North Pacific Fisheries Commission

NPFC-2022-TWG CMSA06-WP07

4th Meeting of the Small Working Group on Operating Model for Chub Mackerel Stock Assessment
12 August 2022 (9 am – 1 pm Tokyo time)
Webex

Summary

Agenda Item 1. Opening of the Meeting

The 4th intersessional meeting of the Small Working Group on Operating Model for Chub Mackerel Stock Assessment (SWG OM) commenced at 9 AM on 12 August 2022, Tokyo time in the format of video conferencing via Webex. The meeting was attended by China (Qiuyun Ma, Libin Dai), European Union (Karolina Molla Gazi), Japan (Shota Nishijima, Kazuhiro Oshima, Momoko Ichinokawa, Akihiro Manabe), Russia (Igor Chernienko, Viktor Zamyatin) as well as the Secretariat (Alex Zavolokin, Sungkuk Kang, Mervin Ogawa). Dr. Joel Rice attended the meeting as an invited expert. The meeting was opened by Dr. Shota Nishijima (Japan) who served as the SWG OM Lead.

Agenda Item 2. Adoption of Agenda

There were no amendments to the agenda.

Agenda Item 3. Report of progress since SWG-OM03

Members reported their progress since SWG-OM03 held on 30 June 2022 and presented their results of fitting the candidate stock assessment models to pseudo data. The performance measure outputs from ASAP, VPA, SAM and BSSPM were submitted to the invited expert (Dr. Joel Rice) by 9 August 2022 before this (SWG OM04) meeting, whereas Russia faced an error problem in applying the OMutility package to the calculation of performance measures from KAFKA.

China (Quiyun Ma) presented the results of fitting ASAP and BSSPM models to the pseudo data for scenarios A and B with 400 iterations for each iteration. For ASAP, the data extracted is based on catch-at-age with six (6) abundance indices scaled by mean. Twenty (20) iterations or 5% of all iterations did not converge both for scenarios A and B. Compared with ASAP based on true data, the results had higher biomass, lower fishing mortality and exploitation rate but similar recruitment. Some BRP-Bmsy had two peaks which is similar to SSBmsy. Most median values of base fishing

2nd Floor Hakuyo Hall
Tokyo University of Marine Science and Technology
4-5-7 Konan, Minato-ku, Tokyo
108-8477 JAPAN

TEL +81-3-5479-8717
FAX +81-3-5479-8718
Email secretariat@npfc.int
Web www.npfc.int

mortality for pseudo data are close to the values of true data except for a couple of iterations. ASAP's BRP-RFI is more concentrated and with no remarkable difference compared with the base value. For BSSPM, the data extracted is based on the total catch (weight) with four (4) abundance indices without recruitment. All iterations for scenario A converged and only one (1) iteration or 0.25% of the total iterations in scenario B did not converge. Scenarios A and B had lower biomass and higher exploitation rate, particularly in recent years. BRP-depletion and BRP-RFI values were close to the true data values.

The Lead asked China why the observed exploitation rate and fishing mortality had different trajectories for ASAP. China replied that they simply used the OMutility package provided by Japan, fitted the model, and presented the results. The Lead requested China to check if there is a bug in the OMutility package by manually calculating the weighted average of fishing mortality and compare the results using the OMutility package.

Japan (Shota Nishijima) presented on fitting VPA and SAM models to the pseudo data and the results of performance measures. The two models were fitted to the pseudo data under all six scenarios (A, B, C, D, E, F) and submitted the performance measures except for Mohn's rho in retrospective analysis. Since VPA and SAM cannot be applied to zero catch, a small constant value (a half of minimum positive value) was added to zero catch samples. The same model configuration used with true data was applied. All iterations converged for VPA. VPA based on true data obtained higher estimates in total biomass and recruitment. However, scenario E showed large uncertainties of fishing mortality and biomass even for past years. For VPA's BRP-fishing mortality, Fmsy_HS had the largest variations under scenarios A and E. For VPA's BRP-total biomass and SSB, extremely large values were estimated in some pseudo datasets and almost zero values were estimated for SBmsy_HS in some pseudo datasets. Japan also noted that SBmsy and Bmsy were sometimes estimated as zero or a negative value when Fmsy was estimated at the maximum value (10). For VPA, extremely low relative fishing impacts were estimated for some pseudo datasets, particularly scenario E, and the depletion ratio to maximum showed large variations when abundances were high (in the 1970s, 1980s, and 2010s). For SAM, some iterations did not converge in the pseudo datasets. For iterations that did not converge, Japan adjusted the model and applied seven (7) simplified model configurations. In the original model, more than 90% of the iterations converged. However, the seven simplified model configurations made it possible to converge all iterations. Compared with VPA, total biomass using SAM showed the median to be close to the true data. However, similar to VPA's fishing mortality, Fmsy_HS using SAM had relatively larger variations under scenarios A and E. For SAM's BRP-total biomass and SSB, almost zero values

were estimated for SBmsy_HS in some pseudo datasets. For SAM's BRP-relative fishing impact, extremely low relative fishing impacts were estimated for some pseudo datasets, particularly scenario E, and the depletion ratio to maximum showed large variations when abundances were high (in the 1970s, 1980s, and 2010s).

The invited expert asked Japan which metrics to recommend (for 16-18 or 17-19) based on the analysis showed by the models. Japan responded that similar results were obtained for those cases, but they do not have any recommendations.

China clarified how to deal with iterations that did not converge and suggested Members to discuss and agree on how to deal with such cases in the next agenda.

Russia (Igor Chernienko) presented on fitting the KAFKA model to the pseudo data and the result of performance measures. Due to some technical problems, Russia presented a single slide. Although all iterations converged, Russia commented that the graph of scenario B appeared strange and that it might have been a result of some technical issues.

The invited expert asked if there were residual patterns of predicted and observed CPUE to help interpret the results which was relatively different from those of the other scenarios with respect to the years between 2010 and 2019. The invited expert also suggested to send this question to Russia via email.

The Lead commented that there should be some degree of variations for the years prior to 2002 because of the different catch-at-age data.

Russia responded that this might be due to a bug in their codes and promised to recheck their calculations.

The Chair speculated that the error might be due to KAFKA requiring the input of the total number of estimates in OMutility to calculate total biomass or SSB and other related quantities. The Lead also asked Russia to inform him as soon as possible if this issue cannot be resolved.

Agenda Item 4. Discussion of performance accuracy and model prioritization

The invited expert presented a summary of the results of submitted performance measures from each Member. Due to time restrictions only the model fits for scenario A were summarized. The presentation covered (1) the review of model fits to scenario A, (2) discussion of priority graphs and tables for TWG CMSA06, (3) recommendation of particular performance measures, (4) discussion of priority performance metrics, and (5) discussion of the development of model fitting to pseudo data.

The invited expert reminded participants that the four models (VPA, ASAP, KAFKA, SAM) were

run with the "True Data" and were used to generate parameters that went to POP-SIM A, which in turn were used to make pseudo data. Individual models were then run against these scenarios to show fits. Fit is summarized by deviates (estimated value − true value) and performance metrics (e.g., RMSE, CV) for each of the performance measures.

As example of fits to data, data from the VPA, KAFKA, SAM and ASAP parameterizations were used to generate simulations that were then run in the VPA model. Deviates ($X_{EST} - X_{TRUE}$) could be used to visually show the range of results for each performance measure. The invited expert also showed other F-related performance measures, biomass-related performance measures, depletion statistics, and deviates from total biomass, in particular, the box plots of deviates where VPA is fitted to data from other models and noted that KAFKA has the least variability.

The performance metrics used were (1) Root mean squared error (RMSE), (2) Median relative percent bias (MRPB), (3) Median absolute relative bias (MARB), and (4) CV.

The invited expert also showed plots of biomass-based performance metrics and fishing mortality-based performance metrics (i.e., which Bmsy(or Sbmsy)-related metrics or which F-related metrics to use?) and presented bi-variate plots of the deviates to show the correlation between two variables, posing the question of what other bi-variate or multivariate plots to consider.

The invited expert showed F-related and biomass depletion and trend box plots of deviates for ASAP, which is the ASAP model fitted to data from other models and commented that ASAP fits the VPA data the most.

It was noted that KAFKA summary is still under development and some years were apparently missing, so there might be a possible error in interpreting the model values for KAFKA biomass trend or the error might be due to the code. KAFKA's BRP-Depletion was similar across the board. For this meeting, only VPA, ASAP and KAFKA were discussed. (SAM analysis will follow).

The invited expert concluded that Members need to discuss the choice of reference performance measures (e.g., biomass, biological reference points), performance metrics (e.g., RMSE, CV) and the results distribution before the TWG CMSA06 and promised to double check the calculations.

The invited expert suggested Members to rank the performance measure by tiers.

The invited expert finally reminded participants of the memo of performance measure and prioritization discussed during the SWG OM03 meeting held on 30 June 2022. (Meeting Summary of SWG OM03)

The Lead requested the invited expert to share his code and results with all Members. With the permission from all Members, the invited expert agreed to share his results and summary for scenario A including all pseudo data used.

China supported the suggestion of the invited expert to remove iterations that failed to converge, to save time and effort.

Japan asked whether to exclude the pseudo data for all estimation model or just ignore the iteration of specific estimation model and what model produced the data without convergence, and it is confirmed that the former procedure is applied.

Japan agreed to exclude iterations that did not converge using the original SAM model.

Japan asked how to determine non-convergence in ASAP and China responded that it is when the system returns a warning that the model did not converge, and no results are displayed.

The Lead asked Russia if there were non-convergence of iterations in KAFKA and Russia responded that they did not have any problems with convergence.

Members discussed the ranking of performance measures and initially agreed on the following: Tier 1 – Fmsy and Bmsy, Mohn's rho; Tier 2 – SSB/median(SSB) (recent periods), B/median(B) (recent periods), Exploitation Rate (recent years, but need to check later), Relative fishing impacts, F%SPR, Fmax, F0.1; Tier 3 –, Biomass, Recruitment, F values, Whole years. Japan noted that, because performance measures with higher rank are more difficult to be interpreted, it is convenient to start the review of results with the measures with lower rank, i.e., in the order of Tier 3, 2, and 1. TheLead noted that these tiers are just initial prioritization and can be changed by the group in the future.

Agenda Item 5. Timeline up to the TWG CMSA06 meeting

Members confirmed and discussed future timeline up to the TWG CMSA06 meeting (5-8 September 2022).

- For SAM, Japan will send the data of performance measure results excluding iterations without convergence to the invited expert by 18 August 2022.
- For KAFKA and ASAP, Russia and China will check whether the performance measures based on true data were computed using the new version of OMutility by 18 August 2022.
- Japan will check and verify the new version of OMutility by 18 August 2022.
- For KAFKA, Russia will check how to use OMutility and estimates in past years (i.e., whether catch-at-age of pseudo data were varied among iterations) by 18 August 2022.
- Members have agreed to share the data of the performance measures based on true and pseudo data under scenario A by 20 August 2022.
- Members have agreed to submit all other performance measures (scenarios C-F and Mohn's rho) by 22 August 2022.
- Members have agreed to submit a document reporting their results of fitting their models to

pseudo data and performance measures by 29 August 2022.

- TheLead requested the Members to submit, on a voluntary basis, a document detailing their ideas about the metrics and methods to be used for evaluating model accuracy by 29 August 2022.
- The Lead requested the invited expert to submit a document detailing his ideas about the metrics and methods to be used for evaluating model accuracy and to present it at the TWG CMSA06 meeting.

Agenda Item 6. Other matters

Kazuhiro Oshima from Japan informed everyone that he will serve as interim Chair of TWG CMSA06 and asked for everyone's cooperation.

Agenda Item 7. Closing of the Meeting

The meeting closed at 1:43 pm Tokyo time.