Food and Agriculture Organization of the United Nations

FIAF/R1262 (En)

FAO Fisheries and Aquaculture Report

ISSN 2070-6987

Report of the

FAO EXPERT CONSULTATION WORKSHOP ON THE "DEVELOPMENT OF METHODOLOGIES FOR THE GLOBAL ASSESSMENT OF FISH STOCK STATUS"

Rome, Italy, 4–6 February 2019

FAO Fisheries and Aquaculture Report No.1262 FIAF/R1262 (En)

Report of the

FAO EXPERT CONSULTATION WORKSHOP ON THE "DEVELOPMENT OF METHODOLOGIES FOR THE GLOBAL ASSESSMENT OF FISH STOCK STATUS"

Rome, Italy, 4–6 February 2019

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2019

Required citation:
FAO. 2019. *Report of the Expert Consultation Workshop on the Development of methodologies for the global assessment of fish stock status, 4–6 February 2019*. FAO Fisheries and Aquaculture Report No. 1262. Rome. Licence: CC BY-NC-SA 3.0 IGO

ISBN 978-92-5-131425-8

PREPARATION OF THIS DOCUMENT

This document is the report of the FAO Expert Consultation Workshop on the “Development of methodologies for the global assessment of fish stock status” held in Rome, Italy, from 4 to 6 February 2019. The workshop is an FAO initiative aimed at improving its ability to understand global stock status by means of a stock assessment tool that can be used to assess currently unassessed stocks. This report summarizes the presentations and main discussions of the workshop and provides recommendations. The document was prepared by Mr Yimin Ye (Chief of the Marine and Inland Fisheries Branch, FIAF), Mr Edoardo Mostarda (FAO Consultant for the Marine and Inland Fisheries Branch, FAO) and Mr Nicolas Gutierrez (FAO Fishery Resources Officer), and reviewed by the experts involved in the workshop.

ABSTRACT

The FAO Expert Consultation Workshop on the “Development of methodologies for the global assessment of fish stock status” took place in Rome, Italy, from 4-6 February 2019. The overall objective of the workshop was to present to a group of experts in stock assessment, fishery management and policy development a new tool (now called **Sraplus**) that can be used to assess the status of currently unassessed stocks. More in detail, participants were asked to i) review the technical design, soundness and practical applicability of the new methodology, ii) evaluate its performance with simulation data and real fisheries data, iii) discuss its advantages and disadvantages as well as any necessary adjustments, and iv) design a plan for steps forward towards its application in global fish stock assessment. The workshop participants contributed in their individual capacities to the discussions which resulted in a set of recommendations on how to improve the tool and the data that is provided to the assessments.

CONTENTS

ABBREVIATIONS AND ACRONYMS

CBD	Convention on Biological Diversity
CV	Coefficient of Variation
CPUE	Catch Per Unit Effort
EEZ	Exclusive Economic Zone
FAO	Food and Agriculture Organization of the United Nations
FMI	Fisheries Management Index
FPI	Fisheries Performance Index
FSSI	Fish Stock Sustainability Index
ISSCAAP	International standard statistical classification of aquatic animals and plants
LIFDC	Low-Income Food-Deficit Countries
MCMC	Markov chain Monte Carlo methods
MSY	Maximum Sustainable Yield
NEI	Not Elsewhere Included
NOAA	National Oceanic and Atmospheric Administration
OHI	Ocean Health Index
RFBs	Regional Fishery Bodies
RFMOs	Regional Fisheries Management Organizations
RMSE	Root-mean-square error
RPs	Reference points
SAR	Swept Area Ratio
SDGs	Sustainable Development Goals
SIR	Sampling-importance-resampling algorithm
SOFIA	State of World Fisheries and Aquaculture
SPR	Spawning per recruit
SRA	Stock Reduction Analysis
TMB	Template Model Builder
UN	United Nations
WOA	World Ocean Assessment
YPR	Yield per recruit

PARAMETERS

B	Biomass
B/B_{MSY}	Biomass relative to the biomass that produces MSY
B_{MSY}	Biomass that would produce the MSY
F	Instantaneous coefficient of fishing mortality
F_{MSY}	F that would produce the MSY
K	Population's carrying capacity
m	Shape parameter
q	Catchability coefficient
r	Intrinsic growth rate of the population
U	Fishing mortality rate
U/U_{MSY}	Fishing mortality rate (U) scaled relative to the level that would achieve MSY
σ_{obs}	Observation error
σ_{proc}	Process error

DAY 1

Opening of the Workshop

1. The Food and Agriculture Organization of the United Nations (FAO) organized the Expert Consultation Workshop on the "Development of new methodologies for the global assessment of fish stock status" which took place from 4–6 February 2019 at FAO headquarters in Rome, Italy.

2. The workshop was attended by 23 participants with specific expertise in stock assessment, fishery management and policy development and by 13 FAO fishery officers (Appendix 1).

3. The meeting was opened by Mr Yimin Ye, Chief of the Marine and Inland Fisheries Branch, who welcomed the participants and introduced the agenda which was adopted with no amendments as under Appendix 2.

4. Mr Manuel Barange, Director of the Fisheries and Aquaculture Policy and Resources Division welcomed the participants and explained that the purpose of the workshop was to help FAO develop a new methodology for stock status classification. Stock status is reported every two years in the State of World Fisheries and Aquaculture (SOFIA) report, and it is one of the most iconic and most widely used indicators that FAO produces. For this reason, the outcomes of the workshop will have important implications for FAO and for how stock status sustainability is interpreted and used. The context of the workshop lies within FAO's main objectives, or rather, to put an end to hunger and poverty while maintaining resources to sustainable levels. FAO has documented an increase in the number of people suffering from hunger in the last two years mainly due to political and social conflicts and climate extremes. Fish plays a crucial role in food security and nutrition especially in the developing world and 22 out of the top 30 fish consuming countries are Low-Income Food-Deficit Countries (LIFDC). Fish also plays a crucial role in livelihood, employment and cultural identity in many countries. Capture fisheries is the only major food production system in the world that relies completely on exploiting natural biodiversity. This means that fish resources must be managed properly in order to avoid impacts on biodiversity, habitats and the functioning of the ecosystems. FAO is the sole agency in the UN system with a responsibility over fisheries, guiding the world in the direction of sustainable use. With the goal of informing and supporting fishery policy and management, FAO started monitoring the status and trend of the world fisheries resources in 1974, and this has been done every two years since 2004. The indicator of stock status is based on about 450 fish stocks worldwide accounting for about 70 percent of the global catch. However, almost half of this catch is actually from stocks with different degrees of data limitations. Over time, substantial advances in the ability to assess stock status have been made and the time is right to update FAO's methodologies. This would also provide countries with a standard methodology to assess stock status as specifically requested by Sustainable Development Goal (SDG) 14. Following the principle of using the best available scientific information, FAO is making an effort, in collaboration with the University of Washington, to develop this new methodology which is expected to be technically more robust and produce more objective estimations of global fish stock status. This expert consultation is required to scrutinize this new methodology, evaluate its performance both technically and operationally and offer further recommendations, bearing in mind that it must work in all situations, and especially in data-poor conditions.

5. Subsequently, Mr Ye invited all participants to introduce themselves and then presented the work that FAO has been doing in relation to the global assessment of fish stock status.

FAO's work on global stock status assessment

6. FAO has monitored the state of the world's fishery stocks since 1974 and currently carries out the assessment of about 450 stocks by FAO statistical area. This means that all stocks of the same species occurring in a FAO area are treated as a single stock. Stock status is currently classified into three categories, e.g. underfished, maximally sustainably fished and overfished. The proportion of stocks falling into these categories provides a means of monitoring progress and changes in the exploitation and management of global fishery resources as a direct measure of sustainability. The criteria used to determine the status of each stock include i) stock abundance, based on estimates of current stock biomass, catch rates (CPUE) and/or survey abundance indices, ii) spawning potential as the percentage biomass of spawning stock relative to the unfished level, iii) catch

trend and iv) size/age composition. The process of stock status classification follows a decision tree (Figure 1) by means of which FAO can either i) adopt existing assessment results at the regional or country level with adjustments to FAO's criteria, ii) carry out assessments using qualitative or quantitative methods for non-assessed stocks, iii) combine information from various sources and iv) consult local experts for data and information and results validation.

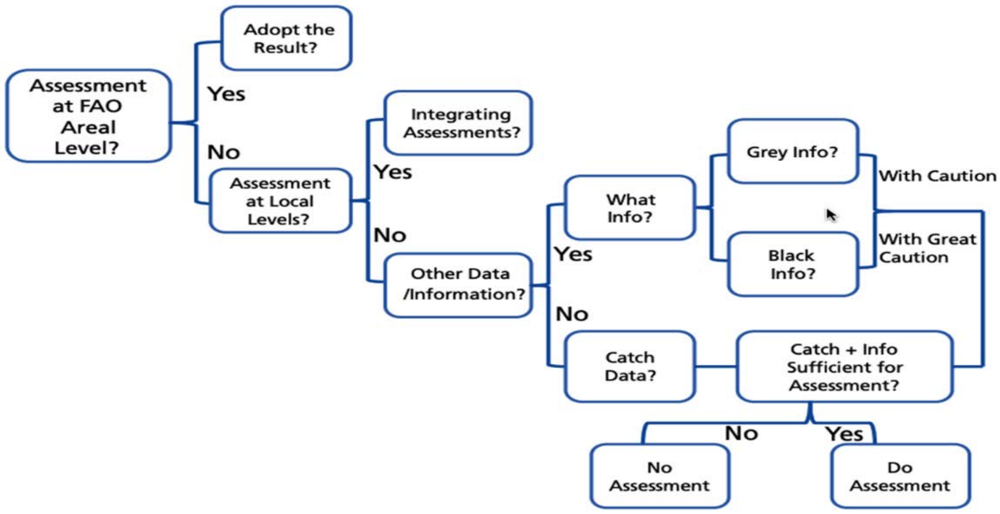

Figure 1. Decision Tree of the stock status classification

7. After assessing all stocks included in FAO's reference list, a global indicator of fish stock status is computed. Other institutions and projects produce indicators of global stock status covering different percentages of total catches, using different methods and having different degrees of quality. Yet, FAO's indicator is the one that is more widely used to measure progress towards a number of global objectives related to fisheries resources (for example, SDG 14, UN WOA, OHI, Target 6 of the CBD, etc.). FAO's methodology for deriving this indicator has some weaknesses. First, it generally uses both quantitative and qualitative information, the latter of which can be rather subjective. Then, there are often issues of consistency when various sources and methods of assessments from different countries of the same area need to be combined. Finally, changes over time in both the methods and number of stocks assessed and, in the experts involved, can generate issues of replicability and transparency. Methods for assessing stocks have changed so that in the 70s and 80s they were mainly based on survey data, catch patterns and catch rates. Only in the 90s and 00s more quantitative assessments were carried out (quantitative pattern analysis). From 2010 onwards, semi-quantitative assessments, e.g. formal data-based assessments together with qualitative opinion-based assessments were carried out and currently the objective is to move towards a wider use of quantitative assessments with a wider coverage of unassessed stocks. To achieve this goal, there is a need to update FAO's assessment methodology also taking advantage of the substantial progress in this field. Advantages of this update would include i) coping with the demand for stock assessments at the country level, ii) supporting member states requests for information on stock status in their waters, iii) making the methodology more widely accepted, iv) an increased transparency and replicability, v) a better valuation of policy effectiveness and management achievements, and vi) increased consistency between species as well as between areas.

8. To conclude, Mr Ye outlined the objectives and expected outputs of the workshop. It was explained that the main objective was to help develop a new stock assessment tool (now called **Sraplus**) that is technically sound and operationally feasible. This means ensuring that the tool and its methodology are i) free of major defects, ii) reliable, especially in data-limited situations, iii) applicable to global fisheries and, iv) user-friendly, especially in countries with limited capacity. The expected outcomes were a set of recommendations for improving the model, the software and the data that is provided to the assessments.

9. After the presentation, a number of points were raised by the participants. Given that catch is the most important source of data of the models under review, FAO were asked to provide more information about the quality of catch data that is reported by countries. Staff of the Statistics and Information Branch explained that the collection and dissemination of statistical information on fisheries and aquaculture is a crucial part of FAO's mission and provided details on the process of data collection, compilation, validation, and dissemination. FAO is the only source of global fisheries and aquaculture statistics, which represent a unique global asset for sector analysis and monitoring. In the 70s, the production of statistics in the agricultural and fisheries sectors was a priority, and several organizations, including FAO, were supporting countries through projects aimed at establishing statistical systems and building capacity in statistics, data collection, and stock assessment. This happened until the mid-80s when the donors decided to leave the responsibility for the collection of fisheries statistics to the countries. The result was a decline in the quality of statistical data. Despite this, FAO has been collaborating with the countries to improve fisheries statistics and has established a series of mechanisms to ensure that the best available information be submitted, revised and validated, either directly or indirectly. Participants also asked FAO to clarify the purpose of reporting the proportion of overfished stocks at a global level considering the substantial differences in the trends of overfished stocks among regions. It was explained that the original purpose of developing an indicator of stock status was to make countries aware of the state of their resources so that they could support and improve their data collection systems. Currently, the main reasons for reporting sustainability are linked to the SDGs process, where countries are supposed to report on the status of their stocks, and there is also an increased focus on any issues related to the sustainability of the natural environment.

Assessing currently unassessed stocks

10. Mr Ray Hilborn presented the background work that has led to the development of the stock assessment tool **Sraplus**. Even though the global trend of overfished stocks is increasing, the pattern is very different at a regional level. These differences are highlighted in SOFIA, but they are often neglected by a number of users. Another issue concerns the number of stocks that are assessed worldwide. FAO in 2011 was assessing 381 stocks (stock here refers to taxon by FAO Area) and currently assessed about 450 stocks, whereas the RAM Legacy Stock Assessment Database, e.g. a compilation of stock assessment results for commercially exploited marine populations from around the world, has assessed about 560 stocks. The FAO catch database includes about 20 000 stocks at a country level (that is taxon by country by FAO Area), but in terms of total catch, only a small portion of these accounts for most of the catch (173 stocks account for 50 percent of total catches, 2093 stocks for 90 percent of total catches). This means that the majority of the stocks has very small annual catches. FAO assesses all major stocks but also a number of minor stocks giving each the same importance. In terms of geographic coverage, all major stocks in Europe (except for the Mediterranean), North America and Western South America are assessed whereas, with the exception of Japan, only a limited number of stocks in Asia and Southeast Asia have been assessed. Regarding Africa, stocks from Morocco and South Africa are relatively well assessed, while several other countries that have small catches, that are however important for food security and livelihood reasons, don't have any information on the status of their stocks. This seems to be related to the lack of data necessary to carry out a formal stock assessment from these countries. This has led authors to develop models based on catch-data only. Catch-based methods by Costello et al. (2016) and Rosenberg et al. (2018) show that there are no significant differences in the status of stocks from developed and developing regions. This indicates that catch-only methods are not able to identify the regions where stock status is known to be poor. However, studies have shown that a wide range of data, in addition to that on catch, is available from the largely unassessed regions. These data include a number of indexes of stock status from expert judgment, such as the Fisheries Management Index (FMI) (Melnychuk et al., 2017), which characterizes attributes of research, management, enforcement, and socioeconomics of fisheries management systems in 28 major fishing countries and which shows a positive correlation with the status of fish stocks (the higher the management intensity, the higher the status of stocks). The Fisheries Performance Index (FPI) (Anderson et al., 2015) is a rapid assessment instrument for measuring the fishery-derived benefits being created not only in the fish stock in the water, but also in the harvest and post-harvest sectors and fishing communities. Additional data are a) the Swept Area Ratio (SAR) (the annual trawled area

divided by the total area of the continental shelf) which has been estimated by Amoroso et al. (2018) at a regional level and related to the relative rate of fishing mortality (U/U_{MSY}), b) time series of fishing capacity and fishing effort at a regional level (Bell, Watson and Ye, 2017), c) scientific surveys catch trends, d) CPUE trends, and for many regions e) time series of abundance and fishing mortality rate for assessed stocks from the same region. These data can be used together with catch-data methods to estimate stock status from unassessed stocks with a higher degree of certainty.

11. After the presentation, a general discussion on the need to define stocks as overfished by means of reference points (RPs) took place and FAO confirmed that the current approach and indicator will not change. Participants also asked whether the new methodology (e.g. **Sraplus**) was only going to be applied to currently unassessed stocks. It was explained that the idea was to use the new tool on unassessed stocks, but it also depended on the outcomes of the workshop and on the feedback provided by the group of experts. Finally, it was pointed out that there is an increasing pressure from the Regional Fisheries Bodies (RFBs) to use their advice regarding the status of stocks that are not assessed by FAO and that methods used by these RFBs based on different models could produce contrasting results.

Purpose and structure of Sraplus

12. Mr Dan Ovando delivered his first presentation outlining the purpose and structure of **Sraplus**. First, it was explained that **Sraplus** is not meant to be a replacement for stock assessment. This tool is meant to be a user-friendly tool for exploring value-of-information in global fisheries, that is to pull together whatever data is available, rapidly get an estimate of stock status for a high number of stocks, and try to see what can be learned from the input data. **Sraplus** builds off of stock reduction analysis (SRA) and quickly provides estimates of stock status as B/B_{MSY}, U/U_{MSY}, and MSY using a variety of potential data sources. These include catch histories, indices of abundance, effort series or changes of U/U_{MSY} or U over time, trawl intensity, and fisheries management indices. There are several other packages for various forms of data-limited stock assessment (e.g. LIME, LBSPR, CCSRA, JABBA, CMSY) which have advantages and disadvantages and given the right data (and applied in appropriate circumstances) may provide good results. The value of **Sraplus** is that it allows users to quickly and easily produce stock status estimates using a variety of data likely to be available to FAO at global scales, e.g. catches, FMI, SAR, and effort series. **Sraplus** was developed to explore the value of including different sources of information in global fisheries assessment.

13. Most of the methods that have been used to assess stocks at the global scale are catch-only methods as catch data is the only easily available source of data for most countries. However, catch-only models require users to make assumptions that are in fact the real assessment and catch data help provide biological context to these assumptions. If no assumptions were made, catch data would likely be misinterpreted for a number of reasons. For example, low catches in a time series could depend on different factors such as the implementation of management measures, low market demand, or to an actual depletion of the stock. Conversely, high catches could be a result of an increase in either the fishing effort or stock biomass. As a result, catch-only models have tried to integrate assumptions or other data to reduce their uncertainty level. One of the most commonly used catch-only model is called Catch-MSY (Martell and Froese, 2012) which is a form of SRA. The conceptual base of this model is that if we know how many fish there are today, and how much we caught before, we can estimate how many there were. At its simplest level, it uses a catch time series and prior distributions of life history parameters (e.g. *r* or the intrinsic growth rate of the population, and K or the carrying capacity). It projects the population forward from some assumption about initial and final population depletion using the randomly selected *r* and K values and eliminates any samples that “crash” the population (for example results with a $B < 0$). It then keeps samples that fall within assumed bounds of initial and final depletion and collects distributions of "successful" parameters. It is critical to note though that the decisions on the assumptions predetermine the range of B/B_{MSY} values that the model “estimates” as current status. In the case of Catch-MSY, the baseline model assumes a Schaefer model of biomass growth, which assumes that $B_{MSY}/K = 0.5$. The model also requires some bounds on plausible B_{final}/K values. Putting these two assumptions together we find that once we have decided on the ratio of B_{MSY}/K (which the Schaefer model sets at 0.5), and provided bounds on initial and final depletion, we have exactly set the range of possible initial and final

B/B_{MSY}. Catch-MSY then helps us interpret our assumptions about initial and final depletion in the context of B/B_{MSY} but does not provide new insight into what B/B_{MSY} may be beyond those provided by the user's assumptions. A natural response to this problem might be to simply reflect uncertainty in final stock status by providing a wide range of plausible final depletion values. However, since there will always be more ways to produce higher depletion levels than lower depletion levels (since lower final depletion levels will be more likely to crash the population given the catch history), being "conservative" by setting very diffuse priors on final depletion will by necessity positively bias any final "estimates" of B/B_{MSY}; the wider a user sets the bounds on final depletion, the higher the values of B/B_{MSY} produced by Catch-MSY will be (see also Froese et al., 2017). **Sraplus** is built around a Pella-Tomlinson production model which provides flexibility in the shape of the production curve model. Details about the structure of the model are presented in Appendix 3. Compared to Catch-MSY, **Sraplus** can use proxies for stock status such as the FMI, SAR or information from nearby assessments and "translates" them into priors on stock status that can be used by the model. An example of this process for the FMI was presented and is described in Appendix 4. **Sraplus** can also incorporate time series of U/U_{MSY} such as U/U_{MSY} data from nearby stock assessed fisheries as priors for unassessed fisheries and data such as abundance indices. In this case, **Sraplus** is equivalent to a traditional surplus production model and is, in fact a stock assessment. The final part of the presentation focused on the model's fitting approaches. More details about these methods are available in Appendix 3.

14. During the presentation, several points were made by the participants. The question was raised as to whether the priors could be fitted to U instead of U/U_{MSY} and then set U_{MSY} as an output of the model. Mr Ovando explained that the main reason why U/U_{MSY} was chosen was to build off the SAR database which relates this parameter to U/U_{MSY}. However, the option of fitting the priors to U could be explored. Participants were also interested in knowing for how many stocks the SAR is calculated and it was explained that the SAR is not by stock but by region and that in each region there are sets of stocks with known U/U_{MSY} which can be related to those values of SAR. It was also noted that since the stocks linked to the SAR must mainly be demersal, as this ratio is based on bottom trawled areas, caution should be taken when applying this prior to non-demersal stocks. An observation was made about the fact that the relationship between SAR and U can change over time. It was explained that the information that both the SAR and FMI give is only a snapshot and a constant relationship over time must be assumed. Another point of debate concerned the different turnover time of trawled species and how this could influence the impact of the SAR on different stocks. Finally, participants initiated a discussion on how to test the performance of the model. Two options are simulation testing and using the estimates from the RAM Legacy Database, both of which have advantages and disadvantages. Therefore, it was recommended to test the model on formally assessed stocks (both well and badly managed) pretending that only catch data and other less-valuable information (e.g. SAR, CPUE) is available and assuming that the formal assessments are correct In general, testing the performance of the model in areas where no assessments are conducted will be an issue.

15. The last presentation of the day was given by Mr Ovando and focused on testing the value of information in **Sraplus** against RAM data. The process involves taking RAM stocks, randomly sampling different types of data from these stocks, using the model to predict stock status, comparing the result with that measured by RAM and assessing the relative value of different types of information. Testing against RAM data instead of simulation testing was chosen since the goal was to test real world proxies (e.g. FMI, SAR, etc.) which can't be feasibly generated by simulations. A first case study, the RAM assessment of the Atlantic cod *Gadus morhua* in the North Sea was presented and more details are available in Appendix 5. The results of this exercise are that a) minor changes in data can have a big impact on results, b) **Sraplus** allows users to easily try different kinds of data and c) model complexity without information is not always helpful. The methodology for testing the value of different sources of data was also illustrated and is described in detail in Appendix 6. Finally, an example of how the model performs when conflicting data is available was presented. The first example referred to the RAM dataset of biomass and exploitation rate U/U_{MSY} for the fourspotted megrim in the North Sea with both data unexpectedly showing a decreasing trend and a second example concerned the Red Fish from Eastern Australia with a known decreasing trend in biomass coupled with expert opinion on the final depletion level.

16. During the presentation, participants discussed a number of points, such as a) the need to consider the lower reliability in catch data for small stocks compared to major stocks, b) the need to take into account the influence of environmental factors over time, especially when dealing with very long time series, and how potential productivity (e.g. carrying capacity) may have shifted to a new baseline level as a consequence, c) the issue of using priors estimated at a regional level and applying them to country-level stocks (a well-managed stock assessed with priors from a poorly managed region, would likely result in bad condition), and d) how to consider relatively short time series which don't show significant changes in biomass.

DAY 2

Use of regional priors for the assessment of data-limited stocks

17. The second day of the workshop was opened by a presentation delivered by Mr Ovando on the use of regional priors for the assessment of data-limited stocks. In the previous day's presentation, performance was tested against RAM data and the "truth" was therefore measured against the estimates by RAM. This route was chosen in order to compare estimates produced by proxies such as the FMI to those from "gold standard" assessments. For unassessed stocks, there is no empirical benchmark, but results can be compared to expert opinion and see whether different data tell different stories. Three regions were chosen to demonstrate **Sraplus** when not explicitly fitting to RAM data: Asia, Europe and East India. More details about these practical applications are available in Appendix 7.

18. After the presentation, in order to better understand the applicability of **Sraplus** to data-limited stocks, the attending experts who are working in countries that seem to have data gaps and no publicly available assessments, were asked to report on the type of data collected in their region that could be used to carry out the assessments. In particular, they were asked to provide information on the availability of time series of abundance, catch and exploitation rates by stock and of additional information such as SAR and survey data. Mr Sathianandan, Mr Valinassab, Mr Tuda, Ms. Palomares, and Mr Chacate explained what type and how fisheries data is collected in India, Iran, Kenya, Philippines, and Mozambique, respectively.

19. Towards the end of the session, participants discussed the strong effect of the priors on the assessment compared to the information provided by catch data only. The main concern was that if the objective is to derive B/B_{MSY} and the priors of B/B_{MSY} are set by the users, an assessment is already being made *a priori* and the assessment is prior-dependent more than catch-series dependent. This suggests the use of caution in the process of selection of the priors and users should also be advised of the fact that, in some instances, the input data may not be informative and **Sraplus** should not be used. For this reason, it was recommended that FAO continue to support countries, collaborate with local experts and provide guidance on the use and selection of data, and on the overall assessment process. Moreover, the risks related to applying regional-level priors to country-level assessments should be emphasized. In general, participants pointed out that care should be taken towards not giving the idea that these are real stock assessments. This might be clear to fisheries scientists, but decision and policy makers would most likely interpret and use the outcomes of these exercises as formal assessments. The term assessment should be used carefully when using these models and it must be very clear what information is going into the model and what is coming out.

20. Before delivering his last presentation, Mr Ovando asked the participants from the United States of America to explain how the process of assessment of data-limited stocks has evolved in the USA. It was explained that the fisheries management system requires that an annual yield be set for each stock. Therefore, the goal of these assessments is not to understand stock status but to estimate the yield that can be harvested. This is sometimes done by running a model very similar to that used in **Sraplus**, using a catch time series as input data, setting a prior on stock status and defining life-history parameters driving productivity. As these tiers are not considered true assessments, they are not included in the Fish Stock Sustainability Index (FSSI) index used to report on SDG indicator 14.4.1.

Validity and reliability of the model

21. The following presentation focused on the validity and reliability of the model. It was explained that the goal of the workshop was to advance methods for the understanding of global fish stocks and that **Sraplus** had been used to demonstrate the value of different kinds of information, and its ability to adapt to different kinds of locations. In the last exercise, two questions were tested: how the choice of data affects the perception of global stock status and how these estimates compare to current assessments. The tests were carried out by pairing each of the stocks presented in FAO report number 569 (FAO, 2011) with various forms of alternative data (see Appendix 8). At the end of the presentation and in relation to the conflicting results of this final exercise, the group of experts raised a number of points, such as a) the need to incorporate, into the model, a sensitivity analysis to determine how much each prior and its variation influences the outcomes, b) the possibility of using this methodology with a standard protocol throughout time in order to look at stock status trends, being aware that high levels of uncertainty do not allow for detecting changes over time, c) the need to get data that is adequate to the stocks under assessment (e.g. regional priors may not work for country-level assessments), d) the need to be as transparent as possible regarding the way the assessments are carried out and their level of uncertainty at a regional level, e) the opportunity of keeping the indicator of global stock status separated based on a specific sets of stocks from the regional and country-level indicators that might be based on different priors and stocks, f) the risk of aggregating catch data by region without evidence of the homogeneity in the catch trends among countries, g) the fact that catch data alone is not informative for the purpose of assessing stock status unless it is coupled with an appropriate index of abundance (such as a reliable CPUE index), h) the need to provide guidance to local stakeholders on the use of this tool and on the weight of different types of information, i) the need to clarify that this model is primarily meant to be used to estimate stock status of currently unassessed stocks according to FAO's criteria.

22. In the afternoon, a hands-on session on how to install and use **Sraplus** was conducted by Mr Ovando.

General considerations on the performance of the model and sources of data

23. In the last session, participants were asked to discuss a number of points that could help draft recommendations on the use of the model for global stock assessment. The following considerations were made:

a) A list of technical suggestions on how to improve the model and its use from both a user's and developer's perspective should be compiled and circulated among the team of experts involved in the workshop who are competent in R, before confirming the model to be a useful tool for stock status assessment.
b) If all the users have is catch data, the use of the model should not be recommended. In case additional information is available, guidance should be given on which type of information is to be considered sufficient to run the model and derive stock status.
c) Additional information is almost always available, but it is often of a non-quantitative nature (e.g. expert opinion, local knowledge). A way must be found to formalize this additional information so that it can be incorporated into the model as a prior. Moreover, FAO should make an effort to document how and what information is used to assess a stock so that countries can better understand the assessment process and know what further information can be provided.
d) Given the high degree of variability in the type of information provided by experts, a protocol or guiding principles on how to incorporate this information in the assessments in a more standardized manner is recommended.
e) Although expert opinion is important, priority should be given to the collection of quantitative data (e.g. survey data, effort, etc.) if funding is available.
f) An indicator of the quality of the assessments by region could be developed. For example, knowing the percentage of stocks that were formally assessed and those whose assessment relied on catch data and other sources of information could be useful.
g) Formally assessed stocks and those assessed with less robust methodologies could be presented separately.

h) Improving the collection of data should be a priority for countries.
i) Recognizing that expert opinion often reflects local knowledge regarding changes in effort by a specific fishery, countries should be encouraged to allocate funding for the collection of effort data.
j) An alternative approach for providing information on the status of unassessed stocks could be to look into their trends in biomass estimated from survey data and make assumptions on their status.
k) The exchange of information between FAO and local experts aimed at gathering more data should be enhanced.

DAY 3

Selection of the reference list of stocks

24. The first presentation of the day, focusing on the reference list of fish stocks for global monitoring was given by Mr Ye. The main topics covered were a) statistical facts of FAO landings data, b) how to determine a reference list of stocks for global monitoring, c) the differences in the reference lists among country, regional and global indices, and d) guiding principles for establishing a reference list. A closer look at FAO landings database shows that total catches have increased since the 50s and that only a small portion of them is composed of species items (this term is used to identify the statistical taxonomic unit, which can correspond to species, genus, family or to higher taxonomic levels) with catches lower than 3 000 tonnes/year. The number of species items by FAO area has increased linearly from 1950 to 2016 reaching about 1 300 species items in 2016 and accounting for 99 percent of the total catch. The number of species items by country has also increased linearly, reaching 9 000 units in 2016, 3 000 of which account for 99 percent of total landings (excluding tuna, migratory and straddling species). Species items are classified according to the International Standard Statistical Classification of Aquatic Animals and Plants (ISSCAAP) in taxonomic groups and within these groups species items are identified at the species, genus, family and order level. An important part of the catches is composed of groups referred to as "not elsewhere included" (nei) which have a poor taxonomic resolution (e.g. marine fishes nei, mackerels nei, crustaceans nei, etc.). A high number of countries reports less than 10 species items, and these generally account for more than 90% of their catches. This would lead to thinking that the selection of stocks representative of the country's catches can be relatively easy for many of them, but a closer look shows that most of these stocks are identified as nei groups at different levels (and often as marine fishes nei), which are difficult to assess. Peru's catches provide an example of the risks associated with considering the percentage of total catch the main criterion for building the reference list of stocks representative of each country's catches. In fact, a single species, the anchoveta, accounts for more than 85 percent of or its landings and assessing only this stock would likely not reflect the status of Peruvian stocks. A number of suggestions were made by the speaker with regards to the criteria for compiling a reference stock list for monitoring sustainability at both the country/regional and global level. An indicator of stock status at the country/regional level could be based on a list of stocks a) covering between 50 and 80 percent of the total catch, b) of major catch, economic value, and social/cultural importance, c) of concern from various stakeholders, and d) shared with other countries.

25. Participants were asked to discuss and offer input on several questions related to the criteria for stock selection. In particular, feedback was sought on whether a) nei groups should be uniformly included, b) a minimum number (10–15) of stocks should be established, c) all stocks should be equally weighted or weighted by their annual catches or both. As for the global level index, it was asked whether a) ISSCAAP groups should be equally represented, b) the index should be weighted by the country's landings, c) the reference list should remain relatively stable over time, and d) discrepancy should be allowed between global and country/regional reference lists.

26. A general discussion followed in which it was stressed that the criteria largely depend on what the index wants to reflect. If the index is related to the SDG indicator 14.4.1, the proportion of fish stocks within biologically sustainable levels, then the number of stocks, and not necessarily their importance in weight should be considered. If the index's goal is to show the performance of a country on stocks' sustainability, then considerations on their economic and social value, or ecosystem importance should be discussed on a country-by-country basis. Mr Ye clarified that the index under discussion is only about fisheries resources

sustainability. That said, participants agreed that different lists of stocks for country, region and global assessments should be compiled and they should ideally include a selection of stocks that are important for different reasons.

27. In the following presentation, Mr Ye addressed two issues that need particular consideration: how to deal with shared stocks and how to use assessments based on fishing mortality as indicators of stock status. The need to deal with shared stocks was illustrated by showing the example of the Mediterranean Sea where 30 countries harvest 354 stocks most of which are shared among two or more countries. If the assessment unit is stock by region, then 354 stocks can be assessed, but if the unit is stock by country then more than 2000 stocks can be assessed. Therefore, a decision must be taken on whether shared stocks should be considered as one assessment unit or not. Regarding stock status and sustainability, this is usually defined by the value of B/B_{MSY} and FAO sets the reference point of overfished to below 0.8 B/B_{MSY}. Stocks with B/B_{MSY} between 0.8 and 1.2 are considered maximally sustainably fished and those with B/B_{MSY} higher than 1.2 are underfished. In relation to the SDGs, stocks fished at biologically sustainable levels comprise the underfished and maximally sustainably fished stocks. Other reference points, which are more useful for management, are based on the fishing mortality F (e.g. F-based yield per recruit (YPR) or spawning per recruit (SPR)), but these can hardly be converted into B-based reference points. Several authors have tried to relate F and B (Carruthers and Agnew, 2016; Gabriel and Mace, 1999) but this relationship is still not well defined. In some cases, for example in the Mediterranean Sea where a constant fishing effort over time can be assumed and where F/F_{MSY} is available for many stocks, an estimate of B/B_{MSY} can be derived. **Sraplus** also would allow users to use estimates of F/F_{MSY} together with catch data to inform stock status (with the F/F_{MSY} priors helping separate out process error from effort changes). In any case, it is important to assess the status of the stocks for which only F-based assessments are available. For example, the USA and Australia report all stocks as having F less than F_{MSY} but still have a relevant percentage of overfished stocks.

Recommendations and next steps

28. During the last session of the workshop, the participants discussed and drafted recommendations for improving the **Sraplus** stock assessment tool and the data that is provided to the assessments. The following recommendations were made:

Model recommendations

1) Catch-only data and heuristic methods are not suitable for determining stock status if no additional information is available. A minimum level of information or data to use as input in the models must be identified in order to get the desired level of confidence in stock status classification. Moreover, guidance should be given on what information and data should be prioritized to inform these methods (e.g. initial and final depletion levels). It is strongly recommended to include additional data (e.g., survey data, standardized CPUE, fishing effort, and estimates of F). Data or information concerning other data-rich stocks occurring in the same area (country or region) should only be used when a connection with the stock under assessment can be demonstrated, and this relationship must be documented.

2) The **Sraplus** software provides a flexible way to incorporate auxiliary data or information into the assessment when only limited data time series are available. Expert opinions play an important role in deciding what data to use and how to use it. Therefore, the involvement of local experts when using such methods is important.

3) It is essential that all assumptions and input information be made explicit in each model run as specific outputs of the software, so that users can’t avoid documenting them. Guidance should be developed and provided in order to minimize risks of misuse (i.e., getting outputs that make no sense).

4) Additional sensitivity tests with regards to the classification of stock status should be designed and performed, including:

    a. A sensitivity analysis on each data input and priors, as well as on qualitative and experts judgement information. Detailed and extensive sensitivity analyses should be done relative to all inputs.
    b. Sensitivity analyses and diagnostics relative to the input data and assumptions, and these should be reported in a standardized way.
    c. Additional model performance tests should be carried out and model performance metrics should be e.g. percentage of miss-classification rather the RMSE. In addition, validation exercises should be conducted, for example by applying the method to fish stocks in data rich countries simulating the absence of important sources of data, and the results compared.
    d. Model performance should be expanded by using other life history scenarios. For example, short-lived species should be included in order to test the performance of **Sraplus** in all situations.
    e. Further investigation on the value of information within FMI and SAR databases as well as the performance of the model when using these databases by themselves (i.e. not in combination with other information) is required since results don't seem to always be consistent with "true" status at the regional level.
    f. It should be evaluated whether the general performance of **Sraplus** compared to RAM improves when the m (shape parameter) is fixed inside the model and when process error is eliminated.

Information and data recommendations

5) The user interface of the package should be improved, including developing a user manual.

6) The Experts' group agreed that the most crucial step with which to improve stock status classification and coverage is to identify additional existent data sources, databases and information. Data improvements (collection, collation and access) should be done at the local level and when possible in cooperation with RFMOs and RFBs. Catch data only is not sufficient, therefore a more formalized process to get additional information, including expert opinions is needed. This information should be included in a standardized way into the methods. This process needs to be done jointly with the countries.

7) A data input and collation protocol should be drafted to guide countries and their stakeholders on how to get the information needed to improve priors and ultimately stock status classification.

8) Strategies for the preparation and collation of data should be rolled out in those regions, countries and fisheries where the information is most limited (in terms of quantity, quality and/or access) and fisheries most critical. This data improvement step should consider a more long-term strategy on improving data for countries (not just by countries).

9) These processes should include a consultation stage in terms of information and outputs at the national level.

Method and metric recommendations

10) In relation to SOFIA's assessment units, and keeping the objective of the metric clearly in mind, guiding principles should be drafted for:

a. Consistency in unit of assessments (what constitutes a “stock”).
b. Selecting reference stocks at the national and/or regional level (by consulting countries, RFMOs or RFBs trying to minimize risks of selection bias (available/good). To ensure representativeness of fish stock status, specific guidelines on various aspects such as catch coverage, livelihood, social factors, minimum number of stocks should be drafted.
c. Identifying aggregation criteria at the global, regional and national levels: e.g. weighting (by catch or other factors) or not weighting.
d. Dealing with “nei” categories, considering they might be a substantial proportion of the landings (e.g. refine groups by collecting info as part of country data protocol).

11) Shared stocks should be included using the existing assessment results. Ways to integrate shared stocks into the aggregated index should be explored.

12) A way to categorize stock status based on local experts opinion should be found, as long as the process is transparent, documented and validated by FAO.

13) FAO should keep performing both the new and old assessments for a few years to understand the differences in both absolute values and trends between the two indices. Some guidance for what time period the new index should be applied to should be developed.

14) There is a need to explore and understand how to integrate the outputs of the different methodologies into a single index (for example, how to integrate different definitions of ‘overexploited’, different reference points, different methods).

15) The indicator should be characterized in terms of coverage (e.g., % of total landings) and reliability (e.g., % from formal or official assessments, % from expert opinion, etc.).

16) It must be emphasized that the objective of this project is to improve the methods for stock status classification for SOFIA (in terms of accuracy, representativeness and coverage, replicability, and transparency) without changing the nature and purpose of the indicator (i.e., measuring the biological sustainability of fishery resources).

17) FAO should invest more resources to improve stock status classification including data and information and capacity building at a national level.

REFERENCES

Amoroso, R.O., Pitcher, C.R., Rijnsdorp, A.D., McConnaughey, R.A., Parma, A.M. Suuronen, P., Eigaard, O.R. et al. 2018. Bottom Trawl Fishing Footprints on the World's Continental Shelves. *Proceedings of the National Academy of Sciences*, 115(43): E10275-E10282.

Anderson, J.L., Anderson, C.M., Chu, J., Meredith, J., Asche, F. Sylvia, G., Smith, M.D. et al. 2015. The Fishery Performance Indicators: A Management Tool for Triple Bottom Line Outcomes." *PLOS ONE*, 10(5): e0122809.

Anderson, S.C., Cooper, A.B., Jensen, O.P., Minto, C., Thorson, J.T., Walsh, J.C., Afflerbach, J. et al. 2017. Improving Estimates of Population Status and Trend with Superensemble Models. *Fish and Fisheries*, 18(4): 732–41.

Bell, J.D., Watson, R.A. & Ye, Y. 2017. Global Fishing Capacity and Fishing Effort from 1950 to 2012. *Fish and Fisheries*, 18(3): 489–505.

Carruthers, T. & Agnew, D. 2016. Using simulation to determine standard requirements for recovery rates of fish stocks. *Marine Policy*. 73. 146-153.

Cope, J.M. 2013. Implementing a Statistical Catch-at-Age Model (Stock Synthesis) as a Tool for Deriving Overfishing Limits in Data-Limited Situations. *Fisheries Research*, 142: 3–14.

Cope, J.M., Thorson, J.T., Wetzel, C.R. & DeVore, J. 2015. Evaluating a Prior on Relative Stock Status Using Simplified Age-Structured Models. Fisheries Research, 171: 101-109.

Costello, C., Ovando, D., Clavelle, T., Strauss, C.K., Hilborn, R., Melnychuk, M.C., Branch, T.A. et al. 2016. Global Fishery Prospects Under Contrasting Management Regimes. *Proceedings of the National Academy of Sciences* 113(18): 5125–5129.

Dick, E.J. & MacCall, A.D. 2011. Depletion-Based Stock Reduction Analysis: A Catch-Based Method for Determining Sustainable Yields for Data-Poor Fish Stocks. *Fisheries Research*, 110(2): 331–41.

FAO. 2011. *Review of the State of World Marine Fishery Resources*. FAO Fisheries and Aquaculture Technical Paper 569. Rome, Italy.

Froese, R., Demirel, N., Coro, G., Kleisner, K. M. & Winker, H. 2017. Estimating fisheries reference points from catch and resilience. *Fish and Fisheries*, 18: 506– 526.

Gabriel, W.L. & Mace, P.M. 1999. A review of biological reference points in the context of the precautionary approach. In V.R. Restrepo, ed. *Proceedings of the Fifth National NMFS Stock assessment Workshop: Providing Scientific Advice to Implement the Precautionary Approach Under the Magnusson-Stevens Fishery Conservation and Management Act*, pp. 34-45. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-F/SPO-40.

Kimura, D.K., Balsiger, J.W. & Ito, D.H. 1984. Generalized Stock Reduction Analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 41(9): 1325–33.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. & Bell., B.M. 2016. TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, 70 (5).

Martell, S. & Froese, R. 2012. A simple method for estimating MSY from catch and resilience. *Fish and Fisheries*, 14: 504-514.

Melnychuk, M.C., Peterson, E., Elliott, M. & Hilborn R. 2017. Fisheries management impacts on target species status. *Proceedings of the National Academy of Sciences*, 114(1): 178-183.

Monnahan, C.C. & Kristensen, K. 2018. No-U-Turn Sampling for Fast Bayesian Inference in ADMB and TMB: Introducing the Adnuts and Tmbstan R Packages. *PLOS ONE*, 13(5): e0197954.

Osio, G.C., Orio, A. & Millar, C.P. 2015. Assessing the Vulnerability of Mediterranean Demersal Stocks and Predicting Exploitation Status of Un-Assessed Stocks. *Fisheries Research*, 171: 110–121.

Pella, J.J. & Tomlinson, P.K. 1969. A Generalized Stock Production Model. *Inter-American Tropical Tuna Commission Bulletin*, 13(3): 416–97.

Rosenberg, A.A., Kleisner, K.M., Afflerbach, J., Anderson, S.C., Dickey-Collas, M., Cooper, A.B., Fogarty, M.J. et al. 2018. Applying a new ensemble approach to estimating stock status of marine fisheries around the world. *Conservation Letters*, 11(1): p. e12363.

Thorson, J.T. & Cope, J.M. 2014. Catch Curve Stock-Reduction Analysis: An Alternative Solution to the Catch Equations. *Fisheries Research*, 171. 10.1016/j.fishres.2014.03.024.

Thorson, J.T., Munch, S.B., Cope, J.M. & Gao, J. 2017. Predicting Life History Parameters for All Fishes Worldwide. *Ecological Applications*, 27(8): 2262-2276.

Thorson, J.T., Rudd, M.B. & Winker, H. 2018. The Case for Estimating Recruitment Variation in Data-Moderate and Data-Poor Age-Structured Models. *Fisheries Research*, 10.1016/j.fishres.2018.07.007.

Walters, C.J., Martell, S.J.D. & Korman, J. 2006. A Stochastic Approach to Stock Reduction Analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 63(1): 212–23.

Winker, H., Carvalho, F. & Kapur, M. 2018. JABBA: Just Another Bayesian Biomass Assessment. *Fisheries Research*, 204: 275–88.

Workshop, Trawl. 2014. Report of the Workshop on the Use of Best Available Science in Developing and Promoting Best Practices for Trawl Fishing Operations in South and Southeast Asia." Bangkok, Thailand.

APPENDIX 1

List of participants (in alphabetical order by surname)

Francisco ARREGUIN-SANCHEZ
Instituto Politécnico Nacional - IPN · Centro Interdisciplinario de Ciencias Marinas
Mexico

Jilali BENSBAI
Chef URD Exploitation et Gestion Des Pêches
Institut National de Recherche Halieutique (INRH-Casa)
Morocco

Osvaldo Ernesto CHACATE
Instituto Nacional de Investigação Pesqueira (IIP)
Maputo, Mozambique

Fiona DAVID
Research Chair
Minderoo Foundation
Australia

Edward DICK
Institute of Marine Sciences
University of California, Santa Cruz,
USA

Rod FUJITA
Environmental Defense Fund
Washington D.C., USA

Serge GARCIA
Chair
IUCN Fisheries Experts Group (IUCN-CEM-FEG)

Ray HILBORN
Professor
University of Washington
Seattle, USA

James INNES
Organisation for Economic Co-operation and Development (OECD)
Paris, France

Ernesto JARDIM
Joint Research Center, European Commission
Ispra, Italy

Katie LONGO
Impacts researcher
Marine Stewardship Council (MSC)
London, United Kingdom

Pamela MACE
Chief advisor
Ministry of Primary Industries
New Zealand

Stephan MUNCH
Scientist
NOAA, USA

Miguel ÑIQUEN CARRANZA
Investigaciones de Recursos Pelágicos, Neríticos y Oceánicos, Instituto Del Mar Del Peru (IMARPE)
Callao, Peru

Hiroshi OKAMURA
Scientist
National Research Institute of Fisheries Science, Japan Fisheries Research and Education Agency

Dan OVANDO
Scientist
University of Washington
Seattle, USA

Maria Lourdes Deng PALOMARES
Senior Scientist and Project Manager
Sea Around Us Project
Vancouver, Canada

Ana PARMA
Center on Food Security and the Environnent
Argentina

Mohammad J. RAHMAN
Scientist
ECOFISH-Bangladesh Project, WorldFish,
Bangladesh and South Asia
Dhaka, Bangladesh

T.V. SATHIANANDAN
Principal Scientist & Head
Central Marine Fisheries Research Institute
(CMFRI)
India

David TICKLER
Minderoo Foundation
Australia

Paul TUDA
Leibniz Centre for Tropical Marine Research
(ZMT)
Bremen, Germany

Tooraj VALINASSAB
Iranian Fisheries Science Research Institute
Teheran, Iran

FAO

Vera AGOSTINI
Deputy Director
Fisheries and Aquaculture Policy and Resources
Division

Tarub BAHRI
Fishery Resources Officer
Marine and Inland Fisheries Branch
(FIAF)

Manuel BARANGE
Director
Fisheries and Aquaculture Policy and Resources
Division

Pedro BARROS
Senior Fishery Resources Officer
Marine and Inland Fisheries Branch
(FIAF)

Miguel BERNAL
Fishery Resources Officer
General Fisheries Commission for the
Mediterranean (GFCM)

Kim FRIEDMAN
Senior Fishery Resources Officer
Marine and Inland Fisheries Branch
(FIAF)

Nicolas GUTIERREZ
Fishery Resources Officer
Marine and Inland Fisheries Branch
(FIAF)

Edoardo MOSTARDA
Fisheries Information Consultant
Marine and Inland Fisheries Branch
(FIAF)

Hassan MOUSTAHFID
Senior Fishery Resources Officer
Marine and Inland Fisheries Branch
(FIAF)

Marc TACONET
Senior Fishery Officer
Statistics and Information Branch
(FIAS)

Stefania VANNUCCINI
Senior Fishery Officer
Statistics and Information Branch (FIAS)

Marcelo VASCONCELLOS
Fishery Resources Officer
Marine and Inland Fisheries Branch
(FIAF)

Yimin YE
Chief
Marine and Inland Fisheries Branch
(FIAF)

APPENDIX 2

Expert Consultation Workshop on the "Development of the Methodologies for the Global Assessment of Fish Stock Status"
FAO Headquarters – Mexico Room – February 4-6, 2019

Agenda

Day 1: Monday 4 February 2019	

09:00 – 10:30
 	
Chair: Yimin Ye
	Opening and introduction of the expert consultation workshop
	
	Welcome remarks by the Director of FAO Fisheries Policy and Resources Division (Manuel Barange)
	
	- Self-introduction of participants
	- Global stock status assessment: the current FAO approach and need for developing a new methodology (Yimin Ye)
	    • Issues of the existing methods
	    • Challenges to produce country level assessment
	- Objectives of the workshop (Yimin Ye)
	- Discussion: significance of the global assessment and the new requirements

10:30 – 11:00	*Refreshment break*

11:00 – 13:00
 	
Chair: Pedro Barros
	***The new methodology of stock status classification* – Sraplus** *(Dan Ovando/Ray Hilborn)*
	
	- Brief review: the current methods and practices of global stock status classification
	- The new methodology
	    • Model structure
	    • Use of multiple sources of data
	    • Fitting approaches: SIR-based vs PML-based
	- Discussion: weaknesses vs strengthts

13:00 – 14:00	*Lunch*

14:00 – 15:30
 	
Chair: Edward Dick
	Performance of the new model *(Ray Hilborn & Dan Ovando)*
	
	- Cases study – Atlantic cod
	- Measurement of accuracy - RMSE compared with RAM database
	- Effects of different data sources on RMSE
	    • Scientific survey data
	    • CPUE data
	    • Fishing effort data
	    • Expert opinion on trends in abundance
	    • Expert opinion on stock status
	    • Trends in fishing mortality rates from other assessed stocks
	    • Data on local fisheries management
	- Gains and issues in using auxiliary data in stock status classification
	- How to decide on what data to use if multiple sources are available
	- Discussion

15:30 – 16:00	*Coffee break*

16:00 – 17:00
 	
Chair:
Stephan Munch
	Performance of the new model cont. *(Dan Ovando/Ray Hilborn)*
	
	- Use of regional priors
	  • Fishery management index
	  • Fish prices
	  • Regional U/U_{MSY} priors
	  • Swept area ratio
	- Examples demonstrating the effects of different sources of data
	  • European stocks
	  • Asia stocks
	  • East India stocks

17:00	- *Day closure*

Day 2: Tuesday 5 February 2019	
09:00 – 10:30 Chair: Stephan Munch	***Performance of the new model cont.*** *(Ray Hilborn & Dan Ovando)* Discussion: advantages and challenges in using non-stock-specific information/data as priors
10:30 – 11:00	*Refreshment break*

11:00 – 13:00
 	
Chair: Pedro Barros
	Validity and reliability of the model – its application to FAO data *(Ray Hilborn & Dan Ovando)*
	
	- Comparison of results
	  • FAO classification (Tech. Paper 569)
	  • Classification with U/U_{MSY}
	  • Classification with Survey Area Ratios
	  • Classification with Heuristic settings for initial and final years' depletion
	  • Classification with CPUE data
	- Discussion:
	  • Overall performance of the model
	  • Feasibility of incorporating auxiliary data
	  Key issues to overcome for practical application

13:00 – 14:00	*Lunch*
14:00 – 15:30 Chair: Dan Ovandao	***Hands-on trial of* Sraplus** *(all participants)* - Using the example data or your own fishery data - Discussion: model reliability when applied to real fisheries
15:30 – 16:00	*Coffee break*
16:00 – 17:00 Chair: Ana Parma	***Recommendations on the use of the model for global stock status assessment*** *(all participants)* - Evaluation of model performance - Flexibility of using multiple sources of data

	- Use of various priors: informative vs diffuse - How to incorporate local knowledge and informal info/data to improve per model performance? - Immediate revision/changes necessary
17:00	*Day closure*

Day 3: Wednesday 6 February 2019	
09:00 – 10:30 Chair: Serge Garcia	***Selection of the reference list of fish stocks*** *(Yimin Ye)* - How to decide on the reference list for representativeness? - Criteria for selecting species - Consistency issues among countries and between years - Discussion
10:30 – 11:00	*Refreshment break*
11:00 – 13:00 Chair: Ernesto Jadim	***Shared stocks and formally assessed stocks*** *(Yimin Ye)* - Exclusion of shared stocks from the EEZ assessment - Incorporation of formally assessed stocks - Discussion
13:00 – 14:00	*Lunch*
14:00 – 15:30 Chair: Yimin Ye	***General Discussion*** *(all participants)* - Strategies to move forward the global stock status assessment - Recommendations for practical application: country vs regional levels **Conclusions of the consultation workshop** (all participants) - Recommendations for technical refinement and process to roll-out to global level - Plans for follow-up actions
15:30	*Workshop closure*

APPENDIX 3

Structure of the model

By Dan Ovando, Ray Hilborn, Cole Monnahan, Merrill Rudd, and James Thorson

The method, depending on the data made available to the model, approximates existing assessment methods such as Catch-MSY (Martell and Froese, 2012) and simple stock synthesis (Cope 2013). The model was built conceptually around the Stock Reduction Analysis (SRA) ideas first presented in Kimura, Balsiger, and Ito (1984), and extended by works such as Walters, Martell, and Korman (2006) and Dick and MacCall (2011). For this reason, the general tool is called **Sraplus**, where the "plus" refers to the model's ability to easily incorporate multiple types of information. The model itself is built around a Pella-Tomlinson (Pella and Tomlinson, 1969) production model. The key difference between this model and a Schaefer model is that the Pella-Tomlinson method provides flexibility in the ratio of B_{MSY}/K, while a Schaefer model fixes this ratio at 0.5. A surplus production model over an age structured approach was chosen due to the type of data explored in this report which do not contain much information on the age structure of the population. The model estimates up to seven parameters, as well as a vector or process errors equal in length to the length of the time series minus one (since initial depletion is also estimated). The estimated parameters are the initial depletion at the start of the catch history (where depletion is a multiplier of carrying capacity, such that 1 means that the initial biomass is equal to carrying capacity), the intrinsic growth rate *r*, the shape parameter m, carrying capacity k, the standard deviation of the process errors, a catchability coefficient q that links provided abundance indices to estimated population size, and a vector of process errors. Priors for the intrinsic growth rate of the population are pulled from FishLife (Thorson et al. 2017, the most recent version of which contains estimates of intrinsic growth rates), for the species in question if available, and for increasingly broad taxonomic groups as needed (e.g. growth rate for the genus if species level data are not available, and so on). Process error is included to allow for a more flexible model but can easily be turned off as needed.

The model has two methods with which it can provide estimates: a sampling-importance-resampling (SIR) algorithm or maximum likelihood (which can be adapted to an MCMC based method as needed). The SIR algorithm works by running a large number of draws (the default value is 1 million). For each draw, the algorithm tries a set of possible parameter values (e.g. a growth rate and carrying capacity) sampled from prior distributions supplied by the user. Given these draws, the model first projects the fishery using the supplied catch history and the biological operating model and checks whether the given set of catches and model parameters would have crashed the fishery (resulted in negative biomass values) or resulted in U_{MSY} values greater than 1. If neither of these are true, the model then calculates a weighting for each run based on the difference between the final depletion (and if available final U/U_{MSY}) and the prior values of final depletion and U/U_{MSY} supplied by the user. Once these draws are collected, we then resample the draws in proportion to each of their weights, such that draws that do not crash the population and conform closely to the supplied priors are selected more frequently than draws that do not crash the population but also produce values far from the supplied priors. The result then is a "filtering" of the prior values through the criteria of not crashing the population given the observed catch history. This SIR based approach is similar to Martell and Froese (2012) and Walters, Martell, and Korman (2006), with the key difference from catch-MSY (a used in Anderson et al. 2017) being that we allow for different B_{MSY}/K ratios and introduce the ability to include process error. The SIR based approach is a useful tool for interpreting data on life history and local knowledge of stock status in the context of B/B_{MSY} and U/U_{MSY}. As pointed out by Walters, Martell, and Korman (2006), there is no reason that this SIR implementation of stochastic-SRA could not be extended to also include additional data such as indices of abundance, and indeed our SIR implementation of stochastic SRA is capable of including these data. However, the challenge of fitting this model with SIR increases rapidly as the length and complexity of the abundance indices increases. Given that process error can be included, finding the exact combination of process errors and catchability coefficients, along with the other model parameters, that do not crash the population but also fit the abundance index by random sampling alone becomes quite challenging, requiring such a large number of SIR runs in many cases as to become computationally intractable (or at least inconvenient). To resolve this problem, an option to utilize maximum likelihood to estimate the model parameters is provided when some years of an abundance index are available. This estimates the same parameters as the SIR based approach, but treats prior knowledge on life history, initial and final depletion, etc. as penalties to the likelihood rather than priors (though we will refer to them broadly as priors throughout the paper for clarity's sake). Uncertainty around final estimates of B/B_{MSY} and U/U_{MSY} are produced by the delta-method, as implemented in Template Model Builder (TMB, Kristensen et al. 2016). This route for the more complicated model fits was chosen as it produces similar estimate to MCMC based approaches (Monnahan and Kristensen 2018), but generally at much faster speeds, which is a substantial advantage in the context of this project, which requires running thousands of assessments in a reasonable span of time. It should be noted that the model can also be fit using Hamiltonian Monte Carlo if desired. The population growth equation is:

$$f(x) = \begin{cases} B_{t+1} = \left(B_t + B_t \frac{r}{m-1}\left(1 - \left(\frac{B_t}{K}\right)^{m-1}\right) - \hat{c}_t\right) p_t, & \text{if } B_t > 0.25 \times K. \\ B_{t+1} = \left(B_t + \frac{B_t}{0.25 \times K}\left(B_t \frac{r}{m-1}\left(1 - \left(\frac{B_t}{K}\right)^{m-1}\right) - \hat{c}_t\right)\right) p_t, & \text{otherwise.} \end{cases} \quad (2)$$

Where B_t is biomass at time t, r is the intrinsic growth rate, m is the scaling parameter that allows for the ratio of B_{MSY}/K to shift (when m is 2 the $B_{MSY}/K = 0.5$ and the model acts as a Schaefer model, lower values of m shift the production function left, higher values right), c is catch, and p is process error. The Pella-Tomlinson model does have an issue where growth rates can become unrealistically large when the population reaches low sizes. This problem is dealt with by following the methods described in Winker, Carvalho, and Kapur (2018) to reduce the production of the population when it falls below a threshold of 25% of carrying capacity. The choice of 25% is arbitrary and a sensitivity analysis on this choice should be evaluated. Incorporation of process errors is useful for two reasons: (1) when an abundance index is available, process errors can reduce bias arising from lack of fit in a deterministic SRA whenever dynamics are poorly explained by catch-history alone, and (2) with or without an abundance index (or other info), the stochastic portion is necessary to get good uncertainty intervals (i.e., with close to nominal coverage, see Thorson, Rudd, and Winker (2018)). When the SIR method is used the supplied catch history is employed. However, for the maximum likelihood approach, convergence is greatly facilitated by estimating catches ($\hat{c}_t$) rather than using the observed catches directly. $\hat{c}_t$ is given as:

$$\hat{c}_t = u_t B_t \quad (3)$$

where u_t is estimated through:

$$u_t = \frac{1}{1 + e^{-u_t^*}} \quad (4)$$

Equation (4) ensures that u_t is between 0 and 1, while the model estimates u^*_t to improve performance. $\hat{c}_t$ is then fit to the observed catches through the likelihood:

$$log\left(c_t\right) \sim normal\left(log\left(\hat{c}_t\right), 0.05\right) \quad (5)$$

where an observation coefficient of variation of 5% is assumed. Catch is estimated rather than directly using the observed catch in order to improve model performance. As priors on depletion, or indices of abundance, pull the estimated population closer to zero, there is a greater and greater chance that the observed catch history will result in a collapse of the population (catch in time t being greater than biomass in time t); It is much harder to find parameters that produce a final depletion of 1% of carrying capacity while not crashing the population prior to then given a catch history, than it is to produce a final depletion of 99% of carrying capacity. Restrictions and penalties on population crashes can be imposed, but this distorts the gradients of the model, causing the model to often become stuck once a few crashes have occurred. Since u_t is constrained between 0 and 1, $\hat{c}_t$ will always be "viable", and by then fitting to the observed catches, the model has an easier time finding parameters that produce a near match to the observed catches while not crashing the population. As of now, none of the data used in this exercise have missing years of catch data, but the model can handle those cases in the future due to the random walk priors imposed on u_t detailed below.

In order to help separate out observation error in the catch and process error in the population model, when catch is being estimated fishing mortality u_t is assumed to follow a random walk on average, assigning each u_t a prior of:

$$log(u_t) \sim normal(log(u_{t-1}), 0.1)) \tag{6}$$

Process error is also allowed (in the manner of the stochastic stock reduction analysis error suggested by Walters, Martell, and Korman 2006). This allows the population dynamics to deviate from the exact values given by the Pella-Tomlinson operating model, while still conforming to the assumptions of this model on average. Process error p is assumed to be log-normally distributed, such that**:**

$$p_t \sim e^{normal(0,\sigma_{proc})} \tag{7}$$

The mean of p is $-0.5*\sigma_{proc}^2$. It is critical to note that as previously discussed, without an abundance index or some years of U/U_{MSY} data, this method is more or less a transformation of the supplied priors: The B/B_{MSY} values produced are a direct function of the supplied priors (with some slight fluctuation allowed since we allow the ratio of B_{MSY}/K to shift in the model). While this approach cannot be viewed as data-driven estimate of stock status, this approach does allow users to interpret local knowledge around life history and relative depletion in terms of B/B_{MSY} and U/U_{MSY}. The accuracy of these estimates will depend on the accuracy of the supplied priors. Beginning at the most data-limited version of the model, when the only information made available is some prior information on initial and final depletion of the fishery, the model approximates Catch-MSY, in that there is not a true likelihood, but instead the model filters out combinations of life history values that would crash the population or produce final initial and final depletion that fall far from the supplied values. The model allows for these initial and final depletion to either be specified by the user, passed from regression models based on variables such as catch histories, fisheries management index values, or trawl footprints, or supplied by the catch heuristics underpinning the version of Catch-MSY utilized in Anderson et al. (2017) (if catch in the first year is less than 20% of maximum catch, initial depletion is assumed to be between 50% and 90% of carrying capacity, otherwise it is assumed to be between 20% and 60% For final depletion, the heuristic assumes that if final catch is greater than 50% of max catch, final depletion is between 30%-70%, 1%-50% otherwise). The initial and final depletion priors are given by equations (8)-(9).

$$log\left(\frac{b_{t=0}}{k}\right) \sim normal\left(log\left(d^*_{t=0}\right), \sigma_{d^*_{t=0}}\right) \tag{8}$$

$$log\left(\frac{b_{t=T}}{k}\right) \sim normal\left(log\left(d^*_{t=T}\right), \sigma_{d^*_{t=T}}\right) \tag{9}$$

The model to this point is an analogue to Catch-MSY. Adding a bit more information, the catch history and prior estimates of initial and final depletion can be augmented with priors on U/U_{MSY} for an arbitrary number of years (e.g. just in the final year, or up to all years), they can enter through a prior ($U/U_{MSY}\hat{}*$) on the U/U_{MSY} values estimated by the model. When additional data are available, the priors on initial and final depletion can be either turned off or left if there is a feeling that they contribute to the model fit (e.g. if the index data are suspect and robust local knowledge on final depletion is available).

$$log\left(\frac{u_t}{u_{msy\,t}}\right) \sim normal\left(log\left(\frac{u_t}{u_{msy\,t}}^*\right), \sigma_u\right) \tag{10}$$

$\sigma d^*_{t=0}$, $\sigma d^*_{t=T}$, and $\sigma_u$are all specified by the user, allowing them to assign weights to different priors. The default value for each of these is 0.2.

From there, the model also allows for the input of index data, which could be obtained through fishery independent research surveys, or through catch-per-unit-effort (CPUE) data. The given index enters the model through the likelihood:

$$log\left(index_t\right) \sim normal\left(log\left(qB_t\right), \sigma_{obs}\right) \tag{11}$$

Abundance indices can be supplied for any subset of the years of available catch data, though their usefulness will depend on having sufficient years with contrasting index levels. It is also important to note that the model does not, as of now, integrate standardization of the CPUE index (see Maunder and Punt 2004 for a summary of CPUE standardization). As such, any indices of abundance must be assumed to be standardized, either statistically or by only using indices from consistently collected means (e.g. a consistent research survey of fishery dependent CPUE from the same type of vessels fishing in the same manner across the same region). The model developed for this project is a flexible tool that at its most data-limited approximates Catch-MSY and at its most data-rich fits a surplus production model given index data and priors on U/U_{MSY}. No individual component of this model is unique per-say. Specific combinations of the kinds of data discussed here could be passed to LIME (Rudd and Thorson 2017), simple stock synthesis (Cope 2013), JABBA or JABBA-select (that allows users to correct for biases arising from changes in fishery selectivity, as well as selectivity mis-matches between fisheries and surveys, Winker, Carvalho, and Kapur 2018), CCSRA (Thorson and Cope 2014), and any number of other available assessment packages. As we develop a better sense of what types of data are truly going to be available and used by FAO, we may choose to utilize one of these more mature packages. However, the tool used here allows us to quickly examine the differences in stock status produced by a wide range of available data, making it an ideal starting place for this project.

APPENDIX 4

Sources of Information

By Dan Ovando, Ray Hilborn, Cole Monnahan, Merrill Rudd, and James Thorson

This section describes how data from the Fisheries Management Index (FMI) (as presented by Melnychuk et al., 2017) is transformed into priors on recent U/U_{MSY} and/or B/B_{MSY} values. It is critical to remember that in the absence of an index of abundance or some other form of data to fit to, any assessment will simply be a reflection of the priors. So, if the priors are passed on initial and final depletion to the most data-limited SIR version of this model, the actual assessment is the process by which the priors are developed. For example, in the case of the default Catch-MSY heuristics, the true assessment, in terms of stock status, are the catch heuristics. The Catch-MSY process simply helps us interpret this assessment in more useful units. Similarly, though, in the absence of other data if final depletion is estimated from Fisheries Management Index data, the process of translating FMI to final depletion as should be thought as the assessment itself, and **Sraplus** as a process for placing this assessment in biological context.

Fisheries Management Index

The Fisheries Management Index (FMI) utilizes surveys filled out by regional experts to score a fishery against a set of 46 specific questions for individual species about what elements of fisheries management were in place. These questions are then aggregated into broader categories of science, enforcement, management, and socioeconomics. The higher the score, the better the expert judges that a given metric is met in that fishery. Importantly, FMI surveys can be filled out in the absence of stock assessments. This allows us to explore how FMI values map onto stock status and explore the ability then to use FMI scores to produce priors on stock status for unassessed fisheries (in a manner similar to Osio, Orio, and Millar (2015) and Cope et al. (2015)). Looking at the raw data, there appears to be a negative correlation between mean FMI across the four metrics and U/U_{MSY}, and a positive correlation for B/B_{MSY}, where reference points were drawn from the RAM legacy stock assessment database (Fig. 1).

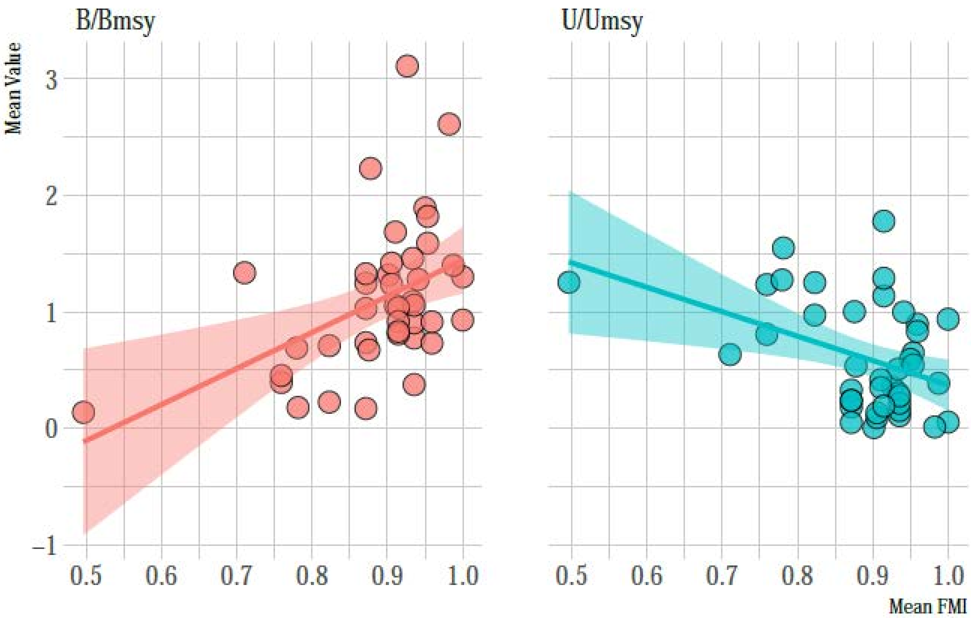

Fig. 1. Mean FMI scores plotted against B/B_{MSY} and U/U_{MSY} for individual RAM stocks. Solid colored line shows a linear model of FMI and status, shaded area are 95% confidence intervals around the mean trend

For all cases, only references points over the last five years of the fishery are included, to try and constrain stock status values to the time period covered by the FMI values. This visual relationship is clearly highly dependent on the small cluster of low FMI fisheries. And, even for the better-represented ranges of the FMI, a wide range of status estimates are possible given the same FMI. This visual relationship can be explored more deeply using a simple model:

$$log(\frac{u}{u_{msy}}) = normal(\beta_0 + \beta_1 research + \beta_2 management + \beta_3 research + \beta_4 socioeconomics, \sigma) \quad (15)$$

The outcome (U/U_{MSY}) is log transformed here as the mean and standard error of the predicted $log(U/U_{MSY})$ values are what would be fed as a prior to the assessment model. The model was fit as a Bayesian GLM using the `rstanarm` package in R. Since the end goal of this model is to use predictions from the estimated regression to provide priors for the assessment model, the primary interest is in the predictive power of the model (as opposed to inference around model coefficients). The predictive power can be examined by using posterior predictive distributions and R^2 values. The "posterior predictive" from a Bayesian model provides a measure of the uncertainty around our prediction for a given observation (as opposed to the uncertainty around the fit to a given data point, which will be much smaller).

$$y_{pp} \sim normal(\beta X, \sigma) \quad (16)$$

By generating numerous posterior predictive distributions for each predicted point, distributions of R^2 values between the posterior predicted point and the true observed value can then be generated. This gives a sense of the predictive skill of the model. Again, the focus is on predictive R^2 here since the interest is in predictive skill, not in the precision of our estimated model coefficients (acknowledging that there are clear risks to considering R^2 values). Following discussions with co-authors, any values from the FMI/status database with U/U_{MSY} less than 0.13 and an FMI enforcement score greater than 0.6 were excluded. The model is able to identify some relationships between FMI attributes and U/U_{MSY}, but the predictive ability is low (Fig. 2-A). Each blue point in the graph shows the U/U_{MSY} value from RAM on the x-axis, the mean posterior predictive value on the Y. The black dashed line in Fig. 2-A is the 1:1 line, and the solid blue line is a simple linear model fit between the observed and predicted points. The fitted line between the observed and predicted points follows the 1:1 line, meaning that on average the model is unbiased, but the large spread of the points around the 1:1 line show that the prediction for any one point is highly error prone (reflected in the relatively low mean R^2 values in Fig. 2-B, which tells us that on average the model only accounts for approximately 20% of the variance in U/U_{MSY}). The unbiased nature of the predictions though suggests that our prediction for the mean $log(U/U_{MSY})$ across multiple fisheries for a given country may be reasonably accurate. The data used in the regression present another challenge in using FMI to predict U/U_{MSY} though. Examining the observed U/U_{MSY} values (y axis of Fig. 2-A), very few values have U/U_{MSY} values above 1, reflective of recent low exploitation rates across many stock assessed fisheries. This presents a lack of contrast though. The data have a broad range of FMI values, but those FMI values are associated with a small range of U/U_{MSY} values. As such, the model favors predicting relatively low U/U_{MSY} values.

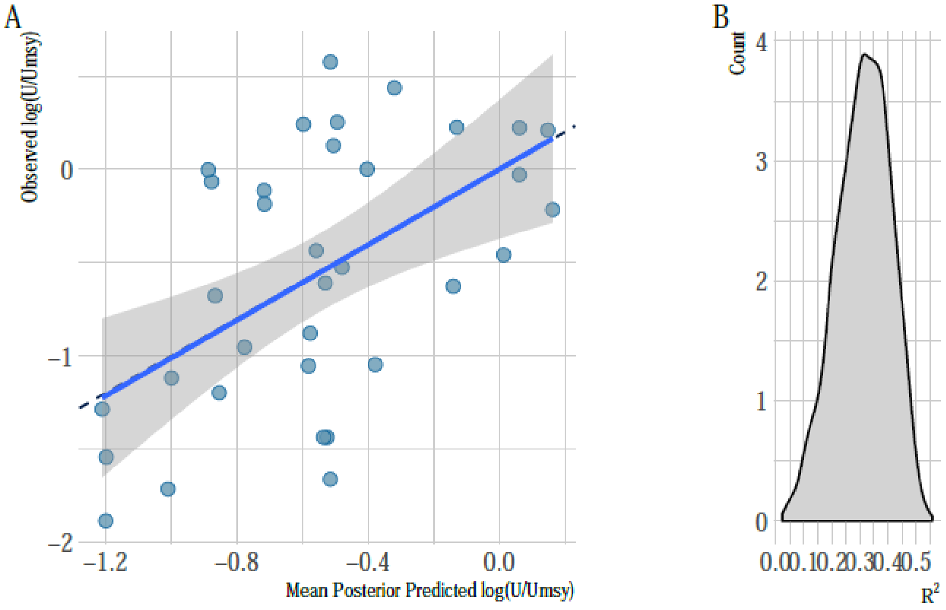

Fig. 2. Mean posterior predicted U/U_{MSY} vs observed U/U_{MSY} for regression of Fisheries Management Index components against U/U_{MSY} (A), and distribution of posterior predictive R^2 values (B)

Examining the marginal effects of the FMI components, the correlations between FMI measure and U/U_{MSY} is clearest for enforcement and socioeconomics attributes (better enforcement/socioeconomics associated with lower U/U_{MSY}), while research and management have little clear relationship (once the other FMI components are controlled for) (Fig. 3).

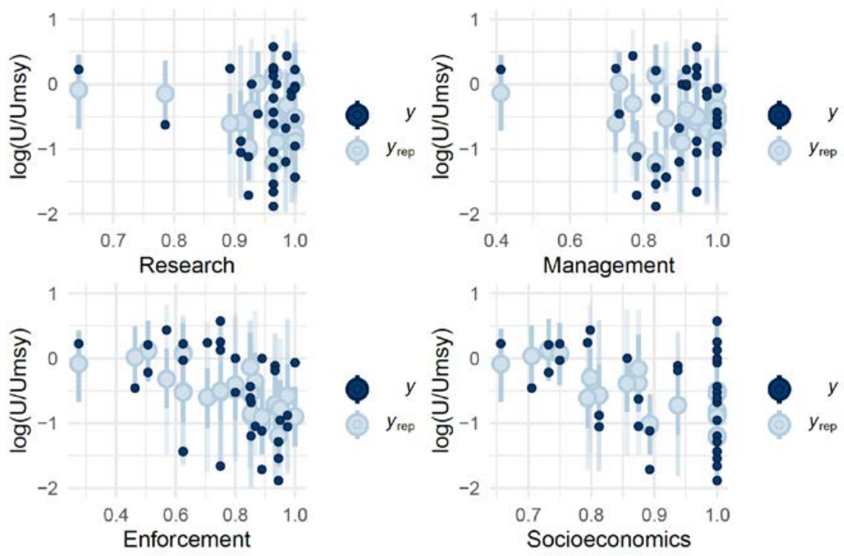

Fig. 3. FMI metric (x axes) vs fitted (dark blue points) and posterior predicted distributions (light blue intervals) of U/U_{MSY} on the y axes (A), and distribution of posterior predictive R^2 values (B).

Better results were obtained repeating the same exercise but now looking to predict log(B/B~MSY) as a function of FMI. The overall relationship was still unbiased, but the model now explains much more of the variance in B/B_{MSY} (mean posterior predictive R^2 above 0.5, Fig. 4). The marginal effects of management, enforcement, and socioeconomics were all positive, with the relationship between higher management and higher B/B_{MSY} being especially clear (Fig. 5). One important consideration here though is that it is unclear how independent the management scores are from the B/B_{MSY} values reported in RAM. If survey respondents are aware of the B/B_{MSY} values in RAM and allow their knowledge of high B/B_{MSY} values to affect their scoring of the “management” attribute for example, this could inflate our perception of the ability of “management” to predict B/B_{MSY} in places without assessments.

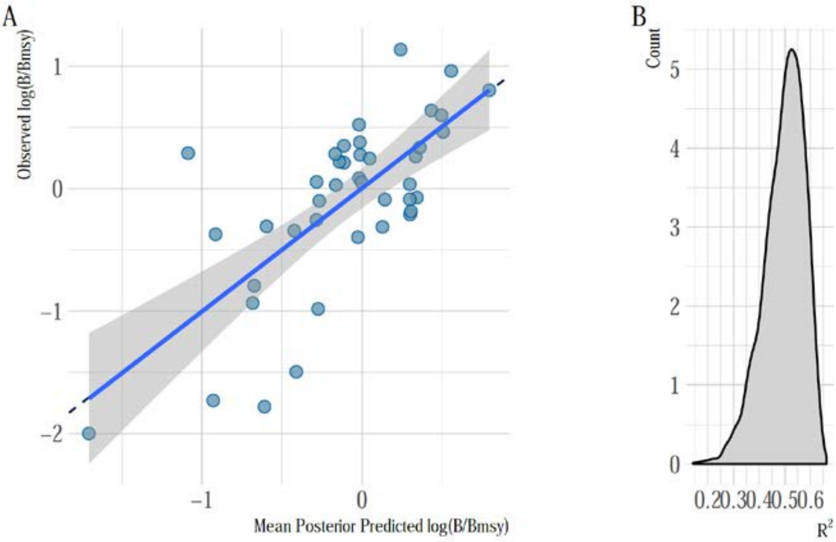

Fig. 4. Mean posterior predicted B/B_{MSY} vs observed B/B_{MSY} for regression of Fisheries Management Index components against U/U_{MSY} (A), and distribution of posterior predictive R^2 values (B).

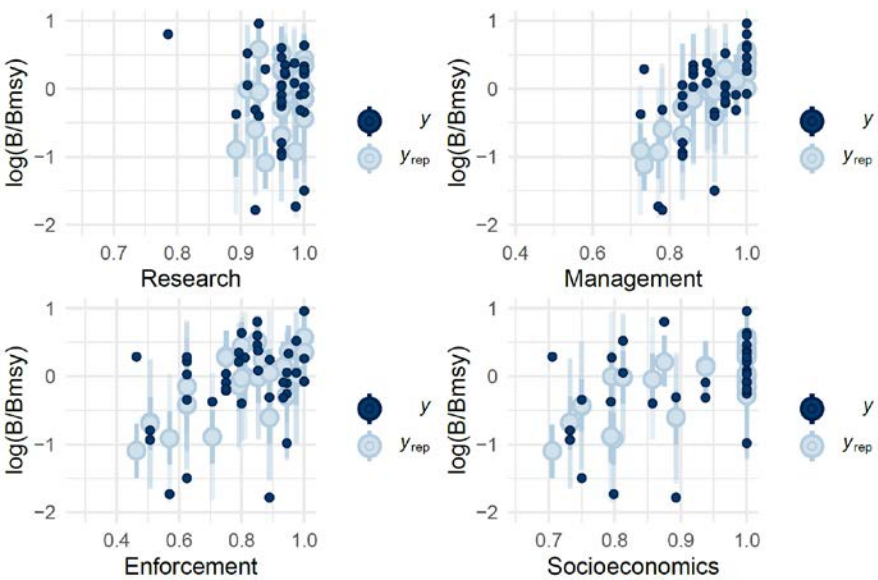

Fig. 5. FMI metric (x axes) vs fitted (dark blue points) and posterior predicted distributions (light blue intervals) of B/B_{MSY} on the y axes (A), and distribution of posterior predictive R^2 values (B).

APPENDIX 5

Case study

By Dan Ovando, Ray Hilborn, Cole Monnahan, Merrill Rudd, and James Thorson

We can illustrate **Sraplus**'s capabilities using a simple case study demonstrating different outcomes in estimated stock status using different kinds of data and assessments. We used both Catch-MSY, as used in Anderson et al. (2017), and **Sraplus** to estimate stock status over time for Atlantic cod in the North Sea as an example fishery. We then ran a number of "experiments" supplying each assessment method with different types of data. For the "heuristic" experiment, we use the same catch-based heuristics as Catch-MSY to specify bounds on initial and final depletion. These heuristics are that if catch in the first year is less than 20% of maximum catch, initial depletion is assumed to be between 50% and 90% of carrying capacity, otherwise it is assumed to be between 20% and 60% For final depletion, the heuristic assumes that if final catch is greater than 50% of max catch, final depletion is between 30%-70%, 1%-50% otherwise. For the "bookends" experiment, we simply use true values of initial and final depletion (as supplied by RAM), with a CV of 20%. The "bookends" is intended to reflect a scenario where robust local expert opinion on stock status is available. For the "survey" experiment, we approximate an accurate abundance index available for the time series of the fishery, intended to reflect a fishery-independent survey available in a region (drawn from RAM as simply total biomass times 1e-3). Lastly, the "swept area" experiment is inspired by Amoroso et al. (2018) and utilizes data on U/U_{MSY} across the time series, as well as a CPUE index created from dividing the RAM catch by RAM U/U_{MSY}. These experiments allow us to examine how the estimates provided by **Sraplus** differ depending on the kinds of data made available to the model, and how the results contrast to Catch-MSY. We sample all data from RAM without observation error, as the purpose here is simply to illustrate how different models and different data produce different outcomes.

Catch-MSY and the bookends/heuristic experiments both capture the overall trend in initial and final B/B_{MSY} (and interestingly in this particular example the Catch-MSY heuristic almost perfectly mirrors the "true" values from RAM). However, they, both of these scenarios miss some of the dramatic increases in abundance through the 1970s. Both the swept-area and survey methods do a much better job at estimating the complete time series of B/B_{MSY}. All methods correctly capture the recent increase in B/B_{MSY} reported in RAM, but none are able to capture how 374 rapidly B/B_{MSY} is reported to have increased in RAM. For the bookends and the heuristics example, the more flexible **Sraplus** (with process error and the ability flexible B_{MSY}/K ratio), appears to do a worse job of estimating B/B_{MSY} than Catch-MSY. This is a useful example of why increasing model flexibility does not automatically improve all estimates. Remember that in the case of bookends and heuristics, the priors passed to the model are on initial and final depletion, not on B/B_{MSY}. In the absence of any other data, there is no real information that would allow the model to update its priors on either the B_{MSY}/K ratio, or on the process errors over time, so long as the priors on initial and final depletion are met (which they are in this case). So, in this case, the default B_{MSY}/K ratio of 0.5 used by the Catch-MSY example here appears to produce a better estimate of B/B_{MSY} when the model is tuned to depletion levels, as opposed to the **Sraplus** version whose flexibility allows it to reflect the depletion priors while providing less accurate estimates of B/B_{MSY} (Fig. 1). Further analysis can be done to assess how often this is the case.

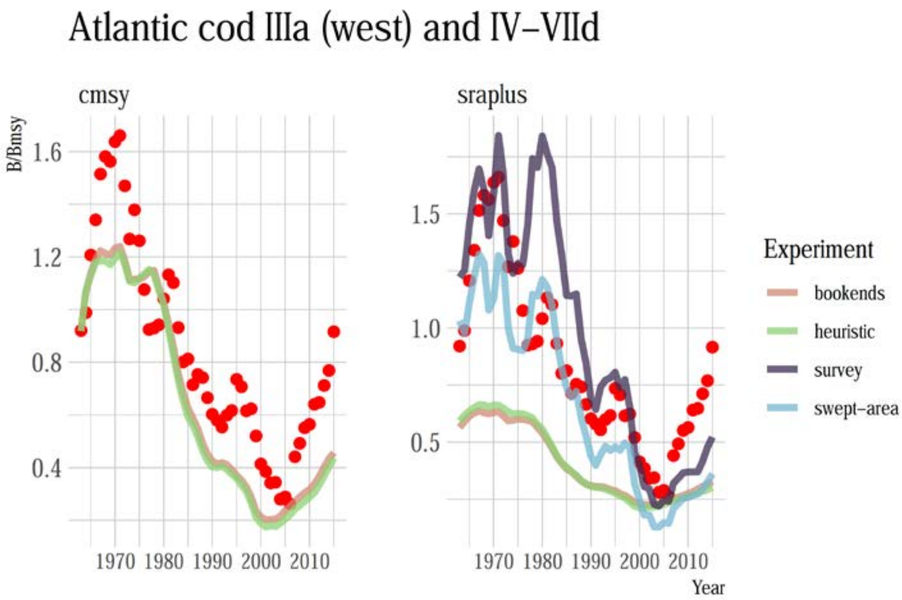

Figure 1. Case study experiment results. Red points are reported values from RAM. Solid colored lines are mean estimated B/B_{MSY} values for each of the experiments for both Catch-MSY and **Sraplus**.

APPENDIX 6

Value of information

By Dan Ovando, Ray Hilborn, Cole Monnahan, Merrill Rudd, and James Thorson

Sraplus makes it relatively simple to incorporate many different kinds of data. Given finite resources, which sources of data would provide the highest probability of reducing uncertainty around stock status if investment was made in a globally curated database? To address this question, different kinds of data were sampled from the RAM database, and then used to estimate stock status over the last five years of available data (though this can easily be extended to shorter or longer time periods). For example, for a run in which the model is fit to an abundance index, an abundance index is created from RAM (approximated as the total abundance from RAM times a constant catchability coefficient) and passed to the estimation model. If the run calls for a sporadically sampled abundance index over the last five years, sampling can be done only every two years of the last ten years of the generated abundance index. U/U_{MSY} values can also be sampled from RAM from any portion of the catch history (though for this analysis only the effect of having an estimate of U/U_{MSY} in the final year of the catch series is tested), along with depletion or B/B_{MSY} in the initial and/or final year of the catch series. All data can be sampled with or without error (where error is defined by a CV, and the sampled data is data ~ data $\times\ e^{norm(0,CV)}$).

Value of information is judged by calculating the Root Mean Squared Error (RMSE), as:

$$rmse = \sqrt{\frac{\sum_{t=1:N}\left(observed_t - predicted_t\right)^2}{N}} \qquad (12)$$

Where N encompass both simulation replicates and number of years. If only RMSE for the final year of the time series is calculated, and each model run produces 2000 successful draws of life history values, then N is 2000. If the final five years are used (still with 2000 simulation replicates), then N is 10,000. The default window is five years, in order to capture the ability of the model to reflect the overall trend in the metric in question over the last five years. An estimate that correctly fits the last five years will perform much better than an estimate that gets the final year correct but misses the mark wildly in the previous four years. It is important to note that RMSE does not capture bias: A model could be completely unbiased (on average correct), while still having a very large RMSE. Subsequent analyses can also examine the effect of different kinds of data on bias.

APPENDIX 7

Applications

By Dan Ovando, Ray Hilborn, Cole Monnahan, Merrill Rudd, and James Thorson

The contrast in status estimated for European and Asian stocks, when the effort data presented in Bell, Watson and Ye (2017) are considered, followed by estimates of stock status along eastern India based on catch heuristics, CPUE, and swept area ratio, are presented

European and Asian Stock Status

One of the questionable results of earlier attempts at global stock status has been that regions such as Asia and Europe have been estimated to have near identical stock status, despite expert opinion, and what stock assessments are available, suggesting otherwise. Part of the problem here is that simply looking at the in the overall catch data, Asia and Europe do not appear to be greatly different (Fig. 1). Both show a steady increase in catches, a slight drop and then plateau. Methods based on catch heuristics then are likely to provide nearly identical stock status estimates for both regions. However, looking at the effort data compiled by Bell, Watson, and Ye (2017), the two regions have very different effort histories, with Europe showing sharp declines in effort in recent years, while Asia has seen a steady increase in fishing effort. Simply dividing the total catch by the effort produces a crude CPUE index that now paints starkly different pictures between the two regions (Fig. 1).

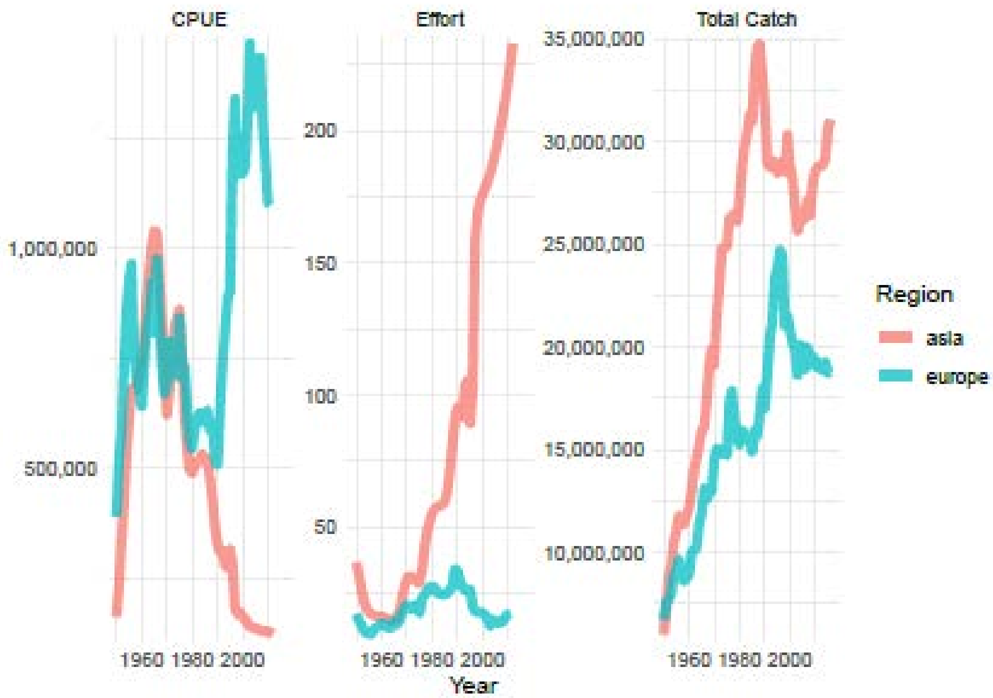

Fig. 1. CPUE, effort, and total catch across rough measures of Asia and Europe.

To explore these differences on a stock-by-stock basis, two sets of assessments on each of the stocks in the region with cumulative lifetime catches over 10,000 tons were run. In order to better match the effort data, a "stock" was defined by summing all of the capture statistics for a given taxonomic group within a large region (e.g. "Asia"). So, under this assumption there is only one stock for each species within the "Asia" region, as opposed to separate stocks for each country within that region. Given these catch histories, the SIR based stochastic-SRA mode of **Sraplus** was used to produce B/B_{MSY} and U/U_{MSY} values using the default catch heuristics of Catch-MSY (as implemented by Anderson et al. (2017)). The catch histories of each of these "stocks" were then divided by the effort trend reported by Bell, Watson and Ye (2017) for that region to create a CPUE index for each stock, which was treated as an index of abundance and passed to our estimation model.

In order to account somewhat for changes in technology, only post-1980 effort data from Bell, Watson and Ye (2017) was included in the CPUE index. Assessed status for Asian and European stocks are presented in Fig. 2. Using the catch heuristics, the two regions are estimated to have nearly identical stock status distributions. In addition, the clear bimodality of the catch heuristic can be seen in that stocks always are estimated with a B/B_{MSY} of ~ 1.5 or of ~ 0.5, though the associated U/U_{MSY} values can vary widely. However, incorporating the "CPUE" data creates starkly different pictures of the two regions. The assessed Asian stocks are now estimated to be mostly in the upper left quadrant (overfished and overfishing), while European stocks are clustered to the lower right (Fig. 2).

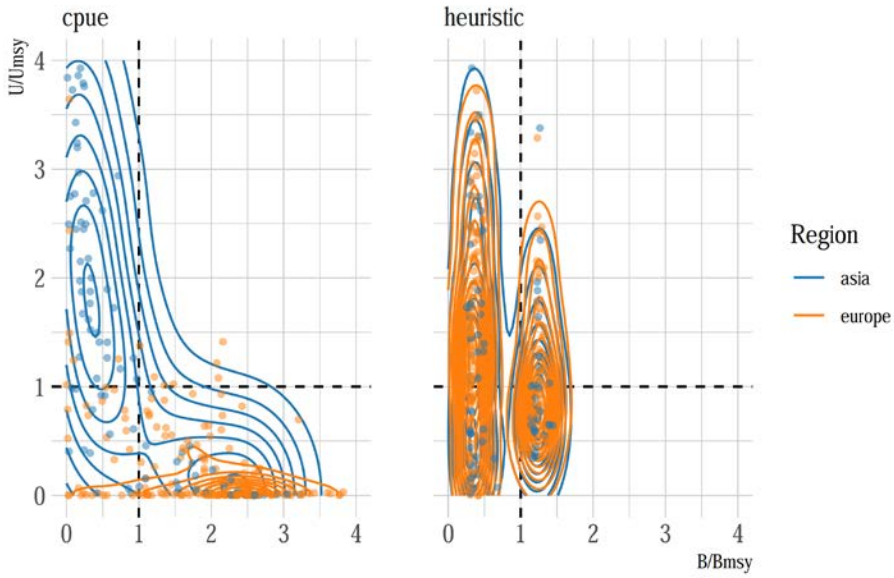

Fig. 2. Stock status plot of mean status (B/Bmsy and U/Umsy) in 2016 using the CPUE data approximated from Bell, Watson and Ye (2017) (left panel) and the Catch-MSY catch heuristic (right panel)

Substantial work remains in appropriately pairing the right effort data with the right catch data, but this demonstrates how inclusion of rough effort data available at a global scale can help differentiate status of two regions with similar catch histories. When paired with local catch histories, the European-wide effort series from Bell, Watson, and Ye (2017) suggests that abundances have been increasing dramatically in recent years. Since we only used post 1980 effort data, the model tends to estimate that stocks were depleted prior to 1980, and then experienced rapid rebuilding over the last 30 years, in order to match the CPUE based abundance index. This results in the implausibly high estimated B/B_{MSY} values for many of the European stocks. This exercise then both demonstrates the stark differences in stock status that inclusion of effort data can provide (that would be masked by use of catch heuristics), but also highlights the need to carefully construct appropriate abundance indices for the fisheries in question.

East India

The previous example focuses on using broadly available databases (e.g. global catch and effort statistics). However, local databases are often available with a little bit of digging. For example, work with collaborators has uncovered data on CPUE by trawlers in East India. Pairing these data with FAO catch statistics from the same region, a similar exercise to the Asia/Europe contrast presented in Fig. 2 can be repeated, except now with more locally tailored CPUE data. A local estimate of swept area ratio of over 18 (which is capped at 4 here to make the case of an extremely high prior on U/U_{MSY}, since values of 18 produce posterior predictive U/U_{MSY} values near 20) as a prior on recent U/U_{MSY}, per the SAR to U/U_{MSY} regression estimated in this report (and built off of Amoroso et al. (2018) can be used. All together then estimates of stock status in Eastern India are compared using catch heuristics, SAR as a prior on recent U/U_{MSY}, and SAR plus a CPUE index. Only stocks in the upper 25% percentile of total landings in the region are included to remove some erratic catch histories of very small fisheries. For the cases where SAR provides a prior on terminal U/U_{MSY}, this prior is used to update the prior on terminal B/B_{MSY} as well (as opposed to leaving it to the same prior specified by the catch heuristic). We set the prior on terminal B/B_{MSY} to the equilibrium B/B_{MSY} that would be produced by the SAR derived U/U_{MSY} prior at equilibrium. Using catch heuristics to estimate stock status in East India produces a very optimistic picture of fisheries in the region, with most fisheries estimated as having a B/B_{MSY} over 1. The high SAR suggests a very high prior on recent U/U_{MSY} and use of the SAR data as a prior on U/U_{MSY} indeed paints a more pessimistic picture of regional stocks. Including both SAR and the CPUE series since 1990 produces yet another picture, with most stocks experiencing overfishing (mostly due to the SAR values), but a greater spread in B/B_{MSY} values (Fig. 3).

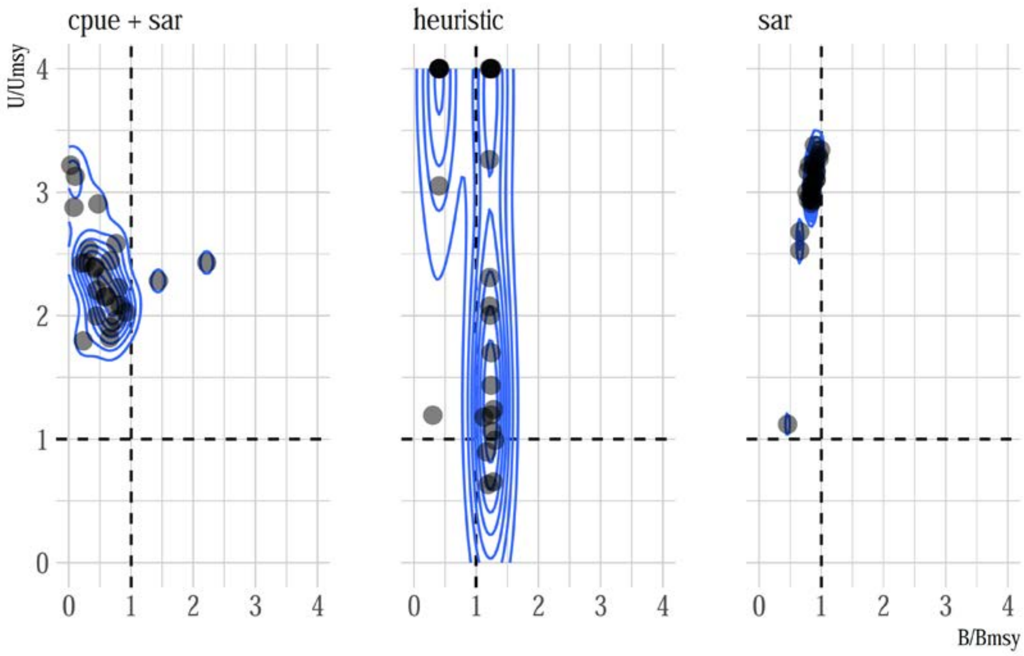

Fig. 3. Kobe plots of recent status in East India. Heuristic panel uses the Catch-MSY heuristic to produce status points, cpue panel uses a local CPUE estimate as an index for overall abundance in the region, that the model is fit to.

This exercise again demonstrates the sensitivity of regional estimates to the data made available to it, and the need for local experts to identify which data are most likely to be informative for a given stock. This exercise also demonstrates some of the challenges of utilizing locally available data. The CPUE data presented for East India shows a low period in the 1980s, followed by a sharp increase through 1990, followed by a crash and stabilization (Fig. 4).

Fig. 4. Reported catch rates of trawlers in Eastern India (Workshop 2014)

Taking these data at face value as abundance indices, this suggests that the stock in the 1990s was roughly four times as abundant as in the early 1980s or recent years. It is possible that the stock had been heavily exploited in the 1980s, followed by a recovery and then increase pressure. Or that the targeted stocks experienced a recruitment boom in the years leading up to 1990. However, many other (more plausible) explanations exist, from shifts in fishing skill in technology, to discovery of new fishing grounds, to shift in the species composition of the CPUE. Interpreting the effects of these shifts in the CPUE as shifts in abundance can lead to highly misleading results. Addition of effort data will often be an improvement over use of catch statistics alone, but clearly blind use of CPUE indices carries great risk. The contrasting outcomes between catch-only, CPUE, and SAR indicate that careful work will be needed in many cases to determine which data are most likely to be truly informative for a region, since there is no guarantee that additional data will tell the same story. For now, we will only include the CPUE data post 1990 in the estimation model, under the assumption that the pre-1990 changes in CPUE reflect technology, skill, or some other non-abundance related confounder of the CPUE data.

APPENDIX 8

Application to FAO Fisheries and Aquaculture Technical Paper 569

By Dan Ovando, Ray Hilborn, Cole Monnahan, Merrill Rudd, and James Thorson

One of the goals of this project phase was to test the validity and reliability of this method with the stocks listed in FAO Fisheries and Aquaculture Technical Paper 569 (FAO 2011). To accomplish this, each of the stocks presented in report number was paired with various forms of alternative data, including:

- cpue: a CPUE index created from continental effort series
- ram_u_umsy: a time series of mean RAM U/U_{MSY} values in the major FAO statistical area
- fmi: aggregated posterior predicted U/U_{MSY} and B/B_{MSY} from FMI regression for all fisheries in that FAO area
- sar: posterior predicted U/U_{MSY} from SAR regression (assumed to be 0.2 for FAO area 67, 4 for FAO area 57)
- cpue_plus: a CPUE index created from continental effort series combined with the posterior predicted distributions from the FMI and SAR regressions.

Note that this is a proof of concept, and much more detailed work would be required to carefully pair each of these data sources with the appropriate stock. For visual clarity, only estimates are presented for two regions at the moment: the Pacific Northeast (FAO area 67), and the Eastern Indian Ocean (FAO area 57). FAO (2011) assigns stocks in this region to different broad status categories of underfished, maximally sustainably fished and overfished. Each of these different types of information was used within **Sraplus** to assess each of the stocks in area 67 and 57 included in FAO (2011) and the percent of stocks classified as “overfished” was compared using different kinds of data. Similar to the East India example, for the FMI and SAR cases, the prior on terminal B/B_{MSY} was modified to reflect the estimated B/B_{MSY} that the prior on U/U_{MSY} would produce at equilibrium. The results show stark differences in classified status across regions depending on what data are used. for the Northeast Pacific region (FAO area 67), FAO (2011) classified 9% of stocks as overfished. Using regional U/U_{MSY} from RAM in the region as priors, the **Sraplus** model estimates slightly over 5% overexploited. Using the SAR data (assumed to be 0.2) or the FMI data produces an estimate of 0% of stocks overfished in the Pacific Northeast. Using the CPUE data along produces an estimate of over 25% overfished in the region, due largely to the increasing trend in efforts reported in Bell, Watson and Ye (2017), and reflecting the problem of using effort data that do not match local dynamics. However, pairing the inaccurate CPUE trend with the SAR and FMI data (cpue_plus) lowers the estimated percent overfished to 10%. In stark contrast, use of the catch heuristics estimates that over 50% of the stocks in the Pacific Northeast are overfished. For the Eastern Indian Ocean region, SAR and FMI data produced an estimate of roughly 100% stocks being overfished, while the heuristic and CPUE data produce estimates closer to the 25% estimated by FAO (2011) (Fig. 1). No local stock assessments are available for the Eastern Indian Ocean, so no estimates are provided for the area by the u_umsy data.

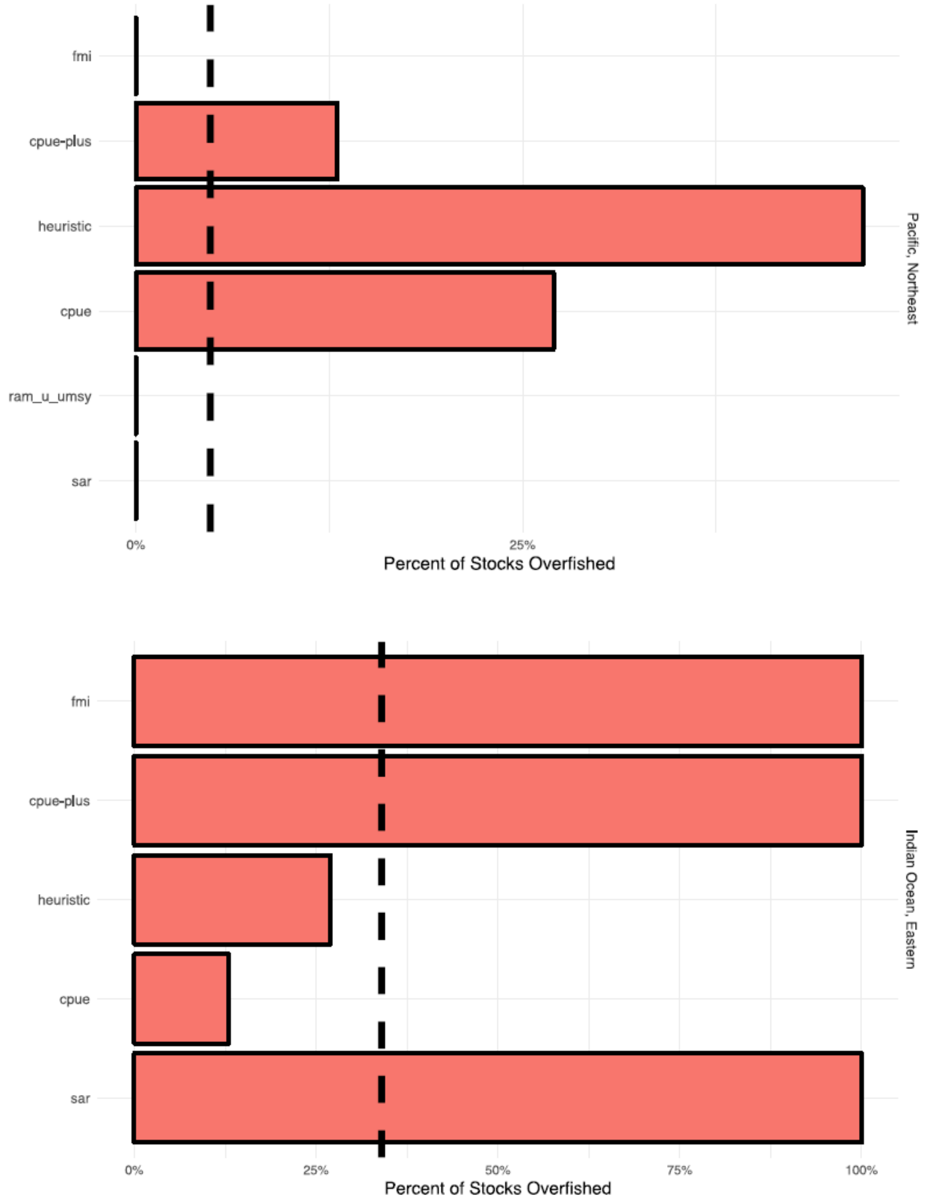

Fig. 1. Percent of stocks whose mean B/B_{MSY} values is less than 0.8, resulting in an 'overfished' classification. y axes denote the type of data used in the assessment, where sar = swept area ratio, heuristic uses the Catch-MSY heuristics, cpue uses continental scale effort data from Bell, Watson and Ye (2017), fmi uses the mean posterior prediction of U/U_{MSY} and B/B_{MSY} for the statistical area from the FMI regression, sar using the mean posterior predicted U/U_{MSY} for the statistical area from the SAR regression, ram_u_umsy uses the mean U/U_{MSY} series for stock assessed fisheries in the region as a prior, and cpue_plus uses the cpue data along with FMI and SAR regression priors. The dotted line shows the percentage of overfished stocks according to FAO (2011). Note that missing points for a given data type or region means that no stocks were on average estimated to be overexploited there.

The FAO Expert Consultation Workshop on the "Development of methodologies for the global assessment of fish stock status" took place in Rome, Italy, from 4-6 February 2019. The overall objective of the workshop was to present to a group of experts in stock assessment, fishery management and policy development a new tool (now called sraplus) that can be used to assess the status of currently unassessed stocks. More in detail, participants were asked to i) review the technical design, soundness and practical applicability of the new methodology, ii) evaluate its performance with simulation data and real fisheries data, iii) discuss its advantages and disadvantages as well as any necessary adjustments, and iv) design a plan for steps forward towards its application in global fish stock assessment. The workshop participants contributed in their individual capacities to the discussions which resulted in a set of recommendations on how to improve the tool and the data that is provided to the assessments.

ISBN 978-92-5-131425-8 ISSN 2070-6987

9 789251 314258

CA4355EN/1/05.19