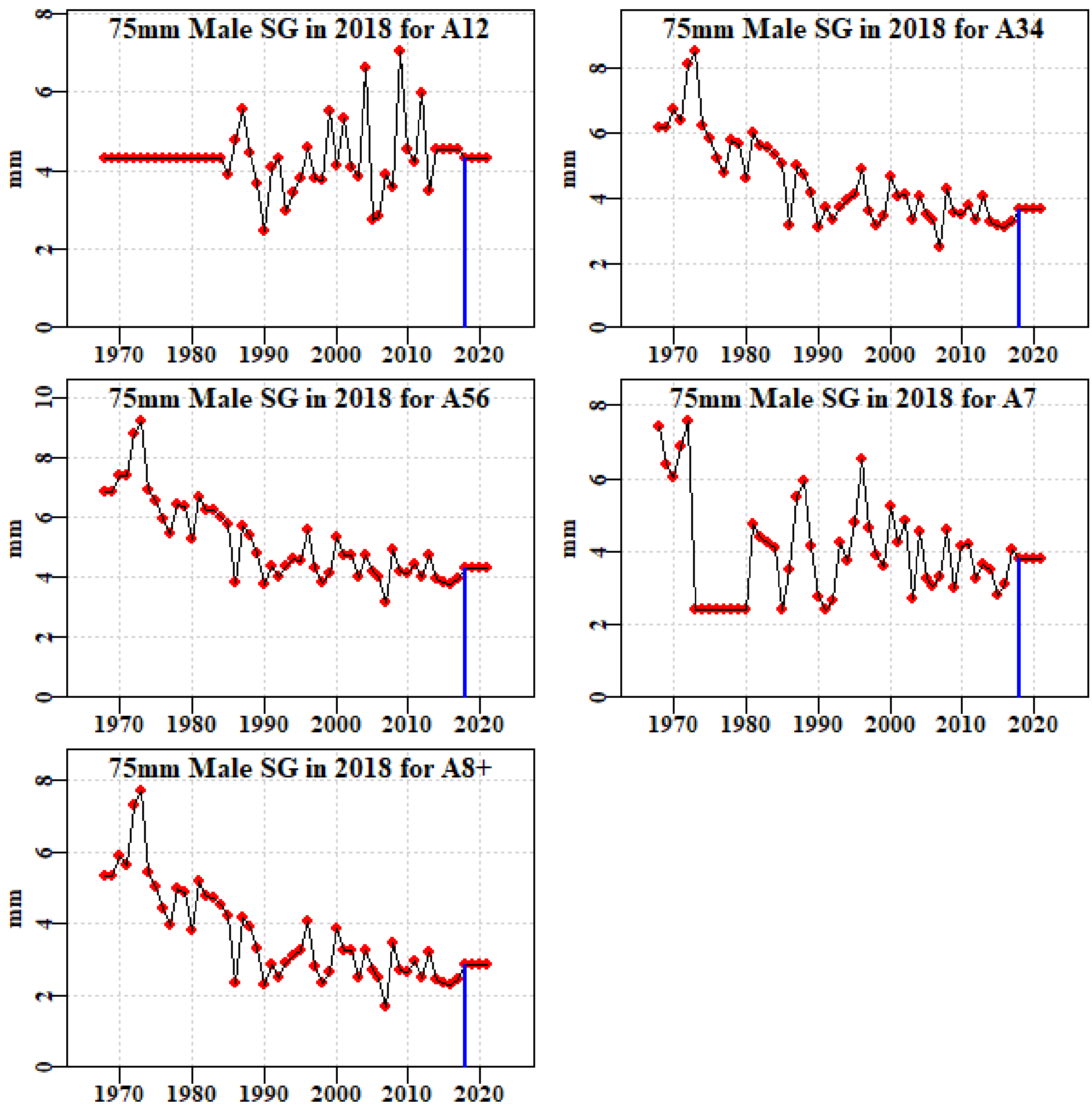

Figure 3. West Coast rock lobster male somatic growth rates for each super-area, with values assumed for projections shown after the arrows.

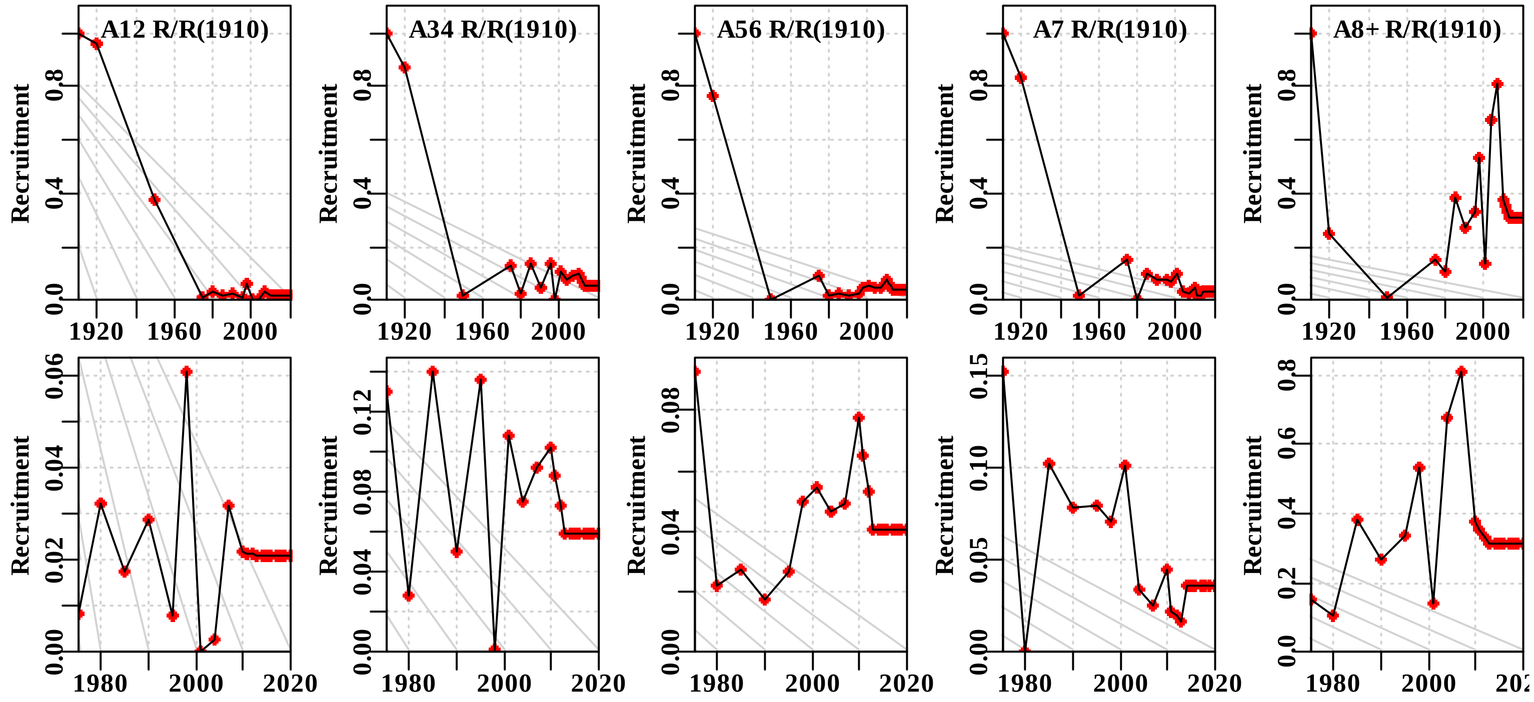

Figure 4: West coast rock lobster: estimated recruitment trends for each super-area; values for 2011+ are as assumed for projections.

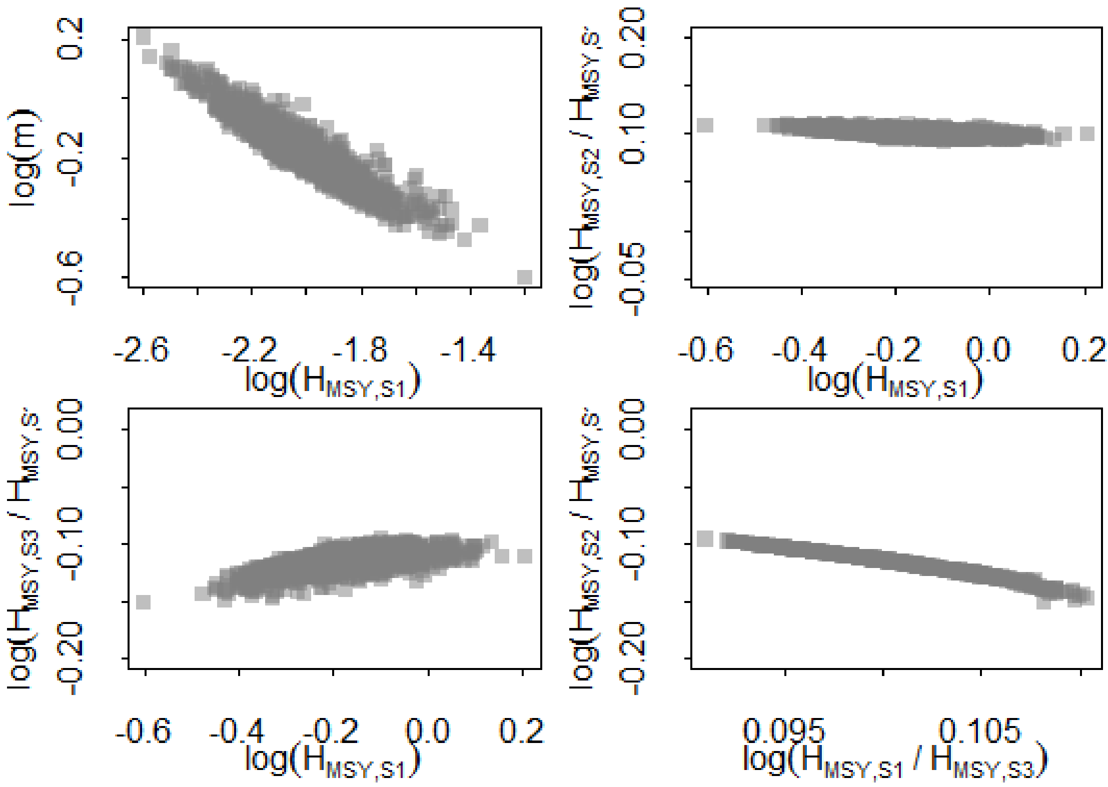

Figure 5. Correlations among model parameters based on JABBA-Select.

Appendix 1: General process for assessment review

Terms of Reference to set the stage for assessments and ultimately peer review:

- Review input data (automated graphs <u>and</u> tables)
  - Estimate catch from all sources:
    - Fishery (commercial, recreational), bycatch/discard, catch at age/length composition, …
    - Uncertainty/bias in each source
  - Review survey data
    - Multiple surveys, begin and end of each series, age-length data, …
    - Fishery CPUE series standardization and methods
    - Uncertainty/bias in each source
  - Data inclusion criteria (representativeness; we want but don't have this yet, suggested)
    - Spatial coverage for entire stock
    - Temporal coverage (minimum series length? current/maintained?)
    - Statistical design for surveys
- Review life history, other biology, ecosystem conditions
  - Growth, mortality, maturity, spatial distribution, with plots for gut checks
    - What are sources of these? Updates?
    - Change over time?
  - Ecological or other factors affecting productivity/recruitment
    - Habitat requirements/impacts
    - Species interactions
    - Change over time? To determine appropriate risk levels, etc.
- Model review: F, recruitment, B, estimated appropriately, with uncertainty
  - Bridge from previous assessment to starting point for new model(s)
    - Review previous assessment, major assumptions
    - Corrections? (code, previous input data)→changes in F, recruitment, B
    - Impact of previous data updates to current→changes in F, recruitment, B
    - Impact of new/candidate datasets→changes in F, recruitment, B
    - Parameter/structure changes as a result of above, done as below
  - Convergence? Iterate to achieve:
    - Parameters at bounds? Iteratively fix, see what else breaks
    - Likelihood profiles for key parameters (h, M typical)
      - Is there a minimum or is it flat? (if flat, no info on parameter)
      - How do F, rec, B, etc. change over flat range (how influential)
    - Ability to estimate parameters determines internal SR vs Rbar w/devs
  - Correlations between parameters? Assumes converged
  - Biological realism of estimated parameters: Length at age, Maturity; time varying? Standard plots
  - Fits to data series and compositions:
    - Patterns in residuals? Standard plots from automated software
    - Influence: Change in fit leaving out series one and keeping the rest
  - Retrospective pattern? F, recruitment, B. How big and what direction? Standard plots
  - Support for other estimated parameters: external evidence? How sensitive are results to changes? Suggests sensitivity runs/robustness/alternative OMs

        - Selectivities (e.g. if domed, what is rationale for invisible big fish? defend choice)
        - Catchabilities (e.g. explain >1)
        - Where is maturity curve relative to selectivity?
    - Clear, documented rationale for selected model at the end
- Reference point review (after model(s) are considered robust enough)
    - Define method: from SR or proxy, depends on selected model
        - Which SRs to keep? Weighting?
        - Which proxy to use? Biological realism, meta-analysis, laws
    - What recruitment is/not included to define stock productivity
        - Exclude/downweight if evidence productivity regime not relevant
        - Exclude/downweight poorly estimated (too recent/CV threshold)
- Status of stock (current B and F relative to reference from models)
    - Done **after** justifying methods for all of the above
    - Compare with simple (non-model based) indicators of current stock status
        - Evidence for any estimated recent low/high recruitment in catch at age, survey indices?
        - Trends in surveys similar to model, or explain why different?
- Projections (OMP specific)
- Research recommendations

Appendix 2: Thresholds used in the management of some other small pelagic species

	Pacific Sardine	**Pacific Herring**
Directed Fishery closed	150,000t	0.1 B_{MSY}
Overfished (and in need for a rebuilding plan)	50,000t	0.5 B_{MSY}
Key risk measure in the MSE	P(B_{1+} < 400,000t)	P($B < 0.5B_{MSY}$)