NPFC-2021-TWG CMSA04-WP14

# Candidate Performance Measures for testing Chub Mackerel Stock Assessment Models

Joel Rice[[1]](#footnote-1)

**Technical Working Group on Chub Mackerel Stock Assessment**

**22-25 June 2021**

**Virtual Working Group Meeting**

# Introduction

The North Pacific Fisheries Council (NPFC) is a regional fisheries management organization established in 2015 to manage fisheries resources in the Convention Area and protect the marine ecosystems of the North Pacific Ocean in which these resources occur. The NPFC aims to develop management plans for its target fisheries. As a part of this process, the Scientific Committee (SC) has to assess stock status of the priority species, set management standards including reference points and recommend harvest control rules to the Commission.

The SC established the Technical Working Group for Chub Mackerel (*Scomber japonicus*) Stock Assessment (TWG CMSA) to assess the status of the chub mackerel stock and provide scientific advice on the management of chub mackerel fisheries to the Commission by 2022. At this point the TWG CMSA has assembled data for the stock assessment, and several candidate stock assessment models have been prepared. Additionally, an agreed upon list of scenarios with alternative assumptions regarding the structural uncertainty of the assessment has been developed based on changing the natural mortality, the weight at age and the maturity at age. The TWG CMSA plans to develop an operating model to select a base case stock assessment model from the five candidate models (SAM, VPA, ASAP, KAFKA and state-space production model).

The principal objectives of the combined process of stock assessment and fisheries management could be described as ensuring sustainable harvests and viable fishing communities, while maintaining healthy ecosystems ensuring the sustainability of the resource and the overall health of the marine environment. As part of this process fisheries scientists are often confronted with the difficult task of comparing multiple models with different structures and selecting one which could provide advice on the stock status and trajectory. The situation with chub mackerel is a hurdle often encountered for highly migratory species that are subject to multiple international fleets. Before selecting between any number of models it is important that the individual models undergo diagnostic tests to ensure that they are internally consistent (Carvalho et al., 2021, Lee et al., 2014).

Model diagnostics can help determine the robustness of estimates for management, identify data conflicts and strengthen the interpretation of the model results. Model diagnostics can help assigning weights to an ensemble of models (Maunder et al., 2020) and in some cases discriminate between them. However traditional diagnostics that are based on residuals, likelihoods, and information theoretic values such as Akaike’s Information Criteria (AIC, Akaike, 1998) require that the models be configured with the same likelihood function and data. Methods to compare the performance of different of stock assessments focus on the fit to observational data and the ability to cross validate predictions and data (Deroba et al., 2015, Carvalho et al., 2021, Kell et al., 2021)

To evaluate the relative strengths of different stock assessment models (modeling platforms), it is important that assessment scientists have objective measures for model evaluation. These objectives can be incorporated into the stock assessment process in the form of performance metrics, which provide the quantitative benchmarks which are used to compare the relative performance of different stock assessment models given the underlying operating model configurations. It is important that models have internal consistency, the ability to fit the available data, and to have models that are robust to uncertainty with respect to key parameters. Because current data availability and management needs (i.e. stock status or relative abundance) may not reflect future data or management needs, is important to consider which models are robust to uncertainty and understand the limitations of each modeling platform.

# Methods

The NPFC is considering management approaches for chub mackerel (NPFC-2019-TWG CMSA02-WP02), based on the results of the developing stock assessment models and future studies of management strategies. Currently there is no management framework in place for chub mackerel, therefore management information needs and reference points have yet to be developed. However, given that the intent of the stock assessment process is to assess the health of the stock and the impact of fisheries on that stock it is possible to utilize (among other performance measures) common fisheries reference points to evaluate the performance of the stock assessment model.

The TWG CMSA proposes to use the methods of Deroba et al. (2015) to test candidate stock assessment models and modeling approaches by simulation using the POPSIM-A fishery simulator as an operating model (NPFC-2019-TWG CMSA02-Final Report). This approach would utilize POPSIM-A as a general data simulator (operating model) to allow the comparison of the stock assessments models ability to fit datasets with known underlying parameters and error structure.

At the 3rd Meeting of the Technical Working Group on Chub Mackerel Stock Assessment (TWG CMSA) the working group proposed a number of performance measures for evaluating the stock assessment models (Annex A). Among the quantities listed in Annex A there are measures that quantify the stock status (e.g. B/BMSY, F/FMSY), as well as basic biological parameters (i.e. steepness). The combined need for discrimination between stock assessment platforms and for stock assessment models to adequately model the observational data suggests the inclusion of additional performance measures related to how well the model fits the data.

The list contained in Annex A is fairly comprehensive, and likely larger than necessary. Suggested priority performance measures include measures related to how well the data is fit as well as measures that quantify the stock status (Table 1). Performance metrics that assess the goodness of fit of the performance measure (i.e. % bias and RMSE) are also included as a point of discussion. Recent studies (Chong et al., 2019) have used the relative errors between the estimated and true value as performance metrics for each derived output in order to quantify bias using the median relative error (MRE) and precision using the median absolute relative error (MARE):

And

where *xest*is the estimated value (calculated from the respective estimation model) and *xtrue* is the true value (calculated from the operating models). Values closer to zero represent the least biased (MRE) and most precise (MARE) results.

# Discussion

Deroba et al. (2015) used the time-series estimates of stock biomass (spawning stock in most cases, total biomass in others) and fishing mortality as performance measures. Other studies (Chong et al. 2019) have used F/FMSY and SPR as performance measures when evaluating alternative stock assessment platforms fit to the same data. Other methods exist such as hindcasting (Carvalho et al., 2021) to assess how well models fit to observational data, and their utility in explaining bias in model estimates of parameters and derived quantities. Here we suggest that the performance measures for evaluating the fit of the stock assessment be based on the measures listed in Table 1, and be preceded by individual model diagnostics to establish an internally consistent model.

The rational for including individual model diagnostics is straightforward, misspecification of key parameters or assumptions stock assessment models can strongly impact the estimates of quantities such as stock depletion and biomass at maximum sustainable yield (Mangel et al., 2013). The performance measures recommended follow some of the same logic, the assessment model should be able to capture the population trend, adequately detect directional changes in the population and recruitment. Additionally, the stock status related quantities (F/FMSY & B/BMSY) should be well estimated. The interpretation of the performance measures should be guided by the following questions;

1. How well does the modeling platform (stock assessment model) replicate the population trends?
2. Can the modeling platform can detect downturns, particularly in the last years of the assessment?
3. Can the assessment model detect relative year class strengths?
4. Can the assessment model detect recruitment failure?
5. How well does the assessment model replicate F/FMSY & B/BMSY?
6. How well does the assessment model estimate the B0 parameter (or similar overall scale)?

The results of the simulation testing by Deroba et al. (2015) were summarized by the 10th, 50th and 90th quantiles and then plotted against the ‘true’ fit. Subsequent qualitative examination for consistency between the time series based on the real data and the trend from the fits to the pseudo-data was the main performance metric. The TWG CMSA has suggested that the median, mean, SE, % bias and RMSE be used as performance metrics to judge the performance measures, here we suggest MRE and MARE as objective measures of precision and bias recently used in studies that have (Carvalho et al. 2021, Chong et al. 2019). Careful consideration as to which of these metrics is to be prioritized for model selection is a topic for consideration of the TWG GMSA.

# Tables



# ****References****

Akaike, H.. 1998. Information theory and an extension of the maximum likelihood principle. Pages 199-213 of: Selected Papers of Hirotugu Akaike. Springer.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I., Wetzel, C.R., Doering, K., Johnson, K.F., Methot, R.D., 2021. A Cookbook for Using Model Diagnostics in Integrated Stock Assessments. Fish. Res. 240, 105959. https://doi.org/10.1016/j.fishres.2021.105959

Chong, L., Mildenberger, T. K., Rudd, M. B., Taylor, M. H., Cope, J. M., Branch, T. A., Wolff, M., and Sta¨bler, M. 2019 Performance evaluation of data-limited, length-based stock assessment methods. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsz212.

Deroba, J.J., Butterworth, D.S., Methot, R.D. Jr., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O’Malley, J.M., Chang, Y-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., Hulson, P.-J. F., 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, 72: 19–30

Kell, L., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., Fu, D., 2021. Validation of stock assessment methods: is it me or my model talking? In Prep.

Lee, H.-H., Piner, K.R., Methot, R.D., Maunder, M.N., 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fish. Res. 158, 138–146. <https://doi.org/10.1016/j.fishres.2013.12.017>.

Mangel, M., MacCall, A.D., Brodziak, J., Dick, E.J., Forrest, R.E., Pourzard, R., Ralston, S., 2013. A perspective on steepness, reference points, and stock assessment. Can. J. Fish. Aquat. Sci. 70, 930–940. https://doi.org/10.1139/cjfas-2012-0372.

NPFC, 2020. 3rd Meeting of the Technical Working Group on Chub Mackerel Stock Assessment. 2020. 3rd Meeting Report. NPFC-2020-TWG CMSA03-Final Report. 25 pp. (Available at [www.npfc.int](http://www.npfc.int))

Stewart, I. J., and Martell, S. J. D. Reconciling stock assessment paradigms to better inform fisheries management. – ICES Journal of Marine Science, 72: 2187–2196

Annex A**. Performance Measures for evaluating stock assessment models, from CMSA03**.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measure | Necessity | Statistics | | | | |
| *State Variables* |  |  |  |  |  |  |
| B (whole years) | Compulsory | median | mean | SE | %bias | RMSE |
| SSB (whole years) | Compulsory if possible | median | mean | SE | %bias | RMSE |
| R (whole years) | Compulsory if possible | median | mean | SE | %bias | RMSE |
| F (whole years) | Compulsory if possible | median | mean | SE | %bias | RMSE |
| Selectivity at age (whole years) | Compulsory if possible | median | mean | SE | %bias | RMSE |
| Catch | Compulsory if possible | median | mean | SE | %bias | RMSE |
| Exploitation rate  (catch/total biomass) | Compulsory | median | mean | SE | %bias | RMSE |
|  |  |  |  |  |  |  |
| *Basic Biological Parameters* |  |  |  |  |  |  |
| B0 | if possible | median | mean | SE | %bias | RMSE |
| Steepness | if possible | median | mean | SE | %bias | RMSE |
|  |  |  |  |  |  |  |
| *Biological Reference Points* |  |  |  |  |  |  |
| Bmsy | if possible | median | mean | SE | %bias | RMSE |
| SBmsy | if possible | median | mean | SE | %bias | RMSE |
| Fmsy | if possible | median | mean | SE | %bias | RMSE |
| F%SPR | Compulsory if possible | median | mean | SE | %bias | RMSE |
| F0.1, Fmax | Compulsory if possible | median | mean | SE | %bias | RMSE |
|  |  |  |  |  |  |  |
| *Depletion Statistics* |  |  |  |  |  |  |
| SSB/max(SSB) (periods\*) | Compulsory if possible | median | mean | SE | %bias | RMSE |
| SSB/SSB0 (periods\*) | if possible | median | mean | SE | %bias | RMSE |
| SSB/SSBmsy (periods\*) | if possible | median | mean | SE | %bias | RMSE |
| B/max(B) (periods\*) | Compulsory if possible | median | mean | SE | %bias | RMSE |
| B/B0 (periods\*) | if possible | median | mean | SE | %bias | RMSE |
| B/Bmsy (periods\*) | if possible | median | mean | SE | %bias | RMSE |
|  |  |  |  |  |  |  |
| *Relative fishing mortality* |  |  |  |  |  |  |
| F/Fmsy (e.g. average for the last 3 years\*) | if possible | median | mean | SE | %bias | RMSE |
| F/F%SPR (periods\*) | Compulsory if possible | median | mean | SE | %bias | RMSE |
| F/F0.1, F/Fmax (periods\*) | Compulsory if possible | median | mean | SE | %bias | RMSE |
|  |  |  |  |  |  |  |
| Retrospective analysis (e.g. Mohn's rho) | if possible |  |  |  |  |  |
|  |  |  |  |  |  |  |

\* To be proposed intersessionally and determined at the TWG CMSA04 meeting

1. Joel Rice, Rice Marine Analytics (ricemarineanalytics@gmail.com) [↑](#footnote-ref-1)