North Pacific Fisheries Commission

NPFC-2023-TWG CMSA07-Final Report

North Pacific Fisheries Commission
7th Meeting of the Technical Working Group on Chub Mackerel Stock Assessment

4–7 September 2023
Port Vila, Vanuatu

REPORT

Agenda Item 1. Opening of the Meeting

1. The 7th Meeting of the Technical Working Group on Chub Mackerel Stock Assessment (TWG CMSA) was held in a hybrid format, with participants attending in-person in Port Vila, Vanuatu, or online via WebEx. The meeting was attended by Members from Canada, China, the European Union (EU), Japan, the Russian Federation, the United States of America, and the Republic of Vanuatu. An invited expert, Dr. Joel Rice, participated in the meeting.

2. The meeting was opened by Dr. Kazuhiro Oshima (Japan), the TWG CMSA Chair, who welcomed the participants. He expressed his appreciation to be able to hold the meeting in beautiful Vanuatu and thanked the Government of the Republic of Vanuatu for hosting the meeting.

3. Mr. Sompert Gereva, Director of Fisheries, welcomed the participants to Port Vila on behalf of the host Member. He expressed Vanuatu's pleasure to host the TWG CMSA meeting and welcomed the successful conclusion of the two Pacific saury-related meetings in the previous week. Mr. Gereva expressed his hope that the discussions would be productive and that the participants' joint efforts would foster progress and collaboration in the chub mackerel stock assessment work.

4. The Executive Secretary, Dr. Robert Day, hoped that the TWG CMSA meeting would be as successful and productive as the previous week's Pacific saury-related meetings and expressed his gratitude to Vanuatu for its continued hospitality and support.

5. The Science Manager, Dr. Aleksandr Zavolokin, outlined the procedures for the meeting.

6. Mr. Alex Meyer was selected as rapporteur.

Agenda Item 2. Adoption of Agenda

7. The agenda was adopted without revision (Annex A). The List of Documents and List of Participants are attached (Annexes B, C).

Agenda Item 3. Overview of the recommendations and outcomes of previous NPFC meetings relevant to chub mackerel

3.1 6th TWG CMSA

8. The Chair provided an overview of the outcomes and recommendations of the 6th TWG CMSA meeting and the progress made in the preceding meetings of the TWG CMSA and the Small Working Group on Operating Model for Chub Mackerel Stock Assessment.

3.2 Intersessional meetings of TWG CMSA

9. The Chair provided an overview of the 1st and 2nd intersessional meetings of the TWG CMSA held in May and June 2023 (NPFC-2023-TWG CMSA07-RP01 and RP02).

3.3 COM07

10. The Science Manager presented the outcomes from the 7th Commission meeting and highlighted the interest in the timeline for the stock assessment of chub mackerel expressed by Commission Members, who considered the work on chub mackerel to be a priority and looked forward to its timely completion.

3.3.1 NPFC Performance Review

11. The Science Manager presented an overview of the NPFC Performance Review and outlined some recommendations from the Performance Review report that concern chub mackerel.

12. The Chair will draft the proposed response to the recommendations of the Performance Review for consideration by Members at the next TWG CMSA meeting.

3.3.2 Resolution on Climate Change

13. The Science Manager presented an overview of the Resolution on Climate Change.

14. The TWG CMSA requested Members to share relevant fisheries and research information on the impact of climate change, such as the impact of climate change on the distribution of chub mackerel, at future meetings.

15. The TWG CMSA agreed to reflect its response to the Resolution on Climate Change in the Stock Assessment Protocol and the TWG CMSA Work Plan and held further discussions under

agenda items 6.4 and 9, respectively.

16. The TWG CMSA noted that the biological parameters of chub mackerel are affected by environmental factors, such as climate change, and also density-dependent factors. The TWG CMSA agreed to include consideration of the impact of density-dependent and density-independent factors on the biological parameters of chub mackerel in its agenda at future meetings.

17. The invited expert pointed out that the combination of direct and indirect effects of climate change, confounded with historical fishing effects, is producing a density-dependent effect, and that both involve complex spatio-temporal effects. He suggested that spatio-temporal effects should be considered in the standardization of catch-per-unit-effort (CPUE) indices, and recommended that Members take a collaborative approach and pool their data to explore the effects of climate change.

Agenda Item 4. Member's fisheries status and research activities

18. China presented a review of its chub mackerel fishery and research activities (NPFC-2023-TWG CMSA07-IP02). In 2022, China operated 105 purse seine vessels and 2 trawl vessels in the Convention Area. Most catch occurred between 42 and 44 degrees north latitude. The catch in 2022 was approximately 83,000 MT, a decrease from 2021 but an increase from 2020. CPUE has been decreasing in recent years, perhaps due to the effects of climate change. The average length of caught individuals was 221 mm. The trend in average fork length from 2016–2021 was a gradual increase to a stable level. The main ages at catch in 2022 were from 1+ to 3+. China collects and analyzes fishing logbooks every year, sends specialist research staff to fishing vessels or ports to collect sample data, monitors the monthly ratio of chub mackerel and blue mackerel in catch, and conducts monitoring of biological features.

19. Japan presented a review of the recent fishery and stock status of chub mackerel (NPFC-2023-TWG CMSA07-IP04). Japan first reminded the TWG CMSA that chub mackerel makes a northward migration for feeding from April to July, followed by a southward migration for wintering from August to March. Japan's catch comes from large-scale purse seine vessels, small-scale purse seine vessels, set nets, and dip nets and other gears. The majority of catch is from large-scale purse seine vessels. Catch declined substantially in fishing year (FY) 2022 (July 2022 to June 2023) to approximately 95,000 MT, which was half of the level in FY2021 and one-third of the level in FY2020. Purse seine catch in FY2022 fell to a 30-year low. Japan conducts its stock assessment by fishing year and doing so enables inclusion of a single peaked catch season within a year. The substantial catch of chub mackerel is between November and spring months. The age classes for Japanese fleets' catch-at-age data were evenly distributed

in 2021. Based on the footprint data, catch of mackerels by Russia and Japan declined from 2021 to 2022. Meanwhile, chub mackerel catch by Chinese fleets has stayed at a similar level, but CPUE has declined. Japan's 2022 summer midwater trawl survey found that the recent expansion of the chub mackerel distribution eastward has result in increased appearance of age 1+ fish in the survey area. The 2022 autumn midwater trawl survey showed that chub mackerel is broadly distributed in the survey area. The 2022 egg survey found that egg density was high between March to June, with a similar to the patten in 2021.

20. The TWG CMSA requested Japan to share any available gear-specific catch and effort information at TWG CMSA08.

21. Russia presented a review of its chub mackerel fishery and research activities in 2022 (NPFC-2023-TWG CMSA07-IP01). In 2022, the main fishing grounds were in the Japanese exclusive economic zone (EEZ) from January to March, before shifting to the Russian EEZ in June, and then back to the Japanese EEZ in December. Monthly CPUE was highest in January, February, March, and December. Monthly catch was highest in January, February, and December. From 2016 to 2022, total annual catch was highest in 2018, followed by 2021 (87,388 MT), and has declined significantly in 2022. In terms of research activities, Russia conducted two multipurpose and multispecies trawling surveys in the upper epipelagic zone of the Northwestern Pacific Ocean in 2022, the first in June and the second in September.

22. The TWG CMSA requested Members to present and explain the methods they use to estimate catch-at-age data.

23. The TWG CMSA requested Members to present more detailed fleet descriptions, including gear specifications/configuration.

Agenda Item 5. Selection of stock assessment model(s) for chub mackerel

5.1 Project overview and methods development for the testing and evaluation of stock assessment models

24. Japan presented a study on simulated responses of summary performance metrics to varying model complexity (NPFC-2023-TWG CMSA07-WP06). Japan found that simple simulation testing shows that summary metrics (or performance metrics) responded to varying model complexity differently and that the median relative bias (MedRB) and the median absolute relative bias (MedARB) showed better scores when the model complexity is correct than the coefficient of variation (CV) and root mean squared error (RMSE). Considering the bias-variance tradeoff, Japan recommended that the first priority be place on MedARB as the main summary metric.

25. The EU informed the TWG CMSA that it would provide more detailed usage examples in Europe for inclusion in the table of qualitative attributes of the candidate stock assessment models (Annex D) following the meeting.

5.2 Summary of performance and ranking of the stock assessment model candidates and their characteristics

5.2.1 Recommendations from the external expert

26. The invited expert presented a consolidation and review of performance measures, a detailed summary of model performance, and ranking of the stock assessment model candidates (NPFC-2023-TWG CMSA07-WP02) and his recommendation on a stock assessment model platform based on work completed as part of the TWG CMSA assessment modeling project (NPFC-2023-TWG CMSA07-WP03). The invited expert explained that the TWG CMSA conducted an analysis of uncertainty following the methods of Deroba et al (1995) to test the robustness of four stock assessment models to error via simulation, the results of which could help develop guidelines for the selection of a stock assessment model, and that the details of the ranking of performance measure and choice of summary metrics were developed through intersessional work conducted since the TWG CMSA06 meeting in September 2022.

27. Based on the comparison, with respect to the models investigated in this project (Age Structured Assessment Program (ASAP), cohort analysis with Kalman filter (KAFKA), state-space assessment model (SAM), virtual population analysis (VPA)), the invited expert recommended the SAM as the model for stock assessment of chub mackerel.

28. The TWG CMSA's report for the TWG CMSA assessment modeling project is attached as Annex E.

5.3 Final agreement on stock assessment model(s) and procedures in TWG CMSA

29. The TWG CMSA endorsed the report presented by the invited expert and agreed to use SAM as the model for stock assessment of chub mackerel.

30. In the model selection process, the TWG CMSA noted the following issues, which should be considered in the finalization of the stock assessment model:
    (a) Mohn's rho calculated based on the fits to the pseudo data was not considered in the final ranking of the stock assessment models.
    (b) Among the models considered, there are different levels of performance among the variable types (i.e. State, BRP, RFI etc.).
    (c) There appear to be trends in the chosen Summary Metric (MedARB) for some annual

performance measures (i.e. total biomass) among some assessment models.

i. This may be due to the nature of time series analysis (i.e. the most recent years often have less data to inform estimation).

ii. The uncertainty in recent years may have impact on fisheries management for this stock.

iii. Interpretation of the basis from which the stock status is estimated should take uncertainty in recent years into account as is done in other RFMOs.

(d) It is important that the Biological Reference Points (MSY related) and Depletion as well as state variables be well estimated, because they will impact the Harvest Control Rule. For the estimation of scale, total biomass is important.

(e) Some models failed to estimate state variable without bias even in self-test, indicating a lack of consistency, and a need for further model improvement.

31. As future work, the TWG CMSA agreed to conduct sensitivity analyses to better understand the effect of catch uncertainty, given that chub mackerel catch is taken from mixed fisheries.

32. The TWG CMSA thanked the invited expert for his great support and contributions to the stock assessment model evaluation project.

Agenda Item 6. Preparations for stock assessment of chub mackerel

6.1 Review of biological parameters

33. Japan presented its chub mackerel weight-at-age and maturity-at-age data, as well as a comparison with China’s weight-at-age and maturity-at-age data (NPFC-2023-TWG CMSA07-WP13). Japan found that in some year classes, Chinese weight-at-age data from a particular cohort exhibit a similar pattern to Japanese weight-at-age data from a previous year class. As for maturity-at-age, Japanese data show maturation from around age 3 with full maturation at age 4 at a weight of around 300-400g, while Chinese data show maturation from age 1-2 and a weight-at-maturity of around 80-190g. Japan suggested that it is necessary to continue to conduct comparisons to determine how the data are prepared for the weight-at-age analysis, to understand how the age analysis of samples is performed among Members, and to clarify maturity-at-age methodologies among Members are consistent.

34. China explained that its chub mackerel samples are taken at a different time and from a different area to Japan’s samples.

35. At the request of the TWG CMSA, Japan presented an updated comparison of Japan and China’s quarterly weight-at-age data up to 2022. The TWG CMSA noted that for ages 0-4, there is much overlap between Chinese weight-at-age data at age t and Japanese weight-at-age

at age t-1 up to 2020, but Japanese weight-at-age data are still larger than Chinese weight-at-age data in 2021–2022. The TWG CMSA noted that for ages 4+, there is a similar pattern among Chinese and Japanese weight-at-age data for ages 4–5, but Chinese weight-at-age data for ages 5–6 is much larger than Japanese weight-at-age data for ages 7+.

36. To develop a common understanding on the difference of weight-at-age among Members, the TWG CMSA requested Members to share their available quarterly or monthly catch-at-length data, length-weight relationship parameters, and length-frequency data in proportion in the intersessional period. The TWG CMSA agreed to use the data-sharing template that Japan used for blue mackerel and Japanese sardine and requested Japan to upload the template to the NPFC Collaboration site.

37. China presented a description of its available data (NPFC-2023-TWG CMSA07-IP03). China explained its methodologies for sampling, ALK development, and estimating catch-at-age from the ALK, and presented its data for length and age distribution, length-weight relationship, catch-at-age, and number-at-age.

38. China presented the monthly catch data and distribution of chub mackerel fishing grounds for its purse seine fleet (NPFC-2023-TWG CMSA07-WP09), and the monthly catch data for its purse seine and pelagic trawl fleets (NPFC-2023-TWG CMSA07-WP10).

39. The TWG CMSA noted the importance of Members using the same methodologies to measure/observe biological parameters. The TWG CMSA requested Members to share their methodologies, including for aging and determining maturity, and to identify and discuss differences among them, as a first step towards developing a protocol for common methodologies.

40. The TWG CMSA reviewed and updated the table of data potentially available for stock assessment of chub mackerel (Data availability for CMSA).

6.2 Intersessional works on fishery data (catch-at-age, weight-at-age, maturity-at-age, if possible)

6.2.1 Calendar to be applied to stock assessment on chub mackerel

41. Japan presented the results of simple simulations for determining the timing for aging and defining fishing year in the stock assessment of chub mackerel (NPFC-2023-TWG CMSA07-WP08). Japan used biological parameters and fisheries scenarios similar to those of chub mackerel in the Pacific to observe how and to what extent potential biases can occur in the abundance estimation when different fisheries use different definitions of fishing year and timing for aging in creating annual catch-at-age data. Based on the results, Japan recommended

that a consistent definition of fishing year and timing for aging be used across fisheries to avoid potential biases in abundance estimation. In addition, because average annual weight-at-age is likely to differ among fisheries due to different fishing seasons even though the weight-age relationships of the population are consistent, the single weight-at-age representing the population should be used when evaluating total abundance, while fishery-specific weight-at-age could be used in calculating total catch weight by fishery. Quarterly catch-at-age data between calendar and fishing years can be converted as long as quarterly catch-at-age data are available. Furthermore, total catches by fishing year can be converted into catches by calendar year when quarterly catch-at-age and weight-at-age data are available. Therefore, it is important to continuously collect quarterly-based catch-at-age and weight-at-age data.

42. The TWG CMSA agreed to consider and compare the application of different definitions of year as follows:
    (a) Application of fishing year (July-June) will be used as a base case. For Chinese quarterly catch data, which are not available from 2014 to 2017, a method to convert the annual catch for this period into the fishing year basis should be developed prior to the TWG CMSA08 meeting.
    (b) A case where each fleet applies their own calendar will be a sensitivity case as a backup of the base case.

6.3 Intersessional works on abundance indices

43. Russia presented research on new predictors for tracking the habitat of chub mackerel (NPFC-2023-TWG CMSA07-WP05 (Rev. 1)). Using data from scientific trawl tracks from February 2021 to May 2023 that record the occurrence or absence of chub mackerel, Russia estimated the variable importance and confidence intervals of a variety of environmental characteristics related to Lagrangian water properties and sea surface temperature (SST) for chub mackerel encounter probability. Russia found that SST from the 0 level from the NEMO and HYCOM models (wT00 and wT0, respectively) had lower importance than water temperature at the 1 level of the NEMO model (wT1), which is approximately at 1.5 m depth, and that Lagrangian characteristics (Lyapunov exponent (L) and the length of passive tracers' trajectories back calculated in time for 15 days (S)) had higher importance than other speed-based variables, but lower importance than variables related to productivity. Russia hoped that the further development of this work would help in the monitoring of the stability of suitable areas for fishing of chub mackerel, which should be useful during interpretation of Russian CPUE fluctuations. Russia requested other Members to compare its estimates of encounter probability with their actual catch data and share the comparison results.

44. Russia presented a standardization of CPUE data for chub mackerel caught by its trawl fishery

from 2015 to 2021 using generalized additive models (GAM) (NPFC-2023-TWG CMSA07-WP04). Russia recommended using the standardized CPUE derived from GAM as input for the stock assessment.

45. The TWG CMSA suggested some technical improvements for Russia's CPUE standardization in relation to:
    (a) The method for selecting data to be used with consideration for targeting;
    (b) The method for computing the index and confidence interval;
    (c) Including the annual spatial distribution of catch and effort;
    (d) Including a table showing estimated fixed-effect parameters and their standard deviation.

46. China presented a standardization of CPUE data for chub mackerel from 2014 to 2022 (NPFC-2023-TWG CMSA07-WP11) using a generalized linear model (GLM) and a GAM. China recommended using the standardized CPUE derived from GAM as input for the stock assessment.

47. Japan noted that the catch proportion of Japanese sardine has increased largely, even higher than chub mackerel, in recent years for Members. Therefore, Japan requested Members to consider this influence during their CPUE standardization.

48. The TWG CMSA noted that Members seemed to have difficulty reviewing each other's CPUE standardizations because not enough information was included in the working papers for understanding and assessing them. The TWG CMSA noted the need to agree on a standardized set of information that should be provided when sharing CPUE standardizations and discussed this further when reviewing and updating the CPUE Standardization Protocol. The TWG CMSA agreed that the sharing of CPUE standardization code would also be useful.

6.3.1 Review and update of the CPUE Standardization Protocol

49. The TWG CMSA reviewed and updated the CPUE Standardization Protocol (Annex F). The TWG CMSA began work to develop a template for presenting Members' CPUE standardizations (NPFC-2023-TWG CMSA07-WP15) and agreed to finalize it during the intersessional period. The template will be attached to the CPUE Standardization Protocol.

6.4 Review of the Stock Assessment Protocol for Chub Mackerel

50. The TWG CMSA reviewed and updated the Stock Assessment Protocol for Chub Mackerel (Annex G).

6.5 Possible settings and specification of stock assessment model

51. Japan presented a detailed description of SAM and an R package of SAM ("frasam") developed for the stock assessment of chub mackerel, and provided a demonstration of SAM's flexibility (NPFC-2023-TWG CMSA07-WP07). Japan explained that SAM can mimic the structures and assumptions of the other candidate models (ASAP, KAFKA, and VPA) and can incorporate the uncertainty of fixed parameters using different scenarios of natural mortality as an example.

52. Japan presented a draft table of possible settings and specification of SAM (NPFC-2023-TWG CMSA07-WP14).

53. The TWG CMSA reviewed and revised the draft table of possible settings and specification of SAM (Annex H). The TWG CMSA agreed to continue to discuss the settings and specification, conduct a preliminary run prior to TWG CMSA08 and present the results at TWG CMSA08, and finalize the settings and specification of SAM at TWG CMSA09.

6.6 Recommendations on preparations for stock assessment

54. The TWG CMSA agreed to:
   (a) submit input data such as quarterly catch-at-age, weight-at-age, and maturity-at-age (or annual data if quarterly data are not available) to be analyzed and aggregated through collaborative work among Members in a transparent manner with a written report, including methodology, with participants from China, Japan, and Russia as authors.
   (b) share available quarterly or monthly catch-at-length data, length-weight relationship parameters, and length-frequency data in proportion in the intersessional period using the template on the NPFC Collaboration site to develop a common understanding on the difference of weight-at-age among Members.
   (c) consider and compare the application of different definitions of year as follows:
      i. Application of fishing year (July-June) will be used as a base case. For Chinese quarterly catch data which are not available from 2014 to 2017, a method to convert the annual catch for this period into the fishing year basis should be developed prior to the TWG CMSA08 meeting.
      ii. A case where each fleet applies their own calendar will be a sensitivity case as a backup of the base case.
   (d) finalize the template for presenting CPUE standardizations during the intersessional period.
   (e) present updated CPUE standardizations at TWG CMSA08 using the CPUE standardization template and following the updated CPUE Standardization Protocol.
   (f) follow the updated Stock Assessment Protocol.
   (g) continue to discuss the settings and specification of SAM, conduct a preliminary run prior to TWG CMSA08 and present the results at TWG CMSA08, and finalize the settings and

specification of SAM at TWG CMSA09.

Agenda Item 7. Future projection of chub mackerel

7.1 Review of the table of options for the basic specifications of conducting future projections for chub mackerel

55. The TWG CMSA reviewed and updated the table of possible options for the basic specifications for conducting future projections for chub mackerel (Annex I). The TWG CMSA agreed to continue to discuss and develop the table and determine provisional specification and setting towards TWG CMSA09.

Agenda Item 8. Biological reference points

8.1 Review of the table of candidate biological reference points for chub mackerel

56. The TWG CMSA reviewed the table of candidate biological reference points for chub mackerel drafted by the invited expert and TWG CMSA06. The TWG CMSA agreed to base its future discussions on the following candidate biological reference points:
    (a) F-based reference points
        i. F_{MSY}
        ii. F%SPR
        iii. $F_{0.1}$, F_{max}
    (b) Biomass-based reference points (including SSB, summary biomass, etc.)
        i. B_{MSY}
        ii. $\%B_0$
        iii. Certain historical level of B

Agenda Item 9. Review of the Work Plan of the TWG CMSA

57. The TWG CMSA reviewed and updated the Work Plan of the TWG CMSA (NPFC-2023-TWG CMSA07-WP01 (Rev. 1)). The TWG CMSA confirmed its intention to complete the first chub mackerel assessment in 2024.

Agenda Item 10. Other matters

10.1 Timeline and intersessional activities before TWG CMSA08

58. The TWG CMSA drafted a timeline and activities from the conclusion of TWG CMSA07 to February 2024 (Annex J).

59. The TWG CMSA agreed to create input data such as catch-at-age, weight-at-age and maturity-at-age data in a collaborative manner towards the TWG CMSA08 meeting. Counterparts from Members will develop the input data through email communication/online meeting.

60. The TWG CMSA discussed the schedule of its meetings in 2024 financial year. The TWG CMSA confirmed that it will hold a meeting in autumn 2024 and may hold one more meeting in 2024, if it will be needed to finalize the stock assessment for chub mackerel.

61. The TWG CMSA discussed developing an online private git repository to develop and share code. The TWG CMSA agreed to use and update the git repository and the NPFC Collaboration site in parallel to ensure that all Members have access to the latest code. Members agreed to develop a working paper for developing general protocols and guidelines for using git repositories for joint data analysis projects and present it at SC08.

62. The EU offered to create a manual that would provide simple directions as to how Members could use git.

10.2 Observer Program

63. The Science Manager reminded the TWG CMSA of information he had previously presented regarding the establishment of a regional NPFC observer program and summarized the relevant discussions from the TWG CMSA05 and TWG CMSA06 meetings.

64. The TWG CMSA noted that after it conducts its first chub mackerel stock assessment, it would have a better understanding of potential data gaps and which of these gaps could be filled by a regional NPFC observer program.

10.2.1 Review data or data description on fisheries bycatch in the chub mackerel fisheries

65. China presented a data description of the fisheries bycatch in its chub mackerel fisheries (NPFC-2023-TWG CMSA07-WP12). China explained that it catches chub mackerel and Japanese sardine as part of mixed-species fisheries. Most of its chub mackerel and sardine catches were harvested by the lighting purse seine fishery. The Japanese sardine catch increased from a very low level in 2014 to a peak (266,615 MT) in 2022. Squid and saury are bycatch or inevitable catch in the mackerel fisheries, and the annual output and proportion were very low, whether in the purse seine or trawl fisheries. The catch of other pelagic fish species was also very low.

66. Japan informed the TWG CMSA that it would provide information on bycatch from its chub mackerel fisheries at TWG CMSA08. Japan explained that it would be able to provide more accurate information for its purse seine fleet operating in northern waters, which is its main fleet, and that it would be difficult to provide accurate information about its other fisheries, such as its set net and dip net fisheries, which are very small in scale.

10.3 Species summary

67. The Chair explained that, due to unforeseen circumstances, he had not been able to draft a species summary for chub mackerel. The TWG CMSA agreed to develop the species summary intersessionally for submission to SC08.

10.4 Other issues

68. The EU explained the importance of sharing qualitative information on Members' biological data collection programs (NPFC-2023-TWG CMSA07-IP05) and presented a draft template for describing details of sampling design and estimation of chub mackerel catch (NPFC-2023-TWG CMSA07-IP06).

69. The TWG CMSA thank the EU for initiating this work and invited Members to work collaboratively with the EU to develop the template further.

Agenda Item 11. Recommendations to the Scientific Committee

70. The TWG CMSA agreed to:
    - (a) share relevant fisheries and research information on the impact of climate change, such as the impact of climate change on the distribution of chub mackerel, at future meetings.
    - (b) include consideration of the impact of density-dependent and density-independent factors on the biological parameters of chub mackerel in its agenda at future meetings.
    - (c) present and explain the methods used by Members to estimate catch-at-age data.
    - (d) present more detailed fleet descriptions, including gear specifications/configuration.
    - (e) share Members' methodologies for measuring/observing biological parameters, including for aging and determining maturity, and to identify and discuss differences among them, as a first step towards developing a protocol for common methodologies.
    - (f) use SAM as the model for stock assessment of chub mackerel.
    - (g) conduct sensitivity analyses to better understand the effect of catch uncertainty, given that chub mackerel catch is taken from mixed fisheries.
    - (h) continue to make preparations for the chub mackerel stock assessment as described in paragraph 54.
    - (i) continue to work intersessionally in accordance with the agreed timeline (Annex J).
    - (j) complete the first chub mackerel stock assessment in 2024.
    - (k) task the Secretariat, working with Members and the SC Chair, to set up an online private git repository to develop and share the TWG CMSA's code.

71. The TWG CMSA recommended that the SC:
    - (a) adopt the Work Plan of the TWG CMSA (NPFC-2023-TWG CMSA07-WP01 (Rev. 1)).
    - (b) endorse the TWG CMSA meeting schedule for 2023-2024 financial years: TWG

CMSA08 on 22-25 January 2024 and TWG CMSA09 in autumn 2024.

(c) hire an invited expert to support the TWG CMSA in the future stock assessment project.

(d) develop general protocols and guidelines for using git repositories for joint data analysis projects.

Agenda Item 12. Adoption of Report

72. The report was adopted by consensus.

Agenda Item 13. Close of the Meeting

73. The Chair thanked the participants for their cooperation and the good progress they had made, Vanuatu for hosting the meeting, the Secretariat and the rapporteur for their support, and the invited expert for his hard work and guidance.

74. The TWG CMSA thanked the Chair and the Vice-Chair for facilitating a smooth and productive meeting.

75. The meeting closed at 11:25 on 7 September 2023, Port Vila time.

Annexes

Annex A – Agenda

Annex B – List of Documents

Annex C – List of Participants

Annex D – Table of qualitative attributes of the candidate stock assessment models

Annex E – Consolidation and review of performance measures, detailed summary of model performance and ranking of the stock assessment model candidates

Annex F – Revised CPUE Standardization Protocol for Chub Mackerel

Annex G – Revised Stock Assessment Protocol for Chub Mackerel

Annex H – Possible settings and specification of SAM

Annex I – Possible options for the basic specifications for conducting future projections for chub mackerel

Annex J – Timeline and activities for intersessional work from the conclusion of TWG CMSA07 to February 2024.

Annex A

Agenda

Agenda Item 1. Opening of the Meeting

Agenda Item 2. Adoption of Agenda

Agenda Item 3. Overview of the recommendations and outcomes of previous NPFC meetings relevant to chub mackerel

- 3.1 6th TWG CMSA
- 3.2 Intersessional meetings of TWG CMSA
- 3.3 COM07 meeting
  - 3.3.1 NPFC Performance Review
  - 3.3.2 Resolution on Climate Change

Agenda Item 4. Member's fisheries status and research activities

Agenda Item 5. Selection of stock assessment model(s) for chub mackerel

- 5.1 Project overview and methods development for the testing and evaluation of stock assessment models
- 5.2 Summary of performance and ranking of the stock assessment model candidates and their characteristics
  - 5.2.1 Recommendations from the external expert
- 5.3 Final agreement on stock assessment model(s) and procedures in TWG CMSA

Agenda Item 6. Preparations for stock assessment of chub mackerel

- 6.1 Review of biological parameters
- 6.2 Intersessional works on fishery data (catch-at-age, weight-at-age, maturity-at-age, if possible)
  - 6.2.1 Calendar to be applied to stock assessment on chub mackerel
- 6.3 Intersessional works on abundance indices
  - 6.3.1 Review and update of the CPUE Standardization Protocol
- 6.4 Review of the Stock Assessment Protocol for Chub Mackerel
- 6.5 Possible settings and specification of stock assessment model
- 6.6 Recommendations on preparations for stock assessment

Agenda Item 7. Future projection of chub mackerel

- 7.1 Review of the table of options for the basic specifications of conducting future projections for chub mackerel

Agenda Item 8. Biological reference points

8.1 Review of the table of candidate biological reference points for chub mackerel

Agenda Item 9. Review of the Work Plan of the TWG CMSA

Agenda Item 10. Other matters

10.1 Timeline and intersessional activities before TWG CMSA08

10.2 Observer Program

10.2.1 Review data or data description on fisheries bycatch in the chub mackerel fisheries

10.3 Species summary

10.4 Other issues

Agenda Item 11. Recommendations to the Scientific Committee

Agenda Item 12. Adoption of Report

Agenda Item 13. Close of the Meeting

Annex B

List of Documents

MEETING INFORMATION PAPERS

Symbol	Title
NPFC-2023-SSC PS11-MIP01 (Rev. 1)	Meeting Information
NPFC-2023-TWG CMSA07-MIP02	Provisional Agenda
NPFC-2023-TWG CMSA07-MIP03 (Rev. 1)	Annotated Indicative Schedule

REFERENCE DOCUMENTS

Document Number	Title
	NPFC Performance Review
	Resolution on Climate Change
NPFC-2022-TWG CMSA06-Final Report	6th TWG CMSA meeting report
NPFC-2023-TWG CMSA07-RP01	1st intersessional TWG CMSA meeting summary
NPFC-2023-TWG CMSA07-RP02	2nd intersessional TWG CMSA meeting summary

WORKING PAPERS

Symbol	Title
NPFC-2023-TWG CMSA07-WP01 (Rev. 1)	TWG CMSA Work Plan, 2023-2027
NPFC-2023-TWG CMSA07-WP02	Consolidation and review of performance measures, detailed summary of model performance and ranking of the stock assessment model candidates
NPFC-2023-TWG CMSA07-WP03	Recommendation to TWG CMSA07 on a stock assessment model platform based on work completed as part of the TWG CMSA assessment modeling project
NPFC-2023-TWG CMSA07-WP04	Standardized CPUE of Chub mackerel (*Scomber japonicus*) caught by the Russia's trawls fishery up to 2021
NPFC-2023-TWG CMSA07-WP05 (Rev. 1)	New predictors for tracking the habitat of chub mackerel (*Scomber japonicus*)
NPFC-2023-TWG CMSA07-WP06	Simulated responses of summary performance metrics to varying model complexity: Which metric to be used?
NPFC-2023-TWG CMSA07-WP07	On the description and flexibility of state-space assessment model
NPFC-2023-TWG CMSA07-WP08	Simple simulations for determining the way for

	aging and defining fishing year in the stock assessment of chub mackerel
NPFC-2023-TWG CMSA07-WP09	Monthly catch data and distribution of chub mackerel fishing grounds by Chinese purse seine fleet
NPFC-2023-TWG CMSA07-WP10	Monthly catch data on the purse seine and pelagic trawl fleet in China
NPFC-2023-TWG CMSA07-WP11	Standardized CPUE of Chub mackerel (*Scomber japonicus*) caught by the China's lighting purse seine fishery up to 2022
NPFC-2023-TWG CMSA07-WP12 (Rev. 1)	Data description on fisheries bycatch in the chub mackerel fisheries in China
NPFC-2023-TWG CMSA07-WP13	The comparison of weight-at-age and maturity-at-age of chub mackerels
NPFC-2023-TWG CMSA07-WP14	SAM configuration options
NPFC-2023-TWG CMSA07-WP15	CPUE Standardization Protocol for Chub Mackerel
NPFC-2023-TWG CMSA07-WP16	Stock Assessment Protocol for Chub Mackerel

INFORMATION PAPERS

Document Number	Title
NPFC-2023-TWG CMSA07-IP01	Russian Mackerel fishery in the northwest Pacific Ocean in 2022
NPFC-2023-TWG CMSA07-IP02	Review of chub mackerel fishery in China and research activities
NPFC-2023-TWG CMSA07-IP03	Content of the document for data description on the chub mackerel in China
NPFC-2023-TWG CMSA07-IP04	Recent fishery and stock status of chub mackerel from Japan
NPFC-2023-TWG CMSA07-IP05	Fisheries Sampling Designs
NPFC-2023-TWG CMSA07-IP06	Details on sampling design and estimation of chub mackerel commercial catch

Annex C

List of Participants

CHAIR

Kazuhiro OSHIMA
oshima_kazuhiro28@fra.go.jp

CANADA

Janelle CURTIS
Janelle.Curtis@dfo-mpo.gc.ca

Chris ROOPER
chris.rooper@dfo-mpo.gc.ca

CHINA

Qiuyun MA
qyma@shou.edu.cn

Libin DAI
644318716@qq.com

Yongchuang SHI
1024731143@qq.com

Heng ZHANG
zhangh1@ecsf.ac.cn

EUROPEAN UNION

Karolina MOLLA GAZI
karolina.mollagazi@wur.nl

JAPAN

Shuya NAKATSUKA
nakatsuka_shuya49@fra.go.jp

Takumi FUKUDA
takumi_fukuda720@maff.go.jp

Momoko ICHINOKAWA
ichinokawa_momoko09@fra.go.jp

Ken ISHIDA
ishida_ken79@fra.go.jp

Akihiro MANABE
manabe_akihiro97@fra.go.jp

Chiaki MIZUGAKI
chiaki_mizugaki010@maff.go.jp

Shota NISHIJIMA
nishijima_shota02@fra.go.jp

Yukiya UCHIDA
yukiya_uchida230@maff.go.jp

Kyutaro YASUMOTO
kyutaro_yasumoto890@maff.go.jp

Ryuji YUKAMI
yukami_ryuji41@fra.go.jp

RUSSIA

Vladimir KULIK
vladimir.kulik@tinro.ru

Igor CHERNIENKO
chernienko.igor@gmail.com

USA

Erin BOHABOY
erin.bohaboy@noaa.gov

VANUATU

Lucy Andrea JOY
ljoy@fisheries.gov.vu

Falma AIVIJI
falma.pala19@gmail.com

Jayven HAM
jham@fisheries.gov.vu

Jeyalda NGWELE
njeyalda@fisheries.gov.vu

Mei-Chin JUAN
meichin.mdfc@gmail.com

Kevin LIN
kevin.mdfc@msa.hinet.net

INVITED EXPERT

Joel RICE
ricemarineanalytics@gmail.com

RAPPORTEUR

Alex MEYER
meyer@urbanconnections.jp

SECRETARIAT

Robert DAY
rday@npfc.int

Alex ZAVOLOKIN
azavolokin@npfc.int

Sungkuk KANG
skang@npfc.int

Natsuki HOSOKAWA
nhosokawa@npfc.int

Annex D

Table of qualitative attributes of the candidate stock assessment models

Qualitative property	**ASAP**	**VPA**	**SAM**	**KAFKA**	**BSSPM**
Extensibility*	Yes	VPA is not a flexible model partly because of the strong assumption of no catch measurement error and its extensibility is not so high	SAM is estimated by a complex mixed-effects (latent-variable) model and its configurations can be flexibly customized for improvements.	Yes	Yes
Shareability**	Yes	The computer codes for VPA that Japan uses are publicly available via GitHub and can be shared as an R package	The computer codes for SAM that Japan uses are publicly available via GitHub and can be shared as an R package	It is necessary to solve some internal issues	Yes
Reproducibility***	Yes	The population dynamics of VPA is described as backward deterministic equations and its optimization is conducted by the (penalized) likelihood method. It is reproducible even by a spreadsheet like Excel.	The population dynamics of SAM is described as stochastic equations and its optimization is conducted by the (marginal) likelihood method. SAM uses a C++ code for TMB to estimate a mixed-effects model and requires moderate technical skill.	Yes	Yes
Computer language****	ADMB	R, TMB	R, TMB	C#, Python, C	R

Usage example*****	Eastern Atlantic and Mediterranean bluefin tuna (Carrano et al. 2022); ATLANTIC BLUEFIN TUNA-ICCAT 2020 Southern Flounder in the South Atlantic, 1989–2017 (Flowers et al. 2019)-NCDMF SAP-SAR-2019-01 Eastern Georges Bank Atlantic Cod (L. O'Brien and Y. Wang, 2013) Bluefish-NOAA (Shepherd and Nieland 2010) PACIFIC MACKEREL-NOAA (Crone et al. 2006)	VPA is used for about 20 stocks in USA1 and about 30 stocks in Japan2. The ridge VPA3 is used for several stocks in Japan including the Pacific stock of chub mackerel	SAM is used for approximately 33 stocks in Europe (ICES)	East Sakhalin pollock, Kuril-Kamchatka navaga, Southern Kurils navaga, bream, zander, roach, sabre carp of Curonian Lagoon and Vistula Lagoon	NPFC Pacific saury Indian Ocean albacore (Li et al. 2016)
Others******		Links: 1, 2, 3	Links: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	Suitable for noisy data and unreported catch situation	

*Whether model configurations can flexibly be extended for model improvements
**Whether computer codes can be shared and publicly available
***Whether anyone can re-run model based on model description
****e.g., R, ADMB, TMB
*****Examples of stock assessments using the model
******Other qualitative properties (optional, important properties can be listed)

Annex E

Consolidation and review of performance measures, detailed summary of model performance and ranking of the stock assessment model candidates

SUMMARY

This paper presents work conducted since TWG CMSA 06 (2022), via correspondence and two intersessional meetings to finalize the selection of methods for the selection of performance measures, summary metrics, their consolidation, evaluation and the resulting ranking of the candidate stock assessment models.

Introduction

At the end of TWG CMSA 06 (2022) participants were unable to agree on which Performance Measures (PM) should be used to rank candidate stock assessment models. Therefore, in preparation for TWG CMSA 07 a draft ranking of PMs was forwarded to the participants, who were then asked to revise and submit the rankings. The goal of this exercise was to begin discussions toward finalizing performance measures and metrics for selecting a candidate stock assessment model in preparation for TWG CMSA 07 (see Appendix A for the timeline of work).

Member nations were requested to rank the performance measures on a scale of 1-4, where lower numbers (i.e. 1) indicate higher priority performance measures. The requested ranking was not considered final, with the assumption that the finalization of PMs would happen via email communication in the month of March 2023. Upon receipt of the draft ranking's, summary statistics of mean, median, standard deviation (S.D.), and 1/sum of estimates were compiled. The use the SD of the ranking was to help interpret the rankings, in part based on magnitude of the S.D. Where a S.D. of 0 means that all countries are in agreement on the priority, a larger S.D. indicates lower consensus. on the measure.

In preparation for TWG CMSA 07 the Group worked via email to develop methods and ensure that all data submitted was correct. Following development of ranking methods and the performance of stock assessment candidate models the group convened the 1st intersessional meeting of the TWG CMSA (May 2023). The methods and discussion of this is outlined in the meeting report, this report covers the majority of agenda and focuses on;

- Methods for Performance Measure Rankings
- Evaluation and Summary of Model Performance and Ranking of the Stock Assessment Model Candidates and Their Characteristics

1. Methods

1.1 Summary of Online Discussion Regarding PM Ranking

Since the initial distribution of the candidate ranking several questions or topics were raised regarding the potential methods. The largest part of the discussion related to the utilization of retrospective analysis and the value of Mohn's Rho.

The retrospective analysis is the only indicator that has been chosen as a performance metric, and can be a helpful diagnostic tool for stock assessment. Typically, the severity of a retrospective pattern was based on the range provided by Hurtado-Ferro et al. (2015), with values higher than 0.20 and lower than -0.15 used as an indication for problematic retrospective patterns. Mohn's rho is a useful proxy that can reflect model inconsistency with respect to the stability of estimates, but it is difficult to interpret its meaning in simulation work. Within the framework of this analysis, any model inconstancy means that the assumed structure of the assessment model used differs from the true process, and that the indicator is not exactly equal to what we want to know. Regarding the retrospective analysis, a recent article explained that it cannot be used as a model validation tool (Kell et al. 2021). Specifically, Kell et al. (2021) note that;

> "A diagnostic tool to check the potential future stability of stock assessment models is retrospective analysis (Mohn, 1999).. Retrospective analysis is widely used to evaluate the stability of model outputs, and in Europe is often the key diagnostic for accepting or rejecting a model (Orio et al., 2019)... However, stability and a reduction in variance can be achieved at the expense of bias by shrinking terminal estimates towards recent historical values. It is impossible to validate a model if bias is unknown, as is the case for unobservable quantities, such as SSB and F (Hodges and Dewar, 1992); since in such cases, the simplest way to remove a retrospective pattern is to ignore the data."

Due to the fact that this project uses a simulation approach in which state variables such as SSB and F are known and biases between estimated and true values can be measured the group considered that higher priority should be placed on relative errors than Mohn's rho because evaluation of biases via simulation is possible. Model validation with Mohn's Rho only is not straightforward. Complicating this is that retrospective analysis will be affected by variability in catch, CPUE, and other model inputs.

Upon receipt of the draft ranking's summary statistics of mean median and 1/sum of estimates were compiled. A suggestion is that to interpret the rankings in part based on the S.D. Where S.D. of 0 means that all countries get agreement on the priority, a larger S.D. means more needs to be discussed on the measure

Based on the average ranking the best summary statistic was the Total Biomass in years 2010 – 2019 (Tby2010 - TB2019), followed by the recruitment (Ry2010-Ry2019) the Fishing Mortality (AFy2010 - AFy2019), and the Exploitation Rate(Ey2010 - Ey2019). biological reference point of Bmsy_0.7_BH_1618, and the relative fishing impact RF_Fmsy_0.7_BH_1618.

The method selected is to follow the rankings of the highest ranked performance measures, namely the:

- Total Biomass in years 2010 – 2019 (Tby2010 -TB2019),
- Recruitment (Ry2010-Ry2019),
- Fishing Mortality (AFy2010 - AFy2019),
- Exploitation Rate (Ey2010 - Ey2019)
- Biological reference point of Bmsy_0.7_BH_1618, T
- Relative fishing impact RF_Fmsy_0.7_BH_1618.

These measures can be summarized by the median absolute relative bias (MedARB).

1.2 Summary of Intercessional work on the methods and ranking of performance measures, as well as metrics

Details of the ranking of performance measure and choice of summary metrics were developed through intersessional work conducted since the TWG CMSA06 meeting in September 2022 (See NPFC TWG CMSA WP 01, 2023) Remaining discussions points decided during the intercessional meeting were related to the methods for selecting the Performance Measures and Summary Mmetrics, as well as aggregating Summary Metrics. The following Topics (A-G) covered the main decisions needed to rank and evaluate the models, some of which were discussed intercessionally:

- *Topic A. Which scenarios to use (A-E)?* The group had recalled that previous discussion recommended the use of the Scenarios A and B (see **Table 1** Scenarios for chub mackerel assessment models) which are the two base case scenarios for scoring the performance metrics. These scenarios differ in that Scenario A uses a single natural mortality estimate for all ages while Scenario B age specific natural mortality estimates

- *Topic B Which Summary Metrics (MedARB, RMSE, MedRB, or CV) to use?* The group noted that the decision to use the median absolute relative bias (MedARB) which is a metric of both bias and precision, and robust to outliers was discussed intersessionally (TWG_CMSA INT 01_WP01). Additionally, a presentation on simulated responses of the candidate summary metrics to varying model complexity to inform decision on which metrics (Nishijima and Ichinokawa, 2023) was discussed. This paper indicated that summary metrics responded differently to varying model complexity, and (2) the median relative bias (MedRB) and the median absolute relative bias (MedARB) showed better scores when the model complexity is correct than the coefficient of variation (CV) and root mean squared error (RMSE). The group agreed that using MedARB as the priority metric for ranking of the models, and subsequently using MedRB to understand the likelihood of risky situations from the direction and magnitude of bias if warranted.

- *Topic C. How to Aggregate Performance Measures for a chosen Summary Metric (i.e. median, mean sum, etc.).* Participants agreed that for each scenario, a geometric mean will be used over the chosen PMs and data models to aggregate SMs (see the flowchart below for details). In the case of zero value of any SM, median value will be applied.

- *Topic D Confirm the Performance Measures to use in the raking of the assessment models.* Participants reviewed weighing of different tiers of the PMs based on the responses from Members submitted intersessionally and confirmed the list of performance measures to include in the ranking (Table 2).

- *Topic E How do we evaluate the difference between self and cross tests, should they carry equal weight?* Participants discussed how to evaluate the difference between self and cross tests. The consensus was to prioritize the self-test because lack of robustness in self-tests may be indicative of bias, whereas lack of robustness in cross-tests might be expected due to differences between models. Participants agreed to give self-test a priority, therefore models would be ranked based on their performance as an estimation model, with the weights assigned only to the self-test. Participants agreed to use weighted geometric mean (Eq. 1) with different weighting of self-test of 1 to 3 with 0.5 intervals (1, 1.5, 2, 2.5, 3), as a diagnostic to investigate the effect of weighting the self-tests.

Equation 1.

$$Weighted\ Geomean\ of\ MedARB\ (X)\ of\ model\ j = \exp\left[\frac{\sum_k\left[\sum_i\left\{w_{i,j} \times \log\left(X_{i,j,k}\right)\right\}\right]}{\sum_k\left[\sum_i w_{i,j}\right]}\right]$$

$$w_{i,j} = \begin{cases} 1, 1.5, 2, 2.5 \text{ or } 3, & i = j \text{ (self test)} \\ 1, & i \neq j \text{ (cross test)} \end{cases}$$

Where X is the MedARB (lower values indicate better performance), the subscripts i and j represent a data model and an estimation model, respectively., the subscript k represents a performance measure.

- Topic F. How to interpret the values of Mohn’s Rho from the model runs. Participants considered retrospective analysis (Mohn’s Rho as a summary metric) as an additional performance measure. Members discussed recent papers (Hurtado -Ferro et al 2014, Kell et al, 2021) and noted the proposed rule of thumb for values of Mohn's Rho which can be used to determine whether a stock assessment shows a retrospective pattern. Participants recognized the importance of retrospective analysis in model diagnostics when conducting stock assessment, and noted that retrospective analysis is a measure of the stability of model estimates, and not necessarily model fit. The group decided not to include Mohn’s Rho in the list of performance measures for the selection of a stock assessment model.

- *Topic G. How to deal with model iterations which do not converge*. The group discussed how to include or exclude model runs that did not converge for any given estimation model. Participants noted that in stock assessment modeling models that did not initially converge may converge with minor revisions to parameters or estimation techniques. The confirmed the suggestion made at SWG OM04 in August 2022 to exclude iterations that never converged on a model-specific basis. The information about the percentage of converged initially non-converged iterations is in the summary table of candidate model properties.

2. Results

Ranking of performances of the SA model candidates

The stock assessment model candidates were ranked based on the methods agreed to by the TWG CMSA. The results of the comparison of the performance metrics for Scenario A and B are shown in Tables 3 and 4, detailed results are shown in Appendix B, and are available on the Collaboration Website. Based on the overall summary weighted geometric mean of the Med. RAB, which was the summary metric chosen by the group, model rankings (from least to most biased) were as follows; 1) SAM, 2) VPA, 3) KAFKA, 4) ASAP, (Table 3). These rankings were calculated under equal weighting and ranking results were the same across the range of weights investigated.

In addition to having the lowest overall summary score, the SAM assessment model also had the lowest self-test value in both Scenarios A and B. In the cross test for both scenarios, where the SAM model was fit to data from other models, it outperformed the ASAP model (i.e. fitting to data from ASAP), but not the KAFKA model. The SAM model essentially tied the VPA in scenario B when both were fit to data from the VPA model. With respect to cross tests the SAM model fit the data best to the VPA and then ASAP and then KAFKA (Scenario A) and VPA, KAFKA, ASAP (Scenario B).

3. Conclusions

Qualitative attributes of the stock assessment model candidates were discussed, along with the strengths and weaknesses of the candidate stock assessment models. The group recommendation of selection of the SA model was based on the rankings and results of the pseudo data analysis. The results indicated that the stock assessment models ranked SAM, VPA, KAFKA, ASAP in ascending order of MedARB across both base case scenarios (A & B). This indicates that SAM is the least biased model considered in this analysis.

The group noted that the project strengths and weaknesses included both extensive analysis in the assessment model assumptions, simulations and high confidence in the comparability and accuracy of the results (strength), however overall, there was a relatively low sample size of the total number of simulations (weakness). Based on the work completed as part of this project, the SAM model has the best performance and is therefore the least biased model (of those considered) for stock assessment of chub mackerel.

Based on the comparison, with respect to the models investigated in this project and the goal of identifying a modeling platform, the SAM should be the recommended model for stock assessment of chub mackerel. The participants recognized that aggregation of summary metrics (across scenarios) was not needed because the ranking results did not change between scenarios A and B.

4. References

Deroba, J., Butterworth, D., Methot, Rick….Hulson, P.. (2015). Simulation testing the robustness of stock assessment models to error: Some results from the ICES strategic initiative on stock assessment methods. ICES Journal of Marine Science: Journal du Conseil. 72. 10.1093/icesjms/fst237.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., ... & Punt, A. E. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 72(1), 99-110. https://academic.oup.com/icesjms/article/72/1/99/2804338

Hodges J. S., Dewar J. A. 1992. Is it you or your model talking?: A framework for model validation. Santa Monica, CA: Rand.

Kell, L., Sharma, R., Kitakado, T., Winker, H., Mosquiera, I, Cardinale, M., Fu, D. 2023. Validation of stock assessment methods: is it me or my model talking?, ICES Journal of Marine Science, Volume 78, Issue 6, September 2021, Pages 2244–2255, https://doi.org/10.1093/icesjms/fsab104

Nishijima, S. and Ichinokawa, M. 2023. Simulated responses of summary performance metrics to varying model complexity: Which metric to be used? Available on the TWG CMSA intersessional meetings group Collaboration website. TWG CMSA INT 01 WP02.

NPFC TWG CMSA. 2023. Final Methods for Performance Measure Ranking and Evaluation. Meeting Document. TWG CMSA INT 01 WP01.

Orio A., Karpushevskaia A., Nielsen A., Sundelöf A., Berg C.W., Albertsen C.M., Stralka C. et al. 2019. Benchmark workshop on Baltic cod stocks (wkbaltcod2). ICES Reports 1.

5. Tables

Table 1. Alternative scenarios for the assessment of chub mackerel.

DATASET	SCENARIO	M	WEIGHT-AT-AGE	MATURITY-AT-AGE	CATCH (AT-AGE)	ABUNDANCE INDEX	FLEET
A	Base-case 1	0.41	Average	Average	Average	All six	Single
B	Base-case 2	Gislason	Average	Average	Average	All six	Single
C	Highest weight and maturity 1	0.41	Highest	Highest	Average	All six	Single
D	Highest weight and maturity 2	Gislason	Highest	Highest	Average	All six	Single
E	Lowest weight and maturity 1	0.41	Lowest	Lowest	Average	All six	Single
F	Lowest weight and maturity 2	Gislason	Lowest	Lowest	Average	All six	Single

Table 2. Performance Measures selected for ranking the candidate stock assessment models

Category	Performance Measure (PM)
state	Total Biomass in years 2010 – 2019 (Tby2010 -TB2019)
state	Recruitment (Ry2010-Ry2019)
state	Fishing Mortality (AFy2010 - AFy2019)
state	Exploitation Rate (Ey2010 - Ey2019)
BRP	Bmsy_0.7_BH_1618
BRP	Fmsy_0.7_BH_1618
Depletion	Deple_SSB_median_2010
Depletion	Deple_B_median_2010
RFI	RF_Fmsy_0.7_BH_1618
RFI	RF_F40SPR_1618
RFI	RF_F0.1_1618

Table 3. Summary metrics (Med ARB) from the performance measures selected.

		Overall Summary Weighted Geometric Mean of Med ARB	
Rank	**Estimation Model**	**Scenario A (Base Case 1)**	**Scenario B (Base Case 2)**
1	SAM	0.396	0.325
2	VPA	0.443	0.345
3	KAFKA	0.587	0.533
4	ASAP	0.599	0.614

Table 4. Summary metrics (Med ARB) from the performance measures selected. Diagonals show the self-tests and off diagonals show the cross tests for the various Estimation and data model comparisons. Bold numeric values show the lowest (least biased) self test values. Warmer (closer to red) colors indicate higher values.

Scenario A (Base Case 1)

	Data Model			
Estimation Model	**ASAP**	**KAFKA**	**SAM**	**VPA**
ASAP	0.68	0.49	0.70	0.54
KAFKA	0.75	0.46	0.48	0.72
SAM	0.64	0.62	**0.19**	0.33
VPA	0.52	0.46	0.52	0.31

Scenario B (Base Case 2)

	Data Model			
Estimation Model	**ASAP**	**KAFKA**	**SAM**	**VPA**
ASAP	0.70	0.56	0.74	0.49
KAFKA	0.63	0.48	0.47	0.56
SAM	0.52	0.61	**0.18**	0.19
VPA	0.42	0.47	0.38	0.19

Table 5. Percentage of model runs that did not initially converge by scenario and estimation model. Note that initial convrgeance does not reflect final convergence.

	Percentage Non-Convergence	
Estimation Model	Scenario A	Scenario B
VPA	0%	0%
SAM	12%	4%
KAFKA	8%	10%
ASAP	5%	5%

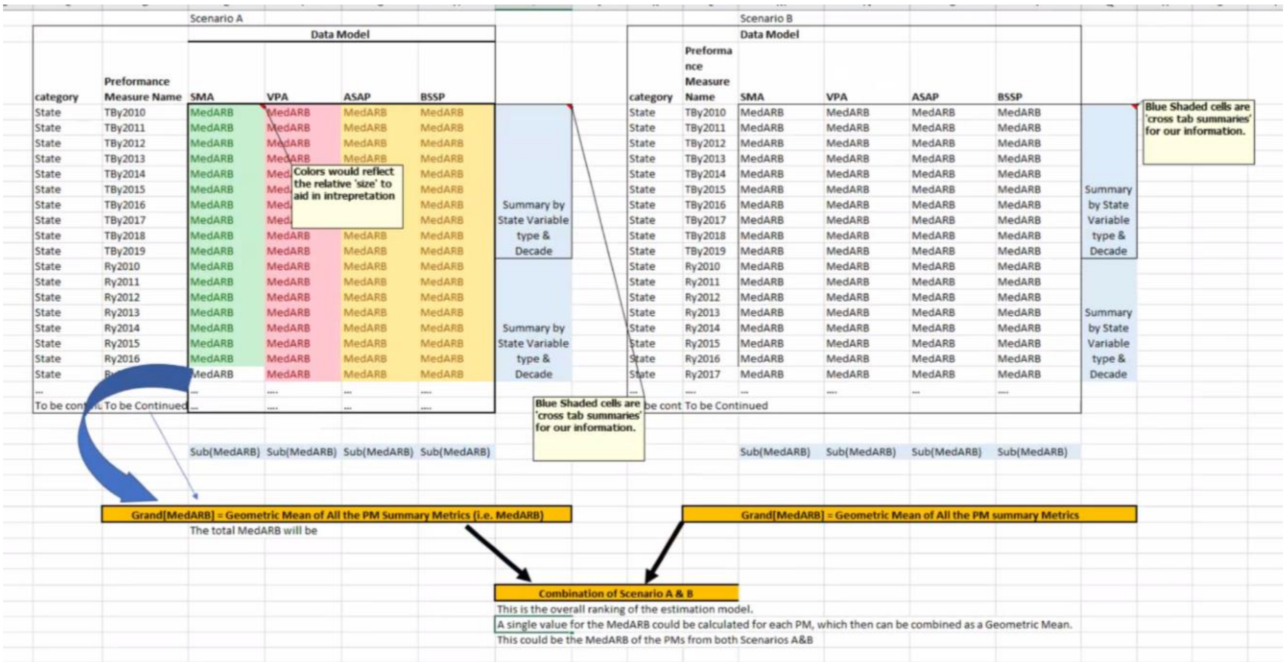

Figure 1. Diagram of how the performance measures and summary metrics (SMs) could be aggregated. Note: The list of SMs is incomplete and presented for demonstration purposes.

Appendix A. Updated timeline of intersessional work toward the selection of a model for stock assessment of chub mackerel

Month	Events	Notes
2023 Jan	Email communication	Start discussions toward finalizing performance measures/metrics
Feb		
Mar	Email communication	Finalize performance measures/metrics **BY MARCH 31**
April 15		Circulate **Final <u>Draft</u> 'Methods for Performance Measure Ranking and Evaluation',** methods for ranking candidate models based on tiered performance measures and chosen metrics.
April 21		Group members Submit Comments on **Final <u>Draft</u> 'Methods for Performance Measure Ranking and Evaluation'**
April 28 - May 15	Email communication	Group works to incorporate comments and creates a **Final 'Methods for Performance Measure Ranking and Evaluation'**
May 15 (approximately)		Joel Submit presentation on application of **'Methods for Performance Measure Ranking and Evaluation'** for model selection.
May 17-19	Intersessional meeting (2-3 days)	Topic & Goal: Rank the candidate stock assessment models and rank candidate models based on Performance Measures. 1) Discuss /present application of **'Methods for Performance Measure Ranking and Evaluation'** for model selection. 2)Draft report on model selection, make recommendation for any future MSE use as an operational model, or stock assessment modeling. NOTE: The choice of the assessment model(s) will take place During the Assessment Phase.
June	Email communication	Finalize report
July	1- or 2-day intersessional	Possible topics

(Early July?)	meeting for finalization of report	• Presentation of report, address any other comments or concerns. LAST CHANCE TO COMMENT , • Possible use of ensemble modeling in the 'Assessment Phase' ○ Model Weighting, Structural Uncertainty, other ways to characterize uncertainty • Model sensitivity • Stability of parameters vs time varying parameters
Aug		**Meeting paper due one month before TWG CMSA07**
Sep 4-7th	TWG CMSA07 (Vanuatu)	Formalize the selection of the stock assessment model; Data preparation

Appendix B. Agenda of 1st Intersessional Meeting of the Technical Working Group on Chub Mackerel Stock Assessment, May 17-19, 2023 (9am – 1pm Tokyo time)

Agenda Item 1. Opening of the Meeting

Agenda Item 2. Adoption of Agenda

Agenda Item 3. Short summary of the TWG CMSA06 meeting and review of timeline

- Review of TWG CMSA06 data submissions
- Intersessional works done before the 1st intersessional meeting
    - Submission of input data to OMutility and retrospective analysis (Mohn's Rho)

Agenda Item 4. Review of the report on the ranking of performance measure and metrics

- Interpretation of self vs cross tests
- Weighting of PMs (Which PMs to use)
- Treatment of Mohn's rho (Review of strengths and weaknesses)
- Ranking of performance measures
    - How to interpret CV
- Performance metrics to use
    - How to aggregate Performance Metrics (i.e. mean or sum, how to treat self vs cross tests)

Agenda Item 5. Ranking of performances of the SA model candidates

Agenda Item 6. Qualitative attributes of the SA model candidates

Agenda Item 7. Wrap-up discussion

Agenda Item 8. Draft report on the SA model selection

8.1 Recommendation of selection of the SA model

8.2 Consideration of stock assessment approaches in other RFMOs (i.e. single model, reference case with sensitivities, ensemble modeling, etc.)

Agenda Item 9. Other matters

- Work assignment
    - Selection of SA model(s)
    - Input data and abundance indices for SA
- Dates of the 2nd intersessional meeting in July

Agenda Item 10. Close of the Meeting

Appendix C Details of the Performance measure summary for Scenario A

		Data Model				
Est.Model	Performance Measure	**ASAP**	KAFKA	SAM	VPA	PartialScore
ASAP	AFy2010	3.97	0.17	1.44	1.50	
ASAP	AFy2011	2.42	0.70	1.91	2.20	
ASAP	AFy2012	1.98	1.61	0.24	0.23	
ASAP	AFy2013	1.93	0.14	0.96	1.12	
ASAP	AFy2014	3.83	0.27	3.19	5.50	
ASAP	AFy2015	4.40	0.76	2.89	4.32	
ASAP	AFy2016	4.84	0.51	3.93	7.76	
ASAP	AFy2017	2.12	0.85	2.09	4.22	
ASAP	AFy2018	1.27	0.31	0.94	2.83	
ASAP	AFy2019	2.57	0.36	0.19	0.97	1.33
ASAP	Bmsy_0.7_BH_1618	0.53	0.18	0.43	0.61	0.40
ASAP	Deple_B_median_2010	0.28	0.04	0.72	0.14	0.18
ASAP	Deple_SSB_median_2010	0.54	0.18	1.41	0.36	0.48
ASAP	Ey2010	0.52	0.28	0.66	0.63	
ASAP	Ey2011	0.42	0.38	0.37	0.25	
ASAP	Ey2012	0.11	0.32	0.13	0.32	
ASAP	Ey2013	0.66	1.84	0.36	0.06	
ASAP	Ey2014	0.54	0.03	0.65	0.46	
ASAP	Ey2015	0.67	0.58	0.76	0.69	
ASAP	Ey2016	0.74	0.65	0.77	0.59	
ASAP	Ey2017	0.68	0.86	0.79	0.61	
ASAP	Ey2018	0.59	0.93	0.87	0.54	
ASAP	Ey2019	0.11	0.67	0.69	0.30	0.44
ASAP	Fmsy_0.7_BH_1618	0.04	0.48	0.04	0.27	0.12
ASAP	RF_F0.1_1618	0.18	0.41	0.34	0.36	0.31
ASAP	RF_F40SPR_1618	0.19	0.25	0.35	0.33	0.27
ASAP	RF_Fmsy_0.7_BH_1618	0.18	0.29	0.34	0.34	0.28
ASAP	Ry2010	0.53	0.23	0.24	0.19	
ASAP	Ry2011	1.91	0.24	0.34	0.26	
ASAP	Ry2012	0.67	0.20	0.27	0.14	
ASAP	Ry2013	0.80	0.84	0.25	0.16	
ASAP	Ry2014	2.49	0.15	0.26	0.10	
ASAP	Ry2015	0.30	0.10	0.35	0.12	
ASAP	Ry2016	0.34	0.61	0.45	0.51	
ASAP	Ry2017	0.15	4.55	0.53	0.23	
ASAP	Ry2018	0.54	7.65	0.72	0.68	
ASAP	Ry2019	0.24	1.74	0.84	1.00	0.42
ASAP	TBy2010	1.07	0.39	1.60	1.69	
ASAP	TBy2011	0.71	0.60	0.54	0.33	
ASAP	TBy2012	0.12	0.24	0.13	0.24	
ASAP	TBy2013	0.40	0.65	0.51	0.07	
ASAP	TBy2014	1.15	0.03	1.61	0.84	
ASAP	TBy2015	2.04	1.40	3.96	2.22	
ASAP	TBy2016	2.80	1.87	3.70	1.46	
ASAP	TBy2017	2.14	6.12	4.16	1.58	
ASAP	TBy2018	1.43	13.24	6.84	1.17	
ASAP	TBy2019	0.12	2.05	1.92	0.25	0.92
	Partial Score	0.68	0.49	0.70	0.54	
Overall Summary Weighted Geo Mean		0.599429				

Est.Model	name	Data Model				PartialScore
		ASAP	**KAFKA**	SAM	VPA	
KAFKA	AFy2010	2.54	0.44	2.28	2.19	
KAFKA	AFy2011	4.64	1.53	1.07	0.46	
KAFKA	AFy2012	3.48	0.63	1.03	1.13	
KAFKA	AFy2013	6.07	2.88	1.52	2.18	
KAFKA	AFy2014	4.95	0.65	2.75	5.04	
KAFKA	AFy2015	1.43	0.25	2.61	6.19	
KAFKA	AFy2016	4.76	0.56	1.96	2.02	
KAFKA	AFy2017	1.82	0.17	1.77	2.15	
KAFKA	AFy2018	3.16	1.99	0.86	2.53	
KAFKA	AFy2019	5.85	0.13	1.12	2.06	1.62
KAFKA	Bmsy_0.7_	0.40	0.63	0.35	0.58	0.48
KAFKA	Deple_B_m	0.34	0.59	0.15	0.53	0.36
KAFKA	Deple_SSB_	0.32	0.59	0.10	0.45	0.31
KAFKA	Ey2010	0.30	0.31	0.36	0.20	
KAFKA	Ey2011	0.32	0.19	0.33	0.18	
KAFKA	Ey2012	0.51	0.29	0.35	0.18	
KAFKA	Ey2013	0.69	0.99	0.38	0.31	
KAFKA	Ey2014	0.41	0.19	0.33	0.18	
KAFKA	Ey2015	0.31	0.31	0.25	0.53	
KAFKA	Ey2016	0.41	0.24	0.27	1.14	
KAFKA	Ey2017	0.45	0.28	0.35	1.89	
KAFKA	Ey2018	0.50	0.23	0.37	3.62	
KAFKA	Ey2019	0.26	0.60	0.34	2.03	0.39
KAFKA	Fmsy_0.7_I	0.80	0.08	1.50	1.65	0.63
KAFKA	RF_F0.1_16	2.31	0.08	1.86	4.15	1.11
KAFKA	RF_F40SPR	1.40	0.21	0.85	2.73	0.91
KAFKA	RF_Fmsy_0	1.63	0.17	1.00	3.02	0.96
KAFKA	Ry2010	0.38	0.77	0.29	0.23	
KAFKA	Ry2011	0.67	0.60	0.24	0.32	
KAFKA	Ry2012	0.87	0.75	0.27	0.35	
KAFKA	Ry2013	0.90	0.94	0.25	0.18	
KAFKA	Ry2014	1.21	0.69	0.28	0.29	
KAFKA	Ry2015	0.68	0.59	0.27	0.53	
KAFKA	Ry2016	0.74	0.42	0.31	0.83	
KAFKA	Ry2017	0.61	1.13	0.44	0.81	
KAFKA	Ry2018	0.78	1.46	0.46	0.94	
KAFKA	Ry2019	0.40	0.28	0.65	0.37	0.51
KAFKA	TBy2010	0.31	0.55	0.29	0.23	
KAFKA	TBy2011	0.33	0.52	0.27	0.16	
KAFKA	TBy2012	0.35	0.64	0.35	0.21	
KAFKA	TBy2013	0.61	0.85	0.57	0.47	
KAFKA	TBy2014	0.31	0.68	0.24	0.18	
KAFKA	TBy2015	0.31	0.61	0.20	0.34	
KAFKA	TBy2016	0.30	0.68	0.26	0.53	
KAFKA	TBy2017	0.25	0.42	0.22	0.65	
KAFKA	TBy2018	0.33	0.42	0.27	0.78	
KAFKA	TBy2019	0.24	0.21	0.29	0.67	0.36
	Partials	0.75	0.46	0.48	0.72	
Overall Summary		0.58731				

		Data Model				
Est.Model	name	ASAP	KAFKA	**SAM**	VPA	PartialScore
SAM	AFy2010	0.46	0.24	0.11	0.02	
SAM	AFy2011	0.61	0.03	0.14	0.14	
SAM	AFy2012	0.62	0.03	0.12	0.04	
SAM	AFy2013	0.58	0.26	0.17	0.09	
SAM	AFy2014	0.65	0.74	0.19	0.53	
SAM	AFy2015	0.74	0.33	0.16	0.30	
SAM	AFy2016	0.78	0.30	0.25	0.29	
SAM	AFy2017	0.80	0.80	0.29	0.45	
SAM	AFy2018	0.82	0.93	0.41	0.83	
SAM	AFy2019	0.85	0.90	0.62	0.54	0.30
SAM	Bmsy_0.7_	0.24	0.46	0.07	0.26	0.21
SAM	Deple_B_m	0.44	0.56	0.17	0.43	0.36
SAM	Deple_SSB_	0.49	0.43	0.19	0.38	0.35
SAM	Ey2010	0.40	0.13	0.13	0.22	
SAM	Ey2011	0.54	0.09	0.14	0.17	
SAM	Ey2012	0.51	0.69	0.12	0.43	
SAM	Ey2013	0.53	1.77	0.19	0.44	
SAM	Ey2014	0.60	0.38	0.19	0.47	
SAM	Ey2015	0.60	0.36	0.14	0.20	
SAM	Ey2016	0.66	0.64	0.19	0.64	
SAM	Ey2017	0.67	0.85	0.23	0.86	
SAM	Ey2018	0.72	0.87	0.33	2.21	
SAM	Ey2019	0.78	0.80	0.52	2.70	0.42
SAM	Fmsy_0.7_	0.77	0.23	0.15	0.38	0.32
SAM	RF_F0.1_16	0.99	0.66	0.44	0.90	0.71
SAM	RF_F40SPR	0.78	0.77	0.32	0.78	0.62
SAM	RF_Fmsy_0	0.79	0.77	0.31	0.80	0.63
SAM	Ry2010	0.45	0.76	0.13	0.06	
SAM	Ry2011	1.34	0.59	0.13	0.16	
SAM	Ry2012	0.68	0.71	0.11	0.24	
SAM	Ry2013	0.82	0.93	0.22	0.29	
SAM	Ry2014	2.54	0.56	0.18	0.14	
SAM	Ry2015	0.54	0.10	0.19	0.28	
SAM	Ry2016	0.63	1.78	0.22	0.66	
SAM	Ry2017	0.63	12.77	0.23	0.51	
SAM	Ry2018	0.65	28.03	0.34	0.79	
SAM	Ry2019	0.70	0.73	0.30	0.28	0.47
SAM	TBy2010	0.48	0.63	0.07	0.09	
SAM	TBy2011	0.48	0.64	0.07	0.13	
SAM	TBy2012	0.48	0.72	0.09	0.19	
SAM	TBy2013	0.58	0.89	0.15	0.27	
SAM	TBy2014	0.58	0.78	0.15	0.26	
SAM	TBy2015	0.58	0.58	0.17	0.28	
SAM	TBy2016	0.65	0.26	0.19	0.43	
SAM	TBy2017	0.67	1.69	0.24	0.49	
SAM	TBy2018	0.71	4.80	0.28	0.68	
SAM	TBy2019	0.69	4.39	0.29	0.69	0.40
	Partials	0.64	0.62	0.19	0.33	
Overall Summary		0.395724				

Est.Model	name	ASAP	KAFKA	SAM	**VPA**	PartialScore
VPA	AFy2010	0.59	0.53	0.19	0.03	
VPA	AFy2011	0.65	0.10	0.30	0.12	
VPA	AFy2012	1.02	0.40	0.31	0.14	
VPA	AFy2013	2.29	0.02	0.34	0.23	
VPA	AFy2014	1.97	1.27	0.28	0.27	
VPA	AFy2015	0.89	0.30	0.36	0.30	
VPA	AFy2016	0.36	0.48	0.38	0.33	
VPA	AFy2017	0.52	0.61	0.49	0.35	
VPA	AFy2018	1.37	0.48	0.50	0.39	
VPA	AFy2019	0.63	0.63	0.47	0.46	0.38
VPA	Bmsy_0.7_	0.10	0.55	0.46	0.24	0.28
VPA	Deple_B_m	0.19	0.46	0.46	0.49	0.37
VPA	Deple_SSB_	0.24	0.65	0.69	0.39	0.45
VPA	Ey2010	0.27	0.24	0.26	0.10	
VPA	Ey2011	0.28	0.10	0.31	0.15	
VPA	Ey2012	0.28	0.45	0.34	0.21	
VPA	Ey2013	0.45	1.49	0.35	0.24	
VPA	Ey2014	0.33	0.53	0.44	0.25	
VPA	Ey2015	0.37	0.19	0.42	0.28	
VPA	Ey2016	0.34	0.21	0.43	0.32	
VPA	Ey2017	0.42	0.39	0.46	0.33	
VPA	Ey2018	0.42	0.21	0.54	0.49	
VPA	Ey2019	0.62	0.68	0.58	0.53	0.34
VPA	Fmsy_0.7_	0.69	0.76	0.39	0.18	0.44
VPA	RF_F0.1_16	1.03	0.69	0.44	0.36	0.58
VPA	RF_F40SPR	0.54	0.47	0.38	0.34	0.43
VPA	RF_Fmsy_0	0.54	0.51	0.38	0.34	0.43
VPA	Ry2010	0.26	0.73	0.42	0.15	
VPA	Ry2011	1.58	0.56	0.56	0.28	
VPA	Ry2012	0.77	0.65	0.46	0.35	
VPA	Ry2013	0.83	0.93	0.40	0.34	
VPA	Ry2014	1.89	0.70	0.89	0.36	
VPA	Ry2015	0.51	0.51	0.88	0.37	
VPA	Ry2016	0.44	0.30	0.93	0.50	
VPA	Ry2017	0.74	1.72	1.07	0.51	
VPA	Ry2018	0.39	2.61	1.45	1.25	
VPA	Ry2019	2.30	0.64	1.51	0.48	0.65
VPA	TBy2010	0.32	0.59	0.38	0.14	
VPA	TBy2011	0.28	0.60	0.52	0.20	
VPA	TBy2012	0.26	0.68	0.52	0.28	
VPA	TBy2013	0.56	0.88	0.50	0.33	
VPA	TBy2014	0.26	0.78	0.50	0.36	
VPA	TBy2015	0.25	0.66	0.66	0.42	
VPA	TBy2016	0.38	0.66	0.96	0.48	
VPA	TBy2017	0.50	0.30	1.17	0.51	
VPA	TBy2018	0.43	0.14	1.56	0.93	
VPA	TBy2019	1.01	0.14	1.78	1.09	0.48
	Partials	0.52	0.46	0.52	0.31	
Overall Summary		0.443212				

Appendix D Details of the Performance measure summary for Scenario B

Scenario B (Base Case 2)		Data Model				
Est.Model	name	ASAP	KAFKA	SAM	VPA	PartialScore
ASAP	AFy2010	3.63	0.70	1.34	1.38	
ASAP	AFy2011	2.43	1.86	2.11	2.06	
ASAP	AFy2012	1.80	4.40	0.27	0.12	
ASAP	AFy2013	1.75	1.56	1.05	0.80	
ASAP	AFy2014	3.59	1.33	3.43	4.24	
ASAP	AFy2015	3.81	0.12	2.82	3.12	
ASAP	AFy2016	3.87	0.50	3.68	5.58	
ASAP	AFy2017	2.10	0.66	2.16	3.02	
ASAP	AFy2018	1.47	0.63	1.20	2.10	
ASAP	AFy2019	2.43	1.53	0.29	0.53	1.49
ASAP	Bmsy_0.7_	0.68	0.63	0.66	0.76	0.68
ASAP	Deple_B_m	0.22	0.53	0.68	0.12	0.31
ASAP	Deple_SSB_	0.66	0.59	1.80	0.85	0.88
ASAP	Ey2010	0.55	0.03	0.66	0.66	
ASAP	Ey2011	0.51	0.17	0.39	0.32	
ASAP	Ey2012	0.11	0.93	0.15	0.28	
ASAP	Ey2013	0.83	3.22	0.30	0.14	
ASAP	Ey2014	0.58	0.27	0.65	0.52	
ASAP	Ey2015	0.69	0.48	0.75	0.73	
ASAP	Ey2016	0.74	0.57	0.77	0.67	
ASAP	Ey2017	0.71	0.84	0.81	0.70	
ASAP	Ey2018	0.64	0.91	0.90	0.66	
ASAP	Ey2019	0.10	0.52	0.69	0.18	0.47
ASAP	Fmsy_0.7_l	0.05	0.32	0.07	0.18	0.12
ASAP	RF_F0.1_16	0.15	0.03	0.26	0.14	0.12
ASAP	RF_F40SPR	0.20	0.29	0.28	0.10	0.20
ASAP	RF_Fmsy_0	0.19	0.19	0.28	0.11	0.18
ASAP	Ry2010	0.48	0.56	0.32	0.22	
ASAP	Ry2011	1.96	0.10	0.28	0.20	
ASAP	Ry2012	0.68	0.21	0.27	0.08	
ASAP	Ry2013	0.80	0.91	0.20	0.07	
ASAP	Ry2014	2.52	0.39	0.29	0.08	
ASAP	Ry2015	0.28	0.14	0.32	0.11	
ASAP	Ry2016	0.35	0.18	0.52	0.41	
ASAP	Ry2017	0.16	3.07	0.57	0.13	
ASAP	Ry2018	0.51	4.74	1.09	0.58	
ASAP	Ry2019	0.19	1.09	0.64	1.41	0.38
ASAP	TBy2010	1.24	0.03	1.53	1.93	
ASAP	TBy2011	1.05	0.21	0.60	0.48	
ASAP	TBy2012	0.11	0.48	0.19	0.22	
ASAP	TBy2013	0.45	0.76	0.36	0.16	
ASAP	TBy2014	1.39	0.21	1.57	1.09	
ASAP	TBy2015	2.23	0.91	3.75	2.74	
ASAP	TBy2016	2.85	1.31	3.88	1.99	
ASAP	TBy2017	2.47	5.11	4.56	2.39	
ASAP	TBy2018	1.80	9.92	8.84	1.98	
ASAP	TBy2019	0.11	1.08	1.85	0.17	0.95
	Partials	0.70	0.56	0.74	0.49	
Overall Summary Weighted Geo Mean		0.613624				

Scenario B (Base Case 2)		Data Model				
Est.Model	name	ASAP	KAFKA	SAM	VPA	PartialScor
KAFKA	AFy2010	1.77	0.72	1.49	1.58	
KAFKA	AFy2011	3.67	1.71	0.63	0.06	
KAFKA	AFy2012	2.64	2.66	0.66	0.50	
KAFKA	AFy2013	3.60	3.61	1.07	1.06	
KAFKA	AFy2014	3.66	0.70	2.30	3.16	
KAFKA	AFy2015	0.79	0.77	1.78	2.51	
KAFKA	AFy2016	1.51	0.26	0.84	0.55	
KAFKA	AFy2017	1.42	0.07	1.43	1.12	
KAFKA	AFy2018	2.55	1.64	0.74	1.28	
KAFKA	AFy2019	4.32	0.34	0.70	0.85	1.13
KAFKA	Bmsy_0.7_BH_	0.43	0.70	0.48	0.66	0.56
KAFKA	Deple_B_medi	0.37	0.62	0.19	0.47	0.38
KAFKA	Deple_SSB_me	0.32	0.63	0.11	0.36	0.30
KAFKA	Ey2010	0.27	0.16	0.33	0.31	
KAFKA	Ey2011	0.31	0.30	0.31	0.22	
KAFKA	Ey2012	0.42	0.61	0.38	0.35	
KAFKA	Ey2013	0.61	2.24	0.39	0.48	
KAFKA	Ey2014	0.46	0.44	0.26	0.21	
KAFKA	Ey2015	0.24	0.22	0.30	0.24	
KAFKA	Ey2016	0.24	0.18	0.33	0.70	
KAFKA	Ey2017	0.31	0.31	0.27	1.08	
KAFKA	Ey2018	0.31	0.14	0.49	2.22	
KAFKA	Ey2019	0.30	0.62	0.29	0.74	0.37
KAFKA	Fmsy_0.7_BH_	0.64	0.05	1.15	1.06	0.44
KAFKA	RF_F0.1_1618	2.16	0.09	1.97	3.27	1.06
KAFKA	RF_F40SPR_16	1.28	0.11	0.73	1.90	0.67
KAFKA	RF_Fmsy_0.7_E	1.47	0.10	0.90	2.13	0.73
KAFKA	Ry2010	0.30	0.84	0.31	0.20	
KAFKA	Ry2011	1.10	0.70	0.25	0.20	
KAFKA	Ry2012	0.82	0.74	0.22	0.21	
KAFKA	Ry2013	0.89	0.97	0.23	0.13	
KAFKA	Ry2014	1.59	0.80	0.28	0.18	
KAFKA	Ry2015	0.59	0.71	0.30	0.38	
KAFKA	Ry2016	0.67	0.62	0.29	0.77	
KAFKA	Ry2017	0.57	0.41	0.33	0.74	
KAFKA	Ry2018	0.73	0.72	0.49	0.91	
KAFKA	Ry2019	0.40	0.20	0.70	0.93	0.47
KAFKA	TBy2010	0.26	0.63	0.41	0.39	
KAFKA	TBy2011	0.35	0.62	0.34	0.29	
KAFKA	TBy2012	0.33	0.65	0.69	0.56	
KAFKA	TBy2013	0.58	0.88	0.67	0.95	
KAFKA	TBy2014	0.21	0.73	0.26	0.24	
KAFKA	TBy2015	0.27	0.66	0.35	0.23	
KAFKA	TBy2016	0.22	0.74	0.39	0.40	
KAFKA	TBy2017	0.20	0.48	0.27	0.51	
KAFKA	TBy2018	0.23	0.52	0.53	0.68	
KAFKA	TBy2019	0.22	0.20	0.31	0.42	0.40
	Partials	0.63	0.48	0.47	0.56	
Overall Summary Weighted Geo Mean		0.532565				

Scenario B (Base Case 2		Data Model				
Est.Model	name	ASAP	KAFKA	SAM	VPA	PartialScore
SAM	AFy2010	0.31	0.04	0.08	0.03	
SAM	AFy2011	0.42	0.27	0.12	0.05	
SAM	AFy2012	0.48	0.29	0.12	0.13	
SAM	AFy2013	0.45	0.63	0.12	0.10	
SAM	AFy2014	0.51	1.51	0.22	0.25	
SAM	AFy2015	0.66	0.87	0.17	0.07	
SAM	AFy2016	0.65	0.03	0.27	0.07	
SAM	AFy2017	0.71	0.65	0.26	0.12	
SAM	AFy2018	0.71	0.90	0.43	0.38	
SAM	AFy2019	0.63	0.88	0.70	0.18	0.26
SAM	Bmsy_0.7_	0.14	0.62	0.08	0.25	0.21
SAM	Deple_B_m	0.31	0.37	0.18	0.37	0.30
SAM	Deple_SSB_	0.48	0.49	0.20	0.24	0.32
SAM	Ey2010	0.29	0.33	0.12	0.13	
SAM	Ey2011	0.44	0.68	0.11	0.04	
SAM	Ey2012	0.40	1.62	0.11	0.22	
SAM	Ey2013	0.29	3.73	0.20	0.27	
SAM	Ey2014	0.49	1.33	0.20	0.30	
SAM	Ey2015	0.58	0.03	0.13	0.05	
SAM	Ey2016	0.56	0.48	0.21	0.38	
SAM	Ey2017	0.61	0.80	0.19	0.51	
SAM	Ey2018	0.61	0.84	0.29	1.61	
SAM	Ey2019	0.62	0.71	0.53	2.02	0.35
SAM	Fmsy_0.7_I	0.73	0.01	0.14	0.19	0.13
SAM	RF_F0.1_16	0.99	0.53	0.63	0.46	0.63
SAM	RF_F40SPR	0.76	0.56	0.33	0.38	0.48
SAM	RF_Fmsy_0	0.77	0.58	0.34	0.40	0.49
SAM	Ry2010	0.38	0.85	0.10	0.06	
SAM	Ry2011	1.54	0.70	0.13	0.05	
SAM	Ry2012	0.70	0.73	0.15	0.11	
SAM	Ry2013	0.83	0.97	0.20	0.21	
SAM	Ry2014	2.70	0.74	0.13	0.09	
SAM	Ry2015	0.34	0.38	0.16	0.15	
SAM	Ry2016	0.40	0.44	0.21	0.59	
SAM	Ry2017	0.53	6.53	0.20	0.41	
SAM	Ry2018	0.33	14.79	0.31	0.73	
SAM	Ry2019	0.72	0.19	0.27	0.26	0.38
SAM	TBy2010	0.31	0.73	0.10	0.03	
SAM	TBy2011	0.41	0.73	0.09	0.04	
SAM	TBy2012	0.29	0.78	0.11	0.06	
SAM	TBy2013	0.56	0.93	0.14	0.17	
SAM	TBy2014	0.39	0.85	0.14	0.15	
SAM	TBy2015	0.36	0.72	0.16	0.17	
SAM	TBy2016	0.52	0.57	0.19	0.34	
SAM	TBy2017	0.62	0.66	0.22	0.38	
SAM	TBy2018	0.60	2.44	0.27	0.62	
SAM	TBy2019	0.65	1.92	0.29	0.62	0.32
	Partials	0.52	0.61	0.18	0.19	
Overall Summary Weighted Geo Mean		0.32489				

Scenario B (Base Case		Data Model				
Est.Model	name	ASAP	KAFKA	SAM	VPA	PartialScore
VPA	AFy2010	0.51	0.06	0.20	0.01	
VPA	AFy2011	0.30	0.79	0.20	0.04	
VPA	AFy2012	0.78	0.01	0.22	0.05	
VPA	AFy2013	1.85	0.93	0.31	0.10	
VPA	AFy2014	1.30	2.53	0.25	0.16	
VPA	AFy2015	0.60	0.38	0.28	0.20	
VPA	AFy2016	0.40	0.04	0.36	0.25	
VPA	AFy2017	0.52	0.37	0.30	0.30	
VPA	AFy2018	0.93	0.34	0.45	0.34	
VPA	AFy2019	0.48	0.28	0.49	0.47	0.27
VPA	Bmsy_0.7_	0.11	0.70	0.41	0.17	0.27
VPA	Deple_B_m	0.17	0.55	0.28	0.30	0.30
VPA	Deple_SSB_	0.21	0.69	0.40	0.32	0.37
VPA	Ey2010	0.17	0.29	0.21	0.02	
VPA	Ey2011	0.19	0.80	0.24	0.06	
VPA	Ey2012	0.22	1.55	0.22	0.10	
VPA	Ey2013	0.72	4.40	0.33	0.13	
VPA	Ey2014	0.28	1.72	0.32	0.15	
VPA	Ey2015	0.23	0.33	0.29	0.20	
VPA	Ey2016	0.27	0.13	0.28	0.25	
VPA	Ey2017	0.32	0.21	0.35	0.30	
VPA	Ey2018	0.32	0.11	0.42	0.37	
VPA	Ey2019	0.53	0.42	0.41	0.39	0.29
VPA	Fmsy_0.7_I	0.45	0.43	0.36	0.15	0.32
VPA	RF_F0.1_16	1.00	0.46	0.65	0.35	0.57
VPA	RF_F40SPR	0.74	0.23	0.31	0.28	0.35
VPA	RF_Fmsy_0	0.73	0.27	0.34	0.28	0.37
VPA	Ry2010	0.27	0.86	0.25	0.06	
VPA	Ry2011	1.66	0.71	0.36	0.12	
VPA	Ry2012	0.72	0.72	0.30	0.17	
VPA	Ry2013	0.83	0.97	0.38	0.19	
VPA	Ry2014	2.00	0.80	0.40	0.22	
VPA	Ry2015	0.36	0.69	0.58	0.40	
VPA	Ry2016	0.28	0.58	0.68	0.37	
VPA	Ry2017	0.44	0.57	0.78	0.48	
VPA	Ry2018	0.30	0.50	0.80	0.61	
VPA	Ry2019	1.69	0.18	0.71	0.39	0.47
VPA	TBy2010	0.20	0.71	0.30	0.05	
VPA	TBy2011	0.20	0.73	0.31	0.08	
VPA	TBy2012	0.18	0.78	0.34	0.13	
VPA	TBy2013	0.62	0.93	0.39	0.18	
VPA	TBy2014	0.24	0.86	0.37	0.20	
VPA	TBy2015	0.22	0.79	0.49	0.27	
VPA	TBy2016	0.22	0.80	0.59	0.32	
VPA	TBy2017	0.29	0.55	0.77	0.44	
VPA	TBy2018	0.28	0.53	0.88	0.53	
VPA	TBy2019	0.73	0.48	0.98	0.58	0.38
	Partials	0.42	0.47	0.38	0.19	
Overall Summary Weighted Geo Mean		0.344678				

Annex F

Revised CPUE Standardization Protocol for Chub Mackerel

CPUE is catch per unit effort obtained either from fishery independent or fishery dependent data. The use of CPUE in a stock assessment implicitly assumes that CPUE is proportional to stock abundance/biomass. However, many factors other than stock abundance/biomass may influence CPUE. Thus, any other factors, other than stock abundance/biomass, that may influence CPUE should be removed from the CPUE index. The process of reducing/removing the impacts of these factors on CPUE is referred to as CPUE standardization.

The following protocol is developed for the CPUE standardization:

(1) Provide a description of the type of data (logbook, observer, survey, etc.), and the "resolution" of the data (aggregated, set-by-set etc.). This description should also include the representativeness of the data in two tables: (1st table) Number of observations, % Coverage of CPUE fleet(catch), % Coverage of CPUE fleet(effort), Total Catch CPUE fleet (mt), Total Effort CPUE fleet, Percentage of overall catch by member (across all fleets/gears); and (2nd table) Number of records remaining, Number removed, Number of records with chub mackerel catch >0;

(2) Conduct a thorough literature review to identify potential explanatory variables (i.e., spatial, temporal, environmental, and fisheries variables) that may influence CPUE values;

(3) Plot annual/monthly spatial catch, effort and nominal CPUE distributions and determine temporal and spatial resolution for CPUE standardization;

(4) Make scatter plots (for continuous variables) and/or box plots (for categorical variables) and present correlation matrix if possible to evaluate correlations between each pair of those variables;

(5) Describe selected explanatory variables based on (2)-(4) to develop full model for the CPUE standardization;

(6) Specify model type and software (packages) and fit the data to the assumed statistical models (i.e., GLM, GAM, Delta-lognormal GLM, Neural Networks, Regression Trees, Habitat based models, and Statistical habitat based models);

(7) Evaluate and select the best model(s) using methods such as likelihood ratio test, information criterions, cross validation etc.;

(8) Provide diagnostic plots to support the chosen model is appropriate and assumption are met (QQ plot and residual plots along with predicted values and important explanatory variables, etc.);

(9) Present estimated values of parameters and uncertainty in the parameters in table;

(10) Present relationship between dependent variable and independent variables. Check whether it is interpretable;

(11) Extract yearly standardized CPUE and standard error by a method that is able to account for spatial heterogeneity of effort, such as least squares mean or expanded grid. If the model includes area and the size of spatial strata differs or the model includes interactions between time and area, then standardized CPUE should be calculated with area weighting for each time step. Model with interactions between area and season or month requires careful consideration on a case by case basis. Provide details on how the CPUE index was extracted;

(12) Calculate uncertainty (SD, CV, and/or CI) for standardized CPUE for each year. Provide detailed explanation on how the uncertainty was calculated;

(13) Provide a table and a plot of nominal and standardized CPUEs over time. When the trends between nominal and standardized CPUE are largely different, explain the reasons (e.g. spatial shift of fishing efforts), whenever possible.

Annex G

Revised Stock Assessment Protocol for Chub Mackerel

(1) Identify the data that will be needed and available to the stock assessment;

(2) Evaluate quality, quantity, and potential error sources of available data (e.g., catch at age, weight at age, length at age), life-history parameters (e.g., natural mortality, growth, and maturity), and abundance indices;

(3) Determine the framework of operating model for extensive simulation tests with the inclusion of potential uncertainties of observed data and life-history parameters;

(4) Create base case scenarios and alternative scenarios for the stock assessment models by the operating model;

(5) External review of the operating model and improvement of the operating model, if needed;

(6) Develop multiple stock assessment models and conduct the performance tests by applying the models to the data generated from the operating model;

(7) Select the best candidate model(s) for the full stock assessment of chub mackerel;

(8) Determine candidate scenarios for biological parameters and input data;

(9) Apply the stock assessment model to the data and determine model specifications for the assessment through the following processes:

- Conduct diagnostics of model convergence, plot and evaluate residual patterns, compare prior and posterior distributions for key model parameters (if using Bayesian approach), and evaluate biological implications of the estimated parameters;
- Develop retrospective analysis to verify whether any possible systematic inconsistencies exist among model estimates of biomass and fishing mortality; conduct likelihood profiles by each key model component is also useful to find systematic inconsistencies;
- Explore the method for representing uncertainties.

(10) Determine the final base case;

(11) Review and finalize stock assessment results;

(12) Review and estimate biological reference points or MSY-based reference points and associated uncertainties;

(13) Provide stock status relative to biological reference points or MSY-based reference points;

(14) Consider methods to include relevant ecosystem considerations regarding the stock in future assessment documents, including data and results from other scientific studies regarding potential impacts on the stock [assessment] due to climate change, predator-prey dynamics, or impacts of distribution and phenological changes on assessment data.

As the next steps, it is recommended that the SC, in cooperation with managers, conduct the following:

(1) Identify target and limit reference points;

(2) Determine if the stock is “overfished” and “overfishing” occurs, for example using the Kobe plot;

(3) Develop alternative harvest control rule (HCR) for the projection (e.g., 5-year projection);

(4) Conduct risk analysis for each level of fishing impacts and each HCR to develop decision tables with alternative state of nature;

(5) Provide stock status, decision tables, and scientific advice on HCR.

Possible settings and specification of SAM

Model configuration	Parameter	Original option (used in the OM process)	Option(s) to be addressed by TWG CMSA08 (short-term work)	Option(s) to be addressed after input data fixed (mid-term work)	Potential option(s) requiring revision or development (long-term work)	Note
Recruitment	$N_{0,y}$	Random walk	Beverton-Holt stock-recruitment relationship (SRR)	Other SRRs	・Hockey-stick (HS) SRR ・Consider other structures of random errors	Analyzing HS SRR is difficult in SAM
Nonlinear coefficient for abundance indices	b_k	・Estimated for the two recruitment indices and Chinese and Russian fishery-dependent indices, but assumed a common value for these two fishery-dependent indices due to short time periods ・Fixed at 1 (not estimated) for the two SSB indices	・Fix at 1 (not estimated) for all abundance indices ・Estimate different values for each abundance indices	Searching the best option about how constraints are imposed on which indices based on AIC etc		
Years of F random walk	-	Exclude the Markov process from 2010 to 2011	Include the Markov process for all years			
Correlation of age classes in F random walk	ρ	A simple function of age difference	?	?		

Process errors in numbers older than age 0	ω_a $(a>0)$	Fix at a very small value (0.01)	・Estimate a common value for all age classes except for age 0	Searching the best option about how constraints are imposed on which indices based on AIC etc	Consider other structures of random errors	
SD in F random walk	σ_a	Impose constraints of common values for some age classes based on AIC	・Estimate a common value for all age classes ・Estimate different values for each age class if converged	Searching the best option about how constraints are imposed on which age classes based on AIC etc	Consider other structures of random errors	
SD in measurement errors of catch at age	τ_a	Impose constraints of common values for some age classes based on AIC	・Estimate a common value for all age classes ・Estimate different values for each age class if converged	Searching the best option about how constraints are imposed on which age classes based on AIC etc	Consider other structures of random errors	
SD in measurement errors of abundance indices	ν_a	Impose constraints of common values for some age classes based on AIC	・Estimate a common value for all age classes ・Estimate different values for each age class if converged	Searching the best option about how constraints are imposed on which age classes based on AIC etc	Consider other structures of random errors	
Number of fleets	-	Single			Multiple	・A relatively large revision is required ・Extension to multi-fleets may be useful

						in fitting fishery-dependent CPUE and for a management purpose
Usage of catch at age	$C_{a.y}$	Fit to all data			Put different weights based on data uncertainty	SAM allows missing data in catch at age
Growth and maturity	?	Single conditional weight/maturity at age			Incorporate density dependence in weight growth and maturity	

Possible options for the basic specifications for conducting future projections for chub mackerel

Items	Option A	Option B	Option C	Option D	Issue to be clarified
Type of simulation	Stochastic (how many times?)	Deterministic			Model uncertainty, Management objective
Duration	Short (<5 years)	Medium (5-10 years)	Long (>10 years)	Equilibrium	Ask the COM to consider management objective and methods Consider appropriate duration for chub mackerel
Type of uncertainties					
Other parameters (not recruitment)	Parameter estimates without uncertainty	Parameter estimates with uncertainty			
Recruitment level	Model-based approach using S-R relations (BH/Ricker/HS/Others)	Empirical approach by resampling past recruitments (what duration?)			Model uncertainty
Error structure in recruitment	Parametric (log-normal?)	Non-parametric (resampling of deviations)	Recruitment (Process error)		
Catch	F-based (Current F/Mean F for reference period)	C-based (What is HCR?)	Other MP?	Include terminal year's F or not	Management Method, HCR

Estimation of catch from terminal year to current year	Terminal year	Last year of harvest	Average of 2 or 3 recent years		

Annex J

Timeline and activities for intersessional work from the conclusion of TWG CMSA07 to February 2024

<table>
<tr><th colspan="2">Month</th><th>Catch@Age</th><th>Weight@Age, Maturity@Age</th><th>Abundance Indices</th><th>SAM</th></tr>
<tr><td rowspan="2">Sep</td><td>Mid</td><td></td><td></td><td>Email communication on draft CPUE document template</td><td>Share specification table</td></tr>
<tr><td>Late</td><td>Japan shares ALK</td><td></td><td></td><td>Email communication</td></tr>
<tr><td rowspan="4">Oct</td><td>Early</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Mid</td><td></td><td></td><td>Determine CPUE document template</td><td></td></tr>
<tr><td>23 Oct</td><td>Share Catch@Age data</td><td>Share Weight@Age and Maturity@Age data</td><td></td><td></td></tr>
<tr><td>Late</td><td>China shares ALK</td><td>Share Catch@Length (Catch@Size)
Exchange of age determination rule (timing of aging)</td><td></td><td></td></tr>
<tr><td rowspan="3">Nov</td><td>Early</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Mid</td><td colspan="4">Intersessional meeting (one day in the week of 13-17 Nov)</td></tr>
<tr><td>Late</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="4">Dec</td><td>Early</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Mid</td><td></td><td></td><td></td><td></td></tr>
<tr><td>23 Dec</td><td>Submission of Catch@Age</td><td>Submission of Weight@Age and Maturity@Age</td><td>Submission of documents
Share CPUE standardization code</td><td></td></tr>
<tr><td>Late</td><td></td><td></td><td></td><td></td></tr>
</table>

Jan	Early				Trial run using tentative input data
	Mid				Share code and tentative input data
	22-25 Jan	TWG CMSA08			
	Late				
Feb	Early				
	Mid				
	25 Feb	Finalization of input data	Finalization of input data	Finalization of input data	