INTERNATIONAL REVIEW PANEL REPORT FOR THE 2017 INTERNATIONAL FISHERIES STOCK ASSESSMENT WORKSHOP
27 November – 1 December 2017, UCT

[Sean Cox[1], Daniel Howell[2], André E Punt[3]]

Introduction

The Panel recognised the very high quality of the research presented at the 2017 International Fisheries Stock Assessment Review Workshop. This included research on Southern African hake, sardine and rock lobster, as well as methods for conducting assessments for relatively data-poor species with a focus on linefish. 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 programmes reviewed, and then focuses on addressing the questions posed by DAFF Research, a more detailed technical review and finally recommendations for further work concerning each. 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.

Summary of general issues

The Panel was pleased with the "fishery description" documents for each species that were provided prior to the meeting. This helped Panel members unfamiliar with South African fisheries and fisheries management techniques to prepare better for the review. The Panel was again provided with a large number of documents that mainly became available only shortly before the workshop. However, the Panel appreciated that the documents were well linked to questions that the Panel was asked to address, and that there were fewer primary documents that were not referenced during the workshop than during previous reviews.

Hake

The Panel reviewed the technical basis for revising the OMP used to set TACs for the hake fishery. The development of an OMP for hake is particularly complicated because of the two-species (*M. capensis* and *M. paradoxus*) nature of the hake resource off South Africa. The Panel made various recommendations related to the operating models under consideration, the choice of the Reference Set of operating models, the Robustness Tests, the OMP variants to consider and how trials within the Reference Set might be weighted.

The Panel noted previous Panel conclusions that models for *M. capensis* and *M. paradoxus* resources should consider hake in Namibia as well as South Africa. Unfortunately, this was again infeasible owing to a lack of data-sharing by Namibia. The Panel once again **strongly recommends** that efforts be made to allow assessment analysts to have access to all hake-related data from the entire southern African region to create opportunities for progress on models that use all of the available information.

Sardine

The Panel identified an approach for specifying "risk" for the revision process to develop the joint anchovy and sardine OMP-18. This involves selecting a threshold level of spawning biomass and an acceptable probability of the resource dropping below this level. This "risk criterion" cannot be implemented immediately as additional calculations need to be

[1] School of Resource and Environmental Management, Simon Fraser University, Canada
[2] Institute of Marine Research, Norway
[3] School of Aquatic and Fishery Sciences, University of Washington, USA

conducted and then discussed by the Small Pelagic OMP Task Team and the Small Pelagic Scientific Working Group.

In relation to whether there is a need for spatial management, the Panel **recommends** simulations in which candidate MP variants are tested when the split of the TAC to the south and west coasts follows a model based on past fishing fleet behaviour. The results of these simulations should be used to determine whether the recommended risk criterion is satisfied or not under the default scenario of no explicit spatial management. If the risk criterion is not satisfied by all of the initial set of candidate MPs tested, calculations should be undertaken to identify MP variants for which (a) there is no explicit spatial management, but TACs are sufficiently low that risk remains acceptable, or (b) there is some spatial management to avoid risk becoming unacceptable. The final selection between candidate MPs that satisfy the risk criterion should balance flexibility (i.e. fewer restrictions on fishing operations) with average catches and other measures of performance, such as the fraction of years with a severely reduced fishery.

Given the current uncertainty regarding whether or not there is a need for spatial management, but also initial calculations suggesting that spatial management may be needed, it would be consistent with the precautionary approach to implement some form of spatial management until OMP-18 is finalised and adopted.

West Coast Rock Lobster

The West Coast Rock Lobster resource is subject to large levels of illegal catch (poaching), which has increased substantially recently and may continue to do so. Illegal catch impacts the sustainability of the fishery, and while the Panel discussion focused on estimation of historical illegal catch and methods for projecting future illegal catch, **the existence of illegal catch at the levels reported to the Panel is of very considerable concern for the future of this valuable resource**. The Panel reviewed the methodology, and agreed that confiscation numbers adjusted for enforcement effort was an appropriate basis to infer trends in illegal catch, although the relationship is inexact. Furthermore, the TRAFFIC data provide a basis for estimating the minimum level of illegally caught West Coast Rock Lobster that are exported from South Africa.

The Panel identified improvements to the estimation methodology that should be implemented before further conclusions regarding trends in poaching rates can be drawn. The Panel also provided suggestions for future developments related to monitoring poaching levels. The Panel **strongly recommends** that effort be made to link confiscation data to the type of enforcement that resulted in the confiscations, because the current lack of this linkage is a major source of uncertainty in estimates of trends in the extent of poaching. The Panel was encouraged by the involvement of DAFF's Compliance Section in West Coast Rock Lobster Scientific Working Group, and **strongly endorses** their continued involvement given that the most appropriate interpretation of the current data requires use of not only the quantitative, but also the qualitative information that they are able to provide.

The Panel also **recommends** that a **high priority** be given to developing a way to quantify the magnitude (and ideally trend) of illegal catches that are sold domestically.

South Coast Rock Lobster

The Panel concluded that there is no evidence to suggest an urgent need to revise the current OMP for South Coast Rock Lobster, *inter alia* as most catch-rate values since OMP-2014 were implemented have been within the expected range. The Panel **recommends** further work that should be taken into account in the next review of the OMP, which accordingly is better delayed until this work has been conducted.

Linefish
The Panel reviewed a new approach for conducting stock assessments for species for which the primary data source is fishery catch-per-unit-effort. The Panel endorsed the method in principle for general use in data-moderate situations, and provided recommendations for how the method could be improved, advised specific tests which need to be undertaken to determine its suitability for a particular stock, and suggested how it can be compared with alternative approaches used for stocks off South Africa.

General considerations
Plots of the difference between the spawning biomass trajectory in the absence of catch (the so-called "dynamic B0" trajectory) and that given the actual historical catch that occurred, both of which are based on the parameters estimated during an assessment (e.g. MARAM/IWS/2017/Sardine/P8), provide a way to distinguish between the effects of catches and the environment. These types of plots should be reported routinely in South African stock assessments. The information that such plots provide on the consequences of catches would be easier to assess if, in addition, the ratio of spawning biomass with catch to that in the absence of catch was plotted against spawning biomass in the absence of catch.

A. Hake

A.1 (*). Is the new selectivity model adequate/appropriate

The current formulation of selectivity should be used in the Reference Set of models. However, robustness tests should be developed to examine (a) the consequences of fishery selectivity changing in the future, and (b) an operating model in which at least one fleet or one survey has a near asymptotic selectivity pattern for *M. capensis*. The Panel does not, however, expect that OMP performance will be very sensitive to these factors because (a) the proposed OMPs use only abundance index data and not fishery age- or length-composition data, and (b) the selectivity for several of the surveys is close to asymptotic already.

The extent of the 'cryptic biomass' modelled should be reported for each survey.

A.2. Are the stock-recruitment models used adequate / appropriate, including the extent of annual variation abundance about these relationships?

The Panel explored evidence that the stock-recruitment relationship for *M. capensis* is dome-shaped, and hence that reducing spawning biomass will increase expected recruitment from current levels, and agreed that the spawning biomass depletion and subsequent recovery rate implied by the historical catch is sufficient support for a dome-shaped relationship over an asymptotic one. The Panel noted that the steepness parameter was hitting its upper bound of 1.5. However, increasing this bound led to an estimate of steepness that did not differ much from 1.5.

A.3 (*). Could the shrinkage procedure used for estimation of recent recruitments be improved?

The procedure could be improved when conducting projections by generating values for the recruitments that are currently shrunk to better capture uncertainty (i.e., implicitly using the full σ_R, similar to the approach applied for South Coast Rock Lobster). However, this is not a high priority unless the species is assessed to be highly depleted. The impact of recent recruitment estimates on the evaluation of an OMP for hake will be limited because performance statistics are based on multi-decade projections.

A.4. Currently OMP testing projections are initiated from MPLE estimates. How might taking estimation uncertainty into account best be achieved?

The Reference Set of trials on which OMP-2014 was based captured a wide range of biomass and depletion values, such that allowing for parameter uncertainty is unlikely to increase the range already considered in projections. Moreover, it is currently not possible to invert the Hessian matrix, which would be needed to initiate a procedure that could be used to generate parameter vectors. As for item A.3, the multi-decadal nature of the forecasts limits the impact of this uncertainty on the results of the OMP testing.

A.5. Was the basis for the previous Reference Set (RS) selection adequate, and if not how should future selections be made?

The Reference Set can be reduced in size by eliminating the factor for natural mortality (see item A.7 for additional details). Consideration should be given to including a scenario in which the model projection starts in 1978 rather than 1917 in the Reference Set, as this scenario would exclude assumptions related to earlier years that are highly uncertain. If this scenario is added, the Reference Set will consist of 12 trials (four ways to handle the uncertainty of the split of the historical catches between *M. paradoxus* and *M. capensis*, and three options for the stock-recruitment relationship).

A.6. Was the previous set of robustness tests adequate?

During its consideration of other questions, the Panel identified additional robustness tests (see items A.1, A.8, A.9, A.12 and A.13).

A.7. Should a change be made to use of *M*-at-age estimates from the hake predation model – an average over last 2-3 decades, or time varying by year since commencement of fishery?

The *M*-at-age vectors in the current Reference Set were selected semi-arbitrarily. In contrast, the estimates of *M*-at-age in MARAM/IWS/2017/Hake/P5 are based on an approach that explicitly accounts for time-varying predation due to hake. The Panel therefore **recommends** that future assessments and operating models be based on the average *M*-at-age by species from this predation model. The average should be based on the years from 1984 to the present, when data to inform year-class strength were available. The predation model is still a work-in-progress, but its use is still preferable to a somewhat arbitrary selection of *M*-at-age values.

A.8 (*). Should a penalty function on the survey *q*'s be included (e.g. to restrict to values below 1).

The Panel notes that several of the estimated survey q's exceed 1 for the Reference Case operating model. This is unexpected because the surveys do not cover the full spatial distribution of *M. paradoxus*. Moreover, the surveys are for one coast only whereas the model predictions apply to the combined abundance on both coasts. Both of these factors would lead one to expect that the estimated survey q values would be below 1. The apparent discrepancy is greatest for the west coast summer survey and the south coast spring survey for *M. paradoxus*. The Panel considered possible reasons for the unexpected high survey q's, including model mis-specification and herding, but there is no clear support for any of the reasons considered. Thus, the Panel **recommends**:

- the operating model should be fitted restricting survey $q < 1$ and examining which data sets are fitted appreciably worse than before, suggesting a conflict between the model with survey $q < 1$ and those data; and

- the model with survey q's restricted in this way should be considered a robustness test.

A.9 (*). Should a penalty function on von Bertalanffy L_∞ values be included?

The parameterization of the von Bertalanffy growth curve should be changed from L_5, $\ln(\kappa)$ and t_0 to L_1, L_5 and $\ln(\kappa)$, which should improve convergence. In addition, a robustness test should be conducted in which a lower bound is imposed on κ.

A.10. How best should (differing?) q values for surveys be estimated given gear changes and sometime use of industry vessels?

The Panel agreed with the analysts that the data for industry vessels should be downweighted to reflect lack of knowledge of survey q (except for the one vessel, *Andromeda*, which had conducted sufficient surveys to allow a separate q to be estimated). In addition, operating model projections should be undertaken in which the survey q for each future year is generated from a distribution that reflects uncertainty regarding possible q's for industry vessels acting as survey vessels.

A.11. Need the ageing error matrices used be reconsidered?

The existing set of ageing error matrices should be adequate for the current OMP revision unless data from a new age-reader are included in the assessment, in which case the ageing error matrices should be updated using all available double-read information.

A.12. For surveys might changing abundance estimation from the current random stratified to a geostatistical approach constitute an improvement?

The CVs of the estimates of abundance from the geostatistical approach tend to be lower than for standard design-based methods, as there is often spatial auto-correlation in density.

The DAFF Working Groups should establish a standard set of diagnostics to examine when reviewing survey results from the geostatistical approach. For example, the causes for large differences in point estimates from the geostatistical and design-based approaches should be understood, the residuals should appear random in time and space, and anisotropy estimates should be consistent with estimates based on simple analyses of the raw data.

In principle, the geostatistical approach should be applied to the survey size-composition data. However, this is not a high priority at present.

Operating models that are fitted to the geostatistical estimates of biomass (to 500m) should be included in the robustness tests. These estimates should be based on analysing data for each year separately, rather than relying on temporal as well as spatial auto-correlation. This is to avoid further complexity in the assessment methodology. Otherwise it will be necessary to modify assessment models to allow for temporal autocorrelation in survey indices by adopting a multivariate likelihood function if abundance estimates are based on the geostatistical approach.

A.13. How best might results from the extension of surveys into deeper water be taken into account?

A robustness test should be developed that includes the data (index and size-composition) from the region from 500-750m as a separate time-series, with the index derived from a design-based analysis of the data. It seems unlikely that the results will be very sensitive to including these data, so conducting this robustness test should be considered low priority if it requires considerable recoding of the operating model.

A.14. Should attempts be made to allow for some explicit movement, either as the basis for an updated RC or as a robustness test?
Implementing the movement model as an operating model requires the development of approaches to allocate future catches spatially. However, this is a substantial exercise and given that the movement model is not final yet, the development of even a robustness test should be deferred to the next time the OMP for hake is reviewed.

A.15 (*). Is there any way of independently checking the *M. paradoxus* / *M. capensis* biomass ratio implied by the assessments?
Evaluation of this question is not straightforward because it is not clear that the *M. paradoxus* /*M. capensis* ratio is implausible given that only spawning stock biomass values are being provided in assessment reports. The Panel offers the following ways to explore this issue:

- Compare estimates of exploitable (or total) biomass in addition to spawning biomass.
- Compare estimates of biomass at size between predators and prey of suitable size.
- Examination of the ratio of catches to biomass over time, in the context of catch-rate and stock trends, could also potentially give information on absolute biomass estimates. However care would need to be taken that appropriate fishable biomasses are used in this analysis.
- Explore the *M. paradoxus* /*M. capensis* ratio using various measures of biomass from the model for hake that includes predation explicitly.
- Compare estimates of density for *M. paradoxus* and *M. capensis* from surveys.
- Evaluate the sensitivity of the *M. paradoxus* /*M. capensis* ratio in models in which catchability is restricted to be less than 1 for all surveys.

A.16 (*). How should the different Reference Set OMs be weighted in reporting performance statistics?
The choice of an OMP can depend on how the trials within the Reference Set are weighted (MARAM/IWS/2017/Hake/P9). Thus, work to evaluate weighting schemes could be highly consequential. However, identification of the best approach for weighting is still an area of research globally, as well as in South Africa. The Panel has the following **recommendations** in regard to weighting of trials within the Reference Set:

- Use of AIC weighting is **not recommended** because this type of weighting relies on assumptions that are unlikely to be valid, such as that all data are independent.
- If Multidimensional Scaling methods such as those in MARAM/IWS/2017/Hake/P9 are adopted, the discrepancy metric should account for the absolute as well as the relative scale of biomass and be based on historical and not future trends.
- Consider methods for model weighting based on predictive performance because this relates most directly to the reliability of projections.

The Panel notes that several groups are exploring ensemble approaches for integrating the results of multiple models (e.g., Robert Thorpe and Mike Spence at CEFAS, UK). The South African analysts should consult with these groups, and the Panel has provided email addresses to begin this cooperation. Overall, it may be prudent to base the selection of OMP-18 on equal weighting, but to nevertheless examine how sensitive the final selection would have been to alternative weighting methods, with a view towards adopting a new weighting method for a future OMP revision.

A.17. Should slope as well as target approaches be used in the OMP's fundamental HCR? Should HCRs that react more rapidly to the most recent data be explored further (this is a particular concern of industry)?
It is ideal to consider a wide range of OMP variants to allow the best OMP to be identified. Slope-based HCRs may be capable of responding more rapidly to recent data. However, such HCRs may "follow noise" and lead to higher TAC variance. The lag between data being collected and changes in TAC being applied may also lead to poor OMP performance if slope estimates change rapidly. These trade-offs would need to explored in the analyses.

A.18. Should HCRs that investigate the use of some recruitment index (probably from younger fish in survey) be explored further?
This option should be assigned low priority for this OMP revision due to the low likelihood that use of a recruitment index will improve performance, and also because the work to modify (and test) the additions to the operating model may be substantial. The Panel notes that if this approach is followed in future, it may be more robust to base OMPs on consecutive years with good (or poor) recruitment rather than basing management action on single year classes observed at small sizes.

A.19. Other recommendations
A.19.1. (H) Explore the reasons for the inability to obtain a positive definite Hessian matrix when fitting the hake model. Ways to enhance the likelihood of obtaining such a Hessian matrix could include (a) replacing Pope's approximate by the "Hybrid method", (b) imposing soft (rather than hard) bounds on the parameters, and (c) setting the values of parameters that are clearly equal to their bounds (e.g., the residual standard deviations for the ICSEAF catch-rate series) rather than trying to estimate them. Experience with other models that have initially failed to provide a positive definite Hessian matrix suggests an approach of initially estimating only a few parameters (e.g., R_0 and the recruitment deviations), fixing the remaining parameters, and checking for convergence (i.e., here a positive definite Hessian matrix). Repeating this procedure over an increasing number of estimated parameters is a good way to identify the parameters that create such problems in the model.

A.19.2. (H) Whether the non-linear minimizer is converging to the true minimum of the objective function can be investigated using jittering (i.e., randomly perturbing the initial values of the parameters and re-fitting the model). Jittering should be a standard part of South African assessments based on complex models.

A.19.3 (H) There are some large changes (since 2013) to the results for the scenarios that form the Reference Set of trials (MARAM/IWS/2017/Hake/P4). While both data and model specifications differ between the new and old scenarios, some of the changes appear unexpectedly large and any such changes should be checked for possible convergence problems. In general, no results should be presented that have not had convergence verified.

A.19.4. (L) Consider applying the method developed by Methot and Taylor (2011) for specifying year-specific bias-correction factors for the stock-recruitment residuals once the assessment is able to provide a positive definite Hessian matrix.

B. Sardine

B.1. Have we an adequate Reference Set of operating models for sardine? How do we best report performance statistics for this set?

The Panel **recommends** that the stock-recruitment relationship be estimated outside of the stock assessment (i.e., option (i) in item B.2.2b from the 2016 Panel report) as this is both an acceptable approach and should speed the OMP revision process.

The Reference Set should include alternative values for *p* as well as the MoveR option. It is desirable to explore another model of west to east movement. However, the evidence for the current alternative (0.5MoveR) is not sufficiently strong to include it in the Reference Set.

Weighting of factors within the Reference Set to provide weighted outcomes should make the selection of an OMP easier (e.g., narrow the ranges of candidate MPs). However, the selection process must also involve examining the results for each individual trial in detail. The Panel lacked a basis for commenting on the factor weights presented in Table 1 of MARAM/IWS/2017/Sardine/P6.

A robustness trial should be developed in which the November biomass on the south coast in 2016 is forced to be low, consistent with the associated survey estimate.

B.2. How do we best choose risk threshold biomass for sardine and anchovy?

There is much less consistency worldwide in the choice of the threshold biomass level for small pelagic species such as sardine and anchovy than for longer-lived species such as cods and hakes. In addition, there is considerable variation in the basis for the threshold biomass selection among jurisdictions; these include impacts on average catches and on the broader ecosystem. The Panel **recommends** that the threshold biomass for defining risk be set to the lowest level of spawning biomass provided that recruitment near this biomass was 'reasonable' (e.g., recruitment at half the asymptotic level for the Beverton-Holt or Hockey Stick stock-recruitment relationship). This approach reflects that there is no basis to infer that recruitment will be adequate at levels of spawning biomass lower than lowest spawning biomass ever observed. Alternatively, a higher level of spawning biomass could be selected so that recruitment at that biomass would be expected to be 'reasonable'.

The Panel notes that the risk identified relates to spawning biomass falling below a pre-specified risk threshold. Given the natural variability of the sardine stock, the extent of precaution of the candidate MPs rests on how the harvest control rules reduce fishery catch. Therefore, it is critical that the OMP actually implements what is likely to occur in practice when biomass is below the harvest control rule biomass threshold.

B.3 (*). How do we best select the acceptable probability of dropping below a risk threshold?

Risk relates to a threshold level of spawning biomass and an acceptable probability of the resource dropping below this level. There are two ways to define this acceptable probability: (a) by analogy with such probabilities that were accepted previously, which could potentially maintain consistency in risk tolerance over time, and (b) by examining afresh the trade-off between risk and other performance statistics. Option (a) involves attempting to keep risk at that for OMP-02 even though changes to the specifications of the operating model (e.g., the extent of variation in recruitment) may have changed over time. This level of risk was selected around the time a joint sardine-anchovy OMP was first adopted, by considering the trade-off between the probability of dropping below a threshold biomass value and other metrics of performance. Option (a) will be difficult to implement at present, given the very marked changes in the structure of the operating model.

For option (b), the Panel suggested examining the additional risk imposed by the directed fishery compared to the case with a zero directed fishery, which is compatible with Working

Group proposals in ICES for similar stocks (ICES WKGMSE 2013). The following algorithm could be used to define such an acceptable probability:

- Identify a "tuning trial" – the Panel suggests the trial with p=0.08 and MoveR as a trial for which risk should be non-zero but not too high, and thus provides contrast in results for different candidate management procedures.
- Identify a management procedure that has a tuning parameter such that increasing values for the tuning parameter leads to higher catch limits for the same stock size (such as OMP-14).
- Construct a zero future catch scenario.
- Conduct simulations for a range of values for the tuning parameter that is sufficiently wide that the median long-term (years 11-20) catch stops increasing if the tuning parameter is increased any further.
- Plot (minimally) the median and lower 5th percentile of short-term (years 3-5 given pre-specified catches for 2016 and 2017) and long-term (years 11-20) catch against the probability of biomass dropping below the threshold level (the "risk" – computed via method (i) or (ii) as defined below).
- Select a risk level (R_*) given the trade-off among the various summary statistics.
- Compute ΔR_1 as the ratio of R_* to the risk for zero catch (R_0), and ΔR_2 as the difference between R_* and R_0.

The process for selecting OMP-18 would then involve comparing the performance statistics for alternative candidate MPs such that risk does not exceed the maximum possible risk, which is defined for any given trial either as the risk when catch is zero plus ΔR_2 or this risk multiplied by ΔR_1. These performance statistics would include short- and long-term catches, the probability of a severely reduced fishery and a measure of annual variation in catch.

The first four steps of this algorithm should be applied to the base case trial on which OMP-14 was based to allow the trade-off among the various summary statistics on which OMP-14 was based to be determined.

The Panel was presented with three methods to compute the probability of biomass dropping below the risk threshold: (i) the proportion of simulation trials in which biomass drops below the risk threshold at least once during the projection time period; (ii) the average probability of biomass being below the threshold over all projection years; and (iii) the probability of biomass being below the threshold in the final year of the projection. Of these alternatives, the Panel eliminated method (i) because it lacks discriminatory power for species such as sardine that exhibit high natural variability and particularly given the current low sardine biomass. The approach suggested above therefore could use either method (ii) or (iii).

B.4. Can one dispense with risk and simply consider catch over the medium-to-long term as sufficient to incorporate any negative consequences of undue depletion of the population?

This approach to evaluating management procedures should be considered for use in future OMP evaluations, and may be more appropriate for long-lived species that are not as heavily subject to large process error variation. The Panel does **not** though **recommend** using this approach for the current OMP revision both because of local inexperience with interpreting results in this form, and because this approach depends heavily on having a stock-recruitment relationship that is reliable at low biomass (at a minimum, sensitivity to alternative assumptions for that relationship, including ones involving depensation, would need to be checked). An alternative to evaluating the risk of undue depletion explicitly would be to report how often fishery catches are unduly low, which may be more robust to the uncertainty

associated with the stock-recruitment relationship at low biomass. Furthermore, choosing to focus on average yield over only the medium-to-long term would lose information on the proportion of years with a severely reduced fishery. If a small probability of a very low TAC is considered a potential management objective, then one cannot dispense with analysing this risk.

B.5 (*). What would be the best way to simulation test the impact of a single area directed sardine TAC in a situation of two spatially distinct sardine population components?

The approach in MARAM/IWS/2017/Sardine/P4 is an appropriate way to model the impact of a single area directed sardine TAC. Uncertainty in the relationship between the proportion of the catch west of Cape Agulhas and the ratio of the biomass off the west coast to the TAC should be taken into account by fitting this relationship to each draw from the posterior.

B.6. The 2016 panel recommended OMP variants that include spatial management be considered (recommendation B.1.3). Is spatial management of the sardine TAC necessary? If we consider explicit spatial management to be necessary during "concerning periods" only, how do we best determine the "flags" for switching such spatial management on and off?

The approaches outlined in MARAM/IWS/2017/Sardine/P7 are sensible ways to include spatial management considerations in an OMP. These options should be explored if the calculations to determine whether or not there is a need for spatial management confirm that the risk criterion cannot be satisfied without some spatial management (or that TACs need to be very low in the absence of explicit spatial management).

B.7. How might one best specify the November survey estimate of abundance below which the directed sardine fishery should be closed?

The Panel response to this question is framed in the context of a short-term (December 2017) management decision, rather than in the context of a threshold biomass in an OMP harvest control rule. The latter should be based on the values for performance statistics from simulation trials, some of which will depend on the choice of risk threshold biomass. In relation to the short-term need, the Panel **recommends** that the approach outlined for item B.2 be adopted, except that the lowest level of estimated spawning biomass at which estimated recruitment was 'reasonable' be converted into survey biomass. In addition, it may be appropriate to increase the value so computed to account for the impacts of observation error – the latter is not relevant to item A.2 as the threshold biomass there is computed using the operating model rather than an actual estimate of biomass from a survey.

C. West Coast Rock Lobster

C.1 (*). For the data available, how might the analysis methods being used be improved?

The analysis of the MCS confiscations and effort data is complicated by the inability to link confiscations to the enforcement type (coastal, FPE, restaurant, sea, slipway and vehicles). Thus, GLM methods have been used to synthesize the data to enable a confiscations per-unit-enforcement effort index of the extent of poaching to be developed. Although the current GLM approach to analysing the effort data (MARAM/IWS/2017/WCRL/P3; MARAM/IWS/2017/WCRL/P4) is broadly sound in principle, it relies on the tenuous assumption that the efficiency ("q") for each enforcement type is roughly the same. For example, equivalent efficiency assumes that the number of confiscations per sea investigation is roughly the same as the number of confiscations per FPE investigation. Unfortunately, information provided by Compliance to the Panel suggests that this is not the case. Moreover,

the trend in effort for enforcement types with low average effort (which are down-weighted in the current analysis approach) differs from trends in high and average effort enforcement types.

Efforts for the various enforcement methods need to be corrected for their variable efficiency before being combined into an overall effort index. This efficiency should be estimated using all available information (quantitative and qualitative), and the analysis re-run using the original GLM procedures presented to the Panel. DAFF Compliance should also work with the analysts to identify events that might have changed the relationship between confiscations per-unit-effort policing effort and the magnitude of poaching over time.

The Panel **strongly recommends** that there be ongoing efforts to link confiscations to the enforcement method, and the resultant information provided to analysts to improve future monitoring of poaching.

C.2. Have the results obtained from the current method been accurately summarised?

The approach taken to summarize the trends in poaching-related catch (MARAM/IWS/2017/WCRL/P6) is appropriate given the data available to the West Coast Rock Lobster Scientific Working Group (but see discussion under item C.1).

C.3. The DAFF MCS confiscations and effort data, and the TRAFFIC analyses, suggest different trends in the extent of poaching over time; how best might these differences be taken into account?

It is somewhat premature to address differences in trend given the need for additional analysis (item C.1). However, it should not be surprising that the trends from the DAFF MCS confiscations and effort data differ from those for the TRAFFIC analyses, as the TRAFFIC data pertain only to exports while the MCS data relate also to illegal catches that would enter domestic as well as export markets.

MARAM/IWS/2017/WCRL/P5 provides an analysis of data on international trade in lobster from South Africa. The data from this analysis have been used to specify the extent of poaching in 2008. The West Coast Rock Lobster Scientific Working Group based the value used in assessment models on the difference between South African exports and South African landings (SAE-SAL), while an alternative estimate is provided by the difference between world imports and South African landings (WI-SAL). The Panel notes that the SAE-SAL values are based on information that is easier to validate. However, the SAE-SAL estimates of illegal removals are almost certainly under-estimates owing to (a) some legal landings being consumed in South Africa, (b) the method not accounting for illegal catches sold in South Africa, and (c) missing illegal exports that are not declared as being South African exports. The WI-SAL information suggests much higher levels of poaching, but relies on the imports indicated as coming from South Africa being correctly assigned to country. There is no direct evidence that imports of lobsters are being incorrectly labelled as being from South Africa, although the generally higher prices of South African lobster may motivate this.

C.4 (*). How best might one determine reliably whether and to what extent the magnitude of poaching might change in the future?

In a monitoring context, an assessment of whether and to what extent the magnitude of poaching has changed should be based on compliance confiscations and effort data. The Panel encourages investigation into the ability of future poaching trend analyses to detect changes in actual poaching effort and/or amount of lobsters poached. Such analyses could be

used to improve the power of the monitoring system and hence the ability to resolve alternative poaching scenarios in the future.

The Panel also considers that in regard to future poaching scenarios, Reference Set models for the OMP evaluation should be based on evidence (as with their other components). Such scenarios could be based either on continuation of current levels of poaching, or on projecting trends identified from analysing the historical poaching data. Possible responses to additional enforcement effort are not easy to model, because it is not possible to predict to what extent a given management measure might impact on poaching levels. Scenarios involving such improved enforcement should be investigated as a research project as they arise, and incorporated into future OMP evaluations once their efficacy has been demonstrated.

C.5. If additional compliance measures of a different nature are introduced, with possible impacts on the behaviour of poachers, how best might the existing poaching trend index derived from DAFF MCS data be calibrated?

The Panel encourages investigation, development and implementation of new compliance measures as necessary to improve compliance. It is critically important that if such measures are introduced, information is collected on both the effort level and the number of lobster recovered per unit-of-enforcement effort. This information is of critical importance for scientists to be able to analyse ongoing trends in poaching levels, as well as the response to enforcement.

C.6. Are there any suggestions for other possible approaches to quantify poaching magnitudes and trends, either with the data currently available, or with additional data which could be practically collected?

Although subject to uncertainty as noted in item C.3, the TRAFFIC data provide a way to estimate the magnitude of poaching that enters the international market. However, TRAFFIC data provide no information on domestic consumption of illegal catches. The Panel **recommends** that a **high priority** be given to developing a way to quantify the magnitude (and ideally trend) of illegal catches that are sold domestically.

D. South Coast Rock Lobster

D.1 (*) The Panel reviewed a comparison of the predicted distribution of future catch-rates using the operating model on which OMP-2014 was based, with the observed catch-rates for 2012, 2013, 2014, and 2015 for each of the sub-areas A1E, A1W, and A2+3. Nine of the 12 catch-rates are within the predicted distributions, whilst those in which this was not the case are for sub-area A1W, which has the lowest catches. The Panel therefore does not see a need for an immediate in-depth review of the OMP for South Coast Rock lobster. Rather, focus for future work for this species should examine the spatial distribution of catch size-composition and catch-rates to examine whether fishing in areas with high catch rates might lead to dome-shaped selectivity patterns as are evident in MARAM/IWS/2017/SCRL/P2, as well as to a non-linear (hyperstable) relationship between catch rates and lobster abundance. The results of this examination could be used to refine the operating models used for South Coast Rock Lobster when the OMP is next reviewed.

D.2. Other recommendations

D.2.1 (H). Future OMP development work should involve Exceptional Circumstances rules.

D.2.2 (M). There is value in representing parameter uncertainty using samples from an appropriately constructed distribution, given that there is currently only one operating model structure, which means that the level of uncertainty considered when evaluating candidate OMPs is relatively low.

E. Linefish

E.1 (*). Review and comment upon the refined Bayesian surplus production assessment method put forward in document MARAM/IWS/2017/Linefish/P2.

JABBA-Select involves generating values for $H_{MSY,S1}$ (the proportion sustainably harvested at B_{MSY} for S1, the first fleet included in the assessment) and *m* (the shape parameter of the Pella-Tomlinson surplus production function) independently. However, there is no guarantee that these two quantities are actually independent. This remains the case even given independent priors on *M* and *h*, as was confirmed during the workshop. Thus, the sampled values for $H_{MSY,S1}$ and *m* should be routinely displayed on a scatter plot, along with the values for $H_{MSY,S2}$ / $H_{MSY,S1}$ and $H_{MSY,S2}$ / $H_{MSY,S1}$, to check the extent to which the assumption of *a priori* independence of these inputs is valid. In addition, the fits of the gamma distributions to the sampled values for these parameters should be shown. If the values are not independent, an appropriate multivariate distribution should be used.

The extent to which priors are updated depends on the quality, quantity, and contrast in the CPUE data. The posteriors for $H_{MSY, S1}$ and *m* should be displayed, along with their priors (the results for Carpenter and Kob suggest that the CPUE data are informative about $H_{MSY, S1}$, but not *m*). In addition, assessments based on JABBA-Select should display the ratios of the posterior variances to the prior variances for the model parameters to assess further which parameters are being informed by the data. Care should be taken to justify prior distributions, particularly when the assumed priors are highly informative (such as those for *M* and initial depletion in the example applications of MARAM/IWS/2017/Linefish/P2). In general, if none of the priors are non-trivially updated, then this should be taken as an indication that the data may be insufficiently informative to form the basis of an assessment.

Assessments based on JABBA-Select should report the posteriors for the observation and process error variances. Ideally, the posterior for the process error variance should be comparable to the variation in spawning biomass obtained by projecting the age-structured model forward for many years without catches, but with process error.

Assessments should also perform sensitivity tests to modifications of prior variances, given the non-linear transformations applied which involve parameters that have priors.

The paper documenting the method was not complete. If it is to form a reference for future work, it needs to specify the method fully. For example, it needs to be documented that the relationship between exploitable and spawning biomass is based on values for the biological parameters at their prior medians, and how the values for *m* from different selectivity patterns are pooled.

JABBA-Select imposes process error on the sum of biomass and production less catch. Sensitivity should be explored to the catch term not being affected by process error.

The observed correlation between $H_{MSY, S1}$ and *m* implies that these parameters should not be generated independently as is the case at present. However, the simulation study suggests that the effect of ignoring this correlation is unlikely to be substantial.

The methodology is able to model changes in overall selectivity resulting from changes in selectivity of individual fleets where length data are available to document the changes, and from variations in the fraction of the catch taken by different fleets. If future selectivity changes were to occur within the different fleet sectors, then model performance would likely degrade substantially if length information was not available from the catches.

E.2 (*). Review and comment broadly upon the simulation approach put forward in MARAM/IWS/2017/Linefish/P3

The Panel was pleased to see the simulation study, which showed that for some fairly informative cases (CPUE linearly proportional to abundance and high contrast in abundance) JABBA-Select can provide close-to-unbiased estimates of depletion and of the exploitation

rate relative to H_{MSY}. In addition, the coverage probabilities for the reported model outputs were close to the nominal levels.

JABBA-Select outperformed the Bayesian deterministic age-structured production model, but the simulation study does not reflect a "fair" comparison of Bayesian production and age-structured models because the inclusion of process error in JABBA-Select allows the production model to better mimic the dynamics of the underlying population that is simulated.

The simulation study focused primarily on "relative quantities". Future simulation studies should include performance metrics related to estimation of spawning biomass in absolute terms.

E.3 General issues and recommendations

E.3.1 (*). The Panel agrees that JABBA-Select is, in principle, an appropriate approach to apply to moderately data-poor (data on catch and CPUE as well as sufficient data to inform biological parameters and selectivity) linefish stocks to estimate "relative" quantities such as H/H_{40} and SB/SB_{40}. However, it is necessary to (a) carefully evaluate the standard convergence diagnostics, (b) apply diagnostics to check that the priors are updated given the data, (c) confirm that the CPUE indices are likely to be reliable indicators of abundance, (d) qualitatively evaluate that no major selectivity changes had occurred within the fleet components, and (e) evaluate if the life history characteristics are suitable for a production model approach.

E.3.2 (*). The Panel notes that the approach presented relies on an underlying assumption that selectivity has remained unchanged since the length composition data were last collected. The applicability of this assumption should be under continuous review.

E.3.3 (*). There have reportedly been difficulties implementing a stochastic age-structured model fitted to CPUE data (i.e., a model that is an age-structured structured analogy of the age-aggregated production model). The Panel has difficulty understanding why it should not be possible to implement a stochastic age-structured model, with fixed values for the biological parameters except for M and steepness, either within a Bayesian or frequentist framework. Such a model would provide a useful comparison to JABBA-Select and should be explored further, including in future simulation studies. Developing such a model would also provide a framework to incorporate potential future data sources, and aid in developing competence in age-structured modelling.

E.3.4 (M). Parameterizing a Bayesian age-structured population dynamics model in terms of "leading parameters" such as MSY and F_{MSY} rather than M or h should lead to improved numerical stability, as well as providing a more appropriate comparison to JABBA-Select.

References

ICES. 2013. Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE); 21 -23 January 2013, ICES HQ, Copenhagen, Denmark

Methot, R.D. and I.G. Taylor. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*. 68: 1744-1760.